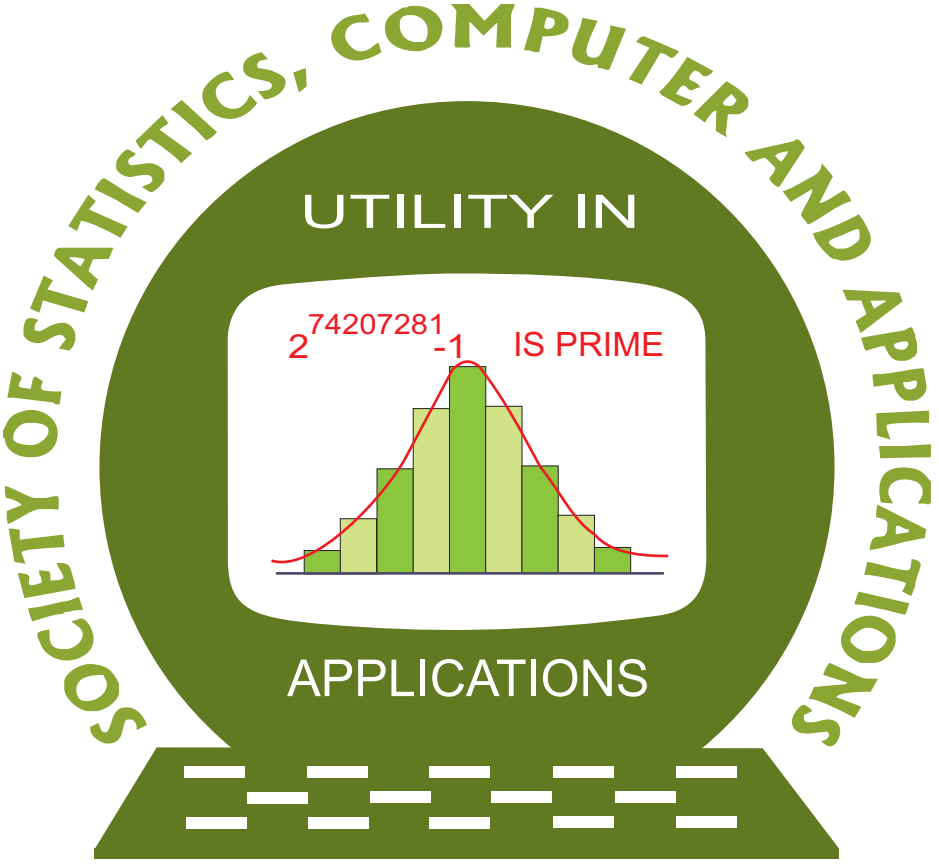
ISSN 2454-7395(online)

STATISTICS AND APPLICATIONS



FOUNDED 1998

Journal of the Society of Statistics, Computer and Applications

https://ssca.org.in/journal.html

Vol. 18, No. 2 (Special Issue), 2020 (New Series)

Society of Statistics, Computer and Applications

Council and Office Bearers

Founder President
Late M.N. Das

President
V.K. Gupta

Executive President
Rajender Parsad

Patrons

A.C. Kulshreshtha, A.K. Nigam, Bikas Kumar Sinha, D.K. Ghosh
K.J.S. Satyasai, P.P. Yadav, Pankaj Mittal, R.B. Barman
R.C. Agrawal, Rajpal Singh

Vice Presidents

A. Dhandapani, Lal Mohan Bhar, P. Venkatesan
S.D. Sharma, V.K. Bhatia

Secretary
D. Roy Choudhury

Foreign Secretary
Sudhir Gupta

Treasurer
Ashish Das

Joint Secretaries

Aloke Lahiri, Hukum Chandra, Shibani Roy Choudhury

Council Members

Alka Arora, Anil Kumar Yadav, Anshu Bhardwaj, Manish Sharma, Manish Trivedi
Manisha Pal, P. Rath, Piyush Kant Rai, Praggya Das, R. Vishnu Vardhan
Ranjit Kumar Paul, Rao Saheb Latpate, S.A. Mir, Sanjeev Panwar, V. Srinivasa Rao

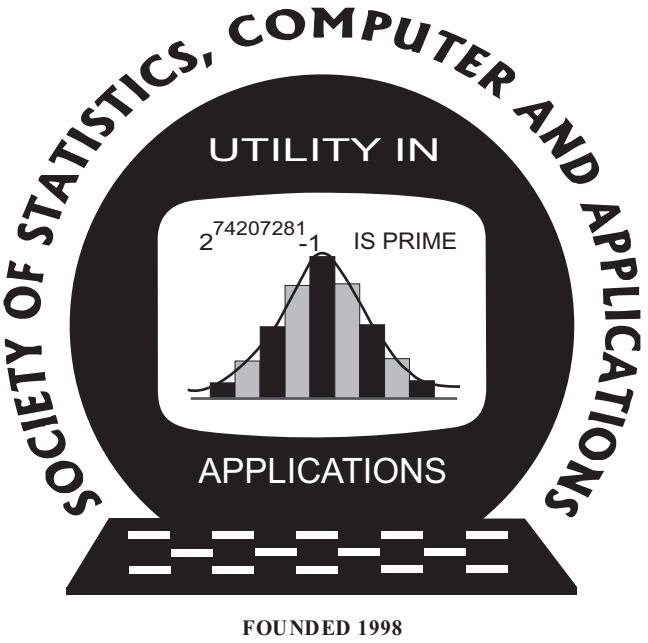
Ex-Officio Members (By Designation)

Director General (Statistics), MOSPI, New Delhi
Director, ICAR-Indian Agricultural Statistics Research Institute, New Delhi
Chair Editor, Statistics and Applications
Executive Editor, Statistics and Applications

Society of Statistics, Computer and Applications, B - 133, Ground Floor, C.R. Park, New Delhi - 110019

Statistics and Applications

ISSN 2454-7395(online)



Journal of the Society of Statistics, Computer and Applications

https://ssca.org.in/journal.html

Volume 18, No. 2 (Special Issue), 2020 (New Series)

Special Issue "Challenges and Opportunities in Statistical Data Designing and Inference in the Emerging Global Scenario" in honour of twin statisticians (brothers) Bimal and Bikas Sinha on their 75th Birthday (16 March 2021)

Statistics and Applications

Volume 18, 2 (Special Issue), 2020 (New Series)

Editorial Panel

Statistics and Applications ISSN 2454-7395 (online)
Volume 18, No. 2 (New Series) 2020

CONTENTS

Special Issue "Challenges and Opportunities in Statistical Data Designing and Inference in the Emerging Global Scenario" in honour of twin statisticians (brothers) Bimal and Bikas Sinha on their 75th Birthday

FOREWORD

It is indeed a pleasure to write this Foreword note in favor of two outstanding statisticians of our time: Bimal and Bikas Sinha, whom we affectionately address as the "*Statistical Twins*". They were born in1946 in the village Atgharia in Pabna District of Bengal Province in undivided India, a year before the region was engulfed in the newly created East Pakistan (which later gave birth to Bangladesh in 1971 December). The Sinha family migrated to Kolkata, West Bengal, India, in 1958 when they were just 12 years old. For 12 years to follow, they had hard economic time, albeit both Bimal and Bikas excelled in their college education and earned their Ph.D. degrees in Statistics in1972-1973 under the able guidance of [Late] Professor Hari Kinkar Nandi in Calcutta University. I have known them since 1968 when they started working on their dissertations in two disjoint areas: Bimal on Bayesian Inference, and Bikas on Optimal Designs. During the past fifty years their research and organizational accomplishment may simply be categorized as outstanding. Both stretched their research work in diverse areas with little overlap to each other, and they have taken leading roles in the development of Statistical Science all over the world. Bimal has been more visible in USA, Europe, Africa, Thailand and Australia, while Bikas in Brazil, USA, Europe and more actively in India, Bangladesh and the South Asian Countries. In Bangladesh, they regard Bikas and Bimal as their native sons and owe to Bikas a deep sense of gratitude for his tireless efforts to promote and organize Int'l Conferences in Statistics at Rajshahi University. I wish [Late] Professor H.K. Nandi were alive to witness the fundamental contributions of his outstanding advisees Bimal and Bikas. I am really deputizing [Late] Professor Nandi, my advisor too, in writing this foreword note. There is an accompanying detailed document in this volume on their vitae and significant contributions which depicts the details. There are some twenty five contributors of high academic stand to this volume; all of these contributors have known Bimal and Bikas, have mostly collaborated with them in research and organizational tasks and I would like to thank each of them for showing this support and respect for the statistical twins. Finally, I would like to extend my deep appreciation to the Guest-Editors (Vinod Gupta, Sanat Sarkar and Ayanendranath Basu) for their time and efforts to make this volume a true collection of notes of appreciation from all of us. Bimal and Bikas are now about 75 years old, and they are in excellent academic stand; we wish them a continuation of long and active career in the years ahead too.

Pranab K. Sen
Cary C. Boshamer Professor Emeritus
Biostatistics, and Statistics & Operations Research,
University of North Carolina at Chapel Hill, NC 27599-7420, USA

A Clique in Sight ... suddenly disappears ... to resurface again … in a milder form …

Gour Mohan Saha
Indian Statistical Institute (Retired)

It is an honour to ruminate over my years of friendship with Bimal and Bikas – known as "Statistical Twins" in the statistical community at large. I have known them since the summer of 1962, when we entered Asutosh College, Calcutta, as first year Stat. Hons. students in B.Sc. Degree Course. In those days, they were very much look alike, in appearance, and also in outfits. And, surprisingly, our classmates found similarity in my physique too with those of Sinha Brothers and we were a classic example of a 'clique' – in the terminology of graph theory! And this continued even in later years when we were at ISI, Kolkata as faculty. We were labeled as an example of 'uniform distribution'! We were, at times, sources of 'joy and confusion', without any clarification as to "who is who" and we very much enjoyed this episode amongst us. After completion of college program in 1965, Sinha Brothers joined the Calcutta University Department of Statistics [CUDS] for their Master's program, while I left for the Indian Agricultural Statistics Research Institute (IASRI) [called IARS at that time], under the I.C.A.R., in New Delhi, for my Master's program. Our friendship temporarily ended at that moment. That was in 1965 and I lost contact with them. Our 'clique' was reduced to an 'edge' between them while I was left as an 'isolate'. Truly I was in "isolation" in a new place and I had to find my new friends and, against my timid nature, a few unexpected foes as well!

At IASRI, fortunately, everything was smooth for me. By 1970, I had the M.Sc. and Ph.D. Degree of IARI (IASRI) and I was inducted as a junior faculty over there. My specialization was in the area of Design of Experiments (DoE). In December 1970, there was an Annual Conference of IASRI which was held in the premises of University of Madras. I attended the same and delivered a technical talk as well. While I was addressing the gathering, I was pleasantly surprised to see my 'old friends' among the audience!

Sinha Brothers were up to attending Science Congress in Bangalore University the following week. It was a moment of immense joy for us to sit and talk and talk, with another friend Srijib Bagchi [of Asutosh College] being with them. They were in the doctoral programme at CUDS and Bikas was working in the area of Optimal Experimental Designs. They both had [Late] Professor H.K. Nandi as their supervisor while I had worked with [Late] Prof. M.N. Das. All of them were at the final stages of their research for Ph.D. This surprising encounter acted as an inspirational moment for them since my nametag was prominently displaying Ph.D. Degree! We exchanged addresses and phone numbers.

In due course, Sinha Brothers obtained their Ph.D. Degrees, and Bimal [Thesis in the area of Bayesian Inference] joined the Research and Training School [RTS] of ISI, Kolkata in November, 1971, while Bikas joined CUDS soon thereafter. They both entered as Lecturers at their respective places.

At some point of time, I had expressed my inclination to join ISI to Bimal and he immediately spoke with the then Dean of Studies Professor Hanurao (now deceased). It so happened that ISI, Kolkata, also was then looking for a strong faculty on DoE! The Dean arranged for my interview with Professor C.R. Rao at Delhi ISI around 1972. I was

interviewed and offered a position and I took no time to opt for Kolkata Centre. After a long gap of almost 7 years (1965 – 1972), I returned to Kolkata and on October 3, 1972, I joined the Indian Statistical Institute, Calcutta, as a Senior Lecturer in the Research and Training School (RTS). My earlier stamp as an 'isolate' was re-labeled as a member of the clique with Sinha Brothers, but I must hasten to add that it was very much weighted network. Of course, the strongest tie [heaviest weight] was between them. To start with, I had 'weak' ties with them but very soon Bimal and I had worked together to strengthen the weight of the arc connecting us. My direct contact with Bikas could not be improved and it remained as a weak tie for a while. Distance between ISI Centre and CUDS/Sinha Residence was a major factor. How long would it remain so? We both wished to 'come closer' but who would be the catalyst?

Sometime in 1974, to our utter surprise, Professor C.R. Rao called upon Bikas at ISI; also called me at his office and told both of us about Visiting Faculty Positions in one Brazilian University, initially for a period of 2 years. Primary responsibilities would be to focus on Faculty Development Programme in Statistical Theory and Applications. He would be ready to recommend if we were willing to go. We spontaneously agreed.

Professor Rao had been Ph.D. Thesis Examiner for Bikas at CUDS. Both Bimal and Bikas have been extremely energetic and systematic in their thoughts and actions. Honestly, my position is almost in the 'opposite pole'. I am, most of the time and in most cases, slow-moving – by nature. Bikas joined The Institute of Mathematics, Federal University at Bahia, Brazil sometime around November, 1975 while I took my own time to reach there sometime in April, 1977. By the twist of Professor Rao's hands and his blessings, this time I developed a very strong tie with Bikas while the tie with Bimal started fading away! Bikas and I were together in Brazil for more than 2 years.

In September 1979, Bikas returned back home and joined Stat-Math Division of ISI, Kolkata. I was told – [Late] Professor J.K. Ghosh was interested in absorbing Bikas at ISI and Professors C.R. Rao and [Late] D. Basu made the recommendations! I returned in April, 1980 to my position in the same division of ISI, Kolkata. Our tie continued to strengthen over all these years till the time of our retirement around 2008-2011. I may add that one node of the clique has been designated as 'Prabhu' by the other two and this labeling has been quite popular among our friends and relations – far and near. We leave it to the readers to solve the mystery behind the special designation "Prabhu" – why so!

Meanwhile, Bimal was shuttling between ISI, University of Montreal (Canada) and University of Pittsburgh (USA). Eventually, Bimal left ISI for USA in August 1980. In 1985, Bimal joined University of Maryland – Baltimore County [UMBC] as Founder-Professor of Statistics Graduate Programme.

Regarding our marriage and nurturing respective families, I got married first, then Bimal and then Bikas. I have two daughters, Bimal has two sons while Bikas has one daughter and one son.....you may draw your own conclusions accordingly – whom to rely to play a game with 50-50 chance of win. Naturally Bikas became smarter after observing our outcomes and played unbiasedly!

During the period of our stay in Brazil, Bikas and I had to learn Portuguese language for teaching Brazilian students and Faculty and for carrying on our social lives. We discussed various topics for teaching and preparing notes. Somehow we could work on one research

problem in the area of DoE and it was eventually published. Much later, I had another opportunity to interact with Bikas at ISI, Kolkata on another research problem – formulated by him, along with S.B. Rao and Prasad Rao.

With Bikas, I had developed mutual friendship and regard to the extent that we used to [and still] address each other as “Chefe’ – meaning ‘Chief” in Portugese. In this connection, I should mention that my “Chefe” wrote a book on sampling methods in Portugese while in Brazil. And I had prepared some Lecture Notes on DoE in Portuguese while in Brazil.

I must say that Bikas had many areas of research interests, while I had interest only in the construction and analysis of Design of Experiments in a variety of application areas like agricultural experiments, industrial experiments, animal experiments, bioassays etc. I may also mention here that on a request from my dear ‘Chefe’, I wrote the Foreword of his Monograph on "Optimal Covariate Designs”, co-authored by Premadhis Das, Ganesh Dutta, Nripes Mandal and Bikas K. Sinha in the year 2015.

On a very personal note of gratitude, I must add that my Chefe had helped me in my visits to Penn State University during 1987-88, and for a subsequent visit to Western Michigan University at Kalamazoo, during 1988-89.

Sinha Brothers were brilliant as students all along, and as we all know now, later on, they turned out to be brilliant researchers and teachers too, in their fields of interest in Statistics. This is why the publication of a Volume in their honor is very much desirable, and welcome on their 75^{th} birthday! This collection is a reflection of their collective research interests in Statistics and the chapters are contributed by their students / fellow researchers / research collaborators and admirers. I am sure it will serve as a reference of great use to the future research scholars in so many areas of statistical theory and applications.

May my friends cherish their retired lives to their fullest satisfaction academically and socially.

PREFACE

This Issue (No. 2) of the Volume 18 of *Statistics and Applications* has been brought out to felicitate the twin statisticians, Bimal Kumar Sinha and Bikas Kumar Sinha (twin brothers) and honour them on their 75th birthday that falls on 16 March 2021.

In March, 2019, Prof. Pranab K. Sen had suggested to Prof. Vinod K. Gupta that a special issue of the journal *Statistics and Applications* may be brought out to honour and felicitate the twin statisticians on their 75th birthday. The proposal was accepted with pleasure by the Editorial Board of the journal. It was also decided that this special issue would be titled *"Challenges and Opportunities in Statistical Data Designing and Inference in the Emerging Global Scenario."*

Prof. Sen also suggested some possible names for consideration as Guest Editors for this issue of the journal. Accordingly, in a discussion between Prof. Sen, the (then) Chair Editor Late Prof. Aloke Dey and the Editorial Board of the journal, it was decided that Aloke Dey, Sanat K. Sarkar, Ayanendranath Basu and Vinod K. Gupta would act as the Guest Editors of this special issue. In consultation with the twin statisticians, the Guest Editors prepared a list of authors to whom invitations were sent for possible contributions to the special issue. The list contained, by and large, authors who were students/collaborators/colleagues of the twin statisticians. It was not easy to tell the total number of research collaborators these twin statisticians have, but surely this would be no less than 180 by any stretch of imagination. There was an overwhelming response to our invitations. Almost all of those who were contacted happily agreed to contribute to this issue and offer their felicitations to these towering stalwarts who have made remarkable contributions towards the growth of statistical sciences and their innovative applications. At the end, we are immensely proud to report that this felicitation issue contains **26** excellent papers from authors spread all over the globe. Of special interest is another article prepared by Bimal and Bikas describing their life and achievements and also a list of their research publications limited to books and research papers in reviewed journals. We strongly feel that this would benefit the younger generation of researchers immensely.

It has indeed been a great pleasure and a matter of pride and honour for us to work on this project and prepare this issue as a mark of respect to these two great scientists who have made significant contributions towards the advancement of statistical thinking. Not only are they great researchers, they are also brilliant teachers and mentors who have influenced the thoughts of many young researchers and inspired them along the journey of their research. With their unparalleled love and affection and unflinching willingness to help, they have made indelible mark on the lives of many students and budding young researchers. Beyond their professional accomplishments and recognitions, they are two great human beings who are generously endowed with the qualities of gentleness, humbleness and kindness, and liked by all who make their acquaintance.

We would like to express our sincere thanks to all the authors who have responded positively to our request. While initially we were not entirely sure about what to expect, we are pleasantly overwhelmed by the actual level of response from the authors. The reviewers, the unobservable layer without which the process of journal publication cannot function properly, have also been prompt and thorough; we are indebted to all of them and thank them sincerely for their support. We would like to place on record our highest admiration for Dr. Rajender Parsad, Executive Editor and Dr. B.N. Mandal, Managing Editor, and the Editorial Board of *Statistics and Applications* for their support and for entrusting their faith on the Guest Editors

for bringing out this special issue in honour of the twin statisticians. The Guest Editors, in turn, are greatly honoured by this responsibility. The help received from Dr. B.N. Mandal for bringing the papers in the format of the journal is highly appreciated. The template provided by Prof. Jyotirmoy Sarkar for preparing articles in LATEX was a tremendous help and we generously thank Prof. Sarkar for his support. Prof. Sarkar and Dr. B.N. Mandal have later modified the template further.

The issue contains papers of high academic standards covering a wide spectrum of statistical research. We are confident that the readers would find these papers enjoyable to read and generating newer ideas for advancing research in statistical sciences. It is hoped that the volume will be a fitting commemoration of the 75 year milestone of the illustrious and celebrated Sinha brothers. We wish Bimal and Bikas a long and healthy life so that the statistics community continues to benefit from their brilliance, knowledge, experience and wisdom for a long time.

V.K. Gupta
Sanat K. Sarkar
Ayanendranath Basu

Unfortunately, on 10 February 2020, Prof Aloke Dey left for his heavenly abode after a brief illness. It was decided that we shall not seek any replacement and the remaining three of us would continue as Guest Editors and complete the project in time. The responsibility of Prof. Aloke Dey was discharged jointly by the three guest editors adequately and the project could be completed well within the stipulated time, in fact ahead of time. Vinod K. Gupta shared the key responsibility of coordinating with the authors, the reviewers and the other guest editors. We take this opportunity to express our grief and condolences to the bereaved family and pray for the liberation of the noble soul. Om Shanti!!

GUEST EDITORS

November 2020

The Sinha Brothers (Bimal and Bikas) – The Twin Statisticians

BIMAL KUMAR SINHA

BIKAS KUMAR SINHA

0. Early Days of Twin Brothers

The twins were born to Birendra Nath Sinha and Jogmaya Sinha on 16 March 1946 in Village Atgharia, District Pabna in Bengal Province of undivided India (now in Bangladesh). They were christened as Bimal and Bikas, with Bimal being elder of the twins.

They had their early education in village school before the Sinha Family migrated to Calcutta (now Kolkata) in 1958.

The twins were extremely notorious and gave hard time to their mother, elder sister and near relations with day-time harmful/destructive activities.

1. General Information About Twin Brothers

Over to Kolkata in 1958: Sinha Brothers got admission in Class VIII in a local school. They showed early promise and secured top positions in school exams. They passed School Final Exam [Class X] in 1961 and got admission into One-year Pre-University Course. In 1962, they got admission into 3-year Degree Course in Asutosh College, Calcutta under Calcutta University with Honours in Statistics, and Mathematics and Physics as combination subjects. They felt highly privileged to be under the tutelage of Late Professors S.B. Chaudhuri and K.K. Mukherjee. In 1965 they passed out with Statistics Honours in the Calcutta University Exam, holding First Class First and First Class Second positions. In 1967, they passed out in M.Sc. [Statistics] Exam under Calcutta University – again holding top two positions. Bimal is senior of the twin brothers and he held 1st position in both the exams! During 1968 - 1971, they carried out doctoral research under the guidance of [Late] Professor H.K. Nandi in the Department of Statistics, Calcutta University. *Bimal* worked in the area of *Bayesian Inference* and *Bikas* worked in the area of *Optimal Designs*. They were awarded Ph.D. Degree [Statistics] of Calcutta University in 1972-1973.

Soon Bimal was inducted as a Lecturer at the Research and Training School [RTS], Indian Statistical Institute, Calcutta on November 1, 1971. And, thereafter, Bikas was

Corresponding Author: Bimal K. Sinha and Bikas K. Sinha
Email: sinha@umbc.edu; bikassinha1946@gmail.com

inducted as the youngest faculty member in Calcutta University Department of Statistics [CUDS] in early 1972. While they were together in the Department of Statistics engaged as research scholars, they worked on a problem in the area of Design of Experiments and this resulted in their first joint publication in 1969 in Calcutta Statistical Association Bulletin.

Subsequently, their work places were virtually different and they were together as faculty by 'choice' on only 1-2 occasions. They have worked together on a total of **8** more research papers – covering such topics as Sequential estimation / Linear regression / Multivariate power series distribution / Bivariate exponential models / Ranked set sampling and Multiple Criteria Decision Making. This list is compiled in Section 6.

2. Bimal K. Sinha [BMS]

Bimal [BMS] joined the Indian Statistical Institute (ISI) Kolkata in 1971 and stayed until 1974, and left for the University of Montreal (UM) in Canada for his post-doctoral research. After staying for one year (1974-1975) at UM, BMS joined the Department of Mathematics/Statistics at the University of Pittsburgh in the fall of 1975 and stayed there for about ten years (with a brief return to ISI/Kolkata for a couple of years: Fall, 1977- Summer, 1980). In 1985 BMS joined the University of Maryland Baltimore County (UMBC) as the Founder of Statistics Graduate Program.

BMS worked extensively on a number of research topics during the last fifty (50) years, published numerous original papers in many international journals, produced thirty (30) doctoral students and became IMS/ASA Fellow. Due to his singular contributions in statistics, UMBC honored him with the title: UMBC Presidential Research Professor. The University System of Maryland (USM) in recognition of BMS's tremendous contributions in statistics offered him the title: USM Research Professor.

BMS has been an ardent devotee of the discipline and uses every opportunity to promote it throughout the globe. His outreach and research collaboration is indeed global with collaborators from many countries: Australia, Canada, India, Thailand, Japan, Taiwan, Germany, Sweden, Portugal. Through an MOU with Mahidol University in Thailand, BMS jointly supervised five (5) doctoral dissertations.

At UMBC, under BMS's initiative/leadership, a series of highly successful annual statistics conferences have been ongoing for the last fifteen (15) years, covering many topics of contemporary relevance. Recently under BMS's pioneering leadership, a series of African International Conferences (AIC) took place in several African countries: Senegal, Cameroon, Ethiopia, Botswana, South Africa, Ghana. This unique initiative has been uniformly praised by the entire statistics community and, in particular, by ASA with commitment for funding.

BMS made pioneering contributions in a number of important topics in mathematical statistics: asymptotic theory (higher order efficiency), decision theory, multivariate analysis, ranked set sampling, statistical meta-analysis, advanced inference in linear models, risk analysis in environmental science, data analysis under confidentiality protection.

BMS is the coauthor of four books [Academic Press, John Wiley (2), Springer], and has written more than 130 original research papers with his national and international collaborators and students, including: Pranab Sen (Chapel Hill), Malay Ghosh (Univ. Florida), (Late) Jayanta Ghosh, Nitis Mukhopadhyay (UConn, Storrs), Anirban Dasgupta (Purdue), Martin Klein (Census Bureau / FDA), Takeaki Kariya (Japan), Jerzy Filar

(Australia), Wei-Shing Shen (Taiwan), Dietrich von Rosen (Sweden), Carloc Coelho (Portugal), Montip Tiensuwan (Thailand), Thomas Mathew / Neerchal Nagaraj (UMBC), Sanat Sarkar (Temple/USA), Zidong Bai / Zehua Chen / Rita Das (Singapore), Roman Zmyslony (Poland)). BMS is also extremely grateful to Dr. Barry Nussbaum [US EPA] and Dr. Tommy Wright [US Census Bureau] for their sincere friendship and research support for a long period of thirty years!

To top it all, BMS is thankful to his twin brother BKS for creating and sharing very rewarding and healthy research atmosphere in some university departments and for giving collaboration opportunities to the students with his rich and diverse research experience.

3. Bikas K. Sinha [BKS]

Bikas [BKS] joined CUDS in early 1972. BKS doctoral thesis was examined by two giants: Late Prof. Jack Kiefer and Prof. C.R. Rao. At the initiative of Prof. C.R. Rao, BKS got an opportunity and accepted a Visiting Faculty Position in one Brazilian University [1975-1979]. This gave him an Int'l exposure and an opportunity to master the Portuguese language.

At the personal initiative of Late Prof. J.K. Ghosh and upon recommendation from Prof. C.R. Rao and Late Prof. D. Basu, BKS was inducted in the Stat-Math Division, ISI, Kolkata, as an Associate Professor in September, 1979 upon his return from Brazil. BKS was promoted to Full Professor at ISI *w.e.f.* January 1, 1985. He retired from ISI on March 31, 2011.

Joining ISI opened up a 'vista' for BKS! At the initiative of Prof. P.K. Sen, BKS was invited as a Visiting Faculty at North Carolina State University, Raleigh, NC, for the academic year 1982 and then again, during July 1, 1985 – June 30, 1986.

Since then there was no looking back for BKS. He has comfortably collaborated with researchers in many different topics in statistical theory and applications and so far he has an impressive list of 105 collaborators worldwide and around 160 research papers/publications. Notable among the collaborators are : Pranab Sen [UNC-Chapel Hill], (Late) Jayanta Ghosh, Malay Ghosh [Univ. Florida, Gainesville], Samad Hedayat [UIC, Chicago], Kirti Shah [Univ. Waterloo, Canada], Erkki Liski [Tampere Univ., Finland], Talluri J Rao [ISI], Rahul Mukerjee [IIMC], (Late) Samin Sengupta [CUDS], Nitis Mukhopadhyay [Uconn., Storrs], Arup Bose [ISI], Montip Tiensuwan [Mahidol Univ., Thailand], Tapio Nummi [Tampere Univ., Finland], Nripes Mandal [CUDS], Friedrich Pukelsheim [Univ. Augsburg, Germany], Thomas Mathew [UMBC, USA], Siddani B Rao [ISI], Prasad Rao [ISI], Manisha Pal [CUDS], Jyoti Sarkar [IUPUI, USA] and Sobita Sapam [Manipur Univ., Imphal]..

Because of expertise in Portuguese language, BKS got an opportunity to serve the UN as an 'Expert on Mission' in the summer of 1991 in Guinea Bissau, West Africa for 3 weeks.

In India, BKS took every opportunity to organize / attend national / international conferences, workshops and he visited almost all corners within the country. This included: special courses structured, developed, organized and conducted in N-E States and there again BKS took a leading role. He visited all these N-E States with great enthusiasm and participated in such activities. He was instrumental to setting up of Tezpur Centre of ISI.

BKS was awarded PCM Gold Medal in 1980 by Sadharan Brahmo Samaj, Kolkata. He has been an Elected Member of Int'l Stat Inst. since 1985. He became Sectional President

(Statistics) in Indian Science Congress Association 89th Annual Meeting held at Lucknow University in 2002. BKS held the prestigious position of Member/Chairman, National Statistical Commission, GoI during the first term [2006-2009]. BKS was also awarded "Centenary Medal for Excellence" by the School of Tropical Medicine, Kolkata in 2014.

BKS continues to have a special bondage with IASRI, Pusa Institute since the time of [Late] Prof. M.N. Das in mid 1970's. He is closely attached to the Research Group in the area of DoE and has visited this institute any number of times as a Resource Person in workshops / conferences. He is a Founder-cum-Life Member of the Society of Statistics and Computer Applications (SSCA).

Special mention must be made of the soft corner BKS has expressed towards CUDS and the country of his origin. He has a great bondage with both the Department of Statistics, CU and Rajshahi University [RU], Bangladesh. Over there in RU, he has organized and participated in three Int'l Conferences with lot of enthusiasm and taking along a good number of Int'l participants in the category of Keynote Speakers / Plenary Speakers / Invited Speakers / Invited Discussants from within Kolkata / India and other foreign countries.

Needless to say, BMS has been instrumental to introducing BKS in a few Int'l Conferences to showcase his twin brother's presentation styles in topics not covered by him! BKS is proud of BMS, being his twin brother !!!

Inherently and interestingly, Sinha Brothers have always maintained a healthy competition in the academic world and in the process, both of them have excelled beyond any bounds.

4. Research Publications of Bimal K. Sinha

4.1. Books

1. Chen, Z., Bai, Z. and Sinha, Bimal K. (2003). *Ranked Set Sampling*: *Theory and Applications*. Springer-Verlag Lecture Notes in Statistics.
2. Hartung, J., Knapp, G. and Sinha, Bimal K. (2008). *Statistical Meta-Analysis with Applications*. John Wiley Series in Probability and Statistics.
3. Kariya, T. and Sinha, Bimal K. (1989). *Robustness of Statistical Tests*. Academic Press, Boston.
4. Khuri, A. I., Thomas, M. and Sinha, Bimal K. (1998). *Statistical Tests in Mixed Linear Models*. John Wiley Series in Probability and Statistics.

4.2. Edited Books

1. Sinha, Bimal K. (1993). *Probability and Statistics*, Volume I. (Joint Editor), Narosa Publishing House, New Delhi.
2. Sinha, Bimal K. (1996). *Probability and Statistics*, Volume II. (Joint Editor), Narosa Publishing House, New Delhi, India.
3. Sinha, Bimal K. (1998). *Applied Statistical Science* III. (Joint Editor), Nova Science Publishers, Inc.
4. Sinha, Bimal K. (2000). *Perspectives in Statistical Sciences*. (Joint Editor), Oxford University Press.

4.3. Research Articles in Journals

1. Ahmed, M., Giri, N. and Sinha, Bimal K. (1983). Estimation of mixing proportion of two known distributions. *Sankhya,* **A45**, 357–371.
2. Ahmed, M., Chaubey, Y. P. and Sinha, Bimal K. (1991). Estimation of a common mean of several inverse Gaussian distributions. *Annals of Institute of Statistical Mathematics*, **43**, 357–367.
3. Banerjee, P. K. and Sinha, Bimal K. (1979). Generating an event with probability $p\propto$, $\propto > 0$. *Sankhya,* **B41**, 282–285.
4. Banerjee, P. K. and Sinha, Bimal K. (1985). Optimal and adaptive strategies in discovering new species. *Sequential Analysis*, **4**, 111–122
5. Basu, S. K. and Sinha, Bimal K. (1975). Locally best invariant and locally minimax test of independence. *Canadian Journal of Statistics*, **3**, 222–12.
6. Chuiv, N., Wu, Z. and Sinha, Bimal K. (1998). Estimation of the location parameter of a Cauchy distribution using a ranked set sample. *Applied Statistical Science* III*,* 297–308.
7. Clement, B., Giri, N. and Sinha, Bimal K. (1978). Effect of additional variates on the power function of multiple correlation R^2-test. *Sankhya*, **B40,** 74-82.
8. Clement, B., Giri, N and Sinha, Bimal K. (1985). Tests for means with additional information. *Communications in Statistics*, **14**, 1427–1452.
9. Dasgupta, A. and Sinha, Bimal K. (1980). On the admissibility of polynomial estimators in the one-parameter exponential family. *Sankhya,* **B29**, 129–142.
10. Dasgupta, A. and Sinha, Bimal K. (1984). Admissibility of generalized Bayes and Pitman estimates in non-regular families. *Communications in Statistics*, **13**, 1709–1722.
11. Dasgupta, A. and Sinha, Bimal K. (1986). Estimation in the multiparameter exponential family: Admissibility and inadmissibility results. *Statistics and Decisions*, **4**, 101–130.
12. Dasgupta, A. and Sinha, Bimal K. (1999). A new general interpretation of the Stein estimate and how it adapts: Applications. *Journal of Statistical Planning and Inference*, **75**, 247–268.
13. Das, R. and Sinha, Bimal K. (1986). Detection of outliers with dispersion slippage in elliptically contoured distributions. *Annals of Statistics*, **14**, 1619–1624.
14. Das, R. and Sinha, Bimal K. (1988). Optimum invariant tests in random MANOVA models. *Canadian Journal of Statistics*, **16**, 193–200.
15. Das, R. and Sinha, Bimal K. (1990). Robust optimum invariant tests of covariance structures useful in linear models. *Sankhya,* **A52**, 244–258.
16. Fei, H., Wu, Z. and Sinha, Bimal K. (1994). Estimation of a gamma mean based on a ranked set sample. *Pakistan Journal of Statistics,* **10(1)A**, 235–249.
17. Fei, Z., Wu, Z. and Sinha, Bimal K. (1994). Estimation of parameters in two-parameter Weibull and extreme-value distributions using ranked set sample. *Journal of Statistical Research*, **28**, 149–162.
18. Fox, J., Frazier, E. and Sinha, Bimal K. (2011). Dose-response modeling for continuous responses: Alternative variance models. *International Journal of Statistical Sciences*, **11**, 189-206.
19. Ghosh, J. K., Joshi, S. M. and Sinha, Bimal K. (1980). On a property of the maximum likelihood estimator. *Sankhya,* **B29**, 143–152.
20. Ghosh, J. K. and Sinha, Bimal K. (977). Multivariate power series distributions and Neyman's properties for multinomial. *Journal of Multivariate Analysis*, **7**, 397–408.

21. Ghosh, J. K., Sinha, Bimal K. and Subramanyam, K. (1977). Edgeworth expansions for Fisher- consistent estimators and second order efficiency. *Calcutta Statistical Association Bulletin*, **28**, 1–18.
22. Ghosh, J. K. and Sinha, Bimal K. (1980). A necessary and sufficient condition for second order admissibility with applications to Berkson's problem. *Annals of Statistics*, **9**, 1334–1338.
23. Ghosh, J. K., Sinha, Bimal K. and Weiand, H. S. (1980). Second order efficiency of the MLE *wrt* any bounded Bowl-shaped loss function. *Annals of Statistics*, **8**, 506–521.
24. Ghosh, J. K. and Sinha, Bimal K. (1982). Third order efficiency of the MLE- a counterexample. *Calcutta Statistical Association Bulletin,* **31**, 151–158.
25. Ghosh, M., Banerjee, P. K. and Sinha, Bimal K. (1981). An admissibility result and its applications. *Aligarh Journal of Statistics*, **1**, 19–22.
26. Ghosh, M., Sinha, Bimal K. and Mukhopadhyaya, N. (1976). Multivariate sequential point estimation. *Journal of Multivariate Analysis*, **6,** 281–294.
27. Ghosh, M. and Sinha, Bimal K. (1980b). On the robustness of least squares procedures in regression models. *Journal of Multivariate Analysis*, **10**, 332–342.
28. Ghosh, M. and Sinha, Bimal K. (1987). Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix, and the generalized variance under entropy loss. *Statistics and Decisions,* **5**, 201–227.
29. Ghosh, M. and Sinha, Bimal K. (1988). Empirical and hierarchical Bayes competitors of preliminary test estimators in two sample problems. *Journal of Multivariate Analysis*, **27**, 206–227.
30. Ghosh, M. and Sinha, Bimal K. (2002). A simple derivation of the Wishart distribution. *American Statistician*, **56**, 100–101.
31. Giri, N., Clement, B., Chakraborty, S. R. and Sinha, Bimal K. (1981). Tests for the mean vector under intraclass covariance structure. *Journal of Statistical Computation and Simulation*, **12**(**3–4**), 237–245.
32. Giri, N. and Sinha, Bimal K. (1975). On the distribution of a random matrix. *Communications in Statistics – Theory and Methods*, **4**, 1057–1063.
33. Giri, N. and Sinha, Bimal K. (1976). On the optimality and non-optimality of some multivariate normal test procedures. *Sankhya*, **B38**, 116–123.
34. Giri, N. and Sinha, Bimal K. (1987). Robust tests of mean vector in symmetrical multivariate distributions. *Sankhya,* **A49**, 254–263.
35. Huang, W. T. and Sinha, Bimal K. (1993). On optimum invariant tests of equality of several intraclass correlation coefficients. *Annals of Institute of Statistical Mathematics,* **45**, 579–597.
36. Kariya, T., Sinha, Bimal K. and Subramanyam, K. (1984a). First, second and third order efficiencies of the estimators for a common mean. *Hitotsubashi Journal of Economics,* **25**, 61–69.
37. Kariya, T., Sinha, Bimal K. and Subramanyam, K. (1984b). Berksons's bioassay problem - revisited. *Sankhya,* **A46**, 408– 415.
38. Kariya, T., Sinha, Bimal K. and Krishnaiah, P. R. (1984). On multivariate left $O(n)$ invariant distributions. *Hitotsubashi Journal of Economics*, **25,** 155–159.
39. Kariya, T. and Sinha, Bimal K. (1985). Nonnull and optimality robustness of some multivariate tests. *Annals of Statistics*, **13**, 1182– 1197.
40. Kariya, T. and Sinha, Bimal K. (1987). Optimality robustness of tests in two population problems. *Journal of Statistical Planning and Inference*, **15**, 167–176.
41. Klein, M., Mathew, T. and Sinha, Bimal K. (2014). Noise multiplication for disclosure limitation of extreme values in log-normal samples. *Journal of Privacy and Confidentiality*, **6**, 77-125.

42. Klein, M., Nagaraj, N., Weihsueh, C., White, P. and Sinha, Bimal K. (2013). Statistical inferences from serially correlated methylene chloride data. *Sankhya,* **B74**, 211-237.
43. Klein, M. and Sinha, Bimal K. (2013). Statistical analysis of noise multiplied data using multiple imputation. *Journal of Official Statistics*, **29**, 425-465.
44. Klein, M. and Sinha, Bimal K. (2015a). Likelihood-based inference for synthetic data based on a normal model. *Statistics and Probability Letters*, **105**, 168-175.
45. Klein, M. and Sinha, Bimal K. (2015b). Inference for singly imputed synthetic data based on Posterior Predictive Sampling under multivariate normal and multiple linear regression models. *Sankhya*, **B77**, 293-311.
46. Klein, M. and Sinha, Bimal K. (2015c). Likelihood-based finite sample inference for singly imputed synthetic data under multivariate normal and multiple linear regression models. *Journal of Privacy and Confidentiality*, **7**(**1**), 43-98.
47. Klein, M., Sinha, Bimal K. and Zylstra, J. (2019). Exact Inference of multiply imputed synthetic data based on plug-in and posterior predictive sampling under a multiple linear regression model. *Calcutta Statistical Association Bulletin,* **71**, 63-82.
48. Knapp, G., Klein, M., Subramanyam, R. and Sinha, Bimal K. (2011). Statistical inferences from Formaldehyde DNA-Protein Cross-Link data: Improving methods for characterization of uncertainty. *Journal of Biopharmaceutical Statistics*, **70**, 42-55.
49. Knapp, G., Sasima, T., Sinha, Bimal K. and Wattanachayakul, S. (2010). On testing equality of several normal means. *Journal of Statistical Theory and Applications,* **9**, 459–477.
50. Knapp, G., Xu, D. and Sinha, Bimal K. (2009). Extracting within-experiment precision of horticultural experiments useful for meta-analysis. *Journal of Applied Horticulture*, **11**, 10-17.
51. Kopylev, L. and Sinha, Bimal K. (2011). The asymptotic distribution of LRT when parameters lie on boundaries. *Sankhya,* **B73**, 20-41.
52. Lam, K., Wu, Z. and Sinha, Bimal K. (1994). Estimation of parameters in a two-parameter exponential distribution using ranked set sample. *Annals of Institute of Statistical Mathematics*, **46**, 723–736.
53. Lam, K., Yu, P. and Sinha, Bimal K. (1997). Estimation of a normal mean based on an unbalanced ranked set sample. *Applied Statistical Science* (Edited by Ahsanullah), 87–97.
54. Lam, K., Yu, P. and Sinha, Bimal K. (1999). Estimation of normal variance based on balanced and unbalanced ranked set samples. *Environmental and Ecological Statistics*, **6**, 23–45.
55. Li, T. and Sinha, Bimal K. (1995). Tests of ordered hypotheses for gamma means. *Journal of Statistical Planning and Inference*, **43**, 311–321.
56. Maitra, R., Ross, P., Lee, H. and Sinha, Bimal K. (2003). On some aspects of data integration techniques with applications. *Computational Mathematics and Modeling*, **14**, 241–256.
57. Mathew, T., Rao, C. R. and Sinha, Bimal K. (1984). Admissible linear estimation in singular linear models. *Communications in Statistics – Theory and Methods*, **13**, 3033–3045.
58. Mathew, T. and Sinha, Bimal K. (1992). Exact and optimum tests in unbalanced split-plot designs under mixed and random models. *Journal of the American Statistical Association*, **87**, 192–200.
59. Mathew, T., Sinha Bimal K. and Sutradhar, B. (1992a). Improved estimation of error variance in general balanced mixed models. *Statistics and Decisions*, **10**, 227–238.

60. Mathew, T., Sinha, Bimal K. and Sutradhar, B. (1992b). Nonnegative estimation of variance components in unbalanced mixed models with two variance components. *Journal of Multivariate Analysis*, **42**, 77–101.
61. Mathew, T., Sinha, Bimal K. and Zhou, L. (1993). Some statistical procedures for combining independent tests. *Journal of the American Statistical Association*, **88**, 912–919.
62. Mathew, T. and Sinha, Bimal K. (1988a). Optimum tests for fixed effects and variance components in balanced models. *Journal of the American Statistical Association*, **83**, 133–135.
63. Mathew, T. and Sinha, Bimal K. (1988b). Optimum tests in unbalanced two-way models without interaction. *Annals of Statistics*, **16**, 1727–1740.
64. Mathew, T. and Sinha, Bimal K. (1991). Towards an optimum test for non-additivity in Tukey's model. *Journal of Multivariate Analysis*, **36**, 68–94.
65. Mathew, T., Sinha, Bimal K. and Niyogi, A. (1994). Improved nonnegative estimation of variance components in balanced multivariate mixed models. *Journal of Multivariate Analysis*, **51**, 83–101.
66. Mitra, P. and Sinha, Bimal K. (2007a). On some aspects of estimation of a common mean of two independent normal populations. *Journal of Statistical Planning and Inference*, **37**, 184-193.
67. Mitra, P. and Sinha, Bimal K. (2007b). On generalized *P*-value approach to inference on common mean. *Journal of Statistical Planning and Inference*, **37**, 3634-3645.
68. Moura, R., Klein, M., Coelho, C. A. and Sinha, Bimal K. (2017). Inference for multivariate regression model based on synthetic data generated under posterior and fixed posterior predictive sampling: comparison with plug-in sampling. *REVSTAT Statistical Journal*, **15**(**2**), 155-186.
69. Nayak, T., Sinha, Bimal K. and Zayatz, L. (2011a). Statistical properties of multiplicative noise masking for confidentiality protection. *Journal of Official Statistics*, **27**, 527-544.
70. Nayak, T., Sinha, Bimal K. and Zayatz, L. (2011b). Privacy protection and quantile estimation from noise multiplied data. *Sankhya,* **B73**, 297-315.
71. Nayak, T. and Sinha, Bimal K. (2012). Some aspects of minimum variance unbiased estimation in presence of ancillary statistics. *Statistics and Probability Letters*, **82**, 1129-1135.
72. Nayak, T. and Sinha, Bimal K. (2015). An appreciation of balanced loss functions via regret loss. *Communications in Statistics - Theory Methods*, **44**, 607-616.
73. Nussbaum, B., Li, X. and Sinha, Bimal K. (2000). A statistical analysis of Hillsdale lake data. *Calcutta Statistical Association Bulletin*, **50**, 293–305.
74. O'Brien, R., Smith, W. J. and Sinha, Bimal K. (1991). A statistical procedure to evaluate cleanup standards. *Journal of Chemometrics*, **5**, 249–261.
75. O'Brien, R., Smith, W. J. and Sinha, Bimal K. (1994). Some aspects of simultaneous inference. *Communications in Statistics - Theory and Methods,* **23**, 59–87.
76. O'Brien, R. and Sinha, Bimal K. (1993). On shortest confidence interval for product of gamma means. *Calcutta Statistical Association Bulletin,* **43,** 181-190.
77. Odom, B. and Sinha, Bimal K. (1995). Asymptotic semiparametric procedures to evaluate cleanup standards. *Journal of Applied Statistical Science* II *(edited by Ahsanullah)*, **10**, 89–102.
78. Pal, N. and Sinha, Bimal K. (1990). Estimation of a common location of several exponentials. *Statistics and Decisions*, **8**, 27–36.
79. Pal, N. and Sinha, Bimal K. (1996). Estimation of a common mean of several normal populations: a review. *Far East Journal of Math. Science,* **71,** 97–100.

80. Pal, N., Shen, W. S. and Sinha, Bimal K. (2000). Estimation of the support of a discrete distribution. *Statistics and Probability Letters*, **46**, 283-286.
81. Pal, M. and Sinha, Bimal K. (2002). Estimation of survival function based on multiple cycle ranked set samples in normal population. *Journal of Statistical Studies*: A special volume in honor of Professor M.M. Ali, **19**, 239–252.
82. Park J., Shah, A. and Sinha, Bimal K. (2014). Testing interval hypotheses for scale parameters in Gamma distributions. *Statistics and Probability Letters*, **85***,* 2168-2171.
83. Park, J., Shah, A., Sinha, Bimal K., Lin, X. and Xu, D. (2015). Likelihood ratio tests for interval hypotheses with applications. *Communications in Statistics - Theory and Methods*, **44***,* 2351-2370.
84. Perron, F., Li, T. and Sinha, Bimal K. (1999). Random selection in ranked set sampling and its applications. *Journal of Statistical Planning and Inference*, **15**, 185–201.
85. Perron, F. and Sinha, Bimal K. (2004). Estimation of variance based on a ranked set sample. *Journal of Statistical Planning and Inference*, **120**, 21–28.
86. Rao, B. R., Clement, B. and Sinha, Bimal K. (1978). Behrens-Fisher problem under the assumption of homogeneous coefficients of variation. *Communications in Statistics – Theory and Methods*, **7**, 637-656.
87. Rukhin, A. and Sinha, Bimal K. (1990-1991). On decision-theoretic estimation of product of gamma scales and generalized variance. *Calcutta Statistical Association Bulletin,* Nandi Memorial Special Issue, **40**, 257–262.
88. Sarkar, S., Sinha, Bimal K. and Krishnaiah, P. R. (1983). Some testing problems with unbalanced data. *Annals of Institute of Statistical Mathematics*, **35**, 63–75.
89. Sarkar, S. and Sinha, Bimal K. (1984). Invariant confidence sequences for some parameters in a general linear regression model. *Annals of Statistics*, **12**, 301–310.
90. Sen, P. K. and Sinha, Bimal K. (1988). Nonparametric estimation of the generalized variance. *Statistics and Decisions*, **6**, 69–78.
91. Shen, W. and Sinha, Bimal K. (1991). Statistical tests for structural relationship. *Statistics and Probability Letters*, **12**, 545–559.
92. Sinha, Bimal K. (1970a). On the equivalence of confidence sets based on Bayesian and non-Bayesian procedures. *Calcutta Statistical Association Bulletin*, **19**, 33–40.
93. Sinha, Bimal K. (1970b). A Bayesian approach to optimum allocation in regression problems. *Calcutta Statistical Association Bulletin*, **19**, 45–52.
94. Sinha, Bimal K. (1971). On the equivalence of Bayesian and non-Bayesian procedures. *Calcutta Statistical Association Bulletin*, **20**, 1–22.
95. Sinha, Bimal, K. (1976a). On unbiasedness of Mann-Wald-Gumbel $\chi 2$ test. *Sankhya*, **A38**, 124–130.
96. Sinha, Bimal K. (1976b). Approximately minimax tests for testing hypotheses about a random parameter with unknown distribution. *Annals of Institute of Statistical Mathematics*, **28**, 249–25.
97. Sinha, Bimal K. (1976c). On improved estimators of the generalized variance. *Journal of Multivariate Analysis*, **6**, 617–625.
98. Sinha, Bimal K. (1980). Is the mle of the common mean of several normal populations admissible? *Sankhya,* **B42**, 192– 196.
99. Sinha, Bimal K. (1984). Detection of multivariate outliers in elliptically contoured distributions. *Annals of Statistics*, **12**, 1558–1565.
100. Sinha, Bimal K. (1985). Unbiased estimation of the variance of the Graybill - Deal estimator of the common mean of several normal populations. *Canadian Journal of Statistics*, **13**, 243–247.

101. Sinha, Bimal K. and Mitra, P. (2007). Inference about a common mean in a one-way random effects model. *Festschrift in honor of Professor Mir Masoom Ali* (Ball State University), 116-140.
102. Sinha, Bimal K. and Mouqadem, O. (1982). Estimation of the common mean of several univariate normal populations. *Communications in Statistics*, **11**, 1604–1614.
103. Sinha, Bimal K. and Mukhopadhyay, N. (1974). Some results on sequential prediction. *Calcutta Statistical Association Bulletin*, **23**, 97–106.
104. Sinha, Bimal K. and Mukhopadhyay, N. (1976). Sequential estimation of a bivariate normal mean vector. *Sankhya*, **B38**, 219–230.
105. Sinha, Bimal K., Rao, C. R. and Subramanyam, K. (1982). Third order efficiency of the maximum likelihood estimator in the multinomial distribution. *Statistics and Decisions*, **1**, 1–16.
106. Sinha, Bimal K. and Wieand, H. S. (1976). Admissibility and minimaxity of MVUE when the parametric space is restricted to integers. *Calcutta Statistical Association Bulletin*, **25**, 165-168.
107. Sinha, Bimal K. and Weiand, H. S. (1977a). Admissibility and inadmissibility of the mle when the parameter space is restricted to integers. *Calcutta Statistical Association Bulletin*, **26,** 113–116.
108. Sinha, Bimal K. and Weiand, H. S. (1977b). Bounds on the efficiencies of four commonly used nonparametric tests of location. *Sankhya*, **B39**, 121–129.
109. Sinha, Bimal K. and Weiand, H. S. (1977c). Multivariate nonparametric tests for independence. *Journal of Multivariate Analysis*, **7**, 272–283.
110. Sinha, Bimal K. and Weiand, H. S. (1979). Union intersection test for the mean vector when the covariance matrix is totally reducible. *Journal of the American Statistical Association*, **74**, 340–343.
111. Sun, Y. and Sinha, Bimal K. (1999). Exact confidence intervals for a common location of several truncated exponentials with different scales. *Communications in Statistics*, **29**, 683–692.
112. Tiensuwan, M., Satinee, L. and Sinha, Bimal K. (2004a). A statistical approach to combining environmental indices with an application to air pollution data from Bangkok, Thailand. *Pakistan Journal of Statistics*, **20(2)**, 245–261.
113. Tiensuwan, M., Satinee, L. and Sinha, Bimal K. (2004b). On a comparison of two standard estimates a binomial proportion based on multiple criteria decision making method. *Journal of Statistical Theory and Applications*, **3**, 141–149.
114. Tiensuwan, M., Satinee, L. and Sinha, Bimal K. (2004c). On a comparison of thre estimators of binomial variance by multiple criteria decision making method. *International Journal of Statistical Sciences*, **3**, 105–117.
115. Tiensuwan, M., Satinee, L. and Sinha, Bimal K. (2004d). A statistical approach to combining environmental indices with an application to air pollution data from Bangkok, Thailand. *Pakistan Journal of Statistics*, **20**, 245–261.
116. Tiensuwan, M., Satinee, L. and Sinha, Bimal K. (2005). On a comparison of four estimates of a common mean by multiple criteria decision making method. *Journal of Statistical Research,* **39,** 23-35**.**
117. Tiensuwan, M., Sukuman, S. and Sinha, Bimal K. (2003). Estimation of reliability based on exponential distribution and ranked set sampling. *Pakistan Journal of Statistics,* **20**, 31–36.
118. Yu, P., Sun, Y. and Sinha, Bimal K. (1999). On exact confidence intervals for the common mean of several normal populations. *Journal of Statistical Planning and Inference*, **15**, 263–277.

119. Yu, P., Sun, Y. and Sinha, Bimal K. (2002). Estimation of the common mean of a bivariate normal population. *Annals of Institute of Statistical Mathematics*, **54,** 861–878.
120. von Rosen, D., Meng, S. Y. and Sinha, Bimal K. (2003). Nonnegative estimation of variance components in multivariate unbalanced mixed linear models with two variance components. *Journal of Statistical Planning and Inference*, **19**, 215–234.
121. Wuttichai, S., Tiensuwan, M. and Sinha, Bimal K. (2011a). On an asymptotic comparison of the maximum likelihood and Berkson's minimum chisquared estimators in dose-response models with one unknown parameter. *Model Assisted Statistics and Applications*, **6**, 21-38.
122. Wuttichai, S., Tiensuwan, M. and Sinha, Bimal K. (2011b). On an asymptotic comparison of maximum likelihood and Berkson's modified minimum chi-square estimates in the two parameter dose-response models. *Journal of Applied Statistical Science* III *(Edited by Ahsanullah)*, **18,** 385-405.

5. Research Publications of Bikas K. Sinha

5.1. Books

1. Bandyopadhyay, S., Rao, A. R. and Sinha, Bikas K. (2011). *Models for Social Networks with Statistical Applications*. Sage Publications.
2. Das, P., Dutta, G., Mandal, N. and Sinha, Bikas, K. (2015). *Optimal Covariate Designs – Theory and Applications*. Springer Publication.
3. Hedayat, A. S. and Sinha, Bikas K. (1991). *Design* and *Inference in Finite Population Sampling*. John Wiley and Sons Publication.
4. Mukherjee, S.P., Sinha, Bikas K. and Chattopadhyay, A. K. (2018). *Statistical Methods in Social Science Research***.** Springer Publication.

5.2. Research Monographs in Optimal Designs [under Lecture Notes in Statistics Series, Springer-Verlag]

1. Liski, E., Mandal, N. K., Shah, K. R. and Sinha, Bikas K. (2002). *Topics in Optimal Design*. **163**.
2. Shah, K. R. and Sinha, Bikas K. (1989). *Theory of Optimal Designs.* **54.**
3. Sinha, Bikas K., Mandal, N. K., Pal, M. and Das, P. (2014). *Optimal Mixture Experiments*. **1028**.

5.3. Research Articles in Journals

1. Abt, M., Gaffke, N., Liski, E. and Sinha, Bikas, K. (1998). Optimal designs in growth curve models II. Correlated model for quadratic growth: optimal designs for parameter estimation and growth prediction. *Journal of Statistical Planning and Inference*, **67**, 287-296.
2. Abt, M., Liski, E., Mandal, N. K. and Sinha, Bikas K. (1997). Optimal designs in growth curve models I. Correlated model for linear growth: optimal designs for slope parameter estimation and growth prediction. *Journal of Statistical Planning and Inference*, **64**, 141-150.

3. Adhikary, S., Ghosh, J. K. and Sinha, Bikas K. (2001). Group testing for identification of defective components: an application of orthogonal arrays. *Statistics and Applications*, **3**, 89-96.
4. Adhikary, B. and Sinha, Bikas K. (1976). On group divisible rotatable designs. *Calcutta Statistical Association Bulletin*, **25**, 79-93.
5. Ahsanullah, M. and Sinha, Bikas K. (1986). On normality *via* conditional normality. *Calcutta Statistical Association Bulletin*, **35**, 199-202.
6. Alam, M., Imon, R. and Sinha, Bikas K. (2006). Maximum likelihood estimation of a finite population size. *Journal of Statistical Theory and Applications,* **5**, 306-315.
7. Bagchi, S., Mukhopadhyay, A. C. and Sinha, Bikas K. (1990). A search for optimal nested row-column designs. *Sankhya*, **B52**, 93-104.
8. Bagchi, S., Sengupta, S. and Sinha, Bikas K. (2009). Estimation of mean life and reliability for exponential life distribution using time-censored life data. *Statistics and Probability Letters*, **79**, 899-906.
9. Bagchi, S. and Sinha, Bikas K. (2020). Some inferential aspects of mixture sampling designs. *Thai Statistician*. To appear.
10. Basu, S., Dewanji, A., Ganguly, K. S., Mohanto, P. P., Samanta, T. and Sinha, Bikas K. (2002). A study of pulsed electro-magnetic field therapy for rheumatoid arthritis patients : analysis of adjustment period data. *Calcutta Statistical Association Bulletin*, **53**, 289-302.
11. Bhandari, S. K. and Sinha, Bikas K. (1992). Nonadaptive hypergeometric group-testing designs for detection of two defective items - some combinatorial aspects. *Sankhya* **A54**, 57-62, Special Issue.
12. Bhaumik, D., Bhaumik, R., Nordgren, R., Jie, F. and Sinha, Bikas K. (2018). A mixed-effects model for detecting disrupted connectivities in heterogeneous data. *IEEE Transactions on Medical Imaging*, **37**, 2381-2389.
13. Bose, A. and Sinha, Bikas K. (1984). Sequential Bernoulli sampling plans re-examined. *Calcutta Statistical Association Bulletin*, **33**, 109-120.
14. Bose, A. and Sinha, Bikas K. (1985). Unbiased sequential estimation of $1/p$: settlement of a conjecture. *Annals of the Institute of Statistical Mathematics*, **37**, 455-460.
15. Chang, C., Chaudhuri, G., Pal, N. and Sinha, Bikas K. (1995). Estimation of a multivariate normal mean vector and local improvements. *Statistics*, **26**, 1-17.
16. Chattopadhyay, G., Pal, M. and Sinha, Bikas K. (2002). On strong and weak confounding in a complex system: characterization and Examples. Forum for Interdisciplinary Mathematics Proceedings on Statistics, Combinatorics and Related Areas (Bombay, 2000), *American Journal of Mathematical Management Sciences*, **22**, 303-320.
17. Chattopadhyay, G. and Sinha, Bikas K. (2017). Coverage probability and exact inference. *Journal of Statistical Theory and Practice* (Online Publication), **12**, 93-99.
18. Das, K., Mandal, N. K. and Sinha, Bikas K. (2003). Optimal experimental designs for models with covariates. *Journal of Statistical Planning and Inference*, **115**, 273-285.
19. Das, P., Mandal, N. K. and Sinha, Bikas K. (2007). Optimal designs in linear regression models with heteroscedastic errors. *Intenational Journal of Statistical Sciences*, **6**, 259-268.
20. Das, P., Mandal, N. K. and Sinha, Bikas K. (2008). Optimum mixture designs: A pseudo Bayesian approach. *Journal of Indian Society of Agricultural Statistics*, **62**, 174-182.
21. Das, P., Mandal, N. K., Pal, M. and Sinha, Bikas K. (2008). Optimum mixture designs under constraint on mixing components. *Statistics and Application*, **6**, 173-190.

22. Das, P., Mandal, N. K., Pal, M. and Sinha, Bikas K. (2010). Optimum mixture designs under random regression coefficients. *Journal of Applied Probability and Statistics*, **5**, 15-29.
23. Das, P., Mandal, N. K., Pal, M. and Sinha, Bikas K. (2015). Optimum mixture designs in a restricted region. *Statistical Papers*, **56**, 105-119.
24. Das, K., Mukhoti, S. and Sinha, Bikas K. (2008). On some aspects of unbiased estimation of parameters in quasi-binomial distributions. *Communications in Statistics - Theory and Methods*, **37**, 3023-3028.
25. Das, P., Pal, M., Mandal, N. K. and Sinha, Bikas K. (2010). Parameter estimation in linear and quadratic mixture models: A Review. *Pakistan Journal of Statistics*, **26**, 77-96.
26. Das, P., Pal, M., Mandal, N. K. and Sinha, Bikas K. (2011). Optimum designs in linear mixture models with synergistic effects. *Journal of Statistical Theory and Applications*, **10**, 305-322.
27. Das, P. and Sinha, Bikas K. (2017). Mixture designs with orthogonal blocks. *Communications in Statistics - Theory and Methods*, **16**, 8155-8165.
28. Dharmadhikari, A. D., Sabnis, S. and Sinha, Bikas K. (2007). Construction of bounds in prototype testing. *Journal of Statistical Theory and Applications*, **6**, 302-312.
29. Dharmadhikari, A. D., Sabnis, S. and Sinha, Bikas K. (2020). A new family of probability density functions with deflated tail. *Statistics and Applications* (New Series) **18(2)**, Special Volume in honour of twin statisticians Bimal and Bikas Sinha on their 75th birthday. To appear.
30. Dhungana, A. R., Sinha, Bikas K. and Tiensuwan, M. (2010a). Sufficiency in bivariate and trivariate normal populations. *Far East Journal of Theoretical Statistics*, **32**, 59-80.
31. Dhungana, A. R., Sinha, Bikas K. and Tiensuwan, M. (2010b). Sufficiency in linear and quadratic regression models. *Journal of Statistical Theory and Applications*, **9**, 387-404.
32. Domagni, F., Hedayat, A. S. and Sinha, Bikas K. (2020). On the nature of saturated 2^*k*-factorial designs for unbiased estimation of non-negligible parameters. *Journal of Statistical Planning and Inference*. To appear.
33. Dong, Y., Hedayat, A. S. and Sinha, Bikas K. (2008). Surveillance strategies for detection of change point in incidence rate based on exponentially weighted moving average method. *Journal of the American Statistical Association*, **103**, 843-853.
34. Dutta, G. and Sinha, Bikas K. (2013). Some further aspects of assessment of agreement involving bivariate normal responses. *International Journal of Statistical Sciences*, **13**, 1-19.
35. Dutta, G. and Sinha, Bikas K. (2017a). Nearly optimal covariate designs – Part I. *Communications in Statistics - Theory and Methods*, **46**, 9691-9702.
36. Dutta, G. and Sinha, Bikas K. (2017b). Nearly optimal covariate designs – Part II. *Communications in Statistics - Theory and Methods*, **46**, 9703-9725.
37. El-Neweihi, E. and Sinha, Bikas K. (2000). Reliability estimation based on ranked set sampling. *Communication in Statistics - Theory Methods*, **29**, 1583-1595.
38. Gerig, T. and Sinha, Bikas K. (1984). On a further characterization of dispersion matrices based on the properties of regression. *Journal of Statistical Planning and Inference*, **10**, 265- 271.
39. Gerig, T. and Sinha, Bikas K. (1985). On Neyman's conjecture: a characterization of the multinomial. *Journal of Multivariate Analysis*, **16**, 440-450.
40. Ghosh, J. K., Maiti, P., Rao, T. J. and Sinha, Bikas K. (1999). Evolution of Statistics in India. *International Statistical Review*, **67**, 13-34.

41. Ghosh, M., Khuri, A., Mukherjee, B. and Sinha, Bikas K. (2006). Design issues for generalized linear models: A review. *Statistical Science*, **21**, 379-399.
42. Ghosh, M. and Sinha, Bikas K. (1990). On the consistency between model- and design-based estimators in survey Sampling. *Communication in Statistics -Theory Methods*, **19**, 689-702.
43. Goswami, A., Sengupta, S. and Sinha, Bikas K. (1990). Optimal strategies in sampling from a social network. *Sequential Analysis*, **9**, 1-18.
44. Goswami, A. and Sinha, Bikas K. (2006). Some probabilistic aspects in the discovery of species. *Sequential Analysis,* **25***,* 103-115.
45. Hedayat, A. S., Keyu, N. and Sinha, Bikas K. (2017). Unbiased estimation of reliability function from a mixture of two exponential distributions based on a single observation. *Statistics and Probability Letters*, **127**, 7-13.
46. Hedayat, A. S., Keyu, N. and Sinha, Bikas K. (2018a). Revisiting the general pattern of domination using the Graybill-deal estimator. In "J.N.K. Rao Felicitation Volume". *Statistics and Applications* (*New Series*), **16**, 91-103.
47. Hedayat, A. S., Keyu, N. and Sinha, Bikas K. (2018b). Performance of the Graybill-deal estimator *via* Pitman closeness criterion. *Journal of Statistical Theory and Applications*, **17**, 291-306.
48. Hedayat, A. S., Li, W. and Sinha, Bikas K. (2012). On some aspects of conditional power evaluation in two-phase clinical trials under linear regression. *Journal of Statistics and Operations Research* [Statistics in the Twenty-First Century Special Volume in Honor of Dr. Mir Masoom Ali], **8**, 393-413.
49. Hedayat, A. S., Lin L., Sinha, Bikas K. and Yang, M. (2002). Statistical methods in assessing agreement: models, issues and tools. *Journal of the American Statistical Association*, **97**, 257-270.
50. Hedayat, A. S., Lou, C. and Sinha, Bikas K. (2009). A statistical approach to assessment of agreement involving multiple raters. *Communications in Statistics - Theory and Methods*, **38**, 2899 - 2922.
51. Hedayat, A. S. and Sinha, Bikas K. (2003). On a sampling design for estimation of negligible accident rates involving electronic toys. *American Statistics*, **57**, 249-252.
52. Hedayat, A. S., Sinha, Bikas K. and Zhang, W. (2011). Some aspects of inference on a normal mean with known coefficient of variation. *International Journal of Statistical Sciences* (Special Issue in Honor of P.C. Mahalanobis), **11**, 159–181.
53. Heiligers, B. and Sinha, Bikas K. (1995). Optimality aspects of Agrawal's designs - II. *Statistica Sinica*, **5**, 599-604.
54. Huda, S. and Sinha, Bikas K. (1986). On cylindrically rotatable designs of type 1. *Journal of Statistical Research*, **20**, 53-63.
55. Koskela, L., Nummi, T. and Sinha, Bikas K. (2005a). On some statistical properties of the apportionment index. *Revista Investigacion Operacional*, **26**, 169-179.
56. Koskela, L., Nummi, T. and Sinha, Bikas K. (2005b). Statistical properties of the apportionment degree and alternative measures in bucking outcome. *Revista Investigacion Operational*, **26**, 259-267
57. Koskela, L., Nummi, T. and Sinha, Bikas K. (2005c). On a family of apportionment indices and its limiting properties**.** *IAPQR Transactions*, **30**, 65-87.
58. Koskela, L., Nummi, T. and Sinha, Bikas K. (2007). Some aspects of the sampling distribution of the apportionment index and related inference. [*A Quarterly Journal of Forest Science*], **41**, 699-715.
59. Lacayo, Herber, Neerchal, N. and Sinha, Bikas K. (2002). Ranked set sampling from a dichotomous population. *Journal of Application and Statistical Sciences*, **11**, 83-90.

60. Liski, E., Luoma, A. and Sinha, Bikas K. (1996). Optimal designs in random coefficient linear regression models. *Calcutta Statistical Association Bulletin*, **46**, 211-229.
61. Liski, E., Luoma, A., Mandal N. K. and Sinha, Bikas K. (1996/97). Optimal design for an inverse prediction problem under random coefficient regression models. *Journal of the Indian Society of Agricultural Statistics*, **49**, 277-288.
62. Liski, E., Luoma, A., Mandal N. K. and Sinha, Bikas K. (1998). Optimal designs for prediction in random coefficient linear regression models. J. N. Srivastava Felicitation Volume, *Journal of Combintorics Information System Sciences*, **23**, 1-16.
63. Liski, E., Luoma, A., Mandal N. K. and Sinha, Bikas K. (1998). Pitman nearness, distance criterion and optimal regression designs. *Calcutta Statistical Association Bulletin*, **48**, 179-194.
64. Luoma, A., Mandal N. K. and Sinha, Bikas K. (2001). A-optimal cubic and quadratic regression designs in asymmetric factor spaces. *Statistics and Applications*, **3**, 147-153.
65. Luoma, Arto, Nummi, T. and Sinha, Bikas K. (2007). Optimal designs in random coefficient cubic regression models. Special Issue in Celebration of the Centennial of The Birth of Samarendra Nath Roy (1906-1964). *Journal of Statistical Planning and Inference*, **137**, 3611- 3617.
66. Maiti, P., Pal, M. and Sinha, Bikas K. (1992). Estimating unknown dimensions of a binary matrix with applications to estimation of the size of a mobile population. *Statistics and Probability*. Edited Volume S. K. Basu and Bimal K. Sinha. Narosa Publishing.
67. Maiti, P., Sarkar, J. and Sinha, Bikas K. (2018). Estimation of a finite population proportion in light of randomized reporting. *Journal of Statistical Theory and Applications*, **17**, 597-605.
68. Maiti, P., Sengupta, S. and Sinha, Bikas K. (2006). Unbiased estimation in a finite population: case of multiple indirect identifiability of population units. *Journal of Statistical Theory and Applications*, **5**, 81-90.
69. Maiti, P. and Sinha, Bikas K. (2014). Interactive linear models in surveys. *Journal of Statistical Theory and Applications*, **13**, 263 -272.
70. Mamunur, R. and Sinha, Bikas K. (2016). Data integration for techniques in agricultural sciences. *Journal of the Society Application of Statistics in Agriculture and Allied Sciences*, **1**, 5-10.
71. Mandal, N. K., Pal, M. and Sinha, Bikas K. (2013). Some finer aspects of the de la Garza Phenomenon. *Statistics and Applications*, **11**, 1-14.
72. Mandal, N. K., Pal, M. and Sinha, Bikas K. (2015). D-optimal designs and model uncertainty in mixture experiments. *International Journal of Statistical Sciences*, **15**, 1-12.
73. Mandal, N. K., Pal, M. and Sinha, Bikas K. (2017). Growth models for repeated measurement mixture experiments: optimal designs for parameter estimation and growth prediction. In "*Advances in Growth Curve and Structural Equation Modeling*: Topics from the Indian Statistical Institute Proceedings 2017)", Ed. Ratan Dasgupta, Springer, 81-94**.**
74. Mandal, N. K., Pal, M. and Sinha, Bikas K. (2018). Optimum designs for pharmaceutical experiments with relational constraints on the mixing components. In "*Advances in Growth Curve and Structural Equation Modeling*: Topics from the Indian Statistical Institute" on the 125th Birth Anniversary of P. C. Mahalanobis. Ed. Ratan Dasgupta, Springer, 49-57.
75. Mandal, N. K., Sapam, S. and Sinha, Bikas K. (2018). Combined intra- and inter- block analysis of balanced ternary designs. *Journal of Statistical Theory and Applications*, **17**, 91-100.

76. Mandal, N. K., Sapam, S. and Sinha, Bikas K. (2019a). Latin square designs with neighbor effects. *Journal of the Indian Society of Agricultural Statistics*, **73**, 91-98.
77. Mandal, N. K., Sapam, S. and Sinha, Bikas K. (2019b). Latin square designs with neighbor effects - Part II. *Communications in Statistics-Theory and Methods*, Published online, https://doi.org/10.1080/03610926.2019.1702694.
78. Mandal, N. K., Shah, K. R. and Sinha, Bikas K. (2000). Comparison of test *vs*. control treatments using distance optimality criterion. *Metrika*, **52**, 147-162.
79. Mandal, N. K., Shah, K. R. and Sinha, Bikas K. (1990). Uncertain resources and optimal designs: problems and perspectives. *Calcutta Statistical Association Bulletin*, **40**, 267-282.
80. Marcovitz, M., Nandy, K. and Sinha, Bikas K. (2016). Eliciting information on sensitive features: Block total response technique and related inference. In "*Handbook of Statistics*, Special Volume on Golden Jubilee Celebration on Randomized Response Technique', Chapter **19**, 317-330.
81. Mathew, T., Rao, P., Rao, S. B. and Sinha, Bikas K. (2014). A new class of optimal designs in the presence of a quantitative covariate. *International Journal of Statistical Sciences*, **14**, 1-16.
82. Mathew, J. and Sinha, Bikas K. (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference*, **93**, 295-307.
83. Mukerjee, R., Shah, K. R. and Sinha, Bikas K. (1992). E-optimal minimally connected block designs under mixed effects model. *Metrika*, **39**, 359-364.
84. Mukerjee, R. and Sinha, Bikas K. (1982/83). A note on the universal optimality criterion for full rank models. *Journal of Statistical Planning and Inference*, **7**, 97- 100.
85. Mukerjee, R. and Sinha, Bikas K. (1990). Almost saturated D-optimal main effect plans and allied results. *Metrika*, **37**, 301-307.
86. Mukhopadhyay, A. C., Saharay, R. and Sinha, Bikas K. (1990). Non-additive linear models: estimability and efficient estimations of interaction parameters. *Communication in Statistics – Theory and Methods*, **19**, 739-764.
87. Mukhopadhyay, A. C. and Sinha, Bikas K. (1998). Two stage designs for locating the optimum of second order response function and related results. J. N. Srivastava Felicitation Volume. *Journal of Combinatorics Information and System Sciences*, **23**, 59-69.
88. Mukhopadhyay, B. B. and Sinha, Bikas K. (1980). A note on a result of M. Ghosh and Bimal K. Sinha: "On the robustness of least squares procedures in regression models". [Journal of Multivariate Analysis, 10 (1980), 332-342]. *Calcutta Statistical Association Bulletin*, **29**, 169-171.
89. Mukhopadhyay, N., Sen, P. K. and Sinha, Bikas K. (1988). Bounded risk estimation of a finite population mean: optimal strategies. *Sequential Analysis Journal*, **7**, 91-109.
90. Mukhopadhyay, N., Sen, P. K. and Sinha, Bikas K. (1989). Stopping rules, permutation invariance and sufficiency principle. *Annals Institute Statistical Mathematics*, **41**, 121-138.
91. Mukhoti, S., Sengupta, S. and Sinha, Bikas K. (2006). Unbiased estimation of the distribution function of an exponential population using order statistics with application in ranked set sampling. *Communication in Statistics – Theory and Methods*, **35**, 1655-1670.
92. Nandy, K. and Sinha, Bikas K. (2020). Block total response technique for quantitative sensitive features in a finite population. *Statistics and Applications* (*New Series*), **18(2)**. Special Volume in honour of twin statisticians Bimal and Bikas Sinha on their 75th birthday. To appear.

93. Nordhansen, K. and Sinha, Bikas K. (2006). ML estimation for a Pareto distribution of the second kind in a sequential design. *Journal of the Indian Statistical Association,* **44**, 61-72.
94. Padmawar, V. R. and Sinha, Bikas K. (2016). Estimating the size of a population comprising of individuals with indirect accessibility. *Journal of Indian Statistical Association*, **53**, 175-200.
95. Pharita, Em-Ot, Tiensuwan, M. and Sinha, Bikas K. (2008). Some aspects of stochastic modeling of dyadic relations in social networks: Theory and applications. *Journal of Statistical Theory and Applications*, **7**, 303-322.
96. Pornpis, Y., Tiensuwan, M. and Sinha, Bikas K. (2006). Cohen's Kappa statistic: a critical appraisal and some modifications. *Calcutta Statistical Association Bulletin*, **58,** 151-169.
97. Pukelsheim, F. and Sinha, Bikas K. (1995). Optimal block designs revisited: an approximate theory detour. *Statistics and Probability Letters*, **24**, 375-380.
98. Raghavarao, D., Sethuraman, V. S. and Sinha, Bikas K. (2006). "Optimal *s^n* factorial designs when observations within-blocks are correlated. *Computational Statistics and Data Analysis*, **50**, 2855-2862.
99. Rao, B. V. and Sinha, Bikas K. (1985). A characterization of Dirichlet distributions. *Journal of Multivariate Analysis*, **25**, 25-30.
100. Rao, P., Rao, S. B., Saha, G. M. and Sinha, Bikas K. (2003). Optimal designs for covariates' models and mixed orthogonal arrays. *Electronic Notes in Discrete Mathematics, Elsevier*, *Amsterdam*, **15**, 157-160.
101. Rao, P., Rao, S. B. and Sinha, Bikas K. (2006). Improved bounds for group testing designs. *Journal of Statistical Planning and Inference*, **136**, 260-269.
102. Rao, P., Rao S. B., Saha, G. M. and Sinha, Bikas K. (2006). Some combinatorial aspects of a counterfeit coin problem. *Linear Algebra and Applications [LAA]:* Special Issue.
103. Rao, P., Rao, S. B. and Sinha, Bikas K. (2013). Group testing designs - A combinatorial marvel. *Statistics and Applications*, **11**, 111-126.
104. Rao, P. and Sinha, Bikas K. (2013). Correlated bivariate linear growth models: optimal designs for slope parameter estimation. In "*Advances in Growth Curve Models*: Topics from the Indian Statistical Institute". Editor: Ratan Dasgupta, 89-100.
105. Rao, P. and Sinha, Bikas K. (2019). Some aspects of optimal covariate designs in factorial experiments. *Statistics and Applications*, **17**, 77-86.
106. Rao, T. J., Sarkar, J. and Sinha, Bikas K. (2016). Randomized response and new thoughts on politz-simmons technique. In "*Handbook of Statistics: Special Volume on Golden Jubilee Celebration of RRT*", Ch. 15, 233-251.
107. Rao, T. J., Sengupta, S. and Sinha, Bikas K. (1991). Some order relations between selection and inclusion probabilities for PPSWOR sampling scheme. *Metrika*, **38**, 335-343.
108. Rao, T. J., Sinha, Bikas K. and Srivenkataramana, T. (2003). On order relations between selection and inclusion probabilities in RHC sampling scheme. *Journal of Appllied Statistical Science*, **12**, 67-73.
109. Rao, T. J. and Sinha, Bikas K. (2011). A brief history of statistics and its development in the indian subcontinent. *International Journal of Statistical Sciences* (Special Issue in Honor of P. C. Mahalanobis), **11**, 1-15.
110. Roy, D. and Sinha, Bikas K. (1992). Characterization of probability laws using paired comparison models. *Order Statistics and Nonparametrics: Theory and Applications*, Alexandria, 1991), 431-437, North-Holland, Amsterdam.

111. Saba, M. and Sinha, Bikas K. (2014). SuDoKu as an experimental design – The traditional Latin square design. *Statistics and Applications*, **12**, 15-20.
112. Saha, G. M. and Sinha, Bikas K. (1985). On the equivalence of two approaches to simplicity in the analysis of block designs and some related results. *Journal of Statistical Planning and Inference*, **11**, 237-240.
113. Saha, Rita and Sinha, Bikas K. (1983). Optimal weighing designs with a string property. *Journal of Statistical Planning and Inference*, **8**, 365-374.
114. Sapam, S. and Sinha, Bikas K. (2020). On the status of variance balanced block designs in the presence of both-sided neighbor effects: two examples. *Statistics and Applications* (*New Series*), **18**(**2**). Special Volume in honour of twin statisticians Bimal and Bikas Sinha on their 75th birthday. To appear.
115. Sapam, S., Meitei, Singh K.K. and Sinha, Bikas K. (2021). Randomized block designs, balanced incomplete block designs and Latin square designs with neighbor effects in the presence of covariates. *Statistics and Applications* (*New Series*), **19**(**1**), 19-31. Special volume in memory of Late Aloke Dey. To appear.
116. Sapam, S. and Sinha, Bikas K. (2020). Graeco-Latin Square Designs with Neighbor Effects. *In "Celebrating the Centenary of Professor C.R. Rao" Special Volume of The Journal of Statistical Theory and Practice. (2020/2021)* To appear.
117. Sarkar, J. and Sinha, Bikas K. (2014). Variants of SuDoKu as experimental designs. *Statistics and Applications*, **12**, 35-60.
118. Sarkar, J. and Sinha, Bikas K. (2015). The SuDoKu family of puzzles–A combinatorial marvel. *Resonance*, 711-725.
119. Sarkar, J. and Sinha, Bikas K. (2016). Weighing designs to identify a single counterfeit coin. *Resonance*, 25-50.
120. Sastry, P. and Sinha, Bikas K. (1986). Linear invariance and admissibility in sampling finite populations. *Sankhya*, **B48**, 246-257.
121. Sastry, P., Sengupta, S. and Sinha, Bikas K. (1987). Some inferential aspects of finite population sampling with additional resources. *Journal of Statistical Planning and Inference*, **16**, 203-211.
122. Sen, M. and Sinha, Bikas K. (1986). A statistical analysis of serially balanced sequences: first order residual proportional to direct effects. *Calcutta Statistical Association Bulletin*, **35**, 31-36.
123. Sen, P. K. and Sinha, Bikas K. (1989). On averaging over distinct units in sampling with replacement. *Sankhya*, **B51**, 65-83.
124. Sengupta, S. and Sinha, Bikas K. (1990). Coherent mixtures of SRSWOR sampling schemes for bounded risk estimation of a finite population mean. *Sequential Analysis*, **9**, 269-282.
125. Sengupta, S. and Sinha, Bikas K. (1993). Estimation of the probability of discovering a new category in finite population Sampling. *Calcutta Statistical Association Bulletin*, **43**, 75-84.
126. Shah, K. R. and Sinha, Bikas, K. (1988). Optimality aspects of 3-concurrence most balanced designs. *Journal of Statistical Planning and Inference*, **20**, 229-236.
127. Shah, K. R. and Sinha, Bikas, K. (1989). On the choice of optimality criteria in comparing statistical designs. *Canadian Journal of Statistics*, **17**, 345-348.
128. Shah, K. R. and Sinha, Bikas, K. (1993a). Optimal designs with component-wise orthogonal row-column structures. *Journal of Combinatorics Information and System Sciences*, **18**, 68-78.
129. Shah, K. R. and Sinha, Bikas, K. (1993). Optimality aspects of row-column designs with non-orthogonal structure. *Journal of Statistical Planning and Inference*, **36**, 331-346.

130. Shah, K. R. and Sinha, Bikas, K. (1996b). Row-column designs. Design and analysis of experiments. *Handbook of Statistics*, **13**, 903-937. North-Holland, Amsterdam.
131. Shah, K. R. and Sinha, Bikas, K. (2001). Discrete optimal designs: criteria and characterizations. In "*Recent Advances in Experimental Designs and Related Topics* (Philadelphia, PA, 1999)", Nova Science Publishers, Huntington, NY,133-151.
132. Shah, K. R. and Sinha, Bikas, K. (2002a). Universal optimality of completely randomized designs. In "*Recent Advances in Statistical Methods* (Montreal, QC, (2001))". IR.mp. Coll. Press, London. 290-295.
133. Shah, K. R. and Sinha, Bikas, K. (2002b). On some aspects of data integration techniques with environmental applications. *Journal of Environmetrics*, **14**, 409-416.
134. Sinha, Bikas K. (1970a). On the optimality of some designs. *Calcutta Statistical Association Bulletin*, **19**, 1- 22.
135. Sinha, Bikas, K. (1970b). Invariant problems of linear inference and related designs. *Calcutta Statistical Association Bulletin*, **19**, 103-122.
136. Sinha, Bikas, K. (1972). On sampling schemes to realize pre-assigned sets of inclusion probabilities of first two orders. *Calcutta Statistical Association Bulletin*, **22**, 89-110.
137. Sinha, Bikas, K. (1973). Comparison of some experiments from sufficiency consideration. *Annals of the Institute Statistical Mathematics*, **25**, 501-520.
138. Sinha, Bikas, K. (1974). On sampling schemes to realize invariant pre-assigned sets of inclusion probabilities of first two orders. *Calcutta Statistical Association Bulletin*, **23**, 45-72.
139. Sinha, Bikas, K. (1975a). On some optimum properties of serially balanced sequences. *Sankhya*, **B37**, 173-192.
140. Sinha, Bikas, K. (1975b). Regular cyclically invariant problems of linear inference and related designs. *Sankhya*, **B37**, 79-90.
141. Sinha, Bikas, K. (1976). On balanced sampling schemes. *Calcutta Statistical Association Bulletin*, **25**, 129-138.
142. Sinha, Bikas, K. (1981). Some further combinatorial and constructional aspects of generalized Youden designs. *Combinatorics and Graph Theory*, (Calcutta, 1980), pp. 488-495, Lecture Notes in Mathematiocs, 885, Springer, Berlin-New York.
143. Sinha, Bikas, K. (1982). Some aspects of simplicity in the analysis of block designs. *Journal of Statistical Planning and Inference*, **6**, 165- 172.
144. Sinha, Bikas, K. (1982/83). On complete classes of experiments for certain invariant problems of linear inference. *Journal of Statistical Planning and Inference*, **7**,171-180.
145. Sinha, Bikas, K. (1991). Sequential methods for finite populations. In *Handbook of Sequential Analysis*, **118**, 313-329. Dekker, New York.
146. Sinha, Bikas, K. (1997). Some inference aspects of a social network. *Applied Statistical Science*, II, 77-86, Nova Science Publishers, Commack, NY.
147. Sinha, Bikas, K. (2009). A reflection on the choice of covariates in the planning of experimental designs. *Journal of the Indian Society of Agricultural Statistics*, **63**, 219-225.
148. Sinha, Bikas, K. (2010). On some aspects of social network models. *Journal of the Indian Society of Agricultural Statistics*, **64**, 289-293.
149. Sinha, Bikas, K. (2017). Some refinements of block total response technique in the context of RRT methodology. *Statistics and Applications*, **15**, 167-171.
150. Sinha, Bikas K. (2020). Linear model perspectives of fMRI studies. *Special Proceeding of the 22nd Annual Conference of SSCA* held at Savitribai Phule Pune University, Pune, (January 2–4, 2020), 39-51.

151. Sinha, Bikas K., Tiensuwan, M. and Sunthornworasiri, N. (2009). Bayesian approach to reliability estimation for the exponential distribution using order statistics. *Journal of Statistical Theory and Applications*, **8**, 207-231.
152. Sinha, Bikas K. and Zielinski, R. (1997). Estimating $P[X > Y]$ in exponential model revisited. *Statistics*, **29**, 299-316.

6. Research Publications with both Sinhas as co-authors

1. Das, K., Sinha, Bikas K. and Sinha, Bimal K. (1993). Variance components estimation—On the status of Grubbs' estimators. In *Quality Through Engineering Design*. W. Kuo (Editor), Elsevier Science Publishers B.V., 141-157.
2. Ghosh, J. K., Sinha, Bimal K. and Sinha, Bikas K. (1977). Multivariate power series distributions and Neyman's properties for multinomials. *Journal of Multivariate Analysis*, **7**, 397-408.
3. Purkayastha, S., Sinha, Bimal K. and Sinha, Bikas K. (1996). On some aspects of ranked set sampling for estimation of normal and exponential parameters. *Statistics and Decisions*, **14**, 223- 240.
4. Sinha, Bimal K. and Sinha, Bikas K. (1969). Comparison of some experiments from sufficiency consideration. *Annals of the Institute of Statistical Mathematics*, **25**, 501-520.
5. Sinha, Bimal K. and Sinha, Bikas K. (1975). Some problems of unbiased sequential binomial estimation. *Annals of the Institute of Statistical Mathematics*, **27**, 245-258.
6. Sinha, Bimal K. and Sinha, Bikas K. (1976). On a characterization of the dispersion matrix based on the properties of regression. *Communication in Statistics – Theory Methods*, **A5**, 1215-1224.
7. Sinha, Bimal K. and Sinha, Bikas K. (1992). Unbiased sequential binomial estimation. *Current Issues in Statistical Inference.* Essays in honor of D. Basu. *IMS Lecture Notes Monograph Series*, **17**, 75-85. *Institute of Mathematical Statistics*, Hayward, CA.
8. Sinha, Bimal K. and Sinha, Bikas K. (1995). An application of bivariate exponential models and related inference. *Journal of Statistical Planning and Inference*, **44**, 181-191.
9. Sinha, Bimal K. and Sinha, Bikas K. (2002). On some aspects of MCDM with environmental applications. In Honour of the 65th Birthday of Professor Mir Masoom Ali. *Journal of Statistical Studies* (Suppl.), 139-144.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 1-14

Wanted! A Mathematician

Jyotirmoy Sarkar
Department of Mathematical Sciences, Indiana University-Purdue University Indianapolis, USA

Received: 21 February 2020; Revised 11 March 2020; Accepted: 14 March 2020

Abstract

How will a mathematician identify a single poisonous bottle of wine from among 1000, if she is permitted only one opportunity to make the fewest number of test subjects drink small extracts from these bottles?

Key words: Optimization; Precision; Resource allocation; Duality principle; Binary numbers; Design of experiments.

PREAMBLE

I revisit a puzzle that has proliferated the Internet in its many different incarnations. Not all sites report the solution to the puzzle; and those that do, do so in a matter-of-fact manner, without explaining how the solution was discovered or why it is optimal. This includes Coldwell (2019), which I liked the most. My objective here is to derive the optimal solution starting from first principle. Additionally, I adopt a story-telling style in hope of exposing a vast array of readers to the secrets of how a mathematician goes about practicing the creative art of Mathematical Sciences. I conclude the paper inviting the reader to solve another optimization problem.

To my family and friends, a reassurance: The story here is entirely fictitious, with no hidden agenda to promote either wine drinking, gambling or calculated killings.

WHAT'S THE PROBLEM?

1. Travel to 20 CE

Hop on a time machine, travel back to 20 CE (common era), and visit the kingdom of the mythical Irish King Conchobar mac Nessa of Ulster. The king is facing an unprecedented predicament. Consequently, he has made an edict inviting all and sundry to participate in a contest in which the winner (to be determined if no one else beats the participant's performance within the next 24 hours) will receive as reward ten thousand gold coins, and any loser (beaten by someone else within the 24 hours limit) will not only lose face, but also lose his head. Will you join the contest?

I think you should not forgo this golden opportunity - after you have derived the optimal solution (with proof) or you have carefully read this paper.

Corresponding Author: Jyotirmoy Sarkar
Email: jsarkar@iupui.edu

2. The King's Conundrum

King Conchobar amassed 1000 bottles of exotic wine, which he had collected from lands far and near, and preserved in a heavily guarded cellar. He had curated the bottles for the express purpose of indulging and impressing his select guests at the *Coronation Anniversary Celebration* coming up in five weeks' time. Unfortunately, his treasured possession was stealthily invaded by a neighboring queen's clever spy, who managed to inject one bottle with poison so lethal that anyone who drinks just a single drop will surely die — though not immediately, but in exactly 30 days. As fate would have it, the spy was quickly caught by the king's elite guards who demanded to know which bottle he had poisoned. However, the spy was unwilling to identify the contaminated bottle, even when offered one thousand gold coins as reward, for he could not trust the guards' offer. Moreover, preferring to demonstrate his total loyalty to his queen even unto death, he swallowed a fatal pill which he had brought with him for a situation just as this one and committed instant suicide — hurling the king in a conundrum.

This suicidal death of the only person who knew which single wine bottle was contaminated with poison left the king first to ponder about how to identify the offending bottle and save the remaining 999 bottles for his prestigious party; then to become progressively puzzled, bewildered and hopelessly perplexed; and eventually to write an edict offering ten thousand gold coins to anyone who could identify the poisonous bottle. He would let the identifier devise a clever experiment in which a few of his 1000 prisoners of war, whom he had captured a year ago when he had invaded the neighboring kingdom, would be forced to drink a concoction extracted from one or more bottles at least 31 days before his Anniversary. He would reward the proposer who is properly trained in the science of mathematics and in the art of exposition who could explain to him, though he himself was not a mathematician by any stretch of the imagination, that indeed the experiment would involve as few prisoners as absolutely necessary. You see, the king wanted to save as many prisoners as possible to serve as slaves, and yet with a very high probability identify the poisonous bottle. In fact, the king had resolved in his mind that on the eve of his celebration, avoiding any spectacle and arousing no suspicion from his subjects and guests, every single experimental prisoner who would survive the forced drinking would be put to death in complete secrecy.

The king's edict also included a rejoinder: Within 24 hours of a proposed solution, if someone else would discover a better solution, which would either increase the probability of correctly identifying the poisonous bottle with the same number of experimental prisoners or fewer, or reduce the number of prisoners without lowering the probability of identification, then the prize would go to the latter solver; and the former proposer would be taunted, humiliated and publicly beheaded in the infamous *Field of Gallows*.

The king sent his emissaries all over the kingdom proclaiming his edict and inviting potential contestants who would design for him the most ideal solution to identify the poisonous bottle with a high probability subject to minimizing the number of experimental subjects. Posters proliferated the marketplace, public squares and sports arena: "Wanted! A Mathematician."

On arrival at Ulster, you learn about this edict from your host family who do their best to dissuade you from participating; but you, who has the benefit of two thousand more years of accumulated human knowledge than the then citizens of Ulster, are not going to give up so easily, are you? Having realized that in order to earn the reward and to save your head (along with your face) you must not only find a solution to the puzzle, but also have the utmost

confidence (via a mathematical proof) that no one else will beat your solution either by lowering the number of experimental subjects or by increasing the probability of correct identification, will you accept the king's challenge?

THINK LIKE A MATHEMATICIAN

3. Put Your Thinking Cap On

Not wishing to give up the great, albeit dangerous, opportunity, you put your thinking cap on and start to ponder over the challenge: You begin with a naïve solution that matches each wine bottle with a unique prisoner, and makes each prisoner drink one shot from the bottle allocated to him. You even think of assigning a different bottle to 999 prisoners and leaving one bottle unassigned, since if no prisoner dies, then the unassigned bottle must be the poisonous one. However, within a short time you rule out this solution because although it would identify the poisonous bottle with 100% certainty, it would also engage too many prisoners in the experiment and expose you to a risk that someone else would easily reduce the number of prisoners. Likewise, you also must discard a second solution which uses only half as many prisoners and makes each experimental prisoner drink a concoction made of one-half shot from each of the two bottles allocated to him. For in this case, while on the eve of the Anniversary you will know for sure which pair of bottles includes the contaminated one, you will not know for sure which one of this pair is the truly poisonous one. Admittedly, compared to detecting one poisonous bottle from among 1000 bottles, it is a much simpler task to detect one bottle out of two. Nonetheless, it is impossible to do so with probability exceeding 1/2, for there remains only one night before the celebration party, rendering it unfeasible to conduct a follow-up experiment!

Proceeding in this manner, you reject a whole family of designs which allocate disjoint batches of b bottles to each of $\lceil 1000/b \rceil$ prisoners (with the last prisoner perhaps being allocated fewer than b bottles), and make each experimental prisoner drink one shot made by mixing $1/b$ fraction of a shot extracted from each of the bottles allocated to him, for while the number of experimental prisoners decreases as the batch size b increases, the probability of correctly identifying the poisonous bottle decreases to only $1/b$, since you will only identify the batch that contains the poisonous bottle, but not the poisonous bottle itself.

As you ponder more over the above family of designs, all at once it dawns on you that you have inadvertently imposed an additional constraint over the solution that was neither explicitly mentioned in the king's edict, nor implied by it: While you permitted a prisoner to drink from multiple bottles, you have allowed only one prisoner to drink from each bottle! Surely someone must necessarily drink from each bottle, save perhaps one (so that at most one bottle is excluded from the experiment); but there was no requirement to restrict each bottle to only one prisoner. How can you construct a more efficient experiment (that is, involve fewer prisoners) that allocates each bottle to a multiplicity of prisoners allowing each prisoner to drink a small extract from that bottle along with extracts from all other bottles allocated to that prisoner and still identify the poisonous bottle?

4. A Sudden Inspiration

While you keep pondering over how to allocate "bottles to prisoners" and "prisoners to bottles," you hear some commotion out in the street caused by people going to the *Field of Gallows* to witness two prisoners who would be hanged, for they had broken into the king's

cellar and during the chase that followed to catch them they had knocked off one bottle of wine — shattering it into a thousand pieces and ruining its content. Although curious as a cat, you resist the urge to follow the mob to the *Gallows*. Instead, you put multiple thinking caps on and come to realize two features that would affect your solution: (1) You no longer have 1000 prisoners to engage in your experimental study — your precious resource has depleted to 998 prisoners; and (2) either the contaminated bottle is among the 999 bottles still intact, or it has been already destroyed! That is, *at most one* bottle among 999 is poisonous. You say to yourself: "The number of bottles and the number of prisoners have changed; and these numbers might change again! Therefore, I must be prepared to solve the king's conundrum not only for 1000 bottles and 1000 prisoners (or for 999 bottles and 998 prisoners), but also for any number of bottles B and any number of prisoners P."

With these realizations, should you feel happy or sad? On the surface, it looks like your task has exploded out of proportion compared to the one you began with — as if the challenge has become almost insurmountable. However, on deeper reflection, a light bulb goes on over your head (this is a purely fictitious idiomatic construction, since there wasn't any light bulb around in the first century; but remember you have time traveled from the twenty-first century): "Perhaps I can solve the problem for small values first, then detect a pattern among the solutions, and eventually extend the solution to any pair (B, P)." A much harder challenge seems to have given birth to a wonderful new opportunity!!

5. Solve Some Simpler Problems First

Suppose that among B bottles *exactly one* is poisonous. You can identify the poisonous bottle for small values of B, say for 1, 2 and 3. Then if you notice a systematic pattern among the solutions, perhaps you can conjecture the solution for an arbitrary value of B, and thereafter prove that conjecture.

In fact, for $B = 1$, the problem is already solved: The only available bottle is poisonous.

For $B = 2$, hopefully the king himself could solve the problem based on his own daily experience, without having to pay a mathematician! At every meal, as the king cautiously watches, his butler takes a portion from the king's plate and eats, ensuring the king that his food is safe to eat. Translated to the problem at hand: If $B = 2$, it suffices to enlist $P = 1$ prisoner and have him drink a shot from Bottle 1. If he dies (in 30 days), Bottle 1 is poisonous and the other bottle (labelled as Bottle 0) is safe; if he survives, Bottle 1 is safe, and Bottle 0 must be poisonous.

Had the king made one prisoner drink a little from each of the two bottles, then surely the prisoner would die; and the king would not know which bottle killed him. On the other hand, if the king had enrolled two prisoners and made each prisoner drink a little from a different bottle and kept track of who drank from which bottle, he would have surely identified the poisonous bottle, but he would have acted sub-optimally according to the terms of his own edict.

What if there is *exactly one* poisonous bottle among $B = 3$ bottles? Then one prisoner is not enough; but $P = 2$ prisoners suffice. Label the bottles with serial numbers 1, 2, 3. Assign Bottle 1 to Prisoner 1, Bottle 2 to Prisoner 2, and Bottle 3 to both prisoners. Let each prisoner drink from the two bottles assigned to him. Surely, at least one prisoner must die. If both prisoners die, then Bottle 3 is poisonous; otherwise, if only Prisoner 1 dies, then Bottle 1 is poisonous; and if only Prisoner 2 dies, then Bottle 2 is poisonous.

For $B = 4$ bottles, the same reasoning above shows that two prisoners suffice to detect the single poisonous bottle with 100% certainty: Just label the newest bottle as 0 and assign it to neither prisoner. If both prisoners survive, Bottle 0 must be poisonous. Thus, in the presence of three bottles, an additional fourth bottle did not make the problem more complex: We simply do nothing to the fourth bottle. Alternatively, having learned the solution to $B = 4$, we can construct the solution to $B = 3$ simply by eliminating any one of the four bottles. Thus, we discover a multiplicity of solutions for $B = 3$. For instance, we could assign one bottle to each of the two prisoners, and set aside the third bottle, assigning it to neither prisoner. Now at most one prisoner may die. If neither prisoner dies, then the bottle that was set aside is poisonous; otherwise, whichever bottle the dead prisoner had drunk from is poisonous. Although there are multiple solutions to $B = 3$ bottles and $P = 2$ prisoners, the solution to $B = 4$ is unique.

How are the solutions to $B = 2$ and $B = 4$ interrelated? Starting from the solution to either problem, can we construct the solution to the other problem? Notice that for $B = 2$, we set aside one bottle and make one prisoner drink from the other bottle. Likewise, for $B = 4$, we set aside one bottle and make each of the two prisoners drink from exactly two bottles, giving them a common bottle to drink from and then another bottle unique to each. In the next paragraph we describe an alternative way to understand this allocation of bottles to the two prisoners that will reveal how the solution for $B = 4$ can arise out of the solution for $B = 2$.

Imagine that the four bottles are rearranged into two bundles of two bottles each — very much like two bottles are packaged together to promote a buy-one-get-one-free deal in a twenty-first century grocery store. Set aside one bundle and assign the other bundle to Prisoner 1. Then the fate of Prisoner 1 will detect which bundle contains the contaminated bottle. This is exactly the solution to the $B = 2$ case. Next, to determine which member of the suspected bundle is the contaminated bottle, we need to experiment again using a second prisoner, except that such sequential experimentation is expressly disallowed. Fortunately, we can pick one bottle from each bundle and assign the two chosen bottles to Prisoner 2 at the same time we start to experiment with Prisoner 1, and then the responses from the two prisoners will be available at the same time. Thus, each of the 4 bottles is matched to *a unique subset* of the two prisoners. Accordingly, the death of a specific subset of prisoners ($\emptyset$, $\{1\}$, $\{2\}$, $\{1,2\}$) uniquely identifies the poisonous bottle.

Now we are ready to move on to the next step in the generalization: Among $B = 8$ bottles, *exactly one* is poisonous. In this case, simply form 4 pairs; allocate the pairs to two prisoners using the above solution to the $B = 4$ case, by bundling two pairs together, etc. Remember that assigning a bundle to a prisoner is the same as assigning all bottles within the bundle to that prisoner. Their fate will determine which pair contains the contaminated bottle. Simultaneously, allocate one bottle from each pair to Prisoner 3, whose fate will determine which member of the detected bundle is the contaminated bottle. More specifically, pair up Bottles 1-2, 3-4, 5-6, 7-8. To Prisoner 1 assign Bottles 5-6-7-8, to Prisoner 2 assign Bottles 3-4, 7-8, and to Prisoner 3 assign the even-numbered Bottles 2, 4, 6, 8. You may permute the bottles and/or permute the prisoners any way you like.

In this manner, for any value of $B = 2^k$, a power of 2, we can extend the above method of allocating $B = 2^k$ bottles to k prisoners.

If B is not a power of 2, simply augment some more bottles, filled with harmless water (or even keeping them empty), until there is a total of $B = 2^k$ bottles. For example, suppose that there were 15 bottles, one of which is poisonous. How will you conduct the experiment to

detect the poisonous bottle? Augment a bottle of water; label it 1; and label the other bottles with serial numbers 2 through $16 = 2^4$. You enroll four prisoners, labelled 1-4. Give Prisoner 1 extracts from even numbered bottles; that is, alternately skip a bottle, include a bottle. For Prisoner 2, alternately skip two bottles, then include two bottles; that is, give Bottles 3-4, 7-8, 11-12, 15-16. To Prisoner 3, alternately skip four bottles, then include four bottles; that is, give Bottles 5-8, 13-16. To Prisoner 4, give extracts from the last eight bottles 9-16. Note that no prisoner got anything from Bottle 1, which you had augmented, and is surely not poisonous. It is straightforward to verify that depending on which bottle is poisonous, the subset of dead prisoners after 30 days will be different.

Reversing the logic, once you know which prisoners have died 30 days later, you can identify the poisonous bottle X uniquely! For example, suppose that Prisoners 1, 2 and 4 die, but Prisoner 3 is alive. Since Prisoner 4 died, X is among 9-16 (the latter half); since Prisoner 3 is alive, X is among 9-12 (the beginning half of the candidate bottles from the previous step); since Prisoner 2 died, X is among 11-12 (why?); and since Prisoner 1 died, X is 12 (since it must be even). Eureka!

6. Binary Codes to Allocate Bottles to Prisoners

For $B = 2^4$ bottles and $P = 4$ prisoners, to smartly conduct the experiment and to confidently identify the offending bottle, you may want to label the bottles with four-digit binary codes 0000 to 1111 (representing numbers 0 through 15, the previously stated serial numbers 1 through 16 reduced by one). Using these binary codes, assign to Prisoner 1 extracts from all eight bottles that have 1 in the rightmost digit; to Prisoner 2 assign all eight bottles that have a 1 in the second digit from right; etc. After 30 days, when you know the fates of all prisoners, summarize that information by writing a 0 for a live prisoner and a 1 for a dead prisoner, starting from the rightmost digit for Prisoner 1 and moving leftward prisoner by prisoner. This summary code *is* the label of the poisonous bottle!

The above strategy of allocating bottles to prisoners is easily extended to P prisoners and $B = 2^P$ bottles, when exactly one bottle is poisonous.

7. Proving Optimality

Can you prove that indeed four is the fewest number of prisoners needed when there is exactly one poisonous bottle among $B = 15$ bottles? For if you cannot, you will have no confidence that your head will remain in its proper place if King Conchobar is still reigning.

To prove optimality of our proposed solution, we utilize a duality principle at play here. It changes the original problem into an equivalent dual problem, whose solution may be easier.

The Duality Principle: Optimization problems may be viewed from either of two perspectives — the primal problem and the dual problem. It suffices to solve either problem; the other problem is immediately solved. Moreover, the solution to the primal (minimization) problem provides an upper bound to the solution of the dual (maximization) problem; likewise, the solution to the dual (maximization) problem provides a lower bound to the solution of the primal (minimization) problem.

Returning to our detection of the single poisonous bottle out of B bottles, let us state the primal and dual problems.

Primal Problem: Given B bottles, with *exactly one* poisonous among them, to determine the fewest number of prisoners P needed to detect the poisonous bottle with the highest probability.

Dual Problem: Given P experimental prisoners, to find the largest number of bottles B so that the single poisonous bottle from among B can be identified with the highest probability.

8. Solving the Dual Problem

The dual problem can be easily solved for small values of P, say for 1, 2 and 3. Then having noticed a systematic pattern in the solutions, one may conjecture a reasonable solution for an arbitrary value of P, and prove the conjecture. We follow this strategy below.

If $P = 1$ prisoner is available, we can have him drink a shot from one bottle. If he dies (in 30 days), the bottle is poisonous; if he survives, the bottle is safe. If there are 2 bottles and it is known that *exactly one* of them is poisonous, then also $P = 1$ prisoner suffices. Let him drink from one bottle and set aside the other bottle: If he dies in 30 days, then the bottle he drank from is poisonous and the other bottle is safe; if he survives, then the bottle he drank from is safe and the other bottle is poisonous. Making him drink from both bottles is futile: For then, he will surely die; and we would not know which bottle killed him.

Next, we must explain that if there are three bottles with *exactly one* of them poisonous, then $P = 1$ prisoner is not sufficient to detect the poisonous bottle. If the prisoner drinks from two or more bottles and dies, we cannot identify which bottle killed him; if he drinks from only one bottle and survives, we cannot tell which of the remaining two bottles is poisonous. Thus, with $P = 1$ prisoner, we can detect the single poisonous bottle from among at most $B = 2$ bottles.

Now consider the situation when there are two bottles of wine and *at most one* of them is poisonous. In this case, one prisoner will not suffice, you will need two prisoners. Here is why. With only one prisoner available, we have two choices: (1) Make him drink from one bottle. If he dies on the 30th day, we know the bottle he drank from is poisonous; and the other bottle is safe. If he survives beyond the 30 days, we know the bottle he drank from is safe; and the second bottle may be either safe or poisonous, but we will not know the complete truth. (2) Make the prisoner drink a little from each of the two bottles. If he survives, then both bottles are safe. If he dies, then one of the bottles is poisonous; but we do not know which one. Thus, in each case, we fail to discover complete information about the two bottles. Therefore, we must enroll a second prisoner in the experiment; assign one bottle to each; make them drink a portion from the assigned bottle. If both prisoners survive beyond 30 days, then both bottles are safe. If not, the dead prisoner must have drunk from the poisonous bottle and the surviving one from the safe bottle. Note that both prisoners cannot die since *at most one* bottle is poisonous.

Let us return to the case when *exactly one* of the bottles is poisonous. If $P = 2$ prisoners are available, we can double the number of bottles to $B = 4$. Pair up the bottles to form two bundles. Simply use Prisoner 1 to detect the bundle with the poisonous bottle (ensuring that the prisoner drinks from both bottles within the bundle assigned to him). Simultaneously, choose one bottle from each bundle and assign them to Prisoner 2 to detect which member of the

bundle is the poisonous bottle. Label the bottles with binary codes 00, 01, 10, 11. Then let Prisoner 1 drink from the second and the fourth bottles (which in binary code have 1 in the rightmost digit), and Prisoner 2 from the third and the fourth bottles (which have 1 in the leftmost digit). If both prisoners die, then the fourth bottle is poisonous; otherwise, if only Prisoner 1 dies, then the second bottle is poisonous; if only Prisoner 2 dies, then the third bottle is poisonous; finally, if none of the prisoners dies, then the first bottle, from which neither prisoner drank, is poisonous.

We should also check that with $P = 2$ prisoners available, it is not possible to detect the single poisonous bottle from among five bottles. To prove this impossibility, for each bottle, ask yourself: "To whom is the bottle assigned?" There are exactly four possible answers: The bottle is assigned to both prisoners, only to Prisoner 1, only to Prisoner 2, to neither prisoner. Therefore, by the pigeonhole principle [see Wikipedia (2019)], at least two bottles must be assigned to the exact same subset of prisoners. Should every member of that subset of prisoners die and no other prisoner die, then we would not know which of the two or more bottles assigned to that subset of prisoners is poisonous.

We leave to the reader to study the situation when there are four bottles of wine and *at most one* of them is poisonous. Two prisoners will not suffice, you will need a third prisoner.

By now a clear pattern has emerged, which we state as a Theorem.

Theorem 1: Exactly P prisoners suffice to detect the single poisonous bottle from among $2^{P-1} < B \leq 2^P$ bottles; but fewer than P prisoners do not suffice. If among 2^P bottles *at most one* is poisonous, then we must enroll $(P + 1)$ prisoners.

Proof: Suffices it to prove the theorem for the largest value of B, namely, 2^P. (For fewer than 2^P bottles, fill $(2^P - B)$ additional bottles with safe-to-drink water and conduct the experiment for 2^P bottles.) Label the bottles (after permuting them randomly) with serial numbers 0 through $(2^P - 1)$ written in binary codes consisting of P digits ranging from $(000 \ldots 0)$ to $(111 \ldots 1)$. Then assign to Prisoner j all those bottles that have 1 in the j-th digit from right. In other words, there is a one-to-one correspondence between the bottles and all possible subsets of P prisoners. Therefore, the subset of prisoners who die in 30 days identifies the poisonous bottle with 100% accuracy: The binary code for the poisonous bottle has in digit j from right the value 1 if Prisoner j is dead, and the value 0 if Prisoner j is alive.

If fewer than P prisoners are available, by the pigeonhole principle multiple bottles will have to be assigned to the same subset of (fewer than P) prisoners. Should that subset of prisoners and no other prisoner die, then we would not know which one of these multiple bottles is poisonous.

When there are 2^P bottles of which *at most one* is poisonous, P prisoners will not suffice: One more prisoner must be enrolled and made to drink a shot from Bottle 000…0 (from which none of the previous P prisoners drank) to determine whether this bottle is safe or poisonous, just in case no other prisoner dies.

This completes the proof of the theorem. □

Applying Theorem 1, we conclude that it suffices to enlist 10 prisoners in King Conchobar's experiment to detect with complete certainty *at most one* poisonous bottle from among 999 bottles, since $2^9 < 999 < 2^{10}$; but 9 prisoners will not do. Go ahead and accept King Conchobar's challenge; just remember to pass on as royalty 15% of your reward to yours truly when you do safely return to the twenty-first century.

9. Executing the Experiment in Practice

To maintain complete secrecy and absolute control over the experiment, the King himself should decide who will drink from which bottle (after he learns the strategy from the mathematician). In complete secrecy of his cellar, he should prepare 10 cups with distinct IDs monogrammed on them so that he would know who drank the cup. He should arrange these cups in random order in 10 positions. Then he should make tags with labels ranging from 0 to 1023, written in ten-digit binary codes such as 0000001101(=13) or 1101110100 (=884), but discard the $22 = \binom{10}{0} + \binom{10}{1} + \binom{10}{9} + \binom{10}{10}$ tags that have 0, 1, 9 or 10 ones in them (that is, discard serial numbers 0 to 10 and 1013 to 1023). Although not necessary, in order to achieve a perfect balance, the king should augment the bottles of wine with a few more bottles of water for a total of 1002 bottles; and assign a unique tag to each of the 1002 bottles in a random order.

The king should prepare what goes into each of the 10 cups, where each cup corresponds to a digit (position) of the binary code. From each bottle, labeled with a unique binary code, he should draw a small amount of wine (say, 1/4 ml if each bottle contains 1 liter) to put into each cup that corresponds to a digit (position) with value 1, and not into the other cups that correspond to digits (positions) with value 0. (He can use syringes to extract wine from the bottle without opening the cork, provided he carefully washes any syringe clean before reusing it.) Each cup will contain the concoction made up of portions drawn from exactly half of the 1002 bottles, thereby containing only 125.25 ml total. This is because every digit (position) has as many 1's as 0's — the balance we referred to earlier.

Thereafter, the king should make a public announcement that he will not only set free but also elevate to nobility ten prisoners on the auspicious occasion of his anniversary — the ten who are judged winners in a series of athletic competitions to be held immediately. This will ensure that the prisoners enrolled in his crafty experiment are healthy, will likely not die of any other cause in the next one month, and will participate willfully and joyfully, oblivious to his devious scheme. The king will invite these 10 athletic winners to a royal dinner, where they will be served the cup with the secret ID matched to each experimental prisoner.

For such an experiment to be successful (from the king's perspective), the king must ensure that none of his experimental subjects dies during these 30 days for any other reason. Perhaps he should invite them to dinner every evening for the next 30 days on pretext of teaching them proper manners of nobility, but truly for keeping attendance and checking on their health. To ensure absolute certainty that no one will kill himself or another participant enrolled in the experiment, he should assign guards and physicians to look after their total wellbeing. It is of paramount importance that he knows exactly which subset of the 10 prisoners died because of unknowingly drinking from the poisoned bottle — for that subset of dead prisoners will uniquely identify the poisonous bottle.

On the 31st day, the king will know which experimental participants have died. Whichever unique bottle was assigned to this subset of dead prisoners is the poisonous one!

While this subset may be of size 2 to 8, on average 5 subjects are expected to die of poisoning. Of course, to rule out any future information leak, the king will very likely renege on his promise; and kill all surviving experimental participants. The king can now enjoy the remaining bottles of wine (minus the 1/2-2 ml drawn out of each) and send the one "special bottle" as a gift to the neighboring queen, with "PEACE" inscribed on it.

ACT LIKE A STATISTICIAN

10. Connection to *Design of Experiments*

We narrated the above fictitious short story hoping to inspire students to learn the beautiful and useful art and science of experimental designs. How is the story of King Conchobar related to *Design of Experiments*?

First, we find it astonishing that back in 20 AD, King Conchobar literally heeded the sage advice of our modern-day statisticians:

"Experimentation is an essential part of any problem of decision-making. Whenever one is faced with the necessity of accepting one out of a set of alternative decisions, one has to undertake some experiments to collect observations on which the decision has to be made."

— Shah and Sinha (2012)

In the story, we can substitute some terminologies from *Design of Experiments*: For instance, each bottle of wine can be thought of as a treatment to be assigned to one or more prisoners, each of whom can be thought of as an experimental unit (on which we can apply as many treatments as we wish).

Since only one treatment is fatal and all other treatments are innocuous, we are essentially conducting a hypothesis test among 1000 hypotheses (each stating one particular bottle is poisonous or all 999 bottles are innocuous), based on data consisting of a single dichotomous response variable — the prisoner is either dead or alive after 30 days. Indeed, since the king has diluted the poisonous drink by a factor of 1 in 501, and each cup either contains ¼ ml of poisonous wine or none at all, the poison remains potent; and it will surely kill any unfortunate soul that drinks it.

In fact, our design is so well thought out that we need no sophisticated analyses: The responses from the ten subjects (almost magically) suffice to identify the poisonous bottle! Thus, the hypothesis test is 100% accurate, with zero probability of Type I error (declaring a bottle poisonous when it is not) and zero probability of Type II error (declaring a bottle safe when it is poisonous), provided that no one dies from a cause other than drinking from the poisoned bottle.

Our story illustrates the following two quotes from leading experts in *Design of Experiments* on the importance of choosing the experimental design carefully:

"If the experimental design is wisely chosen, a great deal of information in a readily extractable form is usually available, and no elaborate analysis may be necessary."

— Box, *et al*. (2005)

"If you do the pre-experiment planning carefully and select a reasonable design, the analysis will almost always be relatively straight-forward. In fact, a well-designed experiment will sometimes almost analyze itself!"

— Montgomery (2013)

Our story also demonstrates some best practices propounded by experts in *Design of Experiments*: An appropriate experimental design is a solution to an optimization problem that expends the least amount of resources and still extracts enough information to resolve an issue with the highest possible precision. One must be mindful of utilizing resources to their maximum potential; practice all kinds of safeguards to reduce biases in the study; and above all, one must not compromise the quality of knowledge one seeks to discover.

A well-known strategy to reduce biases in an experiment is to incorporate proper randomization (that is, to the extent permitted, units must be chosen at random to receive a treatment or a combination of treatments). To accomplish this, we advised the king to randomly assign the binary codes to the wine bottles and to randomly permute the monogrammed cups in positions 1 through 10. Another useful concept in experimental design is balance; for example, each experimental unit must receive the same number of treatments. In the king's experiment, we advocated augmenting two bottles of water to ensure that every cup receives extracts from 501 bottles and therefore contains the same amount of wine (125.25 ml). On the other hand, every treatment (bottle) was applied to 2-8 experimental units (prisoners) according as the binary code assigned to the treatment. Another key concept in implementing a designed experiment is to permit replication (that is, multiple units receive the same treatment combinations) with an aim to reduce associated statistical errors. Since the king's experimental design already achieves a 100% accuracy, there is no need for further reduction of error. Hence, no replication is needed, or recommended.

Lastly, the sanctity of the response variable must be preserved. In the king's experiment, the cause of death must be none other than consumption of poisonous wine. Therefore, we advised the king to identify the healthiest prisoners through athletic competitions, to offer them freedom and a bright future to ensure their cooperation and desire to survive, and to keep them under watch by guards to prevent any homicide and to appoint physicians to treat them of any other ailment.

11. A Variation on the Detection Problem

Recall that the prisoners who drink from the poisonous bottle die not immediately, but 30 days later. Implicitly we are assuming that death can occur at a random time before the 30 days are over; that is, during the time period (0, 30]. Other than knowing the support, the exact probability distribution of the delay time between drinking and death is unknown. This was the reason for restricting the experiment to only one opportunity; that is, make all experimental subjects drink wine at the same time.

However, suppose that death will occur sometime during the period (29½, 30] days after drinking the poisonous wine. Then the experiment can be conducted on four successive days. In such a case, the king can get by with engaging only 8 prisoners in his experiment: On Day 1, he will extract wine from Bottles 1-256 to assign to the 8 prisoners according to the binary rule described in Section 6. On Day 2, he will extract wine from Bottles 257-512 to give to the same 8 prisoners. On Day 3, he will use Bottles 513-768 to assign to the same 8 prisoners. On Day 4, he will use Bottles 769-1000 (plus 24 water bottles) to assign to the same 8 prisoners.

If some prisoners die on Day 31 minus half a day, then using the subset of dead prisoners, the king will identify the poisonous bottle from among 1-256. Otherwise, if all prisoners survive on Day 31, then all these bottles are innocuous, and the king must wait to check the survival status on Day 32. If some prisoners die on Day 32 minus half a day, then using the subset of dead prisoners, the king will identify the poisonous bottle from among 257-512. Otherwise, if all prisoners survive on Day 32, all these bottles are innocuous. And so on.

Referring to *Design of Experiments* literature, we are reminded of a crossover design, in which the same unit receives different treatment combinations in different time periods provided that the response is attributable to the correct treatment combination. For the king's experiment, the response on each prisoner is no longer a binary variable taking values 1 or 0; rather it is a quintenary variable taking values 1, 01, 001, 0001, 0000, according as the time of death is Day 31, 32, 33, 34 or no death at all respectively. Thus, when the time of death after drinking from the poisonous bottle is within half a day of the 30^{th} day mark, we have reduced the number of experimental units to 8, without compromising the inference.

Carrying this argument further, if anyone drinking from the poisonous bottle will surely die within 23 hours, then the king can conduct his devious experiment on 32 nights, requiring only 5 prisoners and utilizing 32 distinct bottles each night. Each prisoner's status will be one of 33 possible outcomes: Death before Day 2, 3, …, 33 or Survival. Thus, with more precise information on the response variable, the sample size can be reduced without sacrificing the quality of inference. Moreover, the experiment can be terminated as soon as at least one prisoner dies.

THINK SOME MORE

12. Further Study

We invite the astute reader to solve another optimization problem.

Exercise

Suppose that a building has 1000 floors above ground. If you drop a marvelous marble from floor N or above, the marble will surely break; but if you drop it from any lower floor, there will be absolutely no effect of the impact. Being as good as new, it can be dropped again (from a higher floor). Every time you want to drop a marble, you must pay ₹10 (with a coupon) to take the elevator to the desired floor. Taking the down-elevator to check whether the marble is intact or broken costs you nothing. At the start of the experiment, you can buy any number of marbles for ₹50 each and any number of coupons for ₹10 each. At any other time, you cannot buy or sell a marble or a coupon. What is the least amount of money you must spend to determine N with complete certainty?

Note that you must minimize the *maximum* amount of money you may spend, and not minimize the expected amount of money.

My answer to the Exercise is ₹330; and I offer this answer in good faith that in case you beat my solution within 24 hours, you won't demand my head. Partial explanation of my answer is given in the Appendix. Can you find a better solution? Or, can you prove that my solution is indeed optimal? Please email me (at jsarkar@iupui.edu) a better solution or a proof that my solution is the best.

Acknowledgements

The author is indebted to Professor George D. Gopen, who through his workshop, articles and books taught him Gopen's *Reader Expectation Approach* to writing science; and who inspired him to incorporate writing into the classroom. The author thanks all students who participated in a writing assignment on this topic and suggested some refinements. The author expresses gratitude to Monica Pena Pardo for giving detailed feedback on an earlier draft. Finally, the author sincerely thanks the Editor for his generous suggestions to improve the paper's impact.

References

Coldwell, N. *A Collection of Quant Riddles with Answers, Riddle 61*. Retrieved on November 07, 2019 from http://puzzles.nigelcoldwell.co.uk/.

Wikipedia, The Free Encyclopedia. *Pigeonhole Principle.* Retrieved on November 07, 2019 from https://en.wikipedia.org/wiki/Pigeonhole_principle.

Shah, K. R. and Sinha, B. K. (2012). *Theory of Optimal Designs*. In: J. Berger, S. Fienberg, J. Gani, K. Krickeberg and B. Singer (Editors), *Lecture Notes in Statistics*, **54**, New York, Springer-Verlag.

Box, G. E. P., Hunter, J. S. and Hunter, W. G. (2005). *Statistics for Experimenters: Design, Innovation and Discovery*. 2nd Edition, Wiley, New York.

Montgomery, D. C. (2013). *Design and Analysis of Experiments*. 8th Edition, Wiley, New York.

APPENDIX

My Answer to the *Exercise* in Section 12

Do not read this Appendix until after you have tried to solve the Exercise.

I will buy four marbles; and I plan to drop the first marble from floors (in order)

$$286, 506, 671, 791, 875, 931, 966, 986, 996, 1000.$$

(To understand where these floor numbers came from, study their successive differences). If the marble does not break at all, then N exceeds 1000; and I will have three unused marbles and three unused coupons. Otherwise, if the marble breaks during any one of the above ten drops, then logic establishes that we need a total of 13 drops to determine N with certainty using the remaining three marbles. Let me illustrate one such situation (leaving all the rest to the reader): Say, the first marble breaks after the 4th drop. Then $671 < N \leq 791$; and the problem reduces to three marbles and 120 floors. In this case, identifying N requires 9 more drops, as explained below (and so a total of $4 + 9 = 13$ drops are needed).

Drop the second marble from floors (in order)

$$707, \quad 735, \quad 756, \quad 771, \quad 781, \quad 787, \quad 790.$$

Say, the second marble breaks after the 3rd drop (all other possibilities are left to the reader). Then $735 < N \leq 756$. So, the problem reduces to two marbles and 21 floors, which requires 6 more drops (which justifies the required 3+6 = 9 drops after the first marble breaks): Drop the third marble from floors $741, 746, 750, 753, 755$. If the third marble breaks during

the second drop, then $741 < N \leq 746$ (again, all other possibilities are left to the reader). So, the problem reduces to one marble and 4 floors, requiring 4 more drops (from floors $742, 743, 744, 745$) and justifying the required 2+4 = 6 drops after the second marble breaks). Thus, in the worst case, the total number of drops of all four marbles is 4+9 = 13, and I must be prepared to spend a maximum total of $4 \times ₹50 + 13 \times ₹10 = ₹330$.

I claim that my choice of buying four marbles, followed by the above strategy of sequentially determining which floor to drop the marbles from, is indeed wise. To justify my claim, let me document what my prospect will be if I buy fewer than four marbles. First, if I buy only one marble, I must be prepared to spend at most ₹10,050 (dropping the marble from floors 1, 2, 3, …). I cannot risk skipping any floor: For if I do and the marble breaks, then N can be any one of the floors I have skipped or the one from which I dropped the marble last. However, in this case, I have no marble left to determine N with certainty! Second, if I buy two marbles, then using the best possible strategy, I may require up to 45 drops (why?). Therefore, I must spend ₹550 in the worst case. Third, if I buy three marbles, then using the best possible strategy, I may have to drop the marbles a total of at most 19 times (why?). Hence, I must spend ₹340 in the worst case. All these options lead to spending more than ₹330, which I agreed to spend to buy four marbles and 13 drops.

What if I buy more than four marbles? If I buy five marbles, then using the best possible strategy, I may need a maximum of 12 drops (why?). So, I must spend ₹370 in the worst case. If I buy six marbles, then using the best possible strategy, I may have to drop the marbles up to 12 times (why?). Hence, I must spend ₹420 in the worst case. Thus, compared to the best strategy using five marbles, the best strategy with a sixth marble does not reduce the number of drops! It was a waste to buy the sixth marble. Buying seven or more marbles will already cost me more than ₹330 even before I buy any elevator coupons! Hence, I recommend buying four marbles and 13 coupons. Can you beat my choice or prove that it is the optimal choice?

The above solution is intricately associated with the relative cost of a marble to a coupon for each elevator ride up. When this relative cost changes, the answer may change. For example, if the cost of each marble decreases to ₹10 but the cost of each coupon remains at ₹10, then I have *two* best choices: Either buy four marbles and 13 coupons; or buy five marbles and 12 coupons. For each choice I will incur a total cost of ₹170. On the other hand, if the cost of each marble increases to ₹100 but the cost of each coupon remains at ₹10, then my best choice is to buy three marbles and 19 coupons incurring a total cost of ₹490. What if the cost of each marble is ₹1000, but the cost of each coupon remains at ₹10? I leave the discovery of the best solution(s) to the reader. In every case, I invite the reader to find a better solution or to prove the optimality of my solution.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 15-29

On the Status of Variance Balanced Block Designs in the Presence of Both-sided Neighbour Effects: Two Examples

Sobita Sapam[1] **and Bikas Kumar Sinha**[2]
[1]*Department of Statistics, Manipur University, Imphal*
[2]*Indian Statistical Institute, Kolkata [Retired Faculty]*

Received: 27 February 2020: Revised: 14 March 2020; Accepted: 18 March 2020

Abstract

The research work presented in this paper is geared towards analysis of variance balanced [VB] block designs in the presence of both-sided neighbour effects. There is a vast literature on VB designs of which the BIBDs are simplest examples. We shall take up two such designs and examine their behaviours in respect of (i) estimates of treatment contrasts, (ii) estimates of block contrasts, and (iii) linear error functions - in the presence of both-sided neighbour-effects. We shall assume a circular model. Estimability issues regarding treatment effects contrasts and block effects contrasts point towards discouraging notes.

Key words: Block designs; Variance balance; Neighbour-effects; Estimability issues; Linear model; ANOVA.

1. Introduction

Variance balance and efficiency balance are two choice-based criteria for selection of designs in many contexts. Block designs, row-column designs and higher dimensional designs have been extensively studied with respect to these two criteria. Combinatorial designs have been characterized utilizing these requirements and this, undoubtedly, forms a fascinating area of research. Some of the works are Hedayat and Stufken (1989), Mishra (2016), Morgan and Uddin (1995), Khatri(1982), Raghavarao (1971), Sinha, Jones and Kageyama (1997).

On the other hand, neighbour-designs, incorporating neighbour-effects, have been studied at length and the concept of balancing has also been introduced. However, though combinatorial balance has been introduced and studied, it seems that there is a gap in this kind of study. From data analysis point of view, no serious attention seems to have been paid for understanding the nature of (i) error functions, (ii) estimable treatment- and block-contrasts, in the presence of NEffects [both Left-sided and Right-sided]. We attempt to fill up this gap. We shall take up two variance-balanced block designs and carefully examine their status with respect to the above - mentioned features.

2. BIBD(7, 7, 3, 3, 1) and Neighbour-Effects

The blocks of the design are obtained by starting with the initial block (1, 2,4) and expanding it, modulo (7). The blocks are (1,2,4); (2,3,5); (3,4,6); (4,5,7); (5,6,1); (6,7,2);

Corresponding Author: Sobita Sapam
Email: sobita1@yahoo.com

(7,1,3). Equivalently, the blocks can be represented in the form of a 7 × 3 matrix as shown in Table 1 below:

Table 1

Block	Col1	Col2	Col3
1	1	2	4
2	2	3	5
3	3	4	6
4	4	5	7
5	5	6	1
6	6	7	2
7	7	1	3

In the absence of any Neighbour Effects [NEs], a complete set of all the eight linearly independent error functions is easily identifiable and each one is shown as difference of two Terms in Table 2 below:

Table 2: Error functions : Term 1 and Term 2

Error Function 1: Term 1(=Column Sum 1) – Term 2(=Column Sum 2)
Error Function 2: Term 1(=Column Sum 1) – Term 2 (=Column Sum 3)
Error Function 3: Term 1 = y(1,1) – y(1,2); Term 2 = [y(7,2) – y(7,3)] + [y(2,2) – y(2,1)]
Error Function 4: Term 1 = y(7,2) – y(7,3); Term 2 = [y(1,1) – y(1,3)] + [y(3,2) – y(3,1)]
Error Function 5: Term 1 = y(1,1) – y(1,3); Term 2 = [y(5,3) – y(5,1)] + [y(4,2) – y(4,1)]
Error Function 6: Term 1 = y(5,3) – y(5,1); Term 2 = [y(7,2) – y(7,3)] + [y(2,2) – y(2,3)]
Error Function7: Term 1 = y(5,3) – y(5,2);Term 2 = [y(7,2) – y(7,3)] + [y(3,1) – y(3,3)]
Error Function 8: Term 1 = y(7,2) – y(7,1); Term 2 = [y(5,3) – y(5,1)] + [y(4,2) – y(4,3)]

It is further verified that the 8 × 21 matrix of the coefficients in these error functions, which is shown below in Table 3, has rank 8.

Remark 1: Thus far we have found out a set of 8 linearly independent error functions which correspond to the error df in the model. This holds under the assumption that there are no neighbour-effects of the plots in the blocks. Below we embark on the problem of examining the status of these error functions in the presence of both the left-and right-sided neighbour effects [LNEs and RNEs].

Table 3: Matrix L of coefficients in linear error functions

EF1	1	−1	0	1	−1	0	1	−1	0	1	−1	0	1	−1	0	1	−1	0	1	−1	0
EF 2	1	0	−1	1	0	−1	1	0	−1	1	0	−1	1	0	−1	1	0	−1	1	0	−1
EF 3	1	−1	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	1
EF 4	−1	0	1	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	1	−1
EF 5	1	0	−1	0	0	0	0	0	0	1	−1	0	1	0	−1	0	0	0	0	0	0
EF 6	0	0	0	0	−1	1	0	0	0	0	0	0	−1	0	1	0	0	0	0	−1	1
EF 7	0	0	0	0	0	0	−1	0	1	0	0	0	0	−1	1	0	0	0	0	−1	1
EF 8	0	0	0	0	0	0	0	0	0	0	−1	1	1	0	−1	0	0	0	−1	1	0

Note: EF – Error Function

2.1. Nature of error functions in the presence of LN Effects and RN Effects

Towards understanding the status of error functions in the presence of LN- and RN-Effects, it is almost immediate to realize that error functions 1 and 2 are free from these effects. For the rest, we need to carry out the exercises. These are shown below in terms of their expectations under the model with NEs :

Error 3= Term1 – Term 2

Term 1 = y(1,1) – y(1,2)
=(τ_1+ LN4 + RN2) − (τ_2+ LN1 + RN4) = $\tau_1 - \tau_2$ + LN4 − LN1 + RN2 − RN4

Term 2 = [y(7,2) – y(7,3)] + [y(2,2) – y(2,1)]
=(τ_1 + LN7 + RN3) − (τ_3 + LN1 + RN7) + (τ_3 + LN2 + RN5) − (τ_2 + LN5 + RN3)
= $\tau_1 - \tau_2$ + (LN7 + LN2 − LN1 − LN5) + (RN3 + RN5 − RN7 − RN3)

Error 3 = Term1 –Term 2 = (LN4 + LN5 – LN7 – LN2) + (RN2 + RN7 – RN4 – RN5)

Error 4 = Term 1 – Term2

Term 1 = y(7,2) – y(7,3)
= (τ_1 + LN7 + RN3) − (τ_3 + LN1 + RN7)
= $\tau_1 - \tau_3$ + LN7 − LN1 + RN3 − RN7

Term 2 = [y(1,1) – y(1,3)] + [y(3,2) – y(3,1)]
= (τ_1 + LN4 + RN2) − (τ_4 + LN2 + RN1) + (τ_4+ LN3 + RN6) − (τ_3+ LN6 + RN4)
= $\tau_1 - \tau_3$ + LN4 − LN2 + LN3 − LN6 + RN2 − RN1 + RN6 − RN4

Error 4 = Term 1 – Term2

= (LN2 + LN6 + LN7 – LN1 – LN3 – LN4) + (RN1 + RN3 + RN4 – RN2 – RN6 – RN7)

Error 5 = Term1 –Term 2

Term 1 = y(1,1) – y(1,3)
= (τ_1 + LN4 + RN2) − (τ_4 + LN2 + RN1)
= $\tau_1 - \tau_4$ + LN4 − LN2 + RN2 − RN1

Term 2 = [y(5,3) – y(5,1)] + [y(4,2) – y(4,1)]
= (τ_1 + LN6 + RN5) − (τ_5 + LN1 + RN6) + (τ_5 + LN4 + RN7) − (τ_4 + LN7 + RN5)
= $\tau_1 - \tau_4$ + (LN6 + LN4) − (LN1 + LN7) + (RN7 − RN6)

Error 5 = Term1 – Term 2 = (LN1 + LN7 – LN2 – LN6) + (RN2 + RN6 – RN1 – RN7)

Error 6 = Term1 – Term 2

Term 1 = y(5,3) – y(5,1)

= (τ_1 + LN6 + RN5) − (τ_5 + LN1 + RN6) = $\tau_1 - \tau_5$ + LN6 − LN1 + RN5 − RN6

Term 2 = [y(7,2) – y(7,3)] + [y(2,2) – y(2,3)]
= (τ_1 + LN7 + RN3) − (τ_3 + LN1 + RN7) + (τ_3 + LN2 + RN5) − (τ_5 + LN3 + RN2)
= $\tau_1 - \tau_5$ + LN7 + LN2 − LN1 − LN3 + RN3 + RN5 − RN2 − RN7

Error 6 = Term1 – Term 2 = (LN3 + LN6 – LN2 – LN7) + (RN2 + RN7 – RN3 – RN6)

Error 7 = Term1 – Term 2

Term 1 = y(5,3) – y(5,2)
= (τ_1 + LN6 + RN5) − (τ_6 + LN5 + RN1)
= $\tau_1 - \tau_6$ + LN6 − LN5 + RN5 − RN1
Term 2 = [y(7,2) – y(7,3)] + [y(3,1) – y(3,3)]

$= (\tau_1 + LN7 + RN3) - (\tau_3 + LN1 + RN7) + (\tau_3 + LN6 + RN4) - (\tau_6 + LN4 + RN3)$

$= \tau_1 - \tau_6 + LN7 + LN6 - LN1 - LN4 + RN3 + RN4 - RN7 - RN3$

Error 7 = Term1 – Term 2 = (LN1 + LN4 – LN5 – LN7) + (RN5 + RN7 – RN1 – RN4)

Error 8 = Term1 – Term 2

Term 1 = y(7,2) – y(7,1)

$= (\tau_1 + LN7 + RN3) - (\tau_7 + LN3 + RN1) = \tau_1 - \tau_7 + LN7 - LN3 + RN3 - RN1$

Term 2 = [y(5,3) – y(5,1)] + [y(4,2) – y(4,3)]

$= (\tau_1 + LN6 + RN5) - (\tau_5 + LN1 + RN6) + (\tau_5 + LN4 + RN7) - (\tau_7 + LN5 + RN4)$

$= \tau_1 - \tau_7 + LN6 - LN1 + LN4 - LN5 + RN5 - RN6 + RN7 - RN4)$

Error 8 = Term1 – Term 2

= (LN1 + LN5 + LN7 – LN3 – LN4 – LN6) + (RN3 + RN4 + RN6 – RN1 – RN5 – RN7)

Now we consider the LN effects only and develop all these 6 equations, that is, from Error 3 to Error 8 into a matrix. Further, we append the row vector $(1,1,...,1)'$ to make it a square matrix of order 7. This is shown in Table 4 below:

Table 4: Coefficients of LN Effects in expectations of observational contrasts for errors

Error Sl. No.	Co-efficient of LNE						
Special row	1	1	1	1	1	1	1
3	0	–1	0	1	1	0	–1
4	–1	1	–1	–1	0	1	1
5	1	–1	0	0	0	–1	1
6	0	–1	1	0	0	1	–1
7	1	0	0	1	–1	0	–1
8	1	0	–1	–1	1	–1	1

It follows that this matrix is of full rank. Moreover, the matrix underlying R-sided NEs is obtainable from the above by simply changing the signs of elements in each row. Consequently, error df will remain intact at 8 df if and only if all the Left-sided NEs are equal and at the same time all the Right-sided NEs are also equal. Unless this is satisfied, we cannot go for the usual ANOVA Table-based data analysis. Once this is satisfied, we see no effect whatsoever of these LN and RN effects on the analysis of data. Further to this, we also find that there are 2 df for error - no matter what happens to the NEs. Thus ANOVA F-tests can be carried out for estimable treatment contrasts and estimable block contrasts - even in the presence of NEs - provided such estimable treatment/block contrasts are available.

2.2. Estimable treatment contrasts in the BIBD with NEffects

We list below in Table 5 simple-minded unbiased estimators of a set of elementary treatment contrasts based on the observations underlying the BIBD - in the absence of LNEs and RNEs. For later use, we have also indicated model expectations of these observational contrasts in the table - assuming the presence of NEs - both Left-sided and Right-sided. Suppose now that Error 3 is a valid error *i.e.*, E[Error3] = 0. That means

LN4 + LN5 = LN2 + LN7.....(LNC1)

and, at the same time,

RN4 + RN5 = RN2 + RN7.....(RNC1)

We now demonstrate that under the condition (LNC1), there is an observational contrast whose expectation is free from LNEs and at the same time, it involves one treatment contrast. We re-write (LNC1) as :

Table 5: Expectations of observational contrasts in terms of treatment contrasts

E[y(1,1) − y(1,2)] = $\tau_1 - \tau_2$+ LN4 − LN1 + RN2 − RN4
E[y(7,2) − y(7,3)] = $\tau_1 - \tau_3$ + LN7 − LN1 + RN3 − RN7
E[y(1,1) − y(1,3)] = $\tau_1 - \tau_4$ + LN4 − LN2 + RN2 − RN1
E[y(5,3) − y(5,1)] = $\tau_1 - \tau_5$+ LN6 − LN1 + RN5 − RN6
E[y(5,3) − y(5,2)] = $\tau_1 - \tau_6$ + LN6 − LN5 + RN5 − RN1
E[y(7,2) − y(7,1)] = $\tau_1 - \tau_7$ + LN7 − LN3 + RN3 − RN1

LN4 – LN2 = LN7 – LN5

or, LN4 – LN2 = (LN7 – LN1) + (LN1 – LN6) + (LN6 – LN5).

We now examine both sides of the expression, which are expressed in terms of LNEffects contrasts. We refer to the Table 5 of treatment contrasts. This yields:

$(\tau_1 - \tau_4) = (\tau_1 - \tau_3) + (\tau_5 - \tau_1) + (\tau_1 - \tau_6)$.

This leads to

$(\tau_3 - \tau_4 - \tau_5 + \tau_6)$.....(TC1).

The message is clear. Under the condition that Error 3 is a valid error, there is an observational contrast, *viz*.,

[y(1,1) − y(1,3)] − [y(7,2) − y(7,3)] + [y(5,3) − y(5,1)] − [y(5,3) − y(5,2)]

whose expectation is free from LNEs and, moreover, it involves the treatment contrast (TC1). We have yet to verify the status of this observational contrast in the presence of the RN Effects. It is easy to check that the RN Effects contrast [RN2 + RN7 – RN3 – RN6] remains present along with the treatment contrast. The condition (RNC1) is different from this and hence, E[Error 3] = 0 alone does not provide any positive result towards estimability of any treatment contrast. The condition

RN2 + RN7 = RN3 + RN6.....(RNC2)

is also needed. That means: (LNC1), (RNC1) and (RNC2) together ensure estimability of (TC1) along with existence of a valid error *viz*., Error 3 [This holds, besides the errors: Error 1 and Error 2].

We now analyse (RNC2) and readily observe that E[Error 6] = 0 whenever (RNC2) holds in addition to LN2 + LN7 = LN3 + LN6.....(LNC2).

Further to this,

$\tau_3 + \tau_6 - \tau_2 - \tau_7$.....(TC2)

becomes estimable and an unbiased estimator is given by

[y(1,1) − y(1,2)] + [y(7,3) − y(7,1)] + [y(5,2) − y(5,3)].

Combining the results, we have the following:

Based on the assumptions (LNC1), (LNC2), (RNC1) and (RNC2), there are two valid errors [Error 3 and Error 6] and also there are two estimable treatment contrasts (TC1) and (TC2).

Likewise, we made an attempt to identify a pair of error functions, which together would produce similar result on a different pair of treatment contrasts. However, we are partially successful. Assuming *E*[Error 4] = *E*[Error 7] = 0, we end up with the conditions:

(a) LN2 + LN6 = LN3 + LN5 and RN2 + RN6 = RN3 + RN5,

(b) LN1 + LN4 = LN5 + LN7 and RN1 + RN4 = RN5 + RN7.

In view of (a) and (b), it turns out that

E[[y(7,2) − y(7,3)] − [y(1,1) − y(1,2)] + [y(5,3) − y(5,1)] − [y(5,3) − y(5,2)]]

$$= \tau_2 + \tau_6 - \tau_3 - \tau_5.$$

We failed to find out another estimable treatment contrast based on (a) and (b). At this stage, we did not make any further attempt with the last two error functions *viz*., Error 5 and Error 8.

Remark 2: It is interesting to observe that the errors [Error 3 to Error 8]remain as valid errors even in the presence of RNEs and LNEs provided RNE of every treatment is the same as the corresponding LNE. However, this does not ensure estimability of treatment contrasts / block contrasts without further unusual conditions, as indicated above [for treatment contrasts].

2.3. Estimable block contrasts in the BIBD with NEffects

It is well-known that a treatment-connected block design is also automatically block-connected. However, when the NEs are present, we have to analyse the block contrasts separately. At first, we display in Table 6 elementary block contrasts and their simple-minded estimates under the assumption of absence of LNEs and RNEs. These are shown in columns 1 and 2. Further, assuming that the LNEs and RNEs are present, we show in columns 3 and 4 of the same table their effects on the chosen observational contrasts. We start with Error 3 which is a valid error whenever (LNC1) and (RNC1) both hold simultaneously. Upon re-writing (LNC1) as: LN4 − LN7 = LN2 − LN5, we find that the LHS corresponds to $\beta_1 - \beta_7$. To find a 'matching' for the RHS, we re-write it as (LN2 − LN7) + (LN7 − LN1) + (LN1 − LN5) which is again expressed as $(\beta_1 - \beta_4) + (\beta_6 - \beta_1) + (\beta_1 - \beta_2)$ and this simplifies to $(\beta_1 + \beta_6 - \beta_2 - \beta_4)$. Therefore, combining the two, we infer that for the block contrast given by $\beta_2 + \beta_4 - \beta_6 - \beta_7$, there is an observational contrast *viz*.,
[y(1,1) − y(7,2)] − [y(1,3) − y(4,1)] + y(1,2) − y(6,3)] − [y(1,2) − y(2,1)]

Table 6: Expectations of observational contrasts in terms of block contrasts

Block contrast	Observational contrasts	LNE(+)(−)	RNE(+)(−)
$\beta_1 - \beta_2$	y(1,2) − y(2,1)	1,5	4,3
$\beta_1 - \beta_3$	y(1,3) − y(3,2)	2,3	1,6

$\beta_1 - \beta_4$	y(1,3) − y(4,1)	2,7	1,5
$\beta_1 - \beta_5$	y(1,1) − y(5,3)	4,6	2,5
$\beta_1 - \beta_6$	y(1,2) − y(6,3)	1,7	4,6
$\beta_1 - \beta_7$	y(1,1) − y(7,2)	4,7	2,3

which eliminates the LN Effects in expectation. This is so far as LN Effects are concerned. We now examine the nature of involvement of RN Effects. An analysis similar to the case of estimation of treatment contrasts suggests:

Under the conditions required for the validity of Error 3 *i.e.*, under (LNC1) and (RNC1):

There is a block contrast *viz.*, $\beta_2 + \beta_4 - \beta_6 - \beta_7$, which is estimable provided further that

RN1 − RN6 = RN2 − RN5.....(RNC3)

holds. When this holds, another condition

LN2 − LN4 = LN3 − LN6.....(LNC3)

ensures estimation of a second block contrast *viz.*, $\beta_3 - \beta_5$ and an estimator is given by [y(1,1) − y(5,3)] − [y(1,3) − y(3,2)]. These are uninteresting conditions on the LN Effects and RN Effects. We do not pursue the matter anymore. To summarize, in the presence of LN Effects and RN Effects, only under certain conditions [like (LNC1) and (RNC1)], we can identify error function(s) on the top of the basic two errors [Error 1 and Error 2]. However, we need further conditions on the LN Effects and RN Effects to provide estimable treatment contrasts and estimable block contrasts - that too - only 1 or 2. There is no substantial promise for the analysis of VB Designs (also possibly for EB Designs) - in the presence of LN Effects and RN Effects.

3. VB Design With Unequal Replications

Mishra and Sarvate (2019, Private Communication) studied a block design which is variance-balanced but based on unequal replication numbers. The blocks of the design are given as:

B1 = (1,1,2,3); B2 = (1,1,4,5); B3 = (1,1,6,7); B4 = (1,2,4,6); B5 = (1,3,5,7); B6 = (2,3,5,6); B7 = (2,3,4,7); B8 = (2,4,5,7); B9 = (2,5,6,7); B10 = (3,4,5,6); B11 = (3,4,6,7).

Here, the total df = 44−1= 43 is decomposed as: 10 df for blocks, 6 df for treatments and 27 df for error. Below in Table 7 we present a full set of 27 linearly independent error functions. Errors 1 to 3 are readily derived as observational contrasts within blocks whereas Errors 4 to 27 are identified as differences of two terms based on observational contrasts.

Table 7: Error functions: Term 1 and Term 2

Error Function 1 : [y(1,1) − y(1,2)]
Error Function 2 : [y(2,1) − y(2,2)]
Error Function 3 : [y(3,1) − y(3,2)]

Error Function 4 : Term 1 = [y(1,2) − y(1,3)]; Term 2 = [y(4,1) − y(4,2)]
Error Function 5 : Term 1 = [y(1,2) − y(1,4)]; Term 2 = [y(5,1) − y(5,2)]
Error Function 6 : Term 1 = [y(2,2) − y(2,3)]; Term 2 = [y(4,1) − y(4,3)]
Error Function 7 : Term 1 = [y(2,1) − y(2,4)]; Term 2 = [y(5,1) − y(5,3)]
Error Function 8 : Term 1 = [y(3,2) − y(3,3)]; Term 2 = [y(4,1) − y(4,4)]
Error Function 9 : Term 1 = [y(3,1) − y(3,4)]; Term 2 = [y(5,1) − y(5,4)]
Error Function 10 : Term 1 = [y(1,3) − y(1,4)]; Term 2 = [y(6,1) − y(6,2)]
Error Function 11 : Term 1 = [y(6,1) − y(6,2)]; Term 2 = [y(7,1) − y(7,2)]
Error Function 12 : Term 1 = [y(4,2) − y(4,3)]; Term 2 = [y(7,1) − y(7,3)]
Error Function 13 : Term 1 = [y(4,2) − y(4,3)]; Term 2 = [y(8,1) − y(8,2)]
Error Function 14 : Term 1 = [y(6,1) − y(6,3)]; Term 2 = [y(8,1) − y(8,3)]
Error Function 15 : Term 1 = [y(6,1) − y(6,3)]; Term 2 = [y(9,1) − y(9,2)]
Error Function 16 : Term 1 = [y(4,2) − y(4,4)]; Term 2 = [y(6,1) − y(6,4)]
Error Function 17 : Term 1 = [y(4,2) − y(4,4)]; Term 2 = [y(9,1) − y(9,3)]
Error Function 18 : Term 1 = [y(7,1) − y(7,4)]; Term 2 = [y(8,1) − y(8,4)]
Error Function 19 : Term 1 = [y(7,1) − y(7,4)]; Term 2 = [y(9,1) − y(9,4)]
Error Function 20: Term 1 = [y(7,2) − y(7,3)]; Term 2 = [y(10,1) − y(10,2)]
Error Function 21: Term 1 = [y(7,2) − y(7,3)]; Term 2 = [y(11,1) − y(11,2)]
Error Function 22: Term 1 = [y(5,2) − y(5,3)]; Term 2 = [y(6,2) − y(6,3)]
Error Function 23: Term 1= [y(5,2) − y(5,3)]; Term 2=[y(10,1) − y(10,3)]
Error Function 24: Term 1 = [y(6,2) − y(6,4)]; Term 2 = [y(10,1) − y(10,4)]
Error Function 25: Term 1 = [y(6,2) − y(6,4)]; Term 2 = [y(11,1) − y(11,3)]
Error Function 26 : Term 1 = [y(7,2) − y(7,4)]; Term 2= [y(5,2) − y(5,4)]
Error Function 27 : Term 1 = [y(7,2) − y(7,4)]; Term 2 = [y(11,1) − y(11,4)]

It is further verified that the 27×44 matrix of the coefficients of error functions has rank 27. This holds under the assumption that there are no NEffects.

