Journal of Statistical Planning and Inference 28 (1991) 271-289
North-Holland

Designing experiments for selecting the largest normal mean when the variances are known and unequal: Optimal sample size allocation

Robert E. Bechhofer*

School of OR & IE, Cornell University, Ithaca, NY 14853, USA

Anthony J. Hayter

School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Ajit C. Tamhane**

Department of Statistics, Northwestern University, Evanston, IL 60208, USA

Received 10 August 1989; revised manuscript received 8 October 1990
Recommended by T.J. Santner

Abstract: We consider the problem of 'optimally' allocating a given total number, N, of observations to $k\geq 2$ normal populations having unknown means but known variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, when it is desired to select the population having the largest mean using a natural single-stage selection procedure based on sample means. Here 'optimal' allocation is one that maximizes the infimum of the probability of a correct selection (P(CS)) over the so-called preference zone of the parameter space (Bechhofer (1954)). The solution of this problem enables us to find the smallest possible N and the associated optimal allocation of the sample sizes, viz. $n_1, n_2, \ldots, n_k$ such that $\sum n_i = N$, required to guarantee a specified $\{\delta^*, P^*\}$ probability requirement. We prove that for $k\geq 3$, the allocation $n_i \propto \sigma_i^2$ (which is convenient to implement in practice) is locally (and for $k=3$, numerically checked to be globally) optimal iff $P^* \leq P_{\mathrm{L}}$ or $P^* \geq P_{\mathrm{U}}$, where P_{L} and P_{U} depend on the largest and the smallest relative variances, respectively. For $P_{\mathrm{L}} < P^* < P_{\mathrm{U}}$, the globally optimal allocation is found by numerical search for $k=3$ and found to be approximately given by $n_i \propto \sigma_i$, the allocation that is known to be globally optimal for $k=2$.

AMS Subject Classification: Primary 62F07; secondary 62K99.

Key words and phrases: Ranking and selection; indifference-zone approach; optimal allocation; normal populations.

* Research of this author was partially supported by the U.S. Army Research Office through the Mathematical Sciences Institute of Cornell University.

** This paper was completed while this author was at the University of Bath during the summer of 1988 on a Visiting Research Fellowship Grant from the British Science and Engineering Research Council.

1. Introduction and summary

Let Π_i denote a normal population with unknown mean μ_i and known variance σ_i^2 $(1 \leq i \leq k)$. Without loss of generality we label the populations so that $\sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_k^2$. To avoid trivialities, we assume that at least one of these inequalities is strict. Let Ω denote the space of all parameter points $\boldsymbol{\omega} = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$. Let $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$ denote the ordered values of the μ_i. We assume that the correct pairing of Π_i (and hence of σ_i^2) with $\mu_{[j]}$ $(1 \leq i, j \leq k)$ is completely unknown. The experimenter's *goal* is to select the population with mean $\mu_{[k]}$ (referred to as the 'best' population and assumed to be unique). If the decision procedure selects this population then a correct selection (CS) is said to have been made.

We adopt the *indifference-zone approach* of Bechhofer (1954) for this selection problem. In this approach, consideration is restricted to those procedures which guarantee the *probability requirement*:

$$\inf_{\Omega(\delta^*)} P(\text{CS}) \geq P^* \tag{1.1}$$

where

$$\Omega(\delta^*) = \{\boldsymbol{\omega} \in \Omega \mid \mu_{[k]} - \mu_{[k-1]} \geq \delta^*\} \tag{1.2}$$

is the so-called *preference zone* (complement of the *indifference zone* in Ω), and $\delta^* > 0$ and $P^* \in (1/k, 1)$ are prespecified constants.

Throughout this article we consider only the 'natural' single-stage selection procedure R, which takes independent random samples $\{X_{ij}\ (1 \leq j \leq n_i)\}$ from the Π_i $(1 \leq i \leq k)$ and selects the population that yields the largest sample mean, $\max_{1 \leq i \leq k} \bar{X}_i$, where $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$ $(1 \leq i \leq k)$. The globally optimal choice of the sample sizes n_i to guarantee the specified probability requirement (1.1) is the problem considered in the present article.

The optimization problem that we wish to solve is the following.

Exact Discrete Optimization Problem: For given $\boldsymbol{\sigma}^2$ and total sample size N, and specified δ^*, find the allocation $\boldsymbol{n} = (n_1, \ldots, n_k)$ which achieves

$$\max \inf_{\Omega(\delta^*)} P(\text{CS}) \tag{1.3}$$

where the max is taken over all allocation vectors $\boldsymbol{n}$ subject to $\sum_{i=1}^k n_i = N$; here the $n_i \geq 0$ are integer valued sample sizes to be used in the procedure R. We denote the solution to (1.3) by $\hat{\boldsymbol{n}} = (\hat{n}_1, \ldots, \hat{n}_k)$ and refer to it as the *globally optimal allocation*. (For the sake of conciseness, we will drop the prefix 'globally' from now on. Thus an allocation referred to simply as optimal will be understood to be globally optimal.) It is easy to see that $\hat{\boldsymbol{n}}$ also solves the dual of this optimization problem, namely, it guarantees (1.1) with the smallest possible total sample size $\hat{N} = \sum_{i=1}^k \hat{n}_i$ for specified P^*. We primarily address the former problem (or rather a continuous approximation of it given in (2.6)) in the present article.

A convenient choice of the n_i (ignoring the integer restriction on them) is one that makes $\mathrm{Var}(\bar{X}_i)$ $(1 \le i \le k)$ equal, i.e.,

$$\frac{\sigma_1^2}{n_1} = \frac{\sigma_2^2}{n_2} = \cdots = \frac{\sigma_k^2}{n_k} . \tag{1.4}$$

This allocation has the advantage that standard tables such as Table I in Bechhofer (1954) or Table A1 in Gibbons, Olkin and Sobel (1977) can be used to determine the n_i necessary to guarantee (1.1) using R; see (2.12). In Bechhofer (1954) it was pointed out that the allocation (1.4) is *not* optimal for $k=2$, the optimal allocation (again ignoring the integer restriction on the n_i) being

$$\frac{\sigma_1}{n_1} = \frac{\sigma_2}{n_2} . \tag{1.5}$$

Dudewicz and Dalal (1975) have studied for $k=2$ the relative efficiency of the allocation (1.4) with respect to the optimal allocation (1.5). They have shown that as σ_1^2/σ_2^2 approaches zero, the allocation (1.4) requires twice as many observations as that required by the allocation (1.5) to guarantee (1.1).

For $k \ge 3$ the optimal allocation has not yet been determined. Tong and Wetzell (1984) have given some asymptotic results but their emphasis is on the sequential setting. Gupta and Miescke (1988) have considered this problem in a decision theoretic framework. In Bechhofer (1954, p. 24) (where the problem was first posed), in Hall (1959, p. 965), and in Dudewicz and Dalal (1975, p. 34) it is stated that for $k \ge 3$ the optimal allocation appears to be too complicated for practical application, while Gibbons et al. (1977, p. 68) remark that (1.4) may not be optimal for $k \ge 3$. In this article we prove that for $k \ge 3$, the allocation (1.4) is in fact *locally optimal* for certain ranges of values of the parameters of the problem. More precisely, let

$$\lambda = \delta^* \sqrt{N}/\bar{\sigma} \tag{1.6}$$

where

$$\bar{\sigma}^2 = \frac{1}{k} \sum_{i=1}^{k} \sigma_i^2. \tag{1.7}$$

We show that for given variances $\sigma_1^2, \ldots, \sigma_k^2$, the allocation (1.4) is *locally optimal* except for $\lambda_{\mathrm{L}} < \lambda < \lambda_{\mathrm{U}}$ where λ_{L} and λ_{U} are two critical constants which can be determined explicitly by solving a simple equation for each. Since $\inf_{\Omega(\delta^*)} P(\mathrm{CS})$ for the allocation (1.4) is a strictly increasing function of λ, the above limits on λ imply corresponding limits on P^*, namely, $P_{\mathrm{L}} < P^* < P_{\mathrm{U}}$. Furthermore, letting

$$\beta_i = \sigma_i^2/\bar{\sigma}^2 \quad (1 \le i \le k), \tag{1.8}$$

we show that $\lambda_{\mathrm{L}}(P_{\mathrm{L}})$ depends only on β_k while $\lambda_{\mathrm{U}}(P_{\mathrm{U}})$ depends only on β_1. Thus the determination of P_{L} and P_{U} requires only the specification of the largest and smallest *relative* variances (with respect to the average variance), respectively. In most practical cases of interest, P_{L} is quite small ($0.30 \sim 0.50$), and so it is only P_{U} that needs to be determined. We show that $\lambda_{\mathrm{L}} < \infty$ $(P_{\mathrm{L}} < 1)$ always, while $\lambda_{\mathrm{U}} < \infty$

