

OPTIMAL DESIGNS FOR LINEAR CALIBRATION PROBLEMS

Roger R. Davidson, University of Victoria

and

Daniel L. Solomon, Cornell University

Summary

A Bayesian analysis is given for choosing the number of observations and values of the independent variable in the context of simple linear regression through the origin. The model assumes normal errors, quadratic loss and conjugate prior. Optimal designs are characterized for situations in which the cost associated with collecting a set of observations is a Schur-convex or Schur-concave function of the set of independent variables.

Key words: Bayesian design, Linear calibration, Majorization, Optimal design, Regression through the origin, Schur-convexity

---

This research was supported by the National Research Council of Canada under grant No. A7166 held at the University of Victoria by Roger R. Davidson.

## 1. Introduction

In certain experimental settings an appropriate mathematical model is the linear regression model constrained to pass through a specific point usually taken to be the origin. For example, in the calibration of a measurement instrument, samples containing none ( $x = 0$ ) of the substance being measured should show an instrument reading ( $Y$ ) of zero. The problem of estimating the slope  $\beta$  of the regression line  $E(Y) = \beta x$  will be referred to as the linear calibration problem. The classical (least-squares, maximum likelihood) estimators and their properties are well known and widely documented. In addition, results are available on the optimum location of the independent variable  $x$ . Relatively unexplored, however, are Bayesian decision theoretic approaches to the design of a linear calibration study, and that is the focus of this work.

As a key feature of the mathematical formulation which follows, we shall propose a cost  $c(x_1, \dots, x_m)$  to be associated with obtaining a set of values  $Y_{x_1}, \dots, Y_{x_m}$  of the response variable at respective values  $x_1, \dots, x_m$  of the independent variable. This cost will be formally taken into account when selecting values of  $m$  and  $(x_1, \dots, x_m)$  upon which to base the calibration. As in most problems of optimal design in which the cost of experimentation is made explicit, there is a tradeoff between that cost and the precision of the inference to be drawn. An example which illustrates the reality of including the cost of experimentation is one in which the collection of an observation is destructive. Consider in this context a chemical assay employing an instrument to measure the amount of some precious substance in a sample. If the procedure for preparing the sample for assay destroys its value, then the cost of assaying a sample might be taken proportional to the amount of the substance present. Similar considerations apply to titration experiments with costly reagents.

The general Bayesian decision theoretic model for optimal choice of experiment can be briefly summarized as follows. The decision maker must specify a loss function  $\ell(\theta, a, e)$  which reflects the penalty associated with conducting the experiment  $e \in E$  and then taking the action (inference)  $a \in A$ , when  $\theta$  is the operative state of nature. The probability distribution of the data  $Y$  observed depends upon the experiment performed and the state of nature. The class of decision rules  $d \in D(e)$ , which associate actions with observed data, thus also depends on the experiment. Finally, a (prior) probability distribution must be specified over the set of states of nature. Inference proceeds from a specific experiment  $e$ , by selecting from  $D(e)$  a decision rule  $d_e^*$ , called a Bayes rule, which minimizes the Bayes risk

$$E_{Y, \theta} \ell(\theta, d_e^*(Y), e) = \inf_{d \in D(e)} E_{Y, \theta} \ell(\theta, d(Y), e).$$

An optimal experiment (design)  $e^*$  is then defined to be one which minimizes over  $E$ , the Bayes risk of the Bayes rule  $d_e^*$ , that is

$$E_{Y, \theta} \ell(\theta, d_{e^*}^*(Y), e^*) = \inf_{e \in E} E_{Y, \theta} \ell(\theta, d_e^*(Y), e).$$

In the linear calibration problem, the states of nature are the values of the slope  $\beta$  while actions are slope estimates. In section 2, we identify the other components of the general formulation of the optimal design problem in such a way as to specify the linear calibration model.

We feel that an important aspect of the contribution made in this paper is the use of the mathematical theory of majorization (a partial order). Majorization has proved a useful tool in a wide area of applications, and we hope that this work demonstrates its potential power in problems of optimal design. Some key results about majorization are outlined in section 2, which are then applied to establish some properties of the optimal calibration design for a general class of cost functions.

