Statistica Sinica 4(1994), 603-615

EFFICIENCY ROBUST EXPERIMENTAL DESIGN AND ESTIMATION USING A DATA-BASED PRIOR

Blaza Toman and Joseph L. Gastwirth

George Washington University

Abstract: In scientific research major studies are often designed after a pilot study has been carried out. This paper uses Bayesian methodology to incorporate the results of the first study (or studies) into the design of the follow-up study. It is assumed that the error variances of the pilot and of the follow-up studies are unknown but that the experimenters are able to give intervals for the possible values of the two variances. This assumption, together with a noninformative prior distribution of the treatment means of the pilot study, leads to a class of prior distributions for the new experiment.

Since the properties of an experimental design can only be judged in reference to a particular estimator, and since the choice of an estimator is in itself a very important problem, we combine the two tasks and simultaneously search for the estimator and the design which achieve the maximin efficiency over the class of posterior distributions.

The method is illustrated on data from "6-Month Drinking Water Studies of Sodium Fluoride", a study, which showed that rats exposed to very high levels of sodium fluoride had diminished growth rates.

Key words and phrases: Analysis of variance, Bayesian experimental design, efficiency robust design and estimation.

1. Introduction

The process of acquiring scientific knowledge is a sequential one. After an experiment yields an interesting result, others are conducted to confirm the findings and further explore the implications of the original experiment. When planning the subsequent experiments, an investigator will wish to incorporate the information available from all related prior studies into the design of the new experiment. This paper proposes a method, based on the Bayesian paradigm, for accomplishing this task.

We consider the situation where data from past experiments come from an analysis of variance model. These data can be combined to form a single prior distribution for the new experiment by employing Bayesian analysis with noninformative priors as carried out by Box and Tiao (1973). When the error variances of all past studies are equal to a known constant, and the error variance of the

planned study is known, the resulting distribution is a multivariate normal. Existing literature on Bayesian experimental design for a wide variety of normal ANOVA models, such as for example Chaloner (1984), Giovagnoli and Verdinelli (1983, 1985), Owen (1970), Piltz (1984), Smith and Verdinelli (1980), Toman and Notz (1991) and Verdinelli (1983) can be applied to derive experimental designs for the new study. Unfortunately, in the more realistic case when the error variances are unknown, the full Bayesian experimental design problem becomes quite intractable. Rather than assume a full prior distribution on the error variances, the method of this article relies on the partial prior information that the error variances lie in known intervals. Using this assumption results in an entire class of possible prior distributions for the new experiment rather than in a single prior as would be the case if a full Bayesian analysis was used. The experimental design problem, which is to find a design which performs well over this class of distributions, is solvable in this framework.

Classes of priors have been used to obtain Bayesian robust estimators by Berger (1985), Berger and Berliner (1986) and many others. Their use in experimental design is relatively recent; a related work is DasGupta and Studden (1990). This paper differs from previous literature in that we derive both the robust estimator and the corresponding robust design, as in this context, the two problems are linked. The criterion we use to select the optimum design and estimator is motivated by the efficiency robustness literature (Gastwirth (1966, 1985), and Birnbaum and Laska (1967)). Since a confidence region for the means may be required, we propose one which is based on the optimal estimator. Finally, we illustrate our method by designing a follow-up experiment to the "6-Month Drinking Water Studies of Sodium Fluoride" study described in U.S. Dept. of HHS (1990).

The model and background information are given in Section 2. The optimal design and estimator is derived in Section 3. Section 4 contains the example and conclusions.

2. Model and Background

We consider the case of the analysis of variance model where $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_k)'$ are the unknown treatment effects. As the "treatment" may mean a particular combination of factor levels, this model is appropriate for one-way analysis of variance as well as multi-factor ANOVA represented by a cell-means model, also called the "μ_{ij}" model by Searle (1971, p.324). The purpose of the experiment is the estimation of $\boldsymbol{\tau}$. We have a total of n experimental units available to be allocated to the k treatment groups. The vector of observations $\mathbf{X}$ the experiment will yield is assumed to have a normal distribution with mean $\mathbf{F}\boldsymbol{\tau}$

and a variance-covariance matrix equal to $\sigma^2\mathbf{I}$, where σ is known to lie in a finite interval $[\sigma_1, \sigma_2]$, $\mathbf{F}$ is the $n \times k$ design matrix, and $\mathbf{I}$ is the $n \times n$ identity matrix. We assume that the observed values $\mathbf{y}$ from a previous related experiment are available. The vector $\mathbf{y}$ is assumed to have come from a normal distribution with mean $\mathbf{G}\tau$ and a variance-covariance matrix equal to $\omega^2\mathbf{I}$, where ω is known to lie in the interval $[\omega_1, \omega_2]$.

