REVSTAT – Statistical Journal
Volume 19, Number 1, January 2021, 145–160

MIXED DOUBLE-RANKED SET SAMPLING: A MORE EFFICIENT AND PRACTICAL APPROACH

Authors: Monjed H. Samuh
– Mathematics & Statistics Department, King Fahd University of Petroleum & Minerals, Saudi Arabia
monjedsamuh@ppu.edu

M. Hafidz Omar
– Mathematics & Statistics Department, King Fahd University of Petroleum & Minerals, Saudi Arabia
omarmh@kfupm.edu.sa

M. Pear Hossain
– Mathematics & Statistics Department, King Fahd University of Petroleum & Minerals, Saudi Arabia
pearhossain35@bsmrstu.edu.bd

Received: April 2018 Revised: July 2018 Accepted: December 2018

Abstract:

- A new modification of ranked set sampling (RSS) is investigated to estimate the mean of the study population. This modified approach is a double-stage approach and a kind of combination between RSS and median RSS (MRSS). It is shown that this new modification is more efficient than of RSS, MRSS, and simple random sampling. The Hellinger distance is used to show that the new approach is more practical than any other double-stage RSS.

Keywords:

- *efficiency; Hellinger distance; median; practicality; ranked set sampling.*

AMS Subject Classification:

- 62D05, 62G08.

1. INTRODUCTION

Ranked set sampling (RSS), a data collection scheme, was first implemented by [9] as a good competitor to simple random sampling (SRS) scheme to estimate the mean of Australian pasture yields in agricultural experimentation. Due to its importance to other situations and for a variety of applications in statistics [9] is reprinted in [10]. RSS scheme has recently been getting some attention from researchers working in statistical process control. [11] and [12] for example, proposed different run rules for control charts under different RSS schemes. [19] studied the EWMA control chart for monitoring linear profiles under various RSS schemes. For discussions of some other situations where RSS found applications, see [17], [4], [18], [14], [5], and [13].

[9, 10] claimed that the RSS mean is an unbiased estimator of the population mean and the variance of the RSS mean is smaller than in simple random sampling (SRS) with equal measurement elements. This sampling scheme is useful when it is difficult to measure large number of elements but visually (without inspection) ranking some of them is easier. It involves randomly selecting m sets (each of size m elements) from the study population. The elements of each set are ordered with regards to the study variable, say $\mathbf{X}$, by any negligible cost method or visually without measurements. Finally, the i^{th} minimum from the i^{th} set, $i = 1, 2, ..., m$, are identified for measurement. The obtained sample is called a ranked set sample of set size m. It is worth to observe that visual ranking with large set size is prone to ranking errors. In practice, the set size should be small ($m = 2$, 3, or 4). For more details see [1], [8], and [21].

[25] provided the mathematical theory behind the claims of [9, 10]. They proved the following identities:

1. $$f(x) = \frac{1}{m}\sum_{i=1}^{m} f_{X_{(i)}}(x),$$

2. $$\mu = \frac{1}{m}\sum_{i=1}^{m} \mu_i,$$

3. $$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} \sigma_i^2 + \frac{1}{m}\sum_{i=1}^{m} (\mu_i - \mu)^2,$$

where μ is the mean and σ^2 is the variance of the study population $f(x)$ and μ_i and σ_i^2 are the mean and the variance of the i^{th} ordered statistic. They also showed that the efficiency of the RSS mean *with respect to* (w.r.t.) SRS, defined by the ratio of the variances of the two sample means, is bounded by 1 and $\frac{m+1}{2}$. In particular, when the study population is degenerate then the efficiency is 1, and when the study population is uniform then the efficiency is $\frac{m+1}{2}$.

As claimed by [9, 10] it is later shown in the literature that estimators calculated based on RSS are more efficient than their counterpart in SRS. For example, [24] showed that the empirical distribution function based on RSS is more efficient than its counterpart in SRS. Some authors estimate the parameters of a specific distribution using RSS, see for example [2] and [22].

For improving the efficiency of estimators, some variations of RSS were proposed. [1] suggested double RSS (DRSS), as a method that improves efficiency of the RSS estimators while keeping m fixed. They reported that the RSS mean estimator is less efficient than that based on DRSS. Median RSS (MRSS) is a modification of RSS proposed by [15] to decrease ranking error and to improve the efficiency of the estimators being estimated. The procedure of MRSS is similar to RSS but in lieu of identifying the i^{th} minimum from the i^{th} set only the median of each set is identified. Given odd set size m, the $\left(\frac{m+1}{2}\right)^{\text{th}}$ smallest element is identified from each set for measurement. When m is even, from the first $\frac{m}{2}$ sets the $\left(\frac{m}{2}\right)^{\text{th}}$ smallest element is identified for measurement and from the second $\frac{m}{2}$ sets the $\left(\frac{m}{2}+1\right)^{\text{th}}$ smallest element is identified for measurement. [20] suggested a double MRSS (DMRSS) as an alternative procedure to improve the efficiency of the sample mean. They compared the DMRSS with SRS, RSS, DRSS, and some other sampling schemes and found that DMRSS is the most efficient scheme.

