SOME BAYESIAN CONSIDERATIONS OF THE CHOICE OF DESIGN FOR RANKING, SELECTION AND ESTIMATION

G. C. TIAO AND B. AFONJA

(Received Mar. 13, 1972)

A Bayesian decision theory approach is used to study the multipurpose choice of design for ranking, selection and estimation. Various utility functions, some of which lead to such standard criteria as the expected mean and probability of correct selection, the trace and determinant, and others leading to a new ranking and a modified determinant criteria are considered. Sensitivity of the optimal design to changes in these utility functions is investigated. In particular, a balanced design is seen to be in general not necessarily better than an unbalanced one. But when choosing among balanced designs, optimality of a design is invariant over all the criteria.

1. Introduction

In recent years, there has been a resurgence of interest in Bayesian statistics—see the review by Lindley [20] and the references therein. Much of the work has been concentrated on the problem of inference. For the equally important problem of the choice of design, Box and Hunter [2] and Draper and Hunter [8] have proposed solutions for a number of linear and non-linear models for estimation purpose, and Ericson [11], [12] and Draper and Guttman [7] have investigated the allocation problem for various sampling schemes. In addition, Dunnett [9] and Raiffa and Schlaifer [23] have considered the question of optimal sample size for the selection of the best treatment. In all these works the solutions are obtained on the basis of some (implicitly or explicitely) chosen utility or loss criteria.

In practice, we are often in a situation where, due to physical or economic limitations, the range of available experiments are very limited and a choice must be made from them. Moreover, such experiments are rarely conducted for a single preconceived purpose which can be well represented by a rigidly defined utility criterion. For example, in a randomized block experiment involving k treatments, one objective may simply be to make inferences about the k means in a general way.

At the same time, certain linear functions of the means may be more important than others. In addition, it may also be desirable to know which of the treatments has the largest mean or to order the magnitude of all the means. While each of these objectives might be adequately represented by a simple utility function, the problem will rapidly become intractable if the multitude of objectives are to be simultaneously considered. A natural way out of such difficulty is to investigate the sensitivity of design selection to changes in utility or loss functions corresponding to different objectives. One is then led to choose designs which are optimal or near optimal with respect to as many as possible relevant utility criteria.

This paper adopts a Bayesian decision theory approach to study the choice of design with respect to a variety of utility functions for ranking, selection and estimation. Section 2 provides a summary of the decision framework and distributional results. Section 3 discusses various utility functions for ranking and selection and Section 4 deals with the problem of estimation. The criteria proposed in these last two sections are then applied in Section 5 to the problem of choosing among a class of balanced designs which include the usual completely randomized complete blocks and balanced incomplete block designs. The paper ends with a comparison of some balanced and unbalanced designs.

2. Basic framework

To facilitate subsequent discussion, we provide below a brief summary of the Bayesian decision analysis, Raiffa and Schlaifer [23] and DeGroot [6], and some key distributional results.

Suppose a choice is to be made from a set of available designs $D=(d_1, \cdots, d_t)$. Associated with a specific design $d \in D$, let $\boldsymbol{y}$ be a set of observations whose distribution $p(\boldsymbol{y}|\boldsymbol{\theta}, d)$ depends upon d and a vector $\boldsymbol{\theta}$ of parameters of interest. Then, given $(\boldsymbol{y}, d)$, solutions to all problems of *inference* or *decision* concerning $\boldsymbol{\theta}$ must be based upon the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}, d) \propto p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta}, d)$ where $p(\boldsymbol{\theta})$ is the prior distribution. In selecting d, one needs to specify a utility function $\bar{U}(d, \boldsymbol{y})$ (or loss function $-\bar{U}(d, \boldsymbol{y})$) which is related to features of the posterior distribution of primary concern. For instance, in an inference problem $\bar{U}$ might be the reciprocal of the generalized variance. The optimal design d^o is then the one for which

$$\bar{\bar{U}}(d^o)=\max_d \bar{\bar{U}}(d)\,, \qquad \bar{\bar{U}}(d)=\mathrm{E}_{\boldsymbol{y}}\, \bar{U}(d, \boldsymbol{y})\,, \tag{2.1}$$

where E is taken over the marginal distribution $p(\boldsymbol{y})=\int p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta}, d)d\boldsymbol{\theta}$.

In problems of decision, including decision theory approach to inference, one faces the further task of making a choice from a set of available actions A whose consequences depend upon $\boldsymbol{\theta}$. It is then necessary to specify a utility (or loss) function $U(d, \boldsymbol{y}, a, \boldsymbol{\theta})$, where $a \in A$. The function $\bar{U}(d, \boldsymbol{y})$ is now the expected utility of the optimal act $a^\circ = a^\circ(d, \boldsymbol{y})$ such that

$$\bar{U}(d, \boldsymbol{y}) = U(d, \boldsymbol{y}, a^\circ) = \max_a \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{y}} U(d, \boldsymbol{y}, a, \boldsymbol{\theta}) \tag{2.2}$$

where E is taken over the posterior distribution of $\boldsymbol{\theta}$. Thus, in choosing a design, the needed specifications are (i) the joint distribution $p(\boldsymbol{y}, \boldsymbol{\theta}|d)$ and (ii) the utility function $\bar{U}(d, \boldsymbol{y})$, or in a decision problem, $U(d, \boldsymbol{y}, a, \boldsymbol{\theta})$.

Throughout the paper we assume, for each design, a linear model

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times k}{\boldsymbol{A}}\ \underset{k\times 1}{\boldsymbol{\theta}} + \underset{n\times 1}{\boldsymbol{e}} \tag{2.3}$$

where $\boldsymbol{A}$ is a fixed matrix of rank k and $\boldsymbol{e}$ is distributed as $N(\boldsymbol{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega}$ assumed known. The structural specification of a design is then given by the triplet $(n, \boldsymbol{A}, \boldsymbol{\Omega})$. We suppose that, *a priori*, $\boldsymbol{\theta}$ is distributed as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_0 = \{\sigma_{ij}^{(0)}\}$ are assumed known and independent of the choice of design. Such a prior distribution might arise as the result of a previous experiment. It follows that (see e.g. Raiffa and Schlaifer [23])

(i) the likelihood function depends only on the two sample quantities

$$\boldsymbol{\Sigma}_s = (\boldsymbol{A}'\boldsymbol{\Omega}^{-1}\boldsymbol{A})^{-1}, \qquad \hat{\boldsymbol{\theta}}_s = \boldsymbol{\Sigma}_s\boldsymbol{A}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}\ ; \tag{2.4}$$

(ii) *a posteriori*

$$\boldsymbol{\theta} \sim N(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \tag{2.5}$$

where

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu} + \boldsymbol{\Sigma}_s^{-1}\hat{\boldsymbol{\theta}}_s) \qquad \text{and} \qquad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0(\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_s)^{-1}\boldsymbol{\Sigma}_s = \{\sigma_{ij}\}\ ;$$

(iii) $\boldsymbol{\theta}$ and the posterior mean vector $\hat{\boldsymbol{\theta}}$ are jointly distributed as multivariate normal with

$$\mathrm{E}(\boldsymbol{\theta}', \hat{\boldsymbol{\theta}}') = (\boldsymbol{\mu}', \boldsymbol{\mu}') \qquad \mathrm{Cov}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \begin{bmatrix} \boldsymbol{\Sigma}_0 & \boldsymbol{V} \\ \boldsymbol{V} & \boldsymbol{V} \end{bmatrix} \tag{2.6}$$

where

$$\boldsymbol{V} = \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}\ . \tag{2.7}$$

For subsequent discussion, it will be useful to define the following quantities

$$\delta_{ij}=v_{ii}+v_{jj}-2v_{ij}\ , \qquad h_{ij}=\sigma_{ii}^{(0)}+\sigma_{jj}^{(0)}-2\sigma_{ij}^{(0)}\ ,$$
$$g_{ij}=\sigma_{ii}+\sigma_{jj}-2\sigma_{ij}\ , \qquad (\delta_{ij}+g_{ij}\equiv h_{ij})\ . \tag{2.8}$$

While the above distributional specification might represent a number of practical situations, the problem of choosing a realistic and mathematically simple utility function is usually much more difficult. In what follows we consider a variety of utility functions for ranking and estimation of $\boldsymbol{\theta}$.