The first step in designing the new experiment is the derivation of the distribution of τ after $\mathbf{y}$ was observed. Using a noninformative first stage prior for τ we obtain (see Box and Tiao (1973))

$$\tau|\mathbf{y} \sim N((\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{y}, \omega^2(\mathbf{G}'\mathbf{G})^{-1}), \quad \omega_1 \leq \omega \leq \omega_2. \tag{2.1}$$

Note that expression (2.1) defines a class of distributions indexed by ω and that the mean of τ given $\mathbf{y}$ is simply the vector of sample treatment means from the first experiment. Let $\mathbf{M} = \mathbf{G}'\mathbf{G} = \text{diag}(m_1, m_2, \ldots, m_k)$, where the m_i are the number of units allocated to treatment i in the first experiment. The available prior information for use in planning of the new experiment can be summarized as follows:

The vector τ of means has a normal prior distribution, denoted by π.

We assume that π is a member of the class:

$$\Gamma = \{\pi : \pi \text{ is normal and } E(\tau) = \theta = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{y},$$
$$\text{Var}(\tau) = \omega^2\text{diag}(1/m_i), i = 1, \ldots, k,\ \omega_1 \leq \omega \leq \omega_2\} \tag{2.2a}$$
$$\sigma_1 \leq \sigma \leq \sigma_2. \tag{2.2b}$$

Expression (2.2a) restates (2.1), and (2.2b) reflects the prior information on σ. When relevant data on different subsets of the experimental treatments are available from several past experiments, one may subsume this information in a more general class of priors. This class is described as

$$\Gamma = \{\pi : \pi \text{ is normal and } E(\tau) = \theta = (\theta_1, \theta_2, \ldots, \theta_k)',$$
$$\text{Var}(\tau) = \mathbf{T} = \text{diag}(t_i),\ t_{iL} \leq t_i \leq t_{iU},\ i = 1, \ldots, k\}. \tag{2.3}$$

The vector θ is the vector of sample treatment means from the past experiments. The values of t_{iL} and t_{iU}, $i = 1, \ldots, k$, are chosen on the basis of the ranges of values of the error variances of the previous experiments and on the sample sizes of each treatment group. The prior information available for the design of the new experiment is given by (2.2b) and (2.3).

Since the distribution of $\mathbf{X}|(\tau, \sigma)$ is normal, for each value of σ and for each prior π in Γ, the corresponding posterior distribution of $\tau|\mathbf{x}$ is also normal with

mean $\bar{\mathbf{x}}-(\sigma^{-2}\mathbf{N}+\mathbf{T}^{-1})^{-1}\mathbf{T}^{-1}(\bar{\mathbf{x}}-\theta)$, and covariance matrix $\mathbf{V}=(\sigma^{-2}\mathbf{N}+\mathbf{T}^{-1})^{-1}$. The vector $\bar{\mathbf{x}}$ is the $k \times 1$ vector of the sample treatment means, the matrix $\mathbf{N}=\mathbf{F}'\mathbf{F}=\text{diag}(n_1, n_2, \ldots, n_k)$ is the allocation matrix of the new experiment. It will be shown in the next section that the optimal estimator and design depend on the prior covariance matrix $\mathbf{T}$ only through the ratios t_i/σ^2. Therefore it is mathematically convenient to express the prior covariance matrix as $\sigma^2\mathbf{S}$, where $\mathbf{S}=\text{diag}(s_i)$, $s_i=t_i/\sigma^2$. The class of prior distributions given in (2.3) together with the prior information that $\sigma_1 \leq \sigma \leq \sigma_2$ therefore leads to the class of posterior distributions

$$\begin{aligned}\Gamma_p = \{\, &\pi(\tau|\mathbf{x}) : \pi(\tau|\mathbf{x}) \text{ are normal with mean } \bar{\mathbf{x}} - (\mathbf{N}+\mathbf{S}^{-1})^{-1}\mathbf{S}^{-1}(\bar{\mathbf{x}}-\theta),\\ &\text{covariance matrix } V=\sigma^2(\mathbf{N}+\mathbf{S}^{-1})^{-1},\\ &\sigma_1 \leq \sigma \leq \sigma_2, s_{iL}(\sigma) \leq s_i \leq s_{iU}(\sigma), i=1,\ldots,k \,\}, \end{aligned} \tag{2.4}$$

where $s_{iL}(\sigma)=t_{iL}/\sigma^2$ and $s_{iU}(\sigma)=t_{iU}/\sigma^2$.

