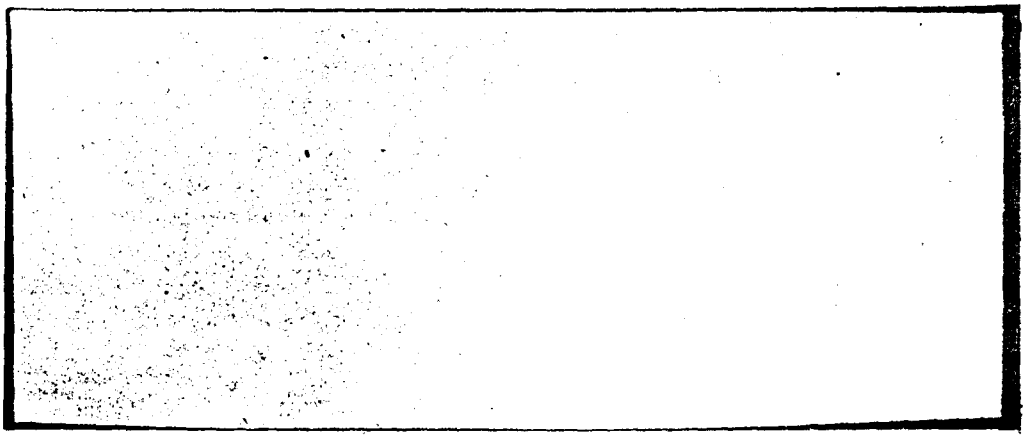

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

R

92

Unclassified

Security Classification

DOCUMENT CONTROL DATA · R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Purdue University	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

On Multiple Decision (Subset Selection) Procedures

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, December 1971

5. AUTHOR(S) (Last name, first name, initial)

Gupta, Shanti S. and Panchapakesan, S.

6. REPORT DATE December 1971	7a. TOTAL NO. OF PAGES 90	7b. NO. OF REFS 73
8a. CONTRACT OR GRANT NO. N00014-67-A-0226-00014 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Mimeo Series #273 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research Washington, D. C.

13. ABSTRACT

This report is a survey of developments and significant results in the area of multiple decision procedures under the subset selection formulation. Section 2 deals with procedures for location and scale parameters. A general theory of the subset selection problem and a decision-theoretic formulation are discussed in Section 3. Sections 4 through 7 deal with parametric and non-parametric procedures for discrete populations, multinomial cells and multivariate normal populations using single-stage sampling, inverse sampling and sequential sampling. Section 8 describes procedures applicable to restricted families of distributions such as the increasing failure rate (IFR) and increasing failure rate on the average (IFRA) distributions. Bayes and empirical Bayes procedures are discussed in Section 9. The last section summarizes briefly several modifications of the basic problem and goal.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

On Multiple Decision (Subset Selection) Procedures

by

Shanti S. Gupta and S. Panchapakesan

Purdue University

and

Southern Illinois University

Department of Statistics
Division of Mathematical Science
Mimeograph Series #273

December 1971

---

*Research supported in part by the Office of Naval Research Contract N00014-67-A-0226-00014 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

On Multiple Decision (Subset Selection) Procedures*

Shanti S. Gupta
Purdue University

and

S. Panchapakesan
Southern Illinois University

1. Introduction

In many of the experimental situations the experimenter is confronted with the problem of making decisions regarding k populations, which, for example, may be categories of wheat, manufactured items coming out of k factories or candidates who are contenders for an award. The classical tests of homogeneity which have been applied in these situations do not supply the information the experimenter really seeks, whether or not the tests yield significant results. In fact, the experimenter's problems begin when he obtains a significant result which goes to reject the null hypothesis that the populations are identical. As a partial answer to the need for a more realistic formulation overcoming the inadequacy of the tests of homogeneity, Mosteller (1948) tested homogeneity against slippage alternatives. Since then many authors have contributed to the theory of slippage tests.

The initial efforts in the direction of multiple decision problems were made by Paulson (1949) who considered the problem of classifying the given populations into a "superior" and an "inferior" group. Later he (1952) investigated the problem of selecting the "best" of k categories when comparing (k-1) experimental categories with a standard or control. Bahadur (1950)

---

*Research supported in part by the Office of Naval Research Contract N00014-67-A-0226-00014 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

has made some early contributions to the theory of k sample problems. Bahadur and Robbins (1950) obtained some minimax rules for selecting from two populations the one with the greater mean. The multiple decision problems that are now known as the ranking and selection problems have been formulated mainly in two ways. The first one is known as the indifference zone formulation due to Bechhofer (1954). This formulation, in its simplest form, selects one of the populations as the best with a guarantee that the true best population is selected with at least a preassigned probability P^* whenever the best and the second best populations are "sufficiently" far apart. For an exposition of this formulation the reader is referred to the excellent monograph by Bechhofer, Kiefer and Sobel (1968). The main investigations surveyed in the present paper are under the second formulation due to Gupta (1956) known as the subset selection formulation. The goal here is to select a non-empty subset of the given populations so that the selected subset includes the best population with at least a preassigned probability P^*. It is usually desired that this be accomplished by selecting a subset as small as possible and without any knowledge of the true values of the parameters.

Suppose that $\pi_1,\ldots,\pi_k$ are k independent populations and $\pi_i (i = 1,\ldots,k)$ is characterized by the distribution function F_{λ_i}, where λ_i is a real valued (unknown) parameter, which is assumed to be a measure of the quality of π_i. Let $\lambda_{[1]} \leq \lambda_{[2]} \leq \ldots \leq \lambda_{[k]}$ be the ordered values of the λ_i. The correct pairing of the ordered and the unordered λ's is not known. The population associated with $\lambda_{[i]}$ is denoted by $\pi_{[i]}$ and the population $\pi_{[k]}$ (or $\pi_{[1]}$) is

usually defined as the best population. In the case of a tie, we assume that one of the populations with $\lambda_i = \lambda_{[k]}$ (or $\lambda_i = \lambda_{[1]}$) is tagged as the best. The selection of any subset which includes the best population is called a correct selection (CS) and $P\{CS|R\}$ denotes the probability of a correct selection using the rule R. Thus we are interested in defining a rule R such that

$$P\{CS|R\} \geq P^* , \quad k^{-1} < P^* < 1 , \tag{1.1}$$

regardless of the true parameter point $\lambda = (\lambda_1,\ldots,\lambda_k)$ in the parameter space $\Omega = \{\lambda\}$. If the distributions are not indexed by the values of any parameter λ, Ω denotes the space of the k-tuples $\{F_1,\ldots,F_k\}$, where F_i is the distribution function of π_i. In order that (1.1) be met, we want

$$\inf_{\Omega} P\{CS|R\} \geq P^* . \tag{1.2}$$

The requirement (1.2) is usually referred to as the basic probability requirement or the P^*-condition.

2. Selection in terms of Location and Scale Parameters.

Many of the early investigations relate to ranking and selection of populations in terms of either location or scale parameters. The ranking of normal means and gamma shape parameters are examples of this type.

Let us first suppose that $\pi_i (i = 1,\ldots,k)$ has the continuous distribution $F_{\lambda_i}(x) = F(x-\lambda_i)$, $-\infty < \lambda_i < \infty$ and x_i is an observation from π_i. In order to select a subset containing the population associated with $\lambda_{[k]}$, we define the following rule R_1.

(2.1) R_1: Select π_i iff $x_i \geq x_{max} - d$

where $x_{max} = \max(x_1,\ldots,x_k)$ and d is a positive constant chosen so as to satisfy the basic probability requirement. It is easy to see that

$$(2.2) \qquad P\{CS|R_1\} = \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} F(y+d+\lambda_{[k]}-\lambda_{[j]})\, dF(y) .$$

Clearly, the infimum of $P\{CS|R_1\}$ is attained when $\lambda_1 = \ldots = \lambda_k$ and hence d is given by

$$(2.3) \qquad \int_{-\infty}^{\infty} F^{k-1}(y+d)\, dF(y) = P^* .$$

Denoting by S the number of populations included in the selected subset, we can see that

$$(2.4) \qquad E(S) = p_1 + \ldots + p_k ,$$

where p_i is the probability that the population associated with $\lambda_{[i]}$ is included in the subset. In the present case

$$(2.5) \qquad p_i = \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j\neq i}}^{k} F(y+d+\lambda_{[i]}-\lambda_{[j]})\, dF(y) .$$

It has been shown by Gupta (1965) that $\sup_{\Omega} E(S)$ is attained when $\lambda_1 = \ldots = \lambda_k$ provided that the density $f_\lambda(x) = f(x-\lambda)$ has a monotone likelihood ratio in x and in that case the supremum is kP^*. The procedure R_1 has also been shown to be montone in the sense that $p_i \geq p_j$ for $\lambda_{[i]} \geq \lambda_{[j]}$.

As an application of the above results, we consider selecting a subset containing the population with the largest mean from k independent normal populations with unknown means $\mu_1,\ldots,\mu_k$ and a common known variance σ^2.

If $\bar{y}_i (i = 1,\ldots,n)$ is the sample mean based on n observations from π_i, the rule R_1 in this case selects π_i iff $\bar{y}_i \geq \max_{1\leq j\leq k} \bar{y}_j - d_1$ where d_1 will depend on n and k. By letting $d_1 = d\sigma/\sqrt{n}$, the constant d is given by

$$\int_{-\infty}^{\infty} \Phi^{k-1}(u+d)\phi(u)du = P^* , \tag{2.6}$$

where, unless otherwise stated, Φ and ϕ denotes here and in the sequel the cdf and the density of the standard normal distribution. If σ^2 is unknown, one will naturally use s^2, the poolod estimate of σ^2 based on k(n-1) degrees of freedom. In this case we can show that d is given by

$$\int_0^{\infty}\int_{-\infty}^{\infty} \Phi^{k-1}(u+yd)\phi(u)g_\nu(y)dudy = P^*, \tag{2.7}$$

where $g_\nu(y)$ is the density of χ_ν/ν with $\nu = k(n-1)$.

Rizvi (1963) considered the goal of selecting a non-empty subset from k normal populations so as to include the one with the largest $\theta_i = |\mu_i|$. He uses a rule of the type R_1 based on $w_i = |x_i|$. For his procedure

$$\sup_{\Omega} E(S) = 2k \int_0^{\infty} [2\,\Phi(u+d)-1]^{k-1}\, d\Phi(u), \text{ where} \tag{2.8}$$

d is given by (2.6). This bound for E(S), however, exceeds kP^*.

Suppose the populations π_i, $i = 1,\ldots,k$, have the continuous distributions $F_{\lambda_i}(x) = F(x/\lambda_i)$, $\lambda_i > 0$, $x_i > 0$. To select a subset containing the population associated with $\lambda_{[k]}$, we define the procedure R_2 as follows:

$$R_2\text{: Select } \pi_i \text{ iff } x_i \geq c^{-1} x_{max} \tag{2.9}$$

where x_i is an observation from π_i and $c > 1$ is determined so that the basic

probability requirement is satisfied. It is easily seen that $\inf_{\Omega} P\{CS|R_2\}$ is attained when $\lambda_1 = \ldots = \lambda_k$ and the constant c is given by

$$\int_0^\infty F^{k-1}(cy)\, dF(y) = P^* . \tag{2.10}$$

The rule R_2 is monotone and if the density $f_\lambda(x) = \frac{1}{\lambda} f(x/\lambda)$ has a monotone likelihood ratio in x, then $\sup_{\Omega} E(S)$ is attained when $\lambda_1 = \ldots = \lambda_k$ and is equal to kP^*.

A specific example of interest is the selection from k gamma populations with densities

$$f_{\lambda_i}(x) = e^{-x/\lambda_i} x^{r-1} / \Gamma(r)\, \lambda_i^r,\ x > 0,\ \lambda_i > 0,\ i = 1,\ldots,k. \tag{2.11}$$

In order to select a subset containing the population with $\lambda_{[k]}$, we use the rule R_2 based on n observations from each population, namely,

$$R_2:\ \text{Select } \pi_i \text{ iff } \bar{x}_i \geq b^{-1} \max_{1 \leq j \leq k} \bar{x}_j , \tag{2.12}$$

where $b > 1$ is determined so as to satisfy the basic probability requirement. This procedure has been studied by Gupta (1963). The analogous problem of selecting the gamma population with the smallest λ_i has been discussed by Gupta and Sobel (1962a, 1962b). This problem arises in the context of selecting a subset containing the normal population with the smallest variance and the rule is an obvious modification of R_2 and is based on the estimates s_i^2 $(i=1,\ldots,k)$ of the population variances σ_i^2 using equal sample size.

For the problem of ranking and selection from normal population in terms of their means, Seal (1 55) considered a class of procedures satisfying the basic probability requirement. Assuming that the populations have a common unknown variance, let $\bar{x}_1,\ldots,\bar{x}_k$ be the sample means from the populations, each based on n independent observations. Let $\underline{c} = (c_1,\ldots,c_{k-1})$ be a vector whose components are arbitrary non-negative numbers such that $c_1 +\ldots+ c_{k-1} = 1$. Let $\bar{x}_{[1]} \leq \ldots \leq \bar{x}_{[k]}$ be the ordered sample means. The class C of rules $D_{\underline{c}}$ defined by Seal is as follows:

$D_{\underline{c}}$: Include in the selected subset the population corresponding to $\bar{x}_{[i]}$ iff

$$(2.13)\quad \bar{x}_{[i]} \geq c_1\bar{x}_{[1]}+\ldots+c_{i-1}\bar{x}_{[i-1]}+c_i\bar{x}_{[i+1]}+\ldots+c_{k-1}\bar{x}_{[k]}-t(P^*,\underline{c})s/\sqrt{n},$$

where s^2 is the usual pooled estimate of the common variance σ^2, and $t(P^*,\underline{c})$ satisfying the P^*-condition is given by the upper $100(1-P^*)$ percent point of the distribution of $Y = (\sum_{i=1}^{k-1} c_i Z_{(i)}-Z_k)/s$ where $z_i = 1,\ldots,k$ are random observations from $N(0,\sigma^2)$ and $z_{(1)} < z_{(2)} <\ldots< z_{(k-1)}$ are the ordered $z_1,\ldots,z_{k-1}$.

The rules of this class possess certain desirable properties. For example, the rule $D_{\underline{c}}$ is unbiased, that is, P{rejecting any population not having the largest mean} $\geq$ P{rejecting the population with the largest mean}. Also the rule has the property of gradation, namely, corresponding to any P^*, there exists a constant μ_0 (depending on the decision rule, the unknown means and the common variance σ^2) such that P{retaining the population with mean μ_i} $\gtrless P^*$ according as $\mu_i \gtrless \mu_0$.

If we now assume that σ is known, we can take $\sigma=1$ with no loss of generality and the rule $D_{\underline{c}}$ will be (2.13) with s=1. We define a subclass $\mathbf{C}'$ of $\mathbf{C}$ by the restriction $c_j = 1$ for some $j = 1,\ldots,k-1$. The procedure R(called R_1 earlier in this section) studied by Gupta (1965) is a member of $\mathbf{C}'$ with $c_{k-1}=1$. It has been shown by Deely and Gupta (1968) that the rule R has the smallest expected subset size among the rules of the class $\mathbf{C}'$ provided that the parametric configuration is $\mu_{[1]}\leq\ldots\leq\mu_{[k-1]}\leq\mu=\mu_{[k]}-\delta(\delta>0)$ and δ is sufficiently large. If we consider a slippage configuration $(\mu,\ldots,\mu,\mu+\delta)$, $\delta>0$, Seal (1955) shows that in the class $\mathbf{C}$, the rule $\bar{D}$ with $c_1=\ldots=c_{k-1}=1/(k-1)$ maximizes (approximately) the probability of including the population with mean $\mu+\delta$. Deely and Gupta show that $E\{S|R\} < E\{S|\bar{D}\}$ except when δ is near zero.

Seal (1958) defined a class of rules similar to $\mathbf{C}$ for the problem of selection from gamma populations given by (2.11). Let $\underline{c} = (c_1,\ldots,c_{k-1})$ be as before a vector of non-negative components such that $\sum_{i=1}^{k-1} c_i = 1$. Let $x_1,\ldots,x_k$ be a set of observations from the k populations and $x_{[1]}\leq x_{[2]}\leq\ldots\leq x_{[k]}$ be the ordered observations. Then, in order to select a subset containing the population with the smallest λ_i, Seal proposed the class of rules $D'_{\underline{c}}$ defined below.

$D'_{\underline{c}}$: Include in the selected subset the population corresponding to $x_{[i]}$ iff

$$(2.14)\qquad x_{[i]} \leq b(c_1\bar{x}_{[1]}+\ldots+c_{i-1}\bar{x}_{[i-1]} + c_i\bar{x}_{[i]}+\ldots+c_{k-1}\bar{x}_{[k]}),$$

where b satisfying the basic probability requirement is given by upper $100(1-P^*)$ percent point of the distribution of $Y_k/\sum_{i=1}^{k-1} c_i Y_{(i)}$, where $Y_1,\ldots,Y_k$ are k random observations from a gamma population with $\lambda=1$ and $Y_{(1)}\leq\ldots\leq Y_{(k-1)}$ are the ordered $Y_1,\ldots,Y_{k-1}$. Seal (1958) has obtained results similar to his earlier ones for the class of rules $D_{\underline{c}}$.

3. General Theory of Subset Selection.

In this section we will describe a class of subset selection rules applicable to populations from a family of stochastically ordered distributions and therefore in particular to populations characterized by a location or scale parameter. Many of the specific selection problems discussed in the subsequent sections fall under this general frame work. We also discuss a decision-theoretic formulation of the problem.

We assume that $\pi_1, \pi_2, \ldots, \pi_k$ have the associated absolutely continuous distributions F_{λ_i} $(i = 1, \ldots, k)$, where $\lambda_i \in \Lambda$, an interval on the real line. The family $\{F_\lambda\}$, $\lambda \in \Lambda$, is assumed to be stochastically increasing (SI) in λ, i.e., for $\lambda < \lambda'$ in Λ, F_λ and $F_{\lambda'}$ are distinct and $F_\lambda(x) \geq F_{\lambda'}(x)$ for all x. For selecting a subset containing population associated with $\lambda_{[k]}$, Gupta and Panchapakesan (1970) have discussed a class of procedures R_h defined by a class of real valued functions $h \equiv h_{c,d}$, $c \geq 1$, $d \geq 0$, possessing the following properties: For every x belonging to the support of F_λ, (i) $h_{c,d}(x) \geq x$, (ii) $h_{1,0}(x) = x$, (iii) $h_{c,d}(x)$ is continuous in c and d, and (iv) $\lim_{d \to \infty} h_{c,d}(x) = \infty$ (c fixed) and/or $\lim_{c \to \infty} h_{c,d}(x) = \infty$ (d fixed), $x \neq 0$. If $x_1, \ldots, x_k$ is a set of observations from $\pi_1, \ldots, \pi_k$, respectively, the rule R_h is defined as follows.

R_h : Include the population π_i iff

$$h(x_i) \geq \max_{1 \leq r \leq k} x_r \,. \tag{3.1}$$

Letting $x_{(r)}$ denote the observation from the population with distribution $F_{[r]} \equiv F_{\lambda_{[r]}}$, we obtain

$$P\{CS|R_h\} = \int \{ \prod_{r=1}^{k-1} F_{[r]}(h(x)) \} \, dF_{[k]}(x) \,. \tag{3.2}$$

Because of the stochastic ordering of $\{F_\lambda\}$, we can see that

$$\inf_{\Omega} P\{CS|R_h\} = \inf_{\lambda \in \Lambda} \psi(\lambda;\, c,d,k) \,, \tag{3.3}$$

where $\psi(\lambda;\, c,d,t+1)$ is given by

$$\psi(\lambda;\, c,d,t+1) = \int F_\lambda^t(h(x))\, dF_\lambda(x) \,. \tag{3.4}$$

In all the specific cases considered earlier in the literature, the general approach is to show that $\psi(\lambda;\, c,d,k)$ is monotonic in λ and use this fact to evaluate $\inf \psi(\lambda;\, c,d,k)$ and find the values of the constants such that the P*-condition is met. One of the main results of Gupta and Panchapakesan (1970) is the following theorem which leads to a sufficient condition for the monotonicity of $\psi(\lambda;\, c,d,k)$.

Theorem 3.1. Let $\{F_\lambda\}$, $\lambda \in \Lambda$, be a family of absolutely continuous distributions on the real line and $\psi(x,\lambda)$ be a real valued function possessing continuous first partial derivatives ψ_x and ψ_λ w.r.t. x and λ, respectively. Then, $E_\lambda \psi(x,\lambda)$ is non-decreasing in λ provided that

$$f_\lambda(x)\, \psi_x(x,\lambda) - \psi_x(x,\lambda)\, \frac{\partial}{\partial \lambda} F_\lambda(x) \geq 0 \quad \text{for all } x \,, \tag{3.5}$$

where $f_\lambda(x)$ is the density corresponding to $F_\lambda(x)$. Further, $E_\lambda \psi(x,\lambda)$ is strictly increasing in λ if (3.5) holds with strict inequality on a set of positive Lebesgue measure.

The above theorem is a generalization of a result of Lehmann (1959, p. 112) which states essentially that, if $\{F_\lambda\}$ is an SI family and $\psi(x)$ is an increasing function of x, then $E_\lambda \psi(x)$ is non-decreasing in λ. As we can see, this comes out as a special case of Theorem 3.1, by letting $\psi(x,\lambda) = \psi(x)$

for all λ and verifying the condition (3.5) to be true.

