A Two-Stage Elimination Type Selection Procedure for Stochastically Increasing Distributions: with an Application to Scale Parameters Problem

by

Seung-Ho Lee *
Purdue University and Ajou University

Technical Report #87-35

Department of Statistics
Purdue University

August 1987

* Research supported in part by the Korea Science & Engineering Foundation, Daejeon 171, Korea. This research was done while the author was visiting the Department of Statistics at Purdue University.

★

ABSTRACT

The purpose of this paper is to extend the idea of Tamhane and Bechhofer (1977, 1979) concerning the normal means problem to some general class of distributions. The key idea in Tamhane and Bechhofer is the derivation of the computable lower bounds on the probability of a correct selection. To derive such lower bounds, they used the specific covariance structure of a multivariate normal distribution. It is shown that such lower bounds can be obtained for a class of stochastically increasing distributions under certain conditions, which is sufficiently general so as to include the normal means problem as a special application. As an application of the general theory to the scale parameters problem, a two-stage elimination type procedure for selecting the population associated with the smallest variance from among several normal populations is proposed. The design constants are tabulated and the relative efficiencies are computed.

1. Introduction

If k populations $\pi_1, \pi_2, \ldots, \pi_k$ are given and we wish to decide on the basis of a properly chosen sampling scheme which one of these populations is the best one, various approaches and methods have been studied up to now. A more detailed overview is provided by Gupta and Panchapakesan (1979). Among those, two-stage procedures with screening in the first stage seem to be quite appropriate, since they are more economical than single stage procedures but still technically not so complicated as sequential ones.

Cohen (1959) was the first to combine Gupta's (1956) maximum mean procedure in the first stage and Bechhofer's (1954) natural decision procedure in the second stage. Later, Alam (1970) proposed a minimax criterion in determining design constants of such two-stage procedures. But these results were mostly confined to the special case of $k = 2$ normal populations with a common known variance. Tamhane and Bechhofer (1977, 1979) extended Alam's (1970) work to the general case of $k \geq 2$ populations with some optimization criterion. They studied in detail a two-stage elimination type procedure using a u-minimax design criterion, and the two-stage procedure was found to be more efficient than the single-stage procedure of Bechhofer (1954). It should be noted that their work was also restricted to the normal means problem. For the normal means problem with a common unknown variance, it is well known that there does not exist any single-stage procedure satisfying the required minimum probability of correct selection. Bechhofer, Dunnett and Sobel (1954) were the first to use Stein's (1945) idea in devising two-stage selection procedures. Unlike the case with a known variance, they estimate the unknown variance at the first stage and select the sample best as the true best in the second stage. Later, Tamhane (1976), and Hochberg and Marcus (1981) considered three-stage procedures with the second stage set for elimination. Gupta and Kim (1984) proposed a two-stage procedure, in which they estimate the unknown variance and eliminate the bad ones in the first stage. For the two parameter exponential populations with a common unknown scale parameter, Desu, Narulla and Villareal (1977) studied a non-elimination type two-stage procedure, and Lee and Kim (1985) proposed an elimination type two-stage procedure.

For the two-stage procedures mentioned so far, optimality results are missing up to now since they are hard to find. The situations become somewhat fairer if there exits a control population. Miescke (1980) studies two-stage procedures for finding populations better than a control in the framework of Neyman-Pearson theory, and showed that optimality of tests carries over to optimality of two-stage procedures. The recent references concerning multi-stage selection procedures can be found in Miescke (1982) and Gupta (1985).

The purpose of this paper is to extend the idea of Tamhane and Bechhofer (1977, 1979) concerning the normal means problem to some general class of distributions and to illustrate the extended theory by using some specific examples. The key idea in Tamhane and Bechhofer is the derivation of the computable lower bounds on the probability of correct selection over the preference-zone. To derive such lower bounds, they used the specific covariance structure of a multivariate normal distribution which heavily depends on the normality assumption. However, it is found that such lower bounds can be obtained for a class of stochastically increasing distributions under certain conditions, which is sufficiently general so as to include the normal means problem as a special application.

In Section 2, the formulation of the problem is given. A two-stage elimination type procedure for selecting the largest parameter value and a design criterion following the lines of Tamhane and Bechhofer (1977, 1979) are described. The main analytical results are contained in Section 3 and 4 which deal with the probability of correct selection and the expected total sample size, respectively. Section 5 treats a dual problem of selecting the population with the smallest parameter value. As an application of the general theory to the scale parameters problem, the problem of selecting the population associated with the smallest variance from among several normal populations is treated in Section 6. The design constants are tabulated and the relative efficiences of the two-stage procedures with respect to the corresponding single-stage procedures are computed.

2. A two-stage procedure and a design criterion

Let $\pi_i (1 \leq i \leq k)$ be k populations, where the probability distribution of π_i depends only on an unknown parameter θ_i in an interval Θ of the real line $(1 \le i \le k)$. Let $\theta_{[1]} \le \theta_{[2]} \le \ldots \le \theta_{[k]}$ denote the ordered values of the unknown parameters $\theta_1, \theta_2, \ldots, \theta_k$. We assume that the correct pairing between θ_i and $\theta_{[i]}$ is unknown. Any population associated with the largest parameter value $\theta_{[k]}$ is called the "best" population.

Following Santner (1975), an indifference-zone will be defined in the entire parameter space $\Omega = \{\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) | \theta_i \in \Theta, 1 \le i \le k\}$ by means of a real valued function δ on Θ having the following properties:

(i) $\delta(\theta) < \theta$ for all $\theta \in \Theta$

(ii) δ, restricted on Θ', is a function onto Θ where $\Theta' = \{\theta \in \Theta | \delta(\theta) \in \Theta\}$.

Define the so-called preference-zone by

$$\Omega(\delta) = \{\underline{\theta} \in \Omega | \theta_{[k-1]} \le \delta(\theta_{[k]})\} \tag{2.1}$$

where the best and the second best are sufficiently far apart so that the experimenter desires to insure the detection of the best with high probability. The complement of $\Omega(\delta)$ is called the indifference-zone. The following preference-zones have been used in the literatures of selection and ranking.

Example 2.1 (a) A location type preference-zone defined by $\delta_1(\theta) = \theta - \delta^*(\delta^* > 0)$, i.e.,

$$\Omega(\delta^*) = \{\underline{\theta} | \theta_{[k]} - \theta_{[k-1]} \ge \delta^*\} \tag{2.2}$$

(b) A scale type preference-zone defined by $\delta_2(\theta) = \theta / \delta^*(\delta^* > 1)$, i.e.,

$$\Omega(\delta^*) = \{\underline{\theta} | \theta_{[k]} \ge \delta^* \theta_{[k-1]}\} \tag{2.3}$$

The goal of the experimenter is to select the best population. The event of correctly selecting the best population is denoted by CS. Following the indifference-zone approach, the attention is restricted to selection procedures $\mathcal{R}$ which guarantee the basic probability requirement,

$$P_{\underline{\theta}}\{CS | \mathcal{R}\} \ge P^* \text{ for all } \underline{\theta} \in \Omega(\delta) \tag{2.4}$$

where $P^*(1/k < P^* < 1)$ is specified prior to the experiment.

Procedures and a Design Criterion

We describe two-stage elimination type selection procedures and a design criterion. At the first stage, the noncontending populations will be screened out using the statistics $T_i^{(1)} = T(X_{i,1}, \ldots, X_{i,n_1})$ $(1 \le i \le k)$ based on n_1 independent observations $X_{i,1}, \ldots, X_{i,n_1}$ from each of $\pi_i (1 \le i \le k)$. At the second stage, we compute the statistics $T_i^{(2)} = T(X_{i,n_1+1}, \ldots, X_{i,n_1+n_2})$ based on n_2 additional independent observations from each of the retained populations, and selection is made using the statistics $T_i = u(T_i^{(1)}, T_i^{(2)})$ based on the overall sample with an appropriate function u.

Evidently the screening process will be done using the following Gupta-type procedure:

Include π_i in the retained populations if and only if $h(T_i^{(1)}) \ge \max_{1 \le j \le k} T_j^{(1)}$, where $h(\cdot)$ is a real valued function such that

(1) $h(x) > x$ for each x, and

(2) $h(x)$ is continuous and strictly increasing in x.