3. Ranking and selection problems

Experiments are often conducted to make inferences about the ordering of a set of parameters $\boldsymbol{\theta}$. For example, k drugs may be compared and we may wish to order the effectiveness of the drugs. Sometimes, the primary concern is simply to determine which of the k drugs is most effective (best). Problems of this type are known as ranking and selection problems or best population problems. Beginning with the work of Bechhofer [1], a vast literature exists on ranking and selection procedures derived from sampling theory. Bayesian consideration of such problems has been given by Raiffa and Schlaifer [23], and Dunnett [9], following some earlier work by Grundy, Healy and Rees [13] in a fiducial framework. Their main concern was on the determination of optimal sample sizes rather on making a choice among available designs.

3.1. *Linear utility for selecting the best treatment*

Following these latter authors, we begin by considering the problem of selecting the best of k (treatments) parameters $\boldsymbol{\theta}$ in relation to the linear utility function $U(d, \boldsymbol{y}, a_i, \boldsymbol{\theta})=b_i+c_i\theta_i$, $(i=1,\cdots,k)$ where a_i is the act that treatment i is chosen. With no loss in generality, we may set $b_i=0$, $c_i=1$ so that

$$U(d, \boldsymbol{y}, a_i, \boldsymbol{\theta})=\theta_i\ , \qquad i=1,\cdots,k\ . \tag{3.1}$$

It follows that given $(d, \boldsymbol{y})$ the optimal act is $a^o=a_j$ when $\hat{\theta}_j$ is the largest posterior mean. Thus $\bar{U}(d, \boldsymbol{y})=\max(\hat{\boldsymbol{\theta}})$ so that

$$\bar{\bar{U}}(d)=\mathrm{E}\max(\hat{\boldsymbol{\theta}}) \tag{3.2}$$

and the choice of design reduces to the evaluation of $\mathrm{E}\max(\hat{\boldsymbol{\theta}})$ for each of the available designs and choosing the one which maximizes $\bar{\bar{U}}(d)$. We shall call this the *max mean* criterion. Using (2.6)–(2.8), it is shown in Appendix 1 that

$$\text{(3.3)} \qquad \mathrm{E}\max(\hat{\boldsymbol{\theta}}) = \mu_k + \sum_{i=1}^{k-1}(\mu_i - \mu_k)\Phi_{k-1}(\boldsymbol{\alpha}_i, \boldsymbol{R}_i) + \sum_{i<j}^{k}\delta_{ij}^{1/2}\phi_1(\alpha_{ij}, 1)\Phi_{k-2}(\boldsymbol{\alpha}_{ij}, \boldsymbol{R}_{ij})$$

where

$$\alpha_{ij} = (\mu_j - \mu_i)\delta_{ij}^{-1/2}; \qquad \boldsymbol{\alpha}_i = \{\alpha_{is}\}, \qquad \boldsymbol{R}_i = \{\rho_{st\cdot i}\},$$

$$\rho_{st\cdot i} = \frac{1}{2}(\delta_{is} + \delta_{it} - \delta_{st})(\delta_{is}\delta_{it})^{-1/2}, \qquad (s, t = 1, \cdots, k;\ \neq i);$$

$$\boldsymbol{\alpha}_{ij} = \{\alpha_{s\cdot ij}\}, \quad \boldsymbol{R}_{ij} = \{\rho_{st\cdot ij}\}, \qquad (s, t = 1, \cdots, k;\ \neq i \neq j),$$

$$\alpha_{s\cdot ij} = (\alpha_{is} - \rho_{sj\cdot i}\alpha_{ij})(1 - \rho_{sj\cdot i}^2)^{-1/2}$$

$$\rho_{st\cdot ij} = (\rho_{st\cdot i} - \rho_{sj\cdot i}\rho_{tj\cdot i})(1 - \rho_{sj\cdot i}^2)^{-1/2}(1 - \rho_{tj\cdot i}^2)^{-1/2};$$

$$\Phi_q(\boldsymbol{\alpha}, \boldsymbol{R}) = \int_{\alpha_1} \cdots \int_{\alpha_q} \phi_q(\boldsymbol{Z}, \boldsymbol{R})d\boldsymbol{Z}, \qquad \Phi_0(\boldsymbol{\alpha}, \boldsymbol{R}) = 1$$

and $\phi_q(\boldsymbol{Z}, \boldsymbol{R})$ is the joint density of q normal variables with zero means and correlation matrix $\boldsymbol{R}$. Note that except for μ_k, $\mathrm{E}\max(\hat{\boldsymbol{\theta}})$ is a function of the $k-1$ contrasts $(\mu_i - \mu_k)$ and the $\binom{k}{2}$ quantities δ_{ij} which are the variances of the differences $(\hat{\theta}_i - \hat{\theta}_j)$. Since $\boldsymbol{\mu}$ are the prior means, the influence of a design is completely determined through δ_{ij}.

For $k=2$, Raiffa and Schlaifer [23] show that $\mathrm{E}\max(\hat{\boldsymbol{\theta}})$ is monotonically increasing in δ_{12} so that, the best design is the one which maximizes δ_{12}. This is a very reasonable result because we see from (2.8) that δ_{12} is the difference of the prior and posterior variances of $\theta_1 - \theta_2$. Thus, the optimal design is the one which minimizes the posterior variance g_{12}.

For $k>2$, determination of $\mathrm{E}\max(\hat{\boldsymbol{\theta}})$ would necessitate evaluation of normal integrals of $(k-1)$ and $(k-2)$ dimensions. When $k=3$, this can be done using available univariate and bivariate tables. For $t=4$, Steck [24] gives a table from which $\Phi_3(\boldsymbol{\alpha}, \boldsymbol{R})$ can be evaluated. The general case can be handled through reduction formulae given by Plackett [22] and John [16]. Various simplifications occurs for special forms of $\boldsymbol{\mu}$ and $\boldsymbol{V}$, Ihm [15] and Curnow and Dunnett [5].

An interesting special case is when $\boldsymbol{\mu} = \mu\mathbf{1}$. This may for example arise when the elements of $\boldsymbol{\theta}$ are regarded as a random sample from a normal population with mean μ. In this case, (3.3) reduces to

$$\text{(3.4)} \qquad \mathrm{E}\max(\boldsymbol{\theta}) = \mu + (\sqrt{2\pi})^{-1}\sum_{i<j}^{k}\delta_{ij}^{1/2}\Phi_{k-2}(\mathbf{0}, \boldsymbol{R}_{ij}).$$

Exact evaluation can thus be done for $k \leqq 5$ using the formulae

$$\Phi_1(0, 1)=\frac{1}{2}, \qquad \Phi_2(\mathbf{0}, \boldsymbol{R})=\frac{1}{4}+(2\pi)^{-1}\sin^{-1} r_{12}$$

(3.5)

$$\Phi_3(\mathbf{0}, \boldsymbol{R})=(4\pi)^{-1}[2\pi-\cos^{-1} r_{12}-\cos^{-1} r_{13}-\cos^{-1} r_{23}]$$

(Moran [21]).