($P_U < 1$) only when

$$\beta_1 > \frac{k}{2(k-1)}, \tag{1.9}$$

i.e., when σ_1^2 is 'sufficiently' large with respect to $\bar{\sigma}^2$; otherwise $\lambda_U = \infty$.

Although we have been able to derive analytically only local optimality results for $k \geq 3$, nevertheless these results are valuable for the following reasons: (a) Numerical searches for $k = 3$ indicate that local optimality of (1.4) indeed corresponds to global optimality; we conjecture that this is true for $k > 3$. (b) These results yield insight into the nature of difficulties and the structure of the solution. As indicated above, this problem has been studied for more than 35 years by many researchers, and a complete analytical solution, particularly for $k > 3$, appears very difficult. The present work represents the most significant stride that has been made toward the solution.

The outline of the paper is as follows. Section 2 gives a mathematical formulation of the optimization problem. Section 3 gives the main theoretical results of the paper. The special case $k = 2$ is discussed in Section 3.1. The new results for $k \geq 3$ are summarized in Theorems 1-3 in Section 3.2. The proofs of all of the theorems are given in the Appendix. Section 3.3 gives a table of critical values of β_1 for selected values of k and P^*; this table is useful in determining whether allocation (1.4) is or is not locally optimal. Section 4 gives the results of numerical searches for the optimal allocation that we carried out for $k = 3$ and for selected $\boldsymbol{\sigma}^2$ configurations when allocation (1.4) is not optimal. Section 5 gives concluding remarks.

2. Problem formulation

Let

$$\Omega_i(\delta^*) = \{\boldsymbol{\omega} \in \Omega(\delta^*) \mid \mu_i = \mu_{[k]}\} \quad (1 \leq i \leq k), \tag{2.1}$$

i.e., $\Omega_i(\delta^*)$ is that part of the preference zone $\Omega(\delta^*)$ where the population having the variance σ_i^2 is the best population. It was shown in Bechhofer (1954) that for procedure R with any choice of $\boldsymbol{n}$ and for any fixed known $\boldsymbol{\sigma}^2$,

$$\inf_{\Omega_i(\delta^*)} P(\mathrm{CS}) = P_{\boldsymbol{\mu}_i(\delta^*)}(\mathrm{CS}) \tag{2.2}$$

where $\boldsymbol{\mu}_i(\delta^*)$ is any $\boldsymbol{\mu}$ satisfying

$$\mu_1 = \cdots = \mu_{i-1} = \mu_{i+1} = \cdots = \mu_k = \mu_i - \delta^*,$$

i.e., $\boldsymbol{\mu}_i(\delta^*)$ is the so-called slippage configuration with $\mu_i = \mu_{[k]}$ $(1 \leq i \leq k)$. Denoting $P_{\boldsymbol{\mu}_i(\delta^*)}(\mathrm{CS})$ by P_i $(1 \leq i \leq k)$ we see that

$$\inf_{\Omega(\delta^*)} P(\mathrm{CS}) = \min_{1 \leq i \leq k} P_i. \tag{2.3}$$

If we let

$$\gamma_i = n_i/N \quad (1 \le i \le k) \tag{2.4}$$

then it is easy to show that for fixed $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_k)$,

$$P_i = P_i(\gamma \mid \lambda, \beta) = \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j \neq i}}^{k} \Phi[\sqrt{\gamma_j/\beta_j}(x\sqrt{\beta_i/\gamma_i} + \lambda)]\phi(x)\,\mathrm{d}x \quad (1 \le i \le k) \tag{2.5}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal c.d.f. and p.d.f., respectively, and λ is given by (1.6).

For given N, $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$, and specified δ^*, each P_i is a function of the discrete valued argument γ since each $\gamma_j \ge 0$ is a multiple of $1/N$ with $\sum_{j=1}^{k} \gamma_j = 1$. For any given k, σ^2, δ^* and N, the *exact* integer-valued optimal allocation $\hat{n} = (\hat{n}_1, \hat{n}_2, \ldots, \hat{n}_k)$ that maximizes (2.3) where $\sum_{i=1}^{k} \hat{n}_i = N$ can be found by enumeration. However, this is only feasible for small values of N. Moreover, the integer solution has the disadvantage that a separate answer is needed for each $(\delta^*/\bar{\sigma}, N)$.

In the sequel we seek an *approximation* to this integer programming problem that does not depend on $\delta^*/\bar{\sigma}$ and N separately but rather only on $\lambda = \delta^*\sqrt{N}/\bar{\sigma}$. To this end we henceforth regard the $\gamma_j \ge 0$ as *continuous* variables summing to unity and ignore their dependence on N. This continuous approximation obviously will become more accurate as N increases. This same device was employed in Bechhofer (1969). Thus $P_i = P_i(\gamma \mid \lambda, \beta)$ can be regarded as a continuous function of γ for given β and specified $\lambda \ge 0$. We refer to any γ in the $(k-1)$-simplex

$$\Gamma = \left\{ \gamma : \gamma_i \ge 0, \sum_{i=1}^{k} \gamma_i = 1 \right\}$$

as an *allocation*.

We now state the continuous optimization problem (which is an approximation to the exact discrete optimization problem (1.3)) as follows:

Approximate Continuous Optimization Problem: For given β and specified $\lambda \ge 0$ find $\gamma \in \Gamma$ which achieves

$$\max_{\gamma \in \Gamma} \min_{1 \le i \le k} P_i(\gamma \mid \lambda, \beta). \tag{2.6}$$

We denote the solution to (2.6) by $\hat{\gamma} = \hat{\gamma}(\lambda, \beta)$ and refer to it as the *optimal allocation*; we denote the corresponding max–min probability by $\hat{P} = \hat{P}(\hat{\gamma})$. For fixed given β we will be interested in studying the behavior of $\hat{\gamma}$ and $\hat{P}$ as functions of λ.

