Subset Selection Procedures for
Restricted Families of Probability Distributions†

by

Shanti S. Gupta* and Ming-Wei Lu**

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #518
November 1977

*Department of Statistics, Purdue University, W. Lafayette, Indiana

**Department of Mathematics, Indiana University, Bloomington, Indiana

†This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.
Subset Selection Procedures for
Restricted Families of Probability Distributions
by Shanti S. Gupta and Ming-Wei Lu

ABSTRACT

In this paper we are interested in studying multiple decision procedures for \( k (k \geq 2) \) populations which are themselves unknown but which one assumed to belong to a restricted family. We propose to study a selection procedure for distributions associated with these populations which are convex-ordered with respect to a specified distribution \( G \) assuming that there exists a best one. The procedure described here is based on a statistic which is a linear function of the first \( r \) order statistics and which reduces to the total life statistics when \( G \) is exponential. The infimum of the probability of a correct selection and an asymptotic expression for this probability are obtained using the subset selection approach. Some other properties of this procedure are discussed. Asymptotic relative efficiencies of this rule with respect to some selection procedures proposed by Barlow and Gupta (1969) for the star-ordered distributions and by Gupta (1969) for the gamma populations with unknown shape parameters are obtained. A selection procedure for selecting the best population using the indifference zone approach is also studied.
SELECTION PROCEDURES FOR RESTRICTED FAMILIES
OF PROBABILITY DISTRIBUTIONS*

1. Introduction

In many problems, especially those in reliability theory, one is interested in using a model for life length distribution which is not completely specified but belongs, for example, to a family of distributions having increased failure rate (IFR), or increasing failure rate on the average (IFRA). Such distributions form special cases of what are now commonly known as restricted families of probability distributions. The idea of using such families stems from the fact that in many cases the experimenter cannot specify the model (distribution) exactly but is able to say whether it comes from a family of distributions such as IFR, IFRA. Families of probability distributions of these types have been studied by several authors, see, for example, Barlow, Marshall and Proschan [4], Barlow and Proschan [5,6] and Barlow and Doksum [1].

In this paper we are interested in studying multiple decision procedures for \( k (k \geq 2) \) populations which are themselves unknown but which are assumed to belong to a restricted family. We now give some definitions of interest to us (see Barlow and Gupta [3]).

(i) \( F \) is said to be convex with respect to \( G \) (written \( F \leq_C G \)) if and only if \( G^{-1}F(x) \) is convex on the support of \( F \).

(ii) \( F \) is said to be star-shaped with respect to \( G \) (written \( F \preceq G \)) if and only if \( F(0) = G(0) = 0 \) and \( \frac{G^{-1}F(x)}{x} \) is increasing in \( x \geq 0 \) on the support of \( F \).

If \( G(x) = 1 - e^{-x} \), \( x \geq 0 \), then \( F \preceq_C G \) is equivalent to saying that \( F \) has increasing failure rate (IFR). Again if \( G(x) = 1 - e^{-x} \), \( x \geq 0 \), \( F \preceq_C G \) is

*This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.
equivalent to saying that \( F \) has increasing failure rate on average (IFRA).

In the statistical literature, selection problems for restricted families were first investigated by Barlow and Gupta [3]. Some further results in this direction and a review of some important results concerning inequalities for restricted families and problems of inference for such families have been given by Gupta and Panchapakesan [10,11] and Patel [16].

In Section 2, we propose and study a subset selection rule for distributions which are \( \preceq \) ordered with respect to a specified distribution \( G \) assuming there exists a best one. Some properties of this rule are discussed. The infimum of the probability of a correct selection is obtained and an asymptotic expression is also given. We also study the asymptotic relative efficiencies of this rule with respect to some selection procedures. Section 3 deals with selecting the best population using the indifference zone approach.

2. Selection rules for distributions \( \preceq \) ordered with respect to a specified distribution \( G \).

Before discussing the selection problem, we give some preliminary known results for sake of completeness. Let \( F \) be the class of absolutely continuous distribution functions \( F \) on \( \mathbb{R} \) such that \( F(0) = 0 \) with positive and right-(or left-) continuous density \( f \) on the interval where \( 0 < F < 1 \). For \( F \in F \), we take \( F^{-1}(1) \) to be equal to the right hand endpoint of the support of \( F \) and we define \( F^{-1}(0) = 0 \). For \( F, G \in F \), consider the following transformation (see Barlow and Doksum [1])

\[
H_F^{-1}(t) = \int_0^{F^{-1}(t)} g[G^{-1}F(u)]du \quad 0 \leq t \leq 1,
\]

where \( g \) denotes the density of \( G \).
We assume that $G$ is always fixed. Since $H_F^{-1}$ (the inverse of $H_F$) is strictly increasing on $[0,1]$, $H_F$ is a distribution. Barlow and Doksum [1] have shown that $F \leq G$ if and only if $H_F$ is convex on the interval where $0 < H_F < 1$. Since $G$ is assumed known we can estimate $H_F^{-1}$ by substituting the empirical distribution $F_n$ of $F$; that is

$$(2.2) \quad H_F^{-1}(t) = H_F^{-1}(F_n(t)) = \int_0^{F_n(t)} g[G^{-1}F_n(u)]du$$

$$(2.3) \quad H_F^{-1}(X_n) = \int_0^{X_n} g[G^{-1}F_n(u)]du = \frac{X_n}{n} g[G^{-1}F_n(u)]X_{i,n} - X_{i-1,n}$$

where $X_{i,n}$ is the $i$th order statistic in a sample of size $n$ from $F$ and $X_{0,n} = 0$.

If $G(x) = 1 - e^{-x}$ for $x \geq 0$, then (2.3) can be written as

$$(2.4) \quad H_F^{-1}(X_n) = \frac{1}{n} [X_{1,n} + \ldots + X_{r-1,n} + (n-r+1)X_{r,n}]$$

We say that $X_{1,n} + \ldots + X_{r-1,n} + (n-r+1)X_{r,n}$ is the total life statistic until $r$th failure from $F$.

(A) Selection procedure and its properties

Let $\pi_1, \ldots, \pi_k$ be $k$ populations. The random variable $X_i$ associated with $\pi_i$ has distribution function $F_i$, $i = 1, 2, \ldots, k$, where $F_i \in F$. Let $F_{[k]}$ denote the cumulative distribution function (c.d.f) of the "best" population. We assume that (a) $F_i(x) \geq F_{[k]}(x)$ for all $x$, $i = 1, \ldots, k$ and (b) there exists a distribution $G$ such that $F_i \preceq G$, $i = 1, \ldots, k$, where $\preceq$ denotes a partial ordering relation on the space of probability distributions. We are given a sample of size $n$ from each $\pi_i (i = 1, \ldots, k)$.