Section 3 provides a procedure for the explicit calculation of the optimal number and location of values of the independent variable when the cost function is restricted to an important subclass of those considered in the previous section. The final section offers some results for an extension to the case in which the minimum cost for taking observations does not occur at the origin.

## 2. Characterization of Optimal Designs

### 2.1 The linear model and associated distribution theory

We assume that a response variable  $Y$  arises from a simple linear model with known ordinate, which is set equal to zero, at  $x = 0$ . The resulting model is  $Y_x = \beta x + \epsilon$  where the random error  $\epsilon$  is normally distributed according to the probability law  $L[\epsilon|\beta] = N(0, \sigma^2)$ , or equivalently  $L[Y_x|\beta] = N(\beta x, \sigma^2)$ , with  $\sigma^2$  independent of  $x$ . The maximum likelihood estimator of the slope parameter  $\beta$  based on  $m$  independent observations

$$(\underline{Y}, \underline{x}) = ((y_1, x_1), \dots, (y_m, x_m))$$

is

$$\hat{\beta} = \frac{\sum_{i=1}^m y_i x_i}{\sum_{i=1}^m x_i^2}.$$

Next we introduce a prior distribution for  $\beta$ , namely  $L[\beta] = N(\beta_0, \sigma_0^2)$ . This prior is a member of the family which is conjugate to the joint sampling distribution of the  $m$  observable responses. The parameter  $\beta_0$  is our prior expectation of the slope  $\beta$ , and the variance  $\sigma_0^2$  is inversely proportional to the precision of our prior information. When  $\sigma^2$  is known the precision of the prior relative to the sampling distribution can be represented by  $n_0 = \sigma^2/\sigma_0^2$ . It follows directly (cf. Raiffa and Schlaifer (1961) p.337) that  $\beta$  has posterior distribution  $L[\beta | (\underline{Y}, \underline{x})] = N(\beta_1, \sigma_1^2)$  with

$$\beta_1 = (n_0 \beta_0 + \sum_{i=1}^m x_i^2 \hat{\beta}) / (n_0 + \sum_{i=1}^m x_i^2)$$

$$\sigma_1^2 = \sigma^2 / (n_0 + \sum_{i=1}^m x_i^2)$$

Note that the mean of the posterior is the weighted average of the mean of the prior and the maximum likelihood estimator where the weights are proportional to the prior precision of  $\beta$  and sampling precision of  $\hat{\beta}$  respectively.

The criterion for the selection of a design point  $\underline{x} = (x_1, \dots, x_m)$  is to be based on the total Bayes risk defined by

$$b(\underline{x}) = r(\underline{x}) + c(\underline{x})$$

where  $r(\underline{x})$  is the minimum Bayes risk and  $c(\underline{x})$  is the cost of making observations at  $\underline{x}$ . For each  $m$  we wish to determine a design point  $\underline{x}$  which minimizes the total Bayes risk. We denote an optimal  $\underline{x}$  for fixed  $m$  by  $\underline{x}_m^*$ . In addition, we wish to find an optimal value for  $m$ .

Under the assumption of quadratic loss  $L(\beta, \tilde{\beta}) = K(\beta - \tilde{\beta})^2$ , the Bayes estimator is the posterior mean  $\beta_1$  and the minimum Bayes risk is

$$r(\underline{x}) = K\sigma_1^2 = K\sigma^2 / (n_0 + \sum_{i=1}^m x_i^2).$$

If we assume that the cost  $c(\underline{x})$  is on a per unit loss basis, then setting  $K = 1$  results in no loss of generality.

It is convenient to be able to restrict attention to nonnegative design points  $\underline{x}$ . This will be possible provided that  $c(\underline{x})$  is an even function of each component of  $\underline{x}$ , that is for each  $\underline{x}$ ,  $c(\underline{x}) = c(|\underline{x}|)$  where  $|\underline{x}| = (|x_1|, \dots, |x_m|)$ . Indeed, it suffices to make the less restrictive assumption that for each  $\underline{x}$ ,  $c(\underline{x}) \geq c(|\underline{x}|)$ . Under this assumption the character of the optimal design point  $\underline{x}_m^*$  can be determined for a wide class of cost functions using the technique of majorization. In this connection it should be noted that one of the common objections to the decision theoretic approach to statistical

problems is the difficulty of specifying realistic loss and cost functions. Thus results which are insensitive to the specific choice of the cost and/or loss functions should have considerable appeal.

