

ON SOME NONPARAMETRIC SELECTION PROCEDURES*

by

Shanti S. Gupta
Purdue University

Sayaji N. Hande
Purdue University
Technical Report #90-60C

Department of Statistics
Purdue University
November 1990
(Revised and updated version of #90-26C)

* This research was supported in part by NSF Grants DMS-8702620, DMS-8923071, and DMS-8717799 at Purdue University.

On Some Nonparametric Selection Procedures*

Shanti S. Gupta and Sayaji N. Hande

Department of Statistics, Purdue University

Abstract

In this paper we study the selection and ranking problem in a nonparametric setup when the populations $\Pi_1, \Pi_2, \dots, \Pi_k$ are characterized by functionals of the associated distribution functions $\theta(F_1), \theta(F_2), \dots, \theta(F_k)$, where $\theta(F_i) = \int g_i dF_i$, for $i = 1, 2, \dots, k$ and g_1, g_2, \dots, g_k are known bounded functions. The problems of selecting the best population under the indifference zone approach and the subset selection approach are considered. Approximate non-randomized rules are obtained. Finally, some simulation studies concerning these procedures are given.

Key Words: Selection and ranking, Nonparametric.

AMS 1985 subject classification: Primary 62G99, 62F07. Secondary 62C20.

*This research was supported in part by NSF Grants DMS-8702620, DMS-8923071, and DMS-8717799 at Purdue University.

1 Introduction

In many practical situations, the experimenter often faces the problem of comparing several competing populations, treatments in clinical trials or processes. The selection and ranking methodology of ranking and selection provides the useful techniques for solving such problems. There have been two main approaches to selection and ranking problems, the indifference zone approach due to Bechhofer (1954) and the subset selection approach due to Gupta (1956). In the indifference zone approach a single population is chosen and is guaranteed to be the best (worst) with probability at least equal to P^* . However, in this formulation it is assumed that the best population is sufficiently apart from the remaining $k - 1$ populations. In the subset selection approach no such restriction on the parameter space is assumed. A random size subset of k populations is chosen which is guaranteed to contain the best (worst) population with probability at least equal to P^* . In this approach the data or the outcome of the experiment is used to decide on how many populations to select. For an extensive review of these formulations see Gupta and Panchapakesan (1979) and Gupta and Panchapakesan (1986).

Often in practice, especially for the new treatments, or for expensive products there is not much information (the past data) which could lead us to assume a parametric model. In this paper we consider a ranking and selection problem in a non-parametric setup. Considerable amount of work has been done on the problems of selecting the population associated with the largest α th quantile (or the largest location parameter) or selecting a subset of the populations which contains the population

associated with the largest α th quantile (or location parameter). Some references are Barlow and Gupta (1969), Gupta and McDonald (1970), Gupta and Huang (1974), Rizvi and Sobel (1967), and Sobel (1967). An extensive review of non-parametric selection and ranking procedures is in Desu and Bristol (1986).

To formulate the problem, let $\Pi_1, \Pi_2, \dots, \Pi_k$ be the k independent populations. The population Π_i is associated with the cumulative distribution function $F_i(\cdot)$ on R^p , for $i = 1, 2, \dots, k$. The population Π_i is characterized by the real-valued functional,

$$\theta(F_i) = \int_{R^p} g_i(x) dF_i(x) ,$$

where g_i is a known, real-valued bounded function on R^p . In this paper we obtain the “optimal” classical type procedures. Non-randomized procedures are proposed. It is also shown that the proposed non-randomized selection procedures are “close” to the optimal procedures. A lower bound for the probability of a correct selection is also obtained. The non-parametric procedures which are developed in this paper are robust and may also be used to do the preliminary analysis. We believe these procedures would be of use in many selection and ranking problems where the distribution functions associated with the populations do not possess “nice” properties. Some applications, examples and Monte Carlo results for the parametric models are presented in Section 4.

2 Indifference Zone Approach

In this section we consider the problem of selecting the best (worst) population under the indifference zone approach. The goal is to select the best population with proba-

bility at least P^* , provided that the “distance” between the best population and the remaining $k - 1$ populations is at least d , where d is some positive number specified by the experimenter.

As defined before, let $\Pi_1, \Pi_2, \dots, \Pi_k$ be the k populations. First we consider the problem of selecting the best population among k populations when the population Π_i is characterized by the functional $\theta(F_i) = \int g dF_i$, for $i = 1, 2, \dots, k$ and we are interested in selecting populations associated with large (small) values of θ . If necessary, we make the transformation

$$g \longrightarrow \frac{g - \inf g}{\sup g - \inf g},$$

and, without any loss of generality assume that $\sup g(x) = 1$ and $\inf g(x) = 0$. Let $\theta_{[1]} \leq \theta_{[2]} \leq \dots \leq \theta_{[k-1]} \leq \theta_{[k]}$ be the ordered values of $\theta_1, \theta_2, \dots, \theta_k$. The correct pairing between ordered and unordered θ 's is completely unknown. The population corresponding to $\theta_{[k]}$ is called the best population, in case of ties we assume that one of them is tagged to be the best population. Our goal is to select the best population with probability of a correct selection at least P^* . We need to define some notations.

Let

$$\mathcal{F} = \{ \underline{F} = (F_1, F_2, \dots, F_k) : F_i \text{ is a distribution on } R^p \}.$$

In general, if we allow \underline{F} to take any value in \mathcal{F} then there does not exist a procedure which would satisfy the P^* condition. Hence we need to restrict the space. Let d be a real number in the interval $(0,1)$ and define, following Bechhofer (1954),

$$\Theta' = \{ (\theta_1, \theta_2, \dots, \theta_k) : \theta_{[k]} - \theta_{[k-1]} \geq d \}$$

and

$$\mathcal{F}' = \{\underline{F} : (\theta(F_1), \theta(F_2), \dots, \theta(F_k)) \in \Theta'\}.$$

Correct selection (CS) : Selecting the best population

Goal: For given P^* ($1/k < P^* < 1$), derive a procedure R and $n_0(d)$ such that for $\forall n \geq n_0(d)$;

$$P_{\underline{F}}(CS|R, n) \geq P^* \text{ for every } \underline{F} \in \mathcal{F}', \quad (1)$$

where $P_{\underline{F}}(CS|R, n)$ denotes the probability of a correct selection for the procedure R . The above condition is called the P^* -condition.

