The Canadian Journal of Statistics
Vol. 21, No. 1, 1993, Pages 59–70
La Revue Canadienne de Statistique

Designs for approximately linear regression: Maximizing the minimum coverage probability of confidence ellipsoids

Douglas P. WIENS*

University of Alberta

Key words and phrases: Regression designs, D-optimality, robustness, confidence ellipsoids.
AMS 1985 subject classifications: Primary 62K05; secondary 62F35, 62J05.

ABSTRACT

The classical D-optimality principle in regression design may be motivated by a desire to maximize the coverage probability of a fixed-volume confidence ellipsoid on the regression parameters. When the fitted model is exactly correct, this amounts to minimizing the determinant of the covariance matrix of the estimators. We consider an analogue of this problem, under the approximately linear model $\mathcal{E}[y|\mathbf{x}] = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}) + f(\mathbf{x})$. The nonlinear disturbance $f(\mathbf{x})$ is essentially unknown, and the experimenter fits only to the linear part of the response. The resulting bias affects the coverage probability of the confidence ellipsoid on $\boldsymbol{\theta}$. We study the construction of designs which maximize the minimum coverage probability as f varies over a certain class. Explicit designs are given in the case that the fitted response surface is a plane.

RÉSUMÉ

L'utilisation du principe classique de D-optimalité en régression peut être justifiée par le souci de maximiser la probabilité de recouvrement d'un ellipsoïde de confiance de volume fixe pour les paramètres de régression. Lorsque le modèle ajusté est exactement correct, cela revient à minimiser le déterminant de la matrice des covariances des estimateurs. Nous considérons un problème analogue, sous le modèle approximativement linéaire $\mathcal{E}[y|\mathbf{x}] = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}) + f(\mathbf{x})$. La perturbation non-linéaire $f(\mathbf{x})$ est essentiellement inconnue et la personne faisant l'expérience s'en tient uniquement à la partie linéaire de la réponse. Le biais qui en découle affecte la probabilité de recouvrement de l'ellipsoïde de confiance pour $\boldsymbol{\theta}$. Nous étudions l'élaboration de plans d'expérience qui maximisent la probabilité de recouvrement minimale, lorsque f varie dans une certaine classe. Des plans explicites sont donnés lorsque la surface de réponse ajustée est un plan.

1. INTRODUCTION AND SUMMARY

Consider the following situation. An experimenter fits, by least squares, a regression model with

$$\mathcal{E}[y|\mathbf{x}] = \mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta}_0. \tag{1.1}$$

Here, the regressors $\mathbf{z} \in \mathbb{R}^p$ are given functions of $\mathbf{x}$, with $\mathbf{x}$ varying freely over a design space $S \subset \mathbb{R}^q$. The problem is to choose n (not necessarily distinct) design points

*Research supported by the Natural Sciences and Engineering Research Council of Canada.

$\mathbf{x}_i \in S$. The estimate $\hat{\boldsymbol{\theta}}$, determined from observations $(y_i, \mathbf{x}_i)$, is to be robust, in a certain minimax sense, against departures from (1.1) in the true model

$$\mathcal{E}[y|\mathbf{x}] = \mathbf{z}^{\mathsf{T}}(\mathbf{x})\boldsymbol{\theta}_0 + f(\mathbf{x}), \qquad f \in \mathcal{F}_\eta. \tag{1.2}$$

The class $\mathcal{F}_\eta$ is an $\mathcal{L}_2$-neighbourhood, defined at (1.5) below. We assume additive, uncorrelated errors with common variance σ^2.

This situation is one commonly faced in practice. The experimenter is typically well aware that (1.1) is only a convenient approximation, but is unable to determine, or unwilling to fit, a more appropriate but more complicated model. Under the model (1.2), however, $\hat{\boldsymbol{\theta}}$ is biased.

Box and Draper (1959) made apparent the dangers of designing an experiment which assumes that (1.1) is exactly correct. By analyzing the relative importance of errors due to bias and to variance, they found that very small deviations from (1.1) can eliminate any supposed gains arising from the use of a design which minimizes variance alone. Beginning with Box and Draper (1959), designs for versions of (1.2) have been constructed by various authors. See Kiefer (1973), Huber (1975), Marcus and Sacks (1976), Sacks and Ylvisaker (1978), Li and Notz (1982), Pesotchinsky (1982), Notz (1989), Wiens (1990, 1991).

Wiens (1992) studied analogues of the classical D, A, E, Q, and G optimality problems. The first of these requires a design which minimizes the maximum, over $\mathcal{F}_\eta$ of the determinant of the mean-squared-error matrix of $\hat{\boldsymbol{\theta}}$. Recall that in the classical problem, one assumes that (1.1) is exactly correct, and works instead with the covariance matrix. In either case the motivation is clear — one would like a small upper bound on the loss of precision of $\hat{\boldsymbol{\theta}}$, as measured by the determinant of this matrix. There is, however, an alternative way of motivating the classical D-optimality problem, as in Kiefer (1958) and Fedorov (1972). The analogous problem, under (1.2), is the subject of this paper.

