

# A Gentle Introduction to Optimal Design for Regression Models

Timothy E. O'BRIEN and Gerald M. FUNK

---

This article demonstrates and underscores the equivalence between a variance-maximization exercise and the methodology involved in obtaining and verifying the optimal design for a key model function. It thus provides an alternate solution to the variance exercise as well as a means to introduce and illustrate the concepts of optimal design theory and practice in a simple and clear manner.

**KEY WORDS:** D-optimality; Design measure; G-optimality; General equivalence theorem; Optimal design theory; Variance-maximization.

---

## 1. INTRODUCTION

With only rare exception, instructors of experimental design and other courses in academic and industrial settings typically avoid addressing issues of optimal experimental design possibly due to the complexity of some of the material and publications in this field. This theory was developed by Smith (1918), Chernoff (1953), Kiefer (1959, 1961, 1974, 1975), Kiefer and Wolfowitz (1960), White (1973), and Dette and O'Brien (1999), and was discussed by Silvey (1980), Atkinson and Donev (1992), and Pukelsheim (1993). Thus, although this rich literature exists for optimal design theory, it is only rarely encountered by undergraduate and graduate students of statistics in their coursework. In the applied spirit of Brady and Allen (2002), because researchers are often very interested in obtaining efficient designs in practice, we feel that students should be exposed to their usefulness.

One reason instructors may find optimal design theory to be a difficult topic to cover in a course is that its methodology does not easily build on established results, and thus requires a good deal of preliminary definitions and results before meaningful applications can be made. This article shows how results from optimal design theory can be easily introduced and used to solve an interesting exercise from basic statistical theory, and thus also serves to bridge this gap in their coursework. We have had surprising success and positive feedback in distributing and discussing earlier versions of this note to students enrolled in undergraduate and graduate statistical theory and experimental design courses and in devoting roughly one or two class periods covering its contents.

## 2. A FAMILIAR STARTING POINT

Consider the following exercise encountered by students enrolled in a post-calculus undergraduate course in basic statistical

theory (see, e.g., DeGroot and Schervish 2002, p. 236, exercise 14). The discrete random variable  $X$  has probability mass function (pmf)  $\omega(x)$  which takes the value zero outside of the unit interval,  $[0, 1]$ , and the exercise is to find the pmf which maximizes the variance. Hence, we seek the (at most countably infinite) points  $x_1, x_2, \dots, x_k$  and corresponding weights  $\omega_1, \omega_2, \dots, \omega_k$ , with  $\omega_j = \omega(x_j)$ , which maximize the criterion function

$$CF = \sum_{j=1}^k \omega_j (x_j - \mu)^2,$$

where the mean  $\mu = \sum \omega_j x_j$ .

Students of elementary statistical theory, using algebraic results and properties of variances, can show that the optimal mass function places mass  $\omega = 1/2$  at each of the points  $x_1 = 0$  and  $x_2 = 1$ . However, it turns out that this rather simple exercise also provides an excellent example useful to introduce and illustrate optimal design theory and methods, thereby giving an alternative proof of the above result and a means to demonstrate uniqueness and optimality of this solution.

## 3. A BASIC INTRODUCTION TO OPTIMAL DESIGN THEORY

This section briefly reviews basic optimal design theory for regression models; readers interested in more extensive developments are referred to Silvey (1980), Box and Draper (1987, chap. 14), Seber and Wild (1989, sec. 5.13), and Atkinson and Donev (1992). For a given linear or nonlinear regression model function,  $E(Y) = \eta(\mathbf{x}, \boldsymbol{\theta})$  with unknown  $p \times 1$  parameter vector  $\boldsymbol{\theta}$ , the optimal design problem typically involves choosing a design with  $k$  design points,  $\xi$ , to best estimate the parameter vector. Here  $\xi$  can be written as

$$\xi = \left\{ \begin{array}{l} x_1, x_2, \dots, x_k \\ \omega_1, \omega_2, \dots, \omega_k \end{array} \right\},$$

