Allocation of Sampling Effort Amongst Several Normal Populations

Roger R. Davidson and Daniel L. Solomon

BU-589-M* August, 1976

Abstract

A Bayesian decision theoretic approach is applied to the problem of allocating observations amongst several independent Normal populations when the goal is estimation of the means of those populations. Quadratic loss is assumed for estimation error, and the cost of sampling each population is assumed proportional to the number of observations taken. Natural conjugate prior distributions are employed. Optimal allocations are found for the case in which the total number of observations is predetermined and for that in which it is not.

---

*In the Mimeo Series of the Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, New York, 14853.

Allocation of Sampling Effort Amongst
Several Normal Populations

Roger R. Davidson and Daniel L. Solomon*

1. INTRODUCTION

This study is concerned with the problem of how one should allocate the sampling effort amongst independent Normal processes when the goal is estimation of the means of those processes. Such an inference objective occurs, for example, in the stratified sampling of a sociological population in which the mean value of some attribute within each subpopulation is of interest. We consider both cases in which the total number of observations on all strata is specified in advance and those in which it is not.

The formulation is Bayesian and decision theoretic with loss suffered for misestimation and cost incurred for sampling. We allow the cost of sampling to differ amongst the populations. The optimality criterion is the minimization of the total expected loss plus cost. Although the goal in this paper is estimation, the approach can be applied to other inference objectives, such as hypothesis testing or ranking and selection, by making appropriate definitions for loss and cost functions.

The first section following introduces the distribution theory and cost structure while the second provides a solution for the optimal sampling plan in a general setting. The third section then establishes explicit solutions for some important special cases, while the last shows how one of the distributional assumptions can be relaxed.

---

* Roger R. Davidson is associate professor, Department of Mathematics, University of Victoria, Victoria, B.C., V8W 2Y2. Daniel L. Solomon is associate professor, Biometrics Unit, Cornell University, Ithaca, N.Y., 14853. The work was sponsored in part by the National Research Council of Canada under grant NRC-A7166.

2. FORMULATION

We begin by introducing the distribution theory and cost structure for the problem.

2.1 The Sampling Distribution

We wish to make inferences about the unknown means of k univariate Normal distributions, $N(\mu_i,\sigma_i^2)$, $\sigma_i^2 > 0$, $i=1,2,\cdots,k$. We suppose that a predetermined number m of observations are to be taken $(m \geq k)$, and that the objective is to allocate those observations amongst the k populations. We write Y_i for the mean of n_i observations from the i^{th} distribution and $\underset{\sim}{Y} = (Y_1,Y_2,\cdots,Y_k)'$, so that for given $\underset{\sim}{\mu}$, $\underset{\sim}{Y}$ has the k-variate Normal distribution

$$N_k(\underset{\sim}{\mu},\underset{\sim}{\Sigma}\underset{\sim}{N}^{-1})$$

where $\underset{\sim}{\mu} = (\mu_1,\mu_2,\cdots,\mu_k)'$, $\underset{\sim}{\Sigma}$ = diagonal (σ_i^2), $\underset{\sim}{N}$ = diagonal (n_i).

2.2 The Prior and Posterior Distributions

Suppose that prior information regarding $\underset{\sim}{\mu}$ can be adequately represented by a natural conjugate distribution, i.e., $\underset{\sim}{\mu} \sim N_k(\underset{\sim}{\mu}_0,\underset{\sim}{\Sigma}_0)$. Having observed $\underset{\sim}{Y} = \underset{\sim}{y}$, the posterior distribution for $\underset{\sim}{\mu}$ is $N_k(\underset{\sim}{\mu}_1,\underset{\sim}{\Sigma}_1)$ where

$$\underset{\sim}{\mu}_1 = (\underset{\sim}{N}\underset{\sim}{\Sigma}^{-1} + \underset{\sim}{\Sigma}_0^{-1})^{-1}(\underset{\sim}{N}\underset{\sim}{\Sigma}^{-1}\underset{\sim}{y} + \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{\mu}_0) \quad \text{and} \quad \underset{\sim}{\Sigma}_1 = (\underset{\sim}{N}\underset{\sim}{\Sigma}^{-1} + \underset{\sim}{\Sigma}_0^{-1})^{-1} .$$

As is often the case in problems such as these, the appeal of using a prior distribution in the family conjugate to the sampling distribution is that it allows for a reasonably tractable mathematical development. In a specific experimental context it is crucial to determine whether or not the prior knowledge of the investigator can be realistically represented by a member of that family.