Our goal is to select a subset from the $k$ populations so as to include
the population with $F[k]$. Let $\Omega = \{ F = (F_1, \ldots, F_k) : \exists a_j \text{ such that } F_i(x) \geq F_j(x) \text{ for all } x \text{ and } i = 1, 2, \ldots, k \}$. Let

$$
(2.5) \quad T_i = \sum_{j=1}^{r} a_{j} X_{i;j,n} \quad \text{for } i = 1, \ldots, k \text{ and }
$$

$$
(2.6) \quad T = \sum_{j=1}^{r} a_{j} Y_{j,n}
$$

where $X_{i;j,n}$ is the $j^{th}$ order statistic from $F_i$, $Y_{j,n}$ is the $j^{th}$ order statistic from $G$, $r$ is a fixed positive integer $(1 \leq r \leq n)$, $a_j = g^{-1}(j-1/n) - g^{-1}(j/r-1/n)$ for $j = 1, \ldots, r-1$, and $a_r = g^{-1}(r-1/n)$.

For selecting a subset containing $F[k]$, we propose the selection rule $R_1$ as follows:

$$
(2.7) \quad T_i \geq c_1 \max_{1 \leq j \leq k} T_j
$$

where $c_1 = c_1(k, p^*, n, r)$ is the largest number between 0 and 1 which is determined as to satisfy the probability requirement

$$
(2.8) \quad \inf_{\Omega} P[CS|R_1] \geq p^*
$$

where CS stands for a correct selection, i.e., the selection of any subset which contains the population with distribution $F[k]$. For a given $\alpha (0 < \alpha < 1)$, we assume each $F_i$ has a unique $\alpha$-quantile. Let $F[i](x) = F[i]$ denote the cumulative distribution function of the population with $i^{th}$ smallest $\alpha$-quantile. Let $T(i)$ be associated with $F[i]$ and let $W_i(x)$ be the c.d.f. of $T(i)$.
Lemma 2.1. Let $F_1, F_2$ be two distribution functions such that $F_1(x) \geq F_2(x) \forall x$ and $T_i = \sum_{j \in \Delta} b_j X_{i,j,n} \quad i=1,2$, where $b_j > 0$ for $j \in \Delta$, $\Delta \subseteq \{1,2,\ldots,n\}$ and $X_{i,j,n}$ is the $j$th order statistic from $F_i$, $i=1,2$, then $P[T_1 \leq x] \geq P[T_2 \leq x]$.

Proof. Let $\Psi(X_{i1},\ldots,X_{i,n}) = \begin{cases} 1 & \text{if } T_i \geq x \\ 0 & \text{otherwise} \end{cases}$ where $X_{i1},\ldots,X_{i,n}$ are $n$ observations from $F_i$ ($i=1,2$). Since $\Psi(X_{i1},\ldots,X_{i,n})$ is nondecreasing in each of its arguments, it follows by induction (Lehmann [15] p. 112) that $E\Psi(X_{11},\ldots,X_{1n}) \leq E\Psi(X_{21},\ldots,X_{2n})$. That is $P[T_1 \geq x] \leq P[T_2 \geq x]$. This proves the lemma.

We now state and prove the following theorem which is more general than that of Patel [16].

Theorem 2.1. If $F_1, G \in F$, $F_i(x) \geq F_{[k]}(x) \forall x$ and $i=1,2,\ldots,k$, $F_{[k]} \leq G$, $a_j \geq 0$ for $j=1,2,\ldots,r$, $g(0) \leq 1$ and $a_r \geq c_1$, then

$$\inf_{\Omega} P[CS|R_1] = \int_0^\infty c_1^{k-1} \left( \frac{x}{c_1} \right) dG_T(x)$$

where $G_T(x)$ is the c.d.f. of $T$.

Proof. $P[CS|R_1] = P[T_{(k)} \geq c_1 T(i), i=1,\ldots,k-1]$

$$= \int_0^\infty \prod_{i=1}^{k-1} w_i \left( \frac{x}{c_1} \right) dW_k(x)$$

$$\geq \int_0^\infty w_k \prod_{j=1}^{k-1} \left( \frac{x}{c_1} \right) dW_k(x) \quad \text{(By Lemma 2.1)}$$

$$= P[Z_{k} \geq c_1 Z_j, j=1,\ldots,k-1]$$

where $Z_1,\ldots,Z_k$ are i.i.d. with c.d.f. $W_k(x)$.

Let $\phi(x) = G^{-1} F_{[k]}(x)$. 

Let $\phi(x) = G^{-1} F_{[k]}(x)$. 

Note that \( \phi(x) \) is nondecreasing in \( x \). Also we can write

\[
(2.10) \quad Z_i = \sum_{j=1}^{r} a_j X_{i;j,n}^* \quad i = 1, \ldots, k,
\]

where \( X_{i;j,n}^* \) is the \( j \)th order statistic in a sample of size \( n \) from \( F_k \), \( i = 1, \ldots, k \).

\[
(2.11) \quad P[Z_k \geq c_1 \max_{1 \leq j \leq k} Z_j] = P[\frac{1}{c_1} Z_k \geq \phi(Z_j), \quad i = 1, \ldots, k-1].
\]

Since \( \sum_{j=1}^{r} a_j = g(0) \leq 1 \), \( a_j \geq 0 \quad \forall j = 1, \ldots, r \), and \( \phi(0) = 0 \), by Lemma 4.1 of Barlow and Proschan [5] and (2.10), then

\[
(2.12) \quad \phi(Z_i) \leq \sum_{j=1}^{r} a_j \phi(X_{i;j,n}^*).
\]

Since \( \frac{1}{c_1} a_i \geq 1 \), \( \frac{1}{c_1} \sum_{j=1}^{r} a_j \geq 1 \) for \( i = 1, \ldots, r \), and \( \phi(0) = 0 \), by Lemma 4.3 of Barlow and Proschan [5] and (2.10), we have

\[
(2.13) \quad \phi(-\frac{1}{c_1} z_k) \geq \frac{1}{c_1} \sum_{j=1}^{r} a_j \phi(X_{k;j,n}^*).
\]

\[
(2.14) \quad \phi(X_{i;j,n}^*) = Y_{i;j,n}
\]

where \( Y_{i;j,n} \) is the \( j \)th order statistic from \( G \), \( i = 1, 2, \ldots, k \). Thus from (2.11), (2.12), (2.13), and (2.14),

\[
P[Z_k \geq c_1 \max_{1 \leq i \leq k} Z_i] \geq P[\sum_{j=1}^{r} a_j Y_{k;j,n} \geq c_1 \sum_{j=1}^{r} a_j Y_{i;j,n}, \quad i = 1, \ldots, k-1]
\]

\[
= \int_{0}^{\infty} G_{k-1}(\frac{x}{c_1}) dG_n(x)
\]
This completes the proof.

The constant \( c_1 = c_1(k, P^*, n, r) \) satisfying (2.8) is the largest number between 0 and 1 determined by

\[
\int_0^\infty G_T^{-1}(x) \left( \frac{x}{c_1} \right) dG_T(x) \geq P^* \quad \text{and} \quad gG_T^{-1}\left( \frac{r-1}{n} \right) \geq c_1.
\]

We now consider two specific distributions \( G(x) \). If \( G(x) = 1 - e^{-x} \), \( x \geq 0 \), then we have following result which slightly generalizes the result of Patel [16].

