UCLA

Department of Statistics Papers

Title

G-Optimal Designs for Multi-Factor Experiments with Heteroscedastic Errors

Permalink

<https://escholarship.org/uc/item/6d8250fz>

Author

Weng-Kee Wong

Publication Date

2011-10-24

G-optimal designs for multi-factor experiments with heteroscedastic errors

Weng Kee Wong

Department of Biostatistics, University of California at Los Angeles, Los Angeles, CA 90024, USA

Received 26 April 1993; revised manuscript received 20 July 1993

Abstract

We consider constructing G-optimal designs for multi-factor experiments when the errors are heteroscedastic. The method is illustrated with examples assuming the Kronecker product models and the additive models. Sufficient conditions are given so that G-optimal designs for the multi-factor experiments can be constructed from the optimal designs for the single-factor experiments.

Key words: Additive models; efficiency function; information matrices; Kronecker product models

1. Introduction

This paper considers constructing multi-factor optimal designs for regression problems where the goal is to estimate the response surface. It is assumed that it is not known in advance which point or points on the response surface might be of interest to estimate before the experiment is run, and the errors are heteroscedastic. The statistical problem here is to determine a good experimental design for this situation.

The proposed optimality criterion is G-optimality: first we consider the variance of the least-squares fit at some point x in the design space and then seek a design which minimizes the maximum variance over all points on the design space. The resulting design is called G-optimal. Interest in such designs when the errors are homoscedastic has been intense especially in the 1960s and 1970s, see Kiefer (1985) and the references therein.

Most works to date for finding optimal designs in multi-factor experiments involves a differentiable optimality criterion, Rafajlowicz and Myszka (1988, 1992), Lau (1988) and Lim and Studden (1988). Optimal designs under a non-differentiable criterion, on

Correspondence to: Dr. W.K. Wong, Department of Biostatistics, University of California at Los Angeles, Los Angeles, CA 90024, USA.

the other hand, are relatively less studied (Gaffke, 1985, 1987; Pukelsheim, 1980) and there appears to be a lack of concrete examples on the construction of such designs in multi-factor experiments. This paper seeks to fill this gap by considering explicit construction of G-optimal designs in multi-factor experiments. While we focus only on G-optimal designs, the method described here can be amended to construct optimal designs for other types of non-differentiable criteria as well, see Wong (1992).

Let the design space be Ω which is assumed to be a Cartesian product of k one-dimensional compact Euclidean subspaces Ω_i . Notationally, we write $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_k$ and $x = (x_1, x_2, \dots, x_k) \in \Omega$ with $x_i \in \Omega_i$, $i = 1, 2, \dots, k$. For expository purposes, we shall assume $\Omega_i = [-1, 1]$ for $i = 1, 2, \dots, k$. Let $f(x)$ denote the regression function and $\lambda(x)$ be a positive function reciprocally proportional to the variance of the response at the point x . Both $\lambda(x)$ and $f(x)$ are known and continuous on Ω . In what follows, it will be convenient to denote a design problem by the triplet $(\Omega, f(x), \lambda(x))$. The model is

$$y(x) = f^T(x)\beta + e/\lambda(x)^{1/2},$$

where $y(x)$ is a univariate response at the point x , β is the vector of parameters, and e is a random error term having mean zero and variance κ . The covariance structure is given by $\text{cov}(y(x_i), y(x_j)) = \kappa/\lambda(x_i)$ if $i = j$ and 0 otherwise.

In this paper, only continuous designs are considered. This means all designs are probability measures on Ω . The analog of the $X^T X$ matrix for a continuous design ξ is its normalized information matrix defined by $M(\xi) = \int_{\Omega} f(x)f^T(x)\lambda(x)\xi(dx)$. Ignoring an unimportant constant, the covariance matrix of the least-squares estimates for β using design ξ is $M^{-1}(\xi)$ and the variance of the fitted value at the point x is $d(x, \xi) = f^T(x)M^{-1}(\xi)f(x)$ if the inverse exists. Otherwise, the variance at that point is assigned the value $+\infty$. For our purpose here, we shall focus on only non-singular designs, i.e. designs for which their information matrices are non-singular. Formally, a design ξ_G is G-optimal if it minimizes $\max_{x \in \Omega} d(x, \xi)$ over the set of continuous designs on Ω .

