# ON SOME RECENT DEVELOPMENTS IN RANKED SET SAMPLING

Sinha, Arun K.
Department of Statistics, Patna University

https://doi.org/10.5109/12596

# ON SOME RECENT DEVELOPMENTS IN RANKED SET SAMPLING

by

**Arun K. Sinha**

**Arun K. Sinha**

FUKUOKA, JAPAN
2005

# ON SOME RECENT DEVELOPMENTS IN RANKED SET SAMPLING

By

**Arun K. Sinha**[*]

**Abstract**

McIntyre (1952) proposed a cost-effective survey sampling method that is currently known as ranked set sampling (RSS) in the literature. In this method a fairly large number of randomly identified sampling units are partitioned into small subsets of the same size. The units of each subset are ranked separately with respect to the characteristic of interest without obtaining their actual measurements. The information used for this purpose is supposed to be easily available and inexpensive, and exactly one unit of each subset with a specified rank is quantified. As the ranking induces stratification on the population it provides a more structured sample than a simple random sample does with the same size. This sample, in turn, yields more efficient estimators of many parameters of interest than a simple random sample of the same size does. Moreover, its implementation needs only the ranking of the randomly selected units, which does not depend upon the method employed for determining the ranking. Thus, one can use any or all the available information (in absence of actual quantification) including subjective judgment for the purpose. Interestingly, by taking advantage of the experience and expertise of the field personnel it exploits the auxiliary information that is not effectively utilized by standard probability survey sampling designs. Even in the presence of ranking error it provides unbiased and more efficient estimators of various population parameters. It has been successfully used in the several areas on interest. Recently, it has been used under the parametric setting for estimating the parameters of several known distributions. This paper discusses its theory, methods and applications. Apart from McIntyre's method of ranked set sampling (MRSS) there are some more RSS methods. In this paper we also discuss them with some newly developed estimators. This work may be of particular interest for those who have been looking for a cost-effective survey sampling technique.

*Key Words and Phrases:* Cost-effectiveness, Equal and Unequal Allocation, Ranking Information and Error, Relative Precision, Simple Random Sampling, Stratification.

## 1. Introduction

McIntyre (1952) proposed a method of sampling that synthesizes the convenience of purposive sampling with the control of simple random sampling (SRS) and referred to it as *a method of unbiased selective sampling using ranked set.* Contrary to classical approaches that presume stratification of a population it works at stratification of samples. Without providing a rigorous mathematical background he pointed out that (i) the mean of the quantifications provided an unbiased estimator of the population mean

---
[*] Department of Statistics, Patna University, Patna 800 005, Bihar INDIA. arunkrsinha@yahoo.com

regardless of the error involved in ranking the randomly drawn units and (ii) for typical unimodal distributions the mean obtained under perfect ranking was about half of the set size plus half times as efficient as the mean of a simple random sample of the same size. But the rigorous mathematical support to this technique, without referring to the McIntyre's work, was given by Takahasi and Wakimoto (1968) and, independently, by Dell (1969) while investigating the McIntyre's contribution. Like McIntyre, Takahasi and Wakimoto also did not provide any name to the method and considered it as a method of *unbiased estimation of the population mean based on the sample stratified by means of ordering*. The McIntyre's method appears to have lain dormant for almost fourteen years until Halls and Dell (1966) conducted a field survey to examine its effectiveness for estimating weights of browse and of herbage in a pine hardwood forest. In fact, they coined the name *ranked set sampling* (RSS) for this technique and found empirically that it was more efficient than SRS. As errors may get involved while ordering due to dependence on the ranker's judgment, Dell and Clutter (1972) showed that the RSS estimator of a population mean remains unbiased and is at least as efficient as the SRS estimator with the same number of quantifications. They pointed out that its performance would depend upon the characteristics of the population and also on the magnitude of errors in ranking. This method is useful where exact quantification of sampling units with respect to the variable of interest is difficult and/or expensive, but the ranking of randomly selected units in subsets could be performed conveniently. The ranking is carried out with respect to the characteristic of interest by means of some crude method like visual perception that does not require the exact measurements. It, in turn, exploits the experience and expertise of the field personnel for improving upon the efficiency of SRS. One of its strengths is flexibility and model robustness regarding the nature of the auxiliary information needed for ranking. For achieving the cost-effectiveness the method presumes that the quantification rather than identification, acquisition and ranking of sampling units is the main component of the total sampling cost. This appears to be a pragmatic approach for a cost-effective survey sampling technique. Though it was initially proposed for estimating the mean pasture yields more efficiently, it has been used in other sampling situations advantageously. In a recent paper Muttlak and Al-Saleh (2000) discussed some recent developments in RSS. Patil, Sinha and Taillie (1999) provided an up-to-date bibliography; Patil, Sinha and Taillie (1994b) presented a comprehensive review of the literature with its reported and suggested applications. They showed its use for estimating the level of concentration of polychlorinated biphenyls (PCB) in soil while Johnson, Patil and Sinha (1993) described its applications in vegetation research and Mode, Conquest and Marker (1999) discussed its relevance for ecological research. Bohn (1996) presented a review of nonparametric RSS methodology. For estimating the average milk yield the technique is used by Al-Saleh and Al-Shrafat (2001). As it presumes sampling from an infinite population Patil, Sinha and Taillie (1995), and Takahasi and Futatsuya (1988) investigated its characteristics under a finite population scenario. Apart from visual perceptions, ranking may be carried out on the basis of remotely sensed information, prior information, results of earlier sampling episodes, rank correlated covariates, expert-opinion/expert systems, etc., or some combinations of these methods. It is interesting to note that it needs only the ordering of the randomly drawn units with respect to the variable of interest in subsets, and this does not depend upon the method employed to achieve the ordering. Contrary to this procedure other techniques like ratio and regression methods that utilize

the auxiliary information for improving upon the SRS estimator of a population mean require very detailed specifications for the outside information. Besides the McIntyre's RSS (MRSS) method, Takahasi (1970) developed an RSS method that uses the ranking information at the estimation stage unlike MRSS that needs the information at the selection stage. But it is slightly less efficient than the former. On ignoring the ranking information one gets a simple random sample from a Takahasi's RSS (TRSS). Further, multiple characteristics may be estimated more efficiently following these RSS methods, see Norris, Patil and Sinha (1995), and Patil, Sinha and Taillie (1994c). Moreover, Norris, Patil and Sinha (1995) proposed some modifications to TRSS and named it as a modified Takahasi's RSS (MTRSS). It is probably the first paper to apply TRSS and MTRSS in real life situations. These, in turn, have enhanced its applicability tremendously to deal with real-life sampling and monitoring situations more effectively. Ridout (2002) extended the use of the sampling technique in this direction further. Sinha et al. (2001) compared the performances of TRSS and MTRSS for estimating a population mean as compared with the linear regression (LR) estimator with and without double sampling under perfect and concomitant ranking scenarios. This investigation finds the RSS methods better than the LR estimators to estimate a population mean when the correlation between the main variable and the concomitant variable is not very high. Though the MRSS estimator is slightly more efficient than the MTRSS estimator, other advantages of the latter offset this loss. Also, like the former this method reveals its maximum potentiality under perfect ranking scenario. Moreover, out of the two TRSS methods the modified technique performs slightly better than the original one, and it is more suited to deal with real life situations. Under the nonparametric set up Bohn and Wolfe (1992,1994), and Hettmansperger(1995), and Koti and Babu (1996) used the sampling method for developing MRSS based nonparametric methods. Recently, MRSS method has been used for estimating parameters of various known distributions. See Lam, Sinha, and Wu (1994 and 1995), Stokes (1995), Bhoj and Ahsanullah (1996), Chuiv and Sinha (1998). While contributing to this area Bhoj (2001a, 2000 a and b, 1999 a and b, 1997 a, b and c) has also introduced some new RSS methods for estimating parameters. These distributions belong to the family of random variables with cumulative distribution function of the form $F[(x - \mu)/\sigma]$ with $\mu$ and $\sigma$ as the location and scale parameters respectively. It has also been used for testing of hypothesis. In this area Shen (1994b) employed it for testing a normal mean, and Shen and Yuan (1996-97) used it for drawing inference about it. It is employed by Abu-Dayyeh and Muttlak(1996) for conducting hypothesis tests on the scale parameter of the exponential and uniform distributions while Muttlak and Abu-Dayyeh (1998) considered it for testing some hypotheses about normal distribution. In this paper an attempt is made to present a critical review of its theory, methods and applications in both parametric and nonparametric framework. This work may be of some particular interest for both theoretical and applied statisticians.

## 2.   Ranked set sampling methods

The two main RSS methods are known as McIntyre's RSS (MRSS) and Takahasi's RSS (TRSS). These are described below.

