RANKING AND SELECTION PROCEDURES FOR LOGISTIC POPULATIONS*

by

S. Panchapakesan
Department of Mathematics
Southern Illinois University
Carbondale, Illinois 62901

Technical Report #90-28C

Department of Statistics
Purdue University

June 1990

* This research was supported in part by the NSF Grant DMS-8702620 at Purdue University. To appear as Chapter 6 in the Logistic Distribution (ed. N. Balakrishnan), Marcel Dekker, New York.

★

RANKING AND SELECTION PROCEDURES FOR LOGISTIC POPULATIONS*

S. Panchapakesan
Southern Illinois University, Carbondale, IL 62901-4408

ABSTRACT

In this paper, our main objective is two-fold, namely, (1) to review available results for logistic distributions, and (2) to provide a *selective overview* of ranking and selection methodology in order to serve as an introduction to the general reader and also to indicate the potential for further investigations in the logistic case. In Section 2, we discuss two basic formulations of ranking and selection problems; these are the subset formulation and the indifference zone formulation. We also mention some modifications and types of procedures relevant to subsequent discussions. The next four sections (3 through 6) deal with procedures for selecting the population with the largest mean from several logistic populations with a common known variance. Of these, all but Section 6 discuss single-stage procedures using the basic subset (Section 3), the indifference zone (Section 4), and the restricted subset (Section 5) formulations. Section 6 is concerned with a two-stage procedure using the indifference zone formulation in which the first stage involves a subset approach to eliminate inferior populations. Logistic distribution has been used to model quantal response in experiments involving quantitative treatment factors; a selection problem that arises in this context is discussed in Section 7. The next section describes procedures for selecting the population having the largest quantile of a given order from distributions that belong to a restricted family defined by tail-ordering with respect to a known distribution G, including special results for logistic G. Finally, we conclude with a brief discussion on future directions for investigations relating to logistic distributions.

Key Words and Phrases: Logistic distributions; subset selection; indifference zone formulation; restricted subset; single-stage and two-stage procedures; quantal response; tail-ordering.

*This research was supported in part by the NSF Grant DMS-8702620 at Purdue University.

1. INTRODUCTION

Problems of statistical inference that are now known as ranking and selection problems first came under systematic investigation by statistical researchers in the early 1950's. The classical techniques for testing homogeneity hypotheses were found inadequate to serve, in many practical situations, the experimenter's real purpose which is often to rank several competing populations (treatments, systems, etc.) or to select the best among them. The attempts to formulate the decision problem to answer such realistic goals set the stage for the development of the ranking and selection theory.

During the last forty years, the ranking and selection literature has steadily grown with developments dealing with various aspects of the theory and applications. An important part of these developments is the study of ranking and selection problems for specific parametric families of distributions including, of course, logistic distributions. It is interesting to note that until recently there has not been much done in the case of logistic distributions. In this paper, our main objective is two-fold, namely, (1) to review available results for logistic distributions, and (2) to provide a *selective overview* of ranking and selection methodology in order to serve as an introduction to the general reader and also to indicate the potential for further investigations in the logistic case.

In Section 2, we discuss two basic formulations of ranking and selection problems; these are the subset formulation and the indifference zone formulation. We also mention some modifications and types of procedures relevant to subsequent discussions. The next four sections (3 through 6) deal with procedures for selecting the population with the largest mean from several logistic populations with a common known variance. Of these, all but Section 6 discuss single-stage procedures using the basic subset (Section 3), the indifference zone (Section 4), and the restricted subset (Section 5) formulations. Section 6 is concerned with a two-stage procedure using the indifference zone formulation in which the first stage involves a subset approach to eliminate inferior populations. Logistic distribution has been used to model quantal response in experiments involving quantitative treatment factors; a selection problem that arises in this context is discussed in Section 7. The next section describes procedures for selecting the population having the largest quantile of a given order from distributions that belong to a restricted family defined by tail-ordering with respect to a known distribution G, including special results for logistic G. Finally, we conclude with a brief discussion on future directions for investigations relating to logistic distributions.

2. RANKING AND SELECTION FORMULATIONS

Ranking and selection problems have generally been studied by using either the *indifference zone* approach of Bechhofer (1954) or the so-called *subset selection* approach due mainly to Gupta (1956). In the former approach the number of populations to be selected is pre-determined, while in the latter it is random. Suppose there are k (≥ 2) populations $\pi_1, \pi_2, \ldots, \pi_k$, where π_i is characterized by the distribution function F_{θ_i} and θ_i is a real-valued parameter taking a value in the set Θ, $i = 1, 2, \ldots, k$. The θ_i are assumed to be unknown. Let $\theta_{[1]} \leq \theta_{[2]} \leq \ldots \leq \theta_{[k]}$ denote the ordered θ_i and $\pi_{(i)}$ denote the population associated with $\theta_{[i]}$, $i = 1, 2, \ldots, k$. The populations are ranked according to their θ-values. To be specific, we define $\pi_{(j)}$ to be better than $\pi_{(i)}$ if $i < j$, that is, if $\theta_{[i]} \leq \theta_{[j]}$. It is assumed that there is no prior information regarding the true pairing of the ordered and unordered θ_i.

Let us consider the basic problem of selecting the *best* population, namely, the one associated with the largest θ_i. In the indifference zone approach, the goal is to select *one* of the k populations and claim it to be the best. Let $\Omega = \{\underset{\sim}{\theta} | \underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k),\ \theta_i \in \Theta,$ $i = 1, \ldots, k\}$ be the parameter space and $\Omega_{\delta^*} = \{\underset{\sim}{\theta} | \delta(\theta_{[k]}, \theta_{[k-1]}) \geq \delta^* > 0\}$, where

$\delta(\theta_{[k]}, \theta_{[k-1]})$ is an appropriate measure of the separation between the best population $\pi_{(k)}$ and the next best $\pi_{(k-1)}$. A *correct selection* (CS) occurs whenever the selected population is the best population. Let $P(CS|R)$ denote the probability of a correct selection (PCS) using the rule R. For a rule R to be valid, it is required that

$$P(CS|R) \geq P^* \text{ whenever } \underset{\sim}{\theta} \in \Omega_{\delta^*}. \tag{2.1}$$

The constants δ^* and $P^*(1/k < P^* < 1)$ are specified in advance by the experimenter. The statistical problem is to define a selection rule which typically has three parts: sampling rule, stopping rule for sampling, and decision rule. For a rule based on a single sample of fixed size n from each population, the design aspect of the experiment is to determine the minimum sample size n for which (2.1) is satisfied. The region Ω_{δ^*} of the parameter space Ω is called the *preference zone*. The complement of Ω_{δ^*} is the *indifference zone* so called because there is no requirement on the PCS when $\underset{\sim}{\theta}$ lies in it.

In the subset selection approach for selecting the best population, the goal is to select a nonempty subset of the k populations so that the best population is included in the selected subset with a minimum guaranteed probability $P^*(1/k < P^* < 1)$. In other words, any valid rule R should satisfy the condition:

$$P(CS|R) \geq P^* \text{ for any } \underset{\sim}{\theta} \in \Omega. \tag{2.2}$$

Here, selection of any subset that includes the best population results in a correct selection. In case of a tie for the best population, it is assumed that one of the contenders is tagged as the best. It should be noted that there is no indifference zone in the above formulation. The size of the selected subset S, denoted by $|S|$, is not specified in advance, but is determined by the data themselves. The expected subset size $E(|S|)$ and the expected number of non-best populations (which is equal to $E(|S|) - PCS$) are natural performance characteristics of a valid rule.

