STATISTICA, anno LXXVIII, n. 1, 2018

INDUCED RANKED SET SAMPLING WHEN UNITS ARE INDUCTED FROM SEVERAL POPULATIONS

P. Yageen Thomas
Department of Statistics, University of Kerala, Trivandrum-695 581, India
Anne Philip [1]
Department of Statistics, University of Kerala, Trivandrum-695 581, India

1. Introduction

Random sampling is well known as an unrestricted method of collecting the units from a population. However, if collecting the units and measuring the characteristic of interest on them is expensive (or risky or painful), then one is compelled to look for alternative sampling methods which are capable of accommodating observational economy considerations. McIntyre (1952) first used a judgement method to rank the randomly chosen units within each of independent groups of units and devised a method of final selection of units from the groups of units based on their ranks and termed this method as ranked set sampling (RSS). Chen *et al.* (2004) have discussed extensively about the observational economy considerations accommodated in RSS as defined by McIntyre (1952). For more details, see Shaibu and Muttlak (2004), Al-Rawwash *et al.* (2010) and Samuh (2017). Ranking by judgement method is not suitable if there is a fear that ranking error which is otherwise known as imperfect ranking creeps in the ranking process of the units. In such situations Stokes (1977) introduced a scheme of sampling the units based on the measurements made on an easily measurable auxiliary variable X which is jointly distributed with the variable Y of primary interest, whose measurement is expensive and thereby defined the RSS in the following manner. Choose n^2 units randomly from the population and arrange them in the random order in n sets of n units each. Rank units within each set based on the measurement made on the auxiliary variable X of the units. Then from the ith set choose the unit ranked i and measure the variable Y of primary interest on this unit for $i = 1, 2, \ldots, n$. Clearly, the observation obtained from the ith set is the concomitant of ith order statistic of X-observations in that set and we write it as $Y_{[i:n]i}$ for $i = 1, 2, \ldots, n$. Then the observations $Y_{[1:n]1}, Y_{[2:n]2}, \ldots, Y_{[n:n]n}$ are said to

[1] Corresponding Author. E-mail: anne04philip@gmail.com

constitute the ranked set sample (rss) as proposed by Stokes. The sampling procedure as described above to obtain the rss is known as the Stokes method of RSS.

Since there is potential difference in the approaches and mathematical preciseness in the RSS methods proposed by McIntyre (1952) and Stokes (1977), rather than calling both procedures as RSS, we prefer to call the RSS method proposed by Stokes as induced ranked set sampling (IRSS) all through this paper. We may further call the sample arising due to IRSS as induced ranked set sample ($irss$).

It is to be noted that contrary to general perception, concomitants of order statistics generated from IRSS are not causing any disadvantage but create advantages by the mathematical rigour we observe on their distributions for devising inference procedures on the parameters involved in the distribution of Y. Distribution theory of concomitants of order statistics in applying IRSS method in inference problems is available in David (1973), Bhattacharya (1984) and David and Nagaraja (1998, 2003). For some recent survey on IRSS, one may refer to Chen *et al.* (2004), Chacko and Thomas (2007, 2008, 2009), Ahmad *et al.* (2010), Lesitha *et al.* (2010), Lesitha and Thomas (2013), Singh and Mehta (2013) and Philip and Thomas (2015).

Taking the advantage of recent developments on the theory of order statistics of independent non-identically distributed (INID) random variables as portrayed in Vaughan and Venables (1972), Balakrishnan (1988), Balakrishnan *et al.* (1992), Bapat and Beg (1989a, 1989b), Beg (1991) and Samuel and Thomas (1998), some applications of these theories in parameter estimation have been illustrated in Sajeevkumar and Thomas (2005) and Thomas and Sajeevkumar (2005). One difficulty experienced in the large scale applications of the results of Thomas and Sajeevkumar (2005) and Sajeevkumar and Thomas (2005) in further inference problems is about the tediousness involved in the evaluation of the covariance between different pairs of order statistics using the permanent expression (for details, see Vaughan and Venables, 1972) for the joint probability density function (pdf) of two order statistics of INID random variables. However, if we make use of McIntyre (1952) method of RSS involving selection of units belonging to different populations for inference problems, then we are redeemed from the burden of obtaining the values of the covariances as those covariances are zero in a rss due to the reason that the observations in the rss occur from independent samples. With this theoretical background Priya and Thomas (2013, 2016) have defined RSS when units from different univariate populations are to be considered and used the resulting rss observations to estimate the common parameters of several univariate distributions.

Eryilmaz (2005) first derived the expression for the cumulative distribution function (cdf) of concomitants of order statistics of INID random variables. It may be noted that for using concomitants of order statistics of INID random variables, we require the pdf of those concomitants. Thus Veena and Thomas (2015) have derived the marginal pdfs and joint pdfs of concomitants of order statistics of INID random variables using permanents. For some of the applications of concomitants of order statistics of INID random variables in estimation, see Veena and Thomas (2015). But there is a limitation in large scale applications of the results of Veena and Thomas (2015) as the expansion of the permanent expression for the joint pdf of different pairs of concomitants of order statistics and computation of those covariances turns out to cause unbearable burden. However, like the extension made by Priya and Thomas (2013, 2016) to McIntyre's RSS to the case when units from different univariate populations are inducted in the sampling scheme, if one extend IRSS scheme and the pdf representation of concomitants of order statistics of INID random variables as given in Veena and Thomas (2015), then the resulting $irss$ relieves the user from the burden of computation of covariances of different pairs of observations while applying IRSS to inference problems. This has motivated the authors to define IRSS to the case when units from several populations are to be inducted in the sample. It is to be noted that unlike the possibility of occurrence of imperfect ranking in the extended method of RSS proposed by Priya and Thomas (2013, 2016), such imperfect ranking never creeps in the proposed extended method of IRSS in this paper.

In Section 2, the procedure for the induced ranked set sampling when there are several populations is described. In Section 3, best linear unbiased estimation of a common threshold parameter θ_2 of two bivariate Pareto distributions is discussed based on an $irss$ of size n, when a sample of size n_1 is drawn from a bivariate Pareto population with shape parameter a_1 and a sample of size n_2 is drawn from another bivariate Pareto distribution with shape parameter a_2, where $n = n_1 + n_2$. The best linear unbiased estimator (BLUE) of θ_2 based on a lower extreme induced ranked set sample ($leirss$) along with its variance is also obtained in this section. Further, in this section, the relative precision of these BLUEs of θ_2 is discussed in relation to an existing BLUE based on concomitants of order statistics. The results of the third section are illustrated with a real life data in Section 4.

2. Induced ranked set sampling when there are several populations

In certain problems of investigation such as crop cutting experiments wherein average yield has to be estimated from a region where different high yielding varieties of a cereal (like rice or wheat) are cultivated so that the mean yield may be more or less identical for those varieties, but the distribution of the yield characteristic differs from one variety to another. For similar descriptions, see Sajeevkumar and Thomas (2005) and Thomas and Sajeevkumar (2005). The extended version of the IRSS which we like to introduce applies to the above mentioned type of situations.

To introduce the IRSS we assume that there are k independent bivariate populations all having a common support set and on the units from each population there are two measurable characteristics X and Y of which X is easily measurable without any cost while the measurement on the characteristic Y of primary interest is costly (or painful). Now we define the following.

DEFINITION 1. *Draw randomly n sets of n units each from the given k populations ($k \geq 2$) such that each set consists of n_i units drawn randomly from the ith population for $i = 1, 2, \ldots, k$ with $n = \sum_{i=1}^{k} n_i$. Measure the characteristic X (an auxiliary variable) on the units which can be measured easily and rank the units within each set based on the measured* *values of X on the units. Now from the jth set choose the unit which is ranked j among its units for the measurement of the characteristic Y of primary interest and let it be denoted by* $Y_{[j:n]j}$ *for* $j = 1,2,\ldots,n$. *Then* $Y_{[j:n]j}$, $j = 1,2,\ldots,n$ *are said to constitute the irss and the sampling scheme which generates this sample is called IRSS from the populations.*

Suppose that each of the given k populations is infinite such that the bivariate characteristic (X_i, Y_i) on the units of ith population has an absolutely continuous bivariate distribution with joint cdf $F_i(x,y)$ and joint pdf $f_i(x,y)$ for $i = 1,2,\ldots,k$. Let the marginal cdf of X_i be $F_{X_i}(x)$ with marginal pdf $f_{X_i}(x)$ and the marginal cdf of Y_i be $F_{Y_i}(y)$ with marginal pdf $f_{Y_i}(y)$, where $i = 1,2,\ldots,k$. If we apply IRSS from the k populations as specified above using Definition 1, then the observation $Y_{[r:n]r}$ is distributed identically as concomitant of the rth order statistic of X-observations in the pooled sample of all $n = \sum_{i=1}^{k} n_i$ observations, comprising of n_i independent observations drawn from $F_i(x,y)$ for $i = 1,2,\ldots,k$. Then from Veena and Thomas (2015), the pdf of $Y_{[r:n]r}$ is given by

$$f_{Y_{[r:n]r}}(y) = \frac{1}{(r-1)!(n-r)!}\int_{-\infty}^{\infty} per\big[\underset{r-1}{\mathbf{a}} \quad \underset{n-r}{\mathbf{b}} \quad \underset{1}{\mathbf{c}}\big]dx, \tag{1}$$

