ROBUST BAYESIAN EXPERIMENTAL DESIGNS
IN NORMAL LINEAR MODELS

BY A. DASGUPTA¹ AND W. J. STUDDEN²

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: Γ₁ is a family of conjugate priors, and a second family of priors Γ₂ is induced by a metric on the space of nonnegative measures. The family Γ₁ has earlier been suggested by Leamer and Polasek, while Γ₂ 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 Γ₂, the same design is optimal for a variety of different problems with different loss structures. In general, the results for Γ₂ 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).

¹Research supported by NSF Grant DMS-87-026-20.
²Research 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 \( \Gamma \).

Consider the usual linear regression problem where \( Y_{n \times 1} \sim N(X\theta, \sigma^2 I) \), where \( X_{n \times p} \) is the design matrix of nonstochastic constants; for ease in exposition, assume \( \sigma^2 > 0 \) to be known; \( \sigma^2 \) 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 \( \mathcal{X} \). The vector of regression coefficients \( \theta_{p \times 1} \) is assumed to have a prior distribution \( \pi(\theta) \) belonging to a suitable class \( \Gamma \).

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

\[
\Gamma_1 = \{ \pi(\theta) : \theta \sim N(\mu, \sigma^2 \Sigma), \mu \text{ fixed}, lI \leq \Sigma^{-1} \leq kI \};
\]

here \( 0 \leq l < k \) and by \( A \succeq B \) 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 \( \mu \) 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

\[
\Gamma_2 = \{ \pi(\theta) : L(\theta) \leq \alpha \pi(\theta) \leq U(\theta) \text{ for some } \alpha > 0 \};
\]

\( L(\theta) \) will be taken as the density of \( N(\mu, \sigma^2 \Sigma_0), \mu, \sigma^2 \) fixed and \( U(\theta) = kL(\theta) \) for a suitable \( k > 1 \). The first works with this family of priors are DeRobertis (1978) and DeRobertis and Hartigan (1981). They define \( \Gamma_2 \) slightly more generally using \( L \) and \( U \) as arbitrary measures. The class \( \Gamma_2 \) 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 $L(\theta) = N(\mu, \sigma^2 \Sigma_0)$ 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 $\Sigma p_i = 1$, and in fact permits $M = n\int x'x d\mu(x)$ where $\mu$ is an arbitrary probability measure.

The general aim of the paper is illustrated using $\Gamma_1$. Interest centers around the Bayes risk, under ordinary squared error loss, given by

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

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 $\Sigma_0$ corresponds to a special or favored prior, let

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

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

$$\Phi_0(M) \geq (1 + \varepsilon)\Phi_0(M_0),$$  

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

For the class $\Gamma_{1}$, the functional $\Phi$ is chosen as either $\Phi_1$, $\Phi_2$ or $\Phi_3$ defined below. Letting $r(\Sigma) = \inf_M r(\Sigma, M)$, define

$$\Phi_1(M) = \sup_{\Sigma \in \Gamma_{1}} \frac{r(\Sigma, M)}{r(\Sigma)}.$$  

Here $\Sigma \in \Gamma_{1}$ means the normal prior in $\Gamma_{1}$ indexed by $\Sigma$. Minimizing $\Phi_1$ corresponds to choosing the design to minimize the maximum inefficiency. The results we have for $\Phi_1$ 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

$$\Phi_2(M) = \text{tr}(M + I)^{-1} - \text{tr}(M + kI)^{-1}.$$  

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

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

$$\Phi_3(M) = \lambda_{\text{max}} \{M^{-1} - (M + kI)^{-1}\}$$  

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 $\lambda_{\text{max}} \{(M + I)^{-1} - (M + kI)^{-1}\}$ 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 \( \Phi_1, \Phi_2 \) and \( \Phi_3 \) 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 \( L(\theta) = N(\mu, \sigma^2 \Sigma_0) \). The results in Section 3 for \( \Gamma_2 \) have two important aspects. The first is that we can handle some natural functionals \( \Phi_i(M) \) defined through the posterior distribution given \( \pi \) and \( y \). As indicated later, these functionals are actually independent of \( y \). The second is that several \( \Phi_i \) are minimized by the same design \( M \) that is best for the above normal prior \( L(\theta) \). To illustrate this, let \( S_0 \), for a given \( M \) and \( y \), be the set of smallest Lebesgue measure among all \( S \) satisfying