If the design matrix

$$\mathbf{Z} = \|\mathbf{z}(\mathbf{x}_1), \ldots, \mathbf{z}(\mathbf{x}_n)\|^{\mathsf{T}}$$

has full rank p, then the least-squares estimate of $\boldsymbol{\theta}_0$ is

$$\hat{\boldsymbol{\theta}} = (\mathbf{Z}^{\mathsf{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathsf{T}}\mathbf{y},$$

and a confidence ellipsoid on $\boldsymbol{\theta}_0$ is given by

$$C = \left\{ \boldsymbol{\theta} \,\middle|\, (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathsf{T}}\mathbf{Z}^{\mathsf{T}}\mathbf{Z}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \le c^2 \left| \frac{1}{n}\mathbf{Z}^{\mathsf{T}}\mathbf{Z} \right|^{1/p} \right\}. \tag{1.3}$$

The volume of C is c^p times that of a sphere of radius $n^{-\frac{1}{2}}$ in $\mathbb{R}^p$. Note that this volume is independent of the choice of the $\mathbf{x}_i$'s, and of the ability of (1.1) to accurately reflect the true state of affairs. The coverage probability may then be used to judge these factors.

If the random errors are normally distributed, the coverage probability of C is an increasing function of the determinant $|\mathbf{Z}^{\mathsf{T}}\mathbf{Z}|$. It is given by

$$P(\boldsymbol{\theta}_0 \in C) = P\left(\chi_p^2 \le \frac{c^2 |(1/n)\mathbf{Z}^{\mathsf{T}}\mathbf{Z}|^{1/p}}{\sigma^2} \right). \tag{1.4}$$

Under mild conditions, (1.4) holds asymptotically for nonnormal errors.

The classical D-optimality problem is to choose the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ so as to maximize (1.4). Equivalently, $|\mathbf{Z}^{\mathsf{T}}\mathbf{Z}|$ is to be maximized.

Now suppose that the statistician fits the model (1.1), when in fact the true model is (1.2), with

$$\mathcal{F}_\eta = \left\{ f \,\middle|\, \int_S f^2(\mathbf{x})\, d\mathbf{x} \le \eta^2,\ \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} = \mathbf{0} \right\}. \tag{1.5}$$

The orthogonality condition in (1.5) is imposed without loss of generality, and ensures the identifiability of $\boldsymbol{\theta}_0$.

An effect of the disturbance in (1.2) is to alter the coverage probability of $\mathcal{C}$. Under (1.2), the χ^2 variable in (1.4) becomes noncentral, with a noncentrality parameter depending on f and on the design. An appropriate analogue of the D-optimality problem is then to determine a design which maximizes the minimum coverage probability of $\mathcal{C}$ as f varies over $\mathcal{F}_\eta$.

The minimization over $\mathcal{F}_\eta$ is carried out in Section 2 below. There, we also show that the admissible designs are those which are invariant under groups of transformations which leave the problem fixed.

It turns out that the admissible design measures are also absolutely continuous. This poses a practical problem, which can be dealt with either by

(1) relaxing a strict adherence to optimality theory, and taking a discrete approximation to the optimal design, or

(2) employing a *randomized* design, whereby the design points are randomly chosen from the optimal design density. In this case, the *expected* minimum coverage probability would be maximized.

See Remark 3 in Section 3, and Wiens (1992), for further discussion of this point.

In Section 3 of this paper, the theory of Section 2 is applied to the case in which the design space is spherical and

$$\mathbf{z}^{\mathsf{T}}(\mathbf{x}) = (1, \mathbf{x}^{\mathsf{T}}), \qquad \mathbf{z}^{\mathsf{T}}(\mathbf{x})\boldsymbol{\theta}_0 = \theta_0 + \sum_{j=1}^{q} \theta_j x_j, \quad p = q + 1. \tag{1.6}$$

The optimal designs are shown to have spherically symmetric densities. For small values of n they resemble smooth versions of the classical D-optimal designs, with the mass highly concentrated near the boundary of S. As $n \longrightarrow \infty$, they tend to uniformity. Their qualitative behaviour is discussed further in the remarks at the end of Section 3.

2. GENERAL THEORY

We first write the coverage probability of $\mathcal{C}$ in terms of the disturbance f and the design measure ξ. The latter is a probability measure on the design space S. We shall assume that S has been linearly transformed in order to have Lebesgue measure 1.

Define

$$\mathbf{B} = \mathbf{B}(\xi) = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^{\mathsf{T}}(\mathbf{x})\, d\xi(\mathbf{x}) \ (= n^{-1}\mathbf{Z}^{\mathsf{T}}\mathbf{Z}),$$

$$\mathbf{b} = \mathbf{b}(f, \xi) = \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x})\, d\xi(\mathbf{x}),$$

$$\delta^2(f, \xi) = \mathbf{b}^{\mathsf{T}}(f, \xi)\mathbf{B}^{-1}(\xi)\mathbf{b}(f, \xi),$$

$$\nu = \sigma^2/n\eta^2.$$

We have absorbed η into ν, allowing us to assume henceforth that $f \in \mathcal{F}_1$. Denote by $\chi_p^2(\lambda^2)$ a noncentral χ^2 random variable, with p d.f. and noncentrality parameter λ^2. The analogue of (1.4) is

$$\mathcal{P}(f, \xi) = P(\boldsymbol{\theta} \in C \,|\, f, \xi) = P\left(\chi_p^2\left(\frac{\delta^2(f, \xi)}{\nu}\right) \leq \frac{c^2|\mathbf{B}(\xi)|^{1/p}}{\sigma^2}\right). \tag{2.1}$$

In practice, any nonrandomized design measure will be discrete, with jumps consisting of integral multiples of n^{-1}. In view of Lemma 1 below, we now drop this restriction. We search for a maximin design ξ_0, i.e., one for which

$$\inf_{\mathcal{F}_1} \mathcal{P}(f, \xi_0) = \sup \inf_{\mathcal{F}_1} \mathcal{P}(f, \xi), \tag{2.2}$$

where the sup is taken over the class of *all* probability measures on S.