where $per\big[\underset{r-1}{\mathbf{a}} \quad \underset{n-r}{\mathbf{b}} \quad \underset{1}{\mathbf{c}}\big]$ denotes the permanent of a square matrix of order n in which the first $(r-1)$ columns are the $(r-1)$ copies of the vector $\mathbf{a} = (F_{X_1}(x),\ldots,F_{X_1}(x),F_{X_2}(x),\ldots,F_{X_2}(x),\ldots,F_{X_k}(x),\ldots,F_{X_k}(x))'$, next $(n-r)$ columns are the $(n-r)$ copies of the vector $\mathbf{b} = (1-F_{X_1}(x),\ldots,1-F_{X_1}(x),1-F_{X_2}(x),\ldots,1-F_{X_2}(x),\ldots,1-F_{X_k}(x),\ldots,1-F_{X_k}(x))'$ and the last column vector is defined by $\mathbf{c} = (f_1(x,y),\ldots,f_1(x,y),f_2(x,y),\ldots,f_2(x,y),\ldots,f_k(x,y),\ldots,f_k(x,y))'$. In each of the above vectors there are k varieties of symbols (functions) repeated for $n_1,n_2,\ldots,n_k$ times respectively. Also, the permanent of a square matrix has an expansion similar to that of the corresponding determinant except that the sign attached to all terms in the expansion is positive. Clearly, for any positive integer l using (1) we define $E\left(Y^l_{[r:n]r}\right)$ as

$$E\left(Y^l_{[r:n]r}\right) = \frac{1}{(r-1)!(n-r)!}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^l\, per\big[\underset{r-1}{\mathbf{a}} \quad \underset{n-r}{\mathbf{b}} \quad \underset{1}{\mathbf{c}}\big]dx\,dy. \tag{2}$$

From (2) we can compute the mean $E(Y_{[r:n]r})$ and $\mathrm{Var}(Y_{[r:n]r})$ for $r = 1,2,\ldots,n$ of all $irss$ observations. It may be noted from the definition of IRSS as given in Definition 1, we observe that $\mathrm{Cov}(Y_{[r:n]r}, Y_{[s:n]s}) = 0$, for $r \neq s$.

In some situations we come across different bivariate distributions having identical value for the location and scale parameters but differ either in the form or in the association (or shape) parameter. For example, suppose we have k high yielding varieties of paddy and for each variety, observations on (X,Y), where Y represents the yield per plot and X, the time taken to reach the reaping level of maturity, are of interest in which X is considered as a easily observable character. Note that as all k varieties belong to high yielding group, the mean yield of all varieties will be equal and the standard deviation of the yield of all varieties again equal and distribution of (X,Y) differs from one variety to another. As another example, suppose we have $k\,(k \geq 2)$ bivariate Pareto distributions on (X,Y) where X represents the salary or professional income and Y represents the income from other sources. Clearly, X is well defined and the threshold parameter of the distribution of X is identically the same for all k income groups in terms of income tax rules whereas the measurement of Y is difficult or under reported. So to study the income distribution in such a situation, estimation of Y based on ranks assigned to observations on X is very useful. We may have in this case different bivariate Pareto models by assuming different values on the shape parameter of the marginal random variables involved.

In this paper we apply IRSS to problems where the above described type of situations arise. Since the examples described above involve k populations with $k \geq 2$, we develop the theory of IRSS by assuming that there are k populations. When we apply the $irss$ to inference problems, to compute the mean and variance of the observations the result of the following theorem is much helpful.

THEOREM 2. *Suppose the IRSS as defined in Definition 1 is applied for k populations which yield the observations* $Y_{[1:n]1}, Y_{[2:n]2}, \ldots, Y_{[n:n]n}$*, where the joint cdf of the bivariate distribution for the ith population is* $D_i(x,y) = F_i(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2})$ *with joint pdf* $d_i(x,y) = \frac{1}{\sigma_1\sigma_2} f_i(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2})$*, for i=1,2, …, k. Suppose for the ith population the marginal distribution of the component random variable X has cdf* $D_i^{(X)}(x) = G_i(\frac{x-\mu_1}{\sigma_1})$ *with pdf* $d_i^{(X)}(x) = \frac{1}{\sigma_1} g_i(\frac{x-\mu_1}{\sigma_1})$ *and the marginal distribution of the component random variable Y has cdf* $D_i^{(Y)}(y) = H_i(\frac{y-\mu_2}{\sigma_2})$ *with pdf* $d_i^{(Y)}(y) = \frac{1}{\sigma_2} h_i(\frac{y-\mu_2}{\sigma_2})$*, for i=1,2, …, k. Then the pdf of* $Y_{[r:n]r}$ *is free from the parameters* μ_1 *and* σ_1 *of the marginal distribution of the random variable X of the populations.*

PROOF. Clearly the pdf of $Y_{[r:n]r}$, the rth observation of $irss$ as given in the statement of this theorem is obtained by replacing **a**, **b**, **c** involved in (1) by
$\mathbf{a} = \left(G_1(\frac{x-\mu_1}{\sigma_1}), \ldots, G_1(\frac{x-\mu_1}{\sigma_1}), G_2(\frac{x-\mu_1}{\sigma_1}), \ldots, G_2(\frac{x-\mu_1}{\sigma_1}), \ldots, G_k(\frac{x-\mu_1}{\sigma_1}), \ldots, G_k(\frac{x-\mu_1}{\sigma_1})\right)'$,
$\mathbf{b} = \mathbf{1} - \mathbf{a}$, where $\mathbf{1}$ is a column vector of n ones and $\mathbf{c} = \frac{1}{\sigma_1\sigma_2}\left(f_1(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2}), \ldots,\right.$
$f_1(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2}), f_2(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2}), \ldots, f_2(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2}), \ldots, f_k(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2}), \ldots,$
$\left.f_k(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2})\right)'$ respectively. If we put $u = \frac{x-\mu_1}{\sigma_1}$ in the right side of the obtained expression for $f_{Y_{[r:n]r}}(y)$, then it reduces to

$$f_{Y_{[r:n]r}}(y) = \frac{\sigma_2^{-1}}{(r-1)!(n-r)!} \int_{-\infty}^{\infty} per\big[\underset{r-1}{\mathbf{a}_1} \quad \underset{n-r}{\mathbf{b}_1} \quad \underset{1}{\mathbf{c}_1}\big] du, \tag{3}$$

where $\mathbf{a}_1 = (G_1(u), \ldots, G_1(u), G_2(u), \ldots, G_2(u), \ldots, G_k(u), \ldots, G_k(u))'$, $\mathbf{b}_1 = \mathbf{1} - \mathbf{a}_1$ and

$\mathbf{c}_1 = \Big(f_1(u, \frac{y-\mu_2}{\sigma_2}), \ldots, f_1(u, \frac{y-\mu_2}{\sigma_2}), f_2(u, \frac{y-\mu_2}{\sigma_2}), \ldots, f_2(u, \frac{y-\mu_2}{\sigma_2}), \ldots,$
$f_k(u, \frac{y-\mu_2}{\sigma_2}), \ldots, f_k(u, \frac{y-\mu_2}{\sigma_2})\Big)'.$

It is then clear to note that $\mathbf{a}_1$, $\mathbf{b}_1$ and $\mathbf{c}_1$ are all free of μ_1 and σ_1 and hence this proves the theorem. □

REMARK 3. *An immediate consequence of the above theorem is that if the other parameters involved in the population bivariate distribution other than $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ are known, then both means and variances of the observations of the irss depend only on μ_2 and σ_2. Further, those values can be obtained using the moment expression*

$$E\left(Y_{[r:n]r}^{l}\right) = \frac{\sigma_2^{-1}}{(r-1)!(n-r)!} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^l \, per\big[\underset{r-1}{\mathbf{a}_1} \quad \underset{n-r}{\mathbf{b}_1} \quad \underset{1}{\mathbf{c}_1}\big] du\, dy. \tag{4}$$

In the next section we illustrate an application of IRSS to estimate the common parameter of two bivariate Pareto distributions.

3. BEST LINEAR UNBIASED ESTIMATION OF A COMMON PARAMETER OF TWO PARETO DISTRIBUTIONS

Bivariate income data such as income of related individuals or income from two different sources (like salary income and income from other sources), etc. are modelled by bivariate Pareto distributions, provided the component random variables involved are positively correlated. For some applications of bivariate Pareto distribution, see Hutchinson (1979). It is also clear to note that marginal random variable on salary income of individuals can be realized exactly whereas income from other sources is complex in nature and hence cannot be measured with high precision and exactness by an assessing agency. So, on income studies to get efficient income estimates in the presence of measurement difficulty on one variable, we propose RSS technique by applying ranking of the units based on the salary income (easily measurable component) observed on the units. For similar applications of RSS one may refer Chacko and Thomas (2007). If we have a population involving two sections of people whose bivariate incomes are described as two bivariate Pareto distributions, then we require a RSS technique as described in Definition 1 to deal with income studies relating to such a population. Thus in this section, we consider two independent populations of units and the distribution of measurement (X_i, Y_i) made on the units of the ith population follows a bivariate Pareto distribution with joint pdf $h_i(x, y)$ given by

$$h_i(x,y) = a_i(a_i+1)(\theta_{1i}\theta_2)^{a_i+1}(\theta_2 x + \theta_{1i} y - \theta_{1i}\theta_2)^{-(a_i+2)}, a_i > 0, x > \theta_{1i} > 0, \quad y > \theta_2 > 0, \tag{5}$$

for $i = 1, 2$, where we assume that the threshold parameter θ_2 associated with the study variate Y_i, is the same for both populations. Let (U_i, V_i), $i = 1, 2$ be two independent non-identically distributed random vectors with (U_i, V_i) having joint pdf $g_i(u, v)$ defined by