In the present case the conjugate family is moderately rich, apart from symmetry, and has parameters which are capable of interpretation. The prior mean μ_0 represents the investigator's "best guess" for μ and Σ_0^{-1} is a measure of the confidence he has in his beliefs and of his beliefs about interrelationships among the components of μ.

2.3 The Loss Frunction

The loss associated with estimating μ by $\hat{\mu}$ is taken to be $(\mu-\hat{\mu})'A(\mu-\hat{\mu})$ where A is a $k \times k$ positive definite and symmetric matrix. The Bayes estimator is then μ_1 and its Bayes risk is trace $A\Sigma_1$.

As another feature of the appeal of the natural conjugate family, we note here that the results derived in the next section can be easily extended to the case in which the loss function is multiplied by

$$\exp\left\{-\frac{\alpha}{2}(\mu-\beta)'\Gamma(\mu-\beta)\right\}$$

for arbitrary β, $\alpha > 0$ and non-negative definite symmetric matrix Γ. Some of the results of the final two sections can be similarly generalized. We choose to retain the original quadratic loss for simplicity of presentation.

2.4 The Cost of Sampling

If the cost of making an observation on the i^{th} population is $d_i \geq 0$, then the total cost of conducting the experiment determined by the allocation $n = (n_1, n_2, \cdots, n_k)$ is $d'n$ where $d = (d_1, d_2, \cdots, d_k)'$. Note that we can write $d'n = \Sigma_{i=1}^k d_i n_i = \Sigma_{i=1}^k \delta_i n_i + \bar{d}m$ where $\bar{d} = \frac{1}{k}\Sigma_{i=1}^k d_i$, $\delta_i = d_i - \bar{d}$ and $\Sigma_{i=1}^k n_i = m$. Note too that $\Sigma_{i=1}^k \delta_i = 0$ and $n_k = m - n_1 - n_2 - \cdots - n_{k-1}$.

Finally, we assume that the loss matrix $\underset{\sim}{A}$ and the sampling costs $\underset{\sim}{d}$ are scaled in such a way that loss and cost are additive. In this event, the total Bayes risk $B(\underset{\sim}{n})$ associated with allocating m observations according to $\underset{\sim}{n}$, and estimating $\underset{\sim}{\mu}$ by the Bayes rule $\underset{\sim}{\mu}_1$ is

$$B(\underset{\sim}{n}) = \mathrm{trace}(\underset{\sim}{A}\underset{\sim}{\Sigma}_1) + \underset{\sim}{d}'\underset{\sim}{n} \ .$$

The problem of optimal allocation is that of choosing $\underset{\sim}{n}$ to minimize $B(\underset{\sim}{n})$ subject to $\Sigma_{i=1}^{k} n_i = m$.

3. GENERAL SOLUTION

To determine the optimal allocation we treat the total Bayes risk $B(\underset{\sim}{n})$ as a function continuous in $\underset{\sim}{n}$. The $\underset{\sim}{n}^0$ which minimizes $B(\underset{\sim}{n})$ will then not necessarily have integral components. However, because of the convexity of $B(\underset{\sim}{n})$ the optimal $\underset{\sim}{n}$ can be determined by testing vectors of integers adjacent to $\underset{\sim}{n}^0$.