Typical examples of $h(\cdot)$ are given by $h(x) = x + d (d > 0)$ and $h(x) = cx (c > 1)$ for location type and scale type procedures, respectively.

Now, the precise definition of a two-stage elimination type procedure $\mathcal{R}$ is given as follows:

(Stage 1) Take n_1 independent observations $X_{i,1}, \ldots, X_{i,n_1}$ from each $\pi_i (1 \le i \le k)$ and compute $T_i^{(1)} = T(X_{i,1}, \ldots, X_{i,n_1})$. We define a set I by

$$I = \{i | h(T_i^{(1)}) \ge \max_{1 \le j \le k} T_j^{(1)}, 1 \le i \le k\} \tag{2.5}$$

and let $|I|$ denote the number of elements in I.

(a) If $|I| = 1$, stop sampling and assert that the population associated with $\max_{1 \le j \le k} T_j^{(1)}$ is the best.

(b) If $|I| \geq 2$, proceed to the second stage.

(Stage 2) Take n_2 additional independent observations $X_{i,n_1+1}, \ldots, X_{i,n_1+n_2}$ from each population in $\{\pi_i | i \in I\}$, and compute $T_i = u(T_i^{(1)}, T_i^{(2)})$ where $T_i^{(2)} = T(X_{i,n_1+1} \ldots, X_{i,n_1+n_2})$. We then assert that the population associated with $\max_{i \in I} T_i$ is the best.

In the definition of the above two-stage procedure, the sample sizes n_1, n_2 and the function $h(\cdot)$ will be chosen so that the procedure guarantees the basic probability requirement (2.4) and different design criteria lead to different choices. We adopt the following unrestricted minimax criterion:

$$\begin{aligned} &\text{minimize} \ \sup_{\underline{\theta} \in \Omega} E_{\underline{\theta}}(\text{TSS}|\mathcal{R}) \\ &\text{subject to} \ \inf_{\underline{\theta} \in \Omega(\delta)} P_{\underline{\theta}}(CS|\mathcal{R}) \geq P^* \end{aligned} \tag{2.6}$$

where TSS is the total sample size needed in the experiment.

3. Lower bounds on the probability of a correct selection

A main problem concerned with the construction of selection procedures using the indifference-zone approach is to find the infimum of the probability of a correct selection over the preference-zone $\Omega(\delta)$. Any parameter configuration achieving such an infimum is called a least favorable configuration (LFC) for the procedure under study.

However, as can be seen from Alam (1970), Tamhane and Bechhofer (1977, 1979), Miescke and Sehr (1980) and Gupta and Miescke (1982), it has been a conjecture since 1970 that the LFC for the elimination type two-stage procedure would be the slippage one. Recently, Bhandari and Chaudhuri (1987) have produced a proof of it for the normal means problem.

Even if the LFC of the parameters were known, the problem of evaluating the infimum of the probability of a correct selection would still remain and it is extremely difficult and costly to evaluate the probability of a correct selection on a computer. Thus, instead of trying to find the LFC, some lower bounds are used to construct a conservative procedure in this paper as in Tamhane and Bechhofer (1977, 1979). To do so, the following assumptions are made regarding the statistics $T_i^{(1)}$ and $T_i = u(T_i^{(1)}, T_I^{(2)})$ used in the procedure $\mathcal{R}$.

Assumption (A1). The distributions of $T_i^{(1)}$ and $T_i^{(2)}$ are stochastically increasing in $\theta_i \in \Theta$ for $i = 1, 2, \ldots, k$.

Assumption (A2). The function $u(t_1, t_2)$, used to define the statistic $T_i = u(T_i^{(1)}, T_i^{(2)})$, is strictly increasing in each variable.

In the sequel, let $F(\cdot|\theta_i)$ and $G(\cdot|\theta_i)$ denote the cdf's of $T_i^{(1)}$ and T_i, respectively and let $H(t_1, t_2|\theta_i)$ denote the joint cdf of $T_i^{(1)}$ and T_i. Then, it follows from the assumptions (A1) and (A2) that $F(\cdot|\theta_i), G(\cdot|\theta_i)$ and $H(\cdot,\cdot|\theta_i)$ are all non-increasing in $\theta_i (1 \leq i \leq k)$. From this fact, we can have the following result.

Lemma 3.1. Under the assumptions (A1) and (A2), the following inequality holds.

$$\inf_{\underline{\theta} \in \Omega(\delta)} P_{\underline{\theta}}(CS|\mathcal{R}) \geq \inf_{\theta \in \Theta'} A(\theta) \tag{3.1}$$

$$\text{where } A(\theta) = E_\theta \left[H^{k-1} \left(h(T_k^{(1)}), T_k | \delta(\theta) \right) \right]. \tag{3.2}$$

Proof. Without loss of generality, we may assume that $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$. Then, for all $\underline{\theta} \in \Omega(\delta)$,

$$\begin{aligned} P_{\underline{\theta}}(CS|\mathcal{R}) &= P_{\underline{\theta}}\{h(T_k^{(1)}) \geq \max_{1 \leq j \leq k} T_j^{(1)}, T_k = \max_{i \in I} T_i\} \\ &\geq P_{\underline{\theta}}\{h(T_k^{(1)}) \geq T_i^{(1)}, T_k \geq T_i \text{ for all } i = 1, \ldots, k-1\} \\ &= \int \prod_{i=1}^{k-1} H(h(x), y|\theta_i) dH(x, y|\theta_k) \end{aligned} \tag{3.3}$$

Thus, the result follows from the facts that $H(\cdot,\cdot|\theta_i)$ is non-increasing in θ_i and $\theta_i \leq \delta(\theta_k)$ for all $i = 1, \ldots, k-1$, whenever $\underline{\theta} \in \Omega(\delta)$.

The lower bound $A(\theta)$ in (3.1) would be difficult to compute in practice due to the dependence between $T_k^{(1)}$ and T_k. Thus it seems reasonable to find a lower bound for

$A(\theta)$, which is slightly less sharp but more easily computable. Such a lower bound can be obtained by the following result.

Lemma 3.2. Suppose that assumptions (A1) and (A2) hold. Then, for all $\theta \in \Theta'$, we have

$$A(\theta) \geq E_\theta\left[F^{k-1}\left(h(T_k^{(1)})|\delta(\theta)\right)\right]E_\theta\left[G^{k-1}(T_k|\delta(\theta))\right]. \tag{3.4}$$

Proof. The assumption (A2) insures that, for each fixed b, there exists a function $v(\cdot, b)$ such that

$$u(T_i^{(1)}, T_i^{(2)}) \leq b \text{ if and only if}$$
$$T_i^{(1)} \leq v(T_i^{(2)}, b).$$

Thus, we have, for each a and b,

$$\begin{aligned}
&\mathcal{P}_\theta\{T_i^{(1)} \leq a, T_i = u(T_i^{(1)}, T_i^{(2)}) \leq b\}\\
&= E_\theta\left[\mathcal{P}_\theta\{T_i^{(1)} \leq a, T_i^{(1)} \leq v(T_i^{(2)}, b)|T_i^{(2)}\}\right]\\
&\geq E_\theta\left[\mathcal{P}_\theta\{T_i^{(1)} \leq a|T_i^{(2)}\}\mathcal{P}_\theta\{T_i^{(1)} \leq v(T_i^{(2)}, b)|T_i^{(2)}\}\right]\\
&= \mathcal{P}_\theta\{T_i^{(1)} \leq a\}\mathcal{P}_\theta\{T_i \leq b\}
\end{aligned}$$

which in turn implies that

$$\begin{aligned}
&E_\theta\left[H^{k-1}\left(h(T_k^{(1)}), T_k|\delta(\theta)\right)\right]\\
\geq &E_\theta\left[F^{k-1}\left(h(T_k^{(1)})|\delta(\theta)\right)G^{k-1}(T_k|\delta(\theta))\right].
\end{aligned}$$

Since $F(h(T_k^{(1)})|\delta(\theta))$ and $G(T_k|\delta(\theta)) = G(u(T_k^{(1)}, T_k^{(2)})|\delta(\theta))$ are nondecreasing in $T_k^{(1)}$, we have

$$\begin{aligned}
&E_\theta\left[F^{k-1}(h(T_k^{(1)})|\delta(\theta))G^{k-1}(T_k|\delta(\theta))\right]\\
\geq &E_\theta\left[F^{k-1}(h(T_k^{(1)})|\delta(\theta))\right]E_\theta\left[G^{k-1}(T_k|\delta(\theta))\right]
\end{aligned}$$

by Tchebyshev's inequality (see, for example, Hardy, Littlewood and Pólya (1934), p. 43). This completes the proof.