LEMMA 1. *In order that* $\inf_{\mathcal{F}_1} \mathcal{P}(f,\xi)$ *be nonzero it is necessary that* ξ *be absolutely continuous.*

For a proof of Lemma 1, see Wiens (1992). Now let $m(x) = \xi'(\mathbf{x})$ be the density of ξ. Define $p \times p$ matrices

$$\mathbf{A} = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^{\mathrm{T}}(\mathbf{x})\, d\mathbf{x}, \qquad \mathbf{C} = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^{\mathrm{T}}(\mathbf{x})m^2(\mathbf{x})\, d\mathbf{x},$$

$$\mathbf{G} = \mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}.$$

Note that

$$\mathbf{G} = \int_S [\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}^{-1}\}\mathbf{z}(\mathbf{x})][\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}^{-1}\}\mathbf{z}(\mathbf{x})]^{\mathrm{T}}\, d\mathbf{x},$$

so that $\mathbf{G}$ is positive definite. Let $\mathbf{G}^{\frac{1}{2}}$ be a positive definite symmetric root of $\mathbf{G}$. Define

$$\mathbf{r}(\mathbf{x}) = \mathbf{G}^{-\frac{1}{2}}\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}^{-1}\}\mathbf{z}(\mathbf{x}). \tag{2.3}$$

Note that for fixed ξ, (2.1) is minimized over $\mathcal{F}_1$ by that f which maximizes $\delta^2(f,\xi)$.

THEOREM 1. *For a fixed design* ξ, *let* μ_ξ *be the maximum eigenvalue of* $\mathbf{G}^{\frac{1}{2}}\mathbf{B}^{-1}\mathbf{G}^{\frac{1}{2}}$, *and let* $\boldsymbol{\beta}_0$ *be the corresponding eigenvector, with* $\|\boldsymbol{\beta}_0\| = 1$. *Then* $\mathcal{P}(f,\xi)$ *is minimized over* $\mathcal{F}_1$ *by*

$$f_0(\mathbf{x}) = \mathbf{r}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}_0, \tag{2.4}$$

and $\inf_{\mathcal{F}_1} \mathcal{P}(f,\xi)$ *is given by*

$$\mathcal{P}(f_0, \xi) = P\left(\chi_p^2\left(\frac{\mu_\xi}{\nu}\right) \leq \frac{c^2|B(\xi)|^{1/p}}{\sigma^2}\right). \tag{2.5}$$

Proof. We will show that the search for an $f \in \mathcal{F}_1$ which maximizes $\delta^2(f,\xi)$ may be restricted to the subclass $\mathcal{H}$ of $\mathcal{F}_1$ defined by

$$\mathcal{H} = \{h(x; \boldsymbol{\beta}) = \mathbf{r}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta};\ \|\boldsymbol{\beta}\| = 1\}.$$

For any $f \in \mathcal{F}_1$, put $\boldsymbol{\beta}_f = \mathbf{G}^{-\frac{1}{2}}\mathbf{b}(f,\xi)/\|\mathbf{G}^{-\frac{1}{2}}\mathbf{b}(f,\xi)\|$. Then we calculate that

$$\mathbf{b}(h(\cdot;\boldsymbol{\beta}_f),\xi) = \mathbf{b}(f,\xi)/\alpha,$$

where

$$0 \le \alpha = \left| \int_S f(\mathbf{x})h(\mathbf{x};\boldsymbol{\beta}_f)\,d\mathbf{x} \right| \le 1$$

by the Cauchy-Schwarz inequality. Thus

$$\delta^2(h(\cdot;\boldsymbol{\beta}_f),\xi) = \delta^2(\mathbf{b}_f,\xi)/\alpha^2 \ge \delta^2(\mathbf{b}_f,\xi),$$

and it suffices to maximize

$$\delta^2(h(\cdot;\boldsymbol{\beta}),\xi) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{G}^{\frac{1}{2}}\mathbf{B}^{-1}\mathbf{G}^{\frac{1}{2}}\boldsymbol{\beta}$$

over $\|\boldsymbol{\beta}\| = 1$. From this observation, the result is immediate. Q.E.D.

When the design problem is invariant under a group of linear transformations of the design space, the invariant design measures dominate the noninvariant measures. Specifically, let Π be a group, under composition, of volume-preserving linear transformations of S given by

$$\mathbf{x} \longrightarrow \pi(\mathbf{x}) = \mathbf{Q}_\pi\mathbf{x}, \qquad |\mathbf{Q}_\pi| = \pm 1.$$

Suppose that:

(H1) $\pi S = S$ for all $\pi \in \Pi$.

(H2) For each $\pi \in \Pi$ there is a nonsingular matrix $\mathbf{P}_\pi$ such that $\mathbf{z}(\pi(\mathbf{x})) = \mathbf{P}_\pi\mathbf{z}(\mathbf{x})$.