3.1. Nature of error functions in the presence of LN and RN Effects

Now we work out expectations of all the 27 Error functions in the presence of left-and right sided neighbor effects in the plots of the blocks, under the assumption that the blocks are circular.

Error 1 = [y(1,1) − y(1,2)]

E[Error 1] = τ_1+ LN3 + RN1 − τ_1 − LN1 − RN2 = LN3 − LN1 + RN1 − RN2

Error 2 = [y(2,1) − y(2,2)]

E[Error 2] = τ_1 + LN5 + RN1 − τ_1 − LN1 − RN4 = LN5 − LN1 + RN1 − RN4

Error 3 = [y(3,1) − y(3,2)]

E[Error 3] = τ_1 + LN7 + RN1 − τ_1 − LN1 − RN6 = LN7 − LN1 + RN1 − RN6

Error 4 = Term 1 − Term 2

Term 1 = [y(1,2) − y(1,3)]

E[Term 1] = τ_1 + LN1 + RN2 − τ_2− LN1 − RN3 = τ_1 − τ_2 + RN2 − RN3

Term 2 = [y(4,1) − y(4,2)]

E[Term 2] = τ_1 + LN6 + RN2 − τ_2 − LN1 − RN4 = τ_1 − τ_2 + LN6 − LN1 + RN2 − RN4

E[Error 4] = LN1 − LN6 + RN4 − RN3

Error 5 = Term 1 − Term 2

Term 1 = [y(1,2) − y(1,4)]

E[Term 1] = τ_1+ LN1 + RN2 − τ_3 − LN2 − RN1 = τ_1 − τ_3 + LN1 − LN2 + RN2 − RN1

Term 2 = [y(5,1) − y(5,2)]

E[Term 2] = τ_1 + LN7 + RN3 − τ_3 − LN1 − RN5 = τ_1 − τ_3 + LN7 − LN1 + RN3 − RN5

E[Error 5] = 2LN1 − LN2 − LN7 + RN2 + RN5 − RN1 − RN3

Error 6 = Term 1 − Term 2

Term 1 = [y(2,2) − y(2,3)]

E[Term 1] = τ_1 + LN1 + RN4 − τ_4 − LN1 − RN5 = τ_1− τ_4 + RN4 − RN5

Term 2 = [y(4,1) − y(4,3)]

E[Term 2] = τ_1 + LN6 + RN2 −τ_4− LN2 − RN6 = τ_1 − τ_4+ LN6 − LN2 + RN2 − RN6

E[Error 6] = LN2 − LN6 + RN4 + RN6 − RN2 − RN5

Error 7 = Term 1− Term 2

Term 1 = [y(2,1) − y(2,4)]

E[Term 1] = τ_1 + LN5 + RN1 − τ_5 − LN4 − RN1 = τ_1 − τ_5 + LN5 − LN4

Term 2 = [y(5,1) − y(5,3)]

E[Term 2] = τ_1 + LN7 + RN3 − τ_5− LN3 − RN7 = τ_1 − τ_5 + LN7 − LN3 + RN3 − RN7

E[Error 7] = LN5 + LN3 − LN4 − LN7 + RN7 − RN3

Error 8 = Term 1 − Term 2

Term 1 = [y(3,2) − y(3,3)]

E[Term 1] = τ_1 + LN1 + RN6 − τ_6 − LN1 − RN7 = τ_1− τ_6 + RN6 − RN7

Term 2 = [y(4,1) − y(4,4)]

E[Term 2] = τ_1+ LN6 + RN2 − τ_6− LN4 − RN1 = τ_1− τ_6 + LN6 − LN4 + RN2 − RN1

E[Error 8] = LN4 − LN6 + RN6 + RN1 − RN2 − RN7

Error 9 = Term 1− Term 2

Term 1 = [y(3,1) − y(3,4)]

E[Term 1] = τ_1 + LN7 + RN1 − τ_7 − LN6 − RN1 = τ_1 − τ_7 + LN7 − LN6

Term 2 = [y(5,1) − y(5,4)]

E[Term 2] = τ_1 + LN7 + RN3 −τ_7− LN5 − RN1 = τ_1 − τ_7 + LN7 − LN5 + RN3 − RN1

E[Error 9] = LN5 − LN6 + RN1 − RN3

Error 10 = Term 1 − Term 2

Term 1 = [y(1,3) − y(1,4)]

E[Term 1] = τ_2+ LN1 + RN3 − τ_3 − LN2 − RN1 = τ_2 − τ_3 + LN1 − LN2 + RN3 − RN1

Term 2 = [y(6,1) − y(6,2)]

E[Term 2] = τ_2 + LN6 + RN3 − τ_3 − LN2 − RN5 = τ_2 − τ_3 + LN6 − LN2 + RN3 − RN5

E[Error 10] = LN1 − LN6 + RN5 − RN1

Error 11 = Term 1− Term 2

Term 1 = [y(6,1) − y(6,2)]

E[Term 1] = τ_2 + LN6 + RN3 − τ_3 − LN2 − RN5 = τ_2 − τ_3 + LN6 − LN2 + RN3 − RN5

Term 2 = [y(7,1) − y(7,2)]

E[Term 2] = τ_2 + LN7 + RN3 − τ_3 − LN2 − RN4 = τ_2 − τ_3 + LN7 − LN2 + RN3 − RN4

E[Error 11] = LN6 − LN7 + RN4 − RN5

Error 12 = Term 1 − Term 2

Term 1 = [y(4,2) − y(4,3)]

E[Term 1] = τ_2 + LN1 + RN4 − τ_4 − LN2 − RN6 = τ_2 − τ_4 + LN1 − LN2 + RN4 − RN6

Term 2 = [y(7,1) − y(7,3)]

E[Term 2] = τ_2 + LN7 + RN3 − τ_4 − LN3 − RN7 = τ_2 − τ_4 + LN7 − LN3 + RN3 − RN7

E[Error 12] = LN1 + LN3 − LN2 − LN7 + RN4 + RN7 − RN3 − RN6

Error 13 = Term 1 − Term 2

Term 1 = [y(4,2) − y(4,3)]

E[Term 1] = τ_2 + LN1 + RN4 − τ_4 − LN2 − RN6 = τ_2 − τ_4 + LN1 − LN2 + RN4 − RN6

Term 2 = [y(8,1) − y(8,2)]

E[Term 2] = τ_2 + LN7 + RN4 − τ_4− LN2 − RN5 = τ_2 − τ_4 + LN7 − LN2 + RN4 − RN5

E[Error 13] = LN1 − LN7 + RN5 − RN6

Error 14 = Term 1 − Term 2

Term 1 = [y(6,1) − y(6,3)]

E[Term 1] = τ_2 + LN6 + RN3 − τ_5 − LN3 − RN6 = τ_2− τ_5 + LN6 − LN3 + RN3 − RN6

Term 2 = [y(8,1) − y(8,3)]

E[Term 2] = τ_2 + LN7 + RN4 − τ_5 − LN4 − RN7 = τ_2 − τ_5 + LN7 − LN4 + RN4 − RN7

E[Error 14] = LN6 + LN4 − LN3 − LN7 + RN3 + RN7 − RN4 − RN6

Error 15 = Term 1− Term 2

Term 1 = [y(6,1) − y(6,3)]

E[Term 1] = τ_2 + LN6 + RN3 − τ_5− LN3 − RN6 = τ_2− τ_5 + LN6 − LN3 + RN3 − RN6

Term 2 = [y(9,1) − y(9,2)]

E[Term 2] = τ_2 + LN7 + RN5 − τ_5 − LN2 − RN6 = τ_2 − τ_5 + LN7 − LN2 + RN5 − RN6

E[Error 15] = LN6 + LN2 − LN3 − LN7 + RN3 − RN5

Error 16 = Term 1 − Term 2

Term 1 = [y(4,2) − y(4,4)]

E[Term 1] = τ_2 + LN1 + RN4 − τ_6− LN4 − RN1 = τ_2 − τ_6 + LN1 − LN4 + RN4 − RN1

Term 2 = [y(6,1) − y(6,4)]

E[Term 2] = τ_2 + LN6 + RN3 − τ_6 − LN5 − RN2 = τ_2 − τ_6+ LN6 − LN5 + RN3 − RN2

E[Error 16] = LN1 + LN5 − LN4 − LN6 + RN4 + RN2 − RN1 − RN3

Error 17 = Term 1 − Term 2

Term 1 = [y(4,2) − y(4,4)]

E[Term 1] = τ_2+ LN1 + RN4 − τ_6 − LN4 − RN1 = τ_2− τ_6 + LN1 − LN4 + RN4 − RN1

Term 2 = [y(9,1) − y(9,3)]

E[Term 2] = τ_2 + LN7 + RN5 − τ_6 − LN5 − RN7 = τ_2− τ_6+ LN7 − LN5 + RN5 − RN7

E[Error 17] = LN1 + LN5 − LN4 − LN7 + RN4 + RN7 − RN1 − RN5

Error 18 = Term 1 − Term 2

Term 1 = [y(7,1) − y(7,4)]

E[Term 1] = τ_2 + LN7 + RN3 − τ_7− LN4 − RN2 = τ_2 − τ_7 + LN7 − LN4 + RN3 − RN2

Term 2 = [y(8,1) − y(8,4)]

E[Term 2] = τ_2 + LN7 + RN4 − τ_7 − LN5 − RN2 = τ_2 − τ_7 + LN7 − LN5 + RN4 − RN2

E[Error 18] = LN5 − LN4 − RN3 − RN4

Error 19 = Term 1 − Term 2

Term 1 = [y(7,1) − y(7,4)]

E[Term 1] = τ_2 + LN7 + RN3 − τ_7 − LN4 − RN2 = τ_2 − τ_7 + LN7 − LN4 + RN3 − RN2

Term 2 = [y(9,1) − y(9,4)]

E[Term 2] = τ_2 + LN7 + RN5 − τ_7 − LN6 − RN2 = τ_2− τ_7 + LN7 − LN6 + RN5 − RN2

E[Error 19] = LN6 − LN4 + RN3 − RN5

Error 20 = Term 1 − Term 2

Term 1 = [y(7,2) − y(7,3)]

E[Term 1] = τ_3 + LN2 + RN4 − τ_4 − LN3 − RN7 = τ_3 − τ_4+ LN2 − LN3 + RN4 − RN7

Term 2 = [y(10,1) − y(10,2)]

E[Term 2] = τ_3 + LN6 + RN4 − τ_4 − LN3 − RN5 = τ_3− τ_4 + LN6 − LN3 + RN4 − RN5

E[Error 20] = LN2 − LN6 + RN5 − RN7

Error 21 = Term 1 − Term 2

Term 1 = [y(7,2) − y(7,3)]

E[Term 1] = τ_3+ LN2 + RN4 − τ_4 − LN3 − RN7 = τ_3 − τ_4 + LN2 − LN3 + RN4 − RN7

Term 2 = [y(11,1) − y(11,2)]

E[Term 2] = τ_3+ LN7 + RN4 − τ_4 − LN3 − RN6 = τ_3 − τ_4 + LN7 − LN3 + RN4 − RN6

E[Error 21] = LN2 − LN7 + RN6 − RN7

Error 22 = Term 1− Term 2

Term 1 = [y(5,2) − y(5,3)]

E[Term 1] = τ_3 + LN1 + RN5 − τ_5 − LN3 − RN7 = τ_3 − τ_5+ LN1 − LN3 + RN5 − RN7

Term 2 = [y(6,2) − y(6,3)]

E[Term 2] = τ_3 + LN2 + RN5 − τ_5 − LN3 − RN6 = τ_3− τ_5 + LN2 − LN3 + RN5 − RN6

E[Error 22] = LN1 − LN2 + RN6 − RN7

Error 23 = Term 1 − Term 2

Term 1 = [y(5,2) − y(5,3)]

E[Term 1] = τ_3 + LN1 + RN5 − τ_5 − LN3 − RN7 = τ_3− τ_5 + LN1 − LN3 + RN5 − RN7

Term 2 = [y(10,1) − y(10,3)]

E[Term 2] = τ_3 + LN6 + RN4 − τ_5 − LN4 − RN6 = τ_3− τ_5 + LN6 − LN4 + RN4 − RN6

E[Error 23] = LN1 + LN4 − LN3 − LN6 + RN5 + RN6 − RN4 − RN7

Error 24 = Term 1 − Term 2

Term 1 = [y(6,2) − y(6,4)]

E[Term 1] = τ_3+ LN2 + RN5 − τ_6 − LN5 − RN7 = τ_3− τ_6 + LN2 − LN5 + RN5 − RN2

Term 2 = [y(10,1) − y(10,4)]

E[Term 2] = τ_3+ LN6 + RN4 − τ_6 − LN5 − RN3 = τ_3− τ_6+ LN6 − LN5 + RN4 − RN3

E[Error 24] = LN2 − LN6 + RN5 + RN3 − RN2 − RN4

Error 25 = Term 1 − Term 2

Term 1 = [y(6,2) − y(6,4)]

E[Term 1] = τ_3+ LN2 + RN5 − τ_6 − LN5 − RN2 = τ_3− τ_6 + LN2 − LN5 + RN5 − RN2

Term 2 = [y(11,1) − y(11,3)]

E[Term 2] = τ_3+ LN7 + RN4 − τ_6 − LN4 − RN7 = τ_3− τ_6+ LN7 − LN4 + RN4 − RN7

E[Error 25] = LN2 + LN4 − LN5 − LN7 + RN5 + RN7 − RN2 − RN4

Error 26 = Term 1 − Term 2

Term 1 = [y(7,2) − y(7,4)]

E[Term 1] = τ_3+ LN2 + RN4 − τ_7 − LN4 − RN2 = τ_3− τ_7+ LN2 − LN4 + RN4 − RN2

Term 2 = [y(5,2) − y(5,4)]

E[Term 2] = τ_3+ LN1 + RN5 − τ_7 − LN5 − RN1 = τ_3− τ_7 + LN1 − LN5 + RN5 − RN1

E[Error 26] = LN2 + LN5 − LN1 − LN4 + RN4 + RN1− RN2 − RN5

Error 27 = Term 1− Term 2

Term 1 = [y(7,2) − y(7,4)]

E[Term 1] = τ_3 + LN2 + RN4 − τ_7 − LN4 − RN2 = τ_3− τ_7 + LN2 − LN4 + RN4 − RN2

Term 2 = [y(11,1) − y(11,4)]

E[Term 2] = τ_3 + LN7 + RN4 − τ_7 − LN6 − RN3 = τ_3−τ_7+ LN7 − LN6 + RN4 − RN3

E[Error 27] = LN2 + LN6 − LN4 − LN7 + RN3 − RN2

For further analysis, in the above, we consider only the LN Effects and develop the underlying matrix. This is shown in Table 8 below.

Table 8: Coefficients of LN Effects in expectations of observational contrasts for errors

Error Sl. No.							
1	−1	0	1	0	0	0	0
2	−1	0	0	0	1	0	0
3	−1	0	0	0	0	0	1
4	1	0	0	0	0	−1	0
5	1	−1	1	0	0	0	−1
6	0	1	0	0	0	−1	0
7	0	0	1	−1	1	0	−1
8	0	0	0	1	0	−1	0
9	0	0	0	0	1	−1	0
10	1	0	0	0	0	−1	0
11	0	0	0	0	0	1	−1
12	1	−1	1	0	0	0	−1
13	1	0	0	0	0	0	−1
14	0	0	−1	1	0	1	−1
15	0	1	−1	0	0	1	−1
16	1	0	0	−1	1	−1	0
17	1	0	0	−1	1	0	−1
18	0	0	0	−1	1	0	0
19	0	0	0	−1	0	1	0
20	0	1	0	0	0	−1	0
21	0	1	0	0	0	0	−1
22	1	−1	0	0	0	0	0
23	1	0	−1	1	0	−1	0
24	0	1	0	0	0	−1	0
25	0	1	0	1	−1	0	−1
26	−1	1	0	−1	1	0	0
27	0	1	0	−1	0	1	−1

Similarly, we may consider the RN effects only and develop the corresponding matrix of coefficients. This is shown in the Table 9 below.

3.2. Estimability of treatment contrasts in the presence of both-sided neighbor effects

Suppose Error 1 is a valid error, that is, *E*[Error 1] = 0. Then LN3 − LN1 = 0. Now we can rewrite it as (LN3 − LN7) + (LN7 − LN1) = 0. Using the Table 10 on expectations of some treatment contrasts, it yields $\tau_5 - \tau_1 + \tau_1 - \tau_3 = 0$. This leads to estimability of the treatment contrast $(\tau_3 - \tau_5)$, provided there are no RN Effects. For this we need the condition

Table 9: Coefficients of RN Effects in expectations of observational contrasts for errors

Error Sl. No.							
1	1	−1	0	0	0	0	0
2	1	0	0	−1	0	0	0
3	1	0	0	0	0	−1	0
4	0	0	−1	1	0	0	0
5	0	0	−1	0	1	0	0
6	0	−1	0	1	−1	1	0
7	0	0	−1	0	0	0	1
8	1	−1	0	0	0	1	−1
9	1	0	−1	0	0	0	0
10	−1	0	0	0	1	0	0
11	0	0	0	1	−1	0	0
12	0	0	−1	1	0	−1	1
13	0	0	0	0	1	−1	0
14	0	0	1	−1	0	−1	1
15	0	0	1	0	−1	0	0
16	−1	1	−1	1	0	0	0
17	−1	0	0	1	−1	0	1
18	0	0	1	−1	0	0	0
19	0	0	1	0	0	−1	0
20	0	0	0	0	1	0	−1
21	0	0	0	0	0	1	−1
22	0	0	0	0	0	1	−1
23	0	0	0	−1	1	1	−1
24	0	−1	1	−1	1	0	0
25	0	−1	0	−1	1	0	1
26	1	−1	0	1	−1	0	0
27	0	−1	1	0	0	0	0

RN5 = RN7. When this happens, we identify Error 20 as a valid error when, in addition, we also require: LN2 = LN6. Under this LN-related condition, we find that $\tau_1 - \tau_4$ becomes estimable if, again, we have: RN2 = RN6. The conditions are not encouraging at all. We stop the analysis and observe that presence of LN and RN Effects really poses estimability problem for treatment effects contrasts and similarly, for block effects contrasts as well.

Table 10: Expectations of observational contrasts in terms of basic treatment contrasts

E[y(4,1) − y(4,2)] = $\tau_1 - \tau_2$ + LN6 – LN1 + RN2 – RN4
E[y(5,1) − y(5,2)] = $\tau_1 - \tau_3$ + LN7 − LN1 + RN3 − RN5
E[y(4,1) − y(4,3)] = $\tau_1 - \tau_4$ + LN6 − LN2 + RN2 − RN6
E[y(5,1) − y(5,3)] = $\tau_1 - \tau_5$ + LN7 − LN3 + RN3 − RN7
E[y(4,1) − y(4,4)] = $\tau_1 - \tau_6$ + LN6 − LN4 + RN2 − RN1
E[y(5,1) − y(5,4)] = $\tau_1 - \tau_7$ + LN7 − LN5 + RN3 − RN1

4. Concluding Remarks

These simple examples illustrate the difficulties in carrying out ANOVA Tests in the presence of LNEs and / or RNEs. We should be careful in handling the data analyses issues.

Acknowledgements

The first author thanks her Mentor Professor K.K. Singh Meitei for providing all facilities towards successfully pursuing her research in the broad area of Design of Experiments and also for arranging academic visits of Professor Bikas Sinha to the Manipur University for collaborative research.

Funding

The first author acknowledges financial support from **DST Women Scientist Scheme - A, Project Sanction order No. SR / WOS-A / PM-98 / 2017(G).**

References

Hedayat, A. and Stufken, J. (1989). A relation between pairwise balanced and variance balanced block designs. *Journal of the American Statistical Association*, **84**(**407**), 753-755.

Mishra, N. (2016). Non-binary variance balanced designs - part I optimality. *Journal of Statistical Theory and Practice*,**11**,63-75.

Morgan, J. P. and Uddin, N. (1995). Optimal non-binary, variance balanced designs. *Statistica Sinica*, **5**, 535-546.

Khatri, C. G. (1982). A note on variance balanced designs. *Journal of Statistical Planning and Inference*, **6**, 173-177.

Raghavarao, D. (1971). *Construction and Combinatorial Problems in Design of Experiments*. John Wiley, New York.

Sinha, K., Jones, B. and Kageyama, S. (1997). Constructions of pairwise-and variance-balanced designs. *A Journal of Theoretical and Applied Statistics*, **29**, 241-250.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 31–34

Simple Proof of the Non-existence of Some Affine Resolvable SRGD Designs

Satoru Kadowaki[1] **and Sanpei Kageyama**[2]
[1]*Department of Science, National Institute of Technology, Matsue College, Matsue, Japan*
[2]*Hiroshima University, Higashi-Hiroshima, Japan*

Received: 21 March 2020; Revised: 04 April 2020; Accepted: 09 April 2020

Abstract

The existence of resolvable and affine resolvable block designs has been discussed in the literature (cf. Clatworthy (1973), Raghavarao (1988)). Recently, a necessary condition for the existence of a certain resolvable pairwise balanced design is provided by Kadowaki and Kageyama (2020). In this paper, through the necessary condition, we can provide a simple proof of the non-existence result of some affine resolvable SRGD design, rather than by the usual methods in combinatorics.

Key words: Affine resolvability; Resolvability; PB design;SRGD design.

1. Introduction

A block design BD(v, b, r, k) with v treatments is said to be resolvable if the b blocks of size k each can be grouped into r resolution sets of b/r blocks each such that in each resolution set every treatment occurs exactly once. A resolvable BD is said to be affine resolvable if every two blocks belonging to different resolution sets intersect in the same number of treatments (cf. Raghavarao (1988)).

A BD(v, b, r, k) is called a group divisible (GD) design with parameters $v = mn, b, r, k$, λ_1, λ_2 if the mn treatments are divided into m groups of n treatments each such that any two treatments in the same group occur together in exactly λ_1 blocks, whereas any two treatments from different groups occur together in exactly λ_2 blocks. The GD designs are further classified into three subclasses: Singular if $r-\lambda_1 = 0$; Semi-Regular (SR) if $r-\lambda_1 > 0$ and $rk - v\lambda_2 = 0$; Regular if $r - \lambda_1 > 0$ and $rk - v\lambda_2 > 0$.

A pairwise balanced (PB) design with parameters v, b, r, K, λ is a collection of b subsets (blocks) of a set of v treatments such that the size of each block is an element of a set K, each treatment occurs in exactly r different blocks and any two treatments occur in exactly λ blocks (cf. Colbourn and Dinitz (2007)).

Kadowaki and Kageyama (2009, 2010) produced comprehensive combinatorial findings on affine resolvable partially balanced incomplete block designs including SRGD designs.

Corresponding Author: Satoru Kadowaki
Email: s.kadowaki@matsue-ct.jp

Recently, Kadowaki and Kageyama (2020) derived a necessary condition for the existence of a certain resolvable PB design. In this paper, through the necessary condition, we can simply prove the non-existence of certain affine resolvable SRGD designs, rather than by using the other usual methods in combinatorics.

2. Statements

Kadowaki and Kageyama (2020) have presented a necessary condition for the existence of a resolvable PB design, as the following shows.

Theorem 1: When $v \geq 3$ and $b = 2r$, if there exists a resolvable PB design with parameters v, b, r, K, λ, then $r - \lambda$ is even.

By use of Theorem 1, the following can be provided.

Theorem 2: In a resolvable GD design with parameters $v = mn, b = 2r, r, k, \lambda_1, \lambda_2$, it holds that

(1) when $n \geq 3, r - \lambda_1$ is even;

(2) when $m \geq 3, r - \lambda_2$ is even.

Proof: For the resolvable GD design, by taking n treatments of the first associates in a group of the GD association scheme as new treatments, a resolvable PB design with parameters $v^* = n, b^* = 2r^*, r^* = r, k_j^*, \lambda^* = \lambda_1$ can be obtained. Then Theorem 1 implies (1), because $r^* - \lambda^* = r - \lambda_1$. Similarly, by taking m treatments of the second associates in different groups of the GD association scheme as new treatments, a resolvable PB design with parameters $v^* = m, b^* = 2r^*, r^* = r, k_j^*, \lambda^* = \lambda_2$ can be obtained. Hence (2) is proved by Theorem 1. □

It should be remarked that the presentation of Theorem 2 is simply in terms of design parameters.

Now we look at affine resolvable SRGD designs. Since an affine resolvable design is also a resolvable design, Theorem 2 will be utilized to discuss the existence of affine resolvable SRGD designs.

More refinement of Theorem 2 is given.

Corollary 2.1: In an affine resolvable SRGD design with parameters $v = mn, b = 2r, r, k, \lambda_1, \lambda_2$, it holds that

(3) m and n are both even;

(4) for even m (≥ 2), the necessary condition (1) is always satisfied;

(5) for even n (≥ 2), the necessary condition (2) is satisfied if and only if $m \equiv 0 \pmod 4$. Then when $m \equiv 2 \pmod 4$, the corresponding affine resolvable SRGD design does not exist. In particular, when $n = 2$, the parameters are expressed by $v = b = 2m, r = k = m, \lambda_1 = 0, \lambda_2 = m/2$ for $m \equiv 0 \pmod 4$.

Proof: In an affine resolvable SRGD design with $v = mn$ and $b = 2r$, it is known that $b = v-m+r, rk = v\lambda_2$ and $r(k-1) = (n-1)\lambda_1+n(m-1)\lambda_2$, which imply that $v = mn, b = 2m(n-1), r = m(n-1), k = mn/2, \lambda_1 = k-m$ and $\lambda_2 = k-k/n$. In an SRGD design, it is known that k is divisible by m. By use of these relations, Corollary 2.1 is proved. In fact, since k/m and k/n with $k = mn/2$ are integers, (3) is obtained. (4) For $m = 2l_1(l_1 \geq 1)$, since $k/m = k/(2l_1)$ is an integer, k is even. Then $r-\lambda_1 = m(n-1)-k+m = 2l_1n-k$ is even, i.e., (1) of Theorem 2. (5) For $n = 2l_2(l_2 \geq 1)$, we have $v = 2l_2m, r = m(2l_2-1), k = l_2m$. Then $\lambda_2 = k-k/n = l_2m-m/2$ which implies that m is even ($=2l_3$, say). Now $r-\lambda_2 = m(2l_2-1)-l_2m+m/2 = l_2m-m/2 = l_3(2l_2-1)$ is even if and only if l_3 is even, i.e., (2) of Theorem 2. Hence $m \equiv 0 \pmod 4$. □

Remark 1: In general, an affine resolvable SRGD design with $b = 2r$ and $\lambda_1 = 0$ has the parameters as $v = mn, b = 2m(n-1), r = m(n-1), k = m, \lambda_1 = 0, \lambda_2 = m-m/n$.

Some existence on affine resolvable SRGD designs with admissible parameters within the scope of $v = mn \leq 100$ and $r, k \leq 20$ has been investigated and tabulated by Kadowaki and Kageyama (2009, 2010). They showed the non-existence of 10 designs of Nos. 14, 16, 17, 18, 25, 27, 32, 34, 35, 38 in their Table 3.4. Such results are given by use of calculation of the Hilbert norm residue symbol (as in Theorem 12.6.2 of Raghavarao (1988)) and the application of existence on some Hadamard matrix and difference matrix (as in Remark 3.3.1 and Corollary 3.3.4 of Kadowaki and Kageyama (2009)). The derivation of these results is slightly complicated.

Among such 10 non-existence results in the Table 3.4, another proof of the non-existence on designs of Nos. 14, 16, 25, 32, 38 can be given by simply applying Theorem 2 or Corollary 2.1 as follows.

No	m	n	v	b	r	k	λ_1	λ_2	non-existence
14	6	2	12	12	6	6	0	3	Corollary 2.1
16	6	4	24	36	18	12	6	9	Theorem 2
25	10	2	20	20	10	10	0	5	Corollary 2.1
32	14	2	28	28	14	14	0	7	Corollary 2.1
38	18	2	36	36	18	18	0	9	Corollary 2.1

If we extend the investigation of affine resolvable SRGD designs of $b = 2r$ over the scope of $v \leq 200$ and $k \leq 100$, then a large number of admissible parameters of the design are found. Some are constructed and some are unknown on existence. Also some are non-existent.

Among such non-existent designs, the following parameters as

$$v = 136, m = 34, n = 4, b = 204, r = 102, k = 68, \lambda_1 = 34, \lambda_2 = 51$$

cannot be applied for Remark 3.3.1 and Corollary 3.3.4 of Kadowaki and Kageyama (2009), and also for Theorem 12.6.2 of Raghavarao (1988). The present Corollary 2.1 (5) can be only applied to show the non-existence.

The above observation reveals that Theorem 2 or Corollary 2.1 is powerful to show the non-existence of designs, rather than by the usual methods in combinatorics.

References

Clatworthy, W. H. (1973). *Tables of Two-Associate-Class Partially Balanced Designs*, NBS Applied Mathematics Series 63, U.S. Department of Commerce, National Bureau of Standards, Washington, D.C. (Library of Congress Catalog Number: 70-187759).

Colbourn, C. J. and Dinitz, J. H. (editors) (2007). *Handbook of Combinatorial Designs, Second Edition*. Chapman & Hall/CRC, New York, pp.231-236 (ISBN: 1-58488-506-8).

Kadowaki, S. and Kageyama, S. (2009). Existence of affine α-resolvable PBIB designs with some constructions. *Hiroshima Mathematical Journal*, **39**, 293-326.

Kadowaki, S. and Kageyama, S. (2010). Erratum, to "Existence of affine α-resolvable PBIB designs with some constructions". *Hiroshima Mathematical Journal*, **40**, 271.

Kadowaki, S. and Kageyama, S. (2020). A necessary condition for the existence of a certain resolvable pairwise balanced design. *International Journal of Theoretical and Applied Mathematics.*, to appear.

Raghavarao, D. (1988). *Constructions and Combinatorial Problems in Design of Experiments*. Dover, New York (ISBN: 0-486-65685-3).

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 35–44

Upper and Lower Bounds of the Dispersion of a Mean Estimator in the Growth Curve Model

Shinpei Imori[1] **and Dietrich von Rosen**[2,3]
[1]*Graduate School of Advanced Science and Engineering*
Hiroshima University, Japan
[2]*Energy and Technology*
Swedish University of Agricultural Sciences
[3]*Department of Mathematics*
Linköping University, Sweden

Received: 26 March 2020; Revised: 15 April 2020; Accepted: 20 April 2020

Abstract

When more repeated measurements than independent observations are available the classical Growth Curve model cannot produce maximum likelihood estimators. In this article we are interested in the estimation of the mean parameters whereas the dispersion parameters are considered to be nuisance parameters. It is possible to produce an unbiased estimator of the mean parameters which is a function of the Moore-Penrose generalized inverse of a singular Wishart matrix. However, its dispersion seems very hard to derive. Therefore, upper and lower bounds of the dispersion are derived. Based on the bounds a general conclusion is that the proposed estimator will work better when the number of repeated measurements is much larger than the number of independent observations than when the number of repeated measurements and the number of independent observations are of similar size.

Key words: Growth curve model; High-dimensional setting; Moore-Penrose generalized inverse.

AMS2020 Subject Classifications: 62H12, 62H99

1. Introduction

High-dimensional statistics (more variables than independent observations) has been considered for many years and it is clear that many problems still are challenging to statistics. Kollo *et al.* (2011) published an article where the classical Growth Curve model (Potthoff & Roy, 1964) was treated and simulations indicated that the proposed ideas were reasonable. One overall conclusion was that problems with the Growth Curve model is more likely to occur when the number of variables is close to the number of independent observations but that the approach of Kollo *et al.* (2011) works when there are many more variables (repeated measurements) than independent observations. Unfortunately, the technical treatment was

Correponding Author: Dietrich von Rosen
Email: dietrich.von.rosen@slu.se

not completely correct since it was utilized that $\mathbf{A}(\mathbf{A}^\top\mathbf{V}\mathbf{A})^+\mathbf{A}^\top = \mathbf{V}^+$ for any non-singular matrix $\mathbf{A}$, $\mathbf{V}$ is positive semi-definite and $^+$ denotes the Moore-Penrose generalized inverse (for a definition of the inverse see *e.g.*, Kollo and von Rosen, 2005; Definition 1.1.5). This relation does however not hold unless $\mathbf{A}$ is an orthogonal matrix. In the paper by Kollo *et al.* (2011) there was a need to calculate moments of expressions involving the Moore-Penrose inverse of a singular Wishart matrix and this took place via the incorrect relation and some invariance arguments. In this article upper and lower bounds of the above mentioned moments will be derived which will support the overall conclusions in Kollo *et al.* (2011), although an exact expression for the dispersion matrix of the mean estimator seems very difficult to obtain.

Throughout the article bold upper cases will denote real matrices, $\mathbf{X} \sim N_{p,n}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{I}_n)$ means that $\mathbf{X}$ of size $p \times n$ is matrix normally distributed with n independent columns which are multivariate normally distributed with mean equal to $\mathbf{0}$ and dispersion $\mathbf{\Sigma}$ which is supposed to be positive definite and $\mathbf{I}_n$ is the identity matrix of size $n \times n$. Note that the dispersion of $\mathbf{X}$ is given by $D[\mathbf{X}] = \mathbf{I}_n \otimes \mathbf{\Sigma}$, where $\otimes$ denotes the Kronecker product. Moreover, $\mathbf{V} \sim W_p(\mathbf{\Sigma}, n)$ denotes that $\mathbf{V}$ is Wishart distributed with n degrees of freedom, which holds if $\mathbf{V}$ can be factored as $\mathbf{V} = \mathbf{X}\mathbf{X}^\top$ (equality in distribution), where $\mathbf{X} \sim N_{p,n}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{I}_n)$ and $^\top$ denotes the transpose. The rank of a matrix $\mathbf{A}$ is denoted $r(\mathbf{A})$.

2. Preparation

In this section three definitions and two useful lemmas are presented. Let $\mathbf{A} \geq 0$ ($\mathbf{A} > 0$) denote that $\mathbf{A}$ is positive semi-definite (positive definite) and let $\mathbf{A} \geq \mathbf{B}$ mean that $\mathbf{A} - \mathbf{B} \geq 0$, where both $\mathbf{A}$ and $\mathbf{B}$ are supposed to be positive semi-definite.

Definition 1:

(i) (Löwner ordering) Let $\mathbf{U}$ and $\mathbf{V}$ be positive semi-definite matrices. If for all vectors $\boldsymbol{\alpha}$ of proper size $\boldsymbol{\alpha}^\top\mathbf{U}\boldsymbol{\alpha} \leq \boldsymbol{\alpha}^\top\mathbf{V}\boldsymbol{\alpha}$ then $\mathbf{V} \geq \mathbf{U}$.

(ii) Let $\mathbf{U}$ and $\mathbf{V}$ be positive semi-definite matrices. If for all vectors $\boldsymbol{\alpha}$ of proper size $\boldsymbol{\alpha}^\top E[\mathbf{U}]\boldsymbol{\alpha} \leq \boldsymbol{\alpha}^\top E[\mathbf{V}]\boldsymbol{\alpha}$ then $E[\mathbf{V}] \geq E[\mathbf{U}]$.

(iii) If for all vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of proper size

$$(\boldsymbol{\alpha} \otimes \boldsymbol{\beta})^\top D[\mathbf{X}](\boldsymbol{\alpha} \otimes \boldsymbol{\beta}) \leq (\boldsymbol{\alpha} \otimes \boldsymbol{\beta})^\top D[\mathbf{Y}](\boldsymbol{\alpha} \otimes \boldsymbol{\beta})$$

then it is written $D[\mathbf{X}] \preceq D[\mathbf{Y}]$, *i.e.*, $D[\boldsymbol{\beta}^\top\mathbf{X}\boldsymbol{\alpha}] \leq D[\boldsymbol{\beta}^\top\mathbf{Y}\boldsymbol{\alpha}]$.

The first lemma is presenting a known result of an explicit expression of a Moore-Penrose inverse of a singular Wishart matrix which can easily be verified via the four defining conditions of the Moore-Penrose inverse.

Lemma 1: Let $\mathbf{V} \sim W_p(\mathbf{\Sigma}, m)$, $p > m$. Then

$$\mathbf{V}^+ = \mathbf{U}(\mathbf{U}^\top\mathbf{U})^{-1}(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top,$$

where $\mathbf{V} = \mathbf{U}\mathbf{U}^\top$ and $\mathbf{U} \sim N_{p,m}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{I}_m)$.

Now a moment relation is presented which will be fundamental for the main results.

Lemma 2: Let $\mathbf{Q}$: $p \times q$, $q < p$, $\mathbf{P}$: $p \times p$ be of full rank and $\mathbf{V} \sim W_p(\mathbf{I}_p, n)$, $p < n$. Then

$$\begin{aligned} &E[(\mathbf{Q}^\top\mathbf{V}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^\top\mathbf{V}^{-1}\mathbf{P}\mathbf{V}^{-1}\mathbf{Q}(\mathbf{Q}^\top\mathbf{V}^{-1}\mathbf{Q})^{-1} \\ &\quad = (\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top\mathbf{P}\mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1} + c_1\mathrm{tr}\{\mathbf{P}(\mathbf{I} - \mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top)\}(\mathbf{Q}^\top\mathbf{Q})^{-1}, \end{aligned}$$

where $c_1 = (n - (p-q) - 1)^{-1}$. **Proof:** Factor $\mathbf{Q}$ as $\mathbf{Q}^\top = \mathbf{H}(\mathbf{I}_q : \mathbf{0})\mathbf{\Gamma}$, where $\mathbf{H}$: $q \times q$ is a non-singular matrix and $\mathbf{\Gamma}$: $p \times p$ is an orthogonal matrix. Then,

$$\begin{aligned} &(\mathbf{Q}^\top\mathbf{V}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^\top\mathbf{V}^{-1}\mathbf{P}\mathbf{V}^{-1}\mathbf{Q}(\mathbf{Q}^\top\mathbf{V}^{-1}\mathbf{Q})^{-1} \\ &\quad = (\mathbf{H}^\top)^{-1}((\mathbf{I}_q : \mathbf{0})\mathbf{\Gamma}\mathbf{V}^{-1}\mathbf{\Gamma}^\top(\mathbf{I}_q : \mathbf{0})^\top)^{-1}(\mathbf{I}_q : \mathbf{0})\mathbf{\Gamma}\mathbf{V}^{-1}\mathbf{P}\mathbf{V}^{-1}\mathbf{\Gamma}^\top(\mathbf{I}_q : \mathbf{0})^\top \\ &\qquad \times((\mathbf{I}_q : \mathbf{0})\mathbf{\Gamma}\mathbf{V}^{-1}\mathbf{\Gamma}^\top(\mathbf{I}_q : \mathbf{0})^\top)^{-1}\mathbf{H}^{-1}. \end{aligned} \tag{1}$$

Moreover, $\mathbf{\Gamma}\mathbf{V}^{-1}\mathbf{\Gamma}^\top = (\mathbf{\Gamma}\mathbf{V}\mathbf{\Gamma}^\top)^{-1}$ follows the same distribution as $\mathbf{V}^{-1}$. Thus, the right hand side of (1) follows the same distribution as

$$(\mathbf{H}^\top)^{-1}(\mathbf{V}^{11})^{-1}(\mathbf{V}^{11} : \mathbf{V}^{12})\mathbf{\Gamma}\mathbf{P}\mathbf{\Gamma}^\top(\mathbf{V}^{11} : \mathbf{V}^{12})^\top(\mathbf{V}^{11})^{-1}\mathbf{H}^{-1} \tag{2}$$

where $\mathbf{V}^{11}$ and $\mathbf{V}^{12}$ are defined via

$$\mathbf{V}^{-1} = \begin{pmatrix} \mathbf{V}^{11} & \mathbf{V}^{12} \\ \mathbf{V}^{21} & \mathbf{V}^{22} \end{pmatrix}, \qquad \begin{matrix} q \times q & q \times (p-q) \\ (p-q) \times q & (p-q) \times (p-q) \end{matrix}$$

and similarly $\mathbf{V}_{12}$ and $\mathbf{V}_{22}$ are defined through

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}, \qquad \begin{matrix} q \times q & q \times (p-q) \\ (p-q) \times q & (p-q) \times (p-q) \end{matrix}.$$

The submatrices satisfy (see *e.g.* Kollo and von Rosen, 2005; Proposition 1.3.4 (i)) $(\mathbf{V}^{11})^{-1}\mathbf{V}^{12} = -\mathbf{V}_{12}\mathbf{V}_{22}^{-1}$. Let $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ be defined through $\mathbf{\Gamma}^\top = (\mathbf{\Gamma}_1^\top : \mathbf{\Gamma}_2^\top)$: $(p \times q : p \times (p-q))$ and note that $\mathbf{\Gamma}_1\mathbf{\Gamma}_1^\top = \mathbf{I}_q$. Then (2) equals

$$\begin{aligned} &(\mathbf{H}^\top)^{-1}(\mathbf{\Gamma}_1 - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{\Gamma}_2)\mathbf{P}(\mathbf{\Gamma}_1^\top - \mathbf{\Gamma}_2^\top\mathbf{V}_{22}^{-1}\mathbf{V}_{12})\mathbf{H}^{-1} \\ &\quad = (\mathbf{H}^\top)^{-1}(\mathbf{\Gamma}_1\mathbf{P}\mathbf{\Gamma}_1^\top - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{\Gamma}_2\mathbf{P}\mathbf{\Gamma}_1^\top - \mathbf{\Gamma}_1\mathbf{P}\mathbf{\Gamma}_2^\top\mathbf{V}_{22}^{-1}\mathbf{V}_{21} \\ &\qquad + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{\Gamma}_2\mathbf{P}\mathbf{\Gamma}_2^\top\mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{H}^{-1}. \end{aligned} \tag{3}$$

It will be utilized that $\mathbf{V}_{12}\mathbf{V}_{22}^{-1/2} \sim N_{q,p-q}(\mathbf{0}, \mathbf{I}_q, \mathbf{I}_{p-q})$ which is independent of $\mathbf{V}_{22}$ (see *e.g.*, Kollo and von Rosen, 2005; Theorem 2.4.12). The expectation of the expression in the right hand side of (3) is to be derived. Since $E[\mathbf{V}_{12}\mathbf{V}_{22}^{-1}] = \mathbf{0}$ it follows from (3) that the next expression should be calculated:

$$(\mathbf{H}^\top)^{-1}\mathbf{\Gamma}_1\mathbf{P}\mathbf{\Gamma}_1\mathbf{H}^{-1} + (\mathbf{H}^\top)^{-1}E[\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{\Gamma}_2\mathbf{P}\mathbf{\Gamma}_2^\top\mathbf{V}_{22}^{-1}\mathbf{V}_{21}]\mathbf{H}^{-1}. \tag{4}$$

Moreover, applying an expectation result for quadratic forms in normally distributed variables (*e.g.*, see Kollo and von Rosen, 2005, Theorem 2.2.9 (i)) implies that the expectation in (4) equals

$$E[\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{\Gamma}_2\mathbf{P}\mathbf{\Gamma}_2^\top\mathbf{V}_{22}^{-1}\mathbf{V}_{21}] = E[\mathrm{tr}\{\mathbf{V}_{22}^{-1}\mathbf{\Gamma}_2\mathbf{P}\mathbf{\Gamma}_2^\top\}]\mathbf{I}_\mathrm{q} \tag{5}$$

which since $\mathbf{V}_{22} \sim W_{p-q}(\mathbf{I}, n)$ (see *e.g.*, Kollo and von Rosen, 2005, Theorem 2.4.14 (iii))

$$E[\mathbf{V}_{22}^{-1}] = c_1\mathbf{I}_{p-q}, \qquad c_1 = \tfrac{1}{n-(p-q)-1}$$

and the right hand side of (5) is identical to $c_1E[\mathrm{tr}\{\mathbf{\Gamma}_2\mathbf{P}\mathbf{\Gamma}_2^\top\}]\mathbf{I}_q$. In order to arrive to the statements of the theorem $\mathbf{\Gamma}_2^\top\mathbf{\Gamma}_2$, $(\mathbf{H}\mathbf{H}^\top)^{-1}$ and $\mathbf{\Gamma}_1^\top\mathbf{H}^{-1}$ have to be expressed in the original matrices. From the definition of $\mathbf{\Gamma}$ and $\mathbf{H}$ it follows that

$$\mathbf{Q}^\top = \mathbf{H}\mathbf{\Gamma}_1, \quad (\mathbf{H}\mathbf{H}^\top)^{-1} = (\mathbf{Q}^\top\mathbf{Q})^{-1}, \quad \mathbf{\Gamma}_1^\top\mathbf{H}^{-1} = \mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}$$

and

$$\mathbf{\Gamma}_1^\top\mathbf{\Gamma}_1 = \mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top, \quad \mathbf{\Gamma}_2^\top\mathbf{\Gamma}_2 = \mathbf{I} - \mathbf{\Gamma}_1^\top\mathbf{\Gamma}_1 = \mathbf{I} - \mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top.$$

These relations establish the lemma. □ Throughout the article let $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ be the ordered eigen values of a symmetric matrix $\mathbf{A}$: $n \times n$.

3. The Growth Curve Model When $p > n$

The classical Growth Curve model (Potthoff and Roy, 1964) has been applied in many areas and is a natural extension of the MANOVA model. Therefore the model is also called GMANOVA model. The model and generalizations of the model together with an almost complete list of references can be found in von Rosen (2018).

Definitiion 2: Let $\mathbf{X}$: $p \times n$, $p \geq n - r(\mathbf{C})$, $\mathbf{A}$: $p \times q$, $q \leq p$, $\mathbf{B}$: $q \times k$, $\mathbf{C}$: $k \times n$, and $\mathbf{\Sigma} > 0$: $p \times p$. Then

$$\mathbf{X} = \mathbf{A}\mathbf{B}\mathbf{C} + \mathbf{E}$$

defines the Growth Curve model, where $\mathbf{E} \sim N_{p,n}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{I})$, $\mathbf{A}$ and $\mathbf{C}$ are known matrices, and $\mathbf{B}$ and $\mathbf{\Sigma}$ are unknown parameter matrices.

Since $p \geq n - r(\mathbf{C})$ we assume a high-dimensional setting. The main purpose of this article is to discuss a specific estimator of $\mathbf{B}$. Note that the size of $\mathbf{B}$, *i.e.*, $q \times k$, does not depend on n and p, and that $\mathbf{\Sigma}$ is thought of being a nuisance parameter. For some details of how to treat the Growth Curve model in a high-dimensional setting see Kollo *et al.* (2011). Alternatively, we can suppose that $\mathbf{\Sigma}$ is known and then from linear models theory it follows that under the assumption $r(\mathbf{A}) = q$, $r(\mathbf{C}) = k$ which will be supposed to hold throughout the article, an estimator of $\mathbf{B}$ equals

$$\widetilde{\mathbf{B}} = (\mathbf{A}^\top\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^{-1}.$$

However, $\mathbf{\Sigma}^{-1}$ has to be estimated. One can use the sums of squares matrix $\mathbf{S} = \mathbf{X}(\mathbf{I} - \mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^{-1}\mathbf{C})\mathbf{X}^\top$ as an estimator of $n\mathbf{\Sigma}$ but if $p > n - r(\mathbf{C})$ the inverse $\mathbf{S}^{-1}$ does not exist and cannot be used to estimate $\mathbf{\Sigma}^{-1}$. Therefore, instead of $\mathbf{S}^{-1}$ the Moore-Penrose inverse $\mathbf{S}^+$ can be used and then the same estimator as in Kollo *et al.* (2011) is obtained:

$$\widehat{\mathbf{B}} = (\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{S}^+\mathbf{X}\mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^{-1}, \tag{6}$$

where it has to be assumed that the column vector space relation $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{S}) = \{\mathbf{0}\}$ is satisfied which implies that $(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}$ holds.

4. $E[\widehat{\mathbf{B}}]$ and Bounds For $D[\widehat{\mathbf{B}}]$

In order to derive the expectation and bounds for the dispersion for $\widehat{\mathbf{B}}$ in (6) it will be utilized that $\mathbf{XC}^\top$ and $\mathbf{S}$ are independently distributed.

Theorem 1: Let $\widehat{\mathbf{B}}$ be defined in (6). Then $E[\widehat{\mathbf{B}}] = \mathbf{B}$.

Proof:

$$E[\widehat{\mathbf{B}}] = E[(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{S}^+]E[\mathbf{XC}^\top(\mathbf{CC}^\top)^{-1}] = E[(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{S}^+\mathbf{AB}] = \mathbf{B}.$$

□

Turning to a discussion of the dispersion matrix for $\widehat{\mathbf{B}}$ it follows that the dispersion matrix for $\widehat{\mathbf{B}}$ in (6) can be presented as

$$\begin{aligned} D[\widehat{\mathbf{B}}] &= E[\mathrm{vec}(\widehat{\mathbf{B}} - \mathbf{B})\mathrm{vec}^\top(\widehat{\mathbf{B}} - \mathbf{B})] \\ &= E[((\mathbf{CC}^\top)^{-1}\mathbf{C} \otimes (\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{S}^+)D[\mathbf{X}](\mathbf{C}^\top(\mathbf{CC}^\top)^{-1} \otimes \mathbf{S}^+\mathbf{A}(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1})] \\ &= (\mathbf{CC}^\top)^{-1} \otimes E[(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{S}^+\mathbf{\Sigma}\mathbf{S}^+\mathbf{A}(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}]. \end{aligned} \tag{7}$$

From Lemma 1 it follows that the expectation in (7) is complicated to express. We will show some calculations but the aim will be to find upper and lower bounds for the expectation similarly to the approach for obtaining bounds for the expectation and dispersion of the Moore-Penrose inverse of a singular Wishart matrix (see Imori and von Rosen, 2020).

When deriving the bounds a number of transformations will take place: $\mathbf{S} = \mathbf{UU}^\top$, where $\mathbf{U} \sim N_{p,n-r(\mathbf{C})}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{I}_{n-r(\mathbf{C})})$, $\mathbf{Y} = \mathbf{\Sigma}^{-1/2}\mathbf{U}$, where $\mathbf{\Sigma}^{-1/2}$ is a symmetric square root; $\mathbf{Y}^\top = \mathbf{TL}$ where $\mathbf{T}$: $(n-r(\mathbf{C}))\times(n-r(\mathbf{C}))$ is a lower triangular matrix and $\mathbf{L}$: $(n-r(\mathbf{C}))\times p$ is a semi-orthogonal matrix, *i.e.* $\mathbf{LL}^\top = \mathbf{I}_{n-r(\mathbf{C})}$; $\mathbf{V} = \mathbf{T}^\top\mathbf{T}$. Firstly it is noted that the expectation in (7) can be expressed as (see Lemma 1)

$$\begin{aligned} &E[(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{S}^+\mathbf{\Sigma}\mathbf{S}^+\mathbf{A}(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}] \\ &\quad = E[(\mathbf{A}^\top\mathbf{U}(\mathbf{U}^\top\mathbf{U})^{-1}(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{U}(\mathbf{U}^\top\mathbf{U})^{-1}(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{\Sigma} \\ &\qquad \times\mathbf{U}(\mathbf{U}^\top\mathbf{U})^{-1}(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{A}(\mathbf{A}^\top\mathbf{U}(\mathbf{U}^\top\mathbf{U})^{-1}(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{A})^{-1}]. \end{aligned}$$

Now the first transformation is applied to this relation and it yields

$$\begin{aligned} E[&(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{Y}(\mathbf{Y}^\top\mathbf{\Sigma}\mathbf{Y})^{-1}(\mathbf{Y}^\top\mathbf{\Sigma}\mathbf{Y})^{-1}\mathbf{Y}^\top\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1} \\ &\times\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{Y}(\mathbf{Y}^\top\mathbf{\Sigma}\mathbf{Y})^{-1}(\mathbf{Y}^\top\mathbf{\Sigma}\mathbf{Y})^{-1}\mathbf{Y}^\top\mathbf{\Sigma} \\ &\times\mathbf{\Sigma}\mathbf{Y}(\mathbf{Y}^\top\mathbf{\Sigma}\mathbf{Y})^{-1}(\mathbf{Y}^\top\mathbf{\Sigma}\mathbf{Y})^{-1}\mathbf{Y}^\top\mathbf{\Sigma}^{1/2}\mathbf{A} \\ &\times(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{Y}(\mathbf{Y}^\top\mathbf{\Sigma}\mathbf{Y})^{-1}(\mathbf{Y}^\top\mathbf{\Sigma}\mathbf{Y})^{-1}\mathbf{Y}^\top\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}]. \end{aligned}$$

Moreover, the second transformation implies that one should consider

$$\begin{aligned} E[&(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top(\mathbf{L\Sigma L}^\top)^{-1}\mathbf{T}^{-1}(\mathbf{T}^\top)^{-1}(\mathbf{L\Sigma L}^\top)^{-1}\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1} \\ &\times\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top(\mathbf{L\Sigma L}^\top)^{-1}\mathbf{T}^{-1}(\mathbf{T}^\top)^{-1}(\mathbf{L\Sigma L}^\top)^{-1}\mathbf{L\Sigma} \\ &\times\mathbf{\Sigma}\mathbf{L}^\top(\mathbf{L\Sigma L}^\top)^{-1}\mathbf{T}^{-1}(\mathbf{T}^\top)^{-1}(\mathbf{L\Sigma L}^\top)^{-1}\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A} \\ &\times(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top(\mathbf{L\Sigma L}^\top)^{-1}\mathbf{T}^{-1}(\mathbf{T}^\top)^{-1}(\mathbf{L\Sigma L}^\top)^{-1}\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}] \end{aligned}$$

and then the third transformation implies an expression which will be studied in detail:

$$\begin{aligned} E[&(\mathbf{A}^\top\boldsymbol{\Sigma}^{1/2}\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{V}^{-1}(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{L}\boldsymbol{\Sigma}^{1/2}\mathbf{A})^{-1} \\ &\times\mathbf{A}^\top\boldsymbol{\Sigma}^{1/2}\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{V}^{-1}(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{L}\boldsymbol{\Sigma} \\ &\times\boldsymbol{\Sigma}\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{V}^{-1}(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{L}\boldsymbol{\Sigma}^{1/2}\mathbf{A} \\ &\times(\mathbf{A}^\top\boldsymbol{\Sigma}^{1/2}\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{V}^{-1}(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{L}\boldsymbol{\Sigma}^{1/2}\mathbf{A})^{-1}], \end{aligned} \tag{8}$$

where it can be shown that $\mathbf{V} \sim W_{n-r(\mathbf{C})}(\mathbf{I}_{n-r(\mathbf{C})}, p)$ which for example follows from the derivation of the Wishart density in Srivastava and Khatri (1979; Theorem 3.2.1) or Imori and von Rosen (2020; Section 3.1). Put

$$\mathbf{Q} = (\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{L}\boldsymbol{\Sigma}^{1/2}\mathbf{A}, \tag{9}$$

$$\mathbf{P} = (\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}\mathbf{L}\boldsymbol{\Sigma}\boldsymbol{\Sigma}\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top)^{-1}. \tag{10}$$

Then (8) is identical to

$$E[(\mathbf{Q}^\top\mathbf{V}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^\top\mathbf{V}^{-1}\mathbf{P}\mathbf{V}^{-1}\mathbf{Q}(\mathbf{Q}^\top\mathbf{V}^{-1}\mathbf{Q})^{-1}]$$

which, since $\mathbf{V} \sim W_{n-r(\mathbf{C})}(\mathbf{I}_{n-r(\mathbf{C})}, p)$, according to Lemma 2

$$E[(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top\mathbf{P}\mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}] + c_1E[\mathrm{tr}\{\mathbf{P}(\mathbf{I} - \mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top)\}(\mathbf{Q}^\top\mathbf{Q})^{-1}], \tag{11}$$

where $c_1^{-1} = p - (n - r(\mathbf{C}) - q) - 1$ and the expectation in (11) is taken over the semi-orthogonal matrix $\mathbf{L}$. Note that it has to be assumed that $c_1 > 0$, *i.e.*, $p > n - r(\mathbf{C}) - q + 1$ but later we need that $p \geq n - r(\mathbf{C})$) . However, it is difficult to perform the integration in (11) and therefore we first focus on finding upper and lower bounds of

$$(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top\mathbf{P}\mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}, \quad \mathrm{tr}\{\mathbf{P}(\mathbf{I} - \mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top)\}$$

which are either functionally independent of $\mathbf{L}$ or are so simplified that only $E[(\mathbf{Q}^\top\mathbf{Q})^{-1}]$ has to be derived.

Lemma 3: Let $\mathbf{P}$ be given by (10). Then

$$\lambda_p(\boldsymbol{\Sigma})\lambda_p(\boldsymbol{\Sigma}^{-1})\mathbf{I}_{n-r(\mathbf{C})} \leq \mathbf{P} \leq \lambda_1(\boldsymbol{\Sigma})\lambda_1(\boldsymbol{\Sigma}^{-1})\mathbf{I}_{n-r(\mathbf{C})}.$$

Proof: The proof is based on a spectral decomposition of $\boldsymbol{\Sigma}$ which yields $\lambda_p(\boldsymbol{\Sigma})\mathbf{I}_p \leq \boldsymbol{\Sigma} \leq \lambda_1(\boldsymbol{\Sigma})\mathbf{I}_p$. Note that $\lambda_p(\boldsymbol{\Sigma})\boldsymbol{\Sigma} \leq \boldsymbol{\Sigma}\boldsymbol{\Sigma} \leq \lambda_1(\boldsymbol{\Sigma})\boldsymbol{\Sigma}$ and therefore $\lambda_1(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top) \leq \lambda_1(\boldsymbol{\Sigma})$, $\lambda_n(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top) \geq \lambda_p(\boldsymbol{\Sigma})$ which jointly establish the lemma. □ Applying Lemma 3 yields that upper and lower bounds for (11) are given by

$$\begin{aligned} &\lambda_p(\boldsymbol{\Sigma})\lambda_p(\boldsymbol{\Sigma}^{-1})(1 + c_1(n - r(\mathbf{C}) - q))E[(\mathbf{Q}\mathbf{Q}^\top)^{-1}] \\ &\quad\leq E[(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top\mathbf{P}\mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}] + c_1E[\mathrm{tr}\{\mathbf{P}(\mathbf{I} - \mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top)\}(\mathbf{Q}^\top\mathbf{Q})^{-1}] \\ &\quad\leq \lambda_1(\boldsymbol{\Sigma})\lambda_1(\boldsymbol{\Sigma}^{-1})(1 + c_1(n - r(\mathbf{C}) - q))E[(\mathbf{Q}\mathbf{Q}^\top)^{-1}]. \end{aligned} \tag{12}$$

Note that $1 + c_1(n - r(\mathbf{C}) - q) = (p - 1)/(n - r(\mathbf{C}) - q)$. Moreover, (12) implies that we now need to find bounds for $E[(\mathbf{Q}\mathbf{Q}^\top)^{-1}]$.

Lemma 4: Let $\mathbf{Q}$ be defined in (9). Then

$$\lambda_p(\mathbf{\Sigma\Sigma})(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1} \le (\mathbf{Q}^\top\mathbf{Q})^{-1} \le \lambda_1(\mathbf{\Sigma\Sigma})(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}.$$

Proof: It is enough to show upper and lower limits for $(\mathbf{L\Sigma L}^\top)^{-1}(\mathbf{L\Sigma L}^\top)^{-1}$ which should be independent of $\mathbf{L}$ and proportional to $\mathbf{I}_{n-r(\mathbf{C})}$. By pre- and post-multiplying by $\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top$ and then taking the inverse establish the lemma. Now

$$\lambda_n((\mathbf{L\Sigma L}^\top)^{-1}(\mathbf{L\Sigma L}^\top)^{-1}) = (\lambda_1(\mathbf{L\Sigma L}^\top))^{-2} \ge (\lambda_1(\mathbf{\Sigma\Sigma}))^{-1}$$

and

$$\lambda_1((\mathbf{L\Sigma L}^\top)^{-1}(\mathbf{L\Sigma L}^\top)^{-1}) = (\lambda_n(\mathbf{L\Sigma L}^\top))^{-2} \le (\lambda_p(\mathbf{\Sigma\Sigma}))^{-1}$$

which yield the inequalities of the lemma. □ From Lemma 4 it follows that we need to calculate $E[(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}]$, where $\mathbf{LL}^\top = \mathbf{I}_{n-r(\mathbf{C})}$. The result is stated in the next lemma.

Lemma 5: Let all matrices be as in Lemma 4. Then, if $p \ge n - r(\mathbf{C}) > q - 1$,

$$E[(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}] = \frac{p-q-1}{n-r(\mathbf{C})-q-1}(\mathbf{A}^\top\mathbf{\Sigma A})^{-1}.$$

Proof: The same transformations as when deriving (8) will now be applied. Let $\mathbf{Y} \sim N_{p,n-r(\mathbf{C})}(\mathbf{0}, \mathbf{I}_p, \mathbf{I}_{n-r(\mathbf{C})})$, $p \ge n - r(\mathbf{C})$. Then $\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Sigma}^{1/2}\mathbf{A} \sim W_q(\mathbf{A}^\top\mathbf{\Sigma A}, n - r(\mathbf{C}))$ and (expectation of an inverse Wishart matrix is applied; *e.g.*, see Kollo and von Rosen, 2005, Theorem 2.4.14 (iii))

$$E[(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}] = \frac{1}{n-r(\mathbf{C})-q-1}(\mathbf{A}^\top\mathbf{\Sigma A})^{-1}. \tag{13}$$

Next the variable substitution $\mathbf{Y}^\top = \mathbf{TL}$ is made, where $\mathbf{T}$: $(n-r(\mathbf{C})) \times (n-r(\mathbf{C}))$, is lower triangular with positive diagonal elements and $\mathbf{L}$ is semi-orthogonal, *i.e.*, $\mathbf{LL}^\top = \mathbf{I}_{n-r(\mathbf{C})}$. The matrices $\mathbf{T}$ and $\mathbf{L}$ are independently distributed. Moreover, $\mathbf{V} = \mathbf{T}^\top\mathbf{T} \sim W_{n-r(\mathbf{C})}(\mathbf{I}_{n-r(\mathbf{C})}, p)$ and given $\mathbf{L}$

$$\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top\mathbf{V}\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A} \sim W_q(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A}, p).$$

Thus,

$$\begin{aligned} E[(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}] &= E[(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top\mathbf{T}^\top\mathbf{T}\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}] \\ &= E[(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top\mathbf{V}\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}] = \frac{1}{p-q-1}E[(\mathbf{A}^\top\mathbf{\Sigma}^{1/2}\mathbf{L}^\top\mathbf{L}\mathbf{\Sigma}^{1/2}\mathbf{A})^{-1}] \end{aligned}$$

and combining this result with (13) establishes the lemma □ From Lemma 3 and Lemma 5 it follows that

$$\begin{aligned} &\lambda_p(\mathbf{\Sigma})\lambda_p(\mathbf{\Sigma}^{-1})\lambda_p(\mathbf{\Sigma\Sigma})(1 + c_1(n - r(\mathbf{C}) - q)\tfrac{p-q-1}{n-r(\mathbf{C})-q-1})(\mathbf{A}^\top\mathbf{\Sigma A})^{-1} \\ &\quad \le E[(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{S}^+\mathbf{\Sigma}\mathbf{S}^+\mathbf{A}(\mathbf{A}^\top\mathbf{S}^+\mathbf{A})^{-1}] \\ &\quad \le \lambda_1(\mathbf{\Sigma})\lambda_1(\mathbf{\Sigma}^{-1})\lambda_1(\mathbf{\Sigma\Sigma})(1 + c_1(n - r(\mathbf{C}) - q)\tfrac{p-q-1}{n-r(\mathbf{C})-q-1})(\mathbf{A}^\top\mathbf{\Sigma A})^{-1}. \end{aligned} \tag{14}$$

Now all preparations are finished and the main result can immediately be presented:

Theorem 2: Let $\widehat{\mathbf{B}}$ be defined in (6) and assume $p \geq n - r(\mathbf{C}) > q - 1$. Then, ($\succeq$ was introduced in Definition 1 (iii))

(i) $D[\widehat{\mathbf{B}}] \succeq (\mathbf{C}\mathbf{C}^\top)^{-1} \otimes \lambda_p(\mathbf{\Sigma})^3 \lambda_p(\mathbf{\Sigma}^{-1}) \frac{p-1}{p-(n-r(\mathbf{C})-q)-1} \frac{p-q-1}{n-r(\mathbf{C})-q-1} (\mathbf{A}^\top \mathbf{\Sigma} \mathbf{A})^{-1}$;

(ii) $D[\widehat{\mathbf{B}}] \preceq (\mathbf{C}\mathbf{C}^\top)^{-1} \otimes \lambda_1(\mathbf{\Sigma})^3 \lambda_1(\mathbf{\Sigma}^{-1}) \frac{p-1}{p-(n-r(\mathbf{C})-q)-1} \frac{p-q-1}{n-r(\mathbf{C})-q-1} (\mathbf{A}^\top \mathbf{\Sigma} \mathbf{A})^{-1}$.

Remark 1: If $p-1$ is close to $n-r(\mathbf{C})-q$ or $n-r(\mathbf{C})$ is close to $q-1$ the dispersion for $\widehat{\mathbf{B}}$ becomes large because the lower bound becomes large. In this case an alternative estimator for $\mathbf{B}$ should be used, *e.g.*, the "unweighted" estimator $(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{X} \mathbf{C}^\top (\mathbf{C}\mathbf{C}^\top)^{-1}$.

If all eigen values of $\mathbf{\Sigma}$ are equal, *e.g.*, $\mathbf{\Sigma} = \mathbf{I}$, the lower and upper bound of Theorem 2 are equal, *i.e.*,

$$D[\widehat{\mathbf{B}}] = \lambda_1(\mathbf{\Sigma}) \frac{p-1}{p-(n-r(\mathbf{C})-q)-1} \frac{p-q-1}{n-r(\mathbf{C})-q-1} (\mathbf{C}\mathbf{C}^\top)^{-1} \otimes (\mathbf{A}^\top \mathbf{A})^{-1}$$

which however is larger than the variance for the unweighted estimator, as it should be according to least squares theory.

5. Simulation Study

In this section a small simulation study is conducted to illustrate Theorem 1. In Remark 1 it was noted that when p is close to n both the upper and lower bounds, for given $(\mathbf{C}\mathbf{C}^\top)^{-1} \otimes (\mathbf{A}^\top \mathbf{\Sigma} \mathbf{A})^{-1}$, depend on

$$\lambda_\bullet(\mathbf{\Sigma})^3 \lambda_\bullet(\mathbf{\Sigma}^{-1}) \frac{p-1}{p-(n-r(\mathbf{C})-q)-1} \frac{p-q-1}{n-r(\mathbf{C})-q-1}, \qquad p \geq n - r(\mathbf{C}),$$

where $\lambda_\bullet(\mathbf{\Sigma})$ denotes either $\lambda_1(\mathbf{\Sigma})$ or $\lambda_p(\mathbf{\Sigma})$ and the same holds for $\lambda_\bullet(\mathbf{\Sigma}^{-1})$. If $p = n - r(\mathbf{C})$ this expression reduces to

$$\lambda_\bullet(\mathbf{\Sigma})^3 \lambda_\bullet(\mathbf{\Sigma}^{-1}) \frac{p-1}{q-1}. \tag{15}$$

Thus, if the largest and smallest eigenvalues of $\mathbf{\Sigma}$ are stable with respect to p (15) increases linearly with p but at the same time $(\mathbf{A}^\top \mathbf{\Sigma} \mathbf{A})^{-1}$ becomes "smaller". Note also that $(\mathbf{C}\mathbf{C}^\top)^{-1}$ becomes "smaller" when n increases.

Instead of studying Theorem 1 we will study (14) since $(\mathbf{C}\mathbf{C}^\top)^{-1}$ is of no interest. In the simulations the following matrices are used: $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, where

$$\mathbf{a}_1 = \mathbf{1}_p, \qquad \mathbf{a}_2^\top = 0.7 \cdot (1, 2, \ldots, p), \qquad \mathbf{a}_3^\top = 0.01 \cdot (1, 4, \ldots, p^2)$$

and

$$\mathbf{C} = \begin{pmatrix} \mathbf{1}_{20}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{20}^\top \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 3 & 7 \\ 2 & 2 \end{pmatrix}.$$

Table 1: The table summarizes the conducted simulation study: $(\mathbf{EST})_{ii}$, $i \in \{1,2,3\}$ is the ith diagonal element of EST defined in (16), $(\mathbf{LB})_{ii}$ and $(\mathbf{UB})_{ii}$, $i \in \{1,2,3\}$, are the ith diagonal element of the lower and upper bounds respectively defined in (18) and (19). Moreover, p is the number of repeated measurements and the data has been generated according to the description in Section 5. In particular $n = 40$, $q = 3$ and $r(\mathbf{C}) = 2$.

p	$(\mathbf{LB})_{11}$	$(\mathbf{EST})_{11}$	$(\mathbf{UB})_{11}$	$(\mathbf{LB})_{22}$	$(\mathbf{EST})_{22}$	$(\mathbf{UB})_{22}$	$(\mathbf{LB})_{33}$	$(\mathbf{EST})_{33}$	$(\mathbf{UB})_{33}$
38	5.1	11.8	26.3	0.14	0.36	0.7	1.8	4.3	9.4
39	3.6	8.2	18.8	0.10	0.22	0.50	1.1	2.7	5.9
40	2.8	6.5	14.5	0.073	0.16	0.37	0.84	1.8	4.3
50	0.95	2.3	5.2	0.016	0.040	0.085	0.11	0.29	0.62
60	0.66	1.7	3.6	0.0077	0.020	0.042	0.040	0.10	0.22
80	0.49	1.2	2.7	0.0032	0.0080	0.017	0.0093	0.023	0.051
100	0.44	1.1	2.4	0.0018	0.0045	0.010	0.0034	0.0085	0.019
150	0.35	0.88	2.0	0.00069	0.0017	0.0039	0.00058	0.0014	0.0033
200	0.33	0.84	1.8	0.00035	0.00091	0.0020	0.00016	0.00043	0.00094

Concerning $\mathbf{\Sigma}$ we randomly generated eigenvectors $\mathbf{\Gamma}$ via another covariance matrix and also randomly generated eigenvalues $\{\lambda_k\}$ uniformly on the interval $[2, 3.1]$. The eigenvalues build up a diagonal matrix $\mathbf{D} = (\lambda_k)$ and then the $\mathbf{\Sigma}$ which has been used in the simulations equals $\mathbf{\Sigma} = \mathbf{\Gamma D \Gamma}^\top$. Note that in the simulations $n = 40$, $r(\mathbf{C}) = 2$ and $q = 3$. The simulations were carried out for $p \in \{38, 39, 40, 50, 60, 80, 100, 150, 200\}$. According to Theorem 1 we have to assume that $p \geq 38$ and it can be shown that the theorem is not true for $p = 37$ and if $p < 37$ our bounds do not even exist. In (14) we have

$$E[(\mathbf{A}^\top \mathbf{S}^+ \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{S}^+ \mathbf{\Sigma} \mathbf{S}^+ \mathbf{A} (\mathbf{A}^\top \mathbf{S}^+ \mathbf{A})^{-1}] \tag{16}$$

which has to be estimated. The simulation data is generated according to the model $\mathbf{X}_i \sim N_{p,n}(\mathbf{ABC}, \mathbf{\Sigma}, \mathbf{I}_n)$, $i = 1, 2, \ldots, 500$, *i.e.*, there are 500 replicates performed in the simulation study. Let $\mathbf{S}_i^+$ denote the $\mathbf{S}^+$ from the ith simulation and we have

$$\mathbf{EST} = \frac{1}{n} \sum_{i=1}^{n} E[(\mathbf{A}^\top \mathbf{S}_i^+ \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{S}_i^+ \mathbf{\Sigma} \mathbf{S}_i^+ \mathbf{A} (\mathbf{A}^\top \mathbf{S}_i^+ \mathbf{A})^{-1}] \tag{17}$$

as an unbiased estimator of the expectation in (16). The results of the simulation study are presented in Table 1. In our case **EST** is of size 3×3. Moreover, we calculated the lower bound, **LB**, and the upper bound, **UB**, as

$$\mathbf{LB} = \lambda_p^3(\mathbf{\Sigma}) \lambda_p(\mathbf{\Sigma}^{-1}) \frac{p-1}{p-34} \frac{p-4}{35} (\mathbf{A}^\top \mathbf{\Sigma} \mathbf{A})^{-1}, \tag{18}$$

$$\mathbf{UB} = \lambda_1^3(\mathbf{\Sigma}) \lambda_1(\mathbf{\Sigma}^{-1}) \frac{p-1}{p-34} \frac{p-4}{35} (\mathbf{A}^\top \mathbf{\Sigma} \mathbf{A})^{-1} \tag{19}$$

which according to the theory should give upper and lower bounds of the expectation in (16). In Table 1 the diagonal elements of **EST**, **LB** and **UB** are presented. The results

follow the theory, *i.e.*, $(\mathbf{LB})_{ii} < (\mathbf{EST})_{ii} < (\mathbf{UB})_{ii}$, $i \in \{1, 2, 3\}$. Moreover, when p increases $(\mathbf{EST})_{ii}$ becomes smaller and the difference $(\mathbf{UB})_{ii} - (\mathbf{LB})_{ii}$ is largest when $p = 38$. Thus, the results of Theorem 1 are in full agreement with the simulation study.

Acknowledgements

Dietrich von Rosen is supported by the Swedish Research Council (2017-03003).

References

Imori, S. and von Rosen, D. (2020). On the mean and dispersion of the Moore-Penrose generalized inverse of a Wishart matrix. Accepted by *Electronic Journal of Linear Algebra*.

Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices. Series: Mathematics and Its Applications*, **579**, Springer, Dordrecht.

Kollo, T., von Rosen, T. and von Rosen, D. (2011). Estimation in high-dimensional analysis and multivariate linear models. *Communications in Statistics: Theory Methods*, **40**, 1241–1253.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.

von Rosen, D. (2018). *Bilinear Regression Analysis. An Introduction. Lecture Notes in Statistics*, **220**. Springer, Cham.

Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*. North-Holland, New York-Oxford.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 45–65

Selection of Divergence Priors in the Presence of Nuisance Parameters

Ruitao Liu[1] **and Malay Ghosh**[2]
[1]*NebulAi Education*
[2]*University of Florida*

Received: 30 March 2020; Revised: 13 April 2020; Accepted: 21 April 2020

Abstract

The selection of priors in the presence of nuisance parameters is an important topic in Bayesian statistics. Bernardo (1979) proposed a stepwise procedure for handling nuisance parameters. He obtained the prior by maximizing the expected Kullback-Leibler divergence between the prior of parameters of interest and the corresponding posterior. His procedure turns out to be very efficient and has been applied to many examples. In this paper, we consider selection of priors in the presence of nuisance parameters under a general divergence criterion, originally introduced by Renyi (1961), later followed by a host of researchers, most notably by Amari (1982) and Cressie and Read (1984). This general divergence measure includes the Kullback-Leibler, Bhattacharyya-Hellinger and Chi-square divergence. It turns out that Bernardo's prior maximizes this divergence in the interior of this class of divergence measures. On the boundary, the Chi-square divergence, the prior turns out to be different from Bernardo's prior for some common families of distributions. Also, outside the boundary, Bernardo's prior turns out to be the minimizer rather than maximizer of the divergence, and there does not exist any prior which maximizes the divergence between the prior and the posterior.

Key words: General divergence; Nuisance parameters; Optimal prior; Characterisation of optimal priors.

1. Introduction

The most important component in Bayesian statistics is the prior for the unknown parameters. The selection of prior has always been a popular topic since the birth of Bayesian statistics.

Ideally, if one has enough historical data, it is possible to elicit an appropriate prior which reflects one's belief about unknown parameters. This is a subjective prior. But the choice of subjective priors is difficult, especially when there is not enough historical information available. In practice, it is common to use the so-called 'objective' priors which are also referred to as 'non-informative priors' or 'default priors'. Those priors are determined

Corresponding Author: Malay Ghosh
Email: ghoshm@ufl.edu

by some objective or structural criterion. For decades, many statisticians worked on this topic. Consequently, numerous criteria for selecting objective priors have been proposed.

One major criterion was proposed by Bernardo (1979). There are two innovations in his paper. First, he introduced the notion of missing information. He used the expected Kullback-Leibler divergence as the measure of the missing information provided by the data and then found the prior which maximizes this divergence. In the absence of nuisance parameters, he found that the desired prior is Jeffreys' general rule prior (Jeffreys 1961) which is proportional to the square root of the determinant of the Fisher information matrix. Second, he used a stepwise procedure to find the prior in the presence of nuisance parameters. His procedure is as follows. First, he assigns a conditional density of nuisance parameters given parameters of interest. Then he obtains the prior for parameters of interest by maximizing the expected Kullback-Leibler divergence between the prior for parameters of interest and the corresponding posterior.

In many cases, Bernardo's procedure produced priors different from Jeffreys' general rule prior. It turns out that his stepwise procedure often yields more reasonable priors than Jeffreys' general rule prior. One good example to show the advantage of Bernardo's stepwise procedure is the Neyman-Scott problem (1948) considered by Berger and Bernardo (1992b), Datta and Ghosh (1995a). The data consist of n pairs of observations: $X_{ij} \sim N(\mu_i, \sigma^2)$, $i = 1, \cdots, n$, $j = 1, 2$. Consider all the parameters to be of equal importance. Then, one gets Jeffreys' general rule prior $\pi(\mu_1, \cdots, \mu_n, \sigma^2) \propto (\sigma^2)^{-n/2-1}$. So the posterior mean is $s^2/(2n-2)$, where $s^2 = \sum_{i=1}^{n} \sum_{j=1}^{2} (x_{ij} - \bar{x}_i)^2$ and $\bar{x}_i = (x_{i1} + x_{i2})/2$. This is an inconsistent estimator of σ^2. On the contrary, by treating σ as the parameter of interest, using Bernardo's procedure, one gets the prior $\pi(\mu_1, \cdots, \mu_n, \sigma^2) \propto \sigma^{-2}$. This gives a posterior mean of $s^2/(n-2)$ which is consistent.

Bernardo's procedure has been applied to many examples. For example, exponential regression (Ye and Berger 1991), multinomial models (Berger and Bernardo 1992a) and AR(1) models (Berger and Yang 1994).

In both cases, with or without nuisance parameters, Bernardo used the expected Kullback-Leibler divergence to develop priors. One may ask questions like: What will happen if we use another divergence? Will we get the same priors as Berdardo did? If not, what do the new priors look like?

Instead of the Kullback-Leibler divergence, Clarke and Sun (1997) considered the expected Chi-square divergence motivated by the classical Chi-square goodness-of-fit statistic. They showed that, for the one-parameter exponential family of distributions with the canonical parameter, maximization of the Chi-square divergence led to a prior different from Jeffreys' prior. For multi-parameter exponential family of distributions, they conjectured that the prior should also be of the same form as they got in one-parameter case. For the case where nuisance parameters are present, they gave brief discussion and left it as an open question.

Recently, for regular one-parameter family of distributions, Ghosh, Mergel and Liu (2011) considered a general divergence between prior and posterior which has been considered

in other contexts by several authors (for example, Renyi, 1961; Amari, 1982; Cressie and Read, 1984). It is a family of divergence measures including the Kullback-Leibler divergence, the Bhattacharyya-Hellinger divergence (Bhattacharyya, 1943; Hellinger, 1909), and the Chi-square divergence. They showed that Jeffreys' general prior is the desired prior under each divergence measure that is in the interior in this class of divergence measures. On the boundary, namely for the Chi-square divergence, the prior turns out to be different from Jefferys' prior for some common families of distributions but still maintains the invariance property. Also, outside the boundary, Jeffreys' prior turns out to be the minimizer rather than maximizer of the divergence, and there does not exist any prior which maximizes the distance between the posterior and the prior. A more comprehensive set of results were later obtained by Liu, Chakrabarty, Samanta, Ghosh and Ghosh (2014) for one parameter family of distributions.

In this paper, we consider prior selection in the presence of nuisance parameters under the general divergence used by Ghosh, Mergel and Liu (2011). This is a generalization of previous work of Bernardo (1979). We characterize optimal priors for every member in this family of divergence measures by using the two-step procedure proposed by Bernardo (1979). Explicit expressions for the optimal priors under every divergence measure (except for the Chi-square divergence) are given. Specifically, for the Kullback-Leibler divergence, we get the same prior as found by Bernardo (1979). Under the Chi-square divergence, we have shown that the objective prior should be the solution to a set of partial differential equations. We also consider a special case when the parameter of interest is one dimensional. In this case, a closed form expression for the optimal prior is provided also under the Chi-square divergence.

The outline of the remaining sections is as follows. In Section 2 of this paper, we have provided a general scheme of deriving the asymptotic expansion of the expected general divergence. Section 3 is devoted to the derivation of optimal prior in the interior of the divergence class and non-existence of optimal priors outside the boundary of this class. Section 4 provides a characterization of optimal priors under the Chi-square divergence class followed by some examples. Some remarks are made in Section 5. The proof of the main result in Section 4 is deferred to the Appendix.

2. Derivation of Priors

Let $\boldsymbol{X}_n = (X_1, \ldots, X_n)$, where the X_i are independent and identically distributed with common pdf $f(x \,|\, \boldsymbol{\theta})$. Parameter vector $\boldsymbol{\theta}$ can be partitioned as,

$$\boldsymbol{\theta}^T = (\theta_1, \cdots, \theta_{d_1}, \theta_{d_1+1}, \cdots, \theta_d) = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T) \in \mathbf{R}^d,$$

where $\boldsymbol{\theta}_1$ are d_1-dimensional nuisance parameters and $\boldsymbol{\theta}_2$ are d_2-dimensional parameters of interest. Here, we assume the parameter space is a compact set in $\mathbf{R}^d$ and consider a prior $p(\boldsymbol{\theta}_2)$ which puts all its mass on a compact set in $\mathbf{R}^{d_2}$. One passes on to the limit eventually in many of the actual examples considered in the literature.

We apply the following two step procedure proposed by Bernardo (1979) to find the divergence priors for the parameters of interest.

First, for fixed $\boldsymbol{\theta}_2$, one assigns a conditional density $\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)$ to the nuisance parameters $\boldsymbol{\theta}_1$.

Second, let $p(\boldsymbol{\theta}_2)$ denote the marginal density function of $\boldsymbol{\theta}_2$. Then the divergence prior $p(\boldsymbol{\theta}_2)$ for the parameter of interest $\boldsymbol{\theta}_2$ is chosen by maximizing the asymptotic expected general divergence $R^\beta(p(\boldsymbol{\theta}_2))$ between the prior $p(\boldsymbol{\theta}_2)$ and the corresponding posterior, that is

$$R^\beta(p(\boldsymbol{\theta}_2)) = \frac{1 - \int \left[\int p^\beta(\boldsymbol{\theta}_2) p^{1-\beta}(\boldsymbol{\theta}_2|\, \boldsymbol{x}_n)\, d\boldsymbol{\theta}_2\right] m(\boldsymbol{x}_n)\mu(d\boldsymbol{x}_n)}{\beta(1-\beta)},$$

where $\mu(d\boldsymbol{x}_n)$ is a dominating measure and $m(\boldsymbol{x}_n)$ is the marginal density of $\boldsymbol{x}_n$

Note 1. The expected general divergence criterion as introduced by Renyi (1961), Amari (1982) and Cressie and Read (1984) is a family of divergences with index parameter β. When $\beta = 1/2$, this is the Bhattacharyya-Hellinger distance, and $\beta = -1$ amounts to the Chi-square distance. For $\beta = 0$ or 1, we need to interpret $R^\beta(p(\boldsymbol{\theta}))$ as its limiting value (when it exists). In particular,

$$R^0(p(\boldsymbol{\theta})) = \iint \left\{\log \frac{p(\boldsymbol{\theta}\,|\,\boldsymbol{x}_n)}{p(\boldsymbol{\theta})}\right\} p(\boldsymbol{\theta}\,|\,\boldsymbol{x}_n) m(\boldsymbol{x}_n)\, d\boldsymbol{\theta}\mu(d\boldsymbol{x}_n), \tag{1}$$

which is the KL divergence between the prior and the posterior considered in Bernardo (1979).

Note 2. In Step 1, for $\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)$, Bernardo recommends using the conditional Jeffreys' general rule prior which is proportional to $\sqrt{|I_{11}(\boldsymbol{\theta})|}$, where $I_{11}(\boldsymbol{\theta})$ is the part of Fisher information matrix $I(\boldsymbol{\theta})$ corresponding to the nuisance parametes.

Note 3. In Step 2, with the choice of $\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)$, we first find an asymptotic expansion of $R^\beta(p(\boldsymbol{\theta}_2))$ and then obtain the priors by maximizing that expansion.

Derivation of divergence priors in the presence of nuisance parameters is complicated. First, we give a general scheme of deriving the asymptotic expansion of the expected general divergence. Then, in Sections (3) and (4), by using the asymptotic expansion with different order of the remainder terms, we consider the prior selection for two cases when $\beta \neq -1$ and $\beta = -1$ separately.

By the relation $f_n(\boldsymbol{x}_n|\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_2) = p(\boldsymbol{\theta}_2|\boldsymbol{x}_n)m(\boldsymbol{x}_n)$, one can rewrite $R^\beta(p(\boldsymbol{\theta}_2))$ as

$$\begin{aligned} R^\beta(p(\boldsymbol{\theta}_2)) &= \frac{1 - \iint p^{\beta+1}(\boldsymbol{\theta}_2) p^{-\beta}(\boldsymbol{\theta}_2\,|\,\boldsymbol{x}_n) f_n(\boldsymbol{x}_n|\boldsymbol{\theta}_2)\, \mu(d\boldsymbol{x}_n)\, d\boldsymbol{\theta}_2}{\beta(1-\beta)} \\ &= \frac{1 - \int p^{\beta+1}(\boldsymbol{\theta}_2) E\left[p^{-\beta}(\boldsymbol{\theta}_2\,|\,\boldsymbol{X}_n)\Big|\,\boldsymbol{\theta}_2\right]\, d\boldsymbol{\theta}_2}{\beta(1-\beta)}, \end{aligned} \tag{2}$$

where $f_n(\boldsymbol{x}_n|\boldsymbol{\theta}_2)$ is the joint density function of $\boldsymbol{x}_n = (x_1, \cdots, x_n)$ given $\boldsymbol{\theta}_2$.

By using the shrinkage argument proposed by Ghosh (1994), one can find the asymptotic expansion to $E\left[p^{-\beta}(\boldsymbol{\theta}_2\,|\,\boldsymbol{X}_n)\Big|\,\boldsymbol{\theta}_2\right]$ and then find the asymptotic expansion to $R^\beta(p(\boldsymbol{\theta}_2))$. The shrinkage argument is discussed in details in Datta and Mukerjee (2004).

Here is the general scheme of deriving the asymptotic expansion:

- **Step 0:** Given the choice of $\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)$, for prior $p(\boldsymbol{\theta}_2)$, find posterior density $p(\boldsymbol{\theta}_2|\boldsymbol{x}_n)$ of $\boldsymbol{\theta}_2$ given $\boldsymbol{x}_n$:

$$p(\boldsymbol{\theta}_2|\boldsymbol{x}_n) = \int_{\boldsymbol{\theta}_1} \pi(\boldsymbol{\theta}|\boldsymbol{x}_n)d\boldsymbol{\theta}_1,$$

where $\pi(\boldsymbol{\theta}|\boldsymbol{x}_n) \propto p(\boldsymbol{\theta}_2)\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)\prod_{i=1}^n f(x_i|\boldsymbol{\theta})$.

- **Step 1:** Consider a proper prior density $\bar{p}(\boldsymbol{\theta}_2)$, such that the support of $\bar{p}(\boldsymbol{\theta}_2)$ is a compact rectangle in the parameter space and $\bar{p}(\boldsymbol{\theta}_2)$ vanishes on the boundary of the support while remaining positive in the interior. Consider the posterior density of $\boldsymbol{\theta}_2$ under $\bar{p}(\boldsymbol{\theta}_2)$, and obtain

$$G(\boldsymbol{x}_n) = \int p^{-\beta}(\boldsymbol{\theta}_2|\boldsymbol{x}_n)\bar{p}(\boldsymbol{\theta}_2|\boldsymbol{x}_n)d\boldsymbol{\theta}_2.$$

- **Step 2:** For $\boldsymbol{\theta}_2$ in the interior of the support of $\bar{p}(\boldsymbol{\theta}_2)$, compute $\lambda(\boldsymbol{\theta}_2)$ defined as

$$\begin{aligned}\lambda(\boldsymbol{\theta}_2) &= \int G(\boldsymbol{x}_n) f_n(\boldsymbol{x}_n|\boldsymbol{\theta}_2)d\boldsymbol{x}_n \\ &= \int G(\boldsymbol{x}_n)\left[\int \prod_{i=1}^n f(x_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\right]d\boldsymbol{x}_n \\ &= \int \lambda_0(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1,\end{aligned}$$

where $\lambda_0(\boldsymbol{\theta}) = \int G(\boldsymbol{x}_n)\prod_{i=1}^n f(x_i|\boldsymbol{\theta})d\boldsymbol{x}_n$.

- **Step 3:** Integrate $\lambda(\boldsymbol{\theta}_2)$ with respect to $\bar{p}(\boldsymbol{\theta}_2)$ and then allow $\bar{p}(\boldsymbol{\theta}_2)$ to converge weakly to the degenerate prior at the true $\boldsymbol{\theta}_2$, supposing that the true $\boldsymbol{\theta}_2$ is an interior point of the support of $\bar{p}(\boldsymbol{\theta}_2)$. This yields $E\left[p^{-\beta}(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)\middle|\boldsymbol{\theta}_2\right]$.

By using the above procedure and equation (2), one can get an asymptotic approximation to $R^\beta(p(\boldsymbol{\theta}_2))$. Furthermore, the divergence priors are obtained by maximizing the approximation.

In the next two sections, according to different values of β, we derive two approximations to $R^\beta(p(\boldsymbol{\theta}_2))$ and call them the first order approximation and the second order approximation respectively. For most of values of β, the derivation of divergence prior only requires the first order approximation and it will be addressed in section (3). In section (4), we will discuss the derivation of divergence prior when $\beta = -1$; that is the only case which needs the second order approximation.

3. Divergence Priors for $\beta \neq -1$

In this section, we consider the prior selection for general expected divergence with $\beta \neq -1$. To begin with, we derive the first order approximation to $E\left[p^{-\beta}(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)\middle|\boldsymbol{\theta}_2\right]$. Then, in view of (2), we get the first order approximation to $R^\beta(p(\boldsymbol{\theta}_2))$. Finally, we discuss the divergence priors according to different values of β such that $\beta \neq -1$.

First, by assuming the same regularity conditions as in Johnson (1970) and Bhattacharya and Ghosh (1978), one gets the following theorem which gives the first order expansion $E\left[p^{-\beta}(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)\Big| \boldsymbol{\theta}_2\right]$.

Theorem 1: For $\beta < 1$ and $\beta \neq -1$, $E\left[p^{-\beta}(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)|\boldsymbol{\theta}_2\right]$ can be expressed as:

$$E\left[p^{-\beta}(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)\Big| \boldsymbol{\theta}_2\right] = n^{-\frac{d_2\beta}{2}}\left[\int K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1 + o(n^{-1})\right], \tag{3}$$

where

$$K(\boldsymbol{\theta}) = \left|I^{22}\right|^{\beta/2}(2\pi)^{d_2\beta/2}(1-\beta)^{-d_2/2},$$
$$I^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix}.$$

Proof of Theorem 1:
The proof uses the shrinkage argument as mentioned before.

Let $\boldsymbol{h}^T = (h_1, \cdots, h_{d_1}, h_{d_1+1}, \cdots h_d) = (\boldsymbol{h}_1^T, \boldsymbol{h}_2^T) = \sqrt{n}(\boldsymbol{\theta}_1^T - \hat{\boldsymbol{\theta}}_1^T, \boldsymbol{\theta}_2^T - \hat{\boldsymbol{\theta}}_2^T)$, where $\hat{\boldsymbol{\theta}}^T = (\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T)$ is MLE of $\boldsymbol{\theta}$. For prior $p(\boldsymbol{\theta})$, from Datta and Mukerjee (1994), one gets the corresponding posterior density

$$p(\boldsymbol{h}|\boldsymbol{x}_n) = \phi_d(\boldsymbol{h}, C^{-1})\left[1 + n^{-\frac{1}{2}}\left\{R_1(\boldsymbol{h}) + \frac{1}{6}R_3(\boldsymbol{h})\right\}\right] + o(n^{-1}), \tag{4}$$

where C is the observed Fisher information matrix, $\phi_d(\boldsymbol{h}, C^{-1})$ is the $d-$variate normal density with the null mean vector and dispersion matrix C^{-1},

$$R_1(\boldsymbol{h}) = \sum_{j=1}^{d}\hat{p}_j h_j/\hat{p}, \quad R_3(\boldsymbol{h}) = \sum_{j=1}^{d}\sum_{r=1}^{d}\sum_{s=1}^{d} a_{jrs}h_j h_r h_s,$$

$$a_{jrs} = \frac{1}{n}\sum_{u=1}^{d}\frac{\partial \log f(X_u|\boldsymbol{\theta})}{\partial\theta_j\theta_r\theta_s} \text{ and } \hat{p} = p(\hat{\boldsymbol{\theta}}), \quad \hat{p}_j = \left.\frac{\partial p(\boldsymbol{\theta})}{\partial\theta_j}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Let $N_{d_2}(\boldsymbol{h}_2|\boldsymbol{\mu}, \Sigma)$ denote the density function of multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and dispersion matrix Σ. Also, corresponding to the partition of the parameter vector $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$, we partition the matrix C^{-1} as:

$$C^{-1} = \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix}.$$

It is easy to establish the relation

$$\phi_d(\boldsymbol{h}, C^{-1}) = N_{d_2}(\boldsymbol{h}_2|0, C^{22}) \times f_{d_1}(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2),$$

where $f_{d_1}(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) = N_{d_1}(\boldsymbol{h}_1|C^{12}[C^{22}]^{-1}\boldsymbol{h}_2, C^{11} - C^{12}[C^{22}]^{-1}C^{21})$.

In **Step 0**, for prior $p(\boldsymbol{\theta}_2)$, one gets

$$p(\boldsymbol{h}_2|\boldsymbol{x}_n) = N_{d_2}(\boldsymbol{h}_2|0, C^{22}) \left[1 + n^{-\frac{1}{2}} \left\{L_1(\boldsymbol{h}_2) + \frac{1}{6} L_3(\boldsymbol{h}_2)\right\}\right] + o(n^{-1}), \tag{5}$$

where

$$L_1(\boldsymbol{h}_2) = \int R_1(\boldsymbol{h}) f_{d_1}(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1, \quad L_3(\boldsymbol{h}_2) = \int R_3(\boldsymbol{h}) f_{d_1}(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1.$$

Step 1. We find an asymptotic expansion for

$$G(\boldsymbol{x}_n) = \int p^{-\beta}(\boldsymbol{\theta}_2|\boldsymbol{x}_n) \bar{p}(\boldsymbol{\theta}_2|\boldsymbol{x}_n) d\boldsymbol{\theta}_2.$$

With the general expansion

$$\left(\frac{1}{b_1 + \frac{b_2}{\sqrt{n}} + \frac{b_3}{n} + o(n^{-1})}\right)^{\beta} = b_1^{-\beta} \left(1 - \beta \frac{b_2}{b_1 \sqrt{n}} + \frac{\beta}{n} \left(\frac{\beta + 1}{2} \frac{b_2^2}{b_1^2} - \frac{b_3}{b_1}\right)\right) + o(n^{-1}),$$

one gets

$$p^{-\beta}(\boldsymbol{h}|\boldsymbol{x}_n) = N_{d_2}^{-\beta}(\boldsymbol{h}_2|0, C^{22}) \left[1 - \beta n^{-\frac{1}{2}} \left\{L_1(\boldsymbol{h}_2) + \frac{1}{6} L_3(\boldsymbol{h}_2)\right\}\right] + o(n^{-1}). \tag{6}$$

Using (5) and (6), for any arbitrary thrice differentiable prior $\bar{p}(\boldsymbol{\theta}_2)$ vanishing outside a compact set, one gets

$$\begin{aligned} p^{-\beta}(\boldsymbol{h}_2|\boldsymbol{x}_n) \bar{p}(\boldsymbol{h}_2|\boldsymbol{x}_n) = N_{d_2}^{1-\beta}(\boldsymbol{h}_2|0, C^{22}) \Big[1 + n^{-\frac{1}{2}} \Big\{\bar{L}_1(\boldsymbol{h}_2) + \frac{1}{6} L_3(\boldsymbol{h}_2) \\ - \beta L_1(\boldsymbol{h}_2) - \frac{\beta}{6} L_3(\boldsymbol{h}_2)\Big\}\Big] + o(n^{-1}), \end{aligned} \tag{7}$$

where

$$\bar{L}_1(\boldsymbol{h}_2) = \int \bar{R}_1(\boldsymbol{h}) f_{d_1}(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1.$$

Here are some observations: (i) $N_{d_2}^{1-\beta}(\boldsymbol{h}_2|0, C^{22}) = K(\hat{\boldsymbol{\theta}}) \times N_{d_2}(\boldsymbol{h}_2|0, \frac{C^{22}}{1-\beta})$,
(ii) $f_{d_1}(\boldsymbol{h}_1|\boldsymbol{h}_2) \times N_{d_2}(\boldsymbol{h}_2|0, \frac{C^{22}}{1-\beta}) = N_d(\boldsymbol{h}|0, Q)$, where

$$K(\hat{\boldsymbol{\theta}}) = (2\pi)^{\frac{d_2 \beta}{2}} |C^{22}|^{\frac{\beta}{2}} (1 - \beta)^{-\frac{d_2}{2}},$$

$$Q = (q^{jr})_{d \times d} = \begin{pmatrix} C^{11} + \frac{\beta}{1-\beta} C^{12} [C^{22}]^{-1} C^{21} & C^{12}/(1-\beta) \\ C^{21}/(1-\beta) & C^{22}/(1-\beta) \end{pmatrix}.$$

With the above observations and the relation $\boldsymbol{\theta} = \boldsymbol{h}/\sqrt{n} + \hat{\boldsymbol{\theta}}$, and noting that

$$\int \bar{L}_1(\boldsymbol{h}_2) d\boldsymbol{h}_2 = \int L_1(\boldsymbol{h}_2) d\boldsymbol{h}_2 = \int L_3(\boldsymbol{h}_2) d\boldsymbol{h}_2 = 0,$$

one gets

$$\begin{aligned} G(\boldsymbol{x}_n) &= \int p^{-\beta}(\boldsymbol{\theta}_2|\boldsymbol{x}_n)\bar{p}(\boldsymbol{\theta}_2|\boldsymbol{x}_n)d\boldsymbol{\theta}_2 \\ &= n^{-\frac{d_2\beta}{2}} \int p^{-\beta}(\boldsymbol{h}_2|\boldsymbol{x}_n)\bar{p}(\boldsymbol{h}_2|\boldsymbol{x}_n)d\boldsymbol{h}_2 = n^{-\frac{d_2\beta}{2}}\left[K(\hat{\boldsymbol{\theta}}) + o_P(n^{-1})\right] \end{aligned} \tag{8}$$

As shown above, $G(\boldsymbol{x}_n)$ can be written as a function of $\hat{\boldsymbol{\theta}}$. Also $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = o_p(n^{-1})$ $(P_{\boldsymbol{\theta}})$. Therefore, by using Taylor expansion, one gets

$$\lambda_0(\boldsymbol{\theta}) = \int G(\boldsymbol{x}_n)\prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})d\boldsymbol{x}_n = n^{-\frac{d_2\beta}{2}}\left[K(\boldsymbol{\theta}) + o(n^{-1})\right], \tag{9}$$

where $K(\boldsymbol{\theta}) = (2\pi)^{\frac{d_2\beta}{2}}|I^{22}|^{\frac{\beta}{2}}(1-\beta)^{-\frac{d_2}{2}}$. Moreover, at the end of **Step 2**, one gets

$$\begin{aligned} \lambda(\boldsymbol{\theta}_2) &= \int \lambda_0(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1 \\ &= n^{-\frac{d_2\beta}{2}}\left[\int K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1 + o(n^{-1})\right]. \end{aligned} \tag{10}$$

In **Step 3**, integrating $\lambda(\boldsymbol{\theta}_2)$ with respect to $\bar{p}(\boldsymbol{\theta}_2)$ and allowing $\bar{p}(\boldsymbol{\theta}_2)$ weakly converge to the degenerate density of true $\boldsymbol{\theta}_2$, we obtain the final asymptotic expansion for $E\left[p^{-\beta}(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)\Big|\boldsymbol{\theta}_2\right]$ as

$$\begin{aligned} E_{\boldsymbol{\theta}_2}\left[p^{-\beta}(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)\right] &= n^{-\frac{d_2\beta}{2}}\left[\int K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1 + o(n^{-1})\right] \\ &= n^{-\frac{d_2\beta}{2}}\left[(2\pi)^{\frac{d_2\beta}{2}}(1-\beta)^{-\frac{d_2}{2}}\int |I^{22}|^{\frac{\beta}{2}}\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1 + o(n^{-1})\right]. \end{aligned} \tag{11}$$

This proves the theorem.

When $\beta < -1$ and $\beta \neq -1$, we can obtain the divergence priors by maximizing the first order approximation to the general expected divergence $R^{\beta}(p(\boldsymbol{\theta}_2))$. The approximation is derived by neglecting the $o(n^{-1})$ term in Theorem 1. That is:

$$R^{\beta}(p(\boldsymbol{\theta}_2)) \approx \frac{1}{\beta(1-\beta)}\left[1 - \left(\frac{2\pi}{n}\right)^{\frac{d_2\beta}{2}}(1-\beta)^{-\frac{d_2}{2}}\int\left[\frac{\phi(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\right]^{-\beta} p(\boldsymbol{\theta}_2)d\boldsymbol{\theta}_2\right], \tag{12}$$

where

$$\phi(\boldsymbol{\theta}_2) = \left[\int \left|I^{22}(\boldsymbol{\theta})\right|^{\frac{\beta}{2}}\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\right]^{-\frac{1}{\beta}}.$$

One may think that the divergence priors should be different as β takes on different values. Amazingly, it turns out that, in most cases, one gets the same prior. Here are the results for different values of β $(\beta \neq -1)$.

CASE I. First consider the case $0 < \beta < 1$. Since then $\beta(1-\beta) > 0$, the problem of prior selection reduces to minimization of

$$\int \left[\frac{\phi(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\right]^{-\beta} p(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2.$$

Noting that $u^{-\beta}$ is a convex function of $u(> 0)$ when $\beta(1-\beta) > 0$, by Jensen's inequality, one gets

$$\begin{aligned}&\int \left[\frac{\phi(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\right]^{-\beta} p(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2 \\ &\geq \left\{\int \left[\frac{\phi(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\right] p(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2\right\}^{-\beta} = \left\{\int \phi(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2\right\}^{-\beta}\end{aligned}$$

with equality if and only if $p(\boldsymbol{\theta}_2) \propto \phi(\boldsymbol{\theta}_2)$.

CASE II. Similarly, when $-1 < \beta < 0$, $\beta(1-\beta) < 0$ and now the problem reduces to maximization of

$$\int \left[\frac{\phi(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\right]^{-\beta} p(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2.$$

Noting that $u^{-\beta}$ is a concave function of $u(> 0)$ when $-1 < \beta < 0$, again by Jensen's inequality, one gets

$$\begin{aligned}&\int \left[\frac{\phi(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\right]^{-\beta} p(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2 \\ &\leq \left\{\int \left[\frac{\phi(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\right] p(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2\right\}^{-\beta} = \left\{\int \phi(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2\right\}^{-\beta}\end{aligned}$$

with equality if and only if $p(\boldsymbol{\theta}_2) \propto \phi(\boldsymbol{\theta}_2)$ which is the same prior developed in the previous case.

CASE III. When $\beta \longrightarrow 0$, using either Theorem 1 or alternatively from Bernardo, one gets the first order approximation of the general expected divergence $R^0(p(\boldsymbol{\theta}_2))$:

$$R^0(p(\boldsymbol{\theta}_2)) \approx K_n + \int p(\boldsymbol{\theta}_2) \log \frac{\phi(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} d\boldsymbol{\theta}_2,$$

where K_n is a constant depending on n. Then, from the property of the Kullback-Leibler distance, $R^0(p(\boldsymbol{\theta}_2))$ is maximized up to first order of approximation by $p(\boldsymbol{\theta}_2) \propto \phi(\boldsymbol{\theta}_2)$.

From the above three cases, we can easily draw the conclusion that when $|\beta| < 1$, the desired divergence prior is proportional to $\phi(\boldsymbol{\theta}_2)$.

CASE IV. Next for $\beta < -1$, writing $\beta = -\lambda$, one can rewrite the first order approximation of the general divergence $R^\beta(p(\boldsymbol{\theta}_2))$ as :

$$R^{\beta}(p(\boldsymbol{\theta}_2)) = \frac{\left(\frac{2\pi}{n}\right)^{-d_2\lambda/2}(1+\lambda)^{-\frac{d_2}{2}}\int\left(\frac{\phi(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\right)^{\lambda}p(\boldsymbol{\theta}_2)\,d\boldsymbol{\theta}_2 - 1}{\lambda(1+\lambda)}, \quad \lambda > 1. \tag{13}$$

Hence it suffices to maximize

$$\int \{\phi(\boldsymbol{\theta}_2)/p(\boldsymbol{\theta}_2)\}^{\lambda}\,p(\boldsymbol{\theta}_2)\,d\boldsymbol{\theta}_2$$

subject to $\int p(\boldsymbol{\theta}_2)\,d\boldsymbol{\theta}_2 = 1$. Again, by Jensen's inequality,

$$\int \{\phi(\boldsymbol{\theta}_2)/p(\boldsymbol{\theta}_2)\}^{\lambda}\,p(\boldsymbol{\theta}_2)\,d\boldsymbol{\theta}_2 \geq \left[\int \{\phi(\boldsymbol{\theta}_2)/p(\boldsymbol{\theta}_2)\}\,p(\boldsymbol{\theta}_2)\,d\boldsymbol{\theta}_2\right]^{\lambda} = \left(\int \phi(\boldsymbol{\theta}_2)\,d\boldsymbol{\theta}_2\right)^{-\beta}$$

since $\lambda > 1$, equality holding if and only if

$$p(\boldsymbol{\theta}_2) \propto \phi(\boldsymbol{\theta}_2).$$

Thus in this case the prior $p(\boldsymbol{\theta}_2) \propto \phi(\boldsymbol{\theta}_2)$ is the minimizer rather then the maximizer of $R^{\beta}(p(\boldsymbol{\theta}_2))$. Also there is no maximizing prior in this case. In fact, one can use similar argument in the previous section to show that

$$\sup_{p}\int \phi(\boldsymbol{\theta}_2)^{\lambda}p^{1-\lambda}(\boldsymbol{\theta}_2)d\boldsymbol{\theta}_2 = +\infty$$

3. Divergence Priors for $\beta = -1$

It remains to consider the case $\beta = -1$, the Chi-square distance as considered in Clarke and Sun (1997) for the one parameter exponential family and in Ghosh, Mergel and Liu (2011) for the general one-parameter family of distributions. Here $p^{\beta+1}(\boldsymbol{\theta}_2) = 1$ so that the first order term appearing in Theorem 1 will not suffice in finding the prior $p(\boldsymbol{\theta}_2)$. We can mimic Theorem 1 to get the second order expansion to $E\left[p(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)|\boldsymbol{\theta}_2\right]$. Here is the new Theorem:

Theorem 2: For $\beta = -1$, $E\left[p(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)|\boldsymbol{\theta}_2\right]$ can be expressed as:

$$
\begin{aligned}
E\left[p(\boldsymbol{\theta}_2 \mid \boldsymbol{X}_n)|\,\boldsymbol{\theta}_2\right] &= n^{\frac{d_2}{2}}\Bigg[\int K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1 + \\
&\frac{1}{n}\Bigg\{\sum_{j=d_1+1}^{d}\sum_{r=1}^{d}\Big[\int\Big(q^{ojr}-\frac{1}{2}I^{jr}\Big)K(\boldsymbol{\theta})\pi_r(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\Big]\frac{p_j(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
&\quad -\frac{1}{2}\sum_{j=d_1+1}^{d}\sum_{r=1}^{d}\frac{\partial}{\partial\theta_r}\Big(\int I^{jr}K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\Big)\frac{p_j(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
&\quad +\frac{1}{6}\sum_{1\le j,r,s\le d}\sum_{u=d_1+1}^{d}\Big[\int A_{jrs}(q^{ojr}q^{osu}+q^{oju}q^{ors}+q^{ojs}q^{oru})K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\Big]\frac{p_u(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
&\quad -\frac{1}{6}\sum_{1\le j,r,s\le d}\sum_{u=d_1+1}^{d}\Big[\int A_{jrs}(I^{jr}I^{su}+I^{ju}I^{rs}+I^{js}I^{ru})K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\Big]\frac{p_u(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \qquad (14) \\
&\quad +\frac{1}{6}\sum_{1\le j,r,s\le d}\sum_{u=d_1+1}^{d}\Big[\int A_{jrs}k_{jrsu}(\boldsymbol{\theta})K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\Big]\frac{p_u(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
&\quad -\sum_{j=d_1+1}^{d}\sum_{r=d_1+1}^{d}\Big[\int\Big\{I^{jr}-\frac{q^{ojr}}{2}\Big\}K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\Big]\frac{p_{jr}(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
&\quad +\frac{1}{2}\sum_{j=d_1+1}^{d}\sum_{r=d_1+1}^{d}\Big[\int I^{jr}K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\Big]\frac{p_j(\boldsymbol{\theta}_2)p_r(\boldsymbol{\theta}_2)}{p^2(\boldsymbol{\theta}_2)} \\
&\quad +S(\boldsymbol{\theta}_2)\Bigg\}+o(n^{-1})\Bigg],
\end{aligned}
$$

where

$$
K(\boldsymbol{\theta}) = \left|I^{22}\right|^{-1/2}(2\pi)^{-d_2/2}2^{-d_2/2}, \quad A_{jrs} = E\left[\frac{\partial^3 \log f(X|\boldsymbol{\theta})}{\partial\theta_j\partial\theta_r\partial\theta_s}\right]
$$

$$
(q^{ojr})_{d\times d} = \begin{pmatrix} I^{11}-\frac{1}{2}I^{12}[I^{22}]^{-1}I^{21} & I^{12}/2 \\ I^{21}/2 & I^{22}/2 \end{pmatrix},
$$

I^{jr} is the jr^{th} element of the Fisher information matrix, $k_{jrsu}(\boldsymbol{\theta})$ involves $p(\boldsymbol{\theta}_2)$ and its derivatives, but $S(\boldsymbol{\theta}_2)$ is only a function of $\boldsymbol{\theta}_2$.

The proof of Theorem 2 is long and involved, and is omitted. The details are available from the authors.

Since $\beta = -1$ so that $\beta(1-\beta) = -2$, neglecting all terms which do not involve $p(\boldsymbol{\theta}_2)$ or its derivatives and using the relation $\frac{p_{jr}(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} = \frac{\partial}{\partial\theta_r}\left(\frac{p_j(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\right) + \frac{p_j(\boldsymbol{\theta}_2)p_r(\boldsymbol{\theta}_2)}{p^2(\boldsymbol{\theta}_2)}$, it suffices to

maximize up to the second order approximation,

$$
\begin{aligned}
R^{(-1)}(p(\boldsymbol{\theta}_2)) \approx \frac{1}{2} n^{\frac{d_2}{2}} \int \Bigg[\int K(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 + \\
\frac{1}{n} \Bigg\{ \sum_{j=d_1+1}^{d} \sum_{r=1}^{d} \Big[\int \Big(q^{ojr} - \frac{I^{jr}}{2} \Big) K(\boldsymbol{\theta}) \pi_r(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \Big] \frac{p_j(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
- \frac{1}{2} \sum_{j=d_1+1}^{d} \sum_{r=1}^{d} \frac{\partial}{\partial \theta_r} \Big(\int I^{jr} K(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \Big) \frac{p_j(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
+ \frac{1}{6} \sum_{1 \le j,r,s \le d} \sum_{u=d_1+1}^{d} \Big[\int A_{jrs} (q^{ojr} q^{osu} + q^{oju} q^{ors} + q^{ojs} q^{oru}) K(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \Big] \frac{p_u(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
- \frac{1}{6} \sum_{1 \le j,r,s \le d} \sum_{u=d_1+1}^{d} \Big[\int A_{jrs} (I^{jr} I^{su} + I^{ju} I^{rs} + I^{js} I^{ru}) K(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \Big] \frac{p_u(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
+ \frac{1}{6} \sum_{1 \le j,r,s \le d} \sum_{u=d_1+1}^{d} \Big[\int A_{jrs} k_{jrsu}(\boldsymbol{\theta}) K(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \Big] \frac{p_u(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \\
+ \sum_{j=d_1+1}^{d} \sum_{r=d_1+1}^{d} \Big[\int \Big\{ \frac{(q^{ojr} - I^{jr})}{2} - \frac{I^{jr}}{2} \Big\} K(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \Big] \frac{\partial p_j(\boldsymbol{\theta}_2)/p(\boldsymbol{\theta}_2)}{\partial \theta_r} \\
+ \sum_{j=d_1+1}^{d} \sum_{r=d_1+1}^{d} \Big[\int \Big\{ \frac{1}{2} (q^{ojr} - I^{jr}) \Big\} K(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \Big] \frac{p_j(\boldsymbol{\theta}_2) p_r(\boldsymbol{\theta}_2)}{p^2(\boldsymbol{\theta}_2)} \Bigg\} \Bigg] d\boldsymbol{\theta}_2.
\end{aligned}
\tag{15}
$$

Let

$$
\begin{aligned}
\boldsymbol{y}(\boldsymbol{\theta}_2) &= (\boldsymbol{y}_{d_1+1}(\boldsymbol{\theta}_2), \cdots, \boldsymbol{y}_d(\boldsymbol{\theta}_2)) = \left(\frac{p_{d_1+1}(\boldsymbol{\theta})}{p(\boldsymbol{\theta}_2)}, \cdots, \frac{p_d(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)} \right) \\
\nabla \boldsymbol{y}(\boldsymbol{\theta}_2) &= \left(\frac{\partial \boldsymbol{y}_{d_1+1}(\boldsymbol{\theta}_2)}{\partial \theta_{d_1+1}}, \cdots, \frac{\partial \boldsymbol{y}_{d_1+1}(\boldsymbol{\theta}_2)}{\partial \theta_d}, \cdots, \frac{\partial \boldsymbol{y}_d(\boldsymbol{\theta}_2)}{\partial \theta_{d_1+1}}, \cdots, \frac{\partial \boldsymbol{y}_d(\boldsymbol{\theta}_2)}{\partial \theta_d} \right).
\end{aligned}
$$

Note that (15) can be expressed as

$$
\int F(\boldsymbol{\theta}_2, \boldsymbol{y}(\boldsymbol{\theta}_2), \nabla \boldsymbol{y}(\boldsymbol{\theta}_2)) d\boldsymbol{\theta}_2, \tag{16}
$$

so we need find $\boldsymbol{y}(\boldsymbol{\theta}_2)$ to maximize the above integral. From Giaquinta (1983), the maximizer should satisfy the Euler-Lagrange equations:

$$
\frac{\partial F}{\partial \boldsymbol{y}_i(\boldsymbol{\theta}_2)} - \sum_{j=d_1+1}^{d} \frac{\partial}{\partial \theta_j} \left(\frac{\partial F}{\partial (\partial \boldsymbol{y}_i / \partial \theta_j)} \right) = 0 \quad i = d_1 + 1, \cdots, d. \tag{17}
$$

Equivalently, the Euler-Lagrange equations are:

$$\begin{aligned}
&\sum_{r=d_1+1}^{d}\left[\int I^{ir}K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\right]\frac{p_r(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_2)}\\
=&-\frac{1}{2}\sum_{1\le j,r,s\le d}\int A_{jrs}m^{ojr}I^{si}K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1\\
&+\frac{1}{2}\sum_{r=d_1+1}^{d}\frac{\partial}{\partial\theta_r}\int I^{ir}K(\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)d\boldsymbol{\theta}_1,\quad i=d_1+1,\cdots,d.
\end{aligned}\tag{18}$$

Here

$$m^{ojr}=\begin{cases} n^{ojr} & \text{when } j,r\in H_1\\ I^{jr} & \text{other}\end{cases},$$

where n^{ojr} is the jr^{th} element of $I^{12}(I^{22})^{-1}I^{21}$. H_1 is the set of indexes of nuisance parameters.

By solving these partial differential equations, one gets the divergence priors. Usually, with multi-dimensional parameters of interest, these equations are so complicated that it is impossible to give a general solution and sometimes, there is no solution to these equations.

In the following, we focus on a special case when the parameter of interest is one-dimensional. In this case, instead of several partial differential equations, we only need to solve one differential equation and easily get a general form of the divergence priors.

When the parameter of interest $\boldsymbol{\theta}_2$ is one dimensional, that is $\boldsymbol{\theta}_2=\theta_d$, then the Euler-Lagrange equation becomes:

$$\begin{aligned}
&\left[\int(I^{dd})^{\frac{1}{2}}\pi(\boldsymbol{\theta}_1|\theta_d)d\boldsymbol{\theta}_1\right]\frac{p'(\theta_d)}{p(\theta_d)}\\
=&-\frac{1}{2}\sum_{1\le j,r,s\le d}\int A_{jrs}m^{ojr}I^{sd}(I^{dd})^{-\frac{1}{2}}\pi(\boldsymbol{\theta}_1|\theta_d)d\boldsymbol{\theta}_1\\
&+\frac{1}{2}\frac{\partial}{\partial\theta_d}\int(I^{dd})^{\frac{1}{2}}\pi(\boldsymbol{\theta}_1|\theta_d)d\boldsymbol{\theta}_1.
\end{aligned}\tag{19}$$

By solving (19), one gets the divergence prior $p(\theta_d)$ which is proportional to

$$Q(\theta_d)^{\frac{1}{2}}\times\exp\left[\int -T(\theta_d)d\theta_d\right],\tag{20}$$

where

$$Q(\theta_d)=\int(I^{dd})^{\frac{1}{2}}\pi(\boldsymbol{\theta}_1|\theta_d)d\boldsymbol{\theta}_1, T(\theta_d)=\frac{\frac{1}{2}\sum_{1\le j,r,s\le d}\int A_{jrs}m^{ojr}I^{sd}(I^{dd})^{-\frac{1}{2}}\pi(\boldsymbol{\theta}_1|\theta_d)d\boldsymbol{\theta}_1}{\int(I^{dd})^{\frac{1}{2}}\pi(\boldsymbol{\theta}_1|\theta_d)d\boldsymbol{\theta}_1}.\tag{21}$$

Here are several examples to illustrate how to find divergence priors for one dimensional parameter of interest.

Example 1: Consider general symmetric location-scale family of distributions with probability density function $f(x|\mu,\sigma) = \frac{1}{\sigma}p(\frac{x-\mu}{\sigma})$ where $p(x) = p(-x)$. Writing $h(x) = \log p(x)$ and noting that $h'(x) = -h'(-x)$, $h''(x) = h''(-x)$ and $h'''(x) = -h'''(-x)$, one gets

$$E\left[\frac{\partial^2 \log f}{\partial \mu^2}\middle|\mu,\sigma\right] = -\sigma^{-2}\int h''(x)p(x)dx, \quad E\left[\frac{\partial^2 \log f}{\partial \mu \partial \sigma}\middle|\mu,\sigma\right] = 0,$$

$$E\left[\frac{\partial^2 \log f}{\partial \sigma^2}\middle|\mu,\sigma\right] = -\sigma^{-2}\left[1 + 2\int xh'(x)p(x)dx + \int x^2h''(x)p(x)dx\right],$$

$$E\left[\frac{\partial^3 \log f}{\partial \mu^3}\middle|\mu,\sigma\right] = E\left[\frac{\partial^3 \log f}{\partial \mu \partial \sigma^2}\middle|\mu,\sigma\right] = 0,$$

$$E\left[\frac{\partial^3 \log f}{\partial \mu^2 \partial \sigma}\middle|\mu,\sigma\right] = -\sigma^{-3}\left[2\int h''(x)p(x)dx + \int xh'''(x)p(x)dx\right]$$

and

$$E\left[\frac{\partial^3 \log f}{\partial \sigma^3}\middle|\mu,\sigma\right] = -\sigma^{-3}\left[2 + 6\int xh'(x)p(x)dx + 6\int x^2h''(x)p(x)dx + \int x^3h'''(x)p(x)dx\right].$$

- If μ is parameter of interest, from (20), the prior should have the following form:

$$p(\mu) \propto \left[\int \sigma\pi(\sigma|\mu)d\sigma\right]^{\frac{1}{2}}.$$

- If σ is parameter of interest, from (20), the prior should have the following form:

$$p(\sigma) \propto \sigma^{\frac{1}{2}+\frac{1+3\int xh'(x)p(x)dx+3\int x^2h''(x)p(x)dx+\frac{1}{2}\int x^3h'''(x)p(x)dx}{1+2\int xh'(x)p(x)dx+\int x^2h''(x)p(x)dx}}.$$

As special case, recall that for the $N(\mu,\sigma^2)$ distribution, $h'(x) = -x$, $h''(x) = -1$ and $h'''(x) = 0$. Hence $p(\sigma) \propto \sigma^3$.

Example 2: Consider the proper dispersion model introduced by Jorgensen (1997). The probability density function of this model is given by

$$f(x|\mu,\lambda) = a(\lambda)c(x)\exp[\lambda t(x,\mu)],$$

where μ and λ are two parameters. Now observe that

$$\frac{\partial \log f}{\partial \mu} = \lambda\frac{\partial t}{\partial \mu}, \quad \frac{\partial \log f}{\partial \lambda} = u(\lambda) + t(x,\mu),$$

where $u(\lambda) = a'(\lambda)/a(\lambda)$. Accordingly,

$$\frac{\partial^2 \log f}{\partial \mu^2} = \lambda\frac{\partial^2 t}{\partial \mu^2}, \quad \frac{\partial^2 \log f}{\partial \mu \partial \lambda} = \frac{\partial t}{\partial \mu}, \quad \frac{\partial^2 \log f}{\partial \lambda^2} = u'(\lambda).$$

Since $E\left(\frac{\partial t}{\partial \mu}|\mu,\lambda\right)=0$, the Fisher information matrix is

$$I(\mu,\lambda)=\ \text{Diag}\ [I_{\mu\mu}(\mu,\lambda),-u'(\lambda)],$$

where $I_{\mu\mu}(\mu,\lambda)=-\lambda E\left[\frac{\partial^2 t}{\partial \mu^2}|\mu,\lambda\right]$. Thus μ and λ are orthogonal in the sense of Cox and Reid (1987). Further, $\frac{\partial^3 \log f}{\partial \mu^3}=\lambda\frac{\partial^3 t}{\partial \mu^3}$, $\frac{\partial^3 \log f}{\partial \mu^2\partial\lambda}=\frac{\partial^2 t}{\partial \mu^2}$, $\frac{\partial^3 \log f}{\partial \mu\partial\lambda^2}=0$, $\frac{\partial^3 \log f}{\partial \lambda^3}=u''(\lambda)$.

- When μ is the parameter of interest, it is easy to get that $I^{dd}=I^{-1}_{\mu\mu}(\mu,\lambda)$, $A_{ddd}=E\left[\frac{\partial^3 \log f}{\partial \mu^3}|\mu,\lambda\right]=\lambda E\left[\frac{\partial^3 t}{\partial \mu^3}|\mu,\lambda\right]$ and all the $m^{jr}=0$ except $m^{dd}=I^{dd}$. Hence, by using (21), one gets

$$Q(\mu)=\int I^{-1/2}_{\mu\mu}(\mu,\lambda)\pi(\lambda|\mu)d\lambda,\ \ T(\mu)=\frac{\int \lambda E\left[\frac{\partial^3 t}{\partial \mu^3}|\mu,\lambda\right] I^{-3/2}_{\mu\mu}(\mu,\lambda)}{2Q(\mu)} \tag{22}$$

Then, by using (20), we can get the prior for μ.

- When λ is the parameter of interest, one gets that $I^{dd}=-\frac{1}{u'(\lambda)}$, $A_{ddd}=E\left[\frac{\partial^3 \log f}{\partial \lambda^3}|\mu,\lambda\right]=u''(\lambda)$ and all the $m^{jr}=0$ except $m^{dd}=I^{dd}$. By using (21), it is easy to check that

$$Q(\lambda)=\left(-\frac{1}{u'(\lambda)}\right)^{1/2},\ \ T(\lambda)=-\frac{u''(\lambda)}{2u'(\lambda)}. \tag{23}$$

Therefore, the prior for λ is proportional to

$$\left(-\frac{1}{u'(\lambda)}\right)^{1/4}\times\exp\left[\int\frac{u''(\lambda)}{2u'(\lambda)}d\lambda\right]=(-u'(\lambda))^{1/4} \tag{24}$$

Now, we consider several special cases of the above general result.

Example 3: Consider the two-parameter Gamma probability density function

$$f(x|\mu,\lambda)=\exp\left(-\frac{\lambda}{\mu}x\right)\frac{\lambda^\lambda x^{\lambda-1}}{\mu^\lambda}\frac{1}{\Gamma(\lambda)}.$$

Here,

$$a(\lambda)=\frac{\lambda^\lambda}{\Gamma(\lambda)},\quad t(x,\mu)=-\frac{x}{\mu}+\log\frac{x}{\mu}.$$

- When μ is the parameter of interest, it is easy to get that

$$I_{\mu\mu}=-\lambda E\left[\frac{\partial^2 t}{\partial \mu^2}|\mu,\lambda\right]=-\frac{1}{\mu^2},\ \ E\left[\frac{\partial^3 t}{\partial \mu^3}|\mu,\lambda\right]=\frac{4}{\mu^3}.$$

Then, by using (22), one gets

$$Q(\mu)=\mu\int\lambda^{-\frac{1}{2}}\pi(\lambda|\mu)d\lambda,\ \ T(\mu)=\frac{2}{\mu}.$$

Therefore, from (20), the prior should have the following form:

$$p(\mu) \propto \mu^{-\frac{3}{2}} \left[\int \lambda^{-\frac{1}{2}} \pi(\lambda|\mu) d\lambda \right]^{1/2}.$$

When $\pi(\lambda|\mu)$ is independent of μ (for example, conditional Jeffreys' general rule prior suggested by Bernardo), the divergence prior for μ is proportional to $\mu^{-\frac{3}{2}}$.

- When λ is the parameter of interest, one gets that $u(\lambda) = 1 + \log\lambda - \frac{d}{d\lambda}\log\Gamma(\lambda)$. Then, from (24), the prior should have the following form:

$$p(\lambda) \propto [-u'(\lambda)]^{\frac{1}{4}}.$$

Example 4: Consider the Inverse Gaussian distribution with probability density function

$$f(x|\mu,\lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right].$$

Here,

$$a(\lambda) = \lambda^{1/2}, \quad t(x,\mu) = -\frac{(x-\mu)^2}{2\mu^2 x}.$$

- When μ is the parameter of interest, it is easy to get that

$$I_{\mu\mu} = -\lambda E\left[\frac{\partial^2 t}{\partial\mu^2}|\mu,\lambda\right] = \frac{\lambda}{\mu^3}, \quad E\left[\frac{\partial^3 t}{\partial\mu^3}|\mu,\lambda\right] = \frac{6}{\mu^4}.$$

Then, by using (22), one gets

$$Q(\mu) = \mu^{3/2} \int \lambda^{-\frac{1}{2}} \pi(\lambda|\mu) d\lambda, \quad T(\mu) = \frac{3}{\mu}.$$

Therefore, from (20), the prior should have the following form:

$$p(\mu) \propto \mu^{-\frac{9}{4}} \left[\int \lambda^{-\frac{1}{2}} \pi(\lambda|\mu) d\lambda \right]^{1/2}.$$

Similar to Example 3, when $\pi(\lambda|\mu)$ is independent of μ, the divergence prior is proportional to $\mu^{-\frac{21}{4}}$.

- When λ is the parameter of interest, one gets that $u(\lambda) = \frac{1}{2\lambda}$. Then, from (24), the prior should have the following form:

$$p(\lambda) \propto \lambda^{-\frac{1}{2}}.$$

Example 5: (Fisher von-Mises) The probability density function

$$f(x|\mu,\lambda) = \frac{\exp[\lambda\cos(x-\mu)]}{2\pi I_0(\lambda)}, \quad \text{where } I_0(\lambda) = \frac{1}{2\pi}\int \exp(\lambda\cos x) dx.$$

Then $t(x,\mu) = \cos(x-\mu)$, $a(\lambda) = I_0^{-1}(\lambda)$. Hence, $\frac{\partial t}{\partial \mu} = \sin(x-\mu)$, $\frac{\partial^2 t}{\partial \mu^2} = -\cos(x-\mu)$, $\frac{\partial^3 t}{\partial \mu^3} = \sin(x-\mu)$, so that $E\left(\frac{\partial^3 t}{\partial \mu^3}|\mu,\lambda\right) = E\left(\frac{\partial t}{\partial \mu}|\mu,\lambda\right) = 0$. Further $u(\lambda) = -\frac{I_0'(\lambda)}{I_0(\lambda)}$ and $u'(\lambda) = -\frac{d}{d\lambda}\left[\frac{I_0'(\lambda)}{I_0(\lambda)}\right]$. • When μ is the parameter of interest, it is easy to get that

$$I_{\mu\mu} = -\lambda E\left[\frac{\partial^2 t}{\partial \mu^2}|\mu,\lambda\right] = \lambda\frac{I_1(\lambda)}{I_0(\lambda)}, \;\; E\left[\frac{\partial^3 t}{\partial \mu^3}|\mu,\lambda\right] = 0.$$

Then, by using (22), one gets

$$Q(\mu) = \int \lambda^{-\frac{1}{2}}\left[\frac{I_0(\lambda)}{I_1(\lambda)}\right]^{1/2}\pi(\lambda|\mu)d\lambda \text{ (which is a constant)}, \;\; T(\mu) = 0.$$

Therefore, from (20), the prior for μ is uniform distribution.

• When λ is the parameter of interest, one gets that $u(\lambda) = -\frac{I_0'(\lambda)}{I_0(\lambda)}$. Then, from (24), the prior should have the following form:

$$p(\lambda) \propto \left[\frac{d}{d\lambda}\left(\frac{I_0(\lambda)}{I_0(\lambda)}\right)\right]^{1/4}.$$

Example 6: Let's consider the selection of priors for the famous Neyman-Scott problem (Berger and Bernardo 1992a, 1992b). This problem can be formalized as a fixed effects one-way balanced ANOVA model.

Let $X_{i1},\cdots,X_{ik}|\mu_i$ be mutually independent $N(\mu_i,\sigma^2)$, $i = 1,\cdots,n, \;\; k \geq 2$. Here, k, the number of observations within each treatment i is fixed, while the number of treatments, n, can grow to infinity. The Fisher Information matrix is

$$I(\mu_1,\cdots,\mu_n,\sigma^2) = k \text{ Diag}(\sigma^{-2},\cdots,\sigma^{-2},\frac{n\sigma^{-4}}{n}).$$

If we consider all the parameters of equal importance, then one gets Jeffreys' general rule prior

$$\pi_J(\mu_1,\cdots,\mu_n,\sigma^2) \propto (\sigma^2)^{-n/2-1}.$$

The corresponding marginal posterior distribution of σ^2 is an Inverse Gamma distribution with two parameters equaling to $\frac{nk}{2}$, $\frac{n(k-1)S}{2}$ respectively. Here $S = \frac{1}{n(k-1)}\sum_{i=1}^n\sum_{j=1}^k(X_{ij}-\bar{X}_i)^2$. Then the posterior mean of σ^2 is given by $n(k-1)S/(nk-2)$, while the posterior mode is $n(k-1)S/(nk+2)$. Both are inconsistent estimators of σ^2.

Now we use the expression (20) to construct prior for this problem and show that the corresponding Bayes estimators of σ^2 are consistent.

Here, σ^2 is the parameter of interest, while $\mu_1,\cdots,\mu_n$ are nuisance parameters. By (21), one gets

$$Q(\sigma^2) = \sqrt{\frac{2}{nk}}\sigma^2, \;\; T(\sigma^2) = \frac{2}{\sigma^2}.$$

Therefore, the desired prior for σ^2 obtained by (20) is proportional to $(\sigma^2)^{-3/2}$.

Letting the conditional prior for $(\mu_1, \cdots, \mu_n)$ given σ^2 equal to the conditional Jeffreys' general rule prior suggested by Bernardo (1979) (in this case, it is the uniform distribution), one gets the two-stage reference prior $\pi_R(\mu_1, \cdots, \mu_n, \sigma^2) \propto (\sigma^2)^{-3/2}$.

The corresponding marginal posterior of σ^2 is an Inverse Gamma distribution with two parameters equaling to $\frac{n(k-1)+1}{2}$, $\frac{n(k-1)S}{2}$ respectively. Then the posterior mean of σ^2 is given by $n(k-1)S/(nk-n-1)$, while the posterior mode is $n(k-1)S/(nk-n+3)$. Both are consistent estimators of σ^2.

Note. Datta and Ghosh (1995a) studied the same problem and developed their two-stage reference prior based on the Kullback-Leibler divergence. Their prior is proportional to $(\sigma^2)^{-1}$ which is slightly different from ours and also leads to consistent Bayes estimators of σ^2.

Example 7: One-way random effects model has been studied by many people through Bayesian approach. Now, we revisit this model and only consider the balanced model.

Let $Y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1, \ldots, k, \quad j = 1, \ldots, n$. Here the common mean μ is an unknown parameter, while α_i's and e_{ij} are mutually independent with α_i's i.i.d $N(0, \sigma_\alpha^2)$ and e_{ij} i.i.d $N(0, \sigma^2)$.

Berger and Bernardo (1992c) first found two-stage reference priors for variance components in this problem. Later Ye (1994) and Datta and Ghosh (1995a, 1995b) found reference priors under different parametrizations. Here, we follow the parametrization used in Ye (1994). Let $\phi = \frac{n\sigma_\alpha^2}{\sigma^2}$. So parameters are μ, σ^2 and ϕ. The likelihood function $L(\mu, \sigma^2, \phi)$ can be written as

$$L(\mu, \sigma^2, \phi) \propto \sigma^{-kn}(1+\phi)^{-k/2} \exp\left\{-\frac{1}{2\sigma^2}\left[S_2 + \frac{S_1 + kn(\bar{Y} - \mu)^2}{1+\phi}\right]\right\},$$

where

$$\bar{Y_i} = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}, \quad \bar{Y} = \frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij},$$

$$S_1 = n\sum_{i=1}^{k}(\bar{Y_i} - \bar{Y})^2, \quad S_2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(Y_{ij} - \bar{Y_i})^2.$$

Then the Fisher information matrix simplifies to

$$I(\mu, \sigma^2, \phi) = \begin{pmatrix} \frac{kn}{(1+\phi)\sigma^2} & 0 & 0 \\ 0 & \frac{kn}{2\sigma^4} & \frac{k}{2(1+\phi)\sigma^2} \\ 0 & \frac{k}{2(1+\phi)\sigma^2} & \frac{k}{2(1+\phi)^2} \end{pmatrix}.$$

The inverse matrix is

$$I^{-1}(\mu, \sigma^2, \phi) = \begin{pmatrix} \frac{\sigma^2(1+\phi)}{kn} & 0 & 0 \\ 0 & \frac{2\sigma^4}{k(n-1)} & -\frac{2(1+\phi)\sigma^2}{k(n-1)} \\ 0 & -\frac{2(1+\phi)\sigma^2}{k(n-1)} & \frac{2n(1+\phi)^2}{k(n-1)} \end{pmatrix}.$$

To illustrate our method, we consider the following two cases.

1. μ is the parameter of interest. The common mean μ is of great relevance in meta analysis (Morris and Normand 1992).

By (21), one gets

$$Q(\mu) = \int \left(\frac{(1+\phi)\sigma^2}{kn}\right)^{1/2} \pi(\sigma^2, \phi|\mu) d\sigma^2 d\phi \text{ (which is just a constant)}, \quad T(\mu) = 0.$$

Therefore, by using (20), the prior for μ is uniform distribution. If we take $\pi(\sigma^2, \phi|\mu)$ as the conditional Jeffreys' general prior which is proportional to $(1+\phi)^{-1}\sigma^{-2}$ in this case, then the two-stage reference prior is

$$\pi(\mu, \sigma^2, \phi) \propto (1+\phi)^{-1}\sigma^{-2}.$$

2. ϕ is the parameter of interest. As pointed out by Ye (1994), the variance ratio σ_α^2/σ^2 is of great interest in genetic studies.

By (21), one gets

$$Q(\phi) = \int \left(\frac{2n(1+\phi)^2}{k(n-1)}\right)^{1/2} \pi(\mu, \sigma^2|\phi) d\sigma^2 d\phi = \sqrt{\frac{2n}{k(n-1)}}(1+\phi), \quad T(\phi) = \frac{2-1/n}{1+\phi}.$$

Therefore, by using (20), the prior for ϕ is proportional to $(1+\phi)^{-\frac{3}{2}+\frac{1}{n}}$. If we take $\pi(\mu, \sigma^2|\phi)$ as the conditional Jeffreys' general prior which is proportional to $(1+\phi)^{-1/2}\sigma^{-3}$ in this case, then the two-stage reference prior is

$$\pi(\mu, \sigma^2, \phi) \propto (1+\phi)^{-2+1/n}\sigma^{-3}.$$

4. Summary

In this paper, We generalize the idea from Bernardo (1979) to handle the problem of selection of priors in the presence of nuisance parameters. Instead of using Kullback-Leibler divergence which is studied by Bernardo (1979), we use a general divergence criterion to develop objective priors. This general divergence criterion is a family of divergence measures between prior and corresponding posterior including the Kullback-Leibler, Bhattacharyya-Hellinger and the Chi-square divergence. An interesting finding is that with one exception (the Chi-square divergence), for every divergence measure in the general divergence family, the desired divergence prior is the same prior as Bernardo found. Under the Chi-square divergence, we have shown that the objective prior should be the solution to a set of partial differential equations. We also consider a special case when the parameter of interest is one dimensional. In this case, the closed forms of the optimal priors are provided and also several examples are given.

References

Amari, S. (1982). Differential geometry of curved exponential families - curvatures and information loss, *The Annals of Statistics,* **10,** 357–387.

Berger, J. O. and Bernardo, J. M. (1992a). Ordered group reference priors with application to the multinomial problem, *Biometrika,* **25,** 25-37,

Berger, J. O. and Bernardo, J. M. (1992b). On the development of reference priors (with discussion), *Bayesian Statistics 4,* Eds J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, pp. 35-60, Springer-Verlag, Berlin.

Berger, J. O. and Bernardo, J. M. (1992c). Reference Priors in a Variance Components Problem, *Bayesian Statistics in Statistics and Econometrics,* Eds P. K. Goel and N. S. Iyengar, pp. 177-194, Springer-Verlag, New York.

Berger, J. O. and Yang, R. (1994). Noninformative priors and Bayesian testing for the AR(1) model, *Econometric Theory,* **10,** 461-482,

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference, *Journal of the Royal Statistical Society, Series B,* **41,** 113-147.

Bhattacharyya, A. K. (1943). On a measure of divergence between two statistical populations defined by their probability distributions, *Bulletin of the Calcutta Mathematical Society,* **35,** 99-109.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion, *The Annals of Statistics,* **6,** 434-451.

Clarke, B. and Sun, D. (1997). Reference priors under the chi-square distance, *Sankhya, Series A,* **59,** 215-231.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inferece (with discussion), *Journal of Royal Statistical Society, Series B,* **49,** 1-39.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society, Series B,* **46,** 440-464.

Datta, G. S. and Ghosh, M. (1995a). Some remarks on noninformative priors, *Journal of the American Statistical Association,* **90,** 1357-1363.

Datta, G. S. and Ghosh, M. (1995b). Hierachical Bayes estimators of the error variance in one-way ANOVA models, *Journal of Statistical Planning and Inference,* **45,** 399-411.

Datta, G. S. and Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics,* Springer, New York.

Ghosh, J. K. (1994). *Higher Order Asymptotics.* NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 4.

Ghosh, M. and Liu, R. (2009). Divergence Priors For Multiparameter Models, *Technical Report,* University of Florida

Ghosh, M., Mergel, V. and Liu, R. (2011). A general divergence criterion for prior selection, *Annals of the Institute of Statistical Mathematics,* **63,** 43-58.

Giaquinta, M. (1983). *Multiple Integrals in the Calculus of Variations and Nonlinear Elliptic Systems* Princeton University Press, Princeton, New Jersey.

Hellinger, E. (1909). Neue Begründung der Theorie quadratischen Formen von unendlichen vielen Veränderlichen. *Journal für Reine und Angewandte Mathematik,* **136** 210-271.

Jeffreys, H. (1961). *Theory of Probability.* (3rd edition) London: Oxford University Press.

Johnson, R.A. (1970). Asymptotic expansions associated with posterior distribution, *The Annals of Mathematical Statistics,* **41,** 851-864.

Jorgensen, B. (1997). *The Theory of Dispersion Models.* Chapman and Hall, New York.

Liu, R., Chakrabarty, A., Samanta, T., Ghosh, J. K. and Ghosh, M. (2014). On Divergwnce Measures Leading to Jeffreys and Other Reference Priors. *Bayesian Analysis* **9**, 331-370.

Morris, C. N. and Normand, S. L. (1992). Hierarchical Models for Combining Information and for Meta-analysis, *Bayesian Statistics 4,* Eds J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, pp. 321-344, Oxford. Univ. Press, New York.

Neyman, J., and Scott, E. L. (1948). Consistent estimates based on partially consistent observations, *Econometrika,* **16,** 1-32.

Renyi, A. (1961). Measures of Entropy and Information. In *Proceedings of the 4th Berkely Symposium„ V1.* Editors: J. Neyman and L. Lecam, pp 547-561.

Ye, K. and Berger, J. (1991). Noninformative priors for inference in exponential regression mdels, *Biometrika,* **78,** 645-656.

Ye, K. (1994). Bayesian reference prior analysis on the ratio of variances for the balanced one-way random effect model, *Journal of Statistical Planning and Inference,* **41,** 267-280.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 67–74

Discriminating Between Superior $UE(\mathbf{s^2})$-optimal Supersaturated Designs

Feng- Shun Chai[1], Ashish Das[2], Rakhi Singh[3] and John Stufken[3]
[1]*Academia Sinica, Taipei, Taiwan*
[2]*Indian Institute of Technology Bombay, Mumbai, India*
[3]*University of North Carolina at Greensboro, Greensboro, NC, USA*

Received: 21 April 2020; Revised: 07 May 2020; Accepted: 12 May 2020

Abstract

For binary factors, a design is supersaturated for the main effects model if the number of runs is smaller than the number of factors. Supersaturated designs (SSDs) cannot have all orthogonal columns, and so, the traditional notions of D-, A-, E-optimality are not applicable here. SSDs are studied under criteria such as $E(s^2)$ or $UE(s^2)$ which are near-orthogonality measures. In this work, following some of the latest works, we provide algorithms to construct better $UE(s^2)$-optimal designs. We also provide a few design examples to demonstrate the proposed algorithms.

Key words: Constructions; Hadamard matrices; Superior designs; $UE(s^2)$-optimal designs.

1. Introduction

Factor-screening experiments are performed in situations when a large number of factors could potentially be affecting the response but only a limited number of runs can be performed. The main goal of these studies is to screen (or, identify) the most important factors. Supersaturated designs are useful in factor-screening experiments and they work under the effect sparsity assumption that only a small number of factors are active. For m binary factors and n runs, under a main-effects model, supersaturated designs require n to be smaller than $m + 1$.

An n-run supersaturated design d for m two-level factors is represented by an $n \times m$ matrix X_d of 1's and -1's, where the ith column of X_d corresponds to the ith factor. Let $Z_d = [1\ X_d]$ be the model matrix of the main-effects model for d. Since $n < m + 1$, it is not possible to have Z_d with mutually orthogonal columns, even though orthogonality is a desirable property. To assess non-orthogonality of supersaturated designs, the available literature on the topic involves finding lower bounds to the popular $E(s^2)$-criterion and constructing designs satisfying these lower bounds. Designs that have an equal number of ± 1s in each column of X_d for even n are called level-balanced designs. An $E(s^2)$-optimal

Corresponding Author: John Stufken
Email: j_stufke@uncg.edu

design is a level-balanced design that minimizes the sum of the squares of the inner products of columns of X_d among all level-balanced designs. A review paper by Georgiou (2014) and the references therein are a good source for the available literature on $E(s^2)$-optimal supersaturated designs.

Jones and Majumdar (2014) extended the class of available designs by removing the imposition of level-balance from the designs. To keep the definition of $E(s^2)$ sensible in the broader class of designs, they included the sums of squares of the inner product of columns of X_d with the column of all 1s in the existing definition and called it unrestricted $E(s^2)$, or $UE(s^2)$. For a design d, the

$$UE_d(s^2) = \frac{1}{\binom{m+1}{2}} \left[\sum_{i=1}^{m} (1^T x_i)^2 + \sum_{1 \le i < j \le m} (x_i^T x_j)^2 \right], \tag{1}$$

where x_i is the ith column of X_d. The $E(s^2)$-criterion minimizes the second quantity in (1) among all level balanced designs (the first quantity is 0 for level-balanced designs), whereas the $UE(s^2)$-criterion minimizes (1) among all possible designs with ± 1s. Jones and Majumdar (2014) obtained lower bounds to $UE_d(s^2)$ and provided constructions of $UE(s^2)$-optimal supersaturated designs. $UE(s^2)$-optimal supersaturated designs are easy to construct and are available for any parameter sets, whereas $E(s^2)$-optimal designs are difficult to construct and are available only for selected parameter sets.

Since many $UE(s^2)$-optimal designs exist, Jones and Majumdar (2014) and Cheng *et al.* (2018) suggested various criteria to choose the better design among all available designs. Using the same notations as in Cheng *et al.* (2018), following are a few definitions:

- $SS = \sum_{i=1}^{m} (1^T x_i)^2 = 1^T X_d X_d^T 1$;
- $LB =$ the number of level-balanced factors for n even;
- $OF =$ the number of orthogonal pairs of factors among the $\binom{m}{2}$ pairs for n even;
- $Q = LB + OF$.

For odd n, these definitions are easily generalized. For example, when n is odd, LB is the number of nearly-level-balanced factors, that is the number of factors with the corresponding column sums of X_d equal to ± 1. Similarly, OF is the number of nearly orthogonal pairs of factors among the $\binom{m}{2}$ pairs, that is, the number of pairs of factors having an inner product equal to ± 1. For even n, Q is half the number of zeros in the matrix $Z_d^T Z_d$, whereas for odd n, Q is half the number of ± 1s in the matrix $Z_d^T Z_d$.

Cheng *et al.* (2018) defined a $UE(s^2)$-optimal design to be a superior $UE(s^2)$-optimal design if it additionally minimizes SS among the class of $UE(s^2)$-optimal designs constructed in a restricted class. Singh *et al.* (2020) then extended the definition of superior $UE(s^2)$-optimal designs in a global class of all $UE(s^2)$-optimal designs, also providing the constructions of the superior $UE(s^2)$-optimal designs in a global class. In this work, we restrict ourselves to the superior $UE(s^2)$-optimal designs constructed in Cheng *et al.* (2018). Since the class of superior $UE(s^2)$-optimal designs is still very large, Cheng *et al.* (2018) further proposed that

Q should be minimized to find a better design among superior $UE(s^2)$-optimal designs. This minimization reduces the spread among the off-diagonal elements of $Z_d^T Z_d$, and because the minimization is restricted to superior $UE(s^2)$-optimal designs it favors such designs with uniformly relatively small correlations between the columns of Z_d. The superior $UE(s^2)$-optimal designs with small Q tend to perform very well on the projection-based measures such as average D-efficiency when only a small number of factors are active.

In this paper, we propose algorithms to find designs with minimum Q among the class of superior $UE(s^2)$-optimal designs constructed by Cheng *et al.* (2018). These algorithms differ based on whether $m = 4t, 4t+1, 4t+2$, or $4t+3$ and are provided in Section 2 along with an example each.

2. Algorithms for Constructing Q-superior $UE(s^2)$-optimal Designs

Constructions of superior $UE(s^2)$-optimal designs (Cheng *et al.* (2018)) differ based on the type of values that m has. Let H be a $4t \times 4t$ normalized Hadamard matrix with all the entries in the first row and first column equal to 1. If $m = 4t-1$, then any $UE(s^2)$-optimal design is superior and these designs are constructed by deleting any $4t-n$ rows and the first column of a normalized Hadamard matrix H. If $m = 4t$, adding a level-balanced column to the $UE(s^2)$-optimal design with $m = 4t-1$ gives a superior $UE(s^2)$-optimal design. If $m = 4t-2$, deleting a column having maximum absolute column sum from a $UE(s^2)$-optimal design with $m = 4t-1$ gives a superior $UE(s^2)$-optimal design. If $m = 4t+1$, two columns are added to a $UE(s^2)$-optimal design with $m = 4t-1$ so that the pairs $(1,1)$, $(-1,-1)$, $(1,-1)$ and $(-1,1)$ appear in these columns as close to equal as possible; moreover, if $n \equiv 2 \pmod 4$, the two columns must be orthogonal. Note that superior $UE(s^2)$-optimal designs can be constructed using other methods which do not necessarily add or delete one or two columns to a Hadamard matrix; some of such construction methods have been studied in Singh *et al.* (2020). We restrict ourselves to superior $UE(s^2)$-optimal designs constructed by Cheng *et al.* (2018).

We provide algorithms to find the designs with minimum Q among the superior $UE(s^2)$-optimal designs. Before we do that, we need the following result due to Singh *et al.* (2020). This result gives the values of SS for superior $UE(s^2)$-optimal designs in a restricted class, that is, the class of designs constructed using the methods of Cheng *et al.* (2018).

Theorem 1 (Theorem 1 of Singh *et al.* (2020))**:** The values of SS for a superior $UE(s^2)$-optimal design d in restricted class are

$$C(m,n) = \begin{cases} n(m-n+1) & \text{for } m = 4t-1 \\ n(m-n)+x & \text{for } m = 4t \\ n(m-n-1)+z & \text{for } m = 4t+1 \\ n(m-2n+2)+4s(n-s) & \text{for } m = 4t-2 \text{ and } s = min(n,2t) \end{cases} \tag{2}$$

where $x = 0$ for n even and $x = 1$ for n odd, and $z = 0$ for $n \equiv 0 \pmod 4$, $z = 4$ for $n \equiv 2 \pmod 4$, $z = 2$ for $m \neq n \equiv 1$ or $3 \pmod 4$, and $z = 4n-10$ for $n = m$.

We now consider four cases depending on whether m is of the form $4t-1$, $4t-2$, $4t$, or $4t+1$, where t is a positive integer.

(A) Construction for $m = 4t - 1$. Let X_b be a matrix obtained by deleting the first column of all 1s and $4t - n$ rows of H. Irrespective of the Hadmard matrix used, and irrespective of the rows deleted, each X_b has $SS = C(n, 4t-1)$ as in Theorem 1, and hence is a superior $UE(s^2)$-optimal design (Cheng *et al.*, 2018; Singh *et al.*, 2020). We call X_b the base matrix. All such X_b's form the set of base design matrices for the other three cases. For X_b, with $m = 4t - 1 = m_b$ (say), we denote the parameters SS, LB, OF, Q by SS^b, LB^b, OF^b, Q^b respectively. To find a superior $UE(s^2)$-optimal supersaturated design with minimum Q, the following steps are proposed:

(i) Find the parameter values for all the $\binom{4t}{n}$ superior $UE(s^2)$-optimal designs obtained from every available non-isomorphic Hadamard matrix H of order $4t$. Collect designs with the same parameters in the same class, thereby forming I classes of designs with distinct parameters $(SS^b,\ LB_i^b,\ OF_i^b, Q_i^b)$, $i = 1, \ldots, I$.

(ii) Without loss of generality, let parameters satisfy $Q_1^b \le Q_2^b \le \cdots \le Q_I^b$. Any superior $UE(s^2)$-optimal design with $Q^b = Q_1^b$ is, therefore, a design as proposed.

For a large m, many non-isomorphic Hadamard matrices exist and it is not possible to do (i) for all non-isomorphic Hadamard matrices. One could then do step (i) for as many non-isomorphic Hadamard matrices as possible. Then, there is a possibility of a design with smaller Q than the proposed design.

Example 1: For $m = 15, n = 12$, there are 5 non-isomorphic Hadamard matrices of order 16, the total number of possible superior $UE(s^2)$-optimal designs are $\binom{16}{12} \times 5 = 1820 \times 5 = 9100$. Among these 9100 possibilities of X_b's, we get only four distinct parameter sets given in Table 1. In Table 1, we also provide the respective number of designs in these classes under the 'Count' column.