In the process of DMRSS, the data points are identified based on the data points of MRSS. For example, if m is odd, the data points of the DMRSS are just the medians of the data points of MRSS; that is, the data points of DMRSS are the medians of the medians of the SRS. It is clear that identifying median of the medians is a hard process, and this contradict the nature of RSS schemes which require visual comparison without inspection (a rationale originally mentioned by [9]). On the other hand, in the process of DRSS, the data points are identified based on the data points of the RSS. For example, the first data point of DRSS is the minimum of the RSS data points, which is easy to be identified visually without inspection. [1] have shown by the degree of distinguishability and the probability of perfect ranking that ranking an independent and identically (iid) data points is harder than ranking ordered (but independent) data points. Thus, getting a DMRSS is harder than a DRSS. In other words, DRSS is more practical than DMRSS.

To improve the efficiency of RSS estimators, we suggest to combine MRSS scheme with RSS scheme; that is, to apply the method of MRSS on the obtained RSS data points. We shall call this method by mixed double-ranked set sampling (MxDRSS).

Section 2 introduces notations and some basic results. MxDRSS is clarified in Section 3. The practicality of this method is discussed and compared with other methods in Section 4. Estimation of the population mean based on MxDRSS is investigated in Section 5. Numerical results for specific distributions are presented in Section 6. Finally, Section 8 concludes the paper.

2. NOTATION AND SOME BASIC RESULTS

Let X be a continuous random variable with cumulative distribution function (cdf) $F(x)$, probability density function (pdf) $f(x)$, mean μ, and variance σ^2. Let $X_1, X_2, ..., X_m$ be a SRS from $f(x)$, then X_i are iid as $f(x)$. Note that when $f(x)$ is infinite, SRS and random sample are used synonymly.

Suppose $Y_1^{(1)}, Y_2^{(1)}, ..., Y_m^{(1)}$ be a RSS; that is $Y_i^{(1)}$ is the i^{th} order statistic of the random sample $X_1, X_2, ..., X_m$, where the superscript (1) represents stage 1. The cdf of Y_i (see for example [3]) is given by

$$F_{Y_i}(y) = F_{X_{(i)}}(y) = \sum_{k=i}^{m} \binom{m}{k} F^k(y)\,(1-F(y))^{m-k}\,,\; i = 1, 2, ..., m, \tag{2.1}$$

and the pdf of Y_i is

$$f_{Y_i}(y) = m\binom{m-1}{i-1} F^{i-1}(y)\,(1-F(y))^{m-i}\, f(y),\; i = 1, 2, ..., m. \tag{2.2}$$

Let $Y_1^{(2)}, Y_2^{(2)}, ..., Y_m^{(2)}$ be a DRSS; that is $Y_i^{(2)}$ is the i^{th} order statistic of the RSS $Y_1^{(1)}, Y_2^{(1)}, ..., Y_m^{(1)}$ and each of $Y_i^{(2)}$ are obtained from independent ranked set samples of size m. Apparently, $Y_1^{(2)}, Y_2^{(2)}, ..., Y_m^{(2)}$ are the order statistics of the independent (not identical) random variables $Y_1^{(1)}, Y_2^{(1)}, ..., Y_m^{(1)}$. Hence, the cdf of $Y_i^{(2)}$ (see for example [6]) is given by

$$F_{Y_i^{(2)}}(y) = \sum_{l=i}^{m} \sum_{S_l} \left(\prod_{k=1}^{l} F_{Y_{j_k}^{(1)}}(y) \prod_{k=l+1}^{m} \left(1 - F_{Y_{j_k}^{(1)}}(y)\right) \right), \tag{2.3}$$

where S_l is the set of the entire permutations $(j_1, j_2, ..., j_m)$, of the integers $(1, 2, ..., m)$ for which $j_1 < j_2 < \cdots < j_l$, and $j_{l+1} < j_{l+2} < \cdots < j_m$ ([6]). The pdf of $Y_i^{(2)}$ is the derivative of $F_{Y_i^{(2)}}(y)$.

Let $W_1^{(1)}, W_2^{(1)}, ..., W_m^{(1)}$ be a MRSS; that is

$$W_i^{(1)} = \begin{cases} X_{(\frac{m+1}{2})} & \text{if } m \text{ is odd } \&\, i = 1, ..., m, \\ X_{(\frac{m}{2})} & \text{if } m \text{ is even } \&\, i = 1, ..., \frac{m}{2}, \\ X_{(\frac{m+2}{2})} & \text{if } m \text{ is even } \&\, i = \frac{m+2}{2}, ..., m. \end{cases} \tag{2.4}$$

The pdf of $W_i^{(1)}$ is

$$f_{W_i^{(1)}}(x) = \begin{cases} f_{X_{(\frac{m+1}{2})}}(x) & \text{if } m \text{ is odd } \&\, i = 1, ..., m, \\ f_{X_{(\frac{m}{2})}}(x) & \text{if } m \text{ is even } \&\, i = 1, ..., \frac{m}{2}, \\ f_{X_{(\frac{m+2}{2})}}(x) & \text{if } m \text{ is even } \&\, i = \frac{m+2}{2}, ..., m. \end{cases} \tag{2.5}$$