In dealing with the above problem, we need to introduce some notations. Let $\underline{p}_i = (p_{i1}, p_{i2}, \dots, p_{in})$; $\underline{p} = (\underline{p}_1, \underline{p}_2, \dots, \underline{p}_k)$, where $p_{ij} \geq 0$ for $i = 1, 2, \dots, k$ and for $j = 1, 2, \dots, n$. Let Z_{ij} be the independent Bernoulli random variables with parameters p_{ij} for $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, n$, and let U_1, U_2, \dots, U_k be the k independent uniform random variables on interval $(0, \frac{1}{2})$. Let

$$S_i = \sum_{j=1}^n Z_{ij} + U_i.$$

Define

$$\psi_i(\underline{p}) = P(S_i = \max_{1 \leq j \leq k} S_j). \quad (2)$$

Now let $\underline{X}_{i1}, \underline{X}_{i2}, \dots, \underline{X}_{in}$ be the observable independent random vectors from the population Π_i , for $i = 1, 2, \dots, k$. Let $\underline{X} = (\underline{X}_{11}, \underline{X}_{12}, \dots, \underline{X}_{kn})$, and let

$$\tilde{g}(\underline{X}) = (g(\underline{X}_{11}), g(\underline{X}_{12}), \dots, g(\underline{X}_{kn})).$$

Now we propose the following selection procedures.

Procedure R_1 :

Select one of the populations $\Pi_1, \Pi_2, \dots, \Pi_k$ with probabilities $\psi_1(\tilde{g}(x)), \psi_2(\tilde{g}(x)), \dots, \psi_k(\tilde{g}(x))$, respectively.

A non-randomized rule related to the rule R_1 is the following:

Procedure R_2 :

Select the population Π_i for which

$$\psi_i(\tilde{g}(x)) = \max_{1 \leq j \leq k} \psi_j(\tilde{g}(x)),$$

randomize in case of ties.

Notice that the procedure R_1 is a randomized procedure while R_2 is a non-randomized procedure related to the procedure R_1 .

First we prove that the decision rule $\delta(X) = (\psi_1(\tilde{g}(X)), \psi_2(\tilde{g}(X)), \dots, \psi_k(\tilde{g}(X)))$ is “optimum” decision rule for selecting the best population among k populations.

Theorem 2.1 :

The procedure R_1 maximizes the infimum of the probability of a correct selection. i.e.

If R' is any other selection procedure then

$$\inf_{F \in \mathcal{F}'} P_F(CS|R') \leq \inf_{F \in \mathcal{F}'} P_F(CS|R_1).$$

Proof:

Let Π_1 be the best population. Observe that $\inf g(x) = 0$ and $\sup g(x) = 1$. Fix $\epsilon > 0$, and get a and b such that $g(a) < \epsilon/2$ and $g(b) > 1 - \epsilon/2$. We let $g(a) = \epsilon_1$ and $g(b) = 1 - \epsilon_2$.

Let P_i be the counting probability measure induced by a distribution function F_i .

Define

$$\mathcal{F}_0 = \mathcal{F}_{0(\epsilon_1, \epsilon_2)} = \left\{ \underline{F} : \begin{array}{l} P_i(\{b\}) = p_i; \quad P_i(\{a\}) = 1 - p_i \\ 0 \leq p_i \leq 1; \quad \text{for } i = 1, 2, \dots, k \end{array} \right\} \cap \mathcal{F}'. \quad (3)$$

Using the earlier notation, we define for $i = 1, 2, \dots, k$,

$$A_i = \{X_{ij} : X_{ij} = b, j = 1, 2, \dots, n\} \quad \text{and} \quad T_i(\underline{X}) = |A_i|.$$

Note that for a class of distribution functions \mathcal{F}_0 , the statistics $T = (T_1, T_2, \dots, T_k)$ is a complete sufficient statistic.

We also note that T_1, T_2, \dots, T_k are independent and they have binomial distributions with parameters $(n, p_1), (n, p_2), \dots, (n, p_k)$, respectively. Since the binomial distribution has the monotone likelihood ratio property, it is easy to see that for every permutation invariant prior on \mathcal{F}_0 , a rule which selects the population associated with the largest T_i (randomized in case of ties) is a Bayes rule for 0-1 valued loss function.

Hence

$$\inf_{\underline{F} \in \mathcal{F}_{0(\epsilon_1, \epsilon_2)}} P_{\underline{F}}(CS|R') \leq \inf_{\underline{p} \in \mathcal{P}(\epsilon_1, \epsilon_2)} P_{\underline{p}}(T_1 + U_1 = \max_j (T_j + U_j)),$$

where

$$\mathcal{P}(\epsilon_1, \epsilon_2) = \{\underline{p} = (p_1, p_2, \dots, p_k) : (1 - \epsilon_2)p_1 + \epsilon_1(1 - p_1) \geq \max_{j \neq 1} (1 - \epsilon_2)p_j + \epsilon_1(1 - p_j) + d\}.$$

Since

$$\inf_{\underline{F} \in \mathcal{F}'} P_{\underline{F}}(CS|R') \leq \inf_{\underline{F} \in \mathcal{F}_{0(\epsilon_1, \epsilon_2)}} P_{\underline{F}}(CS|R')$$

and ϵ is arbitrary, letting $\epsilon \rightarrow 0$ we have

$$\inf_{\underline{F} \in \mathcal{F}'} P_{\underline{F}}(CS|R') \leq \inf_{\underline{p} \in \mathcal{P}'} P(T_1 + U_1 = \max_j (T_j + U_j)),$$

where

$$\mathcal{P}' = \{p : p_1 \geq \max_{j \neq 1} p_j + d\}.$$

But from (2) we get

$$P_{\underline{F}}(CS|R_1) = E_{\underline{F}}\psi_1(\tilde{g}(X)) = E_{\underline{F}}P(S_1 = \max_j S_j|X),$$

where $S_i = \sum_{j=1}^n Z_{ij} + U_i = V_i + U_i$. For given $X = x$, Z_{ij} 's are independent Bernolli random variables with $P(Z_{ij} = 1) = g(x_{ij})$. Hence marginally V_1, V_2, \dots, V_k are independent binomial random variables with parameters $(n, \theta(F_1)), (n, \theta(F_2)), \dots, (n, \theta(F_k))$ repectively. Hence

$$\inf_{p \in \mathcal{P}'} P(T_1 + U_1 = \max_j (T_j + U_j)) = \inf_{\underline{F} \in \mathcal{F}'} P(CS|R_1).$$

This completes the proof of the theorem. □

Remark 2.1 :

From this theorem we see that the procedure R_1 is the “most economical” in the sense that for a given P^* and d there doesn't exist any other procedure which can meet the basic probability requirement with a smaller sample. This was also proved in a special case by Hall (1958).