**Corollary 2.1.** If \( F_i(x) \geq F[k](x) \forall x \) and \( i = 1, \ldots, k \), \( F[k] \preceq G \), \( G(x) = 1 - e^{-x} \), \( x > 0 \) and \( n \geq \max\{r, \frac{r-1}{1-c_1} \} \), then

\[
(2.15) \quad \inf_{\Omega} P[CS|R_1] = \int_0^\infty H_T^{-1}\left( \frac{x}{c_1} \right) dH(x)
\]

where \( H(x) \) is the c.d.f. of a \( \chi^2 \) random variable with \( 2r \) d.f.

**Proof.** If \( G(x) = 1 - e^{-x} \) then \( a_j = \frac{1}{n} \) for \( j = 1, 2, \ldots, r-1 \) and \( a_r = \frac{1}{n} (n-x+1) \).

Also \( \frac{1}{c_1} a_r \geq 1 \) iff \( n \geq \frac{r-1}{1-c_1} \). By Theorem 2.1 and the fact that \( 2nT \) is distributed as \( \chi^2 \) with \( 2r \) d.f., the result follows.

If \( G(x) = x \) for \( 0 < x < 1 \), then we have the following result which is a special case of Theorem 2.1 of Barlow and Gupta [3].

**Corollary 2.2.** If \( F_i(x) \geq F[k](x) \forall x \) and \( i = 1, \ldots, k \), \( F[k] \preceq G \) and \( G(x) = x \) for \( 0 < x < 1 \), then

\[
(2.16) \quad \inf_{\Omega} P[CS|R_1] = n^{n-1} \int_{0}^{\min\{n, n-r\}} \left( \sum_{i=r}^{n} \binom{n}{i} \left( \frac{x}{c_1} \right)^i \left( 1 - \frac{x}{c_1} \right)^{n-i} \right) x^{k-1} (1-x)^{n-r-1} dx.
\]

Actually, the condition \( F[k] \preceq G \) in Corollary 2.2 can be relaxed to
\[ F_k \leq G. \]

We state and prove the following theorem about the asymptotic evaluations of the probability of a correct selection associated with the rule \( R_1 \) in the case where \( r \) is so chosen that \( r \leq (n+1)\alpha < r+1, \; 0 < \alpha < 1 \). This amounts to selecting populations with large values of the \( \alpha \)-quantile for \( \alpha \) (and \( r \)) as defined above. In this case, \( \frac{r}{n} \to \alpha \) as \( n \to \infty \). Note that the result holds for all \( \alpha \).

**Theorem 2.2.** If \( F_i \in F \) for all \( i = 1, \ldots, k \) and

(i) \( F_i(x) \geq F_k(x) \) \( \forall x, \; i = 1, \ldots, k \), \( F_k \leq G \),

(ii) \( G(x) \) has a differentiable density \( g \) in a neighborhood of its \( \alpha \)-quantile \( \eta_\alpha \), \( g(\eta_\alpha) \neq 0 \), and

(iii) \( gG^{-1} \) is uniformly continuous on \([0,1]\), \( G^{-1}(x) \) is convex and there exists an \( \xi, \; 0 < \xi < 1 \), such that for \( \xi \leq y < 1 \), \( \frac{gG^{-1}(y)}{1-y} \) is nondecreasing in \( y \), then as \( n \to \infty \)

\[
(2.17) \quad P[CS|R_1] \geq \int_{-\infty}^{\infty} \phi^{-1}\left[ \frac{x}{c_1} + \frac{1-c_1}{c_1} \eta_\alpha g(\eta_\alpha) \left( \frac{n}{\alpha \bar{\alpha}} \right)^{1/2} \right] d\phi(x)
\]

where \( \bar{\alpha} = 1 - \alpha \) and \( \phi(x) \) is the standard normal c.d.f.

**Proof.** We note that

\[
(2.18) \quad P[CS|R_1] \geq P[Z_k \geq c_1 \max_{1 \leq j \leq k} Z_j]
\]

where \( Z_1, \ldots, Z_k \) are i.i.d. with c.d.f \( W_k(x) \) and \( W_k(x) \) is the c.d.f. of \( T(k) \).
By Theorem (2.2) of Barlow and Van Zwet [7] and condition (iii), then we have (see Barlow and Doksum [1]), for \( n \) large,

\[
Z_i = Y_{i;r,n}
\]

where \( Y_{i;r,n} \) is the \( r \)th order statistic from \( H_{F[k]} \) and \( H_{F^{-1}[k]} \) (the inverse of \( H_{F[k]} \)) is defined in (2.1). Now \( F[k] < G \) if and only if \( H_{F[k]} \) is convex. Since \( G^{-1}(x) \) is increasing and convex, it follows that \( G^{-1}H_{F[k]}(x) \) is convex. Since \( H_{F[k]} \leq G \) and \( G^{-1}(0) = 0 \), then \( H_{F[k]} \leq G \). In a manner similar to the theorem (2.1) of Barlow and Gupta [3], we have

\[
P[Y_{k;r,n} > c_i \max_{1 \leq i \leq k} Y_{i;r,n} | Y_{k;r,n} > c_i Y_{i;r,n}, i \neq k] \geq P[Y_{k;r,n} > c_i Y_{i;r,n} | i \neq k]
\]

where \( Y_{i;r,n} \) is the \( r \)th order statistic from \( G \), \( i = 1, \ldots, k \). From (2.18), (2.20), (2.21) and using the fact that

\[
Y_{i;r,n} \sim N(\eta_\alpha, \frac{\alpha^2}{\text{ng}(\eta_\alpha)})
\]

the theorem follows.

Before we discuss some properties of the selection rule \( R_i \), we introduce some definitions (see Santner [17]).

Define \( P_F(i) = P_F[\pi(i) \text{ is selected} | R] \) where \( \pi(i) \) is associated with \( F[i] \).
Definition 2.1.

(i) A rule $R$ is strongly monotone in $\pi_{(i)}$ if

$$P_{F}(i) \begin{cases} \uparrow \text{in } F_{[i]} \text{ when all other components of } F \text{ are fixed} \\ \downarrow \text{in } F_{[j]}(j \neq i) \text{ when all other components of } F \text{ are fixed.} \end{cases}$$

That means, $P_{F_{1}}(i) \geq P_{F_{2}}(i)$ when $F_{[i]} \succeq F_{[i]}^{*}$ and $P_{F_{2}}(j) \leq P_{F_{1}}(j)$

when $F_{[j]} \succeq F_{[j]}^{*}$ for $j \neq i$, where $F_{1} = (F_{[1]}, \ldots, F_{[i]}, \ldots, F_{[k]})$

$F_{2} = (F_{[1]}, \ldots, F_{[j]}, \ldots, F_{[k]})$ and

$F_{1}^{*} = (F_{[1]}, \ldots, F_{[i]}^{*}, \ldots, F_{[k]})$.

(ii) A rule $R$ is monotone means $P_{F}(i) \leq P_{F}(j)$ for all $F \in \Omega$ with

$F_{[i]}(x) \geq F_{[j]}(x)$.