Table 1: Four sets of parameters for $m = 15, n = 12$

i	SS^b	LB_i^b	OF_i^b	Q_i^b	Count
1	48	6	42	48	7248
2	48	3	57	60	384
3	48	9	51	60	1152
4	48	12	84	96	316
				Total	9100

Any superior $UE(s^2)$-optimal designs corresponding to the first row in Table 1, that is, designs with $Q_1^b = 48$ are the proposed designs. One such example of a superior $UE(s^2)$-optimal design with $Q = 48$ is given below.

```
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
 1 -1  1 -1  1 -1  1 -1  1 -1  1 -1  1 -1  1
-1  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1  1  1
 1  1 -1 -1  1  1 -1 -1  1  1 -1 -1  1  1 -1
-1 -1 -1  1  1  1  1 -1 -1 -1 -1  1  1  1  1
 1 -1  1  1 -1  1 -1 -1  1 -1  1  1 -1  1 -1
-1  1  1  1  1 -1 -1 -1 -1  1  1  1  1 -1 -1
 1  1 -1  1 -1 -1  1 -1  1  1 -1  1 -1 -1  1
-1 -1 -1 -1 -1 -1 -1  1  1  1  1  1  1  1  1
-1 -1 -1  1  1  1  1  1  1  1  1 -1 -1 -1 -1
-1  1  1 -1  1 -1  1  1  1 -1 -1  1 -1  1 -1
 1 -1  1  1 -1 -1  1  1 -1  1 -1 -1  1  1 -1
```

(B) Construction for $m = 4t-2$. For $m = 4t-2$, a column with the maximum absolute column sum is deleted from X_b to get a superior $UE(s^2)$-optimal design. Starting from X_b with $m_b = 4t-1$, we construct a superior $UE(s^2)$-optimal design with $m = 4t-2$ and minimum Q as follows.

(i) For the i-th set, there are $m - LB_i^b$ number of columns in the corresponding X_b's which are not (nearly) level-balanced. Let x_u be the number of columns for a design in the i-th class with the column sums equal to $\pm 2u$ (for n even) and equal to $\pm(2u+1)$ (for n odd), $u = 1, \ldots, k$. Then, for all designs in the i-th class, $i = 1, \ldots, I$, we check whether

$4\sum_{u=1}^{k} u^2 x_u = C(4t-2, n)$ for n even, and

$\sum_{u=1}^{k}(2u+1)^2 x_u = C(4t-2, n) - LB_i^b$ for n odd.

Assume that the conditions are true for $I_1 \leq I$ sets of parameters. Only keep these I_1 sets and number the sets as $i = 1, \ldots, I_1$ such that $Q_1^b \leq Q_2^b \leq \cdots \leq Q_{I_1}^b$.

(ii) Starting from $i = 1$, for the designs in the i-th class, we delete a column with column sum as $\pm n$ for $n \leq 2t$ or as $\pm(4t-n)$ for $n > 2t$ (note that this can always be done, see, Singh *et al.* (2020)). Then for each i, designs could have J_i possible sets of parameters $(C(4t-2, n),\ LB_i = LB_i^b,\ OF_{i(j)}, Q_{i(j)})$, where $OF_i^b - m \leq OF_{i(j)} \leq OF_i^b$, $j = 1, \ldots, J_i$. Without loss of generality, let $Q_{i(1)} \leq Q_{i(2)} \leq \cdots \leq Q_{i(J_i)}$. Define $q_i = min\{Q_{i(1)}, q_{i-1}\}$ for $i = 2, \ldots, I$ and $q_1 = Q_{i(1)}$.

(iii) If $q_i \leq Q_{i+1}^b - m$, then a superior $UE(s^2)$-optimal design with $Q = q_i$ is the proposed design, otherwise the steps (ii)-(iii) are sequentially repeated for $i = 2, \ldots, I_1$. If we reach $i = I_1$, then a superior $UE(s^2)$-optimal design with $Q = q_{I_1}$ is the proposed design.

Example 2: For constructing a superior $UE(s^2)$-optimal design, with $m = 14, n = 12$ having minimum Q, from Table 1, the base design with $m_b = 15, n = 12$ has $I = 4$ sets of parameters. With $C(14, 12) = 32$, the condition in step (i) of the algorithm is not met for $i = 2$. This allows us to reduce the number of sets of parameters to $I_1 = 3$. Now, for the designs in the set $i = 1$, deleting a column with the desired property gives superior $UE(s^2)$-optimal designs with $J_1 = 1$ parameters ($SS = 32$, $LB_1 = 6$, $OF_{1(1)} = 36$, $Q_{1(1)} = 42$). Since, $q = 42 < 60 - 14 = 46$, the design with $Q = q = 42$ is the proposed design and is given below.

```
 1  1  1  1  1  1  1  1  1  1  1  1  1  1
 1  1 -1 -1 -1 -1  1  1  1  1 -1 -1 -1 -1
-1 -1  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1
-1 -1 -1 -1  1  1  1  1 -1 -1 -1 -1  1  1
-1 -1 -1 -1  1  1 -1 -1  1  1  1  1 -1 -1
-1 -1  1  1 -1 -1 -1 -1  1  1 -1 -1  1  1
 1  1 -1 -1 -1 -1 -1 -1 -1 -1  1  1  1  1
 1  1  1  1  1  1 -1 -1 -1 -1 -1 -1 -1 -1
 1 -1  1 -1  1 -1  1 -1  1 -1  1 -1  1 -1
 1 -1 -1  1 -1  1  1 -1  1 -1 -1  1 -1  1
-1  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1  1
-1  1 -1  1  1 -1 -1  1  1 -1  1 -1 -1  1
```

(C) Construction for $m = 4t$. For $m = 4t$, a level-balanced column should be added to X_b to get a superior $UE(s^2)$-optimal design. Our algorithm to construct a superior $UE(s^2)$-optimal design with $m = 4t$, having the minimum value of Q is as follows. We start with X_b's as in Construction (A) corresponding to $m_b = 4t - 1$. Then, the ith set of parameters corresponding to the base matrices is $(SS^b,\ LB_i^b,\ OF_i^b, Q_i^b)$, $i = 1, \ldots, I$. Note that these i sets are such that $Q_1^b \leq Q_2^b \leq \cdots \leq Q_I^b$.

(i) Starting from $i = 1$, for each design in the i-th class (that is, a design with minimum Q), we add all possible balanced columns to the existing $4t-1$ columns. Then for each i, designs could have J_i resultant possible parameters $(C(4t, n),\ LB_i = LB_i^b + 1,\ OF_{i(j)},\ Q_{i(j)})$, where $OF_i^b \leq OF_{i(j)} \leq OF_i^b + m - 1$, $j = 1, \ldots, J_i$. Without loss of generality, let $Q_{i(1)} \leq Q_{i(2)} \leq \cdots \leq Q_{i(J_i)}$. Define $q_i = min\{Q_{i(1)}, q_{i-1}\}$ for $i \geq 2$ and $q_1 = Q_{i(1)}$.

(ii) If $q_i \leq Q_{i+1}^b$, then a superior $UE(s^2)$-optimal design with $Q = q_i$ is the final design. Otherwise steps (i)-(ii) are repeated sequentially for $i = 2, \ldots, I$. If we reach $i = I$, then a superior $UE(s^2)$-optimal design with $Q = q_I$ is the proposed design.

Example 3: For constructing a superior $UE(s^2)$-optimal design with $m = 16, n = 12$ having minimum Q, we make use of Table 1. For the designs in the set $i = 1$ in Table 1, adding level-balanced columns gives $J_1 = 7$ parameter sets of which the one with the minimum $Q_{1(j)}$ is $(SS = 48,\ LB_1 = 7,\ OF_{1(1)} = 42,\ Q_{1(1)} = 49)$. Since, $q = 49 < 60$, the design with $Q = q = 49$, given below, is a proposed superior $UE(s^2)$-optimal design with minimum Q.

```
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
 1 -1  1 -1  1 -1  1 -1  1 -1  1 -1  1 -1  1  1
-1  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1  1  1 -1
 1  1 -1 -1  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1
-1 -1 -1  1  1  1  1 -1 -1 -1 -1  1  1  1  1  1
 1 -1  1  1 -1  1 -1 -1  1 -1  1  1 -1  1 -1 -1
-1  1  1  1  1 -1 -1 -1 -1  1  1  1  1 -1 -1  1
-1 -1 -1 -1 -1 -1  1  1  1  1  1  1  1  1 -1  1
-1 -1  1 -1  1  1  1  1  1  1 -1  1 -1 -1 -1  1
-1  1 -1  1  1  1 -1  1  1 -1  1 -1 -1 -1  1  1
-1  1  1  1 -1 -1  1  1  1 -1 -1 -1  1  1 -1  1
 1 -1 -1  1 -1  1  1  1 -1  1  1 -1  1 -1 -1 -1
```

(D) Construction for $m = 4t + 1$. For $m = 4t + 1$, to get a superior $UE(s^2)$-optimal design, two columns should be added to X_b so that the pairs $(1, 1)$, $(-1, -1)$, $(1, -1)$ and $(-1, 1)$ appear in these columns as close to equal as possible; moreover, if $n \equiv 2 \pmod 4$, the two columns must be orthogonal. Our algorithm to construct a superior $UE(s^2)$-optimal design with $m = 4t$, having the minimum value of Q is an adaptation of the algorithm in Construction (C) in this paper. We again start with X_b's as in Construction (A) corresponding to $m_b = 4t - 1$. Then, the ith set of parameters corresponding to the base matrices is

$(SS^b,\ LB_i^b,\ OF_i^b, Q_i^b)$, $i = 1, \ldots, I$ with $i = 1$ corresponding to the set with the minimum value of Q.

(i) Starting from $i = 1$, for the designs in the i-th class, we add all possible sets of two level-balanced columns with 0 inner product. For each i, designs now have J_i resultant sets of possible parameters $(C(4t+1, n),\ LB_i = LB_i^b + 2,\ OF_{i(j)},\ Q_{i(j)})$, where $OF_i^b \leq OF_{i(j)} \leq OF_i^b + 2m - 3$, $j = 1, \ldots, J_i$. Without loss of generality, let $Q_{i(1)} \leq Q_{i(2)} \leq \cdots \leq Q_{i(J_i)}$. Define $q_i = min\{Q_{i(1)}, q_{i-1}\}$ for $i \geq 2$ and $q_1 = Q_{i(1)}$.

(ii) If $q_i \leq Q_{i+1}^b$, then a superior $UE(s^2)$-optimal design with $Q = q_i$ is the proposed design. Otherwise steps (i)-(ii) are sequentially repeated for $i = 2, \ldots, I$. If we reach $i = I$, then a superior $UE(s^2)$-optimal design with $Q = q_I$ is the proposed design.

Example 4: For constructing a superior $UE(s^2)$-optimal design with $m = 17, n = 12$ having minimum Q, we can again make use of the sets listed in Table 1. Now, for the designs in the set $i = 1$, adding two columns with required properties gives a large number of J_i. However, the set with the minimum $Q_{1(j)}$ is $(SS = 48,\ LB_1 = 8,\ OF_{1(1)} = 43,\ Q_{1(1)} = 51)$. Since, $q = 51 < 60$, the design with $Q = q = 51$, given below, is a proposed superior $UE(s^2)$-optimal design with minimum Q.

```
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
 1 -1  1 -1  1 -1  1 -1 -1 -1 -1  1  1  1  1 -1  1
-1  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1  1  1  1 -1
 1  1 -1 -1  1  1 -1 -1  1 -1  1 -1  1 -1 -1  1  1
-1 -1 -1  1  1  1  1 -1 -1  1  1  1  1 -1 -1  1 -1
 1 -1  1  1 -1  1 -1 -1  1  1 -1 -1  1  1 -1  1  1
-1  1  1  1  1 -1 -1 -1  1  1 -1  1 -1 -1  1 -1 -1
 1  1 -1  1 -1 -1  1 -1  1 -1  1  1 -1  1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1  1  1  1  1  1  1  1  1  1  1
 1 -1  1 -1  1 -1  1  1  1  1  1 -1 -1 -1 -1  1 -1
-1  1  1 -1 -1  1  1  1  1 -1 -1  1  1 -1 -1  1  1
 1 -1  1  1 -1  1 -1  1 -1 -1  1  1 -1 -1  1 -1  1
```

3. Concluding Remarks

In this paper, we have provided algorithms for constructing superior $UE(s^2)$-optimal designs with minimum Q starting from the class of designs in Cheng *et al.* (2018). If it is not feasible to identify the I sets of parameters for all non-isomorphic Hadamard matrices, we propose to run the algorithms on only a selected set of the available Hadamard matrices; however, then there is no guarantee that the designs proposed here would have the minimum value of Q. As mentioned previously, superior $UE(s^2)$-optimal designs also exist outside the class of designs used here (Singh *et al.*, 2020). A future direction is to identify the best designs using better algorithms (or, analytically) to identify superior $UE(s^2)$-optimal designs with minimum Q among a broader class of all superior $UE(s^2)$-optimal designs.

Acknowledgments

The research of F. S. Chai was supported by the grant MOST 109-2118-M-001-004-MY2 from the Ministry of Science and Technology of Taiwan. The research by J. Stufken was partially supported through NSF grant DMS-1935729.

References

Cheng, C. S., Das, A., Singh, R., and Tsai, P. W. (2018). $E(s^2)$- and $UE(s^2)$-optimal supersaturated designs. *Journal of Statistical Planning and Inference*, **196**, 105–114.

Georgiou, S. D. (2014). Supersaturated designs: a review of their construction and analysis. *Journal of Statistical Planning and Inference*, **144**, 92–109.

Jones, B. and Majumdar, D. (2014). Optimal supersaturated designs. *Journal of the American Statistical Association*, **109(508)**, 1592–1600.

Singh, R., Das, A., and Horsley, D. (2020). $SUE(s^2)$-optimal supersaturated designs. *Statistics and Probability Letters*, **158**,1–5.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 75-84

A New Family of Probability Density Functions With Deflated Tails

Avinash D. Dharmadhikari[1], Sanjeev V. Sabnis[2] and Bikas K. Sinha[3]
[1] *(Ex-GM, Tata Motors, Pune) B2/203, Prisam Society, Aundh, Pune, India.*
[2] *Department of Mathematics, Indian Institute of Technology Bombay, Mumbai, India.*
[3] *Ex-Professor, Indian Statistical Institute, Kolkata, India.*

Received: 30 March 2020; Revised: 17 May 2020; Accepted: 20 May 2020

Abstract

The objective of this study is to introduce a new family of probability density functions using a given probability density function and analyze some of its important theoretical properties involving quantiles and failure rate. As an offshoot of this new family of probability density functions, a two-component mixture density that behaves differently at the tail-ends viz., the two densities tend to be negligible to the left / right of two designated tail-end values, respectively, is proposed. This is important as mixture probability distributions have been extensively studied in the literature and their applications in real life are also well-known.

Key words: Mixture distribution; Failure rate; Quantiles; Stochastic ordering.

AMS Subject Classifications: 62E15, 60E15, 62N05

1. Introduction

The subject of mixture distributions is important from theoretical as well as practical points of view. Three major references that deal with this topic are Everittt and Hand (1981), Titterington, Smith and Makov (1985), and McLachlan and Basford (1988). Theoretical aspects are concerned with (i) obtaining parameter estimates within models and those of mixing distributions, (ii) identification of the number of components in a finite mixture, (iii) imputation of the missing indicators of component membership for mixture data, whereas practical aspects deal with areas such as fisheries research, economics, medicine, biology, psychology, palaeontology, geology, botany, agriculture, zoology, reliability and many other fields.

Mathematically, mixture distributions are typically formalized as follows.
Consider a pair $Y = (X, Z)$ of random variables with $g(y)$ as its joint probability density and suppose that

$$g(y) = g(x, z) = f(x|z)\pi(z)$$

Correponding Author: S. V. Sabnis
Email: svs@math.iitb.ac.in

where f and π are, respectively, a conditional and marginal density. Then, in terms of this factorization, the marginal density for X, $p(x)$, is

$$p(x) = \int f(x|z)\pi(z)dz.$$

If the support of π is finite and concentrated on $c_1, c_2, \ldots, c_k$, say, then we get

$$p(x) = \sum_{i=1}^{k} \pi_i f(x_i)$$

where

$$f_i(x) = f(x|Z = c_i)$$

and

$$\pi_i = P(Z = c_i)$$

for $i = 1, 2, \ldots, k$.

In this case, X is said to have a finite mixture distribution, the f_i are called the component densities and the probabilities $\{\pi_i\}$ are called mixing weights.

In this paper, a new family of probability density functions using the given probability density function has been proposed and some of its important theoretical properties involving quantiles and failure rate have been analyzed. As an offshoot of this new family of probability density functions, a two-component mixture density that behaves differently at the tail-ends viz., the two densities tend to be negligible to the left / right of two designated tail-end values, respectively, is proposed. The applications for this version of mixture model are understood to arise in car industries.

The organization of this paper is as follows. Section 2 discusses new models and associated theoretical results. Section 3 contains simulation results, while Section 4 presents conclusions.

2. General Model

This section presents a new class of probability density functions and the same is given in Definition 1 below.

Definition 1: Let X be an absolutely continuous random variable having $f(x)$ as its probability density function and $F(x)$ as its cumulative density function. Then for every pair of real numbers s and t, $s < t$, we may define an induced random variable $X^*(s,t)$ having probability density function $f^*(x|(s,t))$ given by

$$f^*(x|(s,t)) = \begin{cases} \lambda f(x) & x < s \\ (\lambda + \delta) f(x) & s \leq x < t \\ \delta f(x) & x \geq t \end{cases}$$

with $p \in (0,1)$, $\lambda = \dfrac{(1-p)}{D_1}$, $\delta = \dfrac{p}{D_1}$, and $D_1 = (1-p)F(t) + p(1-F(s))$.

The corresponding Cumulative Distribution Function (CDF) is given by

$$F^*(x|(s,t)) = \begin{cases} \lambda F(x) & x < s \\ (\lambda+\delta)F(x) - \delta F(s) & s \le x < t \\ \lambda F(t) - \delta F(s) + \delta F(x) & x \ge t. \end{cases}$$

Remark 1: It may be noted that the probability density function $f^*(x|(s,t))$ defined above in Definition 1 is discontinuous at s and t where s, t, and p are assumed to be fixed. However, further, it may be noted that the corresponding CDF $F^*(x|(s,t))$ is continuous at s and t.

Remark 2: It may be noted that $\{X^*(s,t), s < t\}$ is a family of random variables induced by the given absolutely continuous random variable X and it is such that it is deflated in both the tails and inflated in the middle.

Remark 3: A special case of interest is when s and t, $s < t$, are chosen such that $F(s) = \overline{F}(t)$, where $\overline{F}(t)$ is the survival function of the original absolutely continuous random variable X which gives rise to a new random variable $X^*(s,t)$. In this case, $\overline{F}(s) = F(t) = D_1$ and for any $0 < p < 1$,

$$f^*(x) = \begin{cases} \frac{(1-p)}{\overline{F}(s)} f(x) & x < s \\ \frac{1}{\overline{F}(s)} f(x) & s \le x < t \\ \frac{p}{\overline{F}(s)} f(x) & x \ge t. \end{cases}$$

Further to this, we may also note the following:
(a) For $p = 1$,

$$f^*(x) = \begin{cases} 0 & x < s \\ \frac{f(x)}{\overline{F}(s)} & x \ge s. \end{cases}$$

Note that $f^*(x)$ is a probability density function of a random variable $X^*(s,.)$ which is truncated to the left of s.
(b) For $p = 0$,

$$f^*(x) = \begin{cases} \frac{f(x)}{F(t)} & x < t \\ 0 & x \ge t. \end{cases}$$

Note that $f^*(x)$ is a probability density function of a random variable $X^*(.,t)$ which is truncated to the right of t.

Remark 4: It is easy to verify for selected values of s, t that

$$\lambda \leq 1 \Leftrightarrow p \geq \frac{\overline{F}(t)}{\overline{F}(t)+\overline{F}(s)}$$

and

$$\delta \leq 1 \Leftrightarrow p \leq \frac{F(t)}{F(t)+F(s)}.$$

Here is an interesting result one that makes connection between percentile points of the original probability density function f and those of the new probability density function f^*. We state it in the form of properties of f^* in relation to f.

Let $F^*(x_q^*) = q$ and $F(x_q) = q$, and η_q be such that $F^*(x_q) = q + \eta_q$ where the sign of η_q is positive (negative) when f^* is positively skewed (negatively skewed). Then

Property A: For $x < s$, $x_q^* = F^{-1}\Big(\frac{qD_1}{(1-p)}\Big)$.

Property B: For $s \leq x < t$, $x_q^* = F^{-1}\Big(q + D_1\eta_q\Big)$.

Property C: For $x \geq t$, $x_q^* = F^{-1}\Big(q + \frac{D_1}{p}\eta_q\Big)$.

We provide the proofs of these properties in the Appendix A.

A question regarding stochastic comparison between X^* and X is in order. The following result is geared towards that. The concept of a random variable being stochastically larger than another random variable and the concept of failure rate can be found in Barlow and Proschan (1975).

Theorem 1: The random variable X^* having f^* as its pdf is stochastically larger than the random variable X with pdf f if and only if $p \geq \frac{F(t)}{F(t)+F(s)}$.

Proof: Not to obscure the essential steps of reasoning, we will go through the following Lemmas.

Lemma 1: For $x < s$, the random variable X^* having f^* as its pdf is stochastically larger than the random variable X with pdf f if and only if $p \geq \frac{\overline{F}(t)}{\overline{F}(t)+\overline{F}(s)}$.

Proof: For $x < s$,

$$F^*(x) = \lambda F(x) < F(x) \Leftrightarrow \lambda = \frac{(1-p)}{D_1} \leq 1 \Leftrightarrow p \geq \frac{\overline{F}(t)}{\overline{F}(s)+\overline{F}(t)}.$$

Lemma 2: For $s \leq x < t$, the random variable X^* having f^* as its pdf is stochastically larger than the random variable X with pdf f if and only if $p \geq \frac{F(x)\overline{F}(t)}{\overline{F}(x)F(s)+F(x)\overline{F}(t)}$.

Proof: For $s \leq x < t$,

$$F^*(x) = (\lambda+\delta)F(x) - \delta F(s) < F(x) \Leftrightarrow \frac{F(x)}{D_1} - F(x) < \delta F(s) \quad as \;\; \lambda + \delta = \frac{1}{D_1}$$

$$\Leftrightarrow \frac{(1-D_1)}{D_1}F(x) < \frac{p}{D_1}F(s) \Leftrightarrow p \geq \frac{F(x)\overline{F}(t)}{\overline{F}(x)F(s)+F(x)\overline{F}(t)}.$$

Lemma 3: For $x > t$, the random variable X^* having f^* as its pdf is stochastically larger than the random variable X with pdf f if and only if $p \geq \frac{F(t)}{F(t)+F(s)}$.
Proof: For $x > t$,

$$F^*(x) = \lambda F(t) + \delta F(x) - \delta F(s) \leq F(x) \Leftrightarrow \frac{(1-p)}{D_1}F(t) + \frac{p}{D_1}F(x) - \frac{p}{D_1}F(s) \leq F(x)$$

$$\Leftrightarrow p \geq \frac{F(t)\overline{F}(x)}{[F(t)\overline{F}(x) + F(s)\overline{F}(x)]} = \frac{F(t)}{F(t)+F(s)}.$$

Proof of Theorem 1: The property of F^* being stochastically larger than F implies that simultaneously last inequalities involving p in the proofs of Lemma 1, Lemma 2, and Lemma 3 have to necessarily hold good. This amounts to saying that

$$p \geq \max\Big\{\frac{\overline{F}(t)}{\overline{F}(t)+\overline{F}(s)}, \sup_{x\in[s,t)} \frac{F(x)\overline{F}(t)}{\overline{F}(x)F(s)+F(x)\overline{F}(t)}, \frac{F(t)}{F(t)+F(s)}\Big\}.$$

We establish that this is equivalent to the condition stipulated in the statement of this Theorem.

To this end, we make two claims and prove them.
Claim 1:

$$\frac{\overline{F}(t)}{\overline{F}(t)+\overline{F}(s)} \leq \frac{F(t)}{F(t)+F(s)}$$

$$\Leftrightarrow \overline{F}(t)(F(t)+F(s)) \leq F(t)(\overline{F}(s)+\overline{F}(t))$$

$$\Leftrightarrow \overline{F}(t)F(t) + \overline{F}(t)F(s) \leq \overline{F}(s)F(t) + \overline{F}(t)F(t)$$

$$\Leftrightarrow F(s) \leq F(t) \;\; always\ holds\ as\ s < t.$$

Thus, Claim 1 proved.
Claim 2:

$$\sup_{x\in[s,t)} \frac{F(x)\overline{F}(t)}{\overline{F}(x)F(s)+\overline{F}(t)F(x)} \leq \frac{F(t)}{F(t)+F(s)}.$$

Consider

$$\frac{F(x)\overline{F}(t)}{\overline{F}(x)F(s)+\overline{F}(t)F(x)} \leq \frac{F(t)}{F(t)+F(s)}$$

$$\Leftrightarrow F(x)\overline{F}(t)F(t) + F(x)\overline{F}(t)F(s) \leq F(t)\overline{F}(x)F(s) + F(t)\overline{F}(t)F(x)$$

$$\Leftrightarrow F(x)\overline{F}(t) \leq F(t)\overline{F}(x)$$

$$\Leftrightarrow F(x) \leq F(t) \;\; which \;\; is\ valid\ for\ x \in [s,t).$$

Thus, Claim 2 is proved and the proof is complete.

The following Theorem attempts to make connection between the failure rate of f^* and that of f.

Theorem 2: The probability density function f^* corresponding to the random variable X^* has increasing (decreasing) failure rate η^* over the set C iff for $x \in C$

$$\max\left[(\eta(x) - \frac{d(\ln \eta(x))}{dx}), 0\right] < (>) \min\left[\frac{\delta f(x)}{[1-\lambda F(t) - \delta F(x) + \delta F(s)]}, \frac{\delta f(x)}{[1-\lambda F(x)]}\right]$$

where $\eta(x)$ is the failure rate of the random variable X with f as its probability density function and the set C is such that $C \subseteq \mathcal{R}^+$, and, further, for $x \in C$ $x \neq s, t$.

Proof: The proof of this Theorem is rather long and a bit complicated. We provide details of this proof in the Appendix B.

It is interesting to note that the family of probability distribution functions given in Definition 1 results from the following tweaked version of a mixture of two probability distribution functions having an appealing feature that this mixture density behaves differently at the tail-ends *viz.*, the two densities tend to be negligible to the left / right of two designated tail-end values, respectively.

Definition 2: Let X_i be an absolutely continuous random variable with $f_i(x)$ as its probability density function and $F_i(x)$ as its cumulative density function for $i = 1, 2$. For every $p \in (0, 1)$ and every pair of real numbers s, t $(s < t)$,

$$h_2(u|(s,t)) = \begin{cases} \frac{(1-p)}{D_2} f_1(u) & u < s \\ \frac{1}{D_2}((1-p)f_1(u) + pf_2(u)) & s \leq u < t \\ \frac{p}{D_2} f_2(u) & u \geq t \end{cases}$$

defines a probability density function of a random variable $U(s,t)$, say, for fixed values of s and t. Here $D_2 = (1-p)F_1(t) + p\overline{F}_2(s)$.

The corresponding Cumulative Distribution Function (CDF) is given by

$$H_2(u|(s,t)) = \begin{cases} \frac{(1-p)}{D_2} F_1(u) & u < s \\ \frac{1}{D_2}((1-p)F_1(u) + p(F_2(u) - F_2(s)) & s \leq u < t \\ \frac{1}{D_2}(1-p)F_1(t) + p(F_2(u) - F_2(s)) & u \geq t. \end{cases}$$

Remark 5: It may be noted that Definition 2 reduces to Definition 1 if X_1 and X_2 are identically distributed random variables.

Remark 6: The applications for this version of mixture model are understood to arise in car industries.

Conclusions: In this paper, a new family of probability density functions from a given probability density function is generated, and the relationships of the former to the latter in terms quantiles and failure rate are studied. This family of probability density function results from the tweaked version of a mixture of two probability density functions.

Acknowledgments

The authors would like to thank Dr. Jyotirmoy Sarkar of IUPUI for his comments on the original version of this manuscript. The authors would also like to thank the referee for his/her valuable suggestions which have resulted in improvement in presentation of the contents of the paper.

References

Barlow R. E., and Proschan, F. (1975). *Statistical Testing of Reliability and Life Testing: Probability Models*. Holt, Rinehart and Winston, Inc., New York.

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixtures Distribution*. Chapman and Hall, London.

McLachlan, D. J. and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, New York.

Titterington D. M., Smith A. F. M., and Makov U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, New York.

APPENDIX A

Proof of Property A: For $x < s$, $F^*(x) = \lambda F(x)$.
Let x_q^* be such that $F^*(x_q^*) = q$ and x_q be such that $F(x_q) = q$.
Thus, we have, $F^*(x_q^*) = \lambda F(x_q^*) = q$.
This implies that $F(x_q^*) = \frac{q}{\lambda}$ with $\lambda = \frac{(1-p)}{D_1}$ and $D_1 = (1-p)F(t) + p\overline{F}(s)$.
i.e. $x_q^* = F^{-1}(\frac{q}{\lambda}) = F^{-1}\Big(\frac{qD_1}{(1-p)}\Big)$.
Proof of Property B: For $s \leq x < t$,

$$\begin{aligned} F^*(x) &= \lambda F(x) + \delta F(x) - \delta F(s) \\ &= (\lambda + \delta)F(x) - \delta F(s) \\ &= \frac{1}{D_1}F(x) - \delta F(s). \end{aligned}$$

With x_q^* and x_q as defined above, we have

$$\begin{aligned} F^*(x_q^*) &= \frac{1}{D_1}F(x_q^*) - \delta F(s) \\ &= q \end{aligned} \tag{1}$$

$$\begin{aligned} F^*(x_q) &= \frac{1}{D}F(x_q) - \delta F(s) \\ &= \frac{1}{D_1}q - \delta F(s) \\ &= q + |\eta_q| \end{aligned} \tag{2}$$

where $|\eta_q|$ is such that (i) it will be positive if F^* is positively skewed and (ii) it will be negative if F^* is negatively skewed. From (1) and (2) we have

$$\begin{aligned} |F^*(x_q^*) - F^*(x_q)| &= \Big|\frac{1}{D_1}F(x_q^*) - \delta F(s) - \frac{1}{D_1}F(x_q) - \delta F(s)\Big| \\ &= \Big|\frac{1}{D_1}F(x_q^*) - \frac{1}{D_1}F(x_q)\Big| \\ &= |\eta_q|. \end{aligned}$$

This implies that $\frac{1}{D_1}F(x_q^*) - \frac{1}{D_1}F(x_q) = \eta_q$ i.e. $\frac{1}{D_1}F(x_q^*) = \frac{1}{D_1}F(x_q) + \eta_q$ This, in turn, implies that $x_q^* = F^{-1}\Big(q + D_1\eta_q\Big)$.

Proof of Property C: For $x \geq t$, $F^*(x) = \lambda F(t) - \delta F(s) + \delta F(x)$. Thus, for x_q^* and x_q as defined above, we have

$$\begin{aligned} F^*(x_q^*) &= \lambda F(t) - \delta F(s) + \delta F(x_q^*) \\ &= q \end{aligned} \tag{3}$$

$$\begin{aligned} F^*(x_q) &= \lambda F(t) - \delta F(s) + \delta F(x_q) \\ &= q + |\eta_q|. \end{aligned} \tag{4}$$

From (3) and (4) we have

$$\begin{aligned} |F^*(x_q^*) - F^*(x_q)| &= |\delta F(x_q^*) - \delta F(x_q)| \\ &= |\delta F(x_q^*) - \delta q| \\ &= |\eta_q|. \end{aligned}$$

This yields $x_q^* = F^{-1}\Big(q + \dfrac{D_1}{p}\eta_q\Big)$.

APPENDIX B

From the expressions of $f^*(x)$ and $F^*(x)$, it follows that the corresponding failure rate η^* is given by

$$\eta^*(x) = \begin{cases} \dfrac{\lambda\overline{F}(x)\eta(x)}{1 - \lambda F(x)} & x < s \\ \dfrac{(\lambda+\delta)\overline{F}(x)\eta(x)}{1 - (\lambda+\delta)F(x) + \delta F(s)} & s \leq x < t \\ \dfrac{\delta\overline{F}(x)\eta(x)}{1 - \lambda F(t) - \delta F(x) + \delta F(s)} & x \geq t \end{cases}$$

where $\eta(x)$ denotes the hazard rate of $f(x)$, while $\eta^*(x)$ denotes the hazard rate of $f^*(x)$.

First, let $x < s$. Then $\eta^*(x)$ is increasing in x if and only if its derivative $\frac{d\eta^*(x)}{dx}$ is greater than 0. Note that

$$\begin{aligned}\frac{d\eta^*(x)}{dx} &= \frac{d[\frac{\lambda\overline{F}(x)(\eta(x))}{1-\lambda F(x)}]}{dx} \\ &= \frac{(1-\lambda F(x))(-\lambda f(x)\eta(x)+\lambda\overline{F}(x)\frac{d\eta(x)}{dx})}{(1-\lambda F(x))^2}+\frac{(\lambda)^2\overline{F}(x)f(x)\eta(x)}{(1-\lambda F(x))^2}.\end{aligned}$$

Thus $\frac{d\eta^*(x)}{dx} > 0$ iff $(1-\lambda F(x))(-\lambda f(x)\eta(x)+\lambda\overline{F}(x)\frac{d\eta(x)}{dx})+(\lambda)^2\overline{F}(x)f(x)\eta(x) > 0$
i.e., $\frac{d\eta^*(x)}{dx} > 0$ iff $(\lambda)^2\overline{F}(x)f(x)\eta(x) > \lambda\overline{F}(x)[1-\lambda F(x)][(\eta(x))^2-\frac{d\eta(x)}{dx}]$
i.e., $\frac{d\eta^*(x)}{dx} > 0$ iff $(\lambda)f(x)\eta(x) > [1-\lambda F(x)][(\eta(x))^2-\frac{d\eta(x)}{dx}]$.
Next, whenever $s \le x < t$, we have

$$\eta^*(x) = \frac{(\lambda+\delta)\overline{F}(x)\eta(x)}{1-(\lambda+\delta)F(x)+\delta F(s)}.$$

Then $\eta^*(x)$ is increasing in x iff its derivative $\frac{d\eta^*(x)}{dx}$ is greater than 0.
Note that

$$\begin{aligned}\frac{d\eta^*(x)}{dx} &= \frac{d[\frac{(\lambda+\delta)\overline{F}(x)\eta(x)}{1-(\lambda+\delta)F(x)+\delta F(s)}]}{dx} \\ &= \frac{[1-(\lambda+\delta)F(x)+\delta F(s)][-(\lambda+\delta)f(x)\eta(x)+(\lambda+\delta)\overline{F}(x)\frac{d\eta(x)}{dx}]}{[1-(\lambda+\delta)F(x)+\delta F(s)]^2} \\ &\quad +\frac{(\lambda+\delta)^2\overline{F}(x)f(x)\eta(x)}{[1-(\lambda+\delta)F(x)+\delta F(s)]^2}.\end{aligned}$$

Thus, $\frac{d\eta^*(x)}{dx} > 0$ iff

$$[1-(\lambda+\delta)F(x)+\delta F(s)][-(\lambda+\delta)f(x)\eta(x)+(\lambda+\delta)\overline{F}(x)\frac{d\eta(x)}{dx}]+(\lambda+\delta)^2\overline{F}(x)f(x)\eta(x) > 0$$

$\frac{d\eta^*(x)}{dx} > 0$ iff $(\lambda+\delta)^2\overline{F}(x)f(x)\eta(x) > (\lambda+\delta)\overline{F}(x)[1-(\lambda+\delta)F(x)+\delta F(s)][(\eta(x))^2-\frac{d\eta(x)}{dx}]$
$\frac{d(\eta^*(x))}{dx} > 0$ iff $(\lambda+\delta)f(x)\eta(x) > [1-(\lambda+\delta)F(x)+\delta F(s)][(\eta(x))^2-\frac{d\eta(x)}{dx}]$.
Lastly, for $x \ge t$ we have

$$\eta^*(x) = \frac{\delta\overline{F}(x)\eta(x)}{1-\lambda F(t)-\delta F(x)+\delta F(s)}.$$

Then $\eta^*(x)$ is increasing in x iff its derivative $\frac{d\eta^*(x)}{dx}$ is greater than 0.
Here

$$\begin{aligned}\frac{d\eta^*(x)}{dx} &= \frac{d[\frac{\delta\overline{F}(x)\eta(x)}{1-\lambda F(t)-\delta F(x)+\delta F(s)}]}{dx} \\ &= \frac{[1-\lambda F(t)-\delta F(x)+\delta F(s)][-\delta f(x)\eta(x)+\delta\overline{F}(x)\frac{d\eta(x)}{dx}]}{[1-\lambda F(t)-\delta F(x)+\delta F(s)]^2} \\ &\quad +\frac{(\delta)^2 f(x)\overline{F}(x)\eta(x)}{[1-\lambda F(t)-\delta F(x)+\delta F(s)]^2}\end{aligned}$$

$\frac{d\eta^*(x)}{dx} > 0$ iff $[1-\lambda F(t)-\delta F(x)+\delta F(s)][-\delta f(x)\eta(x)+\delta\overline{F}(x)\frac{d(\eta(x))}{dx}]+(\delta)^2 f(x)\overline{F}(x)\eta(x) > 0$ i.e., $\frac{d\eta^*(x)}{dx} > 0$ iff $(\delta)^2 f(x)\overline{F}(x)\eta(x) > \delta\overline{F}(x)[1-\lambda F(t)-\delta F(x)+\delta F(s)][(\eta(x))^2 - \frac{d(\eta(x))}{dx}]$ i.e., $\frac{d\eta^*(x)}{dx} > 0$ iff $(\delta) f(x)\eta(x) > [1-\lambda F(t)-\delta F(x)+\delta F(s)][(\eta(x))^2 - \frac{d\eta(x)}{dx}]$.
Thus $\eta(x)$ is increasing in x in the respective intervals iff

$$\begin{aligned}
(\lambda f(x)\eta(x) &> [1-\lambda F(x)][(\eta(x))^2 - \frac{d\eta(x)}{dx}] \\
(\lambda+\delta) f(x)\eta(x) &> [1-(\lambda+\delta)F(x)+\delta F(s)][(\eta(x))^2 - \frac{d\eta(x)}{dx}] \\
(\delta) f(x)\eta(x) &> [1-\lambda F(t)-\delta F(x)+\delta F(s)][(\eta(x))^2 - \frac{d\eta(x)}{dx}].
\end{aligned}$$

Note that in the above, if the first and the third inequalities hold then the second inequality automatically holds.

In view of this, $\eta(x)$ is increasing (decreasing) in respective intervals iff

$$\max[((\eta(x))^2 - \frac{d\eta(x)}{dx}), 0] < (>) \min[\frac{\delta f(x)\eta(x)}{1-\lambda F(t)-\delta F(x)+\delta F(s)}, \frac{\lambda f(x)\eta(x)}{1-\lambda F(x)}]$$

that is,

$$\max[(\eta(x) - (\frac{d\ln\eta(x)}{dx})), 0] < (>) \min[\frac{\delta f(x)}{1-\lambda F(t)-\delta F(x)+\delta F(s)}, \frac{\lambda f(x)}{1-\lambda F(x)}].$$

The proof is complete.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 85–95

Block Total Response Technique for Quantitative Sensitive Features in a Finite Population

Karabi Nandy[1] **and Bikas K. Sinha**[2]
[1] *Department of Population and Data Sciences, UT Southwestern Medical Center, Dallas, U.S.A.*
[2] *Formerly at Indian Statistical Institute, Kolkata, India.*

Received: 05 June 2020; Revised: 24 June 2020; Accepted: 26 June 2020

Abstract

The issue of eliciting truthful answers from survey respondents on sensitive questions has always been a challenge. Survey statisticians have developed various techniques to address this issue. Randomized response technique (RRT), originating in 1965 due to Warner, is a popular method in this area.

Block total response technique (BTRT), due to Raghavrao and Federer in 1979, is a method that incorporates experimental design features into RRT with the goal of increasing respondents' anonymity, in addition to producing unbiased estimators of parameters involving sensitive features. In this paper, we have developed an innovative estimator of the population mean of a sensitive feature using a permutation mechanism in the BTRT framework. This enables us also to compute an unbiased estimate of the variance of the proposed estimator.

Key words: Sensitive qualitative feature(s); Sensitive quantitative feature(s); Randomized response; Block total response technique; Random permutations.

1. Measuring Sensitive Characteristics Through Surveys: A Brief Review

Eliciting truthful responses on sensitive issues/characteristics from survey respondents has always been a challenge. During the latter half of the twentieth century, survey statisticians have proposed various methods of conducting surveys that provide anonymity to respondents and encourage them to answer truthfully to such sensitive issues. This enables gauging the level of the sensitive issue in the population, overcoming the biases that arise from false reporting. At the head of these survey methods is the popular randomized response technique (RRT) of interviewing, proposed by Warner in 1965. At the heart of this technique is a randomizing device, which is used to direct the respondent to provide answer to either a random (non-sensitive) question or to the sensitive question. It is of paramount importance that only the respondent knows how the randomization device directed him to respond and he provides only his answer to the interviewer without letting on how the device

Corresponding Author: Karabi Nandy
Email: Karabi.Nandy@UTSouthwestern.edu

directed him to respond. The underlying assumption is that since the respondent provides an answer to the interviewer without exposing his personal situation, any stigma associated with the sensitive question will be abated and the respondent will feel encouraged to respond truthfully. Warner showed that the responses obtained from this process will enable obtaining reliable estimates of the population parameter without direct knowledge of the responses obtained from individual respondents.

The RRT literature has focused on developing innovative randomization devices for both qualitative and quantitative characteristics. Examples of sensitive qualitative characteristics, particularly in the realm of public health, are use of contraceptive methods during sexual activity (yes/no), illicit substance use (yes/no), ever had an induced abortion (yes/no), ever had suicidal thoughts (yes/no) etc. Examples of sensitive quantitative features are the number of sexual partners one has, the number of times a person used illicit substance in the last month, the amount of time spent in a correctional facility etc. In addition to Warner's seminal work, we refer to a few book chapters and journal publications on RRT for details: Chaudhuri and Mukerjee (1987, 1988), Hedayat and Sinha (1991), Chaudhuri (2011) and Chaudhuri and Christofides (2013). Fifty years since it was first introduced, a celebratory Golden Jubilee Volume on RRT was compiled by Chaudhuri *et al.* in 2016 in a volume of the Handbook of Statistics.

In this paper, we are interested in quantitative sensitive characteristics; so we will focus our discussion henceforth for such sensitive characteristics only.

Greensberg *et al.* (1971) presented the first work involving RRT for continuous sensitive characteristics. Several others followed since, such as Eriksson (1973), Pollock and Bek (1976), Anderson (1977), *etc.* In recent years, Diana and Perri (2011) showcased the use of auxiliary information for estimating the mean of quantitative sensitive data and compared different models from both the perspectives of gaining efficient estimators as well as protecting respondents' anonymity. In 2015, Bose's work dealt with estimating the population mean of a sensitive feature wherein it is assumed that the true population values are captured by possibly a superset of M known quantities $[T_1, T_2, \ldots, T_M]$.

In the various approaches to extracting truthful responses on sensitive issues in surveys, an alternative method was proposed by Raghavarao and Federer (1979) where the idea was to incorporate basic experimental design elements in this framework. Block total response technique (BTRT) was suggested, based on the use of supplemented block designs / balanced incomplete block designs / spring balance weighing designs. In the context of a survey design, a "block" may be thought of as a questionnaire, containing a subset of the total number of questions, selected from a pool of questions which includes the sensitive question(s) as well. Of course, a given block may or may not contain the sensitive question(s).

Henceforth, we will closely follow the methodology suggested by Raghavarao and Federer (1979) and adopted in Nandy *et al.* (2016) and elsewhere. We shall use the BTRT framework to develop an estimator of the population mean for a quantitative sensitive characteristic. The entire exercise of sampling and estimation is geared towards unbiased estimation of the parameter under consideration. Layout of the rest of the paper is as follow. In section 2, we present some of the BTRT literature on quantitative sensitive characteristics.

In section 3, we introduce a BTRT version for estimating the mean for a sensitive quantitative item, similar to the methodology presented in Nandy *et al.* (2016) for qualitative items. In section 4, we provide an extension of our methodology that may potentially provide increased protection of respondents' privacy. Finally in section 5, we present some concluding remarks.

2. Methods to Obtain Truthful Responses for Sensitive Quantitative Items Through Survey

2.1. The development of BTRTS for sensitive quantitative items

As mentioned in the introduction, the BTRT method was suggested by Raghavarao and Federer (1979) as an alternative to the RRT. What stands out about this method is the increased protection of respondents' privacy when answering to sensitive questions. In fact, Coutts and Jann (2011) compared various RRT methods to BTR and showed that BTR outperformed the RRTs in terms of increased respondents' trust, better understanding of the interview instructions, lesser time to answer as well as lower non-response rates.

After its introduction, subsequent works in BTR focused on how to incorporate multiple sensitive questions into the design as well as development of various scoring mechanisms, i.e., how to score questions, other than in a binary fashion, so as to further increase respondents' privacy. The latter is rooted in the idea that a total score could be incriminating if that score could only be achieved by answering "yes" to at least one sensitive question. The works of Smith and Street (2003) and Smith (2005) are some examples of these. In 2016, we undertook various meaningful versions/generalizations of the BTRT and introduced empirical Bayes estimators.

All advancements in this area, however, have been for qualitative sensitive characteristics. In this work, we propose to present BTRT for quantitative sensitive characteristics. For this, our starting points are our own work in Nandy *et al.* (2016), Bose (2015) and Mukherjee *et al.* (2018). We present some details related to the latter two in the following subsection.

2.2. Study of a sensitive quantitative item under general sampling scheme

According to Bose (2015), we assume that truthful unknown responses $[Y_1, Y_2, \ldots, Y_N]$ are captured by possibly a superset of M known quantities $[T_1, T_2, \ldots, T_M]$. Therefore, quantitative nature of the sensitive feature is only to the extent of being discrete-valued. In effect, therefore, a finite population Y-distribution refers to a frequency distribution of the T's such as $[N_1, N_2, \ldots, N_M; \sum_i N_i = N]$ or, in other words, it refers to a probability distribution $[w_1, w_2, \ldots, w_M]$ where $w_i = N_i/N; i = 1, 2, \ldots, M$. An RRT is now geared towards unbiased estimation of the w's - using a suitable randomization device as described below.

Choose a fraction δ and a total of R chips such that $R_0 = \delta R$ chips [among the R chips] read as "Report T" in case the respondent happens to choose any of these R_0 chips. Further to this, a set of $R_i = (1 - \delta)R/M$ chips reads as "Report T_i", if the respondent happens to choose any one of these R_i chips $(i = 1, 2, \ldots, M)$. It is now clear that each respondent is supposed to select at random one chip and act accordingly by responding truthfully - without

divulging the type of chip selected. The chosen chip is returned back to the collection each time. That is how we generate the data under RRT.

The observed proportions of the Ts, say $p_1, p_2, \ldots, p_M$ are random and it turns out that

$$E(p_i) = \delta w_i + (1-\delta)/M; i = 1, 2, \ldots, M.$$

This suggests that we can unbiasedly estimate w_i as $[p_i - (1-\delta)/M]/\delta$ for $i = 1, 2, \ldots, M$. Therefore, the population mean $[=\sum_i w_i T_i]$ is unbiasedly estimated. It can be seen that SRSWR sampling of respondents entails us to regard the responses as being independently and identically distributed (iid). This simplifies the data analysis significantly.

Bose also gave an expression for variance of the estimate of the population mean. Variance estimation is not considered there. It follows that this method has an inherent limitation in that it does not address the estimation problem in case the sampling design is fixed size (N, n) sampling design such as SRSWOR (N, n) or any arbitrary sampling design. The iid nature of the responses is highly restrictive to do away with.

Next, Mukherjee *et al.*. (2018) undertook this study in its most general form. They provided formulae for mean estimation, expression for variance of the estimate and a method for variance estimation as well.

The idea is to provide an unbiased estimate of Y_i^* - the true Y-value on the sensitive feature, associated with the i-th respondent - for every selected respondent i in the sample. Once the respondent i has been selected and has been asked to provide RR $[R_i]$ by making a random choice of one chip and acting accordingly, it turns out that

$$E(R_i) = \delta Y_i^* + (1-\delta)\sum_i T_i/M.$$

Therefore,

$$\hat{Y}_i^* = [R_i - (1-\delta)\sum_i T_i/M]/\delta.$$

It is now simple to obtain the Horvitz Thomson Estimator (HTE) $\sum_i \hat{Y}_i^*/N\pi_i$ for the population mean for any arbitrary fixed size (N, n) sampling design - the choice of the sampling design being subject to providing positive first and second order inclusion probabilities *etc.* For the proof of unbiasedness, deduction of the expression for variance as well as estimated variance, we may refer to Mukherjee *et al.* (2018).

It must, however, be noted that the discrete nature of the quantitative feature is still maintained. That is a major limitation of the study so far described.

3. Introducing Block Total Response Technique for Quantitative Sensitive Feature

We now proceed to discuss BTR Technique for unbiased estimation of the population mean for a sensitive quantitative feature Q^* with true values $Y(Q^*)$, apriori known to the respondents. Also are apriori known values of each of a set of v other quantitative ordinary [non-sensitive] features $Q_1, Q_2, \ldots, Q_v$ to the respondents. We draw a random sample of n respondents, following SRSWOR(N, n) sampling. We no longer require that the Y-values be

discrete. At best, it may be convenient to make a choice of the ordinary features Qs such that their ranges broadly cover the range of values of the sentitive feature Q^*.

We employ BTR technique in the following manner. We start with a Binary Proper Equireplicate Block Design [BPEBD] involving b blocks, each of size k, with equal replication number r of each of the v non-sensitive Qs. We then supplement each such block with one additional question, *viz.*, the sensitive question Q^*. Thus each of the b blocks of size k is 'extended' to one of size $(k+1)$. We also introduce an additional block B_0 of size v - incorporating all the v non-sensitive Qs.

The respondents in the sample are randomly split into $(b+1)$ sets of sizes $n^*, n^*, \ldots, n^*, n_0$. We assume that the sample size n has a convenient integer decomposition $n = (bn^* + n_0)$ for suitably chosen integers n^*, n_0. The b blocks each receive n^* respondents and each respondent provides only the sum total of responses to $(k+1)$ questions, the k non-sensitive questions included in the specific block in which the respondent belongs, along with the sensitive feature Q^*. The same is also true of the last block B_0 - although all the v features in this block are non-sensitive in nature. Our goal is to obtain an estimate of $\bar{Y}(Q^*)$.

In this context, we are tempted to quote Raghavarao and Federer (1979): "One early anonymous-direct-question method that was used successfully (e.g. by A. J. King and others at Iowa State University) was to have the respondent complete an unmarked questionnaire in secret and to deposit the questionnaire in a large locked box in which other questionnaires had been deposited; then, the respondent observed that the contents of the box were thoroughly mixed. We shall call this method the 'black box' (BB) method."

In our context, we may refer to $(b+1)$ such black boxes in a meaningful manner.

3.1. Estimation of $\bar{Y}(Q^*)$

At this stage, let us consider an illustrative example with $b = 5, v = 10, r = 2, k = 4, n = 350, n^* = 50, n_0 = 100$. Let further the blocks of the BPEBD be formed as seen in Table 1 below.

Table 1: Blocks in the BPEBD

Block	Non-sensitive Features	Sensitive Feature
B_1	Q_1, Q_2, Q_3, Q_4	Q^*
B_2	Q_5, Q_6, Q_7, Q_8	Q^*
B_3	Q_9, Q_{10}, Q_1, Q_2	Q^*
B_4	Q_3, Q_4, Q_5, Q_6	Q^*
B_5	Q_7, Q_8, Q_9, Q_{10}	Q^*

In the first block, let us now compute the average of Block Total Responses - averaged over all the n^* respondents' BTR scores. Let us denote it by $\bar{BTR}(B_1)$. It follows that its expectation is given by $E[\bar{BTR}(B_1)] = \bar{Y}(Q_1) + \bar{Y}(Q_2) + \bar{Y}(Q_3) + \bar{Y}(Q_4) + \bar{Y}(Q^*)$ where $\bar{Y}(Q)$ refers to the population average of true values for the feature identified through Q. Likewise, we carry out the same for all blocks B_1 to B_5. Additionally, we work it out for the

last block B_0 as well.

Adding the results for the first 5 blocks, we obtain

$$E[\sum_i B\bar{T}R(B_i)] = 2[\bar{Y}(Q_1) + \bar{Y}(Q_2) + \ldots + \bar{Y}(Q_{10})] + 5\bar{Y}(Q^*)$$

while

$$E[B\bar{T}R(B_0)] = \bar{Y}(Q_1) + \bar{Y}(Q_2) + \ldots + \bar{Y}(Q_{10}).$$

From the above, we deduce

$$\hat{\bar{Y}}(Q^*) = \frac{[\sum_i B\bar{T}R(B_i)] - 2[B\bar{T}R(B_0)]}{5}.$$

3.2. Estimation of $V(\hat{\bar{Y}}))$

In order to work out variance estimate of this estimate of the population average of the sensitive feature Q^*, we propose to develop an important representation of the estimate derived above. For this, we assume that the respondents' responses are not associated with others' responses and that, to most extent, the respondents behave independently - so far as the responses are concerned.

We provide below an extensive use of permutation groups. Let P_1 denote a random permutation of size n^* of the integers $1, 2, \ldots, n^*$ associated with the labels of the respondents in Block 1. Likewise, we develop independently all other permutations P_2 to P_b and lastly, P_0 of size n_0 for the block B_0.

Now we group the responses across the $b+1$ blocks in sets of $b+1$ - taking one from each of the b blocks and 2 from the last block B_0. Once more we illustrate this feature by referring to the above example. We choose, for example, $n^* = 50$ and $n_0 = 100$ so that $n = 350$.

P_1 =[44 18 17 14 26 38 19 34 30 37 7 1 20 39 11 3 31 22 46 23 9 28 10 8 12
4 16 27 32 40 29 49 21 48 5 13 15 43 50 2 41 25 35 45 33 36 47 42 6 24];
P_2 =[4 30 16 14 38 46 21 39 32 13 49 19 20 2 48 47 17 31 9 50 27 44 35 6 40
3 10 12 37 11 8 29 1 22 26 24 33 7 34 18 45 23 42 36 43 5 28 15 25 41];
P_3 =[7 8 44 21 39 38 4 43 19 11 45 48 26 3 10 31 15 49 30 25 16 17 46 14 2
5 28 32 1 41 47 40 20 34 27 18 9 13 24 50 36 37 23 33 42 22 12 29 35 6];
P_4 =[13 29 41 11 36 40 46 31 3 48 50 30 7 14 23 21 25 8 9 32 2 37 28 1 42 33
20 45 49 19 12 16 44 43 38 15 39 24 26 4 22 10 17 27 34 6 18 35 5 47];
P_5 =[2 37 13 47 27 21 32 1 22 43 20 33 36 24 28 16 9 35 19 15 31 44 23 41
30 29 5 14 4 49 34 42 48 12 18 6 10 46 17 26 39 7 3 45 25 38 50 11 8 40];

P_0 =[9 10 95 26 18 61 21 60 57 8 67 70 73 2 46 54 100 80 17 40 5 4 77 19 87 52
76 25 81 35 55 14 50 37 29 69 38 89 98 90 59 12 68 7 53 16 75 39 94 48 42
32 56 36 41 96 82 65 78 62 74 93 86 3 97 13 47 49 63 88 85 43 51 30 91 15
58 22 64 71 33 1 27 45 28 20 84 11 23 44 99 34 24 6 31 66 79 92 72 83];

According to the above permutations applied to different blocks, *Set* 1 comprises of responses of 7 respondents labeled (1) 44 in B_1, (2) 4 in B_2, (3) 7 in B_3, (4) 13 in B_4, (5) 2 in B_5 and (6,7) $9, 10$ in B_0. An estimator for $\bar{Y}(Q^*)$ based on this data *Set* 1 is given by

$$\begin{aligned}
1/5 \times [&\{Y(44;Q_1) + Y(44;Q_2) + Y(44;Q_3) + Y(44;Q_4)\} \\
&+ \{Y(4;Q_5) + Y(4;Q_6) + Y(4;Q_7) + Y(4;Q_8)\} \\
&+ \{Y(7;Q_9) + Y(7;Q_{10}) + Y(7;Q_1) + Y(7;Q_2)\} \\
&+ \{Y(13;Q_3) + Y(13;Q_4) + Y(13;Q_5) + Y(13;Q_6)\} \\
&+ \{Y(2;Q_7) + Y(2;Q_8) + Y(2;Q_9) + Y(2;Q_{10})\} \\
&- \{Y(9;Q_1) + Y(9;Q_2) + \ldots + Y(9;Q10)\} \\
&- \{Y(10;Q_1) + Y(10;Q_2) + \ldots + Y(10;Q10)\}].
\end{aligned}$$

We proceed in this manner and obtain 50 estimates of $\bar{Y}(Q^*)$ based on the 50 sets as defined above. Because of the underlying permutation principle, these estimates are also exchangeable in nature. This characterization of the individual estimates lends itself to easy computation of their average, which is the estimate of the population mean. Further, variance estimation becomes a trivial task: $\sum_i (e_i - \bar{e})^2/n(n-1)$ is an unbiased variance estimate of $\bar{e}$ based on iid estimates es.

Even though the respondents are selected according to $SRSWOR(N, n)$, use of permutations within blocks enables us to justify the assumption of iid nature of the estimates based on different sets of data. It is not however clear if the same holds true for any arbitrary fixed size (n) sampling design.

In the above, we assumed the condition: $n = bn^* + n_0$ for suitably chosen integers n^*, n_0. It is possible to relax this condition and instead work with another representation. We reconsider the above example to illustrate this point.

Once again, we start with $n = 350$ but assume the representation: $350 = 30+40+50+60+70+100$. Note that there is a common divisor of 10 among all the respondent group sizes. This time we can assemble the sets so that we have 10 iid estimates of the parameter of interest, *e.g.*, mean of the sensitive feature Q^*. Once these formations are done, the rest is routine in terms of computation of mean and variance of iid estimates.

We describe the essential step below with reference to the first of the 10 sets of estimates. The sizes of the blocks will be the highest common factor, which is 10 in this case. This suggests (i) deriving random permutations of the respondent labels within each block; (ii) forming 10 subsets of equal size within each block. Note that subset sizes will vary across

the blocks; (iii) forming unbiased estimates for the mean of the sensitive feature $[Q^*]$ from subsets collected serially across all the blocks; (iv) using iid sample estimates to arrive at the over-all average *etc.*

We carry out the exercise below. The subsets within each block, after random permutation, are shown within parenthesis. Also the block sizes are indicated in parenthesis.

$P_1(1-30)$ =[(30 7 1); (20 11 3); (22 23 9); (28 10 8); (12 4 16);
(27 29 21); (5 13 15); (2 25 6); (24 18 17); (14 26 19)];

$P_2(1-40)$ =[(13 19 20 2); (17 31 9 27); (35 6 40 3); (10 12 37 11); (8 29 1 22);
(26 24 33 7); (34 18 23 36); (5 28 15 25); (4 30 16 14); (38 21 39 32)];

$P_3(1-50)$ =[(7 8 44 21 39); (38 4 43 19 11); (45 48 26 3 10); (31 15 49 30 25);
(16 17 46 14 2); (5 28 32 1 41); (47 40 20 34 27); (18 9 13 24 50);
(36 37 23 33 42); (22 12 29 35 6)];

$P_4(1-60)$ =[(16 39 48 42 32 56); (36 41 3 13 47 49); (43 51 30 15 58 22);
(33 1 27 45 28 20); (11 23 44 34 24 6 31); (35 55 14 50 37 29);
(38 59 12 7 53 9); (10 26 18 21 60 57); (8 2 46 54 17 40); (5 4 77 19 52 25)]

$P_5(1-70)$ =[(3 13 47 49 63 43 51); (30 15 58 22 64 35 55); (14 50 37 29 69 38 59);
(12 68 7 53 33 1 27); (45 28 20 11 23 44 34); (24 6 31 66 9 10 26);
(18 61 21 60 57 8 67); (70 2 46 54 17 40 5); (4 62 19 52 25 16 39);
(48 42 32 56 36 41 65)].

$P_0(1-100)$ is the 10 subsets formed taking 10 permutations at a time and serially - starting from the left corner.

We display the result based on data analysis for *Set* 1 across all the 6 blocks. For B_1, we consider the first set of 3 respondents labeled $(1, 7, 30)$ and average out the BTRs collected from them. So,

$$E(Set1) = \bar{Y}(Q_1) + \bar{Y}(Q_2) + \bar{Y}(Q_3) + \bar{Y}(Q_4) + \bar{Y}(Q^*).$$

Likewise, we have similar results from the first set of all other blocks. We denote these averaged responses by $R(B_1; S_1), R(B_2 : S_1), \ldots, R(B_6; S_1)$. It now follows that

$$\hat{\bar{Y}}(Q^*; S_1) = [R(B_1; S_1) + R(B_2; S_1) + \ldots + R(B_5; S_1) - 2R(B_6; S_1)]/5.$$

We will be referring to 10 such sample estimates and proceed to compute the combined estimate of the population mean of Q^* and its estimated standard error.

So,

$$\hat{\bar{Y}}(Q^*; S_2) = [R(B_1; S_2) + R(B_2; S_2) + \ldots + R(B_5; S_2) - 2R(B_6; S_2)]/5,$$
$$\hat{\bar{Y}}(Q^*; S_i) = \ldots.$$
$$\hat{\bar{Y}}(Q^*; S_{10}) = [R(B_1; S_{10}) + R(B_2; S_{10}) + \ldots + R(B_5; S_{10}) - 2R(B_6; S_{10})]/5.$$

Therefore,

$$\hat{\bar{Y}}(Q^*) = 1/10[\hat{\bar{Y}}(Q^*; S_1) + \hat{\bar{Y}}(Q^*; S_2) + \ldots + \hat{\bar{Y}}(Q^*; S_{10})].$$

Further, estimated standard error is computed as usual by taking square root of

$$\sum(\hat{\bar{Y}}(Q^*; S_i) - \hat{\bar{Y}}(Q^*))^2/10 \times 9.$$

4. An Extention of the BTRT Method

In the technique presented above, in every block [B_1 to B_b], we are utilizing some k of the v Qs - leaving the rest unutilized. When k is small, respondents may not feel comfortable responding truthfully since responding to Q^* is mandatory. In this section, we provide an extension of the above technique, as follows.

For every block out of B_1 to B_b, we bring a variation in the block composition as: 1. List of k "must respond" Q's - these are kept in Part A. This is the same as before. 2. Remaining $(v-k)$ Q's and Q^* are kept in Part B. From Part B, a respondent is to make a random choice of exactly one question from the total $(v-k+1)$ questions; next the respondent will blend the selected question with those in Part A and provide BTR - without divulging the nature of the question selected from Part B. To simplify the data analysis, it may be assumed that selection from Part B is governed by the rule: Select Q^* with probability δ and any one of the remaining Q's with probability $(1-\delta)/(v-k)$. Further, we use the same selection mechanism in each such block.

Once more, we can study the formation of estimates based on the sets separately and then combine them. We display the result for Set 1 below for the example considered above and with the choice $\delta = 0.4$. Accrued Block Totals provide for the first 5 blocks, the following expressions for their expectations under random choice of the question selected from Part B:

$$[\{Y(44;Q_1)+Y(44;Q_2)+Y(44;Q_3)+Y(44;Q_4)\}+0.1\{Y(44;Q_5)+\ldots+Y(44;Q_{10})\}+0.4Y(44;Q^*);$$

$$\{Y(4;Q_5)+Y(4;Q_6)+Y(4;Q_7)+Y(4;Q_8)\}+0.1\{Y(4;Q_9)+\ldots+Y(4;Q_4)\}+0.4Y(4;Q^*);$$

$$\{Y(7;Q_9)+Y(7;Q_{10})+Y(7;Q_1)+Y(7;Q_2)\}+0.1\{Y(7;Q_3)+\ldots+Y(7,Q_8)\}+0.4Y(7;Q^*);$$

$$\{Y(13;Q_3)+Y(13;Q_4)+Y(13;Q_5)+Y(13;Q_6)\}+0.1\{Y(13;Q_7)+\ldots+Y(13;Q_2)\}+0.4Y(13;Q^*);$$

$$\{Y(2;Q_7)+Y(2;Q_8)+Y(2;Q_9)+Y(2;Q_{10})\}]+0.1\{Y(2;Q_1)+\ldots+Y(2;Q_6)\}+0.4Y(2;Q^*)$$

As for block B_0, we obtain $[Y(9;Q_1)+Y(9;Q_2)+\ldots+Y(9;Q_{10})+Y(10;Q_1)+Y(10;Q_2)+\ldots+Y(10;Q_{10})]$.

From the above, it is routine to obtain an estimate for the average $Y(Q^*)$. Once such estimates are computed from each set, they may be treated as iid sample estimates and hence mean estimation and variance estimation are immediate.

5. Concluding Remarks

In the context of sensitive quantitative features, we have proposed a version of block total response technique which has flexibility in terms of implementation. We expect that the scheme in section 4 will likely provide increased privacy protection to respondents, compared to the BPEBD scheme in Section 3 which involves only the sensitive item Q^* in the supplementary part. We plan to quantify this increase in future work. It may be noted that in most practical surveys, collection of data is on several variables, which are then used to estimate not just marginal distributions but also joint distributions, correlations, regressions *etc*. Deriving joint inferences from data gathered using BTRT would be interesting. We plan to explore this in future study. The innovativeness of the method studied here lies in procuring an estimate as well as estimated standard error by exploiting a permutation method to generate exchangable observations. The results have been deduced under SRSWOR sampling. For a general fixed size sampling design, we have yet to develop a version of BTRT. This is true for both qualitative and quantitative sensitive features.

References

Anderson, H. (1977). Efficiency versus protection in a general randomized response model. *Scandinavian Journal of Statistics*, **4**, 11–19.

Bose, M. (2015). Respondent privacy and estimation efficiency in randomized response surveys for discrete-valued sensitive variables. *Statistical Papers*, **56**, 1055-1069.

Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys.* CRC Press, Boca Raton, FL.

Chaudhuri, A. and Mukerjee, R. (1987). Randomized response techniques: a review. *Statistica Neerlandica*, **41(1)**, 27–44.

Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Responses: Theory and Techniques.* Marcel Dekker, New York, NY.

Chaudhuri, A. and Christofides, T. C. (2013). *Indirect Questioning in Sample Surveys.* Springer, Germany.

Series Volume Editors: Chaudhuri, A., Christofides T. C. and Rao C. R. (2015). *Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits.* Handbook of Statistics; volume **4**.

Chaudhuri A. and Roy D. (1997). Model assisted survey sampling strategies with randomized response. *Journal of Statistical Planning and Inference*, **60**, 61–68.

Coutts, E. and Jann, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods and Research*, **40(1)**, 169–193.

Diana G. and Perri P. F. (2011). A class of estimators for quantitative sensitive data. Statistical Papers, **52**, 633–650.

Eriksson S. A. (1973). A new model for randomized response. *International Statistical Review*, **41**, 101–113.

Guerriero M. and Sandri M. F. (2007). A note on the comparison of some randomized response procedures. *Journal of Statistical Planning and Inference*, **137**, 2184–2190.

Pollock K. H. and Bek Y. (1976). A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association*, **71**, 884–886.

Hedayat, A. S. and Sinha, B. K. (1991). *Design and Inference in Finite Population Sampling.* Wiley, New York, NY.

Mahajan P. K., Gupta J. P. and Singh R. (1994). Determination of optimum strata boundaries for scrambled response. *Statistica*, **54**, 375–381.

Mukherjee, S., Sinha, B. K. and Chatterjee, A. (2018). *Statistical Methods in Social Science Research.* Springer.

Nandy, K., Marcovitz, M. and Sinha, B. K. (2016). Eliciting Information on Sensitive Features: Block Total Response Technique and Related Inference. In 'Handbook of Statistics, Special Volume on Golden Jubilee Celebration on Randomized Response Technique'. Chapter 19, pp 317-330.

Nayak, T. K. and Adeshiyan, S. A. (2009). A unified framework for analysis and comparison of randomized response surveys of binary characteristics. *Journal of Statistical Planning and Inference*, **139**, 2757–2766.

Raghavarao, D. and Federer, W. T. (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of Royal Statistical Society B Methodology*, **41**, 40–45.

Singh S., Joarder A. H. and Kinh M. L. (1996). Regression analysis using scrambled responses. *Australian Journal of Statistics*, **38**, 201–211.

Sinha B. K. (2017). Some refinements of block total response technique in the context of RRT methodology. *Statistics and Applications*, **15(1)**, pp 167–171.

Smith N. F. (2005). The design of scoring schemes for surveys using the block total response method. *Communications in Statistics - Theory and Methods*, **34(11)**, 2157–2168.

Smith, N. F. and Street, D. J. (2003). The use of balanced incomplete block designs in designing randomized response surveys. *Australian and New Zealand Journal of Statistics*, **45(2)**, 181–194.

Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, **60**, 63–69.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 97–119

The Simultaneous Assessment of Normality and Homoscedasticity in One-Way Random Effects Models

Ye Yang and Thomas Mathew
Department of Mathematics and Statistics
University of Maryland Baltimore County, Maryland, USA

Received: 23 May 2020; Revised: 26 June 2020; Accepted: 28 June 2020

Abstract

The article investigates the simultaneous assessment of normality and homoscedasticity in a one-way random effects model. Test procedures are developed assuming a smooth alternative to the normal distribution, specified using Legendre polynomials and Hermite polynomials. Score statistics are derived under both classes of alternatives, and a data driven approach is used to determine the order of the polynomials. Numerical results are reported in order to assess the accuracy of the chisquare distribution as the null distribution of the score statistics. Estimated powers are reported in order to compare the score tests derived under the alternatives based on Legendre polynomials and Hermite polynomials. An example and the corresponding data analysis are reported in order to illustrate the results. Possible extensions to other models involving random effects are briefly indicated.

Key words: Hermite polynomials; Legendre polynomials; Score test; Smooth alternative.

AMS Subject Classifications: 62F03, 62J10

1. Introduction

Mixed and random effects models are among the most widely used tools in applied work. While analyzing data using such models, the standard assumptions include normality of the random effects and the experimental error terms, as well as homoscedasticity of the error distribution. The present work is focused on testing these assumptions under the simplest random effects model, namely, the one-way random effects model. The usual practice is to test these assumptions separately. For example, if normality can be assumed, homoscedasticity can be assessed using a formal test such as Bartlett's test. One can also use a test that is less sensitive to the normality assumption, for example the Levene test and the modified Levene test; see the article by Chang, Pal, and Lin (2017), and Section 3.4 in the book by Montgomery (2020). On the other hand, normality is often assessed using formal tests or using a graphical method such as the normal probability plot, after assuming homoscedasticity. It is certainly desirable to have test procedures that will permit us to simultaneously assess homoscedasticity and normality. The present work aims to develop such test procedures.

Corresponding Author: Thomas Mathew
Email: mathew@umbc.edu

In order to formally introduce the relevant hypothesis, consider data falling into a groups (for example, corresponding to a treatments in a designed experiment). Suppose we have n_l observations available from the lth group, with y_{lj} denoting the jth observation; $j = 1, 2, \cdots, n_l$, $l = 1, 2, \cdots, a$. We are allowing the n_l's to be unequal, so that we can have unbalanced data. The one-way random model for the y_{lj}, along with the normality assumptions, is given by

$$y_{lj} = \mu + \alpha_l + \epsilon_{lj}, \quad \alpha_l \sim N(0, \sigma_\alpha^2),\ \epsilon_{lj} \sim N(0, \sigma_l^2), \tag{1}$$

where all the random variables are assumed to be independent. Note that the error terms (*i.e.*, the ϵ_{lj}'s) have variance σ_l^2, which could differ across the a groups. The α_l's in (1) denote the random effects, $l = 1, 2, \cdots, a$. If we write $\boldsymbol{y}_l = (y_{l1}, y_{l2}, \cdots, y_{ln_l})'$, then the above assumptions imply

$$\boldsymbol{y}_l \sim N\left(\mu \mathbf{1}_{n_l}, \sigma_l^2 I_{n_l} + \sigma_\alpha^2 \mathbf{1}_{n_l} \mathbf{1}_{n_l}'\right), \tag{2}$$

where $\mathbf{1}_r$ is an $r \times 1$ vector of ones. We note that the $\boldsymbol{y}_l$'s are independent for $l = 1, 2, \cdots, a$. Clearly, data analysis based on the one-way random model under the standard assumptions of normality and homoscedasticity amounts to analyzing the data under the multivariate normal model (2), having the structured covariance matrix as specified, and having σ_l^2s all equal. Thus testing homoscedasticity and normality under the one-way random effects model is equivalent to testing the equality of the σ_l^2 along with the multivariate normality of the $\boldsymbol{y}_l$, $l = 1, 2, \cdots, a$, where the covariance matrix has the structure specified in (2). Consequently, our normality assessment is for the multivariate normality of the $\boldsymbol{y}_l$, $l = 1, 2, \cdots, a$, and not for the univariate normality of the random effects and the error terms, even though the latter implies the multivariate normal distribution in (2).

Our development relies on the specification of alternatives to normality to be the class of *smooth alternatives* proposed by Neyman (1937). In general, suppose the problem is to test if a continuous random variable Y follows a specified distribution having density, say $f(y, \boldsymbol{\beta})$, depending on an unknown parameter vector $\boldsymbol{\beta}$. The alternative hypothesis is specified in terms of a smooth alternative involving orthonormal polynomials, say $\{p_i(y; \boldsymbol{\beta})\}$, $i = 1, 2, \cdots$, that are orthonormal with respect to $f(y; \boldsymbol{\beta})$. An order k smooth alternative, say $g_k(y; \boldsymbol{\theta}, \boldsymbol{\beta})$, is given by

$$g_k(y; \boldsymbol{\theta}, \boldsymbol{\beta}) = C(\boldsymbol{\theta}, \boldsymbol{\beta}) \exp\left\{ \sum_{i=1}^{k} \theta_i p_i(y; \boldsymbol{\beta}) \right\} f(y; \boldsymbol{\beta}). \tag{3}$$

In (3), $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)'$ is a vector of unknown parameters, and $C(\boldsymbol{\theta}, \boldsymbol{\beta})$ is a normalizing constant. As already noted, the $\{p_i(y; \boldsymbol{\beta})\}$ are orthonormal polynomials, orthonormal with respect to the null density $f(y; \boldsymbol{\beta})$. It should be clear that if $\boldsymbol{\theta}$ is the null vector, then $g_k(y; \boldsymbol{\theta}, \boldsymbol{\beta})$ in (3) reduces to the null density $f(y; \boldsymbol{\beta})$. In other words, the null density is embedded in the class of alternatives specified in (3), and testing for the goodness-of-fit of the null density is equivalent to testing if the $k-$dimensional vector $\boldsymbol{\theta}$ is the null vector. That is, the goodness-of-fit problem is now reduced to that of testing a hypothesis concerning a finite number of parameters.

Several authors have derived score tests for testing goodness-of-fit under the Neyman (1937) framework, and have investigated the theoretical properties of such tests for a variety of goodness-of-fit problems: see Ledwina (1994), Kallenberg and Ledwina (1995, 1997a, 1997b), Inglot and Ledwina (1996), Inglot, Kallenberg and Ledwina (1997), Kallenberg, Ledwina and Rafajlowicz (1997) and Janic and Ledwina (2009). Most of these articles address testing only the goodness of fit of a particular distribution; however, Kallenberg, Ledwina and Rafajlowicz (1997) address the simultaneous testing of normality and independence in a bivariate scenario. The simultaneous assessment of the various assumptions in a standard linear model set up is taken up in Peña and Slate (2006), and the authors consider the Neyman (1937) framework in order to specify alternatives to normality. More recently, Yang and Mathew (2018) have addressed the simultaneous testing of normality and homoscedasticity in a fixed effects model when we have grouped data, similar to those in an ANOVA context with fixed effects. A book-length discussion of smooth tests is available in Rayner, Thas and Best (2009).

In the next section, we shall derive score tests for testing normality and homoscedasticity in the set up of the model (2), assuming smooth alternatives of the type (3). We shall consider two specifications for the smooth alternative based on two choices for the orthonormal polynomials $\{p_i(y;\boldsymbol{\beta})\}$ in (3), namely, Legendre polynomials and Hermite polynomials. Thus we have two score tests corresponding to the two specifications for the alternatives. While specifying the alternatives, there is obvious arbitrariness in the choice of the order of the Legendre polynomials and Hermite polynomials, i.e., the quantity k in (3). We shall follow a *data driven* approach for choosing the order; an idea developed in Inglot, Kallenberg and Ledwina (1994), and pursued in some of the later papers by the authors, cited earlier. Our tests being based on score statistics, we can think of approximating the null distribution with a chisquare distribution. Thus we shall report numerical results in order to assess the accuracy of the chisquare distribution as the null distribution. The tests will be compared using estimated powers. Data analysis based on an example will be reported in order to illustrate the results.

2. Smooth Alternatives and Score Tests

Before we formally specify the alternative hypothesis, we shall consider an orthogonal transformation of each of the data vectors $\boldsymbol{y}_l$ in (2), $l = 1, 2, \cdots, a$. Let $Q_l = \left(\frac{1}{\sqrt{n_l}}1_{n_l}, Q_l^*\right)$ be an $n_l \times n_l$ Helmert matrix, and consider the transformation

$$\boldsymbol{v}_l = Q_l'\boldsymbol{y}_l,\ l = 1, 2, \cdots, a, \tag{4}$$

$$\text{so that } E(\boldsymbol{v}_l) = (\mu\sqrt{n_l}, 0, \cdots, 0)', \quad V(\boldsymbol{v}_l) = \sigma_l^2 I_{n_l} + \sigma_\alpha^2 \text{diag}(n_l, 0, 0, \cdots, 0).$$

Clearly, testing multivariate normality of the $\boldsymbol{y}_l$'s is equivalent to testing the same for the $\boldsymbol{v}_l$'s. Writing $\boldsymbol{v}_l = (v_{l1}, v_{l2}, \cdots, v_{ln_l})'$, we note that multivariate normality for the $\boldsymbol{y}_l$'s implies

$$v_{l1} \sim N(\mu\sqrt{n_l}, \sigma_l^2 + n_l\sigma_\alpha^2),\ v_{lj} \sim N(0, \sigma_l^2),\ j = 2, 3, \cdots, n_l, \tag{5}$$

$l = 1, 2, \cdots, a$, where the v_{lj}'s for $j = 1, 2, \cdots, n_l$, are also independent (in view of the diagonal covariance matrix of $\boldsymbol{v}_l$, noted above). We also note that the v_{lj}'s for $j = 1, 2, \cdots,$

n_l, are uncorrelated random variables, even if multivariate normality of the $\boldsymbol{y}_l$'s does not hold.

In the remainder of this section we shall test the equality of the σ_l^2's and the univariate normality of the v_{lj}'s against smooth alternatives defined through Legendre polynomials and Hermite polynomials. We shall do so assuming that the v_{lj}'s for $j = 1, 2, \cdots, n_l$, are all independent. That is, we are testing if the the v_{lj}'s for $j = 1, 2, \cdots, n_l$, are all independent normally distributed against the alternative that they are independent having a non-normal distribution defined through a smooth alternative. In other words, the class of smooth alternatives that we are considering is somewhat restricted in view of the independence assumption of the v_{lj}'s under the alternative. The advantage of the independence assumption is that the normality testing is now reduced to testing the univariate normality specified in (5). It should be noted that in an article on normality testing in a two-way random model with and without interaction, Xu, Li and Song (2013) reduced the problem to that of testing univariate normality, after transforming the data to uncorrelated components based on a transformation that depends on the unknown variance components. The authors then replaced the unknown variance components with estimates, and applied standard univariate normality tests, proceeding under the assumption that the transformed univariate components are independent even under the alternative. The transformation that we have used, based on the Helmert matrices Q_l, is of course parameter free.

2.1. Smooth alternatives based on Legendre polynomials

Let

$$z_{l1} = \frac{v_{l1} - \mu\sqrt{n_l}}{\sqrt{\sigma_l^2 + n_l\sigma_\alpha^2}},\ u_{l1} = \Phi\left(z_{l1}\right),\ z_{lj} = v_{lj}/\sigma_l,\ \text{ and }\ u_{lj} = \Phi\left(z_{lj}\right),\ j = 2, 3, \cdots, n_l, \quad (6)$$

where the z_{lj}'s are independent standard normal random variables for $j = 1, 2, \cdots, n_l$ and $l = 1, 2, \cdots, a$, and $\Phi(.)$ denotes the standard normal cdf. In this subsection, we shall specify smooth alternatives based on Legendre polynomials; we recall that these are polynomials that are orthonormal with respect to the uniform distribution in the interval (0, 1). Let $b_i(.)$, $i = 1, 2, \cdots$, denote the system of Legendre polynomials. It is easily verified that if $y \sim N(\mu, \sigma^2)$, then a system of orthonormal polynomials with respect to the $N(\mu, \sigma^2)$ density is obtained as $b_i\left(\Phi\left(\frac{y-\mu}{\sigma}\right)\right)$, $i = 1, 2, 3, \cdots$. In view of this, we conclude that for each fixed l and j, $b_i(u_{lj}) = b_i\left(\Phi(z_{lj})\right)$, $i = 1, 2, \cdots$, form a system of orthonormal polynomials with respect to the standard normal distribution, where $u_{lj} = \Phi\left(z_{lj}\right)$, as defined in (6). While specifying the smooth alternative, we will consider the case of only a common alternative across the a different groups. It is certainly possible to have different alternatives across the different groups, but we shall not consider this case.

In order to specify the likelihood function under a Legendre polynomial based smooth alternative, we note that the smooth alternative in (3) is specified in terms of a parameter vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)'$, and the null density corresponds to $\boldsymbol{\theta}$ being the null vector. Since we need to specify smooth alternatives for v_{l1}, $l = 1, 2, \cdots, a$, and for v_{lj}, $j = 2, 3, \cdots, n_l$, $l = 1, 2, \cdots, a$, where these quantities are defined in (4) and (5), we shall use two parameter vectors similar to $\boldsymbol{\theta}$. Thus let $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{12}, \cdots, \theta_{1k_1})'$, $\boldsymbol{\theta}_2 = (\theta_{21}, \theta_{22}, \cdots, \theta_{2k_2})'$,

and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$. The vector $\boldsymbol{\theta}_1$ will be used to specify the smooth alternative for v_{l1}, $l = 1, 2, \cdots, a$, and the vector $\boldsymbol{\theta}_2$ will be used to specify the smooth alternative for v_{lj}, $j = 2, 3, \cdots, n_l$, $l = 1, 2, \cdots, a$. Recall that we are assuming the independence of all the the v_{lj}'s, within and across the groups. The likelihood function under the assumption of a common alternative across the groups can be specified as

$$L = \prod_{l=1}^{a} \left[C(\boldsymbol{\theta}, \mu, \boldsymbol{\sigma}, \sigma_\alpha) \exp\left\{ \sum_{r=1}^{k_1} \theta_{1r} b_r(u_{l1}) + \sum_{j=2}^{n_l} \sum_{s=1}^{k_2} \theta_{2s} b_s(u_{lj}) \right\} f_{l1}(v_{l1}, \mu, \sigma_l, \sigma_\alpha) \prod_{j=2}^{n_l} f_{l2}(v_{lj}, \sigma_l) \right], \tag{7}$$

where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots, \sigma_a)'$, and for $l = 1, 2, \cdots, a$, $f_{l1}(v_{l1}, \mu, \sigma_l, \sigma_\alpha)$ and $f_{l2}(v_{lj}, \sigma_l)$, respectively, denote the normal density functions of v_{l1} and v_{lj} (j=2, 3, $\cdots$, n_l), when normality holds. The smooth alternatives that represent the departure from the normal distribution have the order k_1 for v_{l1} ($l = 1, 2, \cdots, a$), and order k_2 for v_{lj} ($l = 1, 2, \cdots, a$, j=2, 3, $\cdots$, n_l). The null hypothesis to be tested is that of normality and homoscedasticity. In terms of the parameters in (7) the hypothesis can be stated as

$$H_0 : \ \boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')' = \mathbf{0}, \text{ and } \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2. \tag{8}$$

We shall find it convenient to transform the vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots ., \sigma_a)'$ using an $a \times a$ Helmert matrix $Q = ((q_{lc}))$ having the first column equal to $\frac{1}{\sqrt{a}}\mathbf{1}_a$, as done in Yang and Mathew (2018). We shall denote the transformed vector by $\boldsymbol{\eta}$. That is,

$$\boldsymbol{\eta} = Q'\boldsymbol{\sigma} = (\eta_1, \eta_2, \cdots, \eta_a)'. \tag{9}$$

It is easy to see that $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2$ is equivalent to $\eta_2 = \eta_3 = \cdots = \eta_a = 0$. The log-likelihood function has the expression

$$\begin{aligned} \ln L \ = \ & \sum_{l=1}^{a} \left[\ln\{C(\boldsymbol{\theta}, \mu, \boldsymbol{\sigma}, \sigma_\alpha)\} + \sum_{r=1}^{k_1} \theta_{1r} b_r(u_{l1}) + \sum_{j=2}^{n_l} \sum_{s=1}^{k_2} \theta_{2s} b_s(u_{lj}) + \ln\{f_{l1}(v_{l1}, \mu, \sigma_l, \sigma_\alpha)\} \right. \\ & \left. + \sum_{j=2}^{n_l} \ln\{f_{l2}(v_{lj}, \sigma_l)\} \right]. \end{aligned} \tag{10}$$

In order to obtain the score statistic for testing the hypothesis in (8), we need the elements of the score vector and its variance-covariance matrix, evaluated under the null hypothesis. Explicit expressions can be obtained for these, and are given in the appendix. We note that some of the covariances are zeros. In the case of balanced data (ie., all the n_l, $l = 1, 2, \cdots, a$, are equal having a common value, say n), some additional covariances become zeros. These are also noted in the appendix.

While computing the score statistic, the unknown parameters are obviously replaced with their ML estimates under the null hypothesis. Thus we need the MLEs of μ, σ^2 and σ_α^2, where σ^2 is the common value of the σ_l^2's, under the null hypothesis. In the case of balanced data, we shall use the MLEs computed without imposing the nonnegativity constraint on the σ_α^2. If $\bar{y}_{..}$ denotes the average of all the y_{lj}s, and SS_e and SS_α, respectively, denote the sums

of squares due to error and due to the α_i's under the model (1), the MLEs in the balanced case are given by

$$\hat{\mu} = \bar{y}_{..},\ \hat{\sigma}^2 = \frac{SS_e}{a(n-1)} \ \text{ and } \ \hat{\sigma}^2_\alpha = \frac{1}{n}\left[\frac{SS_\alpha}{a} - \frac{SS_e}{a(n-1)}\right], \tag{11}$$

where σ^2 denotes the common value of the σ^2_ls and n denotes the common value of the n_ls. As we shall see, the lack of the nonnegativity of $\hat{\sigma}^2_\alpha$ will not present any problems for us, since the estimator required in our application will be $\hat{\sigma}^2 + n\hat{\sigma}^2_\alpha$, which is always nonnegative (being equal to SS_α/a). Similar explicit estimates can also be obtained in the unbalanced case as follows. Define

$$\begin{aligned} &\bar{y}_l = \sum_{j=1}^{n_l} y_{lj}/n_l,\ \bar{\bar{y}} = \frac{1}{a}\sum_{l=1}^{a}\bar{y}_l,\ SS_e = \sum_{l=1}^{a}\sum_{j=1}^{n_l}(y_{lj}-\bar{y}_l)^2, \\ &\tilde{n} = a \times \left\{\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_a}\right\}^{-1},\ SS_\alpha = \tilde{n}\sum_{l=1}^{a}(\bar{y}_{l.} - \bar{\bar{y}})^2, \end{aligned} \tag{12}$$

where we note that $\tilde{n}$ is the harmonic mean of the n_l's. Such a formulation is due to Thomas and Hultquist (1978); see also Krishnamoorthy and Mathew (2009, Chapter 4). It can be verified that $E(\bar{\bar{y}}) = \mu$, $E(SS_e) = (N-a)\sigma^2$ and $E(SS_\alpha) = (a-1)(\tilde{n}\sigma^2_\alpha + \sigma^2)$, where $N = \sum_{l=1}^{a} n_l$. The estimates of μ, σ^2 and σ^2_α that we shall use are given by

$$\hat{\mu} = \bar{\bar{y}},\ \hat{\sigma}^2 = SS_e/(N-a) \ \text{ and } \ \hat{\sigma}^2_\alpha = \frac{1}{\tilde{n}}\left[\frac{SS_\alpha}{a} - \frac{SS_e}{N-a}\right]. \tag{13}$$

It should be noted that the estimates in (13) are not the MLEs. In fact the MLEs have no explicit expression and have to be numerically obtained in the unbalanced case. Nevertheless, for convenience we shall use the estimates in (13).

In order to give an expression for the score statistic, let us write the parameters in the order $(\boldsymbol{\theta}', \boldsymbol{\eta}^{*'}, \eta_1, \sigma_\alpha, \mu)'$, where $\boldsymbol{\eta}* = (\eta_2, \eta_3, \cdots, \eta_a)'$ (see (9)) and we recall the partitioning of $\boldsymbol{\theta}$ into the two components $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ of dimensions $k_1 \times 1$ and $k_2 \times 1$, respectively. Thus the null hypothesis in (8) is equivalent to

$$H_0 :\ \boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')' = \mathbf{0},\ \text{and } \boldsymbol{\eta}* = 0. \tag{14}$$

The null hypothesis involves $(k_1 + k_2 + a - 1)$ parameters; in addition, we have three nuisance parameters, namely η_1, σ_α and μ. Now the the score vector has dimension $k_1 + k_2 + a + 2$, which is the total number of parameters under the model (7). Consequently, if V denotes the variance-covariance matrix of the score vector, whose elements are arranged according to the parameter order $(\boldsymbol{\theta}', \boldsymbol{\eta}^{*'}, \eta_1, \sigma_\alpha, \mu)'$, then clearly V has dimension $(k_1 + k_2 + a + 2) \times (k_1 + k_2 + a + 2)$. The elements of the score vector, and those of V are given in the appendix, where the expressions have been simplified assuming the null hypothesis (14).

Let us consider a partitioning of V as

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

where the dimension of V_{11} is $(k_1 + k_2 + a - 1) \times (k_1 + k_2 + a - 1)$, corresponding to the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\eta}*$ in the null hypothesis (14), and the dimensions of the remaining blocks of V should be clear. Let $\boldsymbol{u}$ denote the score vector and $\boldsymbol{u}_1$ denote the first $(k_1 + k_2 + a - 1)$ elements of $\boldsymbol{u}$ corresponding to the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\eta}*$, and define $V_{11.2} = V_{11} - V_{12}V_{22}^{-1}V_{21}$. Recall that $\boldsymbol{u}_1$ and $V_{11.2}$ are functions of the nuisance parameters η_1, σ_α and μ; equivalently σ, σ_α and μ (since, in view of (9), $\eta_1 = \sqrt{a} \times \sigma$ under the null hypothesis). The score statistic for testing the hypothesis (8), or equivalently (14), is given by

$$\hat{S} = \hat{\boldsymbol{u}}_1' \hat{V}_{11.2}^{-1} \hat{\boldsymbol{u}}_1, \tag{15}$$

where we have used the notations $\hat{\boldsymbol{u}}_1$ and $\hat{V}_{11.2}$ to emphasize that the unknown nuisance parameters σ, σ_α and μ have been replaced by estimates computed assuming the null hypothesis; we shall use the estimates of σ^2, σ_α^2 and μ, exhibited earlier in this section. For fixed k_1 and k_2, the score statistic $\hat{S}$ in (15) has an approximate chisquare distribution with df $= (k_1 + k_2 + a - 1)$ under the null hypothesis.

In order to implement the test based on the score statistic given in (15), it is necessary to choose the orders k_1 and k_2 in the likelihood function (7). For this, we shall follow a data-driven approach; *i.e.*, estimate the orders based on the data, as done in Inglot, Kallenberg and Ledwina (1994). Such an approach was also adopted in Yang and Mathew (2018) in a fixed effects linear model. Here we shall only present the relevant expressions that will facilitate the numerical computation of k_1 and k_2, referring to the original articles for details of the methodology and the associated theoretical results. A brief explanation of the methodology is also given in Yang (2016).

Referring to the quantities defined in (6), let

$$\hat{z}_{l1} = \frac{v_{l1} - \hat{\mu}\sqrt{n_l}}{\sqrt{\hat{\sigma}^2 + n_l \hat{\sigma}_\alpha^2}},\ \hat{u}_{l1} = \Phi\left(\hat{z}_{l1}\right),\ \hat{z}_{lj} = \hat{v}_{lj}/\hat{\sigma},\ \text{ and }\ \hat{u}_{lj} = \Phi\left(\hat{z}_{lj}\right),\ j = 2, 3, \cdots, n_l, \tag{16}$$

where $\hat{\sigma}$, $\hat{\sigma}_\alpha$ and $\hat{\mu}$ are the estimates obtained under the null hypothesis, and used in the computation of the score statistic $\hat{S}$ in (15). Let $\hat{H}_{1,k_1}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu})$ and $\hat{H}_{2,k_2}(\hat{\sigma})$ be defined as

$$\begin{aligned} \hat{H}_{1,k_1}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu}) &= \frac{1}{a} \sum_{i=1}^{k_1} \left\{ \sum_{l=1}^{a} b_i\left(\hat{u}_{l1}\right) \right\}^2 \\ \hat{H}_{2,k_2}(\hat{\sigma}) &= \frac{1}{N-a} \sum_{i=1}^{k_2} \left\{ \sum_{l=1}^{a} \sum_{j=2}^{n_l} b_i\left(\hat{u}_{lj}\right) \right\}^2, \end{aligned} \tag{17}$$

where $N = \sum_{l=1}^{a} n_l$, the $b_i(.)$'s are Legendre polynomials and the remaining quantities are defined earlier in this section. We note that the divisor a in the expression for $\hat{H}_{1,k_1}$ is the number of $\hat{z}_{l1}$s, $l = 1, 2, \cdots, a$, and the divisor $N - a$ in the expression for $\hat{H}_{2,k_2}$ is the number of $\hat{z}_{lj}$s, $j = 2, 3, \cdots, n_l$, $l = 1, 2, \cdots, a$. Assuming upper bounds d_1 for k_1 and d_2 for k_2, the orders k_1 and k_2, say $\hat{k}_1$ and $\hat{k}_2$, are determined as follows:

$$\begin{aligned} \hat{k}_1 &= \min\{k_1 : 1 \le k_1 \le d_1, \hat{H}_{1,k_1}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu}) - k_1 \ln(a) \ge \hat{H}_{1,r}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu}) - r \ln(a), r = 1, \cdots, d_1\} \\ \hat{k}_2 &= \min\{k_2 : 1 \le k_2 \le d_2, \hat{H}_{2,k_2}(\hat{\sigma}) - k_2 \ln(N-a) \ge \hat{H}_{2,s}(\hat{\sigma}) - s \ln(N-a), s = 1, \cdots, d_2\}. \end{aligned} \tag{18}$$

In order to determine $\hat{k}_1$ and $\hat{k}_2$ according to the expressions in (18), we need to select the bounds d_1 and d_2. For this, we shall make use of a result mentioned in Kallenberg and Ledwina (1997b), citing Inglot, Kallenberg and Ledwina (1994), which states that the data driven order converges to the value 1 in probability. Even though our set up is different from that under which this result is proved, we proceed under the assumption that the data driven approach is unlikely to yield values of $\hat{k}_1$ and $\hat{k}_2$ that are far removed from the value 1. In our simulations, we chose $d_1 = d_2 = 6$.

Algorithm 1 given below gives a summary of the steps necessary to implement our proposed test for testing the null hypothesis in (8) against a Legendre polynomial based smooth alternative.

Algorithm 1

1. Compute the estimates of $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\sigma}^2_\alpha$ given in (13).
2. Compute the $n_l \times 1$ vectors $\boldsymbol{v}_l = (v_{l1}, v_{l2}, \cdots, v_{ln_l})'$ given in (7).
3. Compute the $\hat{u}_{lj}$ given in (16), which requires the quantities computed in the previous two steps.
4. Compute $\hat{k}_1$ and $\hat{k}_2$ in (18) where $\hat{H}_{1,k_1}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu})$ and $\hat{H}_{2,k_2}(\hat{\sigma})$ are given in (17). For the quantities d_1 and d_2 in (18), we recommend the values $d_1 = d_2 = 6$.
5. Compute the score vector and its variance-covariance matrix using the expressions given in the appendix, and replace μ, σ^2 and σ^2_α with $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\sigma}^2_\alpha$, respectively. Let the quantities so obtained be denoted by $\hat{\boldsymbol{u}}$ and $\hat{V}$, respectively.
6. Partition $\hat{\boldsymbol{u}}$ and $\hat{V}$ as

$$\hat{\boldsymbol{u}} = \left(\hat{\boldsymbol{u}}_1', \hat{\boldsymbol{u}}_2'\right)' \text{ and } \hat{V} = \begin{pmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{pmatrix},$$

   where $\hat{\boldsymbol{u}}_1$ is a $(\hat{k}_1 + \hat{k}_2 + a - 1) \times 1$ vector, and $\hat{V}_{11}$ is a $(\hat{k}_1 + \hat{k}_2 + a - 1) \times (\hat{k}_1 + \hat{k}_2 + a - 1)$ matrix.
7. Compute $\hat{V}_{11.2} = \hat{V}_{11} - \hat{V}_{12}\hat{V}_{22}^{-1}\hat{V}_{21}$ and the score statistic $\hat{S} = \hat{\boldsymbol{u}}_1'\hat{V}_{11.2}^{-1}\hat{\boldsymbol{u}}_1$. Reject H_0 in (8) if the value of the score statistic exceeds the appropriate percentile of the chisquare distribution with df $= (\hat{k}_1 + \hat{k}_2 + a - 1)$.

For the case of balanced data, the n_l's have to be replaced with their common value, say n, in all the expressions. As a result, some of the covariances will become zeros, and these are noted in the appendix. It can then be verified that in the balanced case the matrix $\hat{V}_{11.2}$ is a block-diagonal matrix, with three diagonal blocks having dimensions $\hat{k}_1 \times \hat{k}_1$, $\hat{k}_2 \times \hat{k}_2$ and $(a-1) \times (a-1)$, corresponding to $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\eta}*$, where $\boldsymbol{\eta}* = (\eta_2, \eta_3, \cdots, \eta_a)'$. Accordingly, the score statistic splits into three components. Thus when the null hypothesis is rejected, it is possible to draw conclusions on which component/components contributed to the rejection:

the non-normality of v_{l1}'s ($l = 1, 2, \cdots, a$), the non-normality of v_{lj}'s ($j = 2, 3, \cdots, n_l$, $l = 1, 2, \cdots, a$), or the heteroscedasticity.

It turns out that the distributions of the score vector $\hat{\boldsymbol{u}}_1$ and the submatrix of $\hat{V}_{11.2}$ corresponding to $\boldsymbol{\eta}*$ depend on σ and σ_α. However, in the balanced case we have

$$\mathrm{Var}\left(\frac{\partial \ln L}{\partial \eta_c}\right) = 2\left[\frac{\sigma^2}{(\sigma^2 + n\sigma_\alpha^2)^2} + \frac{(n-1)}{\sigma^2}\right] = 2\left[\frac{1}{\sigma^2(1+n\lambda)^2} + \frac{(n-1)}{\sigma^2}\right],$$

$c = 1, 2, \cdots, a$, where $\lambda = \frac{\sigma_\alpha^2}{\sigma^2}$. We note that the above variance varies between $(n-1)/\sigma^2$ and n/σ^2, as λ varies between 0 and ∞. Thus the impact of σ_α^2 on the score test appears to be fairly small in the balanced case, and we may anticipate this to be so in the unbalanced case as well. Later we shall examine this further through numerical results.

2.2. Smooth alternatives based on Hermite polynomials

We shall now consider the likelihood function under a smooth alternative based on Hermite polynomials. The likelihood is similar to (7) except that Hermite polynomials are used instead of Legendre polynomials. Thus let $h_i(z)$, $i = 1, 2, \cdots$, denote the system of Hermite polynomials. The log-likelihood is now given by

$$\begin{aligned}\ln L \;=\; & \sum_{l=1}^{a}\Bigg[\ln\{C(\boldsymbol{\theta}, \mu, \boldsymbol{\sigma}, \sigma_\alpha)\} + \sum_{r=1}^{k_1}\theta_{1r}h_r(z_{l1}) + \sum_{j=2}^{n_l}\sum_{s=1}^{k_2}\theta_{2s}h_s(z_{lj}) + \ln\{f_{l1}(v_{l1}, \mu, \sigma_l, \sigma_\alpha)\} \\ & + \sum_{j=2}^{n_l}\ln\{f_{l2}(v_{lj}, \sigma_l)\}\Bigg], \end{aligned} \tag{19}$$

where the z_{l1} and z_{lj} are defined in (6). The components of the score vector (under the null hypothesis) can be derived similar to the Legendre case, and are given in the appendix. The elements of the variance covariance matrix of the score vector are also given in the appendix. The score statistic can be worked out similar to (15). It can also be verified that the matrix analogous to $V_{11.2}$ is a block diagonal matrix, having three diagonal blocks. Furthermore, in the balanced case, $V_{11.2}$ will reduce to a completely diagonal matrix. Recall that under the Legendre polynomial based alternative, $V_{11.2}$ simplified to a block diagonal matrix only under balanced data.

We can compute data driven choices of the orders k_1 and k_2 under the Hermite polynomial based smooth alternative also. For this we need to define quantities analogous to those in (17), with the Legendre polynomial terms replaced by the corresponding Hermite polynomial based terms. The orders can then be determined proceeding as in (18). However, when we define the analogous quantities in (17) for the Hermite polynomial case, the summations will be from $i = 3$ to k_1 and $i = 3$ to k_2 (instead of $i = 1$ to k_1 and $i = 1$ to k_2). For this, we need to show that the terms corresponding to $i = 1$ and $i = 2$ are zeros. Actually they are exactly zeros in the balanced case, and we shall choose to ignore them in the unbalanced case since they are likely to be quite small. Such zero-terms in the context of the Hermite polynomial based alternative have been noted, for example, in Rayner, Thas and Best (2009) in the context of testing univariate normality.

Similar to (17), let's define $\tilde{H}_{1,k_1}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu})$ and $\tilde{H}_{2,k_2}(\hat{\sigma})$ as

$$\begin{aligned} \tilde{H}_{1,k_1}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu}) &= \frac{1}{a}\sum_{i=3}^{k_1}\left\{\sum_{l=1}^{a} h_i\left(\hat{z}_{l1}\right)\right\}^2 \\ \tilde{H}_{2,k_2}(\hat{\sigma}) &= \frac{1}{N-a}\sum_{i=3}^{k_2}\left\{\sum_{l=1}^{a}\sum_{j=2}^{n_l} h_i\left(\hat{z}_{lj}\right)\right\}^2, \end{aligned} \tag{20}$$

where the $h_i(.)$'s are Hermite polynomials and $\hat{z}_{lj}$'s are defined in (16). Assuming upper bounds d_1 for k_1 and d_2 for k_2, the orders k_1 and k_2, say $\tilde{k}_1$ and $\tilde{k}_2$, are determined as follows:

$$\begin{aligned} \tilde{k}_1 &= \min\{k_1 : 3 \le k_1 \le d_1, \tilde{H}_{1,k_1}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu}) - k_1 \ln(a) \ge \tilde{H}_{1,r}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu}) - r\ln(a), r = 3, \cdots, d_1\} \\ \tilde{k}_2 &= \min\{k_2 : 3 \le k_2 \le d_2, \tilde{H}_{2,k_2}(\hat{\sigma}) - k_2 \ln(N-a) \ge \tilde{H}_{2,s}(\hat{\sigma}) - s\ln(N-a), s = 3, \cdots, d_2\}. \end{aligned} \tag{21}$$

Let's now consider the case of balanced data and show that the terms corresponding to $i = 1$ and $i = 2$ are zeros in $\tilde{H}_{1,k_1}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu})$ and $\tilde{H}_{2,k_2}(\hat{\sigma})$ in (20), so that in the definition of these quantities the summation can start from $i = 3$, as done in (20). Using (4), (6) and (16), and recalling that the Helmert matrix is an orthogonal matrix with first column being a multiple of a vector of ones, we have the following simplifications in the balanced case, under the null hypothesis,

$$\begin{aligned} & v_{l1} = \sqrt{n}\bar{y}_l, \ \hat{z}_{l1} = \frac{\sqrt{n}(\bar{y}_l - \bar{y}_{..})}{\sqrt{aSS_\alpha}}, \ \sum_{l=1}^{a}\hat{z}_{l1}^2 = \frac{1}{a} \\ & \sum_{j=1}^{n} y_{lj}^2 = \sum_{j=1}^{n} v_{lj}^2 = n\bar{y}_l^2 + \sum_{j=2}^{n} v_{lj}^2, \ \sum_{j=2}^{n} v_{lj}^2 = \sum_{j=1}^{n} y_{lj}^2 - n\bar{y}_l^2 \\ & \sum_{l=1}^{a}\sum_{j=2}^{n} v_{lj}^2 = \sum_{l=1}^{a}\left[\sum_{j=1}^{n} y_{lj}^2 - n\bar{y}_l^2\right] = SS_e \\ & \sum_{l=1}^{a}\sum_{j=2}^{n}\hat{z}_{lj}^2 = \frac{1}{a(n-1)}, \end{aligned} \tag{22}$$

where we have used the expression $SS_\alpha = n\sum_{l=1}^{a}(\bar{y}_l - \bar{y}_{..})^2$. The first two Hermite polynomials are given by

$$h_1(z) = z, \quad \text{and} \quad h_2(z) = \frac{1}{\sqrt{2}}(z^2 - 1).$$

From the observations in (22), it now follows that for balanced data, $\sum_{l=1}^{a} h_i(z_{l1}) = 0$, $i = 1, 2$, and $\sum_{l=1}^{a}\sum_{j=2}^{n} h_i(z_{lj}) = 0$, $i = 1, 2$. It should be noted that this conclusion holds only for the case of balanced data. For unbalanced data, we anticipate that these terms will be small even though they may not be exactly equal to zero.

Once $\tilde{k}_1$ and $\tilde{k}_2$ are determined as in (21), we note that the number of Hermite polynomial terms in $\tilde{H}_{1,\tilde{k}_1}(\hat{\sigma}, \hat{\sigma}_\alpha, \hat{\mu})$ and $\tilde{H}_{2,\tilde{k}_2}(\hat{\sigma})$ are $\tilde{k}_1 - 2$ and $\tilde{k}_2 - 2$, respectively. Hence the score statistic will have an approximate chisquare distribution with df $= (\tilde{k}_1 - 2) + (\tilde{k}_2 - 2) + (a - 1)$.

3. Numerical Results

We shall now report some numerical results to assess the behavior of the score tests we have proposed in the previous sections. Our purpose here is several: to assess the dependence of the score tests on σ_α^2, to examine the accuracy of the asymptotic chisquare distribution (under the null), and to compare the powers of the score tests based on the Legendre polynomial alternative and based on the Hermite polynomial alternative.

In order to the examine the dependence of the tests on σ_α^2, we first estimated the 95th percentiles of the score statistic in the case of balanced data, for various values of a and n, and for $\sigma_\alpha^2 = 0.1$, 2 and 10. We also chose $\mu = 0$ and $\sigma^2 = 1$, where σ^2 denotes the common value of the σ_l^2's. The data driven approach explained in the previous section was used to obtain the order of the polynomials. Most of the time, the data driven choices $\hat{k}_1$ and $\hat{k}_2$ were equal to one in the Legendre polynomial case, and $\tilde{k}_1$ and $\tilde{k}_2$ were equal to three in the Hermite polynomial case, so that the score statistic has an approximate chisquare distribution with df $= a+1$ under the null hypothesis. Table 1 gives the estimated percentiles based on 10^4 simulated samples. We have also included the 95th percentile of the chisquare distribution with df $= a + 1$. We can draw the following conclusions from the numerical results in Table 1: (i) the chisquare distribution approximates the null distribution of the score statistic reasonable well; however, the actual percentiles are slightly larger than that of the chisquare distribution, and (ii) the null distribution is not sensitive to the value of σ_α^2. It appears that in order to have a more accurate test, one can use a Monte Carlo estimate of the corresponding percentile (instead of using the chisquare percentile) after simply assuming $\sigma_\alpha^2 = 1$, regardless of the true value of σ_α^2.

In order to further see the insensitivity of the null distribution with respect to the value of σ_α^2, we estimated the type I error probabilities of the score test when $\sigma_\alpha^2 = 0.1$ and 10, when the test is carried out using the estimated critical value (*i.e.*, 95th percentile) corresponding to $\sigma_\alpha^2 = 2$. The rest of the simulation set up is the same as that used to obtain the results in Table 1. The type I error probabilities are given in Table 2. The insensitivity of the type I error probabilities with respect to the value of σ_α^2 should be clear.

In addition, we looked at Type I error probabilities for unbalanced cases, using critical values estimated from the balanced case with a common value chosen as $\tilde{n}$, which is the harmonic mean of the n_l's; see (13). For this we used a= 10, 20, 30 and 50 groups. We also made three choices in terms of severity of the unbalancedness: mild, moderate and severe. For $a = 10$, our choice of the n_l's to represent mild unbalancedness is (4,4,5,5,5,5,5,5,5,10), which gives the harmonic mean $\tilde{n}$=5. For moderate unbalancedness, we chose the n_l's to be (2,3,4,5,6,7,8,9,10,14), resulting in $\tilde{n}$=4.999008 (we shall take $\tilde{n} = 5$ in this case). For the severe unbalanced case, we made the choice (3,3,4,4,5,5,5,5,40,120) for the n_l's, which also yields $\tilde{n}$=5. For $a = 20$, we adopted the above choices except that each n_l value was repeated twice. For $a = 30$ and 50, the same strategy was followed; *i.e.*, each n_l value chosen for $a =$ 10 was repeated three times and five times each.

The choices of the n_ls just described all resulted in $\tilde{n} = 5$, as already noted. We shall also consider choices that will give $\tilde{n} = 10$, 30 and 50. For this, we multiplied each of the n_ls in the earlier choices with 2, 6 and 10, so as to result in $\tilde{n} = 10$, 30 and 50, respectively.

Table 1: Monte Carlo estimates of the 95th percentiles of the score statistic for σ_α^2= 0.1, 2, 10; $\chi^2_{(a+1,0.95)}$ denotes the 95th percentile of the chisquare distribution with $a+1$ df

		Legendre			Hermite			
a	n	σ_α^2=0.1	σ_α^2=2	σ_α^2=10	σ_α^2=0.1	σ_α^2=2	σ_α^2=10	$\chi^2_{(a+1,0.95)}$
10	5	21.71	21.70	21.20	19.57	20.05	20.17	19.68
10	10	21.02	20.99	20.70	20.09	20.52	20.49	19.68
10	30	20.51	20.49	20.43	20.16	20.33	20.39	19.68
10	50	20.80	20.80	20.79	20.56	20.49	20.49	19.68
20	5	35.25	35.16	34.32	34.56	35.97	35.96	32.67
20	10	33.34	33.34	33.35	35.11	35.42	35.31	32.67
20	30	33.29	33.30	33.25	34.22	34.40	34.46	32.67
20	50	33.25	33.24	33.20	33.95	34.08	34.13	32.67
30	5	47.87	47.78	46.83	48.58	49.45	49.48	44.99
30	10	46.82	46.84	46.83	48.16	48.15	48.23	44.99
30	30	45.51	45.52	45.53	47.43	47.18	47.14	44.99
30	50	45.72	45.72	45.71	46.79	46.63	46.64	44.99
50	5	72.99	72.69	71.64	75.35	76.03	75.88	68.67
50	10	70.67	70.70	70.88	73.54	73.86	73.90	68.67
50	30	69.34	69.34	69.34	71.35	71.39	71.35	68.67
50	50	69.63	69.65	69.64	70.61	70.72	70.56	68.67

Table 2: Estimated type I error probabilities of the score test carried out using the estimated critical value corresponding to $\sigma_\alpha^2 = 2$, for a 5% significance level

		Legendre		Hermite	
a	n	σ_α^2=0.1	σ_α^2=10	σ_α^2=0.1	σ_α^2=10
10	5	0.053	0.050	0.046	0.051
10	10	0.050	0.050	0.046	0.050
10	30	0.049	0.050	0.049	0.051
10	50	0.052	0.050	0.051	0.050
20	5	0.058	0.051	0.044	0.050
20	10	0.045	0.050	0.048	0.050
20	30	0.049	0.050	0.049	0.051
20	50	0.051	0.050	0.049	0.051
30	5	0.051	0.051	0.046	0.050
30	10	0.055	0.050	0.050	0.050
30	30	0.048	0.050	0.052	0.050
30	50	0.053	0.050	0.051	0.050
50	5	0.051	0.051	0.048	0.050
50	10	0.047	0.050	0.049	0.050
50	30	0.049	0.050	0.050	0.050
50	50	0.055	0.050	0.050	0.049

We used the value σ_α^2=2 to estimate the type I error probabilities, and considered both Legendre polynomial-based and Hermite polynomial-based alternatives. The estimated type I error probabilities are given in Table 3 (under Legendre polynomial-based alternatives) and in Table 4 (under Hermite polynomial-based alternatives). The following conclusions can be drawn from the numerical results. Under the Legendre polynomial-based alternative, there is no Type I error inflation for the mild unbalanceness cases; we notice a somewhat mild type I error inflation in the case of moderate unbalancedness when $\tilde{n}$ is small, and a more pronounced type I error inflation in the case of severe unbalancedness when $\tilde{n}$ is small. However, when $\tilde{n}$ is 30 or more, the type I error probabilities are all close to the nominal level of 5%. Under the Hermite polynomial-based alternative, the type I error inflation appears to be a bit more severe, especially in the case of severe unbalancedness and small $\tilde{n}$. However, the results are once again quite satisfactory when $\tilde{n}$ is 30 or more. Overall, using the balanced set up critical value based on $\tilde{n}$ appears to be a satisfactory option, except in the severe unbalanced case and a small $\tilde{n}$.

Table 3: Type I error probabilities for unbalanced data under Legendre polynomial-based alternatives and using the balanced data critical value with $n = \tilde{n}$, for a 5% significance level

		Unbalancedness			Estimated critical value	Chisquare
a	$\tilde{n}$	Severe	Moderate	Mild	when $\sigma_\alpha^2 = 2$	critical value
10	5	0.065	0.047	0.042	21.702	19.675
10	10	0.060	0.052	0.048	20.987	19.675
10	30	0.056	0.048	0.049	20.490	19.675
10	50	0.043	0.045	0.044	20.803	19.675
20	5	0.069	0.054	0.047	35.165	32.671
20	10	0.065	0.055	0.055	33.336	32.671
20	30	0.048	0.052	0.047	33.299	32.671
20	50	0.044	0.049	0.049	33.244	32.671
30	5	0.076	0.062	0.049	47.776	44.985
30	10	0.064	0.051	0.046	46.841	44.985
30	30	0.053	0.048	0.052	45.516	44.985
30	50	0.048	0.041	0.044	45.723	44.985
50	5	0.077	0.061	0.053	72.693	68.669
50	10	0.064	0.057	0.047	70.697	68.669
50	30	0.052	0.050	0.050	69.338	68.669
50	50	0.042	0.046	0.043	69.645	68.669

The type I error probabilities reported in Table 3 and Table 4 were computed when the test was carried out using the estimated critical value corresponding to $\sigma_\alpha^2 = 2$; these critical values are also given in the tables. The tables also give the chisquare critical values. We note that the chisquare critical values are smaller than the estimated critical values, as was noted in Table 1. Thus if the test is carried out using the asymptotic chisquare distribution, one should expect an inflated type I error probability.

Table 4: Type I error probabilities for unbalanced data under Hermite polynomial-based alternatives and using the balanced data critical value with $n = \tilde{n}$, for a 5% significance level

		Unbalancedness			Estimated critical value	Chisquare
a	$\tilde{n}$	Severe	Moderate	Mild	when $\sigma_\alpha^2 = 2$	critical value
10	5	0.093	0.063	0.053	20.052	19.675
10	10	0.066	0.059	0.052	20.516	19.675
10	30	0.058	0.053	0.052	20.327	19.675
10	50	0.046	0.049	0.049	20.495	19.675
20	5	0.082	0.060	0.051	35.972	32.671
20	10	0.065	0.053	0.051	35.420	32.671
20	30	0.051	0.054	0.050	34.400	32.671
20	50	0.046	0.052	0.052	34.080	32.671
30	5	0.080	0.066	0.052	49.451	44.985
30	10	0.066	0.059	0.054	48.149	44.985
30	30	0.054	0.049	0.054	47.175	44.985
30	50	0.050	0.051	0.051	46.626	44.985
50	5	0.075	0.062	0.053	76.031	68.669
50	10	0.058	0.056	0.048	73.862	68.669
50	30	0.053	0.051	0.054	71.387	68.669
50	50	0.046	0.056	0.051	70.715	68.669

Some limited numerical results on the power are reported in Table 5 in the case of balanced data for $a = 10$ and $n = 5$ and $a = 30$ and $n = 5$. We note that the null hypothesis can be violated by having a non-normal distribution for the errors and/or random effects, and/or by having heteroscedasticity of the error distribution. In Table 5, the powers are first reported when random effects are normally distributed and the error terms alone are non-normal and/or heteroscedastic (the first few rows of the table). The last few rows of the table correspond to the alternative scenario where the error terms are normally distributed and could be heteroscedastic, but the random effects are non-normal. The very last row of the table corresponds to the alternative where both the error terms and the random effects follow non-normal distributions, but the errors are homoscedastic. A 5% significance level and estimated critical values are used while computing the power. For introducing heteroscedasticity into the alternative, we proceeded as follows. We split the a groups into two sets, having $a/2$ groups in each set (we have chosen only an even value of a in our simulations). Data are generated from the same alternative error distribution, except that for the data in the second set, the randomly generated error term was multiplied by $\sqrt{2}$, which will result in twice the error variance for the data in the second set, compared to those in the first set. The results on the power show that most of the powers are comparable when the tests are derived using a Legendre polynomial-based alternative or a Hermite polynomial-based alternatives. However, the test derived under the Hermite polynomial-based alternatives appears to have a slight edge in terms of power. Perhaps more extensive simulation are necessary before we can draw firm conclusions.

Table 5: Estimated powers of the Legendre polynomial based and Hermite polynomial based tests using estimated critical values for a 5% significance level

		a=10, n=5		a=30, n=5	
Alternative		Legendre	Hermite	Legendre	Hermite
Error	N(0,1)	0.04	0.05	0.05	0.05
	N(0,1)*	0.11	0.13	0.22	0.21
	t(5)	0.24	0.29	0.56	0.57
	t(5)*	0.30	0.36	0.70	0.70
	Gamma (2,1)	0.28	0.35	0.65	0.68
	Gamma (2,1)*	0.36	0.44	0.78	0.80
Random Effect	t(2)	0.17	0.14	0.41	0.48
	t(2)*	0.23	0.21	0.54	0.59
	Gamma (2,1)	0.09	0.07	0.16	0.18
	Gamma (2,1)*	0.15	0.15	0.36	0.35
Error + Random Effect	Gamma (2,1) + t(2)	0.40	0.43	0.79	0.84

*Heteroscedasticity

4. An Example

We use a quality control data set from clinical chemistry on serum sodium measurements. The data are taken from Andrews and Herzberg (1985), and serum sodium measurements are given for 10 specimens tested by 100 labs. The specimens are from a large homogeneous pool of serum, and one specimen is sent to the labs every two weeks. Here we shall use only a subset of the data, and these data are from 24 labs that used the same analysis method (Method 5 mentioned in Andrews and Herzberg (1985)), and had all 10 specimens tested, so that we have balanced data with a=10 and n=24; the data we used are given in Yang (2016). The results of the data analysis are presented in Table 6. Data driven choices were made for the orders k_1 and k_2. For a 5% significance level, the estimated critical values necessary to carry out the test are given in Table 6. The data driven choices of the orders are also given in the table. The upper bounds d_1 and d_2 were chosen as $d_1 = d_2$=6. We noted earlier that that for one-way random model with balanced data, the matrix $V_{11.2}$ used to compute the score statistic is a block-diagonal matrix consisting of 3 blocks; if the null hypothesis of normality and homoscedasticity is rejected, it is possible to draw conclusions on which component/components contributed to the rejection: normality or homoscedasticity of the error distribution, or the normality of the random effect. We note from Table 6 that the null hypothesis is rejected by the tests based on both Legendre polynomial-based and Hermite-polynomial-based alternatives. In the table, we have also given the decomposition of the score statistic into the three components; the first component corresponds to normality of the random effects, the second and third components correspond, respectively, to normality and heteroscedasticity of the error distribution. The results indicate that there is evidence for both non-normality and heteroscedasticity for the error distribution, but there is no evidence of non-normality of the random effects.

Table 6: Analysis of the serum sodium data using a 5% significance level

Polynomial	Orders	Estimated critical value	Score statistic	Decision
Legendre	$\hat{k}_1 = 1$, $\hat{k}_2 = 2$	20.584	$55.505 = 0.066 + 32.776 + 22.663$	Reject H_0
Hermite	$\tilde{k}_1 = 3$, $\tilde{k}_2 = 6$	20.388	$251.439 = 0.024 + 228.752 + 22.663$	Reject H_0

5. Discussion: Possible Extensions and Limitations

Data analysis in a linear model framework relies on several assumptions: normality, homoscedasticity (especially when the data fall into different groups), independence, and the assumption that the mean vector belongs to a specified subspace. Simultaneous assessment of these assumptions is clearly of interest. An attempt in this direction has been made by Peña and Slate (2006). Their assessment of normality assumes a smooth alternative. The assessment of homoscedasticity assumes that departure from this assumption can be modeled as a function of the mean. However, the violation of the homoscedasticity need not imply that the variance changes with the mean, even though this is a possibility. In our work, we have explored the simultaneous assessment of both normality and homoscedasticity. It should be noted that the smooth tests available in the literature address only the problem of testing the adequacy of a parametric distribution, specified in terms of appropriate orthogonal polynomials. However, Kallenberg, Ledwina and Rafajlowicz (1997) did address the problem of simultaneously testing normality and independence for bivariate data. Such simultaneous testing has been facilitated by having a parametric form under the smooth alternative.

We want to point out several limitations, and some possible generalizations, of the work reported here. In the context of models that involve random effects, if we want to go beyond the one-way random model, difficulties do arise for the simultaenous assessment of normality of the random effects, normality of the error terms, and homoscedasticity. Unbalancd data will add further complications. However, we feel that our methodology in the context of the one-way random model can be generalized to general mixed or random effects models for the simultaneous assessment of normality of the random effects and the error terms, assuming that homoscedasticity holds, provided we have *balanced* data. We shall now illustrate this in the context of the two-way random effects model and the two-way mixed effects models, when the model includes interactions.

5.1. Testing normality in a two-way random model with balanced data

Consider two factors having a and b levels, randomly selected, and suppose we have n observations one each level combination; thus we have balanced data. Let y_{ijm} denote the mth observation corresponding to the ith and jth levels of the two factors; $m = 1, 2, \cdots, n$. The two-way model with interaction is given by

$$y_{ijm} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijm},$$

$i = 1, \cdots, a$, $j = 1, \cdots, b$ and $m = 1, \cdots, n$. We assume that all the effects are random and the random variables α_is, β_js, γ_{ij}s and the error terms ϵ_{ijm}s are all independent. The usual

normality assumptions state that

$$\alpha_i \sim N(0, \sigma_\alpha^2), \quad \beta_j \sim N(0, \sigma_\beta^2), \quad \gamma_{ij} \sim N(0, \sigma_\gamma^2), \quad \text{and} \quad \epsilon_{ijm} \sim N(0, \sigma^2).$$

Note that we are now assuming homoscedasticity. The above normality assumptions imply a multivariate normal distribution with a structured variance-covariance matrix for the data vector consisting of all the y_{ijm}; the mean vector is simply $\mu \mathbf{1}_{abn}$. Our goal is to test if such a multivariate normality assumption holds, by using an appropriate smooth alternative. We recall that in the one-way random model, this was accomplished by transforming the data vector into two uncorrelated components, and simultaneously testing univariate normality for each component assuming independent smooth alternatives. We shall do the same under the two-way random model as well. However, now four sets of independent random variables are involved in the model (α_i's, β_j's, γ_{ij}'s and the ϵ_{ijm}'s). Consequently, we will suitably transform the data vector, and end up with four uncorrelated components. We can then simultaneously test univariate normality of the four components against the assumption of smooth alternatives that are also independent.

Consider the vector $\boldsymbol{y}_{ij} = (y_{ij1}, y_{ij2}, \cdots, y_{ijn})'$, and let $Q_n = \left(\frac{1}{\sqrt{n}} 1_n, Q_n^*\right)$ be an $n \times n$ Helmert matrix. Consider the transformation

$$\boldsymbol{v}_{ij} = Q_n' \boldsymbol{y}_{ij},$$

$i = 1, 2, \cdots, a$, and $j = 1, 2, \cdots, b$. Let $\boldsymbol{v}_{ij} = (v_{ij1}, \boldsymbol{v}_{ij0}')'$, so that $\boldsymbol{v}_{ij0}$ is an $(n-1) \times 1$ vector. It follows that $\boldsymbol{v}_{ij0}$ has the mean vector and covariance matrix given by

$$E(\boldsymbol{v}_{ij0}) = \mathbf{0}, \; Var(\boldsymbol{v}_{ij0}) = \sigma^2 \boldsymbol{I}_{n-1}.$$

Here we have used the facts that $E(\boldsymbol{y}_{ij}) = \mu \mathbf{1}_n$ and Q_n is an $n \times n$ orthogonal matrix with the first column given by $\frac{1}{\sqrt{n}} \mathbf{1}_n$. Denote by $\boldsymbol{v}_0$ the $ab(n-1) \times 1$ vector consisting of the $\boldsymbol{v}_{ij0}'s$, $\forall i = 1, \cdots, a$ and $j = 1, \cdots, b$. Then

$$E(\boldsymbol{v}_0) = \mathbf{0}, \; Var(\boldsymbol{v}_0) = \sigma^2 \boldsymbol{I}_{ab(n-1)}. \tag{23}$$

It should be clear that the components of $\boldsymbol{v}_0$ are the $ab(n-1)$ *error contrasts*; we recall that under the two-way model with interaction and balanced data, the error sum of squares has df $= ab(n-1)$.

Now let's consider v_{ij1}, the first element of the vector $\boldsymbol{v}_{ij}'s$. The model for the y_{ijm}'s imply the following model for v_{ij1}:

$$v_{ij1} = \sqrt{n}\mu + \sqrt{n}(\alpha_i + \beta_j + \gamma_{ij}) + \sqrt{n}\, \bar{\epsilon}_{ij},$$

where $\bar{\epsilon}_{ij} = \frac{1}{n} \sum_{m=1}^{n} \epsilon_{ijm}$. Denote by $\boldsymbol{v}_1$ the vector consisting of the v_{ij1}'s, $j = 1, \cdots, b$ and $i = 1, \cdots, a$. Note that $\boldsymbol{v}_1$ is an $ab \times 1$ vector. We then we have the model

$$\boldsymbol{v}_1 = \sqrt{n}[\mu \mathbf{1}_{ab} + (\boldsymbol{I}_a \otimes \mathbf{1}_b)\boldsymbol{\alpha} + (\mathbf{1}_a \otimes \boldsymbol{I}_b)\boldsymbol{\beta} + \boldsymbol{\gamma} + \bar{\boldsymbol{\epsilon}}],$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\bar{\boldsymbol{\epsilon}}$ are vectors consisting of the α_i, β_j, γ_{ij} and $\bar{\epsilon}_{ij}$, respectively. Thus

$$E(\boldsymbol{v}_1) = \sqrt{n}\mu \mathbf{1}_{ab}$$

$$V(\boldsymbol{v}_1) = n\sigma_\alpha^2(\boldsymbol{I}_a \otimes \mathbf{1}_b\mathbf{1}_b') + n\sigma_\beta^2(\mathbf{1}_a\mathbf{1}_a' \otimes \boldsymbol{I}_b) + n\sigma_\gamma^2\boldsymbol{I}_{ab} + \sigma^2\boldsymbol{I}_{ab}.$$

We now make further transformations of the vector $\boldsymbol{v}_1$. Let $Q_a = \left(\frac{1}{\sqrt{a}}1_a, Q_a^*\right)$ and $Q_b = \left(\frac{1}{\sqrt{b}}1_b, Q_b^*\right)$ be defined similar to Q_n, but with dimensions $a \times a$ and $b \times b$, respectively. Define

$$w_0 = \left(\frac{1}{\sqrt{a}}1_a' \otimes \frac{1}{\sqrt{b}}1_b'\right)\boldsymbol{v}_1,\ \boldsymbol{w}_1 = \left(Q_a^{*'} \otimes \frac{1}{\sqrt{b}}1_b'\right)\boldsymbol{v}_1,$$
$$\boldsymbol{w}_2 = \left(\frac{1}{\sqrt{a}}1_a' \otimes Q_b^{*'}\right)\boldsymbol{v}_1, \quad \text{and} \quad \boldsymbol{w}_3 = \left(Q_a^{*'} \otimes Q_b^{*'}\right)\boldsymbol{v}_1. \tag{24}$$

We note that w_0 is a scalar, and the vectors $\boldsymbol{w}_1$, $\boldsymbol{w}_2$ and $\boldsymbol{w}_3$ are of dimensions $(a-1)\times 1$, $(b-1)\times 1$, and $(a-1)(b-1)\times 1$, respectively. It is readily verified that w_0, $\boldsymbol{w}_1$, $\boldsymbol{w}_2$ and $\boldsymbol{w}_3$ are all uncorrelated. We shall denote the elements of the vectors $\boldsymbol{w}_1$, $\boldsymbol{w}_2$ and $\boldsymbol{w}_3$ as $w_{1i'}$, $w_{2j'}$ and $w_{3i'j'}$, respectively, for $i' = 1, 2, \cdots, a-1$, and $j' = 1, 2, \cdots, b-1$, and we have the following means and variances:

$$E(w_0) = \sqrt{abn}\mu,\ Var(w_0) = \sigma_0^2 = bn\sigma_\alpha^2 + an\sigma_\beta^2 + n\sigma_\gamma^2 + \sigma^2,$$
$$E(w_{1i'}) = 0,\ Var(w_{1i'}) = \sigma_1^2 = bn\sigma_\alpha^2 + n\sigma_\gamma^2 + \sigma^2,\ i' = 1, 2, ..., a-1,$$
$$E(w_{2j'}) = 0,\ Var(w_{2j'}) = \sigma_2^2 = an\sigma_\beta^2 + n\sigma_\gamma^2 + \sigma^2,\ j' = 1, 2, ..., b-1,$$
$$E(w_{3i'j'}) = 0,\ Var(w_{3i'j'}) = \sigma_3^2 = n\sigma_\gamma^2 + \sigma^2,\ i' = 1, 2, ..., a-1, \quad j' = 1, 2, \cdots, b-1.$$

Let $v_{0m'}$ denote the m'th element of the vector $\boldsymbol{v}_0$ defined in (23), so that $E(v_{0m'}) = 0$ and $Var(v_{0m'}) = \sigma^2$, $m' = 1, 2, \cdots, ab(n-1)$. Furthermore, the $v_{0m'}$s are uncorrelated. If the normality assumption holds for all the random effects and the error terms in the two-way random model, then the following four sets of random variables follow independent normal distributions with means all equal to zero, and variances as specified above: (i) $v_{0m'}$, $m' = 1, 2, \cdots, ab(n-1)$, (ii) $w_{1i'}$, $i' = 1, 2, \cdots, a-1$, (iii) $w_{2j'}$, $j' = 1, 2, \cdots, b-1$, and (iv) $w_{3i'j'}$, $i' = 1, 2, \cdots, a-1$, $j' = 1, 2, \cdots, b-1$. In other words, under the assumption of normality, the random variables given in (i), (ii), (iii) and (iv) can be treated as samples of sizes $ab(n-1)$, $a-1$, $b-1$ and $(a-1)(b-1)$ from four independent normal distributions with zero means. Independent smooth alternatives to normality can now be specified for each of the four sets (i), (ii), (iii) and (iv), and score tests can be derived for simultaneously testing normality of the error term and the normality of the random effects.

5.2. Testing normality in a two-way mixed model with balanced data

Now consider the two-way mixed effect model with interaction and balanced data:

$$y_{ijm} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijm},$$

where the α_i, $i = 1, 2, \cdots, a$, are assumed to be the fixed effects, and the rest of the effects are random effects. Here we make the usual assumption: $\sum_{i=1}^a \alpha_i = 0$. The standard normality assumptions that are imposed on the random effects and the error terms are the same as those given in the previous section: $\beta_j \sim N(0, \sigma_\beta^2)$, $\gamma_{ij} \sim N(0, \sigma_\gamma^2)$, and $\epsilon_{ijm} \sim N(0, \sigma^2)$, $i = 1, \cdots, a$, $j = 1, \cdots, b$ and $m = 1, \cdots, n$.

Starting with the transformation based on the Helmert matrix Q_n, we can arrive at the $ab(n-1)$ error contrast vector $\boldsymbol{v}_0$ given in the previous section, having zero mean and the variance-covariance matrix $\sigma^2 I_{ab(n-1)}$; see (23). Let's now consider the quantities w_0, $\boldsymbol{w}_1$, $\boldsymbol{w}_2$ and $\boldsymbol{w}_3$, defind in (24), and denote the elements of $\boldsymbol{w}_2$ and $\boldsymbol{w}_3$ as $w_{2j'}$ and $w_{3i'j'}$, respectively, for $i' = 1, 2, \cdots, a-1$, and $j' = 1, 2, \cdots, b-1$. It can once again be verified that w_0 , $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, and $\boldsymbol{w}_3$ are all uncorrelated, and we have the following means and variances:

$$
\begin{aligned}
&E(w_0) = \sqrt{abn}\mu,\ Var(w_0) = an\sigma_\beta^2 + n\sigma_\gamma^2 + \sigma^2,\\
&E(\boldsymbol{w}_1) = \sqrt{b}Q_a^{*'}\boldsymbol{\alpha},\ Var(\boldsymbol{w}_1) = (n\sigma_\gamma^2 + \sigma^2)\boldsymbol{I}_{a-1},\\
&E(w_{2j'}) = 0,\ Var(w_{2j'}) = an\sigma_\beta^2 + n\sigma_\gamma^2 + \sigma^2,\ j' = 1, 2, \cdots, b-1,\\
&E(w_{3i'j'}) = 0,\ Var(w_{3i'j'}) = n\sigma_\gamma^2 + \sigma^2,\ i' = 1, 2, \cdots, a-1, \quad j' = 1, 2, \cdots, b-1.
\end{aligned}
$$

The vectors $\boldsymbol{v}_0$, $\boldsymbol{w}_2$ and $\boldsymbol{w}_3$ are also uncorrelated, have mean zeros, and covariance matrices $\sigma^2 I_{ab(n-1)}$, $(an\sigma_\beta^2 + n\sigma_\gamma^2 + \sigma^2)\boldsymbol{I}_{b-1}$ and $(n\sigma_\gamma^2 + \sigma^2)\boldsymbol{I}_{(a-1)(b-1)}$, respectively. Three independent smooth alternatives to normality can now be defined, as noted in the previous section, and smooth tests can be derived. Note that the scalar quantity w_0, and the vector $\boldsymbol{w}_1$ have means $\sqrt{abn}\mu$ and $\sqrt{b}Q_a^{*'}\boldsymbol{\alpha}$, respctively, which are unknown nuisance parameters to be estimated. Thus these components will not contribute to the test for normality. In this section and in the previous section, we have not brought up the issue of testing homoscedasticity.

We believe that the approach outlined in this section and the previous section can be adopted to any mixed or random effects model with balanced data. However, the same approach will not go through when we have unbalanced data. Let's briefly indicate why this is so. Consider the case of the random effects model. A key step in the derivations is the transformation in (24) leading to the uncorrelated quantities w_0, $\boldsymbol{w}_1$, $\boldsymbol{w}_2$ and $\boldsymbol{w}_3$. It is not difficult to note that such a transformation leading to uncorrelated quantities is not possible when we have unbalanced data. In short, when we have a linear model with a structured covariance matrix, which is the case for any mixed or random effects model, it is not clear how we can define a smooth alternative by taking the structure into account.

References

Bogdan, M. (1996). Data driven smooth tests for bivariate normality, preprint IM PWr 42/96, Technical University of Wroclaw.

Bogdan, M. (1999). Data driven smooth tests for bivariate normality. *Journal of Multivariate Analysis*, **68**, 26-53.

Chang, C.-H., Pal, N. and Lin, J.-J. (2017). A revisit to test the equality of variances of several populations. *Communications in Statistics - Simulation and Computation*, **46**, 6360-6384.

Inglot, T., Kallenberg, W. and Ledwina, T. (1994). On selection rules, with an application to goodness-of-fit for composite hypotheses. (Memorandum, Faculty of Mathematical Sciences; No. 1242). University of Twente, Faculty of Mathematical Sciences.

Inglot, T., Kallenberg, W. and Ledwina, T. (1997). Data driven smooth tests for composite hypotheses. *The Annals of Statistics*, **25**, 1222-1250.

Inglot, T. and Ledwina, T. (1996). Asymptotic optimality of data-driven Neymans tests for uniformity. *The Annals of Statistics*, **24**, 1982-2019.

Janic, A. and Ledwina, T. (2009). Data-driven smooth tests for a location-scale family revisited, *Journal of Statistical Theory and Practice*, **3**, 645-664.

Kallenberg, W. and Ledwina, T. (1995). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *The Annals of Statistics*, **23**, 1594-1608.

Kallenberg, W. and Ledwina, T. (1997a). Data-driven smooth tests for composite hypotheses: Comparison of powers. *Journal of Statistical Computation and Simulation*, **59**, 101-121.

Kallenberg, W. and Ledwina, T. (1997b). Data-driven smooth tests when the hypothesis is composite. *Journal of the American Statistical Association*, **92**, 1094-1104.

Kallenberg, W., Ledwina, T. and Rafajlowicz, E. (1997). Testing bivariate independence and normality. *Sankhyã, Series A*, **59**, 42-59.

Krishnamoorthy, K. and Mathew, T. (2009). *Statistical Tolerance Regions: Theory, Applications and Computation*. John Wiley, New York.

Ledwina, T. (1994). Data-driven version of Neyman's smooth test of fit. *Journal of the American Statistical Association*, **89**, 1000-1005.

Montgomery, D. C. (2020). *Design and Analysis of Experiments*, 10th Edition, John Wiley, New York.

Neyman, J. (1937) Smooth test for goodness of fit. *Skandinaviska Aktuarietidskrift*, **20**, 149-199.

Peña, E. A. and Slate E. H. (2006). Global validation of linear model assumptions. *Journal of the American Statistical Association*, **101**, 341-354.

Rayner, J. C. W., Thas O. and Best, D. J. (2009). *Smooth Tests of Goodness of Fit : Using R*. John Wiley, New York.

Thomas, J. D. and Hultquist, R. A. (1978). Interval estimation for the unbalanced case of the one-way random effects model. *The Annals of Statistics*, **6**, 582-587.

Xu, W., Li, Y. and Song, D. (2013). Testing normality in mixed models using a transformation method. *Statistical Papers*, **54**, 71-84.

Yang, Y. (2016). *The Simultaneous Assessment of Normality and Homoscedasticity in Some Linear Models*. Doctoral dissertation submitted to the University of Maryland Baltimore County.

Yang, Y. and Mathew, T. (2018). The simultaneous assessment of normality and homoscedasticity in linear fixed effects models. *Journal of Statistical Theory and Practice*, **12**, 66-81.

APPENDIX

The Score Vector and its Variance-Covariance Matrix

Here we shall give only the expressions for the score vector and its variance-covariance matrix (without providing their derivations) under both the Legendre polynomial based and Hermite polynomial based alternatives. In order to derive these, it is necessary to use expressions for the derivatives of $\ln\{C(\boldsymbol{\theta}, \mu, \boldsymbol{\sigma}, \sigma_\alpha)\}$. General results on such derivatives are given in Rayner, Thas and Best (2009, Section 6.1).

A. The score vector, variances and covariances under the Legendre polynomial based alternative

Recall that the elements of the score vector and those of the variance-covariance matrix are to be evaluated under the null hypothesis. For the log-likelihood function given in (10), the components of the score vector (under the null hypothesis) are as follows:

$$\frac{\partial \ln L}{\partial \theta_{1r}} = \sum_{l=1}^{a} b_r(u_{l1}),\ \frac{\partial \ln L}{\partial \theta_{2s}} = \sum_{l=1}^{a}\sum_{j=2}^{n_l} b_s(u_{lj}),\ r = 1, 2, \cdots, k_1,\ s = 1, 2, \cdots, k_2$$

$$\frac{\partial \ln L}{\partial \eta_c} = \sum_{l=1}^{a}(z_{l1}^2 - 1)\frac{\sigma q_{lc}}{\sigma^2 + n_l\sigma_\alpha^2} + \sum_{l=1}^{a}\sum_{j=2}^{n_l}(z_{lj}^2 - 1)\frac{q_{lc}}{\sigma},\ c = 1, 2, \cdots, a,$$

$$\frac{\partial \ln L}{\partial \sigma_\alpha} = \sum_{l=1}^{a}(z_{l1}^2 - 1)\frac{n_l\sigma_\alpha}{\sigma^2 + n_l\sigma_\alpha^2},\ \frac{\partial \ln L}{\partial \mu} = \sum_{l=1}^{a}\frac{\sqrt{n_l}z_{l1}}{\sqrt{\sigma^2 + n_l\sigma_\alpha^2}}$$

where σ^2 is the common variance under the null hypothesis, and q_{lc}'s are the elements of the $a \times a$ Helmert matrix Q defined in Section 2, and $\boldsymbol{\eta} = Q'\boldsymbol{\sigma}$; see (9).

The expressions for the variances and covariances among the components of the score vector involve certain constants c_i and e_i, $i = 1, 2, \cdots$. We shall first give these before giving the variance and covariance terms. Let $f(z)$ denote the density of a standard normal random variable Z. The required constants c_i and e_i are given by

$$\begin{aligned} c_i &= \mathrm{Cov}[b_i(\Phi(Z)), Z] = \int_{-\infty}^{\infty} b_i(\Phi(z))zf(z)dz, \\ e_i &= \mathrm{Cov}[b_i(\Phi(Z)), Z^2] = \int_{-\infty}^{\infty} b_i(\Phi(z))z^2f(z)dz, \end{aligned}$$

$i = 1, 2, 3, \cdots$. By using the expressions for the Legendre polynomials, it can be verified that $c_i = 0$ for i even, and $e_i = 0$ for i odd. When c_i's and e_i's are non-zero, they can be computed numerically. A few such values are given below; see Bogdan (1996, 1999).

$c_1 = 0.977205023801135,\ c_3 = 0.1830082402700861,\ c_5 = 0.0816989764273946,$
$c_7 = 0.04772936798473241,\ c_9 = 0.031880431223894;$
$e_2 = 1.232808888123174,\ e_4 = 0.5211245854593028,\ e_6 = 0.3045144697203598,$
$e_8 = 0.2055889833015625,\ e_{10} = 0.150770690085310.$

The variances and covariances among the components of the score vector are as follows (under the null hypothesis):

$$\mathrm{Var}\left(\frac{\partial \ln L}{\partial \theta_{1r}}\right) = a, \quad \text{and } \mathrm{Cov}\left(\frac{\partial \ln L}{\partial \theta_{1r}}, \frac{\partial \ln L}{\partial \theta_{1r'}}\right) = 0, \ r \neq r'; \ r, r' = 1, 2, \cdots, k_1,$$

$$\mathrm{Var}\left(\frac{\partial \ln L}{\partial \theta_{2s}}\right) = N - a, \text{ where } N = \sum_{l=1}^{a} n_l, \quad \text{and } \mathrm{Cov}\left(\frac{\partial \ln L}{\partial \theta_{2s}}, \frac{\partial \ln L}{\partial \theta_{2s'}}\right) = 0,$$

$$s \neq s'; \ s, s' = 1, 2, \cdots, k_2,$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \theta_{1r}}, \frac{\partial \ln L}{\partial \theta_{2s}}\right) = 0, \ r = 1, 2, \cdots, k_1, \ s = 1, 2, \cdots, k_2,$$

$$\mathrm{Var}\left(\frac{\partial \ln L}{\partial \eta_c}\right) = 2\sum_{l=1}^{a}\left(\frac{\sigma^2}{(\sigma^2 + n_l\sigma_\alpha^2)^2} + \frac{(n_l - 1)}{\sigma^2}\right)q_{lc}^2, \ c = 1, 2, \cdots, a,$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \eta_c}, \frac{\partial \ln L}{\partial \eta_{c'}}\right) = 2\sum_{l=1}^{a}\left(\frac{\sigma^2}{(\sigma^2 + n_l\sigma_\alpha^2)^2} + \frac{(n_l - 1)}{\sigma^2}\right)q_{lc}q_{lc'}, \ c \neq c'; \ c, c' = 1, 2, \cdots, a,$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \theta_{1r}}, \frac{\partial \ln L}{\partial \eta_c}\right) = e_r\sum_{l=1}^{a}\frac{\sigma}{\sigma^2 + n_l\sigma_\alpha^2}q_{lc} \text{ for } r \text{ even, and 0 for } r \text{ odd}, \ c = 1, \cdots, a,$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \theta_{2s}}, \frac{\partial \ln L}{\partial \eta_c}\right) = (e_r/\sigma)\sum_{l=1}^{a}(n_l - 1)q_{lc} \text{ for } s \text{ even, and 0 for } s \text{ odd}, \ c = 1, \cdots, a,$$

$$\mathrm{Var}\left(\frac{\partial \ln L}{\partial \sigma_\alpha}\right) = 2\sum_{l=1}^{a}\frac{n_l^2\sigma_\alpha^2}{(\sigma^2 + n_l\sigma_\alpha^2)^2},$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \theta_{1r}}, \frac{\partial \ln L}{\partial \sigma_\alpha}\right) = e_r\sum_{l=1}^{a}\frac{n_l\sigma_\alpha}{\sigma^2 + n_l\sigma_\alpha^2} \text{ for } r \text{ even, and 0 for } r \text{ odd},$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \theta_{2s}}, \frac{\partial \ln L}{\partial \sigma_\alpha}\right) = 0, \ s = 1, \cdots, k_2,$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \eta_c}, \frac{\partial \ln L}{\partial \sigma_\alpha}\right) = 2\sum_{l=1}^{a}\frac{\sigma\sigma_\alpha}{(\sigma^2 + n_l\sigma_\alpha^2)^2}n_l q_{lc}, \ c = 1, \cdots, a,$$

$$\mathrm{Var}\left(\frac{\partial \ln L}{\partial \mu}\right) = \sum_{l=1}^{a}\frac{n_l}{\sigma^2 + n_l\sigma_\alpha^2},$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \theta_{1r}}, \frac{\partial \ln L}{\partial \mu}\right) = c_r\sum_{l=1}^{a}\frac{\sqrt{n_l}}{\sqrt{\sigma^2 + n_l\sigma_\alpha^2}} \text{ for } r \text{ odd, and 0 for } r \text{ even},$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \theta_{2s}}, \frac{\partial \ln L}{\partial \mu}\right) = 0, \ s = 1, \cdots, k_2,$$

$$\mathrm{Cov}\left(\frac{\partial \ln L}{\partial \eta_c}, \frac{\partial log L}{\partial \mu}\right) = 0, \ c = 1, \cdots, a, \quad \mathrm{Cov}\left(\frac{\partial \ln L}{\partial \sigma_\alpha}, \frac{\partial \ln L}{\partial \mu}\right) = 0.$$

A.1. The case of balanced data

For balanced data, the n_l's have to be replaced with their common value n in all the

expressions. As a result, some of the covariances will become 0, these are:

$$\begin{aligned}
&\text{Cov}\left(\frac{\partial \ln L}{\partial \eta_c}, \frac{\partial \ln L}{\partial \eta_{c'}}\right) = 0, \quad c, c' = 1, 2, \cdots, a \\
&\text{Cov}\left(\frac{\partial \ln L}{\partial \theta_{1r}}, \frac{\partial \ln L}{\partial \eta_c}\right) = 0, \quad r = 1, 2, \cdots, k_1, \ c = 2, ..., a \\
&\text{Cov}\left(\frac{\partial \ln L}{\partial \theta_{2s}}, \frac{\partial \ln L}{\partial \eta_c}\right) = 0, \quad s = 1, 2, \cdots, k_2, \ c = 2, ..., a \\
&\text{Cov}\left(\frac{\partial \ln L}{\partial \eta_c}, \frac{\partial logL}{\partial \sigma_\alpha}\right) = 0, \quad c = 2, \cdots, a
\end{aligned}$$

B. The score vector, variances and covariances under the Hermite polynomial based alternative

The components of the score vector (under the null hypothesis) can be derived similar to those obtained under the Legendre polynomial case. The scores with respect to θ_{1r} and θ_{2s} are given below, and those with respect η_c, σ_α, and μ are not given since they are the same as in the Legendre case. The scores corresponding to θ_{1r} and θ_{2s} can be shown to be equal to zero for $r = 1, 2$ and for $s = 1, 2$.

$$\frac{\partial \ln L}{\partial \theta_{1r}} = \sum_{l=1}^{a} h_r(z_{l1}), \ \frac{\partial \ln L}{\partial \theta_{2s}} = \sum_{l=1}^{a} \sum_{j=2}^{n_l} h_s(z_{lj}), \ r = 3, 4, \cdots, k_1, \ s = 3, 4, \cdots, k_2.$$

Several of the variance and covariance terms are the same as those for the Legendre polynomial case. The terms that are different from the Legendre case are given below, and are in fact zeros.

$$\begin{aligned}
&\text{Cov}\left(\frac{\partial \ln L}{\partial \theta_{1r}}, \frac{\partial \ln L}{\partial \eta_c}\right) = 0, \quad c = 2, \cdots ., a \\
&\text{Cov}\left(\frac{\partial \ln L}{\partial \theta_{2r}}, \frac{\partial \ln L}{\partial \eta_c}\right) = 0 \\
&\text{Cov}\left(\frac{\partial \ln L}{\partial \theta_{2s}}, \frac{\partial \ln L}{\partial \eta_c}\right) = 0, \quad c = 2, \cdots, a.
\end{aligned}$$

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 121–141

A Nonparametric Bayesian Analysis of Response Data with Gaps, Outliers and Ties

Jiani Yin[1] **and Balgobin Nandram**[2]
[1] *Takeda Pharmaceuticals*
[2] *Department of Mathematical Sciences*
Worcester Polytechnic Institute

Received: 29 February 2020; Revised: 27 June 2020; Accepted: 01 July 2020

Abstract

Typically survey data have responses with gaps, outliers and ties, which we call GOT data, and the distributions of the responses might be skewed. Our application is on body mass index (BMI) data, which have these features, and inference is required about the 85^{th} and 95^{th} finite population percentiles. Because the data are collected using a two-stage sample design, usually predictive inference is done using a two-level Bayesian model with normality at both levels (responses and random effects). This is the Scott-Smith (S-S) model and it might not be robust against these GOT features. We use a two-level nonparametric Bayesian model, called the Dirichlet process Gaussian (DPG) model, with several independent Dirichlet processes at the first stage and a normal distribution on the random effects to accommodate the GOT data. The DPG model is different from the more popular two-level Dirichlet process mixture (DPM) model that has a single Dirichlet process on the random effects and independent normal distributions at the first level. Clearly, this DPM model has a shortcoming for survey data because the first stage has a normal distribution. We use our application on BMI GOT data and a very limited simulation study to compare the three models (S-S, DPM, DPG), which show, with the appropriate data, that the DPG model is preferred.

Key words: Bayesian computation; DPM model; GOT data; Predictive inference; Robust model; Survey data.

1. Introduction

When samples are selected from a finite population, the most commonly used method for making inferences in current statistical literature is design-based. This method is nonparametric and it requires large sample sizes for reliable inference. Model-based inference for finite populations has been proposed as an alternative to the design-based theory. Typically survey data have responses with gaps, outliers and ties, and the distributions of the responses might be skewed. Henceforth, we use the acronym, GOT, to describe these features of

Corresponding Author: Balgobin Nandram
Email: balnan@wpi.edu

our data, and here we focus on a nonparametric Bayesian analysis of GOT data. [For convenience, all acronyms, which are used in the paper, are presented in Table 1.] The United States' National Center of Health Statistics has been collecting health data since the 1950's and one of them, body mass index (BMI) data that we discuss, has these features. We consider three two-level models with the first level accommodating the responses and the second level accounting for heterogeneity of groups of data.