In the new experiment we allow the possibility of not observing one or more of the treatments. In such a case the allocation matrix $\mathbf{N}$ may have some rows of zeros, that is $\mathbf{N}=\text{diag}[\mathbf{N}_E, \mathbf{0}]$ where $\mathbf{N}_E$ is the allocation matrix of the treatments in the experiment that are observed, and $\mathbf{0}$ is a zero matrix of appropriate dimension. We allow the elements of $\mathbf{N}$ to have non integer values. This means that after the optimal design is derived, the n_i's must be rounded off to obtain a usable design. The set of all possible designs is defined by this set of allocation matrices denoted by Ξ.

In the next section, we seek the robust estimator and design.

3. Derivation of the Robust Estimator and Design

Our purpose is to arrive at a design and estimator which have desirable efficiency properties over the entire class of prior distributions and for all values of σ in the interval $[\sigma_1, \sigma_2]$. We use Bayes risk (rather than expected posterior loss) as the basis of our criterion since the experimental design must be selected before the data from the new experiment is collected. In classical inference, Gastwirth (1966, 1985), Birnbaum and Laska (1967), Burnett, Krewski and Bleuer (1989) obtain procedures having a high asymptotic relative efficiency compared to the optimal one for each member of a family of distributions which may generate the data. We adopt the analogous criterion of relative efficiency. We limit our search for the best estimator to the class of linear estimators

$$\Delta = \{\delta(\mathbf{x}) : \delta(\mathbf{x}) = \mathbf{A}\bar{\mathbf{x}} + (\mathbf{I}-\mathbf{A})\theta,\ \mathbf{A}=\text{diag}(a_i),\ 0 \leq a_i \leq 1\}. \tag{3.1}$$

Under squared error loss this class contains, but is not limited to, the Bayes estimators of τ, since these are simply the posterior means for the class Γ_p. It is

possible that broadening the class of estimators could produce a more efficient pair $(\delta, \mathbf{N})$. However, our results indicate that we can obtain a reasonably efficient procedure among linear estimators.

A procedure $(\delta_M, \mathbf{N}_M)$ will be called optimal if it achieves

$$e_{\Gamma M} = \inf_{\substack{\mathbf{N}\in\Xi \\ \delta\in\Delta}} \sup_{\substack{\pi\in\Gamma \\ \sigma_1\le\sigma\le\sigma_2}} \frac{r_{\pi,\sigma}(\delta, \mathbf{N})}{r_{\pi,\sigma}(\delta_{\pi,\sigma}, \mathbf{N}_{\pi,\sigma})} = \inf_{\substack{\mathbf{N}\in\Xi \\ \delta\in\Delta}} \sup_{\substack{\pi\in\Gamma \\ \sigma_1\le\sigma\le\sigma_2}} e_{\pi,\sigma}(\delta, \mathbf{N}). \tag{3.2}$$

The quantity $r_{\pi,\sigma}(\delta, \mathbf{N})$ is the Bayes risk of the estimator δ and design $\mathbf{N}$ for a particular σ and the prior π. The pair $\delta_{\pi,\sigma}$ and $\mathbf{N}_{\pi,\sigma}$ are the Bayes estimator and design for the prior π and the particular σ. The quantity $e_{\pi,\sigma}(\delta, \mathbf{N})$ is the inverse of the efficiency of the pair $(\delta, \mathbf{N})$ under the prior π, for the particular σ. The procedure $(\delta_M, \mathbf{N}_M)$ satisfying criterion (3.2) minimizes the maximum Bayes risk relative to the optimum choice $(\delta_{\pi,\sigma}, \mathbf{N}_{\pi,\sigma})$ for each σ and $\pi \in \Gamma$.

We next obtain an expression for $e_{\pi,\sigma}(\delta, \mathbf{N})$. For a particular estimator δ the vector $\mathbf{a} = (a_1, a_2, \ldots, a_k)'$ may be partitioned as $\mathbf{a} = [\mathbf{a}_E, \mathbf{0}]'$, since the elements of τ corresponding to treatments which will not be observed must be estimated purely from the prior information. The risk of δ and a particular design $\mathbf{N}$ for a given τ and σ^2 is

$$R_\sigma(\delta, \mathbf{N}, \tau) = \sigma^2 \mathbf{a}_E' \mathbf{N}_E^{-1} \mathbf{a}_E + (\tau - \theta)'(1 - \mathbf{a})(1 - \mathbf{a})'(\tau - \theta). \tag{3.3}$$

For any σ, taking expectation of this function with respect to a particular prior π in Γ, yields the Bayes risk of the pair $(\delta, \mathbf{N})$

$$r_{\pi,\sigma}(\delta, \mathbf{N}) = \sigma^2[\mathbf{a}_E' \mathbf{N}_E^{-1} \mathbf{a}_E + (\mathbf{1} - \mathbf{a})' \mathbf{S} (\mathbf{1} - \mathbf{a})]. \tag{3.4}$$