When $k=6$, the formulae given by Child [4] reducing $\Phi_4(\mathbf{0}, \boldsymbol{R})$ to a univariate integral can be used. For larger values of k, numerical evaluation of normal integrals seems unavoidable.

3.2. *Probability of correction selection*

The utility function (3.1) implies that, given $(d, \boldsymbol{y})$, one should select treatment j when $\hat{\theta}_j=\max(\hat{\boldsymbol{\theta}})$. A relevant feature of the posterior distribution considered by Dunnett [9] is the posterior probability of correct selection

$$\bar{U}(d, \boldsymbol{y})=\Pr(\theta_j>\theta_i \mid \hat{\theta}_j>\hat{\theta}_i;\ i=1, \cdots, k,\ \neq j). \tag{3.6}$$

It follows that

$$\bar{\bar{U}}(d)=\sum_{j=1}^{k}\Pr(\theta_j>\theta_i, \hat{\theta}_j>\hat{\theta}_i;\ i=1, \cdots, k,\ \neq j) \tag{3.7}$$

and we should thus select the design which maximizes the overall probability of correct selection. This shall be called the *max prob.* criterion. From (2.6)–(2.8),

$$\bar{\bar{U}}(d)=(2\pi)^{-(k-1)}\sum_{j=1}^{k}|\boldsymbol{B}_j\boldsymbol{C}_j|^{-1/2} \cdot \int_{\boldsymbol{\eta}_j}\int_{\boldsymbol{\eta}_j-\boldsymbol{u}} \exp\left(-\frac{1}{2}[\boldsymbol{u}'\boldsymbol{B}_j^{-1}\boldsymbol{u}+\boldsymbol{w}'\boldsymbol{C}_j^{-1}\boldsymbol{w}]\right)d\boldsymbol{w}d\boldsymbol{u} \tag{3.8}$$

where $\boldsymbol{u}'=(u_1, \cdots, u_{k-1})$, $\boldsymbol{w}'=(w_1, \cdots, w_{k-1})$; $\boldsymbol{\eta}_j=\{\eta_{js}\}$, $\eta_{js}=\mu_s-\mu_j$, $\boldsymbol{B}_j=\{b_{st\cdot j}\}$, $b_{st\cdot j}=(1/2)(\delta_{sj}+\delta_{tj}-\delta_{st})$, $\boldsymbol{C}_j=\{c_{st\cdot j}\}$, $c_{st\cdot j}=(1/2)(g_{sj}+g_{tj}-g_{st})$, $(s, t=1, \cdots, k,\ \neq j)$. Similar to $\mathrm{E}\max(\hat{\boldsymbol{\theta}})$ in (3.3), $\bar{\bar{U}}(d)$ in (3.8) is a function of the $(k-1)$ contrasts $\mu_i-\mu_j$ and the $\binom{k}{2}$ variances δ_{ij}.

For $k=2$, it is readily verified by differentiation that $\bar{\bar{U}}(d)$ is monotonically increasing in δ_{12}. Thus, a design which maximizes $\mathrm{E}\max(\hat{\boldsymbol{\theta}})$ also maximizes the probability of correct selection, i.e. the *max mean* and the *max prob* criteria are equivalent. For $k>2$, it will be shown later in Section 5 that the two criteria are also equivalent in the case of balanced designs.

In the above we have introduced the utility (3.6) as a feature of the posterior distribution of $\boldsymbol{\theta}$. In some situations, it may alternatively be justified as the expected utility of a decision problem. In partic-

ular, suppose we have the utility function

$$U(d, \boldsymbol{y}, a_j, \boldsymbol{\theta}) = \begin{cases} 1 & \theta_j = \max(\boldsymbol{\theta}) \\ 0 & \text{otherwise} . \end{cases} \tag{3.9}$$

Then

$$\bar{U}(d, \boldsymbol{y}) = \max_j \Pr(\theta_j > \theta_i, \text{ all } i \neq j \mid \hat{\boldsymbol{\theta}}) . \tag{3.10}$$

In the case $k=2$, we find

$$\max_j \Pr(\theta_j > \theta_i \mid \hat{\boldsymbol{\theta}}) = \Pr(\theta_j > \theta_i \mid \hat{\theta}_j > \hat{\theta}_i) . \tag{3.11}$$

For $k>2$, equality of (3.6) and (3.10) can be established when the posterior covariance matrix $\boldsymbol{\Sigma}$ takes the form $\boldsymbol{\Sigma} = \sigma^2[(1-\rho)I + \rho \mathbf{1}_k \mathbf{1}_k']$ where $\mathbf{1}_k$ is a $k \times 1$ vector of ones. For, in this case, the probability $\Pr(\theta_j > \theta_i, \text{ all } i \neq j \mid \hat{\boldsymbol{\theta}})$ takes the form

$$\Pr \propto f(\eta_1, \cdots, \eta_j, \cdots, \eta_k) = \int_{-\boldsymbol{\eta}} \exp\left(-\frac{1}{2} \boldsymbol{x}' \boldsymbol{R}^{-1} \boldsymbol{x}\right) d\boldsymbol{x} \tag{3.12}$$

where $\boldsymbol{x}' = (x_1, \cdots, x_{k-1})$, $\boldsymbol{\eta}' = (\eta_1, \cdots, \eta_k)$ and $\boldsymbol{R} = (1/2)(\boldsymbol{I} + \mathbf{1}_{k-1} \mathbf{1}_{k-1}')$. The desired result follows by noting that f is a symmetric function and increasing in η_j.

3.3. *Some general ranking criteria*

The utility criteria in (3.1), (3.6) and (3.9) can be readily extended to the problem of selecting two or more out of a set of k treatments. In some situations, we may be interested in a complete ranking of all the treatments. Thus, analogous to (3.6), the utility criterion may be the posterior probability of correct ranking

$$\bar{U}(d, \boldsymbol{y}) = \Pr(\boldsymbol{\theta}_{(l)} \mid \hat{\boldsymbol{\theta}}_{(l)}) ,$$

where $\boldsymbol{x}_{(l)}$ denotes a particular ordering of the k elements of $\boldsymbol{x}' = (x_1, \cdots, x_k)$, $l = 1, \cdots, k!$, so that

$$\bar{\bar{U}}(d) = \sum_{l=1}^{k!} \Pr(\boldsymbol{\theta}_{(l)}, \hat{\boldsymbol{\theta}}_{(l)}) . \tag{3.13}$$

Like the results in (3.3) and (3.8), the expression (3.13) and other similar extensions suffer in general the common practical difficulty of having to evaluate complicated normal integrals in high dimensions. It is thus desirable to seek utility criteria which are more amenable to practical computation.