## 2.1.    McIntyre's RSS methods

For obtaining the McIntyre's RSS first of all $m$ random samples with $m$ units in each sample are selected from a population with mean, $\mu$ and a finite variance, $\sigma^2$. This is the same as drawing $m^2$ units randomly and then these are partitioned into $m$ equal samples. The $m$ units of each subset are ranked with respect to the variable of interest without using their exact measurements. For this purpose some outside information like visual perception, past experience, etc., are used. Using this ranking information the unit with the smallest rank is quantified from the first subset; the unit with the second smallest rank is measured from the second subset, and this process of quantification is continued until the unit with the $m$th rank is measured from the $m$th subset. This yields $m$ measurements with each of the first $m$ ranks, and these constitute an MRSS of size $m$. For obtaining a larger sample of size $mr$ the whole procedure is repeated $r$ times. Here, $m$ is referred to as the set size while $r$ is called as the number of cycles. In terms of usual notations we get the sample of size, $n = mr$ from the population with its size, $N \geq m^2 r$. This procedure is called as MRSS with balanced (equal) allocation because each rank order consists of r observations. Further, one could utilize the prior information, if available, about the population for obtaining a more efficient estimator of the population parameter of interest. For a skewed population the number of quantifications of the $i$th rank, i.e., $r_i$ is taken as proportional to the standard deviation of the rank order, $\sigma_{(i:m)}$. Because of unequal number of quantifications this allocation may be termed as MRSS with unbalanced allocation (MRSSUA).

## 2.2.    Takahasi's RSS Method

After having selected $m^2$ units randomly from an infinite population and arranged them into $m$ sets with $m$ units each, a unit is randomly selected from each set. Each so selected unit is then quantified and a rank between 1 and m (both inclusive) is assigned to its quantification. Obviously, one may not get the same frequency for each rank order as in the case of the MRSS method and also, there is a possibility of zero frequency for a rank order even after selecting $m^2 r$ units from the population. To deal with these predicaments Takahasi (1970) suggested the McIntyre's method for collecting samples in one cycle. This, in turn, ensures that every rank order gets at least one quantification. This works well so long as one is interested only to estimate the population mean. But this does not help while estimating the variance of the estimator because in this case the variance of each rank order is needed. In view of these facts Norris, Patil and Sinha (1995) suggested to use McIntyre's method in two cycles while using TRSS and referred to it as modified TRSS (MTRSS). This increases its adaptation.

## 3.    RSS with concomitant ranking

RSS presumes that the sampling units are correctly ranked with respect to the variable of interest either before (McIntyre's method) or after its quantification (Takahasi's method). This is called perfect ranking (PR) scenario, but this may not be possible always while dealing with real life situations effectively. In these situations one could take help of some other characteristic for ranking, which is supposedly inexpensive, easily available and highly correlated with the main characteristic of interest. For example, while ranking households with respect to their income one may take help of

type/quality/appearance of the houses in which they live in, the number of adult members in households, etc., when perfect ranking is not possible directly. Unlike perfect ranking scenario the ranking so obtained may be referred to as concomitant ranking (CR) because of its dependence on a concomitant variable. In the terminology of classical sampling this characteristic is called as an auxiliary variable, and we may denote it by Y while the main variable is represented by $X$. See Patil, Sinha and Taillie (1994a and b) for a more detailed discussion on concomitant ranking in RSS.

## 4. McIntyre's estimators

Let us consider only one cycle first and suppose that $X_{(i:m)}$ denotes the $i:m$ th order statistic from the population, and the parentheses are used to surround the subscript to show that $X_{(i:m)}$ are independent unlike the usual $i:m$ th order statistic denoted by $X_{i:m}$. Thus the mean of the ranked set sample, $\overline{X}_{(m)}$, is obtained by $\sum_{i=1}^{m} X_{(i:m)}/m$ while the sample mean $(\overline{X})$ based on the same number of independent and identically distributed ($iid$) quantifications is computed by $\sum_{i=1}^{m} X_{i:m}/m$. Here $X_{(i:m)}$ and $X_{i:m}$ have the same marginal distribution and hence equal variances. As the former are independent while the latter are positively correlated it follows that $\overline{X}_{(m)}$ is more efficient than $\overline{X}$ for estimating the population mean. In general suppose that $X_{(i:m)j}$ denotes the ith order statistic based on perfect ranking in the jth cycle, for $i = 1, \ldots, m$ and $j = 1, \ldots, r$. Note that these are not $iid$ in general, but for a given value of $i$ these are so with $E(X_{(i:m)j}) = \mu_{(i:m)}$ and $var(X_{(i:m)j}) = \sigma_{(i:m)}^2$. The McIntyre's estimator, $\hat{\mu}_{MRSS}$, of the population mean, $\mu$, is defined as follows:

$$\hat{\mu}_{MRSS} = \frac{1}{mr} \sum_{i=1}^{m} \sum_{j=1}^{r} X_{(i:m)j}. \tag{1}$$

Also, if $\hat{\mu}_{(i:m)} = \frac{1}{r} \sum_{j=1}^{r} X_{(i:m)j}$ then

$$\hat{\mu}_{MRSS} = \frac{1}{m} \sum_{i=1}^{m} \hat{\mu}_{(i:m)}.$$

Here $E(\hat{\mu}_{(i:m)}) = \mu_{(i:m)}$; $E(\hat{\mu}_{MRSS}) = \mu$ and $var(\hat{\mu}_{(i:m)}) = \sigma_{(i:m)}^2/r$. Thus we get

$$var(\hat{\mu}_{MRSS}) = \frac{1}{m^2 r} \sum_{i=1}^{m} \sigma_{(i:m)}^2. \tag{2}$$

This expression is also expressed as

$$var(\hat{\mu}_{MRSS}) = \frac{1}{mr} \left[ \sigma^2 - \frac{1}{m} \sum_{i=1}^{m} (\mu_{(i:m)} - \mu)^2 \right]. \tag{3}$$

The alternative expression is useful when the variance of the $i$th order statistic is not available. If we denote the SRS estimator of the population mean with the same sample size, $n = mr$ by $\hat{\mu}_{SRS}$ and then $var(\hat{\mu}_{SRS}) = \frac{\sigma^2}{n}$. This suggests that $var(\hat{\mu}_{MRSS}) < var(\hat{\mu}_{SRS})$. See Bickel (1967) and Tukey (1958) for a proof of this inequality, which follows from the well known positively associated property of order statistics.

## 4.1.   Effect of set size

Takahasi and Wakimoto (1968) show that for positive integers $a$, $b$, $c$ and $d$, if $n > a > c > 1$ and $n = ab = cd$ then

$$var(\overline{X}_n) > var(\overline{X}_{(c)d}) > var(\overline{X}_{(a)b}) > var(\overline{X}_{(n)1}).$$

Here $\overline{X}_n$ represents the sample mean of a simple random sample of size $n$ while $\overline{X}_{(c)d}$ denotes the sample mean of a ranked set sample of size $cd$ obtained with the set size $c$ and the number of cycles $d$. This suggests for taking the largest possible set size to derive the maximum advantage of RSS. But this may not be a pragmatic solution to deal with the real life situations because the ranking of sampling units with respect to the variable of interest is carried out without quantifying them. In view of the limitation of outside information and convenience of the sampler usually the set size is taken as three or four. However, one could consider a larger set size if this number does not pose a problem while carrying out ranking.

## 4.2.   Relative precision and its estimators

The relative precision (RP) of the MRSS estimator, $\hat{\mu}_{MRSS}$, as compared with simple random sample (SRS) estimator, $\hat{\mu}_{SRS}$, with the same sample size, $n$, is computed as follows:

$$RP = \frac{var(\hat{\mu}_{SRS})}{var(\hat{\mu}_{MRSS})}.$$

As $var(\hat{\mu}_{SRS}) = \frac{\sigma^2}{mr}$, this leads to

$$RP = \frac{\sigma^2}{\overline{\sigma^2}}; \text{ where } \overline{\sigma^2} = \frac{\sum_{i=1}^m \sigma_{(i:m)}^2}{m}. \tag{4}$$

The variance of $\hat{\mu}_{MRSS}$ given in equation (3) yields the expression for RP as given below:

$$PR = \frac{1}{1 - \frac{1}{m\sigma^2}\sum_{i=1}^m (\mu_{(i:m)} - \mu)^2}. \tag{5}$$

An equivalent and useful measure could be relative cost (RC) and relative savings (RS). These are defined as

$$RC = \frac{1}{RP} \text{ and } RS = 1 - RC.$$

An expression for RS based on equation (4) is given below:

$$RS = \frac{\sigma^2 - \overline{\sigma^2}}{\sigma^2}. \tag{6}$$

An equivalent expression for RS based on equation (5) is obtained as follows:

$$RS = \frac{1}{m\sigma^2}\sum_{i=1}^m (\mu_{(i:m)} - \mu)^2 \tag{7}$$

McIntyre (1952), and Takahasi and Wakimoto (1968) showed that $1 \leq RP \leq \frac{m+1}{2}$ and so, $0 \leq RS \leq \frac{m-1}{m+1}$. The results suggest that there is no loss due to using MRSS instead

of SRS. Obviously, there is no gain when ranking is the same as a random ordering, i.e., $\mu_{(i:m)} = \mu$. For obtaining an estimator of RP we use

$$E\left[\sum_{j=1}^{r}(X_{(i:m)j} - \hat{\mu}_{(i:m)})^2/(r-1)\right] = \sigma_{(i:m)}^2.$$

