Indian Journal of Mathematics and Mathematical Sciences Vol. 6, No. 2, (Dec. 2010) : 233-239

ON SELECTING BEST IFRA POPULATION: A LARGE SAMPLE APPROACH

Parameshwar V. Pandit & C. M. Math

ABSTRACT

The principle problem of ranking and selection is to select the 'best population' among the several populations. The criteria to select best population are usually based on the some parameters of the underlying family. One such way is to select a population that belongs to some parametric family. In this paper, we consider the problem of selecting the 'Best' increasing failure rate(IFRA) population among the several IFRA populations. A procedure to select the 'Best' IFRA population is developed based on a measure of departure from exponentiality towards IFRA. The measure used here was considered by Deshpande(1983) for the problem of testing exponentiality against IFRA alternatives in an one sample setting. Our procedure of selection is based on U-statistics and its large sample properties. Some applications of the selection procedure is also indicated.

Keywords and Phrases: *Increasing failure rate average(IFRA), Selection and ranking, U-statistic, Probability of correct selection.*

1. INTRODUCTION

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k independent populations with continuous distribution functions $F_1, F_2, \ldots, F_k$. The problem is to select the 'Best' population in some sense among the given populations. The criteria to select best population are usually based on the some parameters of the underlying family. One such way is to select a population that belongs to some parametric family. Some selection procedures that use nonparametric approach includes Lehmann's (1963) procedure based on ranks, Barlow and Proschan's (1969) approach based on partial orderings of probability distributions. Further, Patel (1976) gave a selection procedure based on means for selecting from class of increasing failure rate(IFR) distributions. One can refer Gupta and Panchapakesan (1979) for a good review of ranking and selection problems.

Here, we consider selection problems for the IFRA classes of distributions. We first give definition of distribution of increasing failure rate average (IFRA).

Definition1.1: A life distribution F is said to be IFRA, if $\bar{F}(bx) \geq \bar{F}(x)$ $x \geq 0$ and $0 < b < 1$ where $\bar{F} = 1 - F$. (1)

Note that equality in (1) is obtained, when F is exponential. Deshpande (1983) gave

$$\gamma(F) = \int_0^\infty \bar{F}(bx)\,dF(x)$$

as a measure of IFRA-ness and used it to develop a test for testing exponentiality against IFRA distributions. When F is exponential, $\gamma(F) = \frac{1}{b+1}$ and if F is IFRA $\gamma(F) \geq \frac{1}{b+1}$.

Definition 1.2: A life distribution F is said to possess more IFRA-ness property than that of a life distribution G, if $\gamma(F) \geq \gamma(G)$.

The above distribution is applicable to any pair of IFRA distributions. Here we use the value of $\gamma(F)$ as the criterion for selecting the best distribution possessing IFRA-ness property among k IFRA populations. Therefore, for the purpose of validity, we assume that the underlying distributions are continuous. Here the interest is in applications of the procedure to the classes of IFRA distributions.

The selection procedure here is based on the U-statistics $J_n(F)$, an unbiased estimate of $\gamma(F)$ which is defined as below:

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from F. Define

$$J_n(F) = \frac{1}{\binom{n}{2}} \sum\sum_{i \neq j} h(X_i, bX_j)$$

Where $$h(X_i, bX_j) = \frac{1}{2}\,[I(X_i > bX_j) + I(X_j > bX_i)]$$

and $$I(A) = \begin{cases} 1 & \text{if } \ x \in A \\ 0 & \text{otherwise} \end{cases}.$$

Here summation is taken over all $\binom{n}{2}$ combinations of integers (i, j) taken out of integer $\{1, 2, 3, \ldots, n\}$.

The statistics $J_n(F)$ is used by Despande(1983) for testing exponentiality against IFRA and by Zallaikar and Tiwari (1988) to test whether F is more IFRA than G.

In this paper, we utilize the asymptotic normality of $J_n(F)$ for selecting the most IFRA distribution among k IFRA distributions.