The probability requirements (2.1) and (2.2) are known as the P^*-condition. An important step in obtaining constant(s) associated with a proposed rule R so that the P^*-condition is satisfied is to evaluate the infimum of the PCS over Ω or Ω_{δ^*} depending on the approach. The configuration of $\underset{\sim}{\theta}$ for which this infimum is attained is called the *least favorable configuration* (LFC).

There are several variations and generalizations of the basic goal in both indifference zone and subset selection formulations. These are discussed in detail in Gupta and Panchapakesan (1979). One such generalization to be discussed later in Section 5 is the *restricted subset selection* formulation of Santner (1975). The goal in this formulation is to select a nonempty subset of the k populations that contains the best but whose size does not exceed a specified number m ($1 \leq m \leq k$). It is required that the PCS be at least P^* whenever $\underset{\sim}{\theta} \in \Omega_{\delta^*}$. This formulation combines the features of the indifference zone and subset selection formulations discussed earlier.

Besides being a goal in itself, selecting a subset containing the best can serve as a first-stage screening in a two-stage procedure designed to select one population as the best. Tamhane and Bechhofer (1977, 1979) have employed this technique for selecting the population with the largest mean from k normal populations with unknown means and a common known variance σ^2, using the indifference zone approach. We will discuss (Section 6) a similar procedure for logistic populations. It is interesting to note that, when the common variance σ^2 is unknown in the above normal case, a two-stage procedure is necessary in order to meet the P^*-condition.

Families of distributions can be defined through partial ordering relation with respect to a known distribution. Such families have been called *restricted families* of distributions. Partial ordering such as convex ordering, star shape ordering and tail ordering have been considered in the literature. These families are of great importance in reliability theory.

Well-known families such as *IFR* (increasing failure rate) and *IFRA* (increasing failure rate average) distributions are examples of such families. Selection procedures for such restricted families have been considered by a few authors. A review of these and other procedures applicable to reliability models is given by Gupta and Panchapakesan (1988). We will discuss (Section 8) a procedure for selection from a family of distributions which are tail-ordered with respect to a logistic distribution.

As we have pointed out earlier, our objective here is to provide a selective overview of ranking and selection procedures. Several aspects of the theory and related problems have been dealt with in the books by Bechhofer, Kiefer and Sobel (1968), Büringer, Martin and Schriever (1980), Gibbons, Olkin and Sobel (1977), Gupta and Huang (1981), and Gupta and Panchapakesan (1979). The last authors have given a comprehensive survey of developments in the theory with an extensive bibliography. A categorical bibliography is provided by Dudewicz and Koo (1982). Recently, Gupta and Panchapakesan (1985) have given a review and assessment of subset selection procedures.

In the succeeding sections, we discuss specific selection procedures relating to logistic distributions.

3. SINGLE-STAGE (UNRESTRICTED) SUBSET SELECTION OF THE POPULATION WITH THE LARGEST MEAN

Let $\pi_1, \ldots, \pi_k$ be k (≥ 2) independent logistic populations $L(\mu_i, \sigma^2)$, $i = 1, 2, \ldots, k$, where the means μ_i are unknown and the common variance σ^2 is assumed to be known. The distribution function associated with π_i ($1 \leq i \leq k$) is:

$$F(x; \mu_i, \sigma) = \frac{1}{1 + exp\{-\pi(x - \mu_i)/\sigma\sqrt{3}\}}, \qquad -\infty < x < \infty. \tag{3.1}$$

We assume, without loss of generality, that $\sigma = 1$. The population with the largest μ_i is the best. Let $X_{i,m:n}$ denote the median of a random sample of odd size $n = 2m - 1$ from π_i, $i = 1, \ldots, k$. Lorenzen and McDonald (1981) proposed the rule

$$R_1 : \text{ Select } \pi_i \text{ if and only if } X_{i,m:n} \geq \max_{1 \leq j \leq k} X_{j,m:n} - d_1 \tag{3.2}$$

where $d_1 = d_1(k, m, P^*)$ is the smallest positive constant for which the P^*-condition is satisfied.

For convenience, let $\mu_{[1]} \leq \ldots \leq \mu_{[k]}$ denote the ordered μ_i (we continue with the notation of Section 2 for the ranked parameters). Let $X_{(i),m:n}$ denote the sample median from the population associated with $\mu_{[i]}$, $i = 1, \ldots, k$. Then the *PCS* for the rule R_1 is given by

$$P(CS|R_1 = Pr\{X_{(k),m:n} \geq X_{(j),m:n} - d_1, \quad j = 1, \ldots, k-1\}. \tag{3.3}$$

Letting $f_{m:n}(y_m)$ and $F_{m:n}(y_m)$ denote the *pdf* and the *cdf*, respectively, of the median of a random sample of size $n = 2m - 1$ from the standard logistic distribution $L(0, 1)$, we can write (3.3) as

$$P(CS|R_1) = \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} F(y_m + \mu_{[k]} - \mu_{[j]} + d_1) f_{m:n}(y_m) dy_m. \tag{3.4}$$

It is now easy to see that the infimum of the *PCS* over the parameter space Ω is attained when all the μ_i are equal. Thus the constant $d_1 = d_1(k, m, P^*)$ is given by

$$\int_{-\infty}^{\infty} F_{m:n}^{k-1}(y_m + d_1) f_{m:n}(y_m) dy_m = P^*. \tag{3.5}$$

Lorenzen and McDonald (1981) have tabulated the d_1-values for $k = 2(1)10$, $m = 1(1)10$, and $P^* = 0.75, 0.90, 0.95, 0.99$.

Alternatively, one can define the following procedure R_2 based on the sample means $\overline{X}_i,\ i = 1, \ldots, k$.

$$R_2: \text{ Select } \pi_i \text{ if and only if } \overline{X}_i \geq \max_{1 \leq j \leq k} \overline{X}_j - d_2 \tag{3.6}$$

where $d_2 = d_2(k, n, P^*)$ is the smallest positive constant for which the P^*-condition is satisfied. Let $g_n(u)$ and $G_n(u)$ denote the *pdf* and the *cdf*, respectively, of $U = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$ where $\overline{X}$ is the mean of a random sample of since n from $L(\mu, \sigma^2)$. Assuming still $\sigma = 1$, the PCS for the rule R_2 is given by

$$P(CS|R_2) = \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} G_n(u + \sqrt{n}d_2 + \sqrt{n}(\mu_{[k]} - \mu_{[j]}))g_n(u)du. \tag{3.7}$$

The infimum of the PCS over Ω is attained when the μ_i are equal and the constant $d_2 = d_2(k, n, P^*)$ is given by

$$\int_{-\infty}^{\infty} G_n^{k-1}(u + \sqrt{n}d_2)g_n(u)du = P^*. \tag{3.8}$$

Lorenzen and McDonald (1981) have considered this rule R_2 in order to study the efficiency of R_1 relative to R_2; however, they considered only asymptotic ($n \to \infty$) case using convergence to normality. Recently, Han (1987) has studied the rule R_2 and has provided tables of values of $h = \sqrt{n}d_2$ for $k = 2(1)10$, $n = 1(1)10$ and $P^* = 0.75, 0.90, 0.95, 0.975, 0.99$; he used the Edgeworth series expansions to the order $O(n^{-3})$ for $G_n(u)$ and $g_n(u)$, the Gauss-Hermite quadrature with sixty nodes for evaluation of the integral, and a modified regular falsi algorithm for solving non-linear equations.

