

On the Economic Choice of Experiment Sizes for Decision Regarding Certain Linear Combinations

By W. A. ERICSON

University of Michigan

[Received September 1966]

SUMMARY

The author (1965) has presented an algorithm for choosing the sample sizes, n_i ($i = 1, \dots, k$), to be selected from each of k independent normal data-generating processes having unknown means, μ_i , in order to minimize the posterior variance of an arbitrary given linear combination of these means subject to a budget constraint. The joint prior was taken to be a general k -dimensional normal distribution. In the present paper it is shown how, using this algorithm and for a wide class of loss functions, the optimal experimental design (choice of n_1, \dots, n_k) can be reduced to a univariate minimization problem. Several specific loss functions are considered as examples which provide generalization of decision models given by Raiffa and Schlaifer (1961), as well as a generalization of the earlier work of Grundy *et al.* (1956).

1. INTRODUCTION

THE basic model to be assumed here, with the exception of some results briefly mentioned in Section 4, is as follows: Let $\mu = (\mu_1, \dots, \mu_k)$ be a vector of unknown means of k independent normal processes, having known variances σ_i^2 ($i = 1, \dots, k$). Suppose μ is assigned a general k -dimensional normal prior distribution with mean vector \mathbf{m}' and positive definite covariance matrix $\mathbf{V}' \equiv (\mathbf{N}')^{-1}$. Under this specification, if independent samples of sizes n_i , respectively, are drawn from these processes, then, letting $\bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_k)$ denote the vector of observed sample means ($\bar{X}_i \equiv 0$ if $n_i = 0$), the posterior distribution of μ given $\bar{\mathbf{X}}$ is k -dimensional normal with mean $\mathbf{m}'' = (\mathbf{m}'\mathbf{N}' + \bar{\mathbf{X}}\mathbf{N})(\mathbf{N} + \mathbf{N}')^{-1}$ and covariance matrix $\mathbf{V}'' = (\mathbf{N} + \mathbf{N}')^{-1}$, where \mathbf{N} is a diagonal matrix having elements n_i/σ_i^2 . It is additionally assumed that the parameter of interest for decision purposes is the scalar

$$\mu = \mu\pi^t, \quad (1)$$

where the superscript t denotes transpose and $\pi = (\pi_1, \dots, \pi_k)$ is a given vector. It is clear that the prior and posterior distributions of μ are univariate normal defined by the means $m' = \mathbf{m}'\pi^t$ and $m'' = \mathbf{m}''\pi^t$ and the variances $v' = \pi(\mathbf{N}')^{-1}\pi^t$ and $v'' = \pi(\mathbf{N} + \mathbf{N}')^{-1}\pi^t$, respectively. Finally, if a is an act whose utility depends on the unknown parameter value μ , it is assumed that the loss (negative utility) of the sequence of choosing $\mathbf{n} = (n_1, \dots, n_k)$, observing $\bar{\mathbf{X}}$, and choosing a when μ is the "true" state is separable and given by

$$l(\mathbf{n}, \bar{\mathbf{X}}, a, \mu) = l(a, \mu) + \mathbf{c}\mathbf{n}^t, \quad (2)$$

where $l(a, \mu) \geq 0$ and $\mathbf{c} > \mathbf{0}$ is a vector representing the per-unit observation cost.

Consistency in subjectivist decision necessitates choosing $a = a^*$ to minimize

$$l(\mathbf{n}, \bar{\mathbf{X}}, a) \equiv E''_{\mu | \bar{\mathbf{X}}, \mathbf{n}} l(\mathbf{n}, \bar{\mathbf{X}}, a, \mu) = E''_{\mu | \bar{\mathbf{X}}} l(a, \mu) + \mathbf{cn}^t; \tag{3}$$

while $\mathbf{n} = \mathbf{n}^*$ is chosen to minimize

$$l(\mathbf{n}) \equiv E_{\bar{\mathbf{X}} | \mathbf{n}} l(\mathbf{n}, \bar{\mathbf{X}}, a) = E_{\bar{\mathbf{X}} | \mathbf{n}} E''_{\mu | \bar{\mathbf{X}}} l(a^*, \mu) + \mathbf{cn}^t. \tag{4}$$

The main concern here is in finding \mathbf{n}^* under the above model, for various terminal loss functions $l(a, \mu)$.

This model has several immediate interpretations and applications. The first, giving the initial impetus to this work, is that of decision when sample information is provided by means of a stratified sample. In this interpretation the μ_i 's are the unknown stratum means and the π_i 's are taken proportional to the stratum sizes. A second important application is given by taking μ to be the difference between two normal process means, $\mu_1 - \mu_2$, say. Several others will be evident in Section 4 below, where, with $l(a, \mu)$ taken to be quadratic, the results given are easily extended to a number of other distributions generating the data and corresponding prior distributions.

2. A RELEVANT ALLOCATION PROBLEM

As will be shown in Section 3, the class of problems of finding the optimal experiment size, \mathbf{n}^* , under the above model is to a very large extent solved if one has solutions to the following allocation problem.