An expression for the unbiased estimator of the population variance $\sigma^2$ based on a ranked set sample when the number of cycle is more than one is given by

$$\hat{\sigma}_{MRSS}^2 = \left[\frac{mr - m + 1}{m^2 r(r-1)}\right]\sum_{i=1}^{m}\sum_{j=1}^{r}(X_{(i:m)j} - \hat{\mu}_{(i:m)})^2 + \left(\frac{1}{m}\right)\sum_{i=1}^{m}(\hat{\mu}_{(i:m)} - \mu_{MRSS})^2. \quad (8)$$

For this result see Stokes (1976), Patil, Sinha and Taillie (1994a and 1993b), Norris, Patil and Sinha (1995) and Yanagawa (2000). Stokes (1980) proposed an expression for the estimator of the population variance that is mentioned below:

$$s_{(m)r}^2 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{r}(X_{(i:m)j} - \hat{\mu}_{MRSS})^2}{mr - 1}.$$

Unlike the unbiased estimator this estimator of the population variance could be used to estimate the population variance even if the number of cycle is one. It is biased but asymptotically unbiased as $m$ or $r$ increases because

$$E(s_{(m)r}^2) = \sigma^2 + \frac{\sum_{i=1}^{m}(\mu_{(i:m)} - \mu)^2}{m(mr - 1)}.$$

An estimator of RP, based on unbiased estimators of $\sigma^2$ and $\sigma_{(i:m)}^2$, is obtained as follows:

$$\widehat{RP} = \frac{mr - (m-1)}{mr} + \frac{(r-1)\sum_{i=1}^{m}(\hat{\mu}_{(i:m)} - \mu_{MRSS})^2}{\sum_{i=1}^{m}\sum_{j=1}^{r}(X_{(i:m)j} - \hat{\mu}_{(i:m)})^2}. \quad (9)$$

The Stoke's biased estimator of $\sigma^2$ yields an estimator of RP as given below:

$$\widehat{RP} = \left[\frac{m(r-1)}{(mr-1)}\right]\frac{\sum_{i=1}^{m}(X_{(i:m)j} - \mu_{MRSS})^2}{\sum_{i=1}^{m}\sum_{j=1}^{r}(X_{(i:m)j} - \hat{\mu}_{(i:m)})^2}. \quad (10)$$

But these estimators are useful only when the number of cycles, i.e., $r$ is more than one.

## 4.3. Some more efficient estimators

### (a) Skewed distributions

Contrary to the approach of the same sample size for each rank order, McIntyre (1952) suggested to take the sample size of each rank order proportional to its standard deviation while sampling asymmetrical populations .The former may be referred to as equal allocation approach while the latter may be called as unequal allocation method. It means that if $r_i$ denotes the number of sets having quantified units with rank $i$, then $r_i \propto \sigma_{(i:m)}$ for $i = 1, \ldots, m$. This leads to

$$r_i = \frac{n\sigma_{(i:m)}}{\sum_{i=1}^{m}\sigma_{(i:m)}}; \quad i = 1, \ldots, m. \quad (11)$$

The RSS estimator, $\hat{\mu}_{MRSSUA}$, of the population mean, $\mu$, based on the unequal allocation of samples (also called unbalanced allocation) is given by

$$\hat{\mu}_{MRSSUA} = \frac{1}{m} \sum_{i=1}^{m} \frac{T_i}{r_i} \quad \text{and} \quad var(\hat{\mu}_{MRSSUA}) = \frac{1}{m^2} \sum_{i=1}^{m} \frac{\sigma_{(i:m)}^2}{r_i}$$

where $T_i$ shows the sum of the quantification of the $r_i$ units having $i$th rank order. On putting the value of $r_i$ from equation (11) into the expression for $var(\hat{\mu}_{MRSSUA})$ we have

$$var(\hat{\mu}_{MRSSUA}) = \frac{(\overline{\sigma})^2}{n} \tag{12}$$

where $\overline{\sigma} = \frac{\sum_{i=1}^{m} \sigma_{(i:m)}}{m}$.

The relative precision $(RP_{ua})$ of $\hat{\mu}_{MRSSUA}$ relative to $\hat{\mu}_{SRS}$ with the same number of quantifications is given below:

$$RP_{ua} = \frac{var(\hat{\mu}_{SRS})}{var(\hat{\mu}_{MRSSUA})}.$$

This yields that

$$RP_{ua} = \frac{\sigma^2/n}{\sum_{i=1}^{m} \frac{\sigma_{(i:m)}^2}{r_i}/m}. \tag{13}$$

This could also be expressed as

$$RP_{ua} = \left(\frac{\sigma}{\overline{\sigma}}\right)^2. \tag{14}$$

See Patil, Sinha and Taillie (1993b) for these results. Further, it is interesting to note that

$$var(\hat{\mu}_{MRSS}) - var(\hat{\mu}_{MRSSUA}) = \frac{\sum_{i=1}^{m} (\sigma_{(i:m)} - \overline{\sigma})^2}{mn}.$$

This proves that $RP_{ua} \geq RP$. Takahasi and Wakimoto (1968) show that $0 \leq RP_{ua} \leq m$. For showing the relative savings due to these MRSS methods we consider the standard Pareto and the standard lognormal distributions, two well known asymmetrical distributions with wide applications in real life situations. The variances of the order statistics of the former distribution computed by Malik (1966) while those of the latter given by Gupta, McDonald and Galarneau (1974) are used.

Table 1. Relative savings (RS) in percentage due to the RSS estimators of the population mean based on the unequal (Neyman) and the equal allocations (under parentheses) as compared with the corresponding SRS estimator while sampling the standard Pareto population with parameter, $k$.

| Set Size | $k$ | | | | |
|:---:|:---:|:---:|:---:|:---:|:---:|
| $(m)$ | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 |
| 2 | 39 | 37 | 37 | 36 | 36 |
| | (7) | (12) | (14) | (16) | (17) |
| 3 | 56 | 54 | 53 | 53 | 52 |
| | (13) | (19) | (24) | (26) | (28) |
| 4 | 64 | 63 | 62 | 62 | 61 |
| | (17) | (25) | (30) | (33) | (35) |
| 5 | 70 | 69 | 68 | 68 | 68 |
| | (19) | (29) | (35) | (38) | (40) |

Table 2. Relative savings (RS) in percentage due to the RSS estimators of the population mean based on the unequal (Neyman) and the equal allocations as compared with the corresponding SRS estimator while sampling the standard lognormal population.

| Allocation | Set Size | | | |
|:---:|:---:|:---:|:---:|:---:|
| | 2 | 3 | 4 | 5 |
| Unequal | 37 | 53 | 62 | 68 |
| Equal | 16 | 25 | 32 | 37 |

Though the unequal allocation based on Neyman's approach improves the performance of RSS further, this requires the sample size for each rank order to be proportional to its standard deviation that is usually unavailable. This impasse, in turn, makes its application difficult. This scenario appears to be the same as that of stratified sampling, which suggests to use the result of previously conducted similar surveys or the result of a preliminary survey based on a smaller size. Kaur, Patil and Taillie (1997) considered these issues and examined two right-tail allocation models that give more quantification to (i) the largest order statistic and (ii) the two largest order statistics. These perform better than the equal allocation model.

**Estimator of $RP_{ua}$**

For obtaining an estimator of the RPua we use the estimator of the population variance and that of the population variance of the ith rank order based on unequal sample sizes as given below.

$$\hat{\sigma}^2_{MRSSUA} = \sum_{i=1}^{m} \left[ \frac{m(r_i-1)+1}{m^2 r_i(r_i-1)} \right] \sum_{j=1}^{r_i} (X_{(i:m)j} - \overline{X}_{(i:m)})^2$$

$$+ \left( \frac{1}{m} \right) \sum_{i=1}^{m} (\overline{X}_{(i:m)} - \overline{X}_{(m)r})^2. \tag{15}$$

$$\hat{\sigma}^2_{(i:m)UA} = \frac{\sum_{j=1}^{r_i}(X_{(i:m)j} - \hat{\mu}_{(i:m)UA})^2}{r_i - 1} \text{ when } \hat{\mu}_{(i:m)UA} = \frac{1}{m}\sum_{i=1}^{m}\frac{T_i}{r_i}. \tag{16}$$

See Norris, Patil and Sinha (1995).