## 2.2 Majorization and its application to the design problem

Majorization is a relation which determines a partial ordering of the space of  $m$ -dimensional vectors of real numbers. There is a wealth of interesting and diverse applications of majorization, for example Marshall and Olkin (1965), (1974) and Low (1970).

Let  $\underline{u}$  and  $\underline{v}$  be  $m$ -dimensional vectors of real numbers. Then  $\underline{u}$  is said to majorize  $\underline{v}$ , denoted  $\underline{u} > \underline{v}$ , if upon ordering the components

$$u_1 \geq u_2 \geq \dots \geq u_m \quad \text{and} \quad v_1 \geq v_2 \geq \dots \geq v_m$$

it follows that

$$\sum_{i=1}^k u_i \geq \sum_{i=1}^k v_i, \quad k = 1, \dots, m-1 \quad \text{and} \quad \sum_{i=1}^m u_i = \sum_{i=1}^m v_i.$$

A real valued function  $g$  of  $m$  variables is said to be  $S$ -convex (Schur-convex) if  $g(\underline{u}) \geq g(\underline{v})$  whenever  $\underline{u} > \underline{v}$ . In other words  $g$  is  $S$ -convex if and only if  $g$  preserves order under majorization. A function  $g$  is  $S$ -concave if and only if  $-g$  is  $S$ -convex. Because majorization is reflexive,  $\underline{u} > \underline{u}$ , it follows that any  $S$ -convex ( $S$ -concave) function is symmetric in  $\underline{u}$ , i.e. invariant under permutation of  $\underline{u}$ .

If  $\underline{u}$  is any nonnegative vector whose components sum to  $c$  it is clear that

$$(c, 0, 0, \dots, 0) = \underline{u}_1 > \underline{u} > \underline{u}_2 = (c/m, c/m, \dots, c/m).$$

It then follows that for any  $S$ -convex function  $g$  defined on the space of nonnegative vectors,

$$g(\underline{u}_1) \geq g(\underline{u}) \geq g(\underline{u}_2)$$

For  $g$   $S$ -concave, the above inequalities become reversed.

The class of real valued functions of  $\underline{y}$  which are convex and symmetric in  $\underline{y}$  constitute an important subclass of the class of S-convex functions. Within this subclass are the functions  $g(\underline{y}) = \sum_{i=1}^m \phi(u_i)$  where  $\phi$  is any convex function defined on the real line. An important characterization of majorization is that  $\underline{y} > \underline{z}$  if and only if  $\sum_{i=1}^m \phi(u_i) \geq \sum_{i=1}^m \phi(v_i)$  for every convex function  $\phi$ . For further details on majorization the reader is referred to Hardy, Littlewood and Polya (1952), Berge (1963), and Marshall, Olkin and Proschan (1967).

Let us now utilize the ideas of majorization to characterize solutions to the design problem specified earlier. The posterior variance, and hence the minimum Bayes risk  $r(\underline{x})$ , depends on the design point  $\underline{x}$  only through  $t_2 = \sum x_i^2$ . Thus the problem of obtaining an optimal design point  $\underline{x}_m^*$  can be approached by determining the  $\underline{x}$  which minimizes the cost  $c(\underline{x})$  within each class of  $\underline{x}$  having fixed  $t_2$ , and then finding an optimal  $t_2$ .

As noted earlier, we have assumed that  $c(\underline{x}) \geq c(|\underline{x}|)$  for each  $\underline{x}$  so that we can restrict attention to nonnegative design points  $\underline{x}$ . Setting  $\underline{u} = (x_1^2, \dots, x_m^2)$  we can represent the total Bayes risk by

$$B(\underline{u}) = R(\underline{u}) + C(\underline{u})$$

where  $B$ ,  $R$  and  $C$  correspond to  $b$ ,  $r$  and  $c$  defined in section 2.1 when rewritten as functions of  $\underline{u}$ . For each fixed  $t_2 = \sum_{i=1}^m u_i \geq 0$ , the total Bayes risk  $B(\underline{u})$  is minimized at  $\underline{u}_1 = (t_2, 0, \dots, 0)$  when  $C(\underline{u})$  is S-concave in  $\underline{u}$ , and at  $\underline{u}_2 = (t_2/m, \dots, t_2/m)$  when  $C(\underline{u})$  is S-convex in  $\underline{u}$ .