For a given design ξ , let $A(\xi) = \{a \in \Omega \mid d(a, \xi) = \max_{x \in \Omega} d(x, \xi)\}$, which is the set of points in Ω which maximizes the variance function for ξ . As shown in Wong and Cook (1993), this set plays a crucial role in verifying if a design is G-optimal. Specifically, a design $\xi = \xi_G$ is G-optimal if and only if there exists a probability measure γ on $A(\xi)$ such that

$$g(x, \gamma, \xi) \leq 0 \quad \text{for all } x \text{ in } \Omega \quad (1.1)$$

where $g(x, \gamma, \xi) = \int_{A(\xi)} \{\lambda(x)d^2(x, a, \xi) - d(a, \xi)\}\gamma(da)$ and $d(x, a, \xi) = f^T(x)M^{-1}(\xi)f(a)$.

Verification of (1.1) may be tedious even for a relatively simple design problem. Current research is underway to develop an effective strategy for finding a probability measure γ , if in fact the design ξ is G-optimal. Some suggestions for doing this were given in Wong and Cook (1993). Following Wong and Cook (1993), any γ that satisfies (1.1) is called an A -measure for the design ξ .

2. Main results

In this section, G-optimal designs for multi-factor experiments are constructed for a class of regression problems. Specifically if ξ_i is G-optimal for the design problem $(\Omega_i, f_i(x_i), \lambda_i(x_i))$, $i = 1, 2, \dots, k$, we want to find a G-optimal design for $(\Omega, f(x), \lambda(x))$ where

$$\lambda(x) = \prod_{i=1}^k \lambda_i(x_i)$$

and $f(x)$ is one of two forms: (a) $f^T(x) = (f_1^T(x_1), f_2^T(x_2), \dots, f_k^T(x_k))$ and (b) $f(x) = f_1(x_1) \otimes f_2(x_2) \otimes \dots \otimes f_k(x_k)$. The notation $\otimes$ stands for the usual Kronecker product. Models with regression functions of the form (a) and (b) are called additive models and Kronecker Product models, respectively. These two types of models represent important cases for multi-factor experiments, see, for example, Hoel (1965) and Rafajlowicz and Myszka (1988, 1992). Later, we relax the functional form

$$\lambda(x) = \prod_{i=1}^k \lambda_i(x_i)$$

and give a result for a more general model, which includes additive and Kronecker product models as special cases.

The G-optimal designs to follow are all product designs obtained by ‘multiplying’ the G-optimal designs from the single-factor experiments. Product designs are denoted by $\xi = \xi_1 \times \xi_2 \times \dots \times \xi_k$ where ξ_i is a design on Ω_i , $i = 1, 2, \dots, k$. They are typically found first by guesswork followed by verification of its optimality using (1.1). Product designs are also considered in Lau (1988), Lim and Studden (1988) and Rafajlowicz and Myszka (1988, 1992). Our first result concerns Kronecker product models.

Theorem 2.1. *Let $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_k$, $f(x) = f_1(x_1) \otimes f_2(x_2) \otimes \dots \otimes f_k(x_k)$ and ξ_i is a G-optimal design for the design problem $(\Omega_i, f_i(x_i), \lambda_i(x_i))$, $i = 1, 2, \dots, k$. The G-optimal design for the problem $(\Omega, f(x), \prod_{i=1}^k \lambda_i(x_i))$ is $\xi_G = \xi_1 \times \xi_2 \times \dots \times \xi_k$.*