The vector $\mathbf{1}$ is a $k \times 1$ vector of ones. Next, we find the Bayes design $\mathbf{N}_{\pi,\sigma}$ and the Bayes risk $r_{\pi,\sigma}(\delta_{\pi,\sigma}, \mathbf{N}_{\pi,\sigma})$ for a particular prior distribution $\pi \in \Gamma$ and σ. Since for this model and loss function the Bayes risk of the Bayes estimator $\delta_{\pi,\sigma}$ is

$$r_{\pi,\sigma}(\delta_{\pi,\sigma}, \mathbf{N}) = \sigma^2 \text{trace}(\mathbf{N} + \mathbf{S}^{-1})^{-1}, \tag{3.5}$$

the matrix $\mathbf{N}_{\pi,\sigma}$ must minimize expression (3.5). This matrix, called the Bayes A-optimal design, was derived in general form by Owen (1970). The particular solution for our problem, obtained by using the method of Lagrange multipliers, is:

> Let L (obtained using an algorithm presented in the Appendix) be a subset of m of the integers $1, \ldots, k$. Then the matrix $\mathbf{N}_{\pi,\sigma}$ consists of elements n_i:

for all $i \in L$, $n_i = 0$;
for $j \notin L$, if $m = k-1$, $n_j = n$;
if $m < k-1$, $n_j = n/(k-m)$
$$+ \left[\sum_{\substack{i \notin L \\ i \neq j}} s_i^{-1} - (k-m-1)s_j^{-1} \right] \Big/ (k-m). \quad (3.6)$$

Substituting the matrix $\mathbf{N}_{\pi,\sigma}$ into expression (3.5) we obtain the Bayes risk of the Bayes estimator and design

$$r_{\pi,\sigma}(\delta_{\pi,\sigma}, \mathbf{N}_{\pi,\sigma}) = \sigma^2 \left[(k-m)^2 \left(n + \sum_{i \notin L} s_i^{-1} \right)^{-1} + \sum_{j \in L} s_j \right]. \quad (3.7)$$

Dividing (3.4) by (3.7) yields $e_{\pi,\sigma}(\delta, \mathbf{N})$.

Now we find the pair $(\delta_M, \mathbf{N}_M)$. The criterion given by (3.2) is quite complex, and appears to require a search over the infinite classes Ξ, Δ, Γ, and the interval $\sigma_1 \leq \sigma \leq \sigma_2$. The process can be greatly simplified upon noting the following:

Lemma 1. *For a given estimator δ, the Bayes risk $r_{\pi,\sigma}(\delta, \mathbf{N})$ in (3.4) is minimized by the same design denoted by $\mathbf{N}_\delta$, for all σ and $\pi \in \Gamma$. When at least one $a_i \neq 0$, the matrix $\mathbf{N}_\delta$ has elements $n_i = n\, a_i / \Sigma\, a_j$, $i = 1, \ldots, k$. In the case when $\mathbf{a} = 0$, $\mathbf{N}_\delta$ can be any allocation matrix $\mathbf{N}$.*

Proof. This result is a direct consequence of the fact that for a given δ, the vector $\mathbf{a}$ is specified and the term in (3.4) which involves the design $\mathbf{N}_E$ does not depend on the prior distribution. A simple calculation yields the result.

As a result of Lemma 1, it is clear that the solution $(\delta_M, \mathbf{N}_M)$ must achieve

$$e_{\Gamma M} = \inf_{\delta \in \Delta} \sup_{\substack{\pi \in \Gamma \\ \sigma_1 \leq \sigma \leq \sigma_2}} e_{\pi,\sigma}(\delta, \mathbf{N}_\delta). \quad (3.8)$$

Combining (3.4), (3.7) and the result of Lemma 1 yields

$$e_{\pi,\sigma}(\delta, \mathbf{N}_\delta) = \left[\frac{1}{n} \mathbf{a}'\mathbf{J}\mathbf{a} + (\mathbf{1}-\mathbf{a})'\mathbf{S}(\mathbf{1}-\mathbf{a}) \right] \Big/ \left[(k-m)^2 \left(n + \sum_{i \notin L} s_i^{-1} \right)^{-1} + \sum_{j \in L} s_j \right], \quad (3.9)$$

where $\mathbf{J}$ is a $k \times k$ matrix of ones. The optimization can be further simplified by noting that σ does not appear in the expression (3.9). Thus, the particular value σ and the particular prior π enter this expression only through the matrix