Writing $\underset{\sim}{H}_0 = (h_{ij}^0) = \underset{\sim}{\Sigma}_0^{-1}$ and $\underset{\sim}{H}_1 = \underset{\sim}{\Sigma}_1^{-1} = (\underset{\sim}{N}\underset{\sim}{\Sigma}^{-1} + \underset{\sim}{H}_0)$ and recalling that $\underset{\sim}{d}'\underset{\sim}{n} = \Sigma_{i=1}^{k-1}\delta_i n_i + \delta_k(m-n_1-n_2-\cdots-n_{k-1}) + \bar{d}_m$, we have that the partial derivatives of $B(\cdot)$ are

$$\frac{\partial B(\underset{\sim}{n})}{\partial n_i} = \mathrm{tr}\ \underset{\sim}{A}\ \frac{\partial \underset{\sim}{H}_1^{-1}}{\partial n_i} + (\delta_i - \delta_k)$$

$$= -\ \mathrm{tr}\ \underset{\sim}{A}\underset{\sim}{H}_1^{-1}\left(\frac{\partial \underset{\sim}{H}_1}{\partial n_i}\right)\underset{\sim}{H}_1^{-1} + (\delta_i - \delta_k)\ ; \quad i=1,2,\cdots,k-1 \ .$$

But $\underset{\sim}{N}\underset{\sim}{\Sigma}^{-1}$ is a diagonal matrix with entries n_i/σ_i^2 for $i=1,2,\cdots,k-1$ and last entry $(m-n_1-n_2-\cdots-n_{k-1})/\sigma_k^2$. Therefore $\partial\underset{\sim}{H}_1/\partial n_i$ is a diagonal matrix,

$\underset{\sim}{D}_i = \text{diag}\,[0,\cdots,0,1/\sigma_i^2,0,\cdots,0,-1/\sigma_k^2]$ with zeros in all but the i^{th} and k^{th} positions. We have as a condition for the minimizing $\underset{\sim}{n}$,

$$\delta_i - \text{tr}\,\underset{\sim}{A}\underset{\sim}{\Sigma}_1\underset{\sim}{D}_i\underset{\sim}{\Sigma}_1 = \delta_k\ ; \quad i=1,2,\cdots,k-1\ .$$

But

$$\text{tr}\,\underset{\sim}{A}\underset{\sim}{\Sigma}_1\underset{\sim}{D}_i\underset{\sim}{\Sigma}_1 = \text{tr}\,\underset{\sim}{D}_i\underset{\sim}{\Sigma}_1\underset{\sim}{A}\underset{\sim}{\Sigma}_1 = (1/\sigma_i^2)(\underset{\sim}{\Sigma}_1\underset{\sim}{A}\underset{\sim}{\Sigma}_1)_{ii} - (1/\sigma_k^2)(\underset{\sim}{\Sigma}_1\underset{\sim}{A}\underset{\sim}{\Sigma}_1)_{kk}$$

and so the condition may be written

$$\delta_i - (1/\sigma_i^2)(\underset{\sim}{\Sigma}_1\underset{\sim}{A}\underset{\sim}{\Sigma}_1)_{ii} = \delta_k - (1/\sigma_k^2)(\underset{\sim}{\Sigma}_1\underset{\sim}{A}\underset{\sim}{\Sigma}_1)_{kk}\ ; \quad i=1,2,\cdots,k-1\ , \tag{1}$$

or equivalently

$$(1/\sigma_i^2)(\underset{\sim}{\Sigma}_1\underset{\sim}{A}\underset{\sim}{\Sigma}_1)_{ii} - \delta_i = c_m\ ; \qquad i=1,2,\cdots,k\ , \tag{2}$$

where c_m is a constant chosen so that $\Sigma n_i = m$.

We note without proof, that the matrix of second partial derivatives, $\left(\frac{\partial B(\underset{\sim}{n})}{\partial n_j \partial n_i}\right)$, is positive definite. This establishes the minimality of (2) and the asserted convexity of $B(\cdot)$.