Let $W_1^{(2)}, W_2^{(2)}, ..., W_m^{(2)}$ be a DMRSS; that is

$$W_i^{(2)} = \begin{cases} W^{(1)}_{(\frac{m+1}{2})} & \text{if } m \text{ is odd } \&\, i = 1, ..., m, \\ W^{(1)}_{(\frac{m}{2})} & \text{if } m \text{ is even } \&\, i = 1, ..., \frac{m}{2}, \\ W^{(1)}_{(\frac{m+2}{2})} & \text{if } m \text{ is even } \&\, i = \frac{m+2}{2}, ..., m. \end{cases}$$

The pdf of $W_i^{(2)}$ is

$$f_{W_i^{(2)}}(x) = \begin{cases} f_{W^{(1)}_{(\frac{m+1}{2})}}(x) & \text{if } m \text{ is odd } \&\, i = 1, ..., m, \\ f_{W^{(1)}_{(\frac{m}{2})}}(x) & \text{if } m \text{ is even } \&\, i = 1, ..., \frac{m}{2}, \\ f_{W^{(1)}_{(\frac{m+2}{2})}}(x) & \text{if } m \text{ is even } \&\, i = \frac{m+2}{2}, ..., m. \end{cases}$$

Referring to the procedures of MRSS and DMRSS, it is worth observing that both $W_i^{(1)}$ and $W_i^{(2)}$ are independent over i.

3. MIXED DOUBLE-RANKED SET SAMPLNG

MxDRSS scheme is similar to DRSS but in stage 2 MRSS is applied in lieu of RSS. The following steps describe the procedure of MxDRSS:

1. Choose m sets randomly of size m^2 elements each from the study population.
2. Apply the procedure of RSS on each set of Step 1 to acquire a RSS of size m. This produces m ranked sets (each of size m).
3. Apply the procedure of MRSS on each ranked set in Step 2 to acquire a second stage sample, which we call it a MxDRSS of size m.
4. Repeat Steps 1–3 independently h cycles, if needed, to acquire an MxDRSS of size $n = mh$.

In order to clarify this procedure, it is helpful to refer to some illustrations. First let us denote X_{ijk}, $i, j, k = 1, 2, ..., m$ for the units obtained by Step 1, where i is for the number of sets and $j \times k$ is the size of the i^{th} set. X_{ijk} are iid with common distribution function $F(x)$ and density $f(x)$. Second, let $Y_{ij} = X_{(ijj)}$, $i, j = 1, 2, ..., m$ be the units obtained by Step 2 (Y_{ij} denote the j^{th} order statistic from the i^{th} set). Finally, the units obtained in Step 3 are denoted by Z_i, $i = 1, 2, ..., m$. Tables 1 and 2 explain the procedure when $m = 3$ and 4, respectively.

Table 1: Mixed double-ranked set sampling: $m = 3$.

Step 1	Step 2	Step 3
$X_{111}, X_{112}, X_{113}$ $X_{121}, X_{122}, X_{123}$ $X_{131}, X_{132}, X_{133}$	$Y_{11} = X_{(111)}$ $Y_{12} = X_{(122)}$ $Y_{13} = X_{(133)}$	$Z_1 = Y_{(12)}$
$X_{211}, X_{212}, X_{213}$ $X_{221}, X_{222}, X_{223}$ $X_{231}, X_{232}, X_{233}$	$Y_{21} = X_{(211)}$ $Y_{22} = X_{(222)}$ $Y_{23} = X_{(233)}$	$Z_2 = Y_{(22)}$
$X_{311}, X_{312}, X_{313}$ $X_{321}, X_{322}, X_{323}$ $X_{331}, X_{332}, X_{333}$	$Y_{31} = X_{(311)}$ $Y_{32} = X_{(322)}$ $Y_{33} = X_{(333)}$	$Z_3 = Y_{(32)}$

Table 2: Mixed double-ranked set sampling: $m = 4$.

