Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 7–11
http://www.ssca.org.in/journal

Calyampudi Radhakrishna Rao - A Collaborator and a Statistician for the Ages

A. S. Hedayat[1] and John Stufken[2]

[1]*Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago*

[2]*Department of Statistics, George Mason University*

Received: 05 June 2024; Revised: 08 June 2024; Accepted: 09 June 2024

Abstract

C. R. Rao has long been one of the most visible names in statistics and beyond. Even now that he is no longer with us, his work will provide inspiration for researchers for many years to come. After a few broad observations about C. R. Rao's incredible impact on statistics and science, we focus on two areas of contribution by C. R. Rao that are perhaps not as broadly known: Orthogonal arrays and sampling plans for excluding contiguous units. These areas, which overlap with our own interests, demonstrate how C. R. Rao's search for better solutions to statistical problems led him towards defining and studying combinatorial structures.

Key words: Orthogonal arrays; Sampling plans excluding contiguous units

AMS Subject Classifications: 62K05, 05B05

1. Introduction

Statistical science houses many distinguished researchers who have made highly influential contributions. If a hundred experienced statistical researchers were asked to make a list of 20 giants in the field, past or present, these lists would probably all look very impressive. But, most likely, there would only be few names that appeared on virtually every list. Those would be the names of the superstars, the visionaries, the trailblazers, the titans of the discipline. One of those very few names would be that of Calyampudi Radhakrishna Rao, better known to most as C. R. Rao (shortened to CR Rao from hereon).

With his many seminal contributions to statistics, CR Rao has been an inspiration to uncountable number of researchers, and will continue to be so for many more years to come. The staying power of his contributions, many of which were far ahead of their time, is astonishing and a testament to the depth and influence of the contributions. These contributions will live on and many will be taught to future generations of statisticians and scientists. CR Rao will undoubtedly be a statistician for the ages.

Corresponding Author: A. S. Hedayat
Email: hedayat@uic.edu

As others have noted, the impact of CR Rao's contributions were not limited to the field of statistics alone. This is, for example, crystal clear from the citation for his National Medal of Science awarded by US President George W. Bush. It reads:

"For his pioneering contributions to the foundations of statistical theory and multivariate statistical methodology, and their applications, enriching the physical, biological, mathematical, economic and engineering sciences."

It is therefore no surprise that journals like Science (Banks and Clarke, 2023) and Nature (Peddada and Khattree, 2023) are among the many that paid tribute to CR Rao shortly after his death. CR Rao's devotion to science, along with his perspective on statistics, is also captured in the quotation featured on the website of the CR Rao Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS) on the University of Hyderabad Campus:

"We study physics to solve problems in physics, chemistry to solve problems in chemistry, and botany to solve problems in botany. There are no statistical problems that we solve using statistics. We use statistics to provide a course of action with minimum risk in all areas of human endeavor with unavailable evidence."

For those interested to learn more about the life of CR Rao, about the person that he was, about his major contributions, and about his many awards, the AIMSCS web pages at `https://crraoaimscs.res.in/` and various publications (*e.g.*, Bera and Ghosh, 2021, and O'Grady, 2023) provide stories and perspectives that are more interesting and informed than anything we can offer on these aspects. Therefore, while we have benefited greatly and drawn enormous inspiration from many of CR Rao's contributions, we reminisce briefly on two areas of common research interest.

2. Orthogonal arrays

Just as for various other seminal contributions by CR Rao, he introduced orthogonal arrays (OAs) in the 1940s. In fact, as communicated by CR Rao in his Foreword in Hedayat, Sloane and Stufken (1999), he introduced a subclass of OAs in a chapter of his MS thesis in 1943. Rao (1946) reports on this subclass, which he called hypercubes of strength d. This was followed by Rao (1947, 1949), in which the complete class of OAs was studied even though the name orthogonal arrays had not yet been introduced.

Based on CR Rao's writing in the aforementioned Foreword, when he joined ISI in 1941 to study statistics, he was surprised to see a broad research interest in combinatorial structures, which was primarily fueled by the interests of RC Bose and his collaborators. With a strong background in mathematics, CR Rao was quickly able to become an important contributor in this arena, leading to his work on hypercubes of strength d (Rao, 1946) and a collaboration with K.R. Nair (Nair and Rao, 1948). He formulated the more general definition of OAs not until 1947, after having moved to study in Cambridge, UK.