As a consequence of Theorem 1.1, the following theorem is obtained regarding the monotonic behavior of $\psi(\lambda\,;\,c,d,k)$.

Theorem 3.2. For the procedure R_h defined by (3.1), $\psi(\lambda;\,c,d,k)$ is non-decreasing in λ provided that

$$f_\lambda(x)\,\frac{\partial}{\partial\lambda}F_\lambda(h(x)) - h'(x)\,f_\lambda(h(x))\,\frac{\partial}{\partial\lambda}F_\lambda(x) \geq 0 \tag{3.6}$$

for all $\lambda \in \Lambda$ and all x,

where $h'(x) = \frac{d}{dx}h(x)$. Further, $\psi(\lambda;\,c,d,k)$ is strictly increasing in λ if strict inequality holds in (3.6) on a set of positive Lebesgue measure.

Let p_i denote the probability that the population $\pi_{[i]}$ is included in the subset. Then the expected subset size is given by

$$E(S) \equiv E\{S|R_h\} = p_1 + \ldots + p_k\,, \tag{3.7}$$

where

$$p_i = \int \{\prod_{\substack{r=1\\ r\neq i}}^{k} F_{[r]}(h(x))\}\, dF_{[i]}(x), \qquad i = 1,\ldots,k\,. \tag{3.8}$$

Gupta and Panchapakesan (1969b) have shown that, if $\lambda_{[1]} = \lambda_{[2]} = \ldots = \lambda_{[m]} = \lambda \leq \lambda_{[m+1]} \leq \ldots \leq \lambda_{[k]}$, $1 \leq m \leq k$, $E(S)$ is non-decreasing in λ, when $\lambda_{[m+1]},\ldots,\lambda_{[k]}$ are kept fixed, provided that,

$$\frac{\partial}{\partial\lambda_1}F_{\lambda_1}(h(x))\,f_{\lambda_2}(x) - h'(x)\,\frac{\partial}{\partial\lambda_1}F_{\lambda_1}(x)\,f_{\lambda_2}(h(x)) \geq 0 \tag{3.9}$$

for $\lambda_1 \leq \lambda_2$ and all x.

Consequently, the following theorem is obtained.

Theorem 3.3. For the procedure R_h defined by (3.1), the $\sup_{\Omega} E\{S|R_h\}$ is attained when $\lambda_1 = \lambda_2 = \ldots = \lambda_k$ provided that (3.9) holds.

If the condition (3.9) holds, then (3.6) is valid and consequently $\psi(\lambda; c,d,k)$ is non-decreasing in λ. Thus $\sup_{\Omega} E(S) = k \sup_{\lambda} \psi(\lambda; c,d,k)$ can be evaluated. Hence, by verifying the condition (3.9) we are simultaneously assured of the monotonicity of $\psi(\lambda; c,d,k)$, the fact which is used for the evaluation of $\inf_{\Omega} P\{CS|R_h\}$ and $\sup_{\Omega} E\{S|R_h\}$. This connection between the two has been observed by Gupta and Panchapakesan (1970).

It should be pointed out however that condition (3.6) may hold without (3.9) being true. This is the case, for example, when we consider the selection from Cauchy distributions in terms of the location parameter using $h(x) = x+d$, $d > 0$. If (3.6) is satisfied, we have $\inf_{\lambda} \psi(\lambda; c,d,k) = \psi(\lambda_0; c,d,k)$. Then we can evaluate the constants because of the conditions imposed on $h(x)$ provided we assume that $F_{\lambda_0}(x)$ is a distribution function in case $\lambda_0 \notin \Lambda$.

It can be seen that the above results are readily applicable to the cases of location and scale parameters discussed in Section 2. In the case of location parameters the rule R_1 defined earlier uses $h(x) = x-d$, $d \geq 0$, and in the scale parameter case the rule R_2 uses $h(x) = cx$, $c \geq 1$. In both the cases it is easy to see that (3.6) is satisfied and (3.9) reduces to the condition that the density $f_\lambda(x)$ has a monotone likelihood ratio in x.

Another case of importance is that of convex mixtures of distributions. Here the density $f_\lambda(x)$ is of the form $f_\lambda(x) = \sum_{j=0}^{\infty} w(\lambda,j)\, g_j(x)$, where $g_j(x)$, $j = 0,1,\ldots,$ is a sequence of density functions and $w(\lambda,j)$ are non-negative weights such that $\sum_{j=0}^{\infty} w(\lambda,j) = 1$. We assume that the weights are given by

(3.10) $$w(\lambda,j) = a_j \lambda^j / A(\lambda) j!,\ A(\lambda) \geq 0,\ \lambda \geq 0$$

and

(3.11) $$a_{j+1} = (m+\ell j) a_j,\ j = 0,1,\ldots;\ \ell,\ m \geq 0\ .$$

It is easy to see that $A(\lambda) = a_0(1-\lambda\ell)^{-m/\ell}$, provided that $\lambda < 1/\ell$. It has been shown by Gupta and Panchapakesan (1970) that the condition (3.9) is satisfied if, for $\alpha = 0, 1,\ldots,[i/2]$ ([s] denotes the largest integer $\leq s$) and $b \geq 1$,

(3.12) $$b^{i-\alpha}(m+\ell\alpha)\ [g_{i-\alpha}(x)\ \Delta G_\alpha(h(x)) - h'(x) g_{i-\alpha}(h(x))\ \Delta G_\alpha(x)]$$
$$+ b^{\alpha}(m+\ell(i-\alpha))\ [g_\alpha(x)\ \Delta G_{i-\alpha}(h(x)) - h'(x) g_\alpha(h(x))\ \Delta G_{i-\alpha}(x)]$$
$$\geq 0$$

where $\Delta G_\alpha(x) = G_{\alpha+1}(x) - G_\alpha(x)$.

This special case is of interest. If we set $m = 1$, $\ell = 0$, and $a_0 = 1$, we get Poisson weights $w(\lambda,j) = e^{-\lambda}\lambda^j/j!$. Selection problems involving non-central chi-square and non-central F distributions in terms of non-centrality parameter fall under this special case and have been considered earlier by Gupta (1966b), Gupta and Studden (1970), and Gupta and Panchapakesan (1969a). These specific procedures are discussed in Section 5. Again, if we set $\ell = 1$ and $a_0 = 1$, we get densities $g_j(x)$ with negative binomial weights. The distribution of R^2, where R is the multiple correlation coefficient, in the so-called unconditional case is an example of this special case of weights. Selection procedures involving this have been discussed by Gupta and Panchapakesan (1969a) and are described in Section 5. The condition (3.12) with $b = 1$ gives the sufficient condition for the monotonicity of $\psi(\lambda;\ c,d,k)$ obtained by Gupta

and Studden (1970) and Gupta and Panchapakesan (1969a) for proper choices of weight functions.

Let S' be the number of non-best populations included in the selected subset. Then, for the procedure R_h defined by (3.1), $E(S') \equiv E\{S'|R_h\}$ is given by $E(S') = p_1 + \ldots + p_{k-1}$. Panchapakesan (1969) has shown that $\sup_{\Omega} E(S')$ is attained when the distributions are identical provided that (3.9) holds.

It has also been shown that the procedure R_h is monotone, i.e., if $\lambda_i < \lambda_j$ then the probability of π_j being selected is at least as great as the probability of π_i being selected.

In the case of absolutely continuous distributions F_λ, where λ belongs to a discrete set of real numbers, Panchapakesan (1970) has obtained the following theorem corresponding to Theorem 3.1 and has applied it to the case of gamma distributions with integer-valued shape parameters and common scale parameter.

Theorem 3.4. Let $\{F_\lambda\}$ be an absolutely continuous distributions where $\lambda \in \Lambda_d = \{\lambda_1 < \lambda_2 < \ldots\}$ and $\psi(x,\lambda)$ be a real valued function possessing continuous partial derivative ψ_x w.r.t. x. Then, for any positive integer t, $E_\lambda \psi^t(x,\lambda)$ is non-decreasing in λ provided that, for $i = 1,2,\ldots$,

$$(3.13) \qquad \Delta\psi(x,\lambda_i)\, f_{\lambda_j}(x) - \Delta F_{\lambda_i}(x)\, \psi_x(x,\lambda_j) \geq 0, \quad j = i,\ i+1,$$

where $$\Delta\psi(x,\lambda_i) = \psi(x,\lambda_{i+1}) - \psi(x,\lambda_i),\ \Delta F_{\lambda_i}(x) = F_{\lambda_{i+1}}(x) - F_{\lambda_i}(x)\ .$$

Now we present a decision theoretic formulation of the subset selection problem. We are given k populations $\pi_1,\ldots,\pi_k$ where π_i is described by the probability space $(\mathcal{X},\mathcal{B},P_i)$, where P_i belongs to some family $\mathcal{P}$.

We assume that there is a partial order relation $(\succ)$ defined in $\mathcal{P}$. $P_i \succ P_j$ is equivalent to saying that P_i is better than or equal to P_j; or, in other words P_i is preferred over P_j. For example, if $\mathcal{P}$ is a one-parameter family, $P_i(x) = P(\theta_i,x)$, we may define: $P_i \succ P_j$ iff $\theta_i \geq \theta_j$. In many problems $\succ$ denotes stochastic ordering. Other partial orderings that have been considered are: star-shaped ordering, convex ordering, tail ordering.

In the above set-up, we assume that there exists a population π_j such that $\pi_j \succ \pi_i$ for all i. This population π_j will be referred to as the 'best' population. In case of more than one population satisfying the condition we will consider one of them to be tagged as the best.

From each population we observe a random element X_i. The space of observations is: $\mathcal{X}^k = \{\underline{x}=(x_1,x_2,\ldots,x_k),\ x_i \in \mathcal{X},\ i = 1,2,\ldots,k\}$. In most applications $\mathcal{X}^k$ will be a real vector space.

The decision space $\mathcal{D}$ consists of the 2^k subsets d of the set $\{1,2,\ldots,k\}$: to put it formally,

$$\mathcal{D} = \{d \mid d \subseteq \{1,2,\ldots,k\}\} . \tag{3.14}$$

In other words, a decision d corresponds to the selection of a subset of k populations.

A decision $d \in \mathcal{D}$ is called a correct selection (CS) if $j \in d$ which means that the best population π_j is included in the selected subset d. It should be pointed out that in many subset selection procedures investigated earlier, the null set ϕ is excluded from $\mathcal{D}$ to guarantee the selection of a non empty subset.

A measurable function δ defined on $\mathcal{X}^k \times \mathcal{D}$ is called a selection procedure provided that for each $\underline{x} \in \mathcal{X}^k$, we have,

$$(3.15) \qquad \begin{cases} \delta(\underline{x},d) \geq 0 & \text{and} \\ \sum_{d \in \mathcal{D}} \delta(\underline{x},d) = 1 \ , \end{cases}$$

where $\delta(\underline{x},d)$ denotes the probability that the subset d is selected when $\underline{x}$ is observed. The individual selection probability $p_i(\underline{x})$ for the population π_i is then given by

$$(3.16) \qquad p_i(\underline{x}) = \sum_{d \ni i} \delta(\underline{x},d) \ ,$$

where the summation is over all d containing i. If the selection probabilities $p_1(\underline{x}), p_2(\underline{x}),\ldots,p_k(\underline{x})$ take on only the values 0 and 1, then the selection procedure $\delta(\underline{x},d)$ is completely specified.

In general, we can assume that the selection of a subset $d \in \mathcal{D}$ results in a loss. Let us consider the situation where $p_i = p(\theta_i,\underline{x})$ and assume the loss $L(\underline{\theta},d) = L((\theta_1,\theta_2,\ldots,\theta_k),d) = \sum_{i \in d} L_i(\underline{\theta})$ where $L_i(\underline{\theta})$ is the loss if the ith population is selected. We may assume an additional loss L if a correct selection is not made. The overall risk for the nonrandomized rule δ is:

$$(3.17) \qquad R(\underline{\theta},\delta) = \sum_{i=1}^{k} L_i(\underline{\theta}) \, E_{\underline{\theta}} p_i(\underline{x}) + L[1 - P_{\underline{\theta}}\{CS|\delta\}].$$

In many problems it has been assumed that $L_i(\underline{\theta}) = 1$ and $L = 0$, in which case, $R(\underline{\theta},\delta)$ gives the expected size of the selected subset. In general, our aim is to minimize the risk $R(\underline{\theta},\delta)$ which will be done under the usual symmetry condition.

Our goal is to obtain selection rules δ selecting a non-empty subset and satisfying the P*-condition. In general, we wish rules with large probability of a correct selection and a small value of the expected size. The ratio $\eta_\omega(\delta) = k\, P_\omega(CS|\delta)/E_\omega(S|\delta)$ can, among others, be considered as a measure of the efficiency of the procedure δ at $\omega = (P_1,\ldots,P_k)$, $P_i \in \mathcal{P}$. Both $P_\omega(CS|\delta)$ and $E_\omega(S|\delta)$ depend on δ only through the individual selection probabilities and hence if we restrict our attention to these quantities, we can define two rules δ and δ' as equivalent if they have the same individual selection probabilities $p(\underline{x})$ and $p'(\underline{x})$ for all $\underline{x}$. Hence, we can use the following simplified definition, replacing δ by R.

A subset selection rule R is a measurable mapping from $\mathcal{X}^k$ into E^k (k dimensional Euclidean space), namely,

$$R:\ \underline{x} \to (p_1(\underline{x}),p_2(\underline{x}),\ldots,p_k(\underline{x})),\ 0 \le p_i(\underline{x}) \le 1\ ,$$
$$i = 1,2,\ldots,k\ .$$

If p_i's are 0 or 1, the rule is nonrandomized; in this case, R can also be defined by the sets $A_i = \{\underline{x} \in \mathcal{X}^k | p_i(\underline{x}) = 1\}$, $i = 1,2,\ldots,k$. A_i is the set of observations for which π_i is selected. R is said to be unbiased iff

$$\pi_j \succ \pi_i,\ i = 1,2,\ldots,k \Rightarrow P_{\omega,j} \ge P_{\omega,i} \text{ for all } \omega \in \Omega$$

where $P_{\omega,i} = E_\omega p_i(\underline{x})$ = probability that π_i is selected, and is said to be monotone iff

$$\pi_j \succ \pi_i \Rightarrow P_{\omega,j} \ge P_{\omega,i} \text{ for all } i,j \text{ and all } \omega \in \Omega\ .$$

We shall restrict ourselves to selection rules R which are invariant under permutation (or symmetric), i.e., rules R for which

$$(p_1(g\underline{x}),\ldots,p_k(g\underline{x})) = g(p_1(\underline{x}),\ldots,p_k(\underline{x})) \text{ for all } \underline{x}\in\mathcal{X}^k,\ g\in G$$

where G denotes the group of permutations g of the integers $1,2,\ldots k$.

Studden (1967) has discussed the problem of obtaining optimal procedures. He has obtained a necessary and sufficient condition that a rule δ be best invariant, that is, δ is an invariant rule for which $R(\underline{\theta},\delta)$ is minimum. Assume Ω_k to be those permutations of $(\theta_1,\ldots,\theta_k)$ such that the largest parameter value $\theta_{[k]}$ is in the last component and let $\phi_i(\underline{x};\underline{\theta}) = (1/(k-1)!) \sum_{G_i} f(\underline{x},g\underline{\theta})$, $i = 1,\ldots,k$ where $f(\underline{x},\underline{\theta})$ is the joint density of $\underline{x}$(w.r.t. some measure μ) and $G_i = \{g|g^{-1}k = i\}$. The following theorem has been proved by Studden.

Theorem 3.5. A selection rule δ is best invariant iff

$$(3.18)\qquad p_k(\underline{x}) = 1 \text{ if } L\phi_k(\underline{x};\underline{\theta}) > \sum_{i=1}^{k} L_i(\underline{\theta})\ \phi_i(\underline{x};\underline{\theta})$$

$$= 0 \text{ if } L\phi_k(\underline{x};\underline{\theta}) < \sum_{i=1}^{k} L_i(\underline{\theta})\ \phi_i(\underline{x};\underline{\theta})$$

for $\underline{\theta}\in\Omega_k$, almost everywhere μ. The functions $p_i(x)$, $i\neq k$, are defined by the invariant conditions on $p(\underline{x}) = (p_1(\underline{x}),\ldots,p_k(\underline{x}))$. As a corollary, we obtain the result: An invariant selection procedure minimizes $\sum_{i=1}^{k} L_i(\underline{\theta})\ E_{\underline{\theta}}p_i(\underline{x})$ subject to the condition

$$(3.19)\qquad P_{\underline{\theta}}\{CS|\delta\} \geq \gamma(L) \quad \text{for all } \theta\in\Omega$$

iff the individual selection probabilities are determined by (3.18).

The expression given in (3.18) defining the selection probabilities which minimize $R(\underline{\theta},\delta)$ is rather complicated when written down in terms of the

original densities. However, for the slippage situation when the underlying densities are from an exponential family and $L_i(\underline{\theta}) \equiv 1$, the expressions simplify considerably and in this case the following theorem has been obtained by Studden.

Theorem 3.6. Let $f_{\underline{\theta}}(\underline{x}) = \prod_{i=1}^{k} f_{\theta_i}(x_i)$ where $f_\theta(x) = C(\theta)e^{\theta x}$ and $\theta = \theta_{[1]} = \theta_{[2]} = \ldots = \theta_{[k-1]} = \theta_{[k]} - \Delta$ $(\Delta > 0)$. An invariant rule δ minimizes $E_{\underline{\theta}}\{S|\delta\}$ subject to the condition that $P_{\underline{\theta}}\{CS|\delta\} \geq \gamma$ iff for almost all $\underline{x}$

$$P_k(\underline{x}) = 1 \text{ if } \sum_{i=1}^{k-1} e^{\Delta x_i} < Ce^{\Delta x_k}$$

$$= 0 \text{ if } \sum_{i=1}^{k-1} e^{\Delta x_i} > Ce^{\Delta x_k} . \tag{3.20}$$

Studden also considered a simple situation concerning normal populations where the parameters are permitted to vary. It is assumed that $f(\underline{x};\underline{\theta}) = \prod_{i=1}^{k} f(x_i-\theta_i)$ where $f(x)$ is the standard normal density. For fixed Δ let $p(\underline{x};\Delta)$ denote the selection probabilities defined by (3.20) where C is chosen so that $P_{\underline{\theta}}\{CS|p(\underline{x},\Delta)\} = \gamma$ for all $\underline{\theta} = (\theta,\ldots,\theta,\theta+\Delta)$. Let $\Phi(\Delta)$ denote the class of invariant procedures satisfying

$$P_{\underline{\theta}}\{CS|\delta\} \geq \gamma \quad \text{for all } \underline{\theta} \in \Omega(\Delta) \tag{3.21}$$

where $\Omega(\Delta) = \{\underline{\theta} \mid \theta_{[1]} \leq \theta_{[2]} \leq \ldots \leq \theta_{[k-1]} \leq \theta_{[k]} - \Delta\}$.

Theorem 3.7. For any $\underline{\theta}$ with $\theta_{[1]} = \theta_{[2]} = \ldots = \theta_{[k-1]} = \theta_{[k]} - \Delta$ the minimum value of $E_{\underline{\theta}}\{S|\delta\}$ over the class $\Phi(\Delta)$ is attained by $p(\underline{x};\Delta)$, i.e.,

$$\min_{\Phi(\Delta)} E_{\underline{\theta}}\{S|\delta\} = E_{\underline{\theta}}\{S|p(\underline{x};\Delta)\} . \tag{3.22}$$

Now, consider the sequence of selection probabilities defined for $\Delta \in (0,\infty)$ by

$$(3.23) \qquad p_k(\underline{x};\Delta) = 1 \text{ if } \sum_{i=1}^{k-1} e^{\Delta x_i} < C(\Delta)e^{\Delta x_k}$$

$$= 0 \text{ if } \sum_{i=1}^{k-1} e^{\Delta x_i} > C(\Delta)e^{\Delta x_k} .$$

For $\Delta = 0$ we let

$$(3.24) \qquad p_k(\underline{x};0) = 1 \text{ if } \sum_{j=1}^{k-1} x_j/(k-1) < x_k + C(0)$$

$$= 0 \text{ if } \sum_{j=1}^{k-1} x_j/(k-1) > x_k + C( ) ,$$

while for $\Delta = \infty$ we define

$$(3.25) \qquad p_k(\underline{x};\infty) = 1 \text{ if } \max_{1\le j\le k-1} x_j < x_k + C(\infty)$$

$$= 0 \text{ if } \max_{1\le j\le k-1} x_j > x_k + C(\infty) .$$

The values $C(\Delta)$, $\Delta \in [0,\infty]$ are all chosen so that for a fixed set of values $\theta_{[1]} \le \ldots \le \theta_{[k]}$, the probability of a correct selection is equal to a given value γ. The rules defined in (3.24) and (3.25) have been considered by several authors. It has been observed by Studden that $p_k(\underline{x};\Delta)$ has limits $p_k(\underline{x};0)$ and $p_k(\underline{x};\infty)$ almost everywhere μ as Δ approaches zero and infinity, respectively.

In addition to several desirable properties and criteria for selection rules discussed above, another concept was investigated by Nagel (1970). This is concerned with what are called "just" selection rules.