We assume that data are obtained from a two-stage sample survey, for example, a two-stage cluster sampling, stratified or post-stratified sampling that is often seen in small area problems. The sampled values are observed and the nonsampled values are to be predicted using the two-level models. To gain robustness, these models start with a simple idea that uses a random distribution (*e.g.*, a Dirichlet process) in the model instead of some parametric distributions. Assuming a specific parametric form is typically motivated by technical convenience rather than by genuine prior beliefs.

In many surveys, we want to estimate quantities not only for the population as a whole, but also for sub-populations (*e.g.*, to estimate the average income for every county in the United States in order to allocate funds for needed areas). Once a hierarchical model is specified, inferences can be drawn from available data for the population quantities at any level. From a Bayesian perspective, these estimators which can be regarded as posterior means often have better properties than area-specific direct estimators. This makes two-level, and more generally hierarchical Bayesian models, useful in the problem of small area estimation (SAE) (*e.g.*, Rao and Molina 2015). That is, the sample size for a given area or domain may be too small to provide reliable estimates for themselves and it may be needed to borrow information from neighboring areas, or from areas with similar characteristics. Typically, in this two-level model, the first level accommodates the response data and the second level is used to accommodate random effects or means (*i.e.*, the small areas).

BMI is a person's weight in kilograms divided by the square of her/his height in meters and it is used as a screening tool for overweight or obesity. A high BMI can be an indicator of high body fatness. If your BMI is less than 18.5, it falls within the underweight range. If your BMI is 18.5 to 24.9, it falls within the normal or healthy weight range. If your BMI is 25.0 to 29.9, it falls within the overweight range. If your BMI is 30.0 or higher, it falls within the obese range. A child's weight status is determined using an age- and sex-specific percentile for BMI rather than BMI categories used for adults. Overweight is defined as a BMI at or above the 85^{th} percentile and below the 95^{th} percentile for children and teens of the same age and sex. Obesity is defined as a BMI at or above the 95^{th} percentile for children and teens of the same age and sex.

The Expert Committee on Clinical Guidelines for Overweight in Adolescent Prevention Services published criteria for overweight to be integrated into routine screening of adolescents. BMI should be used routinely to screen for overweight and obesity in children and adolescents. Several disorders have been linked to overweight in childhood. A potential increase in type 2 diabetes mellitus is related to the increased prevalence of overweight in children (Fagot-Campagna 2000), as are cardiovascular risk factors, high cholesterol levels, and abnormal glucose levels.

We provide a Bayesian analysis of BMI data from the third National Health and Nutrition Examination Survey (NHANES III), conducted during the period October 1988 through September 1994. Due to confidentiality reasons, the final data set for this study uses only the 35 largest counties with population sizes at least 500,000. The sample sizes are less than 0.02% of the population sizes; see also Flegal *et al.* (2005, 2007) for discussions of other aspects of the NHANES III data. Kuczmarski *et al.* (2002) developed 85^{th} and 95^{th} percentile growth curves for US boys and girls age 2–20 years. Youths with a BMI in at least the 95^{th} percentile for age and sex, or at least 30 (World Health Organization Consultation of Obesity 2000) should be considered overweight and referred for in depth medical follow-up to explore underlying diagnoses. Adolescents with a BMI with at least the 85^{th} percentile (25) but below the 95^{th} percentile (30) should be considered at risk of overweight and referred for a second-level screen; these are different for adults. (See Himes and Dietz 1994 for a summary of this discussion.) Dietz (1998) discussed health consequences of obesity in youth and childhood predictors of adult disease. Currently, obesity is one of the most serious health problems facing the world.

Nandram and Choi (2005) obtained finite population mean for children and young adults under a nonignorable nonresponse model for small domains. But the 85^{th} and the 95^{th} percentiles are more important and informative for BMI data. So Nandram and Choi (2010), using BMI data from NHANES III, showed how to predict these finite population percentiles of BMI for some US counties, incorporating additional measures to minimize possible biases. These measures are the inclusion of survey weights into the nonignorable nonresponse model to reflect the higher probabilities of selection among black, non-Hispanics and Hispanic-Americans. Here, we perform a Bayesian analysis of BMI data from NHANES III to obtain the 85^{th} and 95^{th} finite population percentiles for adults older than twenty years. We do not incorporate survey weights or covariates (age, race, sex) into our analysis. A related objective might be to estimate the proportion of obese individuals using logistic regression (*e.g.*, Nandram, Chen, Fu and Manandhar 2018), but this is far from our main objective in this paper. Here, our main objective is to compare the performance of the three two-level Bayesian models for the analysis of these data.

In Figure 1, we have shown dot plots of the BMI data by county. We can see that there are many gaps, ties and outliers in all counties. A gap occurs because no value exists between two adjacent values. For example, because the BMI values are recorded to one decimal place, there are no values between 20.0 and 20.1 (gaps), there are several values at 24.0 (ties) and there are extreme values in the right tails of the dot plots (outliers). This is why we are troubled by GOT data. Other data, such as income, when they are elicited in surveys, are also GOT data.

However, we often know very little about the specific parametric forms of the distributions, and it is also difficult to completely validate the parametric assumptions. The parametric Bayesian models based on distributional assumptions may be problematic because inferences are sensitive to such assumptions. It may be more appealing to use a nonparametric Bayesian approach because we are interested in extreme percentiles. For example, as stated already, for BMI data interest is usually on the 85^{th} and 95^{th} finite population percentiles. These are in the extreme right tails of the BMI distribution, and because the

BMI data have ties, outliers and gaps, it is dubious for normal distribution to represent them.

Here, we discuss the statistical modeling associated with the analysis of two-level survey data. Our intention is to propose nonparametric Bayesian alternatives using the Dirichlet process (DP). This permits robustification of inference by embedding parametric models in nonparametric models, thereby avoiding critical dependence on parametric assumptions and to allow for heterogeneity, gaps, outliers, ties, *etc.*

The existence of the DP was established by Ferguson (1973). It is a distribution over distributions; each draw from a DP itself is a distribution (*i.e.*, operating on functional spaces). The DP has gained a lot of attention recently. It has nice properties such as clustering and borrowing information which is attractive to SAE, and it can be used to address the nonparametric analysis of GOT data. The Dirichlet process mixture (DPM) model has normality on the responses (not appropriate for GOT data) and a DP on the random effects.

The more appropriate model, introduced by Yin and Nandram (2020), has several Dirichlet processes on the response data and a normal distribution for the random effects; therefore, we call it the DPG model. The Scott-Smith (Scott and Smith 1969, S-S) model has normality on both levels. The difference between the DPM model and the DPG model is that DPM model (normality on the responses) does not accommodate GOT data but its main strength is its clustering property among the small areas (*i.e.*, random effects). The DPM model is actually the opposite of the DPG model, and they are both different from the S-S model that has normality at both levels. In this paper, we compare the analysis of BMI data using the three models (S-S, DPM and DPG); our contribution is not theoretical nor methodological. Incorporating the survey weights into the DPM or the DPG needs new theory and methodology.

In Section 2, we briefly review the Scott-Smith (S-S) model and the DPM model. In Section 3, we discuss the DPG model. Section 3.1 describes the DPG model and its computation. Section 3.2 shows how to do prediction of the finite population quantities under the DPG model. In Section 4, to compare the three models, we present an analysis of BMI GOT data and in the appendix a limited simulation study. In Section 5, we present our conclusions.

2. Scott-Smith Model and Dirichlet Process Mixture Model

In Section 2.1, we first review the S-S model, a two-level parametric model. It is used as a baseline model for the other two nonparametric models that we wish to discuss. In Section 2.2, we present a review of the Dirichlet process mixture (DPM) model. It is worth noting here that the S-S model is not robust against outliers in the data (*e.g.*, Gershunskaya and Lahiri 2018). This is also true for the DPM model. In addition, they do not make any adjustments for gaps and ties in the responses.

We assume that there are ℓ areas, and within the i^{th} area there are N_i (known) individuals. A sample of n_i individuals is available from the i^{th} area, and the remaining $N_i - n_i$

values are unknown. Inference is required for extreme percentiles (85^{th} and 95^{th}) of each area.

It is convenient to momentarily describe some notations. Let y_{ij} denote the value for the j^{th} unit within the i^{th} area, $i = 1, \ldots, \ell, j = 1, \ldots, N_i$. We assume that y_{ij}, $i = 1, \ldots, \ell, j = 1, \ldots, n_i$, are observed, and inference is required for the two finite population percentiles of the i^{th} area. Let $\underset{\sim}{y} = (\underset{\sim}{y}_s, \underset{\sim}{y}_{ns})$, where $\underset{\sim}{y}_s = \{y_{ij}, i = 1, \ldots, \ell, j = 1, \ldots, n_i\}$ is the vector of observed values and $\underset{\sim}{y}_{ns} = \{y_{ij}, i = 1, \ldots, \ell, j = n_i + 1, \ldots, N_i\}$ vector of unobserved values.

2.1. Two-stage Scott-Smith (S-S) parametric model

We describe the Bayesian version of the model of Scott and Smith (1969). This S-S model was developed by Nandram, Toto and Choi (2011) for continuous data y_{ij}, where $i = 1, \ldots, \ell, j = 1, \ldots, N_i$. Letting $\delta^2 = \frac{\rho}{1-\rho}\sigma^2$, our two-level normal model (baseline parametric model) is then

$$y_{ij}|\mu_i \overset{ind}{\sim} N\left(\mu_i, \sigma^2\right), \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, N_i, \tag{1}$$

$$\mu_i \overset{ind}{\sim} N\left(\theta, \frac{\rho}{1-\rho}\sigma^2\right), \tag{2}$$

$$\pi(\theta, \sigma^2, \rho) = \frac{1}{\pi(1+\theta^2)} \frac{1}{(1+\sigma^2)^2}, \quad -\infty < \theta < \infty, \quad \sigma^2 > 0, \quad 0 \le \rho \le 1.$$

Here, we consider a reparameterization of the S-S model, (1) and (2), together with proper non-informative priors that can allow computation of marginal likelihood and Bayes factors. We replace δ^2 by $\frac{\rho}{(1-\rho)}\sigma^2$ to gain some analytical and computational simplicity. Note that $\rho = \delta^2/(\delta^2 + \sigma^2)$ is a common intra-class correlation; see Toto and Nandram (2010), Nandram, Toto and Choi (2011) and Molina, Nandram and Rao (2014). We have used the Cauchy prior centered at 0 for θ; one can use a location-scale Cauchy prior distribution (*e.g.*, Gelman, Jakulin, Pittau and Su, 2008), but one would need to specify the location and scale parameters using the data (double using the data is forbidden in Bayesian statistics). The prior on σ^2 is a standard shrinkage prior (almost noninformative). The sampling importance resampling (SIR) algorithm is used to draw samples from the posterior distribution $\pi(\underset{\sim}{\mu}, \theta, \sigma^2, \rho|\underset{\sim}{y}_s)$; see Yin and Nandram (2020).

2.2. Dirichlet process mixture model

As pointed out by a reviewer, the Dirichlet process (DP) is well known and there is no need to review it. However, to set the stage, we need a brief description. First, we note that Binder (1982) was the first to introduce this model to survey sampling; more recently, see Nandram and Yin (2016 a,b). Let $(\Theta, \mathcal{B})$ be a measurable space, with G_0 a baseline measure (nonrandom) on the space; see Ferguson (1973) for a definition of the DP.

A Dirichlet process, $\mathrm{DP}(\alpha, G_0)$, is defined as the distribution of a random probability measure G over $(\Theta, \mathcal{B})$ such that, for any finite measurable partition of the measur-

able space Θ, $\{A_i\}_{i=1}^n$, $\{G(A_1), \ldots, G(A_n)\} \sim \text{Dirichlet}\{\alpha G_0(A_1), \ldots, \alpha G_0(A_n)\}$. We write $G \sim \text{DP}(\alpha, G_0)$, if G is a random probability measure with a distribution given by the DP, where α is the concentration parameter. Sethuraman (1994) presented an enormously useful representation of DP in the form of a stick-breaking algorithm.

Another implied representation of the DP is the generalized Polya urn scheme, which is obtained by integrating out the random measure, G. Now considering the predictive distribution for θ_{n+1} conditioned on $\theta_1, \ldots, \theta_n$ with G integrated out, we have

$$\theta_{n+1}|\theta_1, \ldots, \theta_n \sim \frac{\alpha}{\alpha + n} G_0(\theta_{n+1}) + \frac{1}{\alpha + n} \sum_{i=1}^{n} \Delta_{\theta_i}(\theta_{n+1}),$$

where $\Delta_a(x)$ is the cdf of a point mass at a. The sequence of predictive distributions for $\theta_1, \theta_2, \ldots$ is called the generalized Polya urn scheme (Blackwell and MacQueen 1973). Here, it is interesting that the probability measure G is discrete with probability one, but the k distinct values $\theta_1^*, \ldots, \theta_k^* \overset{iid}{\sim} G_0$, a continuous measure (*i.e.*, the θ_i are continuous, yet $\theta_i = \theta_j, i \neq j$). There is also a slightly more compressed form that we use for prediction.

In many applications, the almost sure discreteness of the DP measure may be inappropriate. As we noted, the most popular application of the DP is in clustering data using mixture models. There is a set of latent variables, $\{\mu_1, \ldots, \mu_\ell\}$, and as for finite populations, the model is

$$\begin{aligned} y_{ij}|\mu_i, \phi &\overset{ind}{\sim} h(y_{ij}; \mu_i, \phi), \quad j = 1, \ldots, N_i, i = 1, \ldots, \ell, \\ \mu_i|G &\sim G, \\ G &\sim \text{DP}(\alpha, G_0). \end{aligned} \tag{3}$$

This model is referred to as a Dirichlet process mixture (DPM) model; see Lo (1984) where the DPM was introduced. There are numerous applications of the DPM but see Nandram and Choi (2004) and Polettini (2017) for applications on SAE. Each μ_i is a latent parameter modeling y_{ij}, while G is the unknown distribution over parameters modeled using a DP. It can be seen as a Dirichlet process mixture of $h(y_{ij}; \mu_i, \phi)$, where y_{ij}'s with the same value of μ_i belong to the same cluster. The DPM model removes the constraint from discrete measures. It is worth noting that the DPM model for the response data is usually normal, and so it will not fit the GOT data very well. The corresponding parametric baseline model with G_0 replacing the random probability measure G is,

$$\begin{aligned} y_{ij}|\mu_i, \phi &\overset{ind}{\sim} h(y_{ij}; \mu_i, \phi), \quad j = 1, \ldots, N_i, i = 1, \ldots, \ell, \\ \mu_i &\sim G_0. \end{aligned}$$

Kalli, Griffin and Walker (2011), who suggested slice-efficient samplers, gave an improved slice sampling scheme to fit the DPM model that we use in our work, and it is based on the stick-breaking construction (Sethuraman 1994) without truncation error. The idea is to introduce latent variables that permit sampling a finite number of variables at each iteration.

In our context, DPM model is

$$\begin{aligned} y_{ij}|\mu_i,\sigma^2 &\overset{ind}{\sim} \text{Normal}(\mu_i,\sigma^2), \quad j=1,\ldots,N_i, && (4)\\ \mu_i|G &\sim G, \quad i=1,\ldots,\ell, \\ G \mid \theta,\sigma^2,\gamma,\rho &\sim \text{DP}\left\{\gamma, \text{Normal}(\theta, \frac{\rho}{1-\rho}\sigma^2)\right\}, && (5)\\ \pi(\theta,\sigma^2,\gamma,\rho) &= \frac{1}{\pi(1+\theta^2)}\frac{1}{(1+\sigma^2)^2}\frac{1}{(1+\gamma)^2}, && (6) \end{aligned}$$

where $-\infty < \theta < \infty$, $\sigma^2 > 0$, $\gamma > 0$, $0 \le \rho \le 1$, and γ is the concentration parameter. In this formulation the S-S model is a baseline model; the DPM model is centered on the S-S model and γ controls how close DPM model gets to the S-S model. Here, G is a random distribution function, discrete with probability one, with distribution $DP(\cdot,\cdot)$.

3. Dirichlet Process Gaussian (DPG) Model

Since there are gaps, outliers and ties (GOT) in survey data, it is reasonable to use a random distribution drawn from the DP for the sampling population. One drawback of the S-S model is over-shrinkage; the posterior mean of certain areas may be shrunk too much towards the overall mean. Using the DP allows borrowing information moderately within some of the areas instead of all. Moreover, since there are gaps, outliers and ties in the survey data, it is reasonable to use a random distribution drawn from the DP for the sampling population. Thus, it is important to use a nonparametric procedure.

3.1. Model and computation

We consider a nonparametric hierarchical Bayesian extension of the parametric baseline model with the uncertainty on the distribution of our sampling population. Using DPs in the first level and a parametric distribution as prior gives

$$\begin{aligned} y_{ij}|G_i &\overset{ind}{\sim} G_i, \quad i=1,\ldots,\ell, \quad j=1,\ldots,N_i, && (7)\\ G_i|\mu_i &\overset{ind}{\sim} \text{DP}\{\alpha_i, G_0(\mu_i)\}, \\ \mu_i &\overset{iid}{\sim} H_0(\cdot), \end{aligned}$$

where $G_0(\mu_i)$ and $H_0(\cdot)$ can be any parametric distributions. When we have strong beliefs that the area means are from a normal distribution, we may choose to use the normal likelihood in the second level. In particular, we consider $G_0 = N(\mu_i,\sigma^2)$ and $H_0(\cdot) = N(\theta,\delta^2)$, where $\delta^2 = \frac{\rho}{1-\rho}\sigma^2$ in (7) to be consistent with the two-level normal model. A full Bayesian model can be obtained by adding prior distributions. For example, we can use proper non-informative priors,

$$\begin{aligned} \pi(\alpha_i) &= \frac{1}{(\alpha_i+1)^2}, \qquad \alpha_i > 0, \quad i=1,\ldots,\ell, && (8)\\ \pi(\theta,\sigma^2,\rho) &= \frac{1}{\pi(1+\theta^2)}\frac{1}{(1+\sigma^2)^2}, \\ &\quad -\infty<\theta<\infty, 0<\sigma^2<\infty, 0\le\rho\le 1, && (9) \end{aligned}$$

with independence. Generally, it is not sensible to assume that the α_i are identically distributed because they can be very different. As apparent, we have been calling (7), (8) and (9) the DPG model.

Inference of the DPG model can be easily performed. We denote $(\mu, \theta, \sigma^2, \rho)$ as $\underset{\sim}{\psi}$ and $\underset{\sim}{\alpha} = \{\alpha_1, \ldots, \alpha_\ell\}$. The posterior density of α_i are independent with other parameters $\underset{\sim}{\psi}$ in the model, conditioning on only the distinct values. Let k_i denote the number of distinct values for each area in the observed data, $\underset{\sim}{k} = \{k_i, i = 1, \ldots, \ell\}$ be the vector of k_i, $y^*_{i1}, \ldots, y^*_{ik_i}$ be the k_i distinct sample values for each i and $\underset{\sim}{y}^* = \{y^*_{i1}, \ldots, y^*_{ik_i}, i = 1, \ldots, \ell\}$ be the vector of y^*_{ij}. Thus the joint posterior density is

$$\pi(\underset{\sim}{\alpha}, \underset{\sim}{\psi} \mid \underset{\sim}{k}, \underset{\sim}{y}^*) = \left[\prod_{i=1}^{\ell} \pi(\alpha_i \mid k_i)\right] \pi(\underset{\sim}{\psi} \mid \underset{\sim}{y}^*), \tag{10}$$

where $\pi(\alpha_i|k_i) \propto \pi(k_i \mid \alpha_i)\pi(\alpha_i)$. For the other parameters $\underset{\sim}{\psi}$, we have

$$\begin{aligned} y^*_{ij}|\mu_i &\overset{ind}{\sim} N\left(\mu_i, \sigma^2\right), \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, k_i, \\ \mu_i &\overset{iid}{\sim} N\left(\theta, \frac{\rho}{1-\rho}\sigma^2\right), \\ \pi(\theta, \sigma^2, \rho) &= \frac{1}{\pi(1+\theta^2)} \frac{1}{(1+\sigma^2)^2}, -\infty < \theta < \infty, 0 < \sigma^2 < \infty, 0 \le \rho \le 1. \end{aligned} \tag{11}$$

Therefore, the algorithm for the DPG model is
Step 1 : For each i $(i = 1, \ldots, \ell)$, draw α_i from $\pi(\alpha_i|k_i) \propto \alpha^{k_i} \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i+n_i)} \frac{1}{(\alpha_i+1)^2}$ (Antoniak 1974).
Step 2: Draw $\underset{\sim}{\psi}$ from the parametric model (11) which is easy to fit.

Finally, we highlight how the DPG model takes care of GOT responses; this is apparent in the sampling process. When we integrate out the random probability measure (Blackwell and MacQueen, 1973), we get

$$\begin{aligned} f(\underset{\sim}{y_i} \mid \mu_i, \sigma^2, \alpha_i) &= \frac{1}{\sigma}\phi\left(\frac{y_{i1} - \mu_i}{\sigma}\right) \\ &\times \prod_{k=2}^{n_i} \left\{ \frac{k-1}{\alpha_i + k - 1} \frac{\sum_{j=1}^{k-1} \delta_{y_{ij}}(y_{ik})}{k-1} + \frac{\alpha_i}{\alpha_i + k - 1} \frac{1}{\sigma}\phi\left(\frac{y_{ik} - \mu_i}{\sigma}\right) \right\}, \end{aligned} \tag{12}$$

where $\delta_a(b)$ means that b is a point mass at a; so ties are accommodated. Therefore, in each area we are mixing the distributions in (12) using normal mixing distributions in the DPG model. The DPM is different being a Dirichlet process mixture of normals. The DPM model actually produces ties among the random effects or area means (clustering), which is its major strength. But it does not model gaps, outliers, ties and skewness among the responses. By putting DPs on the responses in different areas, we are attempting to accommodate the GOT data.

3.2. Prediction for the finite population

We have a simple random sample of size n_i from a finite population of size N_i, $i = 1, \ldots, \ell$. Let $y_{i1}, \ldots, y_{in_i}$ denote the sampled values. We want to predict $y_{in_i+1}, \ldots, y_{iN_i}$, the nonsampled values, and obtain the posterior predictive distributions for the 85^{th} and 95^{th} finite population percentiles for each area. The sampling process is

$$\begin{aligned} y_{ij}|G_i &\overset{ind}{\sim} G_i, \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, N_i, \\ G_i|\mu_i &\overset{ind}{\sim} \mathrm{DP}\{\alpha_i, G_0(\mu_i)\}. \end{aligned}$$

Predictive inference for the DPG model simply uses the generalized Polya urn scheme (Blackwell and MacQueen 1973) for each area, since all areas are independent (see Nandram and Yin 2016 a,b). Once we have obtained the nonsampled $y_{ij}, j = n_i+1, \ldots, N_i, i = 1, \ldots, \ell$, we can now calculate any finite population quantity of interest. For example, for BMI data, we are interested in the finite population 85^{th} percentile (overweight individuals) and the 95^{th} percentile (obese individuals). The N_i are assumed known, and they are obtained from the 1990 census. To obtain the percentiles, one simply sort all the data (sample values and predicted non-sample values) in increasing order. Then, for the 85^{th} percentile, pick the value at $.85N_i$ (nearest integer) position, and for the 95^{th} percentile, pick the value at $.95N_i$ (nearest integer) position. Also, it is more difficult to estimate the two percentiles because they are in the right tail of the posterior distributions.

Because the N_i are very large ($N_i = n_i/.0002$), it takes relatively more time to compute the percentiles than other finite population quantities. One needs to sort $y_{i1}, \ldots, y_{iN_i}$ at each iteration, and the observed values can take different positions in the sorting. Prediction is relatively easier in the S-S and DPM models because it is done under normality, whereas in the DPG model, it is done under the Polya urn scheme.

4. Application to Body Mass Index Data

We fit the three models (S-S, DPM and DPG) to the BMI data. Our objective is mainly to compare the three models. As we mentioned in previous sections, survey data tend to have gaps, outliers and ties. The BMI data set is an example because in practice, BMI is rounded to one decimal place which creates many ties, and therefore the BMI data are a prime example of GOT data. We present the dot plots for all thirty-five areas (see Figure 1). The observations are more concentrated and having ties within the range around 25. It is also clear that the data are clustered and present gaps. Especially outside the normal weight range, the data become sparse and present bigger gaps.

The Gibbs sampler is needed for only the DPM model; for the S-S model and DPG model, we use random samplers, and therefore no monitoring is required. For the DPM model, we ran $10,000$ iterations, used $5,000$ as a "burn in" and thin every 5^{th} to obtain $1,000$ converged posterior samples. We have computed the p values of the Geweke test and the effective sample sizes for the parameters σ^2, θ, δ^2 and γ for the DPM model. The p values are respectively .48, .41, .46, .62 and the effective sample sizes are respectively

1000, 1000, 698, 1085, thereby showing that the chain is stationary and strongly mixing. Also, trace plots and auto-correlation plots indicate that the chains converge.

For model assessment, we computed the delete-one cross validation (CV) divergence measure (Wang *et al.* 2012). The CV, obtained by predicting y_{ij} when it is deleted to obtain $y_{(ij)}$, is

$$\mathrm{CV} = \frac{1}{\sum_{i=1}^{\ell} n_i} \sum_{i}^{\ell} \sum_{j=1}^{n_i} |y_{ij} - \mathrm{E}(y_{ij}|y_{(ij)})|,$$

$$\mathrm{E}(y_{ij}|y_{(ij)}) = \mathrm{E}_{\Omega|y_{(ij)}} \mathrm{E}(y_{ij} \mid y_{(ij)}, \Omega) = \int \mathrm{E}(y_{ij} \mid y_{(ij)}, \Omega) f(\Omega|y_{(ij)}) d\Omega.$$

where Ω is the set of all parameters. A Monte Carlo estimator of $\mathrm{E}(y_{ij}|y_{(ij)})$ is

$$\widehat{\mathrm{E}(y_{ij}|y_{(ij)})} = \frac{\sum_{h=1}^{M} \{f(y_{ij} \mid y_{(ij)}, \Omega^{(h)})\}^{-1} \mathrm{E}(y_{ij} \mid y_{(ij)}, \Omega^{(h)})}{\sum_{h=1}^{M} \{f(y_{ij} \mid y_{(ij)}, \Omega^{(h)})\}^{-1}},$$

$j = 1, \ldots, n_i, i = 1, \ldots, \ell$. Note that this measure is essentially a weighted average of the $\mathrm{E}(y_{ij} \mid y_{(ij)}, \Omega^{(h)})$ (*i.e.*, a prediction-based measure), it is not based directly on a likelihood function. For the S-S model, DPM model and DPG model the CVs are respectively 0.765, 0.766 and 0.772. So based on this measure, there is virtually no difference among these models.

We have studied other likelihood-based measures. However, when a parametric model is nested in a nonparametric alternative, any likelihood-based diagnostics (*e.g.*, deviance information criterion, Bayesian predictive p values, log-pseudo-marginal likelihood, Bayes factor) will be misleading because we are comparing infinite dimensional distributions. One possible explanation of this fact is that the DP generates discrete distributions with probability one. This phenomenon can arise, more generally, in different contexts (*e.g.*, using the DP in goodness of fit testing). Carota and Parmiginani (1996) and Petrone and Raftery (1997) pointed out that the discreteness of the DP can have a large effect on inferences of posterior distributions and Bayes factors, when the data are partially exchangeable with an unknown partition.

We perform the predictive inference of the 85^{th} and 95^{th} finite population percentiles for each area using the three models (S-S, DPM and DPG). We also use a Bayesian bootstrap (*e.g.*, Yin and Nandram 2020) to do prediction in each county individually without borrowing across counties as a comparison (*i.e.*, the assumption of similarity across counties is not used). Note that for the county level, all sample sizes are roughly 100, about .02% of the population sizes. We have compared the DPG model to the S-S model, the DPM model and Bayesian bootstrap.

In Tables 2 and 3, we present summary statistics, posterior mean (PM) and posterior standard deviation (PSD), of the 85^{th} and 95^{th} finite population percentiles for each county of BMI data under the three models (S-S, DPM and DPG) and Bayesian bootstrap respectively. Again, the bootstrap does not allow for pooling.

First, consider the 85^{th} percentile in Table 2. The PMs are roughly the same with those for the DPG model slightly higher, thereby showing how the DPG model takes care

of the data in the right tail of the data distribution. The PSDs under bootstrap are always larger than those under the three models, sometimes as much as two times. This shows that pooling of information across counties is helpful. However, the PSDs under the three models are roughly in increasing order: S-S model, DPM model and DPG model; the PSDs under the DPG model are always the largest, again showing how the DPG model takes care of the data in the right tail of the data distribution.

Second, we note that there are similar results in Table 3, but the differences are sharper. For example, the PSDs under the bootstrap are much too large and they are much larger under the DPG model than under the S-S model and DPM model.

In short, Tables 2 and 3 are very informative. It is not true that because the PSDs under the DPG model are larger than those under the S-S and DPM model, the DPG is worse. On the contrary, it is true that the DPG model has higher PSDs because it takes care of the GOT features of the BMI data. The S-S model and DPM model understate the PSDs because they do not take care of the gaps, outliers and ties in the data; the DPG model has a stochastic mechanism that accounts for the gaps, outliers and ties.

In Figures 2 and 3, we present plots of the posterior means with credible bands versus direct estimates for BMI data. The predictive inferences of the population percentiles are similar under the S-S model and DPM model. However, the DPG model tends to have higher predictions of the population percentiles with similar credible bands as compared with the other models. We suspect that S-S model and DPM model might underestimate the 85^{th} and 95^{th} finite population percentile when the GOT data are right skewed. In both Figures 2 and 3, we notice that the points under the DPG model are closer to the 45-degree straight line than those under the S-S model and DPM model. This effect is more intense in Figure 3 than in Figure 2, where the points under the S-S model and the DPM model appear to be more horizontally spread out. This is an important point because it shows that there is too much pooling in the S-S model and the DPM model. In particular, it shows that because the DPG model takes care of the GOT features of the BMI data, it is able to represent the pooling effect the best, and without the restrictive parametric assumptions in the sampling process, the DPG model tends to provide less biased estimation.

Finally, in Appendix A, we have performed a limited simulation study. We have generated data like the BMI data using the DPG model and then we fit all three models (S-S, DPM, DPG) to it. All we need from the simulation study is to describe when the DPG model performs better than the S-S model and the DPM model. We have shown that when posterior inference is performed for the 85^{th} and 95^{th} finite population percentiles, the DPG model performs much better than the S-S model and the DPM models, as required; see Appendix A.

Since BMI data are right skewed with gaps, outliers in the right tails and ties, the estimations given by parametric models may be incorrect. Thus based on a belief that the parametric model is too restrictive, we prefer the analysis based on the nonparametric DPG model.

5. Concluding Remarks

Bayesian nonparametric methods are motivated by the desire to avoid overly restrictive parametric assumptions. For GOT data (gaps, outliers and ties), we believe that our DPG model will play an important role for future analysis of "continuous" survey data. The S-S model and the DPM model have normality assumption on the response data, and hence they do not address the problem of GOT data. Based on the analyses, we have positive view of the DPG model.

Our main conclusion is that when data come from the DPG model, it will do much better than the S-S model and the DPM model. The DPM model, while nonparametric (good), it has the DP on the wrong level (random effects) for GOT responses. The S-S model has normality on both levels. Normality on the responses is not appropriate for GOT data. Hence, neither of these two models can accommodate GOT responses. Based on this point, we believe that among the three models (S-S, DPM, DPG), it is the DPG model that is most appropriate to represent the BMI GOT data. The DPG model is more important when interest is on the 85^{th} and 95^{th} finite population percentiles because there are larger gaps in the right tails of the data distributions (*i.e.*, outliers). However, the DPM model is attractive for an important reason; it provides clustering among the small areas but not the data. This clustering among the small areas can be accommodated in the DPG model; see Nandram and Yin (2019).

It is important to include survey weights in the DPG model. This can be done using a standardized composite likelihood. Covariates can also be incorporated into the DPG model. However, our main contribution in this paper has been to demonstrate the superiority of the DPG model when it is fitted to the BMI GOT data.

Acknowledgments

This research was supported by a grant from the Simons Foundation (#353953, Balgobin Nandram). The authors thank the two reviewers, whose comments have changed and improved the original manuscript enormously, and Professor Vinod Gupta for his invitation to honor the "Twin Brothers". The authors also thank Yang Liu and Jai Won Choi for their very careful reading of the manuscript.

References

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, **2(6)**, 1152-1174.

Binder, D. A. (1982). Non-parametric Bayesian models for samples from finite populations. *Journal of the Royal Statistical Society*, Series B, **44(3)**, 388-393.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Polya urn schemes. *The Annals of Statistics*, **1(2)**, 353-355.

Carota, C. and Parmigiani, G. (1996). On Bayes factor for nonparametric alternatives. *Bayesian Statistics*, **5**, 507-511.

Dietz, W. H. (1998). Health consequences of obesity in youth: Childhood predictors of adult disease. *Pediatrics*, **101**, 518–525.

Fagot-Campagna, A. (2000). Emergence of type 2 diabetes mellitus in children: Epidemiological evidence. *Journal of Pediatric Endocrinology Metabolism*, **13**, 1395–1405.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1(2)**, 209-230.

Flegal, K. M., Graubard, B. I., Williamson, D. F., and Gail, M. H. (2005). Excess deaths associated with underweight, overweight, and obesity. *Journal of the American Medical Association*, **293**, 1861–1867.

Flegal, K. M., Graubard, B. I., Williamson, D. F., and Gail, M. H. (2007). Cause-specific excess deaths associated with underweight, overweight, and obesity. *Journal of the American Medical Association*, **298**, 2028–2037.

Gelman, A., Jakulin, A., Pittau, M. P. and Su, Y-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, **2(4)**, 1360-1383.

Gershunskaya, J. and Lahiri, P. (2018). Robust empirical best small area finite population mean estimation using a mixture model. *Calcutta Statistical Association Bulletin*, **69(2)**, 183-204.

Himes, J. H., and Dietz, W. H. (1994). Guidelines for overweight in adolescent preventive services: Recommendations from an expert committee. *American Journal of Clinical Nutrition*, **59**, 307–316.

Kalli, M. and Griffin, J. E. and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, **21(1)**, 83-105.

Kuczmarski, R. J., Ogden, C. L., Guo, S. S., Grummer-Strawn, L. M., Flegal, K. M., Mei, Z., Wei, R., Curtin, L. R., Roche, A. L., and Johnson, C. L. (2002). 2000 CDC growth charts for the united states: Methods and development. *Vital and Health Statistics*, **11(246)**, 1–190.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, **12(1)**, 351-357.

Molina, I., Nandram, B. and Rao, J. N. K., (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *The Annals of Applied Statistics*, **8(2)**, 852-885.

Nandram, B. and Choi, J. W. (2004). Nonparametric Bayesian analysis of a proportion for a small area under nonignorable nonresponse. *Journal of Nonparametric Statistics*, **16(6)**, 821-839.

Nandram, B., Chen, L., Fu, S. and Manandhar, B. (2018). Bayesian Logistic Regression for Small Areas with Numerous Households. *Statistics and Application*, **16(1)**, 171-205.

Nandram, B. and Choi, J. W. (2010). Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection. *Journal of the American Statistical Association*, **105(489)**, 120-135.

Nandram, B. and Choi, J. W. (2005). Hierarchical Bayesian nonignorable nonresponse regression models for small areas: An application to the NHANES data. *Survey Methodology*, **31(1)**, 73-84.

Nandram, B. and Toto, M. C. S. and Choi, J. W. (2011). A Bayesian benchmarking of the Scott-Smith model for small areas. *Journal of Statistical Computation and Simulation*, **81(11)**, 1593-1608.

Nandram, B. and Yin, J. (2016a). Bayesian predictive inference under a Dirichlet process with sensitivity to the normal baseline. *Statistical Methodology*, **28**, 1-17.

Nandram, B. and Yin, J. (2016b). A nonparametric Bayesian prediction interval for a finite population mean. *Journal of Statistical Computation and Simulation*, **86(16)**, 3141-3157.

Nandram, B. and Yin, J. (2019). Hierarchical Bayesian Models for small areas with Dirichlet processes. *JSM Proceedings, 2594-2613, Survey Research Methods Section. Alexandria, VA: American Statistical Association.*

Polettini, S. (2017). A generalized semiparametric Bayesian Fay-Herriot model for small area estimation shrinking both means and variances. *Bayesian Analysis*, **12(3)**, 729-752.

Petrone, S. and Raftery, A. E. (1997). A note on the Dirichlet process prior in Bayesian nonparametric inference with partial exchangeability. *Statistics and Probability Letters*, **36(1)**, 69-83.

Rao, J. N. K. Molina, I. (2015). *Small Area Estimation*, John Wiley and Sons, NY.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**, 639-650.

Scott, A. and Smith, T. M. F. (1969). Estimation in multi-stage surveys. *Journal of the American Statistical Association*, **64(327)**,830-840.

Toto, M. C. S. and Nandram, B. (2010). A Bayesian predictive inference for small area means incorporating covariates and sampling weights. *Journal of Statistical Planning and Inference*, **140**, 2963-2979.

Wang, J. C., Holan, S. H., Nandram, B., Barboza, W., Toto, C. and Anderson, E. (2012). A Bayesian approach to estimating agricultural yield based on multiple repeated surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, **17(1)**, 84-106.

WHO Consultation on Obesity (2000).Obesity: Preventing and managing the global epidemic. *Technical Report 894*, World Health Organization, Geneva, Switzerland.

Yin, J. and Nandram, B. (2020). A Bayesian small area model with Dirichlet processes on responses. *Statistics in Transition, New Series*, **21(3)**, 1–19.

APPENDIX A: Simulation Study Showing the Strength of the DPG Model

We conduct a limited simulation study. We have simulated a single data set from the DPG model and we fitted all three models (S-S, DPM, DPG) to this data set.

We choose $\ell = 50$ and the sample sizes, n_i, for 50 areas. The sample sizes are 35 for each of the first 10 areas, 50 for each of the second 10 areas, 100 for each of the third 10 areas, 200 for each of the fourth 10 areas and 500 for each of the last 10 areas. Then, the population sizes are selected as $N_i = 100n_i, i = 1, \ldots, \ell$. These are comparable to the BMI data. For convenience, to simulate the data set, we have taken $\theta = 0.0$, $\sigma^2 = 0.01$, $\delta^2 = 0.04$, thereby making $\rho = 0.8$. For the concentration parameters of the Dirichlet processes, we have selected $\alpha_i \overset{ind}{\sim} 0.5 + \text{Beta}(5, 5), i = 1, \ldots, \ell$. These choices allow us to have data similar to the BMI data with some flexibility to get gaps, outliers and ties when data are simulated from the DPG model.

We use absolute bias (AB) and posterior root mean squared error (PRMSE) to compare the models. We know the true values of the finite population quantities, denoted by T. Then, $AB = \mid PM - T \mid$ and $PRMSE = \sqrt{(PM - T)^2 + PSD^2}$. We compute these quantities for each of the fifty counties for the 85^{th} and 95^{th} finite population percentiles, and respectively we average them over the fifty counties. We present AB and PRMSE in Table 4 for the case in which data are generated from the DPG model (*i.e.*, for GOT data). It is nice that AB and $PRMSE$ are manyfold smaller under the DPG model than the other two models (S-S, DPM). Therefore, it is clear that the DPG model performs much better than the S-S model and the DPM model, when inference is made about the 85^{th} and 95^{th} finite population percentiles. This is exactly what we want to happen.

Table 1: Acronyms used in the Presentation

Acronym	Meaning
GOT	gaps, outliers and ties
BMI	body mass index
SAE	small area estimation
S-S	Scott-Smith
DP	Diriclet process
DPM	Dirichlet process mixture
DPG	Dirichlet process Gaussian
AB	absolute bias
PRMSE	posterior root mean squared error
PM	posterior mean
PSD	posterior standard deviation
CV	cross validation

Table 2: Comparison of posterior mean (PM) and posterior standard deviation (PSD) of the finite population 85^{th} percentile for each county of body mass index (BMI) data by te three models (S-S, DPM, DPG) and Bayesian bootstrap

	Bootstrap		S-S		DPM		DPG	
	PM	PSD	PM	PSD	PM	PSD	PM	PSD
1	32.14	0.50	32.48	0.35	32.50	0.39	32.46	0.47
2	34.76	1.24	32.93	0.45	32.95	0.43	34.08	0.82
3	30.76	0.78	32.05	0.39	32.00	0.44	31.94	0.63
4	31.57	1.07	31.97	0.43	31.93	0.42	32.48	0.61
5	30.51	0.90	31.75	0.47	31.75	0.45	31.87	0.72
6	33.82	1.22	33.42	0.44	33.35	0.44	33.55	0.67
7	31.59	0.85	32.58	0.36	32.58	0.39	32.45	0.72
8	32.25	0.67	32.46	0.36	32.48	0.42	32.70	0.53
9	32.81	1.18	33.03	0.41	33.01	0.42	33.15	0.75
10	34.01	0.74	33.07	0.39	33.08	0.36	33.73	0.48
11	32.75	0.54	32.78	0.26	32.79	0.27	32.90	0.49
12	30.26	0.80	31.67	0.42	31.67	0.38	31.45	0.53
13	31.91	0.88	32.34	0.36	32.32	0.43	32.64	0.57
14	32.37	0.38	32.80	0.19	32.82	0.20	32.50	0.37
15	33.39	0.50	32.84	0.40	32.85	0.41	33.39	0.42
16	32.21	0.75	32.72	0.37	32.71	0.40	32.73	0.62
17	30.88	0.83	31.95	0.40	31.91	0.42	32.07	0.72
18	31.18	0.80	32.29	0.39	32.28	0.49	32.21	0.85
19	32.03	0.97	32.09	0.38	32.08	0.42	32.77	0.56
20	32.71	0.96	32.50	0.39	32.52	0.42	33.08	0.61
21	33.08	0.98	32.57	0.40	32.58	0.44	33.28	0.56
22	32.06	0.72	32.65	0.36	32.68	0.37	32.57	0.57
23	31.18	0.77	31.85	0.42	31.81	0.42	32.19	0.70
24	32.66	0.66	32.64	0.37	32.68	0.40	32.96	0.52
25	31.63	0.98	32.37	0.39	32.39	0.42	32.47	0.73
26	32.02	0.96	32.34	0.40	32.30	0.45	32.77	0.57
27	31.56	0.44	32.34	0.31	32.36	0.39	32.16	0.50
28	33.51	1.51	32.87	0.39	32.89	0.40	33.33	0.80
29	31.53	0.97	32.30	0.45	32.31	0.49	32.57	0.80
30	30.62	0.94	31.89	0.43	31.83	0.45	32.13	0.71
31	32.36	0.57	33.02	0.38	32.99	0.38	32.72	0.49
32	33.24	0.89	32.96	0.37	32.96	0.37	33.31	0.62
33	30.54	0.51	32.03	0.37	32.01	0.42	31.61	0.57
34	32.48	0.49	32.78	0.31	32.82	0.31	32.71	0.45
35	31.78	1.04	32.40	0.35	32.41	0.42	32.54	0.75

Table 3: Comparison of posterior mean (PM) and posterior standard deviation (PSD) of the finite population 95^{th} percentile for each county of body mass index (BMI) data by the three models (S-S, DPM, DPG) and Bayesian bootstrap

	Bootstrap		S-S		DPM		DPG	
	PM	PSD	PM	PSD	PM	PSD	PM	PSD
1	35.52	1.27	35.79	0.42	35.81	0.45	36.21	0.88
2	40.88	2.32	36.45	0.46	36.47	0.45	38.83	1.54
3	34.90	2.58	35.36	0.47	35.32	0.51	36.16	1.43
4	35.59	1.12	35.31	0.45	35.27	0.45	36.26	0.85
5	35.82	1.61	35.19	0.51	35.19	0.50	36.53	0.92
6	39.32	1.58	37.00	0.44	36.94	0.44	38.45	0.74
7	35.93	1.12	35.95	0.40	35.94	0.44	36.50	0.69
8	37.32	1.49	35.90	0.43	35.92	0.48	37.26	0.86
9	38.76	1.54	36.55	0.45	36.53	0.46	38.02	0.84
10	39.82	1.64	36.48	0.41	36.48	0.41	38.32	1.14
11	37.49	0.94	36.19	0.28	36.21	0.29	37.36	0.71
12	35.84	1.50	35.17	0.47	35.18	0.44	36.46	0.89
13	36.13	1.20	35.68	0.40	35.66	0.45	36.65	0.93
14	36.90	0.80	36.16	0.22	36.19	0.23	36.96	0.69
15	36.04	1.47	36.00	0.48	36.03	0.49	36.64	0.89
16	36.44	1.40	36.08	0.41	36.08	0.44	36.79	0.93
17	34.70	0.99	35.27	0.44	35.23	0.45	35.77	0.83
18	35.57	0.81	35.68	0.38	35.65	0.46	36.16	0.78
19	34.88	0.88	35.31	0.40	35.30	0.44	35.85	0.78
20	37.08	1.89	35.82	0.42	35.84	0.46	37.11	1.14
21	35.75	1.03	35.75	0.44	35.77	0.47	36.30	0.84
22	35.56	1.08	35.94	0.43	35.98	0.42	36.12	0.89
23	36.46	1.46	35.29	0.45	35.24	0.46	36.84	0.92
24	37.80	2.17	36.02	0.44	36.06	0.45	37.40	1.33
25	37.29	2.60	35.76	0.43	35.77	0.46	37.23	1.47
26	36.18	1.92	35.67	0.52	35.62	0.55	36.90	1.10
27	36.09	1.30	35.75	0.38	35.77	0.44	36.51	0.78
28	40.33	1.37	36.50	0.44	36.53	0.46	38.84	0.96
29	35.71	1.10	35.66	0.52	35.67	0.52	36.43	0.78
30	34.57	1.11	35.20	0.48	35.15	0.49	35.87	0.83
31	35.43	1.06	36.28	0.39	36.26	0.39	36.01	0.68
32	39.12	1.40	36.43	0.41	36.43	0.40	38.24	1.00
33	34.10	0.83	35.31	0.42	35.30	0.46	35.32	0.88
34	35.98	1.02	36.09	0.36	36.12	0.36	36.36	0.85
35	37.83	1.13	35.92	0.38	35.92	0.44	37.57	0.92

Table 4: Comparison of absolute bias (AB) and posterior root mean squared error (PRMSE) of the finite population 85^{th} percentile and 95^{th} percentile for the simulated data from DPG model averaged over areas

	S-S Model		DPM Model		DPG Model	
	AB	PRMSE	AB	PRMSE	AB	PRMSE
85^{th} percentile	379.0	385.9	384.1	394.6	18.05	40.0
95^{th} percentile	550.2	556.3	555.3	563.7	35.5	101.0

NOTE: Data are generated from the DPG model, and all three models (S-S, DPM, DPG) are fitted to the data. The numbers in the table must be multiplied by 10^{-4}.

Figure 1: Dot plots of body mass index (BMI) for thirty-five counties

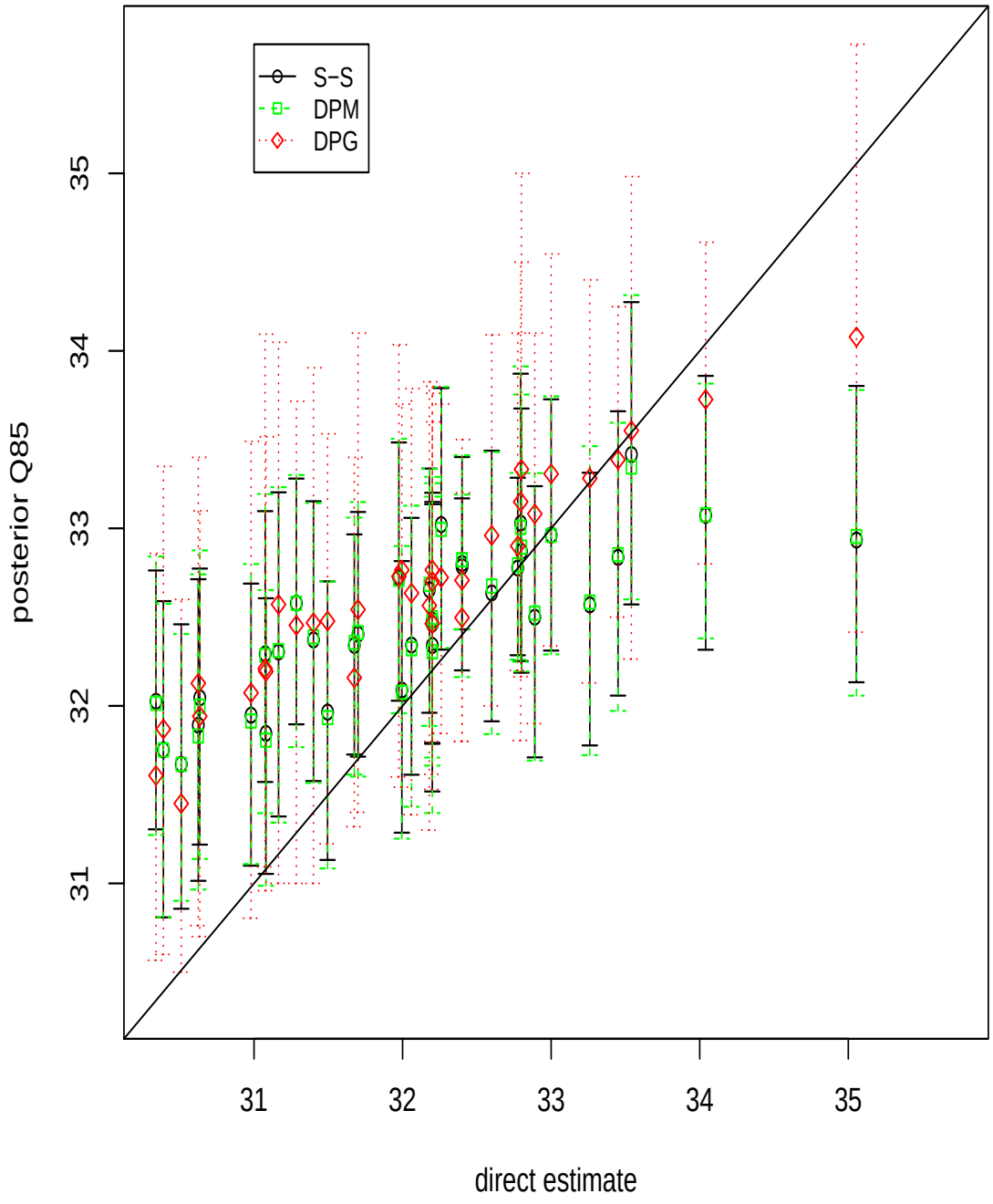


Figure 2: Comparison for body mass index (BMI) data (posterior means with credible bands versus direct estimates): the predictive inference of the finite population 85^{th} percentile for each county under the three models (S-S, DPM, DPG)

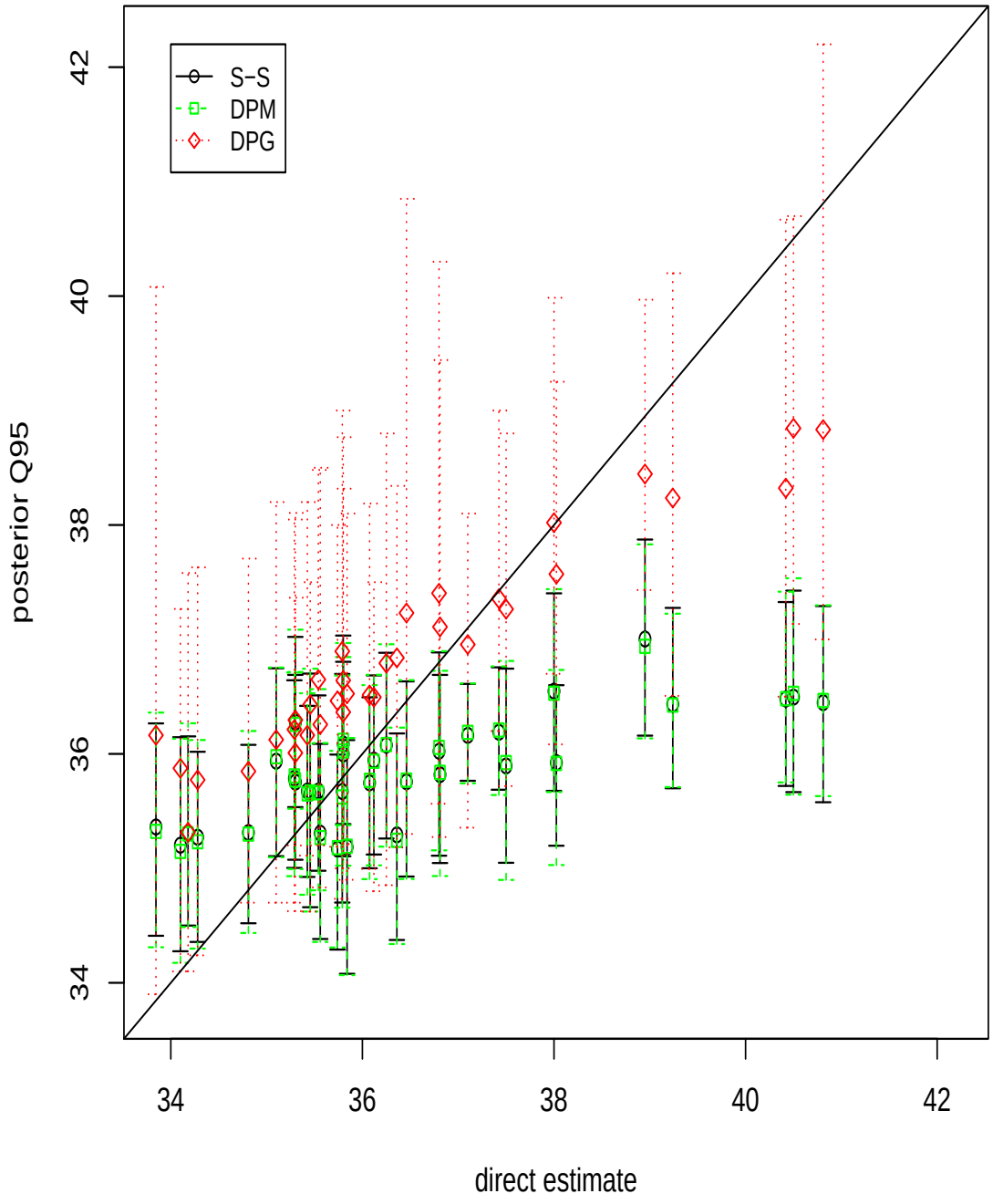


Figure 3: Comparison for body mass index (BMI) data (posterior means with credible bands versus direct estimates): the predictive inference of the finite population 95^{th} percentile for each county under the three models (S-S, DPM, DPG)

Statistics and Applications {ISSN 2454-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 143–154

From Experimental Design to Proportional Representation – A Tribute to Bikas Kumar Sinha and Bimal Kumar Sinha on Their 75th Birthday

Friedrich Pukelsheim
Institute for Mathematics, University of Augsburg, Germany

Received: 21 March 2020; Revised: 06 July 2020; Accepted: 10 July 2020

Abstract

The fields of design of statistical experiments and of proportional representation systems share the problem of approximating virtually continuous weights by distinctly discrete proportions. We explain the common aspects of the problem, review three highlights of the underlying theory, and illustrate the usefulness of the results by examples relating to electoral systems, sampling plans, and experimental design.

Key words: Apportionment rules; Divisor methods; Coherence theorem; Seat bias theorem; Goodness-of-fit theorem.

1. Prologue

The common denominator of Professor Bikas Sinha and Professor Bimal Sinha and myself is our joint research in the design and analysis of statistical experiments, dating back to the last millennium. I have fond memories of the discussions with one or the other of the twin professors when visiting them in Delhi 1988, at UMBC 1991, in Kolkata 1994 and, conversely, playing host in Augsburg 1993. Our relations culminated in the joint paper Pukelsheim and Sinha (1995) which merged Bikas' expertise in exact block designs with my interest in optimal approximate designs.

The two fields, optimality analysis of approximate designs and combinatorial construction of block designs, exhibit a complementary character. The first forms part of continuous mathematics, the second, of discrete mathematics. The transition from the continuous domain to the discrete domain was one of the topics dealt with in Pukelsheim and Sinha (1995). Beyond the statistical origin, the transition problem turned out to be quite intriguing by itself. When I stumbled into the problem I did not know nor preview that it would keep me busy to date. In the sequel I shall review three highlights that I found particularly intriguing.

Corresponding Author: Friedrich Pukelsheim
Email: Pukelsheim@Math.Uni-Augsburg.DE

2. Apportionment Rules

Suppose there is a set of experimental conditions, labeled $j = 1, \ldots, \ell$, for which a (continuous) experimental design indicates that a share w_j of all observations is to be realized under condition j. That is, the shares $w_1, \ldots, w_\ell$ are nonnegative weights (*i.e.* "continuous" real numbers) summing to unity.

Practically, limited funds restrict the experimenter to a finite sample size h, say. What we seek, then, is an (exact) experimental design $x_1, \ldots, x_\ell$ consisting of frequencies (*i.e.* "exact" natural numbers) summing to h such that the proportion x_j/h of observations becomes as equal as can be to the optimal weight w_j, that is,

$$x_1 \approx hw_1, \quad \ldots, \quad x_\ell \approx hw_\ell.$$

If all scaled weights $hw_1, \ldots, hw_\ell$ happen to be natural numbers, the exact solution is $x_1 = hw_1$, ..., $x_\ell = hw_\ell$ and the job is done.

However, the quantities hw_j generally fail to be natural numbers and cannot serve as the frequencies sought. Thus, in general, pure proportionality is impossible. The question arises how to split the sample size h into frequencies $x_1, \ldots, x_\ell$ that are reasonably – in some sense or other – proportional to the weights $w_1, \ldots, w_\ell$.

Evidently the terminology is not geared towards the setting of experimental design. The notation originates from a field that comes with an isomorphic problem, the study of proportional representation systems in parliamentary democracies.

The typcial setting is the following. At an election of a parliament of house size h, ℓ political parties finish with vote shares $w_1, \ldots, w_\ell$. The electoral law stipulates an apportionment rule allotting the h parliamentary seats to the ℓ parties by way of seat contingents $x_1, \ldots, x_\ell$. The apportionment rule should be such that the seat contingent x_j of party j is close to what the party would claim under pure proportionality, $x_j \approx w_j h$. Alas, since seats are assigned to human beings who are indivisible, the seat contingents x_j must be natural numbers and cannot in general become equal to the fractional shares $w_j h$.

The number of parties contesting an election usually varies from one election to the other. Hence an apportionment rule is suitable for an electoral law only if its formulation does not involve the size of the party system, ℓ. To this end a "vote vector" $(v_1, v_2, \ldots)$ is taken to be an infinite sequence of nonnegative numbers that breaks off after a last nonzero term v_ℓ and ends in a tail of zeros. Similarly, a "seat vector" $(x_1, x_2, \ldots)$ is taken to be a sequence of natural numbers terminating with zeros. This convention allows an effectual introduction of apportionment rules.

By definition, an "apportionment rule" A maps every house size h and every vote vector $v = (v_1, v_2, \ldots)$ into a non-empty "solution set" $A(h; v)$ consisting of seat vectors $x = (x_1, x_2, \ldots)$ that have a component sum equal to the house size h and that inherit all zeros of the vote vector v: $v_j = 0 \Rightarrow x_j = 0$ for all j.

The notion of an "allotment method" in Hylland (1978, page 5) is quite similar.

An apportionment rule as defined above is taken to be a set-valued mapping in order that it may accommodate tied situations. A prototype tie arises when three seats are apportioned among two equally strong parties, with 5000 votes each say. Either the first party is allotted one seat and the second two, or the first party two and the second one. As both solutions are equally justified, the solution set comprises both: $A\big(3;(5000,5000)\big) = \{(1,2),(2,1)\}$. See Table 3 below for another example.

Note that vote vectors and seat vectors, which a minute ago were agreed to be infinite sequences that terminate with zeros, are jotted down as vectors of finite length simply by omitting the vanishing tails.

Contemplation of which apportionment rules are practically reasonable or not leads to a subclass of procedures called apportionment methods.

3. Apportionment Methods

The abstract notion of apportionment rules embraces procedures obviously unfit for concrete usage. For instance, whatever the vote vector v, all seats could be allocated to the party listed first, $A(h;v) = \{(h,0,0,\ldots)\}$, a dictatorial rule. The ensemble of all apportionment rules is reduced to a reasonable subset by imposing a set of desirable principles.

There are five basic principles. The first four are anonymity, balancedness, concordance, and decency. They suggest themselves as soon as they are formulated. The fifth principle, exactness, has a more technical flavor.

Anonymity. An apportionment rule A is called "anonymous" when every rearrangement of a vote vector induces the same rearrangement of the accompanying seat vector. Whether a party is listed first or last does not matter, its seat contingent stays the same.

Balancedness. An apportionment rule A is called "balanced" when any two parties that are equally strong differ by at most one seat: $v_i = v_j \Rightarrow |x_i - x_j| \leq 1$. It is unrealistic to insist on equality, but a difference of two or more seats will not be tolerated.

Concordance. An apportionment rule A is called "concordant" when of any two parties the stronger party is allotted at least as many seats as the weaker party: $v_i > v_j \Rightarrow x_i \geq x_j$. A discordant result, giving the stronger party fewer seats than the weaker party, is rejected.

Decency. An apportionment rule A is called "decent" when scalings of the vote vector do not change the solution set: $A\left(h;\frac{1}{d}v\right) = A(h;v)$ for all $d > 0$. Hence absolute vote counts v_j and relative vote shares $w_j = v_j/(v_1 + v_2 + \cdots)$ yield the same solutions.

Exactness. An apportionment rule A is called "exact" when every sequence of vote vectors $v(n)$, $n \geq 1$, that converges to a seat vector x induces sequences of solution vectors $y(n) \in A\big(h;v(n)\big)$, $n \geq 1$, that converge to x, too, provided $x_j = 0 \Rightarrow v_j(n) = 0$ for all $n \geq 1$.

Persuasive as they are the five principles suffer from a common weakness. They are insensitive to the house size h and the size of the party system ℓ. They solely deal with variations in the vote vector $v_1, \ldots, v_\ell$. Anonymity permutes its components, balancedness and concordance compare them by pairs, decency rescales them, and exactness addresses the case when the vote vector coincides with a seat vector or converges to a seat vector. Exactness links the continuum character of the input domain, the space of vote vectors, to the discrete nature of the output range, the grid of seat vectors.

An apportionment rule that is anonymous, balanced, concordant, descent and exact is called an "apportionment method". Almost all procedures that can be found in electoral laws qualify as apportionment methods.

4. Fairness or Coherence Principle

There is a sixth principle, fairness, also known as coherence. Fairness properly incorporates the two parameters missed out so far, the house size h and the size of the party system ℓ. Essentially, given a large house size H and a large party system $1, \ldots, L$ with a solution $x_1, \ldots, x_L$, every subsystem $1, \ldots, \ell$ with its induced seat total $h = x_1 + \cdots + x_\ell$ admits the subvector $x_1, \ldots, x_\ell$ as a solution.

Fairness implements the idea that the whole and its parts must fit together in a coherent way. Balinski and Young (2001, page 141) put it this way: "An inherent principle of any fair division is that *every part of a fair division should be fair*."

Fairness. An apportionment method A is called "fair", or "coherent", when it satisfies (a) coherence of subproblems and (b) coherence of substituted solutions.

(a) "Coherence of subproblems" means that, given a grand seat vector $(x_1, \ldots, x_L) \in A(H; v_1, \ldots, v_L)$ for a system of L parties, the subvector $(x_1, \ldots, x_\ell)$ is a member of the ℓ-subsystem solution set $A(h; v_1, \ldots, v_\ell)$, where $h = x_1 + \cdots + x_\ell$ and $\ell < L$.

(b) "Coherence of substituted solutions" means that, given a grand seat vector $x = (x_1, \ldots, x_L) \in A(H; v_1, \ldots, v_L)$ and an ℓ-subsystem seat vector $y = (y_1, \ldots, y_\ell) \in A(h; v_1, \ldots, v_\ell)$, substitution of y into x yields a grand solution $(y_1, \ldots, y_\ell, x_{\ell+1}, \ldots, x_L) \in A(H; v_1, \ldots, v_L)$, where $h = x_1 + \cdots + x_\ell$ and $\ell < L$.

Coherence of subproblems (a) is a top-down concept. It demands that every subvector that is extracted from a grand solution is a valid solution for the associated subproblem. Coherence of substituted solutions (b) is a bottom-up idea. Tied solutions for subproblems, when substituted into the grand solution, yield tied grand solutions.

The above formalization of fairness makes sense only in the presence of anonymity. With anonymity, the order in which parties are listed is negligible. The system may be rearranged so that the ℓ-subsystem parties are assembled in the initial section $1, \ldots, \ell$. For this reason the notion of fairness asks for apportionment methods, not just for apportionment rules.

5. Divisor Methods

My first highlight of apportionment theory is the Coherence Theorem. It states that the six principles characterize an important class of apportionment rules, divisor methods. Divisor methods scale the votes v_j into interim quotients v_j/d and then round the quotients to a neighboring integer to obtain the seat numbers x_j. The methods differ by which rounding rule they apply. In turn, the applicable rounding rule determines which divisors d are feasible to exhaust the preordained house size, $x_1 + \cdots + x_\ell = h$.

Generally, a rounding rule maps an interim quotient v_j/d that lies in the integer interval $[n-1; n]$ to one of the endpoints. To this end the interval is equipped with a "signpost" $s(n)$. Below $s(n)$, the quotient is rounded downwards to the singleton $\{n-1\}$, above, upwards to the singleton $\{n\}$. If the quotient is equal to $s(n)$, it is rounded ambiguously to the two-element set $\{n-1, n\}$. The ambiguous rounding at the signpost proper turns a "rounding rule" R into a set-valued mapping:

$$R\left(\frac{v_j}{d}\right) = \begin{cases} \{n\} & \text{in case } \dfrac{v_j}{d} \in \big(s(n); s(n+1)\big), \\ \{n-1, n\} & \text{in case } \dfrac{v_j}{d} = s(n) > 0, \\ \{0\} & \text{in case } \dfrac{v_j}{d} = 0. \end{cases}$$

Hence a rounding rule R is specified by its signposts. A general "signpost sequence" $s(0), s(1), s(2), \ldots$ needs to fulfill three properties. (a) It starts with $s(0) = 0$. (b) For $n \geq 1$ the term $s(n)$ is localized in the integer interval $[n-1; n]$. (c) If there is a signpost hitting the left limit of its localization interval, $s(m+1) = m$, then all signposts stay below their right limits, $s(n) < n$, and if there is a signpost hitting the right limit, $s(m+1) = m+1$, then all signposts stay above their left limits, $s(n) > n-1$. The "left-right disjunction" (c) becomes instrumental when verifying exactness of the accompanying divisor method.

By definition, the "divisor method D with rounding rule R" maps a house size h and a vote vector $v_1, v_2, \ldots$ into the set of seat vectors $x = (x_1, x_2, \ldots)$ given by

$$D(h; v) = \left\{ x \middle| x_1 \in R\left(\frac{v_1}{d}\right), x_2 \in R\left(\frac{v_2}{d}\right), \ldots \text{ for some } d > 0, \text{ and } x_1 + x_2 + \cdots = h \right\}.$$

That is, the seat contingent x_j of party j is obtained by scaling its vote count v_j by a divisor d and rounding the interim quotient v_j/d to an adjacent natural number x_j.

The role of the divisor d is to ensure that all h seats are meted out. If d is too small then the interim quotients v_j/d are too large for their roundings to sum to h. If d is too large then the quotients are too small. Thus the divisor acts as a "sliding controller" which is adjusted until the desired total is met, $x_1 + x_2 + \cdots = h$.

Set-valued mappings again emerge due to the handling of ties. Suppose there are two parties whose interim quotients hit the $(m+1)$-st and n-th signposts, $v_1/d = s(m+1)$ and $v_2/d = s(n)$, whence $R(v_1/d) = \{m, m+1\}$ and $R(v_2/d) = \{n-1, n\}$. If the parties' fair apportionment is $m+n$ seats, the first quotient may be rounded downwards and the second upwards, or vice versa. It is not up to mathematics to select one of the two options. The decision is left open by offering two solutions, $D(m+n; v_1, v_2) = \{(m, n), (m+1, n-1)\}$.

Coherence Theorem. *An apportionment rule A is anonymous, balanced, concordant, decent, exact and fair if and only if A is "compatible" with a divisor method D, in the sense that the inclusion $A(h; v) \subseteq D(h; v)$ holds true for all house sizes h and for all vote vectors v.*

The significance of the Coherence Theorem is that it provides helpful practical guidance. If we agree that the six principles are conditions *sine qua non*, there is no need to look outside the class of divisor methods.

Compatibility of a method A with a method D implies that they agree whenever the solution set $D(h; v)$ is a singleton. For, if $D(h; v) = \{x\}$ then $\emptyset \neq A(h; v) \subseteq D(h; v)$ forces $A(h; v) = \{x\}$.

Yet, in the presence of ties, A may differ from D. Then the solution set of D contains two or more seat vectors. In fact, a divisor method D is "complete" in the sense that it enumerates all tied solutions possible. However, a fair apportionment method A may abstain from completeness by implementing a tie resolution strategy.

For example, the electoral law for the Spanish Congreso de los Diputados resolves ties by following the motto "Stronger Parties First". If there are two parties whose interim quotients hit signposts, then the party with more votes is rounded upwards and the party with fewer votes is rounded downwards. Completeness is lost, yet the six principles persist.

The direct part of the proof of the Coherence Theorem is challenging. Starting from an apportionment rule A that satisfies the five basic principles and fairness, a signpost sequence needs to be constructed so that the induced divisor method D is such that A is compatible with D. Conversely, it is easy to verify that every divisor method satisfies the five basic principles and fairness. For details Balinski and Young (2001, page 141) or Pukelsheim (2017, page 168).

6. Stationary Divisor Methods

The multitude of divisor methods still is huge. There are as many divisor methods as there are rounding rules, and there are as many rounding rules as there are signpost sequences. Within this universe there is a one-parameter family, stationary divisor methods, lining up the three apportionment methods that in many respects serve as reference procedures: the divisor method with upward rounding, the divisor method with standard rounding, and the divisor method with downward rounding.

Stationary divisor methods are indexed by a "split" parameter $0 \leq r \leq 1$. The stationary divisor method with split r has signposts $s_r(n) = n-1+r$. As a consequence the interval $[n-1; n]$ is split into the section $[n-1; n-1+r]$ where numbers are rounded downwards to $n-1$, and the section $[n-1+r; n]$ where the rounding is upwards to n. The proper split point $n-1+r$ may be rounded either way. The methods are termed "stationary" because of the stationary position of the signposts in their localization intervals. Whatever the interval, the distance from the signpost to the left endpoint is r, to the right endpoint, $1-r$.

Three members of the family of stationary divisor methods stand out to be of particular importance.

The "divisor method with upward rounding" comes with split $r = 0$. If an interim quotient v_j/d has a nonzero fractional part, it is rounded upwards. If the quotient happens to be a whole number, it may stay as is or it may be rounded to the whole number above.

The "divisor method with standard rounding" belongs to split $r = 1/2$. An interim quotient v_j/d is rounded downwards or upwards according as its fractional part is less than one half or greater than one half. If the quotient happens to have a fractional part equal to one half, it may be rounded either way, downwards or upwards.

The "divisor method with downward rounding" has split $r = 1$. If a quotient v_j/d has a nonzero fractional part, it is truncated to its integer part. If the quotient happens to be a whole number, it stays as is or is rounded to the whole number below.

My second highlight of apportionment theory is the Seat Bias Theorem. Parliaments typically are hesitant to amend an apportionment method once it has found its way into the electoral law. When a method is used repeatedly at several elections, the question arises whether it predictably benefits some participants and disadvantages others. More pointedly, does a method on average favor stronger parties at the expense of weaker parties?

To answer this question we rearrange parties by decreasing vote shares. Some electoral laws stipulate a threshold t lest a party should be dropped from consideration. For example Germany requires at least five percent of all valid votes for a party to participate in the seat apportionment process. Thus parties are taken to be ordered according to $w_1 \geq \cdots \geq w_\ell \geq t$. The key figure for the k-th strongest party is $x_k - hw_k$, the deviation of the actual seat contingent x_k from the proportional seat fraction hw_k. Assuming the vote shares to be uniformly distributed over the probability simplex $\Omega_\ell = \{\, (w_1, \ldots, w_\ell) \in [0;1]^\ell \mid w_1 + \cdots + w_\ell = 1 \,\}$, the expected value of $x_k - hw_k$ for large house sizes and conditional on decreasing vote shares designates the "seat bias" of the k-th strongest party. This seat bias acquires a telling format.

Seat Bias Theorem. *If seats are apportioned using the stationary divisor method with split r and if the threshold is set at t then the seat bias of the k-th strongest party is*

$$\lim_{h\to\infty} \mathrm{E}\left(x_k - hw_k \;\middle|\; w_1 \geq \cdots \geq w_\ell \geq t \right) = \left(r - \frac{1}{2}\right)\left(H_k^\ell - 1\right)(1 - \ell t),$$

where $H_k^\ell = \sum_{n=k}^{\ell}(1/n)$ is a partial sum of the harmonic series.

The seat biases of all parties must sum to zero since $x_1+\cdots+x_\ell-h(w_1+\cdots+w_\ell) = h-h = 0$. That is, if some parties are advantaged, others are disadvantaged. Conversely, if some parties are disadvantaged then others are advantaged. One man's meat is another man's poison.

The three factors of the bias formula mirror three distinct aspects of the problem.

The "method factor" $(r - 1/2)$ reflects the influence of the stationary divisor method under investigation. The factor is positive, zero, or negative according as the split r is larger than one half, equal to one half, or smaller than one half.

The "party factor" $(H_k^\ell - 1)$ captures the impact of the party's rank-order k in a system of ℓ parties. In view of the approximation $H_k^\ell \approx \log \ell - \log k$ the factor changes sign when k passes $\ell/e \approx \ell/3$. Hence the party factor is positive for the top third of stronger parties, and negative for the bottom two thirds of weaker parties.

The "threshold factor" $(1 - \ell t)$ describes the impact of the threshold t when ℓ parties are contesting the election. It affects the size of the bias, but not its sign. For seven parties and a five percent threshold, as in Table 1 below, the factor amounts to $0.65 \approx 2/3$.

All in all a method with split r larger than one half favors stronger parties at the expense of weaker parties. In particular the divisor method with downward rounding (which has $r = 1$) is the procedure most widespread in actual electoral laws. It is also known under the names of D'Hondt, Hagenbach-Bischoff, Jefferson.

A method with split smaller than one half favors weaker participants at the expense of stronger participants. An example is the divisor method with upward rounding (which has $r = 0$). Occasionally the method is used to allocate seats between districts by population figures.

The divisor method with standard rounding ($r = 1/2$) has method factor zero. All seat biases are zero, every party may expect its proportional due. On average no party is advantaged, nor is any party disadvantaged. The divisor method with standard rounding is the unique stationary divisor method that is "unbiased".

The clue to the proof of the Seat Bias Theorem is the identity $x_k - hw_k = (r - 1/2)(\ell w_k - 1) + (x_k - y_k) + u_k(h)$, where $y_k \in R_r(h_r w_k)$ is an auxiliary seat contingent derived from the deterministic multiplier $h_r = h + \ell(r - 1/2)$, and where $u_k(h) = y_k - (h_r w_k - r + 1/2)$ is a rounding residual. It is easy to see that the first term yields the limit formula. The hard part is to show that the other two terms eventually average out to zero. For details see Pukelsheim (2017, page 139).

7. A Closer Look at The Assumptions

The assumptions underlying the Seat Bias Theorem raise suspicion as to its practical usefulness. Nobody would care for parliaments with "large" house sizes h near infinity. Fortunately, when the convergence behavior is scrutinized, the bias formula is seen to fit empirical data perfectly well for all practical purposes provided there are at least twice as many seats as there are parties participating, $h \geq 2\ell$.

Nor would we claim that uniformly distributed vote shares are a realistic model. Strong parties know that they are strong, and weak parties know they are weak. A distribution with pronounced peaks following opinion polls would be more meaningful. Luckily, essential parts of the proof of the Seat Bias Theorem target the rounding residuals $u_k(h)$ for which there is an invariance principle. Rounding residuals transpire to be uniformly distributed assuming no more than that the vote share distribution on the probability simplex Ω_ℓ is absolutely continuous.

In any case, confrontation of the theoretical seat bias formula with practical seat bias data confirms the formula to be an excellent and valid predictor.

Unbiasedness of the divisor method with standard rounding offers a cogent rationale that this is the superior method for use in electoral laws.

Fields other than the political sciences may aim at other features. The many facets of science afford a welcome opportunity for us to return to statistical topics such as sampling allocations and experimental designs.

8. Efficient Rounding of Sampling Allocations and Experimental Designs

The Goodness-of-Fit Theorem is my third highlight. Motivated by statistics and operations research it views the apportionment problem as an approximation task. Let f denote a goodness-of-fit criterion that assesses the quality of an approximation. Given a distribution with virtually continuous weights $w_1, \ldots, w_\ell$ summing to unity, the task is to find a distribution with distinctly discrete weights $x_1/h, \ldots, x_\ell/h$ that provides an f-optimal approximation. The optimization takes place over the set $\mathbb{N}^\ell(h)$ of integer vectors with ℓ components summing to h. Not surprisingly, the answer depends on the goodness-of-fit criterion f selected.

Goodness-of-Fit Theorem.

(a) The divisor method with standard rounding yields solutions $(x_1, \ldots, x_\ell) \in \mathrm{DivStd}(h; w_1, \ldots, w_\ell)$ *that minimize the squared statistical distance criterion*

$$f_a(x_1, \ldots, x_\ell) = \frac{(x_1 - hw_1)^2}{hw_1} + \cdots + \frac{(x_\ell - hw_\ell)^2}{hw_\ell}.$$

(b) The divisor method with downward rounding yields solutions $(x_1, \ldots, x_\ell) \in \mathrm{DivDwn}(h; w_1, \ldots, w_\ell)$ *that minimize the worst-overrepresentation criterion*

$$f_b(x_1, \ldots, x_\ell) = \max\left\{\frac{x_1}{hw_1}, \ldots, \frac{x_\ell}{hw_\ell}\right\}.$$

(c) The divisor method with upward rounding yields solutions $(x_1, \ldots, x_\ell) \in \mathrm{DivUpw}(h; w_1, \ldots, w_\ell)$ *that maximize the worst-underrepresentation criterion*

$$f_c(x_1, \ldots, x_\ell) = \min\left\{\frac{x_1}{hw_1}, \ldots, \frac{x_\ell}{hw_\ell}\right\}.$$

The proof of the theorem is straightforward, see Pukelsheim (2017, page 185). We add a few comments for each of the three parts. The examples in Tables 1–3 are evaluated with the free Java program Bazi (www.th-rosenheim.de/bazi).

In part (a) the criterion f_a resembles the familiar χ^2-statistic. However, the limiting distribution of f_a is a Lévy-stable distribution, not a χ^2-distribution, see Heinrich et al. (2004). Nevertheless, the criterion is in excellent harmony with the constitutional imperative that all voters should contribute equally to the electoral outcome, see Pukelsheim (2017, page 186). Hence the divisor method with standard rounding is the authoritative and unbiased procedure for the apportionment of seats among parties by vote counts. It meets the ideal of "One Person, One Vote" in a superb manner.

Since 2008 the divisor method with standard rounding has been included in the election law for the German Bundestag. See Table 1 for an illustration.

Table 1: Divisor method with standard rounding. Apportionment of 709 seats, election to the 19th German Bundestag, 24 September 2017.

Political Party	Votes	Interim Quotient	Seats [DivStd]
"CDU"	12 447 656	199.8	200
"SPD"	9 539 381	153.1	153
"AfD"	5 878 115	94.4	94
"FDP"	4 999 449	80.2	80
"LINKE"	4 297 270	69.0	69
"GRÜNE"	4 158 400	66.7	67
"CSU"	2 869 688	46.1	46
Sum (Divisor)	44 189 959	(62 300)	709

In part (b) the criterion f_b pops up when allocating observations in stratified sampling schemes, as discussed by Pukelsheim (1997). The reciprocal of the criterion provides a lower bound for the variance efficiency. Maximizing the lower bound is equivalent to minimizing the criterion f_b. Hence given a target sample size n and strata weights $w_1, \ldots, w_\ell$, the number of observations per stratum is determined most efficiently by using the divisor method with downward rounding, $(n_1, \ldots, n_\ell) \in \mathrm{DivDwn}(n; w_1, \ldots w_\ell)$. See Table 2.

Table 2: Divisor method with downward rounding. Efficient proportional sampling plan for 30 observations, see Example 9.5 in Hedayat and Sinha (1991, page 272).

Stratum	Size	Interim Quotient	No. of Obs. [DivDwn]
"Stratum 1"	60	9.5	9
"Stratum 2"	90	14.3	14
"Stratum 3"	50	7.9	7
Sum (Divisor)	200	(6.3)	30

The divisor method with downward rounding is biased in favor of large weights at the expense of small weights, as noticed above. In sampling schemes, the weight w_j relates to the standard deviation in stratum j. The bias behavior means that overly many observations are allocated to strata where the variance is large and uncertainty is high. This rule of conduct appears to be purposive for the allocation of observations.

In part (c) the criterion f_c arises in the optimality theory of experimental designs, see Pukelsheim (2006, page 311). Here, $w_1, \ldots, w_\ell$ are the weights of an optimal design with ℓ support points. If for a given sample size n the weights are discretized into frequencies $n_1, \ldots, n_\ell$, then the smallest term of the likelihood ratios $(n_j/n)/w_j$ turns out to be a universal efficiency bound. Universality means that this lower bound is meaningful for all optimality criteria that are of interest in this context (*i.e.* for all information functions). The best lower bound is the one that is largest. According to part (c) it is obtained using the divisor method with upward rounding.

Hence the divisor method with upward rounding is the recommended procedure to convert an optimum design into an efficient exact design for sample size n, $(n_1, \ldots, n_\ell) \in \mathrm{DivUpw}(n; w_1, \ldots w_\ell)$. See Table 3.

Table 3: Divisor method with upward rounding. Two equally justified (*i.e.* "tied") efficient exact designs #1 and #2 for 9 observations that belong to the A-optimal design for cubic regression on $[-1; 1]$, see Pukelsheim (2006, page 224).

A-Optimal Support Point	A-Optimal Weight	Interim Quotient	No. of Obs. [DivUpw]	
			#1	#2
"-1"	0.151	1	1	2
"-0.464"	0.349	2.3	3	3
"0.464"	0.349	2.3	3	3
"1"	0.151	1	2	1
Sum (Divisor)	1	(0.151)	9	9

According to the previous section the divisor method with upward rounding is biased in favor of small weights at the expense of large weights. The consequence is that small weights are likely to be allocated more observations than pure proportionality would demand. In particular, even the tiniest weight is rounded upwards to at least one observation. Hence the discretization process preserves all support points of the optimal design. Thus the divisor method with upward rounding appears to be a purposive rule of conduct for the discretization of optimal designs.

9. Epilogue

In conclusion we realize that although the five apportionment principles and the notion of fairness were introduced with narratives from the proportional representation world, these concepts make perfectly good sense also in the contexts of sampling schemes and experimental designs.

This proves once again that problems that seemingly are far apart actually share common theoretical underpinnings, just like scientists who live far apart — like Bikas in Asia, Bimal in North America, and myself in Europe — stay united by standards common to all fields of science.

Acknowledgements

The author gratefully acknowledges valuable remarks of the referee and the Editor.

References

Balinski, M. L. and Young, H. P. (2001). *Fair Representation – Meeting the Ideal of One Man, One Vote. Second Edition.* Washington DC: Brookings Institution.

Hedayat, A. S. and Sinha, B. K. (1991). *Design and Inference in Finite Population Sampling.* New York: Wiley.

Heinrich, L., Pukelsheim, F. and Schwingenschlögl, U. (2004). Sainte-Laguë's chi-square divergence for the rounding of probabilities and its convergence to a stable law. *Statistics & Decisions*, **22**, 43–59.

Hylland, A. (1978). *Allotment Methods: Procedures for Proportional Distribution of Indivisible Entities.* Mimeographed Report, John F. Kennedy School of Government, Harvard University. [Reprinted as Working Paper 1990/11, Research Centre, Norwegian School of Management. Sandvika, 1990.]

Pukelsheim, F. (1997). Efficient rounding of sampling allocations. *Statistics & Probability Letters*, **35**, 141–143.

Pukelsheim, F. (2006). *Optimal Design of Experiments. SIAM Classics in Applied Mathematics Volume 50.* Philadelphia PA: SIAM.

Pukelsheim, F. (2017). *Proportional Representation. Apportionment Methods and Their Applications. With a Foreword by Andrew Duff MEP. Second Edition.* Cham CH: Springer.

Pukelsheim, F. and Sinha, B. K. (1995). Optimal block designs revisited – An approximate theory detour. *Statistics and Probability Letters*, **24**, 375–380.

Statistics and Applications {ISSN 2452-7395 (online)}
Volume 18, No. 2, 2020 (New Series), pp 155-162

Identification of Multiple Unusual Observations in Spatial Regression

A.H.M. Rahmatullah Imon[1] and Ali S Hadi[2]
[1]*Department of Mathematical Sciences, Ball State University, USA*
[2]*Department of Mathematics and Actuarial Science*, *The American University in Cairo,Egypt*

Received: 24 February 2020; Revised: 11 June 2020; Accepted: 10 July 2020

Abstract

Traditional outlier detection methods cannot be directly applied to spatial data because of its global nature. Spatial outlier detection methods concentrate on discovering neighborhood instabilities (Shekhar *et al*. 2002). However, most of the traditional detection methods may not accurately locate outliers when multiple outliers exist. Robust spatial *z* test proposed by Hadi and Imon (2018) has largely resolved this issue. But lots of unresolved issues exist in spatial regression where likewise linear or generalized linear models, the entire inferential procedure is generally affected in the presence of unusual observations called outliers (*y*-outliers) and high leverage points (*x*-outliers) or both. A large body of literature are available now for the identification of unusual observations in linear and/or generalized linear regression but this is still an unexplored area in spatial regression. In this paper we propose a new method for the identification of multiple spatial outliers and spatial high leverage points based on robust and clustering algorithms. We also propose a very simple but attractive graphical display to locate these two types of outliers in the same graph.

Keywords: Spatial outlier; Differencing; Masking; High leverage points; Clustering; GP-GSR plot.

1. Introduction

Conceptually spatial outliers are very different from classical outliers. A commonly used definition is that outliers are a minority of observations in a dataset that have different patterns from that of the majority of observations in the dataset. The assumption here is that there is a core of at least 50% of observations in a dataset that are homogeneous (that is, represented by a common pattern) and the remaining observations (hopefully few) have patterns that are inconsistent with this common pattern. Spatial outliers are those observations whose characteristics are markedly different from their spatial neighbors. The identification of spatial outliers is important because it can reveal hidden but valuable knowledge in many applications such as identifying aberrant genes or tumor cells, discovering highway traffic congestion points, locating extreme meteorological events such as tornadoes, and hurricanes *etc.*

Corresponding Author: A.H.M.R. Imon
Email: rimon@bsu.edu

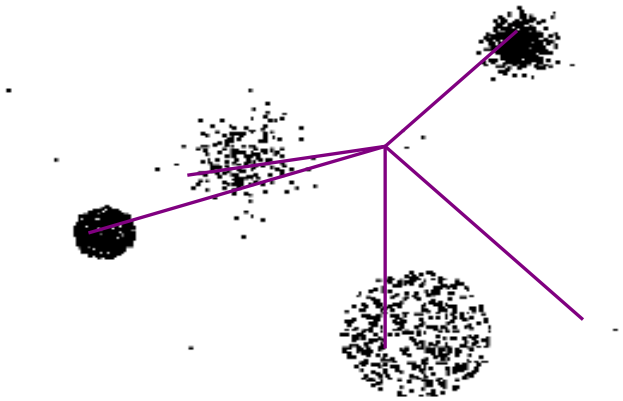


Figure 1: Outliers in data clusters

Although outliers could be easily identified in univariate, bivariate, or even trivariate data through graphical examination of the data, visual inspection does not usually work for more than three dimensions. Not only that automated identification of outliers is tricky even for a two dimensional data if the data form clusters as shown in Figure 1. Here the idea of majority minority simply does not work, bad clusters are identified as outliers (Hadi *et al*., 2009) based on classification techniques. Things could even be cumbersome in regression models where outliers can occur along the *y*-dimension, or along the *x*-dimension, or both and/or among the relationship between *x* and *y*. An excellent review of different aspects of spatial outliers is available in Shekhar *et al*. (2002) and Hadi and Imon (2018). Conceptually, spatial outliers match with outliers in big data and for this reason outlier detection techniques designed for big data are often routinely employed in spatial data. In big data the concept of outlier is local, not global so as in spatial data. The distance and/or density based methods such as *k*–nearest neighbourhood, local outlier factor (LOF), spatial outlier factor (SOF) methods have become more popular. But all these methods are designed to identify outliers along the *y*-axis and hence are not readily applicable for spatial regression. For example, temperatures and amount of rainfall of different regions may vary due to their distances from sea or mountain. Once we fit this relationship by regression we may observe not only strange temperature or rainfall pattern, the distance factor may also be unusual. Attempts have been made to identify outliers based on residuals but it only focuses on the outliers in *y*, but not in *x* or both and the whole concept is rather global than local. To overcome this problem in this paper we propose a method which not only focuses on both *x* and *y* dimensions at the same time, but also considers classification techniques to identify outliers.

2. Methodology

Let us assume that we have *n* pairs of spatial observations (u_i, v_i), $i = 1, 2, \ldots, n$. We further assume that *V* depends on *U* and we are interested to investigate their nature of relationship. In order to understand whether spatial observations are stable in their neighborhood, Shekhar *et al*. (2002) suggested considering the first order differences of the spatial observations. For both *U* and *V* we take the first order differences defined as

$$x_i = u_i - u_{i-1}, y_i = v_i - v_{i-1}; i = 2, 3, \ldots, n \quad (1)$$

Based on the differenced observations obtained in (1), let us consider a standard regression model

$$Y = X\beta + \varepsilon \quad (2)$$

where Y is a vector of observed responses of order $(n-1)$, X is an $(n-1) \times 2$ matrix of explanatory variables including the constant, β is a vector of unknown finite parameters of order 2 and ε is an n-vector of random disturbances with $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma^2 \mathrm{I}$. The traditionally used ordinary least squares (OLS) estimator of β is $\hat{\beta} = (X^T X)^{-1} X^T Y$ and the vector of fitted values is $\hat{Y} = X\hat{\beta} = HY$. The matrix

$$H = X(X^T X)^{-1} X^T \quad (3)$$

is often referred to as weight or leverage matrix whose diagonal elements h_{ii} are termed leverages. The OLS residual vector $\hat{\varepsilon}$ is defined as $\hat{\varepsilon} = Y - \hat{Y}$. Observations corresponding to exceptionally large $\hat{\varepsilon}$ values are termed outliers. However, the question still remains how large is large? For this reason we often consider the standardized version of residuals. One very popular choice is deleted Studentized residuals (DSR) defined as

$$t_i = \frac{y_i - x_i^T \hat{\beta}^{(-i)}}{\hat{\sigma}_{(i)}\sqrt{(1-h_{ii})}},\ i = 2, 3, \ldots, n \quad (4)$$

where $\widehat{\beta}^{(-i)}$ and $\hat{\sigma}_{(i)}$ are the OLS estimates of β and σ respectively with the i-th observation deleted. We call an observation outlier when its corresponding deleted Studentized residual value exceeds 3 in absolute value. Observations corresponding to exceptionally large h_{ii} values are termed high leverage points which are essentially outliers in the X-space. However, since residuals are also functions of leverages, it is better if we identify both outliers and high leverage points simultaneously rather than separately. Gray (1986) proposed the Leverage-Residual (L-R) plot where the leverage value h_{ii} for each observation i, is plotted against the square of a normalised form of its corresponding residual. The bulk of the cases will be associated with low leverage and small residuals so that they cluster near the origin (0, 0). The unusual cases will have either high leverages or large residual components and so will tend to be separated from the bulk of the data. High leverage cases will be located in the upper area of the plot and observations with large residuals will be located in the area to the right.

The L-R plot may be effective in the identification of single outlier but it may be ineffective in the presence of multiple outliers unless we remove a group of suspect outliers prior to fitting the model. Denote a set of cases 'remaining' in the analysis by R and a set of cases 'deleted' by D. Also suppose that R contains $(n-1-d)$ cases after $d \leq (n-1-k)$ cases in D are deleted. Without loss of generality, assume that these observations are the last d rows of X and Y so that we can partition the matrices as

$$X = \begin{bmatrix} X_R \\ X_D \end{bmatrix},\quad Y = \begin{bmatrix} Y_R \\ Y_D \end{bmatrix},\ H = \begin{bmatrix} H_R & H_{RD} \\ H_{DR} & H_D \end{bmatrix} \quad (5)$$

where $H_R = X_R(X^T X)^{-1} X_R^T$ and $H_D = X_D(X^T X)^{-1} X_D^T$ are symmetric matrices of order $(n-1-d)$ and d respectively, and $H_{RD} = X_R(X^T X)^{-1} X_D^T$ is an $(n-1-d) \times d$ matrix. However, $(X_R^T X_R)^{-1}$ can be expressed as

$$(X_R^T X_R)^{-1} = (X^T X - X_D^T X_D)^{-1} = (X^T X)^{-1} + (X^T X)^{-1} X_D^T (I_D - U_D)^{-1} (X^T X)^{-1} \quad (6)$$

where I_D is an identity matrix of order d and $U_D = X_D (X_D^T X_D)^{-1} X_D^T$. Using (6), Imon (2002) defined a group deleted version of high leverage points called generalized potentials defined as

$$p_{ii}^* = \begin{cases} \frac{h_{ii}^{(-D)}}{1 - h_{ii}^{(-D)}} & i \in R \\ h_{ii}^{(-D)} & i \in D \end{cases} \quad (7)$$

where $h_{ii}^{(-D)} = x_i^T (X_R^T X_R)^{-1} x_i$, i= 2, 3, …, n. In other words, $h_{ii}^{(-D)}$ is the i-th diagonal element of $X(X_R^T X_R)^{-1} X^T$ matrix. The vector of estimated parameters after the deletion of d observations, denoted by $\hat{\beta}^{(-D)}$, is obtained using (6) as

$$\hat{\beta}^{(-D)} = (X_R^T X_R)^{-1} X_R^T Y_R = \hat{\beta} - (X^T X)^{-1} X_D^T (I_D - U_D)^{-1} \hat{\varepsilon}_D \quad (8)$$

where $\hat{\varepsilon}_D = X_D \hat{\beta}$. Using (6), (7) and (8), Imon (2005) introduced a group deleted version of residuals called generalized Studentized residuals (GSR) defined as

$$t_{ii}^* = \begin{cases} \frac{y_i - \hat{y}_i^{(-D)}}{\hat{\sigma}^{(-D)} \sqrt{1 - h_{ii}^{(-D)}}} & i \in R \\ \frac{y_i - \hat{y}_i^{(-D)}}{\hat{\sigma}^{(-D)} \sqrt{1 - h_{ii}^{(-D)}}} & i \in D \end{cases} \quad (9)$$

where $\hat{y}_i^{(-D)} = x_i^T \hat{\beta}^{(-D)}$ and $\hat{\sigma}^{(-D)}$ are the fitted values of y and the scale parameter σ respectively after the omission of the suspected outlier group indexed by D. Although the expression of generalized potentials is available for any arbitrary set of deleted cases, D, the choice of such a set is clearly important since the omission of this group determines the weights for the whole set. We call an observation outlier when its corresponding generalized Studentized residual value exceeds 3 in absolute value. No such value exists for generalized potentials. We follow Hadi (1992) to declare an observation as a high leverage point if its corresponding p_{ii}^* exceeds a threshold given as

$$p_{ii}^* > \text{Median}\,(p_{ii}^*) + 3\text{MAD}\,(p_{ii}^*). \quad (10)$$

where MAD stands for the median absolute deviation.

These above results enable us to define a simple graphical display of classifying group deleted leverages and residuals for possible identification of them. Generalized potentials are used as leverages and the generalized Studentized residuals as deletion residuals in a 'generalized potentials –generalized Studentized residuals (GP-GSR)' plot. Since the high leverage points need not to be outliers and outliers may not be points of high leverage we may expect different deletion sets D from the computation of these two quantities. Since D is the group of suspected outliers we prefer to include all observations considered to be suspect either along the y dimension or along the x dimension. We employ the blocked adaptive computationally-efficient outlier nominators (BACON) proposed by Billor *et al.* (2000) as a classifier. Another possibility could be the application of support vector regression for the same, especially when the data is big. The main advantage of the GP-GSR plot is that it is suitable for the data where masking (false negative)

and/or swamping (false positive) make single case diagnostic plots misleading. This plot, unlike the L-R plot retains the signs of residuals, which can be very important when their interpretation is concerned. Since the bulk of the cases will be associated with low leverage and small residuals, most of the pairs (t_{ii}^*, p_{ii}^*) will cluster near the origin (0, 0). The unusual cases will have either high leverages or large residual components and will tend to be separated from the bulk of the cases. High leverage cases will be located at the right corner of the plot and observations with large residuals will be located either at the upper or lower corner of the plot depending on their signs; large positive outliers will be plotted at the upper corner and large negative outliers will be located at the bottom corner of the plot.

3. Results

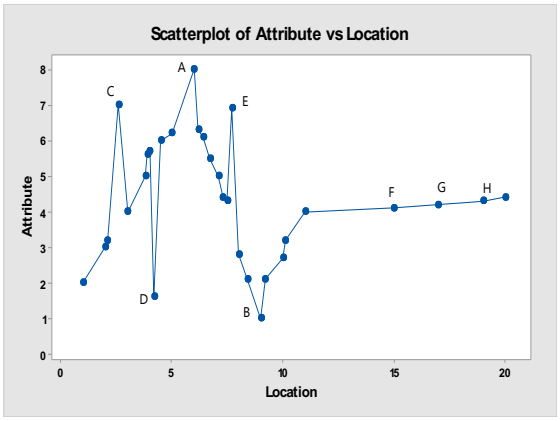
In this section we would like to present an example to demonstrate how our proposed method works in the classification of spatial regression outliers in both *x* and *y* dimensions. Here we consider a spatial outlier data given by Hadi and Imon (2018) extending the idea of Shekhar *et al*. (2002). Although this data is artificial in nature, the use of this type of data is very common in the outlier detection literature (Rousseeuw and Leroy, 1987; Hadi *et al*., 2009) because here we definitely know which observations are genuine outliers. For real data with multiple outliers due to masking and swamping there could be always lots of disagreements regarding which observations are genuine outliers or not. We present the data in Table 1 and also in Figure 2.

Table 1: Hadi and Imon (2018) spatial outlier data

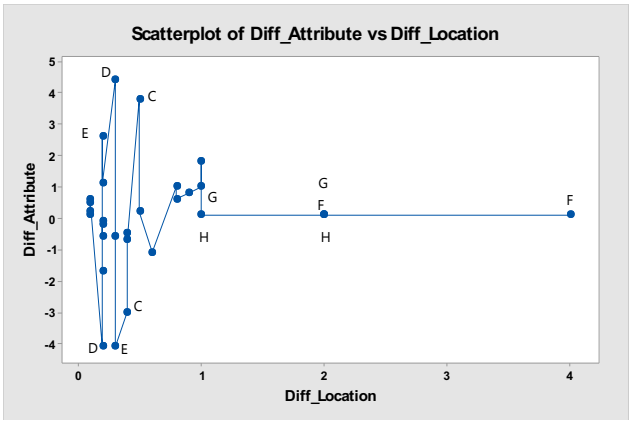
Index	Location	Attribute	Diff_Location	Diff_Attribute
1	1.0	2.0	*	*
2	2.0	3.0	1.0	1.0
3	2.1	3.2	0.1	0.2
4	2.6	**7.0 C**	0.5	**3.8 C**
5	3.0	4.0	0.4	**–3.0 C**
6	3.8	5.0	0.8	1.0
7	3.9	5.6	0.1	0.6
8	4.0	5.7	0.1	0.1
9	4.2	**1.6 D**	0.2	**–4.1 D**
10	4.5	6.0	0.3	**4.4 D**
11	5.0	6.2	0.5	0.2
12	6.0	8.0 A	1.0	1.8
13	6.2	6.3	0.2	–1.7
14	6.4	6.1	0.2	–0.2
15	6.7	5.5	0.3	–0.6
16	7.1	5.0	0.4	–0.5
17	7.3	4.4	0.2	–0.6
18	7.5	4.3	0.2	–0.1
19	7.7	**6.9 E**	0.2	**2.6 E**
20	8.0	2.8	0.3	**–4.1 E**
21	8.4	2.1	0.4	–0.7
22	9.0	1.0 B	0.6	–1.1
23	9.2	2.1	0.2	1.1
24	10.0	2.7	0.8	0.6
25	10.1	3.2	0.1	0.5

26	11.0	4.0	0.9	0.8
27	**15.0 F**	4.1	**4.0 F**	0.1
28	17.0	4.2	2.0	0.1
29	19.0	4.3	2.0	0.1
30	20.0	4.4	1.0	0.1

This example gives a clear distinction between classical outlier and spatial outlier. In Figure 2(a) attribute values are plotted against their locations. For global outliers, traditional statistics will essentially look at the attribute values in the *y* axis and if we do that we observe that the points which are very high such as A or very low such as B. In contrast to that, the spatial outliers are like the spikes C, D and E. They look like spatial outliers because they violate the law of geography that the nearby things should be very similar. When we take the first order difference of the attributes as shown in Figure 2(b) clearly C, D and E look very different than their neighbors. It is also interesting to note that the possible global outliers A and B do not look like outliers anymore. In general, we do not search for outliers along the *x*-axis. But when we carefully look at Figure 2(a), we observe that the point F has a marked difference from its neighbors. Points G and H look unusual too. This difference is visible more clearly when we look at the first order difference of the locations as shown in Figure 2(b). Point F now clearly looks like a high leverage point or an outlier along the *x*-space. Points G and H look more extreme as well.



2(a). The original data



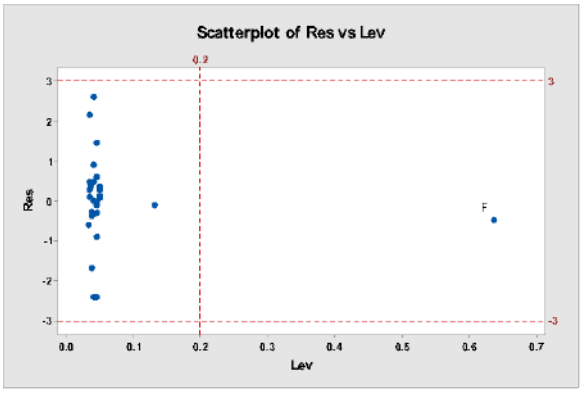
2(b) The first order differenced data

Figure 2: Scatter plot of the original and the first order differenced data

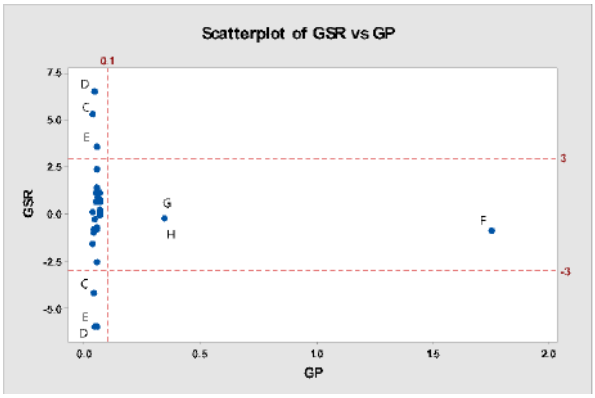
Now we run a spatial regression of attributes on locations. Since our interest is to understand the neighbourhood instability we consider the first order difference of attributes and locations as given in columns 4 and 5 of Table 1. We then run a regression of differences in attributes on differences in location and the resulting deleted Studentized residuals and leverages are given in columns 2 and 3 of Table 2. Although DSR is very popular outlier measure it fails to identify even a single observation as an outlier. Here the cut-off for the leverage is 0.2 and it can identify F as a high leverage point. We see exactly the same picture in the L-R plot as shown in Fig 3(a). Now we compute GSR and GP and the results are presented in columns 4 and 5 of Table 2. We use BACON classifier to obtain the D set first and then compute GSR and GP as outlined in equations (7) and (8). It is worth mentioning that the cut-off value for GP is 0.1 based on equation (10). We also present the GP-GSR plot for this data in Figure 3(b). These results clearly show the merit of our proposed method. It can successfully identify 3 spatial outliers (C, D and E) and 3 spatial high leverage points(F, G, H).

Table 2: Residuals and leverages for the spatial outlier data

Index	Del St. Residual	Leverage	GSR	GP
1	*	*	*	*
2	0.45678	0.040885	1.09925	0.06658
3	0.11424	0.051079	0.20420	0.06290
4	2.15139	0.035779	**5.31765 C**	0.03590
5	–1.69711	0.037989	**–4.20445 C**	0.03835
6	0.47421	0.035612	1.13209	0.04571
7	0.32834	0.051079	0.72348	0.06290
8	0.06085	0.051079	0.07645	0.06290
9	–2.41622	0.045639	**–6.03394 D**	0.05185
10	2.61223	0.041275	**6.49375 D**	0.04367
11	0.07639	0.035779	0.07645	0.03590
12	0.89298	0.040885	0.13783	0.06658
13	–0.92108	0.045639	2.34566	0.05185
14	–0.10826	0.045639	–2.56908	0.05185
15	–0.33030	0.041275	–0.32208	0.04367
16	–0.28573	0.037989	–0.85865	0.03835
17	–0.32172	0.045639	–0.74085	0.05185
18	–0.05503	0.045639	–0.84428	0.05185
19	1.43538	0.045639	**3.58453 E**	0.05185
20	–2.42202	0.041275	**–6.02091 E**	0.04367
21	–0.39244	0.037989	–1.00843	0.03835
22	–0.62533	0.034647	–1.61273	0.03631
23	0.58737	0.045639	1.39415	0.05185
24	0.26077	0.035612	0.59882	0.04571
25	0.27470	0.051079	0.59176	0.06290
26	0.35834	0.037710	0.84471	0.05471
27	–0.49040	**0.636897 F**	–0.91806	**1.75404 F**
28	–0.12121	0.131865	–0.24012	**0.34271 G**
29	–0.12121	0.131865	–0.24012	**0.34271 H**
30	–0.02276	0.040885	–0.06854	0.06658



3(a). L-R plot



3(b). GP-GSR plot

Figure 3: Diagnostic plots for the spatial regression data

4. Discussion and Conclusion

The main objective of our research was to develop a method for the joint identification of outliers and high leverage points for spatial regression. In Section 2 we develop a new method to identify both of them and propose a new graphical display called GP-GSR plot to locate both of them in the same graph. In spatial statistics literature observations with neighbourhood instability are diagnosed as outliers. For this reason we employ our method on the first order difference of x and y. A numerical example clearly shows the advantage of using our proposed method. It clearly shows that the proposed method can successfully identify outliers and high leverage points simultaneously while the existing methods fail to do so.

Acknowledgements

The authors express their thanks and gratitude to the anonymous reviewers for giving some useful suggestions that led to considerable improvement in the methodology and presentation of the results. The authors also express their thanks to the Chief Editor for his suggestions on restructuring the contents for better expression.

References

Billor, N., Hadi, A. S., and Velleman, P. F. (2000). BACON: Blocked adaptive computationally efficient outlier nominators.*Computational Statistics and Data Analysis*, **34**, 279-298.

Gray, J. B. (1984). A simple graphic for assessing influence in regression. *Journal of Statistical Computation and Simulation*, **24**, 121-134.

Hadi, A. S. (1992). A new measure of overall potential influence in linear regression. *Computational Statistics and Data Analysis*, **14**, 1-27.

Hadi, A. S. and Imon, A. H. M. R. (2018). Identification of multiple outliers in spatial data. *International Journal of Statistical Science*, **16**, 87-96.

Hadi, A. S., Imon, A. H. M. R. and Werner, M. (2009). Detection of outliers. *Wiley Interdisciplinary Reviews: Computational Statistics*, **1**, 57–70.

Imon, A. H. M. R**.** (2002). Identifying multiple high leverage points in linear regression. *Journal of Statistical Studies*, **22**, 207–218.

Imon, A. H. M. R. (2005). Identifying multiple influential observations in linear regression. *Journal of Applied Statistics*, **32**, 929–946.

Rousseeuw, P. J. and Leroy, A. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.

Shekhar, S., Lu, C., and Zhang, P. (2002). Detecting graph-based spatial outlier. *Intelligent Data Analysis*, **6**, 451-468.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 163–185

Quantifying Spatial and Temporal Relationships Among Tree-Ring Records

Megan Heyman[1], Scott St George[2] and Snigdhansu Chatterjee[3]
[1]*Department of Mathematics,*
Rose-Hulman Institute of Technology, USA
[2]*Department of Geography, Environment, and Society*
University of Minnesota, USA
[3]*School of Statistics*
University of Minnesota, USA

Received: 04 May 2020; Revised: 12 July 2020; Accepted: 13 July 2020

Abstract

Tree growth rings contain yearly information about climate, extreme weather events, and other growing conditions. In this analysis, we model the relationship strength between tree-ring records with respect to location and time. We employ the discrete wavelet transformation on the ring width records in order to de-correlate the observations within each series while simultaneously retrieving time-scale information. Our model then describes correlations among the resulting wavelet coefficients at different temporal scales by distance. Statistical inference through a new version of the wild bootstrap indicates that the relationship strength decreases linearly as record pair distance increases, but the slopes differ across temporal scales.

Key words: Wavelet; Bootstrap; U-Statistics; Tree ring records.

AMS Subject Classifications: 62G05, 62G09

1. Introduction

Because instrumental measurements of temperature, precipitation, and other aspects of Earth's climate typically span only the past century or so (Harris *et al.*, 2020), we rely on surrogate information recovered from natural climate archives to extend our perspective on recent changes. Fallen snow accumulates on the surface of continental ice sheets or alpine and arctic glaciers, and builds year-by-year a frozen archive of atmospheric chemistry that can extend back several hundreds of thousands of years (Steiger *et al.*, 2017; Brook and Buizert, 2018). Sediment that sinks down to the bottom of lakes often traps windblown pollen and the remains of waterborne organisms, which in turn reflect environmental conditions across the broader region (Holmes *et al.*, 2016; Sandeep *et al.*, 2017). And the water that flows underground to form cave deposits leaves behind physical and chemical clues within the

Corresponding Author: Snigdhansu Chatterjee
Email: chatt019@umn.edu

resulting mineral structures (Arienzo *et al.*, 2017; Affolter *et al.*, 2019). Over the past several decades, networks of proxy climate records have served as the foundation for quantitative estimates of past temperature, drought severity, and other key climate metrics at the local, regional, continental, or global scale (Trouet *et al.*, 2018; Neukom *et al.*, 2019; Cook *et al.*, 2020).

Within the so-called 'Common Era' (the past two thousand years), the growth rings of trees are, by far, the most dominant source of past climate information (Emile-Geay *et al.*, 2017). Every year, trees across the world's temperate and boreal forests form a new layer of wood about their stem. One of the most obvious signs in nature documenting the passage of time, those rings also encode information about the tree's immediate environment within their physical and chemical structure (St George and Ault, 2014). For trees growing in cold alpine or arctic forests, the width and wood density of their annual ring can be excellent surrogates for growing season temperatures (Esper *et al.*, 2018; St George and Esper, 2019). And for trees in warmer and drier environs, because narrow rings tend to follow dry weather, tree-ring records can be used to estimate past changes in precipitation or drought extending back hundreds of years or more (Granato-Souza *et al.*, 2019; Opała-Owczarek and Niedźwiedź, 2019).

Because tree rings and other proxies can extend our perspective on climate change farther back in time than instrumental weather observations, they offer new opportunities to evaluate the time evolution of the dominant modes of climate variability. Towards that purpose, the wavelet transform is commonly applied to study multiscale, nonstationary processes occurring across space and time within tree-ring records (Fan and Bräuning, 2017; McKenzie *et al.*, 2018; Kasatkina *et al.*, 2019).

Climate is a multi-scale phenomenon. Some quasi-periodic and oscillatory patterns like the El Niño Southern Oscillation (ENSO) are sub-decadal in periodicity, while others like the Atlantic multi-decadal oscillation (AMO) or the Pacific decadal oscillation (PDO) have longer time scales. In this paper, our primary goal is to verify that at different temporal scales, the correlation between tree-ring records decay smoothly over space in a climatically homogeneous region. To that end, in this paper we explore the spatio-temporal patterns of paleoclimate records, as exhibited by Ponderosa pine (*Pinus ponderosa* Douglas ex C. Lawson) tree-ring records from four states of the USA. We restrict to only this subspecies of trees in order to eliminate differences due to species variation, and restrict regionally so that we may consider a homogeneous, contiguous region where the climatic patterns are similar. Additional details are provided later in this paper.

Our principal approach is to consider each three ring record as an individual functional observation over time, that exhibits quasi-periodic and oscillatory patterns according to the climatic conditions the tree has experienced in the past. A discrete wavelet transformation of these records allows us to deconstruct such functional time series into various temporal scales. We then consider correlations between the wavelet coefficients from two different trees, at different scales, and study the pattern of such correlations as a function of distance between the two trees.

Notice that the data that we analyze here has complex dependency patterns, hence

it is non-trivial to conduct inference on a statistical model for how correlations, across various wavelet scales, decays over spatial distance. To address this issue, we propose a novel resampling scheme, that generalizes the well-known *wild or external* bootstrap scheme.

The rest of the paper is as follows: In Section 2 we provide a detailed description of the dataset, and then Section 3 provides an overview of the discrete wavelet transform. Next, we present a new approach to describe spatial relationships among time-series records in Section 4. Namely, we apply the discrete wavelet transformation on tree-ring records and calculate three versions of correlation between pairs of wavelet coefficients within each time-scale. Afterwards, it is necessary to re-format the data for analysis; this is described in Section 5. With the re-formatted data, we describe a novel approach to modeling relationship strength across temporal scales in Section 6. The proposed linear model utilizes time-scale, distance, and elevation information among the records to predict strength of relationship across all record pairs. We also describe differences among obtaining estimates utilizing least squares, median regression, and Huber's M estimation techniques. Finally, statistical inference is performed via an extension of the wild bootstrap in Section 7. The wild bootstrap is modified to utilize external random variables which are generated per tree-ring record, instead of for each case within the dataset. The random variables are incorporated together within the method to help account for dependence among cases utilizing the same record(s) while simultaneously producing the sampling distributions for the coefficient estimates. An illustrative theoretical insight into the properties of the proposed extension of the wild bootstrap is presented in Section 9. We collect some concluding remarks and ideas about future research in Section 10.

2. Tree-Ring Record Description

Tree-ring datasets are typically presented in the structure of a 'chronology' — a composite series made by averaging together measurements of tree-ring width, wood density, isotopic composition, or other anatomical or biogeochemical variables from several dozen or more trees at the same location (Cook and Peters, 1997). Compared to the initial measurements made on samples taken from individual trees, tree-ring chronologies offer several advantages (Coulthard *et al.*, 2020). They are easier than sample- or tree-level data to incorporate into regional syntheses, they have been pre-processed to remove the confounding influence of tree size or age (Bunn *et al.*, 2004), and they are often adjusted to minimize or eliminate biologically-driven persistence (Kannenberg *et al.*, 2019).

The International Tree-Ring Databank (ITRDB), an open-access database maintained by the National Oceanic and Atmospheric Administration in Boulder, Colorado, is the largest archive of freely-available tree-ring data worldwide (Zhao *et al.*, 2019). The ITRDB, which was established in 1974 as a permanent repository for digital tree-ring measurements, includes more than four thousand tree-ring datasets (each composed of measurements made on one dozen to more than one hundred trees) from all continents except Antarctica. Here we use a subset of the St George and Ault (2014) dataset, which used the ITRDB's holdings as the foundation for their network of more than 2,200 age-corrected, quality-controlled tree-ring width chronologies from the Northern Hemisphere. Each of these time series, which have annual resolution but varying start and end dates, reports yearly tree growth across

an entire forest stand as a unitless index of tree-ring width (RWI). Our analysis focuses on tree-ring width chronologies derived from Ponderosa pine (*Pinus ponderosa* Douglas ex C.Lawson) forests in the Four Corners region of the southwestern United States (Figure 1). Ponderosa pine (coded as PIPO) is the fourth most common source species for tree-ring width measurements housed by the ITRDB (St George and Ault, 2014) and a range that extends from northern Mexico to southern Canada. Across the United States, Ponderosa pine grow best under high precipitation and low evapotranspirative demand, but throughout its range there are important regional differences in the species' sensitivity to climate (McCullough *et al.*, 2017). For that reason, we restricted our analysis to only include those tree-ring chronologies from Arizona, New Mexico, Colorado, and Utah (yielding a total of 97 series).

A commonly referenced set of guidelines for investigators creating tree-ring records may be found in Cook (1987). Generally, at a given location, multiple trees are chosen to be measured. Within each tree, between 1 and 5 cores (cylinders of heartwood extracted from the trunk of the tree) are obtained in such a way that the tree is not vitally harmed. The entire set of growth rings, combining both early and late wood, is measured within each core. Then, a B-spline is used to remove a growth trend from each individual core. Finally, all of the de-trended core series are combined into a single ring-width index (RWI) for the location. Thus, the records of tree-ring widths are *unitless* and composed of many trees and cores. It is also common practice to further clean records before analysis, by only maintaining observations with strong coherence between cores within the record location at any given time point.

3. Discrete Wavelet Transform

In paleoclimatology and climate science in general, nearly all applications of wavelets are based on the continuous transform (Lau and Weng, 1995; Torrence and Compo, 1998). That method does not place any restriction on series length, but in principle should be applied only to data where the observations are made continuously through time. However, tree-ring records are discrete time series, with observations that are equally spaced and made at specific times (either once or twice within the growing season). For that reason, in this paper, we apply the discrete wavelet transform to a network of tree-ring width chronologies to test the potential of this other method to problems in climate science and paleoclimatology.

Our model aims to quantify spatial and temporal relationships *among* RWI records. However, we cannot ignore the temporal dependencies likely present *within* each record. We choose to employ wavelets to address the *within record* temporal dependency. Wavelets are an ideal choice since they overcome the limitations with respect to the time and frequency domains that are present in Fourier and windowed Fourier analyses. Let $\mathbb{Z}$ be the set of integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Suppose $\psi \in \mathcal{L}_2(\mathbb{R})$ is a given function whose properties will be discussed later. Wavelets constitute the family of functions defined by

$$\psi_{j,k}(u) = 2^{j/2}\psi(2^j u - k), \quad j, k \in \mathbb{Z}. \tag{1}$$

For appropriate choices of $\psi(\cdot)$, these translations and dyadic dilations of $\psi(\cdot)$ constitute an orthonormal basis of $\mathcal{L}_2(\mathbb{R})$. The dilation and shrinkage within the function allow a user to

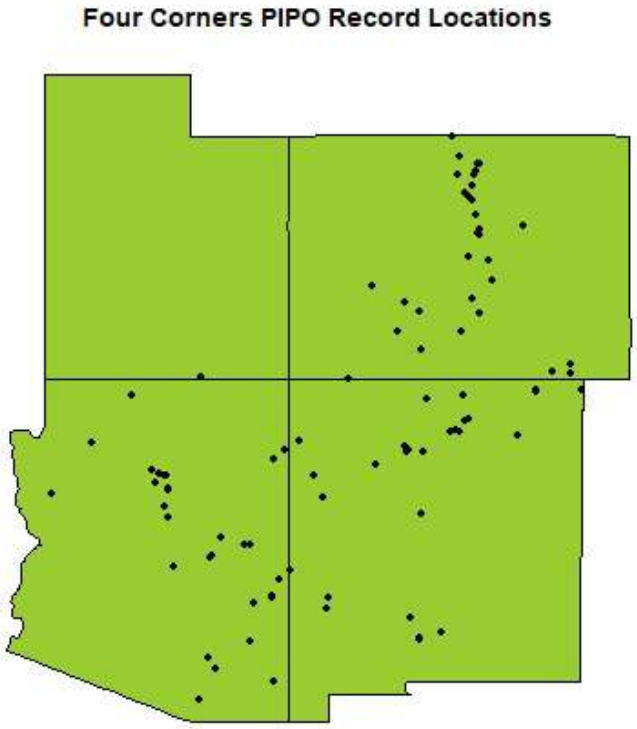


Figure 1: Ponderosa pine record locations. Locations of the 97 Ponderosa pine records in the Four Corners states analyzed by our model.

discover signals which change over the time domain. In many situations, a second function, denoted as $\phi(\cdot)$ and also called the *scaling function* since it may be derived from $\psi(\cdot)$, is used in conjunction with subsets of the family of wavelet functions $\{\psi_{j,k}(\cdot)\}$ for representing functions in a relatively simpler and parsimonious way. The very basic overview of wavelets provided here is necessarily brief; more comprehensive discussions and many additional points of interest may be found in Daubechies (1992); Hubbard (1998); Ogden (1997); Vidakovic (1999); Wasserman (2006); Percival and Walden (2006); Nason (2008) and several other places.

In our case, we are working with discrete time points and are specifically choosing to employ the discrete wavelet transform (DWT). The DWT is a transformation of a vector of data, $\boldsymbol{x}$, which utilizes an orthonormal basis. Assume that the data vector, $\boldsymbol{x} = (x_1, \ldots, x_T)$, is of length $T = 2^J$ where $J \in \mathbb{Z}^+ = \{1, 2, \ldots\}$. Consider the collection of functions

$$\left\{\phi(\cdot), \psi_{j,k}(\cdot);\ j = 0, 1, \ldots, J-1,\ k = 1, 2, \ldots, 2^j\right\}.$$

We evaluate these $T = 2^J$ functions at each of $t = 1, 2, \ldots, T$, and construct the matrix $\boldsymbol{W}$. The first row of $\boldsymbol{W}$ is $(\phi(1), \phi(2), \ldots, \phi(T))$. Then, the remainder of the matrix is generated from the bottom-up. That is, the T^{th} (last) row is given by $(\psi_{0,1}(1), \psi_{0,1}(2), \ldots, \psi_{0,1}(T))$.

Rows $T-2$ and $T-1$ are $(\psi_{1,1}(1), \psi_{1,1}(2), \ldots, \psi_{1,1}(T))$ and $(\psi_{2,1}(1), \psi_{2,1}(2), \ldots, \psi_{2,1}(T))$, respectively. Generally, sets of 2^j rows are generated together for evaluations of the wavelet functions at time-scales, $j = 0, 1, \ldots, J-1$. A concrete example demonstrating this construction is provided in Section 3.1. With some amount of algebra and exploiting known properties of wavelets, for appropriate choice of the function $\psi(\cdot)$ it can be seen that the matrix $\boldsymbol{W}$ is an orthonormal basis system of $\mathbb{R}^T$, that further has all the multi-resolution properties for wavelets as defined in (1).

Based on this, the DWT of the data vector, $\boldsymbol{x}$ is

$$\boldsymbol{\gamma} = \boldsymbol{W}\boldsymbol{x} \tag{2}$$

where the vector, $\boldsymbol{\gamma}$, contains the wavelet coefficients. One important item of note is that the wavelet coefficients convey coarse and fine level information about the series. Specifically, $\boldsymbol{\gamma}$ may be written element-wise as

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_0 & \gamma_{J-1,1} & \cdots & \gamma_{J-1,2^{J-1}} & \cdots & \gamma_{1,1} & \gamma_{1,2} & \gamma_{0,1} \end{pmatrix}^T. \tag{3}$$

Here, γ_0 is the coarse level coefficient corresponding to the scaling function $\phi(\cdot)$, and $\gamma_{j,k}$ is the fine level coefficient corresponding to the wavelet function $\psi_{j,k}(\cdot)$, for $\{(j,k) : \ j = 0, 1, \ldots, J-1, \ k = 1, 2, \ldots, 2^j\}$. A basic example of the DWT is provided in Section 3.1 for readers who are less familiar with the technique.

The DWT has an additional advantage for use in large datasets. The discrete wavelet transformation requires less stringent technical assumptions compared to Fourier transformation since the amplitude corresponding to each frequency does not have to be stable over time. Computationally, the DWT is faster than the fast Fourier transform (see Percival and Walden (2006) for details). As we motivate in Section 6, our model describes correlations among the RWI records by incorporating spatial distance and temporal information. In order to create the correlations between records, we need to begin by de-correlating the individual series. The DWT tends to produce independent (or slightly correlated, Vannucci and Corradi (1999)) coefficients from observed data series which are strongly correlated (see Chapter 10 in Percival and Walden (2006)).

The above framework is based on $T = 2^J$, that is, the length of the observed data is a power of 2. There are several existing techniques to overcome this limitation, see for example Gong *et al.* (2018); Walden and Cristan (1998); Ogden (1997); Nason (2008), however, we choose not to use such advanced techniques here for clarity and tractability of the results.

3.1. Basic example of the DWT

Here we provide a basic example to demonstrate the time-scale information in wavelet coefficients that we have discussed and utilize in our analysis. Consider the most basic wavelet basis, proposed by Haar. The Haar wavelet basis appears as a step function (see Figure 2).

At the coarsest level, the step function spans across the entire vector of data. At finer levels, the step function is shrunk and translated across the data vector. Let's consider the

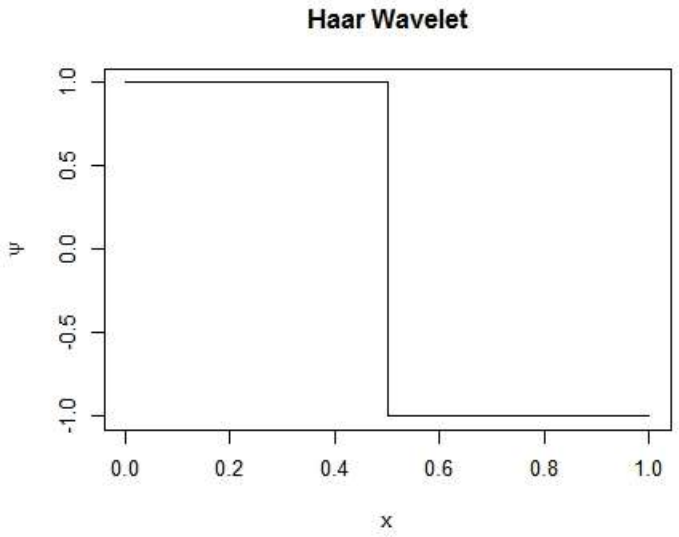


Figure 2: Haar wavelet (on a continuous scale)

the Haar basis when n=8, as generated using the `GenW()` within the `wavethresh` **R** package by Nason (2016).

$$\boldsymbol{W} = \begin{bmatrix} 2^{-3/2} & 2^{-3/2} & 2^{-3/2} & 2^{-3/2} & 2^{-3/2} & 2^{-3/2} & 2^{-3/2} & 2^{-3/2} \\ 2^{-1/2} & -2^{-1/2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2^{-1/2} & -2^{-1/2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2^{-1/2} & -2^{-1/2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2^{-1/2} & -2^{-1/2} \\ 2^{-1} & 2^{-1} & -2^{-1} & -2^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2^{-1} & 2^{-1} & -2^{-1} & -2^{-1} \\ 2^{-3/2} & 2^{-3/2} & 2^{-3/2} & 2^{-3/2} & -2^{-3/2} & -2^{-3/2} & -2^{-3/2} & -2^{-3/2} \end{bmatrix} \tag{4}$$

Since the vector of wavelet coefficients is found using $\boldsymbol{W}\boldsymbol{x}$, we can see that the scaling coefficient, γ_0 is a weighted sum of the vector entries. The coarsest wavelet coefficient, $\gamma_{0,1}$ is a difference in weighted sums between the first and second halves of the vector entries. The coarsest wavelet coefficients capture long-term signals in the data. The finest level coefficients, $\gamma_{2,1}$, $\gamma_{2,2}$, $\gamma_{2,3}$, $\gamma_{2,4}$, are weighted pairwise differences between entries which are next to each other. These capture signal changes that are close in time.

It is clear with the specific choice of Haar basis that $\boldsymbol{W}$ is an orthonormal matrix. Most generally, orthonormality is a wavelet basis property. Although it is easier to visualize and discuss, the Haar basis has an obvious discrete nature which creates non-smooth estimates that are difficult to interpret in practice. Other wavelet bases typically do not have closed-form solutions, but are more often used in statistical analysis. These include Daubechies (1992) and Chui and Wang (1992). With the guidance of several existing works that provide recommendations for selecting wavelet families, including Mandal *et al.* (1996),

Strickland and Hahn (1996), and Mojsilovic *et al.* (2000), the analysis we present here utilizes Daubechies Least Asymmetric family with 8 vanishing moments. Several software packages enable automatic implementation wavelet bases for analysis including `wavethresh` by Nason (2008), `wavelets` by Aldrich (2020), `waveslim` by Whitcher (2020) in **R** and the Wavelet Toolbox in Matlab (The MathWorks, 2020).

4. Strength of Relationship Between RWI Records

Once we have attempted to account for the strong dependence *within* the RWI record by implementing the DWT, we consider modeling the relationships among records. Without assuming a specific type of relationship (for example linear) between two quantitative variables, there are several ways to compute correlation. We considered three correlation measures to describe the RWI data.

4.1. Correlation measures

Generally, correlation measures the strength of relationship between two quantitative variables. Suppose that $\boldsymbol{x} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^T$ and $\boldsymbol{y} = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix}^T$ are two vectors containing real-valued observations. Pearson's correlation (eqn. 5) provides a measure of the linear association between the vectors.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{5}$$

Although commonly utilized, Pearson's correlation is limited to quantifying the linear relationship. Spearman's rank correlation (6) provides a bit more flexibility, by describing the strength of monotonic relationship between two quantitative variables. Spearman's correlation is created by utilizing the *rank* transformation on the observed vectors. In this transformation, the observations within each vector are ordered from smallest to largest, and thus, changed to values within $\mathbb{Z}^+$ where 1 corresponds to the smallest observation within each vector. This transformation doesn't depend on linearity and produces measures less affected by outliers.

$$r_s = \frac{\sum_{i=1}^{n}(rank(x_i) - \overline{rank(x)})(rank(y_i) - \overline{rank(y)})}{\sqrt{\sum_{i=1}^{n}(rank(x_i) - \overline{rank(x)})^2}\sqrt{\sum_{i=1}^{n}(rank(y_i) - \overline{rank(y)})^2}} \tag{6}$$

Kendall's tau correlation (7) also provides a measure of the strength of monotonic relationship between two quantitative variables. It specifically examines the discordance between pairs of observations, as opposed to the orderings among the individual observations as in Spearman's correlation. Discordance is quantified by obtaining the sign, positive or negative, of the difference between observations through the *sgn* transformation. Kendall's tau correlation has a bit of an advantage to Spearman's rank correlation with respect to

interpretability since describing discordant pairs is more straightforward than the sums of ranks.

$$\tau = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} sgn(x_i - x_j)sgn(y_i - y_j)}{n(n-1)} \tag{7}$$

4.2. Creating pairwise correlations of wavelet coefficients

Our model describes correlations between the $\binom{97}{2}$ pairs of PIPO records within the Four Corners states. The correlations are between the series' wavelet coefficients, not the raw observations. This choice allows us to model the relationship strength of relationship with respect to spatial location, while also accounting for temporal separation. Below is an outline of the algorithm utilized to compute the correlations.

1. For the records in a given pair of locations, keep the years when both records have observations.
2. Retain the most recent T observations for each of the series obtained in (1). We require $T = 2^J$ for $J \in \mathbb{Z}^+$.
3. Perform the DWT on each series obtained in (2) and obtain the resulting wavelet coefficient vectors.
4. Within the wavelet coefficients, calculate the time-scale correlation for any levels with 8 or more coefficients. For example, a correlation will be computed for the 2^{J-1} finest scale coefficients. This would correspond to the same year time-scale relationship.

With the Four Corners data, we had enough observations to calculate up to 16-year correlation information. Each of Pearson, Spearman, and Kendall correlations were calculated and considered as candidate responses for the model. The Daubechies least asymmetric wavelet with 8 vanishing moments (see Daubechies (1992)) was selected for the basis in all of our DWT computations.

5. Re-formatting the Data for Analysis

For each of the five available time-scales (same-, 2-, 4-, 8-, and 16-year), the algorithm outlined in subsection 4.2 created $\binom{97}{2}$ correlations. These correlations may be combined to compose vectors of the form

$$\boldsymbol{c}_j = \begin{pmatrix} c_{1,2,j} & c_{1,3,j} & \dots & c_{96,97,j} \end{pmatrix}^T \tag{8}$$

where $c_{i,k,j}$ denotes the correlation of records in the i^{th} and k^{th} locations. Time-scales, same-year (s), 2-year (2), 4-year (4), 8-year (8), 16-year (16), are denoted by the index j in eqn. 8. Many of the record pairings did not result in series long enough to compute the 8- or 16-

year correlations with our constraints. Those entries are missing. To construct the response in our model, we stack all of the time-scale correlation vectors into one vector

$$\boldsymbol{c} = \begin{pmatrix} \boldsymbol{c}_s & \boldsymbol{c}_2 & \boldsymbol{c}_4 & \boldsymbol{c}_8 & \boldsymbol{c}_{16} \end{pmatrix}^T. \tag{9}$$

We also construct the set of predictor variables in a similar manner. The structure of the wavelet coefficients lends to incorporating the time-scale information to the model. A set of indicator variables for each time scale were created with respect to the response vector

$$Y_{i,k,J_0} = \begin{cases} 1, & j = J_0 \text{ in } C_{i,k,j} \\ 0, & j \neq J_0 \text{ in } C_{i,k,j} \end{cases}. \tag{10}$$

The indicator variables for each time scale were stacked in a similar fashion as the correlations in eqn. 9 to create $\boldsymbol{y_s}$, $\boldsymbol{y_2}$, $\boldsymbol{y_4}$, $\boldsymbol{y_8}$, $\boldsymbol{y_{16}}$ vectors, all of the same length as $\boldsymbol{c}$.

In addition to the time-scale information, our model also incorporates spatial information in the form of the distance between locations in each pair. For each pair of records, Lambert's distance, Lambert (1942), in kilometers was calculated. The distance vector for all pairs may be denoted as

$$\tilde{\boldsymbol{d}} = \begin{pmatrix} d_{1,2} & d_{1,3} & \dots & d_{96,97} \end{pmatrix}^T. \tag{11}$$

Five $\tilde{\boldsymbol{d}}$ vectors are stacked as in eqn. 9 to create $\boldsymbol{d}$ of the same length as $\boldsymbol{c}$. Finally, along similar lines, the absolute difference in location elevations for each pair is computed and used to create the $\boldsymbol{h}$ predictor.

6. Model Estimation

Taking the Kendall correlation as an example, the scatterplots in Figure 5 within the appendix indicate a negative relationship between the locations' distance and the observed correlation at each time-scale. Similar results also hold true for the other types of correlations. Therefore, we propose a linear model with an interaction between time-scale and distance and a linear term in elevation to describe correlation in eqn. 12.

$$\begin{aligned} \boldsymbol{c} =& \beta_0 \boldsymbol{y}_s + \beta_1 \boldsymbol{y}_2 + \beta_3 \boldsymbol{y}_4 + \beta_4 \boldsymbol{y}_8 + \beta_5 \boldsymbol{y}_{16} + \\ & \beta_6 \boldsymbol{y} + \beta_7 \boldsymbol{y}_2 \circ \boldsymbol{d} + \beta_8 \boldsymbol{y}_4 \circ \boldsymbol{d} + \beta_9 \boldsymbol{y}_8 \circ \boldsymbol{d} + \beta_{10} \boldsymbol{y}_{16} \circ \boldsymbol{d} + \\ & \beta_{11} \boldsymbol{h} + \boldsymbol{\epsilon}. \end{aligned} \tag{12}$$

With such a large dataset, it was difficult to pinpoint the best estimation method and type of correlation. Even with the pairwise correlations plotted on a gradient scale, (as in Figure 5), it is difficult to determine to see what is happening. Thus, we also considered different estimation techniques for our model – least squares, median regression, and the Huber M estimator.

6.1. Estimation techniques

The three estimation techniques we consider for this analysis are all related through an objective function, ρ which is maximized in order to obtain the estimates. The model in eqn. 12 is a linear model. If we combined all of the predictor vectors into a matrix, $\boldsymbol{X}$, and all of the parameters into a vector, $\boldsymbol{\beta}$, then the model could be written simply as $\boldsymbol{c} = \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$.

Then, the estimates obtained through least squares, median regression, and Huber-M estimation are all solved by minimizing $\rho(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$. Specifically, the least squares estimate is obtained through minimizing the squared $\mathcal{L}_2$ norm as in eqn. 13 while the median regression estimate minimizes the $\mathcal{L}_1$ norm (eqn. 14).

$$\underset{\beta\in\mathbb{R}^{k+1}}{\arg\min}\, ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 \tag{13}$$

$$\underset{\beta\in\mathbb{R}^{k+1}}{\arg\min}\, ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_1 \tag{14}$$

Huber-M estimate (eqn. 15) utilizes a combination of the $\mathcal{L}_1$ and $\mathcal{L}_2$ norms in the minimization problem, dependant on the size of the element within $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$. The $\boldsymbol{I}_k$ within eqn. 15) is an indicator vector, the elements of which take a value of 1 when the magnitude of $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ is no larger than a specified k.

$$\underset{\beta\in\mathbb{R}^{k+1}}{\arg\min}\left[\left(0.5||(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\circ\boldsymbol{I}_k||_2^2\right) + \left(k||(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\circ(\boldsymbol{1} - \boldsymbol{I}_k)||_1 + k^2/2(\boldsymbol{1} - \boldsymbol{I}_k)\right)\right] \tag{15}$$

6.2. Resulting estimates

Tables 1, 2, and 3 contain the estimates for the model proposed in eqn. 12 for each of the correlations and estimation methods. Within each correlation type, estimation method does not seem to make much difference in the obtained values. We do see differences amongst the point estimates across correlation type, but this was not surprising from our data exploration. Generally, we see that each correlation and method contain the same general patterns across the coefficients.

7. Statistical Inference

Continuing forward, the results examined utilize median regression to predict Kendall correlation. In this section, we specifically compare the slopes in record pair distance across time-scales. Figure 3 contains the estimated relationships while the absolute difference in elevation is held constant at 0 km.

Although the linear model seems to be a straight-forward way to describe the RWI correlations, the data structure is actually quite complicated. Each correlation, $c_{i,k,j}$, is based upon the i^{th} and k^{th} records. Thus, each record appears within 5×96 observations within the response vector. There is not an intuitive argument to claim that the correlations

	Pearson			Spearman			Kendall		
	L.S.	Med.	Hub.	L.S.	Med.	Hub.	L.S.	Med.	Hub.
$\hat{\beta}_0$	6.597	6.806	6.663	6.431	6.630	6.480	4.614	**4.695**	4.623
$\hat{\beta}_0 + \hat{\beta}_1$	6.002	6.347	6.144	5.843	6.101	5.973	4.249	**4.297**	4.295
$\hat{\beta}_0 + \hat{\beta}_2$	5.953	6.493	6.295	5.805	6.360	6.113	4.257	**4.538**	4.424
$\hat{\beta}_0 + \hat{\beta}_3$	6.372	7.217	6.940	5.979	6.776	6.557	4.555	**4.926**	4.855
$\hat{\beta}_0 + \hat{\beta}_4$	6.511	7.120	6.930	5.744	6.352	6.074	4.497	**4.716**	4.663

Table 1: Estimated intercepts. Each estimate is multiplied by 1×10^{-1} and bold estimates are associated with the model in which we perform statistical inference. L.S=least squares regression, Med.=Median regression, Hub.=Huber M estimator regression.

	Pearson			Spearman			Kendall		
	L.S.	Med.	Hub.	L.S.	Med.	Hub.	L.S.	Med.	Hub.
$\hat{\beta}_6$	6.23	6.41	6.28	6.15	6.33	6.18	4.48	**4.56**	4.48
$\hat{\beta}_6 + \hat{\beta}_7$	4.70	4.97	4.79	4.63	4.74	4.73	3.46	**3.41**	3.48
$\hat{\beta}_6 + \hat{\beta}_8$	4.36	4.88	4.68	4.48	5.10	4.78	3.32	**3.71**	3.50
$\hat{\beta}_6 + \hat{\beta}_9$	5.77	6.53	6.14	5.30	5.88	5.69	4.10	**4.35**	4.31
$\hat{\beta}_6 + \hat{\beta}_{10}$	3.00	3.35	3.28	2.43	2.86	2.58	2.04	**2.22**	2.16

Table 2: Estimated slopes in distance. Each estimate is multiplied by -1×10^{-4} and bold estimates are associated with the model in which we perform statistical inference. L.S=least squares regression, Med.=Median regression, Hub.=Huber M estimator regression.

are independent. For example, a certain record could happen to be less correlated with all of the other records. Perhaps at that location, an external event caused all of the tree species to have stunted growth.

There are a few existing methods to deal with dependence due to repeated measures. A common choice is to add an individual effect to the model. That is, add a set of indicator variables associated with record ID as fixed or random effects to the model fit. In our analysis, this translates to adding 96 indicator variables to the model. Although possible with the number of cases in our data, the addition of ID is difficult to interpret and our structure is more complicated, since each response is associated with two record IDs. Moreover, treating ID as a random quantity would create a mixed effects model. With few distributional assumptions placed on the model error structure, inference would still be an open area of research.

Our analysis actually continues with the model from eqn. 12 and does not overcomplicate it by adding an explicit record ID variable. Although the fits obtained should be unbiased for a correctly specified model, utilizing the asymptotic normal distribution to estimate parameter error is incorrect with dependent data. To estimate coefficient error, we will

	Pearson			Spearman			Kendall		
	L.S.	Med.	Hub.	L.S.	Med.	Hub.	L.S.	Med.	Hub.
$\hat{\beta}_{11}$	8.461	8.781	8.825	8.566	8.867	8.871	6.598	**6.358**	6.610

Table 3: Estimated slope in Elevation Difference. Each estimate is multiplied by -1×10^{-2} and the bold estimate is associated with the model in which we perform statistical inference. L.S=least squares regression, Med.=Median regression, Hub.=Huber M estimator regression.

instead use a version of the wild bootstrap, originally proposed by Wu (1986), which utilizes random variables indexed by the record ID. We describe how a wild bootstrap response, $\boldsymbol{c_b}$, is obtained in our framework in the subsequent algorithm.

1. Generate n=number of records, independent random variables, each with mean 0 and variance 1. In our analysis, this translates to generating 97 random variables, $U_1, U_2, ..., U_{97}$. Suppose that the index of these random variables corresponds to a location ID.

2. Create two vectors, each of the same length as the response, $\boldsymbol{c}$, containing the random variables generated in (1). The elements of the first vector correspond to the first location (i index in the $\boldsymbol{c}$). The elements of the second vector correspond to the second location (k index in the $\boldsymbol{c}$). Call these vectors $\boldsymbol{u1}$ and $\boldsymbol{u2}$.

3. Create the wild bootstrap response, $\boldsymbol{c}_b$, as $\boldsymbol{c}_b = \hat{\boldsymbol{c}} + \text{diag}(\boldsymbol{u1})\text{diag}(\boldsymbol{u2})\boldsymbol{r}$.

In order to decide whether there was evidence that the linear relationship with record pair distance differed by time-scale, we implemented the described wild bootstrap to obtain bootstrap distributions of all model coefficients estimated in eqn. 12. These bootstrap sampling distributions are displayed in figure 6 within the appendix. Each distribution is centered at the original coefficient estimate (as expected), and all seem to be fairly light-tailed. Specifically, we explored whether there was evidence of a difference in the slopes in distance with respect to the same-year time-scale. With $B = 10,000$ bootstrap samples and Bonferroni adjusted significance level of $\alpha = 0.05$, confidence intervals were created for each of the parameters: $\beta_7, \beta_8, \beta_9$, and β_{10}. These intervals are displayed in figure 4. The black bands in each figure are associated with the bootstrap confidence interval. We can see that all of the bootstrap confidence intervals are shorter than the typical normal intervals, which assume independence across observations. These intervals also indicate evidence that the slopes in distance, across different time-scales, differ from the same-year relationship.

8. Results

The results in Figure 4 imply that the linear relationship between Kendall correlation and record pair distance differs over time. Examination of the scatterplots within Figure 5, indicate a lack of data in the 8- and 16-year time scales, compared to the other scales. Our model inference finds significant relationships at these scales, even though there was little

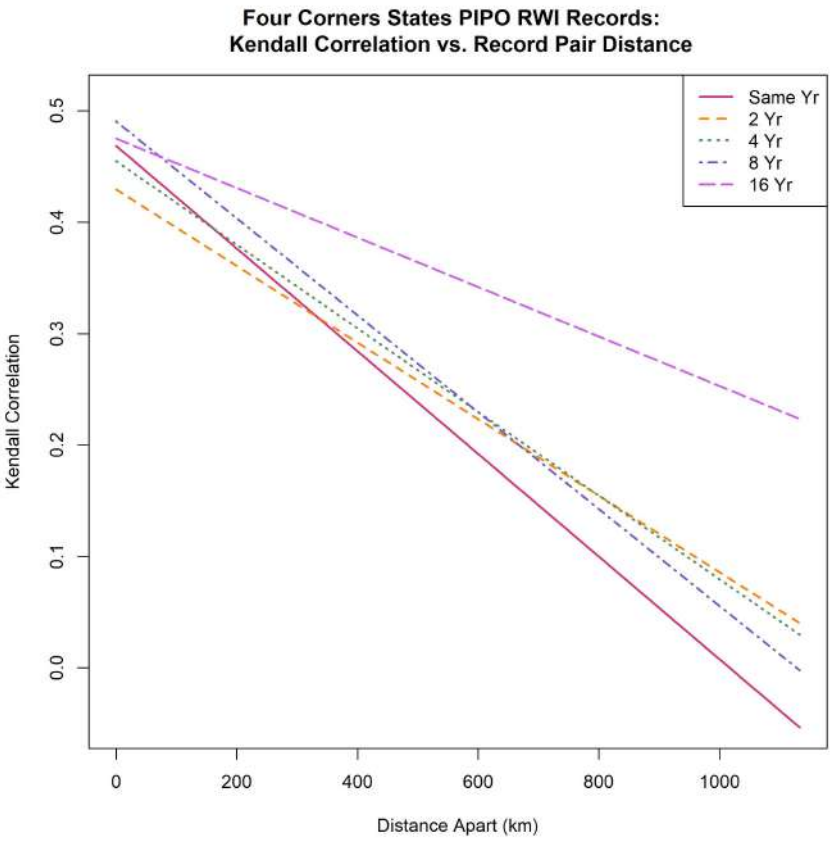


Figure 3: Fitted lines with Kendall correlation as the response and median regression

evidence of a relationship from the exploratory scatterplots. This finding could be due to a lack of power at the higher time-scales with less data. Interpreting the results at these time-scales should be done with caution.

This analysis may have significant implications for researchers examining tree-ring chronologies. First and foremost, this study implies that these *Pinus ponderosa* records contain intricate information, not only with respect to climate, as typically studied, but also with respect to each other. The type information that these records carry also varies over time. The specific analysis presented here was for a subset of the *Pinus ponderosa* in the U.S., and the maximum record length in the set was 551 years, while the shortest record only contained 63 years of RWIs. If older tree-ring records were available, we would be able to expand the analysis further, and draw more concrete conclusions about the longer time-scales. Specifically, if the researcher is interested in studying a 2^k time-scale in the RWI records, then a minimum of 2^{3+k+1} record length in the majority of records would be required. But, as seen in our analysis, this minimum requirement creates an unwieldy correlation estimate at the longest allowable time-scale, so the researcher would likely prefer to have records of length 2^{5+k+1} to seriously study the k^{th} time-scale.

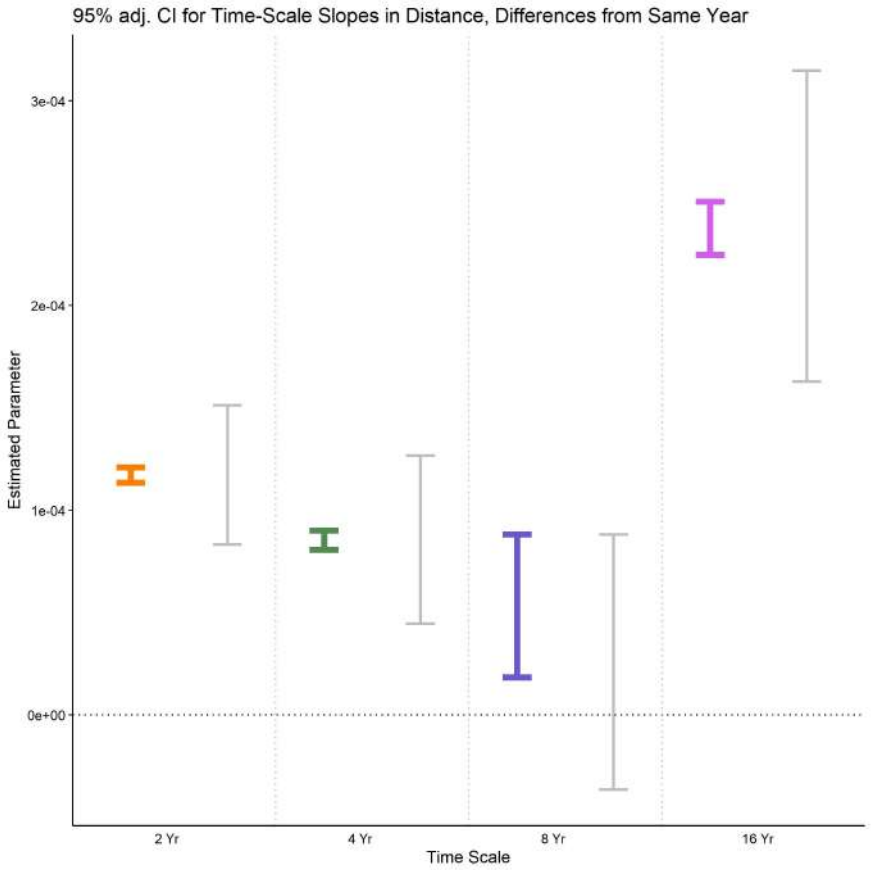


Figure 4: Bonferroni adjusted 95% confidence intervals for the time-scale slope parameters in distance. These are *differences* from the same-year slope. The traditional normal interval appears as light gray for comparison with our bootstrap intervals, in color.

9. The Theoretical Framework

We illustrate the main ideas of the underlying theoretical framework in this section, and omit much of the technical details. Fix a pair of locations. Since the wavelet decomposition de-correlates the data, we consider independent observations $\{(Y_{i,1}, Y_{i,2}, \mathbf{x}_i) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}^p;\ i = 1, \ldots, n\}$ corresponding to this pair of locations. We assume that $\mathbb{E}Y_{i,1} = \mu_1$, $\mathbb{E}Y_{i,2} = \mu_2$, $\mathbb{V}Y_{i,1} = \sigma_1^2$, $\mathbb{V}Y_{i,2} = \sigma_2^2$ and $Cor(Y_{i,1}, Y_{i,2}) = \rho(\mathbf{x}_i)$. In other words, our data consists of bivariate response vectors $(Y_{i,1}, Y_{i,2})$ related to the pair of locations, and associated covariate $\mathbf{x}_i \in \mathbb{R}^p$ that is taken to be non-random in this paper, and we model the correlation between the two responses as a function of the covariate. The bivariate response is allowed to be heteroscedastic, and the two means and two variances may also be functions of the covariate, but that structure is not relevant for present purposes. We assume that $\mathbb{E}Y_{i,j}^{8k} < \infty$ for $j = 1, 2$ and a sufficiently high positive integer k. We also write $\rho \equiv \rho(\mathbf{x}_i)$ below, since there is no cause of confusion here.