Proof. The key step in the proof lies in observing that the variance function for ξ_G is

$$\begin{aligned} d(x, \xi_G) &= f^T(x) M^{-1}(\xi_G) f(x) = \prod_{i=1}^k f_i^T(x_i) M_i^{-1}(\xi_i) f_i(x_i) \\ &= \prod_{i=1}^k d_i(x_i, \xi_i). \end{aligned} \quad (2.1)$$

To complete the proof, let $d_i(x_i, a_i, \xi_i) = f_i^T(x_i) M_i^{-1}(\xi_i) f_i(a_i)$, γ_i be an A -measure for ξ_i , $i = 1, 2, \dots, k$ and note that

$$\begin{aligned}
 g(x, \gamma, \xi_G) &= \int_{A(\xi_k)} \cdots \int_{A(\xi_1)} \prod_{i=1}^k \lambda_i(x_i) d_i^2(x_i, a_i, \xi_i) \gamma_1(da_1) \cdots \gamma_k(da_k) \\
 &\quad - \prod_{i=1}^k d_i(a_i, \xi_i) \leq \int_{A(\xi_k)} \cdots \int_{A(\xi_2)} \left\{ \prod_{i=2}^k \lambda_i(x_i) d_i^2(x_i, a_i, \xi_i) \gamma_2(da_2) \right. \\
 &\quad \left. \cdots \gamma_k(da_k) - \prod_{i=2}^k d_i(a_i, \xi_i) \right\} d_1(a_1, \xi_1) \\
 &\leq \cdots \leq \left\{ \int_{A(\xi_k)} \{ \lambda_k(x_k) d_k^2(x_k, a_k, \xi_k) \gamma_k(da_k) \right. \\
 &\quad \left. - d_k(a_k, \xi_k) \right\} \prod_{i=1}^{k-1} d_i(a_i, \xi_i) \\
 &\leq 0 \quad \text{for all } x \in \Omega.
 \end{aligned}$$

The above result does not apply to additive models with heteroscedastic errors. As a counter-example, one may take $\Omega_1 = \Omega_2 = [-1, 1]$, $f^T(x) = (1, x_1, x_2)$ and, $\lambda(x) = \lambda_1(x_1) \lambda_2(x_2)$ where $\lambda_1(x_1) = x_1 + 2$, and $\lambda_2(x_2) = x_2^2 + 1$. Generally, more restrictive assumptions are required for additive models as the following result shows.

Theorem 2.2. Suppose ξ_i is G-optimal for the design problem $(\Omega_i, f_i(x_i), \lambda_i(x_i))$ and γ_i is an A -measure for ξ_i , $i = 1, 2, \dots, k$. Let $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_k$, $f^T(x_1, x_2, \dots, x_k) = (f_1^T(x_1), f_2^T(x_2), \dots, f_k^T(x_k))$ and $\lambda(x) = \prod_{i=1}^k \lambda_i(x_i)$. Then $\xi_G = \xi_1 \times \xi_2 \times \cdots \times \xi_k$ is G-optimal for the problem $(\Omega, f(x), \lambda(x))$ if

- (i) $\int_{\Omega_i} \lambda_i(x_i) f_i(x_i) \xi_i(dx_i) = 0$, $i = 1, 2, \dots, k$,
- (ii) $\int_{A(\xi_i)} f_i(a_i) \gamma_i(da_i) = 0$, $i = 1, 2, \dots, k$,
- (iii) $\prod_{1 \leq j \neq i \leq k} \max_{x_j \in \Omega_j} \lambda_j(x_j) = \prod_{1 \leq j \neq i \leq k} \int_{\Omega_j} \lambda_j(x_j) \xi_j(dx_j)$, $i = 1, 2, \dots, k$.