We assume that a partial order relation $\succ$ is defined on $\mathcal{X}$ [$\underline{y} \succeq \underline{x}$ or, equivalently, $\underline{x} \prec \underline{y}$ means that $\underline{y}$ is better than $\underline{x}$]. A selection rule R defined by its individual selection probabilities $p_i(\underline{x})$, $i = 1,\ldots,k$, is said to be just iff

$$\left.\begin{array}{r} x_i \prec y_i \\ x_j \succ y_j,\ j \neq i \end{array}\right\} \Rightarrow p_i(\underline{y}) \geq p_i(\underline{x}) . \tag{3.26}$$

For nonrandomized rules determined by acceptance regions $A_1,\ldots,A_k$, we can define a just rule equivalently in terms of increasing sets. A subset $A \subset \mathcal{X}^k$ is said to be increasing iff $\underline{x} \in A$ and $\underline{y} \succ \underline{x} \Rightarrow \underline{y} \in A$. We say that P is stochastically better than Q ($P \succ_{st} Q$) iff $P(A) \geq Q(A)$ for all increasing sets $A \in \mathcal{A}$. We note that if $\mathcal{X}$ is the real line and $\succ$ stands for $>$ (or $\geq$) then the increasing sets are the intervals $[a,\infty)$ and (a,∞) which induce the usual stochastic ordering on the distribution functions. A rule R is said to be just iff

$$\left.\begin{array}{l} \underline{x} \in A_i \\ x_i \prec y_i \\ x_j \succ y_j,\ j \neq i \end{array}\right\} \text{ implies } \underline{y} \in A_i .$$

As mentioned earlier, frequently we require a selection rule to satisfy the basic probability requirement. Hence, a central problem in the subset selection theory is to determine $\inf_{\omega \in \Omega} P_\omega\{CS|R\}$. For many rules investigated in the literature, this infimum is attained in Ω_0 where $\Omega_0 \subseteq \Omega$ is the set of ω where the P_i are identical. This could reasonably be expected of a good rule, because in Ω_0, no statistical information can be employed to find

the arbitrarily tagged population. It has been proved by Nagel (1970) that this property holds for a just selection rule i.e.,

$$\inf_{\omega\in\Omega} P\{CS|R\} = \inf_{\omega\delta\Omega_0} P_\omega\{CS|R\}, \text{ if } R \text{ is just .} \tag{3.27}$$

It is also a reasonable requirement that $P_\omega\{CS|R\}$ be constant over Ω_0 because in stating the P^*-condition, we express that we are content if $P_\omega\{CS|R\}$ is at least P^* and we are not interested in exceeding P^*, at least not in Ω_0 where it can be achieved only by increasing the expected number of populations in the selected subset.

The following lemma can be applied to construct just subset selection rules with constant probability of a correct selection in Ω_0.

Lemma 3.1. Let $X_1, X_2,\ldots,X_k$ be independent and identically distributed random variables with joint distribution P_θ. Let $T(X_1, X_2,\ldots,X_k)$ be a sufficient statistic for θ.

(i) If $E(\delta(X_1,\ldots,X_k)|T) = P^*$ for all T then $E_\theta\delta = P^*$ for all θ.

(ii) If T is complete w.r.t. $\{P_\theta(x)\}$, then $E_\theta(\delta(X_1,\ldots,X_k)|T) = P^*$ is also necessary for $E_\theta\delta = P^*$ for all θ.

Gupta and Nagel (1971) have investigated the problem of constructing just rules in the cases of some discrete distributions such as binomial, Poisson and negative binomial distributions, which are discussed in the next section. They have also discussed the problem of deriving rules with constant $P\{CS|R\}$ in Ω_0 using the likelihood ratio criterion. They consider densities $f(x_i,\theta_i)$, $i = 1,\ldots,k$, where $f(x,\theta)$ is given by

$$f(x,\theta) = c(\theta)\, e^{\theta T(x)}\, h(x) . \tag{3.28}$$

Under the slippage configuration, they derive the rule

$$R: \text{ Select } \pi_i \text{ iff } T_i \geq T_{[k]} - c$$

where $c = c(k, P^*, \theta, \delta)$ is determined from the P^*-condition. This rule is just and the constant c is given by

$$\int_{-\infty}^{\infty} G_\theta^{k-1}(t + c)\, dG_\theta(t) = P^* \tag{3.29}$$

where G_θ is the cdf of T. For the normal distributions with θ as the location parameter, is independent of θ. In general, c depends on θ and, if θ is not known, an estimator of θ may be used. Since ΣT_i is a sufficient statistic for θ, this yields a selection rule of the form

$$\text{Select } \pi_i, \text{ iff } T_i \geq T_{[k]} - c(\Sigma T_i, P^*) \,. \tag{3.30}$$

By Lemma 3.1, this rule has constant probability of a correct selection in Ω_0, if $c(\Sigma T_i, P^*)$ is determined to satisfy

$$P_{\omega_0}\{T_i \geq T_{[k]} - c(\Sigma T_i, P^*) \,|\, \Sigma T_i\} = P^* \tag{3.31}$$

for all ΣT_i, $\omega_0 \in \Omega_0$. However, it is now known whether (3.30) is a just rule.

4. Selection from Discrete Populations

In this section we discuss the results of investigations of procedures for selection from k independent discrete populations. Though selection of the multinomial cell with the largest (smallest) probability where the observations are on integer valued random variables falls under this category, we discuss it in the next section along with problems concerning multivariate normal populations. The case where only the ranks of the observations are considered is discussed in the section on distribution-free procedures. Our present discussion will be mainly concerned with selection from binomial, Poisson and negative binomial populations.

Binomial Case:

We have k independent binomial populations π_i (i=1,...,k) with unknown probabilities of success on a single trial $\theta_1,\ldots,\theta_k$ respectively, where $0 \leq \theta_i \leq 1$, i=1,...,k. The following procedure R based on samples of size n from each population has been proposed by Gupta and Sobel (1960).

R: Select the population π_i iff

$$x_i \geq \max(x_1,\ldots,x_k)-d \tag{4.1}$$

where x_i is the observed number of successes in n observations from π_i and $d=d(n,k,P^*)$ is the smallest non-negative integer that will satisfy the P*-condition.

It is known that $P\{CS|R\}$ is minimized when $\theta_1=\ldots=\theta_k$. Thus, the integer d is the smallest non-negative integer for which

$$\inf_{0\leq\theta\leq 1} \sum_{\alpha=0}^{n} \binom{n}{\alpha}\theta^{\alpha}(1-\theta)^{n-\alpha}\Big[\sum_{j=0}^{\alpha+d}\binom{n}{\alpha}\theta^{j}(1-\theta)^{n-j}\Big]^{k-1} \geq P^* \ . \tag{4.2}$$

The above procedure and another procedure for the case of samples of unequal sizes along with the normal approximations for both these cases have been discussed earlier in the literature and have been briefly summarized by Gupta (1966a). It has been shown by Gupta and Sobel that for k=2, the infimum in (4.2) is attained for $\theta=1/2$, and that, for a fixed k, the value θ_0 at which the infimum takes place tends to 1/2 as $n\to\infty$. However, in general, the value of θ for which the infimum takes place is not known. When $\theta_1=\ldots=\theta_k=\theta$, $P\{CS|R\}$ can be written as a polynomial of degree nk in θ. Let

$$P\{CS|R\} = Q_{k,n,d}(\theta) = \sum_{i=0}^{nk} c_i(k,n,d)\theta^i \quad . \tag{4.3}$$

The minimum of $Q_{k,n,d}(\theta)$ is attained for some θ_0, $0 < \theta_0 < 1$ for which $\frac{dQ}{d\theta}\Big|_{\theta=\theta_0}=0$. Nagel (1966) has evaluated the coefficients $c_i(k,n,d)$ numerically for k=2(1)7, n=2(1)7 and d=0(1)n-1. It is found that the first derivative is of the form

$$\frac{dQ}{d\theta} = [\theta(1-\theta)]^{d-1}T(\theta) \tag{4.4}$$

where $T(\theta)$ is a polynomial in θ. The computations showed that $Q(\theta)$ may have several minima in (0,1). A table of Q values is given for a few selected values of k and n.

Gupta and Nagel (1971) have constructed a rule R_0 for the above binomial problem which overcomes the difficulty of finding the infimum of the probability of a correct selection. Their goal is to construct a just rule such that $P_{\underline{\omega}}\{CS|R\} = P^*$ for all $\underline{\omega} \in \Omega_0$, where $\Omega_0=\{\underline{\omega}:\underline{\omega}=(\theta,\ldots,\theta)\}$. It is clear that

this goal cannot be achieved with a nonrandomized rule, because when $\underline{\omega}=(0,\ldots,0)$ or $\underline{\omega}=(1,\ldots,1)$ the observations will be $\underline{x}=(0,\ldots,0)$ or $\underline{x}=(n,\ldots,n)$ with probability 1, requiring the use of individual selection probabilities $p_i(\underline{x})=P^*$.

The joint density for $\underline{\omega} \in \Omega_0$ is

$$(4.5) \qquad f_{\underline{\omega}}(x_1,x_2,\ldots,x_k) = (1-\theta)^{nk}\exp[(\sum_1^k x_i)\log\frac{\theta}{1-\theta}]\prod_1^k \binom{n}{x_i} .$$

We see that $T = \sum_{i=1}^k X_i$ is a sufficient statistic for θ. Since we are interested in symmetric rules R it is sufficient to know one of the individual selection probabilities, say, p_k. From Lemma 3.1 it follows that

$$(4.6) \qquad E(p_k(\underline{X})|T) = P^* \quad \text{for } T = 0,1,\ldots,kn.$$

The requirement that R be just leads to

$$(4.7) \qquad \begin{cases} y_i \le x_i, \ i=1,2.,,,.k-1 \\ \\ y_k \ge x_k \end{cases} \Rightarrow p_k(x_1,x_2,\ldots,x_k) \le p_k(y_1,y_2,\ldots,y_k).$$

Figure 1 shows the partial ordering induced by (4.7) among the observation vectors for the case k=3, n=2. The indiviaual selection probability $p_3(x_1,x_2,x_3)$ defines a just rule if its values are nondecreasing in the direction of the arrows. Because of symmetry only one of the two permutations (x_1,x_2,x_3) and (x_2,x_1,x_3) is plotted. The numbers underneath the observation vectors denote the corresponding T values.

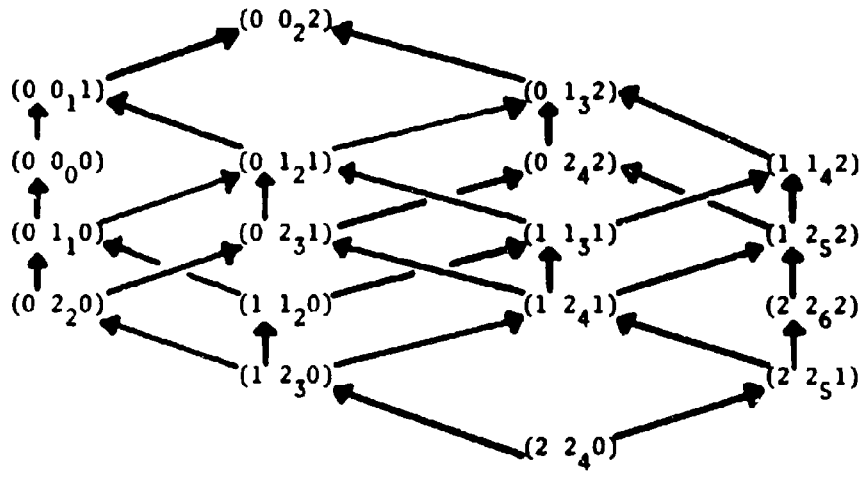

Figure 1. Partial Ordering for Binomial

Observations k=3, n=2.

The conditions (4.6) and (4.7) do not determine a rule uniquely. Gupta and Nagel have proposed the following rule R_0:

$$(4.8) \qquad p_k(\underline{x}) = \begin{cases} 1 & \text{if } x_k > c_T \\ \rho & \text{if } x_k = c_T \\ 0 & \text{if } x_k < c_T \end{cases}$$

where $\rho = \rho(T,P^*,k)$ and $c_T = c_T(P^*,k)$ are determined to satisfy

$$(4.9) \qquad E(p_k(\underline{X})|T) = P\{X_k > c_T|T\} + \rho P\{X_k = c_T|T\} = P^* .$$

The conditional distribution of X_k given T is hypergeometric:

$$(4.10) \qquad P\{X_k = i|T\} = \frac{\binom{n}{i}\binom{(k-1)n}{T-i}}{\binom{kn}{T}} .$$

Let Z_T have the same distribution as X_k given T. Then (4.9) becomes

$$P\{Z_T > c_T\} + \rho P\{Z_T = c_T\} = P^* \tag{4.11}$$

and the constant c_T is smallest integer determined from the inequalities

$$P\{Z_T > c_T\} \leq P^* \tag{4.12}$$

and

$$P\{Z_T \geq c_T\} > P^* \tag{4.13}$$

From (4.11), we have

$$\rho = \frac{P^* - P\{Z_T > c_T\}}{P\{Z_T = c_T\}} . \tag{4.14}$$

It has been established by Gupta and Nagel (1971) that the above rule R_0 is just. They have also tabulated the values of c_T and ρ for k=2,3,5; n=5,10 and P*=.75, .90, .95, .99, in each case T going from 0 to nk.

Since T takes on the values 0,1,...,kn these tables become very extensive for large values of k and n. Therefore it is desirable to find approximations for c_T and ρ. The normal approximation for the hypergeometric distribution gives good results when n is large and T is not extreme (close to 0 or kn). The expectation and variance of Z_T are $\mu = \frac{T}{k}$ and $\sigma^2 = \frac{T(kn-T)(k-1)}{(kn-1)k^2}$ respectively. Using the fact that asymptotically Z_T is $N(\mu,\sigma^2)$, we obtain approximate value $\tilde{c}_T$ given by $\tilde{c}_T = [\frac{1}{2} + \mu - \sigma\, \phi^{-1}(P^*)]$ where ϕ^{-1} is the inverse of the standard normal cdf and [x] is the integral part of x. For ρ we get the approximate value $\tilde{\rho} = \tilde{c}_T + 0.5 - (\mu - \sigma\phi^{-1}(P^*))$. The exact and approximate values of c_T and ρ have been compared by Gupta and Nagel for

k=2,3,5,10; n=5,10,20; and some selected values of T and P*. The results show no change in the values of c_T and $\tilde{c}_T$ and only small deviations in the values of ρ and $\tilde{\rho}$.

The nonrandomized version R_0' of R_0, namely, R_0': Select π_i iff $x_i \geq c_T$, is conservative in the sense of meeting the basic probability requirement. However, R_0' may not be just and it selects large subsets if the θ_i's are close to zero or one. A comparison of R_0 and R is difficult because $\inf_{\Omega} P_{\underline{\omega}}\{CS|R\}$ is not known in the case of R. Since it takes place near $\theta = \frac{1}{2}$, the P*-value for R_0 has been chosen by Gupta and Nagel to satisfy $P_{\underline{\omega}}(CS|R) = P^*$ with $\underline{\omega} = (\frac{1}{2},\frac{1}{2},\ldots,\frac{1}{2})$ which makes the comparison slightly more favorable for R. Under slippage configuration $(\theta,\ldots,\theta, \theta+\delta)$, the numerical computations show that R_0 yields better results for small values of δ, while R is better for large δ. Hence R_0 should be applied if small differences in the success probabilities are expected. This advantage of R_0 becomes more evident in the case of equally spaced configurations, where almost surely more than half of the populations will be retained in the selected subset if the number of observations is increased indefinitely, whereas R will eventually select only the best one.

Gupta and Nagel (1971) have studied rules similar to R_0 defined by (4.8) for the problem of selection from Poisson and negative binomial distributions. The case of Fisher's logarithmic distributions has been discussed by Nagel (1970).

In connection with selection from discrete populations Nagel (1966) considered the problem of minimizing

$$A = \sum_{i=0}^{n} a_i \left(\sum_{j=0}^{i+d} a_j \right)^{k-1} \tag{4.15}$$

under the condition

$$\sum_{i=0}^{n} a_i = 1, \; a_i \geq 0 \text{ for } i = 0,\ldots,n \; . \tag{4.16}$$

Setting

$$A_i = \sum_{j=0}^{i} a_j, \; i = 0,\ldots,n; \; A_i = 0, \; i < n; \; A_i = A_n, \; i > n, \tag{4.17}$$

we have

$$A = \sum_{i=0}^{n} (A_i - A_{i-1}) \, A_{i+d} \; . \tag{4.18}$$

For d = 0, it has been shown that the minimum of A is given by

$$A_{min}(k,n) = \frac{1}{k} + \frac{k-1}{k} a_n \; . \tag{4.19}$$

If $b_k = (k-1)/k^{k/k-1}$, then

$$A_{min}(k,n+1) = 1-b_k/(A_{min}(k,n))^{1/k-1} \; . \tag{4.20}$$

$A_{min}(k,n)$ has been tabulated for k=2(1)8 and n=1(1)25. The case of d > 0 can be handled using the results for d = 0 case.

5. Selection Procedures for Multinomial and Multivariate Normal Distributions.

I. Multinomial Case.

Let $p_1, p_2, \ldots, p_k$ be the unknown cell-probabilities in the multinomial distribution with $\sum_{i=1}^{k} p_i = 1$. Let $x_1, x_2, \ldots, x_k$ be the respective observations in the k cells of the distribution with $\sum_{1}^{k} x_i = N$. Let the ordered cell-probabilities be given by $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]}$. For selecting a subset of the cells containing the cell associated with $p_{[k]}$, Gupta and Nagel (1967) proposed and investigated the following procedure

R_1: Select the cell with observed x_i iff

$$x_i \geq \max(x_1, \ldots, x_k) - D \tag{5.1}$$

where D is a given non-negative integer. Using this rule the probability of a correct selection is given by

$$P\{CS|R_1\} = P(k,N,D;\ p_{[1]}, \ldots, p_{[k]}) = \sum_{\substack{\nu_i = N \\ \nu_i \leq \nu_k + D}} \frac{N!}{\nu_1! \ldots \nu_k!}\, p_{[1]}^{\nu_1} \cdots p_{[k]}^{\nu_k}, \tag{5.2}$$

$$i = 1, 2, \ldots, k$$

Then the following lemma can be established.

Lemma 5.1. (i) If the sum $p_{[i]} + p_{[j]}$, $1 \leq i < j < k$, is kept constant, $P\{CS|R_1\}$ decreases as we pass from the configuration $(p_{[1]}, \ldots, p_{[i]}, \ldots, p_{[j]}, \ldots, p_{[k]})$ to $(p_{[1]}, \ldots, p_{[i]} - \varepsilon, \ldots, p_{[j]} + \varepsilon, \ldots, p_{[k]})$ where $0 < \varepsilon \leq p_{[i]}$.

(ii) If the sum $p_{[i]} + p_{[k]}$, $1 \leq i < k$, is kept constant, $P\{CS|R_1\}$ decreases as we pass from the configuration $(p_{[1]},\ldots,p_{[i]},\ldots,p_{[k]})$ to $(p_{[1]},\ldots,p_{[i]} + \varepsilon,\ldots,p_{[k]} - \varepsilon)$ where $0 < \varepsilon \leq p_{[k]}$.

By using this lemma, the following theorem is obtained.

Theorem 5.1. Let μ be the smallest integer such that $p_{[\mu]} > 0$ and let ν be the largest integer such that $p_{[\nu]} < p_{[k]}$. Then, for a configuration minimizing $P\{CS|R_1\}$, $\mu \geq \nu$. In particular, if $\mu = k-1$, then $\mu > \nu$.

As a consequence of the above theorem, we have

$$\inf_{\Omega} P\{CS|R_1\} = \min_{r=2,\ldots,k} \left(\min_{\frac{1}{r} \leq p \leq \frac{1}{r-1}} F(k,N,D;(0,\ldots,0,s,p,\ldots,p)) \right) \tag{5.3}$$

where $s = 1 - (r-1)p$ and Ω is the space of all configurations of $p_1,\ldots,p_k$.

For the purposes of computations it is not necessary to consider the cases where $r < k$, when the problem is already solved for all smaller values of k for the same N and D. In other words, we need consider only vectors of the type $(s,p,\ldots,p)$, $s = 1 - (k-1)p$. On the basis of numerical evaluations of $F(k,N,D;(s,p,\ldots,p))$ done for $D = 0(1)4$, $k = 2(1)10$ and $N=2(1)15$, it was found that the minimum over p took place either for $p = \frac{1}{k}$ or for $p = \frac{1}{k-1}$ except in the case of $k = 3$, $N = 6$ and $D = 4$ for which the minimum was attained in the interior of the interval $(\frac{1}{k}, \frac{1}{k-1})$.

Consider the configuration $(p,\ldots,p,Ap)$, $A \geq 1$. For any D, the expected subset size is given by

$$E(S) = \sum_{\Sigma\nu_i=N} \frac{N!}{\nu_1!\ldots\nu_k!} p_1^{\nu_1}\ldots p_k^{\nu_k} B_\nu \tag{5.4}$$

where B_ν = number of ν_i's $\geq \nu_{max} - D$. The probability of selecting a non-best population is given by $\frac{E(S) - P\{CS|R\}}{k-1}$. Tables have been provided by Gupta and Nagel (1967) for the values of $P\{CS|R_1\}$, expected proportion of cells selected and the probability of selecting a non-best population corresponding to the configuration $(p,\ldots,p,Ap)$, $A \geq 1$ for $k = 2(1)10$, $N = 2(1)15$, $A = 1(2)5$ and $D = 0(1)2$. Another table gives the minimum D such that $\inf_{\Omega} P\{CS|R_1\} \geq P^*$ for $k = 2(1)10$, $N = 2(1)15$ and $P^* = .75, .90$.