Formally, an OA of strength t based on s symbols, N runs and k factors is an $N \times k$ array with entries from the set of s symbols so that for every $N \times t$ subarray every possible t-tuple based on the s symbols appears equally often as a row of the subarray. Given that there are s^t possible t-tuples, this common number must be N/s^t, which is also known as the index of the OA. Such an array is often denoted by OA(N, k, s, t), while the index is often written as λ.

CR Rao introduced OAs because of their properties for use in fractional factorial experiments. This interest is visible in Nair and Rao (1948), as also in Rao (1947). The bounds for the number of runs in an OA given by Rao (1947) are easily understood if one understands the relationship between the strength of an OA and models for which all parameters are estimable when using the runs of an OA to form a fractional factorial. While statistical properties motivated CR Rao, he was also interested in combinatorial aspects of OAs, as is clearly shown by Rao (1949). Various methods of construction are discussed there, including a simplified construction of OAs that already appeared in Rao (1946). A similar idea appeared shortly afterwards in Hamming (1950) for the construction of error-correcting codes, which led Hedayat, Sloane, and Stufken (1999) to refer to these arrays as Rao-Hamming OAs.

CR Rao returned to his interest in OAs on several later occasions, such as in Rao (1961, 1973).

After so many years since their introduction, OAs remain an active area of research, both in statistics and mathematics. They also continue to be used frequently in factorial experiments.

3. Sampling plans excluding contiguous units

When sampling from a finite population $U = \{1, \ldots, N\}$, a fixed-size n sampling plan, where $n < N$, can be presented as $\{(s_\ell, p_\ell), \ell = 1, 2, ..., m\}$, where the s_ℓ's are distinct subsets of size n from U, p_ℓ is the probability that s_ℓ is the sample outcome, and m is the support size of the sampling plan. The first-order inclusion probability π_i for unit $i \in U$ is defined as the probability that unit i is in the selected sample, and can be computed as

$$\pi_i = \sum_{s_\ell \ni i} p_\ell, \quad i = 1, \ldots, N.$$

Similarly, for two distinct units i and j, the second-order inclusion probability π_{ij} is the probability that both the units are in the selected sample. Thus,

$$\pi_{ij} = \sum_{s_\ell \ni i,j} p_\ell, \quad i, j = 1, \ldots, N, i \neq j.$$

A fixed-size n sampling plan for which $\pi_i = n/N$ and $\pi_{ij} = n(n-1)/(N(N-1))$ has, with respect to the first- and second-order inclusion probabilities, the same characteristics as simple random sampling (SRS). The plan is identical to SRS if $m = \binom{N}{n}$, so that the s_ℓ's consist of all subsets of U of size n, and $p_\ell = 1/m$ for every ℓ.

A common problem would assume that unit $i \in U$ has value Y_i for a certain characteristic (*e.g.*, income or years of secondary education). A study might be interested in the population total $T = \sum_{i=1}^{N} Y_i$ or the population mean T/N. For a large population, observing all Y_i's (a census) might be too time consuming or expensive. Moreover, it would be unnecessary since accurate results can be obtained from a random sample based on a sampling plan with relatively small sample size n. A design-based unbiased estimator of T based on a sample s_ℓ is in that case given by the Horvitz-Thompson estimator

$$\hat{T} = \sum_{i \in s_\ell} \frac{Y_i}{\pi_i}.$$

An unbiased estimator for the variance of $\hat{T}$ exists if and only if $\pi_{ij} > 0$ for all units $i, j \in U, i \neq j$ (*cf.* Hedayat and Sinha, 1991).

If the units can be thought of as naturally being placed on a straight line or circle to depict their proximity (*e.g.*, housing units on a street or street block), then it is conceivable for some characteristics (*e.g.*, household income) that neighboring units provide similar information. One might get a better estimate for the population total by avoiding the selection of contiguous units. This idea was explored in Hedayat, Rao and Stufken (1988a, 1988b). Considering the population units to be ordered on a circle, so that each unit has two contiguous units, these authors define a fixed-size n sampling plan to be a balanced sampling plan without contiguous units if all first-order inclusion probabilities are equal (this common value must be n/N), the second-order inclusion probabilities for contiguous units are 0, and all other second-order inclusion probabilities are equal (this common value must be $n(n-1)/(N(N-3))$. They show that, in terms of the variance of the Horvitz-Thompson estimator, balanced sampling plans without contiguous units are more efficient than SRS if the first-order serial correlation of the Y_i's exceeds $-1/(N-1)$ (which it will if there is any validity to the premise that contiguous units have similar Y_i's).