We conclude this section by showing how to determine the sample sizes necessary to guarantee (1.1) when allocation (1.4) is used. Denote the allocation (1.4) by $\gamma^0 = (\gamma_1^0, \gamma_2^0, \ldots, \gamma_k^0)$ where

$$\frac{\gamma_1^0}{\beta_1} = \frac{\gamma_2^0}{\beta_2} = \cdots = \frac{\gamma_k^0}{\beta_k} = \frac{1}{\sum_{i=1}^{k} \beta_i} = \frac{1}{k}. \tag{2.7}$$

Note from (2.5) that

$$P_1(\boldsymbol{\gamma}^0 \mid \lambda, \boldsymbol{\beta}) = P_2(\boldsymbol{\gamma}^0 \mid \lambda, \boldsymbol{\beta}) = \cdots = P_k(\boldsymbol{\gamma}^0 \mid \lambda, \boldsymbol{\beta}) = P^0(\lambda, \boldsymbol{\beta}) \quad \text{(say)} \tag{2.8}$$

where

$$P^0(\lambda, \beta) = \int_{-\infty}^{\infty} \Phi^{k-1}\left(x + \frac{\lambda}{\sqrt{k}}\right)\phi(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \Phi^{k-1}\left(x + \frac{\delta^*\sqrt{N}}{\bar{\sigma}\sqrt{k}}\right)\phi(x)\,\mathrm{d}x. \tag{2.9}$$

If $c(k, P^*)$ denotes the solution in c to the equation

$$\int_{-\infty}^{\infty} \Phi^{k-1}(x+c)\phi(x)\,\mathrm{d}x = P^*, \tag{2.10}$$

then the total sample size N^0 required to guarantee (1.1) when using the allocation (1.4) (or equivalently (2.7)) is given by

$$N^0 = \left\{\frac{c(k, P^*)}{\delta^*}\right\}^2 \sum_{i=1}^{k} \sigma_i^2. \tag{2.11}$$

The corresponding n_i's (denoted by n_i^0's) are given by

$$n_i^0 = \left\{\frac{c(k, P^*)}{\delta^*}\right\}^2 \sigma_i^2 \quad (1 \le i \le k). \tag{2.12}$$

The critical constant $c(k, P^*)$ is tabulated in the references cited following (1.4).

3. Optimal allocation for $k \geq 2$

3.1. Special case $k = 2$

For $k = 2$, we see that (2.5) reduces to

$$P_1 = P_2 = \Phi[\lambda/\{\beta_1/\gamma_1 + \beta_2/\gamma_2\}^{1/2}],$$

and the optimal allocation for *all* $\lambda \geq 0$ is given by (1.5). The case $k = 2$ has several special simplifying features, which do not extend to the cases $k \geq 3$. These features are:

(i) For any fixed $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$ and $\boldsymbol{\gamma}$, the $P(\mathrm{CS})$ is the same regardless of the association between $\mu_{[i]}$ and σ_j^2 $(i, j = 1, 2)$. Moreover, the allocation (1.5) maximizes this $P(\mathrm{CS})$ at any $\boldsymbol{\mu}$, not just at the slippage configuration.

(ii) This $P(\mathrm{CS})$ (in particular, $P_1 = P_2$) can be expressed as a univariate normal c.d.f., which for given N is maximized by minimizing $\mathrm{Var}(\bar{X}_1 - \bar{X}_2) = \sigma_1^2/n_1 + \sigma_2^2/n_2$ subject to $n_1 + n_2 = N$.

For $k \geq 3$ the P_i are in general different. Furthermore, each P_i is a multivariate normal probability, which depends not only on the $\mathrm{Var}(\bar{X}_i - \bar{X}_j)$ but also on the $\mathrm{corr}(\bar{X}_i - \bar{X}_j, \bar{X}_i - \bar{X}_{j'})$ $(j \neq j' \neq i,\ 1 \leq j, j' \leq k)$.

3.2. General case $k \geq 3$

In this section we determine the range of values of λ for which the allocation γ^0 given by (2.8) is locally optimal when $k \geq 3$. The principal results of this section are summarized in the following theorems:

Theorem 1. *Define*

$$A(\lambda) = \int_{-\infty}^{\infty} x \Phi^{k-2}(x) \phi(x) \phi\left(x - \frac{\lambda}{\sqrt{k}}\right) \mathrm{d}x, \tag{3.1}$$

$$B(\lambda) = \int_{-\infty}^{\infty} \left(x - \frac{\lambda}{\sqrt{k}}\right) \Phi^{k-2}(x) \phi(x) \phi\left(x - \frac{\lambda}{\sqrt{k}}\right) \mathrm{d}x \tag{3.2}$$

and

$$G(\lambda) = \frac{k}{k-1} \frac{A(\lambda)}{A(\lambda) - B(\lambda)}. \tag{3.3}$$

Then the allocation (2.7) *is* locally *optimal iff*

$$G(\lambda) \geq \beta_k \quad or \quad G(\lambda) \leq \beta_1. \tag{3.4}$$

Corollary. *For $\lambda = G^{-1}(1)$ (that G^{-1} exists and is unique follows from Theorem 2 below) the allocation (2.7) is locally (and, in fact globally) optimal iff $\sigma_1^2 = \cdots = \sigma_k^2$.* □

Theorem 2. *For $\lambda > 0$ the function $G(\lambda)$ is continuous and strictly decreasing in λ with* $\lim_{\lambda \to 0} G(\lambda) = \infty$ *and* $\lim_{\lambda \to \infty} G(\lambda) = k/2(k-1)$. *Hence the condition* (3.4) *is equivalent to the condition*

$$\lambda \leq \lambda_{\mathrm{L}} \quad or \quad \lambda \geq \lambda_{\mathrm{U}}, \tag{3.5}$$

respectively. Here λ_{L} is the unique finite solution in λ of the equation

$$G(\lambda) = \beta \tag{3.6}$$

with $\beta = \beta_k$, and if (1.9) is satisfied then λ_{U} is the unique finite solution in λ of (3.6) with $\beta = \beta_1$. If (1.9) is not satisfied then (3.6) with $\beta = \beta_1$ does not have a solution and in that case we define $\lambda_{\mathrm{U}} = \infty$.

Corollary. *The allocation (2.7) is locally optimal iff the specified P^* is $\leq P_{\mathrm{L}}$ or $\geq P_{\mathrm{U}}$ where*

$$P_{\mathrm{L}} = \int_{-\infty}^{\infty} \Phi^{k-1}\left(x + \frac{\lambda_{\mathrm{L}}}{\sqrt{k}}\right) \phi(x)\, \mathrm{d}x \tag{3.7}$$

and

$$P_{\mathrm{U}}=\begin{cases}\displaystyle\int_{-\infty}^{\infty}\Phi^{k-1}\left(x+\frac{\lambda_{\mathrm{U}}}{\sqrt{k}}\right)\phi(x)\,\mathrm{d}x & \text{if } \beta_1>\dfrac{k}{2(k-1)},\\ 1 & \text{if } \beta_1\le\dfrac{k}{2(k-1)}.\end{cases}\tag{3.8}$$ □

We now give (in Theorem 3) an alternative representation for equation (3.6) which is convenient for computing. This representation involves multivariate normal c.d.f.'s for which we use the following notation: Let $X_1, X_2, \ldots, X_p$ have a joint p-variate normal distribution with zero means, unit variances, and common correlation $\varrho=\mathrm{corr}(X_i, X_j)$ for $i\neq j$ $(1\le i, j\le p)$. We denote the equicoordinate multivariate normal probability

$$P(X_1\le x, X_2\le x, \ldots, X_p\le x)$$

by $\Phi_p(x\mid\varrho)$. For $p=1$ this probability is simply the univariate normal c.d.f. denoted by $\Phi(x)$. For $p=0$ we define this probability to be unity.

Theorem 3. *Set $\tau=\lambda/\sqrt{6k}$. Then $\lambda_{\mathrm{L}}=\tau_{\mathrm{L}}\sqrt{6k}$ where τ_{L} is the unique solution in τ of the equation*

$$\frac{\tau\Phi_{k-2}(\tau\mid\frac{1}{3})}{\phi(\tau)\Phi_{k-3}(\tau/\sqrt{2}\mid\frac{1}{4})}=\frac{k(k-2)}{3[2(k-1)\beta-k]}\tag{3.9}$$

with $\beta=\beta_k$. Similarly if condition (1.9) *is satisfied, then $\lambda_{\mathrm{U}}=\tau_{\mathrm{U}}\sqrt{6k}$ where τ_{U} is the unique solution in τ of* (3.9) *with $\beta=\beta_1$.*

Remark 1. For $\tau>0$ the left-hand side of (3.9) is positive which leads to condition (1.9). (Note that $\beta_k>k/2(k-1)$ always.)