**(b) Symmetric Distributions**

Though MRSS with unequal allocation based on Neyman criterion shows a considerable gain for skewed distributions, its gains for symmetrical distributions are very marginal. The important point to note is that all symmetric distributions do not have the same pattern of the variances of order statistics. One could notice two patterns of variances of symmetric distributions. Under one family of symmetric distributions the variances of the order statistics increase with the rank order until the middle position, and then they decrease to the end (for example, Uniform (0,1), Unfolded Weibull (2,0,1), Symmetric Beta (2) etc.) while for the other family the variances decrease with the rank order until the middle and then they increase to have a symmetric pattern (for example Normal (0,1), Logistic (0,1), Laplace (0,1), etc.). Kurtosis seems to discriminate between the two families of symmetric distributions. For these distributions the optimal allocation quantifies the least variable rank order (s) for giving a high precision estimator of the population mean. For the first family of distributions having uniform and other distributions we need to quantify the extreme order statistics to obtain the minimum variance of the estimator of the population mean. For getting the minimum variance in the other family of distributions having normal and other distributions we quantify either the middle order statistic or the two closest middle order statistics in 1:1 proportion depending on whether the set size is odd or even. Because of this reason Yanagawa and Chen (1980) referred to the estimator based on these order statistics as the "median-mean" estimator. For a detailed discussion on these allocations see Patil, Sinha and Taillie (1994a), and Kaur, Patil and Taillie (2000). For the first family of the distributions the $RP_{ext}$ denotes the RP of the RSS estimator of the population mean as compared with the corresponding SRS estimator. It is defined below:

$$RP_{ext} = \frac{\sigma^2}{\sigma^2_{(1:m)}}.$$

Table 3 shows that the performance of equal and Neyman allocation is almost the same while the extreme allocation model performs better than both. Though the values of all the three relative precisions increase with set size, there is no effect of these allocations when the set size is two. For the second family of symmetric distributions $RP_{mid}$ gives the RP of the MRSS estimator of the population mean as compared with the SRS estimator and it is defined below:

$$RP_{mid} = \frac{\sigma^2}{\sigma^2_{(m':m)}}$$

where $\sigma^2_{(m':m)}$ denotes the variance of the middle order statistic when the set size is odd and that of the any of the two closest middle order statistics when the set size is even. Table 4 provides the values of the relative precisions for some symmetric distributions. For a more detailed discussion on these allocations see Kaur, Patil and Taillie (2000).

Bhoj (2001a) proposed a new method of RSS with unequal samples for symmetric distributions. But this method presumes that the ranking of sampling units may not be difficult even if their number is large. This might cause difficulty in its implementation in real life situations because in RSS ranking is carried out on the basis of outside information about the characteristic of interest, and because of this reason the set size is kept small. Though MRSS provides a more efficient estimator with unequal allocation than with equal allocation we need to be careful while using this technique because it

may perform worse than equal allocation if it is not carried out properly. This situation resembles with that of stratified random sampling. For planning proper allocations we need some information about the population being sampled. Arnold, Balakrishnan and Nagaraja (1992, p.187) suggested a simple method based on order statistic for this purpose. According to this method for a symmetric population the plot of $(X_{i:m}, X_{m-i+1:m})$ for $i = 1, \ldots, [\frac{m}{2}]$ gives a slope of -1 but for a positively skewed population it is less than -1 while for a negatively skewed population it is more than -1. According to Wilk and Gnanadesikan (1968) the plot of $X_{i:m} + X_{m-i+1:m}$ against $X_{i:m} - X_{m-i+1:m}$ yields a horizontal configuration for a symmetrical population.

Table 3. Relative precisions $RP$, $RP_{ua}$, $RP_{ext}$ of some selected distributions.

| Distribution | Set Size | $RP$ | $RP_{ua}$ | $RP_{ext}$ |
|---|---|---|---|---|
| Uniform(0,1) | 2 | 1.50 | 1.50 | 1.50 |
| Kurtosis=1.8 | 3 | 2.00 | 2.00 | 2.22 |
| | 4 | 2.50 | 2.53 | 3.12 |
| | 5 | 3.00 | 3.05 | 4.20 |
| | | | | |
| Unfolded | 2 | 1.49 | 1.49 | 1.49 |
| Weibull (2,0,1) | 3 | 1.97 | 1.98 | 2.13 |
| Kurtosis=2.19 | 4 | 2.44 | 2.46 | 2.86 |
| | 5 | 2.91 | 2.93 | 3.59 |
| | | | | |
| Symmetric Beta(2) | 2 | 1.49 | 1.49 | 1.49 |
| Kurtosis=2.14 | 3 | 1.98 | 1.98 | 2.04 |
| | 4 | 2.47 | 2.47 | 2.61 |
| | 5 | 2.96 | 2.96 | 3.19 |

Table 4. Relative precisions $RP$, $RP_{ua}$, $RP_{mid}$ of some selected distributions.

| Distribution | Set Size | $RP$ | $RP_{ua}$ | $RP_{mid}$ |
|---|---|---|---|---|
| Normal(0,1) | 2 | 1.47 | 1.47 | 1.47 |
| Kurtosis=3 | 3 | 1.91 | 1.92 | 2.23 |
| | 4 | 2.35 | 2.36 | 2.77 |
| | 5 | 2.77 | 2.80 | 3.49 |
| | | | | |
| Logistic(0,1) | 2 | 1.44 | 1.44 | 1.44 |
| Kurtosis=4.2 | 3 | 1.84 | 1.86 | 2.55 |
| | 4 | 2.22 | 2.27 | 3.16 |
| | 5 | 2.58 | 2.67 | 4.17 |
| | | | | |
| Laplace(0,1) | 2 | 1.39 | 1.39 | 1.39 |
| Kurtosis=6 | 3 | 1.73 | 1.78 | 3.13 |
| | 4 | 2.04 | 2.16 | 3.84 |
| | 5 | 2.33 | 2.54 | 5.70 |

Tables 3 and 4 are based on Tables 1 and 2 of Kaur, Patil and Taillie (2000).

## 5. Takahasi's estimator

For this estimator we need to obtain a sample of size $(n-m)$ by Takahasi's method and one cycle of data through the MRSS method. The estimator of the population mean is given by

$$\hat{\mu}_{TRSS} = \frac{1}{m} \sum_{i=1}^{m} \frac{S_i}{r_i} \tag{17}$$

where $S_i$ denotes the sum of the quantifications having rank $i$ and $r_i \geq 1$ shows the frequency of the $i$th ordered quantification.

The variance of $\hat{\mu}_{TRSS}$ is given below:

$$var(\hat{\mu}_{TRSS}) = \frac{1}{m(n-m+1)} \left[ 1 - \left( 1 - \frac{1}{m} \right)^{n-m+1} \right] \sum_{i=1}^{m} \sigma^2_{(i:m)} \tag{18}$$

where $\sigma^2_{(i:m)}$ denotes the variance of the $i$th order statistic $X_{(i:m)}$ based on perfect ranking.

### 5.1. Modified Takahasi's RSS Estimators

For the estimator, $\hat{\mu}_{MTRSS}$, we draw an RSS of size $(n-2m)$ based on the Takahasi's method and two cycles of data are obtained following McIntyre's method. The expression, when $r_i \geq 2$, is given as:

$$\hat{\mu}_{MTRSS} = \frac{1}{m} \sum_{i=1}^{m} \frac{S_i}{r_i}. \tag{19}$$

The variance of the estimator $\hat{\mu}_{MTRSS}$ is mentioned below:

$$var(\hat{\mu}_{MTRSS}) = \frac{1}{m(n-2m+1)} \left[ 1 - \frac{m}{(n-2m+2)} \left\{ 1 - \left( 1 - \frac{1}{m} \right)^{n-2m+2} \right\} \right] \sum_{i=1}^{m} \sigma^2_{(i:m)}. \tag{20}$$

See Norris, Patil and Sinha (1995) for this result.