$\mathbf{S}$. Hence the search for the supremum of $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ over the interval $[\sigma_1, \sigma_2]$ and the class Γ in (3.8) is equivalent to the search over the convex and compact set

$$\Pi = \{\mathbf{S} : \mathbf{S} = \text{diag}(s_i),\ s_{Li} \leq s_i \leq s_{Ui}\}. \tag{3.10}$$

The values of s_{Li} and s_{Ui} can be obtained as:

$$s_{Li} = t_{iL}/\sigma_2^2, \quad s_{Ui} = t_{iU}/\sigma_1^2. \tag{3.11}$$

The final simplification of the search process is presented in

Theorem 1. *The unique robust estimator δ_M achieves $e_{\Gamma M} = \inf_{\delta\in\Delta} \sup_{\mathbf{S}\in\Pi_E} e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$, where Π_E is the set of the 2^k extreme points of Π.*

Proof. For each δ, $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ is a function of $\mathbf{S}$ defined on the convex and compact set Π. The function $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ changes form as the set L acquires new members when constraints such as $s_i = (k-1)\left(\sum_{j\neq i} s_j^{-1} + n\right)^{-1}$ are satisfied for some $i \notin L$. The second-order partial derivatives $\frac{\partial^2}{\partial s_i \partial s_j} e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ exist and are finite everywhere, including the boundaries where $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ changes form. The Hessian matrix is positive definite, thus implying that $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ is strictly convex over the set Π. Hence for each δ, $\sup_{\mathbf{S}\in\Pi} e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ is achieved for a matrix $\mathbf{S}$ corresponding to one of the 2^k extreme points of Π.

Now for each $\mathbf{S}$, the function $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ is strictly convex in $\mathbf{a}$ and finite over the convex, compact set of all possible $\mathbf{a}$, that is the set $\{\mathbf{a} : \mathbf{a} = (a_1, a_2, \ldots, a_k)',\ 0 \leq a_i \leq 1,\ i = 1, \ldots, k\}$. Therefore the function $\sup_{\mathbf{S}\in\Pi_E} e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ is also strictly convex in $\mathbf{a}$. Consequently $e_{\Gamma M}$ is achieved, and the estimator δ_M (or equivalently $\mathbf{a}_M$) is unique.

The solution $\mathbf{a}_M$ can now be found by a numerical search procedure, such as the one given by Charalambous and Conn (1978). Once $\mathbf{a}_M$ is found, the design $\mathbf{N}_M$ is computed using Lemma 1.

Given σ, it is possible to identify the normal prior π_M for which δ_M is the Bayes estimator. We may think of this as the "robust" prior for our class Γ. This is useful if, after the data is collected, we wish to obtain credible sets for τ or for linear functions of τ. The set

$$C = \{\tau : (\tau - \delta_M(\mathbf{x}))'\mathbf{V}^{-1}(\tau - \delta_M(\mathbf{x})) \leq \chi_k^2(1-\alpha)\}, \tag{3.12}$$

where $\mathbf{V} = \sigma^2(\mathbf{N}_M + \mathbf{S}_M^{-1})^{-1}$, is a good choice for a $100(1-\alpha)\%$ credible set for τ. The unknown σ^2 can now be estimated from the new data. In the next section we illustrate the derivation of δ_M, $\mathbf{N}_M$ and the region C for a particular example and compare the efficiency of the estimator and design and the coverage probability of the confidence region to other procedures.

4. Application to a Potential Follow-up Study

One of the objectives of the "6-Month Drinking Water Studies of Sodium Fluoride", U.S. Dept. of HHS (1990), was to determine the effect of various doses of sodium fluoride on the body weights of rats. The experimental groups consisted of thirty rats which were the controls and four other groups of ten rats, each of which were given one of four experimental dosages of Sodium Fluoride. The controls were divided into three groups, the first and second of these received the semisynthetic, low fluoride diet which was also fed to the treated rats, the third group received a different diet. Therefore, we considered the first two control groups as the zero dose. Table 1 contains the means and standard errors of the five treatment groups.

Table 1. Mean change in body weights of the male rats

Dose	Sample size	mean $\pm$ standard error
0	20	369 ± 9
10	10	349 ± 7
30	10	359 ± 10
100	10	357 ± 5
300	10	290 ± 8

The results of the study were interesting as there was a significant effect on body weight only at the highest dose level. Since the sample sizes were small, the results of the study should be confirmed on both the same and other species of rats before concluding that fluoride is potentially harmful. We now show how our method could be used to design a follow-up study.