\[
\inf_{\pi \in \Gamma_2} P_\pi(\theta \in S | y) \geq 1 - \alpha.
\]

The Lebesgue measure of \( S_0 \) is independent of \( y \) 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 \( c^T \theta \) 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 \( c^T (M + \Sigma_0^{-1})^{-1} c \). This is discussed in Section 4.

2. Normal priors with a fixed mean. In this section we consider the class \( \Gamma_1 \) defined in (1.1) and the functionals \( \Phi_1, \Phi_2 \) and \( \Phi_3 \).

**Theorem 2.1.** The functionals \( \Phi_1, \Phi_2 \) and \( \Phi_3 \) 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 \( \Sigma \). Both of these properties are preserved under taking a supremum so the statement holds for \( \Phi_1 \). For \( \Phi_2 \), 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 \( \Delta = M_2 - M_1, A = (M_\alpha + I)^{-1}, B = (M_\alpha + kI)^{-1}, C = \Delta A \Delta, D = \Delta B \Delta \) to obtain

\[
g'(\alpha) = -\text{tr} A \Delta A + \text{tr} B \Delta B
\]

and

\[
g''(\alpha) = 2[\text{tr} A \Delta A \Delta A - \text{tr} B \Delta B \Delta B]
= 2[\text{tr} A \Delta A - \text{tr} B \Delta B]
= 2[\text{tr} C(A + B)(A - B) + \text{tr}(BCA - ACB) + \text{tr}(C - D)BB].
\]

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 $\lambda_s$ is the smallest eigenvalue of $M$. The result then follows by standard arguments. This proves the theorem. □

The motivation behind $\Phi_3$ comes from the fact [see DasGupta and Studden (1988)] that the Euclidean diameter $D$ of the set $S$ of all Bayes estimates of $\theta$ under $\Gamma_1$ is given by

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

where $\nu = X'(y - X\mu)$, $\Lambda_2 = M^{-1}$ and $\Lambda_1 = (M + kI)^{-1}$. A general expression for $D$ with bounds $\Sigma_1$ and $\Sigma_2$ on $\Sigma$ is given in the above reference. The problem in working with the $D$ or $D^2$ is that $D$ depends on $y$ 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 $N(\mu, \Sigma)$ has $\Sigma^{-1} \preceq kI$, a direct computation then shows that

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

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 $\Sigma$ in $\Gamma_2$. 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 $\Gamma_1$ defined with bounds $kI$ and $lI$ for $\Sigma^{-1}$ 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 $\Phi_i$ satisfy $\Phi_i(M) = \Phi_i(AMA')$. If $\Phi_i$ 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 $EY = \theta_0 + \theta_1x$, 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 $\pm 1$. Since $\Phi_1$, $\Phi_2$ and $\Phi_3$ (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 $\Sigma_0^{-1} = \begin{pmatrix} r_0 & r_0 \\ r_0 & r_0 \end{pmatrix}$. It is easy to check that the design $M_0$ minimizing $\Phi_0(M) = \text{tr}(M + \Sigma_0^{-1})^{-1}$ is given by $c = c_0 = -r_0/n$ provided $|r_0| \leq n$. If $lI \leq \Sigma_0^{-1} \preceq kI$ this is the case if $n \geq (k - l)/2$. Also the smallest Bayes risk
under $\Sigma_0$ equals
\[
\Phi_0(M) = \frac{2n + r_1 + r_2}{(n + r_1)(n + r_2)}.
\]