Theorem 2.2 :

[1] $P_{\underline{F}}(CS|R_1, n)$ is increasing in n and

[2] $P_{\underline{F}}(CS|R_1, n)$ is an increasing function of $\theta_{[k]}$ provided $\theta_{[1]}, \theta_{[2]}, \dots, \theta_{[k-1]}$ are held fixed.

[3] $\inf_{\underline{F} \in \mathcal{F}'} P_{\underline{F}}(CS|R_1, n) \longrightarrow 1$ as $n \longrightarrow \infty$.

Proof:

It is straightforward to see that

$$P_{\underline{F}}(CS|R_1, n) = P(Y_{kn} + U_1 = \max_{1 \leq j \leq k} (Y_{jn} + U_j)), \quad (4)$$

where $Y_{1n}, Y_{2n}, \dots, Y_{kn}$ are independent binomial random variables with parameters $(n, \theta_{[1]}), (n, \theta_{[2]}), \dots, (n, \theta_{[k]})$ respectively, and U_1, U_2, \dots, U_k are independent uniform random variables over the interval $(0, \frac{1}{2})$. If we consider the problem of selecting the best population among k binomial populations, the procedure which selects the population Π_i for which $Y_{in} + U_i = \max_j (Y_{jn} + U_j)$ is the best invariant and is a Bayes procedure with respect to every invariant prior on Θ' , provided that the underlying loss function is permutation invariant, “monotone” (more loss for selecting bad population) and non-negative. Hence the Bayes risk of the procedure R_1 decreases as n increases for every permutation invariant prior on Θ' . Thus $P_{\underline{F}}(CS|R_1, n)$ is increasing in n . From equation (2) it is clear that $P_{\underline{F}}(CS|R_1, n)$ is an increasing function of $\theta_{[k]}$.

For $\underline{F} \in \mathcal{F}'$

$$P_{\underline{F}}(CS|R_1, n) \geq \inf_{\theta \in \Theta_0} P_{\theta}(Y_k > \max_{1 \leq j \leq k-1} Y_{jn}),$$

where $\Theta_0 = \{\theta : 0 \leq \theta_1; \theta_1 + d = \theta_2 + d = \dots = \theta_{k-1} + d = \theta_k \leq 1\}$. But

$$\begin{aligned} & P_{\theta}(Y_{kn} > \max_{1 \leq j \leq k-1} Y_{jn}) \\ & \geq P_{\theta}^{k-1}(Y_{kn} > Y_{1n}) \quad (\text{Jensen's Inequality}) \\ & = [1 - P_{\theta}(\frac{1}{n}(Y_{1n} - Y_{kn}) > 0)]^{k-1} \\ & = [1 - P_{\theta}(\frac{1}{n}(Y_{1n} - Y_{kn}) - \mu > -\mu)]^{k-1} \quad ; \mu = E[\frac{1}{n}(Y_{1n} - Y_{kn})] = -d < 0 \end{aligned}$$

$$\begin{aligned}
&\geq \left[1 - \frac{\text{Var}\left(\frac{1}{n}(Y_{1n} - Y_{kn})\right)}{d^2}\right]^{k-1} \quad (\text{Chebyshe's inequality}) \\
&\geq \left[1 - \frac{\text{Var}\left(\frac{1}{n}Y_{1n}\right) + \text{Var}\left(\frac{1}{n}Y_{kn}\right)}{d^2}\right]^{k-1} \\
&\geq \left[1 - \frac{1}{2nd^2}\right]^{k-1}.
\end{aligned}$$

Hence for every $\underline{F} \in \mathcal{F}'$

$$P_{\underline{F}}(CS|R_1, n) \geq \left[1 - \frac{1}{2nd^2}\right]^{k-1},$$

i. e.

$$\inf_{\underline{F} \in \mathcal{F}'} P_{\underline{F}}(CS|R_1, n) \geq \left(1 - \frac{1}{2nd^2}\right)^{k-1}.$$

Letting $n \rightarrow \infty$ the [3] follows. This completes the proof of the theorem. \square

The above theorem insures that for a given P^* , there exists $n_0(P^*, k, d)$ such that

$$P_{\underline{F}}(CS|R_1, n) \geq P^* \text{ for every } \underline{F} \in \mathcal{F}', \quad n \geq n_0.$$

The procedure R_1 has nice properties, however it is a randomized procedure. In practice the experimenter would like to use a non-randomized procedure. The procedure R_2 is a non-randomized version of the procedure R_1 . The following theorem gives the relationship between $P_{\underline{F}}(CS|R_1)$ and $P_{\underline{F}}(CS|R_2)$.

Theorem 2.3 :

For every $\underline{F} \in \mathcal{F}$

$$P_{\underline{F}}(CS|R_2) \geq 2 P_{\underline{F}}(CS|R_1) - 1. \quad (5)$$

Proof:

Let Π_1 be the best population, and I be an indicator function then

$$\begin{aligned}
P_{\underline{F}}(CS|R_2) &\geq \int I_{\{\psi_1(\tilde{g}(\underline{x})) > \max_{j \neq 1} \psi_j(\tilde{g}(\underline{x}))\}} d\underline{F}(\underline{x}) \\
&\geq \int [\psi_1(\tilde{g}(\underline{x})) - \max_{j \neq 1} \psi_j(\tilde{g}(\underline{x}))] d\underline{F}(\underline{x}) \\
&= \int \psi_1(\tilde{g}(\underline{x})) d\underline{F}(\underline{x}) - \int \max_{j \neq 1} \psi_j(\tilde{g}(\underline{x})) d\underline{F}(\underline{x}) \\
&= P_{\underline{F}}(CS|R_1) - \int \max_{j \neq 1} \psi_j(\tilde{g}(\underline{x})) d\underline{F}(\underline{x}) \\
&\geq P_{\underline{F}}(CS|R_1) - \int \sum_{j \neq 1} \psi_j(\tilde{g}(\underline{x})) d\underline{F}(\underline{x}) \\
&= P_{\underline{F}}(CS|R_1) - \int (1 - \psi_1(\tilde{g}(\underline{x}))) d\underline{F}(\underline{x}) \\
&= P_{\underline{F}}(CS|R_1) - 1 + P_{\underline{F}}(CS|R_1) \\
&= 2 P_{\underline{F}}(CS|R_1) - 1.
\end{aligned}$$

This proves the theorem. □

Remark 2.2 :

From Theorems 2.2 and 2.3 it follows that

$$\inf_{\underline{F} \in \mathcal{F}'} P_{\underline{F}}(CS|R_2, n) \longrightarrow 1 \quad \text{as } n \longrightarrow \infty.$$

Remark 2.3 :

Observing the method of the proof of the above theorem, we note that, the above theorem holds for any multiple decision problem with 0-1 loss, R_1 is any procedure and the procedure R_2 is a “non-randomized version” of the procedure R_1 .