For comparing the rules R_1 and R_2, Han (1987) considered (1) $E(|S|)$, the expected subset size, (2) $E(|S|) - PCS$, which is the expected number of non-best populations selected, and (3) $E(T)$, where T is the sum of the ranks of the selected populations. His tables include also the comparison of the expected proportion of populations included in the selected subset but this is only $E(|S|)/k$. In order to compare R_1 and R_2 in terms of the above performance characteristics, Han (1987) considered two customary types of configurations of the means: (1) the slippage configuration, $\mu_{[1]} = \ldots = \mu_{[k-1]} = \mu_{[k]} - \delta$, $\delta > 0$, and (2) the equally spaced configuration, $\mu_{[i]} = \mu + (i-1)\delta$, $i = 1, \ldots, k$; $\delta > 0$. His tables of values of the performance characteristics range over $k = 2(1)5, 10$; $n = 3$; $\delta\sqrt{n} = 0.5(0.5)3.0, 4, 5$ (for both configurations); and $P^* = 0.90$. His tables for the ratio of the corresponding performance characteristics of R_1 and R_2 cover $k = 4$; $n = 3, 5$; $\delta\sqrt{n} = 1.5, 3.0$; and $P^* = 0.90, 0.95$. The tables indicate as one would expect that the means procedure R_2 performs better than the medians procedure R_1, the advantage increasing with n.

Lorenzen and McDonald (1981) studied the efficiency of R_1 relative to R_2 for large samples under a slippage configuration described earlier. Let n_1 and n_2 denote the asymptotic sample sizes required by R_1 and R_2, respectively, to satisfy the P^*-condition and to make $E(|S|) - PCS = \epsilon > 0$. The asymptotic efficiency of R_2 relative to R_1 is defined by

$$ARE(R_2, R_1; \delta) = \lim_{\epsilon \downarrow 0}(n_2/n_1). \tag{3.9}$$

Lorenzen and McDonald (1981) have shown that $ARE(R_2, R_1; \delta) = \pi^2/12 \doteq 0.822$. Thus, under a slippage configuration, asymptotically the means procedure requires about 82% of the sample size required by the median procedure to achieve the same expected number of non-best populations in the selected subset. However, the situation can dramatically change in favor of the medians procedure, as Lorenzen and McDonald have shown, when sampling is contaminated in the sense that π_i is logistic with mean θ_i and variance $\alpha + (1-\alpha)\nu^2$. The savings gained by the medians procedure becomes immense as $\nu \to \infty$.

Lorenzen and McDonald (1981) have also compared the medians procedure R_1 with a rank-sum procedure R_3. Let T_i denote the sum of the ranks of the observations from π_i

in the pooled sample obtained from samples of size n from each population (the smallest observation is assigned rank 1 and the largest rank kn). The rule R_3 is defined as follows:

$$R_3: \text{ Select } \pi_i \text{ if and only if } T_i \geq \max_{1\leq j\leq k} T_j - d_3 \tag{3.10}$$

where d_3 is the smallest positive integer so that the P^*-condition is satisfied. This rule R_3 has been studied by Gupta and McDonald (1970) for location and scale parameter families. Lorenzen and McDonald (1981) have studied the asymptotic efficiency of R_3 relative to R_1 under slippage and equal spacing configurations, using Monte Carlo simulations when $k > 2$. Based on their study, the rank-sum procedure outperforms the median procedure when the means are roughly in a slippage configuration, while the reverse is true when the means are equally spaced.

Properties of Rules R_1 and R_2

The rules R_1 and R_2 possess properties considered desirable in a subset selection rule. We discuss these properties below.

(1) Unbiasedness. A rule R is said to be *unbiased* if for all $\underset{\sim}{\mu} \in \Omega$ and $j < k$,

$$P(\pi_{(k)} \text{ is selected by } R) \geq P(\pi_{(j)} \text{ is selected by } R).$$

(2) Monotonicity. A rule R is said to be *monotone* if for $\underset{\sim}{\mu} \in \Omega$ and $i < j$,

$$P(\pi_{(j)} \text{ is selected by } R) \geq P(\pi_{(i)} \text{ is selected by } R).$$

Obviously, monotonicity implies unbiasedness.

(3) Strong Monotonicity. A rule R is said to be *strongly monotone* in $\pi_{(i)}$ if $P(\pi_{(i)}$ is selected by $R)$ is increasing in $\mu_{[i]}$ when all other components of $\underset{\sim}{\mu}$ are fixed and is decreasing in $\mu_{[j]}$ $(j \neq i)$ when all other components of $\underset{\sim}{\mu}$ are fixed.

(4) Consistency. A rule $R(n)$ [based on common sample size n] is said to be *consistent* with respect to $\Omega' \subset \Omega$ if $\inf_{\Omega'} P(CS|R(n)) \to 1$ as $n \to \infty$.

Finally, for the rules R_1 and R_2, the supremum of $E(|S|)$ over Ω is attained when the μ_i are equal. This follows (see Gupta (1965) and Gupta and Panchapakesan (1972)) from the fact that μ_i is a location parameter in the distributions of $X_{i,m:n}$ and $\overline{X}_i$ $(1 \leq i \leq k)$, and that these distributions have the monotone likelihood ratio (MLR) property. Consequently, this supremum is equal to kP^*.

4. SINGLE-STAGE INDIFFERENCE ZONE SELECTION OF THE POPULATION WITH THE LARGEST MEAN

We have k logistic distributions as described in Section 3. Under the indifference zone formulation we select one of the k populations so that the $PCS \geq P^*$ whenever $\underset{\sim}{\mu} \in \Omega_{\delta^*} = \{\underset{\sim}{\mu} : \mu_{[k]} - \mu_{[k-1]} \geq \delta^* > 0\}$. We will consider two procedures R_1' and R_2' which are, respectively, the counterparts of R_1 and R_2 discussed in the previous section. The sampling schemes are same as earlier.

First, we define R_1' based on medians as follows.

$$R_1' : \text{ Select the population that yields the largest } X_{i,m:n}. \tag{4.1}$$

It is easy to see that

$$P(CS|R_1') = \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} F_{m:n}(y_m + \mu_{[k]} - \mu_{[j]}) f_{m:n}(y_m) dy_m. \tag{4.2}$$

Obviously, the infimum of $P(CS|R_1')$ over Ω_{δ^*} is attained when $\mu_{[1]} = \ldots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$. Thus we need to determine the minimum odd sample size n for which

$$\int_{-\infty}^{\infty} F_{m:n}^{k-1}(y_m + \delta^*) f_{m:n}(y_m) dy_m \geq P^*. \tag{4.3}$$

A table of n values satisfying (4.3) for selected values of k, δ^* and P^* is not available.

Alternatively, we can define a rule based on sample means, namely,

$$R_2' : \text{ Select the population that yields the largest } \overline{X}_i. \tag{4.4}$$

In this case,

$$P(CS|R_2') = \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} G_n(u + \sqrt{n}(\mu_{[k]} - \mu_{[j]}))g_n(u)du \tag{4.5}$$

and the infimum of the PCS in (4.5) is attained when $\mu_{[1]} = \ldots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$. Thus the minimum sample size n required to meet the P^*-condition is given by

$$\int_{-\infty}^{\infty} G_n^{k-1}(u + \sqrt{n}\delta^*)g_n(u)du \geq P^*. \tag{4.6}$$

Han (1987) has studied the rule R_2' and has tabulated, using the Edgeworth series expansions to the order of $O(n^{-3})$ for $G_n(u)$ and $g_n(u)$, the values of $\hat{n}$ for which (4.6) is satisfied with equality of both sides. The required values of n are obtained by rounding up $\hat{n}$ to the next higher integer. His table ranges over $k = 2(1)5, 10, 15$; $\delta^* = 0.1, 0.5, 1.0, 2.0, 4.0$; and $P^* = 0.75, 0.90, 0.95, 0.99$.