Notice that if the minimization were not constrained by $\Sigma n_i = m$, then equation (2) would be replaced by

$$\frac{1}{\sigma_i^2}\,(\underset{\sim}{\Sigma}_1\underset{\sim}{A}\underset{\sim}{\Sigma}_1)_{ii} - d_i = 0\ , \qquad i=1,2,\cdots,k\ , \tag{3}$$

where δ_i in (2) is replaced by d_i in (3). Now if $\underset{\sim}{n}^0$ is a solution to (2) and $\underset{\sim}{n}^*$ a solution to (3), and if $\underset{\sim}{\Sigma}_1^0$ and $\underset{\sim}{\Sigma}_1^*$ denote the values of $\underset{\sim}{\Sigma}_1$ at $\underset{\sim}{n}_0$ and $\underset{\sim}{n}^*$ respectively, then

$$\frac{1}{\sigma_i^2}\left(\boldsymbol{\Sigma}_1^0 \mathbf{A} \boldsymbol{\Sigma}_1^0\right)_{ii} - d_i = c_m - \bar{d} \,, \qquad (4)$$

$$\frac{1}{\sigma_i^2}\left(\boldsymbol{\Sigma}_1^* \mathbf{A} \boldsymbol{\Sigma}_1^*\right)_{ii} - d_i = 0 \,,$$

and the similarity of the two problems is apparent. If, furthermore, the unit sampling costs are the same for all populations, then $d_i = \bar{d} \equiv d$, say, and the optimality equations (4) become

$$\left(\boldsymbol{\Sigma}_1^0 \mathbf{A} \boldsymbol{\Sigma}_1^0\right)_{ii} = \sigma_i^2 c_m$$

and $\qquad i=1,2,\cdots,k$.

$$\left(\boldsymbol{\Sigma}_1^* \mathbf{A} \boldsymbol{\Sigma}_1^*\right)_{ii} = \sigma_i^2 d$$

Thus the relationship between the constrained and unconstrained optimal sampling schemes is that

$$\left(\boldsymbol{\Sigma}_1^0 \mathbf{A} \boldsymbol{\Sigma}_1^0\right)_{ii} / \left(\boldsymbol{\Sigma}_1^* \mathbf{A} \boldsymbol{\Sigma}_1^*\right)_{ii} = c_m/d$$

is constant for all $i=1,2,\cdots,k$. This relationship will be explored further in what follows.

4. SOME IMPORTANT CASES

Although the equations (2) may be difficult to solve in general, they simplify in some important special cases. For example, if A is diagonal, then (2) becomes

$$(1/\sigma_i^2)\sum_j a_{jj} \sigma_{ij}^{1\;2} - \delta_i = \text{const.} \qquad i=1,2,\cdots,k \,,$$

where we have denoted by σ^1_{ij} the entries of the posterior covariance matrix Σ_1. Two other cases in which the simplification is substantial are pursued next. They are (1) prior independence of $\mu_1, \mu_2, \cdots, \mu_k$ and (2) k = 2.

4.1 Prior Independence

Although it should not be made lightly, an assumption sometimes tenable, at least approximately, is that prior information has been collected in such a way that $\mu_1, \mu_2, \cdots, \mu_k$ are a priori independent. In this formulation, the prior covariance matrix Σ_0 is then diagonal and so therefore is the posterior covariance matrix

$$\Sigma_1 = \text{diag.} \left[\left(\frac{n_i}{\sigma_i^2} + h_i^0 \right)^{-1} \right] ,$$

where we have written h_i^0 for the reciprocal prior variance of μ_i.