Step 1	Step 2	Step 3
$X_{111}, X_{112}, X_{113}, X_{114}$ $X_{121}, X_{122}, X_{123}, X_{124}$ $X_{131}, X_{132}, X_{133}, X_{134}$ $X_{141}, X_{142}, X_{143}, X_{144}$	$Y_{11} = X_{(111)}$ $Y_{12} = X_{(122)}$ $Y_{13} = X_{(133)}$ $Y_{14} = X_{(144)}$	$Z_1 = Y_{(12)}$
$X_{211}, X_{212}, X_{213}, X_{214}$ $X_{221}, X_{222}, X_{223}, X_{224}$ $X_{231}, X_{232}, X_{233}, X_{234}$ $X_{241}, X_{242}, X_{243}, X_{244}$	$Y_{21} = X_{(211)}$ $Y_{22} = X_{(222)}$ $Y_{23} = X_{(233)}$ $Y_{24} = X_{(244)}$	$Z_2 = Y_{(22)}$
$X_{311}, X_{312}, X_{313}, X_{314}$ $X_{321}, X_{322}, X_{323}, X_{324}$ $X_{331}, X_{332}, X_{333}, X_{334}$ $X_{341}, X_{342}, X_{343}, X_{344}$	$Y_{31} = X_{(311)}$ $Y_{32} = X_{(322)}$ $Y_{33} = X_{(333)}$ $Y_{34} = X_{(344)}$	$Z_3 = Y_{(33)}$
$X_{411}, X_{412}, X_{413}, X_{414}$ $X_{421}, X_{422}, X_{423}, X_{424}$ $X_{431}, X_{432}, X_{433}, X_{434}$ $X_{441}, X_{442}, X_{443}, X_{444}$	$Y_{41} = X_{(411)}$ $Y_{42} = X_{(422)}$ $Y_{43} = X_{(433)}$ $Y_{44} = X_{(444)}$	$Z_4 = Y_{(43)}$

4. PRACTICALITY OF MxDRSS

In this section, Hellinger distance is defined and used as a *measure of added practicality* and applied to some variations of RSS.

Suppose Y and X are two random variables with density functions $f_Y(x)$ and $f_X(x)$, respectively. The Hellinger distance (see for example [16]) between Y and X is defined by

$$H(X,Y) = \left(1 - \int_{-\infty}^{\infty} \sqrt{f_Y(x) f_X(x)} dx\right)^{\frac{1}{2}}.$$

Obviously, for iid random variables, $H(X,Y) = 0$. So the Hellinger distance between any two data points of the SRS $X_1, X_2, ..., X_m$ is zero. Therefore, identifying the ordered data points (for getting either RSS or MRSS) based on the SRS is difficult. That is, obtaining MRSS and RSS are equivalent in terms of practicality.

Now, given the data points of the RSS $(Y_1^{(1)}, Y_2^{(1)}, ..., Y_m^{(1)})$, and using the pdf's of the order statistics, it can be shown after simple calculation that the Hellinger distances between any pair of RSS data points are given in the third column of Table 3. Note that the Hellinger distances in this case are not zeros; that is, the additional work of identifying the ordered data points of DRSS (i.e., for stage 2) based on the RSS data points (stage 1) is simpler now than using SRS data points.

Table 3: Hellinger distances, $m = 2, 3, 4$; 1^{st} and 2^{nd} stage.

m	(k,l)	stage 1	stage 2
2	$(1,2)$	0.4633	0.5920
3	$(1,2)$	0.4086	0.5473
	$(1,3)$	0.7071	0.8625
	$(2,3)$	0.4086	0.5473
4	$(1,2)$	0.3870	0.5306
	$(1,3)$	0.6501	0.8304
	$(1,4)$	0.8399	0.9628
	$(2,3)$	0.3412	0.4889
	$(2,4)$	0.6501	0.8304
	$(3,4)$	0.3870	0.5306

Now, given the data points of the MRSS $(W_1^{(1)}, W_2^{(1)}, ..., W_m^{(1)})$, and suppose m is odd. Due to the iid case, $H\left(W_k^{(1)}, W_l^{(1)}\right) = 0$ for each $k, l = 1, 2, ..., m$. Therefore, getting a DMRSS based on the MRSS practically is the same as obtaining a MRSS based on the SRS. When m is even, the Hellinger distance is given by

$$H\left(W_k^{(1)}, W_l^{(1)}\right) = \begin{cases} H(W_{\frac{m}{2}}^{(1)}, W_{\frac{m+2}{2}}^{(1)}) > 0 & \text{if } k \leq \frac{m}{2} \ \& \ l > \frac{m}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

Now suppose $Y_1^{(2)}, Y_2^{(2)}, ..., Y_m^{(2)}$ be a DRSS, then the Hellinger distance between any pairs of DRSS data points are shown in the last column of Table 3. It is clear that Hellinger distances are higher in stage 2 than in stage 1.

Similarly, for the DMRSS $W_1^{(2)}, W_2^{(2)}, ..., W_m^{(2)}$, the Hellinger distance is zero when m is odd. When m is even, the Hellinger distance is given by

$$H\left(W_k^{(2)}, W_l^{(2)}\right) = \begin{cases} H(W_{\frac{m}{2}}^{(2)}, W_{\frac{m+2}{2}}^{(2)}) > H(W_{\frac{m}{2}}^{(1)}, W_{\frac{m+2}{2}}^{(1)}) > 0 & \text{if } k \leq \frac{m}{2} \,\&\, l > \frac{m}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

To sum up, for a single stage sampling scheme, MRSS and RSS have same practicality, and since it is shown in the literature that MRSS is more efficient than RSS, we recommend to use MRSS. For a double stage sampling scheme, DRSS is more practical than DMRSS. But, it is shown in the literature DMRSS is more efficient. So, to gain the efficiency provided by applying MRSS, we suggest to mix MRSS with RSS by applying the procedure of MRSS on the data points of RSS. That is, in the first stage we apply RSS and in the second stage we apply MRSS. So, the obtained sample is just a combination between RSS and MRSS and it is a double stage approach, and we call it MxDRSS. The practicality of this new MxDRSS scheme is same as DRSS but in Section 6 we show it is more efficient.