$$g_i(u,v) = a_i(a_i+1)(u+v-1)^{-(a_i+2)}, a_i > 0, u > 1, v > 1. \tag{6}$$

Clearly, if (X_i, Y_i) is distributed with joint pdf (5), then $(\frac{X_i}{\theta_{1i}}, \frac{Y_i}{\theta_2})$ is distributed identically as (U_i, V_i) for $i = 1,2$. By Theorem 2, if $Y^{(n_1,n_2)}_{[r:n]r}$, $r = 1,2,\ldots,n$ are the $irss$ observations obtained as per Definition 1 with $k = 2$ described above, then the pdf of $Y^{(n_1,n_2)}_{[r:n]r}$ is free of the threshold parameters θ_{11} and θ_{12}. So without any loss of generality, we take $\theta_{1i} = 1$ in order to work out the means and variances of $Y^{(n_1,n_2)}_{[r:n]r}$. In particular, in order to obtain the means and variances of the observations $Y^{(n_1,n_2)}_{[r:n]r}$, $r = 1,2,\ldots,n$ obtained by IRSS from two Pareto distributions with joint pdfs $h_i(x, y)$, $i = 1,2$ given by (5), it is enough to obtain the means and variances of the observations, $V^{(n_1,n_2)}_{[r:n]r}$; $r = 1,2,\ldots,n$ obtained by IRSS from two Pareto distributions with joint pdfs $g_i(u, v)$, $i = 1,2$ given by (6).

Clearly, $V^{(n_1,n_2)}_{[r:n]r}$ is identically distributed as $V^{(n_1,n_2)}_{[r:n]}$, the rth concomitant in a pooled sample of size n consisting of n_1 observations drawn from $g_1(u, v)$ and n_2 observations drawn from $g_2(u, v)$ such that $n = n_1 + n_2$. Then we have

$$\xi^{(n_1,n_2)}_{[r:n]} = E(V^{(n_1,n_2)}_{[r:n]r}) = E(V^{(n_1,n_2)}_{[r:n]}) \tag{7}$$

and

$$\eta^{(n_1,n_2)}_{[r:n]} = \mathrm{Var}(V^{(n_1,n_2)}_{[r:n]r}) = \mathrm{Var}(V^{(n_1,n_2)}_{[r:n]}). \tag{8}$$

From Veena and Thomas (2015), we obtain the first and second raw moments of the random variable $V^{(n_1,n_2)}_{[r:n]}$, the rth concomitant obtained in a pooled sample of size n consisting of n_1 observations drawn from the joint pdf $g_1(u, v)$ and n_2 observations drawn from the joint pdf $g_2(u, v)$, where $n = n_1 + n_2$ and are given below

$$\begin{aligned} E[V^{(n_1,n_2)}_{[r:n]}] =& 1 + \frac{1}{a_1}\frac{(n_1)!\Gamma(n-r+1-\frac{1}{a_1})}{(n-r)!\Gamma(n_1+1-\frac{1}{a_1})} \\ &+ \sum_{i=1}^{n_2}\binom{n_2}{i}\frac{\prod_{j=1}^{i}(ja_1 - ia_2 + 1)}{{a_1}^{i+1}}\frac{(n_1)!\Gamma(n-r-i+1+\frac{ia_2}{a_1}-\frac{1}{a_1})}{(n-r)!\Gamma(n_1+1+\frac{ia_2}{a_1}-\frac{1}{a_1})} \end{aligned} \tag{9}$$

$$E[V_{[r:n]}^{(n_1,n_2)}]^2 = 1 + \frac{2}{a_1}\frac{(n_1)!\Gamma(n-r+1-\frac{1}{a_1})}{(n-r)!\Gamma(n_1+1-\frac{1}{a_1})} + \frac{2}{a_1(a_1-1)}\frac{(n_1)!\Gamma(n-r+1-\frac{2}{a_1})}{(n-r)!\Gamma(n_1+1-\frac{2}{a_1})}$$
$$+2\sum_{i=1}^{n_2}\binom{n_2}{i}\frac{\prod_{j=1}^{i}(ja_1-ia_2+1)}{{a_1}^{i+1}}\frac{(n_1)!\Gamma(n-r-i+1+\frac{ia_2}{a_1}-\frac{1}{a_1})}{(n-r)!\Gamma(n_1+1+\frac{ia_2}{a_1}-\frac{1}{a_1})}$$
$$+2\sum_{i=1}^{n_2}\binom{n_2}{i}\frac{\prod_{j=0}^{i-1}(ja_1-ia_2+2)}{{a_1}^{i+1}(a_1-1)}\frac{(n_1)!\Gamma(n-r-i+1+\frac{ia_2}{a_1}-\frac{2}{a_1})}{(n-r)!\Gamma(n_1+1+\frac{ia_2}{a_1}-\frac{2}{a_1})}$$
$$+\frac{2n_2}{a_1(a_2-1)}\frac{(n_1)!\Gamma(n-r+\frac{a_2}{a_1}-\frac{2}{a_1})}{(n-r)!\Gamma(n_1+1+\frac{a_2}{a_1}-\frac{2}{a_1})}$$
$$+2n_2\sum_{i=2}^{n_2}\binom{n_2-1}{i-1}\frac{\prod_{j=1}^{i-1}(ja_1-ia_2+2)}{{a_1}^{i}(a_2-1)}\frac{(n_1)!\Gamma(n-r-i+1+\frac{ia_2}{a_1}-\frac{2}{a_1})}{(n-r)!\Gamma(n_1+1+\frac{ia_2}{a_1}-\frac{2}{a_1})}. \tag{10}$$

Then using (9) and (10), we may obtain the variance of $V_{[r:n]}^{(n_1,n_2)}$ from

$$\mathrm{Var}(V_{[r:n]}^{(n_1,n_2)}) = E\left(V_{[r:n]}^{(n_1,n_2)}\right)^2 - \left[E(V_{[r:n]}^{(n_1,n_2)})\right]^2. \tag{11}$$

We make use of the expressions given in (9) to (11) to estimate θ_2 using the $irss$ observations $Y_{[r:n]r}^{(n_1,n_2)}$, $r = 1,2,\dots,n$ and the results are given in the following subsections.

3.1. Best linear unbiased estimator of θ_2 based on an induced ranked set sample

In order to obtain an estimator of θ_2 based on the observations of $irss$, we prove the following theorem.

THEOREM 4. *Suppose that $Y_{[1:n]1}^{(n_1,n_2)}, Y_{[2:n]2}^{(n_1,n_2)}, \dots, Y_{[n:n]n}^{(n_1,n_2)}$ are the n observations in the irss where each ranked set consists of n_1 units from a population with a bivariate Pareto distribution defined by the joint pdf $h_1(x,y)$ and n_2 units from a population with bivariate Pareto distribution defined by the joint pdf $h_2(x,y)$ such that $n = n_1 + n_2$. Let us write $V_{[r:n]r}^{(n_1,n_2)}$, $r = 1,2,\dots,n$ to denote the corresponding $irss$ observations obtained from a very similar set up with $h_i(x,y)$ replaced by $g_i(u,v)$ for i=1,2. Let $\xi_{[r:n]}^{(n_1,n_2)} = E\left(V_{[r:n]}^{(n_1,n_2)}\right)$ and $\eta_{[r:n]}^{(n_1,n_2)} = Var(V_{[r:n]}^{(n_1,n_2)})$ be as defined in (9) and (11). As $V_{[r:n]r}^{(n_1,n_2)}$ and $V_{[s:n]s}^{(n_1,n_2)}$ arise from two independent samples we have $Cov(V_{[r:n]r}^{(n_1,n_2)}, V_{[s:n]s}^{(n_1,n_2)}) = 0$, for $r \neq s$. Let us denote*

$$Y_{[n]}^{(n_1,n_2)} = (Y_{[1:n]1}^{(n_1,n_2)}, Y_{[2:n]2}^{(n_1,n_2)}, \dots, Y_{[n:n]n}^{(n_1,n_2)})' \tag{12}$$

$$\alpha = (\xi_{[1:n]}^{(n_1,n_2)}, \xi_{[2:n]}^{(n_1,n_2)}, \ldots, \xi_{[n:n]}^{(n_1,n_2)})' \tag{13}$$

and

$$B = diag(\eta_{[1:n]}^{(n_1,n_2)}, \eta_{[2:n]}^{(n_1,n_2)}, \ldots, \eta_{[n:n]}^{(n_1,n_2)}). \tag{14}$$

Then the BLUE θ_2^* of θ_2 is given by

$$\theta_2^* = (\alpha' B^{-1} \alpha)^{-1} \alpha' B^{-1} Y_{[n]}^{(n_1,n_2)} \tag{15}$$

and the variance of θ_2^* is given by

$$Var(\theta_2^*) = (\alpha' B^{-1} \alpha)^{-1} {\theta_2}^2. \tag{16}$$

PROOF. Using the notations given in the statement of the theorem, the mean and variance-covariance matrix of the column vector $Y_{[n]}^{(n_1,n_2)}$ of observations in the *irss* may be written as

$$E\left(Y_{[n]}^{(n_1,n_2)}\right) = \alpha\,\theta_2 \quad \text{and} \quad D\left(Y_{[n]}^{(n_1,n_2)}\right) = B\,\theta_2^2, \tag{17}$$

where for $r = 1, 2, \ldots, n$, $\xi_{[r:n]}^{(n_1,n_2)}$ involved in α and $\eta_{[r:n]}^{(n_1,n_2)}$ involved in the diagonal matrix B are given by equations (13) and (14). If a_1 and a_2 involved in $\xi_{[r:n]}^{(n_1,n_2)}$ and $\eta_{[r:n]}^{(n_1,n_2)}$ are known, then by Gauss-Markov theorem, the BLUE θ_2^* of θ_2 is as given in (15) and its variance is as given in (16). Hence the theorem is proved. □

Note that θ_2^* is a linear function of the observations $Y_{[r:n]r}^{(n_1,n_2)}$, $r = 1, 2, \ldots, n$ in the *irss* and hence one can write $\theta_2^* = \sum_{r=1}^{n} b_{r:n}^{(n_1,n_2)} Y_{[r:n]r}^{(n_1,n_2)}$, where $b_{r:n}^{(n_1,n_2)}$, $r = 1, 2, \ldots, n$ are appropriate constants. For convenience, we may write $b_{r:n}$ for $b_{r:n}^{(n_1,n_2)}$ and $Y_{[r:n]r}$ for $Y_{[r:n]r}^{(n_1,n_2)}$. Thus, we have $\theta_2^* = \sum_{r=1}^{n} b_{r:n} Y_{[r:n]r}$.