Remark 2. For $k=3$ the left-hand side of (3.9) reduces to $\tau\Phi(\tau)/\phi(\tau)$ which is very simple to evaluate.

3.3. Table of critical values of β_1 for $k\ge 3$

Table 1 gives values of the lower bound on β_1, say β_1^*, and the values of the associated lower bound on P_{U}, say P_{U}^*, such that for $P^*\ge P_{\mathrm{U}}^*$ (for $P_{\mathrm{U}}^*=0.80$, 0.90, 0.95 and 0.99 and $k=3(1)8$), the allocation γ^0 given by (2.7) is locally optimal if $\beta_1\ge\beta_1^*$. We also have added a row for $P_{\mathrm{U}}^*=1$ in which case $\beta_1^*=k/2(k-1)$.

To illustrate the use of this table, suppose that $k=3$ and $P^*=0.95$. If $\beta_1\ge\beta_1^*=0.806$ then γ^0 is the locally optimal allocation, and the corresponding sample sizes required can be found from (2.12) once δ^* is specified. If $\beta_1<\beta_1^*$ then the optimal allocation is not given by γ^0.

An analogous table could be given for β_k^* for selected values of P_{L}^* such that for

Table 1
Critical values β_1^*

P_U^*	k					
	3	4	5	6	7	8
0.80	0.907	0.854	0.826	0.809	0.798	0.791
0.90	0.838	0.810	0.748	0.731	0.720	0.712
0.95	0.806	0.741	0.708	0.690	0.678	0.670
0.99	0.775	0.670	0.664	0.643	0.629	0.620
1.00	0.750	0.667	0.625	0.600	0.583	0.571

$P^* \le P_L^*$, the allocation γ^0 is locally optimal if $\beta_k \le \beta_k^*$. However, such a table is likely to be of less practical value since large values of P^* are more common.

4. Numerical results for $k=3$

In Section 3 we derived a necessary and sufficient condition for γ^0 to be locally optimal for any $k \ge 3$. In the present section we investigate the nature of the (globally) optimal allocation when that condition fails, i.e., when $P_L < P^* < P_U$. We also investigate the amount of the associated saving in the total sample size in comparison to that required by the allocation γ^0 to guarantee the same probability requirement (1.1).

An analytical characterization of the optimal allocation appears to be very difficult when $P_L < P^* < P_U$ holds. Therefore we decided to investigate numerically the behavior of the optimal allocation as a function of P^* by performing a search in the allocation space Γ. This would be a very formidable computational task for large k, so we confined our attention to $k=3$, in which case the search is only in two dimensions.

For $k=3$, we present the results for a total of six $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ configurations. The first three configurations have $\sigma_3^2/\sigma_1^2 = 3$, while the second three have $\sigma_3^2/\sigma_1^2 = 10$. These configurations and the associated $\boldsymbol{\beta}$ vectors are listed in Table 2. Note that the optimal allocation (and, as will be seen below, the relative saving in the total sample size) depends only on the relative magnitudes of the σ_i^2, not their absolute magnitudes. For each configuration, we have $\beta_1 < k/2(k-1) = 0.75$ and hence $P_U = 1$. The P_L-values associated with each configuration $\boldsymbol{\beta}$ (recall that P_L depends on $\boldsymbol{\beta}$ only through β_k) are also listed in Table 2. For each configuration the optimal allocation was determined numerically for $P^* = 0.80$, 0.90, 0.95 and 0.99; note that this practical range of P^*-values is well in excess of P_L for each configuration.

The numerical search for the optimal allocation was carried out as follows: Let λ^0 be the λ-value required using the allocation γ^0 to guarantee the probability re-

Table 2
Optimum allocation $\hat{\boldsymbol{\gamma}}$ and association relative saving $\widehat{\mathrm{RS}}$

$\boldsymbol{\sigma}^2=(\sigma_1^2,\sigma_2^2,\sigma_3^2)$:			(1, 3, 3)	(1, 2, 3)	(1, 1, 3)	(1, 10, 10)	(1, 4, 10)	(1, 1, 10)
$\boldsymbol{\beta}=(\beta_1,\beta_2,\beta_3)$:			(0.429, 1.286, 1.286)	(0.500, 1.000, 1.500)	(0.600, 0.600, 1.800)	(0.143, 1.430, 1.430)	(0.200, 0.800, 2.000)	(0.250, 0.250, 2.500)
$\boldsymbol{\gamma}^0=(\gamma_1^0,\gamma_2^0,\gamma_3^0)$:			(0.143, 0.429, 0.429)	(0.167, 0.333, 0.500)	(0.200, 0.200, 0.600)	(0.048, 0.476, 0.476)	(0.067, 0.267, 0.667)	(0.083, 0.083, 0.833)
P^*	λ^0	P_L:	0.45	0.50	0.55	0.41	0.44	0.51
0.80	2.862	$\hat{\boldsymbol{\gamma}}$:	(0.206, 0.397, 0.397)	(0.233, 0.333, 0.433)	(0.267, 0.267, 0.467)	(0.143, 0.429, 0.429)	(0.167, 0.300, 0.533)	(0.183, 0.183, 0.633)
		$\hat{\lambda}$:	2.801	2.818	2.795	2.652	2.667	2.514
		$\widehat{\mathrm{RS}}$	4.22%	3.05%	4.63%	14.14%	13.16%	22.84%
0.90	3.863	$\hat{\boldsymbol{\gamma}}$:	(0.206, 0.397, 0.397)	(0.200, 0.333, 0.467)	(0.233, 0.233, 0.533)	(0.111, 0.444, 0.444)	(0.150, 0.300, 0.550)	(0.167, 0.167, 0.667)
		$\hat{\lambda}$:	3.805	3.831	3.810	3.645	3.667	3.492
		$\widehat{\mathrm{RS}}$	2.98%	1.65%	2.73%	10.97%	9.89%	18.29%
0.95	4.694	$\hat{\boldsymbol{\gamma}}$:	(0.206, 0.397, 0.397)	(0.200, 0.333, 0.467)	(0.233, 0.233, 0.533)	(0.111, 0.444, 0.444)	(0.133, 0.300, 0.567)	(0.167, 0.167, 0.667)
		$\hat{\lambda}$:	4.651	4.674	4.654	4.480	4.504	4.300
		$\widehat{\mathrm{RS}}$	1.82%	0.85%	1.70%	8.91%	7.93%	16.08%
0.99	6.265	$\hat{\boldsymbol{\gamma}}$:	(0.175, 0.413, 0.413)	(0.200, 0.333, 0.467)	(0.233, 0.233, 0.533)	(0.095, 0.452, 0.452)	(0.117, 0.300, 0.583)	(0.167, 0.167, 0.667)
		$\hat{\lambda}$:	6.227	6.253	6.231	6.057	6.070	5.806
		$\widehat{\mathrm{RS}}$	1.21%	0.38%	1.08%	6.53%	6.13%	14.12%

quirement (1.1) for specified P^* and for any $\delta^*>0$; from (2.9) we see that $\lambda^0=\sqrt{k}c(k,P^*)$. Let $\hat{\lambda}\le\lambda^0$ be the corresponding λ-value required using the associated optimal allocation $\hat{\gamma}$. Starting with λ^0 we decreased λ in steps of 0.001, determining the optimal allocation $\hat{\gamma}$ and the associated max–min probability $\hat{P}$ for each λ (note that $\hat{P}$ decreases with λ), until the smallest possible λ for which $\hat{P}\ge P^*$ was attained. This is the desired value of $\hat{\lambda}$, which is tabulated together with $\hat{\gamma}$ in Table 2. A mesh size of at most $\frac{1}{60}$ was used for each γ_i in the search over the allocation space Γ.