When the cost  $C(\underline{u})$  is S-concave the optimal design for each  $m$  involves making at most one observation at a non-zero  $x$ . On the other hand, when  $C(\underline{u})$  is S-convex the optimal design involves sampling  $m$  times at a common point. Hence in either event, an optimal design results in observing the response variable  $Y$  at a single non-zero  $x$ . It should be emphasized that the value of  $x$ , or equivalently  $t_2$ , at which the sampling is to be done depends on the actual

cost function. This is illustrated for a special class of cost function in the next section.

### 3. A Special Class of Cost Functions

In this section we consider the class of cost functions defined by

$$c(\underline{x}) = \sum_{i=1}^m c(x_i) \text{ where the cost component}$$

$$c(x) = s + d|x|^\alpha$$

with  $s > 0$ ,  $d > 0$ , and  $\alpha > 0$  specified constants. The constant  $s$  can be interpreted as the setup or overhead cost associated with any observation while the second part represents the way that the cost of an observation depends on the point  $x$  at which it is taken. The function  $c(\underline{x})$  is symmetric in  $\underline{x}$  and is an even function of each component. Observations at  $x = 0$  are the least costly, involving only the setup cost. However, observations at  $x = 0$  have no effect on the posterior distribution and contribute nothing to reducing the minimum Bayes risk. This should be intuitive insofar as the regression line has been constrained to pass through the origin. Thus there is no reason to ever make an observation at  $x = 0$ .

The cost component  $c(x)$  can be rewritten as  $C(u) = s + d|u|^{\alpha/2}$  where  $u = x^2$ . When  $\alpha \leq 2$  the cost component  $C(u)$  is concave and hence the cost function  $C(\underline{y}) = \sum_{i=1}^m C(u_i)$  is S-concave; when  $\alpha \geq 2$   $C(u)$  is convex and the cost  $C(\underline{y})$  is S-convex. In either case the character of an optimal design point is as described in the preceding section. We now consider the problem of finding a specific optimal point  $\underline{x}_m^*$  for a specified  $m$ , and of determining the optimal value of  $m$ .

(i)  $\alpha \leq 2$ . An optimal point has the form  $\underline{x} = (x, 0, \dots, 0)$  where  $x \geq 0$  without loss of generality. The total Bayes risk for points of this form is given by

$$b_m(x) = \sigma^2 / (n_0 + x^2) + mx + d x^\alpha .$$



It is clear that an optimal  $x$  does not depend on  $m$ . Moreover, since observations at  $x = 0$  have no effect on the minimum Bayes risk, there is no reason to make more than one observation. Setting the derivative  $b'_m(x)$  to zero we obtain the equation

$$n_0 + x^2 = \sqrt{2\sigma^2/\alpha d}^{1-\frac{1}{2}\alpha} x$$

When  $\alpha = 2$  it follows that  $b_m(x)$  achieves its minimum at  $x^* = 0$  or

$$\sqrt{(2\sigma^2/\alpha d)^{\frac{1}{2}} - n_0^{\frac{1}{2}}}. \quad \text{The case } \alpha < 2 \text{ requires more detailed analysis. The}$$

left hand side is increasing in  $x$  with positive concavity while the right side is increasing with negative concavity. Thus there are at most two positive solutions, the actual number depending upon the parameters of the prior and of the cost function. For any fixed  $m$  the total Bayes risk  $b_m(x)$  is increasing in a neighbourhood of zero and has continuous first derivative for  $x > 0$ . Thus if there are zero or one solutions then  $b_m(x)$  is minimized at  $x = 0$ ; if there are two positive solutions then  $b$  is minimized either at the larger of the two solutions,  $x^*$ , or at  $x = 0$ . The required solution cannot be found explicitly but can be obtained using a standard iterative procedure such as Newton's method.