(iii) A rule $R$ is unbiased if $P_{F}(i) \leq P_{F}(k)$ for all $F \in \Omega$ with

$F_{[i]}(x) \geq F_{[k]}(x)$.

(iv) A rule $R$ is consistent with respect to $\Omega'$ means

$$\inf_{\Omega'} P[CS|R] \to 1 \text{ as } n \to \infty .$$

It is similar to the proof in Lemma 2.1, we can show the following theorem.

Theorem 2.3. If $a_{i} \geq 0$ for $i = 1, \ldots, r$, then $R_{1}$ is strongly monotone in $\pi_{(i)}$.

Remark 2.1.

(1) If a rule $R$ is strongly monotone in $\pi_{(i)}$ for all $i = 1, \ldots, k$,

then $R$ is monotone and $\inf_{\Omega} P[CS|R] = \inf_{\Omega_{0}} P[CS|R]$ where
\( \Omega_0 = \{ \mathcal{F} = (F_1, \ldots, F_k) \in \Omega : F_1 = \ldots = F_k \} \).

(2) If \( R \) is monotone, then it is unbiased.

(3) If \( F_i(x) = F(x, \theta_i) \), \( i = 1, \ldots, k \) and \( T_i \) is a consistent estimator of \( \theta_i \), then \( R_1 \) is consistent with respect to \( \Omega = \{ \mathcal{F} = (F_1, \ldots, F_k) : \exists a_j \text{ such that } F_i(x) \geq F_j(x) \text{ for all } x \text{ and } i = 1, \ldots, k \} \).

(4) If \( F_i, G \in \mathcal{F}, F_i \leq G, i = 1, \ldots, k \) and the condition (iii) of Theorem 2.2 is satisfied, we can show that \( R_1 \) is consistent with respect to \( \Omega \).

The selection of the population with largest \( F_i \) (\( i = 1, \ldots, k \)) can be handled analogously. We assume \( F_{[1]}(x) \leq F_{[i]}(x), i = 1, \ldots, k \), and \( F_{[1]} \leq G \). The rule for selecting the population with \( F_{[1]} \) is \( R_2 \): Select population \( \pi_i \) if and only if

\[
(2.23) \quad c_2 T_i \leq \min_{1 \leq j \leq k} T_j
\]

where \( c_2 (0 < c_2 < 1) \) is determined so as to satisfy the basic requirement.

In a manner similar to the proof of Theorem 2.1, we have

**Theorem 2.4.** If \( F_i, G \in \mathcal{F}, F_{[1]}(x) \leq F_{[i]}(x) \forall x \) and \( i = 1, \ldots, k \), \( F_{[1]} \leq G \), \( a_j \geq 0 \) for \( j = 1, \ldots, r \), \( g(0) \leq 1 \) and \( a_r \geq c_2 \), then

\[
(2.24) \quad \inf_{\Omega'} P[C | S \in R_2] = \int_{0}^{\infty} \frac{c_{k-1}}{T} (c_2 x) d \xi_T(x)
\]

where \( \xi_T(x) = 1 - G_T(x) \) and \( \Omega' = \{ \mathcal{F} = (F_1, \ldots, F_k) : \exists a_j \text{ such that } F_i(x) \leq F_j(x) \text{ for all } x \text{ and } i = 1, \ldots, k \} \).
(B) **Efficiency of procedure \( R_1 \) under slippage configuration.**

Under the same notations and conditions of Theorem 2.2 and the comments above the Theorem 2.2, we consider slippage configuration \( F_{[i]}(x) = F^{(2)}(\frac{x}{\delta}) \), \( i=1,2,\ldots,k-1 \), and \( F_{[k]}(x) = F(x) \), \( 0 < \delta < 1 \). Let \( E(S|R) \) denote the expected subset size using the rule \( R \). Then \( E(S|R) - P[CS|R] \) is the expected number of non-best populations included in the selected subset. For a given \( \varepsilon > 0 \), let \( n_R(\varepsilon) \) be the asymptotic sample size for which \( E(S|R) - P[CS|R] = \varepsilon \). We define the asymptotic relative efficiency

\[
\text{ARE}(R, R^*, \delta) = \lim_{\varepsilon \to 0} \frac{n_R(\varepsilon)}{n_{R^*}(\varepsilon)}
\]

i.e., \( \text{ARE}(R, R^*; \delta) = \lim_{\varepsilon \to 0} \frac{n_R(\varepsilon)}{n_{R^*}(\varepsilon)} \). Under the slippage configuration we have,

\[
(2.25) \quad E(S|R_1) = P[CS|R_1] + (k-1)P[T(1) \geq c_1 \max_{i \neq 1} T_i]
\]

If \( n \) is large, then from an argument similar to the one in the proof of Theorem 2.2, we have

\[
(2.26) \quad P[T(1) \geq c_1 \max_{i \neq 1} T_i] \approx P[Y_1 \geq c_1 \max_{i \neq 1} Y_i]
\]

where \( Y_1, \ldots, Y_k \) are independent and \( Y_i \) is the \( r \)th order statistic from \( H_{F_{[i]}} \) for \( i=1,\ldots,k \). The right hand side of (2.26) is asymptotically equal to

\[
(2.27) \quad \int_{-\infty}^{\infty} \phi \left( \frac{\delta x}{c_1} - a_\alpha h(a_\alpha) \left( 1 - \frac{\delta}{c_1} \right) \left( \frac{n}{\alpha} \right)^{1/2} \right) d\phi(x)
\]

\[
\phi^{k-2} \left( \frac{x}{c_1} - a_\alpha h(a_\alpha) \left( 1 - \frac{1}{c_1} \right) \left( \frac{n}{\alpha} \right)^{1/2} \right) d\phi(x)
\]

where \( c_1 \) is the constant used in defining \( R_1 \), \( a_\alpha \) is the (unique)
\(\alpha\)-quantile of \(H_{F \mid \mathcal{G}}(x)\) and \(h(x)\) is the density function of \(H_{F \mid \mathcal{G}}(x)\).

For \(k = 2\) and \(n\) large,

\[
(2.28) \quad E(S \mid R_1) - P(\mathcal{G} \mid R_1) \approx \Phi (-h(a_\alpha) a_\alpha (1 - \frac{\delta}{c_1}) (\frac{n}{a_\alpha})^{1/2} (1 + \frac{\delta^2}{c_1})^{-1/2}) .
\]

Let

\[
\int_{-\infty}^{\infty} \phi^{k-1} \left( \frac{x}{c_1} + (1-c_1) \eta_{\alpha} \phi(\eta_{\alpha}) \frac{1}{c_1} (\frac{n}{a_\alpha})^{1/2} \right) \phi(x) = p^* .
\]

Now setting the right side of (2.28) equal to \(\varepsilon\) and using \(c_1 \approx 1 - \frac{2^{1/2}D}{n^{1/2}}\),

where \(D = \Phi^{-1}(p^*)(a\alpha)^{1/2}/\eta_{\alpha} \phi(\eta_{\alpha})\), we obtain

\[
(2.29) \quad n_{R_1}(\varepsilon) \approx \left[ -(a\alpha)^{1/2} \phi^{-1}(\varepsilon)(1 + \delta^2)^{1/2} + \sqrt{2} D\delta a_{\alpha} h(a_{\alpha}) \right]^2 .
\]

\[
[a_{\alpha}^2 h^2(a_{\alpha})(1 - \delta)^2]^{-1} .
\]

Comparison with Barlow-Gupta Procedure

Barlow and Gupta [3] propose a procedure \(R_3\), for the quantile selection problem of star-ordered distributions which is,

\[
R_3: \text{ Select population } \pi_i \text{ if and only if }
\]

\[
(2.30) \quad T_{R,i} \geq c_3 \max_{1 \leq j \leq k} T_{R,j} .
\]

where \(c_3(0 < c_3 < 1)\) is chosen to satisfy \(P(\mathcal{G} \mid R_3) \geq p^*\) and \(T_{R,i}\) is the \(r\)th order statistic from \(F_i\) where \(r \leq (n+1)\alpha < r+1\).