Choose $\mathbf{n} = \mathbf{n}^0 \geq 0$ to minimize

$$v''(\mathbf{n}) \equiv \pi(\mathbf{N} + \mathbf{N}')^{-1} \pi^t \tag{5}$$

subject to the added constraint that

$$\mathbf{cn}^t \leq C. \tag{6}$$

In other words, minimize the posterior variance of μ subject to a budgetary constraint. This auxiliary problem has been solved, in some detail, by the author (1965, 1966). Several useful properties of the function $v''(\mathbf{n})$ and the solution are given below. For complete details of the solution algorithm, explicit formulae and proofs of the results quoted below the reader is referred to these earlier papers.

Over the region $R = \{\mathbf{n} | \mathbf{n} \geq 0, \mathbf{cn}^t \leq C\}$ the function $v''(\mathbf{n})$, where \mathbf{N}' is positive definite symmetric, is twice differentiable with respect to \mathbf{n} , positive, non-increasing and a convex function of \mathbf{n} . Further, for every $\epsilon > 0$ there is an $\mathbf{n}(\epsilon) \geq 0$ such that $\mathbf{n} > \mathbf{n}(\epsilon)$ implies $v''(\mathbf{n}) \leq \epsilon$. These properties established the applicability of the results of Kuhn and Tucker (1951) in deriving a special-purpose algorithm for obtaining the solution to this allocation problem. Also it is thus clear that the constraint (6) may be replaced by an equality constraint. Finally, general-purpose convex programming techniques could be used for obtaining solutions.

Several salient features of the solution to this allocation problem may be briefly described as follows. The interval $C \geq 0$ is partitioned by the points

$$0 = C_r < C_{r-1} < \dots < C_1 < C_0$$

into the intervals I_0, I_1, \dots, I_r , where $I_0 = [C_0, \infty)$, $I_i = [C_i, C_{i-1})$ ($i = 1, \dots, r$). Within each of these intervals some subset, W_i of the k n_j^0 's is zero, the complementary set,

S_i , of the minimizing n_j^0 's being linear functions of C . Over all $C \geq 0$ each n_j^0 is a continuous, piecewise linear function of C . Explicit formulae for the \mathbf{n}^0 and methodology for determining these sub-intervals are given in the papers referred to above.

Further, making the dependence of \mathbf{n}^0 on C explicit by denoting the minimizing value of \mathbf{n} for the budget C by $\mathbf{n}^0(C)$ and letting

$$v^*(C) \equiv v''\{\mathbf{n}^0(C)\}, \tag{7}$$

then it can be shown that $v^*(C)$ is a continuous, strictly decreasing, convex function of C possessing a continuous first derivative. Also for $C \in I_i$, $v^*(C)$ has the functional form:

$$v^*(C) = K_i + \{\gamma_i^2 / (C + \Sigma_i)\}, \tag{8}$$

where $\gamma_i > 0$, $\Sigma_i > 0$ and K_i depend only on the interval I_i . The quantities γ_i and Σ_i can be explicitly expressed for each I_i in terms of π , N^i , \mathbf{c} and the σ_i^2 's. And $K_0 = 0$, $K_r = \pi(N^r)^{-1} \pi^t$, while the remaining K_i 's chosen to make $v^*(C)$ continuous at the end-points of the successive intervals.

With these brief preliminaries some simple, but useful, results concerning the class of decision problems described in Section 1 may be given.

3. GENERAL CHOICE OF \mathbf{n}^* AS A UNIVARIATE PROBLEM

Consider now the original class of problems posed, viz. of choosing that $\mathbf{n} \geq \mathbf{0}$, denoted by \mathbf{n}^* , which minimizes $l(\mathbf{n})$, (4), or

$$l(\mathbf{n}) = E \min_{\bar{X} | \mathbf{n}} E'' l(a, \mu) + \mathbf{c}\mathbf{n}^t. \tag{9}$$

Now the posterior expectation of $l(a, \mu)$, $E''_{\mu | \bar{X}, \mathbf{n}} l(a, \mu)$ will be some non-negative function of a , \mathbf{m}'' and $v''(\mathbf{n})$. Since this quantity depends on \bar{X} only through \mathbf{m}'' , the expectation of the minimum (over all a) of this quantity with respect to the marginal distribution (also normal) of \bar{X} is precisely its expectation with respect to the "prior" distribution of $\mathbf{m}'' = \mathbf{m}'' \pi^t$. Since \mathbf{m}'' is a linear function of \bar{X} , it is easily verified that this prior distribution of the posterior mean is normal with mean m' and variance

$$\pi \mathbf{V}' \pi^t - \pi \mathbf{V}'' \pi^t = v' - v''(\mathbf{n}).$$

It follows that for this model the first term of (9) depends on \mathbf{n} only through $v''(\mathbf{n})$.