By making use of the expressions (9) and (10) for $E(V_{[r:n]}^{(n_1,n_2)})$ and $E(V_{[r:n]}^{(n_1,n_2)})^2$, we have the computed means, $\xi_{[r:n]}^{(n_1,n_2)}$ and variances, $\eta_{[r:n]}^{(n_1,n_2)}$ for $r = 1, 2, \ldots, n$; $a_1 = 2.5(0.5)3.5$; $a_2 = 3(0.5)4$ and for n_1 and n_2 such that $n = n_1 + n_2$ and $2 \leq n \leq 5$ using R programming and are given in Table 1. Using these computed values of the means and variances, we have further obtained the coefficients $b_{i:n}$ in the BLUE $\theta_2^* = \sum_{i=1}^{n} b_{i:n} Y_{[i:n]i}$ and $\theta_2^{-2}\text{Var}(\theta_2^*)$ for $a_1 = 2.5(0.5)3.5$, $a_2 = 3(0.5)4$ and for n_1 and n_2 such that $n = n_1 + n_2$ and $2 \leq n \leq 5$ and these values are presented in Table 2.

3.2. *Best linear unbiased estimator of θ_2 based on a lower extreme induced ranked set sample*

Since the lower terminal of the study variate Y associated with the bivariate Pareto distribution depends on θ_2, it is intuitive to conclude that among the possible Y observations that we may obtain due to ranking made on X, the measurement of Y on the unit with least X- value contains more information on θ_2. Thus on instances such as dealing with the estimation of threshold parameter of the variable of primary interest of bivariate Pareto distributions, instead of the IRSS as described in Definition 1, we define another $irss$ as given below.

DEFINITION 5. *Draw randomly n sets of n units each from the given k populations ($k \geq 2$) such that each set consists of n_i units drawn randomly from the ith population for $i = 1,2,\ldots,k$ with $n = \sum_{i=1}^{k} n_i$. Measure the characteristic X (an auxiliary variable) on the units which can be measured easily and rank the units within each set based on the measured values of X on the units. Now from the jth set choose the unit which is ranked 1 among its units for the measurement of the characteristic Y of primary interest and let it be denoted by $Y_{[1:n]j}$ for $j = 1,2,\ldots,n$. Then $Y_{[1:n]j}$, $j = 1,2,\ldots,n$ are said to constitute the lower extreme induced ranked set sample ($leirss$) and the sampling scheme which generates the above sample is called lower extreme induced ranked set sampling (LEIRSS) from the populations.*

Now if the problem is to estimate the common threshold parameter θ_2 of the two bivariate distributions with joint pdf $h_i(x,y)$, $i = 1,2$ defined in (5) using LEIRSS, then the estimation procedure is described in the following theorem.

THEOREM 6. *Suppose that $Y_{[1:n]1}^{(n_1,n_2)}, Y_{[1:n]2}^{(n_1,n_2)}, \ldots, Y_{[1:n]n}^{(n_1,n_2)}$ are the n observations in the $leirss$ where each ranked set consists of n_1 units from a population with a bivariate Pareto distribution defined by the joint pdf $h_1(x,y)$ and n_2 units from a population with bivariate Pareto distribution defined by the joint pdf $h_2(x,y)$ such that $n = n_1 + n_2$. Let us write $V_{[1:n]r}^{(n_1,n_2)}$, $r = 1,2,\ldots,n$ to denote the corresponding $leirss$ observations obtained from a very similar set up with $h_i(x,y)$ replaced by $g_i(u,v)$ for $i = 1,2$. Let $\xi_{[r:n]}^{(n_1,n_2)} = E\left(V_{[r:n]}^{(n_1,n_2)}\right)$, $\eta_{[r:n]}^{(n_1,n_2)} = Var(V_{[r:n]}^{(n_1,n_2)})$ be as defined in (7) to (11). As $V_{[1:n]r}^{(n_1,n_2)}$ and $V_{[1:n]s}^{(n_1,n_2)}$ arise from two independent samples we have $Cov(V_{[1:n]r}^{(n_1,n_2)}, V_{[1:n]s}^{(n_1,n_2)}) = 0$, for $r \neq s$. Let us denote*

$$Y_{[1]}^{(n_1,n_2)} = (Y_{[1:n]1}^{(n_1,n_2)}, Y_{[1:n]2}^{(n_1,n_2)}, \ldots, Y_{[1:n]n}^{(n_1,n_2)})' \tag{18}$$

$$E(Y_{[1]}^{(n_1,n_2)}) = \xi_{[1:n]}^{(n_1,n_2)} \mathbf{1}' \theta_2 \tag{19}$$

and

$$D(Y_{[1]}^{(n_1,n_2)}) = \eta_{[1:n]}^{(n_1,n_2)} \mathbf{I}_\mathbf{n}' \theta_2^2, \tag{20}$$

where $\mathbf{1}$ is a column vector of n ones and $\mathbf{I_n}$ is the unit matrix of order n. Then the BLUE θ_2^{**} of θ_2 based on the $leirss$ is given by

$$\theta_2^{**} = \frac{1}{n\xi_{[1:n]}^{(n_1,n_2)}} \sum_{r=1}^{n} Y_{[1:n]r}^{(n_1,n_2)} \tag{21}$$

with

$$Var(\theta_2^{**}) = \frac{\eta_{[1:n]}^{(n_1,n_2)}}{n(\xi_{[1:n]}^{(n_1,n_2)})^2}\theta_2^2. \tag{22}$$

PROOF. It is clear from (19) that for each r, $\frac{1}{\xi_{[1:n]r}} Y_{[1:n]r}^{(n_1,n_2)}$ is unbiased for θ_2. Thus the BLUE θ_2^{**} of θ_2 based on $leirss$ is the one as given in (21). Then from (20) and (21), the variance of θ_2^{**} is as given in (22). Hence the theorem is proved. □

Veena and Thomas (2015) have derived the distribution theory of concomitants of order statistics of INID random variables and, as an application, illustrated the method of estimating the common threshold parameter θ_2 on the variable of primary interest involved in two bivariate Pareto distributions with joint pdfs $h_1(x,y)$ and $h_2(x,y)$ by the BLUE of θ_2 based on concomitants of order statistics of the pooled sample of a first sample of size n_1 drawn from $h_1(x,y)$ and a second sample of size n_2 drawn from $h_2(x,y)$ such that $n = n_1 + n_2$. They have obtained the variance of their estimate $\widehat{\theta_2}$ of θ_2 and presented in Table 1 for $a_1 = 2.5(0.5)3.5$, $a_2 = 3(0.5)4$ and for n_1 and n_2 such that $n = n_1 + n_2$ and $2 \leq n \leq 5$. We have reproduced those values of $\text{Var}(\widehat{\theta_2})$ in Table 3 so as to compare our estimate θ_2^* based on $irss$ and our estimate θ_2^{**} based on $leirss$. We have further computed $e_1 = \frac{\text{Var}(\widehat{\theta_2})}{\text{Var}(\theta_2^*)}$ and $e_2 = \frac{\text{Var}(\widehat{\theta_2})}{\text{Var}(\theta_2^{**})}$, relative efficiencies of θ_2^* and θ_2^{**} relative to the estimator $\widehat{\theta_2}$ of Veena and Thomas (2015) for $a_1 = 2.5(0.5)3.5$, $a_2 = 3(0.5)4$ and for n_1 and n_2 such that $n = n_1 + n_2$ and $2 \leq n \leq 5$ and are also presented in Table 3.