Then $\Phi_0(M) \leq (1 + \varepsilon)\Phi_0(M_0)$ if and only if
\[
(nc + r_0)^2 \leq \frac{\varepsilon}{1 + \varepsilon} (n + r_1)(n + r_2) = d \quad \text{(say)}.
\]

Since $\Phi_i$, $i = 1, 2, 3$, 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 closest 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
\[
c^* = \begin{cases} 
\frac{1}{n} (-r_0 + \sqrt{d}) & \text{if } r_0 > 0, \\
\frac{1}{n} (-r_0 - \sqrt{d}) & \text{if } r_0 < 0,
\end{cases}
\]

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
\[
\varepsilon \geq \varepsilon_0 = \frac{r_0^2}{(n + r_1)(n + r_2) - r_0^2}.
\]

Thus if $\varepsilon \geq \varepsilon_0$ 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 $\theta_1, \theta_2, \ldots, \theta_p$ 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 $i$th treatment. For the case $p = 2$, a complete solution is easily given as in Example 1 for general $\Sigma_0^{-1}$. We omit the details. Symmetry considerations using the permutation group show that all three of $\Phi_1$, $\Phi_2$ and $\Phi_3$ (for $l = 0$) are globally minimized by $n_1 = n/p$.

The restricted minimization for $\Phi_1$ appears very difficult so we consider only $\Phi_2$ and $\Phi_3$. For arbitrary $p$, assume that $\Sigma_0^{-1} = R_0$ is diagonal with diagonal elements $r_i$ and assume without loss of generality that $0 \leq r_1 \leq r_2 \leq \cdots \leq r_p$. 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 $i \leq i_0$ and $n_i^0 = 0$ if $i > i_0$ for suitable $C_0$ and $i_0$. Furthermore if $n \geq pr_p - \Sigma r_i$, then $n_i^0 + r_i = (n + \Sigma r_i)/p$ for all $i$. Note that $n_i^0 \geq \cdots \geq n_p^0$. Intuitively, one makes the posterior precisions $n_i + r_i$ as equal as possible (starting with the smallest $r_i$).

The minimum of $\Phi_2$ subject to (1.5) amounts to moving the $n_i^0$ in the "direction" of $n/p$. Lagrangian arguments show that $n_i^* = n_i + r_i$, $i = 1, \ldots, p$, form the required solution if equality holds in (1.5) and for some $u > 0$,
\[
(n_i^* + l)^{-2} - (n_i^* + k)^{-2} + u(n_i^* + r_i)^{-2} = C^*.
\]
where

\[ nC^* = \text{tr}(M^* + lI)^{-2}M^* - \text{tr}(M^* + kI)^{-2}M^* + u \text{ tr}(M^* + \Sigma^{-1}_0)^{-2}M^*. \]

(2.4)

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

Let

\[ \eta(\varepsilon) = \frac{\Phi_0^0 - \Phi_2^2}{\Phi_2^0}, \]

where \( \Phi_0^0 \) is the value of \( \Phi_0 \) at the minimum for \( \Phi_0 \) and \( \Phi_2^2 \) 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 \( \varepsilon \) near 0, the value of \( \eta(\varepsilon) \) is approximately

\[ \eta(\varepsilon) \approx \frac{(n + \sum r_i)\sqrt{s_0^2}}{\Phi_2^0}, \]