One may also wonder when such balanced sampling plans without contiguous units exist. Hedayat, Rao and Stufken (1988a) show by construction that such plans exist for every value of $N \geq 3n$ when $n = 3$ or 4.

In later years, there have been various extensions of these initial results. This includes, for example, requiring second-order inclusion probabilities to be 0 not only for immediate neighbors (*cf.* Stufken, 1993), considering population units to be ordered in a 2-dimensional layout (*cf.* Wright, 2008), and additional existence results, including in the combinatorial literature (*cf.* Guo, Wang, and Feng, 2022). Also, alternative methods have been developed for spatial populations (*cf.* Deville and Tillé, 1998).

4. In conclusion

While C. R. Rao's passing is an enormous loss for the scientific community, his widespread influential contributions provide assurance that his work will be with us for many years to come. In fact, we have no doubt that his contributions will remain an inspiration for budding statistical researchers. In that way, CR Rao is truly a statistician for the ages.

References

Banks, D. and Clarke, J. L. (2023). C. R. Rao (1920–2023). *Science*, **382**, 771–771.

Bera, A. and Ghosh, P. (2021). Glimpses from the life and work of Dr. C. R. Rao: A living legend in statistics. *A Tribute to the Legend of Professor CR Rao: The Centenary Volume*, **1**, 49–65.

Deville, J.-C. and Tille, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika*, **85**, 89–101.

Guo, C., Wang, X., and Feng, T. (2022). Cyclic balanced sampling plans excluding contiguous units with block size four. *Discrete Mathematics*, **345**, 112899.

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell System Technical Journal*, **29**, 147–160.

Hedayat, A., Rao, C., and Stufken, J. (1988a). Designs in survey sampling avoiding contiguous units. *Handbook of Statistics*, **6**, 575–583.

Hedayat, A., Rao, C., and Stufken, J. (1988b). Sampling plans excluding contiguous units. *Journal of Statistical Planning and Inference*, **19**, 159–170.

Hedayat, A. and Sinha, B. K. (1991). *Design and Inference in Finite Population Sampling*. John Wiley & Sons.

Hedayat, A. S., Sloane, N. J. A., and Stufken, J. (2012). *Orthogonal Arrays: Theory and Applications*. Springer Science & Business Media.

Nair, K. and Rao, C. (1948). Confounded designs for asymmetrical factorial experiments. *Journal of the Royal Statistical Society. Series B (Methodological)*, **10**, 109–131.

O'Grady, C. (2023). "A life to enlighten us": Remembering C. R. Rao, 1920–2023. *Significance*, **20**, 8–12.

Peddada, S. D. and Khattree, R. (2023). C. R. Rao, statistician who transformed data analytics (1920–2023). *Nature*, **622**, 691–691.

Rao, C. R. (1946). Hypercubes of strength'd' leading to confounded designs in factorial experiments. *Bulletin of the Calcutta Mathematical Society*, **38**, 67–78.

Rao, C. R. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *Supplement to the Journal of the Royal Statistical Society*, **9**, 128–139.

Rao, C. R. (1949). On a class of arrangements. *Proceedings of the Edinburgh Mathematical Society*, **8**, 119–125.

Rao, C. R. (1961). Combinatorial arrangements analogous to orthogonal arrays. *Sankhyā: The Indian Journal of Statistics, Series A*, **23**, 283–286.

Rao, C. R. (1973). Some combinatorial problems of arrays and applications to design of experiments. In *A Survey of Combinatorial Theory*, pages 349–359. Elsevier.

Stufken, J. (1993). Combinatorial and statistical aspects of sampling plans to avoid the selection of adjacent units. *Journal of Combinatorics, Information and System Sciences*, **18**, 81–92.

Wright, J. H. (2008). Two-dimensional balanced sampling plans excluding adjacent units. *Journal of Statistical Planning and Inference*, **138**, 145–153.