One should naturally compare the minimum sample sizes required by R_1' and R_2', and also compare the performance of the two procedures for various typical parametric configurations. This has not yet been done.

Recently, van der Laan (1989) has considered the rule R_2' with $n = 1$, in which case G_n in (4.6) is the standardized logistic c.d.f. He has tabulated the values of δ^* for $k = 2(1)10$, 25, 50 and $P^* = 0.65$, 0.75, 0.90, 0.95, 0.99, 0.999 and compared these with the corresponding values in the normal case available in Bechhofer (1954). However, it should be noted that, in the case of $n = 1$, the equations (4.6) and (3.5) are the same. So the values of δ^* for $k = 2(1)10$, and $P^* = 0.75$, 0.90, 0.95, 0.99 are readily available in Lorenzen and McDonald (1981).

5. SINGLE-STAGE RESTRICTED SUBSET SELECTION OF THE POPULATION WITH THE LARGEST MEAN

In Section 2, we referred to the restricted subset selection formulation of Santner (1975). The goal here is to select a subset of the k populations whose size does not exceed m ($1 \leq m \leq k-1$) and which includes the population with largest mean μ_i. As we mentioned earlier, in this formulation, we introduce a preference zone $\Omega_{\delta^*} = \{\underset{\sim}{\mu} | \mu_{[k]} - \mu_{[k-1]} \geq \delta^* > 0\}$. We want to define a rule R for which

$$P(CS|R) \geq P^* \text{ whenever } \underset{\sim}{\mu} \in \Omega_{\delta^*}. \tag{5.1}$$

We note that for $m = 1$, this becomes the indifference zone formulation of Bechhofer (1954). If we allow $m = k$ (unrestricted) and $\delta^* = 0$, we will get the usual subset selection formulation.

Now, for the problem at hand, Han (1987) investigated the following rule R_4 based on the sample means $\overline{X}_i$.

$$R_4: \text{ Select } \pi_i \text{ if and only if } \overline{X}_i \geq \max\left\{\overline{X}_{k-m+1:k}, \overline{X}_{k:k} - \frac{d_4\sigma}{\sqrt{n}}\right\} \tag{5.2}$$

where the minimum required sample size n and $d_4 = d_4(k, n, P^*)$ are to be determined so that (5.1) is satisfied. We leave out the expression for $P(CS|R_4)$ for any $\underset{\sim}{\mu} \in \Omega$ in order to avoid further notations. It can be shown without much difficulty [see Han (1987)] that the infimum of $P(CS|R_4)$ over Ω_{δ^*} is attained when $\underset{\sim}{\mu} \in \Omega^0_{\delta^*} = \{\underset{\sim}{\mu} | \mu_{[1]} = \ldots = \mu_{[k-1]} = \mu_{[k]} - \delta^*\}$. Also, $P(CS|R_4)$ is constant over $\Omega^0_{\delta^*}$. Assuming that the common $\sigma = 1$,

$$\inf_{\underset{\sim}{\mu} \in \Omega_{\delta^*}} P(CS|R_4) = \sum_{i=k-m}^{k} \binom{k-1}{i} \int_{-\infty}^{\infty} G_n^i(t + \delta\sqrt{n})\{G_n(t + d_4 + \delta^*\sqrt{n}) - G_n(t + \delta^*\sqrt{n})\}^{k-i-1} g_n(t)dt. \tag{5.3}$$

The minimum sample size n required to satisfy the P^*-condition in (5.1) has been tabulated by Hahn (1987) for $P^* = 0.90$; $k = 5$ with $m = 2(1)4$; $k = 10$ with $m = 2(1)5$; $d_4 = 0.4, 0.7, 1.3, 1.6$ and $\delta = 0.5, 1.0, 2.0$ [the range is for δ/σ if σ is not unity]. Han (1987) has further tables regarding some performance characteristics which are not discussed here.

The rule R_4 is strongly monotone in $\pi_{(i)}$ and so is monotone and unbiased. Han (1987) has shown that R_4 is consistent with respect to Ω_{δ^*}. This implies that, for given P^*, k, m, and δ^* one can meet the P^*-condition by choosing n sufficiently large. Also, not surprisingly, it turns out that, for given δ^*, k, m, and n, the P^*-condition can be met only for $P^* \leq P_1^* = P_1^*(\delta^*, k, m, n) < 1$. The supremum of $E(|S|)$ over Ω is attained when the μ_i are equal. Since we have two constants n and d_4 defining the rule R_4, these can be chosen such that $\inf_{\Omega_{\delta^*}} P(CS|R_4) \geq P^*$ and $\sup_{\Omega} E(|S||R_4) = 1 + \epsilon$ for a specified ϵ. Han (1987) has some tables for choosing n and d_4 for $\epsilon = 0.01$ and selected values of k, m, P^*, and δ/σ.

6. AN ELIMINATION TYPE TWO-STAGE SELECTION PROCEDURE FOR THE POPULATION WITH THE LARGEST MEAN

As before, we assume that the k logistic populations are $L(\mu_i, \sigma^2)$, $i = 1, \ldots, k$, where the common variance σ^2 is known. We want to select the population with the largest μ_i using the indifference zone formulation. As we mentioned in Section 2, we can use a two-stage procedure which screens out bad populations (those with small values of μ_i) by means of a subset selection rule. Such procedures for normal populations with common known variance were studied initially by Cohen (1959) and Alam (1970) whose results were mostly for the case of $k = 2$ populations. Tamhane and Bechhofer (1977, 1979) have studied the problem in depth for $k \geq 2$. Recently, Gupta and Han (1990) have investigated a similar procedure for logistic populations. This procedure R_5 involves the design constants (n_1, n_2, h), where h is a positive constant, and n_1 and n_2 are the sample sizes in the two stages. These constants, depending on k, δ^* and P^*, are chosen so that the P^*-condition is satisfied and they possess a certain minimax property.

<u>Procedure R_5</u>: At Stage 1, take n_1 independent observations from each of the k populations and compute the sample means $\overline{X}_i^{(1)}$, $i = 1, \ldots, k$. Determine the subset of I of $\{1, \ldots, k\}$ where

$$I = \{i | \overline{X}_i^{(1)} \geq \max_{1 \leq j \leq k} \overline{X}_j^{(1)} - h\sigma/\sqrt{n_1}\}. \tag{6.1}$$

Let $\Pi_I = \{\pi_i | i \in I\}$. If Π_I consists of only one population, stop sampling and select as best the population that yielded the largest $\overline{X}_i^{(1)}$. If Π_I consists of more than one population, proceed to Stage 2.

At Stage 2, take n_2 additional observations from each population in Π_I and compute the cumulative sample means $\overline{X}_i$ based on $(n_1 + n_2)$ observations. Select as the best the population that yielded the largest $\overline{X}_i$.