Let Ξ be the class of absolutely continuous design measures on S. For each $\xi \in \Xi$ with density m, each $f \in \mathcal{F}$, and each $\pi \in \Pi$, define ξ_π and f_π by

$$\xi'_\pi(\mathbf{x}) = m(\pi(\mathbf{x})), \qquad f_\pi(\mathbf{x}) = f(\pi(\mathbf{x})).$$

Let Ξ_Π be the class of Π-invariant measures on S, i.e.,

$$\Xi_\Pi = \{\xi \in \Xi \mid \xi_\pi = \xi \text{ for each } \pi \in \pi\}.$$

Similarly, define

$$\mathcal{F}_\Pi = \{f \in \mathcal{F}_1 \mid f_\pi = f \text{ for each } \pi \in \Pi\}.$$

The following technical result will be required for the proof of Theorem 2 below.

LEMMA 2. *If* $\mathbf{V}$, $\mathbf{W}$ *are matrices each of whose elements is a linear function of a real variable* λ, *and if* $\mathbf{W}$ *is positive definite, then*

$$\phi(\lambda) = \mathbf{c}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{V}\mathbf{c}$$

is a convex function of λ *for each* $\mathbf{c}$.

Proof. Denote differentiation with respect to λ by a dot. Put $\mathbf{K} = \mathbf{W}^{-1}\mathbf{V}$. Then

$$\phi'(\lambda) = 2\mathbf{c}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{W}^{-1}\dot{\mathbf{V}}\mathbf{c} - \mathbf{c}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\dot{\mathbf{W}}\mathbf{K}\mathbf{c},$$

$$\phi''(\lambda) = 2\mathbf{c}^{\mathrm{T}}\dot{\mathbf{K}}^{\mathrm{T}}\mathbf{W}\dot{\mathbf{K}}\mathbf{c} \ge 0.$$

Q.E.D.

THEOREM 2.

(i) $\sup_{\Xi} \min_{\mathcal{F}_\Pi} \mathcal{P}(f, \xi) = \sup_{\Xi_\Pi} \min_{\mathcal{F}_\Pi} \mathcal{P}(f, \xi)$.
(ii) *If there exists* ξ_0 *which maximizes* $\min_{\mathcal{F}_\Pi} \mathcal{P}(f, \xi)$ *in* Ξ_Π, *and if* $\mathcal{P}(f, \xi_0)$ *is minimized, over all of* $\mathcal{F}_1$, *by a member of* $\mathcal{F}_\Pi$, *then* ξ_0 *maximizes the minimum coverage probability over all of* $\mathcal{F}_1$ *in the class of all design measures on* S.

Proof. (i): Let $f \in \mathcal{F}_\Pi$, $\pi \in \Pi$ be arbitrary. Then

$$\mathbf{b}(f, \xi_\pi) = \mathbf{P}_\pi^{-1}\mathbf{b}(f, \xi), \qquad \mathbf{B}(\xi_\pi) = \mathbf{P}_\pi^{-1}\mathbf{B}(\xi)\mathbf{P}_\pi,$$

so that

$$\delta^2(f, \xi_\pi) = \delta^2(f, \xi), \qquad |\mathbf{B}(\xi_\pi)| = |\mathbf{B}(\xi)|.$$

Put $\xi_\lambda = (1-\lambda)\xi + \lambda\xi_\pi$, $\lambda \in [0,1]$. By Lemma 2, $\delta^2(f, \xi_\lambda)$ is convex and so

$$\delta^2(f, \xi_\lambda) \le (1-\lambda)\delta^2(f, \xi) + \lambda\delta^2(f, \xi_\pi) = \delta^2(f, \xi).$$

Similarly, since $|\mathbf{B}(\xi_\lambda)|^{-1}$ is convex in λ, we have

$$|\mathbf{B}(\xi_\lambda)|^{1/p} \ge |\mathbf{B}(\xi)|^{1/p}.$$

It follows that

$$\mathcal{P}(f, \xi_\lambda) \ge \mathcal{P}(f, \xi),$$

so that we can improve on ξ unless $\xi = \xi_\lambda$, i.e., $\xi = \xi_\pi$.
(ii): If the stated conditions hold, then

$$\min_{\mathcal{F}_1} \mathcal{P}(f, \xi) \le \min_{\mathcal{F}_\Pi} \mathcal{P}(f, \xi) \le \min_{\mathcal{F}_\Pi} \mathcal{P}(f, \xi_0) = \min_{\mathcal{F}_1} \mathcal{P}(f, \xi_0)$$

for any measure ξ on S. Q.E.D.