By this argument and letting

$$f_i\{v''(\mathbf{n})\} \equiv E \min_{\bar{X} | \mathbf{n}} E'' l(a, \mu),$$

the choice of the design \mathbf{n}^* becomes a problem of choosing that $\mathbf{n} \geq \mathbf{0}$ which minimizes

$$f_i\{v''(\mathbf{n})\} + \mathbf{c}\mathbf{n}^t, \tag{10}$$

where f_i may be a very complicated function or not, depending on the form of $l(a, \mu)$.

The main result of this section may now be demonstrated. This result is that for any function f_i , i.e. any loss function $l(a, \mu) \geq 0$, the vector \mathbf{n}^* which minimizes (10) may be found by minimization of a function of only one variable, assuming that one has available a solution to the corresponding allocation problem of the preceding section.

First, a preliminary dual result regarding the allocation problem, namely that $\mathbf{n}^0(C)$ is also that vector $\mathbf{n} \geq \mathbf{0}$ which minimizes $g(\mathbf{n}) \equiv \mathbf{c}\mathbf{n}^t$ subject to the constraint that $v''(\mathbf{n}) = v^*(C)$. More formally one has:

Lemma 1. The vector solution to the allocation problem of Section 2 for the budget C , $\mathbf{n}^0(C)$, is also that vector \mathbf{n} which minimizes \mathbf{cn}^t subject to the constraints $\mathbf{n} \geq 0$ and $v''(\mathbf{n}) = v^*(C)$.

Proof. Assume the contrary, i.e. suppose there is an $\mathbf{n}' \geq 0$ such that $v''(\mathbf{n}') = v^*(C)$ and also $\mathbf{cn}'^t < \mathbf{c}\{\mathbf{n}^0(C)\}^t$. Then there must be a $\delta > 0$ such that $\mathbf{c}\{\mathbf{n}' + \delta\}^t = \mathbf{c}\{\mathbf{n}^0(C)\}^t$. Hence by the previously asserted properties of $v''(\mathbf{n})$, $v''(\mathbf{n}' + \delta) < v''(\mathbf{n}') = v^*(C)$, contradicting the hypothesis that $\mathbf{n}^0(C)$ minimizes $v''(\mathbf{n})$ among all $\mathbf{n} \geq 0$ for which $\mathbf{cn}^t = C$.

The basic result says, in effect, that that $\mathbf{n} = \mathbf{n}^*$ which minimizes $f_i\{v''(\mathbf{n})\} + \mathbf{cn}^t$ for any given function $f_i \geq 0$ is precisely $\mathbf{n}^0(C^0)$ which minimizes $v''(\mathbf{n})$ subject to $\mathbf{cn}^t = C^0$, where C^0 is that value of $C \geq 0$ which minimizes $f_i\{v^*(C)\} + C$. Thus for any loss function, under the above model, the determination of an optimal experiment \mathbf{n}^* may be accomplished by minimization of $f_i\{v^*(C)\} + C$ by choice of $C \geq 0$ (say C^0) and then taking $\mathbf{n}^* = \mathbf{n}^0(C^0)$ using the solution to the allocation problem of the preceding section. More formally one has:

Theorem 1. If for any function $f_i(x) \geq 0$ defined for $x \geq 0$,

$$g(C) = f_i\{v^*(C)\} + C \tag{11}$$

attains its minimum at $C = C^0$, say, then $l(\mathbf{n}) = f_i\{v''(\mathbf{n})\} + \mathbf{cn}^t$ attains its minimum over all $\mathbf{n} \geq 0$, at $\mathbf{n}^* = \mathbf{n}^0(C^0)$, where $\mathbf{n}^0(C)$ is that value of $\mathbf{n} \geq 0$ minimizing $v''(\mathbf{n})$ subject to $\mathbf{cn}^t = C$.

Proof. It suffices to show that for all $\mathbf{n} \geq 0$

$$f_i\{v^*(C^0)\} + \mathbf{c}\{\mathbf{n}^0(C^0)\}^t \leq f_i\{v''(\mathbf{n})\} + \mathbf{cn}^t. \tag{12}$$

By the properties of $v''(\mathbf{n})$ and $v^*(C)$ given earlier and Lemma 1, for every $\mathbf{n} \geq 0$, $v''(\mathbf{n})$ exists and also there exists some $C \geq 0$ such that $v''(\mathbf{n}) = v^*(C) \equiv v''\{\mathbf{n}^0(C)\}$ and also $C = \mathbf{c}\{\mathbf{n}^0(C)\}^t \leq \mathbf{cn}^t$. Thus for every $\mathbf{n} \geq 0$ there is some $C \geq 0$ such that

$$f_i\{v''(\mathbf{n})\} + \mathbf{cn}^t \geq f_i\{v^*(C)\} + C. \tag{13}$$

However, by hypothesis for all $C \geq 0$

$$f_i\{v^*(C)\} + C \geq f_i\{v^*(C^0)\} + C^0 = f_i\{v^*(C^0)\} + \mathbf{c}\{\mathbf{n}^0(C^0)\}^t. \tag{14}$$

The assertion (12) then follows immediately from (13) and (14).