where \( s_0^2 = p^{-1}\sum_{i=1}^p (a_i - \bar{a})^2, \) \( \bar{a} = \sum a_i/p \) and \( a_i = -(n_i^0 + l)^{-2} + (n_i^0 + k)^{-2} \). Thus the percentage gain is considerable for small \( \varepsilon \). 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 \( \varepsilon \), 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 \( \Phi_3(M) = \lambda_{\text{max}}(M^{-1}(M + kI)^{-1}) \) is very similar. For \( p = 2 \) the restricted problem can again be handled for general \( \Sigma_0^{-1} \). For general \( p \) assume as above that \( \Sigma_0^{-1} = R_0 \) is diagonal with \( 0 \leq r_1 \leq r_2 \leq \cdots \leq r_p \) so that \( n_1^0 \geq n_2^0 \geq \cdots \geq n_p^0 \) 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 \( 0 < n_p^0 < n_p^{0-1} \). For \( \varepsilon \) sufficiently small in (1.5), Lagrangian arguments show that the constrained solution \( n_i^* \) satisfies \( n_i^* = \lambda_i^* + \lambda_i^* = \lambda_i^* + \lambda_i^* = \lambda_i^*, i = 1, \ldots, p - 1 \), where \( \lambda_i^* \) and \( \lambda_i^* \) are determined by \( \Sigma n_i^* = n \) and equality in (1.5). The general solution is to set \( n_i^* = r_i = \lambda_i^* \) for \( i \geq i_0 \) and \( \lambda_i^* \) for \( i < i_0 \) for some \( i_0 \) depending on \( \varepsilon \). The details are omitted.

3. Priors inside a density band. In this section, we consider construction of optimum designs when the family of priors \( \pi \) is given by (1.2); for example, \( \pi \) 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 \( N(\mu; \sigma^2 \Sigma_0) \) 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 \( Y \sim N(X\theta, \sigma^2 I) \) when \( \sigma^2 = 1 \) and \( L \) is \( N(\mu, I) \). The prior mean of
each \( \theta_i \) is in the range \( \mu_i \pm \gamma(k) \), where \( \gamma(k) \) for various values of \( k \) is given
below:

\[
\begin{array}{cccccccccc}
(3.1) & k & 2 & 3 & 4 & 5 & 6 & 8 & 10 \\
\gamma(k) & 0.276 & 0.436 & 0.549 & 0.636 & 0.707 & 0.817 & 0.901 & \\
\end{array}
\]

Thus, for example, if \( \theta_1 \) has mean 0 and variance 1 under \( L \), then the prior
mean varies between \( \pm 0.549 \) 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 \( \Gamma_2 \), however, has some nice robustness proper-
ties.

As mentioned in Section 1, the family of priors \( \Gamma_2 \) is a metric neighborhood
of the prior \( L = N(\mu, \sigma^2 \Sigma_0) \). 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 \( \Phi_0(M) =
\text{tr}(M + \Sigma_0^{-1})^{-1} \).

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

\[
(3.2) \quad \Phi_4(M) = \lambda_{\max}(M + \Sigma_0^{-1})^{-1}
\]

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 \( y \) and
therefore unlike in Section 2, we do not need to take an expected value of \( D_L \n
(or its square). The idea here is that if at the design stage we somehow knew
what the \( y \) data would be, then a Bayesian design should be geared toward
optimum performance for this fixed data. A value of \( D_L \) independent of \( y \)
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 \( \Sigma_0^{-1} = \text{diag}(r_1, \ldots, r_p) \). The problem here
is to minimize \( \lambda_{\max}(M + \Sigma_0^{-1})^{-1} \) subject to \( \Phi_0(M) = \text{tr}(M + \Sigma_0^{-1})^{-1} \) being
near its minimum. In this example the minimum of both functionals is
attained for the same set of \( n_0^1, n_0^2, \ldots, n_0^p \). These values are such that
\( n_0^1 + r_i = \lambda_0 \) and are described in Example 2. Thus the Bayes risk under the
prior \( L \) for estimating the vector of treatment means and the squared diame-
ter 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

$$
(3.3) \quad \Phi_\beta(M) = |M + \Sigma_0^{-1}|^{-1}.
$$

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

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

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

$$
(3.5) \quad S = \{\theta: (\theta - v)'\Lambda^{-1}(\theta - v) \leq \chi^2_{1-\gamma}(p)\},
$$

where $\Lambda = (M + \Sigma_0^{-1})^{-1}$, $v = \Lambda X'(y - X\mu)$ and $\chi^2_{1-\gamma}(p)$ is the 100(1 - $\gamma$)th percentile of the $\chi^2$ distribution with $p$ degrees of freedom. Since the Lebesgue measure of $S$ is proportional to $\Phi_\beta(M)$, 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 $\Phi_\beta(M)$.