As we noticed before we can generalize the procedures R_1 and R_2 to obtain the procedure for selecting the best population with the highest parameter, when the

population Π_i is characterized by the functional

$$\Theta(F_i) = \int g_i dF_i;$$

where g_i is a known real-valued function with $\inf_x g_i(x) = 0$ and $\sup_x g_i(x) = 1$, for $i = 1, 2, \dots, k$. This can be done in the following way.

Define

$$\tilde{g}(x) = (g_1(x_{11}), g_1(x_{12}), \dots, g_1(x_{1n}), g_2(x_{21}), \dots, g_2(x_{2n}), \dots, \dots, g_k(x_{k1}), \dots, g_k(x_{kn})).$$

Let $\psi_1, \psi_2, \dots, \psi_k$ be as defined in equation (2).

Procedure R_3 :

Select one of the populations $\Pi_1, \Pi_2, \dots, \Pi_k$ with probabilities $\psi_1(\tilde{g}(x)), \psi_2(\tilde{g}(x)), \dots, \psi_k(\tilde{g}(x))$, respectively.

A non-randomized version of this procedure is given by:

Procedure R_4 :

Select the population Π_i for which

$$\psi_i(\tilde{g}(x)) = \max_{1 \leq j \leq k} \psi_j(\tilde{g}(x))$$

and randomize in case of ties.

Theorem 2.1 , Theorem 2.2 , Theorem 2.3 and the above remarks hold true for these procedures also.

Theorem 2.3 indicates that the procedure R_2 (R_4) is a “good” approximate non-randomized version of procedure R_1 (R_3), whenever P^* is large, and that is the case in general. For example, if $P_{\underline{F}}(CS|R_1) \geq 0.99$ then $P_{\underline{F}}(CS|R_2) \geq 0.98$. The procedure is good, in the sense that we lose at most $1 - P^*$ due to non-randomization. We also

note that these procedures can be generalized to the problem of selecting the t best populations.

As given by equation (4) the probability of a correct selection can be written in terms of the binomial probabilities. The sample sizes, $n_o(P^*, k, d)$ (exact and approximate) are tabulated by Sobel and Huyett (1957) for $k = 2, 3, 4, 10$, $d = 0.05(0.05)0.5$ and $P^* = 0.5, 0.6, 0.75, 0.90, 0.95, 0.99$. For $k = 2$ they conjectured that the least favorable configuration occurs at $\theta_{[2]} = (1 + d)/2$ and $\theta_{[1]} = (1 - d)/2$. This conjecture is shown to be true by Eaton and Gleser (1989).

3 Subset Selection Approach

In the subset selection approach we select a random size subset of the k populations which contains the best population with probability P^* ($1/k < P^* < 1$). The main feature of selecting a subset of random size is to allow the size to be determined by the observations themselves. Also in the subset selection approach we need not assume any restriction on the “parameter space”.

Now we describe the problem formally, let us assume that there are k populations $\Pi_1, \Pi_2, \dots, \Pi_k$. The random variable associated with population Π_i has the cumulative distribution function $F_i(\cdot)$ on R^p . Again the characterizing function is real-valued as defined earlier. Let $\theta_{[1]} \leq \theta_{[2]}, \dots, \leq \theta_{[k]}$ be the ordered values of $\theta_1, \theta_2, \dots, \theta_k$. The population associated with $\theta_{[k]}$ is called the best population, in case of ties one of them is tagged as the best population. Our goal is to select a non empty subset of these k populations so that the selected subset includes the population associated with $\theta_{[k]}$

with large probability. Let CS denote the event of correct selection and $P_{\underline{F}}(CS|R)$ denote the probability of a correct selection for the procedure R .

CS: Selecting a subset of k populations which contains the best population.

Goal: Derive a subset selection procedure R for which

$$P_{\underline{F}}(CS|R) \geq P^* \text{ for every } \underline{F} \in \mathcal{F}.$$

Let the space \mathcal{D} be the decision space consists of $2^k - 1$ non empty subsets of the set $\{1, 2, \dots, k\}$, *i.e.*

$$\mathcal{D} = \{a : a \subset \{1, 2, \dots, k\} \text{ and } |a| \geq 1\}.$$

Action $a = \{i_1, i_2, \dots, i_r\} \in \mathcal{D}$ corresponds to the selection of the populations $\Pi_{i_1}, \Pi_{i_2}, \dots, \Pi_{i_r}$. A decision “ a ” is called a correct selection (CS) if the best population is included in the selected subset. We implement the procedures established by Gupta and Sobel (1960) for selecting a subset of k binomial populations containing the best population. To define the procedures we need some notation. Let $\underline{p}_i = (p_{i1}, p_{i2}, \dots, p_{in})$, $0 \leq p_{ij} \leq 1$ for $i = 1, 2, \dots, k$ and for $j = 1, 2, \dots, n$. Let $\underline{p} = (\underline{p}_1, \underline{p}_2, \dots, \underline{p}_k)$.