3. RESPONSE SURFACE A PLANE

In this section we construct maximin designs, satisfying (2.2), for the response function (1.6). As design space we take a sphere, in $\mathbb{R}^q$, of unit volume:

$$S = \left\{ \mathbf{x} \,\middle|\, \|\mathbf{x}\| \le r := \frac{\{\Gamma(q/2+1)\}^{1/q}}{\sqrt{\pi}} \right\}.$$

Let Π be the group of orthogonal transformations of S. Then (H1), (H2) of Section 2 are satisfied (we take $\mathbf{P}_\pi = 1 \oplus \mathbf{Q}_\pi$), and Ξ_Π is the class of absolutely continuous, spherically symmetric probability measures on S. Any such measure then has a density which depends on $\mathbf{x}$ only through $\|\mathbf{x}\|$:

$$m(\mathbf{x}) = g(\|\mathbf{x}\|),$$

where

$$\int_0^r \frac{qu^{q-1}}{r^q} g(u)\, du = 1. \tag{3.1}$$

The integrand in (3.1) is the density of $\|\mathbf{X}\| := U$. Define $\gamma = \mathcal{E}X_j^2$, i.e.,

$$\gamma = \int_0^r \frac{u^{q+1}}{r^q} g(u)\, du. \tag{3.2}$$

Then

$$\mathbf{B}(\xi) = 1 \oplus \gamma \mathbf{I}_q, \qquad |\mathbf{B}(\xi)| = \gamma^q. \tag{3.3}$$

A special, limiting role is played by the uniform measure $[m(\mathbf{x}) \equiv 1]$, for which

$$\gamma_0 := \mathcal{E}[x_j^2 \mid m(\mathbf{x}) \equiv 1] = \frac{r^2}{q+2}.$$

THEOREM 3. *For the response function* (1.6), *there is a design* ξ_0 *which maximizes the minimum coverage probability in the class of all probability measures on S. It has the density*

$$m_0(\mathbf{x}) = g_0(\|\mathbf{x}\|; \gamma), \tag{3.4}$$

parametrized by $\gamma \in [\gamma_0, r^2/q]$.

Case (a): *For* $1 \le \gamma/\gamma_0 \le (q+2)^2/q(q+4)$,

$$g_0(u; \gamma) = \begin{cases} 1 + \left(\dfrac{\gamma}{\gamma_0} - 1\right)\left(\dfrac{q+4}{4}\right)\left(\dfrac{u^2}{\gamma_0} - q\right), & 0 \le u \le r, \\ 0 & \textit{otherwise}. \end{cases} \tag{3.5}$$

Case (b): *For* $(q+2)^2/q(q+4) \le \gamma/\gamma_0 \le (q+2)/q$,

$$g_0(u; \gamma) = \begin{cases} \dfrac{(u/r)^2 - b}{K_q(b)}, & r\sqrt{b} \le u \le r, \\ 0 & \textit{otherwise} \end{cases} \tag{3.6}$$

where $K_q(b) = (1-b) - \big(2(1-b^{q/2+1})/q+2\big)$, *and b is the unique solution, in* $[0, 1]$, *to*

$$\frac{\gamma}{\gamma_0} = \frac{K_{q+2}(b)}{K_q(b)}.$$

In each case the value of γ *is related to* $\nu = \sigma^2/n\eta^2$ *by*

$$\gamma = \arg\max P\left(\chi_p^2\left(\frac{J_0(g_0;\, \gamma)}{\nu}\right) \le \frac{c^2\gamma^{q/p}}{\sigma^2}\right), \tag{3.7}$$

where

$$J_0(g_0; \gamma) = \begin{cases} \dfrac{q(q+4)}{4}\left(\dfrac{\gamma}{\gamma_0} - 1\right)^2, & \textit{case (a)}, \\ \dfrac{q\gamma - br^2}{r^2 K_q(b)} - 1, & \textit{case (b)}. \end{cases} \tag{3.8}$$

The least favourable member of $\mathcal{F}_1$ *is*

$$f_0(\mathbf{x}) = \frac{g_0(\|\mathbf{x}\|; \gamma) - 1}{\sqrt{J_0(g_0;\, \gamma)}}. \tag{3.9}$$

Proof. Let $\xi \in \Xi_\Pi$ be arbitrary, with density $g(\|\mathbf{x}\|; \gamma)$ indexed by γ. We find that

$$\mathbf{G} = J_0(g; \gamma) \oplus J_1(g; \gamma)\mathbf{I}_q,$$

where

$$J_0(g;\gamma) = \mathcal{E} g(u;\gamma) - 1, \qquad J_1(g;\gamma) = \mathcal{E}\{U^2 g(U;\gamma)\} - \frac{\gamma^2}{\gamma_0}.$$

Put $\boldsymbol{\beta}_0 = (1, \mathbf{0}^{\mathsf{T}})^{\mathsf{T}}$. We claim that $\mathcal{P}(f,\xi)$ is minimized over $\mathcal{F}_\Pi$ by

$$h(\mathbf{x};\boldsymbol{\beta}_0) = \frac{g(\|\mathbf{x}\|;\gamma) - 1}{\sqrt{J_0(g;\gamma)}}. \tag{3.10}$$

For this recall that, as in the proof of Theorem 1,

$$\mathcal{P}(f,\xi) \geq \mathcal{P}(h(\cdot;\boldsymbol{\beta}_f),\xi).$$

But if $f \in \mathcal{F}_\Pi$, then

$$\mathbf{b}(f,\xi) = \int_S f(\mathbf{x})\, d\xi(\mathbf{x}) \cdot \boldsymbol{\beta}_0,$$

and we find that $\boldsymbol{\beta}_f = \boldsymbol{\beta}_0$. Thus $h(\mathbf{x};\boldsymbol{\beta}_0) \in \mathcal{F}_\Pi$ is less favourable than f. Since this h does not depend on f, the claim is established.

