The Annals of Statistics
1991, Vol. 19, No. 3, 1244–1256

ROBUST BAYESIAN EXPERIMENTAL DESIGNS IN NORMAL LINEAR MODELS

BY A. DASGUPTA[1] AND W. J. STUDDEN[2]

Purdue University

We address the problem of finding a design that minimizes the Bayes risk with respect to a fixed prior subject to being robust with respect to misspecification of the prior. Uncertainty in the prior is formulated in terms of having a family of priors instead of one single prior. Two different classes of priors are considered: Γ_1 is a family of conjugate priors, and a second family of priors Γ_2 is induced by a metric on the space of nonnegative measures. The family Γ_1 has earlier been suggested by Leamer and Polasek, while Γ_2 was considered by DeRobertis and Hartigan and Berger. The setup assumed is that of a canonical normal linear model with independent homoscedastic errors. Optimal robust designs are considered for the problem of estimating the vector of regression coefficients or a linear combination of the regression coefficients and also for testing and set estimation problems. Concrete examples are given for polynomial regression and completely randomized designs. A very surprising finding is that for Γ_2, *the same design* is optimal for a variety of different problems with different loss structures. In general, the results for Γ_2 are significantly more substantive. Our results are applicable to group decision making and reconciliation of opinions among experts with different priors.

1. Introduction. A major problem in the general domain of statistics is the derivation of an experimental design optimal with respect to some criterion consistent with the goal of the study. Typically, the optimality criteria considered by workers in this general area have focused on long-run (frequentist) performance of a design, such as the mean squared error over repeated sampling: the well-known criteria of A, D and E optimality are examples of this kind. It is not unusual though for the experimenter to have nonnegligible prior information about the parameters in the system, information that is sufficiently significant to be of some use but not quite so sharp and precise as to be quantified in terms of a single "prior distribution." The purpose of this article is to address the question of which design should the statistician recommend in the scenario of a collection of plausible, Bayesian prior distributions. This article thus focuses on some experimental design problems from a "robust Bayesian" viewpoint. The subject of robust Bayes methods has, by itself, been a major research area in the recent past; for general exposition and specific results, we refer the reader to Berger (1984), Berger and Berliner (1986), DasGupta and Studden (1989) and Wasserman (1989).

Received August 1988; revised July 1990.
[1]Research supported by NSF Grant DMS-87-026-20.
[2]Research supported by NSF Grant DMS-88-025-35.
AMS 1980 *subject classifications*. 62F15, 62K05, 62F35
Key words and phrases. Robust Bayesian design.

There now exists a vast body of statistical literature on optimal experimental designs (with primarily long-run performance criteria); the pioneering work is due to Jack Kiefer. See Silvey (1980) for many references.

The study of experimental designs in a Bayesian framework has been comparatively limited; some of the important references include Pilz (1979, 1981), Verdinelli (1982), Bandemer (1977), Chaloner (1984) and Ball, Smith and Verdinelli (1989). In this article, optimal experimental designs are derived for the problems of estimation, prediction or testing a null hypothesis in the canonical normal linear model setup when the prior for the parameters belongs to a family of distributions Γ.

Consider the usual linear regression problem where $\mathbf{Y}_{n\times 1} \sim N(X\boldsymbol{\theta}, \sigma^2 I)$, where $X_{n\times p}$ is the design matrix of nonstochastic constants; for ease in exposition, assume $\sigma^2 > 0$ to be known; σ^2 comes out as a proportionality factor in all risk expressions relevant to this paper and consequently will be *ignored* in all risk formulas. The design aspects of the problem enter through the experimenter's choice of the rows of the design matrix X from an available set $\mathscr{X}$. The vector of regression coefficients $\boldsymbol{\theta}_{p\times 1}$ is assumed to have a prior distribution $\pi(\boldsymbol{\theta})$ belonging to a suitable class Γ.

Two different classes of priors will be considered; the first of them is

$$(1.1)\qquad \Gamma_1 = \{\pi(\boldsymbol{\theta}) : \boldsymbol{\theta} \sim N(\boldsymbol{\mu}, \sigma^2\Sigma),\ \boldsymbol{\mu}\text{ fixed},\ lI \le \Sigma^{-1} \le kI\};$$

here $0 \le l < k$ and by $A \ge B$ we mean that $A - B$ is n.n.d. or nonnegative definite. The idea here is that conjugate priors are mathematically attractive and also often provide a rich enough class of priors for a comprehensive Bayesian analysis of the data; the mean of the prior is kept fixed but not the variance–covariance structure because the location of the unknown parameters is usually much easier to elicit subjectively than it is to elicit the higher moments and the strengths of the correlations. Also, as we shall later see, the design problems are reasonably tractable with a family of priors such as (1.1). The family of priors (1.1) was first suggested and used by Leamer (1978, 1982) and Polasek (1985). For an extensive discussion, see DasGupta and Studden (1988).

Normal priors, by definition, are symmetric and unimodal. Moreover, in (1.1) the mean $\boldsymbol{\mu}$ was kept fixed [although we could vary the prior mean as well; see DasGupta and Studden (1988)]. An alternative family of priors that also enjoys mathematical tractability, and yet at the same time allows the mean as well as the variance–covariance to change and in addition includes asymmetric and multimodal priors is the family of priors

$$(1.2)\qquad \Gamma_2 = \{\pi(\boldsymbol{\theta}) : L(\boldsymbol{\theta}) \le \alpha\pi(\boldsymbol{\theta}) \le U(\boldsymbol{\theta}) \text{ for some } \alpha > 0\};$$

$L(\boldsymbol{\theta})$ will be taken as the density of $N(\boldsymbol{\mu}, \sigma^2\Sigma_0)$, $\boldsymbol{\mu}, \sigma^2$ fixed and $U(\boldsymbol{\theta}) = kL(\boldsymbol{\theta})$ for a suitable $k > 1$. The first works with this family of priors are DeRobertis (1978) and DeRobertis and Hartigan (1981). They define Γ_2 slightly more generally using L and U as arbitrary measures. The class Γ_2 is a metric neighborhood of the prior L. A discussion of the metric is given in DeRobertis

(1978). This is further discussed in DasGupta and Studden (1988). The prior $L(\boldsymbol{\theta}) = N(\boldsymbol{\mu}, \sigma^2\Sigma_0)$ will be seen to play a special role in robustness questions.