By this result and earlier ones, it follows that if $f_i(v'')$ is a differentiable function of v'' , with derivative $f'_i(v'')$, then a first-order condition for a local minimum of $g(C)$, (11), at $C = C^0$, is that at C^0

$$(d/dC) [f_i\{v^*(C)\} + C] = f'_i\{v^*(C)\} \{dv^*(C)/dC\} + 1 = 0,$$

or

$$f'_i\{v^*(C)\} = -\{dv^*(C)/dC\}^{-1}, \tag{15}$$

where the derivative $dv^*(C)/dC$ is given for $C \in I_i$ by using the expression (8), as

$$dv^*(C)/dC = -\{\gamma_{ii}/(C + \Sigma_i)\}^2 \quad (C \in I_i). \tag{16}$$

In the remaining sections these results are used to give solutions for \mathbf{n}^* for three different loss functions: for estimation problems under quadratic and linear losses, and for a two-action problem with a linear utility function. In the special case of quadratic losses, the results will be seen to apply far more generally than under the normal distribution model assumed elsewhere.

4. QUADRATIC LOSSES

The simplest application is under a quadratic loss function, i.e. taking

$$l(a, \mu) = k(a - \mu)^2$$

for some $k > 0$. Such a loss function is often a reasonable approximation in estimation-type problems. Here $a^* = m''$ and it is readily verified that $l(\mathbf{n}) = kv''(\mathbf{n}) + \mathbf{cn}'$. By Theorem 1 this is minimized by $\mathbf{n}^* = \mathbf{n}^0(C^0)$ if $kv^*(C) + C$ attains its minimum at $C = C^0$ and $\mathbf{n}^0(C^0)$ is the solution to the allocation problem for the constraint $\mathbf{cn}' = C^0$. The problem is thus solved, given solutions to the allocation problem, by finding C^0 . The condition (15) becomes simply

$$k = \{(C + \sum_i \gamma_i) / \gamma_i\}^2, \tag{17}$$

for C in some I_i . Now it is clear from expression (8) that $v^*(C)$ and hence $kv^*(C) + C$ are continuous convex functions of $C \geq 0$. Thus C^0 is either zero or there is one and only one interval I_i for which the solution C_i of (17) is in I_i and that $C_i = C^0$.

In the case of quadratic losses the methodology given above can be adapted with little alteration to provide the optimal experimental design, \mathbf{n}^* , for inference or decision regarding the unknown linear combination $\mu = \boldsymbol{\mu}\boldsymbol{\pi}'$ under models other than normal with known variances. For, in general, with quadratic losses a^* is the posterior mean of μ while \mathbf{n}^* is chosen (≥ 0) to minimize $l(\mathbf{n}) = kE'v''(\mu) + \mathbf{cn}'$, where $E'v''(\mu)$ denotes the prior expectation of the posterior variance. Under a number of alternative distribution assumptions $E'v''(\mu)$ may be put into the form $\boldsymbol{\pi}(\mathbf{N} + \mathbf{N}')^{-1}\boldsymbol{\pi}'$, where the sample sizes are uniquely determined by the diagonal entries of \mathbf{N} . Whenever this can be done the problem of finding \mathbf{n}^* is formally identical to the normal distribution example of the previous paragraph. This program can be carried out in at least the cases where μ_i ($i = 1, \dots, k$) is taken to be the unknown proportion of a Bernoulli process, the parameter of a Poisson process or the unknown variance (mean known) of a normal process, where, in each case, independent natural conjugate prior distributions (Raiffa and Schlaifer, 1961) are assigned to each μ_i . The same generalization holds for decision on μ where the μ_i 's are normal process means and either the variances σ_i^2 are unknown and independent natural conjugate "normal-gamma" prior distributions are assigned to each $(\mu_i, 1/\sigma_i^2)$ pair, or the informations $1/\sigma_i^2$ are known only up to a multiplicative constant, h , and a k -dimensional normal-gamma prior is assigned to $(\boldsymbol{\mu}, h)$.

5. LINEAR LOSSES

As a second simple example consider decision regarding μ where

$$l(a, \mu) = \begin{cases} k_0(\mu - a) & (\mu \geq a) \\ k_1(a - \mu) & (\mu < a) \end{cases}$$

for positive constants k_0 and k_1 . In this case it is easily verified (see Raiffa and Schlaifer, 1961) that a^* is chosen such that

$$\Phi[(a^* - m'') / \sqrt{\{v''(\mathbf{n})\}}] = k_0 / (k_0 + k_1), \tag{18}$$

where $\Phi(u)$ is the unit normal distribution function. Also it has been shown that

$$\min_{\mathbf{a} \ \mu | \bar{\mathbf{X}}, \mathbf{n}} E'' l(a, \mu) = (k_0 + k_1) \{v''(\mathbf{n})\}^{\frac{1}{2}} \phi(u^*),$$