Equations (2) become for arbitrary loss matrix A,

$$\frac{1}{\sigma_i^2} a_{ii} \left[\left(\frac{n_i}{\sigma_i^2} + h_i^0 \right)^{-2} \right] - \delta_i = c_m \qquad i=1,2,\cdots,k ,$$

or

$$n_i + \sigma_i^2 h_i^0 = [a_{ii}\sigma_i^2/(\delta_i+c_m)]^{\frac{1}{2}} ; \qquad i=1,2,\cdots,k . \tag{5}$$

Note that the equations depend on A only through its diagonal elements. Now summing over i in (5), and recalling that $\Sigma_{i=1}^k n_i = m$, we have an implicit solution for c_m,

$$m + \sum_{i=1}^{k} \sigma_i^2 h_i^0 = \sum_{i=1}^{k} [a_{ii}\sigma_i^2/(\delta_i+c_m)]^{\frac{1}{2}} . \tag{6}$$

4.1.1 Equal Sampling Costs

We shall discuss the general solution of (6) for c_m shortly, but first observe that if the cost associated with making an observation on each population is constant over populations ($d_i \equiv \bar{d}$; $i=1,2,\cdots,k$), then the deviations $\delta_i \equiv 0$; $i=1,2,\cdots,k$ and the value of c_m satisfying (6) is then

$$c_m^0 = \left[\sum_{j=1}^{k} (a_{jj}\sigma_j^2)^{\frac{1}{2}} \Big/ (m + \sum_{j=1}^{k} \sigma_j^2 k_j^0) \right]^2 . \tag{7}$$

Substituting in (5) with $\delta_i = 0$, we conclude that if sampling costs are equal and the μ_i are a priori independent, then the optimal allocation is $\underset{\sim}{n}^0 = (n_1^0, n_2^0, \cdots, n_k^0)'$, where

$$n_i^0 = (a_{ii}\sigma_i^2)^{\frac{1}{2}}(m + \sum_{j=1}^{k} \sigma_j^2 h_j^0) \Big/ \sum_{j=1}^{k} (a_{jj}\sigma_j^2)^{\frac{1}{2}} - \sigma_i^2 h_i^0 .$$

Notice that if the a_{ii} are all equal, so that the unit misestimation loss is the same for each μ_i, then n_i^0 does not depend on their common value. If, furthermore, the sampling and prior variances are constant ($\sigma_i^2 = \sigma^2$, $h_i^0 = h^0$; $i=1,2,\cdots,k$) then $n_i^0 = m/k$ and equal allocation is optimal (assuming that k divides m). Also apparent is that for given $\sigma_1^2, \sigma_2^2, \cdots, \sigma_k^2$, n_i^0 is increasing in a_{ii} and decreasing in h_i^0. That is, (relatively) large unit misestimation loss or prior variance associated with a population favors a (relatively) large sample from that population.

It is instructive here to again compare the optimal allocation, $\underset{\sim}{n}^0$ of the m observations with the optimal sampling scheme $\underset{\sim}{n}^*$ had the constraint $\Sigma_{i=1}^k n_i = m$ not been imposed. In the case of prior independence of the μ_i, the unconstrained formulation reduces to k separate problems, with loss functions $a_{ii}(\hat{\mu}_i - \mu_i)^2$. Still assuming equal costs, $d_i \equiv d$, of sampling, equations (4) become

$$\left(\frac{n_i^0}{\sigma_i^2} + h_i^0 \right)^{-1} = \left(\frac{\sigma_i^2}{a_{ii}} \right)^{\frac{1}{2}} c_m^{\frac{1}{2}}$$

and

$$\left(\frac{n_i^*}{\sigma_i^2} + h_i^0 \right)^{-1} = \left(\frac{\sigma_i^2}{a_{ii}} \right)^{\frac{1}{2}} d^{\frac{1}{2}} \quad .$$

Observe that the left hand sides of these equations are the respective posterior variances of the μ_i under the two sampling schemes and so it is their ratio, $(c_m/d)^{\frac{1}{2}}$ which is constant for all $i=1,2,\cdots,k$.