For ease of exposition we will consider what is now commonly called the approximate design theory. All the design aspects will enter through the "information matrix" $M = X'X$ which can be written as $M = X'X = n\sum p_i x_i x_i'$, where x_i' are the rows of X and $np_i = n_i$ are integers. The approximate theory allows the $p_i \geq 0$ to be arbitrary, subject to $\sum p_i = 1$, and in fact permits $M = n\int x'x\, d\mu(x)$ where μ is an arbitrary probability measure.

The general aim of the paper is illustrated using Γ_1. Interest centers around the Bayes risk, under ordinary squared error loss, given by

$$r(\Sigma, M) = \operatorname{tr}(M + \Sigma^{-1})^{-1}. \tag{1.3}$$

Let $\Phi(M)$ denote some measure of robustness [see (1.5) and (1.6)] of the design M. Useful $\Phi(M)$ will of course be related to $r(\Sigma, M)$. The design M will be chosen to minimize $\Phi(M)$. A restricted optimization problem is also considered. If Σ_0 corresponds to a special or favored prior, let

$$\Phi_0(M) = \operatorname{tr}(M_0 + \Sigma^{-1})^{-1}. \tag{1.4}$$

Then $\Phi(M)$ is minimized subject to the condition

$$\Phi_0(M) \leq (1 + \varepsilon)\Phi_0(M_0), \tag{1.5}$$

where M_0 minimizes $\Phi_0(M)$, and ε is a fixed (usually small) positive number.

For the class Γ_1, the functional Φ is chosen as either Φ_1, Φ_2 or Φ_3 defined below. Letting $r(\Sigma) = \inf_M r(\Sigma, M)$, define

$$\Phi_1(M) = \sup_{\Sigma \in \Gamma_1} \frac{r(\Sigma, M)}{r(\Sigma)}. \tag{1.6}$$

Here $\Sigma \in \Gamma_1$ means the normal prior in Γ_1 indexed by Σ. Minimizing Φ_1 corresponds to choosing the design to minimize the maximum inefficiency. The results we have for Φ_1 also apply to a similar "regret" formulation using $\Phi(M) = \sup_{\Sigma \in \Gamma_1}[r(\Sigma, M) - r(\Sigma)]$.

The functional $\Phi_2(M)$ is defined as

$$\Phi_2(M) = \operatorname{tr}(M + lI)^{-1} - \operatorname{tr}(M + kI)^{-1}. \tag{1.7}$$

For priors in Γ_1, $\operatorname{tr}(M + \Sigma^{-1})^{-1}$ lies between $\operatorname{tr}(M + kI)^{-1}$ and $\operatorname{tr}(M + lI)^{-1}$; so $\Phi_2(M)$ denotes the range of the Bayes risks.

The functional $\Phi_3(M)$ is given by

$$\Phi_3(M) = \lambda_{\max}\left\{M^{-1} - (M + kI)^{-1}\right\} \tag{1.8}$$

and is related to the diameter of the set of Bayes estimates when $l = 0$. Motivation for this is given in Section 2. A more general definition would be $\lambda_{\max}\{(M + lI)^{-1} - (M + kI)^{-1}\}$ but only (1.8) will be discussed in the sequel since the conclusion of Theorem 2.1 may fail for this more general definition.

It is shown in Section 2 that Φ_1, Φ_2 and Φ_3 are nondecreasing and convex in M, the partial order on M being in the sense of positive definiteness. Some simple invariance properties are discussed and some examples are given.

For an arbitrary prior $\pi \in \Gamma_2$ the Bayes risk, denoted by $r(\pi, M)$, will not have an expression as in (1.3) except for the normal priors such as $L(\boldsymbol{\theta}) = N(\boldsymbol{\mu}, \sigma^2\Sigma_0)$. The results in Section 3 for Γ_2 have two important aspects. The first is that we can handle some natural functionals $\Phi_i(M)$ defined through the posterior distribution given π and $\mathbf{y}$. As indicated later, these functionals are actually independent of $\mathbf{y}$. The second is that several Φ_i are minimized by the same design M that is best for the above normal prior $L(\boldsymbol{\theta})$. To illustrate this, let S_0, for a given M and $\mathbf{y}$, be the set of smallest Lebesgue measure among all S satisfying

$$\inf_{\pi \in \Gamma_2} P_\pi(\boldsymbol{\theta} \in S \mid \mathbf{y}) \geq 1 - \alpha.$$

The Lebesgue measure of S_0 is independent of $\mathbf{y}$ and the minimizing M is the one that minimizes the determinant $|M + \Sigma_0^{-1}|^{-1}$. This material is based on results in DasGupta and Studden (1988).

For estimating a fixed vector $\mathbf{c}'\boldsymbol{\theta}$ several appropriate $\Phi_i(M)$ are defined and it is shown that the same M minimizes all of them and that this design corresponds to minimizing $\mathbf{c}'(M + \Sigma_0^{-1})^{-1}\mathbf{c}$. This is discussed in Section 4.