In summary, the optimal sampling scheme is to make a single observation at  $x^*$  if  $b_1(x^*) < b_0 = \sigma_0^2$ , and otherwise to simply estimate  $\beta$  using the prior mean.

(ii)  $\alpha > 2$ . For each  $m$  an optimal point has the form  $\underline{x} = (x, \dots, x)$  where without loss of generality  $x > 0$ . The total Bayes risk for points of this form is

$$b_m(x) = \sigma^2/(n_0 + mx^2) + ms + md x^\alpha$$

When we set the derivative  $b'_m(x)$  to zero we obtain the equation

$$(n_0 + mx^2)^2 = (2\sigma^2/\alpha d)x^{2-\alpha}$$

In that the left hand side is increasing in  $x$  while the right hand side is decreasing from  $+\infty$  to 0, this equation has a unique positive solution  $x_m^*$ , which is easily seen to uniquely yield the minimum of  $b_m(x)$ . As before the value of  $x_m^*$  must be obtained using an iterative procedure. In this case, however,  $x_m^*$  depends on  $m$ . In fact it follows from the derivative equation that  $x_m^*$  is a decreasing function of  $m$  which approaches zero as  $m$  becomes infinite. On the other hand the quantity  $w_m = mx_m^{*2}$  is an increasing function of  $m$  which becomes infinite as  $m$  does. Note that  $x_m^*$  is of order  $m^{-\delta}$  for some  $0 < \delta < \frac{1}{2}$ .

In order to determine an optimal sample size  $m$  we need to consider  $b_m^* = b_m(x_m^*)$ , the total Bayes risk at the optimal point, as a function of  $m$ . Expressing  $b_m^*$  as a function of  $w_m = mx_m^{*2}$  and substituting from the derivative equation yields

$$b_m^* = ms + \sigma^2 \frac{[n_0 + (1+2/\alpha)w_m]}{(n_0 + w_m)^2}$$

where  $m$  is related to  $w_m$  by

$$m = \left( \frac{\alpha d}{2\sigma^2} \right)^{\frac{2}{\alpha-2}} (n_0 + w_m)^{\frac{4}{\alpha-2}} w_m.$$

Note that there is a unique  $w_m$  associated with each  $m$  and that  $w_m$  and  $m$  increase together. Using the chain rule we obtain the derivative

$$\frac{db_m^*}{dm} = s \left[ 1 - \kappa / (n_0 + w_m)^{\frac{2\alpha}{\alpha-2}} \right]$$

where

$$\kappa = \frac{(\alpha-2)}{s} \left(\frac{2}{d}\right)^{\frac{2}{\alpha-2}} \left(\frac{\sigma^2}{\alpha}\right)^{\frac{\alpha}{\alpha-2}} > 0.$$

The derivative is an increasing function of  $w_m$  and hence of  $m$ . Thus there is a unique  $w_m^*$  which minimizes  $b_m^*$  for  $w_m \geq 0$ ,

namely

$$w_m^* = \max\left(0, \kappa^{\frac{\alpha-2}{2\alpha}} - n_0\right).$$

Substituting into the equation for  $m$  we obtain  $m^*$ , the value of  $m$  which minimizes  $b_m^*$  for  $m \geq 0$ . The optimal  $m$  is then one of the integers adjacent to  $m^*$ ,  $[m^*]$  or  $[m^*+1]$ . The corresponding values of  $w_m$  can then be obtained using the numerical procedure alluded to earlier in order to permit a comparison of the two possible values. In practice, however, there will only be a small difference in the total Bayes risk of the optimal  $x_m$  associated with the two candidates.

#### 4. An Extension of the Class of Cost Functions

The character of optimal designs for linear calibration as developed in section 2 is considerably more general than the case developed in the last section would indicate. For instance, a restrictive feature of the class of cost functions in section 3 is that they involve the cost being minimized at  $x = 0$ , the point at which the response is known. Suppose that we alleviate the situation by allowing the minimum cost to occur at  $x_0 \geq 0$ . (Alternatively, we could allow for a general point  $\tilde{x}$  at which the response is known. However, it is clear that there is no loss of generality in allowing at least one of  $x_0$  and  $\tilde{x}$  to be zero.) The class of cost functions in section 3 can then be generalized to  $c(\underline{x}) = \sum_{i=1}^m c(x_i)$  where