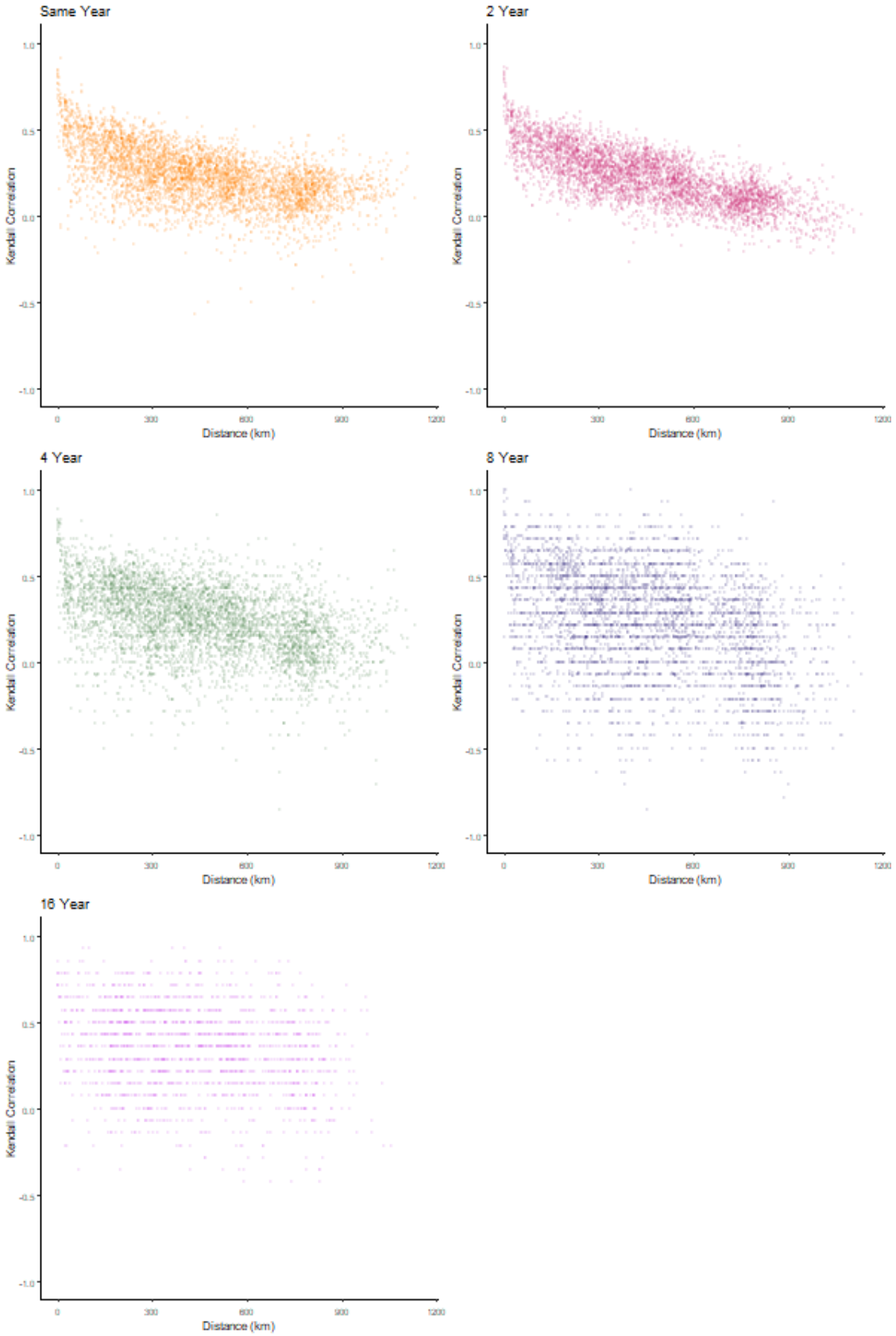


Figure 5: Pairwise Kendall correlations between wavelet coefficients at each time-scale

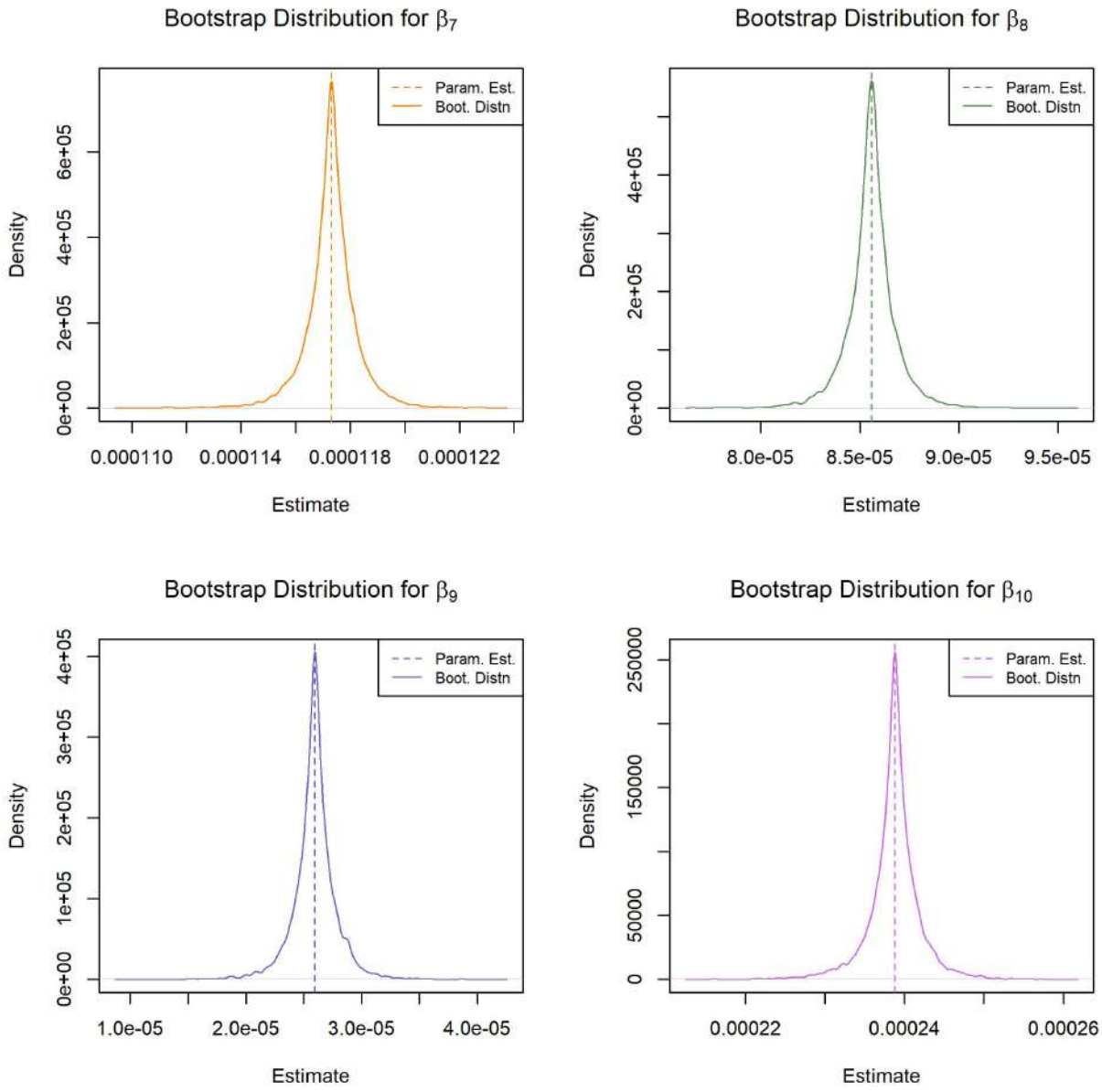


Figure 6: Bootstrap sampling distributions of time-scale slopes in distance. These are estimated *differences* from the same-year time-scale slope. The original estimate is the dashed vertical line.

Let $\epsilon \ll 0.25 \min(\sigma_1^2, \sigma_2^2)$ be a small constant, and let $\mathcal{A}$ be the event

$$\mathcal{A} = \Big\{ \sum_{i=1}^{n} \Big(Y_{i,1} - n^{-1} \sum_{i=1}^{n} Y_{i,1} \Big)^2 > \epsilon, \text{ and } \sum_{i=1}^{n} \Big(Y_{i,2} - n^{-1} \sum_{i=1}^{n} Y_{i,2} \Big)^2 > \epsilon \Big\}.$$

On $\mathcal{A}$ we define our estimator of ρ to be the usual sample correlation

$$\hat{\rho} = \left[\left\{ \sum_{i=1}^{n} \Big(Y_{i,1} - n^{-1} \sum_{i=1}^{n} Y_{i,1} \Big)^2 \right\} \left\{ \sum_{i=1}^{n} \Big(Y_{i,2} - n^{-1} \sum_{i=1}^{n} Y_{i,2} \Big)^2 \right\} \right]^{1/2} \sum_{i=1}^{n} \Big(Y_{i,1} - n^{-1} \sum_{i=1}^{n} Y_{i,1} \Big) \Big(Y_{i,2} - n^{-1} \sum_{i=1}^{n} Y_{i,2} \Big).$$

On it complement of $\mathcal{A}$, we take $\hat{\rho} = 0$. We have to separate the cases where the sample variances are sufficiently high and where they are not using $\mathcal{A}$, since we do not make any distribution assumptions like Gaussianity in this paper. It can be shown that $1 - \mathbb{P}[\mathcal{A}] = O(n^{-4k})$, consequently we only discuss $\hat{\rho}$ under $\mathcal{A}$, the other case is negligible.

Theorem 1: Under the conditions stated above, we have the following results:

(a) For every fixed pair of locations, the correlation estimator $\hat{\rho}$ is consistent and the distribution of $n^{1/2}(\hat{\rho} - \rho)$ converges weakly to a Gaussian distribution.

(b) If $\rho(\mathbf{x}) = \beta^T \mathbf{x}$ for some $\beta \in \mathbb{R}^p$, then the ordinary least squares estimator for β computed from all pairs of locations is consistent and asymptotically Gaussian.

(c) If $\rho(\mathbf{x}) = \beta^T \mathbf{x}$ for some $\beta \in \mathbb{R}^p$, and $\hat{\beta}$ is the ordinary least squares estimator for β, then the distribution of $n^{1/2}(\hat{\beta} - \beta)$ is consistently approximated by the proposed version of the wild bootstrap scheme, conditional on the data, for almost all sample paths.

Proof: [**Proof of Theorem 1**]

Here, we only provide a brief outline of the main arguments for part **(a)** of Theorem 1, in order to not overload this paper with algebraic details.

We define the following terms

$$T_1 = n^{-1} \sigma_1^{-2} \sum_{i=1}^{n} \Big(Y_{i,1} - n^{-1} \sum_{i=1}^{n} Y_{i,1} \Big)^2,$$
$$T_2 = n^{-1} \sigma_2^{-2} \sum_{i=1}^{n} \Big(Y_{i,2} - n^{-1} \sum_{i=1}^{n} Y_{i,2} \Big)^2,$$
$$T_3 = n^{-1} \sigma_1^{-1} \sigma_2^{-1} \sum_{i=1}^{n} \Big(Y_{i,1} - n^{-1} \sum_{i=1}^{n} Y_{i,1} \Big) \Big(Y_{i,2} - n^{-1} \sum_{i=1}^{n} Y_{i,2} \Big).$$

In terms of thee notations, we have under $\mathcal{A}$ that $\hat{\rho} = (T_1 T_2)^{-1/2} T_3$.

In the following, the notation R_a denotes a remainder for all choices of a, with the property that $\mathbb{E}R^{2k} < \infty$. We do not explicitly report the algebra relating to such remainder terms, those are routine. It can be worked out that on $\mathcal{A}$, we have the following

$$\begin{aligned} T_3 &= \rho + n^{-1/2}R_1 + n^{-1}R_2, \\ T_1 &= 1 + n^{-1/2}R_3 + n^{-1}R_4, \\ T_2 &= 1 + n^{-1/2}R_5 + n^{-1}R_6, \end{aligned}$$

consequently we can express $(T_1T_2)^{-1/2}$ as

$$\Big(T_1T_2\Big)^{-1/2} = 1 + n^{-1/2}R_7 + n^{-1}R_8.$$

It is in the above expression, the condition dictating $\mathcal{A}$ is required, for establishing that R_8 is indeed a random variable satisfying $\mathbb{E}R^{2k} < \infty$.

Using the above terms, we now have that

$$\begin{aligned} \hat{\rho} &= \Big(T_1T_2\Big)^{-1/2}T_3 \\ &= \rho + n^{-1/2}R_9, \end{aligned}$$

where R_9 is a smooth function of all of the following terms:

$$\begin{aligned} T_4 &= n^{1/2}\Big(n^{-1}\sum_{i=1}^{n} Y_{i,1} - \mu_1\Big), \\ T_5 &= n^{1/2}\Big(n^{-1}\sum_{i=1}^{n} Y_{i,2} - \mu_2\Big), \\ T_6 &= n^{1/2}\Big(n^{-1}\sum_{i=1}^{n} (Y_{i,1} - \mu_1)^2 - \sigma_1^2\Big), \\ T_7 &= n^{1/2}\Big(n^{-1}\sum_{i=1}^{n} (Y_{i,2} - \mu_2)^2 - \sigma_2^2\Big), \\ T_8 &= n^{1/2}\Big(n^{-1}\sum_{i=1}^{n} (Y_{i,2} - \mu_2)(Y_{i,2} - \mu_2) - \rho\sigma_1\sigma_2\Big). \end{aligned}$$

The rest of the analysis, along with the justification for the novel resampling scheme used in this paper, now follows using fairly routine algebra, and the use of Lyapunov's Central Limit Theorem. □

10. Future Considerations

In the above analysis, we build a model with a response variable containing pairwise correlations between wavelet coefficients at different time-scales. Then, we perform a modified version of the wild bootstrap for statistical inference.

We utilized a small subset of the available data within the international tree-ring databank. The analysis could be further modified or generalized in many ways. Multiple species

within the same area could be included. It could be of interest to correlate records over a larger land area, or even across contents. Finally, incorporating information, such as temperature and precipitation, may help researchers to better understand past climate. However, all these come with additional challenges, for example, the distribution of tree-ring records over the world is patchy and inadequate in many regions, dependence of the data over different tree species and subspecies, between various climatic regions and other features need to be carefully modeled. Additionally, many data records are incomplete, or are unevenly spaced over time. The theoretical details behind the new resampling scheme deserves further study.

Acknowledgments

Support for this collaboration was provided by the United States National Science Foundation (Awards 1622483 (St. G and SC), 1602512 (St. G), 1737918 (SC), 1939916 (SC) and 1939956 (SC)). We thank the Editors and the reviewers for their constructive comments that helped improve this paper.

References

Affolter, S., Häuselmann, A., Fleitmann, D., Edwards, R. L., Cheng, H., and Leuenberger, M. (2019). Central europe temperature constrained by speleothem fluid inclusion water isotopes over the past 14,000 years. *Science Advances*, **5 (6)**, eaav3809.

Aldrich, E. (2020). *wavelets: Functions for Computing Wavelet Filters, Wavelet Transforms and Multiresolution Analyses*. R package version 0.3-0.2.

Arienzo, M. M., Swart, P. K., Broad, K., Clement, A. C., Pourmand, A., and Kakuk, B. (2017). Multi-proxy evidence of millennial climate variability from multiple Bahamian speleothems. *Quaternary Science Reviews*, **161**, 18–29.

Brook, E. J. and Buizert, C. (2018). Antarctic and global climate history viewed from ice cores. *Nature*, **558 (7709)**, 200–208.

Bunn, A. G., Sharac, T. J., and Graumlich, L. J. (2004). Using a simulation model to compare methods of tree-ring detrending and to investigate the detectability of low-frequency signals. *Tree-Ring Research*, **60 (2)**, 77–90.

Chui, C. K. and Wang, J. (1992). On compactly supported spline wavelets and a duality principle. *Transactions of the American Mathematical Society*, **330 (2)**, 903–915.

Cook, E. R. (1987). The decomposition of tree-ring series for environmental studies. *Tree-Ring Bulletin*, **47**, 37–59.

Cook, E. R. and Peters, K. (1997). Calculating unbiased tree-ring indices for the study of climatic and environmental change. *The Holocene*, **7 (3)**, 361–370.

Cook, E. R., Solomina, O., Matskovsky, V., Cook, B. I., Agafonov, L., Berdnikova, A., Dolgova, E., Karpukhin, A., Knysh, N., Kulakova, M., *et al.* (2020). The european russia drought atlas (1400–2016 ce). *Climate Dynamics*, **54 (3)**, 2317–2335.

Coulthard, B. L., St George, S., and Meko, D. M. (2020). The limits of freely-available tree-ring chronologies. *QSRv*, **234**, 106264.

Daubechies, I. (1992). Ten Lectures on Wavelets. In *CBMS-NSF Regional Conference Series In Applied Mathematics*.

Emile-Geay, J., McKay, N. P., Kaufman, D. S., Von Gunten, L., Wang, J., Anchukaitis, K. J., Abram, N. J., Addison, J. A., Curran, M. A., Evans, M. N., *et al.* (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, **4**, 170088.

Esper, J., St George, S., Anchukaitis, K., D'Arrigo, R., Ljungqvist, F. C., Luterbacher, J., Schneider, L., Stoffel, M., Wilson, R., and Büntgen, U. (2018). Large-scale, millennial-length temperature reconstructions from tree-rings. *Dendrochronologia*, **50**, 81–90.

Fan, Z.-X. and Bräuning, A. (2017). Tree-ring evidence for the historical cyclic defoliator outbreaks on Larix potaninii in the central Hengduan mountains, SW China. *Ecological Indicators*, **74**, 160–171.

Gong, K., Braverman, A., and Chatterjee, S. (2018). On a technique for evaluating the quality of earth system models. In *Proceedings of the 8th International Workshop on Climate Informatics: CI 2018*, pages 93 – 96.

Granato-Souza, D., Stahle, D. W., Barbosa, A. C., Feng, S., Torbenson, M. C., de Assis Pereira, G., Schöngart, J., Barbosa, J. P., and Griffin, D. (2019). Tree rings and rainfall in the equatorial Amazon. *Climate Dynamics*, **52 (3-4)**, 1857–1869.

Harris, I., Osborn, T. J., Jones, P., and Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, **7 (1)**, 1–18.

Holmes, N., Langdon, P. G., Caseldine, C. J., Wastegård, S., Leng, M. J., Croudace, I. W., and Davies, S. M. (2016). Climatic variability during the last millennium in western Iceland from lake sediment records. *The Holocene*, **26 (5)**, 756–771.

Hubbard, B. B. (1998). *The World According to Wavelets: The Story of a Mathematical Technique in the Making*. A K Peters, Ltd., Wellesley, MA, 2nd edition.

Kannenberg, S. A., Novick, K. A., Alexander, M. R., Maxwell, J. T., Moore, D. J., Phillips, R. P., and Anderegg, W. R. (2019). Linking drought legacy effects across scales: From leaves to tree rings to ecosystems. *Global Change Biology*, **25 (9)**, 2978–2992.

Kasatkina, E. A., Shumilov, O. I., and Timonen, M. (2019). Solar activity imprints in tree ring-data from northwestern Russia. *Journal of Atmospheric and Solar-Terrestrial Physics*, **193**, 105075.

Lambert, W. D. (1942). The distance between two widely separated points on the surface of the earth. *Journal of Washington Academy of Sciences*, **32 (5)**, 125–130.

Lau, K.-M. and Weng, H. (1995). Climate signal detection using wavelet transform: How to make a time series sing. *Bulletin of the American Meteorological Society*, **76 (12)**, 2391–2402.

Mandal, M. K., Panchanathan, S., and Aboulnasr, T. (1996). Choice of wavelets for image compression. In Chouinard, J.-Y., Fortier, P., and Gulliver, T. A., editors, *Information Theory and Applications II: 4th Canadian Workshop Lac Delage, Québec, Canada, May 28–30, 1995 Selected Papers*, pages 239–249, Berlin, Heidelberg. Springer Berlin Heidelberg.

McCullough, I. M., Davis, F. W., and Williams, A. P. (2017). A range of possibilities: Assessing geographic variation in climate sensitivity of ponderosa pine using tree rings. *Forest Ecology and Management*, **402**, 223–233.

McKenzie, S. M., Slater, G., Kim, S.-T., Pisaric, M. F., and Arain, M. A. (2018). Influence of seasonal temperature on tree-ring δ13c in different-aged temperate pine forests. *Forest Ecology and Management*, **419**, 197–205.

Mojsilovic, A., Popovic, M. V., and Rackov, D. M. (2000). On the selection of an optimal wavelet basis for texture characterization. *IEEE Transactions on Image Processing*, **9 (12)**, 2043–2050.

Nason, G. (2008). *Wavelet Methods in Statistics with R*. Springer, 1st edition.

Nason, G. (2016). *wavethresh: Wavelets Statistics and Transforms*. R package version 4.6.8.

Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., and Werner, J. P. (2019). No evidence for globally coherent warm and cold periods over the preindustrial common era. *Nature*, **571 (7766)**, 550–554.

Ogden, R. T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser, Boston, 1st edition.

Opała-Owczarek, M. and Niedźwiedź, T. (2019). Last 1100 yr of precipitation variability in western central Asia as revealed by tree-ring data from the pamir-alay. *Quaternary Research*, **91 (1)**, 81–95.

Percival, D. B. and Walden, A. T. (2006). *Wavelet Methods for Time Series Analysis*. Cambridge University Press, 1st edition.

Sandeep, K., Shankar, R., Warrier, A. K., Yadava, M., Ramesh, R., Jani, R., Weijian, Z., and Xuefeng, L. (2017). A multi-proxy lake sediment record of Indian summer monsoon variability during the holocene in southern india. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **476**, 1–14.

St George, S. and Ault, T. R. (2014). The imprint of climate within Northern Hemisphere trees. *Quaternary Science Reviews*, **89**, 1–4.

St George, S. and Esper, J. (2019). Concord and discord among Northern Hemisphere paleotemperature reconstructions from tree rings. *Quaternary Science Reviews*, **203**, 278–281.

Steiger, N. J., Steig, E. J., Dee, S. G., Roe, G. H., and Hakim, G. J. (2017). Climate reconstruction using data assimilation of water isotope ratios from ice cores. *Journal of Geophysical Research: Atmospheres*, **122 (3)**, 1545–1568.

Strickland, R. N. and Hahn, H. I. (1996). Wavelet transforms for detecting microcalcifications in mammograms. *IEEE Transactions on Medical Imaging*, **15 (2)**, 218–229.

The MathWorks, I. (2020). *Wavelet Toolbox*. Natick, Massachusetts, United States.

Torrence, C. and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, **79 (1)**, 61–78.

Trouet, V., Babst, F., and Meko, M. (2018). Recent enhanced high-summer north Atlantic jet variability emerges from three-century context. *Nature Communications*, **9 (1)**, 1–9.

Vannucci, M. and Corradi, F. (1999). Covariance structure of wavelet coefficients: theory and models in a Bayesian perspective. *Journal of the Royal Statistical Society B*, **61 (4)**, 971–986.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. Wiley, New York, NY, 1st edition.

Walden, A. T. and Cristan, A. C. (1998). Matching pursuit by undecimated discrete wavelet transform for non stationary time series of arbitrary length. *Statistics and Computing*, **8**, 205–219.

Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer.

Whitcher, B. (2020). *waveslim: Basic Wavelet Routines for One-, Two-, and Three-Dimensional Signal Processing*. R package version 1.8.2.

Wu, C. F. J. (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis (with discussion). *Annals of Statistics*, **14**, 1295.

Zhao, S., Pederson, N., D'Orangeville, L., HilleRisLambers, J., Boose, E., Penone, C., Bauer, B., Jiang, Y., and Manzanedo, R. D. (2019). The international tree-ring data bank (itrdb) revisited: Data availability and global ecological representativity. *Journal of Biogeography*, **46 (2)**, 355–368.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 187-205

Using Conditionally Specified Joint Distribution to Simultaneously Model Discrete and Continuous Data

Nadeesri Wijekoon[1] **and Nagaraj K. Neerchal**[1,2]
[1]*Department of Mathematics and Statistics*
University of Maryland, Baltimore County, USA
[2]*Chinmaya Vishwavidyapeeth, India*

Received: 29 May 2020; Revised: 21 July 2020; Accepted: 24 July 2020

Abstract

Often, in practice, conditionals are easier to model and interpret while the joint distribution itself is either intractable or not available in a closed form. Conditionally specified statistical models offer several advantages over joint models. Conditionally specified models are intuitively appealing and enrich our ability to build interpretable models in practice. In this paper, we derive the likelihood of a joint distribution obtained from Binomial and Bivariate Normal conditionals. Properties of maximum likelihood estimates and pseudolikelihood estimates are explored using a simulation. A conditionally specified model is obtained by assuming that closing prices are conditionally normally distributed and that the buy-sell recommendation by an Analyst follows a logistic regression model given the prices.

Key words: Conditionally specified models; Compatibility; Maximum likelihood; Pseudolikelihood; Gibb's sampling.

AMS Subject Classifications: 60G50, 05C81

1. Introduction

When solving real-world problems, the main difficulty could be selecting a suitable model to reflect the reality being observed (Ghosh and Nadarajah, 2017). Especially when the observed response consists of both continuous and discrete components, it is not very convenient to directly specify a joint distribution. Often, in practice, conditionals are easier to model and interpret while the joint distribution itself is either intractable or not available in a closed form. According to Arnold *et al.* (1999, 2001), although the joint distribution is less tractable, availability of easily handled conditionals enables us to consider computationally more efficient estimation methods such as pseudolikelihood. Arnold *et al.* (1999, 2001) provide an extensive account of conditionally specified joint distributions. They proved theorems describing the conditions for the existence of joint distributions consistent with the given conditionals and provided general forms of such joint distributions. They also consider examples of joint distributions determined by discrete-continuous pairs of conditionals. It

Correponding Author: Nagaraj K. Neerchal
Email: nagaraj@umbc.edu

turns out that the joint distributions obtained by following the theoretical developments are not always tractable. In particular, often the normalizing constant does not lend itself to a closed form. In this paper, we consider a joint distribution for data consisting of a binary random variable and bivariate continuous measurements. The model postulates a joint distribution determined by Binomial (in fact by a logistic regression model) and bivariate Normal conditionals. For this model, we can derive a closed form for the normalizing constant, however, it turns out that the parameter estimation is numerically intensive. Therefore, we consider an alternative method using a pseudolikelihood function. The two approaches are compared in a simulation study.

As an illustrative application, we consider stock price data along with the corresponding buy-sell recommendation of an analyst. Stock market analysts classify a stock as either a "buy" or a "sell" based on their own research into the history of the stock as well as their assessment of other market dynamics which have a bearing on the price of the stock. Here, the distribution of the stock price on the "buy" days is clearly different from the price distribution on "sell" days. Thus, our choice of the model parameters would differ if we knew that the stock has been classified as a buy instead of as a sell. On the other hand, the price history of the stock will influence the classification (buy or sell) decision of the analyst. Thus, even though it is cumbersome and inconvenient to think of a joint distribution of the stock prices (continuous) and the analyst recommendations (binary), it is easier to think of the conditional distribution of the recommendations given the price history and the conditional distribution of the stock price given its buy or sell status. Thus, conditionally specified distribution is a convenient way to model both, expert recommendations and closing prices simultaneously.

Another motivating example can be found in the area of Gerontology. In health care studies involving aged subjects, due to progressively deteriorating health conditions over time, subjects become unable to respond to questions. To avoid the resulting missing data situation, sometimes the study protocol would allow the investigator to collect a proxy response from another person who is familiar with the non-responding subject. Thus, for each subject, we record a pair of responses. One of the variables in the pair is a discrete random variable (0 or 1 depending upon whether the respondent is either the subject or the proxy) and the other is a continuous (a composite score from a mental health related questionnaire) measurement. Note that the ability to respond is usually related to the overall health condition of the patient. A proxy is needed when a subject is not well enough to answer the questions of the study. On the other hand, the distributional properties of the proxy responses and subject responses would be different. Thus, the conditional relationships between responses and self/proxy indication can be specified using commonly used models and a single joint distribution can be derived for analyzing both subject data and proxy data. The reader is referred to Hosseini (2017) for a detailed description of this approach along with a working code implementing the parameter estimation.

The contents of this paper are organized as follows. In Section 2, we briefly introduce conditionally specified models and present the existing theories of deriving the joint distribution and some issues like compatibility of conditionals. In Section 3, we consider the problem of compatibility and present the relevant restrictions on the original problem. In

Section 4, we propose our new joint distributions. In Section 5, we present an illustrative example using the stock data. Finally, some concluding remarks are made in Section 6.

All computations are done using the freeware R.

2. Conditionally Specified Distributions in Exponential Family

In a comprehensive review given by Arnold *et al.*(2001), in *Statistical Science,* it is stated that a bivariate density is easy to understand/visualize in the terms of its conditional densities. In practice, researchers often have better insight into the form of conditional distributions of experimental variables rather than the joint distribution (See Castillo and Galambos, 1989). For instance, instead of providing a model for (X, Y), one can propose families of conditional distributions of X given values of Y, and of Y given values of X. Castillo and Galambos (1989) identified the complete class of such bivariate distributions with given specified conditional distributions. Arnold and Strauss (1988) extended their work to arbitrary exponential family of conditionals. The key result in this area, which gives the form of the joint distribution which is consistent with the given specific pair of conditional distributions is provided in Arnold and Strauss (1988). Below is a brief statement of this key result, which provides a general form of the joint distribution starting with conditionals belonging to the exponential family of distributions. We start with the following notations.

Notations: Define an l_1-parameter family of densities $\{f_1(x;\underline{\theta}) : \underline{\theta} \in \Theta\}$ with respect to μ_1 (frequently, Lebesgue measure or counting measure) on D_1, a subset of Euclidean space of finite dimension, of the form

$$f_1(x;\underline{\theta}) = r_1(x)\beta_1(\underline{\theta})exp\left\{\sum_{i=1}^{l_1}\theta_i q_{1i}(x)\right\} \tag{1}$$

where $q_{1i}(x)$'s (sufficient statistics) are linearly independent, and $\underline{\theta} = (\theta_1, \cdots, \theta_{l_1})^T$. Similarly, we define, an l_2-parameter family of densities $\{f_2(x;\underline{\tau}) : \underline{\tau} \in\}$ with respect to μ_2 (frequently, Lebesgue measure or counting measure) on D_2, a subset of Euclidean space of finite dimension, of the form

$$f_2(y;\underline{\tau}) = r_2(y)\beta_2(\underline{\tau})exp\left\{\sum_{j=1}^{l_2}\tau_j q_{2j}(y)\right\} \tag{2}$$

where $q_{2j}(y)$'s (sufficient statistics) are linearly independent , and $\underline{\tau} = (\tau_1, \cdots, \tau_{l_2})^T$.

Our goal is to identify the class of bivariate densities $f(x, y)$ with respect to $\mu_1 \times \mu_2$ on $D_1 \times D_2$, whose conditionals belong to the above families of densities respectively. That is, we want to find a joint distribution $f(x, y)$ such that $f(x|y) = f_1(x;\underline{\theta}(y))$ and $f(y|x) = f_2(y;\underline{\tau}(x))$. Arnold *et al.*(1988) show the existence and provide a general form of the joint distribution. Their result is stated in the theorem below.

Theorem 1: Let $f(x, y)$ be a bivariate density whose conditional densities satisfy $f(x|y) = f_1(x;\underline{\theta}(y))$ and $f(y|x) = f_2(y;\underline{\tau}(x))$ for some function $\underline{\theta}(y)$ and $\underline{\tau}(x)$ where f_1 and f_2 are as defined in (3) and (4). It follows that $f(x, y)$ is of the form

$$f(x, y) = r_1(x)r_2(y)exp\left\{\underline{q}^{(1)}(x)^T M \underline{q}^{(2)}(y)\right\} \tag{3}$$

where,

$$\underline{q}^{(1)}(x) = (q_{10}(x), q_{11}(x), q_{12}(x), ..., q_{1l_1}(x))^T,$$
$$\underline{q}^{(2)}(y) = (q_{20}(y), q_{21}(y), q_{22}(y), ..., q_{2l_2}(y))^T$$

and where, $q_{10}(x) = q_{20}(y) = 1$ and M is an $(l_1+1)\times(l_2+1)$ matrix of parameters, subject to the requirement that $\int_{D_1}\int_{D_2} f(x,y)d\mu_1(x)d\mu_2(y) = 1$. The family of these joint distributions is referred to as conditional exponential family (CEF). Note that the elements of M may be denoted by m_{ij} for $i = 0, 1, \cdots, (l_1+1)$ and $j = 0, 1, \cdots, (l_2+1)$.

To illustrate the application of Theorem 1 we present an example where one of the conditionals is the Poisson distribution and the other is the Gamma distribution.

Example 1: This example was given in Arnold *et al.*(1999). Suppose we are seeking a joint distribution of a random vector (X, Y) such that, $X|Y = y \sim Poi(y)$ and assume $Y|X = x \sim \Gamma(x+\alpha, \lambda+1)$. Since both conditionals belong to the exponential family of distributions, we can put them in the notations of Theorem 1 as follows: l_1=1 and l_2=2. The M matrix is 2×3. Further, $r_1(x) = \frac{1}{x!}$ and $r_2(y) = \frac{1}{y}$. Similarly, the sufficient statistics $q_{1i}(x)$'s and $q_{2j}(y)$'s can be identified as $q^{(1)}(x) = (1, x)^T$ and $q^{(2)}(y) = (1, -y, ln(y))^T$. Therefore, joint density belongs to CEF and its general form is given by,

$$f(x,y) = \frac{1}{x!y} exp\left(\begin{bmatrix} 1 & x \end{bmatrix} M \begin{bmatrix} 1 \\ -y \\ ln(y) \end{bmatrix} \right) \qquad x = 0, 1, ...; y > 0$$

where,

$$m_{01} > 0, m_{02} > 0, m_{11} \geq 0, m_{12} \geq 0.$$

Note that, when $m_{11} = m_{12} = 0$, X and Y are independent. And when $m_{11} = 0$ and $m_{12} = 1$, it can be shown that the marginal of X is given by

$$f(x) = \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)x!}\left(\frac{\lambda}{\lambda+1}\right)^{\alpha}\left(\frac{1}{\lambda+1}\right)^{x} \qquad x = 0, 1, 2, ...$$

Thus, we obtain the familiar compound Poisson distribution.

An interesting point to be noted here is that, for $m_{11} > 0$ the joint distribution does not yield the compound Poisson distribution as the marginal of X. This indicates that, even though we started with the Poisson and Gamma conditionals, the resulting CEF is a much larger class than that obtained by combining $X|Y = y \sim Poi(y)$ and $Y \sim \Gamma(\alpha, \lambda+1)$.

It turns out that the candidate functions for the conditional distributions will have to satisfy certain conditions for the existence of a corresponding proper joint distribution. This issue is also referred to as the problem of compatibility of conditionals. According to Chen (2010), the incompatibility of the conditionally specified models may lead to serious consequences on the statistical inference and interpretation in the data analysis and on the convergence of the Gibbs sampling. Thus, the compatibility issue is a widely researched area and there are several computational/theoretical approaches in literature to identify the possible compatibility of given families of conditional distributions. Besag (1974), Arnold and

Press (1989), Hobert and Casella (1998), Arnold, Castillo and Serabia (2002) and recently Ghosh and Nadarajah (2017) studied compatibility extensively. We refer the reader to theorems introduced by Arnold and Press (1989) which are used in checking the compatibility of the conditionals in this paper.

Example 2: (Logistic Regression) Suppose X takes values in the set $\{x_1, x_2, ..., x_k\}$, and is real valued. For each x we have $Y|X = x \sim N(\theta_x, \sigma_x^2)$.

And for each y we have,

$$P(X = x|Y = y) = \frac{exp[-(a_x + b_x y)]}{\sum_{x=1}^{k} exp[-(a_x + b_x y)]}.$$

We apply Theorem 4.1 in the Arnold and Press (1989) paper as follows. The Theorem is stated in the Appendix (A.2).

Proof of Compatibility: Let,

$$a(x, y) = f(Y|X = x) \sim N(\theta_x, \sigma_x^2)$$

$$b(x, y) = f(X|Y = y) \sim \frac{exp[-(a_x + b_x y)]}{\sum_{x=1}^{k} exp[-(a_x + b_x y)]}$$

In order to show compatibility, we need to prove that the ratio of $a(x, y)/b(x, y)$ factors into a product such as $U(x) \times V(y)$. Consider,

$$\begin{aligned}\frac{a(x,y)}{b(x,y)} &= \frac{\frac{1}{\sqrt{2\pi}\sigma_x} exp\left\{ -\frac{(y-\theta_x)^2}{2\sigma_x^2} \right\} \sum_{x=1}^{k} exp[-(a_x + b_x y)]}{exp[-(a_x + b_x y)]} \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} exp\left\{ a_x - \frac{\theta_x^2}{2\sigma_x^2} \right\} exp\left\{ -\frac{1}{2\sigma_x^2}\left[y^2 - 2(\theta_x + \sigma_x^2 b_x)y \right] \right\} \sum_{x=1}^{k} exp[-(a_x + b_x y)]\end{aligned}$$

If $b_x = -\frac{\theta_x}{\sigma_x^2}$ and $\sigma_x^2 = \sigma^2$ where a_x's are unconstrained,

$$\begin{aligned}\frac{a(x,y)}{b(x,y)} &= \frac{1}{\sqrt{2\pi}\sigma} exp\left\{ a_x + \frac{\theta_x b_x}{2} \right\} exp\left\{ -\frac{y^2}{2\sigma^2} \right\} \sum_{x=1}^{k} exp[-(a_x + b_x y)] \\ &= \underbrace{exp\left\{ a_x + \frac{\theta_x b_x}{2} \right\}}_{U(x)} \underbrace{\frac{1}{\sqrt{2\pi}\sigma} exp\left\{ -\frac{y^2}{2\sigma^2} \right\} \sum_{x=1}^{k} exp[-(a_x + b_x y)]}_{V(y)} \\ &= U(x)V(y)\end{aligned}$$

where, $\int_{y \in R} V(y)dy < \infty$ and $\sum_{x=1}^{k} U(x) < \infty$.

Therefore, according to Arnold and Press (1989) the two distributions are compatible provided $b_x = -\frac{\theta_x}{\sigma_x^2}$ and $\sigma_x^2 = \sigma^2$ where a_x's are unconstrained.

3. Logistic and Bivariate Conditionals

In this section, we present the derivation of a conditionally specified model starting with Logistic distribution and Bivariate Normal distribution as conditionals. We consider the stock price data as the motivating example. We propose a logistic regression model for the conditional distribution of the binary valued analyst recommendation given the stock prices from the first and last day of the trading week and the conditional distribution of the stock prices given the analyst recommendation as a Bivariate Normal. We first set up the problem in the notations of Theorem 1 and then obtain the form of the joint. It turns out that the normalizing constant can be obtained in a closed form using some results on multivariate normal integrals. We discuss some properties of the resulting joint distribution. We have created a shiny application (an interactive tool in the freeware R) which can be used to explore the structure of the joint distribution for various parameter values.

3.1. Setting up the problem

Suppose we have the distribution of observed beginning and end price vector (2×1) of a single trading week (say y) given the analyst recommendation (say r), $f_{Y|R}(y|R=r)$ and the distribution of analyst recommendation given observed data vector of a week, $f_{R|Y}(R|Y = y) \sim Ber[\pi(y,\alpha)]$, where, $\pi(y,\alpha)$ is a function of y parameterized by α. We assume a logistic link:

$$\text{logit}[\pi(y,\alpha)] = log\left(\frac{\pi(y,\alpha)}{1-\pi(y,\alpha)}\right) = \alpha_0 + \alpha_1 y_1 + \alpha_2 y_2$$

Further, R is a binary variable and Y is a continuous variable such that $f_{Y|R}(y|R=r) \sim N_2(\mu^{(r)}, \Sigma^{(r)})$. where,

$$\mu^{(r)} = \begin{pmatrix}\mu_1^{(r)}\\ \mu_2^{(r)}\end{pmatrix} \text{ and } \Sigma^{(r)} = \begin{pmatrix}\sigma_1^{(r)2} & \rho\sigma_1^{(r)}\sigma_2^{(r)}\\ \rho\sigma_1^{(r)}\sigma_2^{(r)} & \sigma_2^{(r)2}\end{pmatrix}$$

Under this model, the conditional distributions of the stock price given its buy/sell status is assumed to be normal, with different set of parameter values depending upon the classification. As we shall see in the next section, we will need to assume that the variance-covariance matrices of the two conditionals must be the same (that is, $\Sigma^{(0)} = \Sigma^{(1)}$) in order to ensure compatibility.

3.2. Deriving conditionally specified model

Confirming the existence of the joint model given the two conditionals is essential. In other words, the two conditionals should be compatible. Therefore, for this problem, we will start by checking whether $f_{R|Y}(R|Y=y)$ and $f_{Y|R}(y|R=r)$ are compatible. We will use

the Arnold and Press (1989) Theorem given in the Appendix (A.2). Assume,

$$a(r,y) = (2\pi)^{-1}|\Sigma^{(r)}|^{-1/2}exp[-\frac{1}{2}(y-\mu^{(r)})^T\Sigma^{(r)-1}(y-\mu^{(r)})]$$

and

$$b(r,y) = \pi(y)^r[1-\pi(y)].^{(1-r)}$$

Let us form the ratio,

$$\frac{a(r,y)}{b(r,y)} = \frac{(2\pi)^{-1}|\Sigma^{(r)}|^{-1/2}exp[-\frac{1}{2}(y-\mu^{(r)})^T\Sigma^{(r)-1}(y-\mu^{(r)})]}{\pi(y)^r[1-\pi(y)]^{(1-r)}}$$

Let $\Sigma^{(r)} = \Sigma$, be the common variance-covariance matrix. As in Example 2, letting $-\Sigma^{-1}\mu^{(1)} = (\alpha_1, \alpha_2)^T$, and leaving α_0 unconstrained we can rewrite the above ratio as follows:

$$\begin{aligned}\frac{a(r,y)}{b(r,y)} &= \frac{(2\pi)^{-1}|\Sigma|^{-1/2}exp[-\frac{1}{2}(y^T\Sigma^{-1}y - 2y^T\Sigma^{-1}\mu^{(r)} + \mu^{(r)T}\Sigma^{-1}\mu^{(r)}]}{exp(r\alpha_0).exp(\alpha_0+\alpha_1y_1+\alpha_2y_2)^{-1}} \\ &= \underbrace{\frac{(2\pi)^{-1}|\Sigma|^{-1/2}exp[-\frac{1}{2}(y^T\Sigma^{-1}y)}{exp[\alpha_0+\alpha_1y_1+\alpha_2y_2]^{-1}}}_{\text{U(y)}} \cdot \underbrace{\frac{exp[-\frac{1}{2}(\mu^{(r)T}\Sigma^{-1}\mu^{(r)})]}{exp(r\alpha_0)}}_{\text{V(r)}}\end{aligned}$$

Since $\sum_{r=0}^{1} V(r) < \infty$, compatibility of the given family of conditional densities is assured provided the integrability restriction is also satisfied. That is, the two conditionals are compatible under common variance-covariance matrix Σ and under the condition that $\exp[y^T\Sigma^{-1}y - \alpha_1y_1 - \alpha_2y_2]$ integrates to 1.

Now that the compatibility condition is satisfied, we can apply Theorem 1 to obtain the form of the joint distribution. Theorem 1 was used to obtain the joint distribution of $f(y,r)$. Recall that,

$$\ln(f(y,r)) = (1,r)\begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} & m_{04} & m_{05} \\ m_{10} & m_{11} & m_{12} & m_{13} & m_{14} & m_{15} \end{bmatrix}\begin{bmatrix} 1 \\ y_1 \\ y_2 \\ y_1^2 \\ y_1y_2 \\ y_2^2 \end{bmatrix}$$

For simplicity we write,

$$f(y,r) = exp(m_{00}+rm_{10}).exp(Q(y,r)); r = 0,1 \text{ and } y \in \mathbf{R}$$

where, $Q(y,r) = (m_{01}+rm_{11})y_1 + (m_{02}+rm_{12})y_2 + (m_{03}+rm_{13})y_1^2 + (m_{04}+rm_{14})y_1y_2 + (m_{05}+rm_{15})y_2^2$

Obviously, the tiresome part of this derivation is finding the m values. We will now start by finding the solutions for all the m_{ij} values except for m_{00} which is the normalizing constant. We will present a solution for m_{00} at the end of this section. The usual way to find m values (except m_{00}) is comparing $f(Y|R=r)$ and $f(R|Y=y)$ derived from the joint

distribution with the original $f(Y|R=r)$ and $f(R|Y=y)$. Thus, by comparing we obtain

$$f(R=r|Y=y)=\frac{exp(r[m_{10}+m_{11}y_1+m_{12}y_2+m_{13}y_1^2+m_{14}y_1y_2+m_{15}y_2^2])}{1+exp(m_{10}+m_{11}y_1+m_{12}y_2+m_{13}y_1^2+m_{14}y_1y_2+m_{15}y_2^2)} \quad ; \quad r=0,1$$

By comparing the true density and the density of the desired logistic regression model, we have that

$$\pi(y,\alpha)=\frac{exp(m_{10}+m_{11}y_1+m_{12}y_2+m_{13}y_1^2+m_{14}y_1y_2+m_{15}y_2^2)}{1+exp(m_{10}+m_{11}y_1+m_{12}y_2+m_{13}y_1^2+m_{14}y_1y_2+m_{15}y_2^2)}$$

Desired $\pi(y,\alpha)$ can be obtained by setting the quadratic terms into zero. However, having quadratic terms in the general setting reveals that the derived joint density represents a larger class of joint densities. The particular problem that we are interested in is a special case where the logit link is constructed with a linear function only. Thus, by comparing we get

$$m_{10}=\alpha_0 \tag{4}$$

$$m_{11}=\alpha_1 \tag{5}$$

$$m_{12}=\alpha_2 \tag{6}$$

$$m_{13}=m_{14}=m_{15}=0 \tag{7}$$

Similarly, to obtain $f(Y|R=r)$, we first derive $f(R=r)$ using $f(Y=y,R=r)$.

$$\begin{aligned}
&f(R=r)\\
&=\int_{y_2=-\infty}^{\infty}\int_{y_1=-\infty}^{\infty}f(y,r)dy_1dy_2\\
&=\int_{y_2=-\infty}^{\infty}\int_{y_1=-\infty}^{\infty}\frac{f(y,r)}{f(y|r)}\times f(y|r)dy_1dy_2\\
&=exp(m_{00}+rm_{10})(2\pi)|\Sigma|^{1/2}\int_{y_1,y_2=-\infty}^{\infty}exp\left\{\frac{1}{2}(y-\mu^{(r)})^T\Sigma^{-1}(y-\mu^{(r)})+Q(y,r)\right\}f(y|r)d\underline{y}
\end{aligned}$$

In order to obtain a closed form expression for the normalizing constant, we will need to finish the integration. It turns out that an old result involving multivariate normal density function can be used to accomplish this. The complete result is given in the Appendix (A.1) for the reader's convenience. Once the joint distribution becomes available, the marginals of Y and R can be obtained from the joint distribution. Thus, $f(R=r)$ can be written as follows:

$$\begin{aligned}
&(R=r)\\
&=exp(m_{00}+rm_{10})(2\pi)|\Sigma|^{1/2}\int_{y_1,y_2=-\infty}^{\infty}exp\left\{2b\underline{y}+y^TAy\right\}f(y|r)d\underline{y}\\
&=\frac{exp(m_{00}+rm_{10})(2\pi)|\Sigma|^{1/2}}{|I-2A\Sigma|^{1/2}}exp\left\{2b^T\Sigma(I-2A\Sigma)^{-1}b+\mu^{(r)T}(I-2A\Sigma)^{-1}(2b+A\mu^{(r)})\right\}
\end{aligned}$$

where,

$$\mathrm{b}=\left(\frac{2\rho\sigma_1\sigma_2\mu_2^{(r)}-2\sigma_2^2\mu_1^{(r)}}{4(1-\rho^2)}+\frac{m_{01}+rm_{11}}{2},\quad \frac{2\rho\sigma_1\sigma_2\mu_1^{(r)}-2\mu_2^{(r)}\sigma_1^2}{4(1-\rho^2)\sigma_1^2\sigma_2^2}+\frac{m_{02}+rm_{12}}{2}\right)^T$$

$$\text{and}\quad \mathrm{A}=\begin{pmatrix}\frac{1}{2(1-\rho^2)\sigma_1^2}+m_{03}+rm_{13} & m_{04}+rm_{14}\\ \frac{-\rho}{(1-\rho^2)\sigma_1\sigma_2} & \frac{1}{2(1-\rho^2)\sigma_2^2}+m_{05}+rm_{15}\end{pmatrix}$$

Let us now obtain the conditional distribution $f(Y|R=r)$.

$$f(Y|R=r)$$
$$=\frac{f(Y=y,R=r)}{f(R=r)}$$

$$=\frac{|I-2A\Sigma_r|^{1/2}}{(2\pi)|\Sigma_r|^{1/2}}exp\left\{Q(y,r)-2b^T\Sigma_r(I-2A\Sigma_r)^{-1}b-\mu_r^T(I-2A\Sigma_r)^{-1}(2b+A\mu_r)\right\}$$

We compare the conditional distribution $f(Y|R=r)$ expressed in terms of the m_{ij} values to the originally specified form of the same distribution $(Y|R=r)$ expressed in terms of the parameters $\underset{\sim}{\mu}^{(r)}$'s and Σ's etc. to obtain the relationships between the two sets of parameters. These relationships are captured in the following equations.

$$m_{01}+rm_{11}=\frac{\mu_1^{(r)}-\rho_r(\sigma_{11}/\sigma_{22})\mu_2^{(r)}}{(1-\rho^2)\sigma_{11}^2}$$

$$m_{02}+rm_{12}=\frac{\mu_2^{(r)}(\sigma_{11}/\sigma_{22})^2-\rho(\sigma_{11}/\sigma_{22})\mu_1^{(r)}}{(1-\rho^2)\sigma_{11}^2}$$

$$m_{03}+rm_{13}=\frac{-1}{2(1-\rho^2)\sigma_{11}^2}$$

$$m_{04}+rm_{14}=\frac{\rho}{(1-\rho^2)\sigma_{11}\sigma_{22}}$$

$$m_{05}+rm_{15}=\frac{-1}{2(1-\rho^2)\sigma_{22}^2} \tag{8}$$

The above equations can be solved for m_{ij} values in terms of the $\mu_1^{(r)}$ and $\mu_2^{(r)}$ values etc.

Expressions for $m_{01}, m_{02}, m_{03}, m_{04}$ and m_{05} are as follows:

$$\begin{aligned}
m_{01} &= \frac{\mu_1^{(r)} - \rho(\sigma_{11}/\sigma_{22})\mu_2^{(r)}}{(1-\rho^2)\sigma_{11}^2} - r\alpha_1 \\
m_{02} &= \frac{\mu_2^{(r)}(\sigma_{11}/\sigma_{22})^2 - \rho(\sigma_{11}/\sigma_{22})\mu_1^{(r)}}{(1-\rho^2)\sigma_{11}^2} - r\alpha_2 \\
m_{03} &= \frac{-1}{2(1-\rho^2)\sigma_{11}^2} \\
m_{04} &= \frac{\rho}{(1-\rho^2)\sigma_{11}\sigma_{22}} \\
m_{05} &= \frac{-1}{2(1-\rho^2)\sigma_{22}^2}
\end{aligned} \tag{9}$$

From (9), one can also obtain the original parameters $\mu_1^{(r)}, \mu_2^{(r)}, \rho_r, \sigma_{11}^2$ and σ_{22}^2 in terms of m_{ij}'s. Thus, we have,

$$\rho = \tfrac{m_{04}}{2\sqrt{m_{03}.m_{05}}}, \quad \sigma_{22}^2 = \tfrac{2m_{03}}{m_{04}^2 - 4m_{03}.m_{05}} \text{ and } \quad \sigma_{11}^2 = \tfrac{2m_{05}}{m_{04}^2 - 4m_{03}.m_{05}}. \tag{10}$$

According to (10) it is clear that ρ,σ_{11}^2 and σ_{22}^2 do not depend on the value of R. This implies a common variance-covariance matrix for the conditional distributions $f(Y = y|R = 0)$ and $f(Y = y|R = 1)$. Further, $\mu_1^{(r)}$ and $\mu_2^{(r)}$ as follows.

$$\mu_2^{(r)} = \frac{\rho\left\{\frac{m_{01}+r\alpha_1}{m_{03}}\right\} + \left\{\frac{m_{02}+r\alpha_2}{m_{05}}\right\}}{\rho^2\sqrt{\left\{\frac{m_{05}}{m_{03}}\right\} - \left\{\frac{m_{05}}{m_{03}}\right\}}} \qquad \text{and} \qquad \mu_1^{(r)} = \rho\sqrt{\tfrac{m_{05}}{m_{03}}} \times \mu_2^{(r)} - \tfrac{m_{01}+r\alpha_1}{m_{03}} \tag{11}$$

The above solutions are verified using simulation studies.

3.3. Deriving normalizing constant (m_{00})

In general, obtaining closed form solution to m_{00} is known to be difficult and sometimes such a closed form may not even exist. For our problem however, we were able to derive a closed form expression for m_{00}. Since m_{00} should be such that $\int_{y=-\infty}^{\infty} f(y, r) = 1$, it follows that

$$m_{00} = -ln\left[\sum_{r=0}^{1} exp(rm_{10}) \int_{y1=-\infty}^{\infty}\int_{y2=-\infty}^{\infty} exp(Q(y, r))dy_2dy_1\right]$$

As shown in the Appendix A.1, the double integral above is given by,

$$\begin{aligned}
\int_{y_1=-\infty}^{\infty}\int_{y_2=-\infty}^{\infty} exp(Q(y, r)) = \frac{2\pi\sigma_{11}\sigma_{22}\sqrt{1-\rho^2}}{exp(C(r))}|I - 2A\Sigma|^{-1/2} \\
exp\left[2b^T\Sigma(I - 2A\Sigma)^{-1}b + \mu^{(r)T}(I - 2A\Sigma)^{-1}(2b + A\mu^{(r)})\right]
\end{aligned}$$

where,

$$C(r) = \frac{-(\sigma_{22}^2\mu_1^{(r)2} - 2\rho\sigma_{11}\sigma_{22}\mu_1^{(r)}\mu_2^{(r)} + \sigma_{11}^2\mu_2^{(r)2})}{2(1-\rho^2)\sigma_{11}^2\sigma_{22}^2}$$

$$\text{b}=\left(\frac{2\rho\sigma_{11}\sigma_{22}\mu_2^{(r)}-2\sigma_{22}^2\mu_1^{(r)}}{4(1-\rho^2)} + \frac{m_{01}+rm_{11}}{2}, \quad \frac{2\rho\sigma_{11}\sigma_{22}\mu_1^{(r)}-2\mu_2^{(r)}\sigma_{11}^2}{4(1-\rho^2)\sigma_{11}^2\sigma_{22}^2} + \frac{m_{02}+rm_{12}}{2}\right)^T$$

$$\text{and A}=\begin{pmatrix} \frac{1}{2(1-\rho^2)\sigma_{11}^2} + m_{03} + rm_{13} & m_{04} + rm_{14} \\ \frac{-\rho}{(1-\rho^2)\sigma_{11}\sigma_{22}} & \frac{1}{2(1-\rho^2)\sigma_{22}^2} + m_{05} + rm_{15} \end{pmatrix}$$

.

Therefore, m_{00} is given by,

$$m_{00} = -ln\Bigg[2\pi\sigma_{11}\sigma_{22}\sqrt{1-\rho^2}\sum_{r=0}^{1}|I-2A\Sigma|^{-1/2}.$$
$$exp\Big(rm_{10} - C(r) + 2b^T\Sigma(I-2A\Sigma)^{-1}b + \mu^{(r)T}(I-2A\Sigma)^{-1}(2b+A\mu^{(r)})\Big)\Bigg]$$

4. Data Generation and Estimation

In this section, we discuss how to generate data from the new conditionally specified joint distribution. We also discuss estimation of the parameters of the joint distribution using the maximum likelihood estimation (MLE) method and pseudolikelihood estimation (PLE) method. We will also provide some numerical results to investigate the computational effort involved in computing MLEs and PLEs.

4.1. Data generation from the model

Although the proposed joint model has a closed form expression, it is very complex and has a messy normalizing constant. Therefore, generating data directly from the joint model is immensely difficult and may even not be feasible. However, since the model is conditionally specified we can apply other numerical algorithms such as Gibb's Sampling.

Gibbs sampling algorithm, named by Geman and Geman (1984), is a special case of Metropolis-Hastings algorithm. It is a way to generate data from multivariate distributions when the univariate conditional densities are fully specified. For more details the reader is referred to Rizzo(2008). For instance, let us assume that we want to generate data from a bivariate density $f_{X,Y}(x,y)$ and the conditional distributions of the model are fully specified. Let $f_{X|Y}(x/y)$ and $f_{Y|X}(y/x)$ be conditional densities of x given y and y given x respectively. Suppose simulating data from the bivariate density is complicated, Gibbs sampling algorithm is an effective method to generate data from the conditional densities despite having information about the marginal distributions $f_X(x)$ and $f_Y(y)$. Presented below is the algorithm which can be used to generate data from the proposed density.

Gibb's Sampling Algorithm

1. Initial value : Set $R^{(0)} = 1$. So, $Y^{(0)} \sim f_1(y)$.
2. Set $\mu^{(0)}$,$\mu^{(1)}$,Σ and the vector α.
3. Suppose we generated $(Y^{(0)}, R^{(0)}), (Y^{(1)}, R^{(1)}), ..., (Y^{(t)}, R^{(t)})$.
   (a) if $R^{(t)} = 0$ Generate $Y^{(t+1)} \sim f(y/R = 0)$ and $R^{(t+1)} \sim f(r/Y = y^{(t+1)})$
   (b) if $R^{(t)} = 1$ Generate $Y^{(t+1)} \sim f(y/R = 1)$ and $R^{(t+1)} \sim f(r/Y = y^{(t+1)})$

Contour plots and Surface plots (Figure 1) for $f(Y, R = 0)$ and $f(Y, R = 1)$ suggest that the joint model is not a unimodal distribution. We have built an interactive tool for exploring the shape of the joint distribution in Shiny. The reader may download this tool from *https://github.com/nadeesriw/ShinyApp.git*.

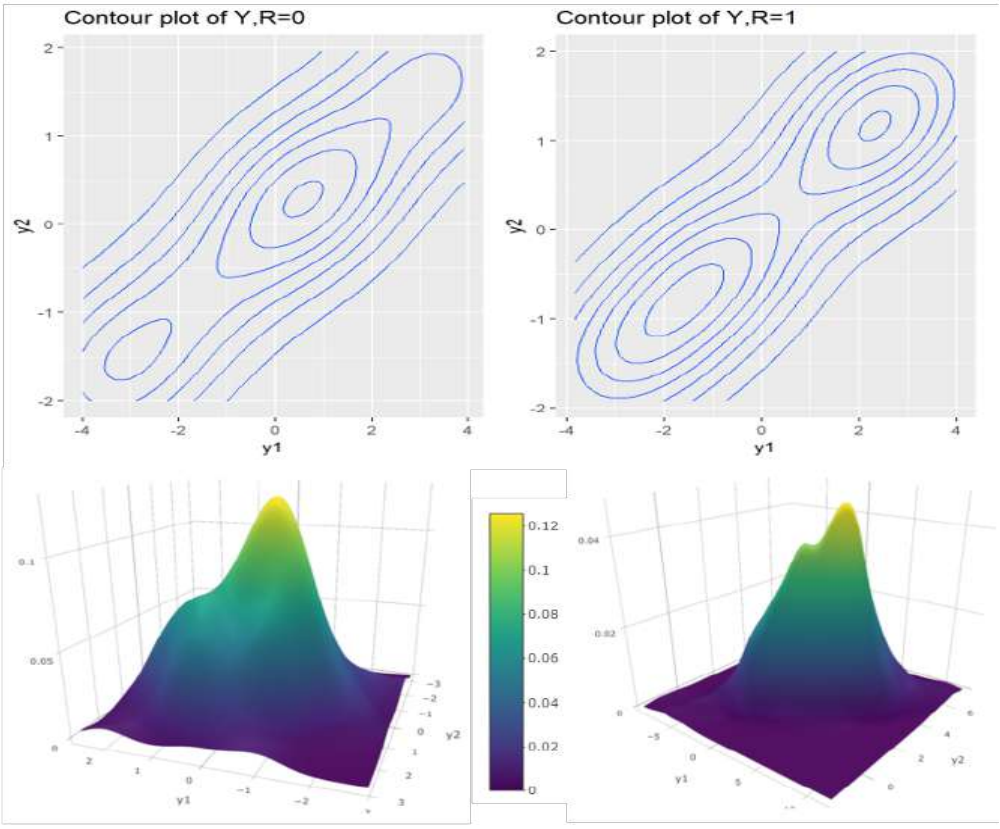


Figure 1: Contour plots and Surface plots of the joint distribution. Top and bottom left: $f(Y, R = 0)$ **and top and bottom right:** $f(Y, R = 1)$**.**

4.2. Estimating parameters

As stated in Arnold *et al.* (1991, 2001) papers, standard estimation methods are often difficult to implement when dealing with conditionally specified models. This is mainly

because of the awkward normalizing constant m_{00} which is often intractable. And, if an explicit expression is available for m_{00}, it is usually complicated, as is the case here. Thus, differentiating the likelihood and deriving the maximum likelihood equations is challenging. Although in practice MLE is the preferred estimation method for parametric models, in our case it comes with a heavy computational burden. Therefore, it behooves upon us to explore other methods such as pseudolikelihood estimation (PLE). In this section, we will compare MLE and PLE in terms of relative efficiency and computational cost.

The likelihood function of the proposed joint distribution is given by,

$$L(f(Y_i, R_i; \hat{\underset{\sim}{M}})) = \prod_{i=1}^{n} \frac{exp(rm_{10} + Q(\underset{\sim}{y}, r))}{\sum_{r=0}^{1} exp(rm_{10}) \int_{y1=-\infty}^{\infty} \int_{y2=-\infty}^{\infty} exp(Q(\underset{\sim}{y}, r)) dy_2 dy_1} \tag{12}$$

where,

$$Q(\underset{\sim}{y}, r) = (m_{01} + rm_{11})y_1 + (m_{02} + rm_{12})y_2 + (m_{03} + rm_{13})y_1^2 + (m_{04} + rm_{14})y_1y_2 + (m_{05} + rm_{15})y_2^2$$

Deriving the score function and obtaining closed form expressions for estimates is not feasible due to the complexity of the model. Therefore, numerical methods were considered to obtain the estimates of parameters.

Arnold and Strauss (1991) proposed PLE as an alternative method to maximum likelihood estimation. Further, they proposed the product of the two conditional distributions as a possible pseudolikelihood function. The advantages of this method are that the function is far more tractable compared to the original likelihood and the awkward normalizing constant m_{00} is absent. The proposed pseudolikelihood function can be written as follows:

$$PL(\mu^{(0)}, \mu^{(1)}, \Sigma, \alpha_0, \alpha_1, \alpha_2) = \prod_{i=1}^{n} f(y_i | R_i = r_i) f(R_i = r_i | y_i)$$

$$\begin{aligned}
&PL(\mu^{(r)}, \Sigma, \alpha_0, \alpha_1, \alpha_2) \\
&= \prod_{i=1}^{n} \frac{1}{2\pi |\Sigma|^{1/2}} \times \\
&exp\left\{ -\frac{1}{2} (\underset{\sim}{y}_i - \mu^{(1)} r_i - \mu^{(0)}(1 - r_i))^T [\Sigma]^{-1} (\underset{\sim}{y}_i - \mu^{(1)} r_i - \mu^{(0)}(1 - r_i)) \right\} \times \\
&[exp\{\underset{\sim}{\alpha}^T y_i^*\}]^{r_i} \frac{1}{1 + exp\{\underset{\sim}{\alpha}^T y_i^*\}}
\end{aligned}$$

where, $\alpha = (\alpha_0, \alpha_1, \alpha_2)^T$ and $\underset{\sim}{y}_i^* = (1, \underset{\sim}{y}_i)^T$. While the computation of PLE is also only feasible numerically, the computations involved are significantly less complex than the computation of the original likelihood.

We carried out a simulation study for different sample sizes varying from 100 to 1000. Estimates of the model parameters were obtained using both MLE and PLE methods. Data were generated using the Gibb's algorithms. 100 such data sets were used for the study. Apart from the estimates, we also calculated the variance, bias and time. The entire analysis was performed using the R 3.5.1 software. The R package called *Rsolnp*, which performs constrained optimization, was used to obtain estimates. Results are shown in Table 1 and Table 2.

Table 1: Wall times of simulation study with respect to sample size and method used: TOL 10^{-3} and 100 data sets

Sample Size (n)	Wall time (in hours)	
	Maximum Likelihood Estimates	Pseudolikelihood Estimates
100	0.24876	0.03046
200	0.48388	0.05411
500	1.01297	0.12695
1000	18.74536	13.94225
10000	29.14374	15.74069

In Table 1, the first column shows the wall times for different sample sizes n, when the maximum likelihood method was used. The second column of Table 1 presents the wall times for pseudolikelihood method. According to results, we can see that pseudolikelihood method is superior to maximum likelihood method in terms of computation efficiency. For n=100, wall time of pseudolikelihood estimation for 100 iterations is 0.0304 hours (1.83 mins) while the wall time of maximum likelihood estimation for 100 iterations is 0.2488 hours ($\sim$ 15 mins). It is apparent that for small sample sizes such as $n = 100$, 200 there is no significant computational cost difference between the two methods. However, for a larger sample size, the computational advantage of PLE surpasses that of MLE quickly. Further, we note that even the PLE method shows large wall times when n increases. In that case, paralleling the code would be more effective. We are planning to explore this in the future.

In Table 2, we present the estimates, the bias and the variances for different sample sizes. By looking at the bias values it is clear that the MLE method has less bias (and nearly zero in some cases) than the PLE method. In both methods, the variances of the estimates decrease as the sample size increases. The MLE of the α vector has less bias compared to the PLE. Overall, it is evident that the MLE outperforms the PLE in terms of accuracy (less bias) and efficiency (lower variance). Therefore, we can conclude that choosing PLE over MLE is a trade off between efficiency and computational cost.

5. Illustrative Example

In this section we provide an illustrative example. We apply the derived joint distribution to a data set with stock prices and a weekly buy/sell recommendation for stocks. All the data has been downloaded from *www.Barchart.com*. The dataset consists of closing prices and recommendations from Monday to Friday for 40 weeks. As the starting step, we only considered the two closing prices from Monday and Friday of each week and the recommendation from Monday of the following week. The data structure is given in Table 3.

Table 2: Maximum likelihood estimates and pseudolikelihood estimates for different sample sizes (n): TOL 10^{-3}

Parameters	True Values	Maximum Likelihood		Pseudolikelihood	
		Estimate (std)	Bias	Estimate (std)	Bias
$n = 100$					
σ_0^2	1.2000	1.1986(0.0100)	-0.0014	1.1836(0.0213)	-0.0164
σ_{01}	0.6481	0.6455(0.0100)	-0.0026	0.6510(0.0135)	0.0029
σ_1^2	1.4000	1.3981(0.0100)	-0.0018	1.4273(0.0213)	0.0273
$\mu_1^{(1)}$	2.0000	2.1828(0.0173)	0.1828	1.9879(0.0176)	-0.0120
$\mu_2^{(1)}$	3.0000	3.2157(0.0141)	0.2157	3.0179(0.0253)	0.0179
$\mu_1^{(0)}$	0.0000	-0.0104(0.0223)	-0.0104	0.0053(0.0182)	0.0053
$\mu_2^{(0)}$	0.0000	-0.0213(0.0173)	-0.0213	-0.0008(0.0218)	-0.0008
α_0	0.0010	0.0756(0.0012)	0.0746	-2.3291(0.2972)	-2.3301
α_1	0.0010	0.0009(0.0000)*	-0.0001	0.5229(0.1117)	0.5219
α_2	0.0010	0.0009(0.0000)*	-0.0001	1.2877(0.1710)	1.2867
$n = 200$					
σ_0^2	1.2000	1.1993(0.0001)	-0.0007	1.2119(0.0085)	0.0119
σ_{01}	0.6481	0.6458(0.0007)	-0.0023	0.6498(0.0053)	0.0017
σ_1^2	1.4000	1.3990(0.0017)	-0.0009	1.4205(0.0073)	0.0205
$\mu_1^{(1)}$	2.0000	2.1824(0.0001)	0.1824	1.9785(0.0076)	-0.0215
$\mu_2^{(1)}$	3.0000	3.2145(0.0001)	0.2145	2.9683(0.0115)	-0.0317
$\mu_1^{(0)}$	0.0000	-0.0124(0.0002)	-0.0124	-0.0049(0.0073)	-0.0049
$\mu_2^{(0)}$	0.0000	-0.0219(0.0001)	-0.0219	0.0013(0.0130)	0.0013
α_0	0.0010	0.0736(0.0000)*	0.0726	-2.2607(0.1565)	-2.2617
α_1	0.0010	0.0009(0.0000)*	-0.0001	0.5220(0.0565)	0.5210
α_2	0.0010	0.0009(0.0000)*	-0.0001	1.1872(0.0994)	1.1862
$n = 500$					
σ_0^2	1.2000	1.1996(0.0000)*	-0.0003	1.2116(0.0009)	0.0116
σ_{01}	0.6481	0.6458(0.0008)	-0.0023	0.6434(0.0014)	-0.0046
σ_1^2	1.4000	1.3995(0.0000)*	-0.0005	1.4109(0.0010)	0.0109
$\mu_1^{(1)}$	2.0000	2.1842(0.0078)	0.1842	1.9995(0.0036)	-0.0005
$\mu_2^{(1)}$	3.0000	3.2158(0.0080)	0.2158	3.0081(0.0056)	0.0081
$\mu_1^{(0)}$	0.0000	-0.0107(0.0112)	-0.0107	0.0035(0.0045)	0.0035
$\mu_2^{(0)}$	0.0000	-0.0210(0.0069)	-0.0210	0.0069(0.0047)	0.0069
α_0	0.0010	0.0755(0.0037)	0.0745	-2.3054(0.0940)	-2.3064
α_1	0.0010	0.0009(0.0000)*	-0.0001	0.4747(0.0171)	0.4737
α_2	0.0010	0.0009(0.0000)*	-0.0001	1.1533(0.0153)	1.1523
$n = 1000$					
σ_0^2	1.2000	1.1998(0.0000)*	-0.0002	1.2039(0.0002)	0.0039
σ_{01}	0.6481	0.6457(0.0009)	-0.0024	0.6434(0.0009)	-0.0046
σ_1^2	1.4000	1.3997(0.0000)*	-0.0003	1.4085(0.0003)	0.0085
$\mu_1^{(1)}$	2.0000	2.1814(0.0088)	0.1814	1.9947(0.0019)	-0.0053
$\mu_2^{(1)}$	3.0000	3.2237(0.0107)	0.2237	2.9945(0.0032)	-0.0055
$\mu_1^{(0)}$	0.0000	-0.0250(0.0096)	-0.0250	-0.0042(0.0021)	-0.0042
$\mu_2^{(0)}$	0.0000	-0.0241(0.0064)	-0.0241	0.0009(0.0019)	0.0009
α_0	0.0010	0.0702(0.0035)	0.0692	-2.3886(0.0976)	-2.3896
α_1	0.0010	0.0009(0.0000)*	-0.0001	0.4790(0.0096)	0.4780
α_2	0.0010	0.0009(0.0000)*	-0.0001	1.1613(0.0094)	1.1603

$*$: very small non zero values

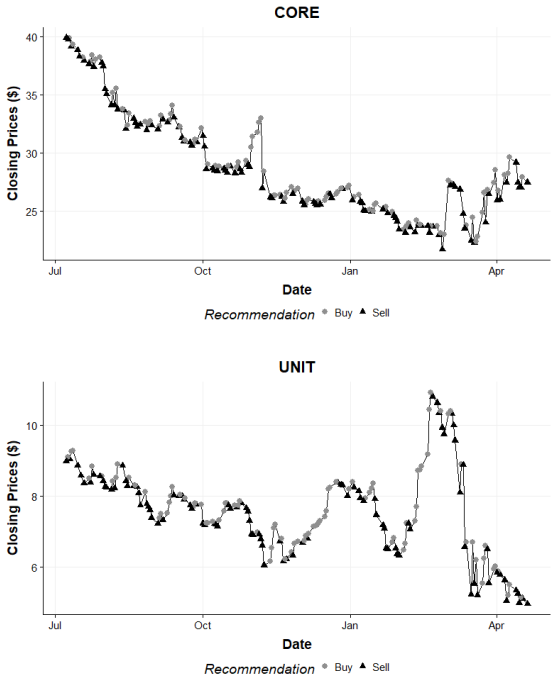


Figure 2: Closing prices, recommendations

Table 3: Data structure

		Closing Prices (Mon.price, Fri.price)	Recommendation
Stock 1	Week 1	$(y_{11}^{(1)}, y_{11}^{(2)})$	Buy
	Week 2	$(y_{12}^{(1)}, y_{12}^{(2)})$	Buy
	⋮	⋮	⋮
	Week 40	$(y_{1,32}^{(1)}, y_{1,40}^{(2)})$	Sell
Stock 2	Week 1	$(y_{21}^{(1)}, y_{21}^{(2)})$	Buy
	Week 2	$(y_{22}^{(1)}, y_{22}^{(2)})$	Sell
	⋮	⋮	⋮
	Week 40	$(y_{2,32}^{(1)}, y_{2,40}^{(2)})$	Sell
Stock_i	week_j	$(y_{ij}^{(1)}, y_{ij}^{(2)})$	r_{ij}

Conditional distributions of closing prices of Monday and Friday of each week given the end of the week recommendation, is assumed to follow a bivariate normal distribution and the end of the week recommendation given the closing prices of Monday and Friday is assumed to follow a standard logistic regression model. The normality assumption is crucial. In this example, we choose stocks where closing prices follow bivariate normal distribution. However, most of the closing prices of stock data do not follow a bivariate normal distribution and hence one might need to transform the data using methods such as Box-Cox transformation. The pseudolikelihood and the maximum likelihood methods are used to estimate the parameters. Results are given in Table 4. Note that the means, variances and correlation estimates

Table 4: Maximum likelihood and pseudolikelihood estimates for stock price data

Stock	Method	$\mu^{(0)}$	$\mu^{(1)}$	σ_1^2	σ_2^2	$\rho\sigma_1\sigma_2$	α_0	α_1	α_2
CORE	Pseudolikelihood	(28.80,28.50)	(29.14,28.59)	19.92	18.29	18.47	-0.36	0.20	-0.19
	Maximum Likelihood	(28.80,28.51)	(29.14,28.60)	19.55	17.96	18.12	-0.54	0.40	-0.39
UNIT	Pseudolikelihood	(7.51,7.33)	(7.70,7.76)	1.73	1.62	1.54	-2.37	-0.807	1.02
	Maximum Likelihood	(7.51,7.33)	(7.70,7.76)	1.72	1.62	1.54	-4.01	-1.64	2.08

obtained by the two methods are very close to each other. However, the estimates of the parameters of the logistic regression model are not so close. The other important point to note is that the α_2 parameter seems to be significantly different from zero. This implies that the product of the Monday and Friday prices is significant in the logistic regression model.

6. Remarks

A researcher can get better insight into the data using conditional specification of experimental variables rather than a joint distribution. Using conditional specification makes the visualization easier. Furthermore, it makes it easier to use Gibbs sampling for data generation. We derived Joint distribution starting from $Y|R = r$ bivariate normal and $R|Y = y$ logistic regression.

In practice, the maximum likelihood method is a preferred estimation method for parametric models. However, if the likelihood function is complicated or contains an awkward normalizing constant, the implementation becomes difficult. Among other methods, pseudolikelihood estimation is intuitively appealing and computationally less burdensome. According to the simulation study, choosing the pseudolikelihood method over the maximum likelihood method is a trade off between the efficiency and accuracy of estimates and the computational cost.

Acknowledgments

The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is partially supported by the U.S. National Science Foundation through the MRI program (OAC–1726023). See *hpcf.umbc.edu* for more information on HPCF and the projects using its resources.

References

Arnold, B. C., Castillo, E. and Sarabia, J. M. (1999). *Conditional Specification of Statistical Models.* Springer.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2001). Conditionally specified distributions: An introduction. *Statistical Science,* **16**, 249–274.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2002). Exact and near compatibility of discrete conditional distributions. *Computational Statistics and Data Analysis,* **40**, 231–252.

Arnold, B. C. and Gokhale, D. (1994). On uniform marginal representation of contingency tables. *Statistic and Probability Letters,* **21**, 311–316.

Arnold, B. C., James, S. and Press, S. J. (1989). Compatible conditional distributions. *Journal of the American Statistical Association,* **84**, 152–156.

Arnold, B. C. and Strauss, D. (1988). Bivariate distributions with exponential conditionals. *Journal of the American Statistical Association,* **83**, 522–527.

Arnold, B. C. and Strauss, D. (1991). Pseudolikelihood estimation: Some examples. *Sankhya: The Indian Journal of Statistics* **B(53)**, 233–243.

Besag, J. (1974). Spatial interaction and the statistical analysis of Lattice systems. *Journal of the Royal Statistical Society,* **B(36)**, 192–225.

Castillo, E. and Galambos, J. (1989). Conditional distributions and the bivariate normal distribution. *Metrika,* **36**, 209–214.

Chen, H. Y. (2010). Compatibility of conditionally specified models. *Statistics and Probability Letters,* **80**, 670–677.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **6(6)**, 721–741.

Ghosh, I. and Nadarajah, S. (2017). On the construction of a joint distribution given two discrete conditionals. *Studia Scientiarum Mathematicarum Hungarica,* **54**, 178–204.

Hobert, J. P. and Casella, G. (1998). Functional compatibility, Markov chains, and Gibbs sampling with improper posteriors. *Journal of Computational and Graphical Statistics,* **7**,42–60.

Hosseini, M. (2017). *Statistical Modeling of Self and Proxy Responses with Gerontological Applications.* PhD thesis, University of Maryland Baltimore County.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rizzo, M. L. (2008). *Statistical Computing with R.* Chapman Hall/CRC.

APPENDIX

1. Moment Generating Function Theorem

This theorem can be used to derive the moment generating function of a linear or a quadratic form in a normal random vector $\underline{x}$ or, more generally, of a second degree polynomial in $\underline{x}$.

Theorem A.1: Let $\underline{x} \sim N_n(\mu, \Sigma)$ and take $q_i = 2b_i\underline{x} + \underline{x}^T A\underline{x}$ where b_i is an $n \times 1$ non random vector and A_i is an $n \times n$ non random symmetric matrix (i=1,2,...,k). Take Γ to be any $r \times n$ matrix such that $\Sigma = \Gamma^T\Gamma$, where $r = \text{rank}\Sigma$. The joint m.g.f. of $q_1, ..., q_k$ is given by,

$$m_{q_1,...,q_k}(t_1, ..., t_k) = |I - 2\sum_{i=1}^{k} t_i\Gamma A_i\Gamma^T|^{-1/2}.exp\Bigg\{2\left[\sum_{i=1}^{k} t_i(b_i + A_i\mu)\right]^T \Gamma^T\left[I - 2\sum_{i=1}^{k} t_i\Gamma A_i\Gamma^T\right]^{-1}\Gamma \left[\sum_{i=1}^{k} t_i(b_i + A_i\mu)\right] + \mu^T\sum_{i=1}^{k} t_i(2b_i + A_i\mu)\Bigg\}$$

$$= |I - 2\sum_{i=1}^{k} t_i\Gamma A_i\Gamma^T|^{-1/2}.exp\Bigg\{2\left[\sum_{i=1}^{k} t_ib_i\right]^T \Sigma\left[I - 2\sum_{i=1}^{k} t_iA_i\Sigma\right]^{-1}\left[\sum_{i=1}^{k} t_ib_i\right] + \mu^T[I - 2\sum_{i=1}^{k} t_iA_i\Sigma\Bigg]^{-1}\sum_{i=1}^{k} t_i(2b_i + A_i\mu)\Bigg\}$$

where $(|t_i| < h_i; i = 1, ..., k)$. for sufficiently small positive constants $h_1, ..., h_k$.

2. Compatibility Theorem: Arnold and Press (1989)

we have,

$a(x,y) = f_{X|Y}(x|y),\ \ x \in S(X),\ \ y \in S(Y),$

$b(x,y) = f_{Y|X}(y|x),\ \ x \in S(X),\ \ y \in S(Y),$

and let, $N_A = (x,y) : a(x,y) > 0$ and $N_B = (x,y) : b(x,y) > 0$

Theorem A.2: A joint density $f(x,y)$, with $a(x,y)$ and $b(x,y)$ as its conditional densities, will exist if and only if,

1. N_A=N_B=N
2. $\exists$ functions u and v such that $\forall$ x,y,$\in$ N

$$a(x,y)/b(x,y) = u(x)v(y)$$

where

$$\int_{S(X)} u(x)d\mu_1(x) < \infty$$

Statistics and Applications (ISSN 2452-7395 (online))
Volume 18, No. 2, 2020 (New Series), pp 207 – 221

Bayesian Mixture Designs for Hypothesis Testing

N.K. Mandal[1] and Manisha Pal[2]
[1]*Former Professor, Department of Statistics, University of Calcutta*
[2]*Department of Statistics, University of Calcutta*

Received: 07 June 2020; Revised: 22 July 2020; Accepted: 28 July 2020

Abstract

Scheffé (1958) first introduced models of different degrees to represent the response function in a mixture experiment. He also introduced corresponding designs for the estimation of the model parameters. Later, several authors studied the problem of finding optimum designs, especially for quadratic and cubic models. In this paper, we consider the problem of designing an optimal experiment for the purpose of performing one or more hypothesis tests in a first-degree mixture model. The Bayesian decision theoretic approach is used for this purpose.

Key words: Mixture experiments; Hypothesis testing; Bayes optimality; Normal prior; Optimum designs.

0. Tribute to Professors Sinha

We feel privileged to be able to contribute to this special issue of Statistics and Applications. We are fortunate to have come in contact with the highly acclaimed statisticians, the Sinha brothers, especially Professor Bikas K. Sinha, with whom we share a very close bond. We have been working with Professor Bikas K. Sinha, whom we fondly call Bikas Da, for more than a decade. His advice, support, positive attitude and, above all, his unbounded energy have been highly inspirational to us.

The first author had the good fortune of being taught by Bikas Da in the Statistics post-graduate course of Calcutta University in the mid seventy's, and, over time, has developed a close brotherly bond with him. The second author feels grateful for the constant encouragement that she has received from him, and for her enhanced knowledge of DoE through her association with him.

1. Introduction

Literature on Bayesian optimal design is generally based on linear models and the loss functions are chosen so as to be appropriate for estimation of the unknown parameters, and also for prediction purpose. See, for example, Chaloner (1982, 1984), El-Krunz and Studden (1991), Chaloner and Verdinelli (1995), Dasgupta (1996). Keeping in mind that the data analyst may also be interested in testing hypotheses regarding the parameters, Toman (1996)

Corresponding Author: Manisha Pal
E-mail: manishapal2@gmail.com

attempted to find optimum designs for hypotheses testing, using a suitable loss function in the discrete set-up. He considered both the cases of single hypothesis and multiple hypotheses testing. In the former case, the optimal design minimizes the Bayes risk, while in the latter case, where more than one decision is to be made, Toman (1996) suggested two approaches – one to minimize a Bayes risk while the other risks are constrained to be less than specified values, and the other to minimize the weighted sum of the Bayes risks, the weights being suitably selected by the experimenter. No further studies along this line have come to our notice. It is also noteworthy that the problem of determining the optimum Bayesian designs for hypotheses tests in the mixture set-up has not been addressed so far.

In this paper, we have considered the first-degree homogeneous mixture model due to Scheffé (1958). With suitably defined loss function, we have determined the Bayes optimal designs for testing both single and multiple hypotheses. The paper has been organized as follows. In Section 2 we have discussed the loss function and the Bayes risk for a linear model. In Section 3 we have obtained Bayes optimal designs for single hypotheses regarding the parameters of the model. Bayes designs for multiple hypotheses are discussed in Section 4. In Section 5, examples have been cited for multiple tests, and a discussion on the article has been given in Section 6.

2. The Loss Function and the Bayes Risk

Consider the linear model $\boldsymbol{Y} \sim N_p(\boldsymbol{X\theta}, \sigma^2 I_p)$, where $\boldsymbol{Y}$ is the vector of observations, $\boldsymbol{X}$ is the design matrixof order $p\times q$ and $\boldsymbol{\theta}$ is the parameter vector of order $q\times 1$. Suppose the prior distribution of $\boldsymbol{\theta}$ is $N_q(\boldsymbol{\tau}, \sigma^2 R)$, where R is a positive definite matrix.

Then, the posterior distribution of $\boldsymbol{\theta}$ is $N_q(\mu, V)$, where

$$\mu = \hat{\boldsymbol{\theta}} - (M + R^{-1})^{-1} R^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\tau})$$
$$V = \sigma^2 (M + R^{-1})^{-1}$$
$$M = \boldsymbol{X'X},$$

and $\widehat{\boldsymbol{\theta}}$ is the least squared estimator of $\boldsymbol{\theta}$.

Suppose that one is interested to test the following k hypotheses:

H_{0i}: $\boldsymbol{c}_i'\boldsymbol{\theta} \geq v_i$ versus H_{1i}: $\boldsymbol{c}_i'\boldsymbol{\theta} < v_i$; $i = 1, 2, \ldots, k$,

where $\boldsymbol{c}_i$ and v_i, $i = 1, 2, \ldots, k$ are specified.

Let $\boldsymbol{d} = (d_1, d_2, ..., d_k)$ be the decision vector corresponding to a given $\boldsymbol{\theta}$, where d_i denotes the decision for the i-th testing problem $i = 1, 2, \ldots, k$. Let us denote by a_{i1} the action favoring $\boldsymbol{c}_i'\boldsymbol{\theta} \geq v_i$ and by a_{i2} that favoring $\boldsymbol{c}_i'\boldsymbol{\theta} < v_i$ in the i-th problem.

DeGroot (1970) showed the following loss function to be appropriate for the i-th problem in the general set-up:

$L_i(\boldsymbol{\theta}, a_{i1}) = 0$, if $\boldsymbol{c}_i'\boldsymbol{\theta} \geq v_i$

$= v_i - \boldsymbol{c}_i'\boldsymbol{\theta}$, if $\boldsymbol{c}_i'\boldsymbol{\theta} < v_i$,

and

$L_i(\boldsymbol{\theta}, a_{i2}) = \boldsymbol{c}_i'\boldsymbol{\theta} - v_i$, if $\boldsymbol{c}_i'\boldsymbol{\theta} \geq v_i$
$= 0$, if $\boldsymbol{c}_i'\boldsymbol{\theta} < v_i$, for $i = 1\ (1)\ k$.

Let $\delta^i(\boldsymbol{y})$ be the Bayes decision rule for the i-th problem. The Bayes risk is obtained by averaging the losses over both $\boldsymbol{Y}$ and $\boldsymbol{\theta}$.

Given a design with information matrix M, the Bayes risk $r_i(\delta^i, M)$ for the i^{th} testing problem is given as (*cf*. DeGroot, 1970):

$$r_i(\delta, M) = \sigma(\boldsymbol{c}_i' R\, \boldsymbol{c}_i)^{1/2}\ \{\Psi(s_i) - (1-p_i)^{1/2}\,\Psi(s_i/(1-p_i)^{1/2})\}, \tag{1}$$

where $\psi(s) = \varphi(s) - s[1-\Phi(s)]$, $\varphi(s)$ and $\Phi(s)$ are the density and cumulative distribution functions respectively of a standard normal variate, $p_i = \boldsymbol{c}_i'(R^{-1}+M)^{-1}\boldsymbol{c}_i/(\boldsymbol{c}_i' R\, \boldsymbol{c}_i)$ is the ratio of the posterior variance to the prior variance of $\boldsymbol{c}_i'\boldsymbol{\theta}$, and $s_i = (v_i - \boldsymbol{c}_i'\boldsymbol{\tau})/\sigma(\boldsymbol{c}_i' R\, \boldsymbol{c}_i)^{1/2}$, the standardized difference between the constant v_i and the prior mean of $\boldsymbol{c}_i'\boldsymbol{\theta}$. For the single hypothesis case, where the hypotheses are H_0: $\boldsymbol{c}'\boldsymbol{\theta} \geq v$ against H_1: $\boldsymbol{c}'\boldsymbol{\theta} < v$, the optimum design is selected so as to minimize the Bayes risk.

Now, $\dfrac{\partial}{\partial p} r(\delta, M) = 2\sigma(\boldsymbol{c}' R \boldsymbol{c})^{1/2}(1-p)^{1/2}\varphi(s/(1-p)^{1/2})$, which is positive, whatever be s. Hence, for every s, the risk function $r(\delta, M)$ is increasing in p. So, minimization of $r(\delta, M)$ can be achieved through minimization of p, which does not involve s, and hence the prior mean. Again, since $p = \boldsymbol{c}_i'(R^{-1}+M)^{-1}\boldsymbol{c}_i/(\boldsymbol{c}_i' R\, \boldsymbol{c}_i)$, and the denominator is free from the design, the Bayes ψ-optimal design will be obtained by minimizing

$$\phi(M) = \text{Trace}\,[\boldsymbol{c}\boldsymbol{c}'(R^{-1} + M)^{-1}]. \tag{2}$$

We consider Scheffé's first order mixture model and work in a continuous design setting.

3. Optimal Mixture Design for a Single Test

Consider the mixture model given below:

$$\eta_{\boldsymbol{x}} = E(Y \mid \boldsymbol{x}) = \sum_{i=1}^{q} \beta_i x_i, \tag{3}$$

where Y denotes the response and $\boldsymbol{x} = (x_1, x_2, \ldots, x_q)$ the mixing proportions of the ingredients. The experimental region is $\Xi = \{(x_1, x_2, \ldots, x_q) : x_i \geq 0, i = 1(1)q, \sum_{i=1}^{q} x_i = 1\}$.

Here, one may be interested in single tests of the following form:

(I) H_0: $\beta_i \geq 0$ versus H_1: $\beta_i < 0$, for some i, $1 \leq i \leq q$
(II) H_0: $\beta_i - \beta_j \geq 0$ versus H_1: $\beta_i - \beta_j < 0$, for some i, j, $1 \leq i < j \leq q$
(III) H_0: $c'\boldsymbol{\beta} \geq 0$ versus H_1: $c'\boldsymbol{\beta} < 0$, where c is a $q \times 1$ real vector.

Before proceeding further, we note an important property of $\phi(M)$, defined in (2).

Property 1: $\phi(M)$, given by (2), is convex in M.

Proof: Let us write

$$\phi(M) = \text{Trace}\,[\boldsymbol{c}\boldsymbol{c}'(R^{-1}+M)^{-1}] = [\boldsymbol{c}'(R^{-1}+M)^{-1}\boldsymbol{c}].$$

We have to show that $\phi(\lambda M_1 + (1-\lambda)M_2) \le \lambda\phi(M_1) + (1-\lambda)\phi(M_2)$.

Now, $\phi(\lambda M_1 + (1-\lambda)M_2) = [\boldsymbol{c}'(R^{-1} + (\lambda M_1 + (1-\lambda)M_2))^{-1}\boldsymbol{c}]$

$$= [\boldsymbol{c}'(\lambda(R^{-1}+M_1) + (1-\lambda)(R^{-1}+M_2))^{-1}\boldsymbol{c}]. \quad (4)$$

But it is known that

$$[\lambda(R^{-1}+M_1) + (1-\lambda)(R^{-1}+M_2)]^{-1} \le \lambda(R^{-1}+M_1)^{-1} + (1-\lambda)(R^{-1}+M_2)^{-1} \quad (5)$$

(*cf.* Fedorov (1972)).

Hence from (4) and (5), we have

$$\phi(\lambda M_1 + (1-\lambda)M_2) \le \lambda\boldsymbol{c}'((R^{-1}+M_1)^{-1}\boldsymbol{c} + (1-\lambda)\,\boldsymbol{c}'(R^{-1}+M_2)^{-1}\boldsymbol{c})$$
$$= \lambda\phi(M_1) + (1-\lambda)\,\phi(M_2),$$

which establishes the convexity of $\phi(M)$. □

For the problem of minimizing $\phi(M)$, given by (2), a necessary and sufficient condition for a design to be Bayes optimal is obtained by applying the generalized equivalence theorem (*cf.* Whittle (1973); Silvey (1980)), which gives:

Theorem1: Any one of the following 3 conditions is necessary and sufficient for a design ξ_0 with information matrix M_0, to be optimal:

(i) $F_\varphi(R^{-1}+M_0, R^{-1}+M) \ge 0$ for all $R^{-1}+M$, for all $M \in \mathsf{M}$

(ii) $F_\varphi(R^{-1}+M_0, R^{-1}+\boldsymbol{x}\boldsymbol{x}') \ge 0$ for all $\boldsymbol{x}\in\Xi$

(iii) $\min_{\boldsymbol{x}\in\Xi}[F_\varphi(R^{-1}+M_0, R^{-1}+\boldsymbol{x}\boldsymbol{x}')] = \max_{R^{-1}+\boldsymbol{x}\boldsymbol{x}'\in\boldsymbol{S}} \min_{\boldsymbol{x}\in\Xi}[F_\varphi(R^{-1}+M, R^{-1}+\boldsymbol{x}\boldsymbol{x}')]$,

where $F_\varphi(M_1, M_2)$ is the Fréchet derivative of M_1 in the direction M_2, M is the class of all information matrices, Ξ is the domain of $\boldsymbol{x}$ and $\boldsymbol{S}$ is the class of all non-singular matrices. Further, if $M_0 = \sum v_i \boldsymbol{x}_i\boldsymbol{x}_i'$, where $\boldsymbol{x}_i \in \Xi$ is the i-th design point of M_0, with mass $v_i > 0$, $i = 1,...,m$, such that $\sum v_i = 1$, then for each i,

$$F_\varphi(R^{-1}+M_0, R^{-1}+\boldsymbol{x}_i\boldsymbol{x}_i') = 0.$$

The proof of Theorem 1 follows along the lines of the proof in Silvey (1980), pages 19- 23. Condition (ii) of Theorem 1 reduces to condition (ii)′ given below:

(ii)′ If ξ_0 be the optimal design with information matrix $M_0 = \sum_{i=1}^{m} v_i \boldsymbol{x}_i\boldsymbol{x}_i'$, where $\boldsymbol{x}_i\in\Xi$ is the i-th support point of ξ_0 with mass $v_i \ge 0$, $i = 1, \ldots, m$, and $\sum v_i = 1$, then, for any $\boldsymbol{x}\in\Xi$,

$$\boldsymbol{x}'(R^{-1}+M_0)^{-1}\psi(R^{-1}+M_0)^{-1}\boldsymbol{x} \le \boldsymbol{x}_i'(R^{-1}+M_0)^{-1}\psi(R^{-1}+M_0)^{-1}\boldsymbol{x}_i, \quad (6)$$

where $\psi = \boldsymbol{c}\boldsymbol{c}'$. It is obvious that equality in (6) holds at the support points of ξ_0.

Inequality (6) helps to identify the nature of the support points of the optimum design. For single hypothesis testing, the left-hand side of (6) is pseudo-convex in $\boldsymbol{x}$, that is, it behaves like a convex function with respect to finding its local minima, but may not actually be convex. As such, the maximum may be attained at the boundary points of Ξ and also at some non-boundary points. This is true irrespective of the form of R.

3.1. Consider problem I

For simplicity sake, let us take $i = 1$ in the above hypothesis testing, so that $\boldsymbol{c}' = (1, 0, 0,\ldots,0)$. Then a Bayes optimal design minimizes $\phi(M) = \boldsymbol{c}'(R^{-1}+M)^{-1}\boldsymbol{c}$, for the given $\boldsymbol{c}$.

To find a closed form solution, let us assume that the prior matrix R is invariant with respect to $\beta_2, \beta_3, \ldots, \beta_q$, so that R^{-1} can be written

$$R^{-1} = \begin{bmatrix} r^{11} & r^{12}\mathbf{1}'_{q-1} \\ & R^{22} \end{bmatrix},$$

where $R^{22} = u_1 I_{q-1} + u_2 J_{q-1}$, for some scalars u_1 and u_2, $\mathbf{1}_{q-1}$ is a $(q-1)$-vector with all elements 1, and $J_{q-1} = \mathbf{1}_{q-1}\mathbf{1}'_{q-1}$. Then, from Property 2 below, the criterion function ϕ will be invariant with respect to $x_2,\ldots,x_q$.

Property2: If M be invariant with respect to the components of $\boldsymbol{x}_{(1)} = (x_1,\ldots,x_{q_1})$ and $\boldsymbol{x}_{(2)} = (x_{q_1+1},\ldots,x_q)$, and R be invariant with respect to the components of $\boldsymbol{\beta}_{(1)} = (\beta_1,\ldots,\beta_{q_1})$ and $\boldsymbol{\beta}_{(2)} = (\beta_{q_1+1},\ldots,\beta_q)$, then ϕ is invariant with respect to the components of $\boldsymbol{x}_{(1)}$ and $\boldsymbol{x}_{(2)}$.

Proof: Consider P_1 and P_2 to be two permutation matrices of orders q_1 and q_2, respectively $(q_1 + q_2 = q)$. Let,

$$P = \begin{bmatrix} P_1 & \mathrm{O} \\ \mathrm{O} & P_2 \end{bmatrix}.$$

Then, P is a permutation matrix of order q.

Now, if we take the same permutations of the components of both $\boldsymbol{x}_{(1)}$ and $\boldsymbol{\beta}_{(1)}$ and similarly of the components of $\boldsymbol{x}_{(2)}$ and $\boldsymbol{\beta}_{(2)}$, then it is clear that, for the new set, the M and R matrices will reduce to PMP' and PRP'. Therefore, the criterion function will reduce to $\boldsymbol{c}'P'P(R^{-1}+M)^{-1}P'P\boldsymbol{c}$, which is same as $\boldsymbol{c}'(R^{-1}+M)^{-1}\boldsymbol{c}$. This establishes Property 2.

Using properties 1 and 2, we get the following theorem:

Theorem 2: A Bayes-optimal design is invariant with respect to $x_2,...,x_q$.

Let us denote the class of all designs invariant with respect to $(x_2,...,x_q)$ by D_1. For a design $\xi \in \mathsf{D}_1$, $M(\xi)$ is of the form

$$M = \begin{bmatrix} m_{11} & m_{12}\mathbf{1}' \\ & M_{22} \end{bmatrix}, \tag{7}$$

where $M_{22} = g\, I_{q-1} + h\, J_{q-1}$, for some scalars g and h. Then,

$$[\boldsymbol{c}'\,(R^{-1} + M)^{-1}\boldsymbol{c}]^{-1} = (m_{11} + r^{11}) - (m_{12} + r^{12})^2\{\mathbf{1}'(M_{22} + R^{22})^{-1}\mathbf{1}\}. \tag{8}$$

From the structure of R^{22}, it is clear that $(R^{22})^{-1}$ will also be of the form $aI_{q-1} + bJ_{q-1}$, where $J_{q-1} = \mathbf{1}_{q-1}\mathbf{1}'_{q-1}$.

The following theorem indicates the Bayes optimal design under certain restriction.

Theorem 3: The Bayes optimal design is a singular design with only one support point at $(1,0,0,\ldots,0)$ *for* $0 \le |r^{12}|[a + (q-1)b] \le 1$.

Proof: Let ξ_0 be the singular design with one support point at (1,0,0,…,0). Then,

$$M(\xi_0) = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{O} \end{bmatrix},\ [\boldsymbol{c}'\,(R^{-1} + M(\xi_0))^{-1}\boldsymbol{c}]^{-1} = (1 + r^{11}) - (r^{12})^2[\mathbf{1}'(R^{22})^{-1}\mathbf{1}],$$

$$\boldsymbol{x}'(R^{-1} + M(\xi_0))^{-1}\boldsymbol{c} = A[x_1 - \boldsymbol{x}'_{(2)}r^{12}(R^{22})^{-1}\mathbf{1}],$$

where $A = [(1 + r^{11}) - (r^{12})^2\{\mathbf{1}'(R^{22})^{-1}\mathbf{1}\}]^{-1}$, $\boldsymbol{x} = (x_1, x_2, ..., x_q)' = (x_1, \boldsymbol{x}'_{(2)})' \in \Xi$.

For ξ_0 to be Bayes optimal design, it must satisfy (3.2) for all $\boldsymbol{x} \in \Xi$, with equality holding at the support point of ξ_0. This is equivalent to satisfying

$$[x_1 - \boldsymbol{x}'_{(2)}r^{12}(R^{22})^{-1}\mathbf{1}]^2 \le 1, \text{ for all } \boldsymbol{x} \in \Xi, \tag{9}$$

with equality holding at (1,0,0,…,0).

For $r^{12} = 0$, (9) holds trivially. For $r^{12} > 0$, by the condition of the theorem, we have

$$\boldsymbol{x}'_{(2)}r^{12}(R^{22})^{-1}\mathbf{1} = r^{12}[a + (q-1)b]\boldsymbol{x}'_{(2)}\mathbf{1} \le 1,$$

since $\boldsymbol{x}'_{(2)}\mathbf{1} \le 1$. Hence, (9) is satisfied.

For $r^{12} < 0$, we can write

$$\text{l.h.s. of (9)} = [x_1 + \boldsymbol{x}'_{(2)} \mid r^{12} \mid (R^{22})^{-1}\mathbf{1}]^2$$
$$= [1 - \boldsymbol{x}'_{(2)}\{I - \mid r^{12} \mid (R^{22})^{-1}\}\mathbf{1}]^2,$$

Now,

$$1-\boldsymbol{x}'_{(2)}[I-|r^{12}|(R^{22})^{-1}]\mathbf{1} = 1-[1-|r^{12}|\{a+(q-1)b\}]\boldsymbol{x}'_{(2)}\mathbf{1}\leq 1,$$

as $\boldsymbol{x}'_{(2)}\mathbf{1} \leq 1$ and from the condition of the theorem. Hence, (9) is satisfied.

Remark: For $|r^{12}|[a+(q-1)b]>1$, condition (9) is violated for $\boldsymbol{x}$ with $0 \leq x_1 \leq \frac{|r^{12}|[a+(q-1)b]}{1+|r^{12}|[a+(q-1)b]}$, when $r^{12}> 0$, and for all $\boldsymbol{x}$, when $r^{12} < 0$. Hence, the singular design with support point (1,0,…,0) will not be optimal.

It is difficult to analytically find the support points of the Bayes optimal design when $|r^{12}|[a+(q-1)b]>1$. We, therefore, obtain the same through computation in the following example with $q = 3$:

Example 1: Suppose $\sigma = 1$ and the prior covariance matrix of the regression coefficients in Scheffé's first order model for $q= 3$ is

$$\mathrm{R} = \begin{bmatrix} 0.70 & 0.75 & 0.75 \\ 0.75 & 3.00 & 1.00 \\ 0.75 & 1.00 & 3.00 \end{bmatrix}.$$

Consider testing of the hypothesis H_0: $\beta_1 \geq 0$ versus H_1: $\beta_1 < 0$. We have

$$R^{-1} = \begin{bmatrix} 2.38806 & -0.44776 & -0.44776 \\ -0.44776 & 0.458955 & -0.04104 \\ -0.44776 & -0.04104 & 0.458955 \end{bmatrix}$$

and

$$R_{22}^{-1} = \begin{bmatrix} 2.196429 & 0.196429 \\ 0.196429 & 2.196429 \end{bmatrix}$$

Hence, $|r^{12}|[a+(q-1)b] = 1.07 > 1$.

Using MATLAB, we obtain the Bayes optimal design as having support points (1,0,0), (0,1,0) and (0,0,1) with masses 0.94, 0.03 and 0.03, respectively. Thus, even when $|r^{12}|[a+(q-1)b]>1$, the example gives the optimum support points at (1,0,0), (0,1,0) and (0,0,1).

Remark: It is interesting to note that the design with support points (1,0,0) and (0,½,½) gives minimum risk for masses 0.94 and 0.06 at (1,0,0) and (0,½,½) respectively, and the risk is same as that obtained for the optimum design with support points at (1,0,0), (0,1,0) and (0,0,1).

3.2. Consider testing of hypothesis II

For simplicity sake, let us take $i = 1$ and $j = 2$in the above hypothesis testing, so that $\boldsymbol{c}' = (1,-1,0,...,0)$.

We assume that the prior parameter matrix R is invariant with respect toβ_1 andβ_2, and with respect to $\beta_3,\beta_4, \ldots,\beta_q$, so that R^{-1} can be written as

$$R^{-1} = \begin{bmatrix} R^{11} & R^{12} \\ R^{12'} & R^{22} \end{bmatrix},$$

where $R^{ii}=s_i I_2+ t_i J_2$, for some scalars s_i and t_i, $i = 1, 2$, are positive definite matrices, $R^{11}$and R^{22} are of orders 2×2 and $(q-2)\times(q-2)$ respectively, and $R^{12} = r_0\begin{pmatrix}\mathbf{1}'_{q-2}\\ \mathbf{1}'_{q-2}\end{pmatrix}$,where r_0 is a scalar.

From property 2, we have the following:

Property 3: ϕ is invariant with respect to (x_1, x_2) and with respect to $(x_3, x_4,...,x_q)$.

Using properties 1 and 3, we get the following theorem:

Theorem 4: A Bayes-optimal design is invariant with respect to (x_1, x_2), and with respect to $(x_3,...,x_q)$.

Let D_2 denote the class of all designs invariant with respect to (x_1, x_2), and with respect to $(x_3,...,x_q)$.

For a design $\xi \in \mathsf{D}_2$, $M(\xi)$ is of the form

$$M(\xi) = \begin{bmatrix} M_{11} & m_0\begin{pmatrix}\mathbf{1}'_{q-2}\\ \mathbf{1}'_{q-2}\end{pmatrix} \\ m_0\left(\mathbf{1}_{q-2}\ \mathbf{1}_{q-2}\right) & M_{22} \end{bmatrix}, \tag{10}$$

where $M_{11} = g_1 I_2 + h_1 J_2$ and $M_{22} = g_2 I_{q-2} + h_2 J_{q-2}$, for some scalars g_1, g_2, h_1 and h_2.

Then,

$$R^{-1} + M = \begin{bmatrix} (u_1 + g_1)I_2 + (v_1 + h_1)J_2 & [m_0 + r_0]\begin{pmatrix}\mathbf{1}'_{q-2}\\ \mathbf{1}'_{q-2}\end{pmatrix} \\ & (u_2 + g_2)I_2 + (v_2 + h_2)J_2 \end{bmatrix}. \tag{11}$$

Theorem 5: The Bayes optimal design is a singular design with two support points at (1,0,0,…,0) and (0,1,0,…,0), each with mass ½.

Proof: Let ξ_0 be the singular design with two support points at (1,0,0,…,0) and (0,1,0,…,0), each with mass ½. Then,

$$M(\xi_0) = \begin{bmatrix} \frac{1}{2} I_2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{O}_{q\text{-}2} \end{bmatrix},$$

$$\boldsymbol{x}'(R^{-1} + M(\xi_0))^{-1}\boldsymbol{c} = \boldsymbol{x}'_{(1)}A\boldsymbol{c} + \boldsymbol{x}'_{(2)}B\boldsymbol{c},$$

where $\boldsymbol{x} = (\boldsymbol{x}'_{(1)}, \boldsymbol{x}'_{(2)})'$, $\boldsymbol{x}'_{(1)} = (x_1, .x_2)$, $\boldsymbol{x}'_{(2)} = (x_3, ..., x_q)$, and A and B are 2×2 and (q-2)×2 matrices, respectively given by

$$A = [(R^{11} + \frac{1}{2})I_2) - R^{12}(R^{22})^{-1}R^{21}]^{-1} = [(u_1 + \frac{1}{2})I_2 + v_1 J_2 - r_0^2 \begin{pmatrix} 1'_{q-2} \\ 1'_{q-2} \end{pmatrix} (R^{22})^{-1}(1_{q-2}\ 1_{q-2})]^{-1}$$

$$= \frac{1}{(u_1 + 1/2)} \left[I_2 - \frac{v_1 - r_0^2 1'_{q-2}\{R^{22}\}^{-1}1_{q-2}}{u_1 + 1/2 + 2[v_1 - r_0^2 1'_{q-2}\{R^{22}\}^{-1}1_{q-2}]} J_2 \right]$$

$$B = -[R^{22} - R^{21}(R^{11})^{-1}R^{12}]^{-1}R^{21}(R^{11})^{-1} = -\frac{1}{u_2}\left[I_{q-2} - \frac{v_2 - r_0^2 1'_2\{R^{11}\}^{-1}1_2}{u_2 + (q-2)[v_2 - r_0^2 1'_2\{R^{11}\}^{-1}1_2]} J_{q-2} \right]$$

$$\times R^{21}(R^{11})^{-1}.$$

For ξ_0 to be Bayes optimal design, it must satisfy (6) for all $\boldsymbol{x} \in \Xi$, with equality holding at the support point of ξ_0. Writing $\boldsymbol{c} = (\boldsymbol{c}'_{(1)}\ \ \mathbf{0}')'$, where $\boldsymbol{c}_{(1)} = (1\text{-}1)'$, and noting that $\mathbf{1}_2'\boldsymbol{c}_{(1)} = 0$, we have, after a little algebraic manipulation, that (6) is equivalent to

$$(x_1 - x_2)^2 \le 1, \text{ for all } \boldsymbol{x} \in \Xi, \tag{12}$$

with equality holding at (1,0,0,…,0) and (0,1,0,…,0).

Clearly, (12) holds for all $\boldsymbol{x} \in \Xi$, with equality at the support points of ξ_0. Thus, ξ_0 is the Bayes optimal design with

$$\phi(\xi_0) = \boldsymbol{c}'(R^{-1} + M(\xi_0))^{-1}\boldsymbol{c} = \boldsymbol{c}'_{(1)}A\boldsymbol{c}_{(1)} = \frac{2}{u_1 + 1/2}.$$

3.3. Consider testing of hypothesis III

The problem of finding optimum design for a general '$\boldsymbol{c}$' in a closed form seems difficult. For this, to find the optimum design, we have considered two specific choices of $\boldsymbol{c}$, namely, (i) $\boldsymbol{c}'$ = (1, 1, …, 1), and (ii) $\boldsymbol{c}' = (c_1, c_2, ..., c_q)$, where c_i = –1 or +1, such that $\sum_{i=1}^{q} c_i = 0$, that is, q is even with $q/2$ of the c_i's equal to +1 and the rest –1. .

3.3.1. Let $\boldsymbol{c}' = (1, 1, \ldots, 1)$.

Since the hypothesis is invariant with respect to the q components, to start with, let us assume that R also has the same invariance property, that is, $R = a\,I + b\,J$, for some scalars a, b, where I and J stand for an identity matrix and a matrix of 1's of appropriate order, respectively. This means that the prior dispersion matrix of $\boldsymbol{\beta}$ is invariant with respect to the coefficients β_i's. Then it is easy to check the following invariance property of the criterion function $\phi\{M(\xi)\}$.

Property 4: $\phi\{M(\xi)\}$ is invariant with respect to the permutation of the components of the mixture.

Because of the properties (1) and (4), we get the following Theorem:

Theorem 6: A Bayes-optimal design is necessarily invariant.

Thus, in view of Theorem 6, we can confine our search for the Bayes-optimal design within the class of ***invariant designs***.

There are three ways out to find the desired design:

(i) Express M in terms of the two moments $\mu_2 = E(x_1^2)$ and $\mu_{11} = E(x_1 x_2)$ of the design, and then show that ϕ is decreasing in μ_{11}.

(ii) Use Lowener Order dominance to find ξ^* such that $M(\xi^*) \geq M(\xi)$ for every invariant design ξ [*cf.* Draper and Pukelsheim (1999)].

(iii) Use Equivalence Theorem to find the optimal design.

Approaches (i) and (ii) fail as soon as the complete symmetry property of the matrix R is violated. In general, it is difficult to find a closed form solution to the problem. However, it may be possible to indicate the nature of the support points of the optimal design using (iii). The following theorem identifies an optimal design satisfying the Equivalence Theorem.

Theorem 7: The Bayes optimal design is a saturated design with support points at (1,0,0,…,0) and its permutations, each with mass $1/q$.

Proof: Let ξ_0 be the saturated design with support points at (1,0,0,…,0), (0,1,0,…,0), …, (0,0,…,1), each with mass $1/q$. Then, $M(\xi_0) = \frac{1}{q} I_q$.

Since both R and M have the complete symmetry property, $(R^{-1} + M(\xi_0))^{-1}$ is also completely symmetric and is of the form $eI_q + fJ_q$, with $e + q\,f > 0$. Then,

$$\boldsymbol{x}'(R^{-1} + M(\xi_0))^{-1}\boldsymbol{c} = \boldsymbol{x}'[eI_q + fJ_q]\mathbf{1}_q = e + qf,$$

since $J_q = \mathbf{1}_q\mathbf{1}_q'$ and $\boldsymbol{x}'\mathbf{1}_q = 1$.

Thus, $\boldsymbol{x}'(R^{-1} + M(\xi_0))^{-1}\boldsymbol{c}$ is a constant, independent of $\boldsymbol{x}$, and therefore (6) is satisfied for all $\boldsymbol{x} \in \Xi$.

Hence, ξ_0 satisfies the Equivalence Theorem, and is, therefore, a Bayes optimal design.

3.1.2. Consider $q = 2k$, k a positive integer, and $\boldsymbol{c} = (c_1, c_2, ..., c_q)'$, with k of the $c_i = -1$ and remaining 1, so that $\sum_{i=1}^{q} c_i = 0$.

Let R be complete symmetric. Intuitionally, we feel that M will also be complete symmetric.

We start with a saturated design ξ_0 which has support points at the extreme points of **X,** each having mass $1/q$. Then, as before, $M(\xi_0) = \frac{1}{q} I_q$.

As $R^{-1}$is a complete symmetric and positive definite matrix, $(R^{-1} + M(\xi_0))^{-1}$will have the form $eI_q + fJ_q$, with $e > 0$, $e + qf > 0$. Then,

$$\boldsymbol{x}'(R^{-1} + M(\xi_0))^{-1}\boldsymbol{c} = \boldsymbol{x}'[eI_q + fJ_q]\boldsymbol{c} = e\,\boldsymbol{x}'\boldsymbol{c},$$

since $\boldsymbol{x}'\mathbf{1}_q = 1$ and $\mathbf{1}_q'\boldsymbol{c} = 0$.

Hence, for each support point of ξ_0, r.h.s. of (6) $= e^2$. The l.h.s. of (6) is $e\,\boldsymbol{x}'\boldsymbol{c}$ for all $\boldsymbol{x} \in \Xi$**,** which is clearly $\le e^2$, and equality holds at the support points of ξ_0.

Hence, ξ_0 is Bayes optimal.

4. Optimal Mixture Design for Multiple Tests

For the first-degree mixture model (3), one may be interested in multiple tests of the form

H_{0i}: $\beta_i \ge 0$ versus H_{1i}: $\beta_i < 0$; $i = 1,2, \ldots, k$, $k \le q$.

For a given design ξ or the corresponding moment matrix $M(\xi)$, let $r_i(\delta^i, M)$denote the Bayes risk for the i-th hypothesis, $i = 1, 2, \ldots, k$. So, now we have a vector of Bayes risks. In order to define a partial ordering of the designs in terms of the moment matrices (*cf.* Kiefer, 1959), we proceed as in Toman (1996), who uses an idea analogous to the classical decision theoretic concept of admissibility. Admissibility is defined with respect of the risk function in the classical decision theory, and not the Bayes risk. In the present case, the index i of the Bayes risk is treated as the parameter in classical risk.

The following definitions are due to Toman (1996).

Definition 1: A design ξ_1 is said to be r-better than design ξ_2 if $r_i(\delta^i, M(\xi_1)) \le r_i(\delta^i, M(\xi_2))$ for $i = 1,2, \ldots, k$, with strict inequality for at least one i.

Definition 2: A design ξ is said to be r-admissible if there exists no r-better design.

From the above it is quite clear that r-admissibility is a desired property of any design.

Toman (1996) indicated two methods of determining *r*-admissible design under multiple optimality criteria. In the present set-up, they are as follows:

I. Minimize $r_k(\delta^k, M)$ subject to $r_i(\delta^i, M) \le a_i$, $i = 1, 2, \ldots, k-1$, where the index k and the scalars a_i, $i = 1,2,.., k-1$ are determined by the experimenter.

II. Define a single risk function by combining the k Bayes risks $r_i(\delta^i, M)$, $i = 1, 2, \ldots, k$ as follows:

$$r(\delta, M) = \sum_{i=1}^{k} w_i r_i(\delta^i, M), \tag{13}$$

where the weights $\{w_i\}$ represent a discrete probability measure on the index i so that $w_i \ge 0$, for $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} w_i = 1$. The weights represent the relative importance of the corresponding decision problems. A design ξ with moment matrix $M(\xi)$ which minimizes (13) will be the *r*-admissible design (*cf*. Toman, 1999).

Method I can be equivalently written as:

Minimize $\boldsymbol{c}_k'(R^{-1}+M)^{-1}\boldsymbol{c}_k$, subject to $\boldsymbol{c}_i'(R^{-1}+M)^{-1}\boldsymbol{c}_i \le b_i$, $i = 1, 2, \ldots, k-1$,

where b_i is some function of a_i. This is because, for each i, $r_i(\delta^i, M)$, given in (1), is an increasing function of $p_i = \dfrac{\mathrm{c}_i'(R^{-1}+M)^{-1}c_i}{\mathrm{c}_i' R c_i}$, and hence of $\boldsymbol{c}_i'(R^{-1}+M)^{-1}\boldsymbol{c}_i$. Thus, we have a constrained optimization problem, which yields an *r*-optimal design (*cf*. Theorem 5 in Toman, 1996).

Lemma 1below shows the equivalence of Methods I and II for some set of weights $\{w_i\}$:

Lemma 1: If M_0 minimizes the combined risk $r(\delta, M)$, given in (13), then it also minimizes $r_k(\delta^k, M)$, subject to the restrictions $r_i(\delta^i, M) \le a_{i0}$, for $i = 1, 2, \ldots, k-1$, where $a_{i0} = r_i(\delta^i, M_0), i = 1,2,...,k-1.$

The lemma is a consequence of the following lemma of Cook and Wong (1994):

Lemma 2: (Cook and Wong, 1994): For $\lambda \in (0,1)$, let ξ_λ maximize the functional $\phi(\xi \mid \lambda) = \lambda\phi_1(\xi) + (1-\lambda)\phi_2(\xi)$, and let $c_\lambda = \phi_1(\xi_\lambda)$, the primary design criterion evaluated at ξ_λ. Then ξ_λ maximizes $\phi_2(\xi)$ subject to the constraint $\phi_1(\xi) \ge c_\lambda$.

The Bayes risk $r(\delta, M)$ depends on the design only through the p_is, which give the ratios of the posterior and prior variances. Further, for any given $\boldsymbol{c}_i$, $p_i \to 0$ as the prior information matrix $R^{-1} \to 0$, provided M is nonsingular.

Approximating the Bayes risks $r_i(\delta^i, M_0), i = 1,2,...,k,$ by a first-order Taylor series expansion around $p_i = 0$, we get

$$r_i(\delta^i, M_0) \cong L_i \boldsymbol{c}_i' (M + R^{-1})^{-1} \boldsymbol{c}_i,$$

where $\varphi(.)$ denotes the standard normal density, and

$$s_i = \frac{v_i - \boldsymbol{c}_i' \boldsymbol{\tau}}{\sigma \sqrt{\boldsymbol{c}_i' R \boldsymbol{c}_i}}, \quad L_i = \frac{\sigma}{2\boldsymbol{c}_i' R \boldsymbol{c}_i} \phi(s_i), i = 1,2,...,k.$$

Then, $r(\delta, M) = \sum_{i=1}^{k} w_i L_i \mathrm{c}_i' (M + R^{-1})^{-1} \mathrm{c}_i = \mathrm{Trace}[B(M + R^{-1})^{-1}],$ (14)

where $B = \sum_{i=1}^{k} w_i L_i \boldsymbol{c}_i \boldsymbol{c}_i'$.

It seems easier to study the problem of multiple hypotheses testing using the second criterion. Some examples are worked out in Section 5.

5. Examples of Multiple Hypotheses Testing

In this section we find Bayes optimal designs in two examples on multiple hypotheses testing.

Example 2: Consider Scheffé's homogeneous mixture model of first degree for a three-component mixture:

$$\eta_{\mathrm{x}} = E(Y \mid \mathrm{x}) = \sum_{i=1}^{3} \beta_i x_i,$$

Suppose σ = 1, and the prior mean and covariance matrix of the regression coefficients are

$$\tau = (5,5,0.1)', \; R = \begin{bmatrix} 5.00 & 0.25 & 0.25 \\ 0.25 & 5.00 & 0.25 \\ 0.25 & 0.25 & 5.00 \end{bmatrix}.$$

(a) Consider testing of the hypotheses H_{0i}: $\beta_i \geq 0$ against H_{Ai}: $\beta_i < 0$, for i = 1, 2, 3.

Using Method 2, we minimize Trace $B(M(\xi) + R^{-1})^{-1}$, given by (14). In the absence of any knowledge about the relative importance of the components, it may be assumed that component problems are equally important. We, therefore, take, $w_i = \frac{1}{3}$, for i = 1, 2, 3.

From the given data, we obtain, $s_1 = s_2 = -2.24$ and $s_3 = -0.045$. Hence, $L_1 = L_2 = 0.007, L_3 = 0.089,$ and $B = \mathrm{Diag}\,(0.007,0.007,0.089)$.

We restrict to the class of saturated designs. Within this class, we get min[Trace $B(M(\xi) + R^{-1})^{-1}$] = 0.045135, which is obtained for a design with mass 0.0874 at each of the support

points (1,0,0) and (0,1,0), and mass 0.8252 at (0,0,1). Comparing this design with an alternative one, say a design which puts equal masses at its three design points, we get the Bayes risk as 0.064301, which is 14.2% more than the optimum Bayes risk.

(b) Now, suppose we are interested to test the hypotheses

$$H_{01}: \beta_1 - \beta_3 \geq 0 \text{ against } H_{A1}: \beta_1 - \beta_3 < 0$$
$$H_{02}: \beta_2 - \beta_3 \geq 0 \text{ against } H_{A2}: \beta_2 - \beta_3 < 0.$$

Here, $s_1 = s_2 = -1.58977$, $L_1 = L_2 = 0.005934$. Then,

$$B = 0.002867 \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

Restricting to the class of saturated designs, the optimum design puts mass 0.272 at each of the points (1,0,0) and (0,1,0), and 0.456 at the point (0,0,1), and the Bayes risk is 0.0205.

6. Discussion

This paper attempts to find Bayes optimal designs for testing of single and multiple hypotheses in Scheffé's homogeneous first-degree mixture model. Interestingly, under the hypotheses considered, the support points of the optimal designs are found to be at one or more of the extreme points of the experimental region. The study can be extended to other testing situations, and also to the cases of quadratic and cubic mixture models.

Acknowledgement

The authors thank the anonymous reviewer for reading the manuscript with great care and making fruitful suggestions, which immensely helped to improve the presentation.

References

Chaloner, K. (1982).*Optimal Bayesian Experimental Design for Linear Models*. Technical Report 238, Carnegie-Mellon University, Department of Statistics.

Chaloner, K.(1984). Optimal Bayesian experimental design for linear models. *Annals of Statistics*, **12(1)**, 283-300.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, **10**, 273-304.

DasGupta, A. (1996). Review of optimal Bayes designs. *Handbook of Statistics*, **13**, 1099-1147.

Draper, N. R. and Pukelsheim, F. (1999). Kiefer ordering of simplex designs for first- and second-degree mixture models. *Journal of Statistical Planning and Inference*, **78**, 325-348.

El-Krunz, S. and Studden, W. J. (1991). Bayesian optimal designs for linear regression models. *The Annals of Statistics*, **19**, 2183-2208.

Fedorov, V. V.(1972). *Theory of Optimal Experiments*. Academic Press. New York.
Scheffé, H. (1958). Experiments with mixtures. *Journal of the Royal Statistical Society*, **B20**, 344-360.
Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.
Toman, B. (1996). Bayesian experimental design for multiple hypothesis testing. *Journal of the American Statistical Association*, **91**, 185-190.
Whittle, P. (1973). Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society*, **B35**(**1**), 123-130.

Statistics and Applications {ISSN 2452-7395 (online)}
Volume 18, No. 2, 2020 (New Series), pp 223-248

A Comparative Study of Approximate Bayesian Computation Methods for Gibbs Point Processes

Jiaxun Chen[1], **Athanasios C. Micheas**[1] **and Scott H. Holan**[1,2]
[1] *Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100*
[2] *U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100*

Received 11 April 2020; Revised 28 July 2020; Accepted 30 July 2020

Abstract

Crucial aspects of applying approximate Bayesian computation (ABC) for Gibbs point processes are the choice of summary statistic and method of constructing the discrepancy measure. In this paper, we present a comparative study of ABC for Gibbs point processes based on various summary statistics and different approaches of constructing the discrepancy measure. We also demonstrate the issue of identifiability of the parameter values for Gibbs point processes and provide a solution for parameter estimation. We further propose robust choices for the discrepancy measures for different point processes through an intensive simulation study. The ABC algorithm, with all of the tested discrepancy measures, is also applied to the Swedish pines data and Chicago crime data to illustrate the feasibility of the proposed approaches.

Key words: Approximate Bayesian computation; Discrepancy measure; Functional characteristic; Gibbs point processes; Informative prior.

1. Introduction

As an important class of spatial point processes, Gibbs point processes have been intensively studied over the past few decades. Since Gibbs point processes can take into account interactions between event locations, they have become reasonable choices for describing this kind of phenomena, *i.e.*, the inhibition and clustering behaviors in point processes. Here, we refer to point processes with inhibition as repulsive spatial point processes and to point processes with clustering as attractive spatial point processes. In the literature, the most well known models for repulsive point processes are the Strauss process (Strauss, 1975) and hardcore process (Ripley, 1981). For attractive point processes, a model that can be used to describe this behavior is the area-interaction process, proposed by Baddeley and Van Lieshout (1995).

Although Gibbs point processes meet the need for describing the underlying process in different fields of science, conducting inference directly using the likelihood of the point process is challenging, and standard Bayesian analysis based on Markov chain Monte Carlo (MCMC) is not feasible due to the intractable normalizing constant. Maximum pseudo-likelihood estimation (MPLE) (Besag, 1975) can be applied to spatial point processes. However, Huang and Ogata (1999) show that the performance of MPLE is

Corresponding Author: Scott H. Holan
E-mail: holans@missouri.edu

poor for repulsive point processes with strong interaction. There are also several estimation methods that have been developed in a Bayesian framework. Geyer (1991) proposed the Markov chain Monte Carlo maximum likelihood estimate (MCMCMLE) for stochastic processes with an intractable normalizing constant. One of the issues with MCMCMLE is that the required computational effort within each MCMC iteration is intensive for Gibbs point processes. Moreover, the estimation result is sensitive to the auxiliary parameters. By implementing an auxiliary variable, Møller *et al.* (2006) proposed an auxiliary variable Markov chain Monte Carlo (AVM) and an exchange algorithm approach, which can cancel out the intractable normalizing constant in the acceptance probability ratio of the Metropolis-Hastings algorithm. Extending the exchange algorithm through "bridging" was proposed by Murray *et al.* (2012).

In addition, the double Metropolis-Hastings (DMH) and an adaptive exchange (AEX) algorithm were proposed by Liang (2010) and Liang *et al.* (2016). Alternatively, approaches based on likelihood approximation were proposed by Beaumont (2003), Andrieu *et al.* (2009), Alquier *et al.* (2016), and Atchadé *et al.* (2013). A comparison of these methods was provided by Park and Haran (2018). The overall conclusion from this comparison recommends that researchers start with DMH since it is computationally efficient and does not require perfect sampling. However, the convergence of DMH cannot be guaranteed, and the auxiliary variable approaches become computationally expensive when the inner sampler for the auxiliary variable is expensive. Recently, Park and Haran (2018) focused on comparing likelihood based sampling methods and provided practical recommendations. In contrast, our study provides evaluations of a likelihood free algorithm and its applications (see details in Section 3).

As an alternative to the approaches based on classic MCMC, Stoica *et al.* (2017) and Shirota and Gelfand (2017) proposed an approximate Bayesian computation (ABC) algorithm for Gibbs point processes. The ABC algorithm provides a likelihood free approach for approximating the posterior distribution and it is straightforward to implement. In Shirota and Gelfand (2017), the simulation results for the Strauss model and the determinantal point process show that the true models can be recovered. Also, the authors compare the ABC with the exchange algorithm of Murray *et al.* (2012) and point out that the ABC is more efficient in terms of the parameter inefficiency factors (IF).

In this study, we propose a more robust approach to constructing discrepancy measures for the ABC and compare the performance of available summary statistics for both repulsive and attractive point processes. Implementing the ABC algorithm for fitting Gibbs point processes has been proven feasible and efficient. However, we still need to thoroughly investigate how to construct the best discrepancy measure for different situations. Inspired by the discussion surrounding the choices of various distance metrics between two point process realizations (Mateu *et al.*, 2010), we use the integrated distance of the available functional summary characteristics to measure the similarity of the interactive structure for two point patterns. This measure takes into account the information of the interactions among all the reasonable scales, which can be more informative compared with the measures only calculated at a given interaction distance.

An important component of the ABC algorithms for Gibbs point processes is the choice of summary statistic. In Stoica *et al.* (2017), the authors discuss the issue, but did

not conduct a comparative study, especially when the sufficient statistics are not available. In our study, we implement and compare the performance of Ripley's K-function, nearest neighbor distance distribution function (D-function), empty-space function (F-function), and J-function (Van Lieshout and Baddeley, 1996) when used in the ABC algorithm, in terms of recovering the parameters of a Gibbs point process. The final recommendation on the most robust summary statistic among the ones tested is formed based on an extensive simulation study. Note that Geyer (1998) proposed to use stochastic approximation as rough estimates of the parameter values. Thus, we implement this estimate as an informative prior and compare it with a vague prior for the ABC algorithms presented herein.

The paper proceeds as follows. Section 2 begins by reviewing the repulsive and attractive Gibbs point processes that we investigate. Section 3 provides the functional summary statistics considered for Gibbs point processes, the proposed structure of the discrepancy measure for the ABC algorithm, and the choices of prior distributions. Sections 4 and 5 include the results of the simulation study, and Section 6 presents applications to the Swedish pines data and Chicago crime data. Finally, in Section 7, we provide concluding remarks.

2. Review of Classic Gibbs Point Processes

The general density function of a Gibbs point process for a finite point pattern $\varphi_n = \{x_1, \ldots, x_n\}$ over a bounded domain $B \in \mathbb{R}^2$ (Cressie, 1993) is given by

$$f(\varphi_n) = c^{-1} \exp\left\{ \sum_{i-1}^{n} g_1(x_i) + \sum_{1 \le i < j \le n} g_{1,2}(x_i, x_j) + \cdots + g_{1,\ldots,n}(x_1, \ldots, x_n) \right\} \tag{1}$$

with respect to the Poisson process with unit intensity, where the function $g_{1,\ldots,k}(\cdot)$ describes the k-level interaction. Hence, c is a normalizing constant and the form of c cannot be provided analytically. In other words, the normalizing constant for the Gibbs point process is typically intractable. Another fundamental functional is the Papangelou conditional intensity,

$$\lambda(u|\varphi_n, \boldsymbol{\theta}) = \frac{f(\varphi_n \cup u|\boldsymbol{\theta})}{f(\varphi_n|\boldsymbol{\theta})},$$

which is the intensity at location u given the point pattern φ_n. Note that the normalizing constant in the likelihood function cancels out when calculating the Papangelou conditional intensity. Thus, it is used to construct the log pseudo-likelihood as follows

$$\log PL(\varphi_n|\boldsymbol{\theta}) = -\int_B \lambda(u|\varphi_n, \boldsymbol{\theta}) du + \sum_{i}^{n} \log\lambda(x_i|\varphi_n, \boldsymbol{\theta}), \tag{2}$$

and the MPLE is obtained by maximizing the latter equation.

In this paper, we consider point processes that are suitable for describing inhibition and attraction. Thus, two classes of Gibbs point processes are investigated: the homogeneous Strauss process and the area-interaction point process.

2.1. Strauss point process

The Strauss process was introduced by Strauss (1975) to model point processes with pairwise interactions; however, Kelly and Ripley (1976) showed that the Strauss process density is only integrable for repulsive interactions. Hence, it is extensively used to model point processes under the regularity assumption. The density function of the homogeneous Strauss model for $\varphi_n = \{x_1, \ldots, x_n\}$ over a bounded domain $B \in \mathbb{R}^2$ is given by

$$f(\varphi_n) = c(\beta, \gamma)^{-1} \beta^n \gamma^{S_r(\varphi_n)}, \varphi_n \in B^n, \tag{3}$$
$$S_r(\varphi_n) = \sum_{1 \le i < j \le n} \mathbb{I}(||x_i - x_j|| \le r),$$

where $\beta > 0$, $\gamma \in [0, 1]$, $c(\beta, \gamma)$ is the normalizing constant, r is the interaction distance and $S_r(\varphi_n)$ is the number of pairs of points that are closer than a distance r. In the Strauss process, β and γ represent the main effect and the interaction effect, respectively. For $\gamma \in (0, 1)$, the point process shows inhibition between points and smaller γ leads to stronger interaction. For $\gamma = 0$, the process is known as the hardcore process which does not allow points to be closer than distance r. For $\gamma = 1$, the process is equivalent to the homogeneous Poisson process with intensity β.

2.2. Area-interaction point process

The area-interaction process (Baddeley and Van Lieshout, 1995) can be used to model point processes with attraction or repulsion. The density function of the homogeneous area-interaction process for $\varphi_n = \{x_1, \ldots, x_n\}$ in B is defined as follows

$$f(\varphi_n) = c(\beta, \gamma)^{-1} \beta^n \gamma^{-m(U_r(\varphi_n))},$$

with respect to the Poisson process with unit intensity, where $\beta > 0$, $\gamma > 0$, $c(\beta, \gamma)$ is the normalizing constant, m denotes the Lebesgue measure and

$$U_r(\varphi_n) = \cup_{i=1}^n B(x_i, r),$$

is the union of discs with radius r centered at the points contained in the point pattern. Similar to the Strauss process, γ controls interaction between points. The area-interaction process generates repulsive point patterns if $\gamma \in (0, 1)$ and clustered point patterns if $\gamma > 1$. For $\gamma = 1$, this process is also equivalent to the Poisson point process with intensity β.

In our study, we implement a canonical scale-free version of the area-interaction process proposed by Baddeley and Turner (2014), since the interpretation is easier. The density function is

$$f(\varphi_n) = c(\kappa, \eta)^{-1} \kappa^n \eta^{-A(\varphi_n)},$$

where κ is the new main effect, η is the new interaction effect and

$$A(\varphi_n) = m(U_r(\varphi_n))/(\pi r^2) - n.$$

In this way, each isolated disc has unit area and contributes a factor κ to the density. Parameter η has the same interpretation as the original parameterization. Notice that the parameter r denotes the disc radius at each point but not the interaction distance. Thus, for the area-interaction process, the interaction distance is $2r$.

3. Approximate Bayesian Computation for Gibbs Point Processes

The ABC algorithm has been widely used for estimating model parameters with an intractable normalizing constant. The fundamental algorithm is known as ABC rejection sampling that was implemented by Pritchard *et al.* (1999) and Sunnåker *et al.* (2013). The specific pseudo-code associated with ABC rejection sampling is shown in Algorithm 1, where φ_{obs} is the observed data, φ_{sim} is the simulated data, $\mathbf{T}(\varphi_n)$ is the summary statistic and $d(\cdot,\cdot)$ is the discrepancy measure that is used to evaluate the similarity between the data. Moreover, $\epsilon > 0$ is the acceptance threshold.

Algorithm 1 ABC rejection sampling

```
for i = 1 to L do
  repeat
    Sample θ* from its prior, θ* ~ π(θ).
    Generate simulated data, φ_sim ~ f(φ|θ*).
  until d(T(φ_obs), T(φ_sim)) ≤ ε
end for
```

Based on Algorithm 1, the proposed θ will be accepted if the discrepancy measure is smaller than the pre-specified threshold, which implies that the data simulated based on θ is similar to the observed one. In this way, the accepted θ is generated from the approximated posterior distribution, since the simulated data approximate the realizations from the observed likelihood. More importantly, the choices of discrepancy measures and summary statistics can determine the efficiency and unbiasedness of the approximation of the posterior distribution.

In order to improve the efficiency of the ABC rejection sampling algorithm, Marjoram *et al.* (2003) proposed the ABC-MCMC, which combines the MCMC with the ABC. This algorithm was applied to repulsive spatial point processes by Shirota and Gelfand (2017). However, the choice of proposal kernel can significantly affect the efficiency of the ABC algorithm in terms of the level of mixing. In addition, choosing the most appropriate proposal kernel is a challenging issue. Thus, we implement the ABC rejection sampling algorithm for our study in order to investigate the performance of different summary statistics, prior distributions, and discrepancy measures.

3.1. Summary Statistics

For homogeneous Gibbs point processes, it is reasonable to evaluate the similarity of point patterns by comparing summary statistics for the main and interaction effects. We use the number of points as a summary statistic of the main effect, since it is also a function of the sufficient statistic. In the literature, several useful functional summary characteristics can be used for describing the interaction effect, *e.g.*, Ripley's K-function, D-function (the nearest neighbor distance distribution function), F-function (the empty space function), and the J-function (Van Lieshout and Baddeley, 1996). The definition of each function for a stationary point process with intensity λ is as follows:

K-function

$$K(r) = \mathbf{E}_o(N(b(o,r) \setminus \{o\}))/\lambda,$$

with $N(b(o,r) \setminus \{o\})$ representing the number of points found within the distance r from the typical point o.

D-function

$$D(r) = \mathbf{P}_o(N(b(o,r) \setminus \{o\}) > 0),$$

where $\mathbf{P}_o$ represents the Palm probability, which is the conditional probability that the point process has the property $N(b(o,r) \setminus \{o\}) > 0$ given that a point of the process is in o (see Illian *et al.* (2008), Page 178).

F-function

$$F(r) = 1 - \mathbf{P}(N(b(o,r)) = 0),$$

where $\mathbf{P}$ represents the distribution for a given point process.

J-function

$$J(r) = \frac{1 - D(r)}{1 - F(r)}.$$

Illian *et al.* (2008) point out that these summary characteristics play an important role in the analysis of point processes, and that each of them has specific advantages that can reveal unique information about the point pattern. Hence, we implement the functions mentioned above and compare their performance within the ABC algorithm. In order to obtain estimates of the various functions, we use the empirical estimators with edge correction using the R package "spatstat" (Baddeley and Turner, 2014).

3.2. Prior distributions

The prior distribution in the ABC is typically used as the proposal kernel or part of the Metropolis-Hastings acceptance probability. Here, we discuss various choices for the prior distribution for the ABC algorithm. Simulation studies detailing their performance are included in Sections 4 and 5.

A common choice of prior distribution for the ABC is a vague prior over a reasonable parameter space, such as a uniform prior with a pre-specified range. However, the choice of parameter space is subjective without strong prior knowledge. Additionally, a uniform prior with a wide parameter space can significantly reduce the efficiency of the ABC algorithm.

Alternatively, a more informative distribution can be applied as a prior for the ABC to improve estimation and efficiency. Geyer (1998) proposed using the stochastic approximation to obtain the starting point for more sophisticated methods. In this context, the estimates provided by the stochastic approximation form a consistent estimator for the model parameters. Thus, we use the estimate from the stochastic approximation as the mean in an informative prior that has an associated large variance. Specifically, we implement the Robbins-Monro (R-M) algorithm (Robbins and Monro, 1951) for Gibbs

point processes. Given the canonical form of the homogeneous Gibbs point process, the model is as follows

$$f(\varphi_n) = \exp\{\boldsymbol{\theta}'\mathbf{T}(\varphi_n) - \log c(\boldsymbol{\theta})\},$$

and the R-M algorithm generates a sequence of estimates $\boldsymbol{\theta}_k, k = 1, 2, \ldots,$ given by

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \frac{A}{k}\{\mathbf{T}(\varphi_{obs}) - \mathbf{T}(\varphi_k)\},$$

where φ_k is a realization generated from the specified Gibbs process given parameters $\boldsymbol{\theta}_k$. We also implement an objective stopping rule proposed by Gu and Zhu (2001),

$$K_1 = \inf\left\{K \geq K_0 : \left\|\sum_{k=K-K_0+1}^{K} \text{sgn}(\theta_{j,k} - \theta_{j,k-1})/K_0\right\| \leq \delta, \forall j = 1, \ldots, m\right\},$$

where K_0 and δ are the pre-specified length of the subsequence of $\boldsymbol{\theta}_k$ and required precision level in order to calculate the stopping rule and $\text{sgn}(x)$ is the sign function, which equals to 1,0 or -1 depending on the sign of x. Also, A is preassigned to ensure that the step function is slowly decreasing. It is difficult to tune these parameters, and iterative adjustment can introduce a computational burden. However, since we only need the sequence to converge to an area that is close to the true value, any reasonable choice of tuning parameters should yield a satisfactory result.

3.3. Construction of the discrepancy measure

We briefly review the details of constructing discrepancy measures and propose several measures that emerge as more robust in ABC estimation. Shirota and Gelfand (2017) implement the approach of constructing discrepancy measures based on linear regression for repulsive point processes (Fearnhead and Prangle, 2012). First, $h(\varphi_u, \varphi_v)$ is defined as a vector of functions of summary statistics for two point patterns, and typically, the length of the vector $h(\varphi_u, \varphi_v)$ is the same as the number of parameters in the specified model. Shirota and Gelfand (2017) used the difference of the number of points and the squared difference of the estimated variance stabilized K-function at a given interaction distance to construct the vector of functions. In contrast, Mateu *et al.* (2010) proposed to construct these functions based on functional summaries over a certain range of interaction $[0, r_{max}]$, *e.g.*, the integrated squared difference between estimates of Ripley's K-function for two point patterns. Inspired by both approaches, we use the difference of the number of points and the integrated absolute difference between estimated functional summaries of the two point patterns to propose and construct our general discrepancy measure, *i.e.*,

$$\begin{aligned} h_1(\varphi_{obs}, \varphi_{sim}) &= n(\varphi_{sim}) - n(\varphi_{obs}), \\ h_2(\varphi_{obs}, \varphi_{sim}) &= \int_0^{r_{max}} |\mathcal{F}_{sim}(r) - \mathcal{F}_{obs}(r)| dr, \end{aligned}$$

where $\mathcal{F}$ represents the estimated summarizing functional, such as those mentioned in Section 3.1 and $h_2(\varphi_{obs}, \varphi_{sim})$ is approximated numerically.

For comparison purposes, following the approach of Shirota and Gelfand (2017), one of the discrepancy measures implemented in our study for a parameter vector of length m is,

$$d_{reg}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim})) = \sum_{j}^{m} (\hat{\xi}_{l,j} - \hat{\xi}_{obs,j})^2 / \widehat{\text{var}}(\hat{\xi}_j), \tag{4}$$

and

$$\hat{\boldsymbol{\xi}}_l = \hat{\mathbf{a}} + \hat{\mathbf{b}} h(\varphi_{sim}, \varphi_{obs}),$$
$$\hat{\boldsymbol{\xi}}_{obs} = \hat{\mathbf{a}} + \hat{\mathbf{b}} h(\varphi_{obs}, \varphi_{obs}) = \hat{\mathbf{a}},$$

where $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ are obtained by fitting linear regression for the pilot run. Moreover, $\widehat{\text{var}}(\hat{\xi}_j)$ is the sample variance of the j-th component of $\hat{\boldsymbol{\xi}}$ and it is also provided by the pilot run. Notice that for Gibbs point processes model, $\boldsymbol{\xi}$ is the log transformed parameter vector. Specifically, $\boldsymbol{\xi}$ is $(\log\beta, \log\gamma)$ for the Strauss process and $(\log\kappa, \log\eta)$ for the area-interaction process.

We further propose and implement another way of constructing the discrepancy measure by rescaling the function $h(\varphi_u, \varphi_v)$ based on the pilot run. The form of the discrepancy measure is

$$d_{rescale}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim})) = w_1 |h_1(\varphi_{obs}, \varphi_{sim})| + w_2 |h_2(\varphi_{obs}, \varphi_{sim})|, \tag{5}$$

where w_1, w_2 are the scales provided by the inverse of the maximum absolute value of the elements of $h(\varphi_u, \varphi_v)$ for the pilot run, *i.e.*, $w_1 = 1/\max(|h_1(\varphi_{obs}, \varphi_{pilot})|)$ and $w_2 = 1/\max(\ |h_2(\varphi_{obs}, \varphi_{pilot})|)$. The goal of this approach is to rescale the effect of each element in the discrepancy measure to a common level.

The surface of the log discrepancy measure from Shirota and Gelfand (2017) and the proposed surfaces based on the pilot run are shown in Figures A1, A2, and A3 (in the Appendix). The value of each pixel in the surfaces is the log discrepancy between the observed point pattern and the simulated point pattern that is generated using the corresponding parameter values on the axes. Smaller discrepancy values indicate that the simulated and observed point pattern are similar. For the simulation study, we expect that the pixel associated with the true parameter value will minimize the discrepancy. However, we can see that the area having minimum discrepancy value is significantly larger than one pixel and that the value of neighboring pixels are also small. This behavior illustrates the identifiability issue associated with the Gibbs point process, *i.e.*, different parameter value combinations can generate similar point patterns in terms of the discrepancy measures discussed in this study. We mitigate the identifiability effect by using an informative prior; simulation results are shown in Section 5. Also, the value of the surfaces from Shirota and Gelfand (2017) are smaller than the values of the surfaces for the proposed method because of the lack of information. The flatness of these surfaces can also cause inefficiencies in the ABC algorithm.

4. Simulation Study: Vague Prior

For this study, we implement a uniform prior for the ABC rejection algorithm. In other words, we generate proposed parameter values from a uniform distribution with

reasonable lower and upper bounds for both the Strauss process and the area-interaction process. The crucial components of the discrepancy measures are the number of points in the point pattern and the estimated summary characteristics over interaction range $[0, r_{max}]$. For the K, D and F-function, we choose a moderately large $r_{max} = 0.18$ based on a sensitivity study. However, r_{max} is 0.08 for the J-function since the value of J-function is either highly unstable or undefined for large r. For the pilot run, we generate 100 parameter values equally spaced in the pre-specified parameter space, *e.g.*, a sequence of β from 50 to 400 and a sequence of γ from 0 to 1. Thus, there are 10,000 combinations of parameter values in total, and the corresponding realizations are generated to calibrate the coefficients $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, w_1 and w_2 for the discrepancy measure. The acceptance threshold ϵ is set to be the 1% percentile of $d_{rescale}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot run. To simulate point patterns from the specified process in the ABC and pilot run, we assume that the interaction distance is known, however it can also be estimated using profile maximum pseudo-likelihood in practice. Finally, 1000 samples are accepted as the posterior realizations. The simulation study is conducted based on 50 realizations from each process, and the average estimate of the model parameters are compared in the following sections.

4.1. Strauss point process

We first consider the homogeneous Strauss process with strong interaction on the unit window, $W = [0, 1] \times [0, 1]$. Following the same parameterization of (3), we set $\beta = 200$, $\gamma = 0.1$ and $r = 0.05$. The simulation algorithm used to generate the point patterns is the *dominated coupling from the past algorithm* (Berthelsen and Møller, 2003). The prior distributions that are used as proposal kernels for β and γ are $\mathcal{U}(50, 400)$ and $\mathcal{U}(0, 1)$, and the estimation results are shown in Table 1. Each metric in this table corresponds to the discrepancy measures based on different summary statistics and structures (Formulas (4) and (5)). Specifically, $K(r)$ represents the discrepancy measure proposed by Shirota and Gelfand (2017). For the remaining metrics, the name of the metric indicates the functional characteristic that is used in the ABC and the method used to construct the discrepancy measure is shown in the parentheses, *e.g.*, D-function (regression) means that the discrepancy measure is constructed by linear regression with summary statistics based on the D-function and number of points.

Considering the point estimate and standard deviation of the posterior distribution for each parameter, the algorithm proposed by Shirota and Gelfand (2017) has the best performance, since the average posterior median provides the smallest bias for the main effect and second smallest bias for the interaction effect. Additionally, the posterior standard deviations are the smallest for both parameters among all of the algorithms. ABC with the discrepancy measure constructed based on the D-function and rescaling method also has a competitive performance, and therefore, both algorithms can be considered in order to fit the Strauss process. However, the large posterior standard deviation indicates that the estimates are affected by the apparent identifiability issue inherited from the underlying mechanism of the Gibbs processes.

4.2. Area-interaction point process

We also conducted a simulation study for the homogeneous area-interaction process with strong interaction on unit window W. By implementing the canonical scale-free

Table 1: The average posterior mean, median and standard deviation for Strauss process with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5). The bold numbers represent the best and second best value in each column.

Metric	Parameter	True value	Average mean	Average median	Average SD
$K(r)$	β	200	**203.5**	**200.4**	**35.65**
	γ	0.1	0.125	**0.113**	**0.070**
D-function	β	200	251.9	248.2	51.37
(regression)	γ	0.1	**0.097**	0.084	**0.063**
D-function	β	200	206.5	**202.9**	41.81
(rescaling)	γ	0.1	**0.120**	**0.103**	0.080
K-function	β	200	226.2	220.7	57.87
(regression)	γ	0.1	0.147	0.123	0.109
K-function	β	200	195.9	192.5	42.18
(rescaling)	γ	0.1	0.173	0.141	0.135
F-function	β	200	182.5	179.4	40.94
(regression)	γ	0.1	0.265	0.228	0.187
F-function	β	200	182.9	180.1	40.14
(rescaling)	γ	0.1	0.248	0.210	0.184
J-function	β	200	**197.0**	194.0	39.50
(regression)	γ	0.1	0.182	0.156	0.128
J-function	β	200	194.9	192.0	**37.70**
(rescaling)	γ	0.1	0.167	0.142	0.119

version, we set $\kappa = 50$, $\eta = 7$ and $r = 0.05$. Here, setting the disc radius of the area-interaction process to be 0.05 implies that the interaction distance is 0.1. For this process, the prior distributions of κ and η are $\mathcal{U}(10, 400)$ and $\mathcal{U}(1, 30)$. The estimation results of 50 realizations from the area-interaction process are shown in Table 2. The average mean and median show that only using the J-function as the summary statistic can provide reasonable estimates for both parameters. Although, the discrepancy measure based on the K-function and D-function can provide a reasonable estimate of one parameter, the posterior standard deviations are significantly larger than using J-function. Moreover, the discrepancy measure based on the J-function outperforms the one proposed by Shirota and Gelfand (2017).

Overall, the most robust estimator in this case is the median of the posterior sample from the ABC algorithm based on the discrepancy measure constructed by the summary statistic J-function and rescaling method. Similar to the Strauss process, the large posterior standard deviations indicate that the identifiability issue still exists for the area-interaction process.

5. Simulation Study: Informative Prior

In order to reduce the effect of the identifiability issue that was illustrated in Figures A1, A2, and A3 (in the Appendix), we implemented a class of informative prior distributions for the ABC. For the Strauss process, the prior of the main effect is a shifted gamma distribution and for the interaction effect a truncated normal distribution. Simi-

Table 2: The average posterior mean, median and standard deviation for area-interaction process with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5). The bold numbers represent the best and second best value in each column.

Metric	Parameter	True value	Average mean	Average median	Average SD
$K(r)$	κ	50	99.95	80.62	69.83
	η	7	**6.760**	3.912	6.647
D-function	κ	50	141.4	122.1	88.04
(regression)	η	7	4.457	2.933	**4.112**
D-function	κ	50	100.3	77.65	72.66
(rescaling)	η	7	**6.737**	4.833	5.703
K-function	κ	50	65.46	**54.12**	42.05
(regression)	η	7	9.811	7.893	6.902
K-function	κ	50	73.15	63.53	41.36
(rescaling)	η	7	7.820	**6.193**	5.646
F-function	κ	50	89.12	75.14	56.49
(regression)	η	7	6.661	4.293	6.146
F-function	κ	50	93.53	80.33	57.59
(rescaling)	η	7	6.155	3.980	5.682
J-function	κ	50	**46.53**	38.53	**29.30**
(regression)	η	7	9.604	8.450	5.387
J-function	κ	50	**58.78**	**49.16**	**35.06**
(rescaling)	η	7	9.000	**7.880**	**5.221**

larly, the shifted gamma distribution is used for both parameters of the area-interaction process. The summary statistics and discrepancy measures considered are the same as in Section 4.

5.1. Strauss point process

We consider an informative prior as the proposal kernel for the ABC. In order to generate enough points to estimate the summary characteristics, we set a lower bound for the prior of the main effect by shifting it to the right. Specifically, the prior distributions for β and γ are

$$\beta - l_\beta \sim G((\hat{\beta} - l_\beta)^2/\sigma_\beta, \sigma_\beta/(\hat{\beta} - l_\beta)), \tag{6}$$

$$\gamma \sim N_{[0,1]}(\hat{\gamma}, \sigma_\gamma^2), \tag{7}$$

where $l_\beta = 50$, $\sigma_\beta = 300$ and $\sigma_\gamma = 0.1$. $\hat{\beta}$ and $\hat{\gamma}$ are the estimated parameter values using stochastic approximation. Thus, the mean and variance of the shifted gamma distribution for β are $\hat{\beta}$ and σ_β. For γ, we choose a standard deviation that is not too large over the parameter space $[0, 1]$ to preserve the information from the stochastic approximation. In order to demonstrate the consistency of the stochastic approximation, we show that the average of the estimated parameter values for β and γ over 50 realizations are 200.7 and 0.098, respectively.