We summarize Lemma 3.1 and Lemma 3.2 into the following theorem.

Theorem 3.1. Under the assumptions (A1) and (A2), the following inequalities hold.

$$\inf_{\underline{\theta}\in\Omega(\delta)} P_{\underline{\theta}}(CS|\mathcal{R}) \geq \inf_{\theta\in\Theta'} A(\theta) \geq \inf_{\theta\in\Theta'} B(\theta) \tag{3.5}$$

where $A(\theta)$ is given by (3.2) and $B(\theta)$ denotes the right hand side of (3.4).

Finally, it should be pointed out that any further simplification of the lower bounds $A(\theta)$ and $B(\theta)$ can not be done without further assumptions on the structure of the statistical model under study. The situation becomes quite simple as can be seen in the following examples.

Example 3.1. (Location Parameters Problem). Suppose that θ_i is a location parameter of the population $\pi_i (1 \leq i \leq k)$ with the preference-zone being as (2.2) in Example 2.1 (a). Suppose further that θ_i is also a location parameter of the distributions of $T_i^{(1)}$ and T_i. Then, for the location-type screening procedure with $h(x) = x + d$, $A(\theta)$ and $B(\theta)$ do not depend on the parameter θ.

In fact,

$$A(\theta) = A(\delta^*) = E_{\theta=0}\left[H^{k-1}\left(T_k^{(1)} + d + \delta^*, T_k + \delta^*\right)\right] \tag{3.6}$$

and

$$B(\theta) = B(\delta^*) = E_{\theta=0}\left[F^{k-1}(T_k^{(1)} + d + \delta^*)\right] E_{\theta=0}\left[G^{k-1}(T_k + \delta^*)\right] \tag{3.7}$$

where F and G denote the cdf's of $T_i^{(1)}$ and T_i, respectively when $\theta_i = 0$ and H denotes the joint cdf of $(T_i^{(1)}, T_i)$ when $\theta_i = 0$.

Remark. As a typical application to the location parameters problem, consider the normal populations $\pi_i's$ with unknown means $\theta_i's$ and a common known variance $\sigma^2 (1 \leq i \leq k)$. Define the two-stage procedure by setting

$$T_i^{(1)} = \overline{X}_i^{(1)} = \sum_{j=1}^{n_1} X_{i,j}/n_i, T_i^{(2)} = \overline{X}_i^{(2)} = \sum_{j=n_1+1}^{n_1+n_2} X_{i,j}/n_2,$$
$$T_i = u(T_i^{(1)}, T_i^{(2)}) = (n_1 T_i^{(1)} + n_2 T_i^{(2)})/(n_1 + n_2) = \overline{X}_i$$

and $h(t) = t + h (h > 0)$, which gives exactly the procedure of Tamhane and Bechhofer (1977, 1979). Clearly, the assumptions (A1) and (A2) hold in this case.

Also, from the definitions of the statistics $T_i^{(1)}, T_i^{(2)}$ and T_i, the corresponding cdf's are given as follows:

$$F(t_1|\delta(\theta)) = P_{\delta(\theta)}\{\overline{X}_i^{(1)} \leq t_1\} = \Phi\{\sqrt{n_1}\ (t_1 - \theta + \delta^*)/\sigma\}$$
$$G(t_2|\delta(\theta)) = P_{\delta(\theta)}\{\overline{X}_i \leq t_2\} = \Phi\{\sqrt{n_1 + n_2}\ (t_2 - \theta + \delta^*)/\sigma\}$$

and

$$H(t_1, t_2|\delta(\theta)) = P_{\delta(\theta)}\{\overline{X}_i^{(1)} \leq t_1, \overline{X}_i \leq t_2\}$$
$$= \Phi_2\left\{\sqrt{n_1}\ (t_1 - \theta + \delta^*)/\sigma, \sqrt{n_1 + n_2}\ (t_2 - \theta + \delta^*)/\sigma|\sqrt{n_1/(n_1 + n_2)}\ \right\}$$

where Φ is the cdf of the standard normal distribution and $\Phi_2\{\cdot, \cdot|\rho\}$ denotes the cdf of the bivariate normal distribution with mean 0, variances 1 and correlation ρ.

Therefore the lower bounds in Theorem 3.1 are given as follows:

$$A(\theta) = E[\Phi_2^{k-1}\{\sqrt{n_1}\ (\overline{X}_k^{(1)} - \theta + \delta^* + h)/\sigma, \sqrt{n_1 + n_2}\ [X_k - \theta + \delta^*)/\sigma|p\}]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi_2^{k-1}\{\sqrt{n_1}\ (\delta^* + h)/\sigma - x, \sqrt{n_1 + n_2}\ \delta^*/\sigma - y|p\}d\Phi_2(x, y|p)$$

and

$$B(\theta) = \int_{-\infty}^{\infty} \Phi^{k-1}\{x + \sqrt{n_1}\ (\delta^* + h)/\sigma\}d\Phi(x) \int_{-\infty}^{\infty} \Phi^{k-1}(y + \sqrt{n_1 + n_2}\ \delta^*/\sigma\}d\Phi(y)$$

with $p = \sqrt{n_1/(n_1 + n_2)}$. The two lower bounds do not depend on the unknown θ, and they are exactly the bounds of Tamhane and Bechhofer (1979), in which the performance of the procedure based on the lower bounds was investigated. The results indicate that the procedures improve upon the single stage procedure of Bechhofer (1954), with the one based on $A(\theta)$ being slightly better than that based on $B(\theta)$. It may be noted that the lower bound $A(\theta)$ in this case can be handled without much difficulty, since the integration involves only a bivariate normal distribution.

Example 3.2. (Scale Parameters Problem). Suppose that θ_i is a scale parameter of the population $\pi_i (1 \leq i \leq k)$. In this case, the preference-zone can be given as that in

Example 2.1 (b). Suppose further that θ_i is a scale parameter of the distributions of $T_i^{(1)}$ and T_i which are assumed to take only positive real numbers. Then, for the scale-type screening procedure with $h(x) = cx(c > 1), A(\theta)$ and $B(\theta)$ are given by

$$A(\theta) = A(\delta^*) = E_{\theta=1}\left[H^{k-1}(c\delta^* T_k^{(1)}, \delta^* T_k)\right]$$

and

$$B(\theta) = B(\delta^*) = E_{\theta=1}[F^{k-1}(c\delta^* T_k^{(1)})]E_{\theta=1}[G^{k-1}(\delta^* T_k)]$$

where F, G and H denote the cdf's of $T_i^{(1)}, T_i$ and $(T_i^{(1)}, T_i)$, respectively when $\theta_i = 1$.

4. Expected total sample size

In order to employ the u-minimax criterion in Section 2, it is necessary to know the set of parameter points in Ω at which the supremum of $E_{\underline{\theta}}(\text{TSS}|\mathcal{R})$ occurs. In this section, it is shown that the supremum is attained when $\theta_1 = \theta_2 = \ldots = \theta_k$, the equal parameter configuration (EPC).