where the  $k$  design support points  $x_1, x_2, \dots, x_k$  are elements of the design space  $\Delta$ , and the associated weights  $\omega_1, \omega_2, \dots, \omega_k$  are nonnegative real numbers which sum to one. For the design  $\xi$ , the  $k \times p$  Jacobian matrix,  $\mathbf{V} = \partial\boldsymbol{\eta}/\partial\boldsymbol{\theta}$  with  $st$ th element  $\partial\eta(x_s)/\partial\theta_t$ , depends on  $\boldsymbol{\theta}$  for nonlinear models and does not for linear models. When the model errors are normally distributed, the corresponding (Fisher) information matrix associated with  $\boldsymbol{\theta}$  is given by

$$\mathbf{M} = \mathbf{M}(\xi, \boldsymbol{\theta}) = \sum_{j=1}^k \omega_j \frac{\partial\boldsymbol{\eta}(x_j)}{\partial\boldsymbol{\theta}} \frac{\partial\boldsymbol{\eta}(x_j)}{\partial\boldsymbol{\theta}^T} = \mathbf{V}^T \boldsymbol{\Omega} \mathbf{V},$$

where  $\boldsymbol{\Omega} = \text{diag}\{\omega_1, \omega_2, \dots, \omega_k\}$ .

Because the asymptotic covariance matrix associated with the maximum likelihood estimate of  $\boldsymbol{\theta}$  is proportional to  $\mathbf{M}^{-1}$ , variance-optimal designs minimize some convex function of  $\mathbf{M}^{-1}$ . The most popular regression design criterion is D-optimality (Silvey 1980, p. 40), where designs are sought to

---

Timothy E. O'Brien is Assistant Professor, and Gerald M. Funk is Associate Professor, Department of Mathematics and Statistics, Loyola University Chicago, Chicago, IL 60626 (E-mail: tobrien@math.luc.edu).

minimize the determinant of  $\mathbf{M}^{-1}$  (or equivalently, to maximize the determinant of  $\mathbf{M}$ ). Reasons for the popularity of this criterion were given by Seber and Wild (1989, p. 250) and include the fact that D-optimal designs are invariant to an (even nonlinear) reparameterization of the model function. The proof of this invariance involves an argument identical to the one given in the proof of Theorem 1 in Dette and O'Brien (1999, p. 97). Further, as discussed by Atkinson and Donev (1992, p. 44), the standardized (predicted) variance function which corresponds to the assumed model function and design is given by

$$d(x, \xi, \theta) = \frac{\partial \eta(x)}{\partial \theta^T} \mathbf{M}^{-1}(\xi, \theta) \frac{\partial \eta(x)}{\partial \theta},$$

and designs that minimize  $\delta(\xi, \theta) = \max_{x \in \Delta} d(x, \xi, \theta)$  are called G-optimal.

The celebrated general equivalence theorem [originally given in Kiefer and Wolfowitz (1960) for linear models and extended to nonlinear models in White (1973)] establishes the equivalence of  $\xi_D$  and  $\xi_G$ , the D- and G-optimal designs. Equally important is a corollary which states that  $\delta(\xi_D, \theta) = p$ , that is, that the maximum predicted variance for the D-optimal design is  $p$ . An important application of this corollary is to graphically check the D-optimality of  $\xi_D$  because  $\delta(\xi, \theta) > p$  for all other (non D-optimal) designs,  $\xi$ . A further result is that  $d(x, \xi_D, \theta) = p$  for  $x$  equal to each of the support points of  $\xi_D$ . These somewhat obscure concepts are illustrated with the following carefully chosen example, which is useful both to illustrate concepts of optimal design and to provide an alternate proof of the variance maximization exercise.

#### 4. A KEY APPLICATION

For the simple linear regression model  $\eta(x, \theta) = \mathbf{f}^T(x)\theta = \alpha + \beta x$  defined over the interval  $\Delta = [0, 1]$ , it is straightforward to show that the corresponding D-optimal design,  $\xi_D$ , associates the weight  $\omega = 1/2$  with each of the points  $x_1 = 0$  and  $x_2 = 1$  and is such that  $|\mathbf{M}^{-1}(\xi_D)| = 4$ ; related results were given by Atkinson and Donev (1992, p. 38 ff). Indeed, the choice of this design to fit this (linear) model function makes intuitive sense because the parameters can clearly be best (jointly) estimated with design points with the greatest spread. As mentioned in the previous section, D-optimality of this de-