REMARK 7. *For obtaining the BLUE of θ_2, we have assumed that both a_1 and a_2 are known in all cases. But in real life situations, these parameters may not be known. The moment type estimators of a_1 and a_2 can be used in such situations as known values of those parameters so that our estimators developed in this paper can be made useful to deal with the estimation process of θ_2. For the Pareto distributions with joint pdf's $h_j(x,y)$ given by* (5), *the correlation between X_j and Y_j is $\rho_j = \frac{1}{a_j}$, provided $a_j > 2$, $j = 1,2$. If r_j is the sample correlation coefficient between X and Y observation pairs available in the $irss$ and $leirss$ which arise from jth population, then a moment type estimator for a_j is given by $\tilde{a}_j = \frac{1}{r_j}$. Note that for negative values of r_1 and r_2, $\tilde{a}_1$ and $\tilde{a}_2$ are also negative, in which case the assumption of the model* (5) *that both $a_1, a_2 > 0$ is violated. Moreover, if $r_j \geq \frac{1}{2}$, $j = 1,2$, then $\tilde{a}_j \leq 2$, in which case variances of the Pareto distributions do not exist. Hence, when a_1 and a_2 are not known, we can make use of the estimators $\tilde{a}_j = \frac{1}{r_j}$, provided $0 < r_j < \frac{1}{2}$, for $j = 1,2$.*

REMARK 8. *It is well known that when n is not small, obtaining the means and variances of* $V_{[r:n]r}^{(n_1,n_2)}$ *for* $r = 1,2,\dots,n$ *involves a large number of terms in the expansion of permanent expression of the pdf of* $V_{[r:n]r}^{(n_1,n_2)}$ *and requires evaluation of integrals of several terms in the expansion which is somewhat cumbersome and hence usually* θ_2^* *and* $Var(\theta_2^*)$ *are obtained for those* n_1 *and* n_2 *such that* $n = n_1 + n_2$ *is small. In spite of this limitation if one requires more precision on the estimate* θ_2^* *one may carry out IRSS with k cycles so that if* $\theta_2^*(m)$ *is the BLUE* θ_2^* *evaluated in the* m*th cycle for* $m = 1,2,\dots,k$ *then the required estimate is* $\tilde{\theta}_2(m) = \frac{1}{k}\sum_{m=1}^{k}\theta_2^*(m)$ *with* $Var(\tilde{\theta}_2) = \frac{Var(\theta_2^*)}{k}$. *Clearly one can choose k in such a manner that the required precision on the estimate is attained.*

REMARK 9. *Generally when n is not small, obtaining the means and variances of* $V_{[1:n]r}^{(n_1,n_2)}$ *involves a large number of terms in the expansion of permanent expression of the pdf of* $V_{[1:n]r}^{(n_1,n_2)}$ *and requires evaluation of integrals of each term which is somewhat cumbersome and hence usually the estimate* θ_2^{**} *and* $Var(\theta_2^{**})$ *are obtained for those* n_1 *and* n_2 *such that* $n = n_1 + n_2$ *is small. However, if one requires more precision on the estimate* θ_2^{**} *one may carry out LEIRSS with k cycles so that if* $\theta_2^{**}(m)$ *is the BLUE* θ_2^{**} *evaluated in the* m*th cycle for* $m = 1,2,\dots,k$ *then the required estimate is* $\tilde{\tilde{\theta}}_2(m) = \frac{1}{k}\sum_{m=1}^{k}\theta_2^{**}(m)$ *with* $Var(\tilde{\tilde{\theta}}_2) = \frac{Var(\theta_2^{**})}{k}$. *Clearly one can choose k in such a manner that the required precision on the estimate is attained.*

TABLE 1

Means and variances of the observations in the irss arising from two bivariate Pareto distributions with pdfs $g_j(u,v)$ for $j=1,2$; $a_1=2.5(0.5)3.5$, $a_2=3(0.5)4$ and n_1, n_2 such that $n=n_1+n_2$ and $2\le n\le 5$.

n	a_1	a_2	n_1	n_2	r	Mean	Variance	n	a_1	a_2	n_1	n_2	r	Mean	Variance
2	2.5	3.0	1	1	1	1.444	0.469	2	2.5	3.5	1	1	1	1.400	0.373
					2	1.722	2.478						2	1.667	2.222
3	2.5	3.0	2	1	1	1.429	0.427	3	2.5	3.5	2	1	1	1.400	0.373
					2	1.532	0.717						2	1.500	0.639
					3	1.873	3.960						3	1.833	3.750
			1	2	1	1.400	0.353				1	2	1	1.353	0.267
					2	1.489	0.569						2	1.427	0.422
					3	1.778	2.741						3	1.686	2.267
4	2.5	3.0	3	1	1	1.421	0.411	4	2.5	3.5	3	1	1	1.400	0.373
					2	1.484	0.562						2	1.462	0.514
					3	1.602	0.948						3	1.577	0.879
					4	1.993	5.320						4	1.962	5.136
			2	2	1	1.364	0.294				2	2	1	1.364	0.294
					2	1.415	0.397						2	1.415	0.397
					3	1.512	0.662						3	1.512	0.662
					4	1.842	3.771						4	1.842	3.771
			1	3	1	1.381	0.311				1	3	1	1.333	0.228
					2	1.432	0.412						2	1.375	0.297
					3	1.526	0.657						3	1.451	0.464
					4	1.827	2.993						4	1.708	2.322
5	2.5	3.0	4	1	1	1.417	0.402	5	2.5	3.5	4	1	1	1.400	0.373
					2	1.462	0.503						2	1.444	0.469
					3	1.532	0.689						3	1.513	0.646
					4	1.662	1.166						4	1.641	1.104
					5	2.094	6.601						5	2.068	6.434
			3	2	1	1.370	0.311				3	2	1	1.370	0.311
					2	1.409	0.387						2	1.409	0.387
					3	1.470	0.528						3	1.470	0.528
					4	1.584	0.893						4	1.584	0.893
					5	1.966	5.148						5	1.966	5.148
			2	3	1	1.385	0.324				2	3	1	1.345	0.256
					2	1.423	0.400						2	1.378	0.314
					3	1.483	0.536						3	1.430	0.420
					4	1.592	0.875						4	1.527	0.689
					5	1.950	4.385						5	1.854	3.801
			4	1	1	1.370	0.290				1	4	1	1.323	0.209
					2	1.406	0.353						2	1.351	0.251
					3	1.460	0.466						3	1.393	0.325
					4	1.559	0.739						4	1.471	0.501
					5	1.872	3.233						5	1.729	2.380
2	2.5	4	1	1	1	1.364	0.312	2	3	3.5	1	1	1	1.364	0.268
					2	1.636	2.151						2	1.546	0.846
3	2.5	4	2	1	1	1.375	0.336	3	3	3.5	2	1	1	1.353	0.249
					2	1.477	0.598						2	1.421	0.376
					3	1.814	3.702						3	1.626	1.215
			1	2	1	1.316	0.214				1	2	1	1.333	0.214
					2	1.381	0.338						2	1.394	0.315
					3	1.636	2.131						3	1.573	0.944
4	2.5	4	3	1	1	1.381	0.346	4	3	3.5	3	1	1	1.348	0.241
					2	1.444	0.485						2	1.390	0.310
					3	1.561	0.848						3	1.467	0.471
					4	1.948	5.099						4	1.695	1.537
			2	2	1	1.333	0.253				2	2	1	1.333	0.216
					2	1.382	0.343						2	1.373	0.276
					3	1.477	0.593						3	1.443	0.412
					4	1.808	3.674						4	1.651	1.292
			1	3	1	1.296	0.179				1	3	1	1.320	0.193
					2	1.331	0.231						2	1.356	0.244
					3	1.397	0.360						3	1.419	0.356
					4	1.642	2.128						4	1.605	1.036

Continued on next page

TABLE 1 - continued from previous page

n	a_1	a_2	n_1	n_2	r	Mean	Variance	n	a_1	a_2	n_1	n_2	r	Mean	Variance
5	2.5	4	4	1	1	1.385	0.352	5	3	3.5	4	1	1	1.345	0.237
					2	1.430	0.447						2	1.376	0.284
					3	1.499	0.622						3	1.422	0.367
					4	1.629	1.078						4	1.505	0.557
					5	2.058	6.404						5	1.753	1.829
			3	2	1	1.345	0.276				3	2	1	1.333	0.217
					2	1.383	0.347						2	1.362	0.260
					3	1.442	0.480						3	1.406	0.334
					4	1.556	0.835						4	1.483	0.502
					5	1.941	5.073						5	1.715	1.602
			2	3	1	1.313	0.213				2	3	1	1.323	0.199
					2	1.343	0.262						2	1.350	0.237
					3	1.390	0.353						3	1.390	0.301
					4	1.482	0.597						4	1.462	0.447
					5	1.807	3.658						5	1.675	1.367
			1	4	1	1.286	0.161				1	4	1	1.313	0.182
					2	1.309	0.192						2	1.338	0.215
					3	1.344	0.247						3	1.375	0.271
					4	1.410	0.380						4	1.441	0.394
					5	1.651	2.133						5	1.633	1.122
2	3	4	1	1	1	1.333	0.222	2	3.5	4	1	1	1	1.308	0.172
					2	1.500	0.750						2	1.426	0.419
3	3	4	2	1	1	1.333	0.222	3	3.5	4	2	1	1	1.300	0.162
					2	1.400	0.340						2	1.349	0.228
					3	1.600	1.140						3	1.485	0.564
			1	2	1	1.300	0.169				1	2	1	1.286	0.143
					2	1.352	0.246						2	1.330	0.198
					3	1.514	0.773						3	1.451	0.465
4	3	4	3	1	1	1.333	0.222	4	3.5	4	3	1	1	1.296	0.158
					2	1.375	0.288						2	1.327	0.195
					3	1.450	0.440						3	1.380	0.275
					4	1.675	1.473						4	1.530	0.685
			2	2	1	1.308	0.183				2	2	1	1.286	0.144
					2	1.344	0.234						2	1.314	0.177
					3	1.409	0.351						3	1.364	0.248
					4	1.607	1.151						4	1.503	0.60
			1	3	1	1.286	0.149				1	3	1	1.276	0.131
					2	1.316	0.186						2	1.302	0.160
					3	1.369	0.267						3	1.348	0.220
					4	1.530	0.800						4	1.474	0.508
5	3	4	4	1	1	1.333	0.222	5	3.5	4	4	1	1	1.294	0.155
					2	1.364	0.268						2	1.316	0.181
					3	1.409	0.347						3	1.349	0.225
					4	1.491	0.530						4	1.407	0.318
					5	1.736	1.772						5	1.567	0.793
			3	2	1	1.313	0.191				3	2	1	1.286	0.145
					2	1.340	0.229						2	1.307	0.168
					3	1.381	0.295						3	1.338	0.208
					4	1.455	0.447						4	1.392	0.292
					5	1.679	1.479						5	1.544	0.715
			2	3	1	1.294	0.163				2	3	1	1.278	0.135
					2	1.318	0.194						2	1.297	0.156
					3	1.354	0.246						3	1.327	0.191
					4	1.419	0.365						4	1.378	0.266
					5	1.615	1.166						5	1.519	0.633
			1	4	1	1.278	0.139				1	4	1	1.270	0.125
					2	1.298	0.162						2	1.289	0.144
					3	1.329	0.201						3	1.316	0.175
					4	1.383	0.287						4	1.364	0.241
					5	1.545	0.827						5	1.494	0.548