There are an infinite number of choices of (n_1, n_2, h) for which the P^*-condition (5.1) is satisfied. Let $|I|$ denote the cardinality of the set I in (6.1) and

$$S = \begin{cases} 0 & \text{if } |I| = 1 \\ |I| & \text{if } |I| > 1. \end{cases} \tag{6.2}$$

Then the total sample size required is given by

$$T = kn_1 + Sn_2. \tag{6.3}$$

Gupta and Han (1990) have adopted an unrestricted minimax criterion to make a choice of (n_1, n_2, h). In other words, for given k, δ^* and P^*, in order to choose (n_1, n_2, h)

$$\text{minimize } \sup_{\underset{\sim}{\mu} \in \Omega} E(T|R_5) \text{ subject to } \inf_{\underset{\sim}{\mu} \in \Omega_{\delta^*}} P(CS|R_5) \geq P^*. \tag{6.4}$$

Gupta and Han (1990) have shown that the supremum of $E(T|R_5)$ over Ω is attained when the μ_i are equal. The LFC for the PCS is a slippage configuration; this follows from the result of Bhandari and Chaudhuri (1988) for two-stage selection for the largest population mean when the sample mean has the MLR property. However, the exact evaluation of the PCS under the LFC for the rule R_5 is complicated. Gupta and Han (1990) have considered a lower bound to $P(CS|R_5)$ for which the infimum over Ω_{δ^*} is easily obtained. Using this, a conservative solution can be obtained for the minimization problem in (6.4). In other words, a conservative solution for (n_1, n_2, h) is obtained by minimizing

$$kn_1 + n_2 \int_{-\infty}^{\infty} \left[G_{n_1}^{k-1}(x+h) - G_{n_1}^{k-1}(x-h)\right] g_{n_1}(x)dx \tag{6.5}$$

subject to

$$\int_{-\infty}^{\infty} G_{n_1}^{k-1}(x+\delta^*\sqrt{n_1}/\sigma+h)g_{n_1}(x)dx \times \int_{-\infty}^{\infty} G_{n_1+n_2}^{k-1}(x+\delta^*\sqrt{n_1+n_2}/\sigma)g_{n_1+n_2}(x)dx \geq P^* \tag{6.6}$$

where $G_n(x)$ and $g_n(x)$ are so defined earlier in the case of the rule R_2 in Section 3. Let $(\hat{n}_1, \hat{n}_2, \hat{h})$ be the solution to the minimization of (6.5) subject to (6.6) when n_1 and n_2 are allowed to be continuous. Then one can use the approximate design constants

$$n_1 = [\hat{n}_1 + 1], \qquad n_2 = [\hat{n}_2 + 1], \qquad h = \hat{h},$$

where $[m]$ denotes the largest integer $\leq m$. Gupta and Han (1990) have tabulated $(\hat{n}_1, \hat{n}_2, \hat{h})$ and $E(T|R_5)$ for $k = 2(1)5, 10, 15$; $P^* = 0.75, 0.90, 0.95, 0.99$ and $\delta^*/\sigma = 0.1, 0.5, 1.0, 2.0, 4.0$; here, the tabulated values of $E(T|R_5)$ are the values of the expression in (6.5).

The performance of the two-stage procedure R_5 can be compared with that of the single-stage procedure R_2' defined by (4.4). Let $\hat{n}$ be the solution of (4.6) with equality. Gupta and Han (1990) considered $E(T|R_5)/k\hat{n}$ as the measure of relative efficiency

(RE). If $RE < 1$, then R_5 is better than R_2'. They have tabulated the values of RE for $k = 2(1)5, 10, 15$ and $P^* = 0.75, 0.90, 0.95, 0.99$ in the cases of the slippage configuration $(\mu, \ldots, \mu, \mu + \delta)$ and the equally spaced configuration $(\mu, \mu + \delta, \ldots, \mu + (k-1)\delta)$ when $\delta/\sigma = 0.1, 0.5, 1.0, 2.0, 4.0$. All the tabulated values of RE are equal to 1 in a very few cases and less than 1 otherwise, thus showing R_5 to be more efficient. The effectiveness of R_5 increases as k increases.

The rule R_5 employs the usual subset selection procedure for eliminating bad populations at the first stage. The size of the selected subset I defined in (6.1) can be k. One can use a restricted subset selection procedure of the type discussed in Section 5 in order to restrict the size of I. Such a two-stage procedure has been investigated by Han (1987) but it will not be discussed here.

7. SELECTION OF THE LOGISTIC QUANTAL RESPONSE WITH THE SMALLEST q-QUANTILE (ED100q)

Consider an experiment in which the treatment factor is quantitative. Each experimental unit is administered a certain "dose" of the treatment to which the unit either responds (a success) or does not respond (a failure). Such experiments are well-known in biological applications as quantal response assays or sensitivity experiments. The probability of response is some unknown function of the dose level x and, denoted by $p(x)$, is called the *quantal response curve.* It is reasonable, in many applications, to assume that $p(x)$ is nondecreasing in x, right-continuous with $p(-\infty) = 0$ and $p(\infty) = 1$. The smallest dose level that induces a response with probability q $(0 < q < 1)$ is called the *100q percent effective dose* (ED100q) and is denoted here by $\mu^{(q)}$.

In a selection problem, we are comparing several different quantitative treatments with unknown associated quantal response curves in order to select the best (to be suitably defined) curve. In a nonparametric set-up, this problem involves subtle difficulties. These have been discussed by Tamhane (1986) who have given an excellent survey of the literature on quantal response curves. He has discussed problems of estimation and multiple comparisons with a view toward the application to the selection problem. Towards this end, he has also considered two parametric models of which the logistic model is one. We discuss below his formulation of and solution to the problem of selecting the best quantal response curve.

Let $\pi_1, \ldots, \pi_k$ be k populations where π_i has the associated quantal response curve $p_i(x)$ given by

$$p_i(x) = \frac{1}{1 + exp\{-(\alpha_i + \beta x)\}}, \quad i = 1, \ldots, k, \tag{7.1}$$

and the common value of β is assumed to be *known.* The $ED100q$ of π_i is given by

$$\mu_i^{(q)} = \left[\log(\frac{q}{1-q}) - \alpha_i\right] / \beta. \tag{7.2}$$

The quantity $\log\{q/(1-q)\}$ is referred to as the *logit transform* of q. The goal is to select the population associated with the smallest $\mu_i^{(q)}$. This goal is meaningful when we have several drugs available for a certain ailment and we want to select the drug that induces a specified success rate q at the lowest dose level. Since all populations have a common β, the problem is equivalent to selecting the population associated with the largest α_i. Consistent with our earlier notations, the ordered α_i are denoted by $\alpha_{[1]} \leq \alpha_{[2]} \leq \ldots \leq \alpha_{[k]}$. Tamhane (1986) has adopted the indifference zone approach taking $\Omega_{\delta^*} = \{\underset{\sim}{\alpha} = (\alpha_1, \ldots, \alpha_k) | \alpha_{[k]} - \alpha_{[k-1]} \geq \beta\delta^*\}$. The P^*-condition is to be satisfied whenever $\underset{\sim}{\alpha} \in \Omega_{\delta^*}$.