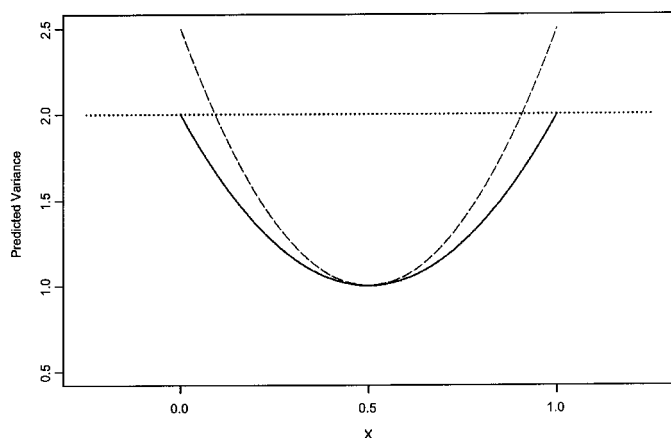


Figure 1. Plot of the variance functions associated with the D-optimal design (solid curve) and a rival design (dashed curve) and the line  $y = 2$ .

sign is easily verified by examining the corresponding variance function. To illustrate, a plot of this predicted variance function,  $d(x, \xi_D, \theta) = 2(1 - 2x + 2x^2) = 4(x - 1/2)^2 + 1$ , given in Figure 1, verifies D-optimality of  $\xi_D$  since the corresponding graph does not exceed the line  $y = p = 2$  over the interval  $\Delta$ . Further, the fact that  $d(x, \xi_D, \theta) = 2$  at the support points ( $x = 0$  and  $1$ ) of  $\xi_D$  also establishes uniqueness of this design. Finally, we can use these results to show that a rival design is *not* D-optimal. Demonstrating that the rival design ( $\xi_R$ ) that associates the weight  $\omega = 1/3$  with each of the support points  $x_1 = 0, x_2 = 1/2$  and  $x_3 = 1$  is indeed not D-optimal for this model function and design space is established by noting both that  $|\mathbf{M}^{-1}(\xi_R)| = 6 > 4$  and by observing that a plot of the corresponding variance function  $d(x, \xi_R, \theta) = 6(x - 1/2)^2 + 1$  crosses above the line  $y = 2$  for several  $x \in \Delta$ ; this too is shown in Figure 1.

Next, consider the same linear model function reparameterized as  $\eta_2(x, \theta_2) = \gamma(x - \delta)$  again defined over the same design space  $\Delta = [0, 1]$ . For an arbitrary design  $\xi$ , the corresponding information matrix is

$$\mathbf{M}(\xi) = \begin{bmatrix} \gamma^2 & -\gamma \sum_{j=1}^k \omega_j (x_j - \delta) \\ -\gamma \sum_{j=1}^k \omega_j (x_j - \delta) & \sum_{j=1}^k \omega_j (x_j - \delta)^2 \end{bmatrix},$$

and the D-optimal design therefore maximizes

$$|\mathbf{M}(\xi)| = \gamma^2 \left[ \sum_{j=1}^k \omega_j (x_j - \delta)^2 - \left\{ \sum_{j=1}^k \omega_j (x_j - \delta) \right\}^2 \right].$$

By the invariance property of D-optimal designs, this determinant is also maximized by the equal-weight design with support points  $x_1 = 0$  and  $x_2 = 1$ . Also, again by the invariance property, because this design does not depend on the choice of the model parameters, we choose  $\gamma = 1$  and  $\delta = \mu$ , so that this same two-point design also maximizes the determinant

$$|\mathbf{M}(\xi)| = \sum_{j=1}^k \omega_j (x_j - \mu)^2.$$

Note that this determinant is identical to the variance criterion function given in Section 2 since  $\sum_{j=1}^k \omega_j (x_j - \mu) = 0$ . It therefore follows that results from optimal design theory provide an alternate proof that this variance is maximized for the equal-weight design with support points  $x_1 = 0$  and  $x_2 = 1$ , and the plot of the corresponding variance function indeed verifies both optimality and uniqueness. More importantly, this exercise also provides a means to introduce students to the basic concepts encountered in optimal design theory which are useful for selecting good designs for *all* linear and nonlinear regression models.