2. Normal priors with a fixed mean. In this section we consider the class Γ_1 defined in (1.1) and the functionals Φ_1, Φ_2 and Φ_3.

THEOREM 2.1. *The functionals Φ_1, Φ_2 and Φ_3 are decreasing and convex on the cone of nonnegative definite matrices.*

PROOF. It is well known that the risk $r(\Sigma, M)$ is decreasing and convex in M for a given Σ. Both of these properties are preserved under taking a supremum so the statement holds for Φ_1. For Φ_2, we let $M_\alpha = (1-\alpha)M_1 + \alpha M_2 = M_1 + \alpha(M_2 - M_1)$, $0 < \alpha < 1$, and follow the usual argument showing that $g(\alpha) = \Phi_2(M_\alpha)$ satisfies $g'(\alpha) \leq 0$ if $M_2 - M_1 \geq 0$ and that $g''(\alpha) \geq 0$. To this end let $\Delta = M_2 - M_1$, $A = (M_\alpha + lI)^{-1}$, $B = (M_\alpha + kI)^{-1}$, $C = \Delta A \Delta$, $D = \Delta B \Delta$ to obtain

$$g'(\alpha) = -\operatorname{tr} A\Delta A + \operatorname{tr} B\Delta B$$

and

$$\begin{aligned} g''(\alpha) &= 2[\operatorname{tr} A\Delta A\Delta A - \operatorname{tr} B\Delta B\Delta B] \\ &= 2[\operatorname{tr} ACA - \operatorname{tr} BDB] \\ &= 2[\operatorname{tr} C(A+B)(A-B) + \operatorname{tr}(BCA - ACB) + \operatorname{tr}(C-D)BB]. \end{aligned}$$

Since $C \geq D$ and $AB = BA$ the last two terms in the above expression for $g''(\alpha)$ are nonnegative. The first term is also nonnegative since $A + B$ and $A - B$ also commute. This shows that $g''(\alpha) \geq 0$. The fact that $g'(\alpha) \leq 0$

follows from the same argument taking $C = D = \Delta \geq 0$. The functional $\Phi_3(M) = k[\lambda_s(k + \lambda_s)]^{-1}$, where λ_s is the smallest eigenvalue of M. The result then follows by standard arguments. This proves the theorem. □

The motivation behind Φ_3 comes from the fact [see DasGupta and Studden (1988)] that the Euclidean diameter D of the set S of all Bayes estimates of $\boldsymbol{\theta}$ under Γ_1 is given by

$$D^2 = \boldsymbol{\nu}'(\Lambda_2 - \Lambda_1)\boldsymbol{\nu} \cdot \Phi_3,$$

where $\boldsymbol{\nu} = X'(\mathbf{y} - X\boldsymbol{\mu})$, $\Lambda_2 = M^{-1}$ and $\Lambda_1 = (M + kI)^{-1}$. A general expression for D with bounds Σ_1 and Σ_2 on Σ is given in the above reference. The problem in working with the D or D^2 is that D depends on $\mathbf{y}$ and consequently an expected value has to be taken in order to address a design problem. If we assume that $l = 0$ so that the normal prior $N(\boldsymbol{\mu}, \Sigma)$ has $\Sigma^{-1} \leq kI$, a direct computation then shows that

$$ED^2 \propto \Phi_3(M) = \lambda_{\max}\{M^{-1} - (M + kI)^{-1}\},$$

where the expectation is taken when $\Sigma^{-1} = kI$. The functional $\Phi_3(M)$ would appear to give less robust designs since the expectation is taken with respect to the most precise Σ in Γ_2. It does not seem likely that $\sup_{\Sigma \in \Gamma_2} E_\Sigma D^2$ is attained for $\Sigma^{-1} = kI$ as one would desire in this case.

Before giving some applications we indicate that Γ_1 defined with bounds kI and lI for Σ^{-1} is amenable to some simple invariance arguments. Suppose that the set of possible information matrices M is closed under a group of matrices A acting on M by AMA'. If, in addition, the group is a subgroup of the orthogonal group so that $A' = A^{-1}$ or $AA' = I$ then simple arguments show that Φ_i satisfy $\Phi_i(M) = \Phi_i(AMA')$. If Φ_i is convex and the group is finite or compact, any minimizing M can then be replaced by an invariant one.

EXAMPLE 1. Consider the simple linear regression model $EY = \theta_0 + \theta_1 x$, where $-1 \leq x \leq 1$. It is well known that any M can be increased in the sense of positive definiteness by using a two-point design that samples only at ± 1. Since Φ_1, Φ_2 and Φ_3 (used when $l = 0$) are decreasing, we restrict ourselves to such designs. The matrices M under consideration are thus of the form $M = n\begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}$, where $|c| \leq 1$. The invariance considerations mentioned above imply that all criteria are minimized by the two-point design with weight $1/2$ at $x = \pm 1$ or $c = 0$.