The simulation results of the same realizations of Section 4.1 are shown in Table 3. From this table, we can see the improvement in terms of the point estimate when compared to the estimate from stochastic approximation, is limited and that the estimated

Table 3: The average posterior mean, median and standard deviation for Strauss process with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5). The bold numbers represent the best and second best value in each column.

Metric	Parameter	True value	Average mean	Average median	Average SD
$K(r)$	β	200	200.2	202.0	16.51
	γ	0.1	0.111	0.116	0.063
D-function	β	200	206.2	**199.8**	16.40
(regression)	γ	0.1	**0.110**	**0.104**	**0.051**
D-function	β	200	200.5	**199.9**	15.87
(rescaling)	γ	0.1	**0.104**	**0.098**	**0.051**
K-function	β	200	202.6	201.9	16.51
(regression)	γ	0.1	0.122	0.116	0.063
K-function	β	200	**200.1**	199.5	15.68
(rescaling)	γ	0.1	0.112	0.105	0.062
F-function	β	200	**199.9**	199.4	**15.36**
(regression)	γ	0.1	0.123	0.118	0.067
F-function	β	200	199.3	198.8	**15.35**
(rescaling)	γ	0.1	0.118	0.111	0.066
J-function	β	200	200.4	**199.8**	15.55
(regression)	γ	0.1	0.117	0.112	0.062
J-function	β	200	199.6	199.0	15.38
(rescaling)	γ	0.1	0.113	0.107	0.060

values are close for different summary statistics and discrepancy measures. However, ABC provides significant reduction of the standard deviation when compared with the standard deviation of the prior distribution, which is an indication of Bayesian learning. The maximum standard deviation reductions for the distribution of β and γ are 11.4% and 49%, respectively. Notice that the discrepancy measure proposed by Shirota and Gelfand (2017) shows good point estimates of the parameters. However, combining the results from the point estimates and the posterior standard deviation, using the D-function as the summary statistic, and constructing the discrepancy measure by rescaling, provides the most robust performance. More importantly, implementing the informative prior can significantly improve point estimation and reduce the posterior standard deviation, in contrast to the ABC with vague prior.

5.2. Area-interaction point process

We also implement the informative prior for the area-interaction process. The right shifted gamma distribution is used as the prior distribution for κ and η,

$$\kappa - l_\kappa \sim G((\hat{\kappa} - l_\kappa)^2/\sigma_\kappa, \sigma_\kappa/(\hat{\kappa} - l_\kappa)),$$
$$\eta - l_\eta \sim G((\hat{\eta} - l_\eta)^2/\sigma_\eta, \sigma_\eta/(\hat{\eta} - l_\eta)),$$

where $l_\kappa = 10$, $l_\eta = 1$, $\sigma_\kappa = 300$, $\sigma_\eta = 100$, $\hat{\kappa}$ and $\hat{\eta}$ are the estimates from stochastic approximation. That is, the mean and variance of the shifted gamma distribution for κ and η are $(\hat{\kappa}, \sigma_\kappa)$ and $(\hat{\eta}, \sigma_\eta)$. We choose to use large prior variances to ensure that the choice of prior will not dominate the results of the analysis. Alternatively, the

Table 4: The average posterior mean, median and standard deviation for area-interaction process with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5). The bold numbers represent the best and second best value in each column.

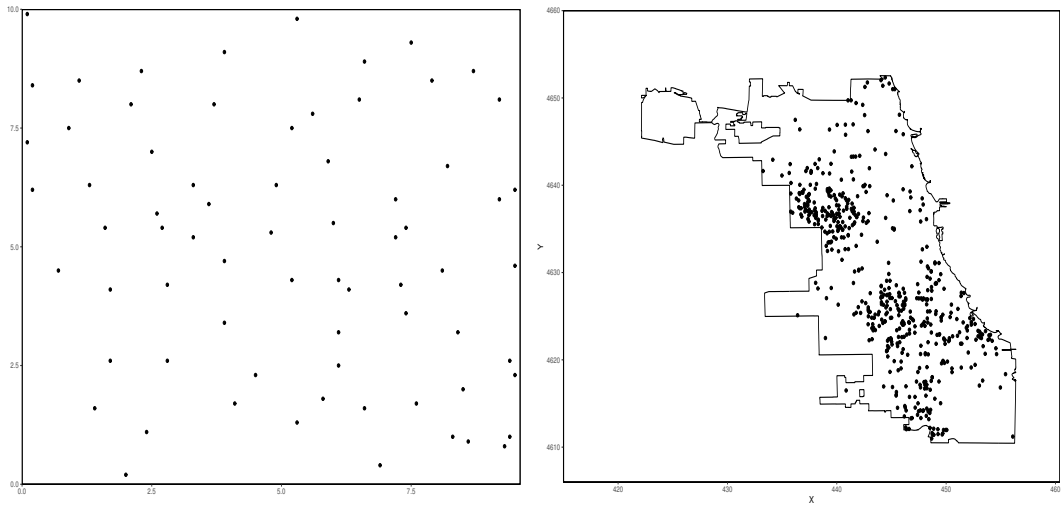
Metric	Parameter	True value	Average mean	Average median	Average SD
$K(r)$	κ	50	58.11	55.64	17.67
	η	7	7.521	6.834	3.310
D-function	κ	50	60.72	58.42	17.52
(regression)	η	7	7.360	6.873	**2.753**
D-function	κ	50	57.18	54.79	16.75
(rescaling)	η	7	7.726	7.158	3.051
K-function	κ	50	56.33	54.60	15.46
(regression)	η	7	7.843	**7.061**	3.924
K-function	κ	50	58.02	56.17	15.63
(rescaling)	η	7	**7.154**	6.504	3.335
F-function	κ	50	59.02	56.64	17.63
(regression)	η	7	7.311	6.642	3.311
F-function	κ	50	59.34	56.92	17.66
(rescaling)	η	7	7.221	6.566	3.241
J-function	κ	50	**51.10**	**49.33**	**13.83**
(regression)	η	7	**7.324**	6.808	**2.967**
J-function	κ	50	**53.74**	**51.90**	**14.32**
(rescaling)	η	7	7.623	**7.098**	3.030

truncated gamma distribution can be used as the proposal kernel to simulate parameters with a specific mean, variance, and positive lower bound. The average estimates from the stochastic approximation of κ and η over the 50 realizations used in Section 4.2 are 56.57 and 7.151, and the simulation results are shown in Table 4. By comparing the average posterior mean, median, and standard deviation with the prior distribution, we see significant improvements in terms of the point estimate and reduction in standard deviation, *i.e.*, the maximum reduction for κ and η are 20.2% and 70.3%, respectively. Among all the possible combinations, using the J-function as the summary statistic to construct the discrepancy measure provides the most robust results for both parameters and using regression or rescaling methods do not drastically affect the performance of the J-function. Similar to the results in Section 4.2, the proposed discrepancy measures based on the J-function outperform the one proposed by Shirota and Gelfand (2017). Also, implementing the informative prior significantly improves point estimation and reduces the posterior standard deviation for all the approaches.

6. Real Data Application

6.1. Swedish pines data

The Swedish pines data, shown in Figure 1(a), contains the locations of 71 pine saplings in a 9.6 by 10 meter window provided (Strand, 1972). Previous analyses on this data include Ripley (1981), Venables and Ripley (1997) and Baddeley and Turner (2000). All the results indicate that the interaction distance r is 0.7. To be proportional with the



(a) (b)

Figure 1: (a) The Swedish pines data: locations of 71 pine saplings in a 9.6 m $\times$ 10 m window first studied by Strand (1972). The data is available in the R package "spatstat". (b) Chicago crime data: locations based on UTM projection of 564 Chicago homicide incidents in 2018. The data is published by the City of Chicago Data Portal.

window size, the maximum interaction distance r_{max} is 0.8 for the J-function and 1.8 for the other summary characteristics. By assuming the Strauss process, the point estimates for β are between 1.49 and 3.29 and for γ are between 0.20 and 0.29 in the previous analyses mentioned above. In our study, we fit the Strauss process and area-interaction process to this dataset by using the ABC algorithm with different discrepancy measures and prior distributions

For the Strauss process, informative and vague priors are implemented in the ABC algorithm. Similar to Section 5, the prior distributions are gamma and truncated normal distribution with the mean provided by the estimate of the parameter values from stochastic approximation, *i.e.*, $\hat{\beta}$ and $\hat{\gamma}$ are 2.262 and 0.197, respectively. The standard deviations of the prior distributions are $\sqrt{30}$ and 0.1, respectively. The results from the ABC are shown in Table 5. For all the ABC algorithms, the acceptance threshold is the 1% percentile of the discrepancy measure in the pilot run. The overall results indicate strong inhibition between points in the point pattern and the values are consistent with the previous studies. Moreover, the standard deviations of the posterior distributions show significant reduction when compared with the prior standard deviations. For the vague priors, $\mathcal{U}(0.5, 40)$ and $\mathcal{U}(0, 1)$ are used for β and γ. The results are included in Table A1 in the Appendix and indicate larger posterior standard deviation. However, the overall results based on the vague prior are still consistent with the ones from the informative prior and previous studies.

We also fit an area-interaction process to the Swedish pines data, and the estimates show strong inhibition as well. However, a goodness-of-fit test shows lack of fit for the area-interaction process compared with the Strauss process. We included the estimates of the area-interaction process in the Appendix.

Table 5: The posterior mean, median, standard deviation and 95% credible interval of Strauss process for the Swedish pines data with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	β	2.877	2.494	1.766	(1.471, 7.175)
	γ	0.203	0.196	0.089	(0.047, 0.381)
D-function	β	1.960	1.732	1.024	(0.694, 4.479)
(regression)	γ	0.211	0.204	0.078	(0.078, 0.379)
D-function	β	2.047	1.878	0.945	(0.782, 4.401)
(rescaling)	γ	0.207	0.201	0.077	(0.077, 0.265)
K-function	β	5.134	4.006	4.190	(1.070, 16.06)
(regression)	γ	0.201	0.196	0.069	(0.080, 0.343)
K-function	β	2.783	2.449	1.462	(0.935, 6.691)
(rescaling)	γ	0.201	0.200	0.077	(0.058, 0.361)
F-function	β	2.894	2.755	0.846	(1.659, 5.029)
(regression)	γ	0.168	0.163	0.089	(0.016, 0.360)
F-function	β	2.881	2.748	0.850	(1.658, 4.951)
(rescaling)	γ	0.167	0.163	0.087	(0.162, 0.356)
J-function	β	2.406	2.234	0.761	(1.311, 4.250)
(regression)	γ	0.156	0.150	0.077	(0.021, 0.311)
J-function	β	2.616	2.479	0.801	(1.465, 4.605)
(rescaling)	γ	0.164	0.158	0.082	(0.022, 0.343)

6.2. Chicago crime data

The Chicago crime data contains locations of reported homicide incidents in Chicago during 2018. The source of the raw data is the City of Chicago Data Portal which is an online resource summarizing incidents of crime that occurred in the city of Chicago (`https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2`). As demonstrated in Figure 1(b), we use the Universal Transverse Mercator (UTM) coordinates for incident locations and the city boundary as the irregular domain of this point pattern. It is obvious that the point pattern shows significant clustering. Thus, we have reason to assume the data is generated from an attractive point process.

In order to be consistent with the simulation study, we assume that the process is homogeneous and proceed to fit the area-interaction process to this point pattern using the ABC algorithm. The disc radius of the area-interaction process is 0.35, which is calculated based on maximizing the profile pseudo-likelihood. Following the same choices in the simulation study, the maximum interaction distance r_{max} is 2 for the J-function and 7 for the other summary characteristics. Also, we implemented the gamma distribution as the informative prior along with the prior mean provided by the estimates from stochastic approximation. The prior mean and standard deviation for κ are 0.3575 and 1 and for η are 24.2648 and 30.

For all the ABC algorithms, the acceptance threshold is the 1% sample percentile

Table 6: The posterior mean, median, standard deviation and 95% credible interval of area-interaction process for the Chicago crime data with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	κ	0.346	0.265	0.114	(0.175, 0.586)
	η	35.74	27.18	26.69	(9.035, 102.4)
D-function	κ	0.495	0.448	0.216	(0.180, 0.994)
(regression)	η	21.68	14.65	22.54	(1.605, 88.19)
D-function	κ	0.542	0.468	0.292	(0.197, 1.300)
(rescaling)	η	20.46	13.40	22.69	(0.605, 77.62)
K-function	κ	0.078	0.012	0.162	(0.001, 0.561)
(regression)	η	18.40	8.497	25.86	(0.046, 90.69)
K-function	κ	0.590	0.447	0.469	(0.159, 1.841)
(rescaling)	η	27.03	15.30	34.82	(0.094, 126.0)
F-function	κ	0.372	0.353	0.131	(0.172, 0.642)
(regression)	η	32.60	24.29	27.39	(5.625, 105.9)
F-function	κ	0.488	0.392	0.316	(0.175, 1.383)
(rescaling)	η	28.57	20.32	28.35	(0.484, 100.6)
J-function	κ	0.381	0.363	0.119	(0.191, 0.661)
(regression)	η	29.32	25.31	17.75	(8.026, 75.16)
J-function	κ	0.417	0.381	0.189	(0.165, 0.912)
(rescaling)	η	31.31	21.59	30.50	(2.258, 114.1)

of the pilot run. The results are shown in Table 6 and indicate significant attraction between points. Similar to the results in Section 5, the ABC algorithm provides significant standard deviation reduction for both parameters. We also use $\mathcal{U}(0,5)$ and $\mathcal{U}(0,150)$ as the vague prior for κ and η. The results are included in Table A4 in the Appendix, which show larger posterior variances for the parameters compared with the ones from the ABC with informative prior. This indicates that the informative prior helps alleviate the lack of identifiability.

7. Concluding Remarks

In this study, we compared the performances of several discrepancy measures for the ABC algorithm and proposed new measures. The simulation results confirm several aspects of our assumptions: the effect of different summary statistics on the estimation results, the improved performance of comparing the functional characteristics over a range of interaction distances and the effect of implementing the informative prior for the ABC.

First of all, the candidate summary characteristics show different performances for different point processes. This indicates that the choices of the most robust summary statistic for the ABC algorithm is point process specific. Although the literature points out that the choice should be a function of the sufficient statistic based on the sufficiency principle, we observed that the nearest neighbor distance distribution function can also provide competitive results for the Strauss process. Also, the F-function and J-function are both sufficient statistics for the area-interaction process, but the simulation shows that the J-function is more robust. Secondly, the discrepancy measure constructed by

the K-function given a specific interaction distance $K(r)$ out-performs the one that is constructed by the K-function over a range of interaction distances for the Strauss process, but not for the area-interaction process. This result indicates that evaluating attractive point processes requires more information than repulsive point processes. Finally, we demonstrated that applying an informative prior can significantly improve the point estimates of the parameters and reduce the posterior standard deviation. However, the identifiability issue of the Gibbs process cannot be completely eliminated by the current approaches. The applications to the Swedish pines data and Chicago crime data show that the ABC algorithm can be easily applied to point processes with regular and irregular windows.

In practice, we recommend that the analysis of the Gibbs point processes using the ABC algorithm should proceed as follows. First, an appropriate model for the point pattern should be specified based on an initial exploratory analysis. The Strauss process can be considered for patterns with strong inhibition and the area-interaction process can be considered for general clustered patterns. After choosing a specific model, one should decide the best discrepancy measure for the chosen model. For the models included in this study, the most robust discrepancy measures are provided. For other models, a similar simulation study, to the ones conducted in this paper, can provide a good indication for the best discrepancy measure. Finally, assuming that an informative prior is implemented, the median of the posterior realizations from the ABC is usually the most robust estimator of the parameters.

Since the Gibbs point process can be uniquely determined by the Papangelou conditional intensity, our future work will be investigating possible choices of the nonparametric estimate of the Papangelou and its implementation as the summary statistic for the ABC algorithm in order to resolve the identifiability issue. Moreover, a non-homogeneous version of the ABC algorithm will be developed. However, examining the non-homogeneous Gibbs point processes can be challenging since the non-homogeneity of the main effect is always confounded with the interaction effect.

Acknowledgements

This research was partially supported by the U.S. National Science Foundation (NSF) under NSF SES-1853096 and through the Air Force Research Laboratory (AFRL) Contract No.19C0067. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors, and not those of the U. S. Census Bureau.

References

Alquier, P., Friel, N., Everitt, R., and Boland, A. (2016). Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels. *Statistics and Computing*, **26(1-2)**, 29–47.

Andrieu, C., Roberts, G. O., *et al.* (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, **37(2)**, 697–725.

Atchadé, Y. F., Lartillot, N., Robert, C., *et al.* (2013). Bayesian computation for statistical models with intractable normalizing constants. *Brazilian Journal of Probability*

and Statistics, **27(4)**, 416–436.

Baddeley, A. and Turner, R. (2000). Practical Maximum Pseudolikelihood for Spatial Point Patterns: (with Discussion). *Australian and New Zealand Journal of Statistics*, **42(3)**, 283–322.

Baddeley, A. and Turner, R. (2014). Package 'spatstat'. *The Comprehensive R Archive Network (CRAN)*.

Baddeley, A. J. and Van Lieshout, M. (1995). Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, **47(4)**, 601–619.

Beaumont, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics*, **164(3)**, 1139–1160.

Berthelsen, K. K. and Møller, J. (2003). Likelihood and non-parametric bayesian mcmc inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, **30(3)**, 549–564.

Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **24(3)**, 179–195.

Cressie, N. (1993). *Statistics for Spatial Data. 2nd Edition*. John Wiley & Sons.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **74(3)**, 419–474.

Geyer, C. (1998). Likelihood inference for spatial point. *Stochastic Geometry: Likelihood and Computation*, **80**, 79.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. Interface Foundation of North America. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/58440.

Gu, M. G. and Zhu, H.-T. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **63(2)**, 339–355.

Huang, F. and Ogata, Y. (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics*, **8(3)**, 510–530.

Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, vol. **70**. John Wiley & Sons.

Kelly, F. P. and Ripley, B. D. (1976). A note on strauss's model for clustering. *Biometrika*, **63**, 357–360.

Liang, F. (2010). A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants. *Journal of Statistical Computation and Simulation*, **80(9)**, 1007–1022.

Liang, F., Jin, I. H., Song, Q. and Liu, J. S. (2016). An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants. *Journal of the American Statistical Association*, **111(513)**, 377–393.

Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. (2003). Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, **100(26)**, 15324–15328.

Mateu, J., Schoenberg, F. P. and Diez, D. (2010). On distances between point patterns and their applications. *American Statistician, in Review*, **9**.

Møller, J., Pettitt, A. N., Reeves, R. and Berthelsen, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants.

Biometrika, **93(2)**, 451–458.

Murray, I., Ghahramani, Z., and MacKay, D. (2012). MCMC for doubly-intractable distributions. *arXiv preprint arXiv:1206.6848*.

Park, J. and Haran, M. (2018). Bayesian inference in the presence of intractable normalizing functions. *Journal of the American Statistical Association*, **113(523)**, 1372–1390.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. and Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular biology and evolution*, **16(12)**, 1791–1798.

Ripley, B. D. (1981). *Spatial Statistics*. vol. **24**. Wiley Online Library.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 400–407.

Shirota, S. and Gelfand, A. E. (2017). Approximate Bayesian computation and model assessment for repulsive spatial point processes. *Journal of Computational and Graphical Statistics*, **26(3)**, 646–657.

Stoica, R. S., Philippe, A., Gregori, P. and Mateu, J. (2017). ABC Shadow algorithm: a tool for statistical analysis of spatial patterns. *Statistics and Computing*, **27(5)**, 1225–1238.

Strand, L. (1972). A model for stand growth. In *IUFRO Third Conference Advisory Group of Forest Statisticians*, vol. **72**. INRA, Institut National de la Recherche Agronomique Paris, p. 3.

Strauss, D. J. (1975). A model for clustering. *Biometrika*, **62(2)**, 467–475.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M. and Dessimoz, C. (2013). Approximate Bayesian computation. *PLoS Computational Biology*, **9(1)**, e1002803.

Van Lieshout, M. and Baddeley, A. (1996). A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica*, **50(3)**, 344–361.

Venables, W. and Ripley, B. D. (1997). Modern Applied Statistic with S-plus. Second Edition. Springer Verlag.

Appendix

Table A1: The posterior mean, median, standard deviation and 95% credible interval of Strauss process for the Swedish pines data with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	β	5.379	2.812	5.909	(1.367, 12.38)
	γ	0.242	0.190	0.201	(0.018, 0.674)
D-function	β	2.176	1.727	1.514	(0.617, 6.450)
(regression)	γ	0.283	0.237	0.177	(0.057, 0.712)
D-function	β	2.226	1.873	1.372	(0.660, 5.633)
(rescaling)	γ	0.272	0.227	0.171	(0.053, 0.674)
K-function	β	11.90	9.761	8.942	(1.345, 34.63)
(regression)	γ	0.181	0.166	0.093	(0.052, 0.409)
K-function	β	3.665	3.118	2.398	(0.900, 9.879)
(rescaling)	γ	0.213	0.189	0.129	(0.029, 0.528)
F-function	β	2.894	2.755	0.846	(1.659, 5.029)
(regression)	γ	0.168	0.163	0.089	(0.016, 0.360)
F-function	β	3.227	3.065	1.194	(1.393, 5.981)
(rescaling)	γ	0.169	0.117	0.158	(0.004, 0.561)
J-function	β	2.810	2.650	0.993	(1.307, 5.263)
(regression)	γ	0.123	0.096	0.103	(0.004, 0.338)
J-function	β	3.011	2.919	1.037	(1.381, 5.417)
(rescaling)	γ	0.137	0.105	0.120	(0.003, 0.437)

A.1. Swedish pines data: fitted by the area-interaction process

The gamma distribution is used as the informative prior, with prior means 1.128 and 0.109 for κ and η. The standard deviations for the priors are $\sqrt{30}$ and $\sqrt{0.1}$, respectively. The estimation results are shown in Table A2. The results also show strong inhibition in the point pattern. We can see that the ABC algorithms with the J-function yield smaller standard deviations for the main function and closer values to the estimate from the Strauss process (Table 5). For the ABC with the vague prior, the prior distributions are $\mathcal{U}(0.1, 40)$ and $\mathcal{U}(0, 1)$. The corresponding results are included in the Appendix, see Table A3.

Table A2: The posterior mean, median, standard deviation and 95% credible interval of area-interaction process for the Swedish pines data with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

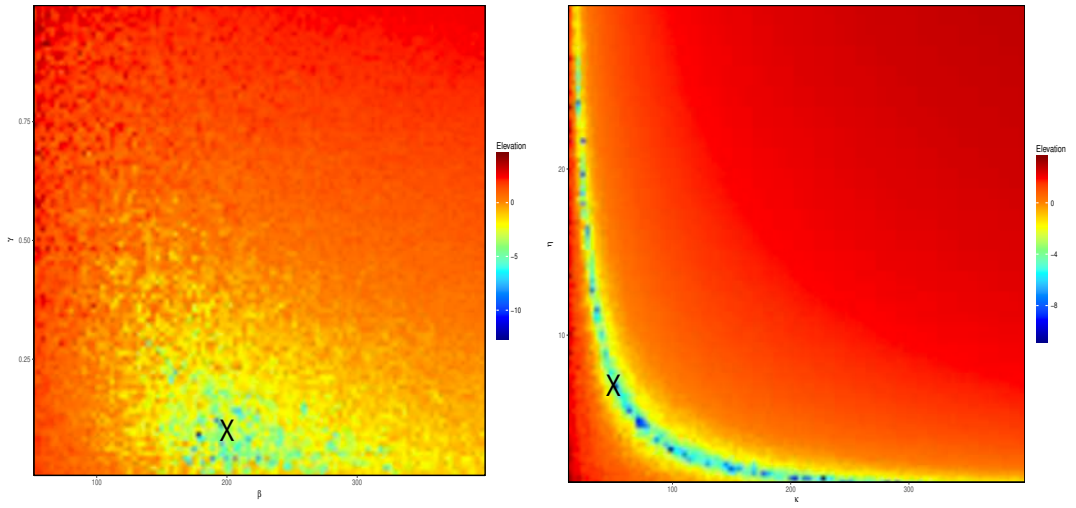
Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	κ	4.557	1.821	7.405	(0.344, 24.80)
	η	0.055	2.410×10^{-4}	0.210	(0.000, 0.450)
D-function	κ	5.343	2.057	8.555	(0.330, 28.88)
(regression)	η	0.046	2.083×10^{-4}	0.181	(0.000, 0.417)
D-function	κ	4.541	1.752	7.609	(0.265, 25.49)
(rescaling)	η	0.057	2.630×10^{-4}	0.236	(0.000, 0.461)
K-function	κ	5.012	1.939	8.166	(0.272, 27.87)
(regression)	η	0.036	1.839×10^{-4}	0.137	(0.000, 0.307)
K-function	κ	4.169	1.605	6.583	(0.252, 23.54)
(rescaling)	η	0.058	2.470×10^{-4}	0.223	(0.000, 0.495)
F-function	κ	3.892	1.018	7.607	(0.071, 25.30)
(regression)	η	0.083	15.08×10^{-4}	0.291	(0.000, 0.704)
F-function	κ	4.689	2.408	6.296	(0.618, 21.54)
(rescaling)	η	0.065	2.640×10^{-4}	0.245	(0.000, 0.610)
J-function	κ	1.785	0.843	2.421	(0.127, 8.619)
(regression)	η	0.090	9.084×10^{-4}	0.247	(0.000, 0.867)
J-function	κ	1.758	0.764	2.345	(0.005, 8.618)
(rescaling)	η	0.075	4.930×10^{-4}	0.263	(0.000, 0.664)

Table A3: The posterior mean, median, standard deviation and 95% credible interval of area-interaction process for the Swedish pines data with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	κ	5.379	2.812	5.901	(1.367, 23.38)
	η	0.242	0.190	0.201	(0.018, 0.674)
D-function	κ	1.982	0.966	3.611	(0.305, 11.06)
(regression)	η	0.308	0.237	0.282	(0.001, 0.942)
D-function	κ	1.726	1.117	2.151	(0.276, 7.722)
(rescaling)	η	0.342	0.290	0.281	(0.002, 0.945)
K-function	κ	2.084	1.455	1.875	(0.300, 7.506)
(regression)	η	0.342	0.290	0.258	(0.004, 0.913)
K-function	κ	1.717	1.281	1.513	(0.270, 6.373)
(rescaling)	η	0.350	0.268	0.289	(0.005, 0.956)
F-function	κ	2.102	1.701	1.573	(0.565, 6.080)
(regression)	η	0.375	0.334	0.287	(0.006, 0.958)
F-function	κ	1.988	1.624	1.394	(0.587, 6.366)
(rescaling)	η	0.373	0.302	0.294	(0.005, 0.965)
J-function	κ	1.553	1.247	1.360	(0.199, 6.024)
(regression)	η	0.390	0.335	0.297	(0.006, 0.965)
J-function	κ	1.420	1.161	1.165	(0.165, 4.511)
(rescaling)	η	0.388	0.348	0.286	(0.008, 0.958)

Table A4: The posterior mean, median, standard deviation and 95% credible interval of area-interaction process for the Chicago crime data with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	κ	0.304	0.273	0.110	(0.173, 0.568)
	η	48.18	44.88	26.90	(9.206, 97.20)
D-function	κ	0.398	0.334	0.208	(0.174, 0.927)
(regression)	η	36.49	29.24	28.82	(1.648, 95.90)
D-function	κ	0.434	0.343	0.272	(0.181, 1.188)
(rescaling)	η	34.72	27.56	28.42	(0.867, 95.26)
K-function	κ	0.123	0.032	0.174	(0.003, 0.653)
(regression)	η	42.18	25.71	43.63	(0.138, 139.9)
K-function	κ	0.370	0.245	0.320	(0.142, 1.372)
(rescaling)	η	61.18	56.44	46.72	(0.557, 145.0)
F-function	κ	0.304	0.266	0.121	(0.171, 0.616)
(regression)	η	47.99	45.41	27.58	(6.391, 97.21)
F-function	κ	0.367	0.282	0.241	(0.177, 1.045)
(rescaling)	η	44.05	41.26	29.05	(1.182, 96.66)
J-function	κ	0.423	0.408	0.111	(0.236, 0.656)
(regression)	η	23.77	20.71	13.10	(7.700, 58.63)
J-function	κ	0.300	0.250	0.160	(0.145, 0.733)
(rescaling)	η	61.35	53.80	41.85	(4.342, 141.7)



(a) (b)

Figure A1: Surfaces of the log discrepancy measure from Shirota and Gelfand (2017) for pilot run of Strauss and area-interaction process. (a) Strauss process; (b) Area-interaction process. The symbol "X" indicates the true values of the parameters.

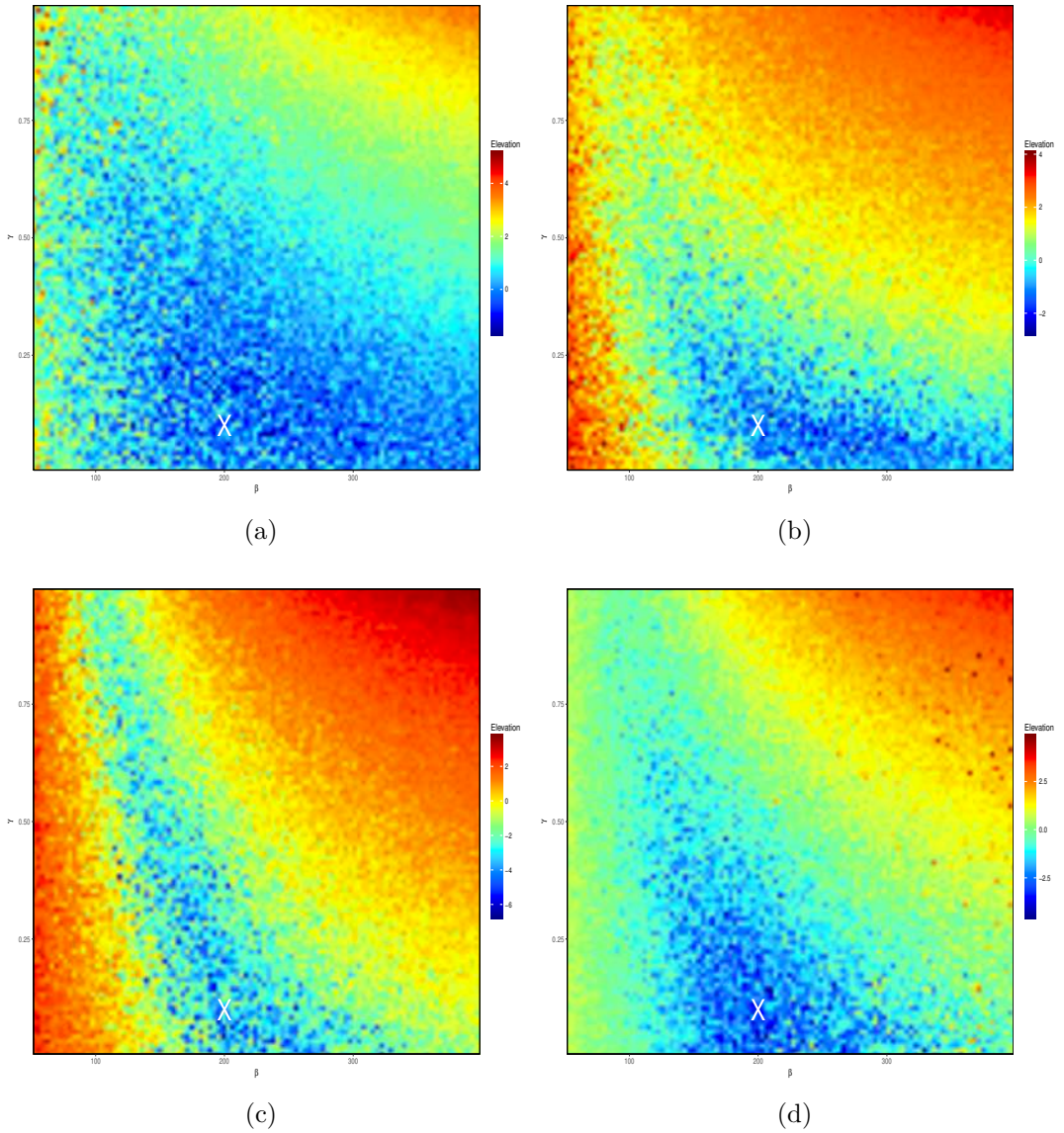
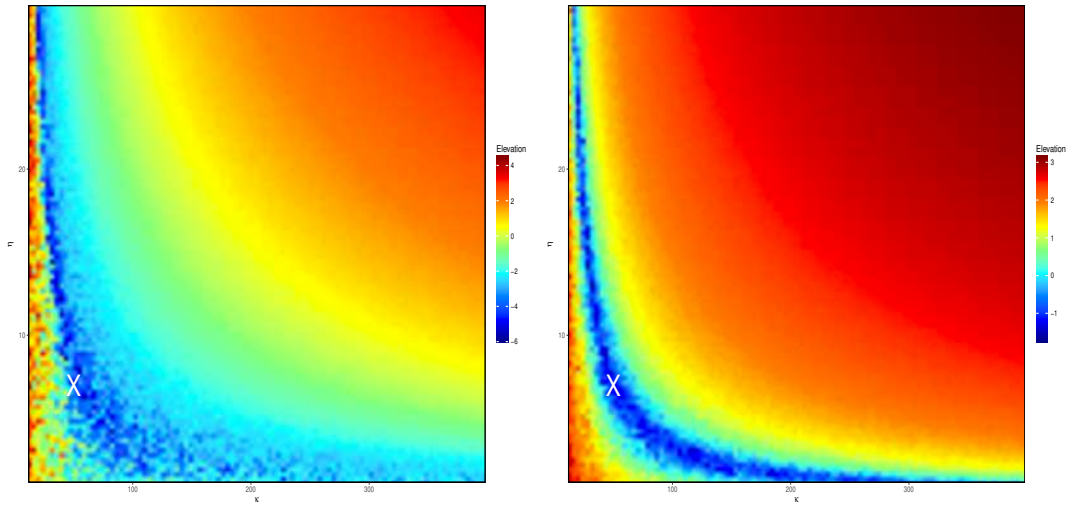
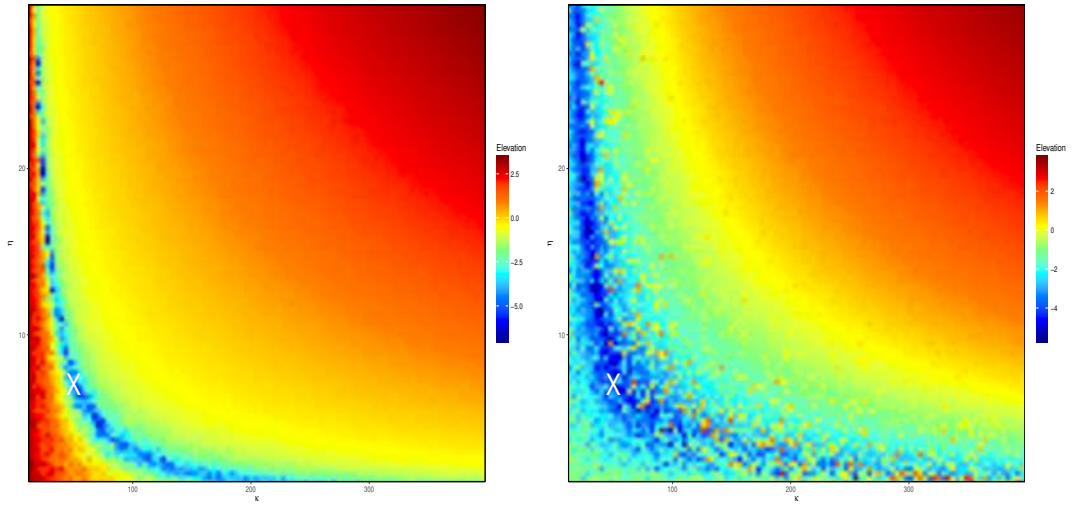


Figure A2: Surfaces of the log discrepancy measure constructed by regression $d_{reg}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot runs of the Strauss process. Each discrepancy measure is constructed based on different summary characteristics: (a) K-function; (b) D-function; (c) F-function; (d) J-function. The symbol "X" indicates the true values of the parameters.



(a) (b)



(c) (d)

Figure A3: Surfaces of the log discrepancy measure constructed by regression $d_{reg}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot runs of the area-interaction process. Each discrepancy measure is constructed based on different summary characteristics: (a) K-function; (b) D-function; (c) F-function; (d) J-function. The symbol "X" indicates the true values of the parameters.

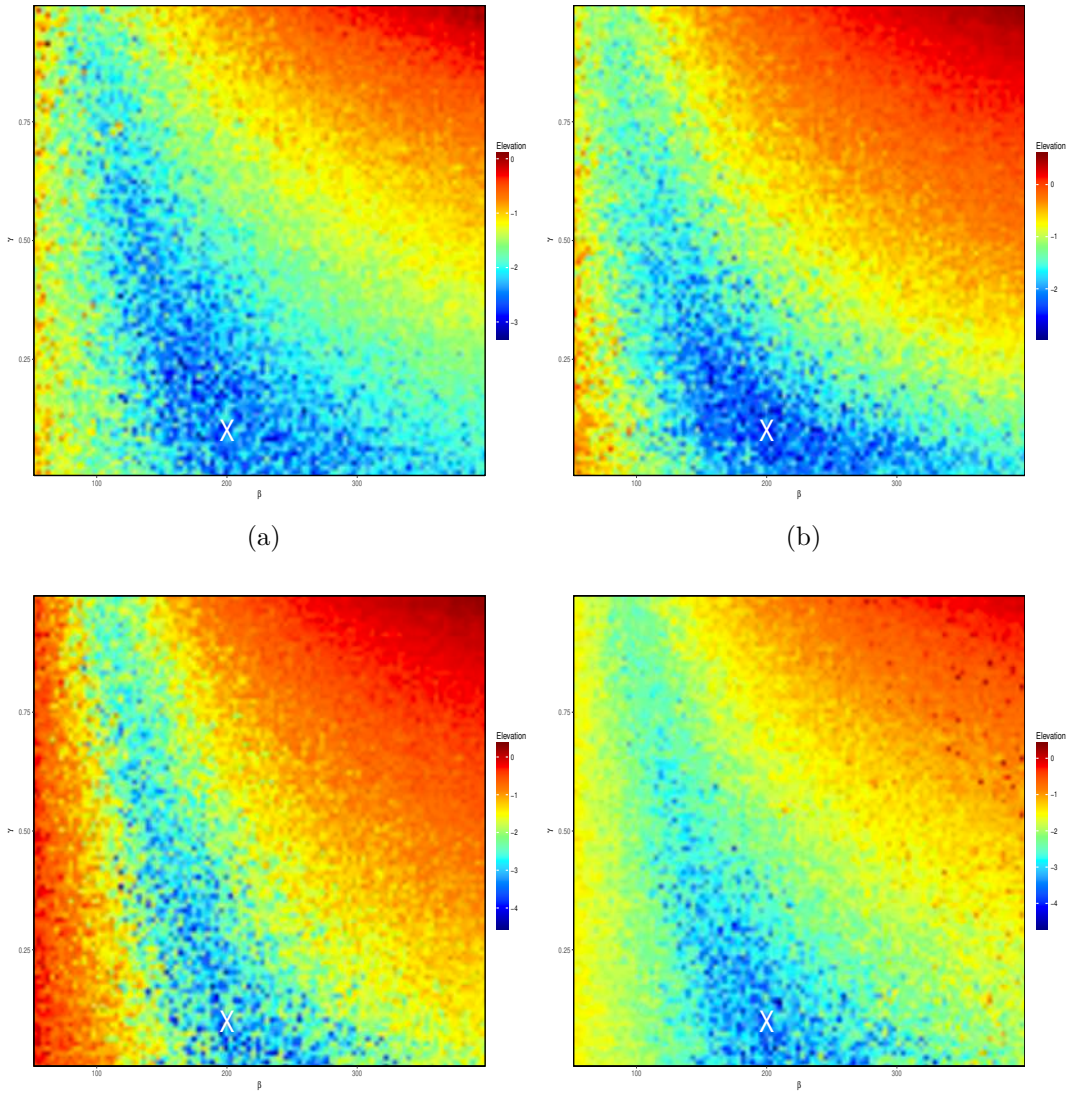
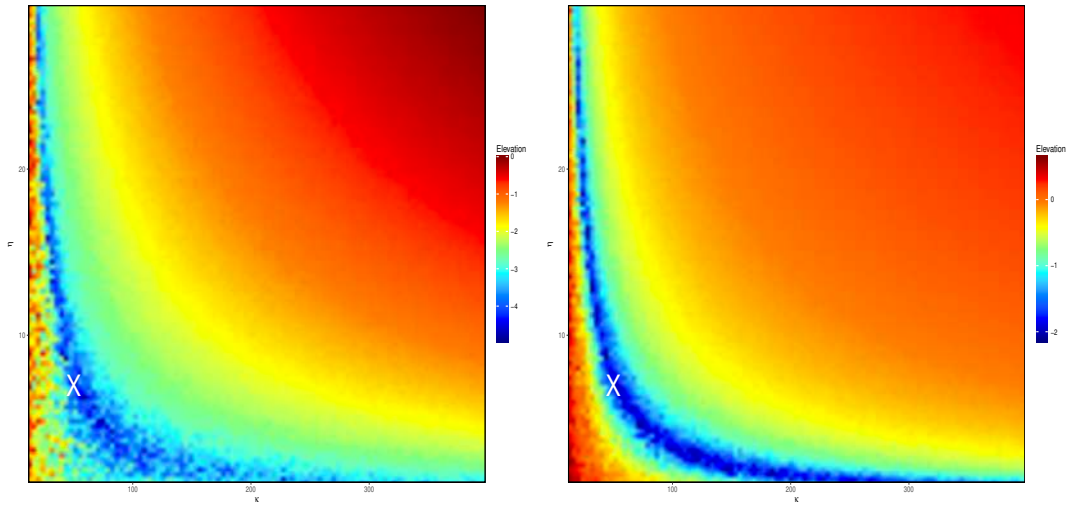
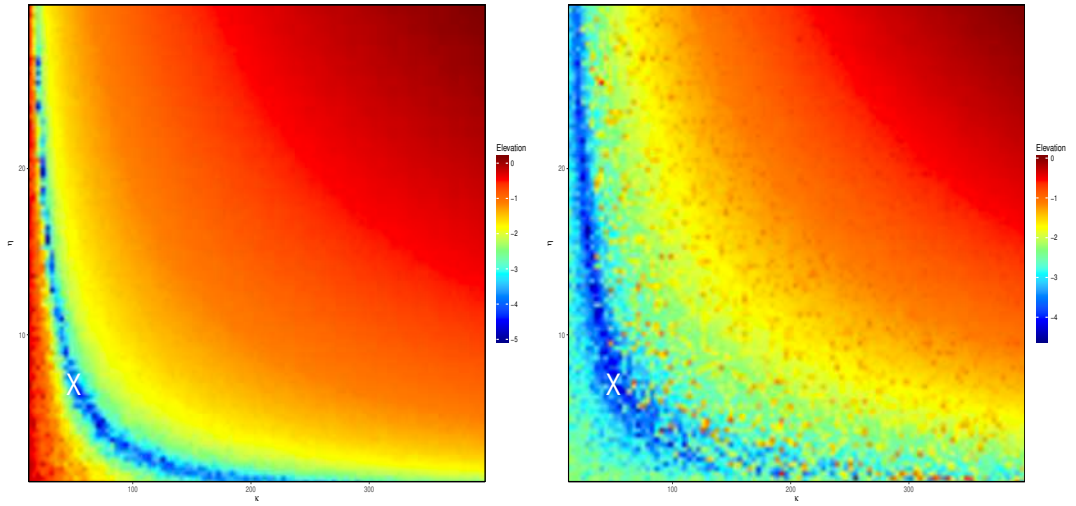


Figure A4: Surfaces of the log discrepancy measure constructed by rescaling $d_{rescale}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot runs of the Strauss model. Each discrepancy measure is constructed based on different summary characteristics: (a) K-function; (b) D-function; (c) F-function; (d) J-function. The symbol "X" indicates the true values of the parameters.



(a) (b)



(c) (d)

Figure A5: Surfaces of the log discrepancy measure constructed by rescaling $d_{rescale}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot runs of the area-interaction model. Each discrepancy measure is constructed based on different summary characteristics: (a) K-function; (b) D-function; (c) F-function; (d) J-function. The symbol "X" indicates the true values of the parameters.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 249–261

Some Computational Issues Related to Common Clustering Techniques

Asis Kumar Chattopadhyay
Department of Statistics
University of Calcutta, Kolkata, India

Received: 29 March 2020; Revised: 02 August 2020; Accepted: 04 August 2020

Abstract

In the present work attempts have been made to highlight different computational problems related to common clustering and dimensionality reduction techniques depending on input data type and underlying model assumptions of the different statistical methods. As clustering and dimensionality reduction techniques are widely used under machine learning and big data analysis, it is very much necessary to highlight the limitations to the user community (especially for the software industry). The effects of directional and missing data have also been considered.

Key words: Clustering; Computation; Directional Data; Missing Data.

1. Introduction

Cluster Analysis, also called data segmentation, has a variety of goals. All relate to grouping or segmenting a collection of objects (also called observations,individuals, cases, or data rows) into subsets or "clusters", such that those within each cluster are more closely related to one another than objects assigned to different clusters. Central to all of the goals of cluster analysis is the notion of degree of similarity (or dissimilarity) between the individual objects being clustered. There are two major methods of clustering - hierarchical clustering and k-means clustering.

Statistical techniques for classification are essentially of two types. Members of the first type are used to construct a sensible and informative classification of an initially unclassified set of data; these are known as cluster analysis methods. The information on which the derived classification is based is generally a set of variable values recorded for each object or individual in the investigation, and clusters are constructed so that individuals within clusters are similar with respect to their variable values and different from individuals in other clusters. The second set of statistical techniques concerned with classification is known as discriminant or assignment methods. Here the classification scheme is known a priori and

Correponding Author: Asis Kumar Chattopadhyay
Email: akcstat@caluniv.ac.in

the problem is how to devise rules for allocating unclassified individuals to one or other of the known classes.

Different Statistical techniques are available for clustering and classification (Fraix Burnet *et al.* (2015), De *et al.* (2013) and references there in). But depending on the nature of the different types of data the following problems often arise and in some cases a proper solution is still not available.

1. Sometimes the data set under consideration has a distributional form (usually normal) and sometimes it is of non normal nature. Based on the above point, there is a justification needed about which clustering or classification technique should be used so that it reflects the proper nature of the data set provided. This problem is more relevant for classification as most of the classification methods are model based. For clustering most of the methods are non parametric in nature and as such the above problem is not very serious. But here also basic assumption is that the nature of the variables under study are continuous where as under practical situations these may be categorical like binary, nominal, ordinal and even directional (particularly for environmental and Astronomical data). Under such situations standard similarity/ dissimilarity measures will not work.

2. The clustering techniques which require an inherent model assumption are known as Model Based Methods, whereas the clustering technique where no modelling assumption or distributional form is needed may be termed as Non-Model based Methods. Hence based on the nature of data set, one has to decide about proper application of the two types of techniques.

3. Even if one decides about the proper methods for the data set at hand, there are several techniques available under both the categories and no predefined criteria can be set to judge which technique is the best for the situation under consideration.

4. The above point arises the need of a comparative study among various available techniques and a computational analysis of all the methods.Once all the methods are implemented, it requires a criterion to decide upon the best technique based on a post classifier. So an appropriate post classification approach is also needed in this regard. For a post classification approach, a pre-classifier or training sample is required. Since in this type of techniques a prior knowledge of classification is provided, these are called Supervised Learning. All other techniques where no prior classification is provided are known as Unsupervised Learning.

5. A comparative validity algorithm may be helpful for predicting the superiority of different techniques.

6. At present big data issues related to data size is quite common. In statistical terms this problems may be tackled in terms of both the number of observations and the variables considered. Many standard clustering techniques fails to deal with such big data sets. Thus some dimension reduction methods may be applied at first and then clustering may be performed on the reduced data set. Some data mining techniques are very helpful under such situations.

9. The above criteria also needs to be validated depending on whether the data is Gaussian or non-Gaussian. That means the dimension reduction techniques may vary according as the data set has a distributional form or not.

10. Finally and most importantly after all these considerations, the similarity of grouping of objects obtained from different methods should be checked in terms of some physical properties .

2. Hierarchical Clustering Technique

Central to all of the goals of cluster analysis is the notion of degree of similarity (or dissimilarity) between the individual objects being clustered. There are two major methods of clustering -hierarchical clustering and k-means clustering. In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object. Hierarchical Clustering is subdivided into agglomerative methods, which proceed by series of fusions of the n objects into groups, and divisive methods, which separate n objects successively into finer groups. Agglomerative techniques are more commonly used. Hierarchical clustering may be represented by a two dimensional diagram known as dendrogram which illustrates the fusions or divisions made at each successive stage of analysis.

2.1. Agglomerative method

An agglomerative hierarchical clustering procedure produces a series of partitions of the data, $C_n, C_{n-1}, \ldots.., C_1$. The first C_n consists of n single object 'clusters', the last C_1, consists of single group containing all n cases.

At each particular stage the method joins together the two clusters which are closest together (most similar). (At the first stage, of course, this amounts to joining together the two objects that are closest together, since at the initial stage each cluster has one object.) Differences between methods arise because of the different ways of defining distance (or similarity) between clusters.

A key step in a hierarchical clustering is to select a distance measure. A simple measure is Manhattan distance, equal to the sum of absolute distances for each variable. The name comes from the fact that in a two-variable case, the variables can be plotted on a grid that can be compared to city streets, and the distance between two points is the number of blocks a person would walk.

A more common measure is Euclidean distance, computed by finding the square of the distance between each variable, summing the squares, and finding the square root of that sum. In the two-variable case, the distance is analogous to finding the length of the hypotenuse in a triangle; that is, it is the distance as the crow flies. A review of cluster analysis in health psychology research found that the most common distance measure in published studies in that research area is the Euclidean distance or the squared Euclidean

distance.

To calculate distance between two clusters it is required to define two representative points from the two clusters. Different linkage measures like "single linkage", "complete linkage", "average linkage" *etc* have been proposed for this purpose.

2.2. Similarity measure for mixed type data

The above mentioned dissimilarity/similarity measures are applicable to continuous type data only. But generally we work with mixed type data sets which includes different types like continuous, discrete, binary, nominal, ordinal *etc.* Gower (1971) has proposed a general measure known as Gower's coefficient of similarity. Two individuals i and j may be compared on a character k and assigned a score s_{ijk}. There are many ways of calculating s_{ijk}, some of which are described below.

Corresponding to n individuals and p variables, Gower's similarity index S_{ij} is defined as

$$S_{ij} = \Sigma_{k=1}^{p} s_{ijk} / \Sigma_{k=1}^{p} \delta_{ijk} (i, j = 1, 2, \ldots n)$$

where δ_{ijk} = 1 when character k can be compared for observations i and j
= 0 otherwise

For continuous (quantitative) variables with values $x_{1k}, x_{2k}, \ldots, x_{nk}$ for the kth variable

$$s_{ijk} = 1 - \mid x_{ik} - x_{jk} \mid / R_k$$

where R_k is the range of the variable k and may be the total range in population or the range in the sample.

For a categorical (qualitative) character with m categories ($m = 2$ for binary variable)

s_{ijk} = 0 if i and j are totally different
= q (positive fraction) if there is some degree of agreement
= 1 when i and j are same

2.3. Linkage measures

To calculate distance between two clusters it is required to define two representative points from the two clusters. Different methods have been proposed for this purpose. Some of them are listed below.

Single linkage: One of the simplest methods is single linkage, also known as the nearest

neighbor technique. The defining feature of the method is that distance between clusters is defined as the distance between the closest pair of objects, where only pairs consisting of one object from each cluster are considered.

In the single linkage method, d_{rs} is computed as $d_{rs} = \text{Min } d_{ij}$, where object i is in cluster r and object j is in cluster s and d_{ij} is the distance between the objects I and j. Here the distance between every possible object pair (i, j) is computed, where object i is in cluster r and object j is in cluster s. The minimum value of these distances is said to be the distance between clusters r and s. In other words, the distance between two clusters is given by the value of the shortest link between the clusters. At each stage of hierarchical clustering, the clusters r and s, for which d_{rs} is minimum, are merged.

Complete linkage: The complete linkage, also called farthest neighbor, clustering method is the opposite of single linkage. Distance between clusters is now defined as the distance between the most distant pair of objects, one from each cluster. In the complete linkage method, $d - rs$ is computed as $d_{rs} = \text{Max } d_{ij}$, where object i is in cluster r and object j is cluster s. Here the distance between every possible object pair (i, j) is computed, where object i is in cluster r and object j is in cluster s and the maximum value of these distances is said to be the distance between clusters r and s. In other words, the distance between two clusters is given by the value of the largest distance between the clusters. At each stage of hierarchical clustering, the clusters r and s, for which d_{rs} is minimum, are merged.

Average linkage: Here the distance between two clusters is defined as the average of distances between all pairs of observations, where each pair is composed of one object from each group. In the average linkage method, d_{rs} is computed as $d_{rs} = Trs/(Nr \times Ns)$ where Trs is the sum of all pairwise distances between cluster r and cluster s. Nr and Ns are the sizes of the clusters r and s respectively. At each stage of hierarchical clustering, the clusters r and s, for which d_{rs} is the minimum, are merged.

Minimax Linkage: This was introduced by Bien and Tibshirani (2011). For any point x and cluster G, define

$$d_{\max}(x, G) = \max_{y \in G} d(x, y)$$

as the distance to the farthest point in G from x. Define the minimax radius of the cluster G as

$$r(G) = \min_{x \in G} d_{\max}(x, G)$$

that is, find the point $x \in G$ from which all points in G are as close as possible. This minimizing point is called the prototype for G. It may be noted that a closed ball of radius

$r(G)$ centered at the prototype covers all of G. Finally we define the minimax linkage between two clusters G and H as

$$d(G, H) = r(GUH)$$

that is, we measure the distance between clusters G and H by the minimax radius of the resulting merged cluster.

It is very important to choose a proper linkage measure in a particular situation. A liberal attitude always leads to single linkage whereas a conservative attitude leads to complete linkage. Minimax is a good choice when one tries to avoid a wrong decision (loss is more important than gain) and without any prior belief, average linkage may give the best answer.

2.4. Optimum number of clusters

Usually the number of clusters are determined from the dendrogram and validated by the physical properties. We specify a horizontal line for a particular similarity/dissimilarity value and the clusters below this line are selected as optimum. But some mathematical rules (thumb rules) are also available which are based on between cluster and within cluster sum of squares values. If we denote by k, the number of clusters and define by $W(k)$ the sum of the within cluster sum of squares for k clusters then the values of $W(k)$ will gradually decrease with increase in k and that "k" may be taken as optimum where $W(k)$ stabilizes. For detailed discussion on may follow the link http://www.cc.gatech.edu/∼hpark/papers/cluster JOGO.pdf (by Jung *et al.* (2002)).

3. Partitioning Clustering - k-means Method

The k-means algorithm (MacQueen, 1967) assigns each point to the cluster whose center (also called centroid) is nearest. The center is the average of all the points in the cluster that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. This method can be used for clustering of objects and not variables.

This method starts with a value of k. We will discuss later the method of selection of the value of k. Then we randomly generate k clusters and determine the cluster centers, or directly generate k seed points as cluster centers. Assign each point to the nearest cluster center in terms of Euclidean distance. Re-compute the new cluster centers. Repeat until some convergence criterion is met *i.e.* there is no reassignment. The main advantages of this algorithm are its simplicity and speed which allows it to run on large data sets. Its disadvantage is that it is highly dependent on the initial choice of clusters. It does not yield the same result with each run, since the resulting clusters depend on the initial random assignments. It maximizes inter-cluster variance and minimizes intra-cluster variance.

The advantages of partitioning method are as follows:

(a) A partitioning method tries to select best clustering with k groups which is not the goal of hierarchical method.

(b) A hierarchical method can never repair what was done in previous steps.

(c) Partitioning methods are designed to group items rather than variables into a collection of k clusters.

(d) Since a matrix of distances (similarities) does not have to be determined and the basic data do not have to be stored during the computer run partitioning methods can be applied to much larger data sets.

For k-means algorithms (Hartigan, 1975) the optimum value of k can be obtained in different ways. On the basis of the method proposed by Sugar and James (2003), by using k-means algorithm first determine the structures of clusters for varying number of clusters taking $k = 2, 3, 4$ *etc.* For each such cluster formation compute the values of a distance measure

$$d_K = (1/p) \min_x E[(x_k - c_k)'(x_k - c_k)]$$

which is defined as the distance of the x_k vector (values of the parameters) from the center c_k (which is estimated as mean value), p is the order of the x_k vector. Then the algorithm for determining the optimum number of clusters is as follows. Let us denote by d'_k the estimate of d_k at the kth point which is actually the sum of within cluster sum of squares over all k clusters. Then d'_k is the minimum achievable distortion associated with fitting k centers to the data. A natural way of choosing the number of clusters is plot d'_k versus k and look for the resulting distortion curve. This curve is always monotonic decreasing. Initially one would expect much smaller drops *i.e.* a levelling off for k greater than the true number of clusters because past this point adding more centers simply partitions within groups rather than between groups.

According to Sugar and James (2003) for a large number of item versus transformed d'_k. Then calculate the jumps in the transformed distortion as

$$J_k = (d_k'^{(-(p/2))} - d_{k-1}'^{(-(p/2))})$$

Another way of choosing the number of clusters is plot J_k versus k and look for the resulting jump curve. The optimum number of clusters is the value of k at which the distortion curve levels off as well as its value associated with the largest jump.

The k-means clustering technique depends on the choice of initial cluster centers (Chattopadhyay *et al.*, 2012). But this effect can be minimized if one chooses the cluster centers

through group average method (Milligan, 1980). As a result, the formation of the final groups will not depend heavily on the initial choice and hence will remain almost the same according to physical properties irrespective of initial centers. In MINITAB package, the k-means method is almost free from the effect of initial choice of centers as they have used the group average method.

3.1. Advantages and disadvantages of k-means algorithm

The main advantages of this algorithm is that it is very fast (in terms of computational speed), robust, easy to understand and interpret. In fact the algorithm has been modified by Hartigan and Wong (1979) which speeds up the algorithm and is used most commonly in the community. The open-source statistical computing environment R (https://cran.r-project.org/); the software which is used in this entire work; has the built-in function kmeans() which implements the above discussed version of the k-means algorithm as its default. The algorithm is very much well suited for the data which are distinct and well-separated from each other. The clusters thus formed are tight and often tighter than the Hierarchical Clustering method, especially when the clusters are globular. But the algorithm suffers due to a number of reasons. k -means depends heavily on the initialization/ seeds. The algorithm assumes the joint distribution of the features within each cluster to have equal variance and to be independent of each other. This assumption is hard to satisfy more than often. Correlation between the features breaks this assumption. k-means cannot find non-convex clusters or the clusters with unusual shapes or overlapping clusters. Finally, this algorithm requires a priori knowledge on the number of clusters/groups to be formed. This is most commonly tackled by using the method proposed by Sugar and James (2003) which has been discussed earlier. Jump-Statistic as a mean of determining the number of clusters "k" is very popular and widely accepted measure. Other possibilities are the uses of gap statistic or silhouette index.

3.2. Example using k-means algorithm

The Fisher's *Iris* data set is a multivariate data set introduced by R.A. Fisher (Fisher (1936)). It is also known as Anderson's *Iris* data set because Edgar Anderson collected the data to quantify the morphologic variation of Iris flowers of three related species. The data set consists of 50 samples from each of three species of Iris (Iris setosa (type-3), Iris versicolor (type-2) and Iris virginica (type-1)). Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres.

We have performed k-means clustering of the data on the basis of the four variables viz. sepal length, sepal width, petal length and petal width. Choosing $k = 3$, we have divided the 150 observations into three groups in order to verify whether we can identify three groups corresponding to three species. From our analysis it is clear that k-means method has correctly identified Iris setosa (type-3) species for all the 50 cases where as there are some errors corresponding to types 1 and 2. For type 2 three cases and for type 1 fourteen cases had wrongly identified. The summary result for k-means clustering is given below:

Number of clusters: 3

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	39	25.414	0.732	1.552
Cluster2	61	38.291	0.731	1.647
Cluster3	50	15.151	0.482	1.248

4. Clustering of Variables

The hierarchical clustering method can also be used for clustering of variables on the basis of the observations. Here instead of the distance matrix one may start with the correlation matrix (higher correlation indicating similarity of variables). The linkage measures as listed in the previous section will not be applicable for variable clustering. In order to measure similarity/dissimilarity between two clusters of variables, one may either use the correlation between first principal components corresponding to the two clusters or the canonical correlations.

Dimensionality reduction techniques like Principal Component Analysis (PCA) or Independent Component Analysis (ICA) could alternatively be used for variable clustering. The variables with larger loading belonging to a particular component may be considered to be in the same cluster.

4.1. Principal Component Analysis (PCA)

In this technique, given a data set of observations on correlated variables, an orthogonal transformation is performed to convert it into a set of uncorrelated variables called the principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance. One rule of thumb is to consider those components whose variances are greater than one in the reduced space. Principal components are guaranteed to be independent only if the variables are jointly normally distributed.

4.2. Independent Component Analysis (ICA)

One of the most recent powerful statistical techniques for analyzing large data sets is independent component analysis (ICA), see Comon (1994) for the original description of ICA. Such data sets are generally multivariate in nature. The common problem is to find a suitable representation of the multivariate data. For the sake of computational and conceptual simplicity such representation is sought as a linear transformation of the original data. Principal component analysis, factor analysis, projection pursuit are some popular methods for linear transformation. But ICA is different from other methods, because it looks for the components in the representation that are both statistically independent and non-Gaussian.

In essence, ICA separates statistically independent component data, which is the original source data, from an observed set of data mixtures. All information in the multivariate data sets are not equally important. We need to extract the most useful information. Independent component analysis extracts and reveals useful hidden factors from the whole data sets. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. See Hyvarinen *et al.* (2001), Comon and Jutten (2010) and Lee (1998) for book length discussions on ICA. ICA can be applied in various fields like neural network (Fiori, 2003), studying EEG data (Bartlett *et al.*, 1995), speech processing (Kumaran *et al.* 2005), brain imaging (McKeown *et al.*, 1997), signal separation (Adali *et al.*, 2009), telecommunications (Hyvarinen *et al.*, 2002), econometrics (Bonhomme and Robin, 2009), *etc.* Chattopadhyay *et al.* (2012) has applied ICA for astronomical data set.

4.3. Conversion of directional data to linear

Note that PCA or ICA has been developed for linear continuous data but if one variable, is circular in nature then the method will not work.. But it is not immediate how to include this type of data for clustering directly or through PCA or ICA. If a density plot of the data show the circular variable has a bimodal distribution and the two modes are near 0^o and 200^o, we may be motivated to consider two main directions, say east and west (approximately), which correspond to 0^o and 180^o.

Chattopadhyay *et al.* (2015) proposed a method of conversion from circular to linear where they considered standard cosine angular distance of an angle θ from a fixed angle ϕ, defined by $d_\phi = 1-\cos(\theta-\phi)$, which is in the linear scale, and $d \in [0,2]$. Thus, for a circular variable θ , we may consider two distances $d_0 = 1-\cos(\theta-0^o)$ and $d_{180} = 1-\cos(\theta-180^o)$, both of which are linear. So, instead of taking θ in our analysis, we may consider the pair $(d_{\max}, d_{sign})$, where $d_{\max} = \max(d_0, d_{180})$ and $d_{sign} = +1$ if $d_{\max} = d_0$ and $d_{sign} = -1$ if $d_{\max} = d_{180}$. Alternately, if we want to ignore the sign we can work with $\theta^* = 2 \times \theta$, which is approximately unimodal with mode near 45^o. We may work with $d^* = 1-\cos(\theta^*-45^o)$.

5. Incomplete Data problems

Statistical analysis with missing data is an important problem as the problem of missing observation is very common in many situations. During the last two decades different methods have been developed to tackle the situation.One possible way to handle missing values is to remove either all features or all objects that contain missing values. Another possibility is imputation where we fill in the missing values by inferring new values for them. The imputation method may not be applicable to some astronomical data sets (Chattopadhyay, 2017) as the missing value may arise from physical process and imputing missing values is misleading and can skew subsequent analysis of data. For example, the Lyman break technique (Giavalisco, 2002) can identify high-redshift galaxies based on the absence of detectable emissions in bands corresponding to the FUV rest frame of the objects. Such high-redshift galaxies were previously unobservable.

Missing values occur for a variety of reasons, from recording problems to instrument limitations to unfavorable observing conditions. In particular, when data are combined

from multiple archives or instruments, it is virtually certain that some objects will not be present in all of the contributing sources. Little and Rubin (1987) identified three models for missing data. When values are Missing At Random (MAR, MCAR), imputation may be a reasonable approach since the values may be predicted from the observed values. The third type of missing values are Not Missing at Random (NMAR), when the value itself determines whether it is missing. This is precisely the case when objects fall below a detector's sensitivity threshold. There is no way to impute these values reliably, because they are never observed.

Under the regression set-up with predictor X and response Y, missing value problems often arise. To decide how to handle missing value problems, primarily we need to know why these values are missing. We may explain the above three general missing mechanisms in the following manner.

A variable value is missing completely at random (MCAR) if the probability of missingness is the same for all units. Under the regression set-up if the missing values are independent of both response and predictor then these are called missing completely at random. Most missingness is not completely at random. A more general assumption,missing at random (MAR), is that the probability of a variable value is missing depends only on variable information. Under the regression set up, if the missing value depends on predictor but not on response then these are called missing at random.

Missingness is no longer at random if it depends on information that has not been recorded and this information also predicts the missing values. In particular, a difficult situation arises when the probability of missingness depends on the variable itself. Under the regression set-up this type of situation arises when probability of response depends on both response and predictor.For statistical inference with missing information, we usually assume that the missingness pattern is MCAR or MAR. But in many situations these assumptions are not valid.

In clustering algorithms, different packages use different types of imputation techniques like mean imputation, hot deck imputation etc. In order to estimate the missing values properly one should take care of this fact. Use of EM algorithm is usually recommended.

6. Conclusion

From the above discussions it is very clear that although clustering and dimension reduction problems are widely used under different disciplines by scientists from several areas, one should always take care of the nature of data in order to apply the methods successfully. In the introduction we have listed several such problems and only a few are discussed in latter sections. It is quite expected that one may identify many other computation based problems which are not listed here.

References

Adali, T., Jutten, C., Romano, J. M. T. and Barros, A. K. (2009). Independent component analysis and signal separation. In: *Proceedings of 8th international conference, ICA 2009*, Conference held in Paraty, Brazil, March 15–18, 2009. Springer,Berlin.

Bien., J. and Tibshirani R. (2011). Hierarchical Clustering with prototypes via minimax linkage. *Journal of the American Statistical Association,* **106(495)**, 1075-1084.

Bonhomme, S. and Robin, J. M. (2009). Consistent noisy independent component analysis. *Journal of Econometrics,* **149**, 12–25.

Chattopadhyay, A. K. (2017). Incomplete data in Astrostatistics. *Wiley StatsRef: Statistics Reference Online,* 1-12. John Wiley & Sons.

Chattopadhyay, A. K., Mondal, S. and Biswas, A. (2015). Independent component analysis and clustering for pollution data. *Environmental and Ecological Statistics,* **22**, 33-43.

Chattopadhyay, A. K., Mondal, S. and Chattopadhyay, T. (2012). Independent component analysis for the objective Classification of globular clusters of the galaxy NGC 5128. *Computational Statistics and Data Analysis,* **56(1)**, 17-32.

Comon, P. (1994). Independent component analysis, a new concept? *Signal Process,* **36**, 287–314.

Comon, P. and Jutten, C. (2010). *Handbook of Blind Source Separation, Independent Component Analysis and Applications.* Academic Press, Oxford, UK

De, T., Chattopadhyay, T. and Chattopadhyay, A. K. (2013). Comparison among clustering and classification techniques on the basis of galaxy data. *Calcutta Statistical Association Bulletin,* **65**, 257–260.

Fiori, S. (2003). Overview of independent component analysis technique with an application to synthetic aperture radar (sar) imagery processing. *Neural Networks,* **16(3–4)**, 453–467.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics,* **7(2)**, 179-188.

Fraix-Burnet, D., Thuillard, M. and Chattopadhyay, A. K. (2015). Multivariate approaches to classification in extragalactic Astronomy. *Frontiers in Astronomy and Space Science,* **2**, 1-17.

Giavalisco, M. (2002). Lyman-Break Galaxies. *Annual Reviews of Astronomy and Astrophysics,* **40**, 579-641.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics,* **27(4)**, 857-871.

Hartigan, J. A. (1975). *Clustering Algorithms.* John Wiley & Sons.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* **28**, 100-108.

Hyvarinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis.* Wiley, New York.

Hyvarinen, A., Karhunen, J. and Oja, E. (2002). *Telecommunications. In: Independent Component Analysis,* Ch 2, Wiley, New York.

Jung, Y., Park, H., Du, D. and Drake, B. L. (2003). A decision criterion for the optimal

number of clusters in hierarchical clustering. *Journal of Global Optimization,* **25(1)**, 91-111.

Kumaran, R. S., Narayanan, K. and Gowdy, J. N. (2005). Myoelectric signals for multimodal speech recognition. *Interspeech*, 1189–1192.

Lee, T. W. (1998). *Independent Component Analysis: Theory and Applications.* Kluwer, Boston, MA.

Little, J. A. R. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* John Wiley & Sons, New York.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In: *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability,* 281-297.

McKeown, M. J., Makeig, S., Brown, G. G., *et al.* (1997). Analysis of fmri data by decomposition into independent components. *American Academy of Neurology Abstract,* **48**, A417.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika,* **45(3)**, 325-342.

Sugar, A. S. and James, G. M. (2003). Finding the number of clusters in a data set: An information theoretic approach. *Journal of the American Statistical Association,* **98**, 750-763.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 263–274

Estimation and Sample Size Calculation for Service Utilization Data

Dulal K. Bhaumik[1] **and Subhash Aryal**[2]
[1]*Department of Psychiatry, University of Illinois at Chicago, Chicago, USA*
[2]*School of Nursing, University of Pennsylvania, Philadelphia, USA*

Received: 01 June 2020; Revised: 29 July 2020; Accepted: 06 August 2020

Abstract

Health service utilization research suffers from lack of statistical methods to analyze routinely obtained zero-inflated correlated outcome data from multilevel longitudinal studies. Parameter estimation suffers from use of maximum likelihood based approach involving cumbersome integration which results in lack of model convergence and utilization of considerable computing resources. Similarly, sample size to conduct randomized controlled trials are estimated using either inappropriate linear models or simplified non-linear models which ignore multiple levels of nesting resulting in severely under powered studies. We propose a robust estimation method based upon Laplace approximation to estimate parameters and derive formula to compute required sample size employing multiple levels of nesting.

Key words: Health services; Zero-inflated data; Laplace approximation; Sample size.

1. Introduction

Health services (HS) researchers are widely using hierarchical mixed-effects models for analysis of their correlated clustered and longitudinal data. Parameters are generally estimated by maximum marginal likelihood, empirical Bayes estimation, fully Bayesian strategies and Generalized Estimating Equations (GEE), and hypotheses are tested using t, χ^2 or F tests. Furthermore, considerable computer software has now been developed and is either freely available over the Internet or commercially available. However, this area is still challenged by a lack of statistical methods appropriate for addressing some unique aspects of health services research data. A major problem in HS data is missing outcomes as well as covariate values. Another equally complex problem is the profusion of zero values in count data such as service units or costs, which results in a highly skewed distribution. To address these issues, in many instances missing values are imputed, and hierarchical zero-inflated mixed models are utilized even though non-convergence issues prevail in estimation. In such models, justification of using random effects in terms of testing its variance components is avoided because of unavailability of user-friendly testing procedures at the boundary value. Another challenge is the determination of sample size, as inadequate sample size runs the

Correponding Author: Dulal K. Bhaumik
Email: dbhaumik@uic.edu

risk of inflated false positive findings (Type I error), while fitting the model with an excessive number of random-effects can mask significant relationships (Type II error). This manuscript addresses issues pertaining to parameter estimation and sample size calculation for HS researches and bridge a critically important gap in the designing stage of health research studies in general, and mental health services research in particular.

In *mental health services research* investigators have studied service utilization, barriers to service utilization, disparities in service utilization and cost associated with service utilization (Hacker, *et al.* 2015). Similarly, service utilization data are also found in research on *general health care* (Gilbert, *et al.* 2012), *dentistry* (Moghimbeigi, *et al.* 2008), *occupational health* (Min and Agresti, 2005) and *substance abuse* (Bandhophadyay, *et al.* 2011). Our careful analysis of the literature revealed that service utilization research studies regularly encounter the problem of missing outcomes and covariates, zero-inflation, over-dispersion, and non-convergence of statistical models. In addition, this area requires feasible parameter estimation techniques and sample size determination methods and user friendly software for analysis of HS data. Particularly there is an lack of suitable software for sample size determination when zero-inflation is expected in a hierarchical design with random-effects. Most of the existing methods either assume linear model or completely ignore the random-effects by using the GEE approach. As such there is a genuine need for sample size methodologies and more importantly software to calculate sample size for service utilization research with zero-inflation.

In Section 2, we present some motivating examples. In Section 3, we discuss methods to model service utilization data. In Section 4, we derive formulae for sample size calculation for studies employing hierarchical designs resulting in zero-inflated outcomes . In Section 5, we present some concluding remarks on service utilization data.

2. Motivating Examples

Next we present two HS research studies to motivate the need for theoretical developments.

2.1. Example 1

The first problem was investigated by Atkins, *et al.* (2015) and compared group differences between Links to Learning (L2L), a school and home-based mental health service model, and Service As Usual (SAU) on several domains including mental health service use, classroom observations of academic engagement, teacher report of academic competence and social skills, parent report of social skills, teacher and parent report of problem behaviors, daily hassles, and curriculum-based measures. Services were Medicaid-funded through 4 social service agencies ($N = 17$ providers) in 7 schools ($N = 136$ teachers, 171 children consists of 124 boys (50 control + 74 Link), and 47 girls (17 control + 30 Link)) in a 2 (Links to Learning vs. services as usual) 6 (pre- and post tests for 3 years) longitudinal design with random assignment of schools to conditions. Services as usual consisted of supported referral to a nearby social service agency. The primary interest was in *differential change over time*. A three-level hierarchical design with multiple observations from students nested within schools was used to analyze the study data. The model included covariates at both

the student level (grade, gender) and classroom and teacher level (classroom assessment scoring system, teacher sense of efficacy scale, organizational health inventory-elementary and Quality of teacher work life survey). The conclusion of the study was that community mental health services targeting empirical predictors of learning can improve school and home behavior for children living in high-poverty urban communities. For a full description of data decomposition, analysis methods, missing value problems and significant results we refer to Atkins, *et al.* (2015). Some key difficulties encountered during the analysis of this dataset were (i) problem of missing data (more than 41.79% in control and 52.88% in L2L data), (ii) differential measurement errors, and (iii) the problem of unreliable measures of some outcomes and covariates. This analysis inspired us to develop novel statistical methods to estimate missing outcomes when corresponding covariates are known, and missing covariates when corresponding outcome measures are known, but in both situations causes of missingness are unknown.

2.2. Example 2

Our second example is based on the work by Cook, *et al.* (2019) and Bhaumik, *et al.* (2019). They recently analyzed data from randomized trial of self-directed care in Texas public mental health system. In this study, the Zero-Inflated Negative Binomial (ZINB) and log-gamma models were used to test the effect of an experimental intervention called self-directed care, in which patients have greater control over service delivery funds and can choose to hire and fire specific service providers. The authors applied the ZINB model to analyze service utilization and log-gamma model for analysis of cost data. A total of 216 subjects with serious mental illness receiving care in the Texas public mental health system were randomly assigned with their consent to receive services as usual (= 102) or the experimental intervention (= 114) and followed for 24 months. The primary hypothesis was that the experimental intervention would produce superior client outcomes at 12 and 24-month follow-up and this proved to be the case. However, since the intervention was intended to be budget neutral (*i.e.*, to cost no more than services delivered through the usual system), secondary analysis of service costs was required. Administrative data were obtained from the local area's managed care company in the form of "shadow claims" and grouped into costs during the first and second years of program participation and for both years combined. Over the two years of the program, experimental participants incurred a total average per person cost of $5,239 (s.d. = 5,500)$ compared to an average of $5,493 (s.d. = 8,268)$ per person in the control group. This difference was non-significant, as expected. However, costs for specific service types had the additional challenge of being zero-inflated, with many non-users of some services. Consequently, the authors used ZINB/log-gamma models for individual services/costs, which model the mixture of the likelihood of having zero service/costs in each category, and the relative amount of service/costs among users. As shown in Table 1, experimental condition subjects were more likely than controls to have zero costs for psychiatric rehabilitation, case management, and skills training, but there were no differences in costs for users of these three services. On the other hand, there was no difference in the likelihood of zero costs for medication management, but among users of this service, costs were significantly lower for the experimental group. For the service of psychotherapy, the experimental group was less likely than controls to have zero costs, and costs were higher for

experimental than control subjects. When the authors used linear mixed-effects regression analysis of these individual service costs adjusting for time, the experimental condition costs were lower for psychiatric rehabilitation, skills training, and medication management, and higher for psychotherapy. The linear mixed-effects model cannot provide information separately for zero costs and costs for users. Clearly, ZINB modeling provided a more complex and complete picture of cost differences where they existed.

Table 1: ZINB analysis modeling first, likelihood of zero costs, and second, costs among service users

	Psychiatric Rehab		Case Mgmt		Skills Training		Medication Mgmt		Psycho-therapy	
	Estimate	p	Estimate	p	Estimate	p	Estimate	p	Estimate	p
Pred. zero cost	0.755	.007	1.183	.001	1.484	<.001	0.731	.076	-1.602	<.001
Pred. costs users	-0.297	.124	0.076	.855	-0.490	.151	-0.439	.001	1.134	.001

There are important implications in these results for health services researchers and the service system administrators and policy makers who use these study findings. First, ZINB allowed us to use a "two-part model" in analyzing utilization data (Manning, *et al.* 2005). As noted by Diehr, *et al.* (1999), the decision to have any use of a service is most likely made by the person and so is primarily associated with personal characteristics, while the cost per user may be more related to features of the health care system. When the goal is understanding the system, a two-part model is preferred because it enables researchers to differentiate between influences on the propensity to use a service, and factors affecting how much of the service is used and at what cost once the individual enters the service delivery system (Diehr, *et al.* 1999).

3. Model

In this section we present some models for analysis of count data inflated with zeros. We begin by positing c centers and n_i subjects nested within the ith center. The total number of subjects $N = \sum_{i=1}^{c} n_i$ are randomized into intervention and control groups. We assume that each subject may utilize mental health services longitudinally over T different time periods. The outcome variable y_{isjt} measures the number of times the jth subject from the ith center nested within the sth intervention group used mental health services for the tth time period. The log likelihood function for all observations $\boldsymbol{y}_i = (y_{i111}, \cdots, y_{i2nT})^t$ nested within the ith center is

$$logit(\pi_{isjt}) = \gamma_0 + \gamma_1 g(t) + \gamma_2 x_{ijk} + \gamma_3 x_{ijk} g(t) + \boldsymbol{\gamma}^{t*} \boldsymbol{w_{ijk}} + \nu_{i0} + \nu_{i1} g(t) + \delta_{isj0} + \delta_{isj1} g(t), \quad (1)$$

$$log(\lambda_{isjt}) = \beta_0 + \beta_1 g(t) + \beta_2 x_{ijk} + \beta_3 x_{ijk} g(t) + \boldsymbol{\beta}^{t*} \boldsymbol{z_{ijk}} + \upsilon_{i0} + \upsilon_{i1} g(t) + \delta^*_{isj0} + \delta^*_{isj1} g(t), \quad (2)$$

Here $f(y_{isjt})$ be the probability mass function of a Poisson distribution for a zero inflated Poisson (ZIP) model, and if the model is zero inflated negative binomial then $f(y_{isjt} = k) = (1 - \pi_{ij}) \frac{\Gamma(k + \lambda_{ij}^{1-d}/\alpha)}{k \Gamma(-\lambda_{ij}^{1-d}/\alpha)} (1 + \alpha \lambda_{ij}^{d})^{-\lambda_{ij}^{1-d}/\alpha} (1 + \lambda_{ij}^{-d}/\alpha)^{-k}$. The dispersion parameter

α is always non-negative and does not depend on covariates. This distribution reduces to the ZIP distribution when $\alpha \to 0$. The constant d is used to identify a particular form of a negative binomial distribution (see Saha and Dong, 1997). For various forms of negative binomial distributions we refer to (McCullagh and Nelder 1989, Ridout, *et al.* 2001, Yau, *et al.* 2003). Xiang, *et al.* (2007) used a score test for testing the over-dispersion of a ZIP regression model against the ZINB alternative (*i.e.*, $\alpha = 0$ in a ZINB model). Both ZIP and ZINB regression models will be inappropriate for fitting data with zero deflation at any settings of the explanatory variables. A useful model for such a situation is the "Hurdle" model proposed first by Mullahy (1986) that separately handles the zero observations and the positive counts. An advantage of the Hurdle model is that it can handle both the zero inflation and zero deflation. The downside of this model is that all zero counts are structural whereas ZINB and ZIP models allow both structural and functional zeros (Pardoe and Durham, 2003). In the Hurdle model, $g_2(y_{ij}) = f(k)/(1 - f(0))$.

In model (1), $\gamma_0 + \gamma_1 g(t)$ and $\gamma_0 + \gamma_1 g(t) + \gamma_2 x_{ijk} + \gamma_3 x_{ijk} g(t)$ are the fixed linear trends for the control group, and for the intervention group, respectively. Thus γ_3 differentiates the slope of the treatment group from the control group of service utilization and β_3 has a similar interpretation of frequency of service utilization. Exponentiation of γ_3 and β_3 provides the odds ratio and risk ratio respectively. Note that these parameter estimates are subject-specific, which indicates the effectiveness of the intervention at the individual level. The interpretation of $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}^*$ is of considerable interest. The gamma parameters ($\boldsymbol{\gamma}^*$) describe the effects of the covariates on the likelihood of service utilization, whereas the beta parameters ($\boldsymbol{\beta}^*$) describe the effects of the same or possibly different covariates on the intensity of service utilization. Also, $\nu_{i0} + \nu_{i1} g(t)$ is the random linear trend for the ith site effect. The correlation between subjects nested within the same site is accounted for by the presence of random site effects. Similarly $\delta_{isj0} + \delta_{isj1} g(t)$ is the random linear trend for the jth subject nested within the ith site, and the random linear trend at the subject level takes care of the correlation between multiple observations nested within the same subject. Similar interpretations hold in model (2). The vectors $\boldsymbol{w_{ijk}}$ and $\boldsymbol{z_{ijk}}$ represent the additional fixed covariates such as age, race, sex etc. for the logit and the log-linear components. A three-level ZIP or ZINB longitudinal mixed-effects model can have a total of 12 variance covariance parameters; six components from the binary part (variance for random intercept, variance for random slope and their covariance for subjects and for communities), and a parallel set of six variance components from the count part of the model. Even though 12 variance components in the above models seems to be a reasonable assumption, in actuality, we do not know how many of them are really significant. Keeping all of them may over-saturate the model. To select an appropriate model we generally use deviance, Akaike information criterion (AIC), and Bayesian information criterion (BIC). Several authors have recently noted that AIC and BIC are not appropriate for model selection when the sample size is small (Kass and Raftery, 1995, Seghouane, 2006, Chen, *et al.* 2008, Tu and Xu, 2012). To resolve this issue, there is a need for alternative approaches to evaluate the significance of variance components.

3.1. Estimation of model parameters

The goal of this section is to derive and use the marginal likelihood function of fixed parameters (*i.e.* γ and β) conditioning on the data and a suitable estimate of random effects (posterior mode). At the initial stage, it is assumed that variance components are known. The numerical integration over the space of random-effects in the estimation process is avoided by approximating the log-likelihood around the starting values of random-effects. In addition, we investigate an alternative procedure based on ordinary Laplace approximation. The convergence rates of both ordinary Laplace and Marginal Maximum Likelihood (MML) (combination of Gaussian quadrature and Newton Rhapson method) is $O(n^{-1})$ (Ghosh, *et al.* 2006, page 206). However, MML requires enormous computational time and often fails to converge for hierarchical Zero-Inflated Data (ZID). Xie, *et al.* (2013) and Gupta, *et al.* (2015) encountered similar convergence problem in their analysis of ZID. On the other hand, the Laplace approximation avoids numerical integration by exploiting a property of the multivariate normal distribution. As a result, this method provides better guarantee of convergence compared to the quadrature methods for hierarchical models. For comparison purposes we also include the penalized quasi likelihood (PQL) approach (Hyede, 1997).

3.2. Comparison of three estimation methods

First, zero-inflated data were simulated under the assumption that all random-effects in the logistic component were stochastically independent from the random-effects in the log linear component. This assumption reduced the complexity of numerical computation. To compare results of parameter estimation obtained by Laplace and quadrature methods for a two-level Poisson and logistic regression mixed-effects models, we set intercept parameters of control and intervention groups at 3 and 0, respectively, and slope parameters at -0.5 and -1, respectively. The variance-covariance matrix of the random slope and random intercept were set at (1, -0.2, 0.05). Based on simulations using a two-level ZI model, we observed in Table that Laplace and quadrature methods produced similar results, whereas results by PQL were unsatisfactory. In addition, we observed in Table that standard errors of these estimates for both Laplace and quadrature methods did not vary significantly. However, the convergence rate obtained by the Laplace method was substantially higher than that of the quadrature method. In addition, the Laplace method required, on average, one-fourth of the computing time required by the MML method (whenever it converged), and the accuracy rates of both Laplace and quadrature were at the same level.

3.3. Computation time and convergence

Another critically important issue in fitting complex models with numerous random-effects is computational time and model convergence. To investigate these issues we used PROC NLMIXED, SAS version 9.4 to fit our models. The computational times for ZIP and ZINB models (i) with fixed-effects, are in terms of seconds, (ii) with mixed-effects having one or two random-effects, are less than 5 minutes for both quadrature and Laplace, (iii) with mixed-effects having three random-effects, are around 80 minutes for quadrature and less than 25 minutes for Laplace, (iv) the quadrature did not converge for both ZIP and ZINB models with four random-effects, whereas, for the same models with four random effects, Laplace converged in two hours. The same data analyzed using GEE took less than

Table 2: Estimation of parameters and standard errors by Laplace, Quadrature and PQL for Poisson and logistic regression models.

	Laplace		Quadrature		PQL	
Parameters	Poisson	Logistic	Poisson	Logistic	Poisson	Logistic
$\beta_0 = \gamma_0 = 3$	2.997 (0.14)	3.31 (0.63)	2.997 (0.14)	3.202 (0.68)	3.022 (0.14)	2.736 (0.87)
$\beta_1 = \gamma_1 = -0.5$	-0.499 (0.04)	-0.558 (0.14)	-0.498 (0.04)	-0.537 (0.14)	-0.478 (0.04)	-0.463 (0.08)
$\beta_2 = \gamma_2 = 0$	0.007 (0.20)	-0.018 (0.64)	0.007 (0.20)	0.039 (0.62)	-0.016 (0.25)	-0.082 (0.72)
$\beta_3 = \gamma_3 = -0.5$	-0.503 (0.07)	-0.543 (0.21)	-0.504 (0.07)	-0.557 (0.25)	-0.490 (0.37)	-0.412 (0.54)
$\sigma^2_{\delta_0} = \sigma^2_{\delta_0^*} = 1$	0.979 (0.14)	1.412 (1.53)	0.981 (0.15)	1.606 (1.60)	0.959 (0.25)	0.219 (0.64)
$\sigma_{\delta_{01}} = \sigma_{\delta^*_{01}} = -0.2$	-0.197 (0.04)	-0.573 (0.53)	-0.197 (0.04)	-0.354 (0.40)	-0.204 (0.09)	-0.005 (0.79)
$\sigma^2_{\delta_1} = \sigma^2_{\delta_1^*} = 0.05$	0.086 (0.01)	0.149 (0.15)	0.086 (0.01)	0.154 (0.21)	0.083 (0.11)	0.012 (0.25)

one minute with an exchangeable correlation matrix. The Bayesian approach with three random-effects took 10 minutes to update 1 chain for 10,000 iterations (5000 burn-in, 5000 update), and 15 minutes to update 2 chains for 10,000 iterations (5000 burn-in, 5000 update). We further repeated the simulation study with various levels of missingness and observed that computational time varied significantly between the methods, and non-convergence became a norm rather than an exception, especially when missingness exceeded more than 30%. An alternative approach when convergence persists is the use of "Maximum A Posteriori (MAP)" estimation that sets the initial value of the parameters to their posterior mode, and uses adaptive quadrature instead of fixed-point quadrature. Yet another alternative is to use the Laplace approximation at each center, and then perform meta analysis to combine results from centers (Bhaumik, *et al.* 2012, Amatya, *et al.* 2015). Convergence rate for this combination approach is expected to be better as random components at the center level are eliminated. Based on this simulation study, we recommend to use Laplace method (or a combination of Laplace and meta analysis) for estimating parameters of zero inflated models when number of random effects is more than two in order to get consistent estimators avoiding non-convergence issues.

4. Sample Size Determination

In this section we address the issue of sample size determination for hierarchical designs with zero-inflated data.

Statistical methods for the analysis of longitudinal data with clustering of subjects are now routinely applied in mental health service utilization studies. The design of such studies often suffers from poorly specified and often inadequate sample sizes. This is because sample size determination methodology is derived based on a single outcome, or based on longitudinal studies which ignore clustering. The determination of sample sizes when subjects are both repeatedly measured over time and clustered within research sites (*e.g.*, multisite Randomized Controlled Trials (RCTs)) can be erroneous unless both factors, and attrition rates are taken into account. Several authors have developed power analysis for cluster-randomized, and/or repeated measurements studies (Roy, *et al.* 2007, Bhaumik, *et al.* 2008, 2013, Amatya, *et al.* 2013, Kapur, *et al.* 2014). Some of the key features of power calculations include (i) type of randomization (participant level, or site level), (ii) cluster and longitudinal variability, (iii) differential attrition rates over time, and also in different

groups (*i.e.* intervention and control groups), and (iv) proportion of allocations of subjects.

4.1. Theoretical foundation for sample size computation using generalized linear models

Denote the outcome of the ith subject nested within the cth cluster measured at the jth time point by y_{cij}, where $i = 1, \ldots, n$, $c = 1, \ldots, C$, and $j = 1, \ldots, T$. Let $\boldsymbol{y}_{ci} = (y_{ci1}, y_{ci2}, \cdots, y_{ciT})^t$ be a column vector of dimension $T \times 1$ composed of outcomes of the ith subject measured at T different time points. Generalized linear mixed model that links the expectation of $\boldsymbol{y}_{ci}$ to the linear predictor has the following expression: $E(\boldsymbol{y}_{ci}|\boldsymbol{\delta}, \boldsymbol{\gamma}) = \mathbf{h}(\boldsymbol{X}_{ci}\boldsymbol{\beta} + \boldsymbol{Z}_{ci}\boldsymbol{\delta}_{ci} + \boldsymbol{W}_{ci}\boldsymbol{\gamma}_c) = \mathbf{h}(\boldsymbol{\eta}_{ci})$, where $\boldsymbol{\eta}_{ci} = \boldsymbol{X}_{ci}\boldsymbol{\beta} + \boldsymbol{Z}_{ci}\boldsymbol{\delta}_{ci} + \boldsymbol{W}_{ci}\boldsymbol{\gamma}_c$, and $\boldsymbol{X}_{ci}$, $\boldsymbol{Z}_{ci}$ and $\boldsymbol{W}_{ci}$ are design matrices associated with fixed-effects ($\boldsymbol{\beta}$), subject-level random-effects ($\boldsymbol{\delta}_{ci}$), and cluster-level random-effects ($\boldsymbol{\gamma}_c$), respectively. Random-effects $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ are independent and assumed to follow multivariate distributions. Denote the number of clusters by C, number of treatments by S, the covariance matrix of the pseudo observation $\boldsymbol{y}^*$ (obtained by linearizing the real observation $\boldsymbol{y}$) by $\boldsymbol{V}_s$, the noncentrality τ parameter of a noncentral F distribution with degrees of freedom a and b by $H(a, b, \alpha, \tau)$. Assume $\mathbf{G}$: $(S-1) \times 1$ is group indicator vector whose sth element is 1 corresponding to treatment s; 0 otherwise. Denote $Cov(\hat{\boldsymbol{\beta}}) = C^{-1}\boldsymbol{\Gamma}^{-\mathbf{1}}$. The focus now is on testing a set of linear hypotheses related to group-by-time (or a function of time) interaction parameters which are expressed in the following general linear hypothesis set up of the fixed-effect parameters $\boldsymbol{\beta}$, $H_0 : \boldsymbol{L}\boldsymbol{\beta} = 0$ *vs.* $H_1 : \boldsymbol{L}\boldsymbol{\beta} \neq 0$.

4.2. Results

Assume that a study wants to compare S treatments in C centers utilizing a longitudinal design of length T, and an allocation vector of $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_S)^t$. Further assume that each center wants to use n subjects and randomization is performed at the center level, i.e. all subjects in a given center receive the same treatment assigned to that particular center. Let the proportion of dropouts in centers receiving the sth treatment be $\boldsymbol{\xi}_s = (\xi_{s,1}, \ldots, \xi_{s,T})^t$. In order to attain at least $(1 - \tau)100\%$ power for the test specified in H_1 at an alternative value of $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, the required number of subjects n per center should maintain the following constraint:

$$n \;\geq\; \min\{j : \hat{\lambda}(j) \geq H(S-1, C-S, \alpha, \tau)/C\}, \tag{3}$$

where $\hat{\lambda}(j) = (\boldsymbol{L}\boldsymbol{\beta}^*)^t(\boldsymbol{L}\hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{L}^t)^{-1}(\boldsymbol{L}\boldsymbol{\beta}^*)$. An arbitrary value of C cannot provide a valid solution of (3). Equation (3) provides a feasible solution only when $C \geq C^*$, where

$$C^* = H(S-1, C-S, \alpha, \tau)/(\boldsymbol{\beta}^{*t}\boldsymbol{L}^t[\boldsymbol{L}((\boldsymbol{U}^t\boldsymbol{\Delta}_\pi\boldsymbol{U}) \otimes \boldsymbol{\Sigma}_\gamma^{-1})^{-1}\boldsymbol{L}^t]^{-1}\boldsymbol{L}\boldsymbol{\beta}^*), \tag{4}$$

where, $\boldsymbol{\Delta}_\pi$ is a diagonal matrix with diagonal elements π_s, $\boldsymbol{U} = (\boldsymbol{u}_1^t, \cdots, \boldsymbol{u}_s^t)^t$ and $\boldsymbol{u}_s = (1 \quad \boldsymbol{G}_s^t)^t$ and $\otimes$ is the Kronecker product. Thus, C^* is the lower bound of C and is independent of n. The proof is mathematically intensive and lengthy, hence is not given here (see Amatya and Bhaumik (2018) for complete derivation). This result suggests that at least C^* clusters are necessary for a cluster randomized study to achieve the desired level of power $1 - \tau$. As C increases (starting from C^*), the requirement for the number of subjects decreases, provided all other parameters remain fixed. An explicit expression of C^*

is given in Amatya and Bhaumik (2018). In order to evaluate the flexibility of this exciting result, we did some robustness studies via simulations (i) changing symmetric distributions (of random effects) to right skewed gamma distributions, (ii) relaxing the constraint of equal sample sample sizes of every center to 10% variations. The simulated power (under various parametric combinations, attrition rates, and model violations of types (i) and (ii)) was never less than 76% when it was fixed at 80%. Comprehensive results are reported in Amatya and Bhaumik (2018). A testing procedure with inflated Type I error rates will require fewer samples, but such a test will often show significance when the intervention effect is actually non-significant. On the other hand, a very conservative test will require more resources to attain the same target power (*e.g.*, 80%) compared to an exact test. Our proposed procedure avoids both scenarios. In order to demonstrate how fatal it can be in terms of power, when inappropriate methods are used for sample size determination we compared our proposed method with two existing methods by Murray (1998) and Heo, *et al.* (2013) designed for linear models. Results are presented in Table where for various values of between-cluster variation in slopes ($\sigma^2_{\gamma 22}$) we compute cluster size and corresponding power. Note that power for both the existing methods is substantially lower than what was targeted at 80%.

Table 3: Comparison of required number of clusters estimated from Murray (1998), Heo, *et al.* (2013), the proposed method, and the power attained in simulated evaluation

$\sigma^2_{\gamma 22}$	Murray (1998)		Heo, *et al.* (2013)		Proposed	
	C	power	C	power	C	power
0.03	8	.287	12	.367	35	.797
0.04	8	.252	12	.361	41	.797
0.05	8	.237	12	.339	46	.773
0.06	8	.243	12	.297	52	.778
0.07	8	.223	12	.275	58	.767
0.08	8	.218	12	.279	64	.795
0.35	10	.153	12	.216	224	.799

Both the existing methods perform well when outcome is linear, however, they are inappropriate for non-linear outcomes. Hence, sample size methodologies should be developed taking into account all complexities (type of outcome, within and between cluster variation, attrition rate) which is incorporated in our proposed method.

5. Conclusions

Health service utilization researchers regularly conduct multi-center studies which are longitudinal in nature. In these studies multiple correlated measurements are obtained from subjects who are nested within hospitals, schools etc. The distribution of the outcome variable usually is highly skewed with a profusion of zero as a large majority of eligible subjects never utilize service either due to lack of need or access, and a long right tail as some subjects are mass consumers of service. Sample size estimation methods used to design these hierarchical longitudinal studies with skewed zero-inflated outcome data either rely on completely inappropriate linear models or employ simple designs ignoring various levels of hierarchy

which can result in severe under-estimation of resulting power. We derive a robust method for sample size estimation that incorporates multiple random-effects in a zero-inflated model. Our simulation study showed the proposed method achieved the desired 80% power consistently whereas the other competing approaches under estimated the power severely. During the data analysis phase researchers are routinely forced to exclude important random-effects from their fitted models due to model convergence issue. We propose a novel technique based upon Laplace approximation which considerably reduces the non-convergence and utilizes less computing resources in comparison to the existing methods.

References

Amatya, A. and Bhaumik, D. K. (2018). Sample size determination for multilevel hierarchical designs using generalized linear mixed models. *Biometrics*, **74**, 673–684.

Amatya, A., Bhaumik, D. K. and Gibbons, R. D. (2013). Sample size determination for clustered count data. *Statistics in Medicine*, **32**, 4162–4179.

Amatya, A., Bhaumik, D. K., Normand, S-L.T., Greenhouse, J., Kaizar, E., Neelon, B. and Gibbons, R. D. (2015). Likelihood-based random-effect meta-analysis of binary events. *Journal of Biopharmaceutical Statistics*, **25**, 984–1004.

Atkins, M. S., Shernoff, E. S., Frazier, S. L., Schoenwald, S. K., Cappella, E., Marinez-Lora, A., Mehta, T. G., Lakind, D., Cua, G., Bhaumik, R. and Bhaumik, D. K. (2015). Re-Designing community mental health services for urban children: supporting schooling to promote mental health. *Journal of Consulting Clinical Psychology*, **83**, 839–852.

Bandyopadhyay, D., DeSantis, S. M., Korte, J. and Brady, K. T. (2011). Some considerations for excess zeroes in substance abuse research. *American Journal of Drug and Alcohol Abuse*, **37**, 376–382.

Bhaumik, D. K., Roy, A., Aryal, S., Hur, K., Duan, N., Normand, S-L.T., Brown, C. H. and Gibbons, R. D. (2008). Sample size determination for studies with repeated continuous outcomes. *Psychiatric Annals*, **38**, 765– 771

Bhaumik, D. K., Amatya, A., Normand, S-L.T, Greenhouse, J., Kaizar, E., Neelon, B. and Gibbons, R. D. (2012). Meta- analysis of rare binary adverse event data. *Journal of American Statistical Association*, **107**, 555–567.

Bhaumik, D. K., Aryal, S., Amatya, A., Kapur, K. and Gibbons, R. D. (2011). Sample size determination for between group comparisons in mixed effects logistic regression models for analysis of longitudinal data. *Journal of Applied Statistical Science*, **19**, 11–22.

Bhaumik, D.K., Aryal, S. and Hur, S. (2019). How to select a suitable model for analysis of mental health service use data. *IAPQR Transactions*, Manuscript accepted for Publication.

Chen, P., Wu, T. J. and Yang, J. (2008). A comparative study of model selection criteria for the number of signals. *IET Radar, Sonar and Navigation*, **2**, 180–188.

Cook, J. A., Shore, S., Burke-Miller, J. K., Jonikas, J. A., Hamilton, M., Ruckdeschel, B., Norris, W., Markowitz, A. F., Ferrara, M. and Bhaumik, D. K. (2019). Mental health self-directed care financing: efficacy in improving outcomes and controlling costs for adults with serious mental illness. *Psychiatric Services*, **70**, 191–201.

Diehr, P., Yanez, D. and Ash, A., Hornbrook, M., and Lin, D. Y. (1999). Methods for analyzing health care utilization and costs. *Annual Review of Public Health*, **20**, 125–144.

Donner, A. and Klar, N. (2002). *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold, London.

Ghosh, S. K., Mukhopadhyay, P. and Lu, J. C. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference*, **136**, 1360–1375.

Gilbert, A. R., Domino, M. E. , Morrissey, J. P. and Gaynes, B. N. (2012). Differential service utilization associated with trauma-informed integrated treatment for women with co-occurring disorders. *Administration and Policy in Mental Health and Mental Health Services Research*, **39**, 426–439.

Gupta, R., Szczesniak, R. D. and Macaluso, M. (2015). Modeling repeated count measures with excess zeros in an epidemiological study. *Annals of Epidemilogy*, **25**, 583–589.

Hacker, K. A., Penfold, R. B., Arsenault, L.S., Zhang, F., Murphy, M. and Wissow, L. S. (2014). Behavioral health services following implementation of screening in massachusetts medicaid children. *Pediatrics*, **134**, 737–746.

Heo, M., Xue, X. and Kim, M. Y. (2013). Sample size requirements to detect an intervention by time interaction in longitudinal cluster randomized clinical trials with random slopes. *Computational Statistics and Data Analysis*, **60**, 169–178.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. Springer, New York, NY.

Kapur, K., Bhaumik, R., Tang, X-C., Hur, K., Reda, D. J. and Bhaumik, D. K. (2014). Sample size determination for longitudinal designs with binary response. *Statistics in Medicine*, **33**, 3781–3800.

Kass, R. E. and Raftery, A. E. (1995). Bayes factor. *Journal of the American Statistical Association*, **90**, 773–795.

Lindsey, M. A., Brandt, N. E., Becker, K. D., Lee, B. R., Barth, R. P., Daleiden, E. L. and Chorpita, B. F. (2014). Identifying the common elements of treatment engagement interventions in children's mental health services. *Clinical Child and Family Pscyhology Review*, **17**, 283–298.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models* (2nd edition). Chapman and Hall, New York.

Min, Y. and Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, **5**, 1–19.

Moghimbeigi, A., Eshraghian, M. R., Mohammad, K. and McArdle, B. (2008). Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. *Journal of Applied Statistics*, **35**, 1193–1202.

Mullahy, J. (1986). Specification and test of ome modified count data models. *Journal of Econometrics*, **3**, 341–365.

Overall, J. E. and Doyle, S. R. (1994). Estimating sample sizes for repeated measurement designs. *Controlled Clinical Trials*, **15**, 100–123.

Pardoe, I. and Durham, C. A. (2003). Model choice applied to consumer preferences. *In Proceedings of the Joint Statistical Meetings*. American Statistical Association.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, **2**, 173–185.

Raudenbush, S. W. and Liu, X. F. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, **5**, 199–213.

Ridout, M., Hinde, J. and Demetrio, C. G. (2001). A score test for testing a zero-in ated poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, **57**, 219–223.

Roy, A., Bhaumik, D. K., Aryal, S. and Gibbons, R. D. (2007). Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics*, **63**, 699–707.

Saha, A. and Dong, D. (1997). Estimating nested count data models. *Oxford Bulletin of Economics and Statistics*, **59**, 423–430.

Seghouane, A. K. (2006). Multivariate regression model selection from small samples using Kullback's symmetric divergence. *Signal Processing*, **86**, 2074–2084.

Tu, S. and Xu, L. (2012). A theoretical investigation of several model selection criteria for dimensionality reduction. *Pattern Recognition Letters*, **33**, 1117–1126.

Xie, H., Tao, J., McHugo, G. J. and Drake, R. E. (2013). Comparing statistical methods for analyzing skewed longitudinal count data with many zeros: An example of smoking cessation. *Journal of Substance Abuse Treatment*, **45**, 99–108.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 275–286

U-Statistics CLT Using Cumulants and a Free Version

Arup Bose[1] and Apratim Dey[2]
[1]*Stat-Math Unit*
Indian Statistical Institute, Kolkata, India
[2]*Department of Statistics*
Stanford University, California

Received: 25 April 2020; Revised: 18 June 2020; Accepted: 11 August 2020

Abstract

It is well-known that a standardised U-statistic based on i.i.d. observations is asymptotically normal. We first give a proof of this using cumulants. Then we consider U-statistics which are based on non-commutative variables. We show that a standardised U-statistics of freely independent identically distributed non-commutative random variables converges to a semi-circle variable. The proof is based on free cumulants. We also discuss briefly the degenerate case.

Key words: U-statistics; Central limit theorem; Degenerate U-statistics; Law of large numbers; Non-commutative probability; Freely independent random variables; Cumulants and free cumulants.

AMS Subject Classifications: 60F05, 46L53

1. Introduction

Let $X_1, X_2, ..., X_n$ be independent and identically distributed (i.i.d.) random variables, defined on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ and with a common distribution F. Suppose $h : \mathbb{R}^m \to \mathbb{R}$ is a Borel measurable function which is symmetric in its arguments. Let

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < ... < i_m \le n} h(X_{i_1}, ..., X_{i_m}). \tag{1}$$

Then U_n is the well-known *U-statistic* with kernel h and has found extensive uses in statistics. Bose and Chatterjee (2018) contains a wealth of material on the properties of U-statistics. A fundamental distributional limit result for U-statistic is the U-statistics Central Limit Theorem (UCLT).

Theorem 1: (UCLT) Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables with a common distribution F and

$$\int_{\mathbb{R}^m} |h(x_1, ..., x_m)|^2 dF(x_1)...dF(x_m) < \infty.$$

Correponding Author: Arup Bose
Email: bosearu@gmail.com

Let $\sigma_1^2 := Var(\mathrm{E}(h(X_1, X_2, ..., X_m)|X_1)) > 0$. Then

$$\sqrt{n}(U_n - \theta) \xrightarrow{w} N(0, m^2\sigma_1^2)$$

where $\xrightarrow{w}$ denotes weak convergence, $N(0, m^2\sigma_1^2)$ denotes the normal distribution with mean 0 and variance $m^2\sigma_1^2$ and $\theta = \int_{\mathbb{R}^m} h(x_1, ..., x_m)dF(x_1)...dF(x_m)$.

The standard proof proceeds by considering the sum of projections of U_n on the subspace $\mathcal{L}_n := \{\psi_1(X_1) + \cdots + \psi_n(X_n) : \psi_1, \ldots, \psi_n \in L^2(F)\}$ and showing that in the limiting case, as $n \to \infty$, the contribution of this sum is most important. The remaining terms are negligible. Then the classical CLT for the sample mean is applied to complete the proof.

With Theorem 1 as the backdrop, our goal in this article is three-fold.

(i) Show how Theorem 1 can be derived using cumulants and thereby avoid projections.

(ii) Establish the CLT for U-statistics in the *non-commutative* set up, by using free cumulants.

(iii) Establish a limit theorem for degenerate U-statistics in the non-commutative set up.

We address these three goals in the next three sections.

2. Proof of UCLT Based on Cumulants

Suppose $Y_1, \ldots Y_n$ are random variables with joint moment generating function

$$M_{Y_1,\ldots,Y_n}(t_1, \ldots, t_n) = \mathrm{E}\left[\exp\left\{\sum_{j=1}^{n} t_j Y_j\right\}\right],\ t_1, \ldots, t_n \in \mathbb{R}.$$

which is finite in a neighbourhood N of 0 in $\mathbb{R}^n$. In that case, the joint cumulant generating function is defined as

$$C_{Y_1,\ldots,Y_n}(t_1, \ldots, t_n) = \log M_{Y_1,\ldots,Y_n}(t_1, \ldots, t_n),\ (t_1, \ldots, t_n) \in N$$

which also has a power series expansion of the form

$$C_{Y_1,\ldots,Y_n}(t_1, \ldots, t_n) = \sum_{k_1,\ldots,k_n=0}^{\infty} \frac{t_1^{k_1} \ldots t_n^{k_n}}{k_1! \ldots k_n!} c_{k_1,\ldots,k_n}(Y_1, \ldots, Y_n), (t_1, \ldots, t_n) \in N.$$

The real numbers $c_{k_1,\ldots,k_n}(Y_1, \ldots, Y_n)$ are called the cumulants of $\{Y_i : 1 \le i \le n\}$. If $k_j \neq 0$ for at least two indices j, then $c_{k_1,\ldots,k_n}(Y_1, \ldots, Y_n)$ is called a mixed cumulant of $\{Y_i : 1 \le i \le n\}$. We shall use the notation

$$c_j(Y_1, \ldots Y_j) \text{ for } c_{1,1,\ldots 1}(Y_1, \ldots Y_j).$$

It is easily seen that the if Y, Y_1, Y_2 are random variables, then

$$c_1(Y) = \mathrm{E}(Y),\ c_2(Y) = \mathrm{Var}(Y) \text{ and } c_{1,1}(Y_1, Y_2) = c_{1,1}(Y_2, Y_1) = \mathrm{Cov}(Y_1, Y_2). \quad (2)$$

In general moments and cumulants are in $1-1$ correspondence and this can be expressed *via* the well-known Möbius function on the set of all partitions of $\{1,\ldots,n\}$, $n \geq 1$. For details of this and other facts, see Nica and Speicher (2006). See also Brillinger (2001) where cumulants have been used extensively to prove limit theorems in time series. We shall need the following facts about cumulants.

Fact 1. Suppose Y_i are independent bounded random variables. Then all their mixed cumulants vanish. This follows easily since the moment generating function factorises.

Fact 2. Cumulants are multi-linear functions of the random variables. This follows from definition.

Fact 3. Y is normally distributed with mean μ and variance σ^2 if and only if its first two cumulants are $c_1(Y) = \mu$ and $c_2(Y) = \sigma^2$ and all other cumulants of Y vanish. This follows immediately from the moment generating function of Y.

Fact 4. Suppose $\{Y_n\}$ is a sequence of random variables such that $c_1(Y_n) \to \mu$, $c_2(Y_n) \to \sigma^2$ and $c_k(Y_n) \to 0$ for all $k \geq 3$. Then $Y_n \xrightarrow{w} N(0,\ \sigma^2)$ where $\xrightarrow{w}$ denotes weak convergence. This is obtained as follows: from the specific nature of the $1-1$ correspondence, it follows that all moments of Y_n converge, and converge to the normal moments. Since the normal distribution is the unique distribution with the normal moments, weak convergence follows.

Proof: [Proof of Theorem 1] We first prove Theorem 1 under the additional assumption that

$$h \ \text{ is bounded.} \tag{3}$$

By linearity of cumulants, $c_1(U_n) = \theta$ and therefore $c_1(\sqrt{n}(U_n - \theta)) = 0$.

Define

$$\zeta_k(h) = \mathrm{Cov}(h(X_1, ..., X_k, X_{k+1}, ..., X_m), h(X_1, ..., X_k, X_{m+1}, ..., X_{2m-k})).$$

Note that $\zeta_1(h) = \sigma_1^2$. The following formula is standard. It can also be proved easily by using the symmetry of h, equation (2) and linearity of cumulants.

$$c_2(\sqrt{n}(U_n - \theta)) = n\binom{n}{m}^{-2} \sum_{k=1}^{m} \binom{n}{k}\binom{n-k}{m-k}\binom{n-m}{m-k}\zeta_k(h).$$

For any fixed natural number a, we have $\binom{n}{a} \sim \dfrac{n^a}{a!}$ as $n \to \infty$. Therefore, for $1 \leq k \leq m$, we have

$$\binom{n}{k}\binom{n-k}{m-k}\binom{n-m}{m-k} \sim \frac{n^k(n-k)^{m-k}(n-m)^{m-k}}{k!(m-k)!(m-k)!} \sim \frac{n^{2m-k}}{k!(m-k)!(m-k)!}.$$

Thus

$$\lim_{n\to\infty} c_2(\sqrt{n}(U_n - \theta)) = \lim_{n\to\infty} \frac{m!m!}{n^{2m-1}} \sum_{k=1}^{m} \frac{n^{2m-k}}{k!(m-k)!(m-k)!}\zeta_k(h).$$

Note that if $1 < k \leq m$ then $\dfrac{n^{2m-k}}{n^{2m-1}} \to 0$. Therefore, only the $k = 1$ term will survive and we get

$$\lim_{n\to\infty} c_2(\sqrt{n}(U_n - \theta)) = m^2\zeta_1(h) > .0$$

Now we will show that $c_j(\sqrt{n}(U_n - \theta)) \to 0$ for all $j \geq 3$. We observe that, by multi-linearity of cumulants,

$$c_j(\sqrt{n}(U_n - \theta)) = \frac{n^{j/2}}{\binom{n}{m}^j} \sum_{I_1, I_2, \ldots, I_j} c_j(h(X_{I_1}), ..., h(X_{I_j})) \tag{4}$$

where $I_1, ..., I_j$ are ordered $m-$tuples $(i_1 < i_2 < \cdots < i_m)$ with each $1 \leq i_r \leq n$ for $1 \leq r \leq m$, and $h(X_I) := h(X_{i_1}, ..., X_{i_m})$ if $I = (i_1, ..., i_m)$.

Now we make the following observations. Fix I_1. Suppose at least one of $I_2, ..., I_j$ does not have any index common with I_1. Then by independence, the corresponding $c_j = 0$.

Let us count the remaining cases. If we fix I_1, then are $O_m(n^{m-1})$ such choices for each of $I_2, ..., I_j$, giving a total of $O_{m,j}(n^{(m-1)(j-1)})$ choices. Finally, I_1 can be chosen in $\binom{n}{m} = O_m(n^m)$ ways. Therefore, the total count of the remaining cases is $O_{m,j}(n^{m+(m-1)(j-1)})$. Note that the we have a common uppper bound for all the cumulants c_j that correspond to these cases.

Hence if $j \geq 3$,

$$c_j(\sqrt{n}(U_n - \theta)) = O_{m,j}(\frac{n^{j/2}}{n^{mj}} \times n^{m+mj-j-m+1}) = O_{m,j}(n^{1-(j/2)}) \to 0 \text{ as } n \to \infty.$$

Hence the proof is complete by an application of Fact 4, under the extra condition (3).

To relax this assumption, we use a standard truncation argument. Define

$$\tilde{h}(x_1, \ldots, x_m) = h(x_1, \ldots, x_m) I(|h(x_1, \ldots, x_m)| \leq B).$$

Let $\tilde{U}_n^{(B)}$ be the corresponding U-statistic. Since $\tilde{h}$ satisfies (3),

$$\sqrt{n}(U_n^{(B)} - \theta_n^{(B)}) \xrightarrow{w} N(0,\ m^2\tilde{\zeta}_1(\tilde{h})) \ \text{ as } \ n \to \infty.$$

It is not hard to show that (use DCT)

$$\zeta_1(\tilde{h}) \to \zeta_1(h) \ \text{ as } \ B \to \infty.$$

Moreover, it is also easy to show, by using the variance formula developed above for any U-statistics, that

$$\lim_{B\to\infty} \lim_{n} V\Big(\sqrt{n}(U_n^{(B)} - \theta_n^{(B)}) - \sqrt{n}(U_n - \theta)\Big) = 0.$$

This completes the proof of Theorem 1. □

Remark 1: (a) The special case of $m = 1$ yields the standard CLT for the mean: if $\{X_i\}$ are i.i.d. with mean 0 and variance 1, then $\sum_{i=1}^{n} X_i/\sqrt{n}$ converges weakly to the standard normal distribution. This cumulant based proof avoids the use of characteristic function.

(b) By extending the above argument, and an appropriate extension of Fact 4 to multivariate normal, it can be shown that if we have several U-statistics then after the needed centering and scaling, they converge jointly to a multivariate normal distribution. We omit the details.

3. UCLT for Free Variables

Free Probability refers to an extension of classical probability to certain non-commutative spaces. One of its central notions is free independence which is a natural notion of independence available in different types of non-commutative probability spaces. Very strong connections between free independence and random matrices were discovered by Voiculescu (see Voiculescu (1991)). A nice combinatorial introduction to free probability is available in Nica and Speicher (2006). Probabilists and statisticians have been increasingly drawn to aspects of free probability, specially in the context of high dimensional random matrices. For some flavour of its application in high dimensional time series, see Bose and Bhattacharjee (2018).

The non-commutative probability space that we shall work with is the $*$-probability space. We shall briefly describe its basic ingredients. For a detailed introduction see Nica and Speicher (2006).

Recall that, an algebra $\mathcal{A}$ over complex numbers is called a $*$-algebra if it contains a unity $1_{\mathcal{A}}$, and is endowed with an antilinear $*$ operation which maps $a \in \mathcal{A}$ to $a^* \in \mathcal{A}$ and which satisfies $(a^*)^* = a$ and $(ab)^* = b^*a^*$ for all $a, b \in \mathcal{A}$.

A *$*$-probability space* is a pair $(\mathcal{A}, \varphi)$ where $\mathcal{A}$ is a $*$-algebra and φ is a linear functional on $\mathcal{A}$ which satisfies $\varphi(1_{\mathcal{A}}) = 1$, $\varphi(a^*) = \overline{\varphi(a)}$, and $\varphi(a^*a) \geq 0$ for all $a \in \mathcal{A}$.

The elements of $\mathcal{A}$ are called *random elements*. An element $a \in \mathcal{A}$ is called *self-adjoint* if $a^* = a$. Recall that the expectation operator is also linear and satisfies $\mathrm{E}(1) = 1$. Thus, it helps to think of φ as an analogue of the expectation operator.

Example 1: Suppose $(\Omega, \mathcal{F}, P)$ is a classical probability space and E is the expectation operator. Let $\mathcal{A}$ be the set of (complex valued) random variables with all moments finite, where random variables that are almost surely equal, are identified as same. Then $(\mathcal{A}, \mathrm{E})$ is trivially a $*$-probability space. In this case elements of $\mathcal{A}$ commute.

Example 2: A typical example of a $*$-probability space is the algebra $\mathcal{A}$ of all $n \times n$ matrices with random variable entries all whose moments are finite, and for any $A \in \mathcal{A}$, $\varphi(A) = n^{-1}\,\mathrm{E}\,\mathrm{Trace}(A)$. The unity is the $n \times n$ identity matrix I, for which $\varphi(I) = n^{-1}\,\mathrm{E}\,\mathrm{Trace}(I) = 1$. With $*$ denoting the usual matrix adjoint, $\varphi(A^*) = n^{-1}\,\mathrm{E}\,\mathrm{Trace}(A^*) = n^{-1}\overline{\mathrm{E}\,\mathrm{Trace}(A)} = \overline{\varphi(A)}$ since the diagonal entries of A^* are complex conjugates of those of A, and $\varphi(A^*A) = n^{-1}\,\mathrm{E}\,\mathrm{Trace}(A^*A) \geq 0$ since all diagonal entries of A^*A are non-negative.

Given random elements $\{a_1, \ldots, a_n\}$, its moments are the quantities $\{\varphi(b_1 \cdots b_k), k \geq 1, b_j \in \{a_1, \ldots, a_n\}$ for all $1 \leq j \leq k\}$. Analogous to cumulants of random variables, there is a concept of *free cumulants* of random elements. These are in $1-1$ correspondence with the moments via the Möbius function on non-crossing partitions of $\{1, \ldots, n\}$, $n \geq 1$. We shall avoid a formal definition. The free cumulants will be denoted by a generic κ. We may note here that for any two random elements a and b, $\kappa_1(a) = \varphi(a)$ and $\kappa_2(a, b) = \varphi(ab) - \varphi(a)\varphi(b)$. Note that in general $\kappa_2(a, b) \neq \kappa_2(b, a)$.

A random element s on a $*$-probability space $(\mathcal{A}, \varphi)$ is said to be a *semi-circle variable* with variance σ^2 if it is self-adjoint and

$$\varphi(s^h) = \begin{cases} \sigma^{2n} C_n = \frac{1}{n+1}\binom{2n}{n}\sigma^{2n}, & \text{if } h = 2n \\ 0, & \text{if } h \text{ is odd.} \end{cases} \tag{5}$$

The numbers $\{C_k, k \geq 1\}$ are known as Catalan numbers and $\{\sigma^{2k} C_k, k \geq 1\}$ define a unique probability measure, known as the *semi-circle distribution* with variance σ^2. It is well-known that $\kappa_1(s) = 0$, $\kappa_2(s, s) = 1$ and all higher order free cumulants of s are 0. Thus, this is the analogue of the standard normal variable X which has the same property for its cumulants.

In the classical set-up, bounded random variables are independent if and only if all their mixed cumulants are 0. Analogously, random elements are "freely independent" or simply *free*, if and only if all their mixed free cumulants vanish.

We also need the notion of convergence of random elements: Suppose we have a sequence of $*$-probability spaces $(\mathcal{A}_n, \varphi_n)$. Suppose $a_n \in \mathcal{A}_n$ are self-adjoint. Then $\{a_n\}$ are said to *converge in distribution* if $\lim \varphi_n(a_n^k)$ exists for all integers k. We visualize a limit $*$-probability space $\mathcal{A}$, generated by an indeterminate (self-adjoint) element a and with the state $\varphi(a^k) = \lim \varphi_n(a_n^k)$ for all k and extended linearly to the entire algebra. Note that convergence in distribution is *not* the same as the usual weak convergence. However, if $\{\varphi_n(a_n^k), k \geq 1\}$ and $\{\varphi(a^k), k \geq 1\}$ determine unique probability measures,$\{\mu_n\}$ and μ with these as their moments, then the above convergence in distribution implies μ_n converges to μ weakly. Analogous notions hold for joint convergence of several variables.

We shall need the following facts about free cumulants. For proofs see Nica and Speicher (2006).

Fact (a). Suppose Y_i are freely independent random elements in some $*$-probability space. Then all their mixed free cumulants vanish.

Fact (b). Free cumulants are multi-linear functions.

Fact (c). Suppose s is a semi-circle variable with mean 0 and variance σ^2. Then the first two free cumulants of s are $\kappa_1(Y) = 0$ and $\kappa_2(Y) = \sigma^2$. Further all other free cumulants of s vanish.

Fact (d). Suppose $\{y_n\}$ is a sequence of self-adjoint random elements such the $\kappa_1(y_n) \to 0$, $\kappa_2(y_n) \to \sigma^2$ and $\kappa_k(y_n) \to 0$ for all $k \geq 3$. Then $\{y_n\}$ converges to a semi-circle variable with variance σ^2.

We are now in position to give a free version of Theorem 1. Since we are working with random elements which are elements of an algebra, we are restricted to working with only polynomials in the variables of the algebra. Hence the statement of Theorem 1 needs to be modified slightly now. Variables are said to be identically distributed of they have the same moments.

Theorem 2: Suppose $X_1, ..., X_n$ are freely independent self-adjoint identically distributed random elements on a $*$-probability space $(\mathcal{A}, \varphi)$. Suppose $h(x_1, ..., x_m)$ is a self-adjoint polynomial in the m variables $x_1, ..., x_m$, symmetric in its arguments. Let U_n be the U-statistics with kernel h. Let $\theta = \varphi(h(X_1, ..., X_m))$. For $k = 0, 1, ..., m$ define

$$d_k = \kappa_2(h(X_1, ..., X_k, X_{k+1}, ..., X_m), h(X_1, ..., X_k, X_{m+1}, ..., X_{2m-k})).$$

Suppose $d_1 > 0$. Then $\sqrt{n}(U_n - \theta)$ converges in distribution to a semi-circle variable with variance $m^2 d_1$.

Example 3: Suppose $\{X_i\}$ are free and identically distributed. Consider the sample variance

$$s_n^2 = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \frac{(X_i - X_j)^2}{2}.$$

Then s_n^2 is a U-statistic. Suppose without loss of generality $\varphi(X_i) = 0$. By an application of Theorem 2,

$$n^{1/2}\Big(s_n^2 - \varphi(X_1^2)\Big) \to s$$

where s is a semi-circle variable with variance $\varphi(X_1^4) - [\varphi(X_1^2)]^2$. Note that this could be 0 (for example if X_i are free Bernoulli ± 1 with probability $1/2$ each) in which case, s is 0.

Example 4: Suppose $\{X_i\}$ are freely independent identically distributed variables where $2[\varphi(X_1)]^2 = \theta$. Let $h(x, y) = xy + yx$ Then

$$\sqrt{n}\Big(U_n - \theta\Big) \to s$$

where s is a semi-circle variable with mean 0 and variance $16\varphi(X_1)^2\Big[\varphi(X_1^2) - [\varphi(X_1)]^2\Big]$.

Proof: [Proof of Theorem 2] The proof is almost a repetition of the proof of Theorem 1. We sketch it. Trivially, $\kappa_1(U_n) = \theta$ and therefore $\kappa_1(\sqrt{n}(U_n - \theta)) = 0$.

As before (we now use the fact that mixed free cumulants vanish for freely independent variables),

$$\kappa_2(\sqrt{n}(U_n - \theta)) = n\binom{n}{m}^{-2} \sum_{k=1}^{m} \binom{n}{k}\binom{n-k}{m-k}\binom{n-m}{m-k} d_k$$

and after similar steps, we get

$$\lim_{n \to \infty} \kappa_2(\sqrt{n}(U_n - \theta)) = m^2 d_1 > 0.$$

Now we will show that $\kappa_j(\sqrt{n}(U_n - \theta)) \to 0$ for all $j \geq 3$. We observe that

$$\kappa_j(\sqrt{n}(U_n - \theta)) = n^{j/2}\binom{n}{m}^{-j} \sum_{I_1, I_2, ..., I_j} \kappa_j(h(X_{I_1}), ..., h(X_{I_j})) \tag{6}$$

where $I_1, ..., I_j$ are ordered $m-$tuples $(i_1 < i_2 < \cdots < i_m)$ with each $1 \leq i_r \leq n$ for $1 \leq r \leq m$, and $h(X_I) := h(X_{i_1}, ..., X_{i_m})$ if $I = (i_1, ..., i_m)$.

Now we count as before and use vanishing of mixed free cumulants when there are at least two freely independent variables, to obtain, for $j \geq 3$,

$$\kappa_j(\sqrt{n}(U_n - \theta)) = O_{m,j}(\frac{n^{j/2}}{n^{mj}} \times n^{m+mj-j-m+1}) = O_{m,j}(n^{1-(j/2)}) \to 0 \text{ as } n \to \infty.$$

The proof is complete once we use Fact (d). □

Remark 2: By extending the above argument, and an appropriate extension of Fact (d), to a semi-circle family (see next section), it can be shown that if we have several U-statistics of non-commutative variables, then they converge jointly to a semi-circle family. We omit the details.

4. Degenerate Case

An obvious question that arises here is what happens under degeneracy *i.e.* when $c_1 = 0$ or $d_1 = 0$? The following result is well-known in the classical case. See Bose and Chatterjee (2018).

Theorem 3: Let $h : \mathbb{R}^m \to \mathbb{R}$ be a symmetric kernel. Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables, such that $\mathrm{E}(h(x_1, X_2, ..., X_m)) = 0$ but

$$\sigma_2^2 := Var(\mathrm{E}(h(X_1, X_2, X_3, ..., X_m)|X_1, X_2))) > 0.$$

Then

$$nU_n \xrightarrow{w} \binom{m}{2} \sum_{k=1}^{\infty} \lambda_k(V_k - 1)$$

where $V_k \overset{i.i.d.}{\sim} \chi_1^2$ and λ_k are eigenvalues of an appropriate integral operator.

The result is first proved for the case $m = 2$. The classical proof (see Bose and Chatterjee (2018)) crucially uses the Fredholm representation: any symmetric kernel $\psi : \mathbb{R}^2 \to \mathbb{R}$ can be written as

$$\psi(x_1, x_2) = \sum_{k=1}^{\infty} \lambda_k f_k(x_1) f_k(x_2)$$

where λ_k are eigenvalues of an appropriate integral operator. The theorem is easy to prove when there are only finitely many non-zero eigenvalues and the general case is tackled by approximation. Then the cases $m \geq 3$ is proved by projections.

The above proof is not suitable for our purposes since our variables are non-commutative. We present a free version of the theorem for degenerate U-statistics based on random elements but with special type of kernels of order 2. It should be possible to extend this result to higher order kernels but we decided to stick to the simplest case.

Theorem 4: Let $h(x_1, x_2) := \sum_{k=1}^{K} a_k(f_k(x_1)f_k(x_2) + f_k(x_2)f_k(x_1))$ where $f_k(x)$ are self adjoint polynomials in the variable x, and a_k are constants. Let $X_1, ..., X_n$ be freely independent self adjoint identically distributed random variables such that the following are true.

(a) For each $1 \leq k \leq K$, $\varphi(f_k(X_1)) = 0$ and $\varphi(f_k^2(X_1)) = 1$.

(b) For each $1 \leq k \neq l \leq K$, $\varphi(f_k(X_1)f_l(X_1)) = 0$.

Define $U_n = \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} h(X_{i_1}, X_{i_2})$. Then,

$$nU_n \to 2\sum_{k=1}^{K} a_k(s_k^2 - 1)$$

where $s_1, ..., s_K$ are freely independent standard semi-circular variables.

Example 5: (Examples 3 and 4 continued) Consider the kernel $h(x, y) = xy + yx$ but we now assume that $\varphi(X_i) = 0$ and $\varphi(X_i^2) = 1$. Then it is easy to see that $K = 1$ and conditions (a) and (b) hold.

$$nU_n = n\binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} (X_{i_1}X_{i_2} + X_{i_2}X_{i_1}) \to 2(s^2 - 1)$$

where s is a semi-circle variable with variance 1.

Recall the sample variance s_n^2. Now suppose that X_i are free and identically distributed as classical Bernoulli ± 1 with equal probability. Then $\varphi(X_i) = 0$ and $\varphi(X_i^4) = [\varphi(X_i^2)]^2$, so that $n^{1/2}(s_n^2 - 1)$ converges to 0. It can be checked that

$$\begin{aligned} n(s_n^2 - 1) &= n\Big[\frac{1}{n-1}[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2] - 1\Big] \\ &= n[\frac{n}{n-1} - 1] - \frac{n}{n-1}\Big(\sqrt{n}\bar{X}\Big)^2 \to -(s^2 - 1) \end{aligned}$$

where s is a semi-circle variable with variance 1.

As preparation for the proof, we need to extend some of the notions introduced earlier. Suppose $(\mathcal{A}_n, \varphi_n)$, $n \geq 1$ is a sequence of $*$-probability spaces. Let $\{a_{i,n}, 1 \leq i \leq k\}$ be random elements from $\mathcal{A}_n$, $n \geq 1$. They are said to converge jointly if $\varphi_n\Big(P(a_{i,n}, a_{i,n}^*, 1 \leq i \leq k)\Big)$ converges for every $k \geq 1$ and every polynomial P. Then we can define a limit $*$-probability space $(\mathcal{A}, \varphi)$ where $\mathcal{A}$ is the $*$-algebra generated by polynomials in indeterminates $\{a_i, 1 \leq i \leq k\}$ and the state φ is determined by the limit. That is, for all $k \geq 1$ and all polynomials

$$\varphi\left(P(a_i, a_i^*, 1 \leq i \leq k)\right) = \lim \varphi_n\Big(P(a_{i,n}, a_{i,n}^*, 1 \leq i \leq k)\Big).$$

We write

$$(a_{i,n}, 1 \le i \le k) \to (a_i, 1 \le i \le k).$$

This is equivalent to saying that for all $j \ge 1$ and for all $1 \le i_1, ..., i_j \le k$,

$$\kappa_j(a_{i_1,n}, \cdots, a_{i_j,n}) \to \kappa_j(a_{i_1}, \cdots, a_{i_j}).$$

A collection $(s_1, ..., s_k)$ of random elements from a $*$-probability space $(\mathcal{A}, \varphi)$ is said to be a semi-circular family if these are self-adjoint semi-circle variables and moreover all mixed free cumulants of order greater than 2 are 0. Note that they are then free if all second order mixed free cumulants are also 0.

We shall also need the following free Central Limit Theorem. This can be proved easily in a few lines by using free cumulants—along the lines mentioned in Remark 1—simply use free cumulants instead of usual cumulants. See Nica and Speicher (2006) for a moment based proof.

Theorem 5: Suppose $\{X_{i,j}, 1 \le i \le k\}, j \ge 1$ are self-adjoint variables which are identically distributed as well as free across $j \ge 1$ in some $*$-algebra such that for all i, j, $\varphi(X_{1,j}) = 0$. Then

$$\frac{1}{\sqrt{n}}(X_{i,1} + \cdots + X_{i,n}, 1 \le i \le k) \to (s_1, \cdots, s_k)$$

which is a semi-circle family in some $*$-probability space $(\mathcal{A}, \varphi_0)$ with $\kappa_2(s_i, s_j) = \kappa_2(X_{i,1}, X_{j,1})$ for all $1 \le i, j \le k$.

Proof: [Proof of Theorem 4]

We observe that

$$\sum_{1 \le i_1 < i_2 \le n} \sum_{k=1}^{K} a_k(f_k(X_{i_1})f_k(X_{i_2}) + f_k(X_{i_2})f_k(X_{i_1})) = \sum_{k=1}^{K} a_k(\sum_{i=1}^{n} f_k(X_i))^2 - \sum_{k=1}^{K}\sum_{i=1}^{n} a_k f_k^2(X_i).$$

Therefore

$$nU_n = \frac{2n}{n-1}\sum_{k=1}^{K} a_k\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_k(X_i)\Big)^2 - \frac{2n}{n-1}\sum_{k=1}^{K} a_k \frac{1}{n}\sum_{i=1}^{n} f_k^2(X_i).$$

By the free Central Limit Theorem 5,

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} f_1(X_i), \cdots, \sum_{i=1}^{n} f_K(X_i), \sum_{i=1}^{n}(f_1^2(X_i) - 1), \cdots, \sum_{i=1}^{n}(f_K^2(X_i) - 1)\right) \to (s_1, \cdots, s_K, t_1, \cdots, t_K)$$

which is a semi-circular family. Moreover, using conditions (a) and (b), $s_1, \cdots, s_K$ are all freely independent identically distributed semi-circle variables with variance 1. The exact parameters for $(t_1, \cdots, t_K)$ shall not be important to us.

Let

$$A_{n,k,1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_k(X_i) \quad \text{and} \quad A_{n,k,0} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f_k^2(X_i) - 1), 1 \le k \le K.$$

Also let

$$b_{k,0} = s_k,\ b_{k,1} = t_k,\ 1 \le k \le K.$$

Then by definition of joint convergence, for any $j \ge 1$, for any $k_1, ..., k_j \in \{1, ..., K\}$ and for any $\epsilon_1, ..., \epsilon_j \in \{0, 1\}$,

$$\kappa_j(A_{n,k_1,\epsilon_1}, \cdots, A_{n,k_j,\epsilon_j}) \to \kappa_j(b_{k_1,\epsilon_1}, \cdots, b_{k_j,\epsilon_j}).$$

If $j > 1$ and $\epsilon_1, ..., \epsilon_j$ are not all 0, say $\epsilon_1 = 1$ without loss of generality, then using the fact that constants are free of everything,

$$\kappa_j(A_{n,k_1,\epsilon_1}, \cdots, A_{n,k_j,\epsilon_j}) = \kappa_j(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_{k_1}^2(X_i), A_{n,k_2,\epsilon_2}, \cdots, A_{n,k_j,\epsilon_j}) \to \kappa_j(t_{k_1}, b_{k_2,\epsilon_2}, \cdots, b_{k_j,\epsilon_j})$$

and therefore, $\kappa_j(\frac{1}{n}\sum_{i=1}^{n} f_{k_1}^2(X_i), A_{n,k_2,\epsilon_2}, \cdots, A_{n,k_j,\epsilon_j}) \to 0$. Further, if $j = 1$, then for any $1 \le k \le K$, $\kappa_1(\frac{1}{n}\sum_{i=1}^{n} f_k^2(X_i) - 1) = 0$ and hence $\kappa_1(\frac{1}{n}\sum_{i=1}^{n} f_k^2(X_i)) \to 1$. This shows the following joint convergence:

$$(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_1(X_i), \cdots, \frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_K(X_i), \frac{1}{n}\sum_{i=1}^{n} f_1^2(X_i), \cdots, \frac{1}{n}\sum_{i=1}^{n} f_K^2(X_i)) \to (s_1, \cdots, s_K, 1, \cdots, 1)$$

Therefore, we have

$$nU_n = \frac{2n}{n-1}\sum_{k=1}^{K} a_k(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_k(X_i))^2 - \frac{2n}{n-1}\sum_{k=1}^{K} a_k\frac{1}{n}\sum_{i=1}^{n} f_k^2(X_i) \to 2\sum_{k=1}^{K} a_k(s_k^2 - 1).$$

□

We have crucially used the representation of the kernel. It is not clear how to obtain a limit theorem for a more general kernel. We intend to pursue this direction in future.

Acknowledgements

We are grateful to the Referees for very constructive comments. These have led to a substantial improvement in the presentation. Most of the work in this paper appeared in Chapter 3 of Dey (2019), the Masters dissertation of Apratim Dey under the guidance of Arup Bose. Research of Arup Bose is supported by a J. C. Bose National Fellowship, Department of Science and Technology, Government of India.

References

Bose, A. and Bhattacharjee, M. (2018). *Large Covariance and Autocovariance Matrices.* CRC Press, Boca Raton.

Bose, A. and Chatterjee, S. (2018). *U-Statistics, M_m-Estimators and Resampling.* Springer Texts and Readings in Mathematics.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory.* SIAM.

Dey, A. (2019). *Random Matrices and Free Probability.* Masters Dissertation, Indian Statistical Institute, Kolkata.

Nica, A. and Speicher, R. (2006). *Lectures on the Combinatorics of Free Probability.* London Mathematical Society Lecture Note Series 335, Cambridge University Press.

Voiculescu, D. (1991). Limit laws for random matrices and free products. *Inventiones Mathematicae*, **104**, **1**, 201-220.

Statistics and Applications (ISSN 2452-7395 (online))
Volume 18, No. 2, 2020 (New Series), pp 287 – 298

Inference on Stress-Strength Reliability of a System with Two Independent Stresses under Generalized Uniform Distribution

Manisha Pal[1] **and Montip Tiensuwan**[2]
[1]*Department of Statistics, University of Calcutta, India*
[2]*Department of Mathematics Faculty of Science, Mahidol University, Thailand*

Received: 19 November 2019; Revised: 24 August 2020; Accepted: 25 August 2020

Abstract

In this paper, we consider a system with strength X and two independent stresses Y and Z working on it. We derive the UMVUEs of the stress-strength reliability $\xi = P[X > \max(Y,Z)]$ and its variance, when X, Y and Z have independent generalized uniform distributions with known shape parameters. We also discuss testing of hypothesis regarding ξ. A comparison of the UMVUE with the MLE has been carried out in terms of the mean squared error. A simulation study has also been indicated.

Key Words: Generalized uniform distribution; Stress-strength reliability; UMVUE; Hypothesis testing.

AMS Subject Classification (2000): Primary 62F30, Secondary 62N05

1. Introduction

The stress-strength model finds applicability in many areas of research, like reliability engineering, psychology, biometry, economics, medicine, environmental risk assessments, etc. The main problem is to infer about $\Pr[X > Y]$, where X denotes the strength and Y the stress. Estimation of $\Pr[X > Y]$ has been addressed by many authors for various distributions of the variables. Some studies along this line are due to Reiser and Guttman (1988), Ivshin (1996), Ali *et al.* (2005), Pal *et al.* (2005), Ng (2006), Krishnamoorthy *et al.* (2007), Kundu and Raqab (2009), Ventura and Racugno (2011), Baklizi (2014), Gunasekera (2015).

In real life situations, a system may have to withstand two or more stresses on it. For example, tension, compression, shear, bending, and torsion are the stresses on the wings, fuselage, and landing gear of an aircraft. In such situations, the stress-strength reliability will be defined by the probability that the strength of the system is more than the maximum of the stresses acting on it. There are very few studies relating to estimation of this reliability. Rinco (1983) initiated a study on the estimation of $\Pr[Y_p > \max(Y_1, Y_2, \cdots, Y_{p-1})]$ when the random variables $Y_1, \ldots, Y_p$ are independent following exponential distributions with unequal location parameters and equal scale parameters, and suggested an estimator. Gupta and Gupta (1988) derived the MLE, MVUE and Bayes estimator of the same for the case of $p = 2$. They carried out simulation studies to compare these estimators. Karaday *et al.* (2011) investigated the MLE of stress-strength reliability, $\Pr[\max(Y_1,Y_2) < X]$, when a component with strength X following a Gamma distribution is exposed to two independent stresses Y_1,Y_2 having exponential distributions with different parameters. Kundu (2017) estimated the reliability

Corresponding author: Montip Tiensuwan
Email: montip.tie@mahidol.ac.th

function $R = \Pr[Y_3 > \max(Y_1, Y_2)]$, when Y_1, Y_2 and Y_3 are independent exponential variables with unknown location parameters. She derived several estimators of R, and compared their performance based on their risks under different loss functions. Park (2010) discussed estimation of reliability in a load sharing system.

In this paper, we find the UMVU estimator of the stress-strength reliability $\Pr[X > \max(Y, Z)]$, where the strength X and the stresses Y and Z are independently distributed, each having a generalized uniform distribution. Force of water flow, stress on venting valve, *etc.* may have generalized uniform distributions.

A generalized uniform (GU) distribution is defined by the density $f(x)$ and cumulative distribution function $F(x)$ as follows:

$$f(x) = \frac{\alpha+1}{\theta^{\alpha+1}} x^{\alpha}, 0 < x < \theta, \quad F(x) = \left(\frac{x}{\theta}\right)^{\alpha+1}, 0 < x < \theta,$$

where $\theta > 0, \alpha > -1$ (see Tiwari *et al.*, 1996). The parameters α and θ are, respectively, the shape and scale parameters of the distribution. We may write the distribution as GU(α, θ).

We also find the UMVU estimator of the variance of the UMVU estimator of $\Pr[X > \max(Y, Z)]$. We further propose a test for the stress-strength reliability, which is uniformly most powerful within the class of tests based on complete sufficient statistics.

2. Stress-Strength Reliability

Consider a system with strength X, which follows the GU(α_1, θ_1) distribution, given by

$$f_X(x) = \frac{\alpha_1+1}{\theta_1^{\alpha_1+1}} x^{\alpha_1}, \quad 0 < x < \theta_1 (< \infty). \tag{1}$$

Suppose there are two independent stresses Y and Z working on the system, which are distributed as GU(α_2, θ_2) and GU(α_3, θ_3), respectively. The system functions as long as it can withstand the two stresses.

Suppose $\theta_2 = \theta_3 = \theta$, say. Let, $\rho = \dfrac{\theta}{\theta_1}$.

The stress-strength reliability of the system is then given by

$$\xi = \Pr[X > \max(Y, Z)] = g(\rho), \text{ say, where}$$

$$g(\rho) = \left(\frac{\alpha_1+1}{\alpha_1+\alpha_2+\alpha_3+3}\right)\rho^{-(\alpha_2+\alpha_3+2)}, \text{ if } \rho \geq 1$$

$$= 1 - \frac{\alpha_2+\alpha_3+2}{\alpha_1+\alpha_2+\alpha_3+3}\rho^{(\alpha_1+1)}, \text{ if } \rho \leq 1. \tag{2}$$

Clearly, ξ is a monotone function of ρ.

Let us assume that α_1, α_2 and α_3 are known, but θ_1 and θ are unknown.

3. MVUE of ξ

Consider independent random sample $(X_1, X_2, ..., X_n)$, $(Y_1, Y_2, ..., Y_m)$, and $(Z_1, Z_2, ..., Z_r)$ of sizes *n, m* and *r* respectively, from the distributions of *X, Y* and *Z*. The statistic

$$\begin{aligned} U &= 1, \text{ if } X_1 > \max(Y_1, Z_1) \\ &= 0, \text{ otherwise.} \end{aligned} \tag{3}$$

is an unbiased estimator of ξ.

Let $(X_{(1)} < X_{(2)} < ... < X_{(n)})$, $(Y_{(1)} < Y_{(2)} < ... < Y_{(m)})$ and $(Z_{(1)} < Z_{(2)} < ... < Z_{(r)})$ be the ordered observations in the samples mentioned above. Then, $X_{(n)}$ is a complete sufficient statistic for θ_1. The following lemma indicates the complete sufficient statistic of θ.

Lemma 1: (i) $W = \max(Y_{(m)}, Z_{(r)})$ is a complete sufficient statistic for θ.

(ii) $\frac{Y_1}{W}$ (and also $\frac{Z_1}{W}$) is distributed independently of θ.

Proof: (i)We have

$$\Pr(W < w) = \Pr(Y_{(m)} < w, Z_{(r)} < w) = \left(\frac{w}{\theta}\right)^{m(\alpha_2+1)+r(\alpha_3+1)}, \ 0 < w < \theta,$$

since $Y_{(m)}$ and $Z_{(r)}$ are independently distributed.

Hence, $W \sim \text{GU}(m(\alpha_2+1)+r(\alpha_3+1), \theta)$.

Now, for any function $h(w)$ of w,

$$E(h(W)) = \int_0^\theta h(w)\{m(\alpha_2+1)+r(\alpha_3+1)\}\frac{w^{m(\alpha_2+1)+r(\alpha_3+1)-1}}{\theta^{m(\alpha_2+1)+r(\alpha_3+1)}}dw = 0$$

$$\Rightarrow \int_0^\theta h(w) w^{m(\alpha_2+1)+r(\alpha_3+1)-1} dw = 0. \tag{4}$$

Differentiating (4) with respect to θ gives $h(\theta) = 0$ for all θ, which implies $h(w) = 0$, for $0 < w < \theta$. Hence, W is a complete statistic.

The sufficiency part follows easily from Neyman-Fisher Factorization Theorem, by considering the joint distribution of $(Y_1, Y_2, ..., Y_m)$ and $(Z_1, Z_2, ..., Z_r)$.

(ii) Since $Y_1 \sim \text{GU}(\alpha_2,\theta)$ and $Z_1 \sim \text{GU}(\alpha_3,\theta)$, we have $\frac{Y_1}{\theta} \sim \text{GU}(\alpha_2,1)$ and $\frac{Z_1}{\theta} \sim \text{GU}(\alpha_3,1)$, which are independent of θ. Similarly, $\frac{W_1}{\theta} \sim \text{GU}(m(\alpha_2+1)+r(\alpha_3+1),1)$, which is independent of θ. Hence, $\frac{Y_1}{W} = \frac{Y_1/\theta}{W/\theta}$ and $\frac{Z_1}{W} = \frac{Z_1/\theta}{W/\theta}$ are distributed independently of θ.

(Proved)

Using Lehmann-Scheffé Theorem, the UMVUE of ξ is, therefore, given by

$$E(U \mid X_{(n)}, W) = \Pr(X_1 > \max(Y_1, Z_1) \mid X_{(n)}, W) = \Pr\left(\frac{Y_1}{W} < \frac{X_1}{X_{(n)}}.\frac{1}{D}, \frac{Z_1}{W} < \frac{X_1}{X_{(n)}}.\frac{1}{D} \mid D \right) \quad (5)$$

where $D = \frac{W}{X_{(n)}}$.

Clearly, (5) is a function of only D, α_1, α_2 and α_3, since the distributions of $\frac{X_1}{X_{(n)}}, \frac{Y_1}{W}$ and $\frac{Z_1}{W}$ are independent of θ_1 and θ. Thus, any unbiased estimator of ξ, which is a function of D will be UMVUE of ξ, for α_1, α_2 and α_3 given.

To find the expression of the UMVUE of ξ, we obtain the density function of the distribution of D, which comes out as

$$\begin{aligned} f_D(d) &= \frac{C}{\rho}\left(\frac{d}{\rho}\right)^{m(\alpha_2+1)+r(\alpha_3+1)-1}, \text{ if } 0 < d < \rho \\ &= \frac{C}{\rho}\left(\frac{d}{\rho}\right)^{-n(\alpha_1+1)-1}, \text{ if } d \geq \rho, \end{aligned}$$

where

$$C = \frac{n(\alpha_1+1)\{m(\alpha_2+1)+r(\alpha_3+1)\}}{n(\alpha_1+1)+m(\alpha_2+1)+r(\alpha_3+1)}. \quad (6)$$

Inspecting possible estimators of ξ based on D, we arrive at the following theorem:

Theorem 1: The UMVU estimator of ξ is given by

$$\begin{aligned} V &= C_1 D^{-(\alpha_2+\alpha_3+2)}, \text{ if } D \geq 1 \\ &= 1 - C_2 D^{\alpha_1+1}, \text{ if } D < 1, \end{aligned}$$

where

$$C_1 = \frac{\{n(\alpha_1+1)+\alpha_2+\alpha_3+2\}\{(m-1)(\alpha_2+1)+(r-1)(\alpha_3+1)\}}{n\{m(\alpha_2+1)+r(\alpha_3+1)\}(\alpha_1+\alpha_2+\alpha_3+3)}$$

$$C_2 = \frac{(n-1)(\alpha_2+\alpha_3+2)\{(\alpha_1+1)+m(\alpha_2+1)+r(\alpha_3+1)\}}{n(\alpha_1+\alpha_2+\alpha_3+3)\{m(\alpha_2+1)+r(\alpha_3+1)\}}.$$

To prove the theorem, it is sufficient to show that V is an unbiased estimator of ξ. For $\rho \geq 1$,

$$E(V) = C[\int_0^1 (1 - C_2 d^{\alpha_1+1}) f_D(d)dd + \int_1^\rho C_1 d^{-(\alpha_2+\alpha_3+2)} f_D(d)dd + \int_\rho^\infty C_1 d^{-(\alpha_2+\alpha_3+2)} f_D(d)dd],$$

which reduces to $\left(\frac{\alpha_1+1}{\alpha_1+\alpha_2+\alpha_3+3}\right)\rho^{-(\alpha_2+\alpha_3+2)}$ on simplification.

Similarly, for $\rho < 1$, it can be shown that V is an unbiased estimator of ξ.

4. UMVUE of ξ^k

Consider $k \geq 2$ to be an integer. We have

$$\xi^k = \{\Pr[X > \max(Y, Z)]\}^k$$
$$= \left(\frac{\alpha_1+1}{\alpha_1+\alpha_2+\alpha_3+3}\right)^k \rho^{-k(\alpha_2+\alpha_3+2)}, \text{ if } \rho \geq 1$$
$$= \sum_{i=0}^{k} (-1)^i \binom{k}{i} \left(\frac{\alpha_2+\alpha_3+2}{\alpha_1+\alpha_2+\alpha_3+3}\right)^i \rho^{(\alpha_1+1)i}, \text{ if } \rho < 1.$$

We find the UMVUE of ξ^k for $k < \min(n, m, r)$ when α_1, α_2 and α_3 are known. An unbiased estimator of ξ^k is given by

$$U_k = 1, \text{ if } X_1 > \max(Y_1, Z_1), X_2 > \max(Y_2, Z_2), \ldots, X_k > \max(Y_k, Z_k),$$
$$= 0, \text{ otherwise.}$$

From Lehmann-Scheffé Theorem, the UMVUE of ξ^k is

$$E(U_k \mid X_{(n)}, W) = \Pr[\frac{Y_1}{W} < \frac{X_1}{X_{(n)}}.\frac{1}{D}, \frac{Z_1}{W} < \frac{X_1}{X_{(n)}}.\frac{1}{D}; \ldots; \frac{Y_m}{W} < \frac{X_n}{X_{(n)}}.\frac{1}{D}, \frac{Z_r}{W} < \frac{X_1}{X_{(n)}}.\frac{1}{D} \mid D],$$

which is again only a function of $D = \frac{W}{X_{(n)}}$. α_1, α_2 and α_3. Hence, an unbiased estimator of ξ^k based on D will be the UMVUE of ξ^k when α_1, α_2 and α_3 are known.