where $\phi(u)$ is the unit normal density function and $u^* = (a^* - m'')/\sqrt{\{v''(\mathbf{n})\}}$. Since a^* is determined so that (18) holds, u^* is independent of \mathbf{n} and \bar{X} and since $v''(\mathbf{n})$ does not depend on \bar{X} , it follows that \mathbf{n}^* is chosen to minimize

$$f_i\{v''(\mathbf{n})\} + \mathbf{c}\mathbf{n}^t = k\{v''(\mathbf{n})\}^{\frac{1}{2}} + \mathbf{c}\mathbf{n}^t,$$

where $k = (k_0 + k_1)\phi(u^*)$. By Theorem 1, \mathbf{n}^* is the solution to the allocation problem, $\mathbf{n}^0(C^0)$, where C^0 minimizes $k\{v^*(C)\}^{\frac{1}{2}} + C$. This function of C is convex, thus either $C^0 = 0$ or there is one and only one interval I_i for which the first-order condition (15) holds for some unique $C \in I_i$ which is the sought C^0 . This condition is simply

$$\{\gamma_i/(C + \Sigma_i)\}^2 = (2/k)\{\gamma_i^2/(C + \Sigma_i) + K_i\}^{\frac{1}{2}},$$

which can be numerically solved for C .

6. A TWO-ACTION PROBLEM

6.1. Definition

As a final example, consider a two-action problem where the utilities of each act, a_1 and a_2 , are linear in the unknown parameter μ . The model put forth is a generalization of that first given by Grundy *et al.* (1956), in the sense that here the underlying parameter is the linear composite, (1); independent samples may be selected to obtain information on each component μ_i , and a general k -dimensional normal prior distribution is assigned to μ . The simple sampling model of Grundy *et al.* is also given in detail in Raiffa and Schlaifer (1961); the notation and results of these latter authors are extensively used in this section. The problem, hereafter termed P_1 , which is discussed by Raiffa and Schlaifer is essentially described as follows: μ is the mean of a normal process with known variance σ^2 . A normal prior with mean m' and variance v' is assigned to the scalar μ . A sample of n observations X_1, \dots, X_n may be drawn where, conditional on μ , the X_i 's are independent $N(\mu, \sigma^2)$. Then with the utility function below they obtain the optimal scalar n^* .

It is assumed that the utilities of the two acts are given by

$$u(a_i, \mu) = K_i + k_i\mu \quad (i = 1, 2),$$

where it is assumed that $k_2 > k_1$. It is then clear that

$$a^* = \begin{cases} a_1 & \text{if } m'' \leq \mu_b, \\ a_2 & \text{if } m'' > \mu_b, \end{cases}$$

where $\mu_b = (K_1 - K_2)/(k_2 - k_1)$. Defining opportunity loss

$$l(a_i, \mu) \equiv \max_i u(a_i, \mu) - u(a_i, \mu),$$

one has the non-negative loss function

$$l(a_1, \mu) = \begin{cases} 0 & (\mu \leq \mu_b), \\ (k_2 - k_1)(\mu - \mu_b) & (\mu > \mu_b), \end{cases}$$

and

$$l(a_2, \mu) = \begin{cases} (k_2 - k_1)(\mu_b - \mu) & (\mu \leq \mu_b), \\ 0 & (\mu > \mu_b). \end{cases}$$

It then follows easily, since the distribution theory is essentially the same as under P_1 (see Chapter 5 of Raiffa and Schlaifer), that

$$\min_{\mathbf{x} | \mathbf{n}} E'' l(a_i, \mu) = k_i \{v''(\mathbf{n})\}^{\frac{1}{2}} L_{N^*} [|(\mu_b - m'') / \sqrt{\{v''(\mathbf{n})\}} |],$$

where

$$k_i = k_2 - k_1 \quad \text{and} \quad L_{N^*}(u) \equiv \int_{-\infty}^u (u-z) \phi(z) dz,$$

the so-called unit normal linear loss integral. Further, it can be shown (Chapter 5 of Raiffa and Schaifer) that

$$f_i \{v''(\mathbf{n})\} + \mathbf{c}\mathbf{n}' = E \min_{\mathbf{x} | \mathbf{n}} E'' l(a_i, \mu) + \mathbf{c}\mathbf{n}' = k_i \{v'\}^{\frac{1}{2}} L_{N^*} \{ |(\mu_b - m') / \sqrt{v'} | \} - k_i \{v' - v''(\mathbf{n})\}^{\frac{1}{2}} L_{N^*} [|(\mu_b - m') / \sqrt{\{v' - v''(\mathbf{n})\}} |] + \mathbf{c}\mathbf{n}'. \quad (19)$$

From this expression and Theorem 1 it follows that, given solutions to the allocation problem, \mathbf{n}^* which minimizes (19) is essentially given once that value, C^0 , of $C \geq 0$ is found which maximizes

$$h(C) = k_i \{v' - v^*(C)\}^{\frac{1}{2}} L_{N^*} [|(\mu_b - m') / \sqrt{\{v' - v^*(C)\}} |] - C. \quad (20)$$

Methodology for finding C^0 can be stated simply, although the demonstration is somewhat lengthy. The final results are stated in Section 6.3, and established by detailed examination of the behaviour of $h(C)$ given below.