We now turn to the restricted optimization criterion described in (1.5). Let $\Sigma_0^{-1} = \begin{bmatrix} r_1 & r_0 \\ r_0 & r_2 \end{bmatrix}$. It is easy to check that the design M_0 minimizing $\Phi_0(M) = \operatorname{tr}(M + \Sigma_0^{-1})^{-1}$ is given by $c = c_0 = -r_0/n$ provided $|r_0| \leq n$. If $lI \leq \Sigma_0^{-1} \leq kI$ this is the case if $n \geq (k - l)/2$. Also the smallest Bayes risk under Σ_0 equals

$$\Phi_0(M) = \frac{2n + r_1 + r_2}{(n + r_1)(n + r_2)}.$$

Then $\Phi_0(M) \le (1 + \varepsilon)\Phi_0(M_0)$ if and only if

$$(2.1)\qquad (nc + r_0)^2 \le \frac{\varepsilon}{1 + \varepsilon}(n + r_1)(n + r_2) = d \quad \text{(say)}.$$

Since Φ_i, $i = 1, 2, 3$, are convex in c and symmetric about 0 each of the restricted minimizations occurs at the root of the equation with equality in (2.1) which is closet to 0, if of course $c = 0$ is not already between the two roots. The required value of c in all three cases is c^* given by

$$(2.2)\qquad c^* = \begin{cases} \dfrac{1}{n}\left(-r_0 + \sqrt{d}\right) & \text{if } r_0 > 0, \\ \dfrac{1}{n}\left(-r_0 - \sqrt{d}\right) & \text{if } r_0 < 0, \end{cases}$$

where d is defined in (2.1). If $r_0 = 0$ then $c = 0$ is between the two roots. One can check that $c = 0$ is between the two roots if

$$\varepsilon \ge \varepsilon_0 = \frac{r_0^2}{(n + r_1)(n + r_2) - r_0^2}.$$

Thus if $\varepsilon \ge \varepsilon_0$ the solution is $c^* = 0$; otherwise c^* is given by (2.2).

EXAMPLE 2. Consider a completely randomized design with p treatments and suppose the treatment means $\theta_1, \theta_2, \dots, \theta_p$ have a prior $\pi \in \Gamma_1$. The information matrix M is now a diagonal matrix with elements n_i = number of measurements on the ith treatment. For the case $p = 2$, a complete solution is easily given as in Example 1 for general Σ_0^{-1}. We omit the details. Symmetry considerations using the permutation group show that all three of Φ_1, Φ_2 and Φ_3 (for $l = 0$) are globally minimized by $n_1 = n/p$.

The restricted minimization for Φ_1 appears very difficult so we consider only Φ_2 and Φ_3. For arbitrary p, assume that $\Sigma_0^{-1} = R_0$ is diagonal with diagonal elements r_i and assume without loss of generality that $0 \le r_1 \le r_2 \le \cdots \le r_p$. It is known (and simple Lagrangian arguments will show) that $\Phi_0(M)$ is minimized when $(n_i^0 + r_i)^{-2} = C_0$ if $i \le i_0$ and $n_i^0 = 0$ if $i > i_0$ for suitable C_0 and i_0. Furthermore if $n \ge pr_p - \Sigma r_i$, then $n_i^0 + r_i = (n + \Sigma r_i)/p$ for all i. Note that $n_1^0 \ge \cdots \ge n_p^0$. Intuitively, one makes the posterior precisions $n_i + r_i$ as equal as possible (starting with the smallest r_i).

The minimum of Φ_2 subject to (1.5) amounts to moving the n_i^0 in the "direction" of n/p. Lagrangian arguments show that n_i^*, $i = 1, \dots, p$, form the required solution if equality holds in (1.5) and for some $u > 0$,

$$(2.3)\qquad (n_i^* + l)^{-2} - (n_i^* + k)^{-2} + u(n_i^* + r_i)^{-2} = C^*,$$

where

$$nC^* = \operatorname{tr}(M^* + lI)^{-2}M^* - \operatorname{tr}(M^* + kI)^{-2}M^* + u\,\operatorname{tr}(M^* + \Sigma_0^{-1})^{-2}M^*. \tag{2.4}$$

The condition on C^* in (2.4) will force $\sum n_i^* = n$. In solving these equations we actually solve (2.3), (1.5) and $\sum n_i^* = n$ for u, C^* and $n_1^*, n_2^*, \dots, n_p^*$.

Let

$$\eta(\varepsilon) = \frac{\Phi_2^0 - \Phi_2^*}{\Phi_2^0},$$

where Φ_2^0 is the value of Φ_2 at the minimum for Φ_0 and Φ_2^* is the value at the constrained minimum. Thus $100\eta(\varepsilon)$ is the percentage gain in robustness for a sacrifice of $100\varepsilon\%$ in the subjective Bayes risk. We remark, and it is not very hard to show, that for ε near 0, the value of $\eta(\varepsilon)$ is approximately

$$\eta(\varepsilon) \approx \frac{(n + \sum r_i)s_a}{\Phi_2^0}\sqrt{\varepsilon},$$

where $s_a^2 = p^{-1}\sum_{i=1}^p (a_i - \bar{a})^2$, $\bar{a} = \sum a_i/p$ and $a_i = -(n_i^0 + l)^{-2} + (n_i^0 + k)^{-2}$. Thus the percentage gain is considerable for small ε. As an example, for $n = 25$, $p = 2$, $r_1 = l = 1$, $r_2 = k = 9$, $\varepsilon = 0.02$ corresponds to $\eta(\varepsilon) = 0.14$, which represents a 14% gain in robustness for a 2% sacrifice in risk. At this point n_1 and n_2 have moved from $n_1^0 = 16.5$ and $n_2^0 = 8.5$ (where $n_1^0 + r_1 = n_2^0 + r_2 = 17.5$) to $n_1^* = 14$ and $n_2^* = 11$. For fixed n, the constant multiplying $\sqrt{\varepsilon}$ appears to be increasing in p. This provides further confirmation that there is generally more gain in robustness for fixed ε, for larger values of p, i.e., more parameters in the model. For $n = 15$ the constants are approximately 1.3, 1.7 and 2.6 for $p = 2$, 3 and 5 respectively.