$$c(x) = s + d|x - x_0|^\alpha$$

Since  $c(\underline{x}) \geq c(|\underline{x}|)$  for all  $\underline{x}$  and  $c(\underline{x})$  is minimized at  $\underline{x}_0 = x_{01} \geq 0$ , we can restrict attention to design points  $\underline{x}$  such that  $\underline{x} \geq \underline{x}_0$ . For if  $0 \leq x_i < x_0$  for some  $i$ , we can reduce both the cost and the minimum Bayes risk by replacing  $x_i$  by  $x_0$ . Now let  $C(\underline{y})$  denote  $c(\underline{x})$  rewritten as a function of  $\underline{y} = (x_1^2, \dots, x_m^2)$ . In the present situation it is easily seen that  $C(\underline{y})$  is S-convex in  $\underline{y}$  when  $\alpha \geq 2$ , while  $C(\underline{y})$  is S-concave on  $\underline{y} \geq \underline{y}_0 = (x_0^2, \dots, x_0^2)$  when  $\alpha \leq 1$ , so that the characterizations of the optimal design points given in section 2 are in effect. Unfortunately, we are unable to use these results for the intermediate cases  $1 < \alpha < 2$ .

Let us now generalize the class of cost functions exemplified by this situation. Suppose that  $c(\underline{x}) \geq c(|\underline{x}|)$  for all  $\underline{x}$  and that  $c(\underline{x})$  assumes its minimum at the point  $\underline{x}_0 \geq 0$  whose coordinates are not necessarily all the same. Then it suffices to restrict attention to design points  $\underline{x}$  such that  $\underline{x} \geq \underline{x}_0$ . As before we can approach the problem of finding the optimal  $\underline{x}_m^*$  for fixed  $m$  by specifying  $t_2 = \sum_i x_i^2 \geq \sum_i x_{0i}^2$ , minimizing  $c(\underline{x})$  among those  $\underline{x} \geq \underline{x}_0$  which yield  $t_2$ , and then finding the optimal value of  $t_2$ . Once again we let  $C(\underline{y})$  denote  $c(\underline{x})$  rewritten as a function of  $\underline{y} = (x_1^2, \dots, x_m^2)$  for  $\underline{y} \geq \underline{y}_0 = (x_{01}^2, \dots, x_{0m}^2)$ . When  $C(\underline{y})$  is S-concave in  $\underline{y}$  for  $\underline{y} \geq \underline{y}_0$  the optimal  $\underline{x}_m^*$  has the form  $(x_m^*, x_0, \dots, x_0)$ , while when  $C(\underline{y})$  is S-convex in  $\underline{y}$  for  $\underline{y} \geq \underline{y}_0$  the optimal  $\underline{x}_m^*$  has the form  $(x_m^*, \dots, x_m^*)$ . In such cases the specific value of  $x_m^*$ , or equivalently the optimal value of  $t_2$ , depends on the specific cost function.

To illustrate this point let us return to the specific class of cost functions introduced earlier and consider the problem of finding a specific optimal point  $\underline{x}_m^*$  for fixed  $m$ . We examine separately the cases  $\alpha \leq 1$  and  $\alpha \geq 2$ .

(i)  $\alpha \leq 1$ . For each  $m$  an optimal design point has the form

$\underline{x} = (x, x_0, \dots, x_0)$  where  $x \geq x_0 \geq 0$ . The total Bayes risk for points of this form is

$$b_m(x) = \sigma^2 / [n_0 + (m-1)x_0^2 + x^2] + ms + d(x-x_0)^\alpha$$

Note that when  $x_0 > 0$  an optimal  $x$  depends on  $m$ . Setting the derivative  $b'_m(x)$  to zero we obtain the equation.

$$n_0 + (m-1)x_0^2 + x^2 = [2\sigma^2 x(x-x_0)^{1-\alpha} / \alpha d]^{1/2}.$$

Proceeding as in section 3 it can be seen that this equation has at most two solutions, and that when  $x_0 > 0$  any solutions are a function of  $m$ .

If there are zero or one solutions then  $b_m$  is minimized at  $x = x_0$ ; if there are two solutions then  $b$  has local minima at the larger solution  $\tilde{x}_m$  and at  $x_0$ .