Both in (3.3) and (3.8), the utilities $\bar{\bar{U}}(d)$ are functions of the contrasts $\mu_i - \mu_j$ and the variances δ_{ij}. If we wish to rank the parameters

$\boldsymbol{\theta}$, it is natural to choose a design which, in some sence, will make the parameters $\boldsymbol{\theta}$ spread out as much as possible *a posteriori*. A reasonable utility criterion may then be the measure of separation

$$\bar{U}(d, \boldsymbol{y})=(\hat{\boldsymbol{\theta}}-\bar{\hat{\theta}}\mathbf{1}_k)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\bar{\hat{\theta}}\mathbf{1}_k)\ , \tag{3.14}$$

where

$$\bar{\hat{\theta}}=(\mathbf{1}_k'\boldsymbol{\Sigma}^{-1}\mathbf{1}_k)^{-1}\mathbf{1}_k'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\theta}}\ ,$$

which, by analogy to Mahalanobis distance measure, may be regarded as the Mahalanobis distance of $\hat{\boldsymbol{\theta}}$ from the central value $\bar{\hat{\theta}}$. Thus,

$$\bar{\bar{U}}(d)=(\boldsymbol{\mu}-\bar{\mu}\mathbf{1}_k)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\bar{\mu}\mathbf{1}_k)+\mathrm{tr}\ \boldsymbol{V}\boldsymbol{\Sigma}^{-1}[I-\mathbf{1}_k(\mathbf{1}_k'\boldsymbol{\Sigma}^{-1}\mathbf{1}_k)^{-1}\mathbf{1}_k'\boldsymbol{\Sigma}^{-1}] \tag{3.15}$$

where $\bar{\mu}=(\mathbf{1}_k'\boldsymbol{\Sigma}^{-1}\mathbf{1}_k)^{-1}\mathbf{1}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, which we shall call the *separation* criterion. It can be shown that $\bar{\bar{U}}(d)$ is again a function of $\mu_i-\mu_j$ and δ_{ij}. The chief advantage here is of course that $\bar{\bar{U}}(d)$ can be conveniently calculated for each available design, involving no more than matrix multiplication and inversion. A notable characteristic of the quadratic utility in (3.14) is that it tends to be dominated by extreme values of $\hat{\theta}_i$. This would be desirable if our main interest were to find out the 'best' and the 'worst' treatments. On the other hand, if we are interested in the ordering of the complete set of parameters, one should employ a criterion which tends to damp out the effect of extreme differences. A utility function possessing such property is

$$\bar{U}(d, \boldsymbol{y})=-\sum_{i<j} q_{ij}\ , \qquad q_{ij}=\exp\left(-\frac{\alpha}{2}(\hat{\theta}_i-\hat{\theta}_j)^2 g_{ij}^{-1}\right) \tag{3.16}$$

where $0<\alpha\leqq 1$. In this case

$$\bar{\bar{U}}(d)=-\sum_{i<j} \mathrm{E}\,(q_{ij})\ , \tag{3.17}$$

where

$$\mathrm{E}\,(q_{ij})=g_{ij}^{1/2}[\alpha h_{ij}+(1-\alpha)g_{ij}]^{-1/2}\exp\left(-\alpha(\mu_i-\mu_j)^2/[\alpha h_{ij}+(1-\alpha)g_{ij}]\right)\ ,$$

which shall be called the *ranking* criterion.

This is an intuitively pleasing criterion. Firstly, $\mathrm{E}\,(q_{ij})$ is increasing in the posterior and prior variance ratio $x_{ij}=g_{ij}/h_{ij}$ so that one should seek a design which makes all the g_{ij}, the posterior variances of the differences, as small as possible. This in turn increases the power to discriminate among the $\boldsymbol{\theta}$ *a posteriori* and hence sharpens the inference that can be made about the ordering of the parameters. Secondly, for sufficiently large value of $(\mu_i-\mu_j)^2/h_{ij}$ the derivative $\partial\,\mathrm{E}\,(q_{ij})/\partial x_{ij}$ is decreasing in $(\mu_i-\mu_j)^2/h_{ij}$ (it is monotonically decreasing if $\alpha=1$). It

follows that greater effort should be devoted to reduce the variances g_{ij} corresponding to those parameters which are closer together and therefore more difficult to order *a priori*. An illustrative example will be given later in Section 5.

4. Estimation

One other important objective of experimentation is to make general inference about the unknown parameters $\boldsymbol{\theta}$. In the traditional sampling theory approaches, the choice of design has usually been made on the basis of some functions (trace, determinant, etc.) of the covariance matrix of the estimates of linear forms of $\boldsymbol{\theta}$, Kiefer [18], [19], Elving [10]. Under the normality assumption, these functions can be viewed as referring to different aspects (lengths of principal axes, volume, etc.) of the confidence hyperellipsoid for the linear forms.

In the Bayesian framework, all inferences about $\boldsymbol{\theta}$ are based on the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}, d)$. One should therefore choose a design to make the distribution, in some sense, as sharp as possible. For $k=1$, this is achieved unambiguously by minimizing the posterior variance σ_{11}. When $k>1$, the ellipsoidal H.P.D. regions of the distribution (Box and Tiao, [3]) are determined by the covariance matrix $\boldsymbol{\Sigma}$ which contains $(1/2)k(k+1)$ elements, and the question then boils down to what function or functions of these elements ought to be considered. In general, we may define two set of k and $(1/2)k(k-1)$ functions, the first are functions of the characteristic roots $(\lambda_1, \cdots, \lambda_k)$ of $\boldsymbol{\Sigma}$ which jointly determine the 'size' and 'shape' of the ellipsoidal regions while the second set are functions of the characteristic vectors which decide the orientation. Even if we leave aside the question of orientation, choice must still be made about functions of the roots. Obvious possibilities are

$$\bar{\bar{U}}(d) = \bar{U}(d, \boldsymbol{y}) = -\sum \lambda_i \qquad (\textit{trace} \text{ criterion}) \tag{4.1}$$

and

$$\bar{\bar{U}}(d) = \bar{U}(d, \boldsymbol{y}) = -\prod \lambda_i \tag{4.2}$$

(*determinant* criterion; Box and Hunter [2])

which correspond to well known criteria in the sampling framework. Such choices can alternatively be viewed from a Bayesian decision theory approach to point estimation. Suppose one wishes to obtain a best point estimate of $\boldsymbol{\theta}$ relative to some utility criterion $U(d, \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{\theta})$. In the terminology of Raiffa and Schlaifer [23], here the action space of $\boldsymbol{a}$ coincides with the parameter space of $\boldsymbol{\theta}$. In this context, the trace

criterion then corresponds to the quadratic utility (or loss)

$$U(d, \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{\theta}) = -(\boldsymbol{a}-\boldsymbol{\theta})'(\boldsymbol{a}-\boldsymbol{\theta}) \tag{4.3}$$

for in this case $\boldsymbol{a}^o(d, \boldsymbol{y}) = \hat{\boldsymbol{\theta}}$ and $\bar{U}(d, \boldsymbol{y}) = -\sum \lambda_i$. On the other hand, the determinant criterion can be obtained from the zero-constant utility

$$U(d, \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{\theta}) = \begin{cases} 0 & |\boldsymbol{a}-\boldsymbol{\theta}| < c \\ -1 & |\boldsymbol{a}-\boldsymbol{\theta}| > c \end{cases} \tag{4.4}$$

where c is an arbitrarily small positive constant. This is because

$$\underset{\boldsymbol{\theta}|\boldsymbol{y}}{\mathrm{E}}\, U = -[1-\Pr\{|\boldsymbol{a}-\boldsymbol{\theta}| < c | d, \boldsymbol{y}\}] . \tag{4.5}$$

Clearly, $a^o(d, \boldsymbol{y}) = \hat{\boldsymbol{\theta}}$ and hence

$$\bar{U}(d, \boldsymbol{y}) = -1 + |\boldsymbol{\Sigma}|^{-1/2} \frac{\Delta\boldsymbol{\theta}}{(2\pi)^{k/2}} \tag{4.6}$$

where $\Delta\boldsymbol{\theta}$ is an arbitrarily small volume element of $\boldsymbol{\theta}$. Maximization of $\bar{\bar{U}}(d) = \bar{U}(d, \boldsymbol{y})$ in (4.6) is thus equivalent to minimization of the determinant $|\boldsymbol{\Sigma}| = \prod \lambda_i$ given by the criterion in (4.2).