The analysis for $\Phi_3(M) = \lambda_{\max}(M^{-1}(M + kI)^{-1})$ is very similar. For $p = 2$ the restricted problem can again be handled for general Σ_0^{-1}. For general p assume as above that $\Sigma_0^{-1} = R_0$ is diagonal with $0 \le r_1 \le r_2 \le \cdots \le r_p$ so that $n_1^0 \ge n_2^0 \ge \cdots \ge n_p^0$ and $\Phi_3(M_0) = [n_p^0(k + n_p^0)]^{-1}$. For illustrative purposes assume $r_{p-1} < r_p$ and n is large enough so that $0 < n_p^0 < n_{p-1}^0$. For ε sufficiently small in (1.5), Lagrangian arguments show that the constrained solution n_i^* satisfies $n_p^* + r_p = \lambda_0''$ and $n_i^* + r_i = \lambda_0'$, $i = 1, \dots, p-1$, where λ_0' and λ_0'' are determined by $\sum n_i^* = n$ and equality in (1.5). The general solution is to set $n_i^* + r_i = \lambda_0''$ for $i \ge i_0$ and λ_0' for $i < i_0$ for some i_0 depending on ε. The details are omitted.

3. Priors inside a density band. In this section, we consider construction of optimum designs when the family of priors π is given by (1.2); for example, π is assumed to be proportional to a function lying between L and $U = kL$ for some $k > 1$, where L is the density of a $N(\boldsymbol{\mu}; \sigma^2\Sigma_0)$ distribution. In contrast to the family of priors (1.1), the mean and variance–covariance structure all change simultaneously as the prior varies in the family (1.2). To give the reader a flavor of how different the prior means can be, we consider the model $\mathbf{Y} \sim N(X\boldsymbol{\theta}, \sigma^2 I)$ when $\sigma^2 = 1$ and L is $N(\boldsymbol{\mu}, I)$. The prior mean of each θ_i is in the range $\mu_i \pm \gamma(k)$, where $\gamma(k)$ for various values of k is given below:

(3.1)

k	2	3	4	5	6	8	10
$\gamma(k)$	0.276	0.436	0.549	0.636	0.707	0.817	0.901.

Thus, for example, if θ_1 has mean 0 and variance 1 under L, then the prior mean varies between ± 0.549 for $k = 4$. The nice feature of our results in this section is that the design which is Bayes with respect to L will be seen to have a number of robustness properties as well.

For a general prior $\pi \in \Gamma_2$ the Bayes risk $r(\pi, M)$ or the posterior risk does not have a closed-form analytical expression and the corresponding functionals are unmanageable. The class Γ_2, however, has some nice robustness properties.

As mentioned in Section 1, the family of priors Γ_2 is a metric neighborhood of the prior $L = N(\boldsymbol{\mu}, \sigma^2 \Sigma_0)$. Consequently, L is a natural choice for the specific prior with respect to which one would like to be nearly Bayes; up to a proportionality constant the Bayes risk under the prior L is $\Phi_0(M) = \operatorname{tr}(M + \Sigma_0^{-1})^{-1}$.

It is shown in DasGupta and Studden (1988) that, in the present setup, the Euclidean diameter of the set of Bayes estimates of $\boldsymbol{\theta}$ (for squared error loss) equals $D_L = 2\gamma(k)\sqrt{\Phi_4(M)}$, where

$$\Phi_4(M) = \lambda_{\max}\left(M + \Sigma_0^{-1}\right)^{-1} \tag{3.2}$$

and $\gamma(k)$ is a fixed constant [see DasGupta and Studden (1988); some values for $\gamma(k)$ were given in (3.1)].

The very attractive feature of this result is that D_L is independent of $\mathbf{y}$ and therefore unlike in Section 2, we do not need to take an expected value of D_L (or its square). The idea here is that if at the design stage we somehow knew what the $\mathbf{y}$ data would be, then a Bayesian design should be geared toward optimum performance for this fixed data. A value of D_L independent of $\mathbf{y}$ enables us to do this.

A reasonable restricted optimization problem would then be to minimize $\Phi_4(M)$ subject to (1.5). Since both of these functionals are decreasing and convex, we have a relatively neat scenario in this case.

Example 3. Consider the completely randomized design with p treatments considered in Example 2. Again let $\Sigma_0^{-1} = \operatorname{diag}(r_1, \dots, r_p)$. The problem here is to minimize $\lambda_{\max}(M + \Sigma_0^{-1})^{-1}$ subject to $\Phi_0(M) = \operatorname{tr}(M + \Sigma_0^{-1})^{-1}$ being near its minimum. In this example the minimum of both functionals is attained for the same set of $n_0^1, n_2^0, \dots, n_p^0$. These values are such that $n_i^0 + r_i = \lambda_0$ and are described in Example 2. Thus the Bayes risk under the prior L for estimating the vector of treatment means and the squared diameter of the set of Bayes estimates are minimized simultaneously (i.e., at the same design).