Tamhane (1986) proposed a single-stage procedure based on n independent observations from each population, those from π_i being taken at equispaced dose levels x_{i1}, x_{i2},

$\ldots, x_{im}$ with $x_{i,j+1} - x_{ij} = d_i$ $(1 \le i \le k,\ 1 \le j \le m-1)$. Let r_{ij} denote the number of successes for population π_i at does level x_{ij} and let $\hat{p}_{ij} = r_{ij}/n$. The maximum likelihood estimator (MLE) $\hat{\alpha}_i$ of α_i $(1 \le i \le k)$ is obtained by solving $\sum_{j=1}^{m} p_{ij} = \sum_{j=1}^{m} \hat{p}_{ij}$, where $p_{ij} = p_i(x_{ij})$ given by (7.1). The rule proposed by Tamhane (1986) is

$$R_6: \text{ Select the population that yields the largest } \hat{\alpha}_i. \tag{7.3}$$

An exact solution for the minimum sample size n required in order to meet the P^*-condition is not available. Tamhane (1986) has obtained a large sample ($m \to \infty$, $n \to \infty$) solution using the fact that, when $d_1 = \ldots = d_k = d$, $\hat{\alpha}_i$ has asymptotically a normal distribution with mean α_i and variance $\beta d/n$. Based on the known results of Bechhofer (1954) for the problem of selecting the largest normal mean when the populations have a common known variance, the solution is given by

$$n = \left\langle \left(\frac{c}{\delta^*}\right)\frac{d}{\beta} \right\rangle \tag{7.4}$$

where $\langle a \rangle$ denotes the smallest integer $\ge a$ and $c = c(k, P^*)$ is given by

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

and Φ is the standard normal $c.d.f.$ The values of c (or a known multiple of it) have been tabulated by Bechhofer (1954), Gupta (1963), Milton (1963), and Gupta, Nagel and Panchapakesan (1973).

Tamhane (1986) has also considered a weighted least squares estimator α_i^* of α_i to be used in the rule R_6 in the place of $\hat{\alpha}_i$. However, in this case also the exact solution is not available. Asymptotically ($m \to \infty$ $n \to \infty$), α_i^* has the same distribution as $\hat{\alpha}_i$ and so a large sample solution is again given by (7.4).

Under the logistic model (7.1), the quantal response curves $p_i(x)$ for different α_i's do not intersect. In this case, one can take observations from all populations at the same dose level x_0. The populations can then be considered to be Bernoulli with success probabilities $p_i(x_0)$, $i = 1, \ldots, k$. The selection problem then reduces to the selection of the Bernoulli population associated with the largest success probability. Many procedures are available in the literature for this classical Bernoulli selection problem. A review of these procedures and a complete bibliography have been given by Bechhofer and Kulkarni (1982), who have themselves made significant contributions to this problem. However, the main obstacle in using any of these procedures to our problem at hand is, as pointed out by Tamhane (1986), the specific choice of x_0. With a hapless choice of x_0, it may turn out that $\mu_{[2]}^{(q)} - \mu_{[1]}^{(q)} < \delta^*$, i.e. $\underset{\sim}{\alpha} \notin \Omega_{\delta^*}$, thus making it impossible to satisfy the P^*-condition.

If we adopt the subset selection approach, then the problem of the specific choice of x_0 discussed above will not arise. In this case one can use the procedures available for the classical Bernoulli subset selection. For details of such procedures, the reader is referred to Gupta and Panchapakesan (1979, Chapter 13).

8. SELECTION FROM A FAMILY OF DISTRIBUTIONS PARTIALLY ORDERED WITH RESPECT TO A LOGISTIC DISTRIBUTION

As we mentioned earlier in Section 2, selection procedures for families of distributions which are partially ordered with respect to a known distribution have been considered in the literature. These procedures are of importance in reliability contexts and a review of these is provided by Gupta and Panchapakesan (1988).

Barlow and Gupta (1969) considered among other things the selection of the population with the largest median (assumed to be stochastically larger than the other populations) from a set of continuous distributions F_i, $i = 1, \ldots, k$, which have *lighter tails* than a specified continuous distribution G with $G(0) = 1/2$. The F_i and G are assumed to have the real line as their support. The definition of F_i having a lighter tail than G used by

Barlow and Gupta (1969) implies that F_i centered at its median Δ_i is *tail-ordered* with respect to G; in other words, $G^{-1}F_i(x + \Delta_i) - x$ is nondecreasing in x. The procedure of Barlow and Gupta (1969) has been shown by Gupta and Panchapakesan (1974) to work for this wider class defined by tail-ordering. In fact, Gupta and Panchapakesan have also shown a generalized version of this by considering tail-ordering of F_i and G when both are centered at their respective α-quantiles. Formally stated, for $0 < \alpha < 1$, F_i is said to be *α-quantile tail-ordered* with respect to G ($F_i \underset{t_\alpha}{<} G$) if $G^{-1}F_i(x + \xi_{i\alpha}) - x - \eta_\alpha$ is nondecreasing in x on the support of F_i, where $\xi_{i\alpha}$ and η_α are the (unique) α-quantiles of F_i and G, respectively. It can be shown (Gupta and Panchapakesan (1974), Lemma 3.1) that

$$P(a + \eta_\alpha \le Y \le b + \eta_\alpha) \le P(a + \xi_{i\alpha} \le X_i \le b + \xi_{i\alpha}) \tag{8.1}$$

for every $a < 0 < b$, where X_i and Y have distributions F_i and G, respectively.

Now, for the discussion of the selection problem, let $\pi_1, \dots, \pi_k$ be k populations with associated absolutely continuous distributions F_i having unique α-quantile $\xi_{\alpha i}$, $i = 1, \dots, k$, for a specified $0 < \alpha < 1$. Let G be a specified absolutely continuous distribution G with α-quantile η_α. We assume that the F_i are α-quantile tail-ordered with respect to G. It is also assumed that there is one population among the F_i that is stochastically larger than the remaining ones; consequently, this particular F_i, denoted by $F_{(k)}$, will have the largest α-quantile and is the best population. Our goal is to select the best population using the subset approach.

Let $X_{i1}, \dots, X_{in}$ be n independent observations from π_i and $X_{i,j:n}$ denote the jth order statistic of the sample from π_i, $i = 1, \dots, k$, where $j \le (n+1)\alpha < j + 1$. Then, for selecting the population with the largest α-quantile, Gupta and Panchapakesan (1974) proposed the following rule.

$$R_7 : \text{ Select } \pi_i \text{ if and only if } X_{i,j:n} \ge \max_{1 \le r \le k} x_{r,j:n} - D \tag{8.2}$$

where $D = D(k, P^*, n, j)$ is the smallest positive integer for which the P^*-condition is satisfied whatever the k-tuple $\{F_1, \ldots, F_k\}$ be.

Since $F_{(k)}$ is stochastically larger than any other F_i

$$\inf_{\Omega} P(CS|R_7) = \int_{-\infty}^{\infty} F_{(k),j:n}^{k-1}(t + D) f_{(k),j:n}(t)dt \tag{8.3}$$

where $f_{(k),j:n}$ is the density associated with $F_{(k),j:n}$. Now, since the partial ordering $\underset{t_\alpha}{<}$ is preserved by the order statistics, we have $F_{(k),j:n} \underset{t_\alpha}{<} G_{j:n}$, where $G_{j:n}$ is the distribution of the jth order statistic in a random sample of size n from G. As a consequence of this,

$$\int_{-\infty}^{\infty} F_{(k),j:n}^{k-1}(t + D) f_{(k),j:n}(t)dt \geq \int_{-\infty}^{\infty} G_{j:n}^{k-1}(t + D) g_{j:n}(t)dt \tag{8.4}$$

where $g_{j:n}$ is the density associated with $G_{j:n}$. Thus the constant $D = D(k, P^*, n, j)$ satisfying the P^*-condition is given by

$$\int_{-\infty}^{\infty} G_{j:n}^{k-1}(t + D) g_{j:n}(t)dt = P^*. \tag{8.5}$$

The values of D have been tabulated by Gupta and Panchapakesan (1974) for $k = 2(1)10$, $n = 5(2)15$, $j = 1(1)n$ and $P^* = 0.75, 0.90, 0.95, 0.99$ when G is chosen to be the logistic distribution $F^*(z) = [1 + \exp(-z)]^{-1}$, $-\infty < z < \infty$, which is $L(0, \pi^2/3)$. In the particular case of the median (i.e. $\alpha = 1/2$ and $j = (n+1)/2$), the equation (8.5) is same as (3.5) for the rule R_1 of Lorenzen and McDonald (1981) in Section 3, except that we now have $F_{j:n}^*$ in the place of $F_{j:n}$ (which is $L(0,1)$). Thus the constants d_1 and D of the rules R_1 and R_7 are related by: $D_1 = d_1\pi/\sqrt{3}$.

