Indian J. pure appl. Math., 12(1) : 1–6, January 1981

SELECTION RULES FOR IFRA POPULATIONS

JAYANT V. DESHPANDE AND GOBIND P. MEHTA

Department of Statistics, Panjab University, Chandigarh 160014

(*Received* 8 *March* 1980)

Let $\pi_1, ..., \pi_k$ be k populations. Let an observation from the population π_i have an IFRA distribution $F(x; \theta_i)$, $i = 1, 2, ..., k$, which are indentical except for the values of a scale parameter θ. The problem considered is that of selecting a subset of fixed size s from the k populations which includes at least c of the t populations associated with t largest (smallest) θ-values. The indifference-zone approach is followed. A generalization of a result of Doksum (1969) is used to show that the infimum of the probability of correct selection (PCS) in the exponential case provides a lower bound for the PCS in case of IFRA distributions.

1. INTRODUCTION

Let $\pi_1, ..., \pi_k$ be k populations. Let the nonnegative random variable X_i associated with π_i have a continuous IFRA (increasing failure rate average) distribution $F_i(x) = F(x; \theta_i)$, $0 < x < \infty$, $0 < \theta_i < \infty$, $i = 1, 2, ..., k$. We assume that for all $i(1 \leqslant i \leqslant k)$ the functional form of the distribution function (d.f.) $F_i(x)$ is the same but unspecified, and they differ only in the scale parameter θ. Here θ is a scale parameter in the sense that the distribution of the random variable X/θ is independent of θ.

Let the ordered values of θ_i's be denoted by $\theta_{(1)} \leqslant ... \leqslant \theta_{(k)}$ and let $\pi_{(i)}$ be the population corresponding to $\theta_{(i)}$, $i = 1, 2, ..., k$. We call the populations $\pi_{(k-t+1)}, ..., \pi_{(k)}$, the t best populations $(1 \leqslant t \leqslant k - 1)$. It is easily seen that the ordering of the populations according to θ_i's is also the same as the ordering according to their mean values. The problem considered is to develop statistical procedures, based on the observations, for selecting a subset of fixed size s from the k populations that includes at least c of the t best populations. This problem is feasible and non-trivial when s, c and t in relation to k are such that

$$\max(s, t) \leqslant k - 1 \quad \text{and} \quad \max(1, s + t + 1 - k) \leqslant c \leqslant \min(s, t).$$

Any selection, using a specified selection procedure, is regarded as correct selection (CS) if the selected subset, in fact, contains at least c of the t best populations. This formulation is the generalization due to Mahamunulu (1967) of the Bechhofer (1954) approach to the selection problem.

In literature, we find that selection with respect to the largest quantile of order $\alpha(0 < \alpha < 1)$ for certain restricted families (which include IFR and IFRA distributions) has been studied by Barlow and Gupta (1969). Their procedures select subsets of random size and are based on sample quantiles (for IFRA distributions) and the sample means (for IFR distributions). Earlier, in an abstract, Patel (1967) suggested a procedure for the problem of selecting a subset containing the population with the largest failure rate average out of several IFRA populations based on the number of failures observed by some fixed common time T. Recently, Patel (1976) has considered the selection problems relating to IFR populations which differ only in their scale parameter.

In this paper we have taken the indifference zone approach. A generalization of a result of Doksum (1969) is used to prove that the infimum of the probability of correct selection (PCS) over the preference-zone for the case of IFRA populations is greater than or equal to the corresponding infimum of the PCS for exponential populations provided the procedure is based on the statistics which satisfy certain conditions.