2. SELECTION PROCEDURE

Let $\pi_1, \pi_2, ..., \pi_k$ denote the k population with unknown IFRA distribution function $F_1, F_2, ..., F_k$ respectively and hence the function form of F_i and the IFRA-ness measure $\gamma(F_i)$ of F_i as defined earlier are unknown, but it is assumed that F_i are continuous. For the sake of convenience, we denote by $\gamma(F_i)$. γ_i The goal is to select the population which has largest $\gamma_{[k]}$, where $\gamma_{[1]} \leq \gamma_{[2]} \leq ... \leq \gamma_{[k]}$ denote the ordered IFRA-ness measure for k distributions. Selecting largest $\gamma_{[k]}$ is to select the most IFRA distribution.

Let $\underline{\gamma} = (\gamma_{[1]}, \gamma_{[2]}, ..., \gamma_{[k]})$ and $\underline{\Omega} = \left\{\underline{\lambda} : \frac{1}{b+1} \leq \gamma_{[1]} \leq \gamma_{[2]} \leq \ ... \leq \gamma_{[k]} \leq 1\right\}$ be the parameter space which is partitioned into a preference zone $\Omega(\delta^*)$ and an indifference zone $\Omega - \Omega(\delta^*)$, where $\Omega(\delta^*)$ is defined by $\Omega(\delta^*) = \{\underline{\gamma} : \gamma_{[k]} - \gamma_{[1]} \geq \delta^*\}$.

The quantity and are pre-assigned by the experiment and selection procedure R is required to satisfy the condition

$$P(CS|R) \geq P^*, \text{ for all } \underline{\gamma} \in \Omega(\delta^*). \qquad \text{(A)}$$

Selection of any population with $\gamma_{[k]}$ is regarded as the correct selection $[CS]$ and condition (A) is referred to as P^* condition.

The selection procedure here is based on the U-Statistics $J_n(F)$ and utilize the large sample properties $J_n(F)$. The asymptotic distribution of $\sqrt{n}\,(J_n(F) - \gamma(F))$ is normal with mean zero and variance $4\xi_1(F)$ (see Deshpande(1983)).

Since, the asymptotic distributions of each $J_n(F)$ is normal with mean $\gamma(F)$ and variance $4\xi_1(F_i)/n$, the problem of selecting the more IFRA population can be treated selection of the largest mean of the normal population.

A strongly consistent estimator of $\xi_1(F)$ is given by

$$\hat{\xi}_1(F) = \frac{1}{n-1}\sum_{i=1}^{n} (\hat{h}(x_i) - J_n(F))^2 ,$$

where $$\hat{h}_1(x_i) = \frac{1}{n-1}\sum_{j} h(x_i, bx_j) .$$

Now, we propose the following two stage selection procedure R, assuming large samples.

Let $t = t(k, p^*) > 0$ be the unique solution of the equation

$$\int_{-\infty}^{\infty} \Phi^{k-1}(z+t)\,d\Phi(z) = p^* ,$$

Procedure *R* : Take an initials sample $X_{i_1}, X_{i_2}, ..., X_{in_0}$ of size n_o from the population π_{i_1}, $i = 1, 2, ..., k$ and compute $J_{n_0}(F_i)$ and $\hat{\xi}_{1n_0}(F_i)$, define $n_i = \max(2n_0, [1/c])$ where $[x]$ denote the smallest integer which is greater than or equal to x and $\frac{1}{c} = 4\hat{\xi}_{1n_0}(F_i)\left(\frac{1}{\delta^*}\right)^2$.

The second stage sample size from π_i denote by n_i is determined as follows:

$$n' = \begin{cases} 0 & \text{if } \left[\frac{1}{c}\right] \le n_0 \\ n_i - n_0 & \text{if } \left[\frac{1}{c}\right] > n_0 \end{cases}.$$

Compute $J_{n_i}(F_i)$ based on n_i, the additional sample taken from π_i and define

$$U_i = a_i J_{n0}(F_i) + (1 - a_i) J_{n_i}(F_i),$$

where $0 < a_i < 1$ is determined to satisfy $4\hat{\xi}_{1n_0}(F_i)\left[\frac{a_i^2}{n_0} + \frac{(1-a_i)^2}{n_i}\right] = \left(\frac{\delta^*}{t}\right)^2$.

Here $4\hat{\xi}_{1n_0}(F_i)\left[\frac{a_i^2}{n_0} + \frac{(1-a_i)^2}{n_i}\right] = \left(\frac{\delta^*}{t}\right)^2$, is strongly consistent estimator for variance of U_i. Note that no second sample is taken from the i^{th} population, that is $n_i = 0$,we define $(1 - a_i)J_{n_i}(F) = 0$ and $(1 - a_i^2)n_i = 0$.