For selecting a subset containing $p_{[1]}$, Gupta and Nagel investigated the rule R_2 which selects the cell with observation x_i iff

$$x_i \leq \min(x_1,\ldots,x_k) + C \tag{5.5}$$

where C is a given non-negative integer. In this case the probability of a correct selection is given by

$$P\{CS|R_2\} = G(k,N,C;\ p_{[1]},\ldots,p_{[k]}) \tag{5.6}$$

$$= \sum_{\substack{\Sigma\nu_i=N \\ \nu_j \geq \nu_1 - c,\ j=1,\ldots,k}} \frac{N!}{\nu_1!\ldots\nu_k!}\, p_{[1]}^{\nu_1}\cdots p_{[k]}^{\nu_k}.$$

The following lemma has been proved.

Lemma 5.2. (i) If the sum $p_{[i]} + p_{[j]}$, $1 < i < j \leq k$, is kept constant, $P\{CS|R_2\}$ decreases as we pass from the configuration $(p_{[1]},\ldots,p_{[i]},\ldots,p_{[j]},\ldots,p_{[k]})$ to $(p_{[1]},\ldots,p_{[i]}-\varepsilon,\ldots,p_{[j]}+\varepsilon,\ldots,p_{[k]})$ where $0 < \varepsilon \leq p_{[i]}$.

(ii) If the sum $p_{[1]} + p_{[j]}$, $1 < j \leq k$, is kept constant, $P\{CS|R_2\}$ decreases as we pass from the configuration $(p_{[1]},\ldots,p_{[j]},\ldots,p_{[k]})$ to $(p_{[1]}+\epsilon,\ldots,p_{[j]}-\epsilon,\ldots,p_{[k]})$ where $0 < \epsilon \leq p_{[j]}$.

As a consequence of Lemma 5.2 the following theorem is obtained.

Theorem 5.2. $P\{CS|R_2\}$ is minimized at a configuration $(p_{[1]},\ldots,p_{[k]})$ given by $(p,\ldots,p,q)$, where $q = 1 - (k-1)p$, $0 < p \leq \frac{1}{k}$.

Numerical evaluation of $G(k,N,C;p,\ldots,p,q)$ for $k = 2(1)10$, $N = 2(1)15$ and $C = 0(1)4$ show that the overall minimum is given by the configuration $(\frac{1}{k},\ldots,\frac{1}{k})$. For the configuration $(p/A,p,\ldots,p)$, $A \geq 1$, tables are available for the expected proportion, $P\{CS|R_2\}$ and the probability of selecting any fixed cell with probability p for $k = 2(1)10$, $N = 2(1)15$, $A = 1(2)5$ and $c = 0(1)2$.

As we have seen above, Gupta and Nagel procedures are based on a fixed sample size. For the problem of selecting the cell with $p_{[k]}$, Panchapakesan (1971) proposed a procedure R_3 which is based on inverse sampling. Observations are taken one at a time until the count in any cell reaches a given number M. Let $x_1, x_2,\ldots,x_k$ be the cell-counts at termination. Then R_3 is defined as follows:

R_3: Select the cell with count x_i iff

$$x_i \geq M - D \tag{5.7}$$

where D is a non-negative integer. For the rule R_3 the probability of a correct selection is given by

$$P\{CS|R_3\} = 1 - \sum_{\alpha=1}^{k-1} L_\alpha \tag{5.8}$$

where

$$L_\alpha = \sum \frac{M(\nu_1+\ldots+\nu_{k-1})!}{\nu_1!\ldots\nu_k!}\, p_{[1]}^{\nu_1} \cdots p_{[k]}^{\nu_k}, \tag{5.9}$$

the summation being over the set of values of $\nu_1,\ldots,\nu_k$ such that $\nu_\alpha = M, 0 \le \nu_k \le M-D-1$ and $0 \le \nu_\beta \le M-1$, $\beta = 1,\ldots,k-1$; $\beta \ne \alpha$. This multiple sum can be expressed in an integral form and we get

$$P\{CS|R_3\} = 1 - \frac{\Gamma((k-1)M+M')}{[\Gamma(M)]^k \Gamma(M')}\, T_\alpha , \tag{5.10}$$

where $M' = M-D$,

$$T_\alpha = \int_{\frac{\phi_1}{\phi_\alpha}} \cdots \int_{\frac{\phi_k}{\phi_\alpha}} \frac{\left(\prod_{i=1}^{k-2} y_i^{M-1}\right) y_{k-1}^{M'-1}}{(1+y_1+\ldots+y_{k-1})^{(k-1)M+M'}}\, dy_1 \cdots dy_{k-1} \tag{5.11}$$

and $\phi_i = p_{[i]}$, $i = 1,\ldots,k$.

It has been established by Panchapakesan that the statement of Lemma 5.1 holds in the case of R_3, and hence that

$$\inf_{\Omega} P\{CS|R_3\} = \min_{r=2,\ldots,k} \left(\min_{\frac{1}{r} \le p < \frac{1}{r-1}} F(k,M,D; (0,\ldots,0,s,p,\ldots,p)) \right) \tag{5.12}$$

where Ω is the space of all configurations of the cell-probabilities, r is the number of positive cell-probabilities in the configuration $(0,\ldots,0,s,p,\ldots,p)$, $0 < s \le p$, and $F(k,M,D; (0,\ldots,0,s,p,\ldots,p))$ is the probability of a correct selection for this configuration. Subject to the condition that $s + (r-1)p = 1$, it has been shown that, for every fixed r, $P\{CS|R_3\}$ increases in p and hence

$$\inf_{\Omega} P\{CS|R_3\} = \min_{r=2,\ldots,k} F_r(k,M,D) \tag{5.13}$$

where $F_r(k,M,D)$ denotes the probability of a correct selection for the configuration $(0,\ldots,0,\frac{1}{r},\ldots,\frac{1}{r})$. It has been recently shown (unpublished) that $F_r(k,M,D)$ is monotonically decreasing in r. Thus

$$\inf_{\Omega} P\{CS|R_3\} = F_k(k,M,D). \tag{5.14}$$

For R_3, the number of observations (n) is a random variable. Exact and asymptotic expressions for E(n) corresponding to the configuration $\phi_1 = \ldots = \phi_k = \frac{1}{k}$ are written down using earlier available results. Specific results have been obtained for the special case k=2.

For selecting the cell associated with $p_{[k]}$, Nagel (1970) constructed a symmetric rule based on N observations, which yields a minimum of PCS when the cell-probabilities are equal and which maximized PCS for the configuration $(\theta,\ldots,\theta,\theta+\delta)$ where $\delta > 0$ and $k\theta + \delta = 1$. His rule R_4 is a randomized rule which selects the cell with observation x_i with probability p_i where

$$p_i = \begin{cases} 1 & \text{if } x_i > d \\ \rho & \text{if } x_i = d \\ 0 & \text{if } x_i < d \end{cases}, \tag{5.15}$$

where $d > 0$ is determined from

$$(\tfrac{1}{k})^N \sum_{i=d+1}^{N} \binom{N}{i} (k-1)^{N-i} < P^* \tag{5.16}$$

and

$$(\tfrac{1}{k})^N \sum_{i=d}^{N} \binom{N}{i} (k-1)^{N-i} \geq P^* . \tag{5.17}$$

It follows from above that

$$\rho = \frac{P^* k^N - \sum_{i=d+1}^{N} \binom{N}{i} (k-1)^{N-i}}{(k-1)^{N-d}} \tag{5.18}$$

II. Multivariate Normal Case.

Selection problems for multivariate normal populations have been investigated when the populations are ranked in terms of (i) generalized variance (ii) distance function and (iii) multiple correlation coefficient. In the following discussion of these investigations, we assume that $\pi_1,\ldots,\pi_k$ are independent p-variate normal populations, where π_i has mean vector μ_i and covariance matrix $\Sigma_i (i = 1,2,\ldots,k)$. Let x_{ij}, $j = 1,2,\ldots,n$, be a sample of size n of vector observations from π_i and $S_i = \frac{1}{n-1} \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)(x_{i\alpha} - \bar{x}_i)'$.

(a) Selection in terms of Generalized Variance, $|\Sigma|$. In this case μ_i and Σ_i are unknown. For selecting a subset containing the population associated with the smallest $|\Sigma_i|$, Gnanadesikan and Gupta (1970) studied the following rule R, based on the sample covariance matrices S_i, $i = 1,..,n$.

R: Select the population π_i iff

$$|S_i| \leq \frac{1}{c} |S|_{\min} , \tag{5.19}$$

where $|S|_{\min} = \min(|S_1|,\ldots,|S_k|)$ and $0 < c \leq 1$. It has been established that

$$\inf_{\Omega} P\{CS|R\} = P\{Y_1 \leq \frac{1}{c} Y_j;\ j = 2,\ldots,k\} , \tag{5.20}$$

where $Y_i (i = 1,\ldots,k)$ are k independent random variables, each being the product of p independent factors, the rth factor being distributed as a chi-square variable with (n-r) degrees of freedom.

The exact distribution of Y_i is unknown except when p=2. In the case of p=2, we get $\inf_{\Omega} P\{CS|R\} = P\{Z_1 \leq \frac{1}{\sqrt{c}} Z_j;\ j = 2,\ldots,k\}$, where Z_i, i = 1,...,k, are k independent random variables each having a chi-square distribution with 2(n-2) degrees of freedom. If, further k=2, then $c^{1/2}$ is the 100(1-P*) percentage point of an F variable with (2n-4, 2n-4) degrees of freedom.

When p > 2, one can use Hoel's approximation for the distribution of Y_i in (5.20) or use the approximation of log χ^2 by the normal distribution. Some study of these approximations were made by Gnanadesikan and Gupta.

Further, the performance of the procedure R was studied in terms of risk functions using three different loss functions. If the ordered generalized variances are denoted by $|\Sigma|_{[1]} \leq |\Sigma|_{[2]} \leq \cdots \leq |\Sigma|_{[k]}$, the different loss functions that were considered for the loss incurred by including the population whose generalized variance is Σ_i, are:

(i) $L_1(\Sigma_i) = |\Sigma|_i / |\Sigma|_{[1]} - 1.0$,

(ii) $L_2(\Sigma_i)$ = (Rank of the population π_i)$/\frac{k(k+1)}{2}$, where the ranks increase along with the generalized variance, and,

(iii) $L_3(\Sigma_i) = \frac{S}{k}$, where S is the number of populations included in the subset.

The computations of the risk functions associated with the above loss functions, for p=2, k=2(1)5, $|\Sigma|_{[i]}/|\Sigma|_{[1]} = a^{2i-2}$, when a = 1.2(.2)2.0(.5)3.0, n = 3(1)7 and P* = .75, indicate that $E(L_2)$ and $E(L_3)$ are sensitive to changes in the values of the parameters and are decreasing functions of a and n. In the case of $E(L_1)$, it increases in the range of values of a considered when n=3 and, for

other values of n, it increases up to a certain point and then decreases as a increases. This lack of monotonicity in the behavior of $E(L_1)$, as the 'best' population moves further away from the other populations, and the difficulty of its interpretation render $E(L_1)$ less suitable than L_2 and L_3. Comparing L_2 and L_3, due to the ease of interpretation, L_3 would be more appropriate as the criterion of performance of the procedure R. Finally, the procedure R is shown to be monotone.

Suppose we consider a partition of the p variables into two sets of q_1 and q_2 components, respectively, where $q_1 + q_2 = p$. The corresponding partition of Σ_i is denoted by

$$\Sigma_i = \begin{pmatrix} \Sigma_{11}^{(i)} & \Sigma_{12}^{(i)} \\ \Sigma_{21}^{(i)} & \Sigma_{22}^{(i)} \end{pmatrix}, \quad i = 1,\ldots,k.$$

Here we assume that Σ_i, $\Sigma_{11}^{(i)}$, $\Sigma_{22}^{(i)}$ are all positive definite. We are interested in selecting a subset containing the population associated with the smallest $|\Sigma_i|/|\Sigma_{11}^{(i)}| \equiv |\Sigma_{22}^{(i)} - \Sigma_{21}^{(i)}\ \Sigma_{11}^{(i)^{-1}}\ \Sigma_{12}^{(i)}| = \sigma_i$, say. In other words, if we consider for each population the conditional distribution of the q_2 set when the q_1 set is fixed, then our criterion of ranking is the conditional generalized variance. If the observations are taken on the variables of the q_2 set, holding the variables of the q_1 set fixed, then the problem reduces to selection in terms of the generalized variance for the conditional normal distributions with dimensionality q_2, a problem solved by Gnanadesikan and Gupta (1970). Let us consider the unconditional case in which all the p variables are random and observations are taken on all of them

and use σ_i as the criterion for ranking. Then consider the partition of the sample covariance matrix S_i denoted by

$$S_i = \begin{pmatrix} S_{11}^{(i)} & S_{12}^{(i)} \\ S_{21}^{(i)} & S_{22}^{(i)} \end{pmatrix}.$$

We compute $s_i = |S_{22}^{(i)} - S_{21}^{(i)} S_{11}^{(i)^{-1}} S_{12}^{(i)}|$. Gupta and Panchapakesan (1969a) studied the following rule R' for selecting the population with smallest σ_i.

R': Select π_i iff

$$s_i \leq \frac{1}{c'} \min(s_1, \ldots, s_k) \tag{5.21}$$

where $0 < c' = c'(k, P^*, n, q_1, q_2) < 1$ is chosen to satisfy the P*-condition. It is shown that

$$\inf_{\Omega} P\{CS|R\} = \int_0^\infty [1 - G(c'x)]^{k-1} dG(x) , \tag{5.22}$$

where G(x) is the cdf of a random variable which is the product of q_2 independent χ^2 variables with degrees of freedom $n-q_1-1$, $n-q_1-2, \ldots, n-q_1-q_2$, respectively.

(b) Selection in terms of distance function.

Suppose the mean vectors μ_i are unknown and $\Sigma_i = \Sigma$(known) for all i. Let $\lambda_i = \mu_i' \Sigma^{-1} \mu_i$, the Mahalanobis distance function of the population π_i from the origin. Let $y_{ij} = x_{ij}' \Sigma^{-1} x_{ij}$; $j = 1, \ldots, n$; $i = 1, \ldots, k$. Then

$y_i = \sum_{j=1}^{n} y_{ij}$ has the non-central χ^2 distribution with np degrees of freedom and non-centrality parameter $\lambda_i' = n\lambda_i$. We are interested in selecting a subset containing the population with the largest λ_i. Gupta (1966b) proposed and studied the following rule R.

R: Select the population π_i iff

$$y_i \geq c \max(y_1, \ldots, y_k) \tag{5.23}$$

where $0 < c = c(k,n,p,P^*) < 1$ is determined to satisfy the P*-condition. The probability of a correct selection is given by

$$P\{CS|R\} = \int_0^\infty \left[\prod_{j=1}^{k-1} F_{\lambda'_{[j]}}\left(\frac{x}{c}\right)\right] dF_{\lambda'_{[k]}}(x) , \tag{5.24}$$

where $\lambda'_{[1]} \leq \lambda'_{[2]} \leq \ldots \leq \lambda'_{[k]}$ are the ordered λ' values and $F_{\lambda'}(x)$ denotes the distribution function of a non-central χ^2 variable with np degrees of freedom and non-centrality parameter λ'. Since $\{F_{\lambda'}\}$ is stochastically increasing in λ',

$$\inf_{\Omega} P\{CS|R\} = \inf_{\lambda' \geq 0} \int_0^\infty F_{\lambda'}^{k-1}\left(\frac{x}{c}\right) dF_{\lambda'}(x) . \tag{5.25}$$

Gupta showed that, for k=2, the integral on the right hand side of (5.25) is non-decreasing in λ' and hence the infimum takes place when $\lambda'=0$. Thus, the constant c satisfies the condition

$$\int_0^\infty G_m\left(\frac{x}{c}\right) dG_m(x) = P^* , \tag{5.26}$$

where $G_m(x)$ is the central χ^2 distribution with np degrees of freedom.

For selecting the population associated with $\lambda'_{[1]}$, a similar procedure was studied, namely,

R': Select π_i iff

$$y_i \leq b \min(y_1,\ldots,y_k) , \tag{5.27}$$

where $b = b(k,n,p,P^*) > 1$ is determined so as to satisfy the P*-condition. In this case, we obtain

$$\inf_{\Omega} P\{CS|R\} = \inf_{\lambda' \geq 0} \int_0^\infty [1-F_{\lambda'}(\tfrac{x}{b})]^{k-1} dF_{\lambda}(x) . \tag{5.28}$$

The integral is shown to be monotonically increasing in λ' for $k = 2$.

For the procedures R and R' defined above Gupta and Studden (1970) established the monotonicity of the integrals appearing in (5.25) and (5.28) w.r.t. λ' in the general case $k \geq 2$. They proved the following theorem for that purpose.

Theorem 5.3. Let $g_j(x)$, $j = 0,1,2\ldots$ be a sequence of density functions on the interval $[0,\infty)$ and define

$$f_\lambda(x) = \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} g_j(x), \quad x \geq 0. \tag{5.29}$$

For a fixed integer $k \geq 2$ and $c > 1$, let

$$I(\lambda) = \int_0^\infty F_\lambda^{k-1}(cx)\, dF_\lambda(x) \tag{5.30}$$

and

$$J(\lambda) = \int_0^\infty [1-F_\lambda(\tfrac{x}{c})]^{k-1} dF_\lambda(x) . \tag{5.31}$$

Let A denote the condition that, for each $\ell \geq 0$

$$(5.32) \qquad \sum_{i=0}^{\ell} \frac{1}{i!(\ell-i)!} [\{G_{i+1}(cx) - G_i(cx)\} g_{\ell-i}(x) - c\, g_i(cx) \{G_{i+1}(x) - G_{\ell-i}(x)\}] \geq 0 .$$

Then, the functions $I(\lambda)$ and $J(\lambda)$ are non-decreasing in λ provided that the the condition A holds. Further, both the functions are strictly increasing in λ if the condition A holds with strict inequality for some integer ℓ.

As pointed out earlier, the condition (5.32) can be obtained from the condition (3.9). In fact, Gupta and Studden verify in the cases of non-central chi-square and non-central F distributions a condition which is stronger than (5.32). This stronger condition states that the sum of the terms in the left hand side of (5.32) corresponding to i and ℓ-i, $i = 0,\ldots,[\ell/2]$, is positive and this is same as the condition (3.12) for proper choices of h(x) and the weight functions.

To be precise, Gupta and Studden considered the case where Σ_i are all not necessarily equal but known. With a slight modification, namely, $y_{ij} = x_{ij}' \Sigma_i^{-1} x_{ij}$, we have essentially Gupta's procedures R and R'. They also studied procedures when Σ_i's are different but all unknown. In this case, let $z_i = \tilde{x}_i' S_i^{-1} \tilde{x}_i$. Then, for the selection of the population with the largest and smallest distance functions, the procedures studied are, respectively,

R_1: Select π_i iff

$$(5.33) \qquad cz_i \geq \max(z_1,\ldots,z_k)$$

and

R_1': Select π_i iff

$$z_i \leq b \min(z_1,\ldots,z_k) \tag{5.34}$$

where $c = c(k,p,n,P^*) > 1$ and $b = b(k,p,n,P^*) > 1$ are determined so that P^*-condition is satisfied. It is known that z_i is essentially distributed as a non-central F variable, whose density is of the form (5.29). Hence Theorem 5.3 applies in this case. It is shown that [illegible] sufficient condition A is satisfied. Thus we obtain the equations to determine the constants c and d, namely,

$$\int_0^\infty F_{p,n-p}^{k-1}(cx)\, dF_{p,n-p}(x) = P^* \tag{5.35}$$

and

$$\int_0^\infty [1-F_{p,n-p}(x|b)]^{k-1}\, dF_{p,n-p}(x) = P^* \,. \tag{5.36}$$

Alam and Rizvi (1966) have also considered the problem of selection in terms of distance function. For Σ_i unknown, their procedure is same as that of Gupta and Studden (which was originally studied in a technical report issued in 1965) but the monotonicity of the integral involved is established rather directly and not by obtaining a sufficient condition applicable to a class of distributions including non-central chi-square and non-central F distributions. Further, in the case of Σ_i known, Alam and Rizvi use the procedure R_1 defined by (5.33) with Σ_i in the place of S_i; in other words, using the statistics $z_i = \bar{x}_i' \Sigma_i^{-1} \bar{x}_i$. This is different from the procedure of Gupta (1966b) and Gupta and Studden (1970), who have observed the undesirability of using $\bar{x}_i' \Sigma_i^{-1} \bar{x}_i$ in the sense that the constant evaluated subject to the P^*-condition is independent of n.

(c) Selection in terms of multiple correlation coefficient.

Let $\rho_i \equiv \rho_{1.2\ldots p}$ be the multiple correlation coefficient between the first variable and the rest in the population π_i. Let $0 \leq \rho_{[1]} \leq \ldots \leq \rho_{[k]} \leq 1$ be the ordered values of the ρ_i. Gupta and Panchapakesan (1969a) investigated the problem of selecting a subset containing the population associated with $\rho_{[k]}$ (or $\rho_{[1]}$). Denote the sample multiple correlation coefficients by $R_i \equiv R^{(i)}_{1.2\ldots p}$. Two cases arise:

(i) The case in which $x_{i2},\ldots,x_{ip}$ are fixed, called the conditional case;

(ii) The case in which $x_{i2},\ldots,x_{ip}$ are random, called the unconditional case.

The following rule R has been investigated by Gupta and Panchapakesan for the selection of $\rho_{[k]}$.