4.1. *The exponential utility function*

The trace criterion has been criticized on the ground that it is not invariant under general linear transformation of $\boldsymbol{\theta}$, see e.g. Kiefer [19]. In practice, however, it is often true that some linear forms of $\boldsymbol{\theta}$ are more meaningful and important to the investigator than others, and so it is not always relevant to insist on invariance as a principle of choice. In the context of point estimation, another criticism of the trace criterion is that the quadratic utility in (4.3) implies too severe a penalty for large deviation from $\boldsymbol{\theta}$. It is difficult to imagine, e.g. in the single parameter case that one should incur a loss when $|a-\theta| = 100\sigma$ which is 100 times greater than the loss when $|a-\theta| = 10\sigma$. Surely it seems more reasonable to have a utility function which damps off for large deviations. On the other hand, the utility function in (4.4) decreases too steeply for small deviations from $\boldsymbol{\theta}$. Also, the resultant determinant criterion, while enjoying the property of invariance, refers only to the volume of the ellipsoidal posterior region of $\boldsymbol{\theta}$, but not to the attenuation of the region.

The above considerations has led us to consider the exponential utility

$$U(d, \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{\theta}) = 1 - \exp\left(-\frac{\alpha}{2}(\boldsymbol{a}-\boldsymbol{\theta})'(\boldsymbol{a}-\boldsymbol{\theta})\right), \qquad (\alpha > 0) . \tag{4.7}$$

The function approximates the quadratic utility in (4.3) for small α and the zero-constant utility in (4.4) for large α and therefore can be regarded as a compromise between the two. In this case, $a^o(d, \boldsymbol{y})=\hat{\boldsymbol{\theta}}$ and

$$\bar{\bar{U}}(d)=\bar{U}(d, \boldsymbol{y})=1-|\boldsymbol{I}+\alpha\boldsymbol{\Sigma}|^{-1/2}. \tag{4.8}$$

This is then equivalent to choosing a design d to minimize the determinant

$$|\boldsymbol{I}+\alpha\boldsymbol{\Sigma}|=\prod_{i=1}^{k}(1+\alpha\lambda_i) \tag{4.9}$$

which shall be called the *modified determinant* criterion.

To see the implications of this criterion, it is useful to write

$$\prod_{i=1}^{k}(1+\alpha\lambda_i)=1+\sum_{j=1}^{k}\binom{k}{j}\alpha^j P_j \tag{4.10}$$

where $P_1=\frac{1}{k}\sum_i \lambda_i$, $P_2=\binom{k}{2}^{-1}\sum_{i<j}\lambda_i\lambda_j, \cdots, P_k=\prod_{i=1}^{k}\lambda_i$. The quantities $(P_1, \cdots, P_k)$ are k symmetric functions of the roots $(\lambda_1, \cdots, \lambda_k)$ describing different aspects of the ellipsoidal region. Specifically, when multiplied by appropriate constants, P_1 is the average of the squared lengths of the principal axes, P_2 is the average of the squared areas of the ellipses spanned by the axes taken two at a time and in general P_s is the average of the squared 'volumes' of the hyperellipsoids spanned by the axes taken s at a time, $s=3, \cdots, k$. Rather than concentrating on a specific aspect (the trace P_1 or the determinant P_k), the criterion in (4.8) attempts to take all the k aspects into account.

As an alternative interpretation, we can write

$$\prod_{i=1}^{k}(1+\alpha\lambda_i)=1+\sum_{j=1}^{k-1} m_j\binom{k}{j}(\alpha\dot{P}_k)^j+\alpha^k P_k \tag{4.11}$$

where $\dot{P}_k=P_k^{1/k}$ is the geometric mean of the λ's, and $m_j=P_j\dot{P}_k^{-j}$. Since $P_1\geqq P_2^{1/2}\geqq P_3^{1/3}\geqq\cdots\geqq P_k^{1/k}$ (Hardy, Littlewood and Pólya [14], p. 51), it follows that $m_j\geqq 1$ where the equality holds if and only if all the λ's are equal. Thus, we may regard the $k-1$ functions $(m_1, \cdots, m_{k-1})$ as measures of 'shape' or 'conditioning' of the ellipsoidal region, in contrast to P_k measuring the size of the region. The criterion in (4.8) aims, therefore, at simultaneously minimizing the size as well as reducing the attenuation of the region. We note that a commonly used measure of ill-conditioning of a matrix proposed by Turing [25], $T=(1/k)[(\sum\lambda_i)(\sum\lambda_i^{-1})]^{1/2}$, is in fact the product $T=m_1 m_{k-1}$.

4.2. *Quadratic regression*

As an illustration, consider the three level quadratic regression model

(4.12) $$y=\theta_1+\theta_2 x+\theta_3 x^2+e$$

where x takes the values $(-1, 0, 1)$. Suppose a total of n observations are to be taken and the design problem is to allocate the n experimental runs to the three levels. We shall discuss the situation where $\boldsymbol{\Omega}=\boldsymbol{I}$ and the prior distribution of $\boldsymbol{\theta}$ is diffuse, i.e. $\boldsymbol{\Sigma}_0=\sigma_0^2\boldsymbol{C}$ where $\boldsymbol{C}$ is a 3×3 positive definite symmetric matrix and $\sigma_0^2\to\infty$. In this case, the posterior covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{\theta}$ is such that

(4.13) $$\boldsymbol{\Sigma}^{-1}\to\boldsymbol{\Sigma}_s^{-1}\boldsymbol{X}\boldsymbol{N}\boldsymbol{X}'$$

where the rows of $\boldsymbol{X}$ are $(1, 1, 1)$, $(-1, 0, 1)$ and $(1, 0, 1)$ and $\boldsymbol{N}$ is a diagonal matrix with elements (n_1, n_2, n_3) corresponding to the three levels $(-1, 0, 1)$. It follows that

(4.14) $$P_1=\frac{1}{6}\left(\frac{1}{n_1}+\frac{4}{n_2}+\frac{1}{n_3}\right), \qquad P_2=\frac{1}{4}\left(\frac{1}{n_1 n_2}+\frac{1}{n_2 n_3}+\frac{1}{3n_1 n_3}\right),$$

$$P_3=\frac{1}{4n_1 n_2 n_3}.$$

As is well known (e.g. Kendall and Stuart [17], pp. 158–161), the determinant criterion leads to the balanced allocation $n_1=n_2=n_3=n/3$. On the other hand, the optimal allocation corresponds to the trace criterion is $n_1=n_3=n/4$ and $n_2=n/2$. With respect to the modified determinant criterion in (4.9), it can be readily verified that, for a given α, the optimal allocation is

(4.15) $$n_1=n_3=(n-n_2)/2, \qquad \frac{n_2}{n}=g\left[1-\left(1-\frac{2}{3}g^{-1}\right)^{1/2}\right]$$

where $g=.375(2+\alpha/n)$ and $1/2\leqq n_2/n\leqq 1/3$. The ratio n_2/n approaches $1/3$ as $\alpha\to\infty$ and $1/2$ as $\alpha\to 0$. Table 4.1 shows the values of n_2/n, the trace, determinant, m_1 and m_2 for various values of α/n. It is seen that if we choose α/n to be around 1, a considerable reduction in the relative values of the trace, m_1 and m_2 is achieved at the expense of a moderate increase of the determinant from its minimum value of 6.75 when $\alpha\to\infty$.