The infimum of $P(CS|R_7)$ in (8.3) can be evaluated asymptotically as $n \to \infty$ with $j/n \to \alpha$ using the asymptotic normality of the sample quantile (under the assumptions that the densities $f_{(k)}$ and g are differentiable in the neighborhood of their respective α-quantiles and that the densities do not vanish at their α-quantiles). Then, corresponding to (8.5), we get

$$\int_{-\infty}^{\infty} \Phi^{k-1}[x + Dg(\eta_\alpha)\{n/\alpha(1-\alpha)\}^{1/2}]\phi(x)dx = P^* \tag{8.6}$$

where Φ and ϕ denote the standard normal *c.d.f.* and density.

When G is taken to be $L(0, \pi^2/3)$, $g(\eta_\alpha) = \alpha(1-\alpha)$ and an approximate value of D is given by

$$\int_{-\infty}^{\infty} \Phi^{k-1}[x + D\{n\alpha(1-\alpha)\}^{1/2}]\phi(x)dx = P^*. \tag{8.7}$$

The D-value satisfying (8.7) can be obtained from the tables of Gupta, Nagel and Panchapakesan (1973); in other words, $D = H[2/n\alpha(1-\alpha)]^{1/2}$, where H is the table value corresponding to $\rho = 0.5$.

It is relevant to note that the left-hand side of (8.5) can be written as $P[Y_k \geq \max_{1 \leq r \leq k} Y_r - D]$, where the Y_i are *i.i.d.* having the distribution function $G_{j:n}$. Further, since $D > 0$, it can be written as $P[\max_{1 \leq r \leq k-1} (Y_r - Y_k) \leq D]$. Thus D given by (8.5) is the $100P^*\%$ point of the distribution of the maximum of the correlated differences $Y_i - Y_k$, $i = 1, \ldots, k-1$. A similar probability of interest is $P[\max_{1 \leq r \leq k} \frac{W_r}{W_k} \leq a]$, $a \geq 1$, where the W_i are *i.i.d.* having the distribution of the *jth* order statistic in a random sample of size n from a distribution of G of a continuous *non-negative* random variable. Such a probability concerning the maximum of correlated ratios arises in the problem of selecting in terms of α-quantiles from k distributions F_i of non-negative random variables which are *star-shaped* with respect to G. This has been investigated by Barlow and Gupta (1969). When G is exponential, the F_i become *IFRA* distributions and the values of $c = a^{-1}$ have been tabulated by Barlow, Gupta and Panchapakesan (1969). When G is the half-normal, the F_i belong to a subclass of *IFRA* distributions because the half-normal is *IFRA* (actually it is *IFR* which implies $IFRA$) and the star-ordering is transitive. In this case, the values of $c = a^{-1}$ have been tabulated by Gupta and Panchapakesan (1975). Another application of immediate interest is to consider the case where G is the half-logistic distribution, i.e. $G(x) = \{2/(1+\exp(-x))\} - 1,\ x \geq 0$. Since G is IFR, it is also $IFRA$. Thus our F_i's belong to yet another subclass of $IFRA$ distributions. In this case, tables of a are not available.

9. CONCLUDING REMARKS

In our selective overview of the ranking and selection theory in Section 2, we have confined ourselves to the basic formulations and some modifications that were relevant to the review of the procedures that have been investigated in the case of logistic populations. Our comments in the present section are meant to indicate the scope for further investigations.

For the problem of selecting the population with the largest mean μ_i using single-stage samples, we have assumed the variances to be *equal and known* and considered *equal sample sizes*. Procedures based on unequal sample sizes are not trivial extensions. Further, the case of unknown variances (known to be equal or not) are important in practice. Also, as we have mentioned earlier, when the variances are unknown (even if they are known to be equal), a single-stage procedure that guarantees the P^*-condition does not exist under the indifference zone formulation. These problems involve questions relating to distribution theory, determination of the LFC, and computations for implementation.

There are many aspects of the ranking problem that have been studied in the literature such as selecting good populations (which are "close" to the best), selecting populations that are better than a standard or a control, estimation of the actual PCS, estimation after selection (such as estimating the mean of the selected population), and confidence interval for the difference between the selected mean and the largest mean. For selecting the populations better than a standard or a control, we may have prior knowledge of the ordering of the experimental populations even though we may not know the values of the parameter of interest. In this case, the procedures have to exhibit an *isotonic* behavior. Some of the above aspects are of current interest relative to selection problems in general. For discussions of these developments and related references, see Gupta and Panchapakesan (1979, 1985). A few additional recent references that might be of interest are Gupta and Liang (1987, 1990), Gupta and Sohn (1990), Gupta and Panchapakesan (1990), and Liang and Panchapakesan (1990).

Gupta and Sohn (1990) have considered subset selection from Tukey's (symmetric) lambda distributions in terms of their location parameters, assuming that they all have *known* common scale and shape parameters. The lambda family of distributions was suggested by Tukey (1960, 1962) as a wide class of symmetric distributions. Later, Ramberg and Schmeiser (1972, 1974) generalized this family so as to include both symmetric and asymmetric distributions. The practical usefulness of this family is highlighted by the fact that it has a simple form for the inverse of the *c.d.f.* and it can be used to approximate a wide class of densities ranging from the uniform to very heavy tailed ones, of course, *including the logistic* [see also Joiner and Rosenblatt (1971)]. Gupta and Sohn (1990) investigated a subset selection rule based on sample medians similar to the rule R_1 in Section 3. Based on approximation to the logistic distribution by proper choices of the scale and shape parameters, they have calculated the constants for the rule R_1 of Lorenzen and McDonald (1981) when $k = 2, 5, 7$; $m = 2, 5, 7, 9$; and $P^* = 0.90, 0.95$. In 13 out of the 24 cases, the d-values, corrected to three decimal places, agree. In the remaining cases, the approximate value using the lambda distribution is one unit more in the third place. A similar comparison can be made in the case of the rule R_2 of Han (1987) described in

Section 3. Sohn (1985) has verified computationally that the approximation to the *c.d.f.* of the sample mean of a logistic distribution by using the lambda distribution is as good as that of Goel (1975) who derived the *c.d.f.* as a series by the method of characteristic functions.

Finally, the use of lambda distribution in approximating a wide class of densities has significant implications in developing versatile software packages for selection and ranking problems. Further, this aspect can be useful in other inference problems related to the logistic distribution.

REFERENCES

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

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

BARLOW, R.E. and 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.

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 (with special reference to Koopman-Darmois populations)*, The University of Chicago Press, Chicago and London.