We begin by defining the class Γ of priors. First, the prior mean of all members of Γ is $\boldsymbol{\theta} = (369, 349, 359, 357, 290)'$. Second, using the fact that this experiment is of the type where one treatment is a control and the rest are actual experimental treatments, and the fact that the previous experiment was balanced in the experimental treatments, the matrix $\mathbf{S}$ can be written as $\mathbf{S} = 1/c$ diag$(1, d, \ldots, d)$ where $c = (\sigma/\omega)^2 m_c$, and $d = m_c/m_t$. Since m_c equals 20 and m_t equals 10, the value of d is 2. We enlarge the class allowing $1.9 \leq d \leq 2.5$. This is reasonable since the control in this experiment is "no treatment", so its prior variance should be less than that of the other treatments. Next, we examine the values of the standard errors. These range between 5 and 10, so we conclude that ω^2 ranged between 250 and 1000. It is now necessary to decide on a range for σ^2, the error variance for the new planned experiment. If the new experiment is done under similar conditions, it is reasonable to expect that σ^2 will be similar to ω^2. Therefore, let σ^2 range between 250 and 1000. If one thought that all possible values of ω^2 were equally likely to pair with each value of σ^2, one would

obtain the range of possible values of their ratio as $0.25 \leq (\sigma/\omega)^2 \leq 4.0$. Since we believe however, that the values of the two error variances are similar, it is much more likely that this range will be narrower. We take $0.75 \leq (\sigma/\omega)^2 \leq 1.35$. The assumed ranges of $(\sigma/\omega)^2$ and d lead to the set

$$\Pi = \{\mathbf{S} : \mathbf{S} = \text{diag}(s_1, s_2, s_2, s_2, s_2), \text{ where } 0.037 \leq s_1 \leq 0.067,\ 0.07 \leq s_2 \leq 0.17\}. \tag{4.1}$$

Next, it is necessary to determine the functional form of $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ for the matrices $\mathbf{S}$ in Π_E. Thus the set L, indicating which treatments will not be used for each of the four matrices in Π_E is needed. The algorithm presented in the Appendix reduces to checking for each matrix whether

(i) $s_1 < 4s_2/(4+n\,s_2)$; in this case the allocation to the control treatment is zero, that is $L = \{1\}$.
or
(ii) $s_2 < s_1/(1+n\,s_1)$; in this case the allocation to the control treatment is n, that is $L = \{2,3,4,5\}$.
or
(iii) neither (i) or (ii) obtains, that is L is empty and all treatments are used.

The $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ are then given by (3.9).

Let us find $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ for the four matrices in Π_E. Since L depends on n, the total sample size of the planned experiment, the results will be presented for $n = 75, 100, 125, 150$. For $n = 75$, $L = \{1\}$ for the matrix which has $s_1 = 0.037$ and $s_2 = 0.17$. For all remaining matrices and sample sizes L is empty. Hence, for $n = 75$, $s_1 = 0.037$, and $s_2 = 0.17$

$$e_{\pi,\sigma}(\delta, \mathbf{N}_\delta) = [1/n\,\mathbf{a}'\mathbf{J}\mathbf{a} + (\mathbf{1}-\mathbf{a})'\mathbf{S}(\mathbf{1}-\mathbf{a})]/[4^2(n+4\,s_2^{-1})^{-1} + s_1]. \tag{4.2}$$

For the other three matrices in Π_E

$$e_{\pi,\sigma}(\delta, \mathbf{N}_\delta) = [1/n\,\mathbf{a}'\mathbf{J}\mathbf{a} + (\mathbf{1}-\mathbf{a})'\mathbf{S}(\mathbf{1}-\mathbf{a})]/[5^2(n+s_1^{-1}+4s_2^{-1})^{-1}]. \tag{4.3}$$

For the larger sample sizes $e_{\pi,\sigma}(\delta, \mathbf{N}_\delta)$ is given by (4.3) for all four matrices in Π_E. Numerical search using the method of Charalambous and Conn (1978) results in the $\mathbf{a}_M$ and $\mathbf{N}_M$ given in Table 2 for the four different sample sizes. Note that the quantities n_c and n_t, that is the allocation of the control and of the experimental treatments, would now be rounded off to obtain a usable design.