They derive an expression for \(n_{R_3}(\varepsilon)\) as follows:

\[
n_{R_3}(\varepsilon) \approx \left[ -(a\alpha)^{1/2} \phi^{-1}(\varepsilon)(1 + \delta^2)^{1/2} + \sqrt{2} D\delta a_{\alpha} f(\xi_{\alpha}) \right]^2 [\xi_{\alpha}^2 f^2(\xi_{\alpha})(1 - \delta)^2]^{-1} .
\]

where \(f\) is the density of \(F\) with unique \(\alpha\)-quantile, \(\xi_{\alpha}\).

\[
(2.31) \quad \text{ARE}(R_1, R_3; \delta) = \lim_{\varepsilon \to 0} \frac{n_{R_1}(\varepsilon)}{n_{R_3}(\varepsilon)} = \frac{\xi_{\alpha}^2 f^2(\xi_{\alpha})}{a_{\alpha}^2 h^2(a_{\alpha})} .
\]
If \( G(x) = 1 - e^{-x} \), \( x > 0 \) and \( F_1(x) = 1 - e^{-x/\delta} \) and \( F_2(x) = 1 - e^{-x} \), \( x \geq 0 \), \( 0 < \delta < 1 \), we have,

\[
(2.32) \quad \text{ARE}(R_1, R_3; \delta) = \frac{(1-\alpha)^2 \log^2 (1-\alpha)}{\alpha^2} \leq 1
\]

\[
= 0.4803, \quad \alpha = 1/2.
\]

Comparison with Gupta Procedure

Gupta [8] gave a selection procedure for gamma populations \( \pi_i \)'s with densities \( \frac{1}{\Gamma(a)} x^{a-1} e^{-x/\theta_i} \), \( x > 0 \), \( \theta_i > 0 \), \( i = 1, 2, \ldots, k \). The procedure \( R_4 \) is

\( R_4: \) Select population \( \pi_i \) if and only if

\[
(2.33) \quad \bar{X}_i \geq c_4 \max_{1 \leq j \leq k} \bar{X}_j
\]

where \( \bar{X}_i \) is the sample mean of size \( n \) from \( \pi_i \) and \( c_4 \) is the largest constant \( (0 < c_4 < 1) \) chosen so that \( P[CS|R_4] \geq P^* \).

For \( k = 2 \), \( \theta_1 = \delta \) and \( \theta_2 = 1 \), (see Barlow and Gupta [3]), we have

\[
(2.34) \quad \text{ARE}(R_3, R_4; \delta) = \lim_{\varepsilon \to 0} \frac{n_{R_3}(\varepsilon)}{n_{R_4}(\varepsilon)} = \frac{a(\log \delta)^2 \alpha \delta (1+\delta^2)}{2(1-\delta)^2 [\xi'_\alpha f(\xi_\alpha)]^2}.
\]

Hence

\[
(2.35) \quad \text{ARE}(R_1, R_4; \delta) = \text{ARE}(R_1, R_3; \delta) \cdot \text{ARE}(R_3, R_4; \delta)
\]

\[
= \left( \frac{\sqrt{\alpha} \log \delta / \sqrt{\alpha} \sqrt{1+\delta^2}}{\sqrt{2} (1-\delta) a_\alpha h(a)} \right)^2.
\]
If \( G(x) = 1 - e^{-x} \) for \( x > 0 \) and \( a = 1 \),

\[
(2.36) \quad \text{ARE}(R_1, R_4; \delta) = \frac{(1-a)(1+\delta^2) \log \delta}{2(1-\delta)^2 a}
\]

\[
(2.37) \quad \text{ARE}(R_1, R_4; \delta+1) = \frac{1-a}{a}.
\]

**Comparison of \( R_1 \) and \( R_5 \) from uniform distribution**

Suppose \( \pi_1 \) and \( \pi_2 \) are two independent uniform populations with distribution functions \( F_i \) \( (i=1,2) \).

\[
(2.38) \quad F_i(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{\theta_i} & 0 \leq x \leq \theta_i \\ 1 & x > \theta_i \end{cases}
\]

where \( \delta = \theta_1 < \theta_2 = 1 \).

A sample of \( n \) independent observations is drawn from each of the two populations. Let \( T_i^* \) be the total life statistic until \( r \)th failure from \( \pi_i \) \( (i=1,2) \) where \( r \leq (n+1)a < r+1 \). The procedure \( R_5 \) is given by

\[
R_5: \text{Select population } \pi_1 \text{ if and only if }
\]

\[
(2.39) \quad T_i^* \geq c_5 \max_{1 \leq j \leq k} T_j^*
\]

where \( c_5 \) is chosen so that \( P[CS|R_5] \geq \pi^* \).

Let \( T_i^*(1) \) be associated with \( \theta_1 \).

\[
(2.40) \quad \text{E}(S|R_5) = P[CS|R_5] = P[T_i^*(1) \geq c_5 T_i^*(2)] = P[T_i^* \geq \frac{c_5}{\delta} T_i]
\]
where \( T_1, T_2 \) are two independent total life statistic until \( r \)th failure from uniform distribution over \((0,1)\). By Gupta and Sobel [12],

\[
(2.41) \quad \frac{T_1 - u}{\sigma} \rightarrow N(0,1) \quad \text{as} \quad n \rightarrow \infty
\]

where \( u = \frac{n \alpha (2n - an + 1)}{2n + 1} \approx \frac{n \alpha (2 - \alpha)}{2} \), \( \sigma^2 = \text{Var} \) and \( A = \frac{\alpha (1 - \alpha) (2 - \alpha)^2}{4} + \frac{\alpha^3}{12} \). Hence \( \frac{u}{\sigma} \approx \frac{u'}{\sigma} = B \sqrt{n} \) where \( B = \frac{\alpha (2 - \alpha)}{2 \sqrt{A}} \). From (2.40), we have

\[
E(S|R_5) - P[CS|R_5] \approx P[Z_1 \geq \frac{c_5}{\delta} Z_2 + (\frac{c_5}{\delta} - 1)B \sqrt{n}]
\]

where \( Z_1, Z_2 \) are i.i.d. with \( N(0,1) \).