Observations made at the least costly point  $x_0 > 0$ , unlike the case  $x_0 = 0$ , result in a decrease in the minimum Bayes risk so that it is possible for the optimal  $m$  to be any nonnegative integer. Unfortunately, we are not able to explicitly determine the optimal  $m$  in terms of the parameters of the prior and of the cost function. However, certain features can be noted.

First there exists an  $M$  such that  $b_m$  is minimized at  $x_0$  for each  $m \geq M$ . An upper bound on  $M$  can be found by determining the real number  $m$  at which the equation above has a single solution at the point  $x_c$  at which the two sides have the same slope. Second, we can determine the  $m_0$  which is optimal when all observations are made at  $x_0$ ;  $m_0$  will be zero when  $x_0 \leq s^{1/2} n_0 / \sigma$  and will be one of the integers adjacent to  $x_0^{-2} (s^{-1/2} \sigma x_0 - n_0)$  when  $x_0 > s^{1/2} n_0 / \sigma$ .

Thus to determine the optimal  $m$ , one could compute  $b_m(\tilde{x}_m)$  for each  $m < M$  and compare with  $b_{m_0}(x_0)$ , a task which is reasonable for modest values of  $M$ .

(ii)  $\alpha \geq 2$ . For each  $m$  an optimal design point has the form  $\underline{x} = (x, \dots, x)$  where  $x \geq x_0 \geq 0$ . The total Bayes risk for points of this form is

$$b_m(x) = \sigma^2 / (n_0 + mx^2) + ms + md(x-x_0)^\alpha$$

Examining the derivative of  $b_m(x)$ , it is easily seen that  $b_m(x)$  assumes its minimum at the unique point  $x_m^* > x_0$  which satisfies

$$(n_0 + m x^2)^2 = (2\sigma^2/\alpha d) x (x - x_0)^{1-\alpha}$$

The value of  $x_m^*$  must be obtained iteratively. It is clear that  $x_m^*$  depends on  $m$ . Indeed it follows from the equation above that  $x_m^*$  decreases to  $x_0$  as  $m$  increases to infinity, while  $w_m = m x_m^{*2}$  increases to infinity as  $m$  increases. However, unlike the case  $x_0 = 0$  we are unable to obtain an explicit expression for  $m$  in terms of  $w_m$  and the parameters of the prior and of the cost function, and hence cannot determine the optimal  $m$  explicitly.

It was noted earlier that for the class of cost functions under discussion, the results based on majorization do not apply in the intermediate cases  $1 < \alpha < 2$ . However, in these cases it can be shown directly that for each  $m$  an optimal design point has one of the two following forms:

$\underline{x} = (x_1, \dots, x_1, x_2)$  where  $x_0 \leq x_1 \leq x_0/(2-\alpha)$  and  $x_2 > x_0/(2-\alpha)$  or

$\underline{x} = (x, \dots, x)$  where  $x_0 \leq x \leq x_0/(2-\alpha)$ . The precise form depends on the parameters of the prior and cost function as well as on  $\alpha$ .

The class of cost functions introduced in section 3 and generalized here in section 4 have been used to illustrate the procedures to be followed once the character of an optimal design point has been determined. It should be clear that such procedures can be examined in detail for any class of cost functions for which the characterizations based on majorization apply.

## References

- [1] Berge, C. (1963). Topological Spaces. MacMillan, New York.
- [2] Hardy, G. H., Littlewood, J. E., and Polya, G. (1952). Inequalities. Cambridge University Press, Cambridge.
- [3] Low, L. (1970). An application of majorization to comparison of variances. Technometrics 12, 141-145.
- [4] Marshall, A. W. and Olkin, I. (1965). Norms and inequalities for condition numbers. Pacific J. Math. 15, 241-247.
- [5] Marshall, A. W. and Olkin, I. (1974). Majorization in multivariate distributions. Ann. Statist. 2, 1189-1200.
- [6] Marshall, A. W., Olkin, I. and Proschan, F. (1967). Monotonicity of ratios of means and other applications of majorization. Inequalities (O. Shisha, ed.), Academic Press, New York.
- [7] Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory. Harvard University Press.