Table 2. The robust designs and estimators, and their relative efficiencies

a_c	a_t	n_c	n_t	n	$e_{\pi,\sigma}^{-1}$	Eff_1	Eff_2
0.17	0.71	4.2	17.7	75	84%	67%	72%
0.32	0.8	9	22.75	100	86%	72%	75%
0.43	0.82	14.5	27.63	125	89%	77%	77%
0.51	0.84	19.8	32.55	150	92%	80%	76%

In Table 2 the column labeled $e^{-1}_{\pi,\sigma}$ gives the minimum efficiency of the procedure $(\delta_M, \mathbf{N}_M)$ over the entire set Π. The column labeled Eff_1 gives the minimum efficiency over the set Π of the standard estimator $\delta = \overline{\mathbf{x}}$ and the design $n_c = n_t = n/5$ which would be used in the absence of the prior information. To assess the potential loss of efficiency when procedures $(\delta, \mathbf{N})$ are used instead of the optimal one, we computed the minimum efficiency over the set Π for the Bayes estimators and designs corresponding to the four extreme matrices $\mathbf{S}$. The smallest of these efficiencies is Eff_2 and reflects the maximum loss over Π of a Bayesian procedure. For moderate sample sizes, comparing Eff_1 and Eff_2 indicates that even the worst Bayesian choice is preferable to ignoring the prior information. Comparing the columns $e^{-1}_{\pi,\sigma}$ and Eff_1 provides a measure of the value of the prior information when used in a robust manner. Comparison of $e^{-1}_{\pi,\sigma}$ and Eff_2 gives a measure of robustness of $(\delta_M, \mathbf{N}_M)$ with respect to other Bayes procedures corresponding to members of the set Π. The results in Table 2 show that using the procedure $(\delta_M, \mathbf{N}_M)$ results in a substantial gain in efficiency for all sample sizes. For the relatively small sample sizes such as $n = 75$, which are used in these experiments, the gain is noteworthy.

After the new data is collected, credible sets for the vector τ should be obtained. For posteriors in Γ_p the Bayes estimator δ corresponds to a vector $\mathbf{a}$, such that $a_i = 0$ if $n_i = 0$, and $a_i = \frac{n_i s_i}{n_i s_i + 1}$ if $n_i \neq 0$. Therefore, given the vector $\mathbf{a}_M$ and $\mathbf{N}_M$ we can solve for $\mathbf{S}_M$. Consider the case when $n = 75$. Here we find that $\mathbf{S}_M = \text{diag}(0.049, 0.138, 0.138, 0.138, 0.138)$. Credible sets such as (3.12) can now be found using an estimated value (from the new data) for σ^2. In simulations, where τ was generated from different priors in Γ, including the extreme priors corresponding to $\mathbf{S}$ in Π_E, the coverage probability of the 95% credible set based on δ_M was between 93% and 97%. On the other hand, when a 95% credible set was computed for one of the extreme priors and τ was generated from another, coverage probabilities were as low as 77%.

To evaluate the sensitivity of the procedure $(\delta_M, \mathbf{N}_M)$ to changes in the specification of the prior information, we computed the procedures and the corresponding efficiencies for the case where $0.5 \leq (\sigma/\omega)^2 \leq 2.0$, and where $0.25 \leq (\sigma/\omega)^2 \leq 4.0$. These results are presented in Tables 3 and 4.

Table 3. The robust designs and estimators and efficiency comparisons when $0.5 \leq (\sigma/\omega)^2 \leq 2.0$

a_c	a_t	n_c	n_t	n	$e^{-1}_{\pi,\sigma}$	Eff_1	Eff_2	Eff_3
0.18	0.72	4.4	17.65	75	80%	65%	69%	78%
0.33	0.81	9.24	22.69	100	83%	70%	70%	81%
0.43	0.82	14.5	27.63	125	87%	76%	75%	87%
0.51	0.84	19.8	32.55	150	90%	79%	77%	90%

Table 3 shows that enlarging the range of $(\sigma/\omega)^2$ slightly does not have much effect on the procedures $(\delta_M, \mathbf{N}_M)$ which are quite similar to those based on the original prior information. A slight increase in the n_c is observed. The efficiencies decreased as expected but the gain in using the robust procedure remains. The column labeled Eff_3 gives the minimum efficiencies over this set Π of the original procedures $(\delta_{M_1}, \mathbf{N}_{M_1})$ presented in Table 2. The fact that the values Eff_3 and $e_{\pi,\sigma}^{-1}$ are very close or equal, seems to indicate that the robust procedure $(\delta_{M_1}, \mathbf{N}_{M_1})$ obtained for the smaller class of priors remained quite robust over the enlarged class.

To explore a large degree of uncertainty about the ratio of the error variances we consider the case when $0.25 \leq (\sigma/\omega)^2 \leq 4.0$ in Table 4.