Theorem 2: For positive integer $k < \min(n, m, r)$, let

$$V_k = G_{1k} D^{-k(\alpha_2+\alpha_3+2)}, \text{ if } D \geq 1$$
$$= \sum_{i=0}^{k} (-1)^i \binom{k}{i} G_{2i} D^{i(\alpha_1+1)}, \text{ if } D < 1,$$

where

$$G_{1k} = \left(\frac{\alpha_1+1}{\alpha_1+\alpha_2+\alpha_3+3}\right)^k \frac{[(m-k)(\alpha_2+1)+(r-k)(n_3+1)][n(\alpha_1+1)+k(n_2+1)+k(n_3+1)]}{n(\alpha_1+1)[m(\alpha_2+1)+r(\alpha_3+1)]}$$

$$G_{2i} = \left(\frac{\alpha_2+\alpha_3+2}{\alpha_1+\alpha_2+\alpha_3+3}\right)^i \frac{(n-i)[i(\alpha_1+1)+m(\alpha_2+1)+r(\alpha_3+1)]}{n[m(\alpha_2+1)+r(\alpha_3+1)]}, i = 0(1)k.$$

Then, V_k is the UMVUE of ξ^k.

Proof: For $\rho \geq 1$,

$$E(V_k) = C[\int_0^1 \sum_{i=0}^{k} (-1)^i \binom{k}{i} G_{2i} d^{i(\alpha_1+1)} f_D(d)dd + \int_1^{\rho} G_{1k} d^{-k(\alpha_2+\alpha_3+2)} f_D(d)dd$$

$$+ \int_{\rho}^{\infty} G_{1k} d^{-k(\alpha_2+\alpha_3+2)} f_D(d)dd]$$

$$= \xi^k + CA\rho^{-m(\alpha_2+1)-r(\alpha_3+1)}, \tag{7}$$

where $A = \sum_{i=0}^{k} (-1)^i \binom{k}{i} \frac{G_{2i}}{i(\alpha_1+1)+k(\alpha_2+\alpha_3+2)} - \frac{G_{1k}}{(m-k)(\alpha_2+1)+(r-k)(\alpha_3+1)}$.

On simplification, the second term of (7) is zero.

Similarly, for $\rho < 1$we get $E(V_k) = \xi^k$.Thus, V_k is an unbiased, and hence UMVU estimator, of ξ^k.

Remark: For $k = 2$, the UMVUE of $Var(V)$is $\hat{V}ar(V) = V^2 - V_2$, provided min $(m,n,r) > 2$, where V_2 is given by

$$V_2 = G_{12} D^{-2(\alpha_2+\alpha_3+2)}, \text{if } D \geq 1$$

$$= G_{20} - 2G_{21} D^{\alpha_1+1} + G_{22} D^{2(\alpha_1+1)}, \text{ if } D < 1,$$

with

$$G_{12} = \left(\frac{\alpha_1+1}{\alpha_1+\alpha_2+\alpha_3+3}\right)^2 \frac{[(m-2)(n_2+1)+(r-2)(n_3+1)][n(\alpha_1+1)+2(n_2+1)+2(n_3+1)]}{n(\alpha_1+1)[m(\alpha_2+1)+r(\alpha_3+1)]}$$

$$G_{20} = 1, \quad G_{21} = \left(\frac{\alpha_2+\alpha_3+2}{\alpha_1+\alpha_2+\alpha_3+3}\right) \frac{(n-1)[(\alpha_1+1)+m(\alpha_2+1)+r(\alpha_3+1)]}{n[m(\alpha_2+1)+r(\alpha_3+1)]}$$

$$G_{22} = \left(\frac{\alpha_2+\alpha_3+2}{\alpha_1+\alpha_2+\alpha_3+3}\right)^2 \frac{(n-2)[2(\alpha_1+1)+m(\alpha_2+1)+r(\alpha_3+1)]}{n[m(\alpha_2+1)+r(\alpha_3+1)]}.$$

It is easy to check that $V^2 - V_2 > 0$.

5. Comparison with MLE of ξ

The maximum likelihood estimator (MLE) of ρ is given by $\tilde{\rho} = \frac{\max(Y_{(m)}, Z_{(r)})}{X_{(n)}}$, where $\tilde{\theta} = \max(Y_{(m)}, Z_{(r)})$ and $\tilde{\theta}_1 = X_{(n)}$ are the MLEs of θ and θ_1, respectively. And, the MLE of ξ is given by $\tilde{\xi} = g(\tilde{\rho})$. Clearly, $\tilde{\xi}$ is a biased estimator of ξ. To compare it with the UMVUE of ξ, a Monte Carlo simulation study has been carried out with 5000 replications. Without loss of generality, we set (α_1, α_2, α_3) = (0.5, 0.25, 0.75), and compute the mean squared errors (MSEs) of the estimators for different settings of the parameters, which are shown in Table 1.

Table 1: Comparison of MSEs of UMVUE and MLE of ξ

		(*n*,*m*,*r*)			
(θ, θ_1)	Estimator	**(5, 5, 5)**	**(5, 10, 10)**	**(10, 10, 10)**	**(10, 5, 5)**
(1, 1.5)	UMVUE	0.01227	0.04535	0.00245	0.00248
	MLE	0.01278	0.08158	0.00476	0.00347
(2, 2)	UMVUE	0.00981	0.01005	0.00156	0. 00241
	MLE	0.01388	0.00825	0.00451	0. 00323
(3, 2)	UMVUE	0.03371	0.03119	0.02912	0. 03433
	MLE	0.03270	0.02756	0.02691	0. 05062

From Table 1 it is clear that the UMVUE does not perform uniformly better than the MLE, though in most situations considered, the MSE of UMVUE is lower than that of the MLE.

4. Test of ξ Based on *D*

As *D* is the key statistic in finding the UMVUE of ξ, we find the best test for H_0: $\xi = \xi_0$ among the class of tests based on *D*, when α_1, α_2 and α_3 are known.

Suppose we want to test the null hypothesis H_0: $\xi = \xi_0$ against $H_A : \xi < \xi_0$, As ξ is monotone decreasing in ρ, this is equivalent to testing H_0^*: $\rho = \rho_0$ against H_A^*: $\rho > \rho_0$, where, from (2), we have

$$\rho_0 = \left(\frac{\alpha_1 + 1}{\xi_0 \{\alpha_1 + \alpha_2 + \alpha_3 + 3\}} \right)^{\frac{1}{(\alpha_2+\alpha_3+2)}}, \text{if } \xi_0 \le \frac{\alpha_1 + 1}{\alpha_1 + \alpha_2 + \alpha_3 + 3}$$

$$= \left((1 - \xi_0) \frac{\alpha_1 + \alpha_2 + \alpha_3 + 3}{\alpha_2 + \alpha_3 + 2} \right)^{\frac{1}{\alpha_1+1}}, \text{if } \xi_0 > \frac{\alpha_1 + 1}{\alpha_1 + \alpha_2 + \alpha_3 + 3}.$$

Within the class of tests based on *D,* the MP test for testing H_0^*: $\rho = \rho_0$ against H_A^*: $\rho = \rho_1 (> \rho_0)$, has the critical region $W = \{d : \lambda(d) > k\}$, where *k* is determined from the size condition, and $\lambda(d)$ is given by

$$\lambda(d)=\frac{f_D(d;\rho_1)}{f_D(d;\rho_0)}=\left(\frac{\rho_1}{\rho_0}\right)^{n(\alpha_1+1)}, \text{ if } d\geq\rho_1$$
$$=\rho_0^{-n(\alpha_1+1)}\rho_1^{-m(\alpha_2+1)-r(\alpha_3+1)}d^{n(\alpha_1+1)+m(\alpha_2+1)+r(\alpha_3+1)}, \text{ if } \rho_0<d<\rho_1$$
$$=\left(\frac{\rho_1}{\rho_0}\right)^{-m(\alpha_2+1)-r(\alpha_3+1)}, \text{ if } d<\rho_0.$$

Clearly $\lambda(d)$ is non-decreasing in d. Hence, $\lambda(d)>k\Leftrightarrow d>d^*$, where d^* satisfies the size condition, *i.e.*, $\Pr[D>d^*\mid H_0^*]=\alpha$, α specified. As $\Pr[D\geq\rho_0\mid H_0^*]=C/n(\alpha_1+1)$, we get

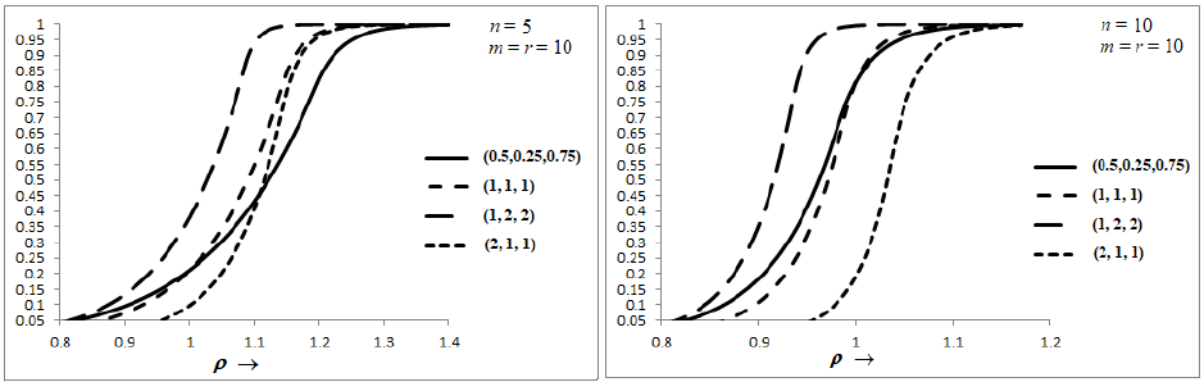
$$d^*=\rho_0\left[\alpha\frac{n(\alpha_1+1)}{C}\right]^{-\frac{1}{n(\alpha_1+1)}}. \text{if } C/n(\alpha_1+1)>\alpha$$
$$=\rho_0\left[(1-\alpha)\frac{m(\alpha_2+1)+r(\alpha_3+1)}{C}\right]^{\frac{1}{m(\alpha_2+1)+r(\alpha_3+1)}}, \text{if } C/n(\alpha_1+1)<\alpha, \tag{8}$$

where $C=\dfrac{n(\alpha_1+1)\{m(\alpha_2+1)+r(\alpha_3+1)\}}{n(\alpha_1+1)+m(\alpha_2+1)+r(\alpha_3+1)}$.

Since d^* is independent of ρ_1, $W^*=\{d\mid d>d^*\}$ will be the critical region of the UMP test of size α among all tests based on D.

For any $\rho>\rho_0$, the power of the test is given by

$$\beta(\rho)=\frac{C}{n(\alpha_1+1)}\left(\frac{d^*}{\rho}\right)^{-n(\alpha_1+1)}, \text{ if } d^*\geq\rho$$
$$=1-\frac{C}{m(\alpha_2+1)+r(\alpha_2+1)}\left(\frac{d^*}{\rho}\right)^{m(\alpha_2+1)+r(\alpha_2+1)}, \text{ if } d^*<\rho.$$



(i) (ii)

Figure 1: Power curves of tests for testing H_0: $\xi=0.5$ against H_A : $\xi<0.5$ for some combinations of (α_1, α_2, α_3), when $\theta=2$, $\theta_1=3$ and $\alpha=0.05$

For testing H_0: $\xi = \xi_0$ against H_A: $\xi > \xi_0$ (or H_A: $\xi \neq \xi_0$), which is equivalent to testing H_0^*: $\rho = \rho_0$ against H_A^*: $\rho < \rho_0$ (or H_A: $\rho \neq \rho_0$), we proceed as above and obtain the UMP test of size α among all tests based on D as follows:

(i) H_A^*: $\rho < \rho_0$

The critical region of the size α test is $W^* = \{d \mid d < d^*\}$, where

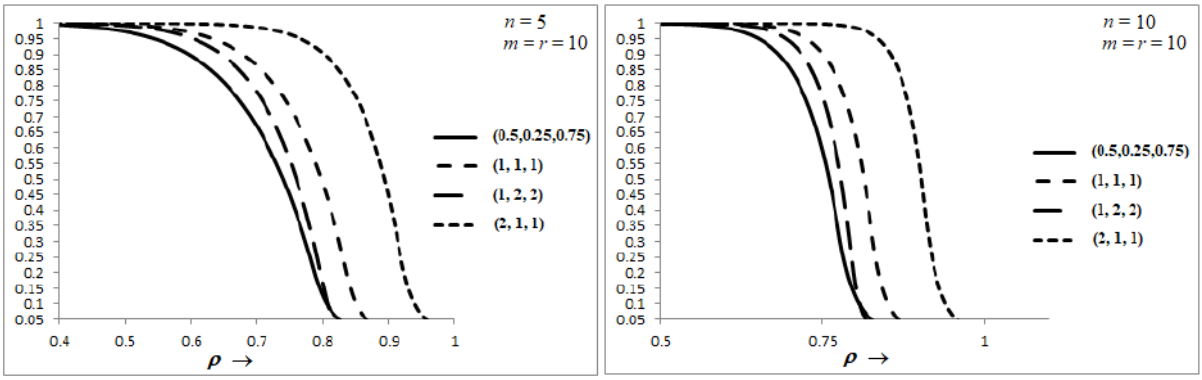
$$d^* = \rho_0 \left[\alpha \frac{m(\alpha_2+1)+r(\alpha_3+1)}{C} \right]^{\frac{1}{m(\alpha_2+1)+r(\alpha_3+1)}}, \text{ if } \frac{C}{m(\alpha_2+1)+r(\alpha_3+1)} > \alpha$$

$$= \rho_0 \left[(1-\alpha) \frac{n(\alpha_1+1)}{C} \right]^{-\frac{1}{n(\alpha_1+1)}}, \text{ if } \frac{C}{m(\alpha_2+1)+r(\alpha_3+1)} < \alpha. \qquad (9)$$

And the power of the test is

$$\beta(\rho) = 1 - \frac{C}{n(\alpha_1+1)} \left(\frac{d^*}{\rho} \right)^{-n(\alpha_1+1)} \text{], if } d^* \geq \rho$$

$$= \frac{C}{m(\alpha_2+1)+r(\alpha_2+1)} \left(\frac{d^*}{\rho} \right)^{m(\alpha_2+1)+r(\alpha_2+1)}, \text{ if } d^* < \rho.$$



(i) (ii)

Figure 2: Power curves of tests for testing H_0: $\xi = 0.5$ against H_A : $\xi > 0.5$ for some combinations of (α_1, α_2, α_3), when $\theta = 2$, $\theta_1 = 3$ and $\alpha = 0.05$

(ii) H_A: $\rho \neq \rho_0$

The critical region of the size α test is $W^* = \{d \mid d < d_1^* \text{ or } d > d_2^*\}$, where d_1^* and d_2^* ($d_1^* < d_2^*$) are determined from the size condition. Assuming $\Pr[D < d_1^* \mid H_0^*] = \Pr[D > d_2^* \mid H_0^*] = \alpha/2$, d_1^* is given by (9) and d_2^* by (8), after replacing α by $\alpha/2$ in each case.

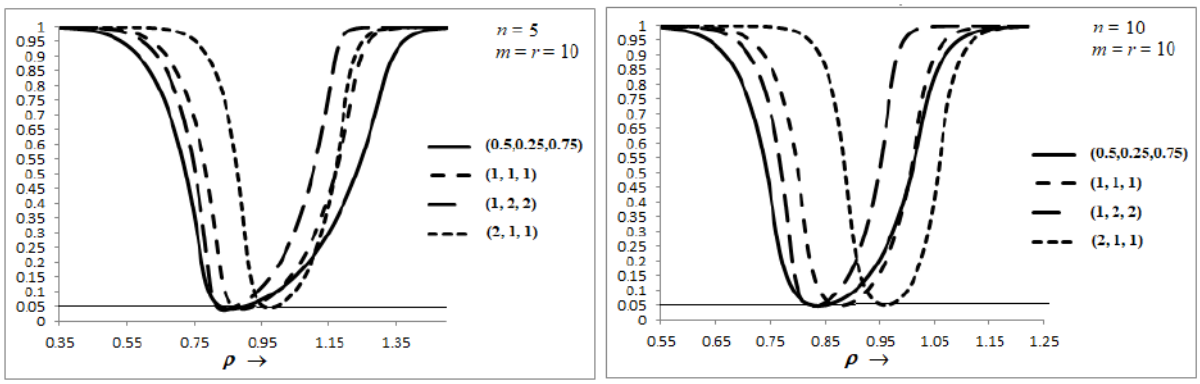
The power of the test is:

$$\beta(\rho) = 1 - \frac{C}{n(\alpha_1+1)}\left(\frac{d_1^*}{\rho}\right)^{-n(\alpha_1+1)} + \frac{C}{n(\alpha_1+1)}\left(\frac{d_2^*}{\rho}\right)^{-n(\alpha_1+1)}, \text{ if } d_2^* > d_1^* \geq \rho$$

$$= 1 + \frac{C}{m(\alpha_2+1)+r(\alpha_2+1)}\left(\frac{d_1^*}{\rho}\right)^{m(\alpha_2+1)+r(\alpha_2+1)} - \frac{C}{m(\alpha_2+1)+r(\alpha_2+1)}\left(\frac{d_2^*}{\rho}\right)^{m(\alpha_2+1)+r(\alpha_2+1)},$$

$$\text{if } d_1^* < d_2^* < \rho$$

$$= \frac{C}{n(\alpha_1+1)}\left(\frac{d_2^*}{\rho}\right)^{-n(\alpha_1+1)} + \frac{C}{m(\alpha_2+1)+r(\alpha_2+1)}\left(\frac{d_1^*}{\rho}\right)^{m(\alpha_2+1)+r(\alpha_2+1)}, \text{ if } d_1^* < \rho < d_2^*.$$



(i) (ii)

Figure 3: Power curves of tests for testing H_0: $\xi = 0.5$ against $H_A : \xi \neq 0.5$ for some combinations of (α_1, α_2, α_3), when $\theta = 2$, $\theta_1 = 3$ and $\alpha = 0.05$

Figures 2 and 3 show that the one-sided tests are unbiased. But Figure 3 indicates that the suggested two-sided test is not unbiased. However, actual calculation shows that the power falls very slightly below the level of significance for some alternative values of ρ, so that the test may be regarded as an almost unbiased test.

6. A Simulation Study

Consider $X \sim$ GU(1, 3), $Y \sim$ GU(1, 2) and $Z \sim$ GU(2, 2). To obtain the UMVUE of ξ, random samples of sizes $m = n = r = 20$ are taken on X, Y and Z respectively. The sample observations are as follows:

Variable	Sample observations									
X	2.1207	2.5148	1.6567	2.9090	2.1738	2.4589	1.2106	0.3185	1.6929	2.9582
	0.8061	2.4919	1.6372	2.8557	2.8237	2.9984	2.9964	0.9302	2.8248	0.1385
Y	1.3415	1.7143	1.0931	1.9151	1.9157	1.0229	0.8568	0.8378	1.6839	1.5374
	1.3346	1.3565	0.1223	1.7003	0.1179	1.9553	1.7329	1.9037	1.2483	0.9266
Z	1.1165	1.3229	1.2098	1.1777	1.8805	1.9025	1.8780	1.7501	1.9573	1.7706
	1.4741	1.9629	1.6615	1.4704	1.7707	1.8960	1.5299	1.6479	0.5454	0.9176

The UMVUE of the stress-strength reliability ξ (= 0.8589 $\mathrm{H_A}: \xi < 0.5$.) is $V = 0.8729$, and the UMVUE of its variance is given by $\hat{Var}(V) = 0.0267$.

Suppose we want to test the hypothesis H_0: $\xi = 0.5$ against H_A:$\xi < 0.5$. The size 0.05 UMP test rejects H_0 if observed $D > 0.8942$. For the given samples, the observed D is 0.6546. Hence, H_0 is accepted.

7. Discussion

The paper studies the UMVU estimator of the stress-strength reliability and its variance when there are two independent stresses acting on a system. The strength of the system and the stresses are assumed to be independent of one another, and follow generalized uniform distributions with known shape parameters, but unknown scale parameters. The UMVU estimator is obtained as a function of the ratio of the complete sufficient statistics of the scale parameters. Tests regarding the stress-strength reliability have been discussed, and the UMP test has been obtained among those based on this ratio. The study has been carried out assuming the scale parameters of the stress distributions to be equal. A natural extension would, therefore, be to assume the scale parameters to be completely unknown. Further, it would be interesting to extend the problem to the case of p (> 2) stresses.

Acknowledgement

The authors thank the anonymous reviewers for their constructive suggestions, which immensely helped to improve the presentation of the paper.

References

Ali, M. M., Pal, M. and Woo, J. (2005). Inference on P(Y<X) in generalized uniform distributions. *Calcutta Statistical Association Bulletin*, **56**(**221-222**), 35-48.

Baklizi, A. (2014). Bayesian inference for Pr (Y<X) in the exponential distribution based on records. *Applied Mathematical Modelling*, **38**(**5–6**), 1698-1709.

Gupta, R. D. and Gupta, R. C. (1988). Estimation of $\Pr\{Y_p > \max(Y_1, Y_2, \cdots, Y_{p-1})\}$ in the exponential case. *Communication in Statistics-Theory and Methods*, **17**, 911-924.

Gunasekera, S. (2015). Generalized inferences of R = Pr(X>Y) for Pareto distribution. *Statistical Papers*, **56**(**2**), 333–351.

Ivshin, V. V. (1996). Unbiased estimation of Pr(X < Y) and their variances in the case of uniform and two-parameter distributions. *Journal of Mathematical Science*, **81**, 2790-2793.

Karaday, N., Saraçoğlu, B. and Pekgör, A. (2011). Stress-strength reliability and its estimation for a component which is exposed two independent stress. *Selçuk Journal of Applied Mathematics* (Special Issue), 131-135.

Krishnamoorthy, K., Mukherjee, S. and Guo, H. (2007). Inference on reliability in two-parameter exponential stress-strength model. *Metrika*, **65**, 261-273.

Kundu, D. and Raqab, M. Z. (2009). Estimation of R = Pr(Y < X) for three-parameter Weibull distribution. *Statistics and Probability Letters*, **79**, 1839-1846.

Kundu, P. (2017). Estimation of $R3 = \Pr(Y_3 > \max(Y_1, Y_2))$ for exponential distribution. *Journal of Advanced Statistics* (Special Issue on Mathematical Statistics and Data Analysis), **2**(**1**), 44-53.

Ng, V. M. (2006). Inference for Pr(Y<X) for exponential and related distributions. *Journal of Modern Applied Statistical Methods*, **5**(**2**), 323-327.

Pal, M., Ali, M. M. and Woo, J. (2005). Estimation and testing of Pr(Y>X) in two-parameter exponential distributions. *Statistics*, **39**(**5**), 415-428.

Park, C. (2010). Parameter estimation for reliability of load sharing systems. *IIE Transactions*, **42**, 753–765.

Reiser, B. and Guttman, I. (1986). Statistical inference for Pr(Y<X): The normal case. *Technometrics*, **28**(**3**), 253-257.

Rinco, S. (1983). Estimation of $Pr(Y_p > \max(Y_1, Y_2, \cdots, Y_{p-1}))$: Predictive approach in exponential case. *The Canadian Journal of Statistics*, **11**(**3**), 239-244.

Tiwari, R. C., Yang, Y. and Zalkikar, W. (1996). Bayes estimation for the Pareto failure model using Gibbs sampling. *IEEE Transactions on Reliability*, **45**(**3**), 471-476.

Ventura, L. and Racugno, W. (2011). Recent advances on Bayesian inference for Pr(X<Y). *Bayesian Analysis*, **6**(**3**), 411-428.

Statistics and Applications {ISSN 2452-7395 (online)}
Volume 18, No. 2, 2020 (New Series), pp 299–316

A Generalized Inferential Model for Meta-Analyses Based on Few Studies

Joyce Cahoon and Ryan Martin[1]
[1]*Department of Statistics*
North Carolina State University, Raleigh, North Carolina, USA

Received: 28 February 2020; Revised: 13 August 2020; Accepted: 01 September 2020

Abstract

Meta-analysis based on only a few studies remains a challenging problem, as an accurate estimate of the between-study variance is apparently needed, but hard to attain, within this setting. Here we offer a new approach, based on the *generalized inferential model* framework, whose success is based on marginalizing out the between-study variance, so that an accurate estimate is not essential. We show theoretically that the proposed solution is at least approximately valid, with numerical results suggesting it is, in fact, nearly exact. We also demonstrate that the proposed solution outperforms existing methods across a wide range of scenarios.

Key words: Confidence interval; Monte Carlo; Normal random effects model; Plausibility function; Profile likelihood.

AMS Subject Classifications: 62F25, 62J05, 62F86

1. Introduction

The most important scientific questions are likely to be pursued by multiple researchers, resulting in separate analyses that may be combined via a single *meta-analysis* to attain stronger and more definitive conclusions. But even when it is appropriate to combine the results from multiple studies, there is often a non-negligible amount of between-study heterogeneity, which is difficult to estimate accurately when the number of studies for meta-analysis is small. Interestingly, meta-analyses with as few as three studies are the norm, not the exception (Davey *et al.* 2011), so there is considerable interest in developing improved methods for inference in this setting of combining results from just a few heterogeneous studies.

To set the scene, consider the classical normal–normal random-effects model where each study k included for meta-analysis provides data (Y_k, σ_k^2). These are modeled as

$$(Y_k \mid M_k) \overset{\text{ind}}{\sim} \mathsf{N}(M_k, \sigma_k^2), \quad M_k \overset{\text{iid}}{\sim} \mathsf{N}(\mu, \nu), \quad k = 1, \ldots, K. \tag{1}$$

Here, M_k denotes the random effect from study k, $\sigma_k^2 > 0$ the variance within study k, μ the underlying population effect and $\nu > 0$ the between-study variance. Note that taking the σ_k^2's as part of the observed data implies that these values are *known*, a common assumption

Corresponding Author: Ryan Martin
Email: rgmarti3@ncsu.edu

in this literature (*e.g.*, DerSimonian and Laird 1986; Michael *et al.* 2019). This hierarchical formulation of the model indicates that the K studies have something in common, namely, a tendency toward μ, but that they are not fully homogeneous and that this degree of heterogeneity, controlled by ν, is unknown. Therefore, the goal is inference on the unknown mean μ, with ν as an unknown nuisance parameter.

In the simple case where ν is *known*, the meta-analysis is straightforward. Marginally, Y_k are independent, distributed as $\mathsf{N}(\mu, \sigma_k^2 + \nu)$, for $k = 1, \ldots, K$. It is easy to check that the minimum variance unbiased estimator of μ is

$$\hat{\mu}(\nu) = \frac{\sum_{k=1}^K w_k(\nu)\, Y_k}{\sum_{k=1}^K w_k(\nu)}, \quad \text{where} \quad w_k(\nu) = (\sigma_k^2 + \nu)^{-1},$$

and its variance is $\{\sum_{k=1}^K w_k(\nu)\}^{-1}$. Of course, the more common scenario in applications is that ν is *unknown* and, somehow, the data must be used to account for that additional uncertainty. A first idea is to estimate ν by some function $\hat{\nu}$ of the data, and then plug that into the formula for the estimator of μ, *i.e.*, $\hat{\mu}(\hat{\nu})$. A well-known strategy is that proposed by DerSimonian and Laird (1986), but there are others, *e.g.*, Paule and Mandel (1982) and Cochran (1954). Unfortunately, the number of studies, K, to be combined is often relatively small (Davey *et al.* 2011), say, $K \leq 7$, and obtaining reliable estimates of the between-study variance based on so few samples is a challenge. Besides, this plug-in style procedure does not naturally account for uncertainty in ν, so any inference drawn can only be (provably) valid in an asymptotic ($K \to \infty$) sense, which may not be especially meaningful in applications where $K \leq 7$. More tangibly, confidence intervals based on the DerSimonian–Laird plug-in style method have shown to perform poorly even when K is as large as 20 (Liu *et al.* 2018); see, also, Viechtbauer (2005), DerSimonian and Kacker (2007), Sidik and Jonkman (2007), Jackson *et al.* (2010), Chung *et al.* (2013), and Veroniki *et al.* (2016). Therefore, there is a need for methods that marginalize out the nuisance parameter, ν, and achieve frequentist error rate control even when the number of studies is small.

To alleviate this problem of estimating ν, likelihood methods have been proposed, *e.g.*, the log likelihood ratio in Goodman (1989), the profile likelihood in Hardy and Thompson (1996), and the signed profile log likelihood in Severini (2000), among others. Unlike previous plug-in methods, these likelihood based procedures introduced an appropriate widening in confidence intervals to deal with the imprecision in estimating ν. In fact, Guolo (2012) was able to improve upon these first-order inference results by introducing a Skovgaard correction to the signed profile log likelihood, making it asymptotically standard normal with an error that vanishes at a rate of $n^{-3/2}$, $n \to \infty$. Resorting to higher-order asymptotics with the Skovgaard correction, however, still failed to modify the likelihood to the extent necessary to achieve the nominal coverage probability when the number of studies available is small.

So-called *exact methods*, featuring such frequentist guarantees, have been offered as alternatives to the plug-in style and likelihood methods described above. These include the methods in Follmann and Proschan (1999), Liu *et al.* (2018), and Wang and Tian (2018) which, in one way or another, are based on permutation distributions. While these permutation methods can produce confidence intervals that achieve nominal frequentist coverage, the discreteness of the permutation distribution makes the results overly conservative, unless K is relatively large. To our knowledge, the most recent work on exact methods for

meta-analysis is Michael *et al.* (2019), and since this shares a number of similarities to our proposal, at least in terms of its construction, we describe this in detail in Section 3.1.

Given that the goal is to develop a method for meta-analysis that controls Type I error, even when K is small, it makes sense to consider the one general, normative framework we are aware of that offers such guarantees. Specifically, Martin and Liu (2013) present a construction of what they call an *inferential model* that leads to provably valid inference, no asymptotic justification required; see Martin and Liu (2016) for a monograph-length introduction, and Martin (2019) for a survey of some recent developments. The distinguishing feature of this approach is the user-specified random set that leads to a sort of "posterior" (but *not* a probability measure) on the parameter space. The distribution of this random set can then be used to visualize the information data provides about the parameter of interest as well as construct inference procedures. We will briefly review the basic construction and its properties in Section 2.1. It may happen that the basic inferential model construction is difficult to carry out in an application, so Martin (2015, 2018) developed a simpler and more direct *generalized* version. Since this is the strategy we follow here for meta-analysis, we provide a brief review in Section 2.2.

This meta-analysis application boils down to a marginal inference problem, *i.e.*, ν is an unknown nuisance parameter to be marginalized out so that we can make inference about μ. The generalized inferential model framework provides at least two strategies for marginalization, and we will show that the method proposed in Michael *et al.* (2019) is itself a generalized inferential model based on one particular choice of marginalization strategy. Our proposed method, on the other hand, is based on a different and arguably more natural choice of marginalization strategy, leading to a method that performs better than theirs in a variety of respects. In Section 3, we describe the construction of a generalized inferential model for meta-analysis, show that the solution in Michael *et al.* (2019) is a special case, and present our proposed method. Details about the computation and theoretical justification are also provided. Numerical comparisons in simulated data experiments demonstrate that our proposed method outperforms existing methods in terms of both validity and efficiency across a broad range of scenarios. Two real applications are presented in Section 5 and some concluding remarks are given in Section 6.

2. Background

2.1. Basic inferential models

Fisher and later Dempster aimed to develop a framework of probabilistic inference without prior distributions, *i.e.*, a prior-free alternative to Bayesian inference. These approaches, however, failed to reach the statistical mainstream, largely because the derived procedures have no frequentist guarantees. To fill that gap, Martin and Liu (2013) argued that frequentist guarantees could be achieved by supplementing the structural, pivotal, or functional model formulation of Fisher (1956), Dempster (2008), Fraser (1968), Barnard (1995), Dawid and Stone (1982), Taraldsen and Lindqvist (2013), and others, with an appropriate user-specified random set. As a consequence of this use of a random set, the inferential output is described by a (data-dependent) non-additive plausibility function (*e.g.*, Shafer 1976) instead of an additive Bayesian, fiducial, or structural posterior probability distribution. The added complexity of non-additivity is not for its own sake, however, it is actually necessary (Balch

et al. 2019) for the strong validity property in (3) that gives the inferential model's belief assignments a meaningful scale for interpretation and, furthermore, leads to the frequentist error rate control guarantees.

In general, we start with a sampling model, $Y \sim \mathsf{P}_{Y|\theta}$, for the observable data Y, depending on some unknown parameter $\theta \in \Theta$. As one would if the goal were to simulate, we express the model as

$$Y = a(\theta, U), \quad U \sim \mathsf{P}_U. \tag{2}$$

where a is a known function, and P_U is a known distribution on the space $\mathbb{U}$. We call the relationship in (2) an *association* between observable data Y, unknown parameter θ, and auxiliary variable U. This association step is the starting point in the construction of a valid inferential model.

A-step. Specify an association of the form (2) and then define the set

$$\Theta_y(u) = \{\theta : y = a(\theta, u)\}, \quad u \in \mathbb{U}.$$

P-step. Specify a predictive random set $\mathcal{S}$, taking values in the power set of $\mathbb{U}$, whose contour function, $f(u) = \mathsf{P}_{\mathcal{S}}(\mathcal{S} \ni u)$, is such that $f(U) \sim \mathsf{Unif}(0,1)$, when $U \sim \mathsf{P}_U$. Remarks on the choice of random set $\mathcal{S}$—and the corresponding contour function f—are provided in Martin and Liu (2016, Ch. 4); see, in particular, Corollary 4.1.

C-step. Combine the ingredients in the A- and P-steps to get a new random set

$$\Theta_y(\mathcal{S}) = \bigcup_{u \in \mathcal{S}} \Theta_y(u),$$

and, for inference about θ, return the distribution of this random set summarized by its plausibility function

$$\mathsf{pl}_y(A) = \mathsf{P}_{\mathcal{S}}\{\Theta_y(\mathcal{S}) \cap A \neq \varnothing\}, \quad A \subseteq \Theta.$$

The distribution of this random set is interpreted as a measure of how *plausible* the hypothesis "$\theta \in A$" is, based on data y and the posited model.

The most unique feature of this construction is the random set $\mathcal{S}$. In our meta-analysis problem, specification of the random set is straightforward as outlined in Section 2.2, but the general details can also be found in Martin and Liu (2016). What matters is that the properties required of the random set make the inferential model *valid*, *i.e.*,

$$\sup_{\theta \in A} \mathsf{P}_{Y|\theta}\{\mathsf{pl}_Y(A) \leq \alpha\} \leq \alpha \quad \text{for all } \alpha \in (0,1) \text{ and all } A \subseteq \Theta. \tag{3}$$

An important consequence of this validity property is the control it provides on the performance of statistical procedures derived from the inferential model output. Indeed, a test that rejects a hypothesis "$\theta \in A$" if $\mathsf{pl}_y(A) \leq \alpha$ will obviously control the frequentist Type I error rate at level α. Similarly, a $100(1-\alpha)\%$ plausibility region for θ, given by

$$\{\theta : \mathsf{pl}_y(\theta) > \alpha\}, \quad \text{where } \mathsf{pl}_y(\theta) := \mathsf{pl}_y(\{\theta\}),$$

has frequentist coverage probability of (at least) $1-\alpha$. These properties are exact in the sense that they do not require any asymptotic approximations. The pointwise plausibility function, $\theta \mapsto \mathsf{pl}_y(\theta)$, is also a useful visualization tool, not unlike a Bayesian posterior density function; see Figure 1 below.

2.2. Generalized inferential models

Since the inferential model construction and validity result is general, efficiency often becomes a concern as the dimension of the auxiliary domain increases with the size of data. To avoid the possible need to specify such a complex, high-dimensional random set, some non-trivial manipulations are required (*e.g.*, Martin and Liu 2015ab) that can be difficult to carry out in a given problem. This motivated Martin (2015, 2018) to develop a construction based on a more general formulation and establish conditions under which the corresponding inferential model is valid. An advantage of this generalized approach is that there is no need for the aforementioned manipulations, hence it is easier to apply.

A generalized inferential model begins with defining a real-valued function $(y, \theta) \mapsto T_y(\theta)$. When $Y \sim \mathsf{P}_{Y|\theta}$, the random variable $T_Y(\theta)$ has a distribution, which we represent with G_θ. The generalized association then extends the notion in Section 2.1 by connecting the data, parameter, and auxiliary variable via the expression

$$T_Y(\theta) = G_\theta^{-1}(U), \quad U \sim \mathsf{Unif}(0,1).$$

In this paper, we will assume (without loss of generality) that $T_y(\theta)$ is large when data y and parameter value θ disagree; therefore, $T_y(\theta)$ is a sort of distance between data y and parameter θ. The first step to our generalized inferential model (A-step) then yields

$$\Theta_y(u) = \Big\{\theta : G_\theta\Big(T_y(\theta)\Big) = u\Big\}, \quad u \in (0,1).$$

The P-step, as before, requires the introduction of some random set in the u-space; but the structure that has been imposed here virtually determines it. We thus take

$$\mathcal{S} = [0, \tilde{U}], \quad \tilde{U} \sim \mathsf{Unif}(0,1).$$

The C-step returns a new random set

$$\Theta_y(\mathcal{S}) = \bigcup_{u \in \mathcal{S}} \Theta_y(u) = \Big\{\theta : G_\theta\Big(T_y(\theta)\Big) \leq \tilde{U}\Big\}, \quad \tilde{U} \sim \mathsf{Unif}(0,1).$$

Note that this set contains those parameter values that agree with y to some degree, and that this degree is calibrated so that validity holds. That is, if

$$\mathsf{pl}_y(\theta) = \mathsf{P}_{\mathcal{S}}\{\Theta_y(\mathcal{S}) \ni \theta\} = \mathsf{P}_{\tilde{U}}\{G_\theta(T_Y(\theta)) \leq \tilde{U}\} = 1 - G_\theta(T_y(\theta)),$$

then we immediately see that $\mathsf{pl}_Y(\theta) \sim \mathsf{Unif}(0,1)$ under $Y \sim \mathsf{P}_{Y|\theta}$ and, hence, the validity property as stated in (3) holds.

The approach described above returns a plausibility function defined on the full parameter space, Θ. From this, one can carry out marginal inference on any feature, $\psi = \psi(\theta)$, of θ via optimization. In particular, following Shafer (1987, Sec. G), the corresponding marginal point plausibility function for ψ is

$$\mathsf{mpl}_y(\psi) = \sup_{\theta : \psi(\theta) = \psi} \mathsf{pl}_y(\theta), \tag{4}$$

and the validity properties associated with pl carry over immediately to mpl.

But there are cases, like the one we consider in Section 3, where interest is exclusively in a specific feature of θ, and it is beneficial to construct a marginal inferential model directly. Express the full parameter as $\theta = (\psi, \eta)$, where ψ and η are the interest and nuisance parameters, respectively. Next, define a function $(y, \psi) \mapsto T_y(\psi)$ that only directly involves the interest parameter, again, with the property that large values of the function correspond to cases where data and the interest parameter disagree. An example of such a function is the negative log relative profile likelihood, as in (6). If the distribution of $T_Y(\psi)$, as a function of $Y \sim \mathsf{P}_{Y|\psi,\eta}$, does not depend on η, then construction of the generalized inferential model for ψ proceeds exactly as above, with any fixed value of η. This is very similar to the developments of Tsui and Weerahandi (1989) in the framework of generalized p-values. In many applications, however, including our meta-analysis problem, the distribution of $T_Y(\psi)$ does depend on the nuisance parameter, so some non-trivial adjustments are required. We discuss this in detail in Section 3.

3. Inferential Models for Meta-Analysis

3.1. Construction

For our meta-analysis case, write $\mathsf{P}_{Y|\mu,\nu}$ for the joint distribution of $Y = (Y_1, \ldots, Y_K)$, where Y_k are independently generated from a $\mathsf{N}(\mu, \sigma_k^2 + \nu)$. It is straightforward to write down an association that links the data Y, the unknown parameter $\theta = (\mu, \nu)$, and a set of auxiliary variables, *e.g.*,

$$Y_k = \mu + (\sigma_k^2 + \nu)^{1/2} U_k, \quad k = 1, \ldots, K,$$

where U_k's are iid $\mathsf{N}(0, 1)$. Following Martin and Liu (2015a), the next step would be to reduce the dimension of $(U_1, \ldots, U_K)$ to match that of θ. This step turns out to be challenging but, fortunately, a generalized inferential model is within reach.

The full parameter is $\theta = (\mu, \nu)$ but, since only μ is of interest, marginalization is desired. As discussed in Section 2.2, there are at least two ways to proceed. The first is to start with a summary $T_Y(\theta) = T_Y(\mu, \nu)$ of the data and full parameter, which takes large values when data and the candidate parameters disagree, and then marginalize to the μ-space *after* constructing the plausibility function on the (μ, ν)-space. That is, we define a plausibility function on the full parameter space as

$$\mathsf{pl}_y(\mu, \nu) = 1 - G_{\mu,\nu}\Big(T_y(\mu, \nu)\Big),$$

where $G_{\mu,\nu}$ is the distribution of $T_Y(\mu, \nu)$ under $Y \sim \mathsf{P}_{Y|\mu,\nu}$. Like in (4), we obtain our desired marginal plausibility by optimization:

$$\mathsf{mpl}_y(\mu) = \sup_\nu \mathsf{pl}_y(\mu, \nu).$$

The corresponding $100(1-\alpha)\%$ plausibility interval for μ is $\{\mu : \mathsf{mpl}_y(\mu) > \alpha\}$, corresponding to a projection of the joint plausibility region for the full parameter onto the μ-space. After some reflection on the solution in Michael *et al.* (2019), one sees that it is precisely a generalized inferential model as just described, with $T_y(\mu, \nu)$ given by

$$T_y(\mu, \nu) = T_w(\mu) + c_0 \log \frac{L_y(\mu, \hat{\nu}_{DL})}{L_y(\mu, \nu)}, \tag{5}$$

a linear combination of DerSimonian and Laird's Wald-type summary statistic $T_w(\mu)$ and a log likelihood ratio, with a constant c_0 controlling its contribution. Those authors do not describe their proposal as a (generalized) inferential model, but we believe that this perspective is beneficial both for developing some intuition about their solution and for comparing with our proposed solution.

Despite the ease of marginalizing out the nuisance parameter from the joint plausibility function, we adopt the second strategy for marginalization discussed in Section 2.2 that eliminates the nuisance parameter *before* constructing the plausibility function. That is, we start with a function, $T_y(\mu)$, that does not directly involve ν. Like in Goodman (1989), we recommend the use of a negative log relative profile in which

$$T_y(\mu) = -\log \frac{\sup_\nu L_y(\mu, \nu)}{\sup_{\mu,\nu} L_y(\mu, \nu)}, \quad \mu \in \mathbb{R}. \tag{6}$$

Here $L_y(\mu, \nu) \propto \prod_{k=1}^K (\sigma_k^2 + \nu)^{-1/2} \exp\{-\frac{1}{2}(\sigma_k^2 + \nu)^{-1}(y_k - \mu)^2\}$ is the likelihood function under the assumed model $\mathsf{P}_{Y|\mu,\nu}$. Note that the ν value at which the maximum is attained in the numerator—call it $\hat{\nu}_\mu$—depends on the specified value of μ. As before, we define the distribution function of $T_Y(\mu)$ under the model $Y \sim \mathsf{P}_{Y|\mu,\nu}$ as

$$G_\nu(t) = \mathsf{P}_{Y|0,\nu}\{T_Y(0) \leq t\}, \quad t > 0.$$

Here we have inserted the default zero value for μ because the location model structure means the distribution of $T_Y(\mu)$ does not depend on the value of μ, when $Y \sim \mathsf{P}_{Y|\mu,\nu}$. Following the remainder of the construction outlined in Section 2.2, we arrive at a marginal point plausibility function

$$\text{“}\mathsf{mpl}_y(\mu)\text{”} = 1 - G_\nu(T_y(\mu)).$$

The quotation marks on the left-hand side are to signal that this is not a function that we can actually work with because the right-hand side depends on the unknown value of the nuisance parameter ν. To overcome this, we will use a plug-in estimate for ν, where it appears in G_ν. Before proceeding, it is important to emphasize that our plug-in proposal is fundamentally different than those mentioned in Section 1; we discuss this in more detail in Section 3.3. For our plug-in estimator, we propose to use that ν value where the maximum in the numerator of the profile likelihood is attained, namely, $\hat{\nu}_\mu$, which implicitly depends on data y. Putting it all together, our proposed marginal point plausibility function for μ is

$$\mathsf{mpl}_y(\mu) = 1 - G_{\hat{\nu}_\mu}(T_y(\mu)), \quad \mu \in \mathbb{R}, \tag{7}$$

which is now just a function of data and the generic argument μ. This function can be plotted to visualize what the data suggests about where the true value of μ is and, more formally, we can read off a $100(1-\alpha)\%$ marginal plausibility interval for μ as follows:

$$\{\mu : \mathsf{mpl}_y(\mu) > \alpha\}.$$

Computation of mpl_y requires an approximation of the analytically intractable distribution $G_{\hat{\nu}_\mu}$, but this is straightforward to do via Monte Carlo; see Algorithm 1. And once $\mathsf{mpl}_y(\mu)$ is available on a grid of values, extracting the plausibility interval for μ is easy, but the endpoints could be targeted more directly using, say, the proposed Monte Carlo method coupled with stochastic approximation.

Algorithm 1: Monte Carlo approximation of $G_{\hat{\nu}_\mu}$

Generate M samples of K study-level errors $e^*_{1m}, \ldots, e^*_{Km} \sim \mathsf{N}(0,1)$ Set a fine grid of μ values **for** *each μ value on the specified grid* **do**
  Find $\hat{\nu}_\mu$ for the observed data y and given μ;
  for $m = 1, \ldots, M$ **do**
    Set $Y^*_{km} = (\sigma_k^2 + \hat{\nu}_\mu)^{1/2} e^*_{km}$ for $k = 1, \ldots, K$;
    Calculate $T^*_m = T_{Y^*_m}(0)$ based on $Y^*_m = (Y^*_{1m}, \ldots, Y^*_{Km})$;
  end
  Approximate $G_{\hat{\nu}_\mu}(t)$ by $M^{-1} \sum_{m=1}^M 1\{T^*_m \leq t\}$;
end

3.2. Illustration

To illustrate our proposed method, two examples are shown in Figure 1. In each, a meta-analysis is carried out on $K = 3$ studies, and each study's variance σ_k^2 generated from a inverse gamma distribution with a shape and scale parameter of 1. The data supplied from each of these hypothetical studies were generated from a normal distribution in which the true population mean was set at $\mu = 0$ and the variance at $\nu + \sigma_k^2$. It is straightforward to construct plausibility functions for the individual studies, with data (Y_k, σ_k^2), for $k = 1, 2, 3$, under the normality assumption (see, *e.g.*, Martin 2017), and these curves are plotted in gray in Figure 1. The black curve corresponds to the marginal plausibility function, $\mu \mapsto \mathsf{mpl}_y(\mu)$, described in the previous section, evaluated using the Monte Carlo method outlined in Algorithm 1. Panels (a) and (b) correspond to $\nu = 1$ and $\nu = 2$, respectively. Note that, as expected, the black curve is a "combination" of the three gray curves, with more influence coming from those gray curves that are tighter, corresponding to a more informative individual study.

3.3. Theoretical properties

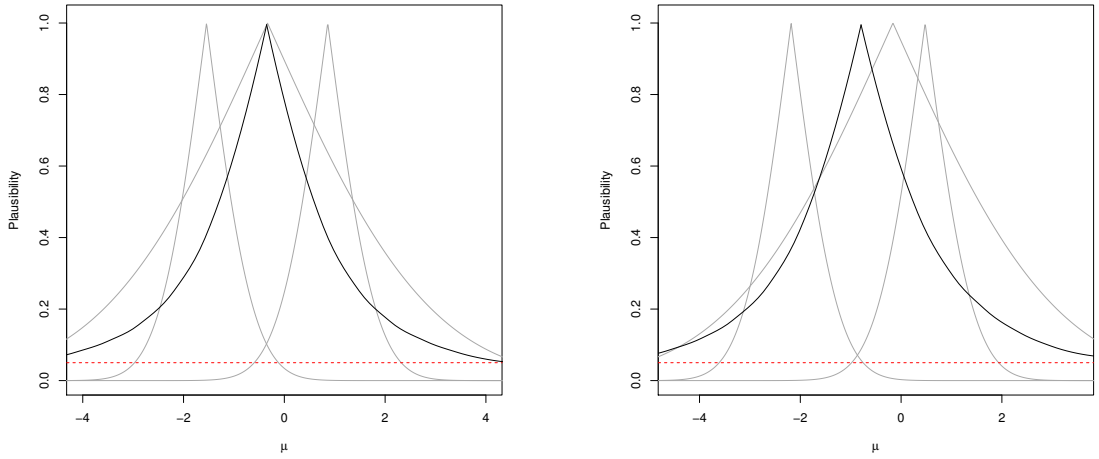
If the value of the nuisance parameter ν were known, and used, in our construction of the (marginal) generalized inferential model for μ, then the validity property, as stated in (3), would be immediate. For the practical case where ν is unknown, we have recommended the inferential model with marginal point plausibility function (7), which involves a plug-in estimator. Our use, however, of this plug-in $\hat{\nu}_\mu$, complicates verification of the validity property. At the very least, under mild assumptions, our proposed generalized inferential model would be valid for large K, and the following theorem confirms this.

Theorem 1: Let $Y^K = (Y_1, \ldots, Y_K)$ be an independent sample from the random effects model, $\mathsf{P}_{Y|\mu,\nu}$, described above, where both μ and ν are unknown, but each within-study variance σ_k^2 is known. Then the marginal plausibility function mpl_{Y^K} in (7) satisfies

$$\mathsf{mpl}_{Y^K}(\mu) \to \mathsf{Unif}(0,1) \quad \text{in distribution under } \mathsf{P}_{Y|\mu,\nu} \text{ as } K \to \infty.$$

In particular, the marginal plausibility region $\{\mu : \mathsf{mpl}_{Y^K}(\mu) > \alpha\}$ has coverage probability approximately equal to $1 - \alpha$, for large K.

Proof: See the Appendix. □



(a) $\nu = 1$ **(b)** $\nu = 2$

Figure 1: Examples of two simulated meta-analyses where the number of studies available $K = 3$. Plausibility functions associated with each individual study (in gray) and the combined plausibility function associated with our proposed inferential methods approach (in black). Marginal 95% plausibility intervals for μ can be obtained where the combined plausibility intersects with $\alpha = 0.05$.

The above theorem only says that the proposed solution is approximately valid for large K, but we do have reason to believe that the support for the proposed solution is actually stronger than this theorem suggests. Indeed, numerically, even for small K, the distribution of $\mathsf{mpl}_Y(\mu)$ is very close to uniform. As shown in Figure 2, based on 10,000 samples of the data pairs (Y_k, σ_k^2) from a small number of studies $K = \{3, 4, 5\}$ and a high level of heterogeneity between studies $\nu = 5$, the distribution is close to uniform. There is some deviation to the left of uniform when the number of studies included for meta-analysis is particularly small, $K = 3$, but this is in the middle of the distribution, not in the lower tails (*e.g.*, around 0.05) where we would naturally be interested. Therefore, the method appears to be not only valid, but nearly exact.

It is natural to ask: why does our proposed method achieve this apparent higher-order of accuracy? At least intuitively, this can be answered by noticing that our proposed method has features in common with both the exact and higher-order asymptotically accurate methods described above. That is, by starting with the relative profile likelihood $T_Y(\mu)$ in (6), we remove almost all of the dependence on the nuisance parameter; that is, by Wilks's theorem, the profile likelihood ratio has a known distribution—no nuisance parameter dependence—up to first order. This means that the exact distribution, G_ν, of our $T_Y(\mu)$ is roughly constant in ν. Therefore, even though there is some remaining dependence on ν, which is why a plug-in estimator is needed, it is not necessary that it be an especially accurate estimate. Ultimately, our final inferential model is built using the plug-in distribution $G_{\hat{\nu}_\mu}$, at each individual μ value, which is very close to the exact distribution. It is this extra accuracy that leads to the superior practical performance in Figure 2 and Section 4, beyond what would be expected from the large-K approximate validity result in Theorem 1.

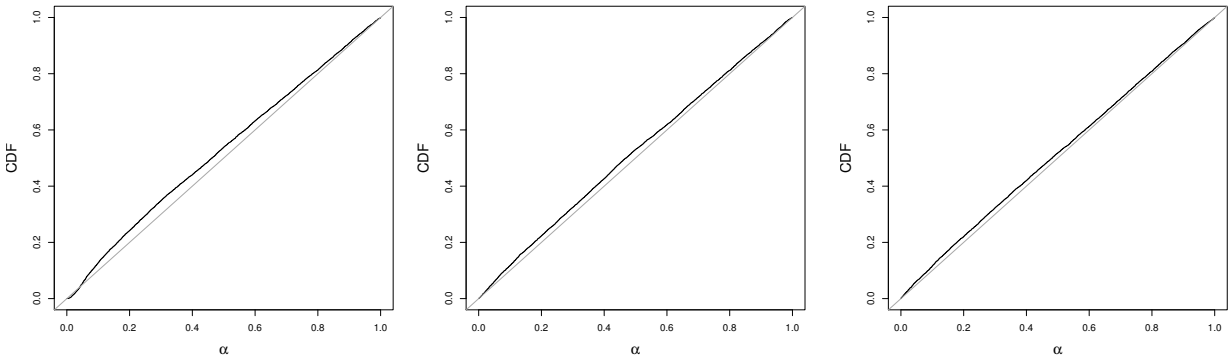


Figure 2: Distribution function $\alpha \mapsto \mathsf{P}_{Y|\mu,\nu}\{\mathsf{mpl}_Y(\mu) \leq \alpha\}$ of $\mathsf{mpl}_Y(\mu)$ (black) compared with that of a $\mathsf{Unif}(0,1)$ (gray) based on 10000 Monte Carlo samples from a normal-normal random effects model.. From left to right, $K = 3, 4, 5$.

4. Simulation Studies

Our simulations examine the performance of our proposed method, compared to that of other existing methods, in the case where $K \leq 7$. We generate our K study-level observations from the normal–normal random effects model, where the within-study variances—the σ_k^2 values, for $k = 1, \ldots, K$—are taken as fixed constants; following Gelman (2006), we generate these "fixed values" from an inverse gamma distribution with a shape and scale parameter of 1. Here we fix the overall effect at $\mu = 0$ but vary the between-study variances as $\nu \in \{1, 3, 5\}$, so that we capture various settings from low to high levels of heterogeneity. For each combination of K and ν, we repeat the experiment 1000 times to get estimates of the coverage probability and mean length of various 95% confidence intervals.

We compare the coverage properties of our approach (IM) against that of DerSimonian and Laird (1986, DL), the exact method in Michael *et al.* (2019, EX), the signed profile log likelihood ratio in Severini (2000, LK), its Skovgaard corrected cousin in Guolo (2012, SV), as implemented in the `metaLik` package in R (Guolo and Varin 2012), a traditional full Bayesian solution with a non-informative Jeffreys prior, as implemented in the `bayesmeta` package (Röver 2012), and, as a benchmark, an *oracle* procedure that knows the true value of ν and uses the classical normal distribution theory for inference on μ. For EX, we set the tuning parameter c_0 to the values based on K as recommended in Michael *et al.* (2019).

As shown in Figure 3, our proposed generalized inferential method outperforms all the other methods—except, of course, the oracle—in terms of both coverage and mean interval length. The Bayes, exact and higher-order likelihood methods tend to have too high nominal coverage, and the others too low. The over-coverage seen here is consistent with the results shown in Michael *et al.* (2019). Oddly, when the between-study variance parameter is set to a higher value $\nu = 5$, so that the average heterogeneity among the 1000 simulations is high, the Skovgaard corrected signed profile log likelihood actually achieves nominal coverage across all small settings of K. This is in line with the results in Guolo (2012), in which the estimator is sensitive to the level of heterogeneity.

To further highlight this sensitivity of the higher-order likelihood approach, we reran our methods above in the same settings used in Guolo (2012); more specifically, we

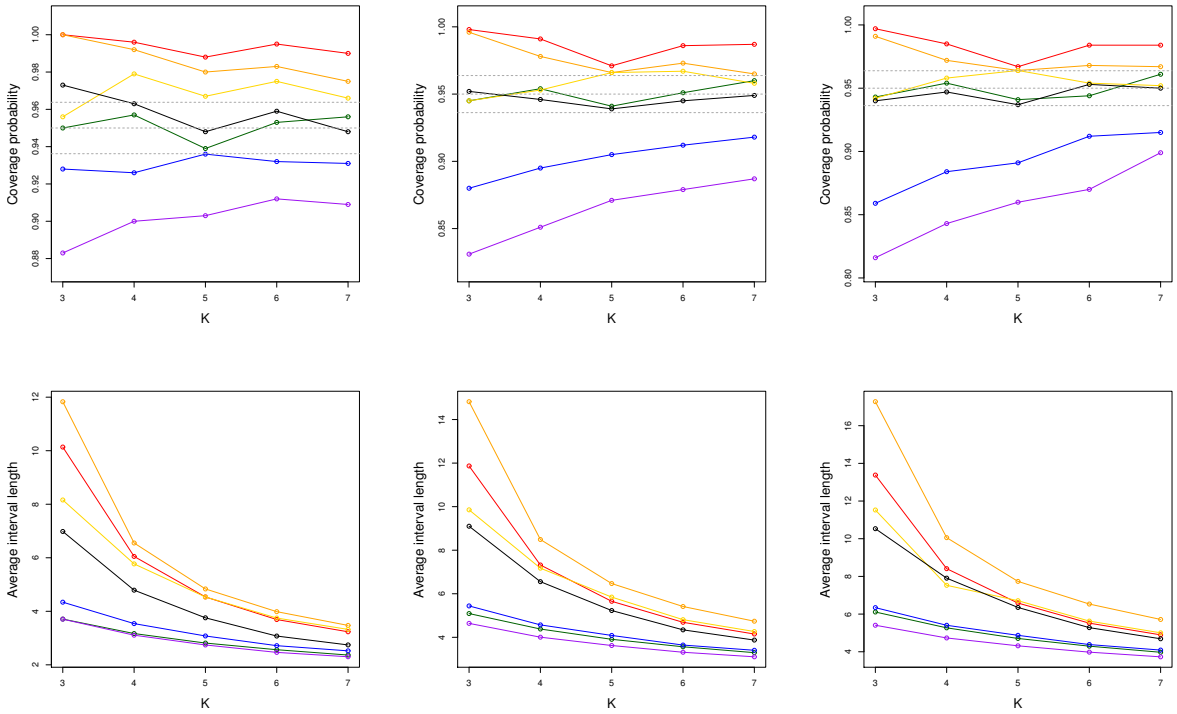


Figure 3: Coverage probabilities and mean interval lengths across 15 different simulation settings for the number of studies available for meta-analysis K and the level of heterogeneity ν. Results for DL (purple), LK (blue), oracle (green), SV (yellow), EX (orange), Bayes (red), and our proposed IM (black). From left to right, data are generated from a fixed between-study variance $\nu \in \{1, 3, 5\}$.

re-examined the performance of our IM approach with $K \in \{3, 4, 5, 6, 7\}$, $\mu = 0.5$, and $\nu = \{0.08, 0.10, 0.12\}$. We also generate the within-study variances for each study K from a uniform distribution on the interval 0.01 and 0.06 as done in Guolo (2012). As shown in Figure 4, our method still outperforms in these settings. Conversely, the under-coverage of DL and LK across these two different simulation settings in Figure 3 and Figure 4 are to be expected, as K is too small for the first-order asymptotic approximations to kick in. Moreover, our proposed method's strong coverage performance is not the result of having overly wide intervals: our mean lengths fall right in between those of the over-and under-coverage methods, and are quite close to that of the oracle as K becomes larger. Remarkably, these patterns hold across different heterogeneity levels as well.

5. Real Data Analyses

5.1. Changes in bone mineral density

To demonstrate how their inference procedure performs against popular meta-analytic techniques, Michael *et al.* (2019) carry out four separate meta-analyses on 59 randomised trials presented in Tai *et al.* (2015). These meta-analyses differ in two categories: (a) the specific bones from which the outcome measure, or the change in bone mineral density (BMD), was measured; and (b) the number studies included. The first meta-analysis consisted of

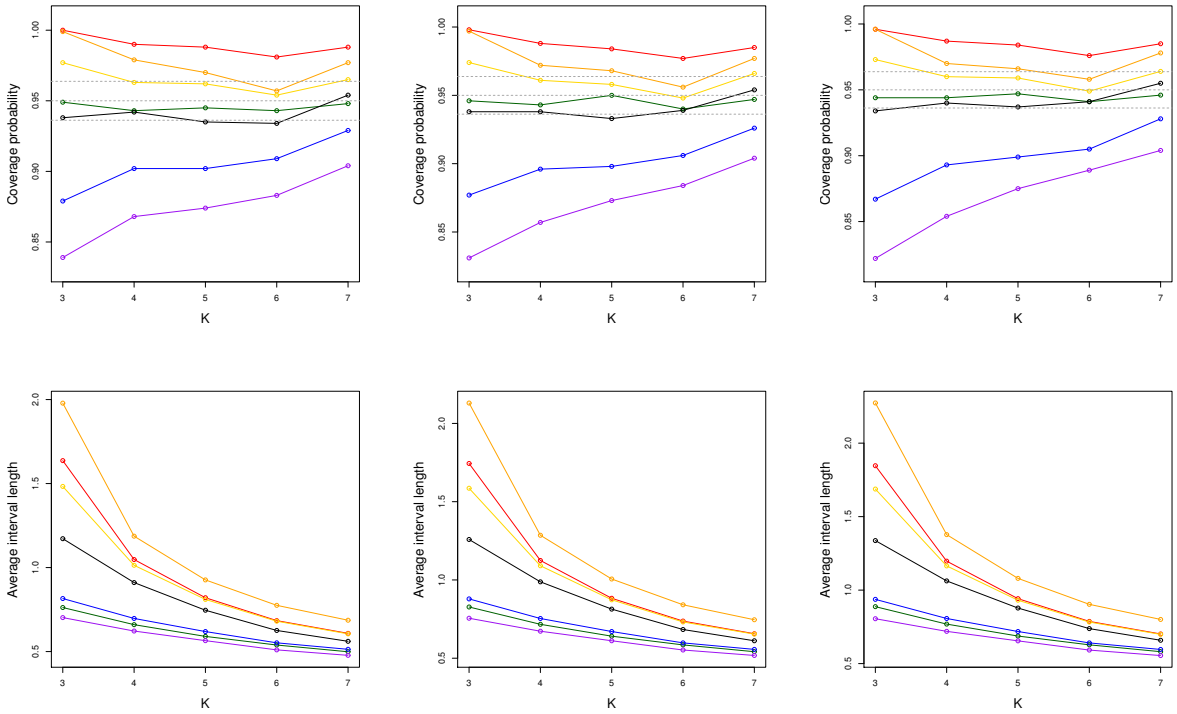


Figure 4: Coverage probabilities and mean interval lengths across 15 different simulation settings as done in Guolo (2012) for the number of studies available for meta-analysis K and the level of heterogeneity ν. Results for DL (purple), LK (blue), oracle (green), SV (yellow), EX (orange), Bayes (ired), and our proposed IM (black). From left to right, data are generated from a fixed between-study variance $\nu \in \{0.08, 0.10, 0.12\}$ across various number of studies, K.

27 separate trials in which BMD changes were taken from the lumbar spine, followed by a meta-analysis of six trials of BMD changes from the hip, five from the forearm, and three from the total body. As shown in Table 1, for the first meta-analysis, our 95% plausibility interval almost matches the exact confidence interval from Michael *et al.* (2019), which also approximately aligns with that of the classical DerSimonian–Laird approach. This is no surprise as the number of studies $K = 27$ itself is large. As for subsequent studies in which $K \leq 6$, the comparison between DerSimonial–Laird and the other two methods changes a lot and, in fact, sometimes leads to different scientific conclusions. For example, one would conclude a significant change in BMD from the forearm and total body meta-analyses based on DerSimonian–Laird, but conclude no significant change based on our method and that of Michael *et al.* (2019). Given that the latter two approaches have stronger theoretical support than the former, the difference in conclusions here might be indicative of the increased risk of false positives when using traditional meta-analytic techniques.

5.2. Risk of acute myocardial infarction

Here we consider a controversial example, one that called to question the use of meta-analyses in general (Egger and Smith 1995; Flather *et al.* 1997). Teo *et al.* (1991) conducted

Table 1: Four meta-analyses on the effect of calcium supplements in changes in bone mineral density from Tai *et al.* (2015). Intervals based on three methods (DL, EX, and our IM) are reported, with interval lengths as subscripts.

Study	K	DL	EX	IM
Lumbar spine	27	$(0.828, 1.669)_{0.841}$	$(0.768, 1.726)_{0.958}$	$(0.811, 1.642)_{0.831}$
Total hip	6	$(0.502, 1.847)_{1.345}$	$(0.159, 2.246)_{2.087}$	$(0.319, 2.131)_{1.812}$
Forearm	5	$(0.209, 3.378)_{3.169}$	$(-0.459, 4.124)_{4.583}$	$(-0.426, 4.625)_{5.052}$
Total body	3	$(0.268, 1.778)_{1.511}$	$(-0.740, 2.796)_{3.536}$	$(-0.486, 2.568)_{3.054}$

a meta-analysis of seven clinical trials that examined mortality across 1301 patients, 657 of which received intravenous magnesium therapy within 12 hours of hospitalization for acute myocardial infarction and 644 of which did not. In the original work, a fixed-effect method was used to combine the results from these seven randomized trials and arrive at a common odds ratio of 0.47, with 95% confidence interval $(0.28, 0.79)$—suggesting magnesium therapy to be highly effective in reducing mortality among this specific patient population. The expected drop in mortality, however, was refuted in a large-scale 58,050-patient follow-up study (Fourth International Infarct Survival Collaborative 1995) that estimated a common odds ratio of 1.06 with 95% confidence interval $(1.00, 1.12)$. As a result, researchers raised concerns about meta-analytic techniques in general, citing issues around publication biases (Yusuf and Flather 1995) and high heterogeneity between studies (Flather *et al.* 1997). To address these problems, the canonical recommendation was to conduct sensitivity analyses via the use of multiple meta-analysis procedures, like that discussed below. Had such a precaution been taken, the fact that the Fourth International Infarct Survival Collaborative study lead to an alternative conclusion would not have been unforeseen.

It is also worth noting here that since the raw observations recorded in Teo *et al.* (1991) are in the form of a dichotomous outcome variable, we take the logarithm of the common odds ratio, between the mortality rate of patients that receive magnesium therapy and that of patients that do not, in order to conduct our meta-analysis as described in the competing procedures above. While there are other simplifications, *e.g.*, Van Houwelingen *et al.* (1993), we subscribe to the rationale in DerSimonian and Laird (1986) that regards the distribution of the log odds as approximately normal. Figure 5 thus compares the resulting interval estimates based on several meta-analytic procedures, namely, those assessed in Section 4. Note that the DerSimonian–Laird and signed profile log likelihood intervals approximate the original results from Teo *et al.* (1991). However, our proposed approach, along with the full Bayesian, the higher-order likelihood, and that in Michael *et al.* (2019), result in an odds ratio interval that suggests magnesium therapy does not significantly affect the short-term mortality of patients with acute myocardial infarction.

6. Conclusion

In this paper, we have considered an important and challenging problem, namely, valid statistical inference for meta-analyses that combine only a few studies. Again, the main obstacle is in dealing with the unknown between-study variance, in which there is only limited information in the few studies being combined. Our proposed solution is based on a recently

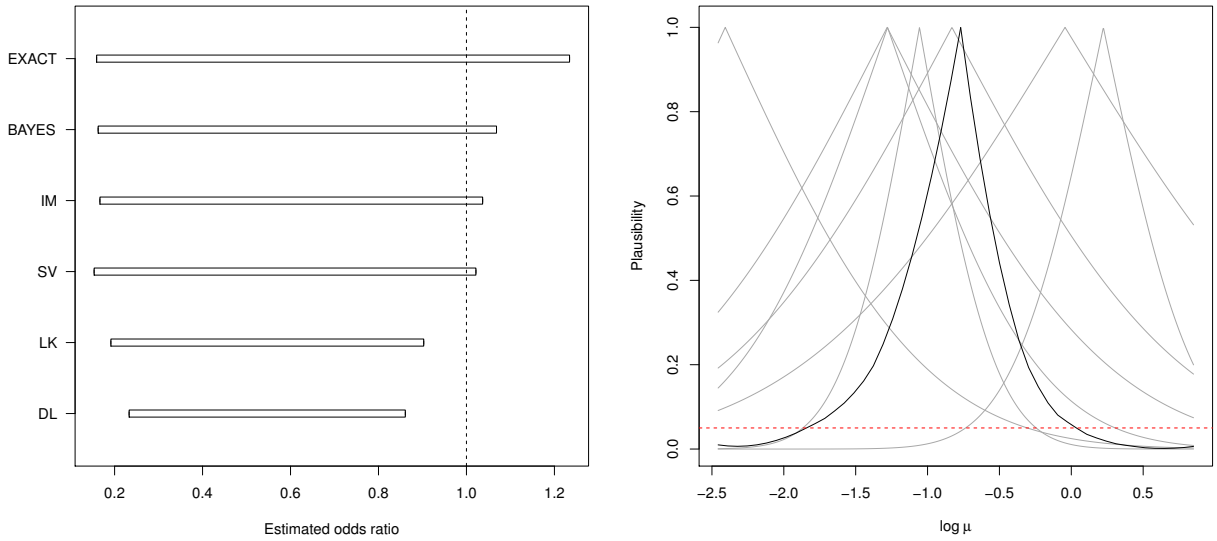


Figure 5: (Left) Estimated odds ratio intervals for various meta-analytic techniques in the example described in Section 5.2. (Right) Combined plausibility function and respective IM interval on the log scale.

proposed generalized inferential model framework, and we harness the power of profiling to construct a generalized association that is "almost" independent of the nuisance between-study variance. From there, we can use the exact distribution of the profile likelihood ratio, as the lack of sensitivity to the nuisance parameter means that it is not necessary to have an accurate plug-in estimator to achieve near-exact inference. In our numerical comparisons, we have demonstrated that the proposed inferential model solution outperforms existing methods in the literature, by being nearly exact and more efficient across a wide range of simulation settings, with few studies and both large and small between-study variance.

Given the strong performance in this meta-analysis application, it is natural to consider using the same generalized inferential model approach to solve other challenging problems. One that we have recently considered is when data come from a parametric model are corrupted by random censoring. The classical solution to this problem relies on the asymptotic normality of maximum likelihood estimators and, therefore, can only give approximately valid inference in an asymptotic sense. But the use of a likelihood ratio effectively marginalizes out the nuisance censoring distribution, so we end up in a position similar to that encountered in the present paper, the key difference being that the nuisance parameter is infinite-dimensional, which creates computational challenges. Preliminary results on this can be found in Cahoon and Martin (2019) and more details are forthcoming.

Acknowledgments

The authors thank two anonymous reviewers for helpful feedback on an earlier version of the manuscript. They are also grateful to the editorial board of *Statistics and Applications* for the opportunity to make a contribution to this special issue in honor of the "Twin Statisticians," Drs. Bikas and Bimal Sinha.

References

Balch, M. S., Martin, R. and Ferson, S. (2019). Satellite conjunction analysis and the false confidence theorem. *Proceedings of the Royal Society*, **A475**(**2227**), 20180565.

Barnard, G. A. (1995). Pivotal models and the fiducial argument. *International Statistical Review*, **63**(**3**), 309–323.

Cahoon, J. and Martin, R. (2019). Possibility measures for valid statistical inference based on censored data. In De Bock, J., de Campos, C.P., de Cooman, G., Quaeghebeur, E., and Wheeler, G. (Eds.), *Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications*, Volume **103** of *Proceedings of Machine Learning Research*, pages 49–58, Thagaste, Ghent, Belgium. PMLR.

Chung, Y., Rabe-Hesketh, S. and Choi, I. -H. (2013). Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in Medicine*, **32**(**23**), 4071–4089.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, **10**(**1**), 101–129.

Davey, J., Turner, R. M., Clarke, M. J. and Higgins, J. P. (2011). Characteristics of meta-analyses and their component studies in the Cochrane database of systematic reviews: a cross-sectional, descriptive analysis. *BMC Medical Research Methodology*, **11**(**1**), 160.

Dawid, A. P. and Stone, M. (1982). The functional-model basis of fiducial inference. *The Annals of Statistics*, **10**(**4**), 1054–1067.

Dempster, A. P. (2008). The Dempster–Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, **48**(**2**), 365–377.

DerSimonian, R. and Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: an update. *Contemporary Clinical Trials*, **28**(**2**), 105–114.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, **7**(**3**), 177–188.

Egger, M. and Smith, G. D. (1995). Misleading meta-analysis. *British Medical Journal Publishing Group*.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference.* Hafner Publishing Co.

Flather, M. D., Farkouh, M. E., Pogue, J. M. and Yusuf, S. (1997). Strengths and limitations of meta-analysis: larger studies may be more reliable. *Controlled Clinical Trials*, **18**(**6**), 568–579.

Follmann, D. A. and Proschan, M. A. (1999). Valid inference in random effects meta-analysis. *Biometrics*, **55**(**3**), 732–737.

Fourth International Infarct Survival Collaborative (1995). ISIS–4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet*, **345**(**8951**), 669–685.

Fraser, D. A. S. (1968). *The Structure of Inference.* Wiley.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**(**3**), 515–534.

Goodman, S. N. (1989). Meta-analysis and evidence. *Controlled Clinical Trials*, **10**(**2**), 188–204.

Guolo, A. (2012). Higher-order likelihood inference in meta-analysis and meta-regression. *Statistics in Medicine*, **31**(**4**), 313–327.

Guolo, A. and Varin, C. (2012). The R package metaLik for likelihood inference in meta-analysis. *Journal of Statistical Software*, **50**(**7**), 1–14.

Hardy, R. J. and Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*, **15**(**6**), 619–629.

Jackson, D., Bowden, J. and Baker, R. (2010). How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? *Journal of Statistical Planning and Inference*, **140**(**4**), 961–970.

Liu, S., Tian, L., Lee, S. and Xie, Min-ge (2018). Exact inference on meta-analysis with generalized fixed-effects and random-effects models. *Biostatistics and Epidemiology*, **2**(**1**), 1–22.

Martin, R. (2015). Plausibility functions and exact frequentist inference. *Journal of the American Statistical Association*, **110**(**512**), 1552–1561.

Martin, R. (2017). A statistical inference course based on p-values. *The American Statistician*, **71**(**2**), 128–136.

Martin, R. (2018). On an inferential model construction using generalized associations. *Journal of Statistical Planning and Inference*, **195**, 105–115.

Martin, R. (2019). False confidence, non-additive beliefs, and valid statistical inference. *International Journal of Approximate Reasoning*, **113**, 39–73.

Martin, R. and Liu, C. (2013). Inferential models: a framework for prior-free posterior probabilistic inference. *Journal of the American Statistical Association*, **108**(**501**), 301–313.

Martin, R. and Liu, C. (2015a). Conditional inferential models: combining information for prior-free probabilistic inference. *Journal of the Royal Statistical Society*, **B77**(**1**), 195–217.

Martin, R. and Liu, C. (2015b). Marginal inferential models: prior-free probabilistic inference on interest parameters. *Journal of the American Statistical Association*, **110**(**512**), 1621–1631.

Martin, R. and Liu, C. (2016). *Inferential Models: Reasoning with Uncertainty*. CRC Press, Boca Raton, FL.

Michael, H., Thornton, S., Xie, M. and Tian, L. (2019). Exact inference on the random-effects model for meta-analyses with few studies. *Biometrics*, **75**(**2**), 485–493.

Paule, R. C. and Mandel, J. (1982). Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards*, **87**(**5**), 377–385.

Röver, C. (2017). Bayesian random-effects meta-analysis using the bayesmeta R package. `https://arxiv.org/abs/1711.08683/`.

Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford University Press.

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *The Annals of Statistics*, **20**(**4**), 1768–1802.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

Shafer, G. (1987). Belief functions and possibility measures. In Bezdek, J., Editor, *The Analysis of Fuzzy Information*, Volume **1**, pages 51–84. CRC Press.

Sidik, K. and Jonkman, J. N. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine*, **26**(**9**), 1964–1981.

Tai, V., Leung, W., Grey, A., Reid, I. R. and Bolland, M.J. (2015). Calcium intake and bone mineral density: systematic review and meta-analysis. *British Medical Journal*, **351**, 41–83.

Taraldsen, G. and Lindqvist, B. H. (2013). Fiducial theory and optimal inference. *The Annals of Statistics*, **41**(**1**), 323–341.

Teo, K. K., Yusuf, S., Collins, R., Held, P. H. and Peto, R. (1991). Effects of intravenous magnesium in suspected acute myocardial infarction: Overview of randomised trials. *British Medical Journal*, **303**(**6816**), 1499–1503.

Tsui, K.-W. and Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, **84**(**406**), 602–607.

Van Houwelingen, H. C., Zwinderman, K. H. and Stijnen, T. (1993). A bivariate approach to meta-analysis. *Statistics in Medicine*, **12**(**24**), 2273–2284.

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P., Langan, D. and Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, **7**(**1**), 55–79.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, **30**(**3**), 261–293.

Wang, Y. and Tian, L. (2018). An efficient numerical algorithm for exact inference in meta analysis. *Journal of Statistical Computation and Simulation*, **88**(**4**), 646–656.

Yusuf, S. and Flather, M. (1995). Magnesium in acute myocardial infarction. *British Medical Journal*, **310**, 751–752.

APPENDIX
Proof of Theorem 1

By definition of mpl_{Y^K} in (7), it is enough to show that $G^K_{\hat{\nu}_\mu}(T_{Y^K}(\mu))$ converges in distribution to $\mathsf{Unif}(0,1)$ under $\mathsf{P}_{Y^K|\mu,\nu}$; note that here we insert the superscript "K" to highlight the dependence on the number of studies, K. Then we can write $G^K_{\hat{\nu}_\mu}(T_{Y^K}(\mu)) = G^K_\nu(T_{Y^K}(\mu)) + \Delta_K$, where

$$\Delta_K = G^K_\nu(T_Y(\mu)) - G^K_{\hat{\nu}_\mu}(T_Y(\mu)),$$

with $\hat{\nu}_\mu$, the maximum likelihood estimate of the heterogeneity parameter at a fixed μ, and ν the true heterogeneity between studies. The key observation is that $G^K_\nu(T_{Y^K}(\mu))$ is exactly uniformly distributed under $\mathsf{P}_{Y^K|\mu,\nu}$, so if we can show $\Delta_K \to 0$ in $\mathsf{P}_{Y^K|\mu,\nu}$-probability, then the claim follows from Slutsky's theorem.

Towards this, we clearly have

$$|\Delta_K| \leq \sup_{t\in[0,1]} |G^K_\nu(t) - G^K_{\hat{\nu}_\mu}(t)|,$$

so we can prove our claim by showing the difference between the two distribution functions vanishes uniformly. But since these are distribution functions, it is enough to show that the difference vanishes pointwise, at each fixed t. Towards this, according to Guolo (2012), the meta-analysis problem is sufficiently regular that the classical first-order distribution theory applies; see, *e.g.*, Severini and Wong (1992, Sec. 4.6). In particular, this implies $\hat{\nu}_\mu = \nu + O_P(K^{-1/2})$ which, in turn, implies that $\mathsf{P}_{Y^K|\mu,\nu}$ and $\mathsf{P}_{Y^K|\mu,\hat{\nu}_\mu}$ are mutually contiguous. Then the classical Wilks's theorem gives us

$$-2\log T_{Y^K}(\mu) \to \mathsf{ChiSq}(1) \quad \text{in distribution, as } K \to \infty, \tag{8}$$

under *both* $\mathsf{P}_{Y^K|\mu,\nu}$ and $\mathsf{P}_{Y^K|\mu,\hat{\nu}_\mu}$. Therefore,

$$G_\nu^K(t) \to G^\infty(t) \quad \text{and} \quad G_{\hat{\nu}_\mu}^K(t) \to G^\infty(t),$$

where G^∞ is the limiting distribution function of $T_{Y^K}(\mu)$ from (8). Now, if we write

$$|G_\nu^K(t) - G_{\hat{\nu}_\mu}^K(t)| \leq |G_\nu^K(t) - G^\infty(t)| + |G_{\hat{\nu}_\mu}^K(t) - G^\infty(t)|,$$

then we see the right-hand converges to 0 in $\mathsf{P}_{Y|\mu,\nu}$-probability as $K \to \infty$. This implies $\Delta_K \to 0$ from which $G_{\hat{\nu}_\mu}^K(T_{Y^K}(\mu)) \to \mathsf{Unif}(0,1)$ follows by Slutsky's theorem.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 317–332

Horseshoe and Strawderman-Berger Estimators for Non-negative Normal Means

Neha Agarwala[1], Junyong Park[2] and Anindya Roy[1]
[1]*Department of Mathematics and Statistics,*
University of Maryland, Baltimore County, Baltimore, MD, USA
[2]*Department of Statistics,*
Seoul National University, Seoul, South Korea

Received: 22 June 2020; Revised: 25 August 2020; Accepted: 06 September 2020

Abstract

High dimensional inference problems are generating considerable interest due to the availability and accessibility of massive amount of data in several fields. Modern statistical problems, however, involve natural constraints on model parameters. For such estimation problems, it is not fitting to apply standard estimates designed for unrestricted alternatives and then to truncate it. Given a n dimensional independent normal observation with common variance, we consider the classical normal mean estimation problem where the mean vector lies in a non-negative orthant. We study the behavior and risk properties of Bayesian estimators under two popular priors, the horseshoe prior and Strawderman-Berger prior, originally developed in the unrestricted mean vector estimation regime and then restrict the distribution of prior to satisfy the parameter constraint. The performance of posterior mean based on the horseshoe prior and the posterior mean and posterior median based on Strawderman-Berger prior is compared with the maximum likelihood estimator, numerically under different sparsity configurations.

Key words: Constrained normal means; Shrinkage estimators; Mixture distribution; Sparsity.

1. Introduction

Traditional statistical theory has mostly focused on methods developed for large samples and a small number of features. The modern scientific world, however, is moving fast towards the regime of high dimensional data. In the high dimensional setting, often one deals with the case when only few variables are relevant. Thus, it has become increasingly important to identify true signals as the data tends to be sparse. Probably the most common of such high dimensional sparse estimation problems is estimation of the mean of a normal distribution when sample size is small compared to the dimension. It is the proverbial needle in a haystack problem that has received much attention in the literature. The setting of the problem is simple. Given data $y_1, \ldots, y_n$, arising independently from the model

$$y_i|\mu_i, \sigma^2 \sim N(\mu_i, \sigma^2),$$

Corresponding Author: Anindya Roy
Email: anindya@umbc.edu

one wishes to estimate the entire vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$. Of course, given that there are only n independent observations for $(n+1)$ unknown parameters, additional assumptions are needed for meaningful estimation of the mean vector. Usually some level of sparsity is assumed for the true mean vector. Both Bayesian and frequentist estimators have been developed for this problem, the most well known being the shrinkage estimators starting with James and Stein (1961), thresholding estimators starting with Donoho and Johnston (1994), penalized estimators such as Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001) and many other variants of them.

In the Bayesian setting, popular approaches include using the spike and slab priors and continuous shrinkage priors for sparse mean estimation. Formulation of sparse mean vector scenarios as a combination of two regimes where the mean values are zero or arising from a measure which allows for possibly large values naturally leads to a mixture prior of the form

$$p(\mu) = p\delta_0 + (1-p)g(\mu).$$

The point mass p as $\mu = 0$ is the *spike* and the probability density $g(\cdot)$ allowing μ to take possibly large non-zero values is the *slab*. Mitchell and Beauchamp (1988) considered it in the context of variable selection in Gaussian regression. Since then such priors have gained popularity in many contexts including variable selection, covariance matrix estimation, false discovery rate estimation. Many authors have advocated the use of such point mixture priors for normal mean estimation. Strawderman-Berger (SB) prior (Strawderman and Berger, 1996) explicitly considered in this article is an example of such a spike-and-slab prior in a hierarchical setting where the hyper-parameters governing the slab $g(\cdot)$ are allowed to change according to some prior for each μ_i. Specifically, they propose the following model

$$\begin{aligned} \mu_i|\tau, \lambda_i &\sim N(0, \tau^2\lambda_i^2), \\ p(\lambda_i) &\propto \lambda_i(1+\lambda_i^2)^{1/2}, \\ p(\tau) &\sim C[\sigma, \sigma]I(\tau > \sigma) \end{aligned}$$

where $C[a, b]$ is the Cauchy density with location and scale equal to a and b, respectively.

A version of the spike-slab prior considered recently is the non-local prior recommended by Johnson and Rossell (2010, 2012) where the slab is well separated from the spike at zero. Being a single component prior, horseshoe type priors are computationally less demanding than the spike-slab priors.

Another class of priors considered for sparse estimation of mean are the shrinkage priors or the global-local priors. Park and Casella (2008) proposed a scale mixture of Gaussian prior that they called the *Bayesian Lasso*. However, these priors do not have sufficient prior mass near zero to work well in the very sparse regime. Carvalho *et al.* (2010) proposed the *horseshoe* (HS) prior defined as

$$\begin{aligned} \mu_i|\tau, \lambda_i &\sim N(0, \tau^2\lambda_i^2), \\ p(\lambda_i) &\propto C[0, 1]_+, \\ p(\tau) &\sim \sigma C[0, 1]_+ \end{aligned}$$

where $C[0, 1]_+$ is the half-Cauchy density, the standard Cauchy truncated to the positive half. The horseshoe prior has only one component as opposed to the two separate com-

ponents of the spike-and-slab priors but overcomes the deficiency of the Bayesian Lasso in sparse regime by allowing infinite prior density at zero.

While full Bayesian analysis is possible, empirical Bayes solutions have also been discussed for the two component mixture priors such as Strawderman-Berger and the single component shrinkage priors such as horseshoe. Empirical Bayes solutions for high-dimensional sparse mean estimation have been also looked at in the literature; see Johnston and Silverman (2004), Brown and Greenshtein (2009).

Often one has prior knowledge on the range of possible values for the mean parameter, such as the parameter is non-negative. One way of estimating such a parameter is to first obtain an unrestricted estimate of the parameter and then truncate it so that the estimate lies in the constrained parameter space. Intuitively, the performance of the estimator is expected to be much better if such constraint conditions are incorporated in the model. Constrained estimation of normal mean restricted to convex cones has been discussed in Sen and Silvapulle (2001). Danaher *et al.* (2012) provides an example of Bayesian estimation of normal mean when the mean is constrained to a convex polytope.

In this paper we particularly look at the case when the dimension is large and the mean vector is assumed to be sparse. We focus on the high dimensional normal means estimation problem where the mean vector is constrained to be in a closed convex polyhedral cone. Let $\boldsymbol{y} = (y_1, \ldots, y_n)' \sim N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$ where the parameter of interest $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ is assumed to belong to the convex cone

$$\mathcal{K} = \{\boldsymbol{\mu} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{\mu} \geq 0\}$$

where $\boldsymbol{A}$ is some fixed $r \times n$ matrix. We assume that $\mathcal{K}$ has non-zero interior volume with respect to the n dimensional Lebesgue measure. Of course, one of the most interesting question is how to specify sparsity in constrained spaces such as $\mathcal{K}$. However, the scope of this paper is very limited. Without getting into a discourse about sparsity in constrained sets such as $\mathcal{K}$, we simply compare the performance of sparsity generating spike-and-slab priors such as Strawderman-Berger and shrinkage priors such as horseshoe, when the priors are defined in terms of scale mixtures of truncated normal instead of normal. This straightforward generalization is probably not optimal, particularly if the conic geometry is very different from that of the entire space. However, given its special importance in the applications, we will only consider $\mathcal{K}$ to be the positive orthant, $\mathbb{R}^d_+$. The geometry of the positive orthant is very similar to the unrestricted linear space, but there are subtle differences in estimation due to the constraint and that is what we explore via numerical investigation.

In Section 2 we discuss the Bayes estimators for the Strawderman-Berger and the horseshoe priors when they are extended to the convex cone case. In Section 3 we present results of a numerical experiment comparing the performance of posterior quantities obtained using different priors along with that of the maximum likelihood estimator (MLE) projected to the convex cone. We end with some discussions in Section 4.

2. Sparse Priors for the Non-negative Orthant

In this paper we consider the restriction $\mu_i \geq 0$ for all i and hence $\boldsymbol{\mu} \in \mathcal{K} = \mathbb{R}^n_+$. For the horseshoe prior and the Strawderman-Berger prior for the non-negative orthant, we simply replace the normal prior for μ_i with normal truncated to the positive half. To judge

the performance of the estimators under different priors in the constrained case, we set forth a list of desirable properties. These are analogous to desirable properties in a sparse mean estimator in the unrestricted case, except adapted to the constrained mean case. For example, one would want the estimators for μ_i to provide considerable shrinkage for small to moderate y_i whereas to leave y_i nearly unperturbed for large positive y_i. In the constrained case, for negative y_i one would expect the estimated mean to be nearly zero, if not exactly zero. The maximum likelihood estimator for μ_i is exactly zero whenever y_i is negative.

Horseshoe prior

The extension for the horseshoe to the positive orthant considered here is then

$$\begin{aligned} \mu_i|\tau,\lambda_i &\sim N(0,\tau^2\lambda_i^2)_+, \\ \lambda_i &\sim C(0,1)_+. \end{aligned}$$

where $N(\theta,v)_+$ represent a $N(\theta,v)$ truncated from below at 0 and $C(0,1)_+$ represent a standard half-Cauchy distribution on the positive reals. We use a Jeffrey's prior on σ and standard half-Cauchy prior with scale equal to σ on τ.

$$\begin{aligned} \pi(\sigma) &\propto \frac{1}{\sigma}, \\ \tau|\sigma &\sim C(0,\sigma)_+. \end{aligned}$$

One could estimate σ and τ using an Empirical Bayes approach. However, here we use a full Bayesian framework. Carvalho *et al* (2010) described λ_i as the local shrinkage parameter and τ the global shrinkage parameter. For the positive orthant, the horseshoe prior that we are considering is essentially a scale mixture of truncated normals, scale being a function of a common variance component, τ and an individual variance component, λ_i for each μ_i.

Conditional on σ, τ and λ_i's, $\mu_i|\boldsymbol{y}$ are independently distributed as

$$\mu_i|\lambda_i,\tau,\sigma,\boldsymbol{y} \sim N(m_i,s_i^2)_+$$

where $m_i = s_i^2\frac{y_i}{\sigma^2}$ and $s_i^2 = [\frac{1}{\sigma^2}+\frac{1}{\tau^2\lambda_i^2}]^{-1}$. Then, we have

$$E(\mu_i|\lambda_i,\tau,\sigma,\boldsymbol{y}) = m_i + \frac{\phi(\frac{-m_i}{s_i})}{1-\Phi(\frac{-m_i}{s_i})}s_i. \tag{1}$$

The Bayes estimator of μ_i is given by

$$\hat{\mu}_i = E(\mu_i|\boldsymbol{y}) = E_{\lambda_i,\tau,\sigma|\boldsymbol{y}}E(\mu_i|\lambda_i,\tau,\sigma,\boldsymbol{y}).$$

From the bounds on the Mill's ratio for the standard normal, we know that for $t>0$,

$$t \;<\; \frac{\phi(t)}{1-\Phi(t)} \;<\; \frac{1+t^2}{t}. \tag{2}$$

This implies $E(\mu_i|\lambda_i,\tau,\sigma,\boldsymbol{y}) > 0$ for all y. Also, for $y_i<0$, $E(\mu_i|\lambda_i,\tau,\sigma,\boldsymbol{y}) < \sigma^2|y_i|^{-1}$.

Moreover, for large positive y_i, $E(\mu_i|\lambda_i,\tau,\sigma,\boldsymbol{y}) \approx [1-\frac{\tau^2\lambda_i^2}{\sigma^2+\tau^2\lambda_i^2}]y_i$. Hence, summarizing we have

Result 1: For the horseshoe prior for the constrained case when the true mean is restricted to the non-negative orthant:

1. $E(\mu_i|\boldsymbol{y}) > 0$ for all $\boldsymbol{y}$.
2. For $y_i < 0$, $E(\mu_i|\boldsymbol{y}) = O(|y_i|^{-1})$.
3. For large positive y_i, $E(\mu_i|\boldsymbol{y}) \approx \mathrm{E}([1-\frac{\tau^2\lambda_i^2}{\sigma^2+\tau^2\lambda_i^2}]|\lambda_i,\tau,\sigma,\boldsymbol{y})y_i$.

Thus, the posterior mean of μ_i acts as a shrinkage estimator and its behavior is similar to what observed in the unrestricted case.

Strawderman-Berger Prior

The extension of Strawderman-Berger prior for the non-negative orthant puts a truncated normal distribution in place of the usual normal distribution.

$$\begin{aligned}\pi(\mu_i) &= p\delta_o + (1-p)\ N(0,\tau^2\lambda_i^2)_+,\\ \pi(\lambda_i) &\propto \lambda_i(1+\lambda_i^2)^{\frac{3}{2}},\\ p &\sim \mathrm{Unif}(0,1).\end{aligned}$$

Similar to horseshoe, we use a Jeffrey's prior on σ and for τ and a truncated Cauchy prior with location and scale both equal to σ bounded below at σ, which are

$$\begin{aligned}\tau|\sigma &\sim C(\sigma,\sigma)\ 1(\tau\geq\sigma),\\ \pi(\sigma) &\propto \frac{1}{\sigma}.\end{aligned}$$

Conditional on λ_i,τ,p,σ, the posterior distribution of μ_i is a mixture distribution

$$\pi(\mu_i|\lambda_i,\tau,p,\sigma,\boldsymbol{y}) = c(\theta_i,y_i)\ \delta_o + \Big(1-c(\theta_i,y_i)\Big)\ N(m_i,s_i^2)_+ \tag{3}$$

where

$$c(\theta_i,y_i) = \frac{\frac{p}{\sigma}\phi(\frac{y_i}{\sigma})}{\frac{p}{\sigma}\phi(\frac{y_i}{\sigma}) + \frac{2(1-p)}{l_i}\phi(\frac{y_i}{l_i})\Phi(\frac{m_i}{s_i})}$$

is the posterior probability of $\mu_i = 0$ which acts as local shrinkage, $\theta_i = \{\lambda_i,\tau,\sigma,p\}$ and $l_i^2 = \sigma^2+\lambda_i^2\tau^2$ for $i = 1,\ldots,n$.
Then, we have

$$E(\mu_i|\boldsymbol{\lambda},\tau,\sigma,p,\boldsymbol{y}) = \Big(1-c(\theta_i,y_i)\Big)\left(m_i + \frac{\phi\Big(\frac{-m_i}{s_i}\Big)}{\Phi\Big(\frac{m_i}{s_i}\Big)}s_i\right).$$

The Bayes estimator for μ_i is the posterior mean, $E(\mu_i|\boldsymbol{y}) = E_{\boldsymbol{\lambda},\tau,p|\boldsymbol{y}}E(\mu_i|\boldsymbol{\lambda},\tau,\sigma,p,\boldsymbol{y})$.

Result 2: The following results hold for the posterior mean computed based on the Strawderman-Berger prior in the constrained case:

1. $\hat{\mu}_i > 0$ since $E(\mu_i|\lambda_i, \tau, p, \sigma, \boldsymbol{y}) > 0$ using the inequality in (2).
2. $\hat{\mu}_i$ is non-decreasing in y_i.
3. For large positive y_i, $E(\mu_i|\boldsymbol{y}) \approx \mathrm{E}\Big[\Big(1 - c(\theta_i, y_i)\Big)\Big(1 - \frac{\tau^2\lambda_i^2}{\sigma^2+\tau^2\lambda_i^2}\Big)|\lambda_i, \tau, \sigma, \boldsymbol{y}\Big] y_i$.

See the appendix for a proof of the above Result 2.

In the two component model, the posterior mean could be computed in a manner similar to that computed for the horseshoe type prior. However, for the spike-and-slab type prior, it is more interesting to look at the component-wise posterior median. For the posterior median, we use the estimator,

$$\hat{\mu}_{iM}(\mu_i|\boldsymbol{y}) = F_i^{-1}(1/2)$$

where

$$F_i(t) = E_{\boldsymbol{\theta}|\boldsymbol{y}}[P(\mu_i \leq t|\theta_i, y_i)]$$

and

$$P(\mu_i \leq t|\theta_i, y_i) = c(\theta_i, y_i) + \Big(1 - c(\theta_i, y_i)\Big)\ \Phi^{-1}\left(\frac{m_i}{s_i}\right)\ \left[\Phi\left(\frac{t - m_i}{s_i}\right) - \Phi\Big(-\frac{m_i}{s_i}\Big)\right].$$

A more specific form of the posterior median is

$$\hat{\mu}_{iM}(\mu_i|\boldsymbol{y}) = \begin{cases} 0 & \text{if } E_{\boldsymbol{\theta}|\boldsymbol{y}}\Big(c(\theta_i, y_i)\Big) \geq 0.5, \\ \inf\{x \geq 0 : F_i(x) \geq 0.5\} & \text{if } E_{\boldsymbol{\theta}|\boldsymbol{y}}\Big(c(\theta_i, y_i)\Big) < 0.5. \end{cases} \tag{4}$$

Thus, for an additive loss

$$L(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = \sum |\mu_i - \hat{\mu}_i|,$$

it makes sense to look at the component-wise posterior median, $\hat{\boldsymbol{\mu}}_M$.

One could also look at the Empirical Bayes estimator of the median which is the expectation of the posterior median expression with respect to $p(\boldsymbol{y}|\boldsymbol{\theta})$. Let

$$G_i(t|\theta_i, y_i) = P(\mu_i \leq t|\theta_i, y_i).$$

Then the expression for the median is

$$\tilde{\mu}_i(\theta_i, y_i) = 1[c(\theta_i, y_i) \leq 0.5]\ G^{-1}_{\theta_i, y_i}\left(\frac{\frac{1}{2} - c(\theta_i, y_i)}{1 - c(\theta_i, y_i)}\right). \tag{5}$$

One could show that the posterior median defined in (5) is a continuous shrinkage soft thresholding rule.

Result 3: For the Strawderman-Berger prior for the normal mean when the true mean is constrained to the non-negative orthant, the component-wise posterior median in (5) satisfies the following properties for a given value of the hyperparameter $\boldsymbol{\theta}$.

1. The posterior $p(\mu_i|\theta_i, y_i)$ is stochastically increasing in y_i and hence the posterior median of μ_i is a monotonically increasing in y_i for each value of the hyperparameter.

2. For each y_i, there exists $T(\theta_i)$ such that $\tilde{\mu}_i(\theta_i, y_i) = 0$ iff $y_i < T(\theta_i)$.

See the appendix for a proof of the above result.

3. Posterior Computation and Numerical Results

Let $\boldsymbol{\theta}$ denote the set of all hyper-parameters. We use $\boldsymbol{\theta}$ interchangeably for horseshoe and Strawderman-Berger prior where $\boldsymbol{\theta} = \{\boldsymbol{\lambda}, \tau, \sigma\}$ for the former and $\boldsymbol{\theta} = \{\boldsymbol{\lambda}, \tau, p, \sigma\}$ for the later. We use a Metropolis within Gibbs algorithm to generate random samples from the marginal posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and thus compute posterior summaries for the posterior $\pi(\mu_i|\boldsymbol{y})$ by averaging the value of the hyperparameters over the randomly generated sample of $\boldsymbol{\theta}$. For the posterior mean $E(\mu_i|\boldsymbol{y})$ we use the estimator

$$E(\mu_i|\boldsymbol{y}) = L^{-1} \sum_{l=1}^{L} \mathrm{E}(\mu_i|\boldsymbol{\theta}_l, \boldsymbol{y}),$$

where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L$ are samples from $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. For the posterior median, $Med(\mu_i|\boldsymbol{y})$, we use the estimator

$$Med(\mu_i|\boldsymbol{y}) = \hat{F}_i^{-1}(1/2),$$

where $\hat{F}_i(t) = L^{-1} \sum_{l=1}^{L} P(\mu_i \leq t|\boldsymbol{y}, \boldsymbol{\theta}_l)$.

The conditional marginal of $\boldsymbol{y}$ can be factorized as

$$\pi(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi(y_i|\lambda_i, \tau, \sigma, p),$$

where $\pi(y_i|\lambda_i, \tau, \sigma) = \frac{1}{l_i}\phi(y_i/l_i)\ \Phi(m_i/s_i)$ for the horseshoe prior and $\pi(y_i|\lambda_i, \tau, p, \sigma) = \frac{p}{\sigma}\phi(y_i/\sigma) + \frac{2(1-p)}{l_i}\ \phi(y_i/l_i)\ \Phi\left(\frac{m_i}{s_i}\right)$ for the Strawderman-Berger prior. The distribution of y_i conditional on the hyperparameters is Skew-Normal for the horseshoe prior and a mixture distribution of Normals for $\mu_i = 0$ and Skew-Normal for $\mu_i > 0$.

Hence the for the Gibbs sampling algorithm, the full conditionals are

1. $\pi(\lambda_i|\tau, \sigma, \boldsymbol{y}) \propto \pi(y_i|\lambda_i, \tau, \sigma)\ \pi(\lambda_i), \quad i = 1, \ldots, n$
2. $\pi(\tau, \sigma|\boldsymbol{\lambda}, \boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\lambda}, \tau, \sigma)\ \pi(\tau|\sigma)\ \pi(\sigma)$.

For the Strawderman-Berger prior we have in addition,

1. $\pi(p|\boldsymbol{\lambda}, \tau, \sigma, \boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\lambda}, \tau, \sigma, p)\ \pi(p)$.

The one-dimensional conditionals can be sampled using a standard Metropolis step.

3.1. Simulation results

We compare the performances of the Strawderman-Berger estimators, horseshoe estimator and Maximum Likelihood Estimator (MLE) under different degrees of sparsity.

The MLE when $\boldsymbol{\mu} \in \mathcal{K} = \mathbb{R}^n_+$ for $\Sigma = \sigma^2 I$ is simply the projection of $\boldsymbol{y}$ onto the non-negative orthant i.e. $\hat{\mu}_i = \max(y_i, 0)$. For a general polyhedral cones, with Σ other than $\sigma^2 I$, the MLE is not straightforward to compute.

We analyze the risk properties of the estimators when the mean vector is simulated under strongly sparse signals and weakly sparse signals. For each of the sparsity level, we further consider two scenarios described below.

Strong sparsity: We use a discrete mixture model to generate exact zero entries for the mean vector using the model below:

$$\begin{aligned} y_i|\mu_i, \sigma^2 &\sim N(\mu_i, \sigma^2), \\ \pi(\mu_i) &= p\delta_o + (1-p)\ G(\alpha, \beta), \end{aligned}$$

where α is taken to be 5, β is 0.5 and 80% of the mean vector has exact zero entries. The major concentration of μ_i's is at 0 with an average concentration of $\mu_i > 0$ at 10 with variance 20. Two possible values of σ are considered: $\sigma = 1$ and $\sigma = 3$. The separation between y_i's at $\mu_i = 0$ and $\mu_i > 0$ is more prominent for $\sigma = 1$ than $\sigma = 3$.

Weak sparsity: For weakly sparse signals, we generate μ_i which decays according to the power law but none of its components are exactly zero. For this, we consider

$$\begin{aligned} y_i|\mu_i, \sigma^2 &\sim N(\mu_i, \sigma^2), \\ \mu_i|\eta, \alpha &\sim \text{Unif}(0, \eta c_i), \\ \eta &\sim \text{Ex}(2), \\ \alpha &\sim \text{Unif}(a, b), \end{aligned}$$

where $c_i = (n/i)^{1/\alpha}$ for $i = 1, \ldots, n$. For simulation purposes, $\sigma = 1$ is chosen and two possible scenarios of $\alpha \sim \text{Unif}(a, b)$ are considered: $a = 0.5, b = 1$ and $a = 1, b = 2$. The first scenario yields relatively large mean entries than the second scenario depending on the randomly generated values of η and α. When $\alpha \sim \text{Unif}(1, 2)$, one can expect the concentration around 0 to be more dense than when $\alpha \sim \text{Unif}(0.5, 1)$ depending on the speed of decay, α.

For each of the scenarios, we simulate 1000 data sets from the corresponding model of dimension $n = 300$ using MCMC with 50000 runs and a burn-in period of 10000. The convergence is assessed using the standard MCMC diagnostic checks and all chains seem to converge. We report the median risk under squared error loss and absolute error loss along with the average risk ratios between the estimators in Table 1 and Table 2.

Figure 1 shows the plots for MLE estimates, posterior mean under horseshoe prior and posterior mean and posterior median under Strawderman-Berger prior for a single realization generated under strongly sparse signals with the variance set to $\sigma = 1$ or $\sigma = 3$. The dimension of the mean vector is 300. Figure 2 presents the same under weakly sparse signals for the two scenarios when $\alpha \sim \text{Unif}(0.5, 1)$ and $\alpha \sim \text{Unif}(1, 2)$.

From Figure 1 and Figure 2, we see that the posterior mean for the horseshoe provides shrinkage near zero, but it is still significantly positive even when the realized y is considerably negative. This is particularly undesirable in the constrained case when the

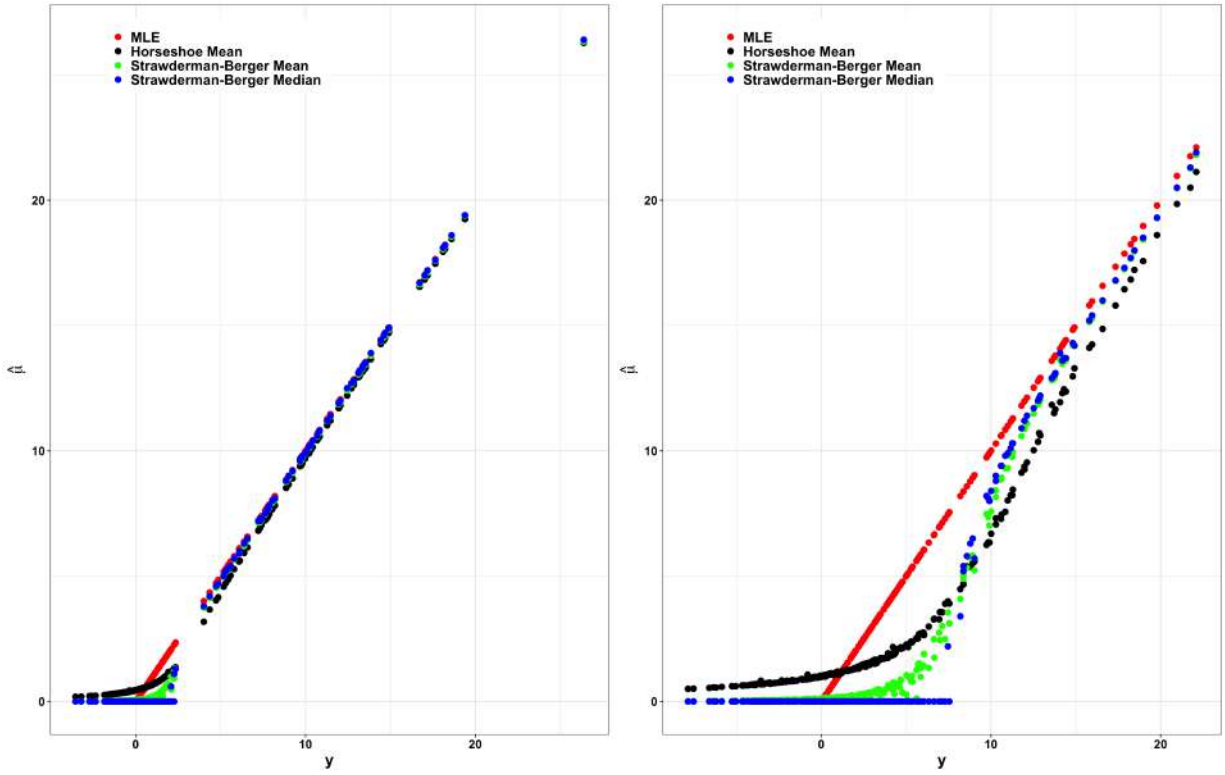


Figure 1: Plots of $\hat{\mu}$ versus y under strong sparsity with $\sigma = 1$ (left) and $\sigma = 3$ (right)

true mean is known to be non-negative. From Result 1, we know that for negative y, the horseshoe estimator decays as $O(|y|^{-1})$. This induces considerable bias. The posterior mean under the Strawderman-Berger prior shrinks more than the horseshoe posterior mean estimator. However, for large positive y the horseshoe estimator seems to perform better, and shrinks less than the posterior mean under the Strawderman-Berger prior.

The posterior median for Strawderman-Berger prior, as expected from the results in Result 3, provides a soft thresholding estimator that is truncated to zero below the truncation point $T(y, p)$ and provides continuous shrinkage for y above the truncation point. All estimators are monotonic in y and the shrinkage factor tends to one as y tends to infinity, thereby satisfying the requirement to not perturbing the big realized values of y.

Table 1 shows that the risk performance of Strawderman-Berger posterior median and posterior mean is better than the MLE and horseshoe posterior mean both in terms of squared error loss and absolute loss for the strong sparsity case. In particular, the horseshoe posterior mean has at least 50% more risk than both the Strawderman-Berger posterior mean and posterior median. However, the risk for horseshoe posterior mean under squared error loss is $20\% - 35\%$ less than the Strawderman-Berger estimators when $\sigma = 3$.

From Table 2, we see that the risk of horseshoe posterior mean is consistently less than that of MLE and Strawderman-Berger posterior mean and posterior median. Specifically, horseshoe posterior mean has of $6\% - 40\%$ more risk than the Strawderman-

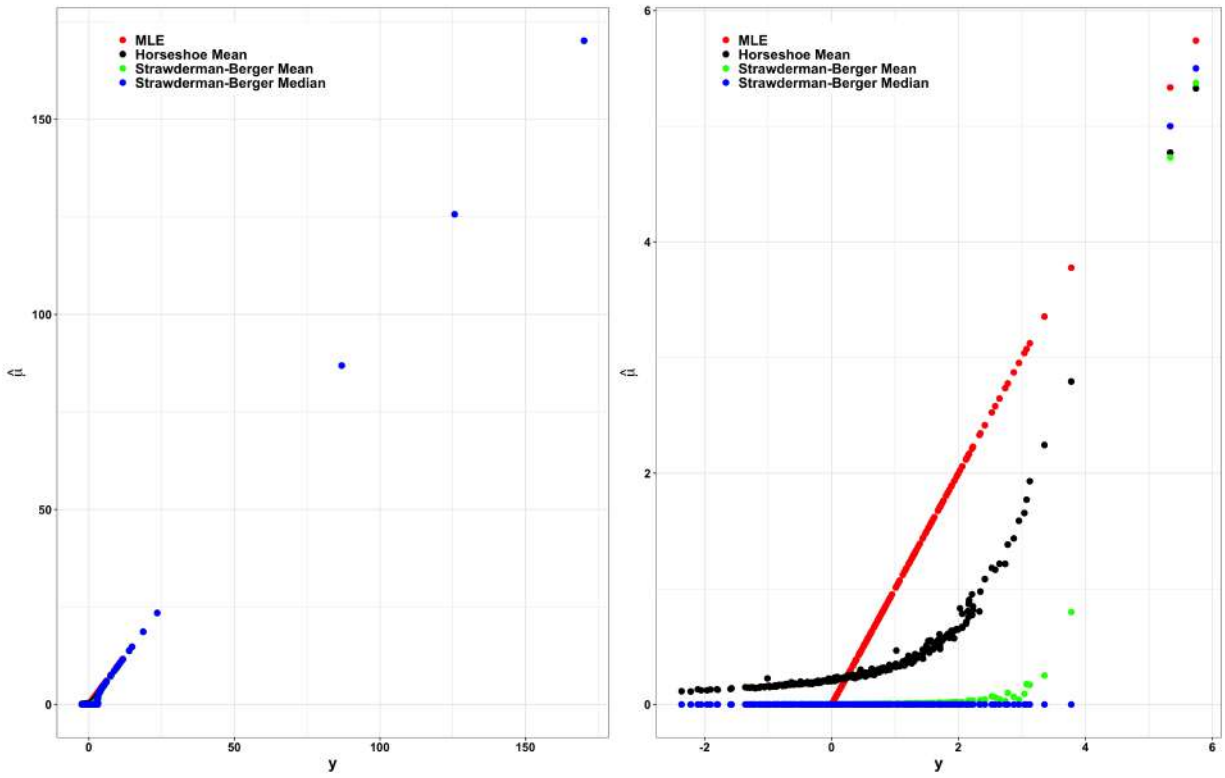


Figure 2: Plots of $\hat{\mu}$ versus y under weak sparsity where $\alpha \sim U(0.5, 1)$ (left) and $\alpha \sim U(1, 2)$ (right)

Berger estimators. However, when $\alpha \sim U(1, 2)$, horseshoe estimator has 63% more risk than the Strawderman-Berger posterior mean and approximately 411% more risk than SB posterior median, although the median squared error risk is less for horseshoe than the other estimators.

4. Real Data Analysis

We studied the performance of the estimators using the childhood acute lymphoblastic leukemia (ALL) data set (GSE412) which includes gene expression information for 110 childhood acute lymphoblastic leukemia samples before and after treatment. From the originally measured 12625 probe sets, genes that were not present in at least one sample were removed to obtain 8280 genes. After cleaning the data, we selected 250 genes for 50 pediatric newly diagnosed children for our analysis. Our goal is to estimate the standardized difference between post-treatment mean, $\boldsymbol{\theta_2}$ and pre-treatment mean, $\boldsymbol{\theta_1}$ regardless of the type of treatment used i.e. $\boldsymbol{\mu} = \frac{\boldsymbol{\theta_2} - \boldsymbol{\theta_1}}{\sigma}$. For illustration purposes, we assume up-regulation of gene expression level in ALL cells so that $\boldsymbol{\mu} \in \mathbb{R}^n_+$. We further assumed that the gene expression levels are uncorrelated and have same variance. The observed data is the standardized difference of the average post-therapy and pre-therapy gene expression.

Table 1: Risk under squared error loss and absolute error loss for strongly sparse signals in two scenarios: $\sigma = 1$ and $\sigma = 3$. The diagonal components are median sum of squared error and absolute error. The off diagonal components are average error ratios of estimator in row by estimator in column.

		$\sigma = 1$				$\sigma = 3$			
		MLE	HS	SB Mean	SB Median	MLE	HS	SB Mean	SB Median
Square Error Loss	MLE	171	1.39	2.23	2.33	1598	1.19	0.98	0.77
	HS		131	1.6	1.67		1361	0.81	0.64
	SB Mean			82	1.04			1636	0.78
	SB Median				78				2129
Absolute Error Loss	MLE	143	0.92	1.97	2.6	428	0.95	1.43	1.42
	HS		156	2.13	2.8		452	1.5	1.49
	SB Mean			73	1.32			295	0.98
	SB Median				56				299

Table 2: Risk under squared error loss and absolute error loss for weakly sparse signals in two scenarios: $\alpha \sim U(0.5, 1)$ and $\alpha \sim U(1, 2)$. The diagonal components are median sum of squared error and absolute error. The off diagonal components are average error ratios of estimator in row by estimator in column.

		$\alpha \sim U(0.5, 1)$				$\alpha \sim U(1, 2)$			
		MLE	HS	SB Mean	SB Median	MLE	HS	SB Mean	SB Median
Square Error Loss	MLE	200.24	2.68	2.91	2.65	179	15.7	128.67	400
	HS		122.52	0.73	0.6		63	1.63	5.11
	SB Mean			185.7	0.81			128	0.92
	SB Median				235.8				136
Absolute Error Loss	MLE	181.8	1.6	1.75	1.63	166	3.09	4.32	4.8
	HS		136.39	0.94	0.85		86	0.91	0.93
	SB Mean			162.67	0.89			128	0.95
	SB Median				186.75				134

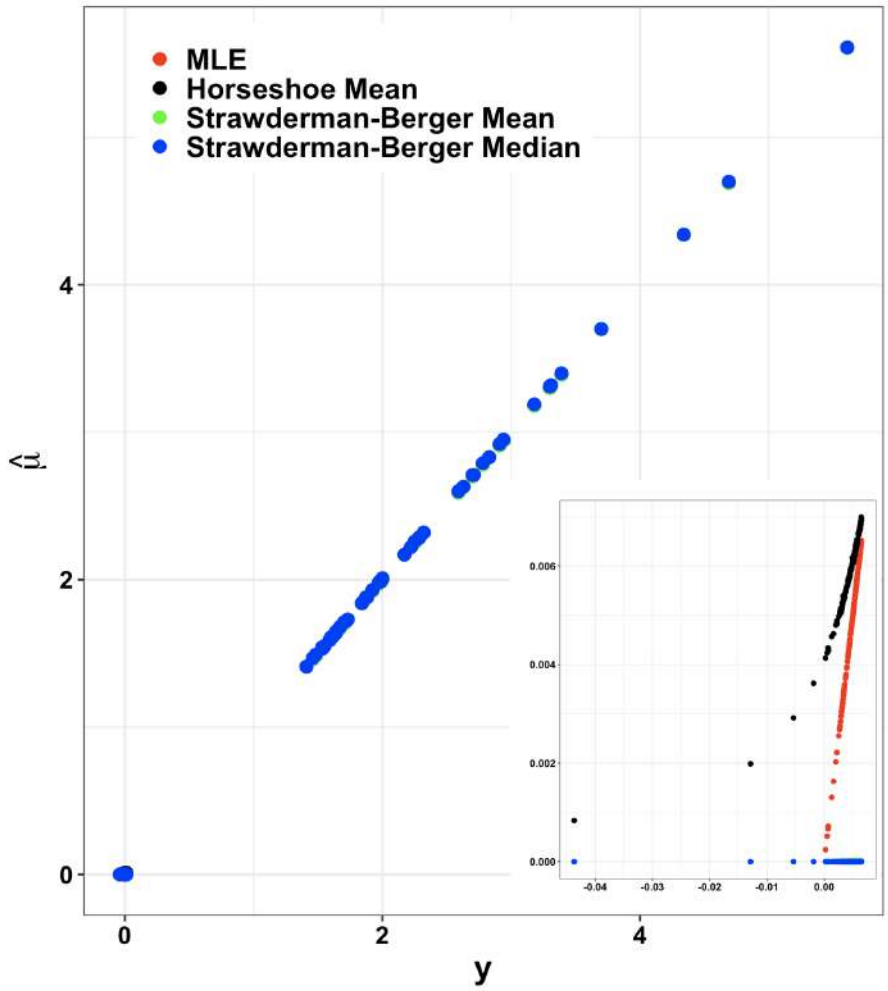


Figure 3: Plot of $\hat{\mu}$ versus y for the standardized difference in post-treatment and pre-treatment gene expression for 250 genes

The estimated MLE, posterior means and posterior median are shown in Figure 3. The summary results for the observed data and the estimates are presented in Table 3. We noticed that Horseshoe posterior mean is always positive whereas the MLE is 0 for negative y's. While the Strawderman-Berger posterior median is exactly 0 for $y < 1.4$, the SB posterior mean is close to 0 for these values of y. All four estimates perform similarly for larger values of y.

5. Discussion

In our simulation studies, we compared the performance of horseshoe posterior mean, Strawderman-Berger posterior mean and posterior median for strongly sparse signals and weakly sparse signals. While the posterior mean for both horseshoe and Strawderman-Berger prior are shrinkage estimators, MLE and Strawderman-Berger posterior median are truncation based estimators with exact zeros for small signals. When the true sparsity regime is strong sparsity, then truncation type estimators maybe preferred. The non-negative constraint does impact the relative performance of the mean and median estimators. It can be shown that the posterior mean under priors considered here are smooth differentiable functions of the observed value. Hence it cannot be

Table 3: Summary statistics for the estimates under Horseshoe (HS) and Strawderman-Berger prior

	$\boldsymbol{y}$	MLE	HS Posterior Mean	SB Posterior Mean	SB Posterior Median
Minimum	-0.0437	0.0000	0.0008	0.0000	0.0000
1st Quartile	0.0043	0.0043	0.0058	0.0000	0.0000
Median	0.0054	0.0054	0.0063	0.0000	0.0000
Mean	0.4618	0.4620	0.4630	0.4583	0.4592
3rd Quartile	0.0064	0.0064	0.0069	0.0000	0.0000
Maximum	5.6094	5.6094	5.6093	5.6094	5.6100

expected to capture the threshold like behavior present in the strongly sparse regime.

In this paper, the numerical studies for non-negative orthant is restricted to horseshoe prior and Strawderman-Berger prior. It would be interesting to consider other scale mixture distributions, similar to Bayesian lasso, with hard thresholding properties for non-negative mean vectors. Another interesting domain is the discrete mixture models where the mixing kernel for the positive means could be chosen in a more flexible manner, belonging to flexible families on the non-negative orthant, e.g. product of gamma densities where heavy tailed priors are used for the hyperparamters. While the scope of this paper is limited to non-negative orthant which has many popular applications, one can think of exploring some of these priors to a general closed convex polyhedral cones. Moreover, the observations maybe allowed to be correlated with a known low-dimensional correlation structure. For example, one could use the model $\boldsymbol{y}|\boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma})$ where the positive definite matrix $\boldsymbol{\Sigma}$ is completely known. The mean vector is again assumed to be in the non-negative orthant. A standard approach to dealing with general $\boldsymbol{\Sigma}$ matrix is to transform the observations to $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}$ so that the problem reduced to the case considered here. However, the transformed mean $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ need not remain in the positive orthant unless $\boldsymbol{\Sigma}$ is an M-matrix with an inverse that admits a positive square-root. Thus, for general $\boldsymbol{\Sigma}$ further investigation is required.

References

Berger, J. O. and Strawderman, W. E. (1996). Choice of hierarchical priors: admissibility in estimation of normal means. *Annals of Statistics*, **24(3)**, 931–951.

Brown, L. D. and Greenshtein, E. (2009). Non-parametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *Annals of Statistics*, **37(4)**, 1685–1704.

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, **97(2)**, 465–480.

Danaher, M. R., Roy, A., Chen, Z., Mumford, S. L. and Schisterman., E. F. (2012). Minkowski–Weyl Priors for models with parameter constraints: An analysis of the BioCycle study. *Journal of the American Statistical Association*, **107(500)**, 1395–1409.

Dharmadhikari, S. and Kumar, J. (1988). Choice of hierarchical priors: admissibility in estimation of normal means. *Unimodality, Convexity, and Applications*, Elsevier Science.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81(3)**, 425–455.

Fan, J. and Li, R. (2001). Variable Selection via non-concave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96(456)**, 1348–1360.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 361-379, University California Press, Berkeley.

Johnson, V. E. and Rossell, D. (2010). On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society Series B*, **72(2)**, 143–170.

Johnson, V. E. and Rossell, D. (2012). Bayesian model selection in high- dimensional settings. *Journal of the American Statistical Association*, **107(498)**.

Johnstone, I. M. and Silverman, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Annals of Statistics*, **32(4)**, 1594–1649.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, **83(404)**, 1023–1032.

Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, **103(482)**, 681–686.

Silvapulle, M. J. and Sen, P. K. (2011). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. John Wiley and Sons.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B*, **58(1)**, 267–288.

https : *//www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc* = *GPL*8300

APPENDIX

Here we prove the claims in Results 2 and 3 about the posterior mean and median of μ_i obtained using the Strawderman-Berger prior.

Result 2:

2. *$\hat{\mu}_i$ is non-decreasing in y_i.*

For notational simplicity, we denote μ_i by μ and y_i by y. Without loss of generality, let us assume $\sigma = 1$.

$$y|\mu \sim N(\mu, 1), \quad \mu \sim g(\mu)$$

$$g(\mu) = \pi\delta_o + (1-\pi)g_1(\mu)$$

An estimator of μ is then given by

$$\begin{aligned} l(y) = E(\mu|y) &= \frac{\int \mu\phi(y-\mu)g(\mu)}{\int \phi(y-\mu)g(\mu)} \\ &= \frac{(1-\pi)\int \mu\phi(y-\mu)g_1(\mu)d\mu}{\pi\phi(y)+(1-\pi)\int \phi(y-\mu)g_1(\mu)d\mu} \\ &= \frac{\int \mu\phi(\mu)g_1(\mu)e^{\mu y}d\mu}{\frac{\pi}{1-\pi}+\int \phi(\mu)g_1(\mu)e^{\mu y}d\mu} \\ &= \frac{a(y)}{b(y)}, \end{aligned}$$

where $a(y) = \int \mu\ \phi(\mu)\ g_1(\mu)\ e^{\mu y}\ d\mu$ and $b(y) = \frac{\pi}{1-\pi} + \int \phi(\mu)\ g_1(\mu)\ e^{\mu y}\ d\mu$.
Then $a'(y) = \int \mu^2\ \phi(\mu)\ g_1(\mu)\ e^{\mu y}\ d\mu$ and $b'(y) = a(y)$

$$\begin{aligned} l'(y) &= \frac{b(y)a'(y) - a(y)b'(y)}{b^2(y)} \\ &= \frac{(\frac{\pi}{1-\pi} + \int \phi(\mu)\ g_1(\mu)\ e^{\mu y}\ d\mu)(\int \mu^2\ \phi(\mu)\ g_1(\mu)\ e^{\mu y}\ d\mu) - (\int \mu\ \phi(\mu)\ g_1(\mu)\ e^{\mu y}\ d\mu)^2}{b^2(y)} \\ &= \frac{\frac{\pi}{1-\pi}\int \mu^2 f^*(\mu)\ d\mu + q(y)\int \mu^2 f^*(\mu)\ d\mu) - (\int \mu f^*(\mu)\ d\mu)^2}{\left(\frac{\pi}{1-\pi} + q(y)\right)^2}, \end{aligned}$$

where $f^*(\mu) = \phi(\mu)\ g_1(\mu)\ e^{\mu y}$ and $q(y) = \int f^*(\mu)d\mu$.
Therefore $l'(y)$ reduces to

$$\begin{aligned} l'(y) &= \frac{\frac{\pi}{1-\pi}\frac{1}{q(y)}\int \mu^2\ \frac{f^*(\mu)}{q(y)}\ d\mu + \int \mu^2\ \frac{f^*(\mu)}{q(y)}\ d\mu) - (\int \mu\ \frac{f^*(\mu)}{q(y)}\ d\mu)^2}{\left(\frac{\pi}{1-\pi}\frac{1}{q(y)} + 1\right)^2} \\ &= \frac{\frac{\pi}{(1-\pi)q(y)}E(\mu^2) + V(\mu)}{\left(\frac{\pi}{1-\pi}\frac{1}{q(y)} + 1\right)^2} \ \geq 0\ \forall y. \end{aligned}$$

Hence $l(y)$ is non-decreasing function of y for any $g_1(\mu)$ defined on positive μ.

Result 3:

1. *The posterior $p(\mu_i|\theta_i, y_i)$ is stochastically increasing in y_i and hence the posterior median of μ_i is a monotonically increasing in y_i for a given value of the hyperparameter $\boldsymbol{\theta}$.*

For notational simplicity, we denote μ_i by μ and y_i by y. Without loss of generality, let us assume $\sigma = \tau = \lambda_i = 1$ for $i = 1, \ldots, n$. From the expression for the posterior of μ, we have,

$$\pi(\mu|y) = c(p, y)\delta_o(\mu) + (1 - c(p, y))f(\mu|y),$$

where

$$\begin{aligned} f(\mu|y) &= h(y)^{-1}e^{\mu y}e^{-\frac{1}{2}\mu^2}g(\mu), \\ h(y) &= \int_0^\infty e^{\mu y}e^{-\frac{1}{2}\mu^2}g(\mu)d\mu, \\ c(p,y) &= [1+\frac{(1-p)}{p}h(y)]^{-1}. \end{aligned}$$

To show that $\pi(\mu|y)$ is stochastically increasing (SI) in y, it is enough to show $f(\mu|y)$ is SI in y since $c(p,y)$ decreases with decrease in y.

Let $\mu_1 < \mu_2$ and $y_1 < y_2$. Then,

$$\frac{f(\mu_1|y_1)f(\mu_2|y_2)}{f(\mu_2|y_1)f(\mu_1|y_2)} = e^{(\mu_2-\mu_1)(y_2-y_1)} \geq 1.$$

Thus,

$$f(\mu_1|y_1)f(\mu_2|y_2) \geq f(\mu_1|y_2)f(\mu_2|y_1).$$

Multiplying both sides by $\pi(y_1)\pi(y_2)$ where $\pi(y)$ is the marginal of y, we have,

$$f(\mu_1,y_1)f(\mu_2,y_2) \geq f(\mu_1,y_2)f(\mu_2,y_1).$$

Hence, $f(\mu,y)$ is Totally Positive of order 2 (TP_2). Hence, μ and y are SI in each other (Theorem 6.1, Dharmadhikari and Joag-Dev 1988).

2. *For each y_i, these exists $T(\theta_i)$ such that $\tilde{\mu}_i(\theta_i,y_i) = 0$ iff $y_i < T(\theta_i)$.*

Since $c(\theta_i,y_i)$ is monotonically decreasing in y_i and

$$\begin{aligned} \lim_{y_i\to-\infty} c(\theta_i,y_i) &= 1, \\ \lim_{y_i\to\infty} c(\theta_i,y_i) &= 0. \end{aligned}$$

For each θ_i, $\exists\ \ \tilde{\mu}_i(\theta_i,y_i) = 0 \iff y_i < T(\theta_i)$.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 333–351

Finite Mixture Modelling of Survival Data: With Applications to Pensioner Lifetime

Jyrki Möttönen[1], Janne Salonen[2], Tapio Nummi[3] and Timothy E. O'Brien[4]
[1]*Department of Mathematics and Statistics, University of Helsinki, P.O.Box 68, FI-00014 University of Helsinki, Finland*
[2]*Finnish Centre for Pensions, 00065 Eläketurvakeskus, Helsinki, Finland*
[3]*Faculty of Information Technology and Communication Sciences, Tampere University, Finland*
[4]*Loyola University Chicago, Department of Mathematics and Statistics, and Institute of Environmental Sustainability, USA*

Received: 01 April 2020; Revised: 09 September 2020; Accepted: 11 September 2020

Abstract

One of the most popular methods for modelling survival analysis data is the ubiquitous and time-honored proportional hazards model of Cox (1972). Popularity notwithstanding, in several cases, however, the proportional hazards assumption is found to be violated. Thus, important model extensions have been developed in the intervening years. One such extension is the so-called Frailty model (see for example Collett, 2015), which is based on the utilization of random effects. Another extension arises from the application of finite mixtures in the context of time-to-event data. Although finite mixture modelling tools are used in many fields of science, they have been less well-developed and used in survival analysis. Thus, a key aim of this article is to provide an interesting application of mixture modeling in survival analysis and to discuss aspects arising in its application. In this paper, we apply these techniques using real data from the research register of the Finnish Centre for Pensions; using pension insurance mortality data, we use the basic Cox proportional hazards model by incorporating finite mixture modelling techniques. Additional comparisons with frailty models are also provided.

Key words: Censoring; Cox model; Disability pension; EM algorithm; Finite mixtures; Mixture models; Mortality.

1. Introduction

Survival analysis (SA) techniques include a set of methods for analyzing time until the occurrence of a pre-specified event of interest such as mortality. In SA, subjects are usually followed over a pre-specified period of time. As such, an event can be, for example, death, the occurrence of a disease or the end of working life. SA can also be used to handle incomplete information. This is called censoring. Observations are censored when the information

Correponding Author: Tapio Nummi
Email: tapio.nummi@tuni.fi

about their survival time is incomplete. A commonly encountered form of censoring is right censoring, which means that the event of interest did not occur during the chosen follow-up period. When studying lifetime data, SA techniques are often considered superior to normal linear regression. SA is applied in many fields of science, including economics (Heckman and Singer, 1985; van den Berg, 2001), medicine and health (Machin *et al.*, 2006; Collett, 2015), amongst other fields.

One of the most popular method or model in SA is the Cox proportional hazards (CPH) model (Cox, 1972). In important cases, however, the required proportional hazards assumption is violated and this has led to key CPH model extensions. For example, in some cases the survival times among individuals are not independent. This situation may arise if individuals falling into specific groups (such as hospitals or clinics) tend to follow similar survival times. These types of model extensions are often addressed in Frailty models (see Collett, 2015). The basic idea in these models is to apply mixed modelling (with random effects) in a SA context. In this type of modelling, the source of correlation is assumed to be known. Frailty modelling includes two main approaches (see *e.g.*, Wienke, 2010). First, it is assumed that the event times include clusters, defined as shared frailties (random effects). Second, frailty can also be defined on an individual level, in which case the corresponding model is called the univariate frailty model. A comprehensive treatment of random effects in survival modelling is given in Ha *et al.* (2017) for example.

In another type of modelling, the observed survival times may be correlated, but the source of correlation cannot be directly measured. Basically, this kind of situation can be addressed using the theory and methods of finite mixtures (*e.g.*, McLachlan and Peel, 2000). One technique used in mixture longitudinal data analysis is trajectory analysis (*e.g.*, Nagin, 2005; Nagin and Odgers, 2010a), where the focus is on the analysis of a sequence of measurements. This technique has been widely applied in the social sciences (Nagin and Odgers, 2010b; Nummi *et al.*, 2017), but applications to survival data have not been that common. Thus, although finite mixture modelling tools are used in many fields of science (*e.g.*, Böhning *et al.*, 2007), they are not that much utilized with CPH model especially. Some related topics are covered in textbooks, like Ng *et al.* (2019) and McLachlan and Peel (2000), but there are surprisingly few published articles of the topic. Obviously, there are some practical obstacles in the combination of these two techniques and we try to figure out what these are with interesting heterogeneous pension insurance survival data that is used for testing and illustration.

In this paper, we employ real data from the research register of the Finnish Centre for Pensions. In the Finnish context, the causes of disability correlate with mortality as highlighted in Polvinen *et al.* (2015) and Sewdas *et al.* (2020). Using these data and focusing on mortality, we test and illustrate several important and practical modelling scenarios. In the first scenario, we apply the basic Cox proportional hazards model. In the second scenario, we apply mixture modelling to survival times, and then we use the identified mixture components as a risk factor in the basic Cox model. Our third scenario includes applying mixture modelling within Cox model. The final scenario involves using a frailty model with individual-level frailties. Our results are summarized in the final section as well as recommendations for future research.

2. Theoretical Background

2.1. Finite mixture analysis

A finite mixture of censored T distribution regression models is defined by letting G denote the number of mixture components (*i.e.*, groups) in the mixture model and Z_i denote the latent class (random) variable which indicates the component (group or sub-population) to which the ith observation (individual) belongs. The conditional density of the outcome variable Y_i, $i = 1, \ldots, n$, given $Z_i = j$, is given by

$$Y_i|\{z_i = j\} \sim T(\beta_0 + \mu_j + \mathbf{x}_i^T\beta, \sigma_j^2, \nu), \quad j = 1, \ldots, G, \tag{1}$$

where $(\beta_0 + \mu_j)$ is the intercept of jth group, $\mathbf{x}_i$ is a p-dimensional predictor vector, β is a p-dimensional parameter vector, σ_j is the scale parameter, ν is the degrees of freedom and $T(\mu, \sigma, \nu)$ is the generalized T distribution with location parameter μ, scale parameter σ and ν degrees of freedom. Note that if $Y \sim T(\mu, \sigma, \nu)$, then the distribution of $(Y - \mu)/\sigma$ is standardized Student's T distribution with ν degrees of freedom. Clearly, the mean of the density within the jth group is $\beta_0 + \mu_j + \mathbf{x}_i^T\beta$.

In the case of right censoring, the observed ith outcome can be defined as follows:

$$Y_i^* = \begin{cases} c_i, & \text{if } \rho_i = 1 \ (i.e.\ y_i > c_i) \\ y_i, & \text{if } \rho_i = 0. \end{cases} \tag{2}$$

Here,

$$\rho_i = \begin{cases} 1, & \text{if the } i\text{th observation is right-censored,} \\ 0, & \text{if the } i\text{th observation is not censored.} \end{cases}$$

Suppose that $P(Z_i = j) = p_j$ and Z_i is independent of predictor variables $\mathbf{x}_i$. The maximum likelihood estimate is then obtained by maximizing the log-likelihood function

$$l(\theta|y_1^*, \ldots, y_n^*, \rho_1, \ldots, \rho_n) = \sum_{i=1}^{n} \log\{p_j[f_{ij}(y_i^*)]^{1-\rho_i}[1 - F_{ij}(y_i^*)]^{\rho_i}\} \tag{3}$$

where f_{ij} and F_{ij} are the probability density function and cumulative distribution function of $T(\beta_0 + \mu_j + \mathbf{x}_i^T\beta, \sigma_j^2, \nu)$ distribution. Correspondingly, $\theta = (\theta_1^T, \ldots, \theta_G^T)$, where $\theta_j = (p_j, \beta_j^T, \sigma_j^2, \nu)^T$. In the R-package CensMixReg, the maximum likelihood estimate of θ is found using an EM-type algorithm (Dempster *et al.*, 1977). Note that it is also possible to fit a finite mixture of censored normal distribution regression models using CensMixReg. However, in our empirical experience and the context of our application, using the T distribution gave more consistent results than the normal distribution.

The posterior probability of the ith individual belonging to the jth mixture component is estimated by

$$p_{ij} = \frac{p_j[f_{ij}(y_i^*)]^{1-\rho_i}[1 - F_{ij}(y_i^*)]^{\rho_i}}{\sum_{h=1}^{G} p_h[f_{ih}(y_i^*)]^{1-\rho_i}[1 - F_{ih}(y_i^*)]^{\rho_i}}, \quad i = 1, \ldots n, \ j = 1, \ldots, G. \tag{4}$$

The individual i is assigned to the group having the highest posterior probability estimate.

Likelihood-based inference can be very difficult in mixture models and this is also the case when mixtures and Cox models are combined. For example, when testing hypotheses $H_0 : \ G = G_0$ against $H_A : G = G_0 + 1$, the usual likelihood ratio testing should not be applied because the clusters may not be nested. Also, other point identification and boundary problems may appear. However, the usual information criteria and some other statistical measures may still be applied.

2.2. The proportional hazards (PH) and frailty model

2.2.1. The hazard function

Let Y be the time to an event. The hazard function can be defined as follows:

$$h(y) = \lim_{\Delta y \to 0} \frac{P(y \leq Y < y + \Delta y | Y \geq y)}{\Delta y}. \tag{5}$$

In this expression, $P(y \leq Y < y + \Delta y | Y \geq y)$ is the conditional probability that the event occurs in a short time-interval, given that the event has not occurred before time y. Note that $h(y)$ is not the probability that the event occurs at time y or before time y. We can interpret the hazard function $h(y)$ as an instantaneous rate of occurrence of an event (*e.g.*, death). We can also approximate the conditional probability with $h(y)\Delta y$, where Δy is a small positive real number.

2.2.2. The Cox PH model with finite mixtures

The Cox proportional hazards model can be defined in terms of the hazard function in the following manner:

$$h_i(y) = \lambda_0(y) \exp(\beta_1 x_{i1} + \ldots + \beta_k x_{ik}), \tag{6}$$

where $h_i(y)$ is the hazard of individual i at time y, $x_{i1}, \ldots, x_{ik}$ are k covariates of the individual i, $\beta_1, \ldots, \beta_k$ are the model regression coefficients and $\lambda_0(y)$ is the baseline hazard function.

The above model is called the proportional hazards model because the hazard ratio relating individual i to individual j,

$$\frac{h_i(y)}{h_j(y)} = \exp\{\beta_1(x_{i1} - x_{j1}) + \ldots + \beta_k(x_{ik} - x_{jk})\}, \tag{7}$$

does not depend on time nor on the base hazard function $\lambda_0(y)$. In his groundbreaking paper, Cox (1972) showed that the regression coefficients can be estimated using partial likelihood methods without knowing the form of the base hazard.

In the mixture modelling context the population density function of time to event has the finite mixture form

$$f(y) = \sum_{j=1}^{G} \pi_j f_j(y), \tag{8}$$

where $f_1(y), \ldots, f_G(y)$ are the densities of the mixture components and $\pi_1, \ldots, \pi_G$ are the mixing proportions which add up to one. Then the survival function of the time to the event has the mixture form

$$S(y) = \sum_{j=1}^{G} \pi_j S_j(y), \tag{9}$$

where $S_1(y), \ldots, S_G(y)$ are the survival functions of the mixture components. However, it has been shown that the hazard function of the time to the event does not have a mixture form under CPH (McLachlan and McGiffin, 1994; Ng *et al.*, 2019). Another possibility is to use a mixture specification of the hazard function

$$h(y) = \sum_{j=1}^{G} \pi_j h_j(y), \tag{10}$$

where $h_1(y), \ldots, h_G(y)$ are the hazard functions of the mixture components, but in that case the survival function of the time to the event does not have a mixture form (McLachlan and McGiffin, 1994; Ng *et al.*, 2019).

Eng and Hanlon (2014) have proposed a method where the mixture components are estimated using EM-algorithm. If the observation y of the jth mixture component follows PH model, it has the density

$$f_j(y, \delta|\mathbf{x}) = [\lambda_{0j}(y) \exp(\mathbf{x}^T \beta_j)]^{\delta} \exp[-H_{0j}(y) \exp(\mathbf{x}^T \beta_j)], \tag{11}$$

where δ is the censoring indicator ($\delta = 1$, if the survival time is observed), $\mathbf{x}$ is the covariate vector, y is the survival time, $\lambda_{0j}(y)$ is the baseline hazard, $H_{0j}(y)$ is the cumulative hazard and β_j is the regression coefficient vector. The density of the complete data can then be written as

$$f(\mathbf{y}, \boldsymbol{\Delta}|\mathbf{x}, \mathbf{U}) = \prod_{i=1}^{n} \sum_{j=1}^{G} [\pi_j f_j(y_i, \delta_i|\mathbf{x}_i)]^{u_{ij}}, \tag{12}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ are the survival times, $\boldsymbol{\Delta} = (\delta_1, \ldots, \delta_n)$ are the censoring indicators ($\delta_i = 1$, if the ith survival time is observed), $x_i = (x_{i1}, \ldots, x_{ik})^T$ is the covariate vector, $\mathbf{U} = (U_1, \ldots, U_n)$ are the latent mixture components ($U_i = j$, if the ith observation belongs to the jth component), $P(U_i = j) = \pi_j$ and $u_{ij} = 1_{\{U_i = j\}}$. Eng and Hanlon (2014) maximized the mixture likelihood using EM-algorithm and called the method *Cox-assisted clustering* (*CAC*).

2.2.3. The shared frailty PH model

The shared frailty model is defined as follows:

$$h_{ij}(y) = h_0(y) \exp(\mathbf{x}_{ij}^T \beta + w_j) = h_0(y) u_j \exp(\mathbf{x}_{ij}^T \beta),\ j = 1, \ldots, G,\ i = 1, \ldots, n_j, \tag{13}$$

where w_j, $j = 1, \ldots, G$, are i.i.d. random effects distributed as $N(0, \sigma^2)$, $h_{ij}(y)$ is the conditional hazard of individual i from the jth component (conditional on w_j), $\mathbf{x}_{ij}$ is the

vector of covariates, β are regression coefficients and $h_0(y)$ is the baseline hazard function. The $u_j = e^{w_j}$ term is the frailty of the jth component, where $j = 1, \ldots, G$. Note that now $\log(u_j)$ is normally distributed. It follows that the hazard ratio is

$$\frac{h_{i_2 j_2}(y)}{h_{i_1 j_1}(y)} = \frac{u_{j_2}}{u_{j_1}} \exp((\mathbf{x}_{i_2 j_2} - \mathbf{x}_{i_1 j_1})^T \beta), \quad j_1 \neq j_2. \tag{14}$$

From this expression, it is seen that the hazard ratio depends on the frailties. We can easily see that if the individual i_2 from the j_2th component and the i_1 from the j_1th component have identical covariate profiles (*i.e.* $\mathbf{x}_{i_2 j_2} = \mathbf{x}_{i_1 j_1}$), then the hazard ratio simplifies to

$$\frac{h_{i_2 j_2}(y)}{h_{i_1 j_1}(y)} = \frac{u_{j_2}}{u_{j_1}}. \tag{15}$$

Note that if the number of components is the same as the number of individuals (*i.e.* $n = \Sigma_j n_j = G$), we get the following individual frailty model or random effects frailty model:

$$h_i(y) = h_0(y) \exp(\mathbf{x}_i^T \beta + w_i) = h_0(y) u_i \exp(\mathbf{x}_i^T \beta), \quad i = 1, \ldots, n. \tag{16}$$

3. Pension Insurance Mortality Data

The practical application of this study relates to pensioners and their mortality. The research data was collected from the research register of the Finnish Centre for Pensions. The register is a national databank of the Finnish population, including both the working age population and retirees. The databank contains comprehensive socioeconomic information on the population and statutory pensions.

The base population consisted of the subset of individuals born in 1940 who were still alive in 1995 (*i.e.*, aged 55). The data is cross-sectional from an analytical point of view but entails longitudinal follow-up information on lifetime from 1995 to 2018 (*i.e.*, from ages 55 to 78). Here, we used a 50-per-cent random sample from the selected cohort and this translated into a total of 10,637 individuals. The remaining lifetime of this cohort underlines the difference in mortality between men and women since, according to Official Statistics of Finland (2020), the expected years alive after 2019 for men was 8.9 years and for women 10.9 years.

The research data contains individual-level information on the following variables:

- Lifetime in years (from 55th birthday until 31 Dec. 2018),
- Gender,
- Pension benefit (Disability Pension or Old-age Pension),
- Cause of permanent disability leading to disability pension (8 classes),
- Employer before retirement (Private Sector, Public Sector or Self-employed),
- Censoring (1=alive or 0=deceased on 31 Dec. 2018),

- Highest Education (Basic Education, Secondary Education, Lower University Degree or Higher University Degree), and
- Age at retirement (a continuous variable while the others are classifying variables).

Lifetime is the response variable that we analyzed in this study, using both finite mixture and survival modelling techniques.

Our analysis focuses on mortality. We followed the individuals of the cohort born in 1940 for the 24-year-period from 1995 to 2018 (*i.e.*, ages 55 to 78). This follow-up time translated into a great share of the cohort alive at the end of the study period (over 72%), disregarding possible illnesses. Furthermore, in the Finnish population (as in many other western countries), the probability of permanent disability, which ends working lives for the majority of the seriously ill people, increases rapidly after age 55. However, a vast majority of the cohort survived without permanent illnesses and could retire on an old-age pension at the agreed retirement age of the pension scheme (65 yrs.). From the perspective of SA, the censoring rate in the population was 72.8 per cent.

The actual data analysis was performed using R software (R Core Team, 2019). Specifically, the CensMixReg package (Sanchez *et al.*, 2018) was used to identify mixture groups, and the Survival (Therneau and Lumley, 2019) and Coxme (Therneau, 2020) survival analysis R packages. Furthermore, we have also used the *cac* R function of Eng and Hanlon (2014) to fit the Cox PH model within the mixture model context.

Table 1 shows some basic descriptive statistics within the respective factor classes. The counts indicate a reasonable number of cases in the classes in the sense that the chosen categories are not so fine as to result in sparse data. Specifically, the smallest class, higher university graduates ($n = 134$), is non-censored, and the other classes include more than 300 individuals.

The shares of the deceased (non-censored) indicate that mortality among men and disability pensioners is high, as high share of these groups face death before age 78. The share of non-censored is significantly lower among women (20%), old-age pensioners (19%) and lower for university graduates (20%).

The same classes are reflected in the average lifetimes of the non-censored. When comparing men and women, the lifetime (by age 78) of the non-censored women is 0.9 years longer than that of men. The same difference can be seen when comparing disability pensioners with old-age pensioners. The lifetime of disability pensioners is 3.9 years shorter than that of old-age pensioners. When comparing classes of education, the lifetimes are relatively similar within the three lowest classes: basic education (15.6 yrs.), secondary education (15.4 yrs.), and lowest university education (15.8 yrs.). The lifetime of high university graduates is 1.1 years longer than that of those with a basic education. The classes of employer before retirement indicate significant but small differences between classes.

Descriptive statistics indicate that we can expect significant differences between genders and pension benefits in the statistical analyses given in the following section.

Table 1: Descriptive statistics of cohort born in 1940

	Non-censored			Censored	Total
	Count	%	Average Lifetime, years*,**	Count	Count
Men	1,803	36	15.3 [15.1–15.5]	3,269	5,072
Women	1,087	20	16.2 [15.9–16.4]	4,478	5,565
Old-age pension	1,300	19	17.8 [17.6–18.0]	5,635	6,935
Disability pension	1,590	43	13.9 [13.6–14.1]	2,112	3,702
Basic ed.	1,721	30	15.6 [15.3–15.8]	3,995	5,716
Secondary ed.	676	27	15.4 [15.1–15.8]	1,856	2,532
Low univ.	359	20	15.8 [15.3–16.3]	1,400	1,759
High univ.	134	21	16.7 [16.0–17.5]	496	630
Private Sector	1,490	30	15.3 [15.0–15.5]	3,481	4,971
Public Sector	1,005	24	15.9 [15.6–16.2]	3,247	4,252
Self-employed	395	28	16.1 [15.7–16.6]	1,019	1,414
Total	2,890			7,747	10,637

*90% confidence limits in parentheses.
**Average lifetime from 55th birthday until 31 Dec. 2018.

4. Data Analysis Using Various Modelling Techniques

4.1. The Cox model

The first step of our study was to analyze our data using a Cox proportional hazards model. The response variable was Lifetime and the explanatory variables were Gender, Pension benefit, Age at retirement, Highest education and Employer. The model estimates (Hazard ratios) are given in Table 4 (under the column titled "Cox PH Fit"). The analysis was done both for the whole data set and for the data set with censored cases excluded. The first analysis includes 10,637 individuals, of which 2,890 are non-censored. The results are reasonable and mirror much of what was demonstrated in Table 1. From the descriptive statistics given in Table 1, significant differences in lifetime in two variables, Gender and Pension benefit, are observed. From the Cox PH model analysis (Table 4) we observe that women's mortality rate is significantly lower (*i.e.*, hazard ratio 0.5) compared to men's mortality rates. Also, disability pensioners have a much higher hazard ratio of mortality (2.42) than old-age pensioners. In comparison with individuals in the basic education level, those with a higher university education have a slightly lower mortality rate (the coefficient for

higher university education (0.82) is statistically significant). Employer is not a statistically significant variable in this model.

The second analysis was done for the non-censored data set to make the analysis and goodness-of-fit measures comparable to the results of the Cox PH with ex ante (in the SA, we call the groups as "ex ante groups" since they are constructed using the survival model underlying equation (17) below. The concordance statistics is an established measure of the goodness-of-fit in survival models (see Harrell *et al.*, 1996); for the basic Cox model, the concordance value is 0.61. Subsequent sections show differences between other models and Cox PH model.

4.2. Modelling mixture components

The second step of our analysis was to first study the distribution of the lifetime. The lifetime is largely centered around 23–24 years, indicating the still living individuals. Furthermore, the long left tail indicates increasing mortality with age. The first step of finite mixture modelling was to search for possible mixture components or sub-groups from the outcome. In general, choice of the number of mixture components is a key central question in finite mixture modelling, and it is usually determined via statistical information criteria (*e.g.*, Bayesian Information Criteria, BIC) and with the subjective consideration of the modeller. In our application, the number of groups was based on the sizes of the mixture components, as we did not want possible artefact or too small groups in the further steps of the analysis.

When using the CensMixReg R package (and included functions), which takes account of right-censoring on the outcome, a regression model must be specified. To this end, we defined the following simple regression model:

$$Lifetime = PensionBenefit + RetirementAge + HighestEducation \tag{17}$$

The distribution of the outcome Lifetime is such that the R implementation standard assumption of the normal distribution was switched to the T distribution, which behaved slightly more stable in our analyses. Some experiments with gender as a factor indicated that the model is somewhat sensitive to the underlying regression model (factors), and so we chose to use a simple model, which provided us with a reasonable number of mixture components and which could be estimated with the EM algorithm. The BIC values for $k = 2, 3, 4$ groups were respectively 28757.09, 28799.61 and 28955.36. As is often the case, choices of a larger number of mixture components/groups led to convergence problems of the EM algorithm, and were therefore not considered here. Overall in the BIC analysis the model solution was stable between several model runs.

The four-component solution yielded the group sizes as shown in Table 2. We named the groups based on average lifetime, counted from the data using the group-assignments. The No risk group included those who were alive at the end of the 23-year study period, and thus had censored measurements in the survival modelling analysis.