6.2. Behaviour of $h(C)$

In examining the behaviour of $h(C)$ one may exploit the fact that in each of the intervals I_i , $h(C)$ mimics the behaviour of a positive linear function of the "dimensionless net gain". This dimensionless net gain is the quantity maximized by the choice of the scalar experiment size in the problem P_1 .

For $C \in I_i$, let

$$v^{**}(C) \equiv v' - v^*(C) = v' - \{K_i + \gamma_i^2 / (C + \Sigma_i)\} > 0, \quad (21)$$

where $v' = \pi V' \pi'$, this quantity being just the variance of the prior distribution of m'' , using the optimal allocation for the budget C . Also, letting

$$K'_i \equiv v' - K_i, \quad (22)$$

$$\rho_i \equiv \{K'_i(C + \Sigma_i) - \gamma_i^2\} / \gamma_i^2, \quad (23)$$

$$D'_i \equiv |(\mu_b - m')| / (K'_i)^{\frac{1}{2}}, \quad (24)$$

$$\lambda_i \equiv k_i \gamma_i \{ (K'_i)^{\frac{1}{2}} / \gamma_i \}^3 \quad (25)$$

and

$$\theta_i \equiv \{ \rho_i / (1 + \rho_i) \}^{\frac{1}{2}} = [\{ K'_i(C + \Sigma_i) - \gamma_i^2 \} / \{ K'_i(C + \Sigma_i) \}]^{\frac{1}{2}}, \quad (26)$$

it is easily verified that for $C \in I_i$

$$h(C) = (\gamma_i^2 / K'_i) \{ \lambda_i \theta_i L_{N^*}(D'_i / \theta_i) - \rho_i \} + \Sigma_i - (\gamma_i^2 / K'_i). \quad (27)$$

Furthermore, from (21) $K'_i > 0$ and $K'_i(C + \Sigma_i) - \gamma_i^2 > 0$. Also for $C \in I_i$, $h(C)$ depends on C only through ρ_i which determines C uniquely. The quantity in curly brackets

above is exactly of the form of the "dimensionless net gain", which when maximized by choice of ρ_i yields the optimal simple sample size for the problem P_1 . (See equation 5-41 of Raiffa and Schlaifer.)

In the form (27) the first two derivatives of $h(C)$ are given for $C \in I_i$ by

$$h'(C) = \frac{1}{2} \lambda_i \rho_i^{-1} (\rho_i + 1)^{-1} \phi(D'_i/\theta_i) - 1 \quad (28)$$

and

$$h''(C) = (K'_i/4\gamma_i^2) \lambda_i \{\rho_i(\rho_i + 1)\}^{-1} \phi(D'_i/\theta_i) \{D_i'^2 + (D_i'^2 - 1) \rho_i - 4\rho_i^2\}, \quad (29)$$

respectively. Using the definitions above and after some manipulation these derivatives may be expressed as

$$h'(C) = (k_i/2) \{v^{**}(C)\}^{-1} \{\gamma_i/(C + \Sigma_i)\}^2 \phi(u) - 1 \quad (30)$$

and

$$h''(C) = (k_i/4) \{v^{**}(C)\}^{-1} \{\gamma_i/(C + \Sigma_i)\}^2 \phi(u) [(u^2 - 1) \{\gamma_i^2/(C + \Sigma_i)\} - 4v^{**}(C)/(C + \Sigma_i)] \quad (31)$$

for $C \in I_i$ and $u = |\mu_b - m'| / \{v^{**}(C)\}^{\frac{1}{2}}$. In this form it is easy to see that since $v^{**}(C)$ and $\{\gamma_i/(C + \Sigma_i)\} = -dv^{**}(C)/dC$ are continuous functions of C over all $C \geq 0$, $h'(C)$ is a continuous function of C , while $h''(C)$ has jump discontinuities at the end-points of the successive intervals, attributed to the last term in (31).

Following the lead of Raiffa and Schlaifer, the behaviour of $h(C)$ in a special case may be seen easily.

Case I ($\mu_b = m'$). In this case it follows that $D'_i = 0$ for all i , and from the definition of $h(C)$, (28) and (29), it is seen that $h(0) = 0$; as $C \rightarrow 0$, $h'(C) \rightarrow \infty$; as $C \rightarrow \infty$, $h'(C) \rightarrow -1$; and $h''(C) < 0$ for all $C \geq 0$. It is then concluded that $h(C)$ is a concave function of C and possesses a unique maximum for some $C > 0$. Typical behaviour in this case is exactly the same as that of the net gain for simple sampling, the problem P_1 , displayed in Fig. 5.7, p. 119, of Raiffa and Schlaifer.