For every $a \in \mathcal{D}$ define

$$\psi_a(\underline{p}) = P(\min_{i \in a} S_i \geq \max_{1 \leq j \leq k} S_j - d > \max_{i \notin a} S_i), \quad (6)$$

where $S_i = \sum_{j=1}^n Z_{ij}$ for $i = 1, 2, \dots, k$. For $i = 1, 2, \dots, k$ and for $j = 1, 2, \dots, n$, Z_{ij} are independent Bernoulli random variables with $P(Z_{ij} = 1) = p_{ij}$. Let $\underline{X}_{i1}, \underline{X}_{i2}, \dots, \underline{X}_{in}$ be the observable random vectors form population Π_i for $i = 1, 2, \dots, k$. Let

$$\tilde{g}(\underline{x}) = (g(\underline{x}_{11}), g(\underline{x}_{12}), \dots, g(\underline{x}_{1n}), \dots, g(\underline{x}_{k1}), g(\underline{x}_{k2}), \dots, g(\underline{x}_{kn})).$$

Procedure R_s :

Having observed $X = x$, select a subset of populations $\Pi_{i_1}, \Pi_{i_2}, \dots, \Pi_{i_r}$ with probability $\psi_a(\tilde{g}(x))$, where $a = \{i_1, i_2, \dots, i_r\}$.

Theorem 3.1 :

$$\inf_{\tilde{F} \in \tilde{\mathcal{F}}} P_{\tilde{F}}(CS|R_s) = \inf_{0 < \theta < 1} P_{\theta}(Y_1 \geq \max_{1 \leq i \leq k} Y_i - d), \quad (7)$$

where Y_1, Y_2, \dots, Y_k are i. i. d. binomial random variables with parameter (n, θ) .

Proof:

Let Π_1 be the best population then

$$\begin{aligned} P_{\tilde{F}}(CS|R_s) &= E_{\tilde{F}} \sum_{a:1 \in a} \psi_a(X) \\ &= E_{\tilde{F}} P(S_1 \geq \max_{1 \leq j \leq k} S_j - d | X), \end{aligned}$$

where $S_i = \sum_{j=1}^n Z_{ij}$ and for given $X = x$, Z_{ij} 's are independent Bernoulli random variables with $P(Z_{ij} = 1 | X = x) = g(x_{ij})$. Hence marginally S_1, S_2, \dots, S_k are independent binomial random variables with parameters $(n, \theta_1), (n, \theta_2), \dots, (n, \theta_k)$, respectively.

Hence we have

$$\inf_{\tilde{F} \in \tilde{\mathcal{F}}} P_{\tilde{F}}(CS|R_s) = \inf_{0 \leq \theta_i \leq \theta_1} P(S_1 \geq \max_{1 \leq j \leq k} S_j - d).$$

From Gupta and Sobel (1960) we know that

$$\inf_{0 \leq \theta_i \leq \theta_1} P(S_1 \geq \max_{1 \leq j \leq k} S_j - d) = \inf_{\theta \in \Omega_0} P(S_1 \geq \max_{1 \leq j \leq k} S_j - d),$$

where $\Omega_0 = \{(\theta_1, \theta_2, \dots, \theta_k) : \theta_1 = \theta_2 = \dots = \theta_k\}$.

This completes the proof of the theorem. □

In the case of $k = 2$, Gupta and Sobel (1960) proved that

$$\inf_{\theta \in \Omega_0} P(S_1 \geq \max_{1 \leq j \leq k} S_j - d)$$

is attained at $\theta_1 = \theta_2 = \frac{1}{2}$. For $k > 2$, the common value θ_0 at which infimum takes place is not known. The conservative values of d based on the normal approximation have been tabulated by Gupta and Sobel (1960) for $k = 2(1)20(5)50$, $n = 1(1)20(5)50(10)100(25)200(50)500$ and $P^* = 0.75, 0.90, 0.95, 0.99$. Gupta, Huang and Huang (1976) obtained conservative values of d when $P^* = 0.75, 0.90, 0.95, 0.99$ and $n = 1(1)4$ when $k = 3(1)15$, and $n = 5(1)10$ when $k = 3(1)5$.

The procedure R_s is randomized, the non-randomized version of this procedure is given by

Procedure R'_s :

After observing $X = \underline{x}$ select a subset of populations $\Pi_{i_1}, \Pi_{i_2}, \dots, \Pi_{i_r}$ if

$$\psi_a(\tilde{g}(\underline{x})) = \max_{a' \in \mathcal{D}} \psi_{a'}(\tilde{g}(\underline{x})),$$

where $a = \{i_1, i_2, \dots, i_r\}$, randomize in case of ties.

As in the previous section we can generalize these subset selection procedures when population Π_i is characterized by the functional $\theta(F_i) = \int g_i dF_i$, for $i = 1, 2, \dots, k$. Unlike the indifference zone approach case, we have not been able to get a lower bound for the probability of a correct selection for the procedure R'_s . We feel however that, there exists a constant $c = c(n)$ and a non trivial subset \mathcal{F}_1 of \mathcal{F} such that

$$P_{\underline{F}}(CS|R'_s) \geq P_{\underline{F}}(CS|R_s) - c(1 - P_{\underline{F}}(CS|R_s)),$$

$\forall F \in \mathcal{F}_1$.

In the next section we will describe a technique to use the nonparametric procedures developed before for the parametric models.

4 Applications and Examples

Let $\Pi_1, \Pi_2, \dots, \Pi_k$ be the k populations. The population Π_i is associated with the cumulative distribution function $F_i(x) = F(x - \mu_i)$. In this univariate case, let $X_{i1}, X_{i2}, \dots, X_{in}$ be the independent observable random variables from population Π_i . Let $\mu_{[1]} \leq \mu_{[2]} \leq \dots \leq \mu_{[k]}$ be the ordered μ_i 's. The population associated with the largest $\mu_{[k]}$ is called the best population. The distribution function $F(\cdot)$ is assumed to be known. We consider the problem of selecting the best population. Let

$$\begin{aligned}\theta(F_i) &= \int g(x) dF_i(x) \\ &= \int g(x + \mu_i) dF(x).\end{aligned}$$

Note that if g is a nondecreasing function then the above transformation is monotone, hence the population associated with the largest θ is the best population. After choosing a proper g we can use the nonparametric procedures developed before for selecting the best population or for selecting a subset containing the best population. In the following examples we take $g(x) = F(x - \mu)$, where $\mu = \mu_{[k-1]} + \frac{3}{4}(\mu_{[k]} - \mu_{[k-1]})$. For large n first μ was estimated and then selection was done using the same data.

Now we present some examples and Monte Carlo results. Standard error for all Monte Carlo estimates was found to be less than 0.035 .