In classical design theory considerable importance is placed on the determinant of M^{-1}. The corresponding Bayes quantity is

$$\Phi_5(M) = |M + \Sigma_0^{-1}|^{-1}. \tag{3.3}$$

This is related to other Bayesian quantities as proved in DasGupta and Studden (1988). For a fixed design M let S be the set of smallest Lebesgue measure such that

$$\inf_{\pi \in \Gamma_2} P(\boldsymbol{\theta} \in S | \mathbf{y}) \geq 1 - \alpha; \tag{3.4}$$

here $0 < \alpha < 1$ is fixed. The set S exists and is simply a Bayes confidence set for the prior $L = N(\boldsymbol{\mu}, \sigma^2 \Sigma_0^{-1})$ for a suitable confidence coefficient $\gamma < \alpha$, that is, $P_L(\boldsymbol{\theta} \in S | \mathbf{y}) = 1 - \gamma > 1 - \alpha$. Since the posterior distribution of $\boldsymbol{\theta}$ under the prior L is $N((M + \Sigma_0^{-1})^{-1} X'(\mathbf{y} - X\boldsymbol{\mu}), (M + \Sigma_0^{-1})^{-1})$, it follows that S is the p-dimensional ellipsoid

$$S = \left\{\boldsymbol{\theta} \colon (\boldsymbol{\theta} - \boldsymbol{\nu})' \Lambda^{-1} (\boldsymbol{\theta} - \boldsymbol{\nu}) \leq \chi^2_{1-\gamma}(p)\right\}, \tag{3.5}$$

where $\Lambda = (M + \Sigma_0^{-1})^{-1}$, $\boldsymbol{\nu} = \Lambda X'(\mathbf{y} - X\boldsymbol{\mu})$ and $\chi^2_{1-\gamma}(p)$ is the $100(1 - \gamma)$th percentile of the χ^2 distribution with p degrees of freedom. Since the Lebesgue measure of S is proportional to $\Phi_5(M)$, the following theorem is proven.

THEOREM 3.1. *The design minimizing the Lebesgue measure of the set S satisfying* (3.4) *is the Bayes D-optimal design with respect to L, that is, the design minimizing $\Phi_5(M)$.*

EXAMPLE 4. Consider the quadratic regression model $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2$ and suppose $-1 \leq x \leq 1$; also let L be the $N(\boldsymbol{\mu}, \sigma^2 \Sigma_0)$ prior where $\boldsymbol{\mu}$ is arbitrary but fixed and $\Sigma_0^{-1} = \operatorname{diag}(r_1, r_2, r_3)$. Then standard monotonicity and convexity arguments and calculus give that the Bayes D-optimal design is of the form

$$M = n \begin{pmatrix} 1 & 0 & c \\ 0 & c & 0 \\ c & 0 & c \end{pmatrix}, \tag{3.6}$$

$$\text{where } c = c_5 = \frac{\dfrac{r_2 - r_1}{n} + 1 + \sqrt{\left(\dfrac{r_2 - r_1}{n} + 1\right)^2 + 3\left(1 + \dfrac{r_1}{n}\right)\left(\dfrac{r_2 + r_3}{n}\right)}}{3}.$$

This amounts to sampling at 0 and ± 1, where the proportion of observations at each of ± 1 is $c/2$. For example, if the prior variances of $\theta_0, \theta_1, \theta_2$ under L are 3, 5 and 1, and if $n = 9$, then c is approximately 0.72. Notice that the optimal design converges to the classical D-optimal design as $n \to \infty$. The corresponding value of c, say, c_0, that minimizes $\Phi_0(M)$ is the root of a cubic equation. For any specific prior Σ_0 this c_0 can be calculated. In considering the restricted minimization of $\Phi_5(M)$ subject to (1.5) the two values c_0 and c_5 can be compared. From the convexity of $\Phi_0(M)$ equality would occur in (1.5) for two values of c. If $c_0 < c_5$ [in (3.6)] then the restricted minimization of $\Phi_5(M)$ would occur for the larger of the two values giving equality in (1.5). This is assuming that ε is sufficiently small. A similar comment holds for $\Phi_4(M)$.

4. This section is slightly different and is concerned with the estimation of a fixed arbitrary linear combination $\mathbf{c}'\boldsymbol{\theta}$, still for the case Γ_2. This enables us to work out optimal designs for estimation of specific regression coefficients, the mean response at fixed levels of the regressor variables or for the extrapolation problem. The prior of main concern is again taken to be $L = N(\boldsymbol{\mu}, \sigma^2\Sigma_0)$. For this prior the Bayes risk for squared error loss is proportional to

$$\Phi_6(M) = \mathbf{c}'\left(M + \Sigma_0^{-1}\right)^{-1}\mathbf{c}. \tag{4.1}$$

Some designs for specific vectors $\mathbf{c}$ and Σ_0^{-1} are given in Chaloner (1984), Pilz (1979) and El-Krunz and Studden (1991). The minimization of (4.1) has interesting robustness implications. We give below four other natural criteria which are equivalent to minimizing $\Phi_6(M)$. The first two criteria concern estimation and rely on results from DasGupta and Studden (1988). The next two concern testing.

C_1: For any vector $\mathbf{c}$ and design M, let $\boldsymbol{\mu}_c$ and σ_c denote the posterior mean and the posterior standard deviation of $\mathbf{c}'\boldsymbol{\theta}$ under a fixed prior $\pi \in \Gamma_2$ and let

$$S_c = \{(\boldsymbol{\mu}_c, \sigma_c) : \pi \in \Gamma_2\}. \tag{4.2}$$

The criterion C_1 is to minimize the range of $\boldsymbol{\mu}_c$ or σ_c. These ranges are actually independent of y [see DasGupta and Studden (1988)].

C_2: For any vector $\mathbf{c}$ and design M, let I_c be the set of smallest Lebesgue measure such that

$$\inf_{\pi \in \Gamma_2} P_\pi(\mathbf{c}'\boldsymbol{\theta} \in I | \mathbf{y}) \geq 1 - \alpha. \tag{4.3}$$

The criterion C_2 is to find the design minimizing the Lebesgue measure of I_c.