Proof. Let $M_i(\xi_i) = \int_{\Omega_i} f_i(x_i) f_i^T(x_i) \lambda_i(x_i) \xi_i(dx_i)$ and $c_i = \prod_{1 \leq j \neq i \leq k} \int_{\Omega_j} \lambda_j(x_j) \xi_j(dx_j)$, $i = 1, 2, \dots, k$. By assumption (i), the information matrix for ξ_G is block diagonal with the i th block equal to $c_i M_i(\xi_i)$. Simple algebra shows the variance function of ξ_G is

$$d(x, \xi_G) = d_1(x_1, \xi_1)/c_1 + d_2(x_2, \xi_2)/c_2 + \cdots + d_k(x_k, \xi_k)/c_k,$$

and $A(\xi_G) = A(\xi_1) \times A(\xi_2) \times \cdots \times A(\xi_k)$. If $\gamma = \gamma_1 \times \gamma_2 \times \cdots \times \gamma_k$, it follows that

$$\begin{aligned}
\int_{A(\xi_G)} \lambda(x) d^2(x, a, \xi_G) \gamma(da) &= \sum_{i=1}^k \left\{ \prod_{1 \leq j \neq i \leq k} \lambda_j(x_j)/c_i \right\} \\
&\times \left\{ \int_{A(\xi_i)} \lambda_i(x_i) d_i^2(x_i, a_i, \xi_i)/c_i \gamma_i(da_i) \right\} \\
&+ \int_{A(\xi)} \lambda(x) \sum_{1 \leq j \neq i \leq k} d_i(x_i, a_i, \xi_i) \\
&\times d_j(x_j, a_j, \xi_j)/(c_i c_j) \gamma_i(da_i) \gamma_j(da_j) \\
&\leq \sum_{i=1}^k d_i(a_i, \xi_i)/c_i = \max_{x \in \Omega} d(x, \xi_G)
\end{aligned}$$

where the last inequality follows from assumptions (ii), (iii) and the optimality of ξ_i , $i = 1, 2, \dots, k$.

Remark 2.1. Conditions (i) and (ii) ensure that the moment matrix of the product design and the subgradient at that matrix are block diagonal matrices built in the individual moment matrices and subgradients, respectively. Condition (iii) implies that the G-optimal design for each of the single-factor experiment is supported on the set $\{x_i \in \Omega_i \mid \max_{y \in \Omega_i} \lambda_i(y) = \lambda_i(x_i)\}$.

An application of the Theorem 2.2 is to the case when $\Omega = [-1, 1]^k$, $f^T(x) = (x_1, x_2, \dots, x_k)$ and $\lambda_i(x_i)$ is symmetric on $[-1, 1]$, $i = 1, 2, \dots, k$. The G-optimal for each of the single-factor experiment can be shown to be symmetric so that assumptions (i) and (ii) in Theorem 2.2 hold trivially. Consequently, if condition (iii) holds, the G-optimal design for the multi-factor experiment is a product of the G-optimal designs for the single-factor experiments. Under further conditions, a result similar to Theorem 2.2 can be established for a more general $\lambda(x)$ and $f(x)$.

Theorem 2.3. Let $\Omega = [-1, 1]^k$ and $\lambda(x)$ is constant on the set $B = \{\beta = (\beta_1, \beta_2, \dots, \beta_k) \in \Omega \text{ and } \beta_i = 1 \text{ or } \beta_i = -1, i = 1, 2, \dots, k\}$. For $j = 1, 2, \dots, k$, let $f_j^T(x)$ be the $\binom{k}{j} \times 1$ vector consisting of all the j -crossproduct terms among the k factors $x_1, x_2, \dots, x_k$, $F_j^T(x) = (1, f_1^T(x), f_2^T(x), \dots, f_j^T(x))$ and,

$$J_j(x) = 1 + \sum_{1 \leq i \leq k} x_i^2 + \sum_{1 \leq i_1 < i_2 \leq k} x_{i_1}^2 x_{i_2}^2 + \dots + \sum_{1 \leq i_1 < \dots < i_j \leq k} x_{i_1}^2 x_{i_2}^2 \dots x_{i_j}^2.$$