Example 4.1 :

Let

$$f_i(x) = \frac{1}{2}e^{-|x-\mu_i|} \quad \text{for } i = 1, 2, \dots, k,$$

where f_i is the density associated with F_i for $i = 1, 2, \dots, k$. We want to select the population associated with $\mu_{[k]}$. Take g as the c.d.f. of a double exponential random variable with parameter μ . The problem of selecting the population with the highest location parameter is same as the problem of selecting the population with the highest functional. Now we will use the nonparametric procedures and make comparisons. Let R_{ij} denote the rank of X_{ij} in the set of pooled observations. Let $T_i = \sum_{j=1}^n R_{ij}$, $V_i = \sum_{j=1}^n \Phi^{-1}(R_{ij}/(N+1))$, where Φ is a distribution function of standard normal variable and $U_i = \sum_{j=1}^n EZ_{(R_{ij})}$, where $Z_{(1)}, Z_{(2)}, \dots, Z_{(N)}$ are order statistics of $N = nk$ independent exponential random variables. Let

R_1 : The nonparametric rule.

R_2 : Non randomized version of the rule R_1 .

R_3 : Selects the population associated with the largest rank score (T_i).

R_4 : Selects the population associated with the largest normal score (V_i).

R_5 : Selects the population associated with the largest exponential score (U_i).

R_{median} : Selects the population associated with the highest median.

R_{mean} : Selects the population associated with the highest mean.

Assume $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, $\mu_5 = 1$, and $n = 13$. Then the following table gives the $P(CS)$ which was computed using Monte Carlo methods.

$P(CS)$	$\mu = 2$	$\mu = 0.75$	$\mu = 0.50$
$P(CS R_1)$	0.4823	0.6815	0.665
$P(CS R_2)$	0.690	0.882	0.880
$P(CS R_{median})$	0.887	0.887	0.887

In practice we do not know the configuration of μ_i 's, then we may estimate $\mu_{[k]}$ and $\mu_{[k-1]}$ by sample medians and set $\mu = \text{estimate of } \mu_{[k-1]} + \frac{3}{4}(\mu_{[k]} - \mu_{[k-1]})$. Then for $n = 23$

$$P(CS|R_1) = 0.871$$

$$P(CS|R_2) = 0.994$$

$$P(CS|R_3) = 0.990$$

$$P(CS|R_4) = 0.983$$

$$P(CS|R_5) = 0.962$$

$$P(CS|R_{median}) = 0.990$$

In this example, in terms of $P(CS)$, the procedure R_2 is better than the procedure R_1 and is as good as other procedures.

Example 4.2 :

Let

$$F_i(x) = \frac{1}{1 + e^{-(x-\mu_i)}}, \text{ for } i = 1, 2, \dots, k.$$

Assume $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, $\mu_5 = 1$ and let g be the c.d.f of logistic, $\mu = 0.75$.

Then for $n = 13$

$$P(CS|R_1) = 0.550$$

$$P(CS|R_2) = 0.806$$

$$P(CS|R_{median}) = 0.748$$

$$P(CS|R_{mean}) = 0.803$$

As in the previous example we will estimate $\mu_{[k]}$ and $\mu_{[k-1]}$ and by sample medians and set $\mu = \text{estimate of } \mu_{[k-1]} + \frac{3}{4}(\mu_{[k]} - \mu_{[k-1]})$. For $n = 23$

$$P(CS|R_1) = 0.694$$

$$P(CS|R_2) = 0.912$$

$$P(CS|R_3) = 0.918$$

$$P(CS|R_4) = 0.916$$

$$P(CS|R_5) = 0.881$$

$$P(CS|R_{median}) = 0.867$$

$$P(CS|R_{mean}) = 0.91$$

In this example, in terms of $P(CS)$, the procedure R_2 is better than the procedure R_1 and is as good as other procedures.

Let $R'_1, R'_2, \dots, R'_5, R'_{median}$ be the associated subset selection rules. Let $P(CS)$, $E(S)$ be the associated probability of a correct selection and expected subset size, respectively. Using $n = 13$, we have

Rule	$P(CS)$	ES	$P(CS)$	ES	$P(CS)$	ES
R'_1	0.787	1.942	0.895	2.732	0.965	3.555
R'_2	0.849	1.209	0.982	2.809	1.000	4.282
R'_3	0.800	1.013	0.825	1.028	0.828	1.052
R'_4	0.902	1.418	0.964	1.865	0.991	2.528
R'_5	0.851	1.408	0.935	1.873	0.975	2.446
R'_{median}	0.950	2.319	0.976	2.607	0.981	2.815

In above examples we observe that $P(CS|R_1) < P(CS|R_2)$. These results indicate that the nonrandomized version of the nonparametric procedure works better when the associated distribution functions are stochastically increasing in the parameters. In these examples we have used the binomial model to develop appropriate selection procedures. Note that if we take g as an indicator function (which has been usually done), the procedures R_1 and R_2 are identical, hence the gain in the $P(CS)$ of R_2 over R_1 , which is observed in the above examples, is not possible in this case. It was also observed that the procedures associated with $g_1(x) = I_{\{x > \frac{\mu_{[k]} + \mu_{[k-1]}}{2}\}}$ and $g_2(x) = I_{\{x > \mu_{[k-1]} + \frac{3}{4}(\mu_{[k-1]} + \mu_{[k]})\}}$ do not perform better than the procedure R_2 .

Now we will prove that $P(CS|R_1) < P(CS|R_2)$ for $k = 2$ and $n = 1$ in the location parameter case. Let $F(\cdot)$ be the associated distribution function, X_1 be the observable random variable from the population Π_1 with location parameter μ_1 and X_2 be the observable random variable from population Π_2 with location parameter μ_2 . Let g be the distribution function of X_2 . Let Π_1 be the best population.

Then

$$\begin{aligned}P(CS|R_1) &= E[g(X_1)(1 - g(X_2)) + \frac{1}{2}g(X_1)g(X_2) + \frac{1}{2}(1 - g(X_1))(1 - g(X_2))] \\&= E[g(X_1) - g(X_1)g(X_2) + \frac{1}{2}g(X_1)g(X_2)] \\&\quad + E[\frac{1}{2} - \frac{1}{2}(g(X_1) + g(X_2)) + \frac{1}{2}g(X_1)g(X_2)] \\&= \frac{1}{2}E[g(X_1) - g(X_2) + 1].\end{aligned}$$

Let Z_1 and Z_2 be the independent random variables with a common distribution function $F(\cdot)$. Set $Z = Z_1 - Z_2$. We have the following;

$$\begin{aligned}P(CS|R_1) &= \frac{1}{2}[P(Z > -\mu_1 + \mu_2) - P(Z > 0) + 1] \\&= \frac{1}{2}[P(Z > -\mu_1 + \mu_2) + \frac{1}{2}].\end{aligned}$$

It is straightforward to see that

$$P(CS|R_2) = P(Z > -\mu_1 + \mu_2).$$

Hence

$$P(CS|R_2) > P(CS|R_1)$$

if and only if

$$P(Z > -\mu_1 + \mu_2) > \frac{1}{2}[P(Z > -\mu_1 + \mu_2) + \frac{1}{2}]$$

which is true if and only if

$$P(Z > -\mu_1 + \mu_2) > \frac{1}{2},$$

which is always true.