**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 $N(\mu, \sigma^2 \Sigma_0)$ prior where $\mu$ is arbitrary but fixed and $\Sigma_0^{-1} = \text{diag}(r_1, r_2, r_3)$. Then standard monotonicity and convexity arguments and calculus give that the Bayes D-optimal design is of the form

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

$$
(3.6) \quad \frac{r_2 - r_1}{n} + 1 + \sqrt{\left(\frac{r_2 - r_1}{n} + 1\right)^2 + 3\left(1 + \frac{r_1}{n}\right)\left(\frac{r_2 + r_3}{n}\right)}
$$

where $c = c_0 = \frac{r_2 - r_1}{n} + 1 + \sqrt{\left(\frac{r_2 - r_1}{n} + 1\right)^2 + 3\left(1 + \frac{r_1}{n}\right)\left(\frac{r_2 + r_3}{n}\right)}$.

This amounts to sampling at 0 and $\pm 1$, where the proportion of observations at each of $\pm 1$ 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 $\Sigma_0$ this $c_0$ can be calculated. In considering the restricted minimization of $\Phi_\beta(M)$ subject to (1.5) the two values $c_0$ and $c_0$ can be compared. From the convexity of $\Phi_\beta(M)$ equality would occur in (1.5) for
two values of \( c \). If \( c_0 < c_5 \) in (3.6) then the restricted minimization of \( \Phi_\theta(M) \) would occur for the larger of the two values giving equality in (1.5). This is assuming that \( \varepsilon \) is sufficiently small. A similar comment holds for \( \Phi_\alpha(M) \).

4. This section is slightly different and is concerned with the estimation of a fixed arbitrary linear combination \( c'\theta \), still for the case \( \Gamma_2 \). 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 \( L = N(\mu, \sigma^2 \Sigma_0) \). For this prior the Bayes risk for squared error loss is proportional to

\[
\Phi_\theta(M) = c'\left(M + \Sigma_0^{-1}\right)^{-1}c.
\]

Some designs for specific vectors \( c \) and \( \Sigma_0^{-1} \) 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 \( \Phi_\theta(M) \). 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 \( c \) and design \( M \), let \( \mu_c \) and \( \sigma_c \) denote the posterior mean and the posterior standard deviation of \( c'\theta \) under a fixed prior \( \pi \in \Gamma_2 \) and let

\[
S_c = \{ (\mu_c, \sigma_c) : \pi \in \Gamma_2 \}.
\]

The criterion \( C_1 \) is to minimize the range of \( \mu_c \) or \( \sigma_c \). These ranges are actually independent of \( y \) [see DasGupta and Studden (1988)].

\( C_2 \): For any vector \( c \) and design \( M \), let \( I_c \) be the set of smallest Lebesgue measure such that

\[
\inf_{\pi \in \Gamma_2} P_{\pi}(c'\theta \in I|y) \geq 1 - \alpha.
\]

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 \( c, c'\theta \) is smaller than or equal to its prior expected value under the basic prior \( L \), that is, \( H_0: c'\theta \leq c'\mu \). 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 \( \delta_{ij} \) 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,

\[
\sup_{\pi \in \Gamma_2} P_{\pi}(H_0|y) - \inf_{\pi \in \Gamma_2} P_{\pi}(H_0|y).
\]

**Theorem 4.1.** The criteria \( C_1, C_2, C_3 \) and \( C_4 \) are all equivalent to minimizing \( \Phi_\theta(M) = c'\left(M + \Sigma_0^{-1}\right)^{-1}c \).
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