First, we derive a general expression of the expected total sample size. Note that the total sample size (TSS) can be written as

$$\text{TSS} = kn_1 + n_2 S \tag{4.1}$$

where S is the number of populations to be sampled at the second stage, i.e., $S = 0$ if $|I| = 1$ and $S = |I|$ otherwise.

$$\text{Since } E_{\underline{\theta}}(S|\mathcal{R}) = E_{\underline{\theta}}(|I||\mathcal{R}) - P_{\underline{\theta}}\{|I| = 1|\mathcal{R}\}$$

we have

$$\begin{aligned} E_{\underline{\theta}}(S|\mathcal{R}) &= \sum_{i=1}^{k}\left[P_{\underline{\theta}}\left\{h(T_i^{(1)}) \geq \max_{1\leq j\leq k} T_j^{(1)}\right\} - P_{\underline{\theta}}\left\{T_i^{(1)} \geq \max_{j\neq i} h(T_j^{(1)})\right\}\right] \\ &= \sum_{i=1}^{k}\left[\int \prod_{j\neq i} F(h(x)|\theta_j)dF(x|\theta_i) - \int \prod_{j\neq i} F(h^{-1}(x)|\theta_j)dF(x|\theta_i)\right] \end{aligned}$$

Thus, a general expression of $E_{\underline{\theta}}(\text{TSS}|\mathcal{R})$ is given by

$$E_{\underline{\theta}}(\text{TSS}|\mathcal{R}) = kn_1 + n_2 \sum_{i=1}^{k} \left[\int \prod_{j \neq i} F(h(x)|\theta_j) dF(x|\theta_i) - \int \prod_{j \neq i} (h^{-1}(x)|\theta_j) dF(x|\theta_i) \right] \tag{4.2}$$

In order to obtain the maximum value of $E_{\underline{\theta}}(\text{TSS}|\mathcal{R})$, we consider, along the lines of Gupta (1965), a parameter configuration $\theta_1 = \ldots = \theta_q (= \theta) \leq \theta_{q+1} \leq \cdots \leq \theta_k$. Then, for such a $\underline{\theta} \in \Omega$, we have

$$E_{\underline{\theta}}(S|\mathcal{R}) = \left(\sum_{i=1}^{q} + \sum_{i=q+1}^{k} \right) \left[\int \prod_{j \neq i} F(h(x)|\theta_j) dF(x|\theta_i) - \int \prod_{j \neq i} F(h^{-1}(x)|\theta_j) dF(x|\theta_i) \right]$$

$$= \sum_{i=1}^{q} \left\{ \int F^{q-1}(h(x)|\theta) \prod_{j=q+1}^{k} F(h(x)|\theta_j) dF(x|\theta) \right.$$

$$\left. - \int F^{q-1}(h^{-1}(x)|\theta) \prod_{j=q+1}^{k} F(h^{-1}(x)|\theta_j) dF(x|\theta) \right\}$$

$$+ \sum_{i=q+1}^{k} \left\{ \int F^{q}(h(x)|\theta) \prod_{\substack{j=q+1 \\ j \neq i}}^{k} F(h(x)|\theta_j) dF(x|\theta_i) \right.$$

$$\left. - \int qF^{q}(h^{-1}(x)|\theta) \prod_{\substack{j=q+1 \\ j \neq i}}^{k} F(h^{-1}(x)|\theta_j) dF(x|\theta_i) \right\}$$

$$= \int qF^{q-1}(h(x)|\theta) \prod_{j=q+1}^{k} F(h(x)|\theta_j) dF(x|\theta)$$

$$- \int qF^{q-1}(h^{-1}(x)|\theta) \prod_{j=q+1}^{k} F(h^{-1}(x)|\theta_j) dF(x|\theta)$$

$$+ \sum_{i=q+1}^{k} \int F^{q}(h(x)|\theta) \prod_{\substack{j=q+1 \\ j \neq i}}^{k} F(h(x)|\theta_j) dF(x|\theta_i)$$

$$- \sum_{i=q+1}^{k} \int F^{q-1}(h^{-1}(x)|\theta) \prod_{\substack{j=q+1 \\ j \neq i}}^{k} F(h^{-1}(x)|\theta_j) dF(x|\theta_i)$$

$$= I_1(\theta) - I_2(\theta) + I_3(\theta) - I_4(\theta), \text{ say.}$$

We need the following regularity condition to know the behavior of $E_{\underline{\theta}}(S|\mathcal{R})$ as a function of θ.

Regularity Condition (C1). For the cdf $F(x|\theta_i)$ of $T_i^{(1)}$, the partial derivatives $f(x|\theta) = \frac{\partial}{\partial x}F(x|\theta)$ and $\dot{F}(x|\theta) = \frac{\partial}{\partial \theta}F(x|\theta)$ exist for all x and for all $\theta \in \Theta$, and for the function $h(\cdot)$, the derivatives $\dot{h}(x) = \frac{d}{dx}h(x)$ and $\dot{h}^{-1}(x) = \frac{d}{dx}h^{-1}(x)$ exist for all x.

Lemma 4.1. Suppose that $\theta_1 = \theta_2 = \ldots = \theta_q (= \theta) \le \theta_{q+1} \le \ldots \le \theta_k$. Then, under the regularity condition (C1),

$$
\begin{aligned}
&\frac{\partial}{\partial \theta} E_{\underline{\theta}}(S|\mathcal{R}) \\
&= \int q(q-1)F^{q-2}(h(x)|\theta)\left\{\dot{F}(h(x)|\theta)f(x|\theta) - \dot{h}(x)f(h(x)|\theta)\dot{F}(x|\theta)\right\} \\
&\times \prod_{j=q+1}^{k} F(h(x)|\theta_j)dx \\
&+ \sum_{i=q+1}^{k} \int qF^{q-1}(h(x)|\theta)\left\{\dot{F}(h(x)|\theta)f(x|\theta_i) - \dot{h}(x)f(h(x)|\theta_i)\dot{F}(x|\theta)\right\} \\
&\times \prod_{\substack{j=q+1 \\ j\neq i}}^{k} F(h^{-1}(x)|\theta_j)dx \\
&- \int q(q-1)F^{q-2}(h^{-1}(x)|\theta)\left\{\dot{F}(h^{-1}(x)|\theta)f(x|\theta) - \dot{h}^{-1}(x)f(h^{-1}(x)|\theta)\dot{F}(x|\theta)\right\} \\
&\times \prod_{j=q+1}^{k} F(h^{-1}(x)|\theta_j)dx \\
&- \sum_{i=q+1}^{k} \int qF^{q-1}(h^{-1}(x)|\theta)\left\{\dot{F}(h^{-1}(x)|\theta)f(x|\theta_i) - \dot{h}^{-1}(x)f(h^{-1}(x)|\theta_i)\dot{F}(x|\theta)\right\} \\
&\times \prod_{\substack{j=q+1 \\ j\neq i}}^{k} F(h^{-1}(x)|\theta_j)dx.
\end{aligned}
$$

Proof. Using the integration by parts, it can be shown that

$$\frac{\partial}{\partial\theta}I_1(\theta) = \int q(q-1)F^{q-2}(h(x)|\theta)\dot{F}(h(x)|\theta)f(x|\theta)\prod_{j=q+1}^{k}F(h(x)|\theta_j)dx$$
$$+\int qF^{q-1}(h(x)|\theta)\frac{\partial}{\partial\theta}f(x|\theta)\prod_{j=q+1}^{k}F(h(x)|\theta_j)dx$$
$$=\int q(q-1)F^{q-2}(h(x)|\theta)\dot{F}(h(x)|\theta)f(x|\theta)\prod_{j=q+1}^{k}F(h(x)|\theta_j)dx$$
$$-\int q(q-1)F^{q-2}(h(x)|\theta)\dot{h}(x)f(h(x)|\theta)\dot{F}(x|\theta)\prod_{j=q+1}^{k}F(h(x)|\theta_j)dx$$
$$-\sum_{i=q+1}^{k}\int qF^{q-1}(h(x)|\theta)f(h(x)|\theta_j)\dot{h}(x)\dot{F}(x|\theta)\prod_{\substack{j=q+1\\j\neq i}}^{k}F(h(x)|\theta_j)dx,$$

$$\frac{\partial}{\partial\theta}I_2(\theta) = \int q(q-1)F^{q-2}(h^{-1}(x)|\theta)\dot{F}(h^{-1}(x)|\theta)f(x|\theta)\prod_{j=q+1}^{k}F(h^{-1}(x)|\theta_j)dx$$
$$-\int q(q-1)F^{q-2}(h^{-1}(x)|\theta)\dot{h}^{-1}(x)f(h^{-1}(x)|\theta)\dot{F}(x|\theta)\prod_{j=q+1}^{k}F(h^{-1}(x)|\theta_j)dx$$
$$-\sum_{i=q+1}^{k}\int qF^{q-1}(h^{-1}(x)|\theta)f(h^{-1}(x)|\theta_i)\dot{h}^{-1}(x)\dot{F}(x|\theta)\prod_{\substack{j=q+1\\j\neq i}}^{k}F(h^{-1}(x)|\theta_j)dx,$$

$$\frac{\partial}{\partial\theta}I_3(\theta) = \sum_{i=q+1}^{k}\int qF^{q-1}(h^{-1}(x)|\theta)f(x|\theta_i)\dot{F}(h(x)|\theta)\prod_{\substack{j=q+1\\j\neq i}}^{k}F(h(x)|\theta_j)dx$$

and

$$\frac{\partial}{\partial\theta}I_4(\theta) = \sum_{i=q+1}^{k}\int qF^{q-1}(h^{-1}(x)|\theta)\dot{F}(h^{-1}(x)|\theta)f(x|\theta_i)\prod_{\substack{j=q+1\\j\neq i}}^{k}F(h^{-1}(x)|\theta_j)dx.$$

Summing up $\frac{\partial}{\partial\theta}I_i(\theta)(1 \le i \le 4)$, the result is obtained.