Due to the properties of order statistics $V_1, ..., V_m$, it can be seen that $H(V_1, V_m)$ is the largest distance and $H(V_{\frac{m}{2}}, V_{\frac{m+2}{2}})$ is the minimum distance. Also note that $H(V_1, V_{1+r}) = H(V_{m-r}, V_m)$, $r = 2, ..., m-1$. Apparently increasing m decreases the Hellinger distances for the same pair of order statistics; which is reasonable in the sense that identifying the ordered data points from a small m is easier than in a large m. It can also be concluded from Table 3 that identifying the ordered data points for stage 2 (DRSS) based on the ordered data points of stage 1 (RSS) is consistently easier than identifying the ordered data points for stage 1 (RSS) based on the identical data points of SRS. This result is consistent with the findings of [1].

5. ESTIMATION OF THE POPULATION MEAN

In this section estimation of the population mean is studied. Particularly, in Section 5.1 the population mean estimation is reviewed under the SRS, RSS, and DRSS schemes. In Sections 5.2 and 5.3 the population mean estimation is reviewed respectively under the MRSS and DMRSS schemes and also the results given in the literature about these schemes are enhanced and some new closed form expressions for the variances of the sample means and efficiencies are provided. Finally, in Section 5.4 the population mean estimation is investigated under the proposed MxDRSS scheme.

5.1. Population mean estimation based on SRS, RSS, and DRSS

Let $X_1, X_2, ..., X_m$ be a SRS from $f(x)$. The mean of the sample $\bar{X} = \sum_{i=1}^{m} X_i/m$ is an unbiased estimator of μ with variance σ^2/m.

Let $Y_1^{(1)}, Y_2^{(1)}, ..., Y_m^{(1)}$ be a RSS. It is shown by [25] (see also [26]) that $\bar{Y}^{(1)} = \sum_{i=1}^{m} Y_i^{(1)}/m$ is an unbiased estimator of μ and $\text{Var}(\bar{Y}^{(1)}) \leq \text{Var}(\bar{X})$. [7] reported that $\text{Var}(\bar{Y}^{(1)}) = \sigma^2/m - \sum_{i=1}^{m} \left(\mu_i^{(1)} - \mu\right)^2 / m^2$, where $\mu_i^{(1)}$ is the i^{th} order statistic's mean.

Let $Y_1^{(2)}, Y_2^{(2)}, ..., Y_m^{(2)}$ be a DRSS. [1] reported that the mean $\bar{Y}^{(2)} = \sum_{i=1}^m Y_i^{(2)}/m$ is an unbiased estimator of μ with $\mathrm{Var}(\bar{Y}^{(2)}) = \sigma^2/m - \sum_{i=1}^m \left(\mu_i^{(2)} - \mu\right)^2/m^2$, where $\mu_i^{(2)}$ is the i^{th} order statistic's mean of the RSS $Y_1^{(1)}, Y_2^{(1)}, ..., Y_m^{(1)}$. They also showed that $\mathrm{Var}(\bar{Y}^{(2)}) \leq \mathrm{Var}(\bar{Y}^{(1)})$.

5.2. Population mean estimation based on MRSS

Let $W_1^{(1)}, W_2^{(1)}, ..., W_m^{(1)}$ be a MRSS. Let $\bar{W}^{(1)} = \frac{1}{m}\sum_{i=1}^m W_i^{(1)}$ be the sample mean of MRSS. Then

$$E\left(\bar{W}^{(1)}\right) = \begin{cases} \mu^{(1)}_{\frac{m+1}{2}} & \text{if } m \text{ is odd,} \\ \frac{1}{2}\left(\mu^{(1)}_{\frac{m}{2}} + \mu^{(1)}_{\frac{m+2}{2}}\right) & \text{if } m \text{ is even,} \end{cases}$$

where $\mu_k^{(1)} = E\left(X_{(k)}\right)$. [15] reported that, for symmetric distribution, $\bar{W}^{(1)}$ is an unbiased estimator of μ.

The variance of $\bar{W}^{(1)}$ can be derived as follows:

$$\mathrm{Var}\left(\bar{W}^{(1)}\right) = \mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^m W_i^{(1)}\right).$$

Since the data points of MRSS are independent, then

$$\mathrm{Var}\left(\bar{W}^{(1)}\right) = \frac{1}{m^2}\sum_{i=1}^m \mathrm{Var}\left(W_i^{(1)}\right).$$