4.1 Asymptotic Relative Efficiency

The efficiency of a subset selection procedure can be measured by $E(S)$, where S denote the selected subset size. In this section we make comparison of the nonparametric procedure developed and Gupta's maximum type procedure based on the sample means. Let $\Pi_1, \Pi_2, \dots, \Pi_k$ be the k populations. The population Π_i is associated with the cumulative distribution function $F_i(x) = F(x - \mu_i)$. Let us denote

$$\Omega_\Delta = \{\mu : \mu_{[1]} = \mu_{[2]} = \dots = \mu_{[k-1]} = 0, \mu_{[k]} = \Delta\}.$$

Definition 4.1 *Let R_1 and R_2 be the two subset selection procedures satisfying the P^* condition, then the asymptotic relative efficiency of R_1 relative to R_2 is*

$$ARE(R_1, R_2; \Delta) = \lim_{\epsilon \rightarrow 0} n_2(\epsilon)/n_1(\epsilon),$$

where $n_1(\epsilon)$ and $n_2(\epsilon)$ are the sample sizes for which $E_\mu(S) - P_\mu(CS) = \epsilon$ for R_1 and R_2 respectively for $\mu \in \Omega_\Delta$.

Let σ^2 be the variance of a random variable from the distribution $F(\cdot)$. Let $X_{i1}, X_{i2}, \dots, X_{in}$ be the observable random variables and \bar{X}_i be the sample mean from the population Π_i . Let R_m be the Gupta's maximum type procedure;

Procedure R_m :

Select Π_i if and only if

$$\bar{X}_i > \max_j \bar{X}_j - \frac{\sigma d_m}{\sqrt{n}},$$

where d_m is chosen to satisfy the P^* condition:

$$\inf_{\mu} P(CS|R_m) \geq P^*.$$

Let R_s be the nonparametric procedure for selecting a subset containing the best population when the population Π_i is characterized by $\theta(F_i) = \int g(x)dF_i(x)$; $g(x) = F(x - \Delta/2)$. To define the procedure we need some more notation. Let,

$$\begin{aligned} Y_{ij} &= 1 \text{ with probability } g(X_{ij}) \\ &= 0 \text{ with probability } 1 - g(X_{ij}). \end{aligned}$$

Let $T_i = \sum_j Y_{ij}$, for $i = 1, 2, \dots, k$. Note that T_i is a binomial random variable with parameters n and $p = E(g(X_{i1}))$.

Procedure R_s :

Select Π_i if and only if

$$T_i > \max_j T_j - \frac{\sqrt{nd_s}}{2},$$

where d_s is chosen to satisfy the P^* condition:

$$\inf_{\mu} P(CS|R_s) \geq P^*.$$

To derive $ARE(R_m, R_s; \Delta)$ we need to find out $n_m(\epsilon)$ and $n_s(\epsilon)$ for arbitrarily small ϵ . Hence to find out d_m and d_s to satisfy the P^* condition we can assume n is large. Using the central limit theorem we see that $d_m = d$ is chosen to satisfy

$$\int \Phi^{k-1}(x+d)d\Phi(x) = P^*. \quad (8)$$

Gupta and Sobel (1960) have shown that for large n , d_s is chosen to satisfy the condition (8).

Now we will obtain asymptotic expression for $E_{\mu}(S) - P_{\mu}(CS)$ for the procedures R_m and R_s . We assume $\mu \in \Omega_{\delta}$.

$$\begin{aligned}
& E(S|R_m) - P(CS|R_m) \\
&= (k-1)P(\Pi_1 \text{ is selected}) \\
&= (k-1)P(\bar{X}_1 > \max_j \bar{X}_j - \frac{\sigma d}{\sqrt{n}}) \\
&\simeq (k-1) \int \Phi^{k-2}(z+d)\Phi(z+d-\sqrt{n}\Delta_m)d\Phi(z) \quad (\Delta_m = \Delta/\sigma) \\
&\simeq \frac{(k-1)}{\sqrt{n}\Delta_m} \int \Phi^{k-2}(z+d)\phi(z+d-\sqrt{n}\Delta_m)d\Phi(z) \\
&\quad (\text{for large } z > 0, \Phi(-z) \simeq \phi(z)/z) \\
&= \frac{(k-1)}{\sqrt{2n}\Delta_m} \phi\left(\frac{d-\sqrt{n}\Delta_m}{\sqrt{2}}\right) \int \Phi^{k-2}\left(\frac{x}{\sqrt{2}} + \frac{d+\sqrt{n}\Delta_m}{2}\right)d\Phi(x) \\
&\simeq \frac{(k-1)}{\sqrt{2n}\Delta_m} \phi\left(\frac{d-\sqrt{n}\Delta_m}{\sqrt{2}}\right) \int [1 - (k-2)\Phi\left(-\frac{x}{\sqrt{2}} - \frac{d+\sqrt{n}\Delta_m}{2}\right)]d\Phi(x) \\
&\quad (\text{for small } x; (1-x)^k \simeq 1 - kx) \\
&\simeq \frac{(k-1)}{\sqrt{2n}\Delta_m} \phi\left(\frac{d-\sqrt{n}\Delta_m}{\sqrt{2}}\right) [1 - (k-2)\Phi\left(\frac{-d-\sqrt{n}\Delta_m}{\sqrt{6}}\right)] \\
&\simeq \frac{(k-1)}{\sqrt{2n}\Delta_m} \phi\left(\frac{d-\sqrt{n}\Delta_m}{\sqrt{2}}\right) \\
&\simeq \frac{(k-1)}{2} \Phi\left(\frac{d-\sqrt{n}\Delta_m}{\sqrt{2}}\right) \\
&\quad \text{for large } z > 0, \Phi(-z) \simeq \phi(z)/z.
\end{aligned}$$

It is straightforward to see that

$$E(S|R_s) - P(CS|R_s) = (k-1)P(T_1 > \max_j T_j - \frac{\sqrt{nd}}{2}).$$

Let X and Y be the independent random variables with a common cumulative distribution function $F(\cdot)$. Let $Z = X - Y$ and $p_1 = P(Z > -\Delta/2)$ and $p_0 = P(Z > \Delta/2)$. Note that $p_1 = 1 - p_0$, T_1, T_2, \dots, T_{k-1} are independent binomial random variables with parameter (n, p_0) and T_k is a binomial random variable with parameter (n, p_1) .