Using (3.10) and (3.3) in (2.1) gives

$$\min_{\mathcal{F}_\Pi} \mathcal{P}(f,\xi) = P\left(\chi_p^2\left(\frac{J_0(g;\gamma)}{\nu}\right) \leq \frac{c^2\gamma^{q/p}}{\sigma^2}\right).$$

To maximize this in Ξ_Π we first fix γ, and find $g_0(u;\gamma)$ to minimize $J_0(g;\gamma)$, subject to (3.1) and (3.2). For this, it suffices for $g_0(u;\gamma)$ to minimize

$$\begin{aligned} J_0(g;\gamma) &+ 2abr^2 \int_0^r \frac{u^{q-1}g(u)}{r^q}\,du - 2a\int_0^r \frac{u^{q+1}}{r^q}\,g(u)\,du \\ &= \int_0^r \frac{u^{q-1}}{r^q}\left\{qg^2(u) - \{2a(u^2 - br^2)\}g(u) - q\left(1 + \frac{\gamma^2}{\gamma_0}\right)\right\} du \end{aligned} \tag{3.11}$$

for some Lagrange multipliers a, b, and to satisfy (3.1), (3.2). We minimize (3.11) by minimizing the integrand pointwise, obtaining

$$g_0(u;\gamma) = a(u^2 - br^2)^+/q, \qquad a > 0, \quad b \leq 1, \quad 0 \leq u \leq r.$$

Solving for a and b in terms of γ, and then maximizing over γ, shows that for $f \in \mathcal{F}_\Pi$ the optimal $\xi_0 \in \Xi_\Pi$ is given by (3.5)–(3.8). Cases (a) and (b) correspond to $b \leq 0$ and $b > 0$ respectively.

To complete the proof we must, by Theorem 2(ii), show that $\mathcal{P}(f,\xi_0)$ is minimized, over all of $\mathcal{F}_1$, by $f_0(\mathbf{x}) = h(\mathbf{x};\boldsymbol{\beta}_0)$. By Theorem 1, this holds as long as $\boldsymbol{\beta}_0$ is the eigenvector corresponding to the maximum eigenvalue of

$$\mathbf{G}^{\frac{1}{2}}\mathbf{B}^{-1}\mathbf{G}^{\frac{1}{2}} = J_0(g_0;\gamma) \oplus \gamma^{-1}J_1(g_0;\gamma)\mathbf{I}_q.$$

Equivalently,

$$J_0(g_0;\gamma) \geq \gamma^{-1}J_1(g_0;\gamma). \tag{3.12}$$

The verification of (3.12) is straightforward. We omit the details. Q.E.D.

3.1. Remarks.

REMARK 1. The solution in case (a) of Theorem 3 holds for small values of ν. In the limiting case $\nu \longrightarrow 0$ ($n \longrightarrow \infty$) we have $\gamma = \gamma_0$, $g_0(u;\gamma) \equiv 1$, and the maximin design is uniform on S. This is to be expected, since as $n \longrightarrow \infty$ the bias becomes the sole contributor to the estimation error. In Box and Draper's study (Box and Draper 1959) of regression designs when the fitted model is incorrect, they note that "the optimal design in typical situations in which both variance and bias occur is very nearly the same as would be obtained if variance were ignored completely and the experiment designed so as to minimize bias alone." They go on to construct designs which minimize bias, and find that the bias-minimizing designs must have all moments, up to a certain order, equal to those of the (continuous) uniform design. Case (b) of Theorem 3 holds for large ν; in the limiting case $\nu \longrightarrow \infty$ ($\eta^2 \longrightarrow 0$) we find $\gamma \longrightarrow r^2/q$ and the optimal design places all mass on the boundary of S, as in the classical D-optimality problem.

REMARK 2. The maximization in (3.7) has been carried out numerically, for selected values of p. We have set

$$c^2 = \sigma^2 \left(\frac{r^2}{q}\right)^{-q/p} \chi^2_{p;0.95}, \tag{3.13}$$

so that the coverage probability is 0.95 for the classically optimal design, under the ideal condition $\eta = 0$. See Table 1 for numerical values. We have denoted by $\nu = \nu_0$ the boundary point between cases (a) and (b). If $\nu > \nu_0$, then $b > 0$, and S contains a sphere, around $\mathbf{0}$, within which no observations are made. The last column in the table gives the coverage probability when in fact the fitted model is correct. The difference between this value and 0.95 may be thought of as the premium paid, in lost coverage at the ideal model, in return for the protection afforded by a more robust design, at the given value of η. The user may then choose the particular design corresponding to the maximum premium which he is willing to pay.

Note that the minimum coverage for the optimal design is never catastrophically less than the nominal value 0.95. In contrast, it is a consequence of Lemma 1 that the minimum coverage probability, in $\mathcal{F}_\eta$, for any $\eta > 0$, of the classically "optimal" design is zero.

REMARK 3. If a random vector $\mathbf{x}$ has a spherically symmetric density, then $\mathbf{x}/\|\mathbf{x}\|$ is distributed uniformly over the surface of the unit sphere, independently of $\|\mathbf{x}\|$. Thus, to randomly choose design points from the optimal design density one could first simulate a value u of U, with density $qu^{q-1}g_0(u;\gamma)/r^q$, as at (3.1), and then simulate values of $u \cdot \mathbf{x}/\|\mathbf{x}\|$. Since the distribution of $\mathbf{x}/\|\mathbf{x}\|$ is the same for *any* spherical density, it suffices, in this last simulation, to let the q elements of $\mathbf{x}$ be i.i.d. normals.