We now state the main result of this section in the following theorem.

Theorem 4.1. Suppose that the regularity condition (C1) holds. Then, the supremum of $E_{\underline{\theta}}(\text{TSS}|\mathcal{R})$ is attained whenever $\theta_1 = \theta_2 = \ldots = \theta_k$ provided that

$$\dot{F}(h(x)|\theta_1)f(x|\theta_2) - \dot{h}(x)f(h(x)|\theta_2)\dot{F}(x|\theta_1) \ge 0 \tag{4.3}$$

and

$$\dot{F}(h^{-1}(x)|\theta_1)f(x|\theta_2) - \dot{h}^{-1}(x)f(h^{-1}(x)|\theta_2)\dot{F}(x|\theta_1) \leq 0 \tag{4.4}$$

for all $\theta_1 \leq \theta_2$ and all x. Thus,

$$\sup_{\underline{\theta}\in\Omega} E_{\underline{\theta}}(\text{TSS}|\mathcal{R})$$
$$= kn_1 + kn_2\left\{\int F^{k-1}(h(x)|\theta)dF(x|\theta) - \int F^{k-1}(h^{-1}(x)|\theta)dF(x|\theta)\right\}. \tag{4.5}$$

Proof. Note that, by Lemma 4.1, (4.3) and (4.4) imply $\frac{\partial}{\partial\theta}E_{\underline{\theta}}(S|\mathcal{R}) \geq 0$ for a parameter configuration $\theta_1 = \ldots = \theta_q = \theta \leq \theta_{q+1} \leq \ldots \leq \theta_k$. Hence the supremum of $E_{\underline{\theta}}(S|\mathcal{R})$ is attained whenever $\theta_1 = \theta_2 = \ldots = \theta_k$. Therefore the result follows from (4.1) and (4.2).

It should be remarked that the conditions (4.3) and (4.4) are reduced to the monotone likelihood ratio property of the density $f(x|\theta)$ of $F(x|\theta)$ in location or scale parameters problem in the framework of Examples 3.1 and 3.2.

5. Problem of selecting the smallest

This section treats a dual problem of selecting the population with the smallest parameter value $\theta_{[1]}$. For this selection problem, the preference-zone is given by

$$\Omega(\delta) = \{\underline{\theta} \in \Omega | \theta_{[2]} \geq \delta(\theta_{[1]})\} \tag{5.1}$$

where the real valued function $\delta(\cdot)$ satisfies the following properties:

(i) $\delta(\theta) > \theta$ for all $\theta \in \Theta$

(ii) δ, restricted on Θ', is a function onto Θ' where $\Theta' = \{\theta \in \Theta | \delta(\theta) \in \Theta\}$.

The statistics $T_i^{(1)}, T_i^{(2)}$ and $T_i = u(T_i^{(1)}, T_i^{(2)})$ in Section 2 are used in the construction of a two-stage procedure in this case. In the problem of selecting the population with the smallest parameter the following Gupta-type procedure will be used for screening purpose:

Include π_i in the retained populations if and only if

$$h(T_i^{(1)}) \leq \min_{1\leq j\leq k} T_j^{(1)} \tag{5.2}$$

where $h(\cdot)$ satisfies the following properties:

(i) $h(x) < x$ for all x, and

(ii) $h(x)$ is continuous and strictly increasing in x.

Now, a two-stage selection procedure in this problem can be constructed in exactly the same manner as that in Section 2 except that the screening procedure in (2.5) is replaced by that in (5.2) and obvious modifications for final decision rules are made. Such a selection procedure in this problem will be denoted by $\mathcal{R}'$.

For the procedure $\mathcal{R}'$, the following results can be obtained with some slight modifications of the arguments for the procedure $\mathcal{R}$ in Section 2.

Theorem 5.1. Under the assumptions (A1) and (A2) in Section 3, the following inequalities hold.

$$\inf_{\underline{\theta}\in\Omega(\delta)} \mathcal{P}_{\underline{\theta}}(CS|\mathcal{R}') \geq \inf_{\theta\in\Theta'} A'(\theta) \geq \inf_{\theta\in\Theta'} B'(\theta) \tag{5.3}$$

where $A'(\theta)$ and $B'(\theta)$ are defined by

$$A'(\theta) = E_\theta\left[M^{k-1}(h(T_1^{(1)}), T_1|\delta(\theta))\right] \tag{5.4}$$

and

$$B'(\theta) = E_\theta\left[\left\{1 - F(h(T_1^{(1)})|\delta(\theta))\right\}^{k-1}\right] E_\theta\left[\left\{1 - G(T_1|\delta(\theta))\right\}^{k-1}\right] \tag{5.5}$$

with $M(x, y|\theta_i) = \mathcal{P}_{\theta_i}[T_i^{(1)} > x, T_i > y], 1 \leq i \leq k$.

Theorem 5.2. Suppose that the regularity condition (C1) in Section 4 holds. Then, the supremum of $E_{\underline{\theta}}(\text{TSS}|\mathcal{R}')$ is attained whenever $\theta_1 = \theta_2 = \ldots = \theta_k$ provided (4.3) and (4.4) are satisfied. Thus,

$$\sup_{\underline{\theta}\in\Omega} E_{\underline{\theta}}(\text{TSS}|\mathcal{R}')$$
$$= kn_1 + kn_2\left[\int\{1 - F(h(x)|\theta)\}^{k-1}dF(x|\theta) - \int\{1 - F(h^{-1}(x)|\theta)\}^{k-1}dF(x|\theta)\right] \tag{5.6}$$

Finally, a remark on the case with nuisance parameters should be in order. It can be easily observed that the characterizations of the procedures $\mathcal{R}$ and $\mathcal{R}'$ depend on the parameters only through statistics $T_i^{(1)}$ and $T_i (1 \leq i \leq k)$. Thus, the results obtained so far remain valid as long as the distributions of $T_i^{(1)}$ and T_i do not depend on the nuisance parameters.

6. Normal variances problem

For the problem of selecting the population associated with the smallest variance from among several normal populations, Bechhofer and Sobel (1954) proposed a single-stage procedure $\mathcal{R}_0$, in the framework of the indifference-zone approach. Gupta and Sobel (1962a, 1962b) investigated the same problem under the framework of subset selection. The values of the sample sizes needed in the single-stage procedure $\mathcal{R}_0$ of Bechhofer and Sobel (1954) can also be obtained from the tables of Gupta and Sobel (1962b). An extended tables are also available from Gibbons, Olkin and Sobel (1977). Later, Tamhane (1975) formulated this problem in the two-stage sampling scheme with screening in the first stage and proposed a lower bound on the probability of correct selection. However, due to the computational difficulties involved, no tables were given.

Let $\pi_i (1 \leq i \leq k)$ denote k normal populations with unknown means $\mu_i (-\infty < \mu_i < \infty, 1 \leq i \leq k)$ and unknown variances $\sigma_i^2 (0 < \sigma_i^2 < \infty, 1 \leq i \leq k)$. The ordered variances are denoted by $\sigma_{[1]}^2 \leq \sigma_{[2]}^2 \leq \ldots \leq \sigma_{[k]}^2$. It is assumed that there is no priori information available about the correct pairing between π_i and $\sigma_{[i]}^2 (1 \leq i \leq k)$. The goal is to select a population associated with $\sigma_{[1]}^2$.

It can be easily shown that this problem falls into the framework of Section 5 with $\delta(\sigma^2) = \sigma^2/\delta^*$ $(0 < \delta^* < 1)$, while $\mu_1, \ldots, \mu_k$ are the nuisance parameters.