The percentage relative saving ($\widehat{\text{RS}}$) in the total sample size resulting from the use of the optimal allocation $\hat{\gamma}$ instead of the allocation γ^0 to guarantee the same probability requirement (1.1) is given by

$$\widehat{\text{RS}}=\left(\frac{N^0-\hat{N}}{N^0}\right)\times 100=\left\{\frac{(\lambda^0)^2-(\hat{\lambda})^2}{(\lambda^0)^2}\right\}\times 100. \tag{4.1}$$

The values of $\widehat{\text{RS}}$ are also listed in Table 2.

From Table 2 we first note that, as one would expect, the relative savings are substantially higher for the configurations with $\sigma_3^2/\sigma_1^2=10$ compared to those for the configurations with $\sigma_3^2/\sigma_1^2=3$. Thus the relative saving in the total sample size from the use of the optimal allocation $\hat{\gamma}$ (in comparison to that required when using the allocation γ^0) appears to increase with $\sigma^2_{\max}/\sigma^2_{\min}$. Of course, the relative saving is not simply a function of $\sigma^2_{\max}/\sigma^2_{\min}$. For example, the relative savings are quite different for the cases $\boldsymbol{\sigma}^2=(1,10,10)$ and $(1,1,10)$. For each configuration, the relative saving is highest for $P^*=0.80$ and decreases as P^* increases. For the configuration $\boldsymbol{\gamma}^2=(1,1,10)$, the relative saving is nearly 23% for $P^*=0.80$. This indicates that there is much to be gained by using the optimal allocation $\hat{\gamma}$ instead of the 'convenient' allocation γ^0, particularly when $\sigma^2_{\max}/\sigma^2_{\min}$ is large and P^* is in between P_{L} and P_{U}. We should, however, stress that although the *relative* savings are small for large P^*, the *absolute* savings, $N^0-\hat{N}=\{(\lambda^0)^2-(\hat{\lambda})^2\}(\bar{\sigma}/\delta^*)^2$, can be quite large, more so when $\delta^*/\bar{\sigma}$ is small.

In practice the numerical search for the optimal allocation $\hat{\gamma}$ can be prohibitively expensive and possibly even infeasible for large k. Therefore it would be desirable to have a simple heuristic rule that would improve upon γ^0 and possibly serve as a reasonable approximation to the optimal allocation $\hat{\gamma}$. With this in mind we now carefully examine the $\hat{\gamma}$-vectors listed in Table 2.

In several cases we note that $\hat{\gamma}$ does not change as we vary P^*. We do not have a simple explanation for this behavior of the optimal allocation. We also observe quite unmistakeably that $\hat{\gamma}_i\lesseqqgtr\gamma_i^0$ iff $\beta_i\gtreqqless 1$. In other words, $\hat{\gamma}$ allocates a smaller (larger) proportion of observations (than that allocated by γ^0) to any population with larger (smaller) variance relative to $\bar{\sigma}^2$ which results in the inequality (generalizing from $k=3$ to $k>3$)

$$\frac{\hat{\gamma}_1}{\beta_1}\ge\frac{\hat{\gamma}_2}{\beta_2}\ge\cdots\ge\frac{\hat{\gamma}_k}{\beta_k}\,. \tag{4.2}$$

Table 3
Allocation $\tilde{\boldsymbol{\gamma}}$ and association relative saving $\widetilde{\mathrm{RS}}$

$\boldsymbol{\sigma}^2=(\sigma_1^2,\sigma_2^2,\sigma_3^2)$:		(1, 3, 3)	(1, 2, 3)	(1, 1, 3)	(1, 10, 10)	(1, 4, 10)	(1, 1, 10)
$\sqrt{\boldsymbol{\beta}}=(\sqrt{\beta_1},\sqrt{\beta_2},\sqrt{\beta_3})$:		(0.655, 1.134, 1.134)	(0.707, 1.000, 1.225)	(0.755, 0.775, 1.342)	(0.378, 1.195, 1.195)	(0.447, 0.894, 1.414)	(0.500, 0.500, 1.581)
$\tilde{\boldsymbol{\gamma}}=(\tilde{\gamma}_1,\tilde{\gamma}_2,\tilde{\gamma}_3)$:		(0.224, 0.388, 0.388)	(0.241, 0.341, 0.418)	(0.268, 0.268, 0.464)	(0.137, 0.432, 0.432)	(0.162, 0.325, 0.513)	(0.194, 0.194, 0.613)
$P^*=0.80$	$\tilde{\lambda}$:	2.803	2.827	2.801	2.657	2.673	2.516
	$\widetilde{\mathrm{RS}}$	4.08%	2.43%	4.22%	13.81%	12.77%	22.72%
$P^*=0.90$	$\tilde{\lambda}$:	3.812	3.853	3.834	3.653	3.687	3.502
	$\widetilde{\mathrm{RS}}$	2.62%	0.52%	1.50%	10.58%	8.90%	17.82%
$P^*=0.95$	$\tilde{\lambda}$:	4.659	4.694	4.663	4.490	4.529	4.315
	$\widetilde{\mathrm{RS}}$	1.49%	0%	1.32%	8.50%	6.91%	15.50%
$P^*=0.99$	$\tilde{\lambda}$:	6.259	6.315	6.296	6.110	6.129	5.850
	$\widetilde{\mathrm{RS}}$	0.19%	−1.60%	−0.99%	4.89%	4.29%	12.81%

We know that $\beta_i \gtreqless \sqrt{\beta_i}$ depending on whether $\beta_i \gtreqless 1$. Therefore the allocation $\tilde{\gamma}$ with

$$\frac{\tilde{\gamma}_1}{\sqrt{\beta_1}} = \frac{\tilde{\gamma}_2}{\sqrt{\beta_2}} = \cdots = \frac{\tilde{\gamma}_k}{\sqrt{\beta_k}}, \tag{4.3}$$

which chooses the n_i's in proportion to the σ_i's, is an allocation that satisfies (4.2). Recall that this allocation is globally optimal for $k=2$. It would be of interest to determine how close this allocation is to the optimum for $k=3$ when $P_{\mathrm{L}} < P^* < P_{\mathrm{U}}$ holds and hence when γ^0 is known *not* to be optimal. To this end we determined the smallest λ-value (denoted by $\tilde{\lambda}$) for the allocation $\tilde{\gamma}$ such that the associated probability is $\geq P^*$ for the $\boldsymbol{\sigma}^2$-configurations and P^*-values listed in Table 2. We also calculated the percentage relative saving ($\widetilde{\mathrm{RS}}$) associated with $\tilde{\gamma}$ relative to γ^0 as in (4.1). The results are given in Table 3.

Inspection of Tables 2 and 3 reveals that in many cases, the $\tilde{\gamma}$ allocation achieves relative savings nearly equal to those achieved by the optimal allocation $\hat{\gamma}$. The $\tilde{\gamma}$ allocation improves upon the γ^0 allocation in all of the cases studied except two (for $\boldsymbol{\sigma}^2 = (1,2,3)$ and $(1,1,3)$ when $P^* = 0.99$), and in those two cases the excess sample size required by $\tilde{\gamma}$ compared to that required by γ^0 is not large, in relative terms.

Recognizing the computational difficulties involved in determining the optimal allocation $\hat{\gamma}$ when $P_{\mathrm{L}} < P^* < P_{\mathrm{L}}$ holds, we recommend the $\tilde{\gamma}$ allocation in this case with little reservation.