To further illustrate the above lifetime analysis, we draw the distribution of lifetimes by

Table 2: Mixture group sizes and lifetimes

	Estimated Counts	%	Average Lifetimes (years)
No Risk (NR)	7,747	72.8	> 23.6
Low Risk (LR)	169	1.6	22.7
At Risk (AR)	325	3.1	20.1
High Risk (HR)	2,396	22.5	14.5

mixture component. The boxplot of lifetimes for the mixture components are shown in Figure 1. As demonstrated by Figure 1 and Table 2, the components or sub-groups have greatly different mortalities. The vast majority of the sample (72.8%) are in the No Risk group, and alive at the end of study period. The Low Risk and At Risk groups show a slightly increased mortality with an average lifetimes of 22.7 years and 20.1 years and relatively narrow range of lifetimes. The average lifetime in High Risk group is only 14.5 years and the Figure shows also a wide range of lifetimes. The overall conclusion from Figure 1 is that the distribution between mixture components barely overlap indicating a clear group assignment of the above mixture analysis.

Although the distribution of lifetimes barely differ for men and women within components, the shares or proportions of men and women within the components are different, as there are significantly more men in the High Risk group (see Table 3 below). These additional results give further assurance for the group-based correlation of survival times and we will utilize this information in conjunction with the Cox PH regression model in the following section.

Table 3: Mixture groups by gender

	Men		Women		Total	
	Count	%	Count	%	Count	%
No Risk (NR)	3,269	64.5	4,478	80.5	7,747	72.8
Low Risk (LR)	104	2.1	65	1.2	169	1.6
At Risk (AR)	196	3.9	129	2.3	325	3.1
High Risk (HR)	1,503	29.6	893	16.0	2,396	22.5
Total	5,072		5,565		10,637	

The mixture groups can be further dissected and parsed by analyzing the cause of retirement on a disability pension. The cause of permanent disability gives a clear indication or a proxy of an individual's health, and some causes of disability are more life-threatening than others. For example, in examining Figure 2, we see that neoplasms are often life-threatening in the sense that those presenting with neoplasms are represented to a large degree in the High Risk mortality group. Indeed, approximately 74 per cent of individuals with neoplasms were in the High Risk mortality group. Conversely, individuals with depression or diseases of the musculoskeletal system were somewhat over-represented in the No Risk group. Interestingly, although the cause of disability was not used in mixture analysis, this analysis revealed substantial agreement with these results.

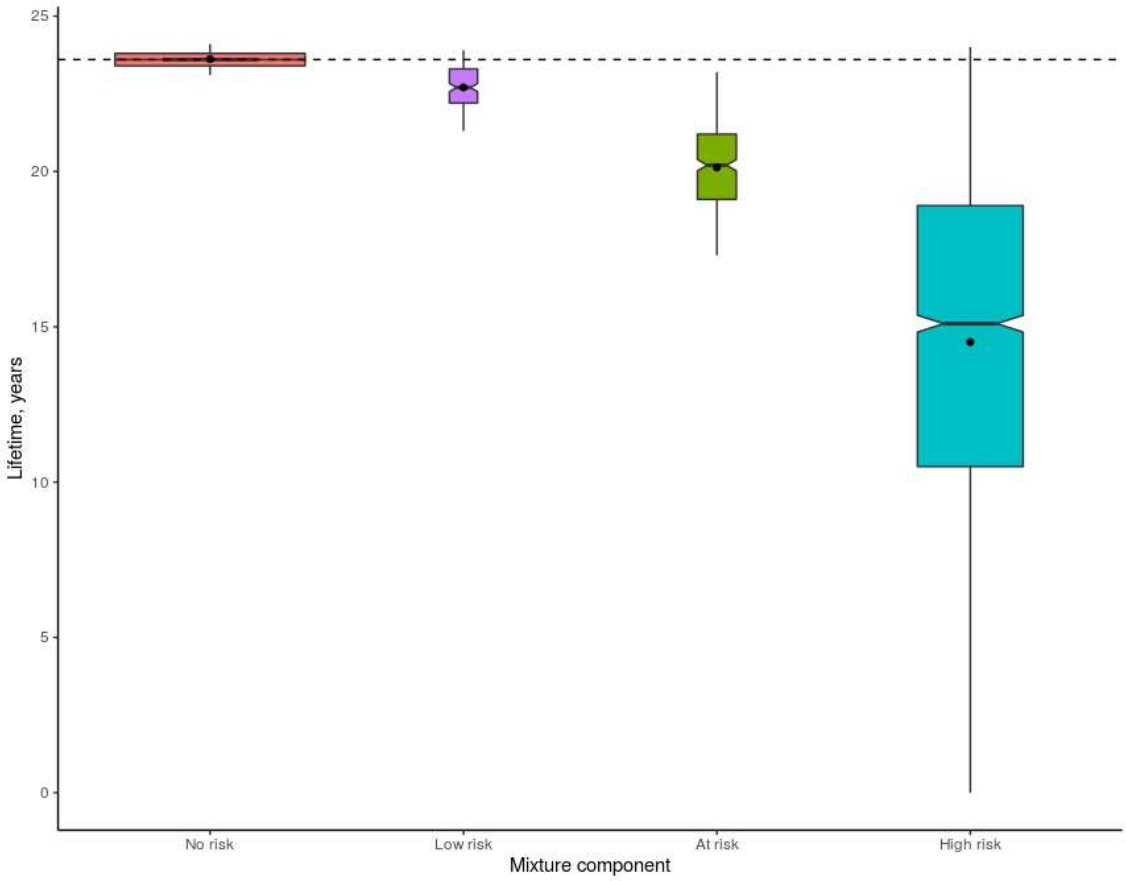


Figure 1: Box plot of lifetime by mixture component. Box widths are proportional to sample sizes. The notches indicate the sample medians and the dots indicate the sample means. The dashed line indicates the sample median of the censored observations.

4.3. The Cox model with mixture components as a factor

The third step in our analysis was to combine the four-class factor obtained in the preceding section with the basic Cox PH model developed in Section 4.1. To achieve this, we used the risk groupings from the information of ex ante defined mixture groups. The hazard ratios are also presented in Table 4 in the column labelled "Cox PH + Ex Ante Fit". For those individuals who died before the end of the 23-year follow-up period, we select one class of the mixture components as a reference group, and in this analysis, the chosen reference group was the Low Risk group. As noted in Table coefficients, the fitted results have changed in some meaningful ways as compared with the Cox PH model fit. For example, now the hazard ratio of disability pensioners is much higher (7.00) compared to old-age pensioners. The mixture-component-based classes are statistically significant in the model. The coefficients are reasonable and confirm that the mixture groups capture interesting sub-populations. The estimates for the At Risk group (3.03), and especially for the High Risk group (24.00), are large compared to the reference (Low Risk) group, as is to be expected.

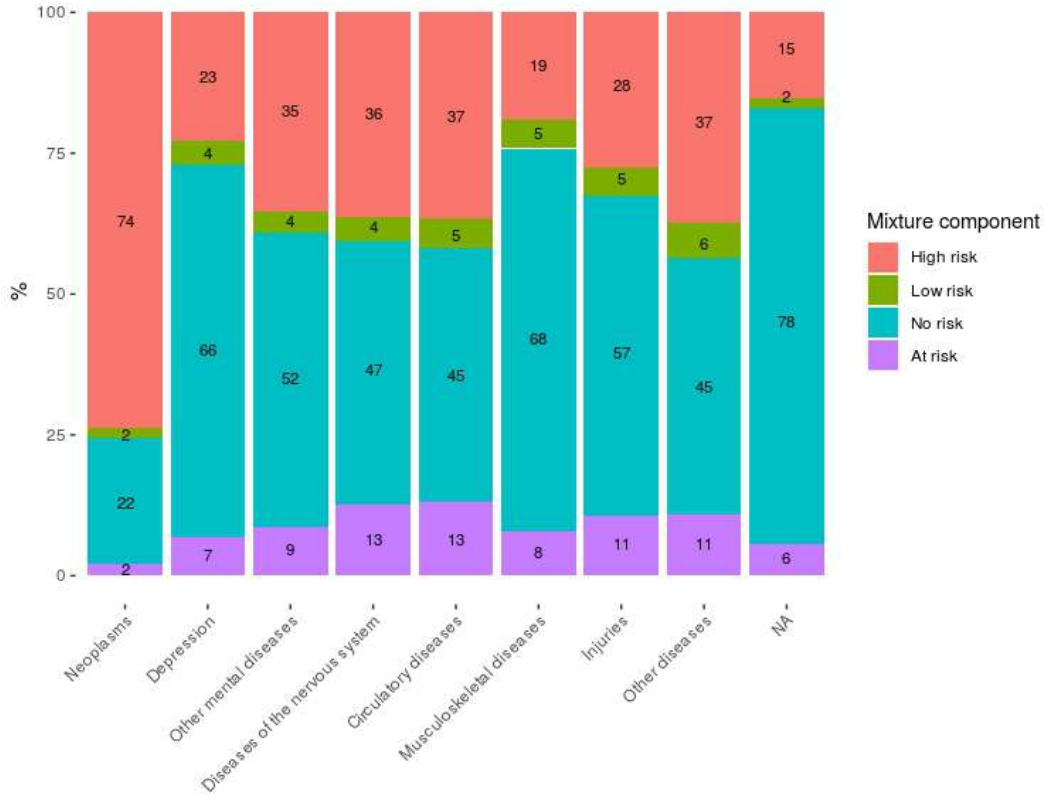


Figure 2: Cause of disability by mixture component, %

In comparing the basic Cox PH model fit with the Cox PH combined with Ex Ante fit (treating groups as a factor), the LR test statistic χ^2 value is 1754.9 on 2 degrees of freedom. This difference is highly significant and this indicates that adding the group component information significantly increases the predictive power of the Cox PH model. Similarly, the concordance statistic (0.73) for the latter model is higher than the value corresponding to the Cox PH model (0.61), which also indicates an improved model fit.

4.4. The frailty model with random effects

The fourth step of our analysis included fitting a frailty model with individual random effects. The results are listed in Table 4 (Column titled "Frailty with Random Effects Fit"). Note that in this analysis we did not use the mixture component factor developed in Section 4.2 and used in Section 4.3. The results can be summarized by noting that the hazard ratios are quite similar to the basic Cox PH model. The coefficient for women (0.47) is slightly smaller than in the Cox PH model (0.50) and significantly smaller than in the Cox model with mixture components as a factor (0.92), and the coefficient for the disability pension (2.67) is slightly higher than in the Cox PH model (2.42). As was the case for the basic Cox PH model, the Frailty model disability pension estimate differs greatly from the Cox PH model with mixture components, which yields the estimated hazard ratio of 7.0.

Table 4: Cox and frailty regression estimates (hazard ratios)

		Cox PH Fit		Cox PH + Ex Ante Fit	Frailty with Random Effects Fit
		Censored included	Censored excluded	Censored excluded	Censored included
Gender:	Women	0.50***	0.91*	0.92***	0.47***
	Men	ref	ref	ref	ref
Pension:	Old-age	ref	ref	ref	ref
	Disability	2.42***	1.50***	7.00***	2.67***
Pension age		0.98***	0.99	0.97***	0.98***
Education:	Basic ed.	ref	ref	ref	ref
	Secondary ed.	0.92	1.05	0.88**	0.91*
	Low univ.	0.77***	0.99	0.78***	0.75***
	High univ.	0.82*	0.95	0.84	0.79*
Employer:	Private sector	ref	ref	ref	ref
	Public sector	1.06	0.97	0.95	1.06
	Self-employed	0.94	0.96	0.94	0.93
Group:	No Risk	N/A	N/A		N/A
	Low Risk	N/A	N/A	ref	N/A
	At Risk	N/A	N/A	3.03***	N/A
	High Risk	N/A	N/A	24.00***	N/A
RE Variance					0.60
Concordance			0.61	0.73	
The number of cases (n)		2890	2890	2890	2890

Significance codes: $***< 0.001$, $**< 0.01$, $*< 0.05$

The basic Cox PH model test against the frailty mode were compared with a LR test. The value of the statistic was 33.978 (1 degrees of freedom), which clearly indicates that the frailty model with individual effects increases the predictive power of the Cox model.

To further develop and illustrate this model, we provide the estimated random coefficients (*i.e.*, individual frailties) in Figure 3 in the Appendix. The estimated values for the No Risk group (mean = -0.16) are all negative highlighting that the individuals in this group survived until the end of the study period (*i.e.*, until age 78). Essentially all of the High Risk (0.48), At Risk (0.24) and Low Risk (0.13) group estimates are positive. Not surprisingly, the range of the coefficients do not differ appreciably between the Low-Risk and At-Risk groups, but the coefficients for the High Risk group are notably higher (mean and median around 0.5) than in the other groups.

4.5. The mixture Cox model

Our last analysis consists of mixture modelling of lifetime using the finite mixture form of the Cox PH model as shown in section 2.2.2. The number of mixture groups is more complicated in our application as far more groups are separated or indicated as compared to the ex ante modelling of lifetime as in section 4.2. The BIC analysis focusing on non-censored cases indicate at least 15 groups with relatively similar survival patterns. The survival times of the groups are similar and near each other. To simplify the results, we selected a three-group solution with relatively equal sizes. Also 4 group solution was tested, but the magnitude of the estimates obtained was no longer at a plausible level.

Table 5 shows the Cox model estimates within the mixture groups. The gender (woman) effect is highly group-specific and estimates also differ compared to basic Cox model (Table 4, Censored excluded). Estimates on disability pension and pension age are in the same range between groups and, similar to basic Cox model. Estimates of level of education differ greatly between groups. Overall the statistical significance of most of the covariates is high, especially in groups two and three. Figure 4 in the Appendix shows the corresponding distribution of lifetime within mixture groups. The figure shows that the lifetimes are relatively similar between groups. In the high mortality group one the average lifetime is 14.1 years and correspondingly 15.8 years and 18.2 years in groups two and three. The model selected here includes three sub-groups (Non-censored only). However, we can see that the groups are different from the ex ante mixture analysis (see Table 2) where groups indicate very different mortalities. The ex ante mixture analysis showed a large High Risk group ($N = 2,396$) with high mortality (average lifetime 14.5 yrs.) and two smaller groups with moderate mortality. In the mixture Cox model the groups are somewhat more similar and not easily comparable with the ex ante groups.

5. Concluding Remarks

Survival analysis (SA) techniques are appealing approaches in many fields of research and application because of their simplicity, their nonparametric nature, and their accommodation of the presence of censoring. The key question regarding the ubiquitous Cox model is whether the resulting model estimates are unbiased. To address such deficiencies, frailty models have been developed to include a random factor for unknown covariates, and the frailties aim to capture effects which are not explicitly included in the model. Furthermore, the merit of the finite mixture modelling approach is that it reveals possible latent classes from a given outcome distribution. As a starting point, as is demonstrated in our work here it is noteworthy and interesting to bring the sub-groups or mixture components into the context of SA. Using these techniques in parallel will continue to provide new insights into the Cox model, especially regarding estimates and dealing with any related biases.

In this study, we analyzed a simple empirical data, including information on lifetimes and some background factors of a Finnish cohort. The focus of the analysis was on modelling lifetime at ages 55 to 78. The basic Cox model revealed differences in mortality with respect to gender, pension benefit and education. The results indicate that women face a smaller likelihood of death compared to men, and disability pensioners face a far greater likelihood of death than old-age pensioners.

Table 5: Mixture Cox model regression estimates by mixture group (hazard ratios)

		Group 1	Group 2	Group 3
Gender:	Women	0.01***	7.91***	14.93***
	Men	ref	ref	ref
Pension:	Old-age	ref	ref	ref
	Disability	4.26***	2.64***	4.63***
Pension age		0.99	0.86***	1.04***
Education:	Basic ed.	ref	ref	ref
	Secondary ed.	0.94	0.02***	24.17***
	Low univ.	3.14***	0.004***	11.35***
	High univ.	0.01***	0.07***	82.79***
Employer:	Private sector	ref	ref	ref
	Public sector	1.04	1.55***	1.16
	Self-employed	0.10***	3.57***	1.01
The number of cases (n)		1037	923	930

More importantly, we analyzed the outcome lifetime with a finite mixture technique and discovered that it consists of four distinct sub-populations with different level of mortalities. In the application considered here, the frailty model with individual random effects yields estimates approximately the same as in the basic Cox PH model. Clearly Cox model estimates change when adding mixture component as a factor. Thus, the results show some discrepancy in parameter estimates, or bias, in basic Cox PH model estimates. Both goodness-of-the fit statistics and likelihood-ratio tests improve using the extensions provided here. Nonetheless, the frailty model yields estimates that are close to the basic Cox model. The results for Cox-mixture analysis show differences between groups regarding gender and level of education. Results on disability pension are similar compared to other models. It is likely that the group-composition, which is different compared to ex ante groups, affects the estimates.

These analyzes indicate substantially that, at this stage of the life course, there are significant differences in mortality between men and women, and indeed the expected lifetime for women of the studied cohort is known to be about two years longer than that for men. The eligibility rules of disability pension are strict because there must be a severe and long-term illnesses in order to get a pension. Therefore, the increased likelihood of premature death indicated by the models is no surprise.

Our ongoing work includes analytically merging the Cox PH model with mixture modelling techniques, with an eye to developing a much-needed open-source (*e.g.*, R) package to facilitate use by practitioners and statistical modellers. As noted previously, the cac function is designed to this kind of analysis and it can be useful in analyzing Cox mixture models.

Acknowledgements

The authors wish to thank the anonymous referee for many thoughtful comments that greatly improved this paper.

References

Böhning, D., Seidel, W., Alfó, M., Garel, B., Patilea, V. and Walther, G. (2007). Advances in mixture models. *Computational Statistics and Data Analysis*, **51**, 5205–5210.

Collett, D. (2015). *Modelling Survival Data in Medical Research.* Third Edition. CRC Press, Boca Raton, FL.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society,* **B34**, 187–220.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society*, **B39**, 1–38.

Eng, K. H. and Hanlon, B. M. (2014). Discrete mixture modeling to address genetic heterogeneity in time-to-event regression. *Bioinformatics*, **30**, 1690–1697.

Ha, I. D., Jeong, J.-H. and Lee, Y. (2017). *Statistical Modelling of Survival Data with Random Effects. H-Likelihood Approach.* Springer Nature, Singapore.

Harrell, F. E., Lee, K. L. and Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, **15**, 361–387.

Heckman, J. J. and Singer, B. S. (Eds.) (1985). *Longitudinal Analysis of Labor Market Data.* Cambridge UK Econometric Society Monographs 10. Cambridge University Press, Cambridge, UK.

Machin, D., Cheung, Y. B. and Parmar, M. (2006). *Survival Analysis: A Practical Approach.* Second Edition. Wiley, Chichester.

McLachlan, G. J. and McGiffin, D. C. (1994). On the role of finite mixture models in survival analysis. *Statistical Methods in Medical Research*, **3**, 211–226.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models.* Wiley, Hoboken, NJ.

Nagin, D. S. (2005). *Group-Based Modeling of Development.* Harvard University Press, Cambridge, MA.

Nagin, D. S. and Odgers, C. L. (2010a). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, **6**, 109–138.

Nagin, D. S. and Odgers, C. L. (2010b). Group-based trajectory modeling (nearly) two decades later. *Journal of Quantitative Criminology*, **26**, 445–453.

Ng, S. K., Xiang, L. and Yau, K. K. W. (2019). *Mixture Modelling for Medical and Health Sciences.* Chapman & Hall/CRC, New York.

Nummi, T., Salonen, J. and O'Brien, T. (2017). Statistical analysis of labor market integration: A mixture regression approach. In D.-G. Chen, Z. Jin, G. Li, Y. Li, A. Liu and Y. Zhao, Editors, *New Advances in Statistics and Data Science*, ICSA Book Series in Statistics, 313–321. Springer, Cham.

Official Statistics of Finland (2020). Population projection (e-publication). Helsinki: Statistics Finland [Referred: 28.2.2020]. http://www.stat.fi/til/vaenn/tau_en.html.

Polvinen, A., Laaksonen, M., Gould, R., Lahelma, E., Leinonen, T. and Martikainen, P. (2015). Socioeconomic inequalities in cause-specific mortality after disability retirement due to different diseases. *Scandinavian Journal of Public Health*, **43**, 159–168.

R Core Team (2019). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Sanchez, L. B., Lachos, V. H. and Moreno, E. J. L. (2018). *CensMixReg: Censored Linear Mixture Regression Models.* Version 3.1. https://cran.r-project.org/src/contrib/Archive/CensMixReg/.

Sewdas, R., de Wind, A., Stenholm, S., Coenen, P., Louwerse, I., Boot, C. and van der Beek, A. (2020). Association between retirement and mortality: Working longer, living longer? a systematic review and meta- analysis. *Journal of Epidemiology and Community Health*, **0**, 1–8.

Therneau, T. M. (2020). *coxme: Mixed Effects Cox Models.* Version 2.2-16. https://CRAN.R-project.org/package=coxme.

Therneau, T. M. and Lumley, T. (2019). *Survival: A Package for Survival Analysis in S.* Version 3.1-8. https://CRAN.R-project.org/package=survival.

van den Berg, G. J. (2001). Duration models: specification, identification and multiple durations. In J.J. Heckman and E.E. Leamer, (Eds.), *Handbook of Econometrics*, First Edition, Volume **5**, Chapter **55**, 3381–3460. Elsevier.

Wienke, A. (2010). *Frailty Models in Survival Analysis.* Chapman and Hall/CRC, Boca Raton, FL.

Zhang, Y. (2008). *Parametric Mixture Models in Survival Analysis with Applications.* Ph.D. thesis. Temple University.

APPENDIX

Figures

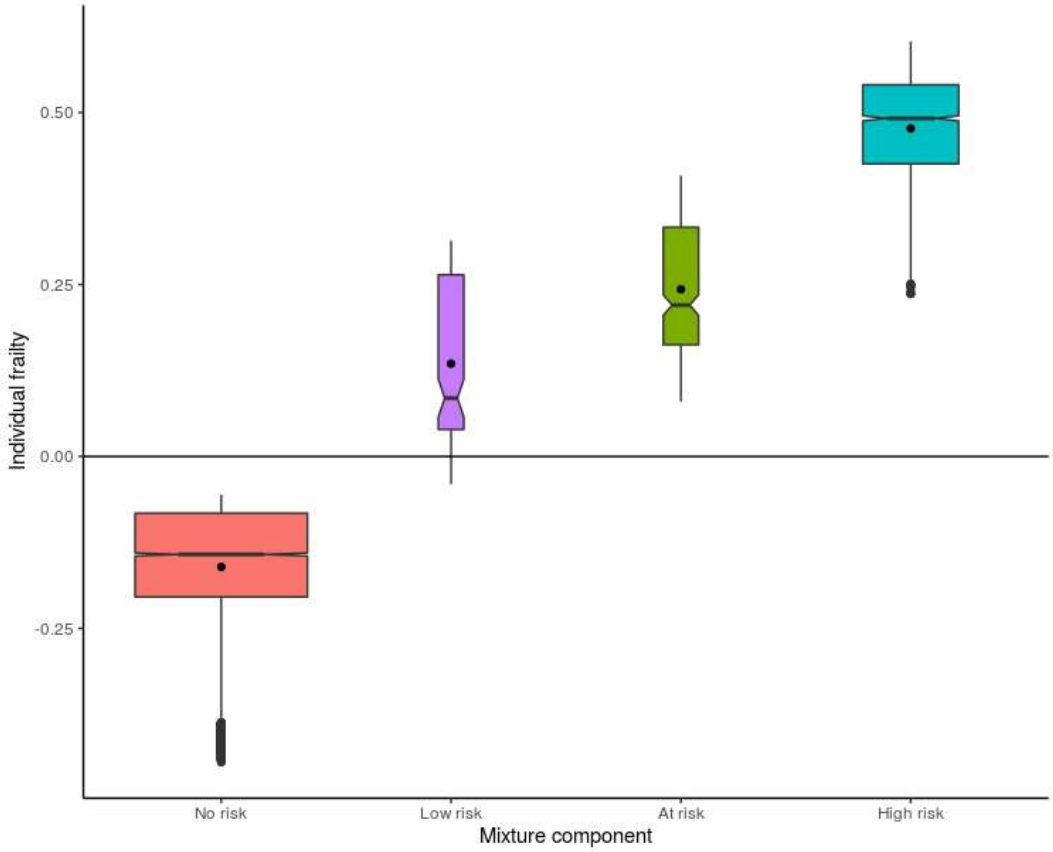


Figure 3: Box plot of individual frailty coefficients by mixture component. Box widths are proportional to sample sizes. The notches indicate the sample medians and the dots indicate the sample means.

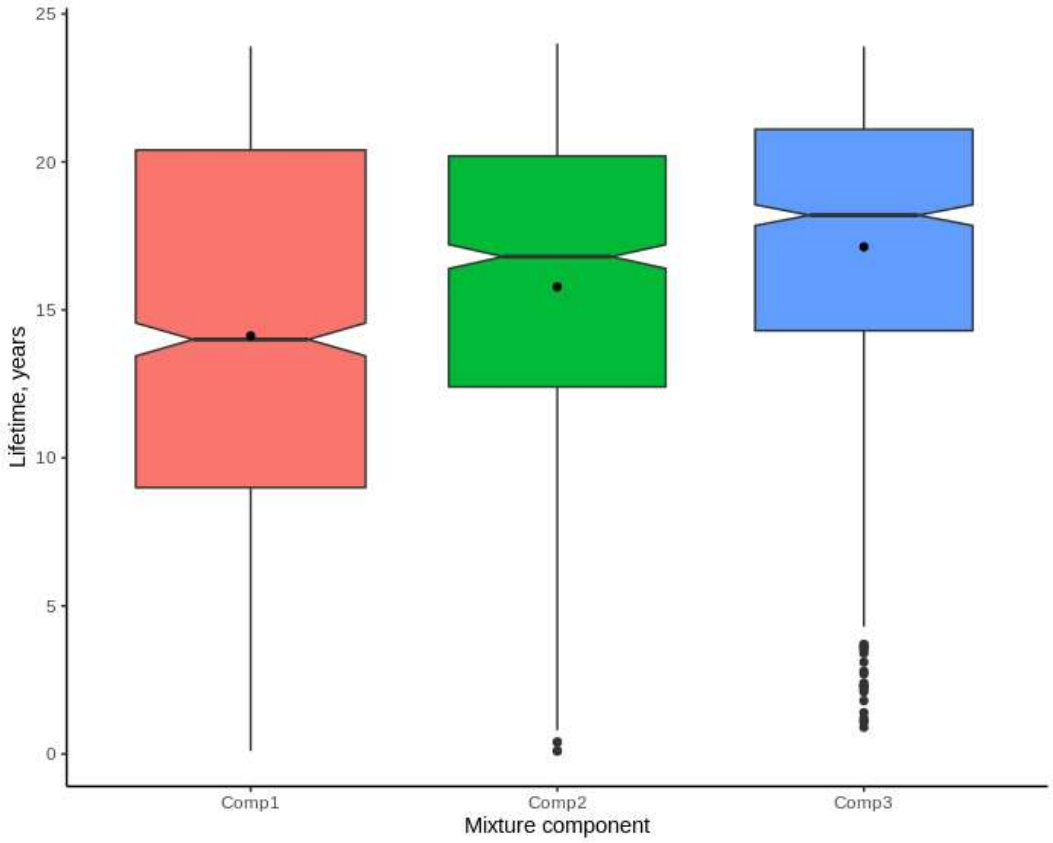


Figure 4: Box plot of lifetimes in mixture Cox model by mixture component. Box widths are proportional to sample sizes. The notches indicate the sample medians and the dots indicate the sample means.

Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 353–368

Estimation of Kurtosis Parameters of Multivariate Populations

Nighat Zahra[1], Muhammad Kashif Ali Shah[1] and S. Ejaz Ahmed[2]
[1]*Department of Statistics, GC University Lahore, Pakistan*
[2] *Department of Mathematics and Statistics, Brock University, St. Catharines, ON, Canada*

Received: 01 June 2020; Revised: 06 September 2020; Accepted: 25 September 2020

Abstract

This paper is based on the estimation of the kurtosis parameters of several multivariate populations under the assumption that all the kurtosis parameters are equal. The shrinkage and preliminary test estimators are suggested for the estimation of the vector of kurtosis parameters. Asymptotic properties of the suggested estimators are presented analytically and compared on the basis of their asymptotic distributional bias and asymptotic quadratic risk. Monte-Carlo simulations are performed in order to explain the analytical results numerically. A real data example is also given to demonstrate the application of the suggested estimators. From the results it can be observed that the Stein-type estimators perform better than all other estimators when the number of populations is greater than four and also when the assumption of homogeneity is suspicious.

Key words: Kurtosis; Shrinkage; Preliminary test; Asymptotic quadratic bias; Asymptotic quadratic risk; Stein-type estimators.

AMS Subject Classifications: 62E20, 62F10, 62F12

1. Introduction

Kurtosis and skewness are often considered as the shape parameters of a probability distribution. First introduced by Karl Pearson in 1905, kurtosis can be defined as a measurement with which to represent the size of a distribution's tails in contrast to a normal distribution. In a contemporary context, kurtosis is widely used in different areas of research, such as finance, space science, economics and signal processing. See Kim and White (2004), Liang *et al.* (2008), Nita and Gary (2010), Lai (2012), Araújo *et al.* (2012), and Echer and Bolzan (2016) for detailed examples.

Mardia (1970) defined the kurtosis parameter of a p-dimensional random variable $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as

$$\beta = \mathbb{E}\left[\left\{(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right\}^2\right]. \tag{1}$$

Corresponding Author: Nighat Zahra
Email: nighatzahra@gcu.edu.pk

For a random sample of n observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ taken from a multivariate population, let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ be the estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively, then the estimator of kurtosis parameter proposed by Mardia (1970) can be defined as

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \left\{ (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}) \, \hat{\boldsymbol{\Sigma}}^{-1} \, (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})^T \right\}^2, \tag{2}$$

where $\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i$ and $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}) \, (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})^T$.

Ahmed *et al.* (2012) worked on the improved estimation of kurtosis parameter β of a multivariate population using uncertain prior information (UPI). They proposed the linear shrinkage and the preliminary test estimators for β and developed the large sample theory for these estimators. Zahra *et al.* (2017 b) presented the improved estimation of kurtosis parameters for two multivariate populations and suggested that the shrinkage pretest estimator performs better when the null hypothesis $\beta_1 = \beta_2$ is uncertain; otherwise, the restricted estimator performs well. In this paper, we have extended their work on multivariate populations to q sample case using the UPI that kurtosis parameters of all populations are homogeneous. The UPI can be presented in the form of null hypothesis as

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_q = \beta_0, \tag{3}$$

where β_0 is unknown.

Let $\boldsymbol{X}_i^{(l)} = \left(\boldsymbol{X}_{1i}^{(l)}, \boldsymbol{X}_{2i}^{(l)}, \cdots, \boldsymbol{X}_{pi}^{(l)} \right)$ where $i = 1, 2, \cdots, n_l$ and $l = 1, 2, \cdots, q$ be a multivariate random sample of size n_l from a p-variate normal distribution with mean vector $\boldsymbol{\mu}_l$ and covariance matrix $\boldsymbol{\Sigma}_l$. We want to estimate the parameter vector of kurtosis coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_q)^T$ using their maximum likelihood estimator (MLE) $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_q)^T$ with sample sizes $(n_1, n_2, \cdots, n_q)$ under the UPI given in equation (3). Our multivariate sampled data $\boldsymbol{X}_i^{(l)}$ may have collected at q different times or spaces and there is natural tendency to combine the data to get efficient estimation results. One has to integrate both the sample information (SI) available in the form of $\boldsymbol{X}_i^{(l)}, l = 1, 2, \cdots, q$ and the UPI in such a way that estimators having optimal properties in terms of smallest risk, can be developed. Such integrated estimation strategies are based on the preliminary test (pretest) and Stein's shrinkage methodologies and are proved to be superior for data polling purposes. Many of these estimation strategies under different contexts have been discussed by Zahra *et al.* (2017 a), Shah *et al.* (2017), Lisawadi *et al.* (2019), Shah *et al.* (2020) and references therein.

The organization of the paper is as follows: Several estimation strategies are given in Section 2 that utilized either SI or the combination of both SI and UPI. Section 3 is concerned with the expressions of the asymptotic distributional quadratic bias (ADQB) and asymptotic distributional quadratic risk (ADQR) of the stated estimators. The results of the Monte-Carlo simulations are given in Section 4. An empirical example is given in Section 5 and the concluding remarks are made in the last section. All computations are done with the latest version of freeware R, and mathematical proofs are given in the Appendix. The matrices and vectors are represented with boldface symbols, while script letters $\mathbb{E}$ and $\mathbb{V}$ are reserved for the operators of expectation and variance.

2. Estimation Strategies

In this section, some improved estimation strategies for the kurtosis parameter vector $\boldsymbol{\beta}$ for q multivariate populations are described. At the first place, Using Mardia's estimator given in equation (2), the unrestricted estimator (UE) of $\boldsymbol{\beta}$ using only the SI is defined as

$$\hat{\boldsymbol{\beta}}^{UE} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_q)^T. \tag{4}$$

The restricted estimator (RE) of the parameter vector $\boldsymbol{\beta}$ under UPI is given as

$$\hat{\boldsymbol{\beta}}^{RE} = \left(\hat{\beta}_0, \hat{\beta}_0, \cdots, \hat{\beta}_0\right)^T = \hat{\beta}_0 \mathbf{1}_q. \tag{5}$$

The point estimator of β_0 denoted by $\hat{\beta}_0$ is calculated as $\hat{\beta}_0 = \hat{\omega}^{-1} \sum_{l=1}^{q} \lambda_{l,n} \hat{\beta}_l^{UE}$, where $\hat{\omega} = \sum_{l=1}^{q} \lambda_{l,n}$ and $\lambda_{l,n} = n_l/n$. An alternative form of $\hat{\boldsymbol{\beta}}^{RE}$, which is further used for the derivation of the mathematical results, is given as

$$\hat{\boldsymbol{\beta}}^{RE} = \hat{\omega}^{-1} \boldsymbol{J}_q \boldsymbol{V}_n^{-1} \hat{\boldsymbol{\beta}}^{UE} = \boldsymbol{H}_n \hat{\boldsymbol{\beta}}^{UE}, \tag{6}$$

where $\boldsymbol{J}_q = \mathbf{1}_q \mathbf{1}_q^T$, $\boldsymbol{H}_n = \hat{\omega}^{-1} \boldsymbol{J}_q \boldsymbol{V}_n^{-1}$, $\boldsymbol{V}_n = \text{diag}\left(\frac{v}{\lambda_{1,n}}, \frac{v}{\lambda_{2,n}}, \cdots, \frac{v}{\lambda_{q,n}}\right)$, and $v = 8p(p+2)$. Assuming $\lim(\lambda_{l,n}) = \lambda_l$ is fixed for $l = 1, 2, \cdots, q$ then $\boldsymbol{V}_n$ converges in probability to $\boldsymbol{V} = \text{diag}\left(\frac{v}{\lambda_1}, \frac{v}{\lambda_2}, \cdots, \frac{v}{\lambda_q}\right)$ as $n \to \infty$ where $n = n_1 + n_2 + \cdots + n_q$. The linear shrinkage (LS) estimator of $\boldsymbol{\beta}$ may be defined as

$$\hat{\boldsymbol{\beta}}^{LS} = \hat{\boldsymbol{\beta}}^{UE} - \pi(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}); \quad \pi \in (0, 1), \tag{7}$$

where π is the degree of trust in the null hypothesis (3). The linear shrinkage estimator becomes a restricted estimator when π is one and an unrestricted estimator when it is zero.

The preliminary test or simply pretest (PT) estimator of the population parameter vector $\boldsymbol{\beta}$ is stated as

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}}^{UE} - \left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right) I(\mathfrak{L}_n < c_{n,\alpha}), \tag{8}$$

where $I(\cdot)$ is an indicator function and $\mathfrak{L}_n$ is a Wald-type test statistic computed as

$$\mathfrak{L}_n = n \left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right)^T \boldsymbol{V}_n^{-1} \left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right). \tag{9}$$

Under the null hypothesis given in equation (3), $\mathfrak{L}_n$ converges in distribution to a χ^2 distribution with $(q-1)$ degrees of freedom. Thus, upper α-level critical values of $\mathfrak{L}_n$ defined by $c_{n,\alpha}$ are approximated by a $\chi^2_{(q-1)}$ distribution. The shrinkage pretest (SP) estimator which incorporates π into equation (8) is given as

$$\hat{\boldsymbol{\beta}}^{SP} = \hat{\boldsymbol{\beta}}^{UE} - \pi \left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right) I(\mathfrak{L}_n < c_{n,\alpha}). \tag{10}$$

It is interesting to note that for $\pi = 1$, the shrinkage preliminary test estimator $\hat{\boldsymbol{\beta}}^{SP}$ is reduced to the preliminary test estimator $\hat{\boldsymbol{\beta}}^{PT}$.

The Stein-type shrinkage (SS) estimator is defined as

$$\hat{\boldsymbol{\beta}}^{SS} = \hat{\boldsymbol{\beta}}^{UE} - (q-3)\mathfrak{L}_n^{-1}\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right), \quad q \geq 4, \tag{11}$$

and an improved Stein-type shrinkage (S+) estimator for $q \geq 4$ is given as

$$\hat{\boldsymbol{\beta}}^{S+} = \hat{\boldsymbol{\beta}}^{UE} - (q-3)\mathfrak{L}_n^{-1}\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right) - \left[1-(q-3)\mathfrak{L}_n^{-1}\right] I(\mathfrak{L}_n < q-3)\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right). \tag{12}$$

Secondly, it is to be noted that the test-statistic given in equation (9) is consistent against fixed $\boldsymbol{\beta}$ such that $\boldsymbol{\beta} \notin H_0$, hence all the estimators involving $\mathfrak{L}_n$ are equivalent to the bench-mark estimator $\hat{\boldsymbol{\beta}}^{UE}$ for the fixed alternatives in a large sample setup. Interested readers are referred to Ahmed (2002) for further details. Therefore, we consider a sequence of local alternatives $\{\mathcal{H}_{(n)}\}$ as

$$\mathcal{H}_{(n)} : \boldsymbol{\beta} = \boldsymbol{\beta}_{(n)}, \tag{13}$$

where $\boldsymbol{\beta}_{(n)} = \beta_0 \mathbf{1}_q + n^{-1}\boldsymbol{\delta}$, and $\boldsymbol{\delta} \in \mathbb{R}^q$ is a fixed real vector. It is to be noted that we are not making any assumption for local alternatives setup either. By virtue of Ahmed *et al.* (2012) and under the sequence of local alternatives defined in equation (13) in the univariate sense, following result holds:

$$\sqrt{n_l}(\hat{\beta}_l - \beta_l) \sim N(\delta, v). \tag{14}$$

Following Appendix A.1 in Zahra *et al.* (2017 a) with aforementioned local alternative and using equation (14), the asymptotic distribution of $\hat{\boldsymbol{\beta}}^{UE}$ is

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta}) \sim N(\mathbf{0}, \boldsymbol{V}). \tag{15}$$

3. Asymptotic Results

In this section, the analytical results regarding the asymptotic properties of the aforementioned estimators are presented. Using the asymptotic framework cited in Section 3 of Shah *et al.* (2020), and under the sequence of local alternatives given in equation (13), following evaluation criterion are used to assess the performance of the estimators under consideration. We have omitted the special cases of estimators (as mentioned above) from discussions in order to save space, as one can deduce their results by using their relations.

3.1. Asymptotic distributional bias

The vector of asymptotic distributional bias (ADB) of an estimator $\hat{\boldsymbol{\beta}}^{*}$ is calculated as

$$\mathfrak{B}(\hat{\boldsymbol{\beta}}^{*}) = \lim_{n\to\infty} \mathbb{E}\left(\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{*} - \boldsymbol{\beta}_{(n)}\right)\right). \tag{16}$$

Theorem 1: Expressions for ADB of various estimators under sequence of local alternatives are given as

$$\begin{aligned}
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{LS}) &= -\pi\boldsymbol{\delta}^{\star},\\
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{SP}) &= -\pi\boldsymbol{\delta}^{\star}\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta),\\
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) &= -(q-3)\boldsymbol{\delta}^{\star}\mathbb{E}\left(\chi^{-2}_{q+1}(\Delta)\right),\\
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{S+}) &= -\boldsymbol{\delta}^{\star}[\Phi_{q+1}(q-3;\Delta) + (q-3)\mathbb{E}\{\chi^{-2}_{q+1}(\Delta)I(\chi^2_{q+1}(\Delta) > (q-3))\}],
\end{aligned}$$

where $\boldsymbol{\delta}^{\star} = \boldsymbol{C}_0\boldsymbol{\delta}$, $\boldsymbol{C}_0 = \boldsymbol{I}_q - \boldsymbol{H}_0$, and $\Phi_{q+1}(\cdot;\Delta)$ is the cumulative distribution function (CDF) of a non-central chi-square distribution with $q+1$ degrees of freedom and non-centrality parameter Δ.

Proof: See Appendix 1 for the proof. □

3.2. Asymptotic distributional quadratic bias

The expressions given in Theorem 1 are in vector form and for comparison purposes, we needed expressions in scalar form. Thus, we applied following quadratic transformation that yielded asymptotic distributional quadratic bias (ADQB) of the competing estimators:

$$\mathfrak{B}^{\star}(\hat{\boldsymbol{\beta}}^{*}) = \left(\mathfrak{B}(\hat{\boldsymbol{\beta}}^{*})\right)^{T} \boldsymbol{V}^{-1}\mathfrak{B}(\hat{\boldsymbol{\beta}}^{*}). \tag{17}$$

Theorem 2: Under the sequence of local alternatives, expressions for ADQB of various estimators are given as

$$\begin{aligned}
\mathfrak{B}^{\star}(\hat{\boldsymbol{\beta}}^{LS}) &= \pi^2\Delta,\\
\mathfrak{B}^{\star}(\hat{\boldsymbol{\beta}}^{SP}) &= \pi^2\Delta\left[\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta)\right]^2,\\
\mathfrak{B}^{\star}(\hat{\boldsymbol{\beta}}^{SS}) &= (q-3)^2\Delta\left[\mathbb{E}\left\{\chi^{-2}_{q+1}(\Delta)\right\}\right]^2,\\
\mathfrak{B}^{\star}(\hat{\boldsymbol{\beta}}^{S+}) &= \Delta\left[\Phi_{q+1}(q-3;\Delta) + (q-3)\mathbb{E}\left\{\chi^{-2}_{q+1}(\Delta)I(\chi^2_{q+1}(\Delta) > (q-3))\right\}\right]^2.
\end{aligned}$$

Proof: Using the transformation mentioned in equation (17) and with the help of following Lemma, the proof is straightforward.

Lemma 3: The test statistic given in equation (9) converges to a non-central χ^2 distribution with $(q-1)$ degrees of freedom and non-centrality parameter $\Delta = (\boldsymbol{\delta}^{\star})^T\boldsymbol{V}^{-1}\boldsymbol{\delta}^{\star}$ as $n \to \infty$. □

Note that the estimators based on the strategies of preliminary test and Stein-type shrinkage methodologies are biased.

3.3. Asymptotic mean square error matrix

The asymptotic mean-square error matrices are needed for the computations of asymptotic distributional quadratic risk (ADQR) expressions of the various estimators mentioned above. Under the sequence of local alternative, the general expression for such matrices is given as

$$\mathfrak{S}(\hat{\boldsymbol{\beta}}^{*}) = \lim_{n\to\infty}\mathbb{E}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{*} - \boldsymbol{\beta}_{(n)}\right)\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{*} - \boldsymbol{\beta}_{(n)}\right)^{T}\right]. \tag{18}$$

Theorem 4: The expressions for the asymptotic mean square error matrices for various estimators, under the sequence of local alternatives and for $\boldsymbol{C} = \boldsymbol{V}\boldsymbol{C}_0^T$, are given as follows:

$$
\begin{aligned}
\mathfrak{S}(\hat{\boldsymbol{\beta}}^{LS}) &= \boldsymbol{V} - \pi(2-\pi)\boldsymbol{C} + \pi^2\boldsymbol{\delta}^\star(\boldsymbol{\delta}^\star)^T,\\
\mathfrak{S}(\hat{\boldsymbol{\beta}}^{SP}) &= \boldsymbol{V} - \pi(2-\pi)\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta)\boldsymbol{C}\\
&\quad + \pi\boldsymbol{\delta}^\star(\boldsymbol{\delta}^\star)^T\left[2\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta) - (2-\pi)\Phi_{q+3}(\chi^2_{q-1,\alpha};\Delta)\right],\\
\mathfrak{S}(\hat{\boldsymbol{\beta}}^{SS}) &= \boldsymbol{V} - (q-3)\boldsymbol{C}\left[2\mathbb{E}\left[\chi^{-2}_{q+1}(\Delta)\right] - (q-3)\mathbb{E}\left[\chi^{-4}_{q+1}(\Delta)\right]\right]\\
&\quad + (q-3)(q+1)\boldsymbol{\delta}^\star(\boldsymbol{\delta}^\star)^T\mathbb{E}\left[\chi^{-4}_{q+3}(\Delta)\right],\\
\mathfrak{S}(\hat{\boldsymbol{\beta}}^{S+}) &= \mathfrak{S}(\hat{\boldsymbol{\beta}}^{SS}) - \boldsymbol{C}\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+1}(\Delta)\right\}^2 I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\right]\\
&\quad + \boldsymbol{\delta}^\star(\boldsymbol{\delta}^\star)^T\begin{bmatrix} 2\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+1}(\Delta)\right\} I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\right] \\ -\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+3}(\Delta)\right\}^2 I\left(\chi^2_{q+3}(\Delta) < (q-3)\right)\right]\end{bmatrix},
\end{aligned}
$$

Proof: Following Appendix 7 of Zahra *et al.* (2017 a), the proof can be completed. □

3.4. Asymptotic distributional quadratic risk

The asymptotic distributional quadratic risk (ADQR) of an estimator $\hat{\boldsymbol{\beta}}^\star$ of the parameter vector $\boldsymbol{\beta}$ is defined as

$$
\begin{aligned}
\mathfrak{R}(\hat{\boldsymbol{\beta}}^\star;\boldsymbol{W}) &= \lim_{n\to\infty}\mathbb{E}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)^T\boldsymbol{W}\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)\right]\\
&= \lim_{n\to\infty}\mathrm{tr}\left[\boldsymbol{W}\mathbb{E}\left\{\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)^T\right\}\right] = \mathrm{tr}\left[\boldsymbol{W}\mathfrak{S}(\hat{\boldsymbol{\beta}}^*)\right], \qquad (19)
\end{aligned}
$$

where $\boldsymbol{W}$ is a $(q\times q)$ positive semi-definite (psd) weight matrix.

Theorem 5: Under the sequence of local alternatives, expressions of ADQR for the various estimators are given as

$$
\begin{aligned}
\mathfrak{R}(\hat{\boldsymbol{\beta}}^{LS};\boldsymbol{W}) &= \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{V}\right) - \pi(2-\pi)\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{C}\right) + \pi^2\Delta_W,\\
\mathfrak{R}(\hat{\boldsymbol{\beta}}^{SP};\boldsymbol{W}) &= \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{V}\right) - \pi(2-\pi)\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{C}\right)\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta)\\
&\quad + \pi\Delta_W\left[2\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta) - (2-\pi)\Phi_{q+3}(\chi^2_{q-1,\alpha};\Delta)\right],\\
\mathfrak{R}(\hat{\boldsymbol{\beta}}^{SS};\boldsymbol{W}) &= \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{V}\right) - (q-3)\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{C}\right)\left[2\mathbb{E}\left[\chi^{-2}_{q+1}(\Delta)\right] - (q-3)\mathbb{E}\left[\chi^{-4}_{q+1}(\Delta)\right]\right]\\
&\quad + (q-3)(q+1)\Delta_W\mathbb{E}\left[\chi^{-4}_{q+3}(\Delta)\right],\\
\mathfrak{R}(\hat{\boldsymbol{\beta}}^{S+};\boldsymbol{W}) &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{SS};\boldsymbol{W})\\
&\quad - \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{C}\right)\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+1}(\Delta)\right\}^2 I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\right]\\
&\quad + \Delta_W\begin{bmatrix} 2\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+1}(\Delta)\right\} I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\right] \\ -\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+3}(\Delta)\right\}^2 I\left(\chi^2_{q+3}(\Delta) < (q-3)\right)\right]\end{bmatrix},
\end{aligned}
$$

where $\Delta_W = (\boldsymbol{\delta}^\star)^T \boldsymbol{W} \boldsymbol{\delta}^\star$ is the non-centrality parameter involving weight matrix $\boldsymbol{W}$.

Proof: Using the asymptotic mean square matrices given in Theorem 4 and the definition of ADQR in equation (19), the proof is trivial. □

Corollary 6: For the choice $\boldsymbol{W} = \boldsymbol{V}^{-1}$, ADQR expressions given above are simplified and are given as

$$
\begin{aligned}
R_1 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{V}^{-1}) = q, \\
R_2 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{RE}; \boldsymbol{V}^{-1}) = 1 + \Delta, \\
R_3 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{LS}; \boldsymbol{V}^{-1}) = q - \pi(2-\pi)(q-1) + \pi^2 \Delta, \\
R_4 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{PT}; \boldsymbol{V}^{-1}) = q - (q-1)\Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta) \\
&\qquad + \Delta \left[2\Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta) - \Phi_{q+3}(\chi^2_{q-1,\alpha}; \Delta) \right], \\
R_5 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{SP}; \boldsymbol{V}^{-1}) = q - \pi(2-\pi)(q-1)\Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta) \\
&\qquad + \pi\Delta \left[2\Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta) - (2-\pi)\Phi_{q+3}(\chi^2_{q-1,\alpha}; \Delta) \right], \\
R_6 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{SS}; \boldsymbol{V}^{-1}) = q - (q-1)(q-3)\Big[2\mathbb{E}\left[\chi^{-2}_{q+1}(\Delta) \right] - (q-3)\mathbb{E}\left[\chi^{-4}_{q+1}(\Delta) \right] \Big] \\
&\qquad + (q-3)(q+1)\Delta \mathbb{E}\left[\chi^{-4}_{q+3}(\Delta) \right], \\
R_7 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{S+}; \boldsymbol{V}^{-1}) = \mathfrak{R}(\hat{\boldsymbol{\beta}}^{SS}; \boldsymbol{V}^{-1}) \\
&\qquad - (q-1)\mathbb{E}\Big[\left\{ 1 - (q-3)\chi^{-2}_{q+1}(\Delta) \right\}^2 I\left(\chi^2_{q+1}(\Delta) < (q-3) \right) \Big] \\
&\qquad + \Delta \begin{bmatrix} 2\mathbb{E}\left[\left\{ 1 - (q-3)\chi^{-2}_{q+1}(\Delta) \right\} I\left(\chi^2_{q+1}(\Delta) < (q-3) \right) \right] \\ -\mathbb{E}\left[\left\{ 1 - (q-3)\chi^{-2}_{q+3}(\Delta) \right\}^2 I\left(\chi^2_{q+3}(\Delta) < (q-3) \right) \right] \end{bmatrix}.
\end{aligned}
$$

Proof: This proof of Corollary 6 can be completed by replacing $\boldsymbol{W} = \boldsymbol{V}^{-1}$ and noting that $\mathrm{tr}(\boldsymbol{W}\boldsymbol{V}) = \mathrm{tr}(\boldsymbol{I}_q) = q$, $\mathrm{tr}(\boldsymbol{W}\boldsymbol{C}) = \mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{C}_0^T) = \mathrm{tr}(\boldsymbol{C}_0^T) = \mathrm{tr}(\boldsymbol{I}_q - \boldsymbol{H}_0)^T = \mathrm{tr}(\boldsymbol{I}_q - \boldsymbol{H}_0) = q - 1$, and $\Delta_W = (\boldsymbol{\delta}^\star)^T \boldsymbol{V} \boldsymbol{\delta}^\star = \Delta$. □

3.5. Risk comparison

In this section, the performance of restricted (RE), linear shrinkage (LS), both preliminary test (PT and SP) and Stein-type shrinkage estimators (SS and S+) is compared with the benchmark unrestricted estimator (UE) using the aforementioned simplified ADQR expressions of Corollary 6. For this purpose, we have defined the notion of asymptotic relative efficiency (AREFF) of an estimator $\hat{\boldsymbol{\beta}}^*$ with reference to $\hat{\boldsymbol{\beta}}^{UE}$ as

$$
\mathrm{AREFF}(\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\beta}}^{UE}) = \frac{\mathfrak{R}(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{V}^{-1})}{\mathfrak{R}(\hat{\boldsymbol{\beta}}^*; \boldsymbol{V}^{-1})} = \frac{R_1}{R_j}; \quad j \leq 7. \tag{20}
$$

An estimator is considered to be more efficient, in asymptotic terms, if AREFF exceeds 1, and vice versa. Since all the risk expressions are the function of a drift parameter Δ, we

have plotted the AREFF of all the competing estimators against Δ in order to compare their performance in Figures 1–4, while fixing $q = 4, 8, 10$, $\pi = 0.50$, and $\alpha = 0.01, 0.05, 0.10, 0.20$.

From Figure 1, it can be seen that both $\hat{\boldsymbol{\beta}}^{RE}$ and $\hat{\boldsymbol{\beta}}^{LS}$ have higher efficiencies than $\hat{\boldsymbol{\beta}}^{UE}$ (which is constant), and that the AREFF decreases as Δ increases but the decay of $\hat{\boldsymbol{\beta}}^{RE}$ is much faster than that of $\hat{\boldsymbol{\beta}}^{LS}$.

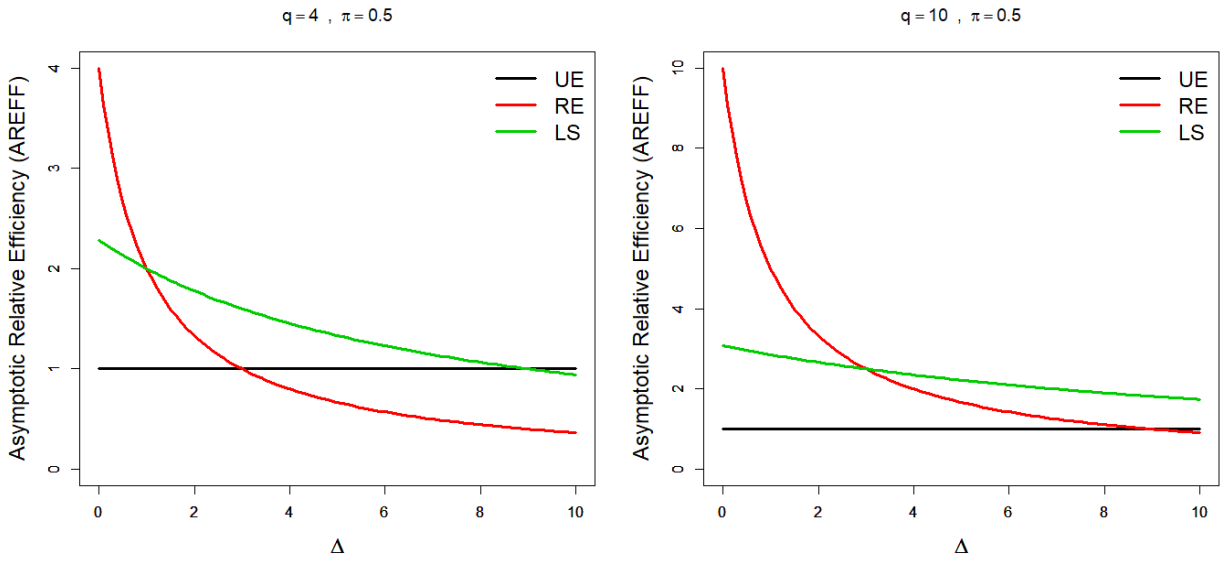


Figure 1: Asymptotic relative efficiencies of $\hat{\boldsymbol{\beta}}^{UE}$, $\hat{\boldsymbol{\beta}}^{RE}$, and $\hat{\boldsymbol{\beta}}^{LS}$

AREFFs of the unrestricted estimator and the preliminary test estimators are compared in Figure 2. It can be observed that the $\hat{\boldsymbol{\beta}}^{PT}$ dominates the $\hat{\boldsymbol{\beta}}^{SP}$ in a region where the drift parameter Δ is smaller *i.e.*, when the null hypothesis (3) is true or nearly true. But for larger values of Δ, the situation is reversed, and it is $\hat{\boldsymbol{\beta}}^{SP}$ that outperforms $\hat{\boldsymbol{\beta}}^{PT}$ for all choices of α. Moreover, curves of AREFFs of $\hat{\boldsymbol{\beta}}^{PT}$ are approaching $\hat{\boldsymbol{\beta}}^{SP}$ for larger values of α. This means that for smaller values of α, the region where $\hat{\boldsymbol{\beta}}^{PT}$ dominates $\hat{\boldsymbol{\beta}}^{SP}$ is more spacious.

From Figure 3, it is evident that $\hat{\boldsymbol{\beta}}^{S+}$ is far superior to $\hat{\boldsymbol{\beta}}^{SS}$ and $\hat{\boldsymbol{\beta}}^{UE}$, uniformly dominating both estimators. Figure 4 establishes the dominance of the restricted estimator $\hat{\boldsymbol{\beta}}^{RE}$ over the Stein-type shrinkage estimator $\hat{\boldsymbol{\beta}}^{SS}$ and shrinkage preliminary test estimator $\hat{\boldsymbol{\beta}}^{SP}$ for smaller values of Δ. Even the shrinkage preliminary test estimator $\hat{\boldsymbol{\beta}}^{SP}$ performs better than $\hat{\boldsymbol{\beta}}^{SS}$ in a reasonable region over Δ.

4. Monte-Carlo Simulations

In this section, we conducted extensive Monte-Carlo simulations to examine the performance of the various estimators discussed earlier for $\boldsymbol{\beta}$, which incorporates UPI into the estimation procedure. The performance of the estimators is investigated by comparing their simulated relative efficiencies (SRE). The SRE of an estimator $\hat{\boldsymbol{\beta}}^{*}$ to a benchmark unre-

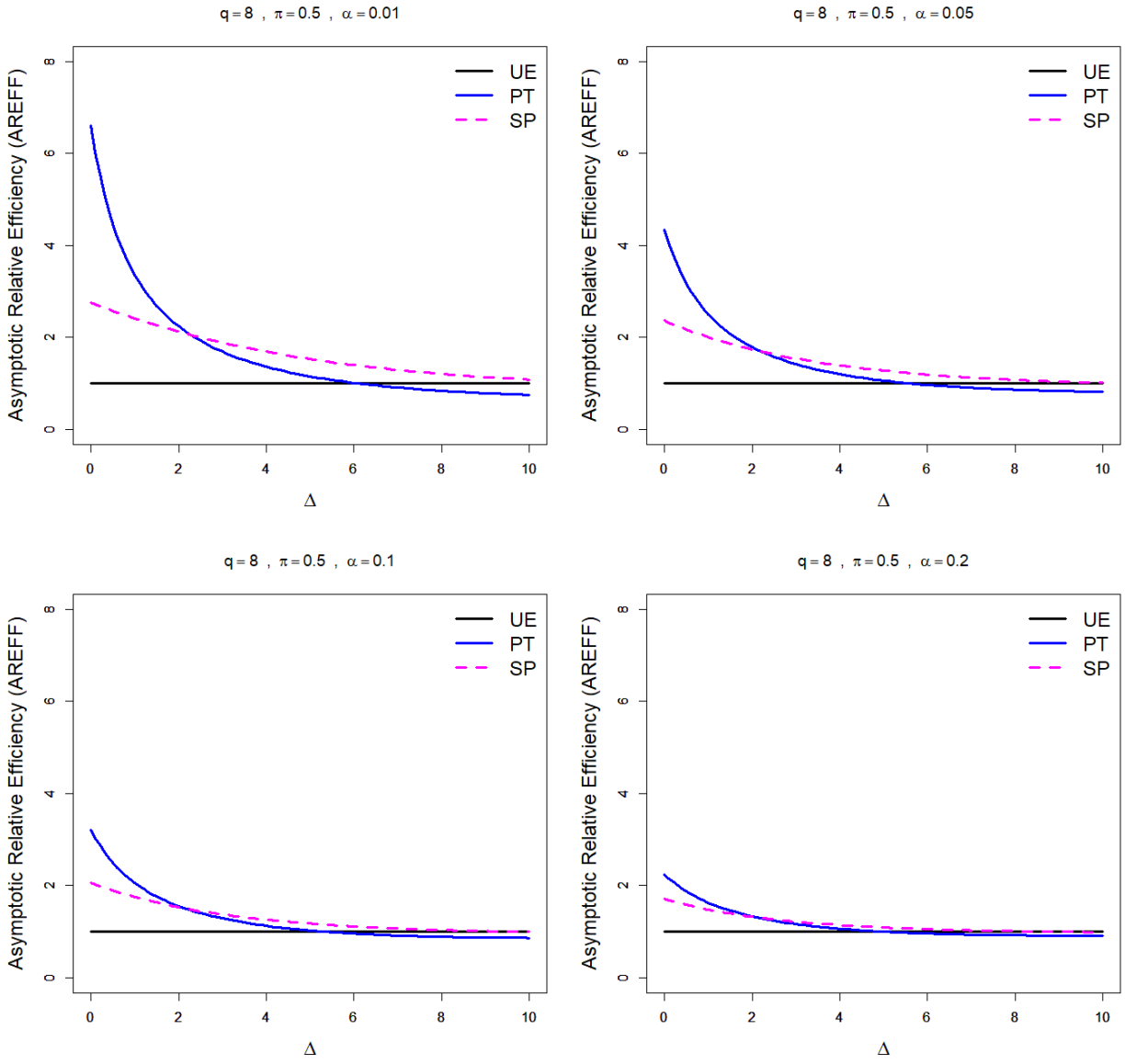


Figure 2: Asymptotic relative efficiencies of $\hat{\boldsymbol{\beta}}^{UE}$, $\hat{\boldsymbol{\beta}}^{PT}$, and $\hat{\boldsymbol{\beta}}^{SP}$

stricted estimator $\hat{\boldsymbol{\beta}}^{UE}$ is defined by the ratio of their simulated risks as

$$\text{SRE}(\hat{\boldsymbol{\beta}}^{*}, \hat{\boldsymbol{\beta}}^{UE}) = \frac{\text{Simulated Risk}(\hat{\boldsymbol{\beta}}^{UE})}{\text{Simulated Risk}(\hat{\boldsymbol{\beta}}^{*})}. \tag{21}$$

The value of an SRE greater than 1 indicates that $\hat{\boldsymbol{\beta}}^{*}$ is superior to $\hat{\boldsymbol{\beta}}^{UE}$. Furthermore, we have defined a parameter Δ^{*} (which is essentially a measure of how far away we deviate from the hypothesized common kurtosis vector $\boldsymbol{\beta_0}$) as $\Delta^{*} = (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{T} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)$.

Ahmed *et al.* (2012) mentioned comparisons for kurtosis of a multivariate normal distribution with t-distribution with degrees of freedom ranging from $\nu = 5$ to $\nu = 60$. Fol-

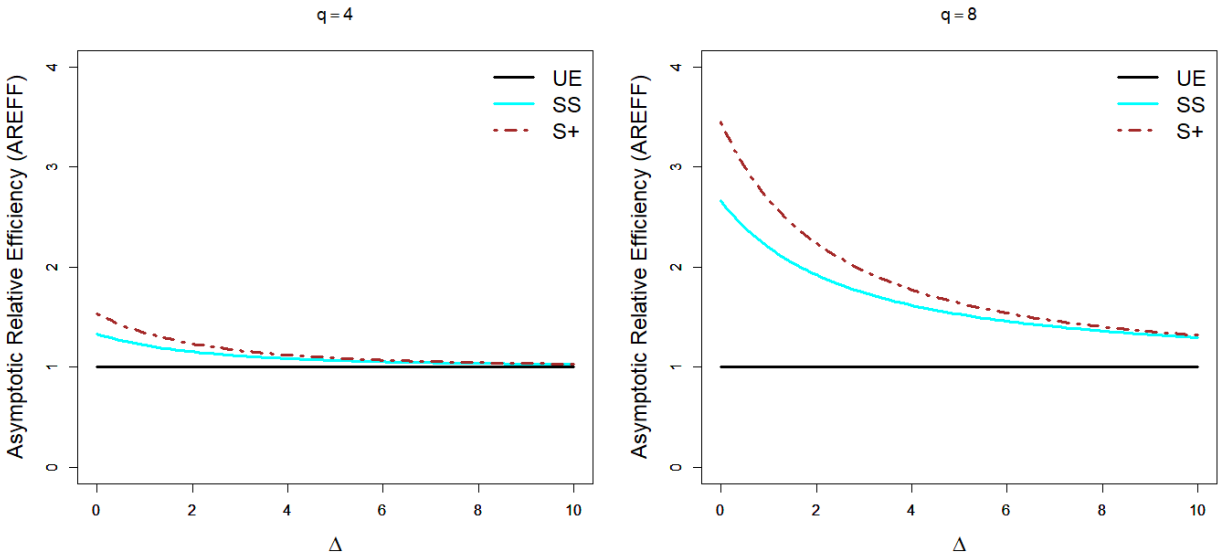


Figure 3: Asymptotic relative efficiencies of $\hat{\boldsymbol{\beta}}^{UE}$, $\hat{\boldsymbol{\beta}}^{SS}$, and $\hat{\boldsymbol{\beta}}^{S+}$

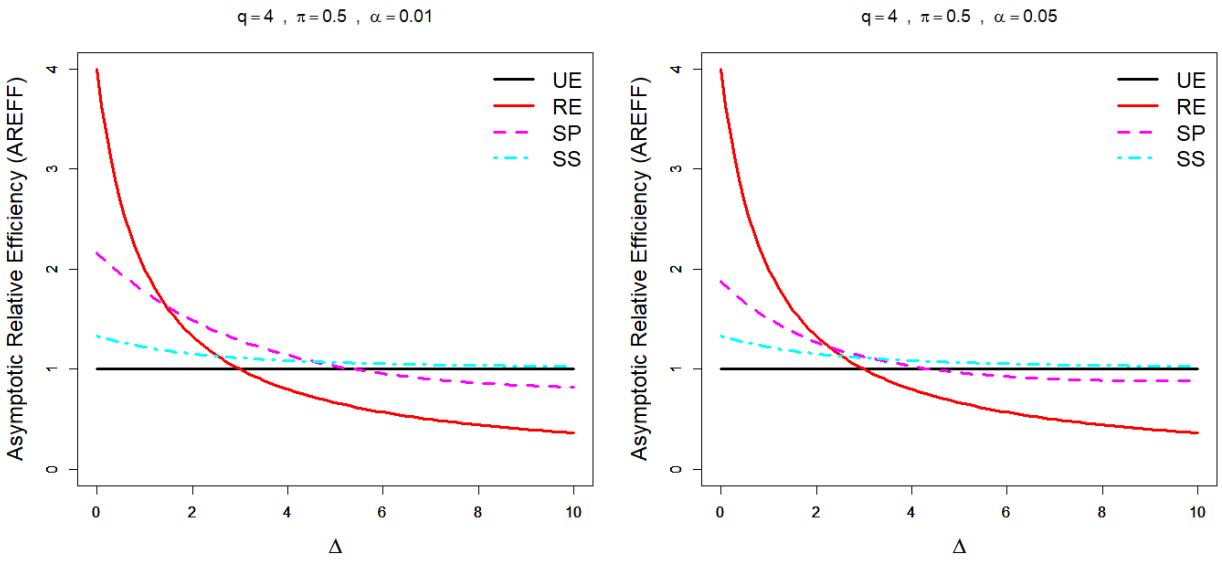


Figure 4: Asymptotic relative efficiencies of $\hat{\boldsymbol{\beta}}^{RE}$, $\hat{\boldsymbol{\beta}}^{SP}$, and $\hat{\boldsymbol{\beta}}^{SS}$

lowing the same idea, we have considered multivariate normal distributions with dimensions $p = 2$ and 4 initially, when the null hypothesis is assumed to be true. In order to study the deviation of the data from the null hypothesis, further samples are taken from multivariate t-distribution with various degrees of freedom between 5 and 60.

Simulation processes are repeated $N = 5000$ times for different choices of $q = 4$, 6, 8,

10 and $n_l = 80, 100, 150, 200$. The choices of α and π are assumed to be 0.05, 0.10 and 0.25, 0.50, 0.75, respectively. Simulated relative efficiencies of the suggested estimators for different configurations of p, q, n_l, α, and π against various values of $\Delta^* \geq 0$ were computed; results are reported in Table 1 for only $p = 4$, $\pi = 0.50$ and $n_l = 100$ to conserve space. A graphical representation is shown in Figure 5.

Table 1: Simulated relative efficiencies of estimators when $p = 4$, $\pi = 0.50$ and $n_l = 100$

q	Δ^*	$\hat{\boldsymbol{\beta}}^{RE}$	$\hat{\boldsymbol{\beta}}^{LS}$	$\hat{\boldsymbol{\beta}}^{PT}$		$\hat{\boldsymbol{\beta}}^{SP}$		$\hat{\boldsymbol{\beta}}^{SS}$	$\hat{\boldsymbol{\beta}}^{S+}$
				$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$		
4	**0.00**	4.00	2.29	2.41	2.00	1.78	1.60	1.35	1.50
	0.58	2.52	2.02	1.60	1.43	1.46	1.34	1.23	1.32
	1.30	1.22	1.55	1.02	1.01	1.12	1.09	1.09	1.13
	1.89	0.82	1.25	0.86	0.89	0.99	0.99	1.02	1.04
	3.46	0.51	0.89	0.81	0.86	0.91	0.94	0.98	0.99
	10.39	0.45	0.69	0.97	0.98	0.99	0.99	0.99	0.99
6	**0.00**	5.91	2.65	3.03	2.39	2.01	1.77	1.86	2.33
	0.50	3.17	2.28	1.97	1.74	1.68	1.54	1.58	1.90
	1.12	1.58	1.86	1.16	1.12	1.24	1.18	1.29	1.37
	1.63	0.96	1.42	0.92	0.93	1.06	1.04	1.09	1.14
	2.98	0.53	0.93	0.80	0.84	0.92	0.94	0.99	0.99
	8.94	0.43	0.69	0.97	0.98	0.98	0.99	0.99	0.99
8	**0.00**	7.84	2.89	3.57	2.77	2.17	1.92	2.43	3.15
	0.44	2.98	2.08	1.62	1.45	1.42	1.32	1.72	1.79
	0.99	1.66	1.85	1.27	1.22	1.34	1.26	1.39	1.55
	1.44	1.07	1.52	0.98	0.98	1.12	1.09	1.18	1.24
	2.65	0.57	0.99	0.81	0.85	0.93	0.95	0.99	0.99
	7.94	0.42	0.68	0.96	0.97	0.98	0.99	0.99	0.99
10	**0.00**	9.91	3.07	3.85	2.97	2.25	1.99	3.05	3.94
	0.40	4.02	2.41	1.88	1.63	1.57	1.43	2.14	2.28
	0.90	2.23	2.05	1.36	1.27	1.32	1.24	1.68	1.74
	1.31	1.30	1.72	1.08	1.06	1.20	1.15	1.34	1.41
	2.40	0.61	1.04	0.83	0.86	0.95	0.96	0.99	0.99
	7.20	0.42	0.68	0.96	0.97	0.98	0.99	0.99	0.99

The restricted estimator $\hat{\boldsymbol{\beta}}^{RE}$ performed better when homogeneity assumption of kurtosis parameters holds, but when it is not true, the SRE of the restricted estimator declined rapidly, and approaches zero for larger values of Δ^*. The SRE of the linear shrinkage estimator $\hat{\boldsymbol{\beta}}^{LS}$ declined slowly and its performance is comparable to shrinkage estimators only when q and Δ^* are small. The SRE of the pretest $\hat{\boldsymbol{\beta}}^{PT}$ and shrinkage pretest $\hat{\boldsymbol{\beta}}^{SP}$ estimators declined as Δ^* increased, but after reaching a minimum value, it attained a value of one again. Both pretest estimators performed well for smaller values of $q \leq 4$ and Δ^*. The Stein-type estimators $\hat{\boldsymbol{\beta}}^{SS}$ and $\hat{\boldsymbol{\beta}}^{S+}$ performed better than all other suggested estimators in the wider range of Δ^*, especially as q increases. In short, simulation study endorsed the analytical findings.

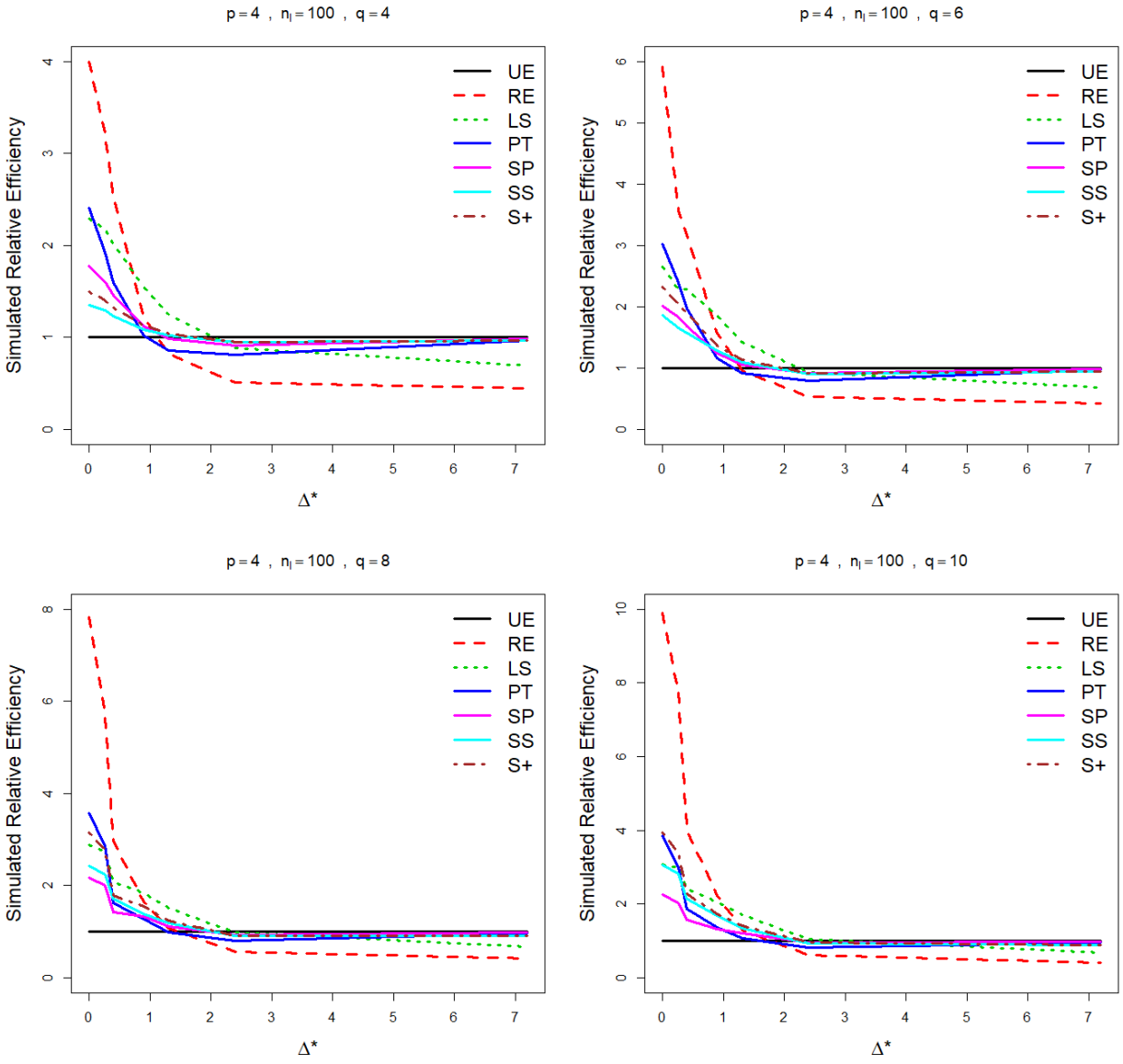


Figure 5: Simulated relative efficiency of the estimators

5. Data Application

Our real-data application considered the four-dimensional multivariate data of geographical regions of Europe, based on monthly long-term interest rates in percentages. Four countries of each region are considered; Central Europe: Austria, Germany, Hungry, and Poland; Southern Europe: Spain, Italy, Portugal, and Slovenia; Western Europe: Belgium, France, Netherlands, and Switzerland; Northern Europe: United Kingdom, Norway, Denmark, and Sweden. The data comprises of 100 observations for each country from August 2007 to November 2015, as reported by the Organisation for Economic Co-operation and Development (OECD) website.

The unrestricted estimator is obtained as $\hat{\boldsymbol{\beta}}^{UE} = (25.16, 27.20, 28.17, 20.45)^T$. Next, we want to test the hypothesis that the kurtosis parameters are same for all four regions, against the alternative that at least one of them is different from others. The test statistic $\mathfrak{L}_n$ is 18.42, therefore, we reject the null hypothesis at $\alpha = 0.05$. The suggested estimators are calculated for $\pi = 0.50$ and $\alpha = 0.05$ as $\hat{\boldsymbol{\beta}}^{RE} = (25.24, 25.24, 25.24, 25.24)^T$; the pretest, shrinkage pretest and Stein-type estimators are equal to the unrestricted estimator, while linear shrinkage and improved Stein-type shrinkage estimators are given as $\hat{\boldsymbol{\beta}}^{LS} = (25.20, 26.22, 26.71, 22.85)^T$ and $\hat{\boldsymbol{\beta}}^{S+} = (25.17, 27.10, 28.01, 20.71)^T$.

Bootstrap methodology is used to assess the performance of the suggested estimators. Samples of equal size $n_l = 100$ are selected from each country with replacement; this process is repeated $N = 5000$ times. All the suggested estimators are computed for $\pi = 0.25, 0.50, 0.75$ and $\alpha = 0.05, 0.10, 0.30$. Simulated relative efficiencies of the estimators under consideration relative to the unrestricted estimator are computed and reported in the following table:

Table 2: Relative efficiencies of estimators for long-term interest rates data based on multivariate bootstrap samples

α	π	$\hat{\boldsymbol{\beta}}^{RE}$	$\hat{\boldsymbol{\beta}}^{LS}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{SP}$	$\hat{\boldsymbol{\beta}}^{SS}$	$\hat{\boldsymbol{\beta}}^{S+}$
	0.25	0.54	1.23	0.93	0.99	1.01	1.01
0.05	0.50	0.54	1.11	0.93	0.97	1.01	1.01
	0.75	0.54	0.80	0.93	0.95	1.01	1.01

It is revealed from the above table that the SRE of $\hat{\boldsymbol{\beta}}^{RE}$ is less than 1, which is in line with the analytical and simulated results; this confirms when the null hypothesis is not true, the restricted estimator performs inferior to all other estimators. The SRE of $\hat{\boldsymbol{\beta}}^{LS}$ declines as the value of π increases. $\hat{\boldsymbol{\beta}}^{PT}$ and $\hat{\boldsymbol{\beta}}^{SP}$ both have SREs smaller than 1 and their SREs remain below 1 as π increases. However, we recommend using the positive part of the Stein-type shrinkage estimator, since its performance is not drastically impacted by departure from the null hypothesis.

6. Concluding Remarks

In this paper, we have discussed the asymptotic theory of simultaneous estimation of kurtosis parameters for q multivariate normal distributions using the UPI that all kurtosis parameters are homogeneous. It is concluded that the performance of restricted and pretest estimators is better when the null hypothesis of equal kurtosis parameters holds, while the risk of the restricted estimator becomes unbounded as we move away from the null hypothesis. The pretest estimator performs better than restricted and Stein-type shrinkage estimators but only in a certain region of the parametric space. Stein-type estimators, outperform the unrestricted estimator in the entire parametric space. The improved Stein-type shrinkage estimator is strongly recommended when the equality of parameters is uncertain for $q \geq 4$, while for small dimensions, shrinkage pretest estimator is a better choice.

Acknowledgements

The research work of Professor S. Ejaz Ahmed was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Ahmed, S. E. (2002). Simultaneous estimation of coefficients of variation. *Journal of Statistical Planning and Inference*, **104(10)**, 31–51.

Ahmed, S. E., Omar, M. H. and Joarder, A. H. (2012). Stabilizing the performance of kurtosis estimator of multivariate data. *Communications in Statistics - Simulation and Computation*, **41(10)**, 1860–1871.

Araújo, T., Dias, J., Eleutério, S. and Louçã, F. (2012). *How Fama Went Wrong: Measures of Multivariate Kurtosis for the Identification of the Dynamics of a N-Dimensional Market*. Working Papers WP21/2012/DE/UECE, Department of Economics, School of Economics and Management, Technical University of Lisbon, Portugal.

Echer, E. and Bolzan, M. J. A. (2016). A comparative study of solar wind and foreshock turbulence near Uranus orbit. *Planetary and Space Science*, **120**, 70–77.

Kim, T. H. and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, **1(1)**, 56–73.

Lai, J. Y. (2012). An empirical study of the impact of skewness and kurtosis on hedging decisions. *Quantitative Finance*, **12(12)**, 1827–1837.

Liang, Z., Wei, J., Zhao, J., Liu, H., Li, B., Shen, J. and Zheng, C. (2008). The statistical meaning of kurtosis and its new application to identification of persons based on seismic signals. *Sensors*, **8(8)**, 5106–5119.

Lisawadi, S., Ahmed, S. E., Reangsephet, O. and Shah, M. K. A. (2019). Simultaneous estimation of Cronbach's alpha coefficients. *Communications in Statistics - Theory and Methods*, **48(13)**, 3236–3257.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.

Nita, G. M. and Gary, D. E. (2010). Statistics of the spectral kurtosis estimator. *Publications of the Astronomical Society of the Pacific*, **122(891)**, 595–607.

Shah, M. K. A., Lisawadi, S. and Ahmed, S. E. (2017). Merging data from multiple sources: pretest and shrinkage perspectives. *Journal of Statistical Computation and Simulation*, **87(8)**, 1577–1592.

Shah, M. K. A., Zahra, N. and Ahmed, S. E. (2020). On the simultaneous estimation of Weibull reliability functions. In: *Proceedings of the Thirteenth International Conference on Management Science and Engineering Management. ICMSEM 2019*, *Advances in Intelligent Systems and Computing*, **1001**, 85–108. Springer, Cham.

Zahra, N., Lisawadi, S. and Ahmed, S. E. (2017 a). Meta-analysis, pretest, and shrinkage estimation of kurtosis parameters. *Communications in Statistics - Simulation and Computation*, **46(10)**, 7986–8004.

Zahra, N., Lisawadi, S. and Ahmed, S. E. (2017 b). Improved estimation of kurtosis parameters for two multivariate populations. *Lobachevskii Journal of Mathematics*, **38(1)**, 110–115.

APPENDIX
Mathematical Proofs

Two important results cited in Lemma 3 of Shah *et al.* (2020) are used, as well as some distributional results crucial for the derivation of mathematical results of appendices, are given in the following theorem:

Theorem 7: Under the sequence of local alternatives $\{\mathcal{H}_n\}$ and assuming independence among q components, following distributional result holds:

$$\begin{aligned}
\boldsymbol{\eta}_{1,n} &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta}\right) \xrightarrow{D} \boldsymbol{\eta}_1 \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{V}),\\
\boldsymbol{\eta}_{2,n} &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta}_0\right) \xrightarrow{D} \boldsymbol{\eta}_2 \sim \mathcal{N}_q(\boldsymbol{\delta}, \boldsymbol{V}),\\
\boldsymbol{\eta}_{3,n} &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta}_0\right) \xrightarrow{D} \boldsymbol{\eta}_3 \sim \mathcal{N}_q(\mathbf{0}, v\boldsymbol{J}_q),\\
\boldsymbol{\eta}_{4,n} &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right) \xrightarrow{D} \boldsymbol{\eta}_4 \sim \mathcal{N}_q(\boldsymbol{\delta}^\star, \boldsymbol{C}),\\
\begin{pmatrix}\boldsymbol{\eta}_{2,n}\\ \boldsymbol{\eta}_{4,n}\end{pmatrix} &\xrightarrow{D} \begin{pmatrix}\boldsymbol{\eta}_2\\ \boldsymbol{\eta}_4\end{pmatrix} \sim \mathcal{N}_{2q}\left\{\begin{pmatrix}\boldsymbol{\delta}\\ \boldsymbol{\delta}^\star\end{pmatrix}, \begin{pmatrix}\boldsymbol{V} & \boldsymbol{C}\\ \boldsymbol{C}^T & \boldsymbol{C}\end{pmatrix}\right\},\\
\begin{pmatrix}\boldsymbol{\eta}_{3,n}\\ \boldsymbol{\eta}_{4,n}\end{pmatrix} &\xrightarrow{D} \begin{pmatrix}\boldsymbol{\eta}_3\\ \boldsymbol{\eta}_4\end{pmatrix} \sim \mathcal{N}_{2q}\left\{\begin{pmatrix}\mathbf{0}\\ \boldsymbol{\delta}^\star\end{pmatrix}, \begin{pmatrix}v\boldsymbol{J}_q & \mathbf{0}\\ \mathbf{0} & \boldsymbol{C}\end{pmatrix}\right\},
\end{aligned}$$

where $\xrightarrow{D}$ means convergence in distribution as $n \to \infty$.

Proof: See Appendices A1–A6 of Zahra *et al.* (2017 a) for detailed proof with some adjustments in notations. □

A1. Proof of Theorem 1

$$\begin{aligned}
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{LS}) &= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}(\hat{\boldsymbol{\beta}}^{LS} - \boldsymbol{\beta}_{(n)})\right] = \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{UE} - \pi(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \boldsymbol{\beta}_0 - \frac{1}{\sqrt{n}}\boldsymbol{\delta}\right\}\right]\\
&= \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{2,n} - \boldsymbol{\delta} - \pi\boldsymbol{\eta}_{4,n}\right] = \mathbb{E}(\boldsymbol{\eta}_2) - \boldsymbol{\delta} - \pi\mathbb{E}(\boldsymbol{\eta}_4) = -\pi\boldsymbol{\delta}^\star.
\end{aligned}$$

$$\begin{aligned}
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{SP}) &= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}(\hat{\boldsymbol{\beta}}^{SP} - \boldsymbol{\beta}_{(n)})\right]\\
&= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{UE} - \pi(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})I(\mathfrak{L}_n < c_{n,\alpha}) - \boldsymbol{\beta}_0 - \frac{1}{\sqrt{n}}\boldsymbol{\delta}\right\}\right]\\
&= \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{2,n} - \boldsymbol{\delta} - \pi\boldsymbol{\eta}_{4,n}I(\mathfrak{L}_n < c_{n,\alpha})\right]\\
&= \mathbb{E}(\boldsymbol{\eta}_2) - \boldsymbol{\delta} - \pi\mathbb{E}[\boldsymbol{\eta}_4 I(\chi^2_{q-1}(\Delta) < \chi^2_{q-1,\alpha})] = -\pi\mathbb{E}[\boldsymbol{\eta}_4 I(\chi^2_{q-1}(\Delta) < \chi^2_{q-1,\alpha})]\\
&= -\pi\boldsymbol{\delta}^\star\Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta).
\end{aligned}$$

$$\begin{aligned}
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) &= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}(\hat{\boldsymbol{\beta}}^{SS} - \boldsymbol{\beta}_{(n)})\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{UE} - (q-3)\mathfrak{L}_n^{-1}(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \boldsymbol{\beta}_0 - \frac{1}{\sqrt{n}}\boldsymbol{\delta}\right\}\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{2,n} - \boldsymbol{\delta} - (q-3)\boldsymbol{\eta}_{4,n}\mathfrak{L}_n^{-1}\right] = \mathbb{E}(\boldsymbol{\eta}_2) - \boldsymbol{\delta} - (q-3)\mathbb{E}[\boldsymbol{\eta}_4\chi_{q-1}^{-2}(\Delta)] \\
&= -(q-3)\boldsymbol{\delta}^{\star}\mathbb{E}[\chi_{q+1}^{-2}(\Delta)].
\end{aligned}$$

$$\begin{aligned}
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{S+}) &= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}(\hat{\boldsymbol{\beta}}^{S+} - \boldsymbol{\beta}_{(n)})\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{SS} - \left(1-(q-3)\mathfrak{L}_n^{-1}\right) I(\mathfrak{L}_n < (q-3))(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \frac{1}{\sqrt{n}}\boldsymbol{\delta}\right\}\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{SS} - \boldsymbol{\beta}_{(n)} - \left(1-(q-3)\mathfrak{L}_n^{-1}\right) I(\mathfrak{L}_n < (q-3))(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})\right\}\right] \\
&= \mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) - \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{4,n} I(\mathfrak{L}_n < (q-3))\right] + (q-3)\lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{4,n}\mathfrak{D}_n^{-1} I(\mathfrak{D}_n < (q-3))\right] \\
&= \mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) - \mathbb{E}\left[\boldsymbol{\eta}_4 I(\chi_{q-1}^2(\Delta) < (q-3))\right] + (q-3)\mathbb{E}\left[\boldsymbol{\eta}_4\chi_{q-1}^{-2}(\Delta) I(\chi_{q-1}^2(\Delta) < (q-3))\right] \\
&= \mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) - \boldsymbol{\delta}^{\star}\Phi_{q+1}(q-3;\Delta) + (q-3)\boldsymbol{\delta}^{\star}\mathbb{E}\left[\chi_{q+1}^{-2}(\Delta) I(\chi_{q+1}^2(\Delta) < (q-3))\right] \\
&= -\boldsymbol{\delta}^{\star}\left[\Phi_{q+1}(q-3;\Delta) + (q-3)\mathbb{E}[\chi_{q+1}^{-2}(\Delta) I(\chi_{q+1}^2(\Delta) > (q-3))]\right].
\end{aligned}$$

Statistics and Applications {ISSN 2452-7395 (online)}
Volume 18, No. 2, 2020 (New Series), pp 369–381

Statistical Validity and Consistency of Big Data Analytics: A General Framework

Bikram Karmakar[1] **and Indranil Mukhopadhyay**[2]
[1]*Department Statistics, University of Pennsylvania, PA, USA*
[2*]*Human Genetics Unit, Indian Statistical Institute, India*

Received: 31 July 2020; Revised: 30 September 2020; Accepted: 05 October 2020

Abstract

Informatics and technological advancements have triggered generation of huge volume of data with varied complexity in its management and analysis. Big Data analytics is the practice of revealing hidden aspects of such data and making inferences from it. Although storage, retrieval and management of Big Data seem possible through efficient algorithm and system development, concern about statistical consistency remains to be addressed in view of its specific characteristics. Since Big Data does not conform to standard analytics, we need proper modification of the existing statistical theory and tools. Here we propose, with illustrations, a general statistical framework and an algorithmic principle for Big Data analytics that ensure statistical accuracy of the conclusions. The proposed framework has the potential to push forward advancement of Big Data analytics in the right direction. The partition-repetition approach proposed here is broad enough to encompass all practical data analytic problems.

Key words: Big Data; Data mining; Partition-repetition; Statistical inference.

AMS Subject Classifications: 62K05, 05B05

1. Introduction

'Big Data' presents itself with unique challenges in retrieving, storing and all the way to analysing the data. Technological breakthrough makes generation and collection of huge volume of data possible in many fields like genetics, genomics, health care, customer service, informatics, to name a few. Among various challenges presented by the abundance of data, analysis of the data is a well recognized hurdle. While the explosion of information allows us to know more about the process, appropriate methods or algorithms are essential to make 'correct' inference or to reveal hidden patterns.

Recent advancements of technology and targeted methods to Big Data analytics give access to ample capacity for storing the data along with the skill of parallel computing. Much effort has been dedicated to extract information from Big Data in an efficient manner. From a practical standpoint, concern remains about the validity of results from analysis of Big Data. As attested by many recent articles, in most cases the inference based on such

Corresponding Author: Indranil Mukhopadhyay
Email: indranil@isical.ac.in

data is unacceptable and unreliable. For example, High dimension conventional classification methods are no better than random guesses (Fan and Fan, 2008). Understanding the output of Big Data analytics than to fixate on the technical aspect of it is the most important issue (Fan *et al.*, 2014; Fan *et al.*, 2018; Huang, 2014), because the future decision making process depends only on this output. The aim of this article is to put forward a framework in order to establish the acceptability of the learning from the Big Data. This framework also fits to the paradigm of parallel computing and at the same time provides a robust statistical basis for practical application.

The classical statistical theory of data analysis has its roots in axioms of probability theory. By some arguments data analysis is more complex than physics, biology and even behavioural science. Formal statistics developed so far can help to tackle the analytics but it will produce realistic results only if we can keep the basic assumptions loose (Tukey, 1962). The velocity of data flow in today's world makes it more challenging for producing meaningful conclusion over time on the same problem (Efron, 2020). Naturally it is not possible to analyze the entire data at the same instance when we have so little time to produce results. Also, subsequent results can make a previous conclusion redundant. With growing complexity of Big Data, statistical theory needs to be revisited (Davidian, 2013), mainly due to the violation of probabilistic independence or exchangeability conditions. Statistics community has raised concerns about how the sound and carefully developed theory can help build a structure around it. Implicitly classical statistics is already equipped with basic mechanisms to deal with big data. Sampling and sufficiency, among other core discoveries of statistics, are extremely useful in analysis of data of large volume, a characteristic of big data (Donoho, 2017). But the analysis of such high volume data needs to be done in presence of variety and high velocity, the two main characteristics of big data. Algorithmic or computational innovations for parallel computing are not the entire solution, but they are important tools when coupled with appropriate statistical methods in order to utilise the entire available information (sufficiency) that is contained in the data flow (large sample) (Donoho, 2017). In this article we exploit an algorithmic architecture used in practice to tackle Big Data and suggest an appropriate mathematical ground for analysis of such architecture.

We propose a partition and repetition approach in a general framework for statistical analysis of Big Data. This approach expands the horizon of standard statistical methods as well as opens new avenues for novel methods to encompass and tackle the challenges arisen due to the specific characteristics of Big Data. With the help of this general framework, we prove consistency and accuracy of the analytic results thus obtained. We have explained this theory through various examples that are usually required in common data analysis paradigm in respect of many fields. We hope that such a framework would help in further development of Big Data analytics. Note that although here we mainly address the problems due to volume, velocity and variety, the 3-V's that occur simultaneously in a typical Big Data problem, the other 3-V's–veracity, validity, and volatility–must always be taken care of in any statistical analysis. A statistical model has to deal with inherent variability which is nothing but volatility especially when considered with respect to time. It is natural that with available data at hand, we have to model the variance and hence the volatility, how it is related with time or other auxiliary variables. This study of volatility would be inbuilt in our proposed framework. Similarly veracity and validity must be ensured properly with appropriate strategies, for example, using back-testing, training-testing protocol *etc.* Our

proposed framework for the basis of data analytics in the Big Data paradigm considers many possible statistical fundamentals and validity requirements when analysing data with the above characteristics.

2. Developing the general framework

2.1. The divide and conquer algorithm

Abundance of digital information is one way to explain what we today understand as 'Big Data'. There are two aspects to the story. Firstly, human intuition suggests that accuracy of the answer to our question increases if we have more and more information. This intuition works backward; we start with a question, try to comprehend what data we might need to answer the question and then realize that relevant information exists somewhere in digitized format. The catch is that, this retrospective thought process assumes that the skill by which human intelligence finds this answer from the data is transferable to mechanical and algorithmic computing. Secondly, with huge volume of data we can find a question of interest from the data itself and then get the answer to the question. But the inherent complexity of available data makes this task difficult. This whole process is advertised as Big Data analytics.

Principle characteristics of Big Data are its volume, velocity, variety and complexity (Katal, 2013). All of them presents as unique challenges at a technical level of dealing with the data. At the hardware level we have reached a saturation point on the achievable clock pulse on a single processor. Rather, the growth in computing capacity is attained by increasing the number of threaded cores. Moreover, while storage capacity is fairly cheap and scalable, the Random Access Memory is not so. Recognizing this hardware restriction the state of the art algorithms (Hadoop, Amazon EC2) for Big Data analytics has adopted a partitioning based method.

However, in view of advancements in computing systems including storage and processing, need for new data analytic tools are required that are adaptive to new technologies (Petcu *et al.*, 2015). Building such statistical tools and algorithms for monitoring and analysis is needed to achieve success in Big Data analytics. Hence standard statistical methods should be revisited, modified, and validated in the light of scalability to extremely large scale data applications (Reed and Dongarra, 2015).

Fisher *et al.* (2012) have identified the standard workflow of data analysis as, (1) acquiring data, (2) choosing an architecture, (3) shaping the data to the architecture, (4) writing and editing the code, and (5) reflecting and iterating on the results. The initial struggle is to adopt a suitable architecture for the data and map the collected data to that architecture. In this article, we are not focusing on this domain of analytics job. Rather the focus is on the later part of analysing the data. To address the problem of huge volume of data, the way is to partition it into small portions that are manageable by the Random Access Memory, process the data in a parallel manner, and finally combine the processed information to produce the final output. This idea of partitioning has been used, although in a subtle way, in other areas of research, *e.g.*, data mining (Buehrer *et al.*, 2015; Calders *et al.*, 2010), Markov Chain Monte Carlo (Wang *et al.*, 2015). An extra benefit of this divide and conquer method is that such an algorithm easily adapts to the velocity of Big Data. Velocity contributes to new partitions which are to be analysed and then the inference is

to be combined with the earlier output (Schifano *et al.*, 2016). The other issues relating to variety and complexity are taken into account by the statistical methods and algorithms that are used in the analysis.

2.2. The framework

Sample space structure: Classical theory of statistical analysis is a well developed area with sound theories. To establish a framework for Big Data analytics we naturally would like to fall back on those works. To begin with, we consider a sample space $(\mathcal{S}, \mathcal{A}_\mathcal{S})$ where $\mathcal{S}$ is the space of realized values of the data and $\mathcal{A}_\mathcal{S}$ is the sigma field associated with the sample space. We denote by $\mathcal{M}(\mathcal{S})$ the set of probability measures on $(\mathcal{S}, \mathcal{A}_\mathcal{S})$. Also let $\mathcal{M}^e(\mathcal{S})\Big(\subset \mathcal{M}(\mathcal{S})\Big)$ be the set of probability measures with finite support. An observed data $X_{n\times p}$ can be identified by a probability measure m_X on $(\mathcal{S}, \mathcal{A}_\mathcal{S})$, with a support having finite cardinality, defined as follows,

$$m_X(A) = \sum_{x_i \in A, i=1}^{n} \frac{1}{n},$$

for any $A \subseteq \mathcal{A}_\mathcal{S}$ and $x_i\,(i = 1, \ldots, n)$ is the i-th data point. To build a theory around it we would require a suitable metric on the space $\mathcal{M}(\mathcal{S})$. For example, if $(\mathcal{S}, \mathcal{A}_\mathcal{S})$ is a polish space then with Prokhorov metric $(d_\mathcal{M})$ we can put weak convergence on $\mathcal{M}(\mathcal{S})$.

Till this point we have not considered any aspect of Big Data par se. Our aim is to build the ideology of Big Data analytics on this sample space structure. Identification of the realized data with an empirical measure on some sample space gives a broader ground to work on. In a Big Data set up, we hardly have any control on the generation of data. Thus unlike in classical statistical theory, where mostly we want to build better experimental designs to apply statistical methods, be it standard or novel, here we want to construct an algorithm that would work with the data generation process. This difference in approaches is subtle but central to how these two ideologies differ.

The problem approach: Main goal of Big Data analytics is to extract information from the data, which is equivalent to getting information from an element in $\mathcal{M}^e(\mathcal{S})$. So we assume that a satisfactory data collection and mapping architecture exists. To develop a full framework, we introduce some definitions about functionality of data analysis. This is necessary to avoid the cumbersome details and technicalities of a particular scenario.

Extracted information of a data analysis can be viewed as an element in the result space $(\mathcal{R})$. A *problem approach* (ρ) is a function from $\mathcal{M}^e(\mathcal{S})$ to $\mathcal{R}$. Based on this formulation of problem approach we can consider two classes of problem approaches as follows.

Definition 1: *Inference Problem*: If the problem approach ρ can be extended to a strictly larger subset of $\mathcal{M}(\mathcal{S})$ than $\mathcal{M}^e(\mathcal{S})$, then such a problem or problem approach is called an inference problem.

Definition 2: *Mining Problem*: If the problem approach ρ can only be defined on $\mathcal{M}^e(\mathcal{S})$, then such a problem or problem approach is called a mining problem.

The usual examples of these two classes of problems are as follows. Parametric estimation and testing problems fall under the class of inference problems where the subset of $\mathcal{M}(\mathcal{S})$ under consideration is $\mathcal{M}^e(\mathcal{S})$ along with the parameter models. Clustering problem or outlier detection problem, on the other hand, are under the class of mining problems. In later sections, we shall discuss both these classes of problem approaches and their solutions in more details.

A technical assumption we need to have is that, one such problem approach is *viable* if the map

$$\rho : \left(\rho^{-1}(\mathcal{R}), d_{\mathcal{M}}\right) \longrightarrow \left(\mathcal{R}, d_{\mathcal{R}}\right) \tag{1}$$

is a continuous map, where $d_{\mathcal{M}}$ and $d_{\mathcal{R}}$ are appropriate metrics on respective spaces. A *viable problem approach (ρ)* then ensures that the problem is consistent in the number of samples and robust in the data points. This means that slight change in the data generation process ($\mathcal{M}(\mathcal{S})$) should not create substantial difference in the result ($\mathcal{R}$). Here consistency indicates the large sample property of converging results as the number of data points increases whereas robustness indicates very little or not significant change in the results from two data sets that are not too different from each other.

The existence of ρ has important implications both in statistical modelling with an underlying stochastic data generation model and also in algorithm modelling with unknown data mechanism (Breiman, 2001). We only emphasis that ρ should be judiciously chosen and it has no conflict with the "two cultures" of statistical modelling (Breiman, 2001). However, in any case, we assume that there is an underlying σ-field behind the data generation process, be it known or unknown, and hence ρ is well defined. The existence of ρ is essential for establishing the sound framework for Big Data analytics that we establish through two theorems in the next section.

2.3. Big Data Algorithm

We now discuss various components of our proposed algorithmic structure of Big Data analytics.

Partitioning: A naturally accepted strategy in analysing huge volume of data is to consider small parts of data at a time. Our formulation for Big Data analytics formulates this method of partitioning the data as a functional,

$$\begin{aligned} &H_L : \mathcal{M}^e(\mathcal{S}) \longrightarrow \mathcal{M}^e(\mathcal{S}) \times \cdots \times \mathcal{M}^e(\mathcal{S}) \quad (L \textit{ such } \mathcal{M}^e(S)\, s) \\ &H_L(m) = (m_1, m_2, \ldots, m_L), \end{aligned} \tag{2}$$

such that $(m_1, m_2, \ldots, m_L)$ is related to m by,

$$\begin{aligned} &supp(m) = \bigcup_{i=1}^{L} supp(m_i); \\ &supp(m_i) \cap supp(m_j) = \emptyset, \;\; 1 \le i \ne j \le L. \end{aligned} \tag{3}$$

where $supp(m)$ denotes the support set of m and $\emptyset$ denotes the empty set.

For convenience we write $supp(m_i) = \left(x_1^{(i)}, x_2^{(i)}, \ldots, x_{n_i}^{(i)}\right)$ for each i. For a fixed data m (or $m \equiv X$) we would be given a problem approach ρ. Then the divide and conquer strategy would choose a partitioning functional H_L.

But to reduce the error in result due to partitioning, the strategy is to repeat $K(>1)$ times the partitioning; denote them by $H_{L,1}, H_{L,2}, \ldots, H_{L,K}$. This type of algorithm we call as the *partition-repetition algorithm*. We now formulate this partition-repetition algorithm in a comfortable manner.

Let $\mathcal{H}_L$ be the set of all partitioning functionals H_L. A σ-field $\mathcal{A}_{\mathcal{H}_L}$ can be defined as the smallest σ-field on $\mathcal{H}_L$ such that the functions $f_{i,j}(\cdot)$ on $\left(\mathcal{H}_L, \mathcal{A}_{\mathcal{H}_L}\right)$ to $(\mathcal{S}, \mathcal{A}_{\mathcal{S}})$ are measurable for any choice of $m \in \mathcal{M}^e(\mathcal{S})$, where

$$f_{i,j}\Big(H_L(m)\Big) = x_j^{(i)} \quad j = 1, 2, \ldots, n_i;\ i = 1, 2, \ldots, L.$$

Then the strategy of analysing data of unmanageable size, in terms of volume, variety and most importantly velocity, by partition-repetition algorithm can be understood as a probability measure P_{H_L} on the measurable space $\left(\mathcal{H}_L, \mathcal{A}_{\mathcal{H}_L}\right)$. More precisely $\{H_{L,1}, H_{L,2}, \ldots, H_{L,K}\}$ would be viewed as a random sample from the probability measure space $\left(\mathcal{H}_L, \mathcal{A}_{\mathcal{H}_L}, P_{H_L}\right)$. For simplicity of notation let us denote by ρ^L the map,

$$\rho^L : (m_1, \ldots, m_L) \longmapsto (\rho(m_1), \ldots, \rho(m_L)) \quad for\ m_i \in \mathcal{M}^e(\mathcal{S});$$

for $i = 1, 2, \ldots, L$. Then a single random sample H_L from the probability distribution P_{H_L} provides us L results $\rho^L\Big(H_L(m)\Big)$, which are L elements from $\mathcal{R}$. With a random sample $H_{L,1}, H_{L,2}, \ldots, H_{L,K}$ from the distribution, the set of results we get using the problem approach ρ is

$$\begin{aligned}\{R^*_{k,l}\}_{k=1,2,\ldots,K;\ l=1,2,\ldots,L} &= \left\{R^*_{k,1}, R^*_{k,2}, \ldots, R^*_{k,L}\right\}_{k=1,2,\ldots,K} \\ &= \left\{\rho^L\Big(H_{L,k}(m)\Big)\right\}_{k=1,2,\ldots,K}.\end{aligned}$$

This framework also encompasses the case where rather than partitioning one chooses to sub-sample. In that case we would get rid of the extra restriction in equation (3) on the functional H_L. Popular algorithms of Bootstrap and Bag-of-Little-Bootstraps (Kleiner *et al.*, 2014) are covered in this framework.

Combining: Next critical part of the algorithm is combining the results obtained above, $\left\{R^*_{k,l}\right\}_{k=1,2,\ldots,K;\ l=1,2,\ldots,L}$ in order to arrive at a final result. Let C_{KL} be the combining map that takes all the results from the collection and gives the final result. The triplet $\left(\rho, P_{H_L}, C_{KL}\right)$ can be called a *solution* to a Big Data problem.

Now it remains to understand the viability of the solution. We have put a stable condition of continuity in equation (1) on problem approach ρ as a *viable problem approach*. Proper behaviour of the pair $\left(P_{H_L}, C_{KL}\right)$ would ensure an accurate solution to the problem ρ for m.

We focus on the case where $C_{KL} := C_K^2 \circ C_L^1$ works in two stages. In the first stage C_L^1 works on each partition (k) to collect the results

$$R_k^* := C_L^1\Big(\{R_{kl}^*\}_{l=1,2,\ldots,L}\Big) \quad for\ k = 1, 2, \cdots, K.$$

This K-tuple is combined by C_K^2. For a fixed data m when C_L^1 is a measurable map, the randomness of $\{H_{L,1}, H_{L,2}, \ldots, H_{L,K}\}$ makes the collection $\{R_1^*, R_2^*, \ldots, R_K^*\}$ an independently and identically distributed (i.i.d.) sample on the measure space $\Big(\mathcal{R}, \mathcal{A}_{\mathcal{R}}\Big)$. This formulation of the solution $\Big(\rho, P_{H_L}, C_K^2 \circ C_L^1\Big)$ provides an opportunity to use rich statistical theory in data analytics.

In the general case, the result space can be quite complicated (we shall give concrete examples in later section). Rather than dealing with the space $\mathcal{R}$ itself it would be better to work with real numbers. This is achieved by an evaluation function $ev : \mathcal{R} \longrightarrow \mathbb{R}^N$ for some fixed N belonging to the set of natural integers. Then, viability of the choice of P_{H_L} can be understood using the evaluation function of the result space $\mathcal{R}$. For a given data m and a problem approach ρ, we call a partitioning probability measure P_{H_L} to be *viable* under the first stage combining operator C_L^1 if,

$$\int ev \circ C_L^1\Big(\rho^L(H_L(m))\Big) dP_{H_L} = ev \circ \rho(m). \tag{4}$$

This condition means that the probability measure P_{H_L} and the combining method C_L^1 are compatible with each other for the problem ρ. If we do infinitely many repetitions of our partition-repetition based algorithm, the combining method C_L^1 will give equivalent performance as the one we would have got if we could apply ρ on the data m.

The second stage of combining method C_K^2 operates on the collection of first stage result by combining $R_1^*, R_2^*, \ldots, R_K^*$ to get the solution

$$R_K^{**} := C_K^2\Big(\{R_k^*\}_{k=1,2,\ldots,K}\Big).$$

Now the viability of C_K^2 is based on the comparison of R_K^{**} with $\rho(m) = R^*$ (say). Here we present the soundness of the algorithm of partitioning and combining through the following theorem.

Theorem 1: For a Big Data solution $\Big(\rho, P_{H_L}, C_K^2 \circ C_L^1\Big)$, if P_{H_L} is a viable partitioning method under combining method C_L^1 (*i.e.*, equation (4) is satisfied) and convergence in ev is equivalent to that of in $\mathcal{R}$, then there exists a second stage combining method C_K^2, such that $R_K^{**} \longrightarrow R^*$ almost surely in P_{H_L}.

Proof: Define C_K^2 on $\mathcal{R} \times \mathcal{R} \times \cdots \times \mathcal{R}$ (K times) as follows,

$$C_K^2(R_1, R_2, \ldots, R_K) := \arg \min_{\{R_k\}_{k=1,2,\ldots,K}} \Big|\Big|ev \circ R_i - ev \circ R^*\Big|\Big|.$$

Let us use the notations $Y_k = ev \circ R_k^*$, $Z_K = ev \circ R_K^{**}$ and $\mu = ev \circ R^*$. Since $\{R_k^*\}_{k \geq 1}$ is an i.i.d. sample, by strong law of large numbers as equation (4) holds, for all $\epsilon > 0$ with

$P_{\mathcal{R}} := P_{H_L} \circ C_L^{1^{-1}} \circ ev^{-1}$,

$$P_{\mathcal{R}}\Big(\cup_{k_0=1}^{\infty} \cap_{K \geq k_0} \Big(\Big|\Big| \frac{1}{K} \sum_{k=1}^{K} Y_k - \mu \Big|\Big| < \epsilon \Big) \Big) = 0.$$

Now using the fact that $||\sum_{k=1}^{K} Y_k/K - \mu|| \geq ||Z_K - \mu||$ and definition of C_K^2, the above holds with $\sum_{k=1}^{K} Y_k/K$ replaced by Z_K. Since convergence in $(\mathcal{R}, d_{\mathcal{R}})$ is equivalent to that in $\Big(ev \circ \mathcal{R}, ||\cdot||\Big)$, rest of the argument follows as by assumption convergence in $(\mathcal{R}, d_{\mathcal{R}})$ is equivalent to that in $\Big(ev \circ \mathcal{R}, ||\cdot||\Big)$. □

The theorem above deals with the volume aspect of Big Data. It says that even if the data is unmanageable to be processed practically, we can adopt partition-repetition approach to get a good solution. It is also not passed our attention that the number of combination rules may be more than two, but the final convergence of results requires some more assumptions and strong theorems in the dependence set up.

Next we also need to answer the question which is more of classical statistical in nature. If the velocity of the data provides us more and more information of specific form, is the partition-repetition algorithm able to extract that information? The following theorem tells us if that is the case, we would be able to choose a partitioning measure and a sequence of combining methods that gives the final result.

Theorem 2: Let $\{m_n\}_{n\geq 1} \in \mathcal{M}^e(\mathcal{S})$ and $m \in$ *domain of* ρ. Suppose the problem approach ρ is viable on its domain and $m_n \longrightarrow m$. If conditions of Theorem 1 hold for the sequence of solutions $\Big(\rho, P_{H_L,n}, C_{K,n}^2 \circ C_{L,n}^1\Big)$, then there exists a sequence of integers $\{k_n : n \geq 1\}$ and a P_{H_L} such that, for $n \geq 1$, $P_{H_L,n}$ is absolutely continuous with respect to P_{H_L} with

$$\Big|\Big| ev \circ C_{k_n,n}^2 \circ C_{L,n}^1 \Big\{ \rho^L(H_{L,k}(m_n)) \Big\}_{k=1,2,\ldots,K} - ev \circ \rho(m) \Big|\Big| \longrightarrow 0,$$

as $n \to \infty$ almost surely in $P_{\mathcal{R}}$.

Proof: Define $P_{H_L}(\cdot) = \sum_{n=1}^{\infty} P_{H_L,n}(\cdot)/2^n$. Let us denote,

$$R_{K,n}^{**} = C_{K,n}^2 \circ C_{L,n}^1 \Big(\Big\{ \rho^L\Big(H_{L,k}(m_n) \Big) \Big\}_{k=1,2,\ldots,K} \Big).$$

Then for every $\epsilon(>0)$, by Theorem 1 and equation (1) there exists a sequence $\{k_n(\epsilon) : n \geq 1\}$ and $N \geq 1$ such that for all $n \geq N$,

$$\Big|\Big| ev \circ R_{k_n(\epsilon),n}^{**} - ev \circ \rho(m) \Big|\Big| < \frac{\epsilon}{2^n},$$

almost surely in $P_{\mathcal{R},n} = P_{H_L,n} \circ C_{L,n}^{1^{-1}} \circ ev^{-1}$. Choosing ϵ as rationals, result follows from Cantor's diagonal argument. □

Both these results are of existential nature rather than being instructive for practice. Although little abstract in their formulation, these theorems form the basis of the methods that would be applied in practice. Study on combining methods is not new to statistics. This

framework enforces the importance of various combining methods along with partitioning methods in the light of Big Data analytics.

The power of this kind of theory is that we do not put any hard and fast regularity condition on the data or the data generation process. Theorem 2 only requires that the data collected eventually amounts to some specific information.

3. Illustrative Examples

An analyst's job and a statistician's work differ in a crucial way. An analyst is more concerned with how to extract information from the data available. This work is referred to as number crunching. A statistician is concerned about the quality of the extracted information sometimes taking for granted the effort of extracting the information. In a Big Data scenario where importance of analyst's job comes more into the limelight, a statistician could provide support by accepting some compromise on their ideology. In this section we illustrate the formulation developed above through some standard data analytic problems.

We first consider a few problems where the solution $\rho(m_n)$ can be calculated without any error from partitioning based algorithm. Here we specify by subscript n the size of the data. In these examples it is enough to consider P_{H_L} to be some degenerate probability distribution of convenience and we only require a single sample ($K = 1$) from it.

Calculating sample mean: Here P_{H_L} can be any distribution that partitions the data into manageable balanced pieces. Then for $\rho(m_n) := (\int x\, dm_n, n)$ the combining method shall be,

$$C_L^1(\{(\bar{x}_i, n_i)\}_{i=1,2,\ldots,L}) = \Big(\frac{\sum_i n_i \bar{x}_i}{\sum_i n_i}, \sum_i n_i\Big).$$

A little tweak in these definitions allows us to calculate many other descriptive statistics like weighted means, dispersion measures and also some robust measures for central tendency.

Sorting: To get a Big Data solution to the sorting problem we can define a partitioning P_{H_L} as a degenerate distribution such that it divides the data m_n into L parts based on a sequence $bound_0 < bound_1 < \cdots < bound_L$ as,

$$bound_{i-1} \le \{x_j^{(i)}\} < bound_i \;\; for\; i = 1, 2, \cdots, L.$$

The choice of the sequence $\{bound_i\}$ should be such that the individual parts are of manageable sizes. With ρ providing us with a sorted array, the combining stage should simply concatenate the ordered parts, *i.e.*,

$$C_L^1(\{R_l^*\}_{l=1,2,\ldots,L}) := (R_1^*, R_2^*, \cdots, R_L^*).$$

Similar solutions of the above type are obvious for problems like searching, calculating extreme statistics $(x_{(1)}, x_{(n)})$, constructing a histogram *etc.* Most of the time these simple problems are only intermediate steps towards more challenging problems of data analytics.

Some solutions to more standard problems of Big Data analytics are discussed in brief below. First few examples are inference problems while the later ones are mining problems.

We assume that the data are cleaned and dressed for the purpose at hand. We avoid discussing the technical aspects of implementing these algorithms in practice, though in a few examples we shall provide references to available literature that has more focus on detailed analysis of the algorithms.

Estimation: The problems of modelling (nonparametric, parametric, time series or even Bayesian) come under the radar of inference problem. Based on the requirements of the solution (*e.g.*, unbiasedness, minimum variance, consistency) there would be different Big Data solutions to the problem approach ρ. Many of the times it suffices to consider P_{H_L} as a random partitioning measure of the data, although while considering spatial and/or temporal data more clever partitioning measure would be required to satisfy viability condition like equation (4).

Let us consider the problem of finding maximum likelihood estimate for a parameter based on some algorithm (say, Expectation-Maximization algorithm or Newton-Raphson or Fisher's Scoring *etc.*). The scenario is that, we have a statistical model in mind where the number of parameters is fixed. Then partitioning the data simply breaks the objective function (log-likelihood function) into L parts. Consequently an intuitive choice of the combining method C_{KL} would be whichever of the results from partitions maximizes the whole objective function. Although this method does not ensure the MLE for the data, but in practice we are hardly concerned about theoretical properties like efficiency; the estimate found by this method is acceptable.

Testing: Consider a test function ρ that provides p-value for testing H_0 against H_1. Then based on random partitioning of the data into balanced parts, a conservative combining algorithm (Tippett, 1931) for the corresponding solution can be

$$R_k^* := C_L^1\Big(\{R_{lk}^*\}_{l=1,2,\ldots,L}\Big) = \min_{l=1,2,\ldots,L} R_{lk}^*, \text{for } k = 1, 2, \ldots, K,$$

and

$$R^{**} := C_K^2\Big(\{R_k^*\}_{k=1,2,\ldots,K}\Big) = median\{R_k^*\}_{k=1,2,\ldots,K}.$$

A large part of recent statistical methods literature focuses on the regime of $p >> n$. Even when the data is not formally a Big data, because it does not inherit the various characteristics discussed in the introduction, the data set can be in this regime; for example, genome sequencing data. Note that the above discussion also encompasses the scenario when $p >> n$. Depending on the testing problem, if we get a p-value or a test statistic for each partitioned dataset, the solution is immediate in the proposed framework. It is valuable to consider specific problems in this regime in depth as they can be helpful to solve important problems in the relevant fields. But data sets solely of the $p >> n$ variety arguably represent a small part of Big data as we consider here.

Variable Selection: The context in which variable selection problem has been addressed in recent literature is sometimes too idealistic for Big Data paradigm, although there are some promising methods. The data generation process is assumed to provide information on a set of response variables and a fixed set of regressors. We might be interested in a subset of these variables which have effect on the responses. The quality of the selected variables

can be assessed by proportions of the variables wrongly selected. In a situation where assumption of homoscedastic uncorrelated linear model is valid, Barbar and Candes (2015) proposed a method to select variables with a control on the proportion of falsely discovered variables. This method is no doubt computationally heavy. The partition-repetition philosophy can be used to adapt this algorithm to achieve the same goal in current context.

If the data generation process is well controlled, the above inference problems and solutions make sense. Some recent works are available in the area of regression (Battey *et al.*, 2015; Chen and Xie, 2014) focusing on divide and conquer methods. Unfortunately spurious correlations, noisy data etc. are very common in Big Data perspective. In that case these naive solutions can be hugely mis-representative of the actual truth. Data mining problems are more relevant in such a scenario. In a mining problem we are interested in the data itself without having to make any modelling assumption. Possible Big Data solutions to a few mining problems are discussed below.

Clustering: An elaborate and critical discussion on clustering problem in view of Big Data analytics can be found in recent article by the authors (Karmakar and Mukhopadhyay, 2016; Karmakar *et al.*, 2019). Karmakar and Mukhopadhyay (2016) provide a detailed example illustrating how the proposed framework fits to the class of Big Data clustering problems where it (a) demonstrates existence and evaluation of the required quantities and (b) proves validity of the final result. In brief, the combing method would identify the unique clusters from the set $\{R^*_{lk}\}_{l=1,2,\ldots,L}$ based on a decision function that tells us to combine two results when they seem to form a single data cloud. The second stage is to make stable clusters based on some measure from the K sets of clusterings $\{R^*_k\}_{k=1,2,\ldots,K}$.

Outliers Detection: Based on a random partitioning measure P_{H_L} and a problem approach ρ that separates the outliers (m^o_n) and the data (m^d_n) section, (i.e., $\rho(m_n) := (m^d_n, m^o_n)$), the combining method C^1_L would check the structure of the outliers from the individual parts and get the outliers from the whole part. The method should check if outliers from one part belongs to the data section of some other part and also if outliers from all the parts together form some data section. Second stage of combining would then pick out the stable outliers from all repetitions.

Ramaswamy *et al.* (2000) discuss another Big Data solution to this mining problem based on a different partitioning method based on clustering the data and van Stein *et al.* (2016) propose local subspace-based solution to outlier detection problem, which applies a combining strategy using global neighbourhoods. These methods can be viewed as special cases of our proposed framework.

Classification: First we consider the k-Nearest Neighbor classifier, where ρ finds the k nearest neighbours of a test data point (x) as,

$$\rho(m_n) := ((x_{(i)}, d(x, x_{(i)}))_{i=1,2,\ldots,k})$$

$$\begin{aligned} \text{such that } \; d(x, x_{(1)}) &\leq \cdots \leq d(x, x_{(k)}) \\ &\leq min\{d(x, x_i); x_i \in X \setminus \{x_{(1)}, x_{(2)}, \ldots, x_{(k)}\}\}. \end{aligned}$$

Based on any partitioning P_{H_L}, then the problem is exactly solvable in a single repetition with a combining operator that picks the k data points nearest to x among the $L \times k$ points.

Subsequently the classifier is contracted on a second algorithm that simply checks for the maximum number of representatives in these k data points from each of the classes.

Another celebrated class of classifiers is decision trees. A relevant combining operator of decision trees based on partition of the data is proposed by Hall *et al.* (1998).

4. Discussion

Data is the lubricant that drives the machinery of statistics. It is no longer a topic of debate that the way data is generated and collected in modern times is drastically different from what statisticians are used to deal with. Statistics should adapt to this change and thereby assist the masses of data analytic work.

The main contribution of this article is suggesting a basis of statistical theory for present day data analytic works. In composing the theory we have tried to stay true to the practical nature of a data science job. This formulation proposes a divide and conquer algorithm (either partition-repetition or subsampling method). More importantly it respects the fact that more often than not we have no control on the data generation process. We have also tried to encompass all possible data analytic problems. A range of such data analytic problems are discussed in perspective of our formulation.

5. Conclusion

Successful use of statistical theory in data analysis would require understanding the field of 'Big Data'. Rather than being insistent on developing methods and elaborate theories based on idealistic assumptions, we have kept their applicability in mind. Our proposed framework encompasses statistical analyses of majority of problems in view of complex characteristics of Big Data and can be extended further keeping its compatibility with modern advances in computational world.

References

Barber, R. F. Candes, E. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, **43(5)**, 2055–2085. doi:10.1214/15-AOS1337.

Battey, H., Fan, F., Liu, H., Lu, J. and Zhu, Z. (2015). Distributed estimation and inference with statistical guarantees. *arXiv preprint:1509.05457v1.*

Breiman, L. (2001). The statistical modelling: the two cultures, *Statistical Science*, **16(3)**, 199-231.

Buehrer, G., de Oliveira, R. L., Fuhry, D. and Parthasarathy, S. (2015). Towards a parameter-free and parallel itemset mining algorithm in linearithmic time. *2015 IEEE 31st International Conference on Data Engineering, Seoul*, 1071–1082.

Calders, T., Garboni, C. and Goethals, B. (2010). Efficient pattern mining of uncertain data with sampling. In M. J. Zaki, J. X. Yu, B. Ravindran, V. Pudi (Eds.), *PAKDD 2010, Part I. LNCS (LNAI)*, **6118**, Springer, Heidelberg, 480–487. doi:10.1007/978-3-642-13657-3_51.

Chen, X. and Xie, M. (2014). A split-and-conquer approach for analysis of extraordinarily large data. *Statistica Sinica*, **24**, 1655–1684. doi:10.5705/ss.2013.088.

Davidian, M. (2013). Aren't *we* data science. *Amstat News*, 433-435.

Donoho, D. (2017). 50 Years of data science. *Journal of Computational and Graphical Statistics*, **26(4)**, 745–766. doi: 10.1080/10618600.2017.1384734.

Efron, B. (2020). Prediction, estimation, and attribution. *Journal of theAmerican Statistical Association*, **115**, 636–655,

Fan, J. and Fan, Y. (2008). High dimensional classification using features annealed independence rules. *The Annals of Statistics*, **36**, 2605–2637. doi:10.1214/07-AOS504.

Fan, J. Fang, H. and Han, L. (2014). Challenges of big data analysis. *National Science Review*, **1(2)**, 293–314. https://doi.org/10.1093/nsr/nwt032.

Fan, J., Shao, Q. M. and Zhou, W. X. (2018). Are discoveries spurious? Distributions of maximum spurious correlations and their applications. *The Annals of Statistics*, **46(3)**, 989–1017.

Fisher, D., DeLine, R., Czerwinski, M. and Drucker, S. (2012). Interactions with big data analytics. *Interactions with Big Data Analysis*, **19(3)**, 50–59.

Hall, L., Chawla, N. and Bowyer, K. W. (1998). Combining decision trees learned in parallel. In *Working Notes of the KDD-97 Workshop on Distributed Data Mining*, 10–15.

Huang, J. P. (2014). Big data need physical ideas and methods. *arXiv preprint:1412.6848.*

Katal, A., Wazid, M. and Goudar, R. H. (2013). Big data: Issues, challenges, tools and good practices. In *2013 Sixth International Conference on Contemporary Computing (IC3)*, 404–409. doi:10.1109/IC3.2013.6612229.

Karmakar, B., Das, S., Bhattacharya, S., Sarkar, R. and Mukhopadhyay, I. (2019). Tight clustering in Big Data Analytics with focus on microarray data. *Scirntific Reports*, **9:3053**, 1-12.

Karmakar, B. and Mukhopadhyay, I. (2016). An efficient partition – repetition approach in clustering of big data. In S. Pyne, B.L.S. Prakasa Rao, S.B. Rao (Eds.), *Big data analytics: Methods and applications*, Springer India, New Delhi, 75–93. doi:10.1007/978-81-322-3628-3_5.

Kleiner, A., Talwalkar, A., Sarkar, P. and Jordan, M. I. (2014). A scalable bootstrap for massive data. *Journal of Royal Statistical Society Series B*, **76**, 795–816. doi:10.1111/rssb.12050.

Petcu, D. et al.. (2015). On processing extreme data. *Scalable Computing: Practice and Experience*, **16(4)**, 467–489.

Ramaswamy, S., Rastogi, R. and Shim, K. (2000). Efficient algorithms for mining outliers from large data sets. *ACM Sigmod Records*, **29(2)**, 427–438. doi:10.1145/335191.335437.

Reed, D. A. and Dongarra, J. (2015). Exascale computing and big data. *Communications of the ACM*, **58(7)**, 56–68. doi:10.1145/2699414.

Schifano, E. D., Wu, J., Wang, C., Yan, J. and Chen, M. H. (2016). Online updating of statistical inference in the Big Data setting. *Technometrics*, **58(3)**, 393-403.

Tippett, L. H. C. (1931). The methods of statistics. Williams & Norgate, London.

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, **33**, 1–67.

van Stein, B., van Leeuwen, M. and Bäck, B. (2016). Local subspace-based outlier detection using global neighbourhoods. *IEEE International Conference on Big Data*, 1136–1142.

Wang, X., Guo, F., Heller, K. A. and Bunson, D. B. (2015). Parallelizing MCMC with random partition trees. *Advances in Neural Information Processing Systems*, 451–459.

Zhao, T., Cheng, G. and Liu, H. (2016). A partially linear framework for massive heterogeneous data. *The Annals of Statistics*, **44(4)**, 1400–1437. doi:10.1214/15-AOS1410.

Statistics and Applications {ISSN 2452-7395 (online)}
Volume 18, No. 2, 2020 (New Series), pp 383-391

Finite Sample Properties of A-Optimal Designs for Binary Response Data

Rajesh Ranjan Nandy[1], Srichand Jasti[1] and Karabi Nandy[2]

[1]*Department of Biostatistics and Epidemiology, University of North Texas Health Science Center*
[2]*Department of Population and Data Sciences, University of Texas Southwestern Medical Center*

Received: 24 September 2020; Revised: 8 October 2020; Accepted: 11 October 2020

Abstract

A-optimality refers to the design that minimizes the sum of variances of the estimators of all parameters in a model. By virtue of the Cramer-Rao bound, for a vector-parameter of k components, k^2 times the trace of the inverse of the information matrix for the parameters serves as a lower bound for the sum of variances of the estimators and the bound is attained asymptotically. Hence, asymptotically, A-optimality is achieved by maximizing the trace of the inverse of the information matrix. For a binary response experiment with a logit model, the asymptotic solution is known to be a two-point design which is point symmetric but not weight symmetric. For nonlinear models, Cramer-Rao bound may be crude for finite samples and hence the asymptotic solution may be different from the design that minimizes the sum of variances. Here we explore the validity of the asymptotic solution by directly minimizing the sum of variances using numerical methods in the space of all 2-points designs as well as more restrictive design spaces. We demonstrate that even in a restrictive search space of point symmetric designs, the theoretical solution is half as efficient for a sample size of 100. Further improvement is achieved by relaxing the restriction of the solution being point symmetric.

Key words: A-optimality; Dose-response model; Information matrix; Logistic regression model.

1. Introduction

Optimal designs are a class of experimental designs that are optimal with respect to some statistical criterion. The context is to provide estimators of unknown model parameters and the optimality criteria seek to maximize or minimize some meaningful statistical functions relevant to the model and criteria. Traditionally, optimality-criteria are functionals of the eigenvalues of the information matrix. Much of the literature on optimal design rests on asymptotic properties of various optimality criteria. We refer to Pukelsheim (1993) for description of different optimality criteria.

Dose-response models have been extensively studied in the optimal design literature (Hedayat *et. al*. 1997). Logistic or logit models and probit models are among the popular ones. In this paper, we focus on optimality for a logistic linear regression model [Abdelbasit

Corresponding Author: Rajesh Ranjan Nandy
E-mail address: Rajesh.Nandy@unthsc.edu

and Plackett (1983), Biederman *et. al*. (2006), Ford *et. al.* (1992), Minkin (1987), Sitter and Wu (1993), Yang and Stufken (2009)]. Among the well-known and commonly used optimality criteria, A-optimality is perhaps the most intuitive. Consider a binary response y_x resulting from a non-stochastic dose level x. Assume that y_x takes the values 0 and 1 and the probability that y_x takes the value 1 is given by

$$P(y_x = 1) = (1 + e^{-(\alpha+\beta x)})^{-1} \tag{1}$$

where α and β are unknown and $\beta > 0$,without loss of generality. A-optimality criterion is simply minimizing the sum of variances of the parameter estimates in the model. For this two-parameter logistic model, the A-optimality criterion seeks to minimize $Var(\hat{\alpha}) + Var(\hat{\beta})$. It is mathematically challenging to directly optimize the sum of the variances for a theoretical solution. Instead, investigators have exploited the Cramer-Rao bound, which presents a lower bound on the sum of variances of unbiased estimators, indicating that the variance of any such estimator is at least as high as the inverse of the Fisher information. Specifically, in the current context, the lower bound is the trace of the inverse of the information matrix. In other words,

$$Var(\hat{\alpha}) + Var(\hat{\beta}) \geq \sum_{i=1}^{m} \xi_i \frac{e^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2}(1 + x_i^2)/|\boldsymbol{I}(\boldsymbol{\alpha}, \boldsymbol{\beta})|, \tag{2}$$

where $\sum \xi_i = 1$ and

$$\boldsymbol{I}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{pmatrix} \sum_{i=1}^{m} \xi_i \frac{e^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2} & \sum_{i=1}^{m} \xi_i x_i \frac{e^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2} \\ \sum_{i=1}^{m} \xi_i x_i \frac{e^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2} & \sum_{i=1}^{m} \xi_i x_i^2 \frac{e^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2} \end{pmatrix}. \tag{3}$$

By minimizing the trace of the inverse of the information matrix, instead of $Var(\hat{\alpha}) + Var(\hat{\beta})$, it is possible to obtain a theoretical solution. The solution to the A-optimal design was first postulated by Mathew and Sinha (2001) under restricted conditions and later established conclusively by Yang (2008). In this context, it should be noted that a major challenge in determining an optimal design for nonlinear models is that it actually depends on the unknown parameters. This presents a conundrum: one is looking for the design with the goal of optimizing the estimation of the unknown parameters, and yet one must know the true values of the parameters to find the best design. This problem has been addressed previously by Nandy and Nandy (2015) and is not the focus of the current article.

However, the Cramer-Rao bound is strict for finite samples and equality is only attained asymptotically. Hence, the A-optimal solution obtained by minimizing the trace of the inverse of the information matrix is only approximate. For non-linear models, the Cramer-Rao bound may be crude with small samples and hence the asymptotic solution can be different from the design that minimizes the sum of variances of the estimates. For finite samples, it is of great importance to examine the differences between the asymptotic and exact solutions. Keeping this in mind, the objective of this article is to focus on A-optimality criterion in a non-linear model, specifically the two-parameter logistic regression model noted in (1) and study its finite sample properties.

2. Methods

2.1. Theoretical asymptotic solution

The theoretical asymptotic solution for the A-optimal design has been shown to be a two-point design that is point-symmetric but not weight symmetric (Yang 2008). Specifically, the design is given by

$d^* = \{(x_1^*, \xi_1^*), (x_2^*, \xi_2^*)\}$ where $x_1^* = \frac{(-c^*-\alpha)}{\beta}$, $x_2^* = \frac{(c^*-\alpha)}{\beta}$;

$\xi_1^* = \xi_{c^*,\alpha,\beta}$, $\xi_2^* = 1 - \xi_1^*$where $\xi_{c,\alpha,\beta} = \frac{\sqrt{(\beta^2+(c+\alpha)^2)}}{\sqrt{(\beta^2+(c+\alpha)^2)}+\sqrt{(\beta^2+(c-\alpha)^2)}}$

and $c^* > 0$ is the only positive solution of the equation

$$\frac{c^2-\alpha^2-\beta^2}{\sqrt{(\beta^2+(c+\alpha)^2)}+\sqrt{(\beta^2+(c-\alpha)^2)}} = 1 + \frac{c(1-e^c)}{1+e^c}. \quad (4)$$

2.2. Exact numerical solution using simulation

We now describe the simulation methodology for obtaining empirically A-optimal designs. It should be noted that in our chosen parametrization, the variance of $\hat{\beta}$depends heavilyon the chosen scale of measurement of$x$, whereas the variance of $\hat{\alpha}$ does not, since it isunit-free. Hence the A-optimal solution is not scale-invariant, and the scale can be chosenarbitrarily to modify the optimal design points. This is a serious weakness of the criterionin the context of logistic regression model. In order to circumvent the arbitrariness of the solutions based on the chosen scale, we fix a scale for which β = 1. For the sake of brevity, we describe the process for $\alpha = 1$. The methods outlined here can be easily applied to any other values of the parameters by appropriate rescaling and shift. In principle, for a given finite sample size, it is possible to find the true A-optimal design in the full unrestricted design space. However, the computational time can be prohibitively expensive. So, the search is conducted in the space of two-point designs, lifting the restriction of point symmetry. This sheds light on how much improvement a restricted search can offer over the asymptotic A-optimal design solution.

In order to facilitate simpler and faster solutions within the restricted search space of two-point designs, we impose different types of additional restrictions as described below.

i. First, we fix the symmetric design points by the doses determined from the theoretical solution, and then search for a weight (ξ_1) that minimizes the A-optimality criterion, *i.e.*, the sum of variances of the estimates. Note that $\sum \xi_i = 1$ and ξ_i's represent the relative frequencies (n_i/n)'s for a given total sample size of n.
ii. Next, we fix the weight (ξ_1) to the theoretical solution, and then search for point-symmetric doses (x_1, x_2) that minimize the sum of variances of the estimates.
iii. We then conduct an exhaustive grid search in the restricted space of two-point, point symmetric designs.
iv. Finally, we complete the investigation by relaxing the point symmetry restriction and conduct an exhaustive grid search in the space of all two-point designs.

It should be noted that even with the additional restrictions, the performance will not be any worse than the theoretical A-optimal design, since the theoretical solution resides within the restricted search spaces.

2.3. Simulation details

Simulations for sample sizes varying from 20 to 1000 were conducted for each of the cases considered. For case (i) in 2.2, we used the theoretical A-optimal design points as obtained from equation 4 and then searched for the weight that minimizes the A-optimality criterion. Hence, the dosage values are $x_1 = \frac{c^* - \alpha}{\beta}$ and $x_2 = \frac{-c^* - \alpha}{\beta}$, where c^* is the theoretical A-optimal design point as obtained from equation 4. For each pair (x_1, x_2), the sample weight ξ_1, *i.e.* the proportion of the total sample allocated to x_1, was varied with the remainder being allocated to x_2. For a given sample size n and a weight $\xi_1, n_1 = \xi_1 * n$ random Bernoulli responses were generated at dosage x_1, with probability of success $p_1 = \frac{1}{1+\exp\left(-1+\frac{c^*-\alpha}{\beta}\right)}$. Similarly, $n_2 = 1 - n_1$random Bernoulli responses were generated at dosage x_2 with probability of success $p_2 = \frac{1}{1+\exp\left(-1+\frac{-c^*-\alpha}{\beta}\right)}$. A logistic regression model was fit to the resulting dataset of $n\ (= n_1 + n_2)$ Bernoulli responses at design points x_1 and x_2. The corresponding estimates of α and β are obtained and $(Var(\hat{\alpha}) + Var(\hat{\beta}))$ calculated by repeating this process 10,000 times from which an empirical estimate for the A-optimality criterion is obtained for a given sample size and design. The final optimal design was chosen to be the one that minimized this criterion.

For case (ii) in 2.2, the weight ξ_1 was determined from the theoretical A-optimal design in equation 4, and c was allowed to vary in the design space of point-symmetric designs. The optimal c and corresponding design points x_1 and $x_2$are obtained by repeating the procedure for case (i). For case (iii), we generalize the process by also allowing sample weight ξ_1 to vary. Finally, we relax the assumption of point-symmetry and conduct a search in a much larger space, where the design points x_1 and $x_2$are also allowed to vary freely. The optimization problems are solved using a grid search, with a search space set up for *c* ranging from 0.1 to 2.0 in 0.05 increments. For ξ_1, the range is setup to be 0.1 to 0.9 in 0.04 increments.

2.4. Computational detail

The programming is completed in R software (R-Project.org, v 3.3.1) using the "doParallel" package to conduct simultaneous simulations on all cores of a hyper-threaded quad-core computer. Efficiencies are obtained during the simulations by minimizing the number of calls to built-in functions. For example, instead of going through the linear process of generating a sample of size '*n*', conducting a logistic regression, saving the parameter estimates, and then generating another dataset, all the datasets (*e.g.* 10,000 $\times n$ size matrix) are generated in one call and the logistic regression model is applied to each dataset and parameters saved, resulting in 10,000 fewer calls to the "rbinom" function to generate the random sample.

3. Results

3.1. Performance of the theoretical A-optimal design

As noted earlier, the Cramer-Rao bound is a lower bound for the true sum of the variances of the estimates. We first compare the true sum of the variances of the estimates for finite samples with the Cramer-Rao bound to assess how far off the asymptotic design is from true A-optimality. With α and β set to 1, the theoretical design points are calculated to be**:**

$\boldsymbol{x_1}$(low dose) = –2.482 and $\boldsymbol{x_2}$ (high dose) = 0.482 and optimal weights are 0.71 and 0.29 respectively. In Table 1, we have summarized the results. It is clear that we need a sample size of at least 300 to attain the bound. For sample sizes of 100 or less, the true sum of variances is much higher than the best case.

Table 1: Comparison of theoretical versus empirical A-optimal design solutions

Sample size (n)	$\boldsymbol{c_A^*}$	$\boldsymbol{\xi_1}$	$A_{opt} = \boldsymbol{Tr(inv(I))/n}$	$\boldsymbol{A_{opt}^*}$
20	1.482	0.71	0.54	29.15
40	1.482	0.71	0.27	4.84
60	1.482	0.71	0.18	1.31
80	1.482	0.71	0.14	0.57
100	1.482	0.71	0.11	0.25
300	1.482	0.71	0.04	0.04
1000	1.482	0.71	0.01	0.01

Note:$\boldsymbol{c_A^*}$ is the theoretical solution $\boldsymbol{c}^*$in (4) and A_{opt}^* is the A-optimality criterion (*i.e.*, sum of the variances of the parameter estimates) for the chosen design points.

3.2. Performance of various finite samples designs compared to theoretical A-optimal design

To compare the performances, we define an improvement (or loss) in efficiency for each design as

$$E = \frac{A_{search}^* - A_{opt}^*}{A_{opt}^*} * 100\%$$

where, A_{search}^* is the minimum value of the A-optimality criterion for the restricted design space. In Tables 2-5, we summarize the performances of the four finite sample designs described in 2.2 with A_{opt}, the A-optimality criterion for the theoretical A-optimal design.

Table 2: Optimal proportion $\boldsymbol{w_1}$ (c is fixed) at various sample sizes and gain in efficiency

Sample size (n)	$\boldsymbol{c_A^*}$	$\boldsymbol{\xi_{search}^*}$	$\boldsymbol{A_{search}^*}$	$\boldsymbol{E}$ **(%)**
20	1.482	0.87	25.00	14.25
40	1.482	0.59	3.85	20.52
60	1.482	0.59	0.62	52.78
80	1.482	0.51	0.26	55.32
100	1.482	0.59	0.14	43.72
300	1.482	0.71	0.04	0.00
1000	1.482	0.71	0.01	0.00

Note:$\boldsymbol{\xi_{search}^*}$is the weight at the left design point for which the minimum value of the A-optimality criterion is attained, as shown in A_{search}^*.

Table 3: Optimal design point c ($\xi_1$is fixed) at various sample sizes and gain in efficiency

Sample size (n)	c_A	ξ_1	A^*_{search}	E (%)
20	1	0.71	17.88	38.68
40	0.95	0.71	1.95	59.74
60	0.65	0.71	0.49	62.53
80	0.95	0.71	0.21	63.35
100	1.25	0.71	0.14	42.11
300	1.50	0.71	0.04	0.00
1000	1.55	0.71	0.01	0.00

Note:c_A is the right design point that minimizes the A-optimality criterion, A^*_{search}.

Table 4: A-optimal design in the restricted class of point symmetric designs

Sample Size (n)	c_A	ξ_1	A^*_{search}	E (%)
20	0.5	0.55	12.94	56
40	0.7	0.55	0.68	86
60	1.15	0.51	0.29	78
80	1.15	0.59	0.17	70
100	1.3	0.63	0.13	48
300	1.45	0.67	0.04	0
1000	1.45	0.71	0.01	0

We can achieve improvements, ranging from 50% to 90%, depending on the sample size.

Table 5: A-optimal design in the class of two-point designs without any additional restrictions

Sample Size (n)	x_1	x_2	ξ_1	A^*_{search}	E (%)
20	–3.0	–0.1	0.87	10.27	65
40	–1.5	–0.1	0.59	0.60	88
60	–1.7	0.3	0.59	0.25	81
80	–1.8	0.6	0.67	0.16	72
100	–1.9	0.5	0.67	0.12	50
300	–2.3	0.5	0.67	0.04	0
1000	–2.3	0.6	0.71	0.01	0

We find further efficiency by relaxing the symmetry requirement, although the improvement is limited and is only significant at the smallest sample sizes

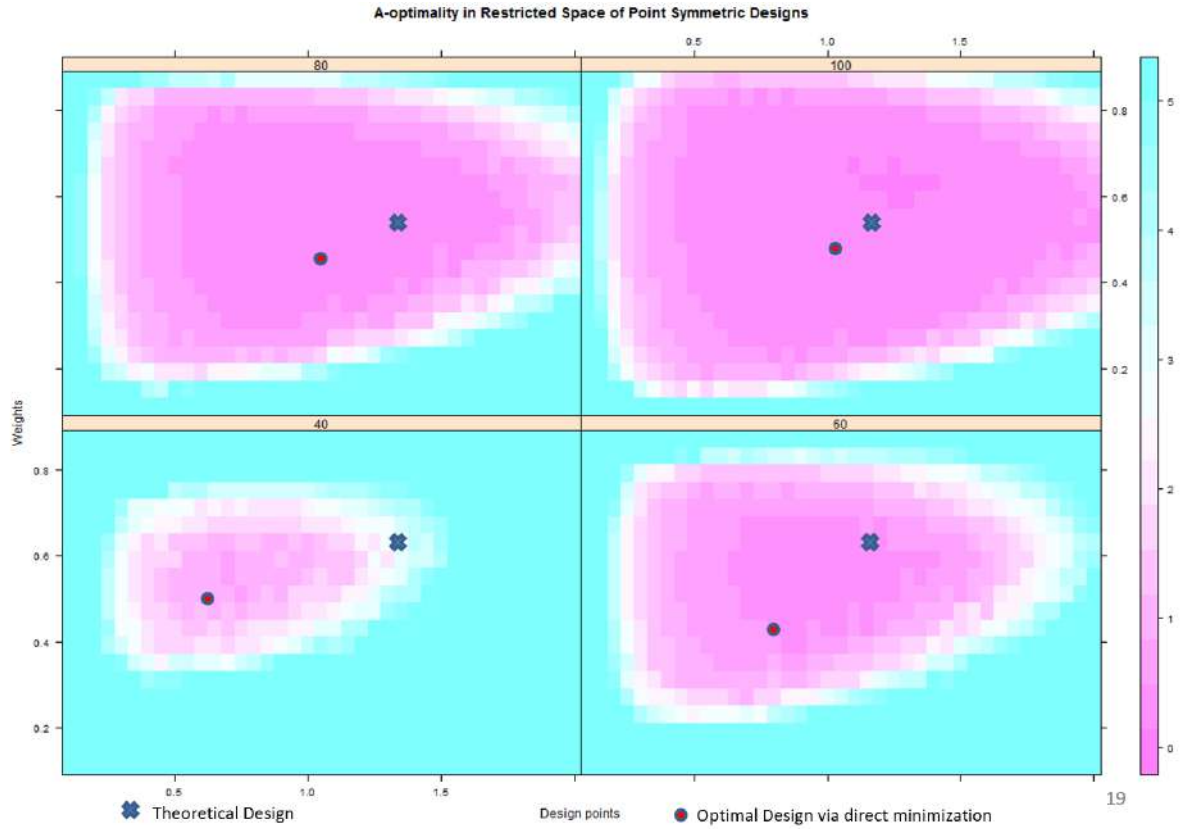


Figure 1: Comparison of the theoretical versus direct minimization in point symmetric design space

In this graph, each segment relates to sample sizes 40, 60, 80 and 100. We can see that as the sample size increases, the distance between the theoretical solution of the A-optimality criterion and the solution via direct minimization decreases.

4. Discussion

Even though we have clearly established that the asymptotic result is inadequate for a sample size of 100 or less, the fundamental reason for the widespread use of asymptotic result in small sample designs is the lack of a theoretical solution. In fact, it is impractical to find the true A-optimal design numerically by searching the entire space of designs. Instead, we obtained the optimal solutions numerically in several restricted design spaces and assessed the improvements over the asymptotic solution.

In the restricted space of all 2-point designs only (without any additional restriction), the optimal solution offers an improvement of up to 88%. The Cramer-Rao bound is attained with a sample size of only 100, whereas the theoretical solution needs approximately 300 samples to reach the Cramer-Rao bound. Hence, even the optimal solution obtained from a restricted design space can offer a vast improvement over the theoretical solution.

If we impose the additional restriction of point symmetry in the design space (weights unrestricted), the optimal solution offers an improvement of up to 86%. Hence, even with the addition of a further restriction of point symmetry in the design space, we observe a vast improvement over the theoretical solution.

To gain computational efficiency, if in addition we fix the weights of the 2-point design to match the weights of the theoretical solution, the optimal solution still offers an improvement of up to 63%. On the other hand, if we fix the symmetric design points to match the theoretical solution, the optimal solution offers an improvement of up to 55%. Nonetheless, irrespective of which design space is chosen, the improvement over the theoretical solution is remarkable. It is also evident from the results that there is a trade-off between computational efficiency and the performance.

If higher performance is a priority, it is preferable to use the point symmetric design, as the performance is very close to the entire 2-point design space but with a much higher computational efficiency. In fact, it can be easily observed in Figure 1 that the optimal solution in the point-symmetric design space is quite different from the theoretical solution for smaller sample sizes. As expected, as sample size grows, the two solutions tend to converge.

If computational efficiency is the priority, it is preferable to use the point symmetric design with fixed weight, as the performance is better than the point symmetric design with fixed weight with similar computational efficiency.

From a practitioner perspective, it may be prohibitive to perform a grid search to obtain the finite-sample optimal design. In a future communication, the authors will provide a comprehensive table for the finite sample A-optimal designs for different values of α and sample sizes. It would suffice to have the table for $\beta = 1$ only, since β can and will be rescaled to 1. We will also address other important optimality problems; for example, the estimation of percentiles, median effective dose, etc.

Finally, it should be noted that in Tables 2−5, the gain inefficiency increases and then decreases with increased sample size. This may appear counter-intuitive as we expect a monotonic behavior with increased sample sizes. However, it can be explained by the fact that when sample size is very small, we frequently encounter singularity issues in a logistic regression framework. This results in a lack of efficiency in terms of A-optimality criterion.

5. Limitations and Conclusions

There are two main limitations of the work. First, we have been unable to provide a theoretical solution to the finite sample problem. However, it is unclear if it is at all possible to obtain a theoretical solution to the problem. The second limitation is that our method does not provide A-optimal design for the entire unrestricted design space. However, the solution from the space of all 2-point designs is close to the true A-optimal solution for a relatively small sample size as evidenced by the A-optimality criterion values being close to the Cramer-Rao bound.

The fundamental conclusion from this article is that the asymptotic theoretical A-optimal solution for a logistic dose response performs poorly in minimizing the sum of variances of the parameters for small finite samples. To our knowledge, this is the first article studying the finite sample characteristics of A-optimality in a dose response model. This finding in of itself is quite significant as it is customary to use the asymptotic theoretical solution in the finite sample case.

Acknowledgements

The authors would like to acknowledge Professor Bikas K. Sinha for many enriching discussions on the topic of optimal design. The concept behind this current work was motivated from such discussions and our previous works.

References

Abdelbasit, K. M. and Plackett, R. L. (1983). Experimental design for binary data. *Journal of the American Statistical Association*, **78**, 90-98.

Biederman, S., Dette, H. and Zhu, W. (2006). Optimal Designs for Dose-Response Models with Restricted Design Spaces. *Journal of the American Statistical Association*, **101(474)**, 747-759.

Ford, I., Torsney, B. and Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society*, **B54**, 569-583.

Hedayat, A. S., Yan, B. and Pezzuto, J. M. (1997). Modeling and identifying optimum designs for fitting dose-response curves based on raw optical density data. *Journal of the American Statistical Association*, **92**, 1132-1140.

Mathew, T. and Sinha, B. K. (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference*, **93**, 295–307.

Minkin, S. (1987). Optimal design for binary data. *Journal of the American Statistical Association*, **82**, 1098-1103.

Nandy, K. and Nandy, R. (2015). A study of the performance of 2-stage adaptive optimal designs in a logistic dose-response model. *Communications in Statistics - Simulation and Computation*, DOI: 10.1080/03610918.2015.1089284.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, Philadelphia (ISBN:978-0-89871-604-7).

Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, **37**: 81–89. MR 0015748

Sitter, R. R. and Wu, C. F. J. (1993). Optimal designs for binary response experiments: D-, and A- criteria. *Scandinavian Journal of Statistics*, **20**, 329-342.

Yang, M. (2008). A-optimal designs for generalized linear model with two parameters. *Journal of Statistical Planning and Inference*, **138**, 624–641.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for non-linear models with two parameters. *Annals of Statistics*, **37**, 518-541.

Publisher
Society of Statistics, Computer and Applications
B - 133, Ground Floor, C.R. Park, New Delhi - 110019
Tele: 011 - 40517662
https://ssca.org.in/
statapp1999@gmail.com
2020

Printed by : Galaxy Studio & Graphics
Mob: +91 9818 35 2203, +91 9582 94 1203