Using these facts and the central limit theorem we have the following:

$$\begin{aligned}
& E(S|R_s) - P(CS|R_s) \\
&= (k-1)P\left(\frac{\sqrt{n}}{\sqrt{p_0(1-p_0)}}\left(\frac{T_1}{n} - p_0\right) \geq \max_j \frac{\sqrt{n}}{\sqrt{p_0(1-p_0)}}\left(\frac{T_j}{n} - p_0\right) - \frac{d}{2\sqrt{p_0(1-p_0)}}\right) \\
&\quad \text{noting that } p_0(1-p_0) = p_1(1-p_1), \\
&\simeq (k-1) \int \Phi^{k-2}\left(z + \frac{d}{2\sqrt{p_0(1-p_0)}}\right) \Phi\left(z + \frac{d}{2\sqrt{p_0(1-p_0)}} - \frac{\sqrt{n}(p_1-p_0)}{\sqrt{p_0(1-p_0)}}\right) d\Phi(z) \\
&\simeq (k-1) \int \Phi^{k-2}\left(z + \frac{d}{2\sqrt{p_0(1-p_0)}}\right) \Phi\left(z + \frac{d}{2\sqrt{p_0(1-p_0)}} - \sqrt{n}\Delta_s\right) d\Phi(z),
\end{aligned}$$

where $\Delta_s = (p_1 - p_0)/\sqrt{p_0(1-p_0)}$. Now using the similar method we see that

$$E(S|R_s) - P(CS|R_s) \simeq \frac{(k-1)}{2} \Phi\left(\frac{\frac{d}{2\sqrt{p_0(1-p_0)}} - \sqrt{n}\Delta_s}{\sqrt{2}}\right).$$

Let $n(\epsilon)$ be the solution of the equation $\Phi(-n(\epsilon)) = \epsilon$. Substituting ϵ for $E(S|R_m) - P(CS|R_m)$ and for $E(S|R_s) - P(CS|R_s)$ we have

$$\begin{aligned}
n_m(\epsilon) &= (\sqrt{2}n(\epsilon') + d)^2/\Delta_m^2 \\
n_s(\epsilon) &= \left(\sqrt{2}n(\epsilon') + \frac{d}{2\sqrt{p_0(1-p_0)}}\right)^2/\Delta_s^2,
\end{aligned}$$

where $\epsilon' = 2\epsilon/(k-1)$. Since $n(\epsilon) \rightarrow \infty$ as $\epsilon \rightarrow 0$, letting $\epsilon \rightarrow 0$ we have

$$\begin{aligned}
ARE(R_m, R_s : \Delta) &= \lim_{\epsilon \rightarrow 0} \frac{n_s(\epsilon)}{n_m(\epsilon)} \\
&= \frac{\Delta_s^2}{\Delta_m^2} \\
&= \frac{\Delta^2 p_0(1-p_0)}{\sigma^2(p_1-p_0)^2}.
\end{aligned}$$

References

- [1] Barlow, R. E. and Gupta, S. S. (1969) "*Selection procedures for restricted families of probability distributions*" Annals of Mathematical Statistics 40, 905-917.
- [2] Bechhofer, R. E. (1954) "*A single sample multiple decision procedure for ranking means of normal populations with known variances*" Annals of Mathematical Statistics 25, 16-39.
- [3] Desu, M. M. and Bristol, D. R. (1986) "*Rizvi and Sobel type non parametric selection procedures : Review and Extensions*" AJMMS 6, 87-107.
- [4] Eaton, M. L. and Gleser L. J. (1989) "*Some results on convolutions and a statistical application*" Contributions to Probability and Statistics, Essays in Honor of Ingram Olkin, (Ed. Gleser et. al.), Springer-Verlag, 75-90.
- [5] Gupta, S. S. (1956) "*On a decision rule for a problem of ranking means*" Mimeograph series No. 150 , Institute of statistics, University of North Carolina, Chapel Hill.
- [6] Gupta, S. S. and Huang, D. Y. (1974) "*Nonparametric subset selection procedures for the 't' best populations*" Bulletin of Mathematics , Academia Sinica 2, 377-386.
- [7] Gupta, S. S. and Huang, D. Y., Huang, W. T. (1976) "*On ranking and selection procedures and tests of homogeneity for binomial populations*" Essays in probability and statistics (Eds. S. Ikeda et al.), Shinko Tsusho Co. Ltd., Tokyo, Japan.

- [8] Gupta, S. S. and McDonald, G. C. (1970) "*On classes of selection procedures based on ranks*" Nonparametric Techniques in statistical Inference , ed M. L. Puri , Cambridge University Press, London, 419-514.
- [9] Gupta, S. S. and Panchapakesan, S. (1986) "*Subset selection procedures : Review and Assessments*" JAMMS , 5, 235-311.
- [10] Gupta, S. S. and Panchapakesan (1979) "*Multiple Decision Procedures* " Wiley, New York.
- [11] Gupta, S. S. and Sobel, M. (1960) "*Selecting a subset containing the best of several binomial populations*" Contributions to Probability and Statistics (Eds. I, Olkin et al.) Stanford University Press, Stanford, California, 224-248.
- [12] Hall, W. J. (1958) "*Most economical multiple-decision rules*" Annals of Mathematical Statistics, 29, 1079-1094.
- [13] Rizvi, M. H. and Sobel, M. (1967) "*Nonparametric procedures for selecting a subset containing the population with largest α th quantile*" Annals of Mathematical Statistics, 38, 1788-1803.
- [14] Sobel, M. (1967) "*Nonparametric procedures for selecting the 't' populations with the largest α th quantile*" Annals of Mathematical Statistics, 38, 1804-1816.
- [15] Sobel, M. and Huyett, M. J. (1957) "*Selecting the best one of the several binomial populations*" Bell System Tech. J., 36, 537-576.