5. Concluding remarks

In this paper we have shown that the convenient allocation γ^0 given by (2.7) is locally optimal for $k \geq 3$ if and only if $P^* \leq P_{\mathrm{L}}$ or $P^* \geq P_{\mathrm{U}}$ where P_{L} and P_{U} can be explicitly determined given β_k and β_1, respectively. The determination of the globally optimal allocation $\hat{\gamma}$ (whether or not it equals γ^0) requires the knowledge of all of the β_i's, and the determination of the associated sample sizes $\hat{n}_i$ needed to guarantee (1.1) for specified $\{\delta^*, P^*\}$ requires the knowledge of all of the σ_i^2's. The optimal allocation is difficult to determine when $P_{\mathrm{L}} < P^* < P_{\mathrm{U}}$ holds. In that case, use of the allocation $\tilde{\gamma}$ given by (4.3) (or some other allocation satisfying (4.2)) is suggested.

There are two matters of concern when the variances are unequal. First, assuming that selection in terms of means is still meaningful, the appropriateness of the procedure R, which bases its decision on the sample means $\bar{X}_i$, may be called into question for the following reason: Suppose that the two largest sample means differ by a very small amount, but the largest sample mean has a much a larger variance than the second largest sample mean. (This is possible even when the σ_i^2's are equal but the n_i's are not.) Intuition suggests that in this case we should select the population yielding the second largest sample mean as the 'best'. This is because the second largest sample mean is a much more reliable estimator of its population mean (which is thus likely to be large and possibly the largest) than the largest sample mean is of its population mean (which is thus less likely to be the largest). Recently Berger and Deely (1988) have given a Bayesian solution to this problem which involves shrinking the sample means toward a central average, the extent of shrinkage being greater for extreme (large or small) sample means having larger variances.

The second matter concerns the appropriateness of the selection goal itself. If the population having the largest mean also has (nearly) the largest variance then the experimenter might wish to select another population with a somewhat smaller mean if it also has a small variance. Santner and Tamhane (1984) have proposed a formulation and a procedure for such a selection goal.

Appendix

We now provide the proofs of the three theorems stated in Section 3.

Proof of Theorem 1. Instead of $\boldsymbol{\gamma}$, it will be more convenient to work in terms of $\boldsymbol{\alpha}=(\alpha_1,\dots,\alpha_k)$ where

$$\alpha_i=\sqrt{\gamma_i/\beta_i} \quad (1\le i\le k).$$

We wish to determine the necessary and sufficient conditions for the allocation $\boldsymbol{\gamma}^0$ given by (2.8), i.e.,

$$\boldsymbol{\alpha}^0=(\alpha_1^0,\alpha_2^0,\dots,\alpha_k^0)=(1/\sqrt{k},1/\sqrt{k},\dots,1/\sqrt{k}),$$

to be locally optimal. Since at $\boldsymbol{\alpha}^0$ we have $P_1=\cdots=P_k=P^0$ as noted in (2.9), and since the objective function to be maximized is $\min_{1\le i\le k}P_i$, it follows that

$\boldsymbol{\alpha}^0$ is locally optimal

$\Leftrightarrow$ for every $\boldsymbol{c}\in\mathscr{C}^k$, $\exists i$ $(1\le i\le k)$ which in general depends on $\boldsymbol{c}\ni$

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}P_i(\gamma_1^0+c_1\varepsilon,\dots,\gamma_k^0+c_k\varepsilon)\bigg|_{\boldsymbol{\alpha}^0}=\sum_{j=1}^{k}c_j\frac{\partial P_i}{\partial\gamma_j}\bigg|_{\boldsymbol{\alpha}^0}\le 0 \tag{A.1}$$

where $\mathscr{C}^k=\{\boldsymbol{c}\colon \sum_{i=1}^k c_i=0\}$ is the space of all k-dimensional contrasts. Note that the quantity on the left-hand side of the equality in (A.1) is the gradient of P_i at $\boldsymbol{\alpha}^0$ along the direction $\boldsymbol{c}$.

To derive a formula for this gradient we require the partial derivatives $\partial P_i/\partial\alpha_j$ $(j\ne i)$ and $\partial P_i/\partial\alpha_i$ evaluated at $\boldsymbol{\alpha}^0$. It can be shown that

$$\frac{\partial P_i}{\partial\alpha_j}\bigg|_{\boldsymbol{\alpha}^0}=\sqrt{k}A(\lambda) \tag{A.2}$$

where $A(\lambda)$ is defined in (3.1), and

$$\frac{\partial P_i}{\partial\alpha_i}\bigg|_{\boldsymbol{\alpha}^0}=-(k-1)\sqrt{k}B(\lambda) \tag{A.3}$$

where $B(\lambda)$ is defined in (3.2). Hence the gradient in (A.1) can be written as

$$\sum_{j=1}^{k} c_j\left(\frac{d\alpha_j}{d\gamma_j}\right)\left(\frac{\partial P_i}{\partial \alpha_j}\right)\bigg|_{\boldsymbol{\alpha}^0} = \sum_{j=1}^{k} c_j\left(\frac{1}{2\alpha_j\beta_j}\right)\left(\frac{\partial P_i}{\partial \alpha_j}\right)\bigg|_{\boldsymbol{\alpha}^0}$$

$$= \frac{\sqrt{k}}{2}\left\{\sum_{\substack{j=1\\ j\neq i}}^{k} \frac{c_j}{\beta_j}\sqrt{k}A(\lambda) - \frac{c_i}{\beta_i}(k-1)\sqrt{k}B(\lambda)\right\}$$

(using (A.2) and (A.3))

$$= \frac{k}{2}A(\lambda)\left\{\sum_{j=1}^{k}\frac{c_j}{\beta_j} - \frac{c_i}{\beta_i}k\left[\frac{G(\lambda)-1}{G(\lambda)}\right]\right\}$$

where $G(\lambda)$ is defined in (3.3). Since $A(\lambda)>0$, condition (A.1) is equivalent to the condition:

$\boldsymbol{\alpha}^0$ is locally optimal

$$\Leftrightarrow \quad \forall \boldsymbol{c}\in\mathscr{C}^k,\ \exists i\ (1\le i\le k)\ \ni \sum_{j=1}^{k} d_j - d_i kH(\lambda)\le 0 \tag{A.4}$$

where for notational convenience we have put

$$d_i = c_i/\beta_i \quad (1\le i\le k)$$

and

$$H(\lambda) = (G(\lambda)-1)/G(\lambda).$$

We consider three cases separately: $G(\lambda)=1$, i.e., $H(\lambda)=0$; $G(\lambda)>1$, i.e., $H(\lambda)>0$; and $G(\lambda)<1$, i.e., $H(\lambda)<0$.

Case 1 ($G(\lambda)=1$, i.e., $H(\lambda)=0$). In this case $\boldsymbol{\alpha}^0$ is locally optimal

$$\Leftrightarrow \quad \forall \boldsymbol{c}\in\mathscr{C}^k,\ \sum_{j=1}^{k} d_j\le 0$$

$$\Leftrightarrow \quad \beta_1=\cdots=\beta_k=1$$

$$\Leftrightarrow \quad \sigma_1^2=\cdots=\sigma_k^2.$$

Case 2 ($G(\lambda)>1$, i.e., $H(\lambda)>0$). In this case we want to show that $\boldsymbol{\alpha}^0$ is locally optimal iff $G(\lambda)\ge\beta_k$ (because we cannot have $G(\lambda)\le\beta_1$ since $\beta_1<1$). We first do the 'only if' part of the proof. Thus suppose $\boldsymbol{\alpha}^0$ is locally optimal. Choose $d_1=\cdots=d_{k-1}=1$ and $d_k=-\sum_{i=1}^{k-1}\beta_i/\beta_k$. From (A.4) we then have,

$$\exists i \ni (k-1) - \frac{\sum_{i=1}^{k-1}\beta_i}{\beta_k} - d_i kH(\lambda)\le 0.$$