Table 4. The robust designs and estimators and efficiency comparisons when $0.25 \leq (\sigma/\omega)^2 \leq 4.0$

a_c	a_t	n_c	n_t	n	$e_{\pi,\sigma}^{-1}$	Eff_1	Eff_2	Eff_3
0.22	0.82	4.6	17.6	75	78%	60%	59%	75%
0.37	0.86	9.6	22.6	100	81%	68%	69%	79%
0.48	0.89	15.0	27.5	125	83%	74%	74%	80%
0.55	0.89	20.08	32.47	150	88%	79%	79%	87%

It can be seen from Table 4 that the efficiency of the robust $(\delta_M, \mathbf{N}_M)$ for this class has decreased relative to the efficiency of the robust $(\delta_M, \mathbf{N}_M)$ for the smaller class considered in Table 2. This was expected since the range of $(\sigma/\omega)^2$ is now quite large. The decrease at all sample sizes is not large, on the order of about 5 to 6%. Furthermore, comparison of $e_{\pi,\sigma}^{-1}$ with Eff_1 and Eff_2 shows that the use of our procedure still results in substantial gain in efficiency, especially in the small to moderate sample size case. The estimators δ_M and especially the allocation matrices $\mathbf{N}_M$ have remained quite similar to the original ones. The comparison of Eff_3 with $e_{\pi,\sigma}^{-1}$ in this more extreme case shows only a slightly larger impact of the possible misspecification in the prior class. On the basis of this numerical example, we conclude that the procedure $(\delta_M, \mathbf{N}_M)$ appears to be quite robust to changes in the set Π.

Appendix

The Algorithm for obtaining the set L, that is the set of treatments which will not be used in the experiment.

1. Find i_1 such that $s_{i_1} < (k-1)\left(\sum_{j \neq i_1} s_j^{-1} + n\right)$. $L = \{i_1\}$.
   If no such i_1 exists then L is empty and STOP.
   If more than one candidate for i_1 exists, pick the one for which

$s_{i_1} - (k-1)\left(\sum_{j \neq i_1} s_j^{-1} + n\right)$ is the smallest. STOP if $k = 2$.

2. Find i_2 such that $s_{i_2} < (k-2)\Big(\sum_{\substack{j \notin L \\ j \neq i_2}} s_j^{-1} + n\Big)$. $L = \{i_1, i_2\}$.

   If no such i_2 exists STOP. If $k = 3$ STOP.

3. Continue till no i_m exists such that $s_{i_m} < (k-m)\Big(\sum_{\substack{j \notin L \\ j \neq i_m}} s_j^{-1} + n\Big)$

   or till $m = k-1$.

Acknowledgement

This research was supported in part by National Science Foundation Grant number SES-9001458.

References

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. Springer-Verlag, New York.

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

Birnbaum, A. and Laska, E. (1967). Efficiency robust two-sample rank tests. *J. Amer. Statist. Assoc.* **62**, 1241-1257.

Box, G. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading.

Burnett, R., Krewski, D. and Bleuer, S. (1989). Efficiency robust score tests for rodent tumourigenicity experiments. *Biometrika* **76**, 317-324.

Chaloner, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* **12**, 283-300.

Charalambous, C. and Conn, R. (1978). An efficient method to solve the minimax problem directly. *SIAM J. Numer. Anal.* **15**, 162-187.

DasGupta, A. and Studden, W. J. (1990). Robust Bayesian experimental designs in normal linear models. Purdue University Technical Report #88-34C.

Gastwirth, J. L. (1966). On robust procedures. *J. Amer. Statist. Assoc.* **61**, 929-948.

Gastwirth, J. L. (1985). The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *J. Amer. Statist. Assoc.* **80**, 380-384.

Giovagnoli, A. and Verdinelli, I. (1983). Bayes D-optimal and E-optimal block designs. *Biometrika* **70**, 695-706.

Giovagnoli, A. and Verdinelli, I. (1985). Optimal block designs under a hierarchical linear model. *Bayes. Statist.* **2**, 655-662.

Owen, R. J. (1970). The optimum design of a two-factor experiment using prior information. *Ann. Math. Statist.* **41**, 1917-1934.

Piltz, J. (1984). *Bayesian Estimation and Experimental Design in Linear Regression.* Teubner-Texte zur Mathematik, Teubner-Verlag, Leipzig.

Searle, S. R. (1971). *Linear Models.* John Wiley, New York.

Smith, A.F.M. and Verdinelli, I. (1980). A note on Bayes designs for inference using a hierarchical linear model. *Biometrika* **67**, 613-619.

Toman, B. and Notz, W. (1991). Bayesian optimal experimental design for treatment-control comparisons in the presence of two-way heterogeneity. *J. Statist. Plann. Inference* **27**, 51-63.

U.S. Department of Health and Human Services (1990). Toxicology and carcinogenesis, status of sodium fluoride. *NIH Publication* No. **90-2848**.

Verdinelli, I. (1983). Computing Bayes D- and E-optimal block designs for a two-way model. *Statistician* **32**, 161-167.

Department of Statistics, George Washington University, Washington, DC 20052, U.S.A.

(Received August 1991; accepted December 1993)