Lemma: There exists a_i satisfying

$$4\hat{\xi}_{1n_0}(F_i)\left[\frac{a_i^2}{n_0} + \frac{(1-a_i)^2}{n_i}\right] = \left(\frac{\delta^*}{t}\right)^2.$$

Proof:

Define
$$a_i = \begin{cases} 1 & \text{if } \frac{1}{c} \le n_0 \\ \frac{1}{2}(1 + \sqrt{(2n_0 c - 1)}) & \text{if } n_0 < \frac{1}{c} \le 2n_0 \\ \frac{n_0}{n_1} & \text{if } \frac{1}{c} > 2n_0 \end{cases}.$$

Then, it is straight forward to show that a_i defined above satisfies

$$4\xi_{1n_0}(F_i)\left[\frac{a_i^2}{n_0} + \frac{(1-a_i)^2}{n_i}\right] = \left(\frac{\delta^*}{t}\right)^2 \text{ because } 4\xi_{1n_0}(F_i)\left(\frac{t}{\delta^*}\right)^2 = \frac{1}{c}.$$

It is to be noted that a_i can also be chosen to be $(1-(\sqrt{2n_0c-1}))/2$ and $(1+(\sqrt{2n_0c-1}))/2$. In such a case, the initial sample size n_0 is equal to the additional sample size n_i and the coefficients a_i and $(1-a_i)$ become interchangeable.

Theorem: For any p^*, $p^* \in \left(\frac{1}{k}, 1\right)$ there exists n_0 large enough such that $\inf_{\Omega(\delta^*)} p(CS|R) \cong p^*$.

Proof: For any i, we can write

$$P\left(\frac{U_i-\gamma_i}{(\delta^*/t)} \le y\right) = P\left[\frac{U_i-\gamma_i}{\sqrt{\left(4\xi_{1n_0}(F_i)\left[\frac{a_i^2}{n_0}+\frac{(1-a_i)^2}{n_i}\right]\right)}} \le y\right]$$

$$= E_{\hat{\xi}_{1n_0}(F_i)} P\left[\frac{U_i-\gamma_i}{\sqrt{\left(4\xi_{1n_0}(F_i)\left[\frac{a_i^2}{n_0}+\frac{(1-a_i)^2}{n_i}\right]\right)}} \Big| \hat{\xi}_{1n_0}(F_i)\right]$$

$$= E_{\hat{\xi}_{1n_0}(F_i)} P\left[\frac{U_i-\gamma_i}{\sqrt{\left(4\hat{\xi}_{1n_0}(F_i)\left[\frac{a_i^2}{n_0}+\frac{(1-a_i)^2}{n_i'}\right]\right)}} \le y\,\frac{\sqrt{\hat{\xi}_{1n_0}(F_i)}}{\sqrt{\xi_{1n_0}(F_i)}} \Big| \hat{\xi}_{1n_0}(F_i)\right]$$

$$\cong E_{\hat{\xi}_{1n_0}(F_i)} \Phi\left[y\,\frac{\sqrt{\hat{\xi}_{1n_0}(F_i)}}{\sqrt{\xi_{1n_0}(F_i)}} \Big| \hat{\xi}_{1n_0}(F_i)\right] \cong \Phi(y) \qquad \text{(B)}$$

since $\left[\frac{\sqrt{\hat{\xi}_{1n_0}(F_i)}}{\sqrt{\xi_{1n_0}(F_i)}}\right] \to 1$.

Let $U_{(i)}$ denote the statistic corresponding to the population having parameter $\gamma_{[i]}$, $i = 1, 2, \ldots, k$.