Case II ($\mu_b \neq m'$). The behaviour of $h(C)$ in this case differs substantially from that of the dimensionless net gain under simple sampling, for, unlike the situation for P_1 , $h(C)$ may possess several relative maxima. It is easily seen that for $C \in I_r$ and as $C \rightarrow 0$, $h'(C) \rightarrow -1$ and for $C \in I_0$, as $C \rightarrow \infty$, $h'(C) \rightarrow -1$. Thus $h(C)$ is decreasing at the origin and also for sufficiently large C .

For $C \in I_i$ ($i = 0, 1, \dots, r$) the sign of $h''(C)$ is dictated by the sign of the term in square brackets in (31). That term may be written as

$$\gamma_i^2 / \{(C + \Sigma_i)^3 v^{**}(C)\} \{- (4K_i'^2 / \gamma_i^2) (C + \Sigma_i)^2 + (\mu_*^2 + 7K_i') (C + \Sigma_i) - 3\gamma_i^2\}, \quad (32)$$

where $\mu_* = |\mu_b - m'|$. Thus the sign of $h''(C)$ for $C \in I_i$ is given by that of the term in the second set of curly brackets in (32), which is positive for $C_i^m < C < C_i^M$ and zero or negative elsewhere where C_i^m and C_i^M are the roots of the quadratic in (32). These roots are always real and for each i the smaller root, C_i^m , falls outside the interval $I_i = [C_i, C_{i-1})$. This may be seen as follows: By solving the quadratic in (32), $C_i^m < C_i$ whenever

$$(C_i + \Sigma_i) 8K_i'^2 (1/\gamma_i^2) > f(\mu_*^2) \equiv (\mu_*^2 + 7K_i') - \{(\mu_*^2 + 7K_i')^2 - 48K_i'^2\}^{\frac{1}{2}}.$$

Now $f(\mu_*^2) \leq f(0) = 6K_i'$ for all $\mu_*^2 \geq 0$, and further, by (21) and (22),

$$\gamma_i^2 / (C_i + \Sigma_i) \leq K_i' < 4K_i' / 3$$

which establishes the desired inequality. It then follows that if

- (a) $C_i^M < C_i$ then $h''(C) < 0$ for all $C \in I_i$,
- (b) $C_i^M > C_{i-1}$ then $h''(C) > 0$ for all $C \in I_i$,
- (c) $C_i < C_i^M < C_{i-1}$ then $h''(C) \begin{cases} > 0 \\ = 0 \\ < 0 \end{cases}$ for $C \begin{cases} < \\ = \\ > \end{cases} C_i^M$ ($C \in I_i$). (33)

If C_i^M equals C_i or C_{i-1} then $h''(C)$ is less than zero or greater than zero, respectively, for all $C \in I_i$, except that it is equal to zero at C_i or C_{i-1} , respectively.

Thus $h(C)$ may be concave throughout I_i (a), or convex (b), or first convex and then concave (c). The conjecture that perhaps there exists some i^* such that (a) holds for all I_i ($i < i^*$) (c) holds for I_{i^*} , while (b) holds for $i > i^*$ is false. The numerical example below demonstrates this. Thus $h(C)$ may possess several relative maxima. However, it is clear that within each I_i , $h(C)$ behaves like some portion of the dimensionless net gain function for P_1 discussed by Raiffa and Schlaifer.

6.3. Use of Nomographs in Finding C^0

A procedure may now be given for finding C^0 which maximizes $h(C)$, using nomographs given in Raiffa and Schlaifer.

Case I ($\mu_b = m'$). In this case $D'_i = 0$ for all I_i , and by the properties of $h(C)$ demonstrated above C^0 is the unique C satisfying $h'(C) = 0$ for $C \in I_i$. Using published nomographs one can easily find ρ_i^0 such that $h'(C)$, (28), is zero. The corresponding C_i^0 is, by (23), given by $C_i^0 = (1/K'_i) \gamma_i^2(\rho_i^0 + 1) - \Sigma_i$. C^0 is that unique C_i^0 falling in its respective interval I_i . To carry out this explicitly let $z = \lambda_i^{\frac{1}{2}}$, $D'_\infty \equiv D'_i = 0$, and find the corresponding η_0 from Chart I of Raiffa and Schlaifer; then take $\rho_i^0 = \eta_0 z^2$.