2. The Proposed Procedure and Its Probability of Correct Selection

Let $X_{i1}, \ldots, X_{in} (i = 1, 2, \ldots, k)$ be k independent random samples from the populations $\pi_1, \pi_2, \ldots, \pi_k$, respectively. Let $T_{in} = T(X_{i1}, \ldots, X_{in})$ be an appropriate statistic based on the sample from π_i. Let the ordered T_{in}'s be denoted by

$$T_{(1)n} \leqslant \ldots \leqslant T_{(k)n}.$$

Let Y_i be the statistic based on the sample from the population $\pi_{(i)}$. So the set $(Y_1, \ldots, Y_k)$ is some permutation of the set $(T_{1n}, \ldots, T_{kn})$ and $Y_{k-t+1}, \ldots, Y_k$ are the statistics associated with the t best populations.

Let $\underset{\sim}{\theta} = (\theta_{(1)}, \ldots, \theta_{(k)})$ denote a point in the parametric space Ω. Let $d^*(d^* > 1)$ be a specified constant. The space Ω is partitioned into a preference zone

$$\Omega(d^*) = \{\underset{\sim}{\theta} \in \Omega \mid (\theta_{(k-t+1)}/\theta_{(k-t)}) \geqslant d^*\}$$

and its complement $\bar{\Omega}(d^*)$, the indifference zone. Let $P^*(0 < P^* < 1)$ be a predetermined constant. A procedure R should then be such that the PCS satisfies the condition

$$P\{CS \mid \underset{\sim}{\theta}, R\} \geqslant P^* \quad \text{for all} \quad \underset{\sim}{\theta} \in \Omega(d^*).$$

We use the following selection procedure.

Procedure R — Select the set of s populations corresponding to $T_{(k-s+1)n}, \ldots, T_{(k)n}$. The PCS using this selection procedure is

$$P\{CS \mid \underset{\sim}{\theta}, R\} = P\{c^{th} \text{ largest of } (Y_{k-t+1}, \ldots, Y_k)$$

$$> (s - c + 1)^{st} \text{ largest of } (Y_1, \ldots, Y_{k-t}) \mid \underset{\sim}{\theta}\}.$$

Let the distribution function of Y_i be $G_n(\cdot\,; \theta_{(i)})$. We make the following assumptions regarding the statistics Y_i, $i = 1, \ldots, k$.

Assumption I — Y_i is an absolutely continuous random variable.

Assumption II — The family $\{G_n(\cdot\,; \theta_{(i)})\}$ of the d.f. is a stochastically increasing family for each positive integer n.

Assumption III — The statistic Y_i is a nondecreasing function of the observations from the population $\pi_{(i)}$, i.e.,

$$Y_i(z_1, \ldots, z_n) \leqslant Y_i(w_1, \ldots, w_n)$$

if $z_j \leqslant w_j$ for all $j (1 \leqslant j \leqslant n)$.

Under the assumptions I and II, by a result of Mahamunulu (1967, p. 1083), it follows that the infimum of the PCS over $\Omega(d^*)$ occurs when the configuration of the parameters is

$$\left.\begin{array}{ll} \theta_{(1)} = \ldots = \theta_{(k-t)} = \theta & \text{(say)} \\ \theta_{(k-t+1)} = \ldots = \theta_{(k)} = \theta^* & \text{(say)} \end{array}\right\} \qquad \ldots(2.1)$$

where $\theta^* = d^*\theta$. Under this configuration, $Y_1, \ldots, Y_{k-t}$ are i.i.d. random variables with distribution $G_n(\cdot\,; \theta)$ and each is based on a sample of size n from $F(x) = F(x; \theta)$ and $Y_{k-t+1}, \ldots, Y_k$ are i.i.d. random variables with distribution $G_n(\cdot\,; \theta^*)$ and each is based on a sample of size n from $F(x) = F(x; \theta^*) = F(x/d^*)$. Thus

$$\inf_{\underline{\theta} \in \Omega(d^*)} P\{CS \mid \underline{\theta}, R\}$$

$$= \inf_{\theta} \frac{t!}{(t-c)!\,(c-1)!} \sum_{\alpha=0}^{s-c} \binom{k-t}{\alpha} \int_0^\infty [G_n(x; \theta)]^{k-t-\alpha}$$

$$\times\, [1 - G_n(x; \theta)]^{\alpha}\, [G_n(x; \theta^*)]^{t-c}\, [1 - G_n(x; \theta^*)]^{c-1}\, dG_n(x; \theta^*). \qquad \ldots(2.2)$$

The following generalization of Theorem 3.1 of Doksum (1969) is used to prove that the infimum of the PCS (2.2) for IFRA populations is greater than or equal to the corresponding infimum of the PCS for exponential populations.

