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,

$$Var(\hat{\alpha}) + Var(\hat{\beta}) \geq \sum_{i=1}^{m} \xi_i \frac{e^{-(\alpha+\beta x_i)}}{\left(1+e^{-(\alpha+\beta x_i)}\right)^2}(1 + x_i^2)/|\boldsymbol{I}(\boldsymbol{\alpha},\boldsymbol{\beta})|, \tag{2}$$

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

$$\frac{c^2-\alpha^2-\beta^2}{\sqrt{(\beta^2+(c+\alpha)^2)}+\sqrt{(\beta^2+(c-\alpha)^2)}} = 1 + \frac{c(1-e^c)}{1+e^c}. \tag{4}$$

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 $\beta = 1$. 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)	$\boldsymbol{c_A^*}$	$\boldsymbol{\xi_{search}^*}$	$\boldsymbol{A_{search}^*}$	***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

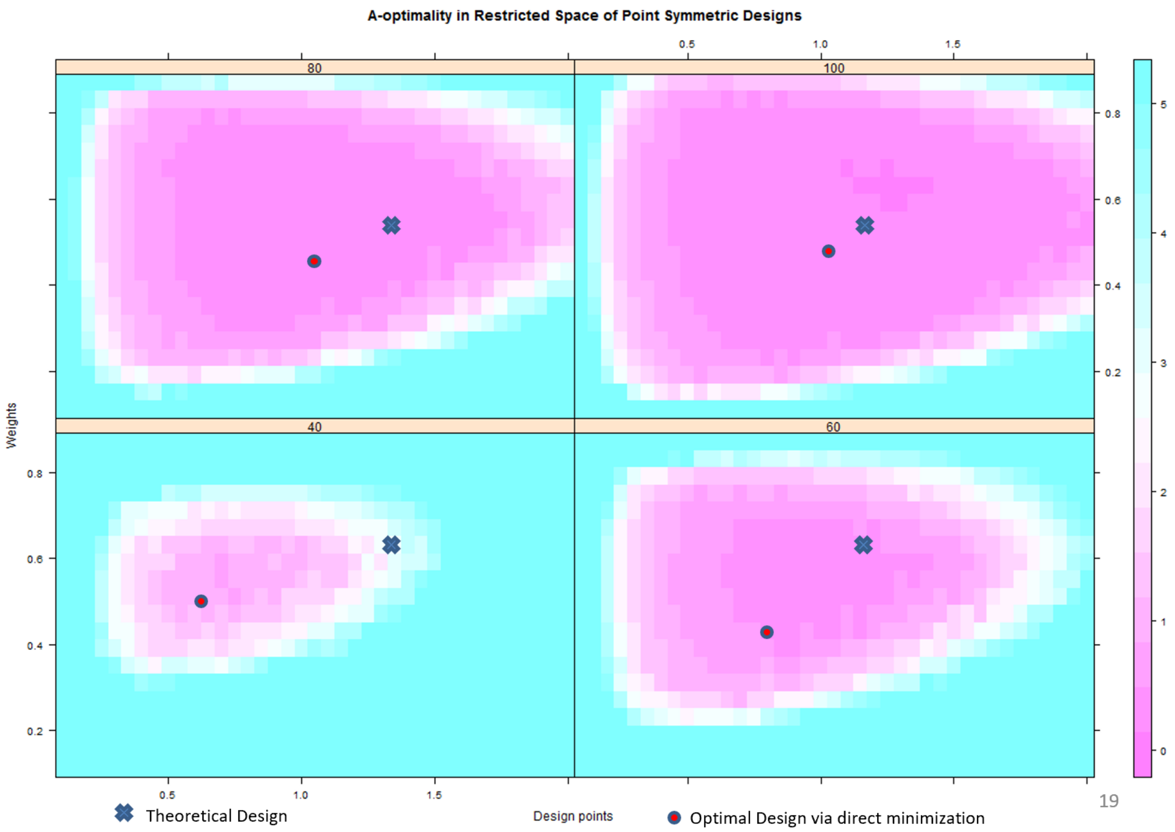

Figure 1: Comparison of the theoretical versus direct minimization in point symmetric design space

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.