However, this is true iff

$$d_{\max} kH(\lambda)\ge k(1-1/\beta_k)$$

$$\Leftrightarrow \quad H(\lambda)\ge 1-1/\beta_k \quad (\text{since } d_{\max}=1)$$

$$\Leftrightarrow \quad G(\lambda)\ge\beta_k. \tag{A.5}$$

For the 'if' part of the proof, we must show that $G(\lambda)\geq\beta_k \Rightarrow$ for any $\boldsymbol{c}\in\mathscr{C}^k$, (A.4) holds for some i. We will show that for a given $\boldsymbol{c}\in\mathscr{C}^k$, this is true for $i=i^*$ where $d_{i^*}=\max_{i\in I} d_i$ and $I=\{i: c_i>0\}$. To prove this we introduce additional notation and derive two inequalities. Let $J=\{j: c_j<0\}$, $\mathrm{card}(I)=s$ $(1\leq s\leq k-1)$, $c_I=\sum_{i\in I} c_i=-c_J$, $d_I=\sum_{i\in I} d_i$ and $\beta_I=\sum_{i\in I}\beta_i$. The first inequality is obtained as follows:

$$\begin{aligned}
&G(\lambda)\geq\beta_k\\
&\quad\Rightarrow\ k\beta_k H(\lambda)\geq\sum_{i=1}^{k}(\beta_k-\beta_i)\quad\text{(by using (A.5))}\\
&\quad\Rightarrow\ k\beta_k H(\lambda)\geq\sum_{i\in I}(\beta_k-\beta_i)=s\beta_k-\beta_I &\text{(A.6)}\\
&\quad\Rightarrow\ s-kH(\lambda)\leq\beta_I/\beta_k. &\text{(A.7)}
\end{aligned}$$

The second inequality is obtained as follows:

$$\begin{aligned}
&d_{i^*}\geq d_i\quad\text{for } i\in I\\
&\quad\Rightarrow\ d_{i^*}(1-\beta_i/\beta_k)\geq d_i(1-\beta_i/\beta_k)\quad\text{for } i\in I\\
&\quad\Rightarrow\ c_i-d_{i^*}\beta_i\geq(d_i-d_{i^*})\beta_k\quad\text{for } i\in I\\
&\quad\Rightarrow\ c_I-d_{i^*}\beta_I\geq(d_I-sd_{i^*})\beta_k\quad\text{(by summing over } i\in I)\\
&\quad\Rightarrow\ \{d_{i^*}-(c_I/\beta_I)\}(s\beta_k-\beta_I)\geq\{d_I-s(c_I/\beta_I)\}\beta_k\quad\text{(by rearranging terms)}\\
&\quad\Rightarrow\ \{d_{i^*}-(c_I/\beta_I)\}k\beta_k H(\lambda)\geq\{d_I-s(c_I/\beta_I)\}\beta_k\quad\text{(by using (A.6))}\\
&\quad\Rightarrow\ d_I-d_{i^*}kH(\lambda)\leq\{s-kH(\lambda)\}(c_I/\beta_I). &\text{(A.8)}
\end{aligned}$$

In the penultimate step above we have used the fact that $d_{i^*}\geq c_I/\beta_I$.

Returning to the 'if' part of the proof, we see that for $i=i^*$, the left-hand side of (A.4) equals

$$\begin{aligned}
&\sum_{j=1}^{k} d_j-d_{i^*}kH(\lambda)\\
&\quad=d_I-d_{i^*}kH(\lambda)+\sum_{j\in J}(c_j/\beta_j)\\
&\quad\leq d_I-d_{i^*}kH(\lambda)+c_J/\beta_k\quad\text{(since } \beta_k\geq\beta_j \text{ and } c_j<0 \text{ for } j\in J)\\
&\quad\leq\{s-kH(\lambda)\}(c_I/\beta_I)-c_I/\beta_k\quad\text{(by using (A.8) and } c_J=-c_I)\\
&\quad\leq(\beta_I/\beta_k)(c_I/\beta_I)-c_I/\beta_k\quad\text{(by using (A.7))}\\
&\quad=0.
\end{aligned}$$

This shows that for given $\boldsymbol{c}\in\mathscr{C}^k$, (A.4) holds for $i=i^*$. This completes the proof for Case 2.

Case 3 ($G(\lambda)<1$, i.e., $H(\lambda)<0$). The proof in this case is analogous to that in Case 2. Here we want to show that $\boldsymbol{\alpha}^0$ is locally optimal iff $G(\lambda)\leq\beta_1$ (because we cannot have $G(\lambda) \geq \beta_k$ since $\beta_k > 1$). The proof of the 'only if' part is obtained by choosing $d_1 = \sum_{i=2}^{k} \beta_i/\beta_1$ and $d_2 = \cdots = d_k = -1$. The proof of the 'if' part is obtained by showing that $G(\lambda) \leq \beta_1 \Rightarrow$ for any given $c \in \mathscr{C}^k$, (A.4) holds for $i = j^*$ where $|d_{j^*}| = \max_{j \in J} |d_j|$. This completes the proof of Case 3 and hence of the theorem. □

The corollary to Theorem 1 follows from Case 1 considered above, or from the fact that $\beta_1 \leq 1 \leq \beta_k$ with equalities holding iff $\sigma_1^2 = \cdots = \sigma_k^2$.

Proof of Theorem 2. The continuity of $G(\lambda)$ follows from the continuity of the integrals $A(\lambda)$ and $B(\lambda)$. The limiting value of $G(\lambda)$ for $\lambda \to 0$ is obtained by noting that

$$A(0) = B(0) = \int_{-\infty}^{\infty} x\Phi^{k-2}(x)\phi^2(x)\,\mathrm{d}x > 0$$

and hence $\lim_{\lambda \to 0} G(\lambda) = \infty$. Next, by combining the $\phi(x)$ and $\phi(x - \lambda/\sqrt{k})$ terms and setting $y = x - \lambda/2\sqrt{k}$ we can write

$$A(\lambda) = \frac{\mathrm{e}^{-\lambda^2/4k}}{2\pi} \int_{-\infty}^{\infty} \left(y + \frac{\lambda}{2\sqrt{k}}\right)\Phi^{k-2}\left(y + \frac{\lambda}{2\sqrt{k}}\right)\mathrm{e}^{-y^2}\,\mathrm{d}y$$

and

$$B(\lambda) = \frac{\mathrm{e}^{-\lambda^2/4k}}{2\pi} \int_{-\infty}^{\infty} \left(y - \frac{\lambda}{2\sqrt{k}}\right)\Phi^{k-2}\left(y + \frac{\lambda}{2\sqrt{k}}\right)\mathrm{e}^{-y^2}\,\mathrm{d}y.$$

Hence we have from (3.3),

$$G(\lambda) = \left(\frac{k}{k-1}\right)\left[\frac{C(\lambda) + (\lambda/2\sqrt{k})D(\lambda)}{(\lambda/\sqrt{k})D(\lambda)}\right] \tag{A.9}$$

where

$$C(\lambda) = \int_{-\infty}^{\infty} y\Phi^{k-2}\left(y + \frac{\lambda}{2\sqrt{k}}\right)\mathrm{e}^{-y^2}\,\mathrm{d}y \tag{A.10}$$

and

$$D(\lambda) = \int_{-\infty}^{\infty} \Phi^{k-2}\left(y + \frac{\lambda}{2\sqrt{k}}\right)\mathrm{e}^{-y^2}\,\mathrm{d}y. \tag{A.11}$$

It is easy to see that $\lim_{\lambda \to \infty} C(\lambda) = \int_{-\infty}^{\infty} y\mathrm{e}^{-y^2}\,\mathrm{d}y = 0$ and $\lim_{\lambda \to \infty} D(\lambda) = \int_{-\infty}^{\infty} \mathrm{e}^{-y^2}\,\mathrm{d}y > 0$. Hence the limit of the square bracketed term in (A.9) as $\lambda \to \infty$ is $\frac{1}{2}$, which yields $\lim_{\lambda \to \infty} G(\lambda) = k/2(k-1)$.