3.2. Examples.

We illustrate the construction of designs in two cases. In Example 1 the optimal design measure is approximated by a discrete measure. In Example 2 we use both a discrete approximation and a randomization as described in Remark 3 above.

EXAMPLE 1. Take $q = 1$—straight-line regression with an intercept. Suppose that we are willing to pay a premium of about 5%, in lost coverage at the ideal model, for the added robustness. From Table 1 we may take $\nu = \nu_0 = 3.1174$. Theorem 3 then gives that the distribution function of the optimal design is

$$\xi_0(x) = 4x^3 + 0.5, \qquad -0.5 \le x \le 0.5.$$

TABLE 1: Numerical values for the designs of Section 3.

$\nu\ (=\sigma^2/n\eta^2)$	γ/γ_0	b	$\min_{\mathcal{F}_\eta} \mathcal{P}(f,\xi_0)$	$\mathcal{P}(0,\xi_0)$
		$p=2,\ \gamma_0=\frac{1}{12}$		
0	1	—	0.8226	0.8226
0.01	1.0044	—	0.8231	0.8233
0.1	1.0392	—	0.8258	0.8285
1.0	1.3224	—	0.8456	0.8632
2.0	1.5708	—	0.8604	0.8856
$\nu_0=3.1174$	1.8000	0	0.8724	0.9018
4.0	1.8834	0.0499	0.8797	0.9069
10.0	2.1849	0.2580	0.9036	0.9224
100.0	2.6649	0.6754	0.9351	0.9406
∞	3.0	1	0.95	0.9500
		$p=5,\ \gamma_0=0.0750$		
0	1	—	0.8442	0.8442
0.005	1.0010	—	0.8443	0.8446
0.10	1.0250	—	0.8484	0.8528
0.30	1.0730	—	0.8563	0.8680
$\nu_0=0.5265$	1.1250	0	0.8645	0.8827
1.0	1.1780	0.2183	0.8778	0.8960
10.0	1.3907	0.7722	0.9243	0.9360
100.0	1.4619	0.9229	0.9419	0.9456
∞	1.5	1	0.95	0.95

This can of course be approximated by a discrete design measure in numerous ways. One is to take design points

$$x_i = \xi_0^{-1}\left(\frac{i-0.5}{n}\right) = 0.5\left(\frac{2i-1}{n}-1\right)^{\frac{1}{3}}, \qquad i=1,\ldots,n.$$

For $n = 10$, this gives design points

$$\pm 0.2321,\ \pm 0.3347,\ \pm 0.3969,\ \pm 0.4440,\ \pm 0.4827.$$

EXAMPLE 2. Take $q = 4$ and an approximate 5% premium. From Table 1, we take $\nu = 1$. By Theorem 3, and with

$$r = \frac{2^{\frac{1}{4}}}{\sqrt{\pi}} \simeq 0.6709, \qquad K_4(b) = (1-b) - \frac{1-b^3}{3} \simeq 0.4518,$$

the d.f. of U is

$$G(u) = \begin{cases} 0, & 0 \le u \le r\sqrt{b} \simeq 0.3135, \\ \dfrac{(u^2-br^2)^2(2u^2+br^2)}{3K_4(b)r^6}, & r\sqrt{b} \le u \le r. \end{cases}$$

As design points we take

$$\mathbf{x}_i = \mathbf{G}^{-1}\left(\frac{i-0.5}{n}\right)\frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}, \qquad i=1,\ldots,n,$$

TABLE 2: Randomized design points for Example 2.

i	X_1	X_2	X_3	X_4	U
1	−0.072	0.417	0.025	−0.019	0.425
2	0.052	−0.038	0.323	−0.345	0.477
3	−0.314	0.121	0.378	0.023	0.507
4	−0.317	−0.404	0.117	−0.044	0.529
5	0.414	−0.099	−0.153	−0.306	0.546
6	0.357	0.271	0.293	0.166	0.561
7	−0.335	0.089	−0.398	−0.223	0.573
8	−0.217	0.324	0.112	−0.420	0.585
9	−0.418	0.368	−0.097	0.185	0.595
10	−0.071	0.218	−0.450	0.332	0.604
11	0.437	−0.284	0.248	0.203	0.612
12	0.406	−0.137	0.441	−0.082	0.620
13	0.505	−0.164	−0.105	0.318	0.627
14	0.303	0.008	−0.188	−0.524	0.634
15	0.330	0.102	0.398	−0.365	0.641
16	0.130	−0.584	−0.164	0.183	0.647
17	−0.194	−0.238	0.014	−0.576	0.653
18	−0.411	0.227	0.360	−0.287	0.658
19	−0.233	0.478	−0.287	0.274	0.663
20	0.277	−0.159	0.263	0.525	0.669

where the $\mathbf{v}_i$ are randomly generated vectors, each consisting of four independent normal r.v.s.

Table 2 gives design points obtained in this manner for $n = 20$. Some selected projections of this design are presented in Figure 1.

REFERENCES

Box, G.E.P., and Draper, N. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.*, 54, 622–654.