## 6. Illustration of RSS methods

For illustrating RSS methods we draw a random sample of 36 units from the data set showing the revenue (in millions of kroner) earned by 40 municipalities in Sweden, see Mukhopadhyay (1998, page 141). The drawn sample in a set of three observations is given below:

| 24 | 386 | 74 | 163 | 37 | 97 | 134 | 128 | 96 | 199 | 290 | 626 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 422 | 101 | 536 | 412 | 250 | 155 | 249 | 240 | 196 | 144 | 249 | 63 |
| 1277 | 230 | 467 | 241 | 159 | 111 | 55 | 623 | 288 | 277 | 720 | 488 |

Suppose we wish to draw an MRSS with equal allocation of size, $n = 12$ while keeping the set size, $m = 3$ and the number of cycles, $r = 4$. For drawing the sample the ranking is carried out vertically in each column and the lowest value is taken from

the first column, the value having the second rank is selected from the second column and the highest value is drawn from the third column. This process is repeated for each set of three columns and the values so selected are given below.

| Rank | | |
|---|---|---|
| 1 | 2 | 3 |
| 24 | 230 | 536 |
| 163 | 159 | 155 |
| 55 | 240 | 288 |
| 144 | 290 | 626 |

As $\hat{\mu}_{(i:m)} = \frac{1}{r} \sum_{j=1}^{r} X_{(i:m)j}$, we obtain $\hat{\mu}_{(1:m)} = 96.5$, $\hat{\mu}_{(2:m)} = 229.8$, $\hat{\mu}_{(3:m)} = 401.3$. Thus an estimate of the population mean, $\hat{\mu}_{MRSS} = \frac{1}{m} \sum_{i=1}^{m} \hat{\mu}_{(i:m)} = 242.433$. As we compute Est.$\sigma_{(1:3)}^2 = 4556.25$, Est.$\sigma_{(2:3)}^2 = 2916$ and Est.$\sigma_{(3:3)}^2 = 47524$, $var(\hat{\mu}_{MRSS}) = 1527.67$. For MRSS with unequal allocation we compute $r_1 = 2$, $r_2 = 2$ and $r_3 = 8$. Thus we have the following values as shown below in a tabular form.

| Rank | Values | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 24 | 163 | | | | | | |
| 2 | 230 | 159 | | | | | | |
| 3 | 536 | 155 | 249 | 623 | 288 | 277 | 720 | 626 |

Here we get $\hat{\mu}_{(1:m)} = 93.5$, $\hat{\mu}_{(2:m)} = 98.3$, $\hat{\mu}_{(3:m)} = 50.2$ with Est.$\sigma_{(1:3)}^2 = 9662.89$, Est.$\sigma_{(2:3)}^2 = 2520.04$, Est.$\sigma_{(3:3)}^2 = 46096.09$. Then we obtain an estimate of the population mean, $\hat{\mu}_{MRSSUA} = 240.733$ and $var(\hat{\mu}_{MRSSUA}) = 1317.05$.

For MTRSS we collect the two cycles of the data set following the MRSS procedure and then the TRSS is used. The final data set is given below:

| Rank | Values | | | | |
|---|---|---|---|---|---|
| 1 | 24 | 163 | 128 | 96 | |
| 2 | 230 | 159 | 134 | 199 | 290 |
| 3 | 536 | 155 | 626 | | |

Here we compute $\hat{\mu}_{(1:m)} = 102.8$, $\hat{\mu}_{(2:m)} = 202.4$, $\hat{\mu}_{(3:m)} = 439$ with Est.$\sigma_{(1:3)}^2 = 3504.64$, Est.$\sigma_{(2:3)}^2 = 3757.69$, Est.$\sigma_{(3:3)}^2 = 62500$. These yield an estimate of the population mean, $\hat{\mu}_{MTRSS} = 248.067$ and $var(\hat{\mu}_{MTRSS}) = 2124.87$. Further, the population variance is estimated for the three samples. The estimates of the population variance based on the MRSS, MRSSUA and MTRSS are obtained as 30767.7, 37212.9 and 32007.3 respectively. These, in turn, give the variances of the estimator of the population mean based on SRS as 2563.97, 3101.08 and 2667.25 respectively. Thus the estimates of the RP of the MRSS, MRSSUA and TMRSS estimators of the population mean relative to the respective SRS estimator are obtained as 1.678, 2.355 and 1.255 respectively.

## 7.  RSS methods with concomitant ranking

In order to obtain the expressions of the variances of the three RSS estimators under concomitant ranking we assume that there is a linear relationship between the

main variable $X$ and the concomitant variable $Y$. This yields that

$$E\{X_{[i:m]}|Y_{(i:m)}\} = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(Y_{(i:m)} - \mu_Y)$$

where $X_{[i:m]}$ denotes the $i$th order statistic of $X$ based on a concomitant ranking whereas $Y_{(i:m)}$ shows the $i$th order statistic of $Y$ based on perfect ranking.

Thus, the expression

$$var(X_{[i:m]}) = E\{var(X|Y_{(i:m)})\} + var\{E(X|Y_{(i:m)})\}$$

leads to the following expression for the $var(X_{[i:m]})$ for the standard bivariate normal distribution:

$$\sigma^2_{[i:m]} = (1 - \rho^2) + \rho^2\sigma^2_{(i:m)}$$

where $\sigma^2_{[i:m]}$ denotes the $var(X_{[i:m]})$ and $\rho$ shows the correlation between $X$ and $Y$. For obtaining the variances of the three RSS estimators under the concomitant ranking scenario while sampling the standard bivariate normal population the expression $\sum_{i=1}^{m}\sigma^2_{(i:m)}$ is replaced by $\sum_{i=1}^{m}\sigma^2_{[i:m]}$ in relevant equations. See Sinha et al. (2001), and Patil, Sinha and Taillie (1993a) for more details.

## 8. RSS under parametric framework

Let $X$ be a random variable with distribution function within the class $F[\frac{(x-\mu)}{\sigma}]$ where $\mu$ and $\sigma$ are the location and scale parameters respectively. Barnett and Moore (1997) showed, following Lloyd (1952), that the best linear unbiased order statistics estimator (BLUE) based on ordering a single random sample $x_1, x_2, \ldots, x_n$ is given by

$$\theta^* = (A'V^{-1}A)^{-1}A'V^{-1}x$$

with variance-covariance matrix as

$$var(\theta^*) = \sigma^2(A'V^{-1}A)^{-1}$$

where $A = (1, \alpha)$ with $\alpha' = (\alpha_i)$ and $\alpha_i = E(U_i)$ when $U_{(i)} = \frac{X_{(i)} - \mu}{\sigma}$. $V = (\sigma_{ij})$ is the variance-covariance matrix of the reduced order statistics $U_{(i)}$, $1' = (1, 1, \ldots, 1)$ and $x'$ is the set of ordered sample values $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$.

For RSS $x'$ is replaced by $(x_{(1:n)}, x_{(2:n)}, \ldots, x_{(n:n)})$. Further, the diagonal matrix $W = \text{diag}(\sigma_{ij})$ replaces $V$ because these sample values are uncorrelated as they come from independent samples. Thus the ranked set BLUE of the parameter vector $\theta'(\mu, \sigma)$ is obtained. Here

$$\mu^* = \sum_{i=1}^{n}\gamma_i x_{(i:n)} \quad \text{and} \quad \sigma^* = \sum_{i=1}^{n}\eta_{(i)} x_{(i:n)}$$

where

$$\gamma_i = \frac{\frac{1}{\sigma_{ii}}\left[\sum_{j=1}^{n}(\alpha_j^2/\sigma_{jj}) - \alpha_i\sum_{j=1}^{n}(\alpha_j/\sigma_{jj})\right]}{B},$$

$$\eta_i = \frac{\frac{1}{\sigma_{ii}}\left[\sum_{j=1}^{n}\alpha_j(1/\sigma_{jj}) - \sum_{j=1}^{n}(\alpha_j/\sigma_{jj})\right]}{B},$$

$$\text{and } B = \sum_{j=1}^{n}\left(\frac{\alpha_j^2}{\sigma_{jj}}\right)\sum_{j=1}^{n}\frac{1}{\sigma_{jj}} - \left[\sum_{j=1}^{n}\left(\frac{\alpha_j}{\sigma_{jj}}\right)\right]^2.$$

Further, we obtain that $var(\mu^*) = \frac{\sigma^2 \sum_{i=1}^n \alpha_i^2/\sigma_{ii}}{B}$ and $var(\sigma^*) = \frac{\sigma^2 \sum_{i=1}^n 1/\sigma_{ii}}{B}$.

These results get simplified substantially for symmetric $X$. Sinha, Sinha and Purkayastha (1996), and Barnett and Moore (1997) consider the case of normal distribution and obtain the following estimators

$$\mu^* = \frac{\sum x_{(i:n)}/\sigma_{ii}}{\sum 1/\sigma_{ii}}, \quad \sigma^* = \frac{\sum \alpha_i x_{(i:n)}/\sigma_{ii}}{\sum \alpha_i^2/\sigma_{ii}}$$

with

$$var(\mu^*) = \frac{\sigma^2}{\sum 1/\sigma_{ii}}, \quad var(\sigma^*) = \frac{\sigma^2}{\sum \alpha_i^2/\sigma_{ii}}.$$

Note that these estimators involve variances of order statistics, which are usually not known.

## 9. Comparison of the RSS estimators of the population mean

In this subsection we examine the variances of the RSS estimators of the population mean of the normal population under BLUE, i.e., $var(\mu^*)$ BLUE, median MRSS estimator represented by $var(\hat{\mu}_{MEDIAN})$, the Neyman allocation based MRSS estimator, i.e., $var(\hat{\mu}_{MRSSUA})$ and the usual MRSS estimator denoted by $var(\hat{\mu}_{MRSS})$. These estimators could arise while sampling a standard normal population following MRSS technique. These values, shown in Table 5, reveal that there is no differences among these variances for the set size two, but when it is more than two we find that $var(\hat{\mu}_{MEDIAN}) < var(\mu^*)_{BLUE} < var(\hat{\mu}_{MRSSUA}) < var(\hat{\mu}_{MRSS})$.