\[
E(S|R_5) - P[CS|R_5] = \int_{-\infty}^{\infty} \phi\left( \frac{\delta}{c_5} x - (1 - \frac{\delta}{c_5})B \sqrt{n} \right) d\phi(x)
\]

\[
= \phi\left[-(1 - \frac{\delta}{c_5})B \sqrt{n} / \sqrt{1 + \left(\frac{\delta}{c_5}\right)^2} \right]
\]

Let \( E(S|R_5) - P[CS|R_5] = \varepsilon > 0 \), we obtain

\[
(2.42) \quad \left(\frac{1}{c_5} - \frac{1}{\delta}\right) \sqrt{n} = \frac{1}{\delta^2} + \frac{1}{\sqrt{2} \phi^{-1}(\varepsilon)} \cdot \frac{\phi^{-1}(\varepsilon)}{B}
\]

Note that \( \inf_{\Omega} P[CS|R_5] = P[T_1 \geq c_5 T_2] \),

\[
\approx \phi\left[-(1 - \frac{1}{c_5})B \sqrt{n} / \sqrt{1 + 1/c_5^2} \right]
\]

where \( T_1 \) and \( T_2 \) are defined as above.

Setting \( \inf_{\Omega} P[CS|R_5] = P^* \) and using \( c_5 \approx 1 - \frac{\sqrt{2} \phi^{-1}(P^*)}{\sqrt{n} B} \) and

\[
\left(\frac{1}{c_5} \approx 1 + \frac{\sqrt{2} \phi^{-1}(P^*)}{\sqrt{n} B} \right)
\]

from (2.42), we obtain
(2.43) \[ n_{R_5}(\varepsilon) \approx \left\{ \frac{\phi^{-1}(\varepsilon) \sqrt{1+\delta^2} - \sqrt{2} \delta \phi^{-1}(p^*)}{B(1-\delta)} \right\}^2. \]

From (2.29) and (2.43),

(2.44) \[ \text{ARE}(R_1, R_5; \delta) = \lim_{\varepsilon \to 0} n_{R_1}(\varepsilon) \cdot n_{R_5}(\varepsilon) = \frac{\alpha \delta B^2}{a_{\alpha}^2 h(a_{\alpha})}. \]

If we assume that \( G(x) = x \) for \( 0 < x < 1 \), then

(2.45) \[ \text{ARE}(R_1, R_5; \delta) = \frac{B^2(1-\alpha)}{\alpha} = \frac{3(1-\alpha)(2-\alpha)^2}{3(1-\alpha)(2-\alpha)^2 + \alpha^2} < 1 \]

\( \text{ARE}(R_1, R_5; \delta) \) is a decreasing function of \( \alpha \) and for \( \alpha = 1/2 \), it is equal to 0.931. Note that in (2.45), \( R_1 \) is based on \( r^{th} \) ordered statistic and \( R_5 \) is based on the total life statistic until \( r^{th} \) failure.

(C) Selection procedure for distribution \( \xi \) ordered with respect to Weibull distribution

Assume that the specified distribution \( G(x) \) is given by

\[
G(x) = \begin{cases} 
1 - e^{-\lambda x^\alpha} & \text{for } x \geq 0 \\
0 & \text{for } x < 0 
\end{cases}
\]

where \( \lambda > 0 \) and attention is restricted to \( \alpha \geq 1 \) which is assumed known.

In this case, we use \( T_i^* \) as our statistic where

\[ T_i^* = \sum_{j=1}^{r-1} x_{i;j,n}^\alpha + (n-r)x_{i;r,n}^\alpha , \quad i = 1, \ldots, k. \]

As before, \( X_{i;j,n} \) denote the \( j^{th} \) order statistic from \( F_1 \), \( i = 1, \ldots, k \). Since \( G(x) \) is convex with respect to the exponential distribution if \( \alpha \geq 1 \) and since the convex ordering is transitive, the family of distributions which are convex
with respect to Weibull \((\alpha \geq 1)\) will have IFR distribution. Thus our interest here is in a special subclass of IFR distributions. The rule for selecting the population which is associated with \(F_{[k]}\) is as follows,

\[
R_6: \text{Select population } \pi_i \text{ if and only if }
\]

\[
(2.46) \quad \max_{1 \leq j \leq k} T^*_j \geq c_6 \quad T^*_i = \frac{g_{\alpha}(x_{i,n})}{g_{\alpha}(x_{j,n})}
\]

where \(c_6 (0 < c_6 < 1)\) is determined so as to satisfy the basic probability requirement.

Using the fact that if \(F < G\) and \(F(0) = G(0) = 0\) then \(F_{\alpha} < G_{\alpha}\) for \(\alpha \geq 1\), where \(F_{\alpha}\) is the c.d.f. of \(X_{\alpha}\), \(F(x)\) is the c.d.f. of \(X, C_{\alpha}\) is the c.d.f. of \(Y_{\alpha}\) and \(G(y)\) is the c.d.f. of \(Y\). Also, \(G_{\alpha}^{-1}F_{\alpha}(x_{i,n})\) is stochastically equivalent to the \(i^{th}\) order statistic from \(G(x) = 1 - e^{-x}\), for \(x \geq 0\), where \(x_{1,n} \leq \ldots \leq x_{n,n}\) are order statistics from \(F\). In a manner similar to the proof of Theorem 2.1, one can prove the following theorem.

**Theorem 2.5.** If \(F_i(x) \geq F_{[k]}(x) \forall x\) and \(i = 1, \ldots, k\), \(F_{[k]}(0) = 0\), \(F_{[k]} < G\), \(G(x) = 1 - e^{-\lambda x}\), \(x > 0\), \(\lambda > 0\) and \(\alpha (\geq 1)\) is known and \(n \geq \max\{r, \frac{r-1}{1-c_6}\}\), then

\[
(2.47) \quad \inf_{\Omega} P[C | R_6] = \int_0^\infty h(x) \frac{e^{-\frac{x}{c_6}}}{\beta^\alpha} \frac{e^{-\lambda x}}{\Gamma(\alpha)} dx
\]

where \(H(x)\) is the c.d.f. of a \(\chi^2\) random variable with 2r d.f.

(D) Selection with respect to the means for Gamma populations

Let \(\pi_1, \ldots, \pi_k\) be \(k\) populations with densities

\[
f_i(x) = \frac{\alpha_i}{\Gamma(\alpha_i)} \beta_i x^{\alpha_i - 1} e^{-\beta x}, \quad x \geq 0, \quad \beta > 0, \quad \alpha_i \geq 1, \quad i = 1, \ldots, k.
\]

Let \(F_i(x)\) be the distribution function of \(\pi_i\), \(i = 1, \ldots, k\). We are given
a sample size of \( n \) from each \( \pi_i \). Let \( T_i^* \) be total life statistic until \( r^\text{th} \) failure from \( \pi_i \). Let \( \alpha_1 \leq \cdots \leq \alpha_k \) be the ordered values of \( \alpha_i \)'s. We are interested in selecting the population with the largest value \( \alpha[k] \) (unknown). Since the mean of \( \pi_i \) is \( \frac{\alpha_i}{\beta} \), selection of the population with largest mean is equivalent to selecting the population with largest value, \( \alpha[k] \). The subset selection rule based on \( T_i \) is:

\[
R_j : \text{ Select population } \pi_i \text{ if and only if } \sum_{1 \leq j \leq k} T_i^* \geq c_7 \max_{1 \leq j \leq k} T_j^*,
\]

where \( c_7 (0 < c_7 < 1) \) is the largest value chosen to satisfy \( P[CS|R_j] \geq P^* \). Since the rule \( R_j \) is scale invariant, we can assume \( \beta = 1 \).