Fedorov, V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Huber, J. (1975). Robustness and designs. *A Survey of Statistical Design and Linear Models* (J.N. Srivastava, *ed.*), North Holland, Amsterdam, 287–303.

Kiefer, J. (1958). On the nonrandomized optimality and randomized non-optimality of symmetrical designs. *Ann. Math. Statist.*, 29, 675–699.

Kiefer, J. (1973). Optimum designs for fitting biased multiresponse surfaces. *Multivariate Analysis III*, (P.R. Krishnaiah, *ed.*) Academic Press, New York, 287–297.

Li, K.C., and Notz, W. (1982). Robust designs for nearly linear regression. *J. Statist. Plann. Inference*, 6, 135–151.

Marcus, M.B., and Sacks, J. (1976). Robust designs for regression problems. *Statistical Theory and Related Topics II* (S.S. Gupta and D.S. Moore, *eds.*), Academic Press, New York, 245–268.

Notz, W. (1989). Optimal designs for regression models with possible bias. *J. Statist. Plann. Inference*, 22, 43–54.

Pesotchinsky, L. (1982). Optimal robust designs: Linear regression in R^k. *Ann. Statist.*, 10, 511–525.

Sacks, J., and Ylvisaker, D. (1978). Linear estimation for approximately linear models. *Ann. Statist.*, 6, 1122–1137.

Wiens, D.P. (1990). Robust minimax designs for multiple linear regression. *Linear Algebra Appl.*, 127, 327–340.

Wiens, D.P. (1991). Designs for approximately linear regression: Two optimality properties of uniform designs. *Statist. Probab. Lett.*, 12, 217–221.

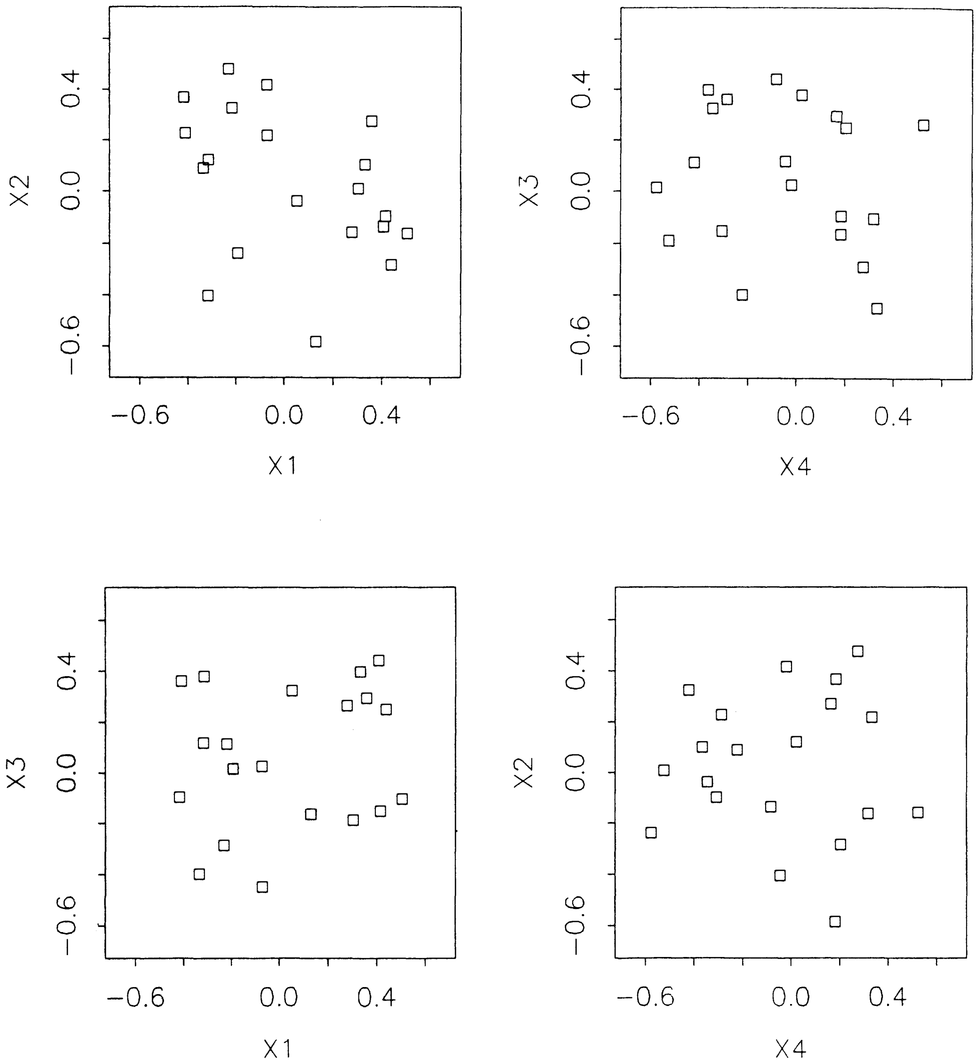

FIGURE 1: Projections of the randomized design of Example 2.

Wiens, D.P. (1992). Minimax designs for approximately linear regression. *J. Statist. Plann. Inference*, 31, 353–371.

Received 17 December 1990
Revised 30 July 1992
Accepted 2 September 1992

Department of Statistics and Applied Probability
University of Alberta
Edmonton, Alberta
T6C 2G1