R: Select π_i iff

$$R^{*2} \geq c \max (R_1^{*2},\ldots,R_k^{*2}) \tag{5.37}$$

where $R_i^{*2} = R_i^2/(1-R_i^2)$, $i = 1,\ldots,k$, and $0 < c = c(k,P^*,p,n) < 1$ is chosen subject to the P*-condition. In the formal statement of R we do not make the distinction between the conditional and unconditional cases.

Letting $\lambda_i = \rho_i^2$, $i = 1,\ldots,k$, the distribution of R_i^{*2} is given by

$$u_\lambda(x) = \sum_{j=0}^{\infty} \frac{\Gamma(q+m+j)\lambda^j}{\Gamma(q+m)j!} (1-\lambda)^{q+m} f_{2(q+j),2m}(x) \tag{5.38}$$

in the unconditional case and by

$$u_\lambda(x) = \sum_{j=0}^{\infty} \frac{e^{-m\lambda}(m\lambda)^j}{j!} f_{2(q+j),2m}(x) \tag{5.39}$$

in the conditional case, where

(5.40) $\qquad q = (p-1)/2 \ , \ m = (n-p)/2$

and $f_{r,s}(x)$ denotes the density of the F-distribution with r and s degrees of freedom. It is easy to show that $u_\lambda(x)$ has a monotone likelihood ratio in x and hence the distribution of R^{*2} is stochastically increasing in λ. Thus we obtain

$$(5.41) \qquad \inf_{\Omega} P\{CS|R\} = \inf_{\lambda} \int_0^\infty U_\lambda^{k-1}(x/c)\, dU_\lambda(x) \ ,$$

where $U_\lambda(x)$ is the cdf corresponding to $u_\lambda(x)$.

In the conditional case, the condition A of Theorem 5.3 is satisfied and hence the infimum takes place for $\lambda = 0$. For the unconditional case the same result is shown by proving the following theorem.

Theorem 5.4. Let $g_j(x)$, $j=0,1,2,\ldots$ be a sequence of density functions on the interval $[0,\infty)$ and define

$$(5.42) \qquad f_\lambda(x) = \sum_{j=0}^{\infty} \frac{\Gamma(q+j)}{\Gamma(q)} \frac{\lambda^j}{j!} (1-\lambda)^q g_j(x), \ x \geq 0, \ 0 \leq \lambda < 1.$$

For a fixed integer $k \geq 2$ and $0 < c < 1$, let $I(\lambda)$ and $J(\lambda)$ be defined as in (5.30) and (5.31). Let B denote the condition that, for each integer $\ell \geq 0$

$$(5.43) \qquad \sum_{i=0}^{\ell} \frac{(q)_i (q)_{\ell-i}}{i!(\ell-i)!} [(q+i)\{G_{i+1}(x|c) - G_i(x|c)\}g_{\ell-i}(x)$$

$$-c^{-1}(q+\ell-i)g_i(x|c)\{G_{\ell-i+1}(x) - G_{\ell-i}(x)\}] \geq 0$$

where $(q)_s = q(q+1)\ldots(q+s-1)$ and $G_j(x)$ is the cdf corresponding to $g_j(x)$.

Then, $I(\lambda)$ and $J(\lambda)$ are non-decreasing in λ if condition B holds and the two functions are strictly increasing in λ if strict inequality holds in condition B for some integer ℓ.

It can be easily verified that the condition B is satisfied in the unconditional case. Thus, in either case, we get

$$\inf_{\Omega} P\{CS|R\} = \int_0^\infty F_{2q,2m}^{k-1}(x/c)\, dF_{2q,2m}(x) , \tag{5.44}$$

where $F_{2q,2m}(x)$ is the cdf corresponding to $f_{2q,2m}(x)$. Since the distribution of R^{*2} when $\lambda = 0$ is the same in both conditional and unconditional cases, the constant c used in the procedure is the same and is given by

$$\int_0^\infty F_{2q,2m}^{k-1}(x/c)\, dF_{2q,2m}(x) = P^* . \tag{5.45}$$

When q and m are integers, i.e., p and n are odd, we can use series expansion for $F_{2q,2m}(x)$ and obtain formulae for computing c for specified values of q,m and P^*. The final result is:

$$P^* = \frac{\Gamma(q+m)}{\Gamma(q)\Gamma(m)(1-c)^m} \times \sum_{\alpha=0}^{qk-1} \sum_{j=0}^{(k-1)(m-1)} (-1)^\alpha \binom{qk-1}{\alpha} a(k-1,j) \left(\frac{c}{1-c}\right)^{\alpha+j} K(c,m,q,\alpha,j) \tag{5.46}$$

where $a(r,j)$ and $K(c,m,q,\alpha,j)$ are given by the following recurrence relations:

$$a(1,j) = \begin{cases} 1 , & j = 0 \\ q(q+1)\ldots(q+j-1), & 1 \le j \le m-1 . \end{cases} \tag{5.47}$$

and for $r > 1$

$$(5.48)\qquad a(r,j) = \begin{cases} 1 & j = 0 \\ \sum_{s=\max(j-(r-1)(m-1),0)}^{\min(m-1,j)} a(1,s)a(r-1,j-s), & 1 \le j \le r(m-1). \end{cases}$$

$$(5.49)\qquad K(c,m,q,\alpha,j)$$

$$= \begin{cases} \dfrac{\Gamma(m+\alpha+j)\Gamma(q-\alpha-j)}{\Gamma(m+q)} I_{1-c}(m+\alpha+j, q-\alpha-j), & q > \alpha+j \\ \sum_{\ell=1}^{m+q-1} \binom{m+q-1}{\ell} (-1)^{\ell} \dfrac{\{1-c^{\ell}\}}{\ell} - \log c, & q = \alpha+j \\ \sum_{\substack{\ell=0 \\ \ell\neq\alpha+j-q}}^{m+\alpha+j-1} \binom{m+\alpha+j-1}{\ell} (-1)^{\ell} \dfrac{\{1-c^{\ell-\alpha-j+q}\}}{\ell-\alpha-j+q} & \\ \quad + \binom{m+\alpha+j-1}{\alpha+j-q} (-1)^{\alpha+j-q+1} \log c, & q < \alpha+j \end{cases}$$

where $I_x(a,b)$ is the incomplete beta function.

For selecting the population associated with $\rho_{[1]}$, the rule proposed is R' which selects π_i iff $d R_i^{*2} \le \min_{1\le j\le k} R_j^{*2}$ where $0 < d = d(k, P^*, q, m) < 1$ is chosen so as to satisfy the basic probability requirement. The constant d is given by

$$(5.50)\qquad \int_0^{\infty} [1 - F_{2q,2m}(xd)]^{k-1} \, dF_{2q,2m}(x) = P^* .$$

Since $1 - F_{2q,2m}(xd) = F_{2m,2q}(1/xd)$, for a given set of q,m,k and P^*, the constant d of the procedure R' is the same as the constant c of the procedure R with q and m interchanged. It can be shown that the procedures R and R' have the monotonicity property.

Govindarajulu and Gore (1971) have discussed selection from bivariate normal populations in terms of their product-moment correlation coefficient. If ρ_i denotes the correlation coefficient in the population $\pi_i (i = 1,\ldots,k)$, then to select a subset containing the population with $\rho_{[k]}$, Govindarajulu and Gore have investigated the following two rules R_1 and R_2 based on the sample product moment correlation coefficients r_i and the transforms $s_i = \frac{1}{2} \log \frac{1+r_i}{1-r_i}$ $(i = 1,\ldots,k)$, respectively. R_1 selects π_i iff

$$r_i \geq \max_{1 \leq j \leq k} r_j - h \tag{5.51}$$

and R_2 selects π_i iff

$$s_i \geq \max_{1 \leq j \leq k} s_j - h \tag{5.52}$$

where $h > 0$ is chosen so as to satisfy the P*-condition. It has been shown that, for large n, h satisfies

$$P(U_i \leq h\sqrt{n/2},\ i = 1,\ldots,k-1) = P^* , \tag{5.53}$$

where the U_i have a multivariate normal distribution with $E(U_i) = 0$, $V(U_i) = 1$, $E(U_i U_j) = 1/2$, $i \neq j$. If we are interested in ranking $|\rho_i|$, then the procedure suggested is to select π_i iff $|r_i| \geq \max |r_j| - h$, where large sample solution of h is given by (5.53). It is to be noted that ranking in terms of $|\rho_i|$ is really a special case of ranking in terms of multiple correlation coefficient investigated by Gupta and Panchapakesan (1969a).

6. Distribution-Free Procedures.

In this section we discuss a non-parametric procedure for selection in terms of quantiles of a given order based on order statistics and some procedures based on ranks and paired comparisons.

(a) Selection in terms of quantiles.

Suppose π_i $(i=1,\ldots,k)$ is a continuous population with distribution function F_i whose form is not known. It is assumed $x_\alpha(F_i)$ is the unique α-quantile of the distribution F_i. Let $F_{[i]}$ denote the distribution with the its smallest α-quantile. The problem of selecting a subset containing the population with the largest α-quantile has been studied by Rizvi and Sobel (1967). Their formulation of the problem requires the P*-condition to be met for the set Ω_1 of all k-tuples $(F_1,\ldots,F_k)$ for which $F_{[k]}$ is stochastically larger than any other population.

For $0 < \alpha < 1$, we take n sufficiently large so that $1 \leq (n+1)\alpha \leq n$ and define a positive integer r by the inequalities $r \leq (n+1)\alpha < r+1$. Then the procedure $R_1 = R_1(c)$ proposed by Rizvi and Sobel is defined in terms of a positive integer $c(1 \leq c \leq r-1)$ and the order statistics $Y_{j,i}$ where $Y_{j,i}$ denctes the jth order statistic from the population F_i based on n independent observations.

R_1: Select F_i iff

$$Y_{r,i} \geq \max_{1\leq j\leq k} Y_{r-c,j} \tag{6.1}$$

where c is the smallest integer with $1 \leq c \leq r-1$ for which $\inf_{\Omega_1} P\{CS|R_1\} \geq P^*$.

For any α and k, it may happen that a value of $c \leq r-1$ does not exist for some pairs (n,P^*). However, if $P^* < P_1 = \binom{n}{r} \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} / \binom{n(i+1)}{r}$, then a value of $c \leq r-1$ exists and is unique. The value of c has to satisfy

$$(6.2) \qquad \int_0^1 G_{r-c}^{k-1}(u)\, dG_r(u) \geq P^* ,$$

where $G_r(u) = I_u(r, n-r+1)$ is the standard incomplete beta function.

It has also been shown that $E\{S|R_1\}$ is maximized in Ω_1 when the populations are identical. Further, we let P_Δ denote the configuration with $\theta_{[k]} - \theta_{[i]} = \Delta (i=1,\ldots,k-1)$ under the assumption that $F_{[i]}(x) = F(x-\theta_{[i]})$. Let $n_1(\epsilon)$ be the approximate sample size (obtained by using asymptotic theory of quantiles) required to satisfy

$$(6.3) \qquad E\{S|R_1, P_\Delta\} \leq 1 + \epsilon .$$

Similarly $n_2(\epsilon)$ denotes the sample size required to satisfy (6.3) when we use the procedure R_2 based on sample means $\bar{x}_i (i=1,\ldots,k)$, which selects the population corresponding to $\bar{x}_i$ iff $\bar{x}_i > \max_{1\leq j\leq k} \bar{x}_j - \delta$ where $\delta > 0$ is chosen to satisfy the P*-condition. Then the asymptotic relative efficiency of R_1 relative to R_2 is defined by

$$(6.4) \qquad ARE(R_1,R_2) = \lim_{\epsilon\to 0} [n_2(\epsilon)/n_1(\epsilon)] .$$

For $\alpha = \frac{1}{2}$ and normal shift alternatives with $\sigma = 1$, $ARE(R_1,R_2) = 2/\pi$. Again, for $\alpha = \frac{1}{2}$ and two-sided exponential shift alternatives with continuous symmetric densities about the median value θ_i, $ARE(R_1,R_2) = 2$.

Desu and Sobel (1971) have discussed non-parametric procedures for quantile selection under a modified goal of selecting a fixed-size subset which is described elsewhere in this paper. Barlow and Gupta (1969) investigated the quantile selection in certain restricted class of distributions and this is also discussed elsewhere.

(b) Paired comparisons procedures.

In the paired comparison approach, we compare all the $k(k-1)/2$ possible pairs of the populations $\pi_1,\ldots,\pi_k$ and we have n replications of each comparison. For $i, j=1,\ldots,k;\ i\neq j$ and $\gamma = 1,\ldots,n$, let

$$x_{ij\gamma} = \begin{cases} 1 & \text{if } \pi_i \to \pi_j \\ 0 & \text{if } \pi_j \to \pi_i \end{cases} \tag{6.5}$$

where $\pi_i \to \pi_j$ means that π_i is preferred to π_j.

It is assumed that the ties are not possible. Let

$$P\{X_{ij\gamma} = 1\} = \phi_{ij} \text{ and } P\{X_{ij\gamma} = 0\} = \phi_{ji} = 1 - \phi_{ij}\ . \tag{6.6}$$

The score a_i of the population π_i is defined by

$$a_i = \sum_{\gamma=1}^{n} a_{i\gamma} = \sum_{\gamma=1}^{n} \sum_{j\neq i} X_{ij\gamma} \tag{6.7}$$

where $a_{i\gamma}$ denotes the (partial) score of π_i in the γth replication. It is easy to see that $\sum_{i=1}^{k} a_{i\gamma} = k(k-1)/2$ and $\sum_{i=1}^{k} a_i = nk(k-1)/2$.

It is assumed that the preference probabilities ϕ_{ij} satisfy a linear model. To be specific, let θ_i be the true "merit" of π_i when judged on some characteristic. Let $y_i (i=1,\ldots,k)$ be the observed merit of π_i on which the comparisons are based. Suppose that $\pi_i \to \pi_j$ if $y_i > y_j$ and $\pi_j \to \pi_i$ otherwise. Then the preference probabilities ϕ_{ij} are said to satisfy a linear model if $\phi_{ij} = P\{y_i - y_j > 0\}$ for all i and j can be expressed as $H(\theta_i - \theta_j)$, where $H(x)$ is a distribution function on the real line with $H(-x) = 1 - H(x)$.

Under the above linear model, Trawinski and David (1963) proposed the following rule R based on the score a_i for selecting a subset containing the population with the largest θ_i.

$$R: \text{ Select } \pi_i \text{ iff } a_i \geq \max_{1 \leq j \leq k} a_j - \nu ,$$

where $\nu = \nu(k,n,P^*)$ is a non-negative integer to be chosen so as to satisfy the P*-condition. Under the linear model, it has been shown that the least favorable configuration is given by $\phi_{ij} = 1/2$ for all i and $j (i \neq j)$ and is denoted by $C(1/2)$. Thus ν is the smallest integer for which

$$P\{CS|R_1: \ C(1/2)\} \geq P^* . \tag{6.8}$$

Trawinski (1969) obtains an approximation for $E\{S|R\}$ in terms of $(k-1)$ variate normal distributions and transforms these into more numerically tractable integrals. His approximation is obtained under a slippage configuration which is specified by

$$\phi_{ij} = 1/2 \quad \text{for } i, j = 1,\ldots,k-1; \ i \neq j , \tag{6.9}$$

$$\phi_{ki} = \phi \quad \text{for } i = 1,\ldots,k-1 .$$

and is valid whenever $\phi < \frac{1}{2} + \frac{1}{2}\{k/(k+1)\}^{1/2}$.

(c) Procedures based on ranks.

Let X_{ij}, $j = 1,\ldots,n_i$, be independent observations from population π_i $(i = 1,\ldots,k)$ whose associated distribution function is $F_{\lambda_i}(x)$. The functional forms of F_λ is not known but it is assumed that $\{F_\lambda\}$ is a stochastically increasing family. All the observations are pooled and R_{ij} denotes the rank of X_{ij} in the combined sample of $N = n_1 + \ldots + n_k$

observations. Let $Z(1) \leq Z(2) \leq \ldots \leq Z(N)$ denote an ordered sample of size N from a continuous distribution G such that $-\infty < a(r) \equiv E_G(Z(r)) < \infty$ $(r = 1,\ldots,N)$. With each of the observations X_{ij} associate the number $a(R_{ij})$ and define

$$H_i = n_i^{-1} \sum_{j=1}^{n_i} a(R_{ij}), \quad i = 1,\ldots,k . \tag{6.10}$$

Using the quantities H_i, Gupta and McDonald (1970) defined the following three classes of procedures for selecting a subset containing the population with the largest θ_i:

$$\begin{aligned} &R_1(G): \text{ Select } \pi_i \text{ iff } H_i + d \geq \max(H_1,\ldots,H_k),\ d \geq 0 \\ (6.11)\quad &R_2(G): \text{ Select } \pi_i \text{ iff } cH_i \geq \max(H_1,\ldots,H_k),\ c \geq 1 \\ &R_3(G): \text{ Select } \pi_i \text{ iff } H_i \geq D, \quad -\infty < D < \infty . \end{aligned}$$

All the three classes of rules are equivalent if $R = 2$. The following theorem is established regarding the infimum of the probability of a correct selection.

Theorem 6.1. For the procedures $R_1(G)$, $R_2(G)$ and $R_3(G)$,

$$\inf_{\Omega} P\{CS|R_i(G)\} = \inf_{\Omega_k} P\{CS|R_i(G)\},\ i = 1,2,3 \tag{6.12}$$

where Ω is the space of all configurations of $\theta = (\theta_1,\ldots,\theta_k)$ and $\Omega_k = \{\theta \in \Omega: \theta_{[k-1]} = \theta_{[k]}\}$. Further, for $R_3(G)$,

$$\inf_{\Omega} P\{CS|R_3(G)\} = \inf_{\Omega_0} P\{CS|R_3(G)\} , \tag{6.13}$$

where $\Omega_0 = \{\theta : \theta_{[1]} = \ldots = \theta_{[k]}\}$.

It should be noted that a result of the type (6.13) is not true in general for $R_1(G)$ and $R_2(G)$. The procedures $R_1(G)$ (and their randomized analogs) have been suggested by Bartlett and Govindarajulu (1968) for continuous distributions differing by a location parameter. The procedures of the type $R_2(G)$ have been proposed by Blumenthal and Patterson (1969). For all these procedures a result of the type (6.13) is not true in general. Rizvi and Woodworth (1970) have given counterexamples to show that the least favorable configuration is not always given by the identical distributions case.

In the cases of $R_1(G)$ and $R_2(G)$, Gupta and McDonald (1970) have obtained bounds on the probability of a correct selection. It has been shown that

$$\inf_{\Omega} P\{H_{(k)} \geq v\} \leq \inf_{\Omega} P\{CS|R_1(G)\} \leq \inf_{\Omega} P\{H_{(k)} \geq u\} \tag{6.14}$$

and

$$\inf_{\Omega} P\{H_{(k)} \geq v'\} \leq \inf_{\Omega} P\{CS|R_2(G)\} \leq \inf_{\Omega} P\{H_{(k)} \geq u'\} , \tag{6.15}$$

where $H_{(k)}$ is the statistic H_i associated with the distribution $F_{\theta_{[k]}}$ and, u' and v' are given by

$$u' = u'(d,k,n) = n^{-1} A[1 + c(k-1)^{-1}] \tag{6.16}$$

and

$$v' = v'(d,k,n) = (nc)^{-1} \sum_{r=N-n+1}^{N} a(r) , \tag{6.17}$$

where $$A = \sum_{r=1}^{N} a(r).$$

For the particular case where $a(r) = r$, $nH_i = T_i$, where the T_i are the rank-sum statistics. In this case we denote $R_i(G)$ by R_i. For this special case, we obtain

$$\inf_{\Omega} P\{CS|R_1\} \geq P\{U \leq nd\} , \tag{6.18}$$

where U is the Mann-Whitney statistic associated with samples of sized n and $(k-1)n$ taken from two identically distributed populations. A similar result is true for R_2.

As regards R_3, we observe that R_3 may not always select a non-empty subset. A sufficient condition for selection of a non-empty subset is that P^* be sufficiently large so that $D \leq A/N$. For large n, this sufficient condition holds if $P^* > \frac{1}{2}$. The constant $D = D(k,n,P^*)$ for the rule R_3 is found such that

$$P\{U \leq n^2(k - \tfrac{1}{2}) - n(D - \tfrac{1}{2})\} \geq P^* . \tag{6.19}$$

Asymptotic expressions were obtained for $E(S|R_1)$ and $E(S|R_3)$. Assuming $n_i = n$, for large n, the distribution of $T' = (T_1,\ldots,T_k)$ is approximately multivariate normal with mean vector $\mu_T' = (\mu_1,\ldots,\mu_k)$ and variance-covariance matrix $\sum_T$. Let A be a $(k-1) \times k$ matrix given by

$$A = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & -1 \\ 0 & 1 & 0 & \ldots & 0 & -1 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{pmatrix} . \tag{6.20}$$

Define $W^{\nu} = A_{\nu}T$, where A_{ν} is the $(k-1) \times k$ matrix obtained from matrix A by moving column j to column $j+1$, $j = \nu, \nu+1,\ldots,k-1$ and replacing column ν by column k. Let $\underline{\mu}_{\nu} = A_{\nu}\underline{\mu}_T$ and $\sum_{\nu} = A_{\nu}\sum_T A_{\nu}'$. Then we have the following theorem.