$\begin{align*}
S_c &= \sqrt{c'(M + \Sigma_0^{-1})^{-1}c} \cdot S_h + \left(c'(M + \Sigma_0^{-1})^{-1}v, 0\right),
\end{align*}$

where $v = X'(y - X\mu)$, and

$\begin{align*}
S_h &= \{(\mu_X, \sigma_X) : X \sim f, \phi \leq \alpha f \leq k\phi \text{ for some } \alpha > 0\}
\end{align*}$

where $\phi$ denotes the standard normal density and $\mu_X, \sigma_X$ denote the mean and standard deviation of $X$. The important point here is that the set $S$ does not depend on either $y$ 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 $\mu = 0$. Since the loss is zero-one, for a fixed design $M$, the Bayes test under $L$ takes action $a_0$ if and only if $P_L(H_0|y) \geq P_L(H_1|y)$. Consequently, the posterior risk equals $g(p) = \min(p, (1 - p))$, where $p = P_L(H_0|y)$ (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 $c'\theta$ under $L$ is $N(c'v, c'\Lambda c)$, where $v$ and $\Lambda$ are as in the proof of Theorem 3.1 (with $\mu = 0$), it follows that $p = \Phi(-c'v/\sqrt{c'\Lambda c})$. Here $\Phi$ denotes the standard normal c.d.f. Let $M_L(t)$ denote the marginal of $y$ under the prior $L$ and $M_L(t)$ denote the marginal of $t = c'v/\sqrt{c'\Lambda c}$ under $L$. Then the Bayes risk of the Bayes test under $L$ is

$\begin{align*}
E_{M_L(y)}(g(p)) &= E_{M_L(t)}(g(\Phi(-t))).
\end{align*}$

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

$\begin{align*}
\tau^2 &= \frac{c'\Sigma_0 M(M + \Sigma_0^{-1})^{-1}c}{c'(M + \Sigma_0^{-1})^{-1}c} = \frac{c'\Sigma_0 c}{c'(M + \Sigma_0^{-1})^{-1}c} - 1.
\end{align*}$

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 $\tau$. This follows since $E[g(\Phi(Z))] = 2\int_{-\infty}^{\infty}g(\Phi(z))n(z; 0, \tau^2)dz$, $g(\Phi(z))$ is decreasing and $2n(z; 0, \tau^2)f(z; 0, \tau^2)f(z; 0, \tau^2)$ has a monotone likelihood ratio. Combining this with (4.8), it follows that (4.7) is increasing in $c'(M + \Sigma_0^{-1})^{-1}c$. This completes the result for $C_3$.

To derive the optimum design for the criterion $C_4$, let $A = \{\theta : c'\theta \leq 0\}$. Then

$\begin{align*}
\sup_{\pi \in \Gamma_2} P_{\pi}(H_0|y) &= \sup_{\pi \in \Gamma_2} \int_A d\pi(\theta|y) \\
(4.9) &= \sup_{\eta} \int_A d\eta(\theta|y) / \int_{R^p} d\eta(\theta|y),
\end{align*}$

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

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

\[
\sup_{\pi} P_{\pi}(H_0|y) - \inf_{\pi} P_{\pi}(H_0|y) = \frac{(k^2 - 1)p(1 - p)}{(1 + (k - 1)p)(k - (k - 1)p)}.
\]

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 \( y \) induced by the prior \( L \)) is increasing in \( c'(M + \Sigma^{-1}_0) \). This follows by a repetition of the argument used to prove that (4.7) is decreasing in \( c'(M + \Sigma^{-1}_0) \). The theorem is now proved. \( \square \)

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 \( \sigma^2 \) is unknown and an appropriate inverse gamma prior is used for \( \sigma^2 \). 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**


**Department of Statistics**  
**Purdue University**  
**West Lafayette, Indiana 47907**