Table 4.1. Values of $(n_2/n, \sum\lambda_i, \prod\lambda_i, m_1, m_2)$ for various choices of α

α/n	$n_2/2$	$\sum\lambda_i$	$\prod\lambda_i$	m_1	m_2
0	.5000	8.0000	8.0000	1.3333	1.3333
.1	.4790	8.0128	7.6923	1.3535	1.3438
.5	.4345	8.1400	7.1994	1.4059	1.3717
1.0	.4069	8.2884	6.9881	1.4450	1.3931
1.5	.3919	8.3928	6.9013	1.4695	1.4064
2.0	.3821	8.4710	6.8540	1.4867	1.4158
∞	.3333	9.0000	6.7500	1.5874	1.4699

5. Balanced designs

In the preceding two sections, we have discussed a number of utility criteria appropriate for various ranking and estimation problems. We now consider a special class of designs for which the matrix $\boldsymbol{\Sigma}_s$ in (2.4) takes the form

$$\boldsymbol{\Sigma}_s=\gamma \boldsymbol{I}+\xi \mathbf{1}_k\mathbf{1}_k', \qquad (\gamma>0,\ \xi>-\gamma k^{-1}). \tag{5.1}$$

We shall call these *balanced* designs and index them by (γ, ξ). They include the usual completely randomized design (CR), randomized complete block (RCB) and balanced incomplete block (BIB) designs. First, we give a general result concerning the optimality of such designs.

LEMMA 5.1. *For the class of balanced design defined in* (5.1), *if* $U(d, \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{\theta})=U(\boldsymbol{a}, \boldsymbol{\theta})$ *and, for every* γ, $\bar{\bar{U}}(d)=f(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0, \gamma)$ *independent of* ξ, *then* $\bar{\bar{U}}(d)$ *is decreasing in* γ.

PROOF. Let (d, d_1) be any two designs indexed respectively by (γ, ξ) and (γ_1, ξ_1) such that $\gamma \leqq \gamma_1$, and with corresponding observations $\boldsymbol{y}$ and $\boldsymbol{y}_1$. We need to prove that $\bar{\bar{U}}(d) \geqq \bar{\bar{U}}(d_1)$. Now there exists $\gamma_2>0$ such that

$$\gamma=\gamma_1\gamma_2(\gamma_1+\gamma_2)^{-1}.$$

Let d_2 be the design with index (γ_2, ξ_2) and observation $\boldsymbol{y}_2$. Then from (2.4), the augmented design $d^*=(d_1, d_2)$ with observation $(\boldsymbol{y}_1, \boldsymbol{y}_2)$ has index (γ, ξ^*). But by hypothesis of the lemma,

$$\bar{\bar{U}}(d)=\bar{\bar{U}}(d^*).$$

Hence,

$$\begin{aligned}\bar{\bar{U}}(d)&=\mathrm{E}_{\boldsymbol{y}_1}\mathrm{E}_{\boldsymbol{y}_2|\boldsymbol{y}_1}\max_{\boldsymbol{a}}\mathrm{E}_{\boldsymbol{\theta}|(\boldsymbol{y}_1,\boldsymbol{y}_2)}U(\boldsymbol{a}, \boldsymbol{\theta})\\ &\geqq \mathrm{E}_{\boldsymbol{y}_1}\max_{\boldsymbol{a}}\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{y}_1}U(\boldsymbol{a}, \boldsymbol{\theta})=U(d_1)\end{aligned}$$

as required.

For the class of balanced designs, when the prior covariance matrix $\boldsymbol{\Sigma}_0$ is also of the form (5.1), it is readily shown that the lemma is immediately applicable to the max mean (3.2), and the max prob (3.7) (through the use of (3.9)) criteria for selection, and to the trace (4.1), determinant (4.2) and modified determinant (4.8) criteria when we are interested in linear contrasts of $\boldsymbol{\theta}$. In addition, it can be verified directly that for the separation (3.14) and ranking (3.17) criteria, $\bar{\bar{U}}(d)$ is decreasing in γ. Thus, we have proved the following theorem.

THEOREM 5.1. *If* $\Sigma_0=\gamma_0 I+\xi_0 1_k 1_k'$, *then for the class of balanced designs* (5.1), *the one for which* γ *is the smallest is optimal with respect to the max mean, max prob, separation and ranking criteria, and to the trace, determinant and modified determinant criteria for estimating linear contrasts of* θ.

In practice, experiments are frequently run in blocks and we are often faced with the problem of making a choice among different feasible blocking arrangements. In general, suppose we are interested in k treatment parameters θ and the model in (2.3) takes the form

$$y=\underset{n\times k}{A}\ \underset{k\times 1}{\theta}+\underset{n\times b}{Z}\ \underset{b\times 1}{\beta}+\underset{n\times 1}{\varepsilon} \tag{5.2}$$

where the error vector e is split into two components, the block effects $Z\beta$ and the within block errors ε. Each row of A and Z contains a single unity as its only nonzero element. The block variables β are assumed to be randomly drawn from a normal population such that $\beta\sim N(0, \sigma_b I)$, and the errors ε are distributed as $N(0, \sigma_\varepsilon^2 I)$. For the moment we shall consider the case of equal replicate and equal block size binary designs, i.e. $A'A=rI$ and $Z'Z=lI$ where r is the number of times each treatment occurs and l the block size. Then, the class of RCB and BIB designs is such that

$$A'ZZ'A=(r-\lambda)I+\lambda 1_k 1_k' \tag{5.3}$$

where $r\geqq\lambda$ and λ is the number of times two treatments appear in the same block. When $r=\lambda$, the design is an RCB and when $\sigma_b^2=0$ the model (5.3) degenerates to a CR. For this class of designs, we have that

$$\gamma=\frac{\sigma_\varepsilon^2}{r}[f+\phi(1-f)]^{-1}, \tag{5.4}$$

with

$$\phi=\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+l\sigma_b^2} \quad \text{and} \quad f=\frac{\lambda k}{rl}, \tag{5.5}$$

where f, the well known efficiency factor in the traditional comparison between BIB and RCB designs, equals 1 for an RCB and ($\phi=1$, $f=1$) for a CR. The optimal design in the sense of Theorem 5.1 can be readily determined. The result (5.4) is, in a sense, to be expected since γ is simply the variance of the best linear combination of the intra and inter block estimators of the contrasts $\theta_i-\bar{\theta}$. The interesting and important point here is the invariance of the criterion γ to the purpose of the experiment and the utility functions considered. We

note that the equivalence of the trace and determinant criteria with respect to intra block estimates of orthogonal contrasts was established by Kiefer [19].