Lemma 2.1 — Let $u_1, \ldots, u_M$ and $v_1, \ldots, v_N$ be two independent random samples from the populations with continuous d.f. $F(\cdot)$ and $F^*(\cdot)$, respectively. Let

$$F^*(x) = F(x/d^*),\ 0 < x < \infty,\ d^* \geqslant 1.$$

Let K be a d.f. such that $K(0) = 0$, $F \underset{*}{<} K$ (that is, $K^{-1}F(x)$ is starshaped on the support of F); $K^*(x) = K(x/d^*)$ and $K(x) < 1$ for each $x < \infty$, or $F(x) < 1$ for each $x < \infty$. Let $\varphi(\mathbf{u}, \mathbf{v}) = \varphi(u_1, \ldots, u_M; v_1, \ldots, v_N)$ be a function such that

$$\text{(i)} \quad \varphi(u_1, \ldots, u_M; v_1, \ldots, v_N) \leqslant \varphi(u_1, \ldots, u_M; v_1^*, \ldots, v_N^*) \text{ if } v_j^* \geqslant v_j, \quad j = 1, 2, \ldots, N,$$

and

$$\text{(ii)} \quad \varphi(u_1, \ldots, u_M; v_1, \ldots, v_N) = \varphi(g(u_1), \ldots, g(u_M); g(v_1), \ldots, g(v_N))$$

where $g(x) = K^{-1}F(x)$.

Then

$$E_{F,F^*} \{\varphi(\mathbf{u}, \mathbf{v})\} \geqslant E_{K,K^*} \{\varphi(\mathbf{u}, \mathbf{v})\}$$

for each $d^* \geqslant 1$.

The proof follows on lines similar to Theorem 3.1 of Doksum (1969).

Using this lemma we prove the following theorem:

Theorem 2.1 — The infimum of the PCS (2.2) for distributions $F(x; \cdot)$ is greater than or equal to the corresponding infimum of the PCS for distributions $K(x; \cdot)$ if (i) $F \underset{*}{<} K$ and (ii) the order of Y_i's is preserved when the observations are transformed by the function $g(x) = K^{-1}F(x)$, i.e.,

$$Y_i(x'_{i1}, \ldots, x'_{in}) \leqslant Y_j(x'_{j1}, \ldots, x'_{jn})$$

implies

$$Y_i(g(x'_{i1}), \ldots, g(x'_{in})) \leqslant Y_j(g(x'_{j1}), \ldots, g(x'_{jn})) \qquad \ldots(2.3)$$

for $i, j = 1, 2, \ldots, k$, where $x'_{i1}, \ldots, x'_{in}$ is a random sample from the population $\pi_{(i)}$, $i = 1, 2, \ldots, k$.

Proof : Under the configuration (2.1), denote the $M = n(k - t)$ observations from F by $u_1, \ldots, u_M$ and $N = nt$ observations from F^* by $v_1 \ldots v_N$, where

$$F^*(x) = F(x/d^*).$$

Define the function $\varphi(\mathbf{u}, \mathbf{v})$ as follows:

$$\begin{aligned} \varphi(\mathbf{u}, \mathbf{v}) &= 1 \text{ if } c^{th} \text{ largest of } (Y_{k-t+1}, \ldots, Y_k) \\ &> (s - c + 1)^{st} \text{ largest of } (Y_1, \ldots, Y_{k-t}), \\ &= 0 \text{ otherwise.} \end{aligned} \qquad \ldots(2.4)$$

Then

$$\inf_{\underline{\theta} \in \Omega(d^*)} P\{CS \mid \underline{\theta}, R\} = E_{F,F^*}\{\varphi(\mathbf{u}, \mathbf{v})\}.$$

Since $F \underset{*}{<} K$ and the function φ defined by (2.4) satisfies the requirements of Lemma 2.1, the result follows.