Case 1: All \( \alpha_i \) are unknown and \( \geq 1 \). Let \( \Omega_1 = \{ \alpha = (\alpha_1, \ldots, \alpha_k) : \alpha_i \geq 1 \forall i \} \).

In this case, by Corollary 2.1 and \( F_{1C} \leq G(x) = 1 - e^{-x} \), \( x \geq 0 \), \( i = 1, \ldots, k \) we have the following result. If \( n \geq \max \{ r, \frac{r-1}{1-c_7} \} \), then \( \inf_{\Omega} P[CS|R_j] = \int_0^\infty H^{-1}(c_7-x) dH(x) \), where \( H(x) \) is the c.d.f. of a \( \chi^2 \) r.v. with \( 2r \) d.f.

Case 2: \( \alpha_i \) are unknown but assume \( 1 \leq \alpha_i \leq \Delta \), \( i = 1, \ldots, k \) and \( \Delta \) is known. Let \( F_\Delta(x) \) be the c.d.f. of \( X \) with density function \( f_\Delta(x) = \frac{1}{\Gamma(\Delta)} x^{\Delta-1} e^{-x} \), \( x > 0 \). Let \( H(x) \) be the c.d.f. of a \( \chi^2 \) r.v. with \( 2r \) d.f. and let \( h(x) \) be its density function. The following theorem is the lower bound for the probability of correct selection without any condition on \( n \).

Theorem 2.6.

\[
P[CS|R_j] \geq \int_0^\infty H^{-1}(\frac{2n}{c_7}x) \frac{2nh(2ny)}{f_\Delta(y)} e^{-x} dx
\]

where \( y = F_\Delta^{-1}(1 - e^{-x}) \).
Proof. \( P[CS|R_7] = P[T^*_k \geq c_7 \max_{1 \leq j \leq k-1} T^*_j] \),

where \( T^*_j \) is associated with \( a_{[i]} \), \( i = 1, \ldots, k \). Since \( F_\Delta(x) \leq F_1(x) \leq G(x) = 1 - e^{-x} \),

(2.50) \( P[CS|R_7] \geq P[T^*_k \geq c_7 \max_{1 \leq j \leq k-1} T^*_j] \)

where \( T^*_k \) is the total life statistic until \( r \)th failure from \( G(x) \) and \( T^*_j(j = 1, \ldots, k-1) \) is the total life statistic until \( r \)th failure from \( F_\Delta(x) \). Since \( \Delta \geq 1 \) then \( F_\Delta \leq G \). Let \( \phi(x) = G^{-1}F_\Delta(x) \),

(2.51) \( P[T^*_k \geq c_7 T^*_j, j = 1, \ldots, k-1] = P[\phi(\frac{1}{n} T^*_k) \geq \phi(\frac{c_7}{n} T^*_j), j = 1, \ldots, k-1] \)

By Lemma 4.1 of Barlow and Proschan [5] with \( a_1 = \ldots = a_{r-1} = c_7/n \),

\( a_r = (n-r+1)c_7/n \), \( a_i = 0 \) for \( i \geq r+1 \) and \( \phi(X) = Y \) where \( X(Y) \) is a r.v. with distribution function \( F_\Delta(G) \) respectively, we have

(2.52) \( P[\phi(\frac{1}{n} T^*_k) \geq \phi(\frac{c_7}{n} T^*_j), j = 1, \ldots, k-1] \geq P[\phi(\frac{1}{n} T^*_k) \geq \frac{c_7}{2n} Y_j, j = 1, \ldots, k-1] \)

where \( Y_j(j = 1, \ldots, k-1) \) is a r.v. with \( \chi^2 \) with \( 2r \) d.f. From (2.50), (2.51) and (2.52), we have

\( P[CS|R_7] \geq \int_0^\infty H^{k-1}(2n \frac{x}{c_7}) dB(x) \), where \( B(x) = P[\phi(\frac{1}{n} T^*_k) \leq x] \)

Since \( B(x) = H[2n F_\Delta^{-1}(1 - e^{-x})] \), then

\( \int_0^\infty H^{k-1}(2n \frac{x}{c_7}) dB(x) = \int_0^\infty H^{k-1}(2n \frac{x}{c_7}) \frac{2nh(2ny)}{f_\Delta(y)} e^{-x} dx \).

This completes the proof.
Let \( S \) denote the size of the selected subset. The expected value of \( S \) when \( R_7 \) is used is given by

\[
E(S|R_7) = \sum_{i=1}^{k} \frac{1}{\alpha_i} \max_{1 \leq j \leq k} T_j^*
\]

Let \( \Omega' = \{ \alpha = (\alpha_1, \ldots, \alpha_k) : 1 \leq \alpha_i \leq \Delta, i = 1, \ldots, k \} \). For \( \alpha \in \Omega' \), since \( F_\Delta(x) \leq F_1(x) \leq G(x) = 1 - e^{-x} \), then

\[
E(S|R_7) \leq k \max_{2 \leq j \leq k} T_j^*
\]

where \( T_j^* \) is the total life statistic until \( r^\text{th} \) failure from \( F_\Delta(x) \) and \( T_j^*(j=2, \ldots, k) \) is the total life statistic until \( r^\text{th} \) failure from \( G(x) \).

Thus

\[
E(S|R_7) = k \int_0^\infty H^{k-1} \left( \frac{2x}{c_7} \right) dS(x)
\]

where \( H(x) \) is the c.d.f. of \( \chi^2 \) r.v. with \( 2r \) d.f. and \( S(x) \) is the c.d.f. of the total life statistic until \( r^\text{th} \) failure from \( F_\Delta(x) \).

Remark 2.2. (i) We can show that the lower bound for case 2 in Theorem 2.6 is less than or equal to the lower bound for case 1.

(ii) Now we are dealing with the problem in case 2. Let

\[
\int_0^\infty H^{k-1} \left( \frac{x}{c_7} \right) dH(x) = P^* \text{, then } c_7 \text{ can be determined. If } n \geq \max \{ r, (r-1)/(1-c_7) \}, \text{ then we should use the lower bound for case 1. If } r \leq n < (r-1)/(1-c_7) \text{, then the lower bound for case 1 cannot be applied. In this case, we can use the lower bound for case 2.}
\]

(iii) Sometimes, the distribution function \( S(x) \) which is defined above the remark 2.2 is hard to compute. From \( E(S|R_7) \leq k \max_{2 \leq j \leq k} T_j^* \)
where \( T_{1}^{**} \) is the total life statistic until \( r^{th} \) failure from \( F_{\Delta} \) and 
\( T_{j}^{**}(j=2,\ldots,k) \) is the total life statistic until \( r^{th} \) failure from \( G(x) \).