Note that the ratio n_i^0/n_i^* is <u>not</u> constant over i so that the optimal allocation $\underset{\sim}{n}^0$ is <u>not</u> obtained from the unconstrained solution $\underset{\sim}{n}^*$ of the component problems by simply scaling the n_i^* to add to m. That is, in general

$$n_i^0 \neq mn_i^*/\Sigma n_j^* \quad .$$

4.1.2 Not Necessarily Equal Sampling Costs

Dropping the assumption that $\delta_i = 0$; $i=1,2,\cdots,k$, we return to a discussion of the solution to (6) for c_m, which when substituted into (5) completely determines the optimal allocation in the case of prior independence.

Writing

$$g(c_m) = \sum_{i=1}^{k} [a_{ii}\sigma_i^2/(\delta_i + c_m)]^{\frac{1}{2}} - M$$

where $M = m + \Sigma\sigma_i^2 h_i^0$, values of c_m satisfying (6) are exactly the zeros of $g(\cdot)$. Note that $g(\cdot)$ is defined for all $c_m + \delta_i > 0$; $i=1,2,\cdots,k$. That is, if we take $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_k$ (so that $\delta_1 < 0$), then the domain of $g(\cdot)$ is $(0 <) - \delta_1 < c_m < \infty$. On this domain we have

$$g'(c_m) = -\frac{1}{2} \sum_{i=1}^{k} (a_{ii}\sigma_i^2)^{\frac{1}{2}} (\delta_i + c_m)^{-\frac{3}{2}} < 0$$

so that $g(\cdot)$ decreases monotonically from $g(-\delta_1) = \infty$ to $g(\infty) = -M < 0$. Thus $g(c_m) = 0$ has exactly one root and this root is positive.

We next determine an interval containing this root and suggest an initial value for a numerical evaluation of c_m over this interval. Since $g(c_m) = 0$,

$$\sum_{i=1}^{k} [a_{ii}\sigma_i^2/(\delta_i + c_m)]^{\frac{1}{2}} = M > 0 \ , \tag{8}$$

and since each term in the sum if positive we have

$$\frac{M}{k} \le \max_i \, [a_{ii}\sigma_i^2/(\delta_i + c_m)]^{\frac{1}{2}} \le [\max_i \, (a_{ii}\sigma_i^2)/(\min_i \, (\delta_i) + c_m)]^{\frac{1}{2}}$$

and also

$$\frac{M}{k} \ge \min_i \, [a_{ii}\sigma_i^2/(\delta_i + c_m)]^{\frac{1}{2}} \ge [\min_i \, (a_{ii}\sigma_i^2)/(\max_i \, (\delta_i) + c_m)]^{\frac{1}{2}} \ .$$

Equivalently,

$$L_1 \equiv (k/M)^2 \min_i \, (a_{ii}\sigma_i^2) - \max_i \, (\delta_i) \le (k/M)^2 \max_i \, (a_{ii}\sigma_i^2) - \min_i \, (\delta_i) \equiv U \ . \tag{9}$$

It also follows from (8) that <u>for each</u> $i=1,2,\cdots,k$, $[a_{ii}\sigma_i^2/(\delta_i + c_m)]^{\frac{1}{2}} < M$, and so $(a_{ii}\sigma_i^2 - M^2\delta_i)/M^2 < c_m$. Summing over i and recalling that $\Sigma\delta_i = 0$, we find

$$L_2 \equiv (1/kM^2) \sum_{i=1}^{k} a_{ii}\sigma_i^2 < c_m \ .$$

Combining this result with (9) we have that

$$\max \, (L_1, L_2) \le c_m \le U \ .$$

Recall from (7), that when the unit observation costs are the same for all populations, the value of c_m which satisfies (6) is $c_m^0 = [\Sigma(a_{ii}\sigma_i^2)^{\frac{1}{2}}]^2/M^2$. By observing that $k \min_i (a_{ii}\sigma_i^2)^{\frac{1}{2}} \le \Sigma_i (a_{ii}\sigma_i^2)^{\frac{1}{2}} \le k \max_i (a_{ii}\sigma_i^2)^{\frac{1}{2}}$, that $\max_i \delta_i \ge 0$, $\min_i \delta_i \le 0$, and that $[\Sigma(a_{ii}\sigma_i^2)^{\frac{1}{2}}]^2 \ge \Sigma a_{ii}\sigma_i^2$; it can be shown that $\max(L_1, L_2) \le c_m^0 \le U$. Therefore, this easily calculated value will serve as an initial estimate for an iterative scheme which searches the interval for the zero of $g(\cdot)$.