Corollary 2.1 — The infimum of the PCS (2.2) for IFRA populations is greater than or equal to the corresponding infimum of the PCS for exponential populations if the statistics Y_i satisfy requirement (2.3) with $g(x) = -\log[1 - F(x)]$.

PROOF : F is an IFRA distribution if and only if F is star ordered with respect to the standard exponential distribution. Hence, the result follows from the above theorem by taking K to be the standard exponential distribution and F to be any IFRA distribution.

3. EXAMPLES AND REMARKS

Example 3.1 — Consider the statistic $T_{in} = X_{i(\alpha)}$, the α^{th} order statistic from the ith sample, $1 \leqslant \alpha \leqslant n$, $i = 1, 2, \ldots, k$. This statistic satisfies assumptions I, II, and III as well as the requirement (2.3). Hence, for the selection procedure R based on these statistics, the infimum of the PCS for IFRA populations will be greater than or equal to the corresponding infimum of the PCS for exponential populations. This lower bound is, in fact,

$$P\{c^{th} \text{ largest of } (d^* Z_{k-t+1(\alpha)}, \ldots, d^* Z_{k(\alpha)}) > (s - c + 1)^{st} \text{ largest of } (Z_{1(\alpha)}, \ldots, Z_{k-t(\alpha)})\} \quad \ldots(3.1)$$

where $Z_{i(\alpha)}(i = 1, 2, \ldots, k)$ is the α^{th} order statistic based on a sample of size n from the standard exponential distribution and are thus independent and identically distributed. Let $K_\alpha(x)$ denote the d.f. of $Z_{i(\alpha)}$, $i = 1, 2, \ldots, k$. Then the d.f. of $d^* Z_{i(\alpha)}$ is $K_\alpha(x/d^*)$, $i = k - t + 1, \ldots, k$. Hence the lower bound (3.1) is the RHS of (2.2) with $G_n(x; \theta)$ and $G_n(x; \theta^*)$ replaced by $K_\alpha(x)$ and $K_\alpha(x/d^*)$, respectively. An alternative way to obtain the result, for the above T_{in}'s, is to extend the approach of Patel (1976) to cover the IFRA situation.

Example 3.2 — Let $T_{in} = \sum_{\alpha=1}^{n} a_\alpha X_{i\alpha}$, with $a_\alpha \geqslant 0 \; \forall \; \alpha$ and $\alpha_\alpha > 0$ for some α, $i = 1, 2, \ldots, k$. It is seen that this statistic satisfies assumptions I, II, and III. If in addition $g(x) = K^{-1}F(x) = px - q$, $p \geqslant 0$, $q \geqslant 0$, then $F \underset{*}{<} K$ and this statistic satisfies requirement (2.3) as well.

REFERENCES

Barlow, R. E., and Gupta, S. S. (1969). Selection procedures for restricted families of distributions. *Ann. Math. Statist.*, **40**, 905–17.

Bechhofer, R. E. (1954). A single sample multiple deiscion procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.*, **25**, 16–39.

Doksum, K. (1969). Star-shaped transformations and the power of rank tests. *Ann. Math. Statist.*, **40**, 1167–76.

Mahamunulu, D. M. (1967). Some fixed-sample ranking and selection problems. *Ann. Math. Statist.*, **38**, 1079–91.

Patel, J. K. (1967). Selecting a subset containing the best one of several IFRA populations—Abstract. *Ann. Math. Statist.*, **38**, 957.

———— (1976). Ranking and selection of IFR populations based on means. *J. Am Statist. Assoc.*, **71**, 143–46.