To show that $G(\lambda)$ is strictly decreasing in λ, we must show that $C(\lambda)/\lambda D(\lambda)$ is strictly decreasing in λ. To demonstrate this it suffices to show that $C(\lambda)/D(\lambda)$ is strictly decreasing in λ, i.e., for $0 \leq \lambda_1 < \lambda_2$ we have

$$C(\lambda_1)/D(\lambda_1) > C(\lambda_2)/D(\lambda_2).$$

However, this is true iff

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\Phi^{k-2}\left(y+\frac{\lambda_1}{2\sqrt{k}}\right)\Phi^{k-2}\left(z+\frac{\lambda_2}{2\sqrt{k}}\right)e^{-(y^2+z^2)}\,dy\,dz$$

$$>\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z\Phi^{k-2}\left(y+\frac{\lambda_1}{2\sqrt{k}}\right)\Phi^{k-2}\left(z+\frac{\lambda_2}{2\sqrt{k}}\right)e^{-(y^2+z^2)}\,dy\,dz$$

$$\Leftrightarrow \int_{y>z}\int (y-z)\left\{\Phi^{k-2}\left(y+\frac{\lambda_1}{2\sqrt{k}}\right)\Phi^{k-2}\left(z+\frac{\lambda_2}{2\sqrt{k}}\right)-\Phi^{k-2}\left(y+\frac{\lambda_2}{2\sqrt{k}}\right)\right.$$

$$\left.\times\Phi^{k-2}\left(z+\frac{\lambda_1}{2\sqrt{k}}\right)\right\}e^{-(y^2+z^2)}\,dy\,dz>0. \qquad \text{(A.12)}$$

Since $y-z>0$, (A.12) will be true if the quantity in the curly brackets is >0, which follows from the strictly increasing montone likelihood ratio property of the normal distribution.

The uniqueness of λ_L and λ_U (when $\beta_1<k/2(k-1)$) follows from the continuity and strictly decreasing property of $G(\lambda)$. This completes the proof of the theorem. □

The corollary to Theorem 2 follows because for $\lambda\leq\lambda_L$ and $\lambda\geq\lambda_U$, γ^0 is locally optimal, and the corresponding max-min probability of a correct selection, P^0, given by (2.9) is a strictly increasing function of λ.

Proof of Theorem 3. In (A.10) set $z=y\sqrt{2}$ to yield

$$C(\lambda)=\sqrt{\tfrac{1}{2}\pi}\int_{-\infty}^{\infty} z\Phi^{k-2}\left(\frac{z}{\sqrt{2}}+\frac{\lambda}{2\sqrt{k}}\right)\phi(z)\,dz.$$

Integrate by parts with $\Phi^{k-2}(z/\sqrt{2}+\lambda/2\sqrt{k})=u$ and $-\phi(z)=v$ (and hence $z\phi(z)\,dz=dv$) to yield

$$C(\lambda)=\sqrt{\tfrac{1}{2}\pi}\left\{\left[-\phi(z)\Phi^{k-2}\left(\frac{z}{\sqrt{2}}+\frac{\lambda}{2\sqrt{k}}\right)\right]_{-\infty}^{\infty}\right.$$

$$\left.+\frac{k-2}{\sqrt{2}}\int_{-\infty}^{\infty}\Phi^{k-3}\left(\frac{z}{\sqrt{2}}+\frac{\lambda}{2\sqrt{k}}\right)\phi\left(\frac{z}{\sqrt{2}}+\frac{\lambda}{2\sqrt{k}}\right)\phi(z)\,dz\right\}. \qquad \text{(A.13)}$$

In (A.13) the first term inside the curly brackets is zero. In the second term, note that

$$\phi\left(\frac{z}{\sqrt{2}}+\frac{\lambda}{2\sqrt{k}}\right)\phi(z)=\phi\left(\frac{\lambda}{\sqrt{6k}}\right)\phi\left(\sqrt{\tfrac{3}{2}}z+\frac{\lambda}{2\sqrt{3k}}\right)$$

and make the change of variables $\sqrt{\tfrac{3}{2}}z+\lambda/2\sqrt{3k}=y$, i.e., $z/\sqrt{2}+\lambda/2\sqrt{k}=y/\sqrt{3}+\lambda/3\sqrt{k}$, to obtain

$$C(\lambda)=\sqrt{\tfrac{1}{6}\pi}(k-2)\phi\left(\frac{\lambda}{\sqrt{6k}}\right)\int_{-\infty}^{\infty}\Phi^{k-3}\left(\frac{y}{\sqrt{3}}+\frac{\lambda}{3\sqrt{k}}\right)\phi(y)\,dy$$

$$=\sqrt{\tfrac{1}{6}\pi}(k-2)\phi(\tau)\Phi_{k-3}(\tau/\sqrt{2}\mid\tfrac{1}{4}). \qquad \text{(A.14)}$$

Here $\tau=\lambda/\sqrt{6k}$, and the middle step above follows from the identity

$$\int_{-\infty}^{\infty} \Phi^{p}(ay+b)\phi(y)\,\mathrm{d}y=\Phi_{p}\left(\frac{b}{\sqrt{1+a^{2}}}\,\middle|\,\frac{a^{2}}{1+a^{2}}\right)$$

where $p>0$ is an integer, a and b are arbitrary reals, and the notation $\Phi_p(x\mid\varrho)$ is defined in Section 3.

In the same way it can be shown that

$$D(\lambda)=\sqrt{\pi}\,\Phi_{k-2}(\tau\mid\tfrac{1}{3}). \tag{A.15}$$

Finally substituting (A.14) and (A.15) in (A.9), and the resulting expression for $G(\lambda)$ in equation (3.6) leads to equation (3.9). This completes the proof of the theorem. □

Acknowledgement

We are indebted to Dr. Prakash Awate who contributed in the early stages of this work while he was a graduate student at Cornell University.

References

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.

Bechhofer, R.E. (1969). Optimal allocation of observations when comparing several treatments with a control. In: P.R. Krishnaiah, Ed., *Multivariate Analysis, II.* Academic Press, New York, 465–473.

Berger, J.O. and J. Deely (1988). A Bayesian approach to ranking and selection of related means with alternatives to analysis-of-variance methodology. *J. Amer. Statist. Assoc.* **83**, 364–373.

Dudewicz, E.J. and S.R. Dalal (1975). Allocation of observations in ranking and selection with unequal variances. *Sankhyā Ser. B.* **37**, 28–78.

Gibbons, J.D., I. Olkin and M. Sobel (1977). *Selecting and Ordering Populations.* John Wiley, New York.

Gupta, S.S. and K.J. Miescke (1988). On the problem of finding the largest mean under heteroscedasticity. In: S.S. Gupta and J.O. Berger, Eds., *Statistical Decision Theory and Related Topics, IV, Vol. 2,* 37–49.

Hall, W.J. (1959). The most economical character of some Bechhofer and Sobel decision rules. *Ann. Math. Statist.* **30**, 964–969.

Santner, T.J. and A.C. Tamhane (1984). Designing experiments for selecting a normal population with a large mean and a small variance, In: T.J. Santner and A.C. Tamhane, Eds., *Design of Experiments: Ranking and Selection.* Marcel Dekker, New York, 179–198.

Tong, Y.L. and D.E. Wetzell (1984). Allocation of observations for selecting the best normal population. In: T.J. Santner and A.C. Tamhane, Eds., *Design of Experiments: Ranking and Selection.* Marcel Dekker, New York, 213–224.