4.2 Two Populations

In the case of allocating m observations among two populations, if n of the observations are taken from the first, then the total Bayes risk is minimized when n satisfies equation (1). Since k=2, this becomes

$$\delta_1 - \frac{1}{\sigma_1^2} (\underset{\sim}{\Sigma}_1 \underset{\sim}{A} \underset{\sim}{\Sigma}_1)_{11} = \delta_2 - \frac{1}{\sigma_2^2} (\underset{\sim}{\Sigma}_1 \underset{\sim}{A} \underset{\sim}{\Sigma}_1)_{22} \tag{10}$$

where now

$$(\underset{\sim}{\Sigma}_1 \underset{\sim}{A} \underset{\sim}{\Sigma}_1)_{11} = \Delta^{-2}\left\{a_{11}\left(\frac{m-n}{\sigma_2^2} + h_{22}^0\right)^2 - 2a_{12}h_{12}^0\left(\frac{m-n}{\sigma_2^2} + h_{22}^0\right) + a_{22}h_{12}^{0\,2}\right\},$$

$$(\underset{\sim}{\Sigma}_1 \underset{\sim}{A} \underset{\sim}{\Sigma}_1)_{22} = \Delta^{-2}\left\{a_{22}\left(\frac{n}{\sigma_1^2} + h_{11}^0\right)^2 - 2a_{12}h_{12}^0\left(\frac{n}{\sigma_1^2} + h_{11}^0\right) + a_{11}h_{12}^{0\,2}\right\},$$

and

$$\Delta = \det(\underset{\sim}{\Sigma}_1^{-1}) = \left(\frac{m-n}{\sigma_2^2} + h_{22}^0\right)\left(\frac{n}{\sigma_1^2} + h_{11}^0\right) - h_{12}^{0\,2} > 0 .$$

Substitution in (10) produces a quartic equation for n, which can then be solved by standard numerical polynomial methods. If however, as assumed earlier, the unit sampling costs are equal, then $\delta_1 = \delta_2 = 0$ and (10) becomes a quadratic equation in n,

$$s_{11}[(m'-y)^2 - t_{12}t_{21}] - 2s_{12}t_{21}[(m'-y) - y] + s_{22}[t_{12}t_{21} - y^2] = 0 \tag{11}$$

where we have put $y = n + t_{11}$ and

$$s_{ij} = a_{ij}\sigma_i^2\ ,\quad t_{ij} = h_{ij}^0\sigma_i^2\ ;\ i,j = 1,2; \quad \text{and} \quad m' = m + t_{11} + t_{22}\ .$$

The quadratic can of course be solved explicitly, and notice that we have not required prior independence or any special structure on the matrix $\underset{\sim}{A}$. However, in the special case (sometimes assumed for generalized least squares) in which $\underset{\sim}{A}$ is proportional to the inverse covariance matrix $\underset{\sim}{\Sigma}^{-1}$ of the sampling distribution, we have $s_{11} = s_{22}$ and $s_{12} = 0$. Equation (11) then becomes linear with solution $y = m'/2$, and the optimal allocation of the m observations is

$$n_1 = \frac{m}{2} + \frac{h_{22}^0\sigma_2^2 - h_{11}^0\sigma_1^2}{2}$$

and

$$n_2 = m - n_1 = \frac{m}{2} + \frac{h_{11}^0\sigma_1^2 - h_{22}^0\sigma_2^2}{2}\ .$$

Note that although this allocation does not depend on h_{12}^0, it does depend on the prior covariance between μ_1 and μ_2 through the elements h_{11}^0 and h_{22}^0 of $\underset{\sim}{\Sigma}_0^{-1}$.