#### 5. AN EXTENSION

Upon being exposed to this example, some students hasten to point out that although the designs considered above have continuous (i.e., possibly nonrational) weights, they have a discrete (countable) set of design support points, and wonder if a larger value of the determinant of the information matrix  $\mathbf{M}(\xi, \theta) = \int (\partial \eta(x) / \partial \theta) (\partial \eta(x) / \partial \theta^T) d\xi(x)$  can be achieved

by considering the space of probability measures (i.e., density functions),  $\Xi = \{\xi(x) \text{ s.t. } \int d\xi(x) = 1 \text{ for } x \in \Delta\}$ . This larger class of design measures is addressed in Kiefer (1974), and for simplicity we consider here only the Beta class of probability distribution functions.

Let the random variable in Section 2 have the probability distribution function

$$\omega = \xi(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

for  $\alpha, \beta > 0$  and  $x \in \Delta = [0, 1]$ . Then by basic calculus, the corresponding variance,

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

is maximized for  $\alpha = \beta$  as  $\alpha, \beta \rightarrow 0$ , and for which the maximal variance is  $1/4$ . It comes as no surprise that as  $\alpha, \beta \rightarrow 0$ , the beta density function approaches one with equal mass at the support points  $x_1 = 0$  and  $x_2 = 1$ . It therefore follows that by the equivalence established in the previous section that this two-point discrete design is also optimal over the space of Beta design measures. Interested students can find additional results on optimal design measures in Pukelsheim (1993).

## 6. DISCUSSION

This article uses the simple exercise of maximizing a variance over a class of probability mass functions to gently expose beginning students to some important concepts of optimal design theory, a topic perhaps encountered only in courses in advanced experimental design. In so doing, students easily appreciate the power and aesthetics associated with this rich theory, and are

thereby equipped to continue their study of optimal design theory and practice.

[Received November 2002. Revised July 2003.]

## REFERENCES

- Atkinson, A. C., and Donev, A. N. (1992), *Optimum Experimental Designs*, Oxford: Clarendon Press.
- Box, G. E. P., and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, New York: Wiley.
- Brady, J. E., and Allen, T. T. (2002), "Case Study Based Instruction of DOE and SPC," *The American Statistician*, 56, 312–315.
- Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics*, 24, 586–602.
- DeGroot, M. H., and Schervish, M. J. (2002), *Probability and Statistics* (3rd ed.), New York: Addison-Wesley.
- Dette, H., and O'Brien, T. E. (1999), "Optimality Criteria for Regression Models Based on Predicted Variance," *Biometrika*, 86, 93–106.
- Kiefer, J. (1959), "Optimum Experimental Designs," *Journal of the Royal Statistical Society, Ser. B*, 21, 272–319.
- (1961), "Optimum Designs in Regression Problems II," *Annals of Mathematical Statistics*, 32, 298–325.
- (1974), "General Equivalence Theory for Optimum Designs (Approximate Theory)," *The Annals of Statistics*, 2, 849–879.
- (1975), "Optimal Design: Variation in Structure and Performance Under Change of Criterion," *Biometrika*, 62, 277–288.
- Kiefer, J., and Wolfowitz, J. (1960), "The Equivalence of Two Extremum Problems," *Canadian Journal of Mathematics*, 12, 363–366.
- Pukelsheim, F. (1993), *Optimal Design of Experiments*, New York: Wiley.
- Seber, G. A. F., and Wild, C. J. (1989), *Nonlinear Regression*, New York: Wiley.
- Silvey, S. D. (1980), *Optimal Design*, London: Chapman and Hall.
- Smith, K. (1918), "On the Standard Deviations of Adjusted and Interpolated Values of an Observed Polynomial Function and its Constants and the Guidance They Give Towards a Proper Choice of the Distribution of Observations," *Biometrika*, 12, 1–85.
- White, L. V. (1973), "An Extension of the General Equivalence Theorem to Nonlinear Models," *Biometrika*, 60, 345–348.