C_3 and C_4: Suppose we want to test the hypothesis that for a fixed vector $\mathbf{c}$, $\mathbf{c}'\boldsymbol{\theta}$ is smaller than or equal to its prior expected value under the basic prior L, that is, H_0: $\mathbf{c}'\boldsymbol{\theta} \leq \mathbf{c}'\boldsymbol{\mu}$. Consider this as a decision problem with a zero–one loss $L(H_i, a_j) = \delta_{ij}$, $i, j = 0, 1$, where a_j denotes the action "accept H_j" and δ_{ij} denotes the usual Kronecker delta. The criterion C_3 is to find the design that minimizes the posterior Bayes risk (of the Bayes test) with respect to L. Finally, the criterion C_4 is to find the design minimizing the range of the posterior probabilities of H_0, that is,

$$\sup_{\pi \in \Gamma_2} P_\pi(H_0 | \mathbf{y}) - \inf_{\pi \in \Gamma_2} P_\pi(H_0 | \mathbf{y}). \tag{4.4}$$

THEOREM 4.1. *The criteria C_1, C_2, C_3 and C_4 are all equivalent to minimizing $\Phi_6(M) = \mathbf{c}'(M + \Sigma_0^{-1})^{-1}\mathbf{c}$.*

PROOF. The results for C_1 and C_2 rely on results from DasGupta and Studden (1988). It is shown there that the set S_c is actually given by

$$S_c = \sqrt{\mathbf{c}'(M + \Sigma_0^{-1})^{-1}\mathbf{c}} \cdot S_h + \left(\mathbf{c}'(M + \Sigma_0^{-1})^{-1}\boldsymbol{\nu}, \mathbf{0}\right), \tag{4.5}$$

where $\boldsymbol{\nu} = X'(\mathbf{y} - X\boldsymbol{\mu})$, and

$$S_h = \{(\mu_X, \sigma_X) \colon X \sim f, \phi \le \alpha f \le k\phi \text{ for some } \alpha > 0\} \tag{4.6}$$

where ϕ denotes the standard normal density and μ_X, σ_X denote the mean and standard deviation of X. The important point here is that the set S does not depend on either $\mathbf{y}$ or M. The result for C_1 is then immediate.

The proof for C_2 is similar to that of Theorem 3.1 and is omitted.

The proof for C_3 follows. Assume without loss of generality that $\boldsymbol{\mu} = \mathbf{0}$. Since the loss is zero–one, for a fixed design M, the Bayes test under L takes action a_0 if and only if $P_L(H_0|\mathbf{y}) \ge P_L(H_1|\mathbf{y})$. Consequently, the posterior risk equals $g(p) = \min(p, (1 - p))$, where $p = P_L(H_0|\mathbf{y})$ (note p will depend on the design M). Notice $g(p)$ is symmetric about $p = \frac{1}{2}$ and also unimodal with mode at $\frac{1}{2}$. Now, since the posterior distribution of $\mathbf{c}'\boldsymbol{\theta}$ under L is $N(\mathbf{c}'\boldsymbol{\nu}, \mathbf{c}'\Lambda\mathbf{c})$, where $\boldsymbol{\nu}$ and Λ are as in the proof of Theorem 3.1 (with $\boldsymbol{\mu} = \mathbf{0}$), it follows that $p = \Phi(-\mathbf{c}'\boldsymbol{\nu}/\sqrt{\mathbf{c}'\Lambda\mathbf{c}})$. Here Φ denotes the standard normal c.d.f. Let $M_L(\mathbf{y})$ denote the marginal of $\mathbf{y}$ under the prior L and $M_L(t)$ denote the marginal of $t = \mathbf{c}'\boldsymbol{\nu}/\sqrt{\mathbf{c}'\Lambda\mathbf{c}}$ under L. Then the Bayes risk of the Bayes test under L is

$$E_{M_L(y)}(g(p)) = E_{M_L(t)}(g(\Phi(-t))). \tag{4.7}$$

Trivially, $M_L(t)$ is the $N(0, \tau^2)$ distribution, where

$$\tau^2 = \frac{\mathbf{c}'\Sigma_0 M(M + \Sigma_0^{-1})^{-1}\mathbf{c}}{\mathbf{c}'(M + \Sigma_0^{-1})^{-1}\mathbf{c}} = \frac{\mathbf{c}'\Sigma_0\mathbf{c}}{\mathbf{c}'(M + \Sigma_0^{-1})^{-1}\mathbf{c}} - 1. \tag{4.8}$$

We now need the fact that if $Z \sim N(0, \tau^2)$ and $g(\Phi(Z))$ is any symmetric unimodal function of Z with mode at 0, then $E[g(\Phi(Z))]$ is decreasing in τ. This follows since $E[g(\Phi(Z))] = 2\int_0^\infty g(\Phi(z))n(z; 0, \tau^2)\,dz$, $g(\Phi(z))$ is decreasing and $2n(z; 0, \tau^2)I_{\{z>0\}}$ has a monotone likelihood ratio. Combining this with (4.8), it follows that (4.7) is increasing in $\mathbf{c}'(M + \Sigma_0^{-1})^{-1}\mathbf{c}$. This completes the result for C_3.