TABLE 2

Coefficients $b_{i:n}$ in the BLUE $\theta_2^* = \sum_{i=1}^{n} b_{i:n} Y_{[i:n]i}$ and $\theta_2^{-2} Var(\theta_2^*)$ based on a the irss for some values of a_1, a_2 and for n_1 and n_2 such that $n = n_1 + n_2$ and $2 \le n \le 5$.

n	a_1	a_2	n_1	n_2	b_1	b_2	b_3	b_4	b_5	$\frac{Var(\theta_2^*)}{\theta_2^2}$
2	2.5	3.0	1	1	0.546	0.123				0.177
3	2.5	3.0	2	1	0.374	0.239	0.053			0.112
			1	2	0.374	0.247	0.061			0.094
4	2.5	3.0	3	1	0.281	0.215	0.138	0.030		0.081
			2	2	0.295	0.227	0.145	0.031		0.064
			1	3	0.282	0.221	0.147	0.039		0.063
5	2.5	3.0	4	1	0.225	0.185	0.142	0.091	0.020	0.064
			3	2	0.234	0.193	0.148	0.094	0.020	0.053
			2	3	0.226	0.189	0.147	0.097	0.024	0.053
			1	4	0.225	0.189	0.149	0.100	0.028	0.048
2	2.5	3.5	1	1	0.577	0.115				0.154
3	2.5	3.5	2	1	0.388	0.243	0.051			0.103
			1	2	0.392	0.261	0.057			0.077
4	2.5	3.5	3	1	0.289	0.219	0.138	0.029		0.077
			2	2	0.295	0.227	0.145	0.031		0.064
			1	3	0.293	0.232	0.157	0.037		0.050
5	2.5	3.5	4	1	0.229	0.188	0.143	0.091	0.020	0.061
			3	2	0.234	0.193	0.148	0.094	0.020	0.053
			2	3	0.236	0.197	0.153	0.100	0.022	0.045
			1	4	0.233	0.198	0.158	0.108	0.027	0.037
2	2.5	4.0	1	1	0.607	0.106				0.139
3	2.5	4.0	2	1	0.403	0.243	0.048			0.098
			1	2	0.410	0.273	0.051			0.067
4	2.5	4.0	3	1	0.297	0.222	0.137	0.028		0.075
			2	2	0.307	0.234	0.145	0.029		0.058
			1	3	0.305	0.242	0.163	0.033		0.042
5	2.5	4.0	4	1	0.235	0.191	0.144	0.090	0.019	0.060
			3	2	0.243	0.199	0.150	0.093	0.019	0.050
			2	3	0.246	0.205	0.158	0.099	0.020	0.040
			1	4	0.242	0.206	0.165	0.112	0.023	0.030
2	3.0	3.5	1	1	0.523	0.187				0.103
3	3.0	3.5	2	1	0.365	0.254	0.090			0.067
			1	2	0.364	0.259	0.097			0.058
4	3.0	3.5	3	1	0.277	0.222	0.154	0.055		0.049
			2	2	0.278	0.224	0.158	0.058		0.045
			1	3	0.277	0.225	0.161	0.063		0.041
5	3.0	3.5	4	1	0.222	0.189	0.152	0.106	0.038	0.039
			3	2	0.223	0.191	0.153	0.108	0.039	0.036
			2	3	0.223	0.192	0.155	0.110	0.041	0.034
			1	4	0.232	0.178	0.145	0.105	0.042	0.029
2	3.0	4.0	1	1	0.545	0.182				0.091
3	3.0	4.0	2	1	0.375	0.257	0.088			0.062
			1	2	0.377	0.270	0.096			0.049
4	3.0	4.0	3	1	0.282	0.225	0.155	0.054		0.047
			2	2	0.286	0.230	0.161	0.056		0.040
			1	3	0.284	0.233	0.169	0.063		0.033
5	3.0	4.0	4	1	0.226	0.192	0.153	0.106	0.037	0.038
			3	2	0.229	0.195	0.156	0.109	0.038	0.033
			2	3	0.230	0.198	0.160	0.113	0.040	0.029
			1	4	0.227	0.198	0.163	0.119	0.046	0.025
2	3.5	4.0	1	1	0.514	0.230				0.068
3	3.5	4.0	2	1	0.360	0.265	0.118			0.045
			1	2	0.359	0.269	0.125			0.040
4	3.5	4.0	3	1	0.274	0.227	0.167	0.074		0.033
			2	2	0.275	0.228	0.170	0.077		0.031
			1	3	0.274	0.230	0.172	0.082		0.028
5	3.5	4.0	4	1	0.221	0.192	0.159	0.117	0.052	0.026
			3	2	0.221	0.193	0.160	0.119	0.054	0.025
			2	3	0.221	0.194	0.162	0.121	0.056	0.023
			1	4	0.220	0.194	0.163	0.123	0.059	0.022

TABLE 3

Variances of θ_2^{**} and $\hat{\theta}_2$, efficiencies e_1 of θ_2^{*} relative to $\hat{\theta}_2$ and e_2 of θ_2^{**} relative to $\hat{\theta}_2$ for some values of a_1, a_2 and for n_1 and n_2 such that $n = n_1 + n_2$ and $2 \leq n \leq 5$.

n	a_1	a_2	n_1	n_2	$\frac{\text{Var}(\theta_2^{**})}{\theta_2^2}$	$\frac{\text{Var}(\hat{\theta}_2)}{\theta_2^2}$	e_1	e_2
2	2.5	3.0	1	1	0.112	0.179	1.009	1.591
3	2.5	3.0	2	1	0.070	0.114	1.016	1.629
			1	2	0.060	0.096	1.015	1.594
4	2.5	3.0	3	1	0.051	0.083	1.020	1.631
			2	2	0.040	0.074	1.159	1.863
			1	3	0.041	0.065	1.018	1.583
5	2.5	3.0	4	1	0.040	0.065	1.023	1.628
			3	2	0.033	0.060	1.123	1.801
			2	3	0.034	0.054	1.020	1.599
			1	4	0.031	0.049	1.020	1.574
2	2.5	3.5	1	1	0.095	0.154	1.000	1.615
3	2.5	3.5	2	1	0.063	0.104	1.006	1.640
			1	2	0.049	0.077	1.003	1.595
4	2.5	3.5	3	1	0.048	0.078	1.012	1.637
			2	2	0.040	0.064	1.006	1.616
			1	3	0.032	0.050	1.006	1.571
5	2.5	3.5	4	1	0.038	0.062	1.016	1.632
			3	2	0.033	0.054	1.009	1.619
			2	3	0.028	0.045	1.006	1.595
			1	4	0.024	0.037	1.008	1.553
2	2.5	4.0	1	1	0.084	0.138	0.991	1.640
3	2.5	4.0	2	1	0.059	0.098	0.997	1.656
			1	2	0.041	0.066	0.991	1.605
4	2.5	4.0	3	1	0.045	0.075	1.003	1.648
			2	2	0.036	0.058	0.993	1.628
			1	3	0.027	0.042	0.994	1.573
5	2.5	4.0	4	1	0.037	0.060	1.009	1.640
			3	2	0.031	0.050	0.997	1.630
			2	3	0.025	0.040	0.992	1.603
			1	4	0.019	0.030	0.996	1.550
2	3.0	3.5	1	1	0.072	0.104	1.009	1.440
3	3.0	3.5	2	1	0.045	0.068	1.015	1.503
			1	2	0.040	0.059	1.014	1.478
4	3.0	3.5	3	1	0.033	0.050	1.019	1.521
			2	2	0.030	0.046	1.017	1.506
			1	3	0.028	0.041	1.017	1.485
5	3.0	3.5	4	1	0.026	0.040	1.021	1.527
			3	2	0.024	0.037	1.019	1.517
			2	3	0.023	0.034	1.018	1.504
			1	4	0.017	0.031	1.096	1.877
2	3.0	4.0	1	1	0.063	0.091	1.000	1.455
3	3.0	4.0	2	1	0.042	0.063	1.007	1.510
			1	2	0.033	0.049	1.003	1.475
4	3.0	4.0	3	1	0.031	0.048	1.012	1.524
			2	2	0.027	0.040	1.006	1.507
			1	3	0.022	0.033	1.006	1.471
5	3.0	4.0	4	1	0.025	0.038	1.016	1.530
			3	2	0.022	0.034	1.010	1.518
			2	3	0.020	0.029	1.007	1.499
			1	4	0.017	0.025	1.009	1.466
2	3.5	4.0	1	1	0.050	0.068	1.008	1.355
3	3.5	4.0	2	1	0.032	0.045	1.014	1.422
			1	2	0.029	0.040	1.013	1.403
4	3.5	4.0	3	1	0.023	0.034	1.017	1.445
			2	2	0.022	0.031	1.015	1.433
			1	3	0.020	0.029	1.015	1.417
5	3.5	4.0	4	1	0.019	0.038	1.445	2.064
			3	2	0.018	0.025	1.017	1.446
			2	3	0.016	0.024	1.016	1.436
			1	4	0.016	0.022	1.017	1.423