Let

$$T_i^{(1)} = \sum_{j=1}^{n_1} (X_{i,j} - \overline{X}_i^{(1)})^2$$

$$T_i^{(2)} = \sum_{j=n_1+1}^{n_2} (X_{i,j} - \overline{X}_i^{(2)})^2$$

$$T_i = u(T_i^{(1)}, T_i^{(2)}) = T_i^{(1)} + T_i^{(2)}$$

and

$$h(t) = ct \quad (0 < c < 1)$$

where $\overline{X}_i^{(1)} = \sum_{j=1}^{n_1} X_{i,j}/n$, and $\overline{X}_i^{(2)} = \sum_{j=n_1+1}^{n_2} X_{i,j}/n_2$.

An elimination type two-stage selection procedure $\mathcal{R}_1$ is proposed as follows:

(Stage 1) Take n_1 independent observations $X_{i,1}, \ldots, X_{i,n_1}$ from each $\pi_i (1 \leq i \leq k)$, compute $T_i^{(1)}$ and determine a subset I of $\{1, 2, \ldots, k\}$ where

$$I = \{i | cT_i^{(1)} \leq \min_{1 \leq j \leq k} T_j^{(1)}\}, 0 < c < 1 \tag{6.1}$$

(a) If $|I| = 1$, stop sampling and assert that the population associated with $\min\limits_{1 \leq j \leq k} T_j^{(1)}$ is the best.

(b) If $|I| \geq 2$, proceed to the second stage.

(Stage 2) Take n_2 additional independent observations $X_{i,n_1+1}, \ldots, X_{i,n_1+n_2}$ from each populations in $\{\pi_i | i \in I\}$, compute $T_i = T_i^{(1)} + T_i^{(2)}$ and assert that the population associated with $\min\limits_{j \in I} T_j$ is the best.

Note that $T_i^{(1)}/\sigma_i^2, T_i^{(2)}/\sigma_i^2$ and T_i/σ_i^2 all have the chi-squared distributions with $\nu_1 = n_1 - 1, \nu_2 = n_2 - 1$ and $\nu = \nu_1 + \nu_2 = n_1 + n_2 - 2$ degrees of freedom, respectively. However, the joint distribution of $T_i^{(1)}$ and T_i is rather complicated and inconvenient to compute in this case. Thus, we use the lower bound $B'(\theta)$ in Theorem 5.1 to determine the design constants (n_1, n_2, c) for the two-stage procedure $\mathcal{R}_1$. By straightforward computation,

$$B'(\delta^*) = \int_0^\infty \{1 - F_{\nu_1}(c\delta^* x)\}^{k-1} dF_{\nu_1}(x) \int_0^\infty \{1 - F_\nu(\delta^* y)\}^{k-1} dF_\nu(y) \tag{6.2}$$

where $F_\nu(\cdot)$ denotes the cdf of chi-squared distribution having ν degrees of freedom.

Remark. Tamhane (1975) proposed almost the same procedure as the procedure $\mathcal{R}_1$. The only difference is the statistic T_i used in Stage 2. His T_i is defined by

$$T_i = \sum_{j=1}^{n_1}(X_{i,j} - \overline{X}_i)^2 + \sum_{j=n_1+1}^{n_2}(X_{i,j} - \overline{X}_i)^2$$

where $\overline{X}_i = \sum_{j=1}^{n_1+n_2} X_{i,j}/(n_1+n_2)$. Hence the degrees of freedom of T_i is $\nu = n_1+n_2-1$. When the population mean $\mu_i(1 \leq i \leq k)$ are all known, with the obvious definitions of the statistics, the two procedures are exactly the same. He also derived a lower bound $C(\delta^*)$, say, on the probability of a correct selection, of the form

$$C(\delta^*) = \int_0^\infty \{1 - F_{\nu_1}(c\delta^* x)\}^{k-1} dF_{\nu_1}(x) + \int_0^\infty \{1 - F_\nu(\delta^* x)\}^{k-1} dF_\nu(x) - 1. \quad (6.3)$$

For the same ν_1 and ν (this is the case when all μ_i's are known), $B'(\delta^*) \geq C(\delta^*)$ since $ab \geq a+b-1$ for $a, b \in (0,1)$, and hence $B'(\delta^*)$ is a less conservative lower bound.

The supremum of the expected total sample size can be obtained from Theorem 5.2 and is given as follows.

$$\sup_{\underline{\theta}\in\Omega} E_{\underline{\theta}}(\text{TSS}|\mathcal{R}_1)$$
$$= kn_1 + kn_2\left[\int_0^\infty \{1 - F_{\nu_1}(cx)\}^{k-1} dF_{\nu_1}(x) - \int_0^\infty \{1 - F_{\nu_1}(x/c)\}^{k-1} dF_{\nu_1}(x)\right]. \quad (6.4)$$

Therefore, the corresponding optimization problem to determine the design constants (n_1, n_2, c) is to minimize (6.4) subject to $B'(\delta^*) \geq P^*$. This is an extremely complicated integer programming problem with a non-linear objective function.

In solving the optimization problem, we have treated n_1 and n_2 as continuous variables, and used the SUMT (Sequential Unconstrained Minimization Technique) algorithm of Fiacco and McCormick (1968). A source program in FORTRAN for SUMT algorithm is given by Kuester and Mize (1973). We denote by $(\hat{n}_1, \hat{n}_2, \hat{c})$ a solution to this continuous version of the optimization problem. The problem has been solved numerically for $k = 2(1)\ 10$, $P^* = 0.90, 0.95$ and $\sqrt{\delta^*} = 0.50\ (0.05)\ 0.70$. The results are given in Table 6.1.

In supplying the objective function (6.4) and the constraint function (6.2) to SUMT algorithm, we used the 32-point Laguerre numerical quadrature formula to evaluate the integrals, and the values of the chi-square cdf's were evaluated using the 32-point Legendre numerical quadrature formula. All the computations were carried out in double precision arithmatic on VAX-11/780 of Purdue University's Statistics Department. The convergence criterion was fixed throughout to be 1×10^{-8}. The tabulated values are rounded off in the fourth decimal places for $\hat{n}_1$, and $\hat{n}_2$, and in the sixth decimal places for $\hat{c}$.

The Performance of $\mathcal{R}_1$ relative to $\mathcal{R}_0$.

In order to get insight into the performance of the two-stage procedure $\mathcal{R}_1$, we consider the ratio (termed relative efficiency RE),

$$\text{RE} = \{E_{\underline{\theta}}(\text{TSS}|\mathcal{R}_1)/kn_0\} \times 100(\%) \tag{6.5}$$

where n_0 is the sample size needed for the single-stage procedure of Bechhofer and Sobel (1954) to satisfy the same probability requirement. Clearly RE depends on $\underline{\theta}, (\delta^*, P^*)$ and k.

Since $\mathcal{R}_0$ is a special case of $\mathcal{R}_1$ (with $c = 1$ or ∞), it immediately follows that $1 \geq$ RE (EPC) $\geq$ RE (LFC) and $\mathcal{R}_1$ is uniformly at least as good as $\mathcal{R}_0$. The values of RE are given in Table 6.2.

From Table 6.2 one finds that the relative savings by applying the two-stage procedure $\mathcal{R}_1$ are not dramatic. However, even a small relative saving means a lot in terms of total sample size when k and/or n_1 is moderately large. Also, it can be observed that the relative saving increases as k becomes large. This is in accordance with one's intuition that the screening process would be helpful when k is large.

To illustrate the use of Table 6.1, suppose that $k = 6$ and that the experimenter specifies $\delta^* = (0.7)^2, P^* = 0.90$. Then the design constants necessary to implement the two-stage procedure $\mathcal{R}_1$ are given by $\hat{n}_1 = 17.775, \hat{n}_2 = 20.145$ and $\hat{c} = 0.62632$. Thus we take $n_1 = 18$ observations from each populations and compute the sample variances $S_i^2 (1 \leq i \leq 6)$. If the number of S_i^2's smaller than $\min_{1 \leq i \leq 6} S_i^2/0.62632$ is one, stop sampling and assert that the population associated with $\min S_i^2$ is the best. If more than one S_i^2's are smaller than $\min S_i^2/0.62632$, take $n_2 = 21$ additional observations from each of the contending populations and assert that the population associated with the smallest sample variance based on the pooled sample of size $n_1 + n_2 = 39$ is the best. In using this two-stage procedure $\mathcal{R}_1$, the average value of the total number of observations is 90.4% at EPC and 77.0% at LFC compared with that of the single-stage procedure $\mathcal{R}_0$ of Bechhofer and Sobel (1954).