Then, we have

$$P(CS|R) = P[U_{(i)} \le U_{(k)}, \quad i = 1, 2, \ldots, k-1]$$

$$= P[U_{(i)} - \gamma_{[i]} \le \underline{U}_{(k)} - \gamma_{[k]} + \gamma_{[k]} - \gamma_{[i]}] \quad i = 1, 2, \ldots, k-1$$

$$= P\left[\frac{U_{(i)}-\gamma_{[i]}}{\delta^*/t} \le \frac{U_{(k)}-\gamma_{(k)}}{\delta^*/t} + \frac{\gamma_{[k]}-\gamma_{[i]}}{\delta^*/t}, \quad i=1,2,3,\ldots,k-1\right]$$

$$\cong \int_{-\infty}^{\infty} \pi_{i=1}^{k-1} \Phi\left(z + \frac{\gamma_{[k]} - \gamma[i]}{\delta^*/t}\right) d\Phi(z) \qquad \text{(B)'}$$

$$\ge \int_{0}^{\infty} \Phi^{k-1}\left(z + \frac{\delta^*}{\delta^*/t}\right) d\Phi(z) \qquad \text{(C)}$$

$$= p^* \qquad (D)$$

Since $U_{(i)}$'s are independent.

The inequality (C) is true since the right hand side of (B') is minimized, when $\gamma_{[1]} = \gamma_{[2]} = \ldots = \gamma_{[k-1]} = \gamma_{[k]} - \delta^*$ and (D) follows from

$$\int_{-\infty}^{\infty} \Phi^{k-1}(z+t)\, d\Phi(z) = p^*.$$

To select the most IFRA distribution, we select the distribution which yields $T_{[k]}$. In this case the preference zone is defined for fixed δ^* as

$$\{\gamma : \gamma_{[k]} - \gamma_{[k-1]} \ge \delta^*\}, \quad \delta^* > \frac{1}{b+1}.$$

The probability of correct selection is given by

$$P[U_{(k)} \ge T_{[i]}; \quad i = 1, 2, \ldots, k].$$

This probability of Correct Selection is minimized when $\gamma_{[1]} = \gamma_{[2]} = \ldots = \gamma_{[k-1]} = \gamma_{[k]} - \delta^*$ and minimum value is the assigned P^* used compute t.

$$\int_{-\infty}^{\infty} \Phi^{k-1}(z+t)\, d\Phi(z) = p^*.$$

3. APPLICATIONS

The application of the procedure R to select the "best" distribution according to its IFRA-ness to a number of interesting situations, for which no other selection procedures are available, makes it very significant. It can be used to select among

k given life distributions, with the same functional form but different shape parameters, the distribution with the largest mean life or the distribution with the largest mean residual life at a fixed time $t_0 > 0$. The procedure can be used to select the best of k IFRA distributions all of which have the same mean. The procedure is applicable in selecting the 'Best' among k distributions with scale and shape parameters, namely. Gamma, Weibull etc and that among k linear failure rate distributions.

REFERENCES

[1] Barlow R. E., and Gupta S. S., (1969), Selection Procedures for Restricted Families of Distributions, *Ann. Math. Statisti.*, **40**, 905-917.

[2] Bechhofer R. E., (1954), A Single Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances, *Ann. Math. Statisti.*, **25**, 16-39.

[3] Deshpande J. V., (1981), A Class of Tests for Exponentiality Against Increasing Failure Rate Average Alternatives, *Biometrika*,**70**(2), 514-518.

[4] Dudewicz E. J., (1971), Non-Existence of a Single-Sample Selection Procedure whose *P*(*CS*) is Independent of the Variances, *S. Afr. Staist. J.*, **5**, 37-78.

[5] Dudewicz E. J., and Dalal S. R.,(1975), Allocation of Observations in Ranking and Selection with Unequal Variances, *Sankhya Ser. B.*, **37**, 28-78.

[6] Gupta S. S., and Panchapakesan S., (1979), Multiple Decision Procedures: *Theory and Methodology of Selecting and Ranking Populations*. Wiley New York.

[7] Patel J. K., (1976), Ranking and Selection of IFR Populations Based on Means, *J. Amer. Statist. Assoc.*, **71**, 143-146.

[8] Park D. H., and Lal Sexena K. M., (1987), Nonparametric Procedure for Selecting the Least NBU Population: Large Sample Approach, *Comm. Statist.-Theory Meth.*, **16**(11), 3359-3377.

[9] Tiwari R. C., and Zalkikar J. N., (1988), Testing whether *F* is 'More IFRA than is *G*', *Microelectronics Reliabilty*, **28**(5), 703-712.

Parameshwar V. Pandit
Department of Statistics,
Bangalore University,
Bangalore -560056, India.
E-mail: panditpv12@gmail.com

C. M. Math
Department of Statistics,
K.L.E.'s G.H.College, Haveri.
E-mail: chandrashekar.math@gmail.com