Theorem 6.2. If Σ_ν is non-singular for $\nu = 1,\ldots,k$; then

$$(6.21)\qquad E\{S|R_1\} = \sum_{\nu=1}^{k} K_\nu \int_{-\infty}^{d} \cdots \int_{-\infty}^{d} \exp\left[-(\underline{W}^\nu-\underline{\mu}_\nu)' \Sigma_\nu^{-1} (\underline{W}^\nu-\underline{\mu}_\nu)/2\right] \prod_{\substack{i=1\\ 1\neq\nu}}^{k} dW_i^\nu ,$$

where $K_\nu = [(2\pi)]^{k-1}|\Sigma_\nu|]^{-\frac{1}{2}}$. For R_3,

$$(6.22)\qquad E\{S|R_3\} = \sum_{\nu=1}^{k} \Phi[(\mu_\nu-D)/\sigma_\nu] .$$

Let π_1 and π_2 be two normal populations with means 0 and $\theta(\geq 0)$ respectively and a common unit variance. The asymptotic relative efficiency of R_1 (which is equivalent to R_2 and R_3 in the case of two populations) relative the rule R based on sample means (see Section 2) is given by

$$(6.23)\qquad ARE\ (R_1,R;\theta) = \{[2\Phi(2^{-1/2}\theta)- 1]/2\theta B(\theta)\}^2$$

where

$$(6.24)\qquad B^2(\theta) = \int_{-\infty}^{\infty} \Phi^2(x+\theta)\, \phi(x)\, dx - \Phi^2(2^{-1/2}\theta) .$$

We see that $\lim_{\theta\downarrow 0} ARE\ (R_1,R;\theta) = 3/\pi$.

In the case of two exponential distributions $F_{\theta_i}(x) = 1 - e^{-x/\theta_i}(x > 0)$, where $\theta_1 = 1$ and $\theta_2 = \theta \geq 1$, a similar comparison of R_2 and the rule R' by Gupta (1963) for gamma populations yields

$$(6.25)\qquad ARE\ (R_2,R';\theta) = [(\theta-1)/4(\theta+1)\ B_1(\theta)\ \log\theta]^2,$$

where

$$(6.26)\qquad B_1^2(\theta) = 1-2(1+\theta)^{-1} + (2\theta+1)^{-1} + \theta(2+\theta)^{-1} -2\theta^2\ (1+\theta)^{-2}.$$

In this case $\lim_{\theta\downarrow 1} ARE\ (R_2,R';\theta) = 3/4.$

Some exact evaluations of the procedures R_1, R_2 and two other proceduros were made in the case of three independent exponential populations by McDonald (1969a). Procedures similar to R_1, R_2 and R_3 were studied by McDonald (1969b) by taking $T_i = \sum_{j=1}^{n} R_{ij}$ where R_{ij} is the rank of X_{ij} among $X_{1j}, X_{2j},\ldots,X_{kj}$. The results for the probability of a correct selection are very similar to those discussed above. In another paper McDonald (1971) has discussed some methods of approximating the constants required to implement the procedures R_1 and R_2.

(d) Selection in terms of measures of association.

Let $F_i(x,y)$ denote the continuous distribution function of $\pi_i (i = 1,\ldots,k)$, a set of k bivariate populations and τ_i denote the rank correlation coefficient for population π_i. Let $(X_{i,j},Y_{i,j})$, $j = 1,\ldots,n$ and $i = 1,\ldots,k$ be n independent observations from each of these populations. The rank R_{ij} of Y_{ij} is the rank of its associated X value among $X_{i1},\ldots,X_{in}$. The sample rank-correlation coefficient is given by

$$(6.27)\quad T_i = \binom{n}{2}^{-1} \sum_{j<j'}^{n} \sum^{n} \operatorname{sign}(R_{ij}-R_{ij'}),\ i = 1,\ldots,k\ .$$

For selecting a subset containing the population with the largest τ, Govindarajulu and Gore (1971) proposed the following rule R.

R: Select π_i iff

$$(6.28)\quad T_i \geq \max_{1\leq j\leq k} T_j - h\ .$$

Using the normality of the T_i and assuming a knowledge of the structure of X_{ij} and Y_{ij} (which implies the same sign for the correlation between any two X's) they have obtained a lower bound on $P\{CS|R\}$ which is used to obtain a suitable value of h. In the absence of any information on the structure of X_{ij}

and Y_{ij}, an approximate value of h is found by using certain consistent estimators of the mean and the variance of the asymptotic distribution of T_i.

For sufficiently small ρ_i the asymptotic efficiency of the procedure R relative to the procedure P_1 defined by (5.51) based on product moment correlation coefficient is found to be $9/\pi^2$ when the underlying populations are bivariate normal. For the p-variate case ($p > 2$) some suitable measures of association have been discussed by Govindarajulu and Gore.

7. Sequential Procedures

Barron and Gupta (1970) investigated a non-eliminating sequential rule, for selecting from k independent normal populations with unknown means $\theta_1,\ldots,\theta_k$ respectively and a common known variance σ^2, a subset containing the population with the largest θ_i. The rule is non-eliminating in the sense that, though the rule selects and rejects populations at various stages, observations are taken from all the populations until the final decision is made. The ordered θ_i are denoted by $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ and it is assumed that the successive differences between the ordered θ_i are known. To select a subset containing the population with $\theta_{[k]}$, the procedure investigated by Barron and Gupta is described below.

We take one observation from each population denoted by $x_1,x_2,\ldots,x_k$. For each population π_i define

$$(7.1) \qquad Y_{i1} = \begin{cases} 1 & \text{if } x_i \geq x_{max} - d\sigma \\ 0 & \text{otherwise,} \end{cases}$$

where $x_{max} = \max(x_1,\ldots,x_k)$ and d is given by

$$(7.2) \qquad \int_{-\infty}^{\infty} \phi^{k-1}(x+d)\, d\phi(x) = P^* .$$

Then we draw a second set of one observation from each population and define $Y_{i2}(i=1,\ldots,k)$ similar to Y_{i1}. Continuing in this manner, after the mth set of observations are drawn, we have Y_{im}, $i=1,\ldots,k$. For each population π_i, we define

$$(7.3) \qquad S_{im} = \sum_{j=1}^{m} Y_{ij} .$$

We have a pair of sequences of real numbers $\eta \equiv \eta_{b,c} = \{\{b_m\}, \{c_m\}\}$ such that for all $m \geq 1$,

(i) $b_m \leq b_{m+1}$, $c_m \leq c_{m+1}$

(ii) $b_m < c_m$

(iii) $\lim_{m\to\infty} b_m = \infty$

(iv) $P\{\bigcap_{m=1}^{\infty} [b_m < S_{im} < c_m]\} = 0$ for all $i=1,\ldots,k$.

The sequential selection procedure is now defined.

ψ : Tag population π_i, $i=1,\ldots,k$, at the first stage $m \geq 1$ such that $S_{im} \notin (a_m, b_m)$ and mark it "rejected" if $S_{im} \leq a_m$ and "accepted" if $S_{im} \geq b_m$. Continue sampling from all k populations until each has been tagged; then accept those marked "accepted" and reject those marked "rejected".

The following observations are made at the outset. For any m, $P\{Y_{im} = 1\} = p_i$ and $P\{Y_{im} = 0\} = 1-p_i$ where

$$p_i = \int_{-\infty}^{\infty} \Big[\prod_{\substack{j=1 \\ j\neq i}}^{k} \Phi(x+d+(\theta_{[i]}-\theta_{[j]})/\sigma)\Big] d\Phi(x), \quad i=1,\ldots,k. \tag{7.5}$$

Also $Y_{i1}, Y_{i2},\ldots,Y_{im}$ are independent and S_{im} is distributed as a binomial random variable with parameters m and p_i. Let $\pi_{(r)}$ denote the population with mean $\theta_{[r]}$. Define

$$a_i(m) \equiv a_i(m,\eta_{b,c}) = P\{\text{accepting } \pi_{(i)} \text{ at stage } m \mid \psi(\eta_{b,c})\},$$

$$r_i(m) \equiv r_i(m,\eta_{b,c}) = P\{\text{rejecting } \pi_{(i)} \text{ at stage } m \mid \psi(\eta_{b,c})\},$$

$$a_i(\eta_{b,c}) = \sum_{m=1}^{\infty} a_i(m) \quad \text{and} \quad r_i(\eta_{b,c}) = \sum_{m=1}^{\infty} r_i(m),$$

where $\mathscr{P}(\eta_{b,c})$ is the procedure using the pair of sequences $\eta_{b,c}$. When there is no ambiguity, $\mathscr{P}(\eta)$ is used for $\mathscr{P}(\eta_{b,c})$.

Definition 7.1. Let $\eta=(\{b_m\}, \{c_m\})$ and $\eta'=(\{b'_m\},\{c'_m\})$ be two pairs of sequences satisfying (7.4). The sequences $\{b_m\}$ and $\{b'_m\}$ are said to be pairwise ordered iff $b_m \leq b'_m$ for all $m \geq 1$. This relation is denoted by $\{b_m\} \prec \{b'_m\}$.

Definition 7.2. The pair η is ordered w.r.t. η' (denoted by $\eta \prec \eta'$) iff $\{b_m\} \prec \{b'_m\}$ and $\{c_m\} \prec \{c'_m\}$.

Definition 7.3. A class of pairs of sequences satisfying (7.4) is said to be ordered if for all $\eta,\eta' \in$ either $\eta \prec \eta'$ or $\eta' \prec \eta$.

The following two theorems have been established by Barron and Gupta.

Theorem 7.3. If $\eta' \prec \eta$ then $a_i(\eta') \leq a_i(\eta)$ and $r_i(\eta') \leq r_i(\eta)$, $i=1,2,\ldots,k$. In particular $P\{CS|\mathscr{P}(\eta')\} \geq P\{CS|\mathscr{P}(\eta)\}$.

Theorem 7.2. The procedure $\mathscr{P}(\eta)$ is monotone and unbiased, i.e., $a_k \geq a_{k-1} \geq \ldots \geq a_1$ and $r_k \leq r_i$, $i=1,2,\ldots,k-1$.

The rest of the investigation of the procedure $\mathscr{P}(\eta)$ has been accomplished by using the following class C_1 of pairs of sequences. Let $b_m = \delta m - \gamma_1$, $c_m = \delta m + \gamma_2$ where δ is a rational number in $(0,1)$ and γ_1,γ_2 are positive integers. For γ_1,γ_2 fixed, the class C_1 is ordered in δ. For this class it is shown that condition (iv) of (7.4) holds. If we set $R_{im} = S_{im} - \delta m$, for any $\eta \in C_1$, the events $[\delta m-\gamma_1 < S_m < \delta m+\gamma_2]$, $[S_m \geq \delta m+\gamma_2]$ and $[S_m \leq \delta m-\gamma_1]$ are equivalent to $[-\gamma_1 < R_m < \gamma_2]$, $[R_m \geq \gamma_2]$ and $[R_m \leq -\gamma_1]$ respectively. By taking $\delta=t/s$ where t and s are relatively prime integers with $t < s$, the problem of evaluating the various probabilities and expectations is reduced to a problem concerning a random walk on the line where the state space is all points of the form $(Ns-Mt)/s$ for all integers $M > N > 0$. It is now possible to relate it to

a random walk on the space of integers. These probabilities and expectations are not always easy to compute and hence some approximations and bounds were obtained. We summarize the results below.

Theorem 7.3. For the sequential procedure $\mathscr{S}(\eta)$ where $\eta = (\{\delta m-\gamma\}, \{\delta m+\gamma\})$ and $\delta = t/s > 0$

$$\lim_{\gamma \to \infty} a_i(\delta,\gamma) = \begin{cases} 0 & \text{if } p_i < t/s \\ \frac{1}{2} & \text{if } p_i = t/s \\ 1 & \text{if } p_i > t/s \end{cases} \tag{7.6}$$

where p_i is given by (7.5).

Theorem 7.4. Let m_i = the smallest $m \geq 1$ such that $\pi_{(i)}$ is accepted or rejected and $M_i = E\{m_i \in \mathscr{S}(\eta)\}$. Then, for the sequential procedure $\mathscr{S}(\eta)$ specified in Theorem 7.3,

$$M_i \doteq \gamma / |p_i - t/s| \tag{7.7}$$

provided γ is sufficiently large and $p_i \neq t/s$.

Numerical evaluations made for $\delta = .75$, $\gamma = 3(1)10$ and $p_i = .4, .6, .8, .9$ indicate that the approximations are good for all the γ values chosen. The approximation in the case of the probability of selecting the populations using the procedure improves as γ increases.

There still remains the problem of choosing the two constants δ and γ. Theorem 7.3 guarantees that for any choice of $\delta \in (p_{k-1}, p_k)$, there exists a $\gamma = \gamma(\delta,\varepsilon)$ such that for any $\varepsilon > 0$,

$$\begin{cases} \text{(i)} \quad a_k(\delta,\gamma) \geq 1-\varepsilon \quad \text{and} \\ \text{(ii)} \quad a_{k-1}(\delta,\gamma) \leq \varepsilon, \end{cases} \tag{7.8}$$

regardless of the configuration of $p_1 \leq p_2 \leq \ldots \leq p_k$ and hence the configuration of $\theta_{[1]} \leq \theta_{[2]} \leq \ldots \leq \theta_{[k]}$. Thus for a sufficiently small ϵ, the P*-condition can always be satisfied by choosing an appropriate $\eta \in C_1$. If we define S to be the size of the selected subset when the procedure terminates then $E(S) = \sum_{i=1}^{k} a_i \leq 1 + (k-1)\, a_{k-1}$. Then we can replace (7.8) by

$$(7.9) \qquad \begin{cases} \text{(i)} & a_k(\delta,\gamma) \geq 1-\epsilon \quad \text{and} \\ \text{(ii)} & 1-\epsilon < E(S) \leq 1+(k-1)\epsilon \end{cases}$$

regardless of the configuration of the means $\theta_1, \theta_2, \ldots, \theta_k$. The experimenter has for any $\delta \in (p_{k-1}, p_k)$ a countably infinite number of procedures η which guarantee (7.9). Given two procedures $\eta, \eta' \in C_1$ which satisfy (7.9), the procedure with the smaller expected number of stages is preferable in some sense. If $M = \max_{1<i<k} M_i$, then the experimentor will want to use a minimax rule, namely, an η which minimizes M over the subclass $C_2 \subset C_1$ of procedures satisfying (7.9). The following theorem has been established using approximate value of M.

Theorem 7.5. For $\delta \in (p_{k-1}, p_k)$,

$$(7.10) \qquad \min_{\delta} M = \begin{cases} \min\limits_{\delta^* \leq \delta \leq \overline{\delta}} \dfrac{\gamma_1(\delta)}{\delta - p_{k-1}}, & \text{for } \delta^* \leq \overline{\delta} \\ \min\limits_{\overline{\delta} \leq \delta \leq \delta^*} \dfrac{\gamma_2(\delta)}{p_k - \delta}, & \text{for } \overline{\delta} \leq \delta^*, \end{cases}$$

where $\gamma_1(\delta)$ is the first positive integer such that $a_k \geq 1-\epsilon$, $\gamma_2(\delta)$ is the first positive integer such that $a_{k-1} \leq \epsilon$, δ^* is the value of δ such that

$$\gamma_1(\delta) = \gamma_2(\delta) \quad \text{and} \quad \bar{\delta} = (p_k + p_{k-1})/2 \ .$$

A lemma shows that the approximate unique value δ^* is given by

$$(7.11) \qquad \delta^* = \begin{cases} \dfrac{\log[(1-p_{k-1})/(1-p_k)]}{\log[p_k(1-p_{k-1})/p_{k-1}(1-p_k)]} \ , & \text{if } p_{k-1} + p_k \neq 1 \\ 1/2 \ , & \text{if } p_{k-1} + p_k = 1 \ . \end{cases}$$

However, there still remains the problem of choosing a specific δ if $\delta^* \neq \bar{\delta}$. It has been found empirically by Barron (1968) that often $\delta^* \approx \bar{\delta}$, so that the experimenter will not be "far" from the minimum for any choice of δ between $\bar{\delta}$ and δ^*. Numerical evidence indicates that if $\bar{\delta}$ and δ^* are significantly apart, the minimum takes place near δ^*. It seems an approximate minimax rule which has certain desirable properties would be $\mathscr{P}(\eta^*)$ where $\eta^* = (\{\delta^* m - \gamma^*\}, \{\delta^* m + \gamma^*\})$.

Some sample size comparisons have been made numerically between the procedure $\mathscr{P}(\eta^*)$ and the fixed sample-size procedure of Gupta (1965) based on means of samples of size n from the k population, which is denoted here by $R(n)$ and defined below.

$$R(n): \quad \text{Select } \pi_i \text{ iff } \bar{x}_i \geq x_{max} - \frac{d\sigma}{\sqrt{n}}$$

where d is given by (7.2).

The comparison was made with $\sigma = 1$ under slippage configuration $\theta_{[1]} = \ldots + \theta_{[k-1]} = \theta$, $\theta_{[k]} = \theta + \tau$, $\tau > 0$, and the equally-spaced configuration $\theta_{[1]} = \theta$, $\theta_{[2]} = \theta + \tau, \ldots, \theta_{[k]} = \theta + (k-1)\tau$, $\tau > 0$. The following ranges of the values of k, τ and P^* were considered:

(i) Slippage configuration: $k = 2(1)10, 25, 50$; $\tau = 0.05, 0{\cdot}10(\cdot 10)0{\cdot}60, 1,2$; $P^* = {\cdot}75, {\cdot}90$.

(ii) Equally-Spaced Configuration: $k = 2(1)5$; $\tau = 0.05, 0{\cdot}10(\cdot10)0{\cdot}60$; $P^* = \cdot75, \cdot90$.

The empirical results indicate that $\mathcal{A}(n^*)$ is preferable when the means are close and $R(n)$ is better when any one mean gets significantly larger than the others.

Guttman (1963) considers a sequential procedure for a goal which is different from the usual one. Suppose that $\Pi_i (i=1,\ldots,k)$ has the density $f_{\theta_i}(x)$ and the quality of the population is characterized by $h_i = g(\theta_i)$ where g is a known function. Let T_i be an appropriate statistic based on a sample of n independent observations in the sense that $E(T)$ is $g(\theta)$ or a monotonic function of $g(\theta)$. Consider the rule R which selects Π_i iff

$$T_i \in \omega_{n,k}(P^*, \underline{T}) \tag{7.12}$$

where $\omega_{n,k}(P^*,\underline{T})$ is a random linear set contained in the sample space of T_i and depends on $\underline{T} = (T_1,\ldots, T_k)$ and is such that $\inf_{\Omega} P\{CS|R\} = P^*$.

Since the size of the selected subset is random, a natural question is how to proceed sequentially so that we could select one population as the best or reduce the size of the subset selected subject to certain cost considerations which restrict the number of stages.

Let t denote the stage of the experiment and k_t denote the number of populations retained at the start of the stage. If M units of capital are available to spend on the procedure and at each stage a sample of n_t independent observations are taken from each population, let t_o be the largest integer for which $\sum_{i=1}^{t_o} k_t n_t d \leq M$ where d is the cost per observation.

The sequential procedure proposed and investigated by Guttman (1963) is defined below.

R': At each stage t, use the rule R with $P^* = P_t^*$ where

$P_t^* = 1 - \frac{1-\beta}{2^t}$ adopting the following stopping rule:

At the end of stage t,

(1) Stop if $t = t_o$.

(2) Stop if $t < t_o$ and $k_{t+1} = 1$

(3) Continue if $t < t_o$ and $k_{t+1} > 1$.

It has been shown that $P\{CS|R'\} \geq \beta$. Suppose that there is infinite capital. We say that the rule R' is in state γ if, at any stage t, we have $k_t = \gamma$. The states form a Markov chain with non-stationary transition probabilities

$$(7.13) \qquad p_{\gamma\alpha} = P\{k_{t+1} = \alpha | k_t = \gamma\}, \quad 1 \leq \alpha \leq \gamma = k_t \leq k \ .$$

These are dependent on $\omega_{n_t,\gamma}(P_t^*, \underline{T})$. We note that $p_{\gamma\alpha} = 0$ if $\gamma < \alpha$ and $\sum_{\alpha=1}^{\gamma} p_{\gamma\alpha} = 1$. The following theorem has been established by Guttman (1963).

Theorem 7.6. Consider the Markov chain with the above structure. Let $p_{\alpha\alpha}(t) = 1 - \delta_\alpha(t)$, $0 < \delta_\alpha(t) < 1$ for $\alpha \neq 1$. Then the Markov chain is absorbed at state 1 (i.e., R' terminates at a finite stage) iff $\sum_{t=1}^{\infty} \delta_\alpha(t)$ diverges for all $\alpha \neq 1$.

It might be possible to find a "reasonable" value of n_t in some special cases. Suppose that the expected subset size $E(S)$ at stage t can be written as a function of n_t, k_t, P_t^* and the differences $h_{[j]} - h_{[i]}$, $i < j$. Since k_t and P_t^* are known, if we have information about the differences of the $h_{[i]}$, we can set $E(S) = 1$ and solve for n_t.

8. Selection from Restricted Families of Distributions.

There are situations where we do not know the actual functional forms of the distributions F_i, $i = 1,\ldots,k$, associated with the populations but have some information about the class of functions to which they belong defined in terms of a partial order relation with respect to a known distribution G. Such families do occur in practical problems. In these cases the evaluation of the necessary constants for the procedures depends on the knowledge of G but not on the forms of the F_i themselves and in this restricted sense the procedures are distribution-free. Barlow and Gupta (1969) have discussed selection procedures for restricted families of distributions mainly in terms of their quantiles. We will briefly discuss here these procedures and indicate certain other related problems.