Now, from Eq (2.4) and Eq (2.5), we have

$$\mathrm{Var}\left(\bar{W}^{(1)}\right) = \begin{cases} \frac{1}{m}\sigma^{2(1)}_{\frac{m+1}{2}} & \text{if } m \text{ is odd,} \\ \frac{1}{2m}\left(\sigma^{2(1)}_{\frac{m}{2}} + \sigma^{2(1)}_{\frac{m+2}{2}}\right) & \text{if } m \text{ is even,} \end{cases}$$

where $\sigma_k^{2(1)} = \mathrm{Var}\left(X_{(k)}\right)$. Using the result of [7],

$$\mathrm{Var}\left(\bar{W}^{(1)}\right) = \begin{cases} \sigma^2 - \frac{1}{m}\sum_{i=1}^m \left(\mu_i^{(1)} - \mu\right)^2 - \frac{1}{m}\sum_{i:i\neq\frac{m+1}{2}}^m \sigma_i^{2(1)} & \text{if } m \text{ is odd,} \\ \frac{1}{2}\sigma^2 - \frac{1}{2m}\sum_{i=1}^m \left(\mu_i^{(1)} - \mu\right)^2 - \frac{1}{2m}\sum_{i:i\neq\frac{m}{2},\frac{m+2}{2}}^m \sigma_i^{2(1)} & \text{if } m \text{ is even.} \end{cases}$$

5.3. Population mean estimation based on DMRSS

Let $W_1^{(2)}, W_2^{(2)}, ..., W_m^{(2)}$ be a DMRSS. Let $\bar{W}^{(2)} = \frac{1}{m}\sum_{i=1}^{m} W_i^{(2)}$ be the sample mean of DMRSS. Then

$$E\left(\bar{W}^{(2)}\right) = \begin{cases} \mu^{(2)}_{\frac{m+1}{2}} & \text{if } m \text{ is odd,} \\ \frac{1}{2}\left(\mu^{(2)}_{\frac{m}{2}} + \mu^{(2)}_{\frac{m+2}{2}}\right) & \text{if } m \text{ is even,} \end{cases}$$

where $\mu_k^{(2)} = E\left(W_{(k)}^{(1)}\right)$. Using the properties of order statistics and for symmetric distribution it can be shown that $E\left(\bar{W}^{(2)}\right) = \mu$ and the variance of $\bar{W}^{(2)}$ is

$$\text{Var}\left(\bar{W}^{(2)}\right) = \begin{cases} \frac{1}{m}\sigma^{2(2)}_{\frac{m+1}{2}} & \text{if } m \text{ is odd,} \\ \frac{1}{2m}\left(\sigma^{2(2)}_{\frac{m}{2}} + \sigma^{2(2)}_{\frac{m+2}{2}}\right) & \text{if } m \text{ is even,} \end{cases}$$

where $\sigma_k^{2(2)} = \text{Var}\left(W_{(k)}^{(1)}\right)$. Using the result of [1],

$$\text{Var}\left(\bar{W}^{(2)}\right) = \begin{cases} \sigma^2 - \frac{1}{m}\sum_{i=1}^{m}\left(\mu_i^{(2)} - \mu\right)^2 - \frac{1}{m}\sum_{i:i\neq\frac{m+1}{2}}^{m}\sigma_i^{2(2)} & \text{if } m \text{ is odd,} \\ \frac{1}{2}\sigma^2 - \frac{1}{2m}\sum_{i=1}^{m}\left(\mu_i^{(2)} - \mu\right)^2 - \frac{1}{2m}\sum_{i:i\neq\frac{m}{2},\frac{m+2}{2}}^{m}\sigma_i^{2(2)} & \text{if } m \text{ is even.} \end{cases}$$

5.4. Population mean estimation based on MxDRSS

Let $Z_1, Z_2, ..., Z_m$ be a MxDRSS; that is

$$Z_i = \begin{cases} Y^{(1)}_{(\frac{m+1}{2})} & \text{if } m \text{ is odd } \& \, i = 1, ..., m, \\ Y^{(1)}_{(\frac{m}{2})} & \text{if } m \text{ is even } \& \, i = 1, ..., \frac{m}{2}, \\ Y^{(1)}_{(\frac{m+2}{2})} & \text{if } m \text{ is even } \& \, i = \frac{m+2}{2}, ..., m. \end{cases}$$

Referring to the procedure of MxDRSS, one may conclude that Z_i are independent over i, and it is worth observing that they are not identical. The pdf of Z_i is

$$f_{Z_i}(x) = \begin{cases} f_{Y^{(1)}_{(\frac{m+1}{2})}}(x) & \text{if } m \text{ is odd } \& \, i = 1, ..., m, \\ f_{Y^{(1)}_{(\frac{m}{2})}}(x) & \text{if } m \text{ is even } \& \, i = 1, ..., \frac{m}{2}, \\ f_{Y^{(1)}_{(\frac{m+2}{2})}}(x) & \text{if } m \text{ is even } \& \, i = \frac{m+2}{2}, ..., m. \end{cases}$$