Case II ($\mu_b \neq m'$). Here D'_i is not zero for each i , and it is clear from the preceding discussion that $h(C)$ may possess a relative maximum only in those intervals, I_i , for which $h(C)$ has a concave segment or in which $h''(C) < 0$ for some C 's. This set of intervals, S , is easily identified using (33) and the fact that

$$C_i^M = [(\mu_*^2 + 7K'_i) + \{(\mu_*^2 + 7K'_i)^2 - 48K_i'^2\}^{\frac{1}{2}}] / \{8K_i'^2(1/\gamma_i^2)\} - \Sigma_i. \tag{34}$$

It follows by analogy with the case treated by Raiffa and Schlaifer that $h(C)$ attains a relative maximum in I_i if and only if $I_i \in S$ and the largest root of $h'(C) = 0$ (C_i^0 , say) is in I_i . Again C_i^0 may be given by $(1/K'_i) \gamma_i^2(\rho_i^0 + 1) - \Sigma_i$, where ρ_i^0 is the largest root obtained by equating (28) to zero. Again ρ_i^0 may be found for each interval using Chart I of Raiffa and Schlaifer. Simply take $D_\infty = D'_i$ and proceed as in Case I. Then C^0 defining the absolute maximum of $h(C)$ is that C_i^0 for which $h(C_i^0)$ is largest, provided $h(C_i^0) > 0$. Otherwise $C^0 = 0$, including the case of no relative maxima existing.

6.4. A Numerical Example

To conclude, consider the following numerical example chosen merely to illustrate the method.

Suppose $\mu = 3\mu_1 + 7\mu_2$, where $\mu = (\mu_1, \mu_2)$ is assigned a bivariate normal prior distribution with mean $m' = (10, 10)$ and variance-covariance matrix

$$V' = \begin{bmatrix} 300 & -150 \\ -150 & 100 \end{bmatrix}.$$

Suppose also that $\sigma_1^2 = 100$, $\sigma_2^2 = 256$, $c_1 = 1$ and $c_2 = 4$. Two acts, a_1 and a_2 , are contemplated; the utility of each is linear in μ with $K_1 = 275,500$, $K_2 = 100,000$, $k_1 = 1,000$ and $k_2 = 2,350$. Then $\mu_b = 130$, $k_i = 1,350$, $m' = 100$, $v' = 1,300$ and $|\mu_b - m'| = 30$. Using the results of the author (1965), the solution to the allocation problem of minimizing $v''(\mu)$ by choice of $n_1 \geq 0$, $n_2 \geq 0$ subject to $n_1 + 4n_2 \leq C$ was found to be as summarized in Table 1.

TABLE 1

i	I_i	γ_i	Σ_i	K_i	K'_i	$n_1^0(C)$	$n_2^0(C)$
3	[0, 0.438)	5	0.33	1225.0	75.0	C	0
2	[0.438, 3.271)	194	29.49	0.0	1300.0	$0.506 - 0.155C$	$-0.127 + 0.289C$
1	[3.271, 10.382)	80	10.24	675.0	625.0	0	$0.25C$
0	[10.382, ∞)	254	55.09	0.0	1300.0	$-1.226 + 0.118C$	$0.308 + 0.221C$

Using (33) and (34) one finds $C_0^M = 13.36$, $C_1^M = 6.73$, $C_2^M = 10.44$ and $C_3^M = 1.20$; hence $h(C)$ is convex for $C \in I_3$ and $C \in I_2$; for $C \in I_1$, $h(C)$ is convex for $C \leq 6.73$ and concave otherwise; while for $C \in I_0$, $h(C)$ is convex for $C \leq 13.36$ and concave otherwise. To locate relative maxima, one need only check I_0 and I_1 . For I_1 , $D'_1 = 1.2$, $\lambda_1 = 3296$ and $z = \lambda_1^{\frac{1}{2}} = 14.88$. From Chart I of Raiffa and Schlaifer one finds $\eta_0 \approx 0.075$ and thus $\rho_1^0 \approx 16.61$ and $C_1^0 \approx 170 \notin I_1$. Thus there is no relative maximum in I_1 . Similarly for I_0 , $D'_0 = 0.83$, $\lambda_0 = 980.8$ and $z = 9.94$, from Chart I of Raiffa and Schlaifer $\eta_0 \approx 0.108$ from which $\rho_0^0 \approx 10.66$ and $C_0^0 \approx 524 \in I_0$. Thus $h(C)$ has a relative maximum in I_0 at $C = 524$. This being the only relative maximum and since $h(524) > 0$, $C^0 = 524$. Finally, using the formulae in the table above for $n^0(C)$ for $C \in I_0$, it is found that $n_1^* \approx 61$ and $n_2^* = 116$, defining the optimal experiment.

ACKNOWLEDGEMENTS

The author expresses his deepest appreciation to Professor Howard Raiffa for suggesting and encouraging this work. Final preparation of this paper was supported by the National Science Foundation under Grant NSF-GP-6008.

REFERENCES

ERICSON, W. A. (1965), Optimum stratified sampling using prior information. *J. Amer. Statist. Ass.*, 60, 750-771.
 — (1966). On the minimization of a certain convex function arising in applied decision theory. (Unpublished manuscript.)
 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.
 KUHN, H. W. and TUCKER, A. W. (1951). Non-linear programming. *Proc. 2nd Berkeley Symp. Math. Statist. & Prob.*, pp. 481-492.
 RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University.