Table 5. Values of $var(\mu^*)_{BLUE}$, $var(\hat{\mu}_{MEDIAN})$, $var(\hat{\mu}_{MRSSUA})$, $var(\hat{\mu}_{MRSS})$ while sampling a standard normal population with set size $(n) = 2(1)6$.

| Set Size $(n)$ | $var(\mu^*)_{BLUE}$ | $var(\hat{\mu}_{MEDIAN})$ | $var(\hat{\mu}_{MRSSUA})$ | $var(\hat{\mu}_{MRSS})$ |
|---|---|---|---|---|
| 2 | 0.34084 | 0.34084 | 0.34084 | 0.34084 |
| 3 | 0.17230 | 0.14956 | 0.17373 | 0.17418 |
| 4 | 0.10399 | 0.09011 | 0.10589 | 0.10652 |
| 5 | 0.06956 | 0.05737 | 0.07151 | 0.07220 |
| 6 | 0.04979 | 0.04103 | 0.05164 | 0.05232 |

Further, using BLUE based on an MRSS Chuiv and Sinha (1998) discuss the estimation of (i) normal mean and variance (ii) an exponential mean (iii) the location and scale parameters of a two-parameter exponential distribution (iv) the location parameter of a Cauchy distribution with an unknown scale parameter (v) the location and scale parameters of a logistic distribution (vi) the location and scale parameters of a Weibull distribution and an extreme-value distribution. This is mainly a survey paper based on some recent works of Professor Bimal K Sinha and his associates. For more details see Fei, Sinha and Wu (1994), Lam, Sinha and Wu (1994 and 1995), Chuiv and Sinha (1994), Sinha, Sinha and Purkayastha (1996). The estimation of parameters of a normal population has also been considered by Stokes (1995), and Barnett and Moore (1997), but unlike other authors Barnett and Moore also discussed the situation of imperfect ordering in addition to perfect ranking scenario. They also estimated parameters of exponential distribution. For the estimation of a lognormal mean see Shen (1994a).

The estimation of parameters of the extreme value distribution is considered by Bhoj (1997b) and Fei, Sinha and Wu (1994). Bhoj and Ahsanullah (1996) obtained more efficient estimators of parameters of the generalized geometric distribution using BLUE with MRSS. Professor Dinesh S. Bhoj has developed some new RSS procedures and obtained more efficient estimators of parameters of rectangular, Laplace, logistic, and exponential distributions using BLUE technique based on MRSS and his new RSS. For more details see Bhoj (2001a and b, 2000a and b, 1999a and b, 1997 a, b and c). Also, see Barabesi and El-Shaarawi (2001) for the efficiency of RSS. Tam, Yu and Fung (1998) examined the sensitivity of the BLUE of the population mean to the misspecification of the underlying distribution. They observed that the relative performance of the BLUE depends on the kurtosis, not on the skewness, of the underlying distribution.

## 10.    Maximum likelihood estimators

Using a ranked set sample the maximum likelihood (ML) method of estimation is employed by Stokes (1995). Kvam and Samaniego (1994) described the nonparametric ML (NPML) estimation while Zheng and Al-Saleh (2000) proposed modified ML estimators using an RSS.The asymptotic properties of the NPML estimator based on the sample are discussed by Huang (1997).

## 11.    Estimating distribution function, reliability of a component and Bernoulli parameter

Stokes and Sager (1988) studied the estimation of distribution function and showed that the empirical distribution function (EDF) based on an RSS was unbiased and more efficient than the EDF based on an equal sized SRS. Using extreme and median MRSS Samawi and Al-Sageer (2001) discussed it but under a generalized RSS scenario it was investigated by Kim (1995). Further, Barabesi and Fattorini (2000) developed the kernel estimators of probability density functions using the MRSS procedure under balanced and unbalanced allocation scenarios. Using an RSS El-Neweihi and Sinha (2000 and 2001) considered the estimation of the reliability of a component whose lifetime is exponentially distributed along with the non-negative unbiased variance estimation. The problem was also referred to by Kvam and Samaniego (1993). Kvam (1999) obtained the RSS estimator of a Bernoulli parameter that is superior to the corresponding estimator based on a simple random sample. Li, Sinha and Chuiv (1999), and Lacayo, Nagaraja and Sinha (1999) discussed the problem with more details.

## 12.    Other RSS designs

We may consider two main families of RSS designs that depend on ranking before (selection stage) and after (estimation stage) quantification of sampling units. The MRSS design comes under the first family that includes most of the variations recently suggested mainly for developing relatively more efficient estimators of parameters of interest. Some of these even ignore the cost and the convenience factors. The unequal (unbalanced) allocation scheme proposed by McIntyre (1952) and developed by Yanagawa and Chen (1980), and Patil, Sinha and Taillie (1994a) for obtaining "median-mean" estimators that was further extended by Kaur, Patil and Taillie (1997and 2000) comes under this category. Recently some more papers have appeared for estimating

the population mean using the median and the extreme order statistics. These include Bhoj (2001a), Muttlak (1997 and 1998), Swami, Ahmed and Abu-Dayyeh (1996) and Sinha, Rodriguez and Patil (1995). Samawi and Al-Sageer (2001) used extreme and median RSS for the estimation of the distribution function. Sinha, Sinha and Purkayastha (1996) introduced the concept of partial MRSS that was extended further by Shen and Yuan (1995). Hossain and Muttlak (1999) extended the concept of "quasi-midrange", suggested by Dixon (1954), to MRSS to obtain more efficient estimators of the population mean and standard deviation. They referred it to as paired ranked set sampling. Bhoj (1997a and c, 1999a abd b, 2000a and b) proposed a number of variations to MRSS for obtaining better estimators using the technique of BLUE. A generalized ranked set sampling design as an extension to MRSS is considered by Kim (1995) and Kim and Arnold(1999). Though the efficiency of MRSS estimator increases with the set size, it cannot be increased arbitrarily because of ranking error that could also increase with the set size. In view of these predicaments Al-Saleh and Al-Kadiri (2000), and Al-Saleh and Al-Omary (2002) introduced the double ranked set sampling and multistage ranked set sampling respectively which increase the efficiency without extending the set size. In fact, multistage RSS is just an extension of double RSS or we could also infer that double RSS is a special case of multistage RSS. But these methods require a larger sample size drawn initially that might, in turn, increase the cost of sampling. Al-Saleh and Al-Odat (2001) discussed a variation of ranked set sampling while Al-Saleh and Samawi (2000) investigated the efficiency of Monte Carlo methods with steady state ranked simulated samples. Chen and Shen (2000) considered two-layer RSS with concomitant variables and Chen (2000) evaluated the efficiency of RSS relative to SRS under multi-parameter family. MRSS under Bayesian set up is discussed by Al-Saleh, Al-Sharafat and Muttlak (2000). TRSS and MTRSS belong to the second family of the RSS design. See Norris, Patil and Sinha (1995), and Ridout (2002) for a detailed discussion on this design.

## 13.   Ratio and regression estimators

The ratio estimators based on MRSS and median MRSS were developed by Samawi and Muttlak (1996 and 2001). These are more efficient than the corresponding estimator based on SRS. Comparing MRSS estimator of the population mean with the SRS based linear regression estimator Patil, Sinha and Taillie (1993a) showed that the MRSS estimator was considerably more efficient than the regression estimator unless correlation is very high. They considered the regression estimator with and without double sampling while for MRSS they assumed perfect and concomitant ranking scenarios. They considered that the main and auxiliary variables jointly followed a bivariate normal distribution. Note that this model is favorable to the regression estimator since the latter is unbiased under the bivariate normality while the RSS estimator is always so. Further, Sinha, Rodriguez and Patil (1995), and Muttlak (1998) showed that the median MRSS estimator gave better performance than the equal allocation based MRSS under perfect and concomitant ranking scenarios. Muttlak (1997) reported that the extreme MRSS estimator of the population mean was more efficient than the regression estimator unless the correlation between the variable of interest and the auxiliary variable was more than 0.80. Also, he (1995) used the MRSS to estimate the slope and the intercept of a simple linear regression equation, which are more efficient than the corresponding estimators based on SRS. The author (1996), further, obtained more efficient estimators of the parameters of the multiple regression model based on this sampling technique. Yu and

Lam (1997) proposed a more efficient RSS regression estimator than usual MRSS and SRS estimators unless the correlation between the variable of interest and the concomitant variable is less than 0.4. The method is illustrated using a real data set. While extending the work of Barnett and Moore (1997) further, Barreto and Barnett (1999) obtained the best linear unbiased estimators in the classes of linear combinations of the ranked set sample values for the simple linear regression models with replicated observations. The efficiency of these estimators relative to ordinary least square estimators is quite high for the slope and intercept parameters for the normal case. Considering TRSS and MTRSS Sinha et al.(2001) examined the effectiveness of the estimators of the population mean as compared with the linear regression estimators and found the former more efficient than the latter.