Assume that each F_i has a unique α-quantile, $\xi_{\alpha i}$. Let $F_{[i]}$ denote the cumulative distribution function (cdf) of the population with the ith smallest α-quantile. We assume that

(8.1)

(a) $F_{[i]}(x) \geq F_{[k]}(x)$, $i = 1,2,\ldots,k$ and all x,

(b) there exists a continuous distribution G such that $F_{[i]} \underset{\sim}{\prec} G$ for all $i = 1,\ldots,k$,

where $\underset{\sim}{\prec}$ denotes a partial ordering relation on the space of distributions. To be precise, $F \underset{\sim}{\prec} F$ for all F and $F \underset{\sim}{\prec} G$, $G \underset{\sim}{\prec} H \Rightarrow F \underset{\sim}{\prec} H$. Note that $F \underset{\sim}{\prec} G$ and $G \underset{\sim}{\prec} H$ do not necessarily imply $F \equiv G$.

Some special cases of partial ordering which are of interest here are:

(i) $F \underset{*}{\prec} G$ iff $F(0) = G(0) = 0$ and $G^{-1}F(x)/x$ is nondecreasing in $x \geq 0$ on the support of F.

(ii) $F \underset{c}{\prec} G$ iff $G^{-1}F(x)$ is convex on the support of F.

(iii) $F \underset{r}{\lesssim} G$ iff $F(0) = G(0) = \frac{1}{2}$ and $G^{-1}F(x)/x$ is increasing (decreasing) for x positive (negative) on the support of F.

If $G(x) = 1 - e^{-x}$, $x \geq 0$, then (i) defines the class of IFRA distributions studied by Birnbaum, Esary and Marshall (1966) while (ii) defines the class of IFR distributions studied by Barlow, Marshall and Proschan (1963). It is easy to see that $\underset{c}{\lesssim}$ ordering implies $\underset{*}{\lesssim}$ ordering. Implications of $\underset{r}{\lesssim}$ ordering have been studied by Lawrence (1966). Van Zwet (1964) investigated the convex ordering and s-ordering (not defined above).

(a) Quantile selection rules for distributions $\underset{*}{\lesssim}$ ordered w.r.t. G.

The distributions $F_{[i]}$ and G satisfy the assumptions in (8.1). Let $T_{j,i}$ denote the jth order statistic based on n independent observations from F_i where $j \leq (n+1)\alpha < j+1$. Then for selecting the population with the largest α-quantile, Barlow and Gupta (1969) proposed the rule

R: Select the population π_i iff

$$T_{j,i} \geq c \max_{1 \leq r \leq k} T_{j,r} , \tag{8.2}$$

where $0 < c = c(k,P^*,n,j) < 1$ is determined so as to satisfy the P*-condition.

It has been shown by Barlow and Gupta that

$$\inf_{\Omega} P\{CS|R\} = \int_0^\infty [G_j(x/c)]^{k-1} \, dG_j(x) , \tag{8.3}$$

where Ω is the space of all the k-tuples $(F_1,\ldots,F_k)$ and $G_j(x)$ is the cdf of the jth order statistic based on n independent observations from G. Thus the constant c of the procedure is determined by

$$\int_0^\infty [G_j(x/c)]^{k-1} \, dG_j(x) = P^* \tag{8.4}$$

and is tabulated by Barlow, Gupta and Panchapakesan (1969) in the case of

$G(x) = 1 - e^{-x}$, $x > 0$ for selected values of n, k, j and P^*. For $j = 1$, the constant c is easily seen to be independent of n.

We discussed earlier in Section 6 a non-parametric procedure R_1 studied by Rizvi and Sobel (1967) for the quantile selection problem. It has been shown by Barlow and Gupta that the rules R and R_1 are asymptotically equally efficient in the sense defined by (6.4) under the scale slippage configuration.

A selection rule R' proposed by Gupta (1963) for gamma populations based on the sample means has been referred to in Section 2. Comparing R and R' under the slippage configuration $\lambda_{[i]} = \delta\lambda_{[k]}$, $0 < \delta < 1$, $i = 1,\ldots,k-1$, we have

$$(8.5) \qquad A(R,R';\delta) \geq 2(1-\delta)^2 \bar{\alpha}^{-2} [-\log \bar{\alpha}]^2 / [r (\log \delta)^2 \alpha\bar{\alpha} (1 + \delta^2)] ,$$

where $\bar{\alpha} = 1 - \alpha$. Consequently we obtain

$$(8.6) \qquad A(R,R';\delta \uparrow 1) \geq 0.493 \quad \text{for } \alpha = 1/2 .$$

Barlow and Gupta (1969) also considered selection in terms of median when the distributions $F_i (i = 1,\ldots,k)$ have lighter tails than G which means that F_i centered at its median, Δ_i, is $\underset{r}{\leq}$ - ordered w.r.t. G $(G(0) = \frac{1}{2})$ and $(d/dx)\, F_i(x + \Delta_i)\big|_{x=0} \geq (d/dx)\, G(x)\big|_{x=0}$. In order to select the population with the largest median, the following rule R_2 was proposed.

R_2: Select π_i iff

$$(8.7) \qquad T_{j,i} \geq \max_{1 \leq r \leq k} T_{j,r} - D, \qquad j \leq (n+1)/2 < j+1 .$$

It was shown that the constant $D > 0$ satisfying the P^*-condition is determined by

$$(8.8) \qquad \int_{-\infty}^{\infty} G_j^{k-1} (t+D)\, dG_j(t) = P^*$$

where G_j is as defined in (8.3).

It is easy to show that, if F has a lighter tail than G, then $G^{-1}F(x)-x$ is increasing in x, which means that F is tail-ordered w.r.t. $G(F \overset{c}{<} G)$ according to a definition of Doksum (1969). As a matter of fact the rule R_2 defined by (8.7) can be used for the larger class of distributions F, which are tail-ordered w.r.t. G.

(b) Selection w.r.t. the means for IFR distributions.

Let μ_i be the mean of the distribution F_i, $i = 1,\ldots,k$, and $F_{[i]}$ denote the distribution with the ith largest mean. We assume that

(a) $F_{[i]}(x) \geq F_{[k]}(x)$ for $i = 1,\ldots,k-1$ and all x;

(b) $F_{[i]} \overset{c}{<} G$ for $i = 1,\ldots,k$

where $G(x) = 1 - e^{-x}$, $x \geq 0$. We also assume that $F_i(0) = 0$ for all i. Let $\bar{x}_i$ be the sample mean based on n independent observations from π_i and $H_i(x)$ be the cdf of $\bar{x}_i$. Let $H_{[i]}$ denote the distribution of the sample mean from $F_{[i]}$. Then

(8.9) $\quad H_{[i]}(x) \geq H_{[k]}(x)$ for $i = 1,\ldots,k-1$ and all x

and

(8.10) $\quad H_{[i]} \overset{c}{<} G$ for $i = 1,\ldots,k$.

The statement in (8.9) is an immediate consequence of the assumption (a) above, while (8.10) follows from (b) and the closure of IFR distributions under convolutions (see Barlow, Marshall and Proschan (1963)). For selecting a subset containing the population $F_{[k]}$, Barlow and Gupta (1969) proposed the rule R_3, namely,

R_3: Select the population π_i iff

(8.11) $$\bar{x}_i \geq c' \max_{1 \leq j \leq k} \bar{x}_j$$

where the constant c' $(0 < c' < 1)$ satisfying the P*-condition is given by

$$\int_0^\infty [G(x/c')]^{k-1}\, dG(x) = P^* . \tag{8.12}$$

The disadvantage of the rule R_3 is that the constant c obtained from (8.12) is independent of n. However, by restricting the class of distributions to the gamma family we can obtain a lower bound for $P\{CS|R_3\}$ which depends on n.

(c) Some results relating to partial orderings of distributions.

The two procedures R and R_1 defined by (8.2) and (8.7) for the two types of ordering provides the motivation for an attempt by Panchapakesan (1969) to unify these two by a general order relation which throws more light on a lemma of Gupta (1966b). We define the general ordering here in a slightly revised form.

Definition 8.1. Let $\mathcal{H} = \{h(x)\}$ be a class of real-valued function on the real line. Then F is said to be $\mathcal{H}$-ordered w.r.t. G if $F(0) = G(0)$ and $G^{-1}F(h(x)) \geq h(G^{-1}F(x))$ for all $h \in \mathcal{H}$.

We note that if $\mathcal{H} = \{ax, a \geq 1\}$ and $F(0) = G(0) = 0$, then we get star-ordering. If $\mathcal{H} = \{x+b, b \geq 0\}$ and $f(0) = G(0) = \frac{1}{2}$, then $\mathcal{H}$-ordering reduces to tail ordering. It has been shown that $\mathcal{H}$-ordering is a partial ordering and that order statistics preserve the ordering. The following lemma is the key result we need to bound below the probability of a correct selection.

Lemma 8.1. If $F \underset{\mathcal{H}}{\leq} G$, then, for any positive integer t,

$$\int F^t(h(x))\, dF(x) \geq \int G^t(h(x))\, dG(x) \tag{8.13}$$

for all $h \in \mathcal{H}$.

Gupta (1966b) proved the following lemma.

Lemma 8.2. X is a random variable having the distribution function $F_\lambda(x)$. Let $h_b(x)$ be a class of functions and suppose there exists a distribution function $F(x)$ such that $h_b(g_\lambda(x)) \geq g_\lambda(h_b(x))$ for all λ and all x, where $g_\lambda(x)$ is defined by $F_\lambda(g_\lambda(x)) = F(x)$ for all x. Then for any $t > 0$,

$$\int F_\lambda^t(h_b(x))\ dF_\lambda(x) \geq \int F^t(h_b(x))\ dF(x). \tag{8.14}$$

It is shown that the assumption of Lemma 8.2 amounts to saying $F_\lambda \underset{h_b}{\leq} F$. A general selection problem discussed by Panchapakesan (1969) is as follows. Let $\pi_1,\ldots,\pi_k$ be k populations and F_i is the distribution function associated with π_i. We assume that there exists one among the k populations which is stochastically larger than any other. Let us denote the distribution of that population by $F_{[k]}$. Thus we have

$$F_i(x) \geq F_{[k]}(x) \text{ for } i = 1,\ldots,k \text{ and all } x. \tag{8.15}$$

It is also assumed that there exists a continuous distribution G and a class of realvalued functions $\mathcal{H} = \{h(x)\}$ such that

$$F_i(x) \underset{\mathcal{H}}{\leq} G \text{ for } i = 1, 2,\ldots,k\ . \tag{8.16}$$

If $\underline{X}_i = (X_{i1}, X_{i2},\ldots,X_{in})$ is the observed sample from π_i, then we confine ourselves to the class of statistics $T_i = T(\underline{X}_i)$ that preserve both the ordering relations (8.15) and (8.16). Let F_{T_i} represent the cdf of $T(\underline{X}_i)$ under F_i and G_T, the cdf of $T(\underline{Y})$ under G, where $\underline{Y} = (Y_1,\ldots,Y_n)$ is a random sample from G. If $h(x) \geq x$, then for selecting a subset containing the population associated with $F_{[k]}$, the following rule R_4 was

proposed.

R_4: Select π_i iff

(8.17) $$h(T_i) \geq \max(T_1,\ldots,T_k) .$$

It has been shown that

(8.18) $$P\{CS|R_4\} = \int_{-\infty}^{\infty} G_T^{k-1}(h(x)) \; dG_T(x) .$$

If $h(x)$ is indexed by the constants c and d ($c \geq 1$, $d \geq 0$) then we can find suitable constants c and d if conditions on $h(x)$ given in the very beginning of Section 3 are satisfied.

9. Bayes and Empirical Bayes Procedures.

Let $y = (y_1, \ldots, y_k) \in E^k$ (Euclidean k-space) be an observation of the random vector $Y = (Y_1, \ldots, Y_k)$ whose components are independent random variables, Y_i having the density $f(y_i|\theta_i)$. The space of action is denoted by $\mathcal{G}$ and it consists of all non-empty subsets of k-populations (Y_i is the random variable associated with the population π_i, $i=1,\ldots,k$). A selection procedure D is a mapping from E^k to $\mathcal{G}$. The loss incurred when $\underline{\theta}' = (\theta_1, \ldots, \theta_k)$ is the true state of nature and $D(y)$ is the subset selected is denoted by $L(D(y), \underline{\theta})$. Let G_i be the a priori distributions of θ_i and $G = \prod_{i=1}^{k} G_i$ denotes the a priori distribution on the parameter space Ω. The Bayes risk of a decision procedure D w.r.t. the a priori distribution G is defined by

$$R(D,G) = \int_{\Omega} \left\{ \int_{E^k} L(D(y),\underline{\theta})\, f(y|\underline{\theta})\, dy \right\} d\, G\, (\underline{\theta}) , \tag{9.1}$$

where

$$f(y|\underline{\theta}) = \prod_{i=1}^{k} f(y_i|\theta_i) .$$

A Bayes procedure w.r.t. G is a procedure D^* for which the Bayes risk is minimum. Suppose we consider the loss function in selecting the subset S_j given by

$$L(S_j,\underline{\theta}) = \sum_{q \in S_j} \alpha_{jq}(\theta_{[k]} - \theta_q) \tag{9.2}$$

where $\alpha_{jq} \geq 0$ and the summation is over all populations q included in S_j.

Deely and Gupta (1968) investigated Bayes procedures with the above formulation.

Before stating the main results of their investigation, we adopt the following notation for the sequal.

S_j denotes the singleton consisting of π_j, $j=1,\ldots, k$. The remaining 2^k-k-1 subsets containing two or more populations will be denoted by S_j, $j=k+1,\ldots, 2^k-1$ with no explicit ordering. Further let

$$\Psi(S_j,y) = \int_{\Omega} L(S_j,\underline{\theta})\, f(y|\underline{\theta})\, dG(\underline{\theta}),\ j=1,2,\ldots, 2^k-1$$

$$a_q = \int_{\Omega} (\theta_{[k]}-\theta_q)\, f(y|\underline{\theta})\, dG(\underline{\theta}),\ q=1,\ldots, k \tag{9.3}$$

$$a_{[1]} = \min_{1\le q\le k} a_q$$

Deely and Gupta have established the following result.

Theorem 9.1. Let the loss function be given by (9.2) in which $a_{jq} = a > 0$ for $j = 1,\ldots, k$. If $\sum_{q\in S_j} a_{jq} \ge a$ for every $j = 1,2,\ldots, 2^k-1$, then the Bayes procedure w.r.t. G for selecting a subset containing the population with $\theta_{[k]}$ is given by $D^* = D^*(y) = S_j$ where j is any positive integer $1,2,\ldots, k$ such that

$$\Psi_G(S_j,y) = \min_{1\le i\le k} \Psi(S_i,y) \quad . \tag{9.4}$$

This result is applied to the normal means problem with G_i as (i) normal with mean λ_i and variance β_i^2 and (ii) uniform on $(\lambda_i-d_i, \lambda_i+d_i)$. In the first case, the Bayes procedure is:

Select π_i for which

$$(9.5) \qquad \frac{n\beta_i^2 \bar{x}_i + \lambda_i}{1+n\beta_i^2} = \max_{1\le j\le k} \frac{n\beta_j^2 \bar{x}_j + \lambda_j}{1+n\beta_j^2}$$

where $\bar{x}_i$'s are sample means based on n observation.

Some other cases like selection for binomial and Poisson populations where the parameters, respectively, have beta and gamma a priori distributions have been discussed by Deely (1965) who has also investigated empirical Bayes procedures for the selection problem which we presently discuss.

In the empirical Bayes approach, only the existence of an a priori distribution G on the parameter space is assumed and not a particular G. Thus the Bayes procedure is not available. Suppose independent observations $(\underline{x}_1^*, \underline{\theta}_1), (\underline{x}_2^*, \underline{\theta}_2), \ldots, (\underline{x}_n^*, \underline{\theta}_n)$ on a random variable $\underline{X}$ are available with $\underline{\theta}_i$'s all being drawn from the same distribution G. (The $*$ indicates that "r" observations from each population have been taken for $i = 1, \ldots, n$). The "prior observations" contain information about G and thus if a decision procedure D_n based upon $\underline{X}_1^*, \ldots, \underline{X}_n^*$ could be found such that $R(D_n, G)$ converges to $R(D_G, G)$ (i.e. the Bayes risk of D_n converges to the Bayes risk of the Bayes procedure D_G which we would use if we knew G at the start) for any G in some family $\tilde{G}$, then the procedure D_n is asymptotically optimal to D_G and is called an empirical Bayes procedure w.r.t. the unknown G. The main theorem of Deely (1965) proves that under certain regularity conditions the Bayes procedure w.r.t. an estimate G_n of G is also empirical Bayes w.r.t. G. In order to apply this theorem, a suitable estimate G_n is required. A completely satisfactory answer to this problem is not available.

Suppose we make an additional sssumption that G belongs to a parametric family $\bar{G}$ with parameter $\underline{\lambda} = (\lambda_1, \ldots, \lambda_k)$. Suppose now an estimate λ_{nj} of λ_j depending on the prior observations from the jth population can be found such that G_{nj} based on the observations converges to G_j with probability one. Then it is shown that $G_{\pi,n} = \prod_{j=1}^{k} G_{j,n}$ converges to $G = \prod_{j=1}^{k} G_j$ with probability one. Further, $G_{\pi,n}$ is also a member of $\bar{G}$. Thus, if the Bayes procedure w.r.t. any G in $\bar{G}$ is available, then in particular $G_{\pi,n}$ is available and thus an empirical Bayes procedure w.r.t. G is obtained. Empirical Bayes procedures have been obtained for several special cases of $f(x|\theta_i)$ and $\bar{G}$, namely, (i) normal-normal, (ii) normal-uniform (iii) binomial-beta, (iv) Poisson-gamma. To illustrate the type of results obtained, we consider the case of normal-normal.

Let π_i $(i=1, \ldots, k)$ have the normal density $f(x|\theta_i)$ with unknown mean θ_i and known variance σ_i^2 and let θ_i be distributed normally with unknown but finite mean λ_i and known variance β_i^2. Let $\underline{x}_1^*, \underline{x}_2^*, \ldots, \underline{x}_n^*$ be independent prior observations and $\underline{x}^*$ the present observation. Then the empirical Bayes procedure under the linear loss function in (9.2) with $\alpha_{jq} = 1, D_{G_{\pi,n}}(\underline{x}^*)$ select the population π_i for which

$$Z_i = \max_{1 \le j \le k} Z_j \tag{9.6}$$

where

$$Z_j = \frac{r\beta_j^2 \bar{x}_j + \sigma_j^2 \bar{\bar{x}}_j}{\sigma_j^2 + r\beta_j^2} \tag{9.7}$$

$\bar{x}_j$ denotes the sample mean from π_j based on present observation and $\bar{\bar{x}}_j$ is the over-all mean of the prior observations from π_j.

Similar procedures have been obtained for the case where G is subject to certain very general conditions. We briefly describe one of the results below for the sake of illustration.

Suppose $f(x|\theta_j)$ be a normal density with mean θ_j and variance σ_j^2. Let θ_j be distributed according to G_j such that $\int_{\Omega} \theta \, d\, G_j(\theta) < \infty$, $j = 1,\ldots,k$. Let $\underline{x}_1^*, \underline{x}_2^*, \ldots, \underline{x}_n^*$ be independent prior observations and $\underline{x}^*$ be the present observations. We denote the mean of the present observations from π_j by $\bar{x}_j$ and the means of the prior observations from π_j by $\bar{x}_{\alpha j}$, $\alpha = 1,\ldots, n$. Let $H_{nj}(\bar{x}_j)$ denote $(n+1)^{-1}$ times the total number of $\bar{x}_{\alpha j}$'s which are $\leq \bar{x}_j$ including the present observation $\bar{x}_j$. Define

$$(9.8) \qquad h_{nj}(\bar{x}_j) = \frac{H_{nj}(\bar{x}_j + n^{-1/5}) - H_{nj}(\bar{x}_j - n^{-1/5})}{2n^{-1/5}}, \quad j = 1,\ldots, k$$

and

$$(9.9) \qquad g_{nj}(\bar{x}_j) = \frac{h_{nj}(\bar{x}_j + n^{-1/5}) - h_{nj}(\bar{x}_j - n^{-1/5})}{2n^{-1/5}} .$$

Then the empirical Bayes procedure under linear loss function (9.2) (with $\alpha_{jq} = 1$) for selecting the best population is the procedure which selects the population π_j $(j = 1,\ldots, k)$ for which $\bar{x}_j + \frac{\sigma_j^2}{r} \frac{g_{nj}(\bar{x}_j)}{h_{nj}(\bar{x}_j)}$ is maximum. The main result used in these cases is a result due to Robbins (1964).

10. Modified Formulations and Goals

In the preceding sections we discussed the general theory of subset selection problems under the usual formulation and described several cases of specific distributions and ranking criteria used. There are, however, a few other cases which were not mentioned earlier. Barr and Rizvi (1966) considered the problem of selecting a subset containing the population with the largest θ from a set of k populations having uniform distributions over $(0, \theta_i)$, $i=1,\ldots, k$. Guttman (1961) investigated selection problems using the coverage probabiiity as the criterion of ranking. If π_i $(i=1,\ldots, k)$ is described by the sample space $(\chi,\mathcal{G},P_{\theta_i})$ where P_{θ_i} is a probability measure belonging to the class $\{P_\theta\}$, $\theta \in \Theta$, the populations are ranked according to $b_i = \int_A dP_{\theta_i}$, where the set $A \in \mathcal{G}$. Guttman has discussed specific procedures for normal and exponential distributions with $A = (-\infty,a)$ where a is known and specified in advance.