Let $\bar{Z} = \frac{1}{m}\sum_{i=1}^{m} Z_i$ be the sample mean of MxDRSS. Then

$$E\left(\bar{Z}\right) = \begin{cases} \mu_{Y^{(1)}_{(\frac{m+1}{2})}} & \text{if } m \text{ is odd,} \\ \dfrac{1}{2}\left(\mu_{Y^{(1)}_{(\frac{m}{2})}} + \mu_{Y^{(1)}_{(\frac{m+2}{2})}}\right) & \text{if } m \text{ is even,} \end{cases}$$

where $\mu_{Y^{(1)}_{(k)}} = E\left(Y^{(1)}_{(k)}\right)$. Using the properties of order statistics and for symmetric distribution it can be shown that $E\left(\bar{Z}\right) = \mu$ and the variance of $\bar{Z}$ is

$$\operatorname{Var}\left(\bar{Z}\right) = \begin{cases} \dfrac{1}{m}\sigma^2_{Y^{(1)}_{(\frac{m+1}{2})}} & \text{if } m \text{ is odd,} \\ \dfrac{1}{2m}\left(\sigma^2_{Y^{(1)}_{(\frac{m}{2})}} + \sigma^2_{Y^{(1)}_{(\frac{m+2}{2})}}\right) & \text{if } m \text{ is even,} \end{cases}$$

where $\sigma^2_{Y^{(1)}_{(k)}} = \operatorname{Var}\left(Y^{(1)}_{(k)}\right)$.

6. NUMERICAL RESULTS FOR SPECIFIC DISTRIBUTIONS

6.1. Results from a uniform distribution

Suppose that the underlying population is uniform U$(0,1)$, then the sample means using SRS, RSS, MRSS, DRSS, DMRSS and MxDRSS of size m are unbiased estimators of μ, while the variances depend on the sampling scheme.

1. For a SRS, $\operatorname{Var}(\bar{X}) = 1/12m$.

2. For a RSS, $\operatorname{Var}(\bar{Y}^{(1)}) = 1/6m(m+1)$, and the relative efficiency (see [25]) w.r.t. SRS is $\operatorname{Eff}(\bar{Y}^{(1)};\bar{X}) = \operatorname{Var}(\bar{X})/\operatorname{Var}(\bar{Y}^{(1)}) = (m+1)/2$.

3. For a MRSS, the variance of the sample mean and the relative efficiency have not been provided in the literature in closed form. However, we find that the following expressions can be obtained for this situation:

$$\operatorname{Var}\left(\bar{W}^{(1)}\right) = \begin{cases} \dfrac{1}{4m(m+2)} & \text{if } m \text{ is odd,} \\ \dfrac{1}{4(m+1)^2} & \text{if } m \text{ is even.} \end{cases}$$

Thus, the relative efficiency w.r.t. SRS is given by

$$\operatorname{Eff}(\bar{W}^{(1)};\bar{X}) = \frac{\operatorname{Var}(\bar{X})}{\operatorname{Var}(\bar{W}^{(1)})} = \begin{cases} \dfrac{m+2}{3} & \text{if } m \text{ is odd,} \\ \dfrac{(m+1)^2}{3m} & \text{if } m \text{ is even.} \end{cases}$$

4. For a DRSS, when $m = 3$, $\text{Var}(\bar{Y}^{(2)}) \approx 0.0092$, and the relative efficiency is $\text{Eff}(\bar{Y}^{(2)}; \bar{X}) = 3.026$. When $m = 4$, $\text{Var}(\bar{Y}^{(2)}) \approx 0.0049$, and the relative efficiency is $\text{Eff}(\bar{Y}^{(2)}; \bar{X}) = 4.281$.

5. For a DMRSS, when $m = 3$, $\text{Var}(\bar{W}^{(2)}) = \sigma_2^{2(2)}/3 \approx 0.0089$, and the relative efficiency is $\text{Eff}(\bar{W}^{(2)}; \bar{X}) = 3.130$. For $m = 4$, $\text{Var}(\bar{W}^{(2)}) = (\sigma_2^{2(2)} + \sigma_3^{2(2)})/8 \approx 0.0047$, and the relative efficiency is $\text{Eff}(\bar{W}^{(2)}; \bar{X}) = 4.422$.

6. For a MxDRSS, when $m = 3$, $\text{Var}(\bar{Z}) = \sigma^2_{Y^{(1)}_{(2)}} \approx 0.0115$, and the relative efficiency is $\text{Eff}(\bar{Z}; \bar{X}) = 2.406$. When $m = 4$, $\text{Var}(\bar{Z}) = (\sigma^2_{Y^{(1)}_{(2)}} + \sigma^2_{Y^{(1)}_{(3)}})/2 \approx 0.0060$, and the relative efficiency is $\text{Eff}(\bar{Z}; \bar{X}) = 3.470$.

So far, we have discussed results for symmetric but rectangular distribution. In the next subsection, we will discuss results for other types of well known distributions.