From Table 2, it can be observed that for each fixed pair of (a_1, a_2) and for each value of n, $\theta_2^{-2}\text{Var}(\theta_2^*)$ decreases as n_2 increases. From Table 3, it is seen that the efficiency e_1 of θ_2^* relative to $\widehat{\theta_2}$ is more than 1 for all pairs (n_1, n_2) except for the case with $a_1 = 2.5, a_2 = 4$. However for this case also the relative efficiency e_1 is very close to 1. Thus the relative efficiency e_1 of the BLUE θ_2^* based on $irss$ is either very close to 1 or more than 1 for all values of $n = n_1 + n_2$ tried in this paper. However, it may be noted that, the efficiency e_2 of θ_2^{**} relative to $\widehat{\theta_2}$ is always greater than 1 and it is uniformly and significantly larger than e_1 for every pair of values of (a_1, a_2) and for all values of n. Hence, the the BLUE θ_2^{**} of θ_2 based on the $leirss$ is relatively more efficient than the estimator $\widehat{\theta_2}$ due to Veena and Thomas (2015) and the estimator θ_2^* based on the observations of $irss$.

4. Application of the results by a real life data

In Kerala, one of the southern states of India, students with low parental income are given some percentage of reservation for admission in various under-graduate and post-graduate programmes of the colleges. In addition to this, students admitted in colleges are eligible for some scholarships based on their low parental income. So in some cases, students have the tendency to under-report their parental income for getting admission in reservation quota or for getting scholarship. However, one may use the reported income as an auxiliary variable in ranking the units of any selected set of units so that units selected based on the assigned ranks can be subjected for further scrutiny in determining the correct income. In particular we devote this section to illustrate the results developed in Section 3 for using IRSS and LEIRSS and estimating the common parameter θ_2 associated with the variable Y of primary interest involved in two bivariate Pareto distributions.

In Kerala's social set-up, it is observed that mostly students from educated families choose science subjects for their higher study so that more accuracy is experienced on the reported annual parental income of those students than the reported annual parental income of students admitted in humanity subjects. Hence if we write (X, Y) to denote a bivariate random variable where X is the (auxiliary) variable representing reported annual parental income and Y is the variable representing actual annual parental income, then we may expect the population distribution of (X, Y) of the students of Science and humanities as two bivariate Pareto distributions $h_1(x, y)$ and $h_2(x, y)$ which are as defined in (5). Since the reporting accuracy of the science and humanity students vary significantly, θ_{11} and θ_{12} may be different. However, both set of students with respect to the marginal random variable Y on actual annual parental income may be identical as it remains untampered. Hence there is justification for retaining the same threshold parameter θ_2 for the distribution of the component random variable Y. Thus the theory developed in Section 3 can be used as such for estimating θ_2 efficiently.

Thus for illustration we consider the students admitted for post-graduate courses during 2015-'16 in the University College, Trivandrum, one of the premier higher education institutions in Kerala. The ratio of post-graduate courses in science and humanity streams in the college is observed to be 1:2. This makes us to propose IRSS and LEIRSS for $n_1 = 2$ and $n_2 = 4$ with $n = 6$ and by three cycles. As indicated in Definition 1 for each cycle we select 6 sets each with six units comprising of two science students and 4 humanity students, note down the reported annual parental income to order the units of each set in a cycle and then order the units based on the noted annual parental income in the $irss$. We then meet personally the selected students, quizzed them and thereby obtained a more closer value of their annual parental income. The reported annual parental income and a closer estimate of the actual annual parental income of the selected students under IRSS in the three cycles are presented in Table 4.

TABLE 4
Reported annual parental income (Rs. in hundreds) and the closer estimate of actual annual parental income (Rs. in hundreds) of students selected under IRSS.

			Cycle 1			
r	1	2	3	4	5	6
$X_{(r:6)r}$	36	180	60	180	840	240
$Y_{[r:6]r}$	96	960	120	600	1080	4200
			Cycle 2			
r	1	2	3	4	5	6
$X_{(r:6)r}$	36	36	72	85.8	360	2520
$Y_{[r:6]r}$	60	1560	672	216	2520	4320
			Cycle 3			
r	1	2	3	4	5	6
$X_{(r:6)r}$	36	72	60	180	480	324
$Y_{[r:6]r}$	240	720	240	780	757.8	6480

We have again identified students in each set with lowest reported annual parental income, made a personal interview and thereby obtained a closer estimate of the actual annual income of their parents. The reported and actual annual parental incomes of the selected students in the $leirss$ with $n_1 = 2$, $n_2 = 4$ and $n = 6$ for each of the $k = 3$ cycles are given in Table 5.

TABLE 5
Reported annual parental income and the closer estimate of actual annual parental income of students selected under LEIRSS.

Cycle 1						
r	1	2	3	4	5	6
$X_{(1:6)r}$	36	72	36	72	36	36
$Y_{[1:6]r}$	96	3600	480	96	180	240

Cycle 2						
r	1	2	3	4	5	6
$X_{(1:6)r}$	36	36	48	18	120	54
$Y_{[1:6]r}$	60	156	168	120	216	6000

Cycle 3						
r	1	2	3	4	5	6
$X_{(1:6)r}$	36	42	36	36	24	36
$Y_{[1:6]r}$	240	1680	96	240	360	174

In order to apply the results of Section 3, we need the values of parameters a_1 and a_2. The distinct pairs of the reported annual parental income X and closer estimate of the actual annual parental income Y of the science students in the $irss$ and $leirss$ are identified as $(180,600)$, $(240,4200)$, $(36,60)$, $(72,672)$, $(72,720)$, $(180,780)$, $(480,757.8)$, $(324,6480)$, $(36,96)$ and the Pearson correlation coefficient r_1 between X and Y based on these bivariate observations is obtained as $r_1 = 0.4676$. Then a moment type estimator of a_1 is given by $\hat{a_1} = \frac{1}{r_1} = 2.14$. The correlation coefficient (r_2) between X and Y based on the remaining 24 distinct pairs of the reported parental income and closer estimate of the actual parental income of the humanity students in the $irss$ and $leirss$ is obtained as $r_2 = 0.4254$ and hence the moment type estimator of a_2 is given by $\hat{a_2} = \frac{1}{r_2} = 2.35$.

If the parent bivariate distribution is $h_1(x,y)$ then from (5) we have

$$\mu_1 = E(Y_1) = \frac{a_1}{a_1 - 1}\theta_2. \tag{23}$$

Similarly, from $h_2(x,y)$ as defined in (5) we have

$$\mu_2 = E(Y_2) = \frac{a_2}{a_2 - 1}\theta_2. \tag{24}$$

Thus in order to estimate μ_1 and μ_2, the mean actual annual parental income of students of science and humanity streams, we may use the estimates $\hat{a_1} = 2.14$ and $\hat{a_2} = 2.35$ in (23) and (24) respectively. To obtain an estimate θ_2^* of θ_2 based on the $irss$ observations $Y_{[r:n]r}$, $r = 1,2,\ldots,6$, we have obtained the coefficients $b_{r:6}$, $r = 1,2,\ldots,6$ in the BLUE $\theta_2^* = \sum_{r=1}^{6} b_{r:6} Y_{[r:6]r}$ based on a single cycle and these coefficients in order are 0.18695, 0.15811, 0.12764, 0.09466, 0.05712, 0.00608. The Var(θ_2^*) is obtained as Var$(\theta_2^*) = 0.07409\,\theta_2^2$. Then using the *irss* observations given in Table 4 and the above coefficients we obtain the estimates $\theta_2^*(1) = 32907$, $\theta_2^*(2) = 53430$, $\theta_2^*(3) = 34586$ in cycles 1, 2 and 3 respectively. Then from Remark 8 we obtain the estimate $\tilde{\theta}_2 =$ Rs. 40,308 with Var$(\tilde{\theta}_2) = 0.02470\,\theta_2^2$. Using this value of $\tilde{\theta}_2$ and $\hat{a}_1 = 2.14$ in (23) we obtain a modified version of the moment type estimate of μ_1 as $\hat{\mu}_1 =$ Rs. 75,666 with Var$(\hat{\mu}_1) \simeq 0.08705\,\theta_2^2$. Similarly, using $\hat{a}_2 = 2.35$ and $\tilde{\theta}_2 =$ Rs. 40,308 in (24) we obtain the modified version of the moment type estimate of μ_2 as $\hat{\mu}_2 =$ Rs. 70,164 with Var$(\hat{\mu}_2) \simeq 0.07484\,\theta_2^2$.