To derive the optimum design for the criterion C_4, let $A = \{\boldsymbol{\theta}\colon \mathbf{c}'\boldsymbol{\theta} \le 0\}$. Then

$$\begin{aligned} \sup_{\pi \in \Gamma_2} P_\pi(H_0|\mathbf{y}) &= \sup_{\pi \in \Gamma_2} \int_A d\pi(\boldsymbol{\theta}|\mathbf{y}) \\ &= \sup_\eta \frac{\int_A d\eta(\boldsymbol{\theta}|\mathbf{y})}{\int_{R^p} d\eta(\boldsymbol{\theta}|\mathbf{y})}, \end{aligned} \tag{4.9}$$

where $\pi(\boldsymbol{\theta}|\mathbf{y})$ denotes the posterior of $\boldsymbol{\theta}$ given $\mathbf{y}$ resulting from $\pi \in \Gamma_2$. Here $\eta(\boldsymbol{\theta}|\mathbf{y}) = \alpha\pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})$, where $\alpha > 0$ is such that $L \le \alpha\pi \le kL$, and thus denotes the "unnormalized" posterior. Clearly, $L(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) \leq \alpha\pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) \leq kL(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta})$. It is easy to see that the ratio in (4.9) is maximized by choosing $\alpha\pi(\boldsymbol{\theta}) = kL(\boldsymbol{\theta})$ if $\boldsymbol{\theta} \in A$ and $L(\boldsymbol{\theta})$ if $\boldsymbol{\theta} \notin A$ implying $\sup_{\pi \in \Gamma_2} P_\pi(H_0 | \mathbf{y}) = kp/[kp + (1-p)]$, where $p = \int_A d\tilde{L}(\boldsymbol{\theta} | \mathbf{y})$.

Similarly, $\inf_{\pi \in \Gamma_2} P_\pi(H_0 | \mathbf{y}) = p/[p + k(1-p)]$ and consequently,

$$\sup_\pi P_\pi(H_0 | \mathbf{y}) - \inf_\pi P_\pi(H_0 | \mathbf{y}) = \frac{(k^2-1)p(1-p)}{(1+(k-1)p)(k-(k-1)p)}. \tag{4.10}$$

The right side of (4.10) is easily seen to be symmetric about $p = 1/2$ and unimodal with mode at $p = 1/2$. Thus the expected range of the posterior probability of H_0 (under the marginal distribution of $\mathbf{y}$ induced by the prior L) is increasing in $\mathbf{c}'(M + \Sigma_0^{-1})^{-1}\mathbf{c}$. This follows by a repetition of the argument used to prove that (4.7) is decreasing in $\mathbf{c}'(M + \Sigma_0^{-1})^{-1}\mathbf{c}$. The theorem is now proved. □

5. Concluding remarks, other models and generalizations. In the present article we have taken a novel approach to designing an experiment when we want to use the available prior information but also want to guard as much as possible against possible misspecification of prior information. Our results allow consideration of several meaningful criteria and especially encouraging are the findings in Section 4 that the user can use the same optimum design for a variety of design situations and that this design corresponds to the basic prior L.

Much more has to be done. Other ways to model prior information have to be considered. The case of an unknown error variance was not considered in this article to keep the setup simple. However, most results of this paper are also valid when the error variance σ^2 is unknown and an appropriate inverse gamma prior is used for σ^2. The practically useful cases of heteroscedastic and/or correlated errors will be considered elsewhere.

Acknowledgments. The authors would like to thank an Associate Editor and a referee for suggestions leading to an improved presentation. In addition, Jim Berger contributed several helpful discussions.

REFERENCES

BALL, F. G., SMITH, A. F. M. and VERDINELLI, I. (1989). Biased coin designs with a Bayesian bias. Technical Report, Dept. Mathematics, Univ. Nottingham.

BANDEMER, H. (1977). *Theorie und Anwendung der Optimalen Versuchsplanung* **1** and **2**. Akademie-Verlag, Berlin.

BERGER, J. (1984). The robust Bayesian viewpoint. In *Robustness in Bayesian Statistics* (J. Kadane, ed.) 63–124. North-Holland, Amsterdam.

BERGER, J. and BERLINER, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14** 461–486.

CHALONER, K. (1984). Optimal Bayesian experimental designs for linear models. *Ann. Statist.* **12** 283–300.

DASGUPTA, A. and STUDDEN, W. J. (1988). Robust Bayesian analysis and optimal experimental designs in normal linear models with many parameters. Technical Report 88-14, Dept. Statistics, Purdue Univ.

DASGUPTA, A. and STUDDEN, W. J. (1989). Frequentist behavior of robust Bayes estimates of normal means. *Statist. Decisions* **7** 333–362.

DEROBERTIS, L. (1978). The use of partial prior knowledge in Bayesian inference. Ph.D. dissertation, Yale Univ.

DEROBERTIS, L. and HARTIGAN, J. A. (1981). Bayesian inference using intervals of measures. *Ann. Statist.* **9** 235–244.

EL-KRUNZ, S. M. and STUDDEN, W. J. (1991). Bayesian optimal designs in linear regression models. *Ann. Statist.* **19**. To appear.

LEAMER, E. E. (1978). *Specification Searches: Ad hoc Inference with Nonexperimental Data.* Wiley, New York.

LEAMER, E. E. (1982). Sets of posterior means with bounded variance prior. *Econometrica* **50** 725–736.

PILZ, J. (1979). Konstruktion von optimalen diskreten Versuchsplänen für eine Bayes-Schätzung im linearen Regressionsmodell. *Freiberger Forschungshefte* **117** 123–152.

PILZ, J. (1981). Robust Bayes and minimax-Bayes estimation and design in linear regression. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 163–177.

POLASEK, W. (1985). Sensitivity analysis for general and hierarchical linear regression models. In *Bayesian Inference and Decision Techniques with Applications* (P. K. Goel and A. Zellner, eds.) 375–387. North-Holland, Amsterdam.

SILVEY, S. D. (1980). *Optimal Design.* Chapman and Hall, London.

VERDINELLI, I. (1982). Computing Bayes D- and A-optimal block designs for a two-way model. Unpublished manuscript.

WASSERMAN. L. (1989). A robust Bayesian interpretation of likelihood regions. *Ann. Statist.* **17** 1387–1393.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907