5.1. *Sensitivity of optimality of balanced designs*

In the above, we have restricted considerations to the class of balanced designs. When the prior means are unequal, one would expect that for ranking and selection purposes a balanced arrangement might be inferior to some unbalanced ones even if the prior covariance matrix Σ_0 takes the balanced form as in Theorem 5.1. As an illustration, consider the simple problem of allocating n observations to three treatments following the model in (5.2) with $\sigma_b^2=0$ and the prior distribution of θ is normal with $\mu=(\mu_1, \mu_2, \mu_3)$ and $\Sigma_0=\sigma_0^2 I$. Let n_i be the number of observation for the ith treatment, $i=1, 2, 3$, $(\sum n_i=n)$. Then, for the elements $(h_{ij}, v_{ij}, \delta_{ij})$ in (2.8),

$$(5.6)\qquad h_{ij}=2\sigma_0^2\,, \qquad v_{ii}=\sigma_0^2(1+wn_i^{-1})^{-1}\,, \qquad v_{ij}=0\,, \qquad \delta_{ij}=v_{ii}+v_{jj}$$

where $w=\sigma_i^2/\sigma_0^2$.

Consider first the problem of selecting the largest θ_i with respect to the utility in (3.1). To obtain the optimal allocation, we may substitute (5.6) into (3.3) and differentiate $\mathrm{E}\max(\hat{\theta})$ with respect to n_i under the constraint $\sum n_i=n$. The resulting two equations are

$$(5.7)\qquad \left(\frac{n_2+w}{n_1+w}\right)^2=\frac{a_{12}+a_{23}}{a_{12}+a_{13}}\,, \qquad \left(\frac{n_3+w}{n_2+w}\right)^2=\frac{a_{23}+a_{13}}{a_{23}+a_{12}}$$

where

$$a_{ij}=\delta_{ij}^{-1/2}\phi(\alpha_{ij}, 1)\Phi_1(\alpha_{s\cdot ij}, 1) \qquad s\neq i\neq j\,.$$

It can be readily shown that equality of the prior means, $\mu_i=\mu_j$, is a sufficient and necessary condition for $n_i=n_j$ to satisfy (5.7). In addition, if $\mu_1<\mu_2<\mu_3$, then for $n_1=n_2$, $a_{23}>a_{13}$ and for $n_2=n_3$, $a_{13}>a_{12}$ so that the solution must satisfies $n_1\leqq n_2\leqq n_3$. Thus, when the prior means are unequal, the balanced allocation is not necessarily optimal. It should be possible to extend these results to any number of treatments. To illustrate the situation, Table 5.1 shows the values of $\mathrm{E}\max(\hat{\theta})$ of various allocations for a combination of values of μ_3/σ_0 and w when $\mu_1=\mu_2=0$ and $n=30$. Although the solution is such that $n_1=n_2>n_3$ for $\mu_3/\sigma_0=-2$ and $n_1=n_2<n_3$ for $\mu_3/\sigma_0=2$, the differences in the utilities of different allocations are extremely slight. Nevertheless, it is still discernable that the larger the value of w (the stronger the prior information compared with that from the sample), the more sensitive is the utility to different allocations.

Table 5.1. Values of $E \max(\hat{\theta})$ for various choices of $(n_i, \mu_3/\sigma_0, w)$ with $\mu_1=\mu_2=0$

$n_1=n_2$	n_3	$w=.5$	1	2	
10	10	.5603	.5463	.5216	$\frac{\mu_3}{\sigma_0}=-2$
12	6	.5617	.5495	.5275	
13	4	.5620	.5500	.5308	
10	10	2.0812	2.0737	2.0609	$\frac{\mu_3}{\sigma_0}=2$
8	14	2.0812	2.0738	2.0615	
6	18	2.0799	2.0718	2.0589	

Consider now the allocation problem with respect to the ranking criterion (3.17) with $\alpha=1$. In this case the equations of the first derivatives are

$$\left(\frac{n_2+w}{n_1+w}\right)^2=\frac{b_{12}+b_{23}}{b_{12}+b_{13}}, \qquad \left(\frac{n_3+w}{n_2+w}\right)^2=\frac{b_{23}+b_{13}}{b_{23}+b_{12}} \tag{5.8}$$

where

$$b_{ij}=\left(\frac{g_{ij}}{h_{ij}}\right)^{1/2}\exp\left(-\frac{1}{2}\left(\frac{\mu_i-\mu_j}{\sigma_0}\right)^2\right).$$

Again, it readily follows that $n_i=n_j$ if and only if $\mu_i=\mu_j$. However, unlike the previous case, if $\mu_1<\mu_2<\mu_3$, then the solution must satisfies $n_2\geqq\max(n_1, n_3)$ and if further $\mu_3-\mu_2>\mu_2-\mu_1$, then $n_1\geqq n_3$. To illustrate the sensitivity of the utility to different allocations, Table 5.2 shows the values of $\bar{\bar{U}}(d)$ of various allocations for selected values of μ_3/σ_0 and w when $\mu_1=\mu_2=0$ and $n=30$. As expected, the stronger the prior information, the more sensitive is the utility to different allocations.

Table 5.2. Values of $-\bar{\bar{U}}(d)$ in (3.17) for various choices of $(n_i, \mu_3/\sigma_0, w)$ with $\mu_1=\mu_2=0$

$n_1=n_2$	n_3	$w=.5$	1	2	
10	10	.4829	.6672	.9033	$\frac{\mu_3}{\sigma_0}=1$
11	8	.4829	.6668	.9024	
12	6	.4933	.6794	.9189	
13	4	.5226	.7138	.9511	
10	10	.2772	.3831	.5186	$\frac{\mu_3}{\sigma_0}=2$
12	6	.2654	.3670	.4977	
13	4	.2661	.3668	.4958	

Although too much should not be read into one or two examples, the results in Tables 5.1 and 5.2 do lend considerable support to the use of balanced designs in the common situation where the prior information is weak compared with that to be expected from the experiment.

Appendix 1. Derivation of $\mathrm{E}\max(\hat{\boldsymbol{\theta}})$

We now provide a brief sketch of the derivation of the $\mathrm{E}\max(\hat{\boldsymbol{\theta}})$ in (3.3). From (2.6), $\hat{\boldsymbol{\theta}}$ is distributed as normal $N(\boldsymbol{\mu}, \boldsymbol{V})$. Thus

$$\text{(A.1)} \qquad \begin{aligned}\mathrm{E}\max(\hat{\boldsymbol{\theta}}) &= \sum_{i=1}^{k} \int_{D_i} \hat{\theta}_i p(\hat{\boldsymbol{\theta}}) d\hat{\boldsymbol{\theta}} \\ &= \sum_{i=1}^{k} \mu_i \int_{D_i} p(\hat{\boldsymbol{\theta}}) d\hat{\boldsymbol{\theta}} + \sum_{i=1}^{k} \int_{D_i} (\hat{\theta}_i - \mu_i) p(\hat{\boldsymbol{\theta}}) d\hat{\boldsymbol{\theta}}\end{aligned}$$

where D_i is the region $\hat{\theta}_i = \max(\hat{\boldsymbol{\theta}})$ and $p(\hat{\boldsymbol{\theta}})$ is the density of $\hat{\boldsymbol{\theta}}$. Since $\sum_{i=1}^{k} \int_{D_i} p(\hat{\boldsymbol{\theta}}) d\hat{\boldsymbol{\theta}} = 1$ it follows that