Large Sample Approximation

In solving the optimization problem involving (6.4) and (6.2), it is extremely tedious to compute the integrals when n_1 and/or n_2 are large. Hence an approximate solution for large sample size is useful. We shall give an approximate solution to the problem based on normal theory.

It is well known that if S^2 is the sample variance associated with the variance σ^2, then $\sqrt{(\nu-1)/2}\ log(S^2/\sigma^2)$ is asymptotically normally distributed with mean zero and variance unity as the number of degrees of freedom ν, associated with S^2, tends to infinity. From this fact, it can be shown that, when ν is large

$$\int_0^\infty \{1 - F_\nu(ax)\}^{k-1} dF_\nu(x) \cong \int_{-\infty}^{\infty} \Phi^{k-1}(x+d)\, d\Phi(x) \tag{6.6}$$

where $d = \sqrt{(\nu-1)/2}\ log(a^{-1})$, and $F_\nu(\cdot)$ is the cdf of chi-square distribution with ν degrees of freedom and $\Phi(\cdot)$ is the cdf of the standard normal distribution.

Replacing the integrals in (6.2) and (6.4) involving the chi-square cdf's by the corresponding integrals of the right hand side of (6.6) involving the normal cdf, we can obtain after slight modifications the following asymptotic version of the optimization problem:

Minimize

$$kc_1^2 + kc_2^2 \int_{-\infty}^{\infty} \{\Phi^{k-1}(x+d) - \Phi^{k-1}(x-d)\}\, d\Phi(x) \tag{6.7}$$

subject to

$$\int_{-\infty}^{\infty} \Phi^{k-1}(x+d+c_1)\, d\Phi(x) \int_{-\infty}^{\infty} \Phi^{k-1}\left(y + \sqrt{c_1^2 + c_2^2}\right) d\Phi(x) \geqq P^* \tag{6.8}$$

If we denote the solutions to (6.7) and (6.8) by $(\hat{c}_1, \hat{c}_2, \hat{d})$, then the approximate values of the design constants $(\hat{n}_1, \hat{n}_2, \hat{c})$ of the procedure $\mathcal{R}_1$ are computed using the following formulas:

$$\hat{n}_1 = 2\left\{\frac{\hat{c}_1}{log(\delta^{*-1})}\right\}^2 + 2log(k-1),$$

$$\hat{n}_2 = 2\left\{\frac{\hat{c}_2}{log(\delta^{*-1})}\right\}^2 + 2log(k-1)$$

and

$$\hat{c} = exp\left\{-\hat{d}\sqrt{2/(n_1-2)}\right\}.$$

The second term in the formula for $\hat{n}_1$ (or $\hat{n}_2$) is a slight correction term based on empirical results cited from Gibbons, Olkin and Sobel (1977). The correction term is added since the first term drifts below the true value of $\hat{n}_1$ (or $\hat{n}_2$) as k increases.

The values of $(\hat{c}_1, \hat{c}_2, \hat{d})$ can be found in the tables of Tamhane and Bechhofer (1979). To illustrate numerically the closeness of the normal approximation we take the values of $(\hat{c}_1, \hat{c}_2, \hat{d})$ out of Table II of Tamhane and Bechhofer (1979) corresponding to $P^* = 0.95, k = 10$, namely, $\hat{c}_1 = 2.452, \hat{c}_2 = 2.744$ and $\hat{d} = 1.322$. Then the approximate values of $(\hat{n}_1, \hat{n}_2, \hat{c})$ for $\delta^* = (0.7)^2$ are (28.025, 33.988, 0.69317). These approximate values are slightly larger than the corresponding exact values (27.292, 31.970, 0.69114) given in Table 6.1.

Acknowledgement:

The author wishes to thank to Prof. Shanti S. Gupta, Woo-Chul Kim and Dr. Tachen Liang for their extensive comments for this work. Also thanks are to the Department of Statistics, Purdue University for giving me an opportunity to visit during 1986-87 and for providing facilities for my research.

Table 6.1
Design Constants for the Two-Stage Procedure $\mathcal{R}_1$

k = 2

	$P^* = 0.90$			$P^* = 0.95$		
$\sqrt{\delta^*}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$
0.50	4.672	6.029	0.99999	6.544	6.354	0.92158
0.55	5.365	3.894	0.76742	6.178	3.825	0.46474
0.60	5.039	2.211	0.17160	7.771	4.964	0.45385
0.65	6.453	3.426	0.21494	10.181	6.975	0.47896
0.70	8.621	5.450	0.27360	13.960	10.354	0.52255

k = 3

	$P^* = 0.90$			$P^* = 0.95$		
$\sqrt{\delta^*}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$
0.50	6.208	6.542	0.90887	6.901	5.003	0.63690
0.55	5.746	3.936	0.43711	8.328	5.527	0.56113
0.60	7.204	4.982	0.42234	10.606	7.296	0.56670
0.65	9.417	6.829	0.44267	14.037	10.107	0.59111
0.70	12.201	10.822	0.47227	19.426	14.711	0.62732

k = 4

	$P^* = 0.90$			$P^* = 0.95$		
$\sqrt{\delta^*}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$
0.50	5.551	4.649	0.54605	7.459	5.377	0.57673
0.55	6.721	5.609	0.51311	9.313	6.936	0.58219
0.60	8.479	7.240	0.51621	11.950	9.270	0.60041
0.65	11.119	9.822	0.53815	15.897	12.840	0.62928
0.70	15.293	14.016	0.57488	21.727	18.912	0.66103

Table 6.1
Design Constants for the Two-Stage Procedure $\mathcal{R}_1$
(continued)

k = 5

	$P^* = 0.90$			$P^* = 0.95$		
$\sqrt{\delta^*}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$
0.50	5.899	5.547	0.53513	7.850	6.194	0.56399
0.55	7.287	6.799	0.52954	9.880	8.087	0.58431
0.60	9.264	8.923	0.54652	12.733	10.866	0.60955
0.65	12.202	12.165	0.57446	16.992	15.071	0.64215
0.70	16.754	17.388	0.60969	23.766	21.753	0.68249

k = 6

	$P^* = 0.90$			$P^* = 0.95$		
$\sqrt{\delta^*}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$
0.50	6.138	6.034	0.51152	8.128	6.906	0.55542
0.55	7.663	7.763	0.53191	10.260	9.061	0.58159
0.60	9.787	10.273	0.55662	13.257	12.206	0.61114
0.65	12.943	14.057	0.58952	17.737	16.930	0.64665
0.70	17.775	20.145	0.62632	24.894	24.289	0.68800

k = 7

	$P^* = 0.90$			$P^* = 0.95$		
$\sqrt{\delta^*}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$
0.50	6.238	6.640	0.49599	8.345	7.534	0.54877
0.55	7.940	8.589	0.53035	10.537	9.900	0.57717
0.60	10.174	11.410	0.55998	13.643	13.344	0.60980
0.65	13.488	15.661	0.59628	18.563	18.491	0.65756
0.70	18.569	22.392	0.63444	25.741	26.644	0.69189