Then if η is the design which places equal mass at 1 and -1 , the G-optimal for the problem $(\Omega, F_j(x), \lambda(x))$ is $\xi_G = \eta \times \dots \times \eta$ (k copies) provided

$$\lambda(x) J_j(x) \leq \lambda(\beta) J_j(\beta) \quad \text{for all } x \in \Omega. \quad (2.2)$$

Proof. Fix an integer j between 1 and k and denote an arbitrary element in the set B by β . Then the information matrix for ξ_G is a

$$\sum_{i=0}^j \binom{k}{i} \times \sum_{i=0}^j \binom{k}{i}$$

diagonal matrix with every element on the diagonal equals to $\lambda(\beta)$. Since ξ_G has

$$2^k \left(\geq \sum_{i=0}^j \binom{k}{i} \right)$$

support points, $M_j(\xi_G)$ is always non-singular and the variance function for ξ_G at the point x is $F_j^T(x)F_j(x)/\lambda(\beta)$. It follows that this is a symmetric quadratic function in each of the x_i and so $A(\xi_G) = \{-1, 1\}^k$. Taking $\gamma = \xi_G$, it is straightforward to check (1.1) and (2.1) are equivalent, thereby completing the proof of the theorem.

Remark 2.2. A trivial sufficient condition for (2.1) to hold is $\lambda(\beta) \geq \lambda(y)$ for all y in Ω .

Remark 2.3. In Theorem 2.3, if $j=1$, $F_1^T(x) = (1, x_1, x_2, \dots, x_k)$ and if $j=k$, $F_k(x) = h_1(x_1) \otimes h_2(x_2) \otimes \dots \otimes h_k(x_k)$, $h_i^T(x_i) = (1, x_i)$, $i=1, 2, \dots, k$. Thus, additive model and the Kronecker product model are included in the class of models.

Remark 2.4. The G-optimal designs constructed from Theorem 2.3 has 2^k support points which means ξ_G has minimal number of support points only when the regression function is $F_k(x)$. If $k-j$ is large, the G-optimal design for $F_j(x)$ will have more support points than are actually needed, see Caratheodory's theorem (Fedorov, 1972, p. 66).

Acknowledgements

This research is partially supported by a NH grant CA16042. The author thanks two referees for their helpful comments and suggestions.

References

- Fedorov, V.V. (1972). Theory of Optimal Experiments. Translated and edited by Studden, W.J. and Klimko, E.M., Academic Press, New York.
- Gaffke, N. (1985). Direction derivatives of optimality criteria at singular matrices in convex design theory. *Statistics* **16**, 373–388.
- Gaffke, N. (1987). Further characterization of design admissibility for partial parameter estimation in linear regression models. *Ann. Statist.* **15**, 942–957.
- Hoel, P.G. (1965). Minimax designs in two dimensional regression. *Ann. Math. Statist.* **36**, 1097–1106.
- Kiefer, J.C. (1985). Jack Carl Kiefer Collected Papers III. Design of Experiments. Springer, New York.
- Lau, T.S. (1988). D-optimal designs on the unit q -ball. *J. Statist. Plann. Inference* **19**, 299–315.

- Lim, Y.B. and Studden, W.J. (1988). Efficient D_s -optimal designs for multivariate polynomial regression on the q -cube. *Ann. Statist.* **16**, 1225–1240.
- Pukelsheim, F. (1980). On linear regression designs which maximize information. *J. Statist. Plann. Inference* **4**, 339–364.
- Rafajlowicz, E. and Myszka, W. (1988). Optimum experimental design for a regression on a hypercube — generalization of Hoel's result. *Ann. Inst. Statist. Math.* **40**, 821–828.
- Rafajlowicz, E. and Myszka, W. (1992). When product type experimental design is optimal? Brief survey and new results. *Metrika* **39**, 321–333.
- Wong, W.K. (1992). A unified approach to the construction of minimax designs. *Biometrika*, **79**, 611–619.
- Wong, W.K. and Cook, R.D. (1993). Heteroscedastic G -optimal designs. *J. Roy. Statist. Soc. Ser. B.*, Vol 55, No. 4, 871–880.