## References

Abu-Dayyeh, W. and Muttlak, H. A. (1996). Using ranked set sampling for hypothesis tests on the scale parameter of the exponential and uniform distributions, *Pak. J. Statist.*, **12** (2), 131–138.

Al-Saleh, M. F. and Al-Kadiri, M. A. (2000). Double ranked set sampling, *Statistics & Probabilty Letters*, **48** (2), 205–212.

Al-Saleh, M. F. and Al-Odat, M. (2000). A variation of ranked set sampling, *Journal of Applied Statistical Science*, **11** (to appear).

Al-Saleh, M. F. and Al-Omary, A. (2002). Multistage ranked set sampling, *Journal of Statistical Planning and Inference*, **102**, 31-44.

Al-Saleh, M. F. and Al-Sharafat, K. (2001). Estimation of average milk yield using ranked set sampling, *Environmetrics*, **12** (to appear).

Al-Saleh, M. F., Al-Sharafat, K. and Muttlak, H. A. (2000). Bayesian ranked set sampling, *Biometrical Journal*, **42**, 489–500.

Al-Saleh, M. F., and Samawi, H. (2000). On the efficiency of Monto Carlo methods using steady state raked simulated samples, *Communication in Statistics (Simulation and Computation)*, **29**, 941–954.

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992). *A First Course in Order Statistics*, Wiley, New York, USA.

Barabesi, L. and El-Shaarawi, A. (2001). The efficiency of ranked set sampling for parameter estimation, *Statistics & Probabilty Letters*, **53**, 189–199.

Barabesi, L. and Fattorini, A. (2000). Kernel estimators of probability density functions by ranked set sampling, Manuscript.

Barnett, V. and Moore, K. (1997). Best linear unbiased estimates in ranked set sampling with particular reference to imperfect ordering, *Journal of Applied Statistics*, **24**, No.6, 697–710.

Barreto, M. C. C. and Barnett, V. (1999). Best linear unbiased estimators for the simple linear regression model using ranked set sampling, *Environmental and Ecological Statistics*, **6**, 119–133.

Bhoj, D. S. (1997a). New parametric ranked set sampling, *Journal of Applied Statistical Science*, **6**, No.4, 275–289.

Bhoj, D. S. (1997b). Estimation of parameters of the extreme value distribution using ranked set sampling, *Commun. Statist. Theory Meth.*, **26**(3), 653–667.

Bhoj, D. S. (1997c). Estimation of parameters using modified ranked set sampling, In : Ahsanullah, M. ed., Applied Statistical Science, II, 145–163. Nova Science Publishers, Inc.USA.

Bhoj, D. S. (1999a). Estimation of parameters of the exponential distribution using three ranked set sampling procedures, *Journal of Applied Statistical Science*, **8**, No. 2/3, 155–172.

Bhoj, D. S. (1999b). Minimum variance linear unbiased estimators of the Rayleigh parameter based on ranked set sampling procedures, In : Ahsanullah, M. and Yildrim ed., Applied Statistical Science, IV, 269–277. Nova Science Publishers, Inc.USA.

Bhoj, D. S. (2000a). New ranked set sampling for one-parameter family of distributions, *Biometrical Journal*, **42**, 5, 647-658.

Bhoj, D. S. (2000b). Estimation of parameters of the Laplace distribution using ranked set sampling procedures, In: Balakrishnan, N. ed., Advances on Methodological and Applied Aspects of Probability and Statistics, 169–182. Gorden and Breach Publishers, Newark, New Jersey, USA.

Bhoj, D. S. (2001a). Ranked set sampling with unequal samples, *Biometrics*, **57**, 957–962.

Bhoj, D. S. (2001b). New ranked set sampling, To appear in *Pakistan Journal of Statistics.*

Bhoj, D. S. and Ahsanullah, M. (1996). Estimation of parameters of the generalized distribution using ranked set sampling, *Biometrics*, **52**, 685–694.

Bickel, P.J. (1967). Some contributions to the theory of order statistics, In: Proc. 5th Berkley Symp., 1, 575–91.

Bohn, L. L. (1996). A review of nonparametric ranked set sampling methodology, *Communications in Statistics, Part A-Theory and Methods*, **25**, 2675–85.

Bohn, L. L. and Wolfe, D. A. (1992). Nonparametric two-sample procedures for ranked set samples data, *Journal of the American Statistical Association*, **87**, 552–561.

Bohn, L. L. and Wolfe, D. A. (1994). The effect of imperfect judgment rankings on properties of procedures based on the ranked set samples analog of the Mann-Whitney- Wilcoxon statistic, *Journal of the American Statistical Association*, **89**, 168–76.

Chen, Z. (2000). The efficiency of ranked set sampling relative to simple random sampling under mutiple-parameter family, *Statistica Sinica*, **10**, 247–263.

Chen, Z. and Shen, L. (2002). Two-layer ranked set sampling with concomitant variables, *Journal of Statistical Planning and Inference*, (to appear).

Chuiv, N.N. and Sinha, B. K. (1994). Estimation of the location parameter of a Cauchy distribution using a ranked set sample, Technical Report, University of Maryland, Baltimore County, USA.

Chuiv, N.N. and Sinha, B. K. (1998). On some aspects of ranked set sampling in parameter estimation, In: Balakrishnan and Rao ed., Handbook of Statistics (Order Statistics: Applications), 17, 337-377. Elsevier, New York.

Dell, T. R. (1969). The theory of some applications of ranked set sampling, Ph. D. Thesis, University of Georgia, Athens, GA, USA.

Dell, T. R., and Clutter, J. L. (1972). Ranked set sampling theory with order statistics background, *Biometrics*, **28**, 545–553.

Dixon, W. J. (1957). Estimates of mean and standard deviation of a normal population, *Annals of Mathematical Statistics*, **31**, 385–391.

El-Neweihi, E. and Sinha, B. K. (2000). Reliability estimation based on ranked set sampling, *Commun. Statist.- Theory Meth.*, **29**(7), 1583–1595.

El-Neweihi, E. and Sinha, B. K. (2001). Some aspects of ranked set sampling in reliability estimation, Manuscript.

Fei, H., Sinha, B. K. and Wu, Z. (1994). Estimation of parameters in two-parameter Weibull and extreme-value distributions using ranked sets sampling, *J. Statist. Res.*, **28**, 149–161.

Gupta, S. S., McDonald, G. C., and Galarneau, D. I. (1974). Moments, product moments and percentage points of the order statistics from the lognormal distribution for samples of size twenty and less, *Sankhya, Series B*, **36**, 230–260.

Halls and Dell (1966). Trial of ranked set sampling for forage yields. *Forest Sci.*, **12**(1), 22–26.

Hettmansperger, T. P. (1995). The ranked set sample sign test. *J. of Nonparametric Statistics*, **4**, 263-270.

Hossain, S. S. and Muttlak, H. A. (1999). Paired ranked set sampling: a more efficient procedure. *Environmetrics*, **10**, 195–212.

Huang, J. (1997). Asymptotic properties of the NPMLE of a distribution function based on ranked set samples, *Annals of Statitistics*, **25**(3), 1036–1049.

Johnson, G. D., Patil, G. P. and Sinha, A. K. (1993). Ranked set sampling for vegetation research, *Abstarcta Botanica*, **17**(1-2), 87–102.

Kaur, A., Patil, G. P., and Taillie, C. (1997). Unequal allocation models for ranked set sampling with skew distributions, *Biometrics*, **53**, 123–130.

Kaur, A., Patil, G. P., and Taillie, C. (2000). Optimal allocation for symmetric distributions in ranked set sampling, *Annals of the Institute of Statistical Mathematics*, **52**, No.2, 239–254.

Kim, Y. H. (1995). Estimation of a distribution function under generalized ranked set sampling, Ph. D.Thesis, University of California, Riverside, USA.

Kim, YoungHee and Arnold, B. C. (1999). Parameter estimation under generalized ranked set sampling. *Statistics & Probabilty Letters*, **42**, 353–360.

Koti, K. M. and Babu, G. J. (1996). Sign test for ranked set sampling. *Communications*

*in Statistics, Part A-Theory and Methods*, **25**, 1617–30.

Kvam, P. (1999). Ranked set sampling for binary data, Manuscript.

Kvam, P. H. and Samaniego, F. S. (1993). On the inadmissibility of empirical averages as estimators in ranked set sampling, *Journal of Statistical Planning and Inference*, **36**, 39–55.