Since LEIRSS provides more efficient estimate on θ_2, we compute $\xi_{[1:6]1}^{(2,4)} = 1.47319$ and $\eta_{[1:6]1}^{(2,4)} = 0.58385$. Using these computed values and the observations of *leirss* given in Table 5 in (21) and (22) to obtain $\theta_2^{**}(1) = 53082$, $\theta_2^{**}(2) = 76026$, $\theta_2^{**}(3) = 31564$ in cycles 1, 2 and 3 respectively. The Var(θ_2^{**}) is obtained as Var$(\theta_2^{**}) = 0.04484\,\theta_2^2$. Consequently, by Remark 9, a better estimate $\tilde{\tilde{\theta}}_2 =$ Rs. 53557 is obtained with Var$(\tilde{\theta}_2) = 0.01495\,\theta_2^2$. Thus on using $\hat{a}_1 = 2.14$ and $\tilde{\tilde{\theta}}_2 =$ Rs. 53557 in (23) we obtain a better estimate $\hat{\hat{\mu}}_1$ of μ_1 as $\hat{\hat{\mu}}_1 =$ Rs. 1,00,537 with Var$(\hat{\hat{\mu}}_1) \simeq 0.05268\,\theta_2^2$. Similarly on using $\hat{a}_2 = 2.35$ and $\tilde{\tilde{\theta}}_2 =$ Rs. 67223 in (24) we obtain $\hat{\hat{\mu}}_2 =$ Rs. 93,229 with Var$(\hat{\hat{\mu}}_2) \simeq 0.04530\,\theta_2^2$.

Finally, for making a comparison between the actual average annual parental income and the reported annual parental income, we find the average $\bar{X}_1$ of all the 36 available X-observations of the science students and $\bar{X}_2$ of all the 72 available X-observations of the humanity students. These are obtained a $\bar{X}_1 =$ Rs. 82,745 and $\bar{X}_2 =$ Rs. 50,533. Clearly, among the estimates $\hat{\mu}_1$ and $\hat{\hat{\mu}}_1$ of μ_1, the most efficient estimate is $\hat{\hat{\mu}}_1 = Rs.1,00,537$, which is more than 1.2 times the estimate of average reported parental income of the science students. Similarly, among the estimates $\hat{\mu}_2$ and $\hat{\hat{\mu}}_2$ of μ_2, the most efficient estimate is $\hat{\hat{\mu}}_2 = Rs.93,229$, which is more than 1.8 times the estimate of the average reported parental income of the humanity students.

We can apply the IRSS and LEIRSS procedures to many similar instances to deal with the estimation problems more efficiently. For example, this methodology can be even used to assess the loss of income tax revenue of governments due to under-reporting of incomes from other sources based on assignment of ranks on people in selected sets based on their salary income.

ACKNOWLEDGEMENTS

The authors are very grateful to the referees and the editor for their critical comments and suggestions which have improved the earlier version of the paper. The first author also acknowledges the Kerala State Council for Science, Technology and Environment, Trivandrum for its financial assistance in the form of Emeritus Scientist Fellowship. The second author also acknowledges the UGC, New Delhi for providing financial support in the form of Senior Research Fellowship.

REFERENCES

M. AHMAD, H. MUHAMMAD, H. A. MUTTLAK (2010). *Ranked Set Sampling*. Cambridge Publishing, New Castle Up on Tyne.

M. AL-RAWWASH, M. T. ALODAT, K. M. ALUDAAT, N. ODAT, R. MUHAIDAT (2010). *Prediction intervals for characteristics of future normal sample under moving ranked set sampling*. Statistica, 70, no. 2, pp. 137-151.

N. BALAKRISHNAN (1988). *Recurrence relations from order statistics from n independent and non-identically distributed random variables*. Annals of the Institute of Statistical Mathematics, 40, pp. 273-277.

N. BALAKRISHNAN, S. M. BENDRE, H. J. MALIK (1992). *General relations and identities for order statistics from non-independent non-identical variables*. Annals of the Institute of Statistical Mathematics, 44, pp. 177-183.

R. B. BAPAT, M. I. BEG (1989a). *Order statistics for non identically distributed variables and permanents*. Sankhya, A, 51, pp. 79-93.

R. B. BAPAT, M. I. BEG (1989b). *Identities and recurrence relations for order statistics corresponding to non-identically distributed random variables*. Communications in Statistics - Theory & Methods, 18, pp. 1993-2004.

M. I. BEG (1991). *Recurrence relations and identities for product moments of order statistics corresponding to non-identically distributed random variables*. Sankhya, A, 53, pp. 365-374.

P. K. BHATTACHARYA (1984). *Induced order statistics: Theory and applications*. In P. R. KRISHNAIAH AND P. K. SEN (eds.), *Handbook of Statistics*, 4, North Holland, Amsterdam, pp. 383-403.

M. CHACKO, P. Y. THOMAS (2007). *Estimation of a parameter of bivariate Pareto distribution by ranked set sampling*. Journal of Applied Statistics, 34, pp. 703- 714.

M. CHACKO, P. Y. THOMAS (2008). *Estimation of a parameter of Morgenstern type bivariate exponential distribution by ranked set sampling*. Annals of the Institute of Statistical Mathematics, 60, pp. 301-318.

M. CHACKO, P. Y. THOMAS (2009). *Estimation of parameters of Morgenstern type bivariate logistic distribution by ranked set sampling*. Journal of the Indian Society of Agricultural Statistics, 63, pp. 77-83.

Z. CHEN, Z. BAI, B. K. SINHA (2004). *Ranked Set Sampling: Theory and Applications.* Springer, New York.

H. A. DAVID (1973). *Concomitants of order statistics.* Bulletin of the International Statistical Institute, 45, pp. 295-300.

H. A. DAVID, H. N. NAGARAJA (1998). *Concomitants of order statistics.* In N. BALAKRISHNAN AND C. R. RAO (eds), *Handbook of Statistics*, 16, Elsevier Science, Amsterdam, pp. 487-513.

H. A. DAVID, H. N. NAGARAJA (2003). *Order Statistics (third edition).* John Wiley & Sons Inc, New Jersey.

S. ERYILMAZ (2005). *Concomitants in a sequence of independent non-identically distributed random vectors.* Communications in Statistics- Theory & Methods, 34, pp. 1925-1933.

T. P. HUTCHINSON (1979). *Four applications of bivariate Pareto distribution.* Biometrical Journal, 21, pp. 553-593.

G. LESITHA AND P. Y. THOMAS (2013). *Estimation of a parameter involved in Morgenstern type bivariate Pareto distribution using ranked set sampling.* In C. S. KUMAR, M. CHACKO AND E. I. A. SATHAR (eds.), *Collection of Recent Statistical Methods and Applications*, Department of Statistics, University of Kerala, pp. 171-82.

G. LESITHA, P. Y. THOMAS, M. CHACKO (2010). *Applications of ranked set sampling in estimating parameters of Morgenstern type bivariate logistic distribution.* Calcutta Statistical Association Bulletin, 62, pp. 71-90.

G. A. MCINTYRE (1952). *A method for unbiased selective sampling using ranked sets.* Australian Journal of Agricultural Research, 3, pp. 385-390.

A. PHILIP, P. Y. THOMAS (2015). *On concomitants of order statistics arising from the extended Farlie-Gumbel-Morgenstern bivariate logistic distribution and its application in estimation.* Statistical Methodology, 25, pp. 59-73.

R. S. PRIYA, P. Y. THOMAS (2013). *Estimating the common mean and common standard deviation of several distributions by ranked set sampling.* IAPQR Transactions, 38, pp. 83-96.

R. S. PRIYA, P. Y. THOMAS (2016). *An application of ranked set sampling when observations from several distributions are to be included in the sample.* Communications in Statistics - Theory & Methods, 45, pp. 7040-7052.

N. K. SAJEEVKUMAR, P. Y. THOMAS (2005). *Applications of order statistics of independent and non-identically distributed random variables.* Communications in Statistics - Theory & Methods, 34, pp. 775-783.

P. SAMUEL, P. Y. THOMAS (1998). *Modified expression for a recurrence relation on the product moments of order statistics.* Statistics & Probability Letters, 37, pp. 89-95.

M. H. SAMUH (2017). *Ranked set two-sample permutation test.* Statistica, 77, no. 3, pp. 237-249.

A. B. SHAIBU, H. A. MUTTLAK (2004). *Estimating the parameters of the normal, exponential and gamma distributions using median and extreme ranked set samples.* Statistica, No.1, 64, pp. 75-98.

H. P. SINGH, V. MEHTA (2013). *An improved estimation of parameters of Morgenstern type bivariate logistic distribution using ranked set sampling.* Statistica, 73, no. 4, pp. 437-461.

S. L. STOKES (1997). *Ranked set sampling with concomitant variables.* Communications in Statistics - Theory & Methods, 6, pp. 1207-1211.

P. Y. THOMAS, N. K. SAJEEVKUMAR (2005). *Estimation of the common parameter of several exponential distributions.* IAPQR Transactions, 30, pp. 125-134.

R. J. VAUGHAN, W. N. VENABLES (1972). *Permanent expressions for order statistics densities.* Journal of Royal Statistical Society: Series B, 34, pp. 308-310.

T. G. VEENA, P. Y. THOMAS (2015). *Application of concomitants of order statistics of independent non identically distributed random variables in estimation.* Communications in Statistics - Theory & Methods, 44, pp. 2530-2545.

SUMMARY

The method of ranked set sampling when units are to be inducted from several bivariate populations is introduced in this work. The best linear unbiased estimation of a common threshold parameter of two bivariate Pareto distributions is discussed based on the n ranked set observations, when a sample of size n_1 is drawn from a bivariate Pareto population with shape parameter a_1 and a sample of size n_2 is drawn from another bivariate Pareto with shape parameter a_2 such that $n = n_1 + n_2$. The application of the results of this paper is illustrated with a real life data.

Keywords: Best linear unbiased estimator; Bivariate Pareto distribution; Concomitants of order statistics; Ranked set sampling.