$$\text{(A.2)} \qquad \sum_{i=1}^{k} \mu_i \int_{D_i} p(\hat{\boldsymbol{\theta}}) d\hat{\boldsymbol{\theta}} = \mu_k + \sum_{i=1}^{k-1} (\mu_i - \mu_k) \int_{D_i} p(\hat{\boldsymbol{\theta}}) d\hat{\boldsymbol{\theta}}$$

which, after standardization, yields the first two terms of (3.3). For the second sum on the extreme right of (A.1), it suffices to illustrate the term $i=k$. Let

$$\text{(A.3)} \qquad \begin{aligned}&\boldsymbol{x}' = (x_1, \cdots, x_{k-1}), \qquad \boldsymbol{b}' = (b_1, \cdots, b_{k-1}) \\ &x_j = \hat{\theta}_k - \hat{\theta}_j + b_j, \quad b_j = \mu_j - \mu_k, \qquad (j = 1, \cdots, k-1), \\ &x_k = \hat{\theta}_k - \mu_k\end{aligned}$$

and partition

$$\boldsymbol{V} = \begin{bmatrix} \underset{k-1}{\boldsymbol{V}_{11}} & \underset{1}{\boldsymbol{V}_{1k}} \\ \underset{k-1}{\boldsymbol{V}_{k1}} & \underset{1}{v_{kk}} \end{bmatrix}^{k-1}_{1}.$$

Since $\mathrm{E}(x_k|\boldsymbol{x}) = (v_{kk}\mathbf{1}_{k-1} - \boldsymbol{V}_{1k})'\boldsymbol{U}^{-1}\boldsymbol{x}$ where $\boldsymbol{U} = \boldsymbol{V}_{11} - \mathbf{1}_{k-1}\boldsymbol{V}_{1k} - \boldsymbol{V}_{k1}\mathbf{1}'_{k-1} + v_{kk} \cdot \mathbf{1}_{k-1}\mathbf{1}'_{k-1}$ is the covariance matrix of $\boldsymbol{x}$, it follows that

$$\text{(A.4)} \qquad \begin{aligned}\int_{D_k} (\hat{\theta}_k - \mu_k) p(\hat{\boldsymbol{\theta}}) d\hat{\boldsymbol{\theta}} &= \int_{b}^{\infty} \mathrm{E}(x_k|\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} \\ &= \sum_{j=1}^{k-1} (v_{kk} - v_{kj}) \int_{b}^{\infty} \boldsymbol{w}'_j \boldsymbol{x} p(\boldsymbol{x}) d\boldsymbol{x}.\end{aligned}$$

where $\boldsymbol{w}'_j$ is the jth row of $\boldsymbol{U}^{-1}$. Using the fact that

$$\boldsymbol{w}'_j \boldsymbol{x} = \frac{1}{2} \frac{\partial}{\partial x_j} \boldsymbol{x}' \boldsymbol{U}^{-1} \boldsymbol{x}$$

we find

$$\text{(A.5)} \qquad \int_{b}^{\infty} \boldsymbol{w}'_j \boldsymbol{x} p(\boldsymbol{x}) d\boldsymbol{x} \propto \int_{b_j}^{\infty} \left[-\exp\left(-\frac{1}{2}\boldsymbol{x}' \boldsymbol{U}^{-1} \boldsymbol{x}\right) \right]_{x_j = b_j}^{\infty} d\boldsymbol{x}_j,$$

where $\boldsymbol{x}_j$ and $\boldsymbol{b}_j$ are $(k-2)\times 1$ vectors obtained from $\boldsymbol{x}$ and $\boldsymbol{b}$ by deleting respectively x_j and b_j. Expression (A.5) is thus proportional to a $k-2$ dimensional normal integral. Repeating this argument for all (i, j) and after some simplification and standardization, we obtain the third term in (3.3). The case when $\boldsymbol{V}$ is of the form (5.1) has been given by Dunnett [9].

UNIVERSITY OF WISCONSIN
UNIVERSITY OF IFE

REFERENCES

[1] Bechhofer, R. E. (1954). A single sample multiple decision procedure for ranking means of normal populations with known variances, *Ann. Math. Statist.*, **25**, 16-39.

[2] Box, G. E. P. and Hunter, W. G. (1965). Sequential design of experiments for nonlinear models, *I.B.M. Computing Symposium on Statistics, Oct. 1963.*

[3] Box, G. E. P. and Tiao, G. C. (1965). Multiparameter problems from a Bayesian point of view, *Ann. Math. Statist.*, **36**, 1468-1482.

[4] Child, D. R. (1967). Reduction of the multivariate normal integral to characteristic form, *Biometrika*, **54**, 293-300.

[5] Curnow, R. N. and Dunnett, C. W. (1962). Numerical evaluation of certain multivariate normal integrals, *Ann. Math. Statist.*, **33**, 571-579.

[6] DeGroot, M. H. (1970). *Optimal Statistical Decision*, McGraw-Hill, New York.

[7] Draper, N. R. and Guttman, I. (1968). Some Bayesian stratified two-phase sampling results, *Biometrika*, **55**, 131-139.

[8] Draper, N. R. and Hunter, W. G. (1967). The use of prior distributions in the design of experiments for parameter estimation, *Biometrika*, **54**, 147-153.

[9] Dunnett, C. W. (1960). On selecting the largest of k normal population means, *J. R. Statist. Soc.*, B, **22**, 1-40.

[10] Elving, G. (1959). Design of linear experiments, in *Probability and Statistics*: Harold Cramer, vol. Ed., U. Grenander, Wiley, New York, 58-74.

[11] Ericson, W. A. (1965). Optimum stratified sampling using prior information, *J. Amer. Statist. Ass.*, **60**, 750-771.

[12] Ericson, W. A. (1970). Subjective Bayesian models in sampling finite populations, (with Discussion), *J. R. Statist. Soc.*, B, **31**, 195-233.

[13] Grundy, P. M., Healy, M. J. R. and Rees, D. H. (1956). Economic choice of the amount of experimentation, *J. R. Statist. Soc.*, B, **18**, 32-55.

[14] Hardy, G. H., Littlewood, J. E. and Pólya, G. (1959). *Inequalities*, Cambridge University Press.

[15] Ihm, P. (1959). Numerical evaluation of certain multivariate normal integrals, *Sankhya*, **21**, 363-366.

[16] John, S. (1959). On the evaluation of the probability integral of a multivariate normal distribution, *Sankhya*, **21**, 367-370.

[17] Kendall, M. G. and Stuart, A. (1966). *The Advanced Theory of Statistics*, vol. 3, Griffin, London.

[18] Kiefer, J. (1958). On the non randomized optimality and randomized nonoptimality of symmetric designs, *Ann. Math. Statist.*, **29**, 675-699.

[19] Kiefer, J. (1959). Optimum experimental designs (with Discussion), *J. R. Statist. Soc.*, B, **21**, 272-319.

[20] Lindley, D. V. (1971). *Bayesian Statistics, A Review*, Society for Industrial and Applied Mathematics, Philadelphia.

[21] Moran, P. A. P. (1968). *An Introduction to Probability Theory*, Clarendon Press, Oxford.
[22] Plackett, R. L. (1954). A reduction formula for the normal multivariate integrals, *Biometrika*, **41**, 351-360.
[23] Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*, Harvard University, Boston.
[24] Steck, G. P. (1958). A table for computing trivariate normal probabilities, *Ann. Math. Statist.*, **29**, 780-800.
[25] Turing, A. M. (1948). Rounding off errors in matrix processes, *Quart. J. Mech. and Applied Math.*, **1**, 287.