5. SAMPLING VARIANCES UNKNOWN

In the formulation of the first section it was assumed that the covariance matrix $\underset{\sim}{\Sigma}\underset{\sim}{N}^{-1}$ = diagonal (σ_i^2/n_i) is known a priori. We here relax that assumption by requiring that only the relative precision of the independent processes be known. Thus we write $\sigma_i^2 = \sigma^2 s_i^2$, $i=1,2,\cdots,k$ where now $\underset{\sim}{S}$ = diagonal (s_i^2) must be known but σ^2 need not. For given $\underset{\sim}{\mu}$ and σ^2, the sampling distribution of $\underset{\sim}{Y}$ is now the k-variate Normal distribution, $N_k(\underset{\sim}{\mu},\sigma^2\underset{\sim}{S}\underset{\sim}{N}^{-1})$.

With both $\underset{\sim}{\mu}$ and σ^2 unknown, the natural conjugate family of prior distributions is the Normal-inverted gamma 2 (cf. Raiffa and Schlaifer [1]). That is, a priori, the conditional distribution of μ for given σ^2 is k-variate Normal, $N_k(\mu_0, \sigma^2 \cdot S_0)$, and the distribution of σ^2 is inverted gamma 2; i.e., the probability density of $h \equiv 1/\sigma^2$ is proportional to $h^{\upsilon/2-1}e^{-\upsilon vh/2}$ for $v > 0$, $\upsilon > 2$ specified parameters.

It follows that the posterior distribution of (μ, σ^2) having observed $Y = y$ is again Normal-inverted gamma 2 and, upon integrating the joint density with respect to σ^2, it can be shown that the marginal posterior distribution of μ for given y is k-variate Student (Raiffa and Schlaifer [1], p. 320). The first two moments are

$$\mu_1^* = E(\mu|y) = (NS^{-1} + S_0^{-1})^{-1}(NS^{-1}y + S_0^{-1}\mu_0)$$

$$\Sigma_1^* = \mathrm{Cov}(\mu|y) = (NS^{-1} + S_0^{-1})^{-1} v\upsilon/(\upsilon-2) \; .$$

Now the Bayes risk of an estimator $\hat{\mu}(Y)$ is

$$E_{\mu,Y,\sigma^2}\left(\mu - \hat{\mu}(Y)\right)' A\left(\mu - \hat{\mu}(Y)\right)$$

$$= E_{\mu,Y}E_{\sigma^2|\mu,Y}\left(\mu - \hat{\mu}(Y)\right)' A\left(\mu - \hat{\mu}(Y)\right)$$

$$= E_{\mu,Y}\left(\mu - \hat{\mu}(Y)\right)' A\left(\mu - \hat{\mu}(Y)\right) .$$

Thus the Bayes estimator is the mean of marginal (with respect to σ^2) posterior distribution of μ given y; i.e., is μ_1^* . The minimum Bayes risk is therefore $\mathrm{tr}\, A\Sigma_1^*$. But inasmuch as a priori, $E\sigma^2 = v\upsilon/(\upsilon-2)$, we may write the minimum Bayes risk as

$$\mathrm{tr}\, A\, [N(E\sigma^2 S)^{-1} + (E\sigma^2 S_0)^{-1}]^{-1}$$

which agrees with the value for the original formulation with $\underset{\sim}{\Sigma} = \sigma^2 \underset{\sim}{S}$ and $\underset{\sim}{\Sigma}_0 = \sigma^2 \underset{\sim}{S}_0$ replaced by their expectations under the prior distribution for σ^2. Thus the results of the preceding sections all extend upon making that substitution.

REFERENCES

[1] Raiffa, Howard and Schlaifer, Robert, *Applied Statistical Decision Theory*, Boston, MA.: Harvard Univ. Press, 1961.