BECHHOFER, R.E. and KULKARNI, R. (1982). Closed adaptive sequential procedures for selecting the best of $k \geq 2$ Bernoulli populations, In *Statistical Decision Theory and Related Topics-III* (S.S. Gupta and J.O. Berger, eds.) *1*, Academic Press, New York, 61-108.

BHANDARI, S.K. and CHAUDHURI, A.R. (1988). On two conjectures about two-stage selection problem, *Sankhyā Ser. B*.

BÜRINGER, H., MARTIN, H. and SCHRIEVER, K.-H. (1980). *Nonparametric Sequential Procedures*, Birkhauser, Boston.

COHEN, D.S. (1959). A Two-Sample Decision Procedure for Ranking Means of Normal Populations with a Common Known Variance, M.S. Thesis, Dept. of Operations Research, Cornell University, Ithaca, New York.

DUDEWICZ, E.J. and KOO, J.O. (1982). *The Complete Categorized Guide to Statistical Selection and Ranking Procedures*, Series in Mathematical and Management Sciences, Vol. 6, American Sciences Press, Columbus, Ohio.

GIBBONS, J.D., OLKIN, I. and SOBEL, M. (1977). *Selecting and Ordering Populations: A New Statistical Methodology*, John Wiley & Sons, New York.

GOEL, P.K. (1975). On the distribution of standardized mean of samples from the logistic population. *Sankhyā Ser. B 37*, 165-172.

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, North Carolina.

GUPTA, S.S. (1963). Probability integrals of the multivariate normal and multivariate t, *Ann. Math. Statist. 34*, 792-828.

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

GUPTA, S.S. and HAN, S. (1990). An elimination type two-stage procedure for selecting the population with the largest mean from k logistic populations, *Amer. J. Math. Management Sci.*, to appear.

GUPTA, S.S. and HUANG, D.-Y. (1981). *Multiple Decision Theory: Recent Developments*, Lecture Notes in Statistics, Vol. 6, Springer-Verlag, New York.

GUPTA, S.S. and LIANG, T. (1987). On some Bayes and empirical Bayes selection procedures, In *Probability and Bayesian Statistics* (R. Viertl, ed.), Plenum Publishing Corporation, New York.

GUPTA, S.S. and LIANG, T. (1990). On a lower confidence bound for the probability of a correct selection: analytical and simulation studies, *Proceedings of the First International Conference on Statistical Computing* held in Turkey, March 30-April 2, 1987, to appear.

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

GUPTA, S.S., NAGEL, K. and PANCHAPAKESAN, S. (1973). On the order statistics from equally correlated normal random variables, *Biometrika 60*, 403-413.

GUPTA, S.S. and PANCHAPAKESAN, S. (1972). On a class of subset selection procedures, *Ann. Math. Statist. 43*, 814-822.

GUPTA, S.S. and PANCHAPAKESAN, S. (1974). Inference for restricted families: (a) multiple decision procedures; (b) order statistics inequalities, In *Reliability and Biometry: Statistical Analysis of Lifelength* (F. Proschan and R.J. Serfling, eds.), SIAM, Philadelphia, 503-596.

GUPTA, S.S. and PANCHAPAKESAN, S. (1975). On a quantile selection procedure and associated distribution of ratios of order statistics from a restricted family of probability distributions, In *Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of System Reliability and Safety Assessment* (R.E. Barlow, J.B. Fussell and N.D. Singpurwalla, eds.), SIAM, Philadelphia, 557-576.

GUPTA, S.S. and PANCHAPAKESAN, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*, John Wiley & Sons, New York.

GUPTA, S.S. and PANCHAPAKESAN, S. (1985). Subset selection procedures: review and assessment, *Amer. J. Math. Management Sci. 5*, 235-311.

GUPTA, S.S. and PANCHAPAKESAN, S. (1988). Selection and Ranking Procedures in Reliability Models, Chapter 9 in *Handbook of Statistics 7: Quality Control and Reliability* (P.R. Krishnaiah and C.R. Rao, eds.), North-Holland, Amsterdam, 131-156.

GUPTA, S.S. and PANCHAPAKESAN, S. (1990). On sequential ranking and selection procedures, In *Handbook of Sequential Methods* (B.K. Ghosh and P.K. Sen, eds.), Marcel Dekker, New York, to appear.

GUPTA, S.S. and SOHN, J.K. (1990). Selection and ranking procedures for Tukey's generalized lambda distributions, In *Frontiers of Modern Statistical Inference Procedures-II* (E.J. Dudewicz and E. Boffinger, eds.), Proceedings of the Second IPASRAS Conference held in Sydney, Australia, August 9-14, 1987, to appear.

HAN, S. (1987). Contributions to Selection and Ranking Theory with Special Reference to Logistic Populations, Ph.D. Thesis (Also Tech. Report No. 87-38), Dept. of Statistics, Purdue University, West Lafayette, Indiana.

JOINER, B.L. and ROSENBLATT, J.R. (1971). Some properties of the range in samples from Tukey's symmetric lambda distributions, *J. Amer. Statist. Assoc. 66*, 394-399.

LIANG, T. AND PANCHAPAKESAN, S. (1990). Isotonic selection with respect to a control: a Bayesian approach, In *Frontiers of Modern Statistical Inference Procedures-II* (E.J. Dudewicz and E. Boffinger, eds.), Proceedings of the Second IPASRAS Conference held in Sydney, Australia, August 9-14, 1987, to appear.

LORENZEN, T.J. and MCDONALD, G.C. (1981). Selecting logistic populations using the sample medians, *Comm. Statist. A-Theory Methods 10*, 101-124.

MILTON, R.C. (1963). Tables of Equally Correlated Multivariate Normal Probability Integral, *Tech. Report No. 27*, Dept. of Statistics, University of Minnesota, Minneapolis, Minnesota.

RAMBERG, J.S. and SCHMEISER, B.W. (1972). An approximate method for generating symmetric random variables, *Comm. ACM 15*, 987-990.

RAMBERG, J.S. and SCHMEISER, B.W. (1974). An approximate method for generating asymmetric random variables, *Comm. ACM 17*, 78-82.

SANTNER, T.J. (1975). A restricted subset selection approach to ranking and selection problems, *Ann. Statist. 3*, 334-349.

SOHN, J. (1985). Multiple Decision Procedures for Tukey's Generalized Lambda Distributions, Ph.D. Thesis (Also Tech. Report No. 85-20), Dept. of Statistics, Purdue University, West Lafayette, Indiana.

TAMHANE, A.C. (1986). A survey of literature on quantal response curves with a view toward application to the problem of selecting the curve with the smallest q-quantile (ED100q), *Comm. Statist. Theory Methods 15*, 2679-2718.

TAMHANE, A.C. and BECHHOFER, R.E. (1977). A two-stage minimax procedure with screening for selecting the largest normal mean, *Comm. Statist. A-Theory Methods 6*, 1003-1033.

TAMHANE, A.C. and BECHHOFER, R.E. (1979). A two-stage minimax procedure with screening for selecting the larges mean (II): an improved PCS lower bound and associated tables, *Comm. Statist. A-Theory Methods 8*, 337-358.

TUKEY, J.W. (1960). The Practical Relationship Between the Common Transformations of Percentages or Fractions and of Amounts, *Tech. Report No. 36*, Statistical Research Group, Princeton.

TUKEY, J.W. (1962). The future of data analysis, *Ann. Math. Statist. 33*, 1-67.

VAN DER LAAN, P. (1989). Selection from logistic populations, *Statist. Neerlandica 43*, 169-174.