Several authors have considered formulations and goals different from the usual ones. In the remaining part of this section we will briefly describe these modifications.

(a) A generalization of subset selection goal.

Suppose that there exists a binary relation $\preceq$ which orders the populations $\pi_1,\ldots, \pi_k$ from worst to best. The ordered populations are denoted by $\pi_{(1)} \preceq \pi_{(2)} \preceq \cdots \preceq \pi_{(k)}$. This gives a unique t-subset comprising the t best populations, namely, $\{\pi_{(k-t+1)}, \pi_{(k-t+2)},\ldots, \pi_{(k)}\}$ for any $t(1 \leq t \leq k)$. The experimenter's goal is to select a subcollection of the collection of all subsets of size s from the k populations such that at least one such selected subset contains at least c of the t best populations. A correct selection is a realization of the experimenter's goal. For a given probability

P^*, a rule R_s is proposed satisfying the condition that $P(CS|R_s) \geq P^*$ no matter what the unknown configuration of the populations $\pi_1, \ldots, \pi_k$. Of course in a meaningful problem, we have constraints on the values of t, s and c, namely, $1 \leq t < k$, $1 \leq s < k$, $\max[1, s+t+1-k] \leq c \leq \min[s,t]$.

Let X_{ij}, $j = 1, \ldots, n_i$, be independent random variables denoting observations from population π_i, $i = 1, \ldots, k$. Let $T_i = T(X_{i1}, X_{i2}, \ldots, X_{in_i})$, $i = 1, \ldots, k$, be independent statistics with absolutely continuous distributions $G_{T_i} \equiv G_i$, $i = 1, \ldots, k$, suitably chosen such that $\pi_i \preceq \pi_j \Rightarrow T_i \underset{st}{\leq} T_j$, $1 \leq i, j \leq k$. Let t_i be an observed value of T_i, $i = 1, \ldots, k$. Then the rule R_s proposed and studied by Gupta and Deverman (1969) is the following.

R_s: Consider all possible s-subsets (subsets of size s) of $\pi_1, \ldots, \pi_k$. Include in the collection of s-subsets the s-subset $\{\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_s}\}$ having the observations $A = \{t_{i_1}, t_{i_2}, \ldots, t_{i_s}\}$ and complementary set of observations $A^c = \{t_{i_{s+1}}, \ldots, t_{i_k}\}$ iff $d[T_{[1]}(A), T_{[k-s]}(A^c)] \geq -d^*$, where $T_{[i]}(A)$ is the ith smallest element in any finite set of real numbers, $d(x,y)$ is a generalized difference such that (i) $d(x,y) = 0 \Leftrightarrow x=y$, (ii) for fixed $y = y_0$, $d(x,y_0)$ is increasing in x and (iii) for fixed $x = x_0$, $d(x_0,y)$ is decreasing in y, and the constant $d^* \geq 0$ is chosen so that the P^* probability condition is satisfied. For the procedure R_s, it has been shown that the infimum of $P(CS|R_s)$ occurs when all the populations are identical w.r.t. the binary relation with which they are ordered.

Gupta and Deverman have also discussed the normal means problem in particular.

(b) Selecting a subset better than a standard

Under this formulation we have $(k+1)$ populations π_i $(i=0,1,\ldots,k+1)$ with the associated distribution functions F_{θ_i}. The parameters $\theta_1, \ldots, \theta_k$

are unknown and the parameter θ_0 of the standard population may or may not be known. The goal is to select a subset containing all the populations π_i for which $\theta_i \geq \theta_0$ (or $\theta_i \leq \theta_0$). Any rule R defined for the purpose is required to satisfy the P*-condition.

The cases of location and scale parameters have been discussed by Gupta (1965). Earlier Gupta and Sobel (1958) have considered the normal means problem where the procedure based on sample means $\bar{x}_i$ (i=0,..., k) selects π_i iff $\bar{x}_i \geq \bar{x}_0 - A/\sqrt{n}$. (It is assumed that all populations have unit variance).

Puri and Puri (1968, 1969) have investigated rules based on ranks for the location and scale parameter cases and have studied the efficiency of these procedures compared to the normal theory procedures. The results and techniques of these investigations are similar to those of Lehmann (1963).

Nonparametric selection procedures for selecting populations better than a standard when the comparison is in terms of α-quantile have been discussed by Rizvi, Sobel and Woodworth (1968). The corresponding subset selection problem under the usual formulation has been investigated by Rizvi and Sobel (1967) and has been discussed in Chapter 6.

In comparing a population with a standard Lehmann (1961) considered a population to be good if it is sufficiently better than the standard. To be precise, let π_i (i=1,..., k) be a population whose quality is characterized by a real-valued parameter θ_i and a population is said to be positive (or good) if $\theta_i \geq \theta_0 + \Delta$ and negative (or bad) if $\theta_i \leq \theta_0$, where Δ is a given positive constant and θ_0 is either a given number or a parameter that may be estimated. A negative population if included in the selected subset is called a false positive, while a good population not included in the subset is

called a false negative. Roughly speaking, the aim of a selection procedure is to seek out the positive populations while holding false positives in the selected subset to a minimum.

Let $S(\theta,\delta)$ and $R(\theta,\delta)$ denote the expected number of true positives and false positives, respectively, using the procedure δ. Then the problem is to determine a procedure for which $\sup_{\theta \in \Omega} R(\theta,\delta)$ is minimum subject to the condition that $\inf_{\theta \in \Omega'} S(\theta,\delta) \geq \gamma$ where Ω denotes the whole parameter space and Ω' denotes the set of parameter-points for which at least one of the populations is positive.

Under certain conditions, Lehmann (1961) shows that a rule minimax in the above sense selects π_i when $T_i \geq c_i$, where T_i is a suitable statistic whose distribution depends only on θ_i and where c_i is a suitable constant. He has also discussed the applications of these to distributions with monotone likelihood ratio in the case where θ_0 is known and to normal distributions where observations on θ_0 are included in the experiment.

Krishnaiah and Rizvi (1966) have considered the problem of selecting multivariate normal populations better than a control on the basis of the linear combinations of the elements of the mean vectors of the populations. Different definitions of positive and negative populations have been used and in each case a selection procedure δ is proposed such that $\inf_{\omega} P(\omega,\delta) \geq P^*$ or $\inf_{\omega} S(\omega,\delta) \geq p^*$ where $P(\omega,\delta)$ denotes the probability of including all positive populations, $S(\omega,\delta)$ denotes the expected proportion of true positives and P^* and p^* are given constants. As an illustration of the type of results obtained by Krishnaiah and Rizvi, consider the set of populations $\pi_1,\ldots,\pi_k$ and the control population π_0, where π_i $(i=0,1,\ldots,k)$ is the p-variate normal distribution $N_p(\underline{\mu}_i,\Sigma_i)$. Let $\theta_{mc} = \underline{a}'_c\underline{\mu}_i$, $(c=1,\ldots,r;\ i=1,\ldots,k)$, where

$\underline{a}_1, \ldots, \underline{a}_r$ are specified vectors. The population π_i is said to be positive if $\theta_{ic} \geq \theta_{0c} + \Delta_c$, $c=1,\ldots, r$, and negative if $\theta_{ic} \leq \theta_{0c}$, $c=1,\ldots, r$, where Δ_c are given positive constants. For the case of known $\Sigma_i (i=0,1,\ldots,k)$, the rule δ proposed selects π_i iff

$$(10.1) \quad \underline{a}_c'(\bar{\underline{x}}_i - \bar{\underline{x}}_0)/[\underline{a}_c'(n_i^{-1}\Sigma_i + n_0^{-1}\Sigma_0)\underline{a}_c]^{1/2} \geq d, \quad c=1,\ldots, r,$$

where $\bar{\underline{x}}_i$ is the sample mean vector from π_i based on n_i observations.

Krishnaiah (1967) investigated similar procedures when the comparison of the multivariate normal populations with the control population is based on linear combinations of elements of the covariance matrices, determinants of the covariance matrices and the largest (smallest) characteristic roots.

Desu (1970) considered the selection problem where the populations are not compared with a standard but rather with the best among them. If $d(\theta_i, \theta_j)$ is a distance measure between θ_i and θ_j and if $\theta_{max} = \max(\theta_1,\ldots,\theta_k)$, population π_i is said to be superior (or good) if $d(\theta_{max}, \theta_i) \leq \delta_1^*$ and inferior (or bad) if $d(\theta_{max}, \theta_i) \geq \delta_2^*$, where δ_1^*, δ_2^* are specified constants such that $0 < \delta_1^* < \delta_2^*$. For the location and scale parameter cases which have been considered, $d(\theta_i, \theta_j)$ is taken to be $\theta_i - \theta_j$ and θ_i/θ_j respectively. The proposed procedure R selects π_i iff $d(Y_{max}, Y_i) \leq d(\delta_2^*, c)$ where Y_i is a real-valued statistic based on a random sample of size n from π_i whose distribution has θ_i as a scale (or location) parameter and the constant c is to be chosen such that the P*-condition is satisfied. The correct selection here is the selection of a subset which contains no inferior population.

(c) A fixed subset size approach to the selection problem.

Mahamunulu (1967) considered a selection problem under the indifference-zone approach with the modified goal of selecting a subset of size s which contains at least c of the t best populations where $\max(1, s+t+1-k) \leq c \leq \min(s,t)$.

Closely related to Mahamunulu's problem of determining the common sample size required for a given subset size s, is the problem investigated by Desu and Sobel (1968). Their goal is to select the smallest possible fixed subset size s that will contain the t best of k populations $(t \leq s \leq k)$, based on any given sample size from each population. The basic probability requirement is met under the usual indifference-zone set-up. The aim in the modification is to avoid the possible inclusion of all the populations in the selected subset. The smallest fixed subset size s is determined as a function of the common sample size n and the specified constants but not of the observations.

Nonparametric procedures for selecting fixed-size subsets when the populations are ranked in terms of α-quantiles have been discussed by Desu and Sobel (1971). The random subset size procedure for the case of $t = 1$ has been earlier studied by Rizvi and Sobel (1967) and has been described in Chapter 6.

Sobel (1969) investigated the problem of selecting from k populations a subset containing at least one of the t-best populations for given t and $k(1 \leq t \leq k)$ under an indifference-zone set-up. For $t = 1$, the problem is related to the problem of Desu and Sobel (1968). The procedures proposed by Sobel select a subset which is either of fixed size or of random size depending on the values of the constants specified.

REFERENCES

Alam, K. and Rizvi, M.H. (1966). Selection from multivariate populations. Ann. Inst. Statist. Math. 18, 307-318.

Bahadur, R.R. (1950). On the problem in the theory of k populations. Ann. Math. Statist. 21, 362-375.

Bahadur, R.R. and Robbins, H. (1950). The problem of the greater mean. Ann. Math. Statist. 21, 469-487.

Barlow, R.E. and Gupta, S.S. (1969). Selection procedures for restricted families of distributions. Ann. Math. Statist. 40, 905-917.

Barlow, R.E., Gupta, S.S. and Panchapakesan, S. (1969). On the distribution of the maximum and minimum of ratios of order statistics. Ann. Math. Statist. 40, 918-934.

Barlow, R.E., Marshall, A.W. and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. Ann. Math. Statist. 34, 375-389.

Barr, D.R. and Rizvi, M.H. (1969). Ranking and selection problems of uniform distributions. Trabajos Estadist. 17, 15-31.

Barron, A.M. (1968). A class of sequential multiple decision procedures. Ph.D. Thesis, Purdue University. Also Mimeograph Series No. 169, Department of Statistics, Purdue University, Lafayette, Indiana.

Barron, A.M. and Gupta, S.S. (1970). A class of non-eliminating sequential multiple decision procedures. Mimeograph Series No. 247, Department of Statistics, Purdue University, Lafayette, Indiana.

Bartlett, N.S. and Govindarajulu, Z. (1965). Some distribution-free statistics and their application to the selection problem. Ann. Inst. Statist. Math. 20, 79-97.

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., Kiefer, J. and Sobel, M. (1968). Sequential Identification and Ranking Procedures. Statistical Research Monographs, Vol. III, The University of Chicago Press, Chicago.

Birnbaum, Z.W., Esary, J.D. and Marshall, A.W. (1966). A stochastic characterization of wear-out for components and systems. Ann. Math. Statist. 37, 816-825.

Blumenthal, S. and Patterson, D.W. (1969). Rank order procedures for selecting a subset containing the population with the smallest scale parameter. Sankhya Ser. A 31, 37-42.

Deely, J.J. (1965). Multiple decision procedures from an empirical Bayes approach. Ph.D. Thesis, Purdue University. Also Mimeograph Series No. 45, Department of Statistics. Purdue University, Lafayette, Indiana.

Deely, J.J. and Gupta, S.S. (1968). On the properties of subset selection procedures. Sankhya Ser. A 30, 37-50.

Desu, M.M. (1970). A selection problem. Ann. Math. Statist. 41, 1596-1603.

Desu, M.M. and Sobel, M. (1968). A fixed subset-size approach to a selection problem. Biometrika 55, 401-410.

Desu, M.M. & Sobel, M. (1971). Nonparametric procedures for selecting fixed-size subsets. In Statistical Decision Theory and Related Topics (ed. S. S. Gupta and J. Yackel), pp. 255-273, Academic Press, N.Y.

Doksum, K. (1969). Starshaped transformations and the power of rank tests. Ann. Math. Statist. 40, 1167-1176.

Gnanadesikan, M. and Gupta, S.S. (1970). Selection procedures for multivariate normal distributions in terms of measures of dispersion. Technometrics 12, 103-117.

Govindarajulu, Z. and Gore, A.P. (1971). Selection procedures with respect to measures of association. In Statistical Decision Theory and Related Topics (ed. S. S. Gupta and J. Yackel), pp. 313-345, Academic Press, N.Y.

Gupta, S.S. (1956). On a decision rule for a problem in ranking means. Mimeograph Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, N.C.

Gupta, S.S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math. 14, 199-216.

Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules. Technometrics 7, 225-245.

Gupta, S.S. (1966a). Selection and ranking procedures and order statistics for the binomial distributions. In Classical and Contagious Discrete Distributions. (ed. G. P. Patil), pp. 219-230, Pergamon Press, N.Y.

Gupta, S.S. (1966b). On some selection and ranking procedures for multivatiate normal populations using distance functions. In Multivariate Analysis (ed. P. R. Krishnaiah), pp. 457-475, Academic Press, N.Y.

Gupta, S.S. and Deverman, J. (1969). On a selection procedure concerning the t-best population. Technical Report, Sandia Corporation, Livermore, California. Also Abstract in Ann. Math. Statist. 40, 1870.

Gupta, S.S. and McDonald, G.C. (1970). On some classes of selection procedures based on ranks. In Nonparametric Techniques in Statistical Inference (ed. M. L. Puri), pp. 491-514, Cambridge University Press, London.

Gupta, S.S. and Nagel, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. Sankhya Ser. B 29, 1-34.

Gupta, S.S. and Nagel, K. (1971). On some contributions to multiple decision theory. In Statistical Decision Theory and Related Topics (ed. S. S. Gupta and J. Yackel), pp. 79-102, Academic Press, N.Y.

Gupta, S.S. and Panchapakesan, S. (1969a). Some selection and ranking procedures for multivariate normal populations. In Multivariate Analysis II (ed. P. R. Krishnaiah), pp. 475-505, Academic Press, N.Y.

Gupta, S.S. and Panchapakesan, S. (1969b). Selection and ranking procedures. In The Design of Computer Simulation Experiments (ed. T. H. Naylor), pp. 132-160, Duke University Press, Durham, North Carolina.

Gupta, S.S. and Panchapakesan, S. (1970). On a class of subset selection procedures. Mimeo Series No. 225, Dept. of Statistics, Purdue Univ., Lafayette, Ind. (To appear in Ann. Math. Statist.).

Gupta, S.S. and Sobel, M. (1958). On selecting a subset which contains all populations better than a standard. Ann. Math. Statist. 29, 235-244.

Gupta, S.S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. In Contributions to Probability and Statistics (ed. I. Olkin et al), pp. 224-248, Stanford University Press, Stanford, Calif.

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

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

Gupta, S.S. and Studden, W.J. (1970). On some selection and ranking procedures with applications to multivariate populations. In Essays in Probability and Statistics (ed. R. C. Bose et al), pp. 327-338, University of North Carolina Press, Chapel Hill, North Carolina.

Guttman, I. (1961). Best population and tolerance regions. Ann. Inst. Statist. Math. 13, 9-26.

Guttman, I. (1963). A sequential procedure for the best population. Sankhyā Ser. A 25, 25-28.

Krishnaiah, P.R. (1967). Selection procedures based on covariance matrices of multivariate normal populations. In Blanch Anniversary Volume, pp. 147-160, Aerospace Research Laboratories, Office of Aerospace Research, U.S. Air Force.

Krishnaiah, P.R. and Rizvi, M.H. (1966). Some procedures for selection of multivariate normal populations better than a control. In Multivariate Analysis (ed. P. R. Krishnaiah), pp. 477-490, Academic Press, N.Y.

Lawrence, M.J. (1966). Inequalities and tolerance limits for s-ordered distributions. Technical Report ORC-66-37, Operations Research Center, College of Engineering, University of California, Berkeley.

Lehmann, E.L. (1959). Testing Statistical Hypotheses, John Wiley, N.Y.

Lehmann, E.L. (1961). Some Model I problems of selection. Ann. Math. Statist. 32, 990-1012.

Lehmann, E.L. (1963). A class of selection procedures based on ranks. Math. Ann. 150, 268-275.

Mahamunulu, D.M. (1967). Some fixed-sample ranking and selection problems. Ann. Math. Statist. 38, 1079-1091.

McDonald, G.C. (1969a). On some distribution-free ranking and selection procedures. Ph.D. Thesis, Purdue University. Also Mimeograph Series No. 174, Department of Statistics, Purdue University, Lafayette, Indiana.

McDonald, G.C. (1969b). Some multiple comparison selection procedures based on ranks. Research Publication GMR-948, General Motors Corporation, Warren, Michigan.

McDonald, G.C. (1971). On approximating constants required to implement a selection procedure based on ranks. In Statistical Decision Theory and Related Topics (ed. S. S. Gupta and J. Yackel), pp. 299-312, Academic Press, N.Y.

Mosteller, F. (1948). A K-sample slippage test for an extreme population. Ann. Math. Statist. 19, 58-65.

Nagel, K. (1966). On selecting a subset containing the best of several discrete distributions. Mimeograph Series No. 66, Department of Statistics, Purdue University, Lafayette, Indiana.

Nagel, K. (1970). On subset selection rules with certain optimality properties. Ph.D. Thesis. Purdue University. Also Mimeograph Series No. 222, Department of Statistics, Purdue University, Lafayette, Indiana.

Panchapakesan, S. (1969). Some contributions to multiple decision (selection and ranking) procedures. Ph.D. Thesis, Purdue University. Also Mimeograph Series No. 189, Department of Statistics, Purdue University, Lafayette, Indiana.

Panchapakesan, S. (1970). On a monotonicity property relating to the gamma distributions. Mimeograph Series No. 224, Department of Statistics, Purdue University, Lafayette, Indiana.

Panchapakesan, S. (1971). On a subset selection procedure for the most probable event in a multinomial distribution. In Statistical Decision Theory and Related Topics (ed. S. S. Gupta and J. Yackel), pp. 275-298, Academic Press, N.Y.

Paulson, E. (1949). A multiple decision procedure for certain problems in the analysis of variance. Ann. Math. Statist. 20, 95-98.

Paulson, E. (1952). On the comparison of several experimental categories with a control. Ann. Math. Statist. 23, 239-246.

Puri, P.S. and Puri, M.L. (1968). Selection procedures based on ranks: scale parameter case. Sankhya Ser. A 30, 291-302.

Puri, P.S. and Puri, M.L. (1969). Multiple decision procedures based on ranks for certain problems in analysis of variance. Ann. Math. Statist. 40, 619-632.

Rizvi, M.H. (1963). Ranking and selection problems of normal populations using the absolute values of their means: fixed sample size case. Ph.D. Thesis, University of Minnesota. Also Technical Report No. 31, Department of Statistics, University of Minnesota, Minneapolis, Minnesota.

Rizvi, M.H. and Sobel, M. (1967). Nonparametric procedures for selecting a subset containing the population with the largest αquantile. Ann. Math. Statist. 38, 1788-1803.

Rizvi, M.H., Sobel, M. and Woodworth, G.C. (1968). Nonparametric ranking procedures for comparison with a control. Ann. Math. Statist. 39, 2075-2093.

Rizvi, M.H. and Woodworth, G.C. (1970). On selection procedures based on ranks: counterexamples concerning least favorable configurations. Ann. Math. Statist. 41, 1942-1951.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. 35, 1-20.

Seal, K.C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist. 26, 387-398.

Seal, K.C. (1958). On ranking parameters of scale in Type III populations. J. Amer. Statist. Assoc. 53, 164-175.

Sobel, M. (1969). Selecting a subset containing at least one of the t best populations. In Multivariate Analysis II (ed. P. R. Krishnaiah), pp. 515-540, Academic Press, N.Y.

Studden, W.J. (1967). On selecting a subset of k populations containing the best. Ann. Math. Statist. 38, 1072-1078.

Trawinski, B.J. (1969). Asymptotic approximation to the expected size of the selected subset. Biometrika 56, 207-213.

Trawinski, B.J. and David, H.A. (1963). Selection of the best treatment in a paired-comparison experiment. Ann. Math. Statist. 34, 75-91.

Van Zwet, W.R. (1964). Convex Transformations of Random Variables, Mathematical Center, Amsterdam.