Using the similar arguments in the proof of Theorem 2.6, we can get

\[
E(S|R_{\gamma}) \leq k \int_{0}^{\infty} H^{k-1} \left[ \frac{2n}{c_{7}} F_{\Delta}^{-1} \left( 1 - e^{-\frac{x}{2n}} \right) \right] dH(x)
\]

where \( H(x) \) is the c.d.f. of a \( \chi^{2} \) r.v. with \( 2r \) d.f. In this case, the upper bound of \( E(S|R_{\gamma}) \) can be computed.


Let \( \pi_{1},\ldots,\pi_{k} \) be \( k \) populations. The random variable \( X_{i} \) associated with \( \pi_{i} \) has an absolutely continuous distribution \( F_{i} \). We assume there exists a \( F_{[k]}(x) \) such that \( F_{[i]}(x) \geq F_{[k]}(\frac{x}{\delta}) \) for all \( x \), \( i=1,\ldots,k-1 \) and \( \delta(0<\delta<1) \) is specified. Let

\[
(3.1) \quad \Omega(\delta) = \{ F = (F_{1},\ldots,F_{k}): \exists a_{j} \text{ such that } F_{i}(x) \geq F_{j}(\frac{x}{\delta}) \forall i \neq j \}.
\]

The correct selection is the choice of any population which is associated with \( F_{[k]} \). We propose the selection rule \( R_{\delta} \): Select population \( \pi_{i} \) if and only if

\[
(3.2) \quad T_{i} = \max_{1 \leq j \leq k} T_{j} \quad \text{where } T_{i} \text{ is defined as in (2.5)}.
\]

We want the \( P[CS|R_{\delta}] \geq P^{*} \), for all \( F \in \Omega(\delta) \), where \( P^{*} \) is specified.

**Theorem 3.1.** If \( F_{i}, \ G \in \bar{F}, \ i=1,\ldots,k, \ F_{[k]} \leq G, \ a_{j} \geq 0, \ j=1,\ldots,r, \)
\( g(0) \leq 1 \) and \( a_{r} \geq \delta \), then

\[
(3.3) \quad \inf_{\Omega(\delta)} P[CS|R_{\delta}] = \int_{0}^{\infty} G_{k-1}^{T} (\frac{x}{\delta}) dG_{T}(x)
\]

where \( G_{T}(x) \) is the c.d.f. of \( T \).
Proof. \[ P[CS|R_8] = P[T_{(k)} \geq \max_{1 \leq j \leq k} T_{(j)}] . \]

Since \( F_i(\delta x) \geq F_k(x) \), \( i = 1, \ldots, k - 1 \) and by Lemma 2.1, then
\[
P[CS|R_8] = P[T_{(k)} \geq \delta \frac{T_{(j)}}{\delta} \forall j \neq k] \geq P[T_{(k)} \geq \delta T^*_j \forall j \neq k]
\]

where \( T^*_1, \ldots, T^*_k \), \( T_{(k)} \) are i.i.d. with c.d.f. \( W_k(x) \). Using the same argument as in Theorem 2.1, we have our theorem.

For given \( k, \delta, P^* \) and \( G(x) \), we can possibly find the values of the pair \( (n, r) \), \( (n \geq r) \) which satisfy

\[
(3.4) \quad a_r \geq \delta \quad \text{and} \quad \int_0^\infty G^{-1}_T \left( \frac{X}{\delta} \right) dG_T(x) \geq P^*
\]

If \( G(x) = x \) for \( 0 < x < 1 \), we can always find the values of the pair \( (n, r) \), \( (n \geq r) \) which satisfy

\[
n \binom{n-1}{r-1} \int_0^\infty \left[ \sum_{i=r}^n \binom{n}{i} \left( \frac{x}{\delta} \right)^i \left( 1 - \frac{x}{\delta} \right)^{n-i} \right]^{k-1} x^{r-1}(1-x)^{n-r} dx \geq P^*
\]

If \( G(x) = 1 - e^{-x} \) for \( x \geq 0 \), we can find the smallest integer \( r \), say \( r_0 \), which satisfies \( \int_0^\infty H^{-1}_T \left( \frac{X}{\delta} \right) dH(x) \geq P^* \) where \( H(x) \) is the c.d.f. of a \( \chi^2 \) random variable with \( 2r \) d.f. Since \( \frac{1}{\delta} a_r \geq 1 \) iff \( n \geq \frac{r-1}{1-\delta} \), we can find the minimum \( n \) satisfying \( n \geq \max\{r_0, (r_0-1)/(1-\delta)\} \).

Acknowledgement.

The authors are very thankful Mr. Woo-Chul Kim of Purdue University for critically reading this paper and making several helpful comments and suggestions leading to an improvement of the paper.
REFERENCES


**Subset Selection Procedures for Restricted Families of Probability Distributions**

Shanti S. Gupta and Ming-Wei Lu

**Purdue University**
Department of Statistics
West Lafayette, IN 47907

Office of Naval Research
Washington, DC

Approved for public release, distribution unlimited.

Selection Procedures, Reliability, Convex, Star-shaped, IFR, IFRA, Total Life Statistic, Probability of Correct Selection.

In this paper we are interested in studying multiple decision procedure for $k(k > 2)$ populations which are themselves unknown but which one assumed to belong to a restricted family. We propose to study a selection procedure for distributions associated with these populations which are convex-ordered with respect to a specified distribution $G$ assuming there exists a best one. The procedure described here is based on a statistic $T_i = \sum_{j=1}^{r} a_j X_{i;j\cdot n}$ for
$i = 1, \ldots, k$ where $X_{i;j,n}$ is the $j$-th order statistic from $F_i$, $r$ is a fixed positive integer (1 ≤ $r$ ≤ $n$), $a_j = gG^{-1}(\frac{j-1}{n}) - gG^{-1}(\frac{j}{n})$ for $j=1, \ldots, r-1$, $a_r = gG^{-1}(\frac{r-1}{n})$ and $g$ is the density of $G$. This statistic $T_i$ was considered by Barlow and Doksum (1972). If $G(x) = 1 - e^{-x}$ for $x > 0$, then $nT_i = \sum_{j=1}^{r} X_{i;j,n}$ is the total life statistic until $r$-th failure from $F_i$. This shows that the procedure based on $T_i$ generalizes Patel's result (1976) for the IFR family.

The infimum of the probability of a correct selection is obtained and the asymptotic expression is also obtained using the subset selection approach. Some other properties of this procedure are discussed. We also study the asymptotic relative efficiencies of this rule with respect to some selection procedures proposed by Barlow and Gupta (1969) for the star-shaped ordered distributions, Gupta (1963) for the gamma populations with unknown shape parameters and etc. An example is given to illustrate the use of the selection procedure for the two independent uniform distributions. Application to quantile selection rules for distributions convex ordered with respect to Weibull distribution is given. A selection procedure for selecting the best population using the indifference zone approach is also studied.