Kvam, P. H. and Samaniego, F. S. (1994). Nonparametric maximum likelihood estimation based on ranked set samples, *Journal of the American Statistical Association*, **89**, 526–537.

Lacayo, H., Nagaraja, N. K. and Sinha, B. K. (1999). Ranked set sampling from a dichotomous population, Manuscript submitted to the Journal of Applied Statistical Sciences.

Lam, K., Sinha, B. K., Wu, Z. (1994). Estimation of parameters in a two-parameter exponential distribution using ranked set sampling, *Ann. Inst. Statist. Math.*, **46**, 723–736.

Lam, K., Sinha, B. K., Wu, Z. (1995). Estimation of location and scale parameters of a Logistic distribution using ranked set sampling, In: Nagaraja, Sen. and Morrison, ed., Papers in Honor of Herbert A. David, 187–197.

Li, D., Sinha, B. K. and Chuiv, N. N. (1999). On estimation of $P(X > c)$ based on a ranked set sample, In U.J. Dixit and M. R. Satam ,ed., Statistical Inference and Design of Experiments, Narosa Publishing House, New Delhi, India.

Lloyd, E. H. (1952). Least squares estimation of location and scale parameters using order statistics, *Biometrika*, **39**, 88–95.

Malik, H. J. (1966). Exact moments of order statistics from the Pareto distribution, *Skandinavisk Aktuarietidskrift*, **49**, 144–157.

McIntyre, G. A. (1952). A method of unbiased selective sampling using ranked sets, *Australian Journal of Agricultural Research*, **3**, 385–390.

Mode, N., Conquest, L. and Marker, D. (1999). Ranked set sampling for ecological research: accounting for the total cost of sampling, *Environmetrics*, **10**, 179–194.

Mukhopadhay, P. (1998). *Theory and Methods of Survey Sampling*, Prentice-Hall of India, New Delhi-110 001, India.

Muttlak, H. A. (1995). Median ranked set sampling, *Journal of Applied Statistical Science*, **6**, No. 4, 245–255.

Muttlak, H. A. (1996). Estimation of parameters in a multiple regression model using ranked set sampling, *Journal of Information & Optimization Sciences*, **17**(3), 521–533.

Muttlak, H. A. (1997). Extreme ranked set sampling: a comparison with regression and ranked set sampling estimators, Manuscript.

Muttlak, H. A. (1998). Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators, *Environmetrics*, **9**, 255–267.

Muttlak, H. A. and Abu-Dayyeh,W. (1998). Testing some hypothese about the normal distribution using ranked set samples: a more powerful test, *Journal of Information*

& *Optimization Sciences*, **19**(1), 1–11.

Muttlak, H. A. and Al-Saleh, M. F. (2000). Recent developments in ranked set sampling, *Pakistan Journal of Statistics*, **16**, 269–290.

Norris, R. C., Patil, G. P. and Sinha, A. K. (1995). Estimation of multiple characteristics by ranked set sampling methods, *Coenoses* (The journal of the International Center for Theoretical and Applied Ecology, Italy), **10**(2-3), 95–111.

Patil, G. P., Sinha, A. K. and Taillie, C. (1993a). Relative precision of ranked set sampling: Comparison with the regression estimator, *Environmetrics*, **4**(4), 399–412.

Patil, G. P., Sinha, A. K. and Taillie, C. (1993b). Estimation of the population variance using ranked set samples, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, PA, USA. Technical Report No. 93–0501.

Patil, G. P., Sinha, A. K. and Taillie, C. (1994a). Ranked set sampling: A novel method to accomplish observational economy in environmental studies, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, PA, USA. Technical Report No. 94–0411.

Patil, G. P., Sinha, A. K. and Taillie, C. (1994b). Ranked set sampling, In: Handbook of Statistics, **12** (G. P. Patil and C. R. Rao eds.), 167–200. North-Holland, Elsevier Science B. V., Amsterdam.

Patil, G. P., Sinha, A. K. and Taillie, C. (1994c). Ranked set sampling for multiple characteristics, *International Journal of Ecology and Environmental Sciences*, **20**, 357–373.

Patil, G. P., Sinha, A. K. and Taillie, C. (1995). Finite population corrections for ranked set sampling, *Annals of the Institute of Statistical Mathematics*, **47**(4), 621–636.

Patil, G. P., Sinha, A. K. and Taillie, C. (1999). Ranked set sampling: a bibliography, *Environmental and Ecological Statistics*, **6**(1), 91–98.

Ridout, M. S. (2002). A note on ranked set sampling for multiple characteristics, Manuscript.

Samawi, H. M., Ahmed, M. S., and Abu-Deyyeh, W. (1996). Estimating the population mean using extreme ranked set sampling, *Biometrical Journal*, **38**(5), 577–586.

Samawi, H. M., and Al-Sageer, O.A. (2001). On the estimation of the distribution function using extreme and median ranked set sampling, *Biometrical Journal*, **43**, 357–373.

Samawi, H. M., and Muttlak, H. A. (1996). Estimation of ratio using ranked set sampling, *Biometrical Journal*, **38**, 753–764.

Samawi, H. M., and Muttlak, H. A. (2001). On ratio estimation using median ranked set sampling, *Journal of Applied Statistical Sciences*, **10**(2), 89–98.

Shen, W. H. (1994a). On estimation of lognormal mean using a ranked set sample, *Sankhya Series B*, **56**, 323–333.

Shen, W. H. (1994b). Use of ranked set sampling for test of a normal mean, *Calcutta Statistical Association Bulletin*, **44**(175-176), 183–193.

Shen, W. H. and Yuan, W. (1995). A test for a normal mean based on a modified partial ranked set sample, *Pakistan Journal of Statistics*, **11**(3), 227–233.

Shen, W. H. and Yuan, W. (1996-97). Inference for a normal mean based on a ranked set sample, *Parisankhyan Samikkha*, **3-4,** 71–87.

Sinha, A. K., Perez-Abreu, V., Patil, G. P. and Taillie, C. (2001). On the effectiveness of the Takahasi's ranked set sample estimator as compared with the regression estimator, *American Journal of Mathematical and Management Sciences*, **21**, Nos. 1&2,145-163.

Sinha, A. K., Rodriguez, J.E. and Patil, G. P. (1995). A comparison of the McIntyre's ranked set sample estimator with the regression estimator, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, PA, USA. Technical Report No. 95-0303.

Sinha, B. K., Sinha, B. K.and Purkayastha, S. (1996). On some aspects of ranked set sampling for estimation of normal and exponential parameters, *Statistics & Decision*, **14**, 223–240.

Stokes, S. L. (1976). An investigation of the consequences of ranked set sampling, Ph. D. Thesis. Department of Statistics, University of North Caroliina, Chapel Hill, North Carolina, USA.

Stokes, S. L. (1980). Estimation of variance using judgment ordered ranked set samples, *Biometrics*, **36**, 35–42.

Stokes, S. L. (1995). Parametric ranked set sampling, *Annals of the Institute of Statistical Mathematics*, **47**, 465–482.

Stokes, S. L. and Sager, T.W. (1988). Characterization of a ranked set sample with applications to estimating distribution functions, *Journal of the American Statistical Association*, **83**, 374–381.

Takahasi, K. (1970). Practical note on estimation of population means based on samples stratified by means of ordering, *Annals of the Institute of Statistical Mathematics*, **22**, 421–428.

Takahasi, K. and Futatsuya, M. (1988). Ranked set sampling from a finite population, *Proceedings of Inst. Statist. Math.*, **36**, 55-68. (Japanese with English summary).

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering, *Annals of the Institute of Statistical Mathematics*, **20**, 1–31.

Tam, C. Y.C., Yu, P.L.H. and Fung, T.W.K. (1998). Sensitivity analysis of BLUE for the population mean based on a ranked set sample, *Commun. Statist.-Simul.*, **27**(4), 1075–1091.

Tukey, J. W. (1958). A problem of Berkson, and minimum variance orderly estimators, *Annals of Mathematical Statistics*, **29**, 588–92.

Wilk, M. B. and Gnanadeskikan, R. (1968). Probability plotting methods for the analysis of data, *Biometrika*, **55**, 1–17.

Yanagawa, T. (2000). K. Wakimoto's contribution to the ranked set sampling···, The Tenth Japan and Korea Joint Conference of Statistics, 179–184.

Yanagawa, T. and Chen, S. H. (1980). The MG procedure in ranked set sampling, *Jour-

*nal of Statistical Planning and Inference*, **4**, 33–44.

Yu, P. L. H. and Lam, K. (1997). Regression estimator in ranked set sampling, *Biometrics*, **53**, 1070–1080.

Zheng, G. and Al- Saleh, M. F. (2000). Modified maximum likelihood estimators based on ranked set sampling, *Annals of the Institute of Statistical Mathematics*, **54**, 641–658.