Table 6.1
Design Constants for the Two-Stage Procedure $\mathcal{R}_1$
(continued)

k = 8

	$P^* = 0.90$			$P^* = 0.95$		
$\sqrt{\delta^*}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$
0.50	6.440	7.118	0.48408	8.528	8.075	0.54346
0.55	8.157	9.308	0.52725	10.754	10.640	0.57246
0.60	10.477	12.403	0.56057	13.946	14.345	0.60751
0.65	13.923	17.032	0.59933	19.344	20.491	0.66845
0.70	19.214	24.432	0.64026	26.431	28.944	0.69637

k = 9

	$P^* = 0.90$			$P^* = 0.95$		
$\sqrt{\delta^*}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$
0.50	6.504	7.693	0.47554	8.685	8.558	0.53907
0.55	8.335	9.949	0.52377	10.931	11.297	0.56789
0.60	10.724	13.274	0.55959	14.192	15.237	0.60478
0.65	14.269	18.254	0.60020	19.834	21.536	0.66514
0.70	19.741	26.317	0.64419	26.924	30.716	0.69566

k = 10

	$P^* = 0.90$			$P^* = 0.95$		
$\sqrt{\delta^*}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$	$\hat{n}_1$	$\hat{n}_2$	$\hat{c}$
0.50	6.609	8.126	0.46947	8.822	8.991	0.53317
0.55	8.484	10.524	0.52004	11.085	11.888	0.56385
0.60	10.932	14.058	0.55798	14.399	16.037	0.60188
0.65	14.560	19.347	0.60004	20.013	22.682	0.66819
0.70	20.125	27.832	0.64840	27.292	31.970	0.69114

Table 6.2
Relative Efficiencies RE of the Procedure $\mathcal{R}_1$

$P^* = 0.90$

k	$\sqrt{\delta^*}$ = 0.50 EPC	0.50 LFC	0.55 EPC	0.55 LFC	0.60 EPC	0.60 LFC	0.65 EPC	0.65 LFC	0.70 EPC	0.70 LFC
2	99.9	94.5	99.9	97.6	99.9	94.9	96.3	91.2	97.5	92.0
3	98.1	92.7	98.3	85.2	95.6	83.6	95.9	84.5	97.5	86.0
4	99.5	85.7	92.7	79.5	94.7	81.3	94.0	80.8	96.4	83.0
5	95.8	82.3	91.9	78.8	95.7	81.9	88.4	75.6	94.3	80.5
6	91.0	78.4	89.4	76.7	89.3	76.5	93.1	79.5	90.4	77.0
7	93.4	80.7	86.2	74.2	87.7	75.2	88.8	76.0	88.5	75.5
8	88.4	76.5	88.5	76.3	85.5	73.5	84.6	72.4	86.2	73.5
9	89.9	77.9	79.6	68.9	87.3	75.2	86.4	74.2	85.8	73.4
10	84.3	73.2	80.9	70.1	84.6	73.0	81.9	70.4	82.5	70.7

$P^* = 0.95$

k	$\sqrt{\delta^*}$ = 0.50 EPC	0.50 LFC	0.55 EPC	0.55 LFC	0.60 EPC	0.60 LFC	0.65 EPC	0.65 LFC	0.70 EPC	0.70 LFC
2	99.9	96.0	99.9	85.8	99.9	84.3	99.9	85.2	98.4	83.4
3	99.9	86.8	99.7	82.7	98.1	81.2	96.6	79.9	96.9	80.0
4	94.9	79.1	95.5	79.1	97.5	80.3	91.9	75.4	93.6	76.5
5	92.6	77.1	94.8	78.6	93.7	77.2	91.0	74.7	92.6	75.7
6	89.1	74.4	92.7	76.9	89.0	73.4	88.7	72.9	89.8	73.6
7	91.5	76.5	90.0	74.8	87.7	72.5	85.8	70.8	86.5	71.0
8	87.1	73.0	87.0	72.5	86.0	71.3	83.6	69.0	86.6	71.2
9	88.5	74.4	84.0	70.2	84.1	69.9	85.4	70.7	84.5	69.6
10	84.2	70.9	81.1	67.9	85.2	71.0	81.0	67.4	82.1	68.0

REFERENCES

[1] Alam, K. (1970). A two-sample procedure for selecting the population with the largest mean from k normal populations. *Ann. Inst. Statist. Math.*, **22**, 127–136.

[2] Bhandari, S. K. and Chaudhuri, A. R. (1987). On two conjectures about two-stage selection problem. (to appear).

[3] 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.

[4] Bechhofer, R. E., Dunnett, C. W. and Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika*, **41**, 170–176.

[5] Bechhofer, R. E. and Sobel, M. (1954). A single-sample multiple decision procedure for ranking variances of normal populations. *Ann. Math. Statist.*, **25**, 273–289.

[6] Cohen, D. S. (1959). A two-sample decision procedure for ranking means of normal populations with a common known variance. M.S. Thesis, Dept. of Operations Research, Cornell Univ., Ithaca, New York.

[7] Desu, M. M., Narulla, S. C. and Villarreal, B. (1977). A two-stage procedure for selecting the best of k exponential distributions. *Commun. Statist.-Theor. Meth.* **A6(12)**, 1223–1230.

[8] Fiacco, A. V. and McCormick, G. P. (1968). *Nonlinear Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, Inc., New York.

[9] Gibbons, J. D., Olkin, I. and Sobel, M. (1977). *Selecting and Ordering Populations: A New Statistical Methodology.* John Wiley and Sons, Inc., New York.

[10] Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Ph.D. Thesis (Mimeo. Ser. No. 150). Inst. of Statist., Univ. of North Carolina, Chapel Hill.

[11] Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics*, **7**, 225–245.

[12] Gupta, S. S. (1985). A short bibliography of recent developments in selection and ranking procedures. Tech. Rep. No. 85-11, Dept. of Statistics, Purdue Univ., West Lafayette, Indiana.

[13] Gupta, S. S. and Kim, W. -C. (1984). A two-stage elimination type procedure for selecting the largest of several normal means with a common unknown variance. *Design of Experiments: Ranking and Selection* (T. J. Santner and A. O. Tamhane, eds.), Marcel Dekker, New York, 77–94.

[14] Gupta, S. S. and Miescke, K. J. (1982). On the least favorable configurations in certain two-stage selection procedures. *Statistics and Probability: Essays in Honor of C. R. Rao.* (G. Kallianpur et al., eds.), Amsterdam - New York - Oxford, North-Holland, 295–305.

[15] Gupta, S. S. and Panchapakesan, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*, John Wiley and Sons, Inc., New York.

[16] Gupta, S. S. and Sobel, M. (1962a). On selecting a subset containing the population with the smallest variance. *Biometrika*, **49**, 495–507.

[17] Gupta, S. S. and Sobel, M. (1962b). On the smallest of several correlated F-statistics. *Biometrika*, **49**, 509–523.

[18] Hardy, G. H., Littlewood, J. E. and Polyá, G. (1934). *Inequalities.* Cambridge Univ. Press, Cambridge.

[19] Hochberg, Y. and Marcus, R. (1981). Three stage elimination type procedures for selecting the best normal population when variances are unknown. *Commun. Statist.-Theor. Meth.*, **A10**, 597–612.

[20] Kuester, J. L. and Mize, J. H. (1973). *Optimization Techniques with Fortran*, McGraw-Hill, Inc.

[21] Lee, S. -H. and Kim, W. -C. (1985). An elimination type two-stage selection procedure for exponential distributions. *Commun. Statist.-Theor. Meth.*, **A14**, 2563–2571.

[22] Miescke, K. J. (1980). On two-stage procedures for finding a population better than a control. Mimeo. Ser., No. 80-28, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

[23] Miescke, K. J. (1982). Recent results on multi-stge selection procedures. Tech. Rep. No. 82-25, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

[24] Miescke, K. J. and Sehr, J. (1980). On a conjecture concerning least favorable configurations in certain two-stage selection procedures. *Commun. Statist.-Theo. Meth.*, **A9**, 1609–1617.

[25] Santner, T. J. (1975). A restricted subset selection approach to ranking and selection problems. *Ann. Statist.*, **3**, 334–349.

[26] Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.*, **16**, 243-258.

[27] Tamhane, A. C. (1975). A minimax two-stage permanent elimination type procedure for selecting the smallest normal variance. Tech. Report 260, Dept. of Operations Res., Cornell Univ., Ithaca, New York.

[28] Tamhane, A. C. (1976). A three-stage elimination type procedure for selecting the largest normal mean (common unknown variance). *Sankhyá*, **B38**, 339–349.

[29] Tamhane, A. C. and Bechhofer, R. E. (1977). A two-stage minimax procedure with screening for selecting the largest normal mean. *Commun. Statist.-Theor. Meth.*, **A6**, 1003–1033.

[30] Tamhane, A. C. and Bechhofer, R. E. (1979). A two-stage minimax procedure with screening for selecting the largest normal mean (II): an improved PCS lower bound and associated tables. *Commun. Statist.-Theor. Meth.*, **A8**, 337–358.