6.2. Results for the normal, exponential, and skew normal distributions

The relative efficiencies of the sample means obtained by RSS, MRSS, DRSS, DMRSS, and MxDRSS w.r.t. SRS for the normal distribution $N(0, 1)$, skew normal distribution $SN(0, 1, 1)$, and exponential distribution $\text{Exp}(1)$ are summarized in Table 4. Also the results of the uniform distribution $U(0, 1)$ are provided. Table 5 shows the bias and variance of the obtained estimators from the skewed distributions. Moreover, to examine the effect of the kurtosis and skewness on the biasedness and relative efficiency of the considered sampling schemes the gamma distribution $\text{Gamma}(\alpha, 1)$ is used, where α is changed from 1 to 6 (note that increasing α decreases the kurtosis and the skewness) and the results are shown in Figures 1 and 2 for $m = 3$ and $m = 4$, respectively. So, from Figures 1 and 2 one may conclude that:

(**a**) bias is a bit higher for skewed distributions than non-skewed distributions;

(**b**) the efficiency is low for highly skewed distributions.

From the results of Tables 4, 5, and Figures 1 and 2 the remarks below can be observed:

1. In terms of efficiency, the best sampling scheme among those studied in this paper is the DMRSS except for highly skewed distribution like the exponential distribution.

2. As m increases, the efficiency also increases except for the $\text{Exp}(1)$ under DMRSS (it decreases when $m > 2$ as shown by [20]) and MxDRSS (it decreases when $m > 3$). Our MxDRSS scheme shows better performance than DMRSS when $m > 3$.

3. The efficiency is lower for those distributions with large skewness and large kurtosis.

4. In terms of biasedness, the MRSS has the smallest bias.

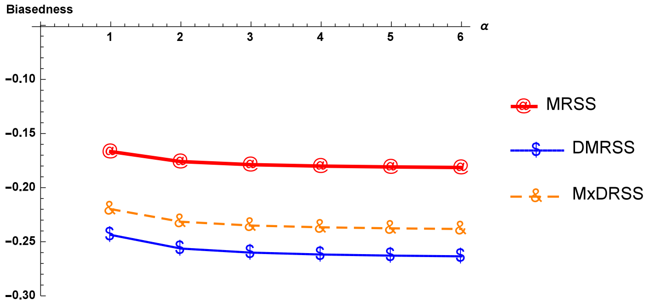
5. The bias is small when the skewness is small.

Table 4: The efficiency in the population mean estimation under the considered sampling schemes w.r.t. SRS.

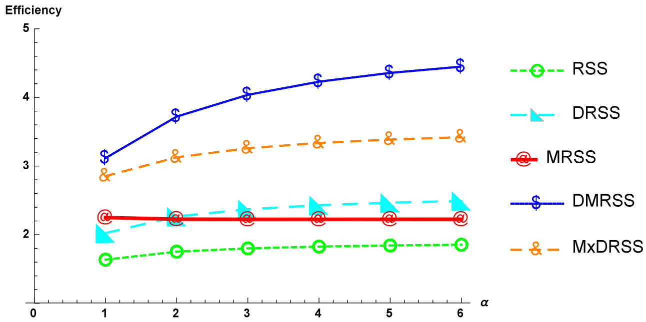
Distribution	(Skewness, kurtosis)	Method	m			
			2	3	4	5
U(0, 1)	(0, −1.2)	RSS	1.500	2.000	2.500	3.000
		MRSS	1.500	1.667	2.083	2.333
		DRSS	1.923	3.026	4.281	5.670
		DMRSS	1.923	3.130	4.422	6.925
		MxDRSS	1.923	2.406	3.470	4.350
N(0, 1)	(0, 0)	RSS	1.467	1.914	2.347	2.770
		MRSS	1.467	2.229	2.774	3.486
		DRSS	1.785	2.633	3.526	4.456
		DMRSS	1.785	4.992	7.091	12.226
		MxDRSS	1.785	3.615	5.046	7.318
SN(0, 1, 1)	(0.137, 0.062)	RSS	1.465	1.909	2.339	2.759
		MRSS	1.465	2.241	2.786	3.500
		DRSS	1.780	2.620	3.503	4.419
		DMRSS	1.780	5.016	7.089	12.030
		MxDRSS	1.780	3.635	5.059	7.290
Exp(1)	(2, 6)	RSS	1.333	1.636	1.920	2.190
		MRSS	1.333	2.250	2.441	2.230
		DRSS	1.516	2.024	2.523	3.016
		DMRSS	1.516	3.116	2.867	2.226
		MxDRSS	1.516	2.854	2.988	2.265

Table 5: The (bias, variance) of the sample mean obtained by MRSS, DMRSS, and MxDRSS for skewed distributions.

Distribution (Skewness, kurtosis)	Method	m		
		3	4	5
SN(0, 1, 1) (0.137, 0.062)	MRSS	(−0.010, 0.101)	(−0.010, 0.061)	(−0.014, 0.039)
	DMRSS	(−0.015, 0.045)	(−0.016, 0.024)	(−0.018, 0.011)
	MxDRSS	(−0.014, 0.062)	(−0.013, 0.034)	(−0.017, 0.018)
Exp(1) (2, 6)	MRSS	(−0.167, 0.120)	(−0.167, 0.075)	(−0.217, 0.043)
	DMRSS	(−0.244, 0.048)	(−0.249, 0.025)	(−0.281, 0.011)
	MxDRSS	(−0.220, 0.068)	(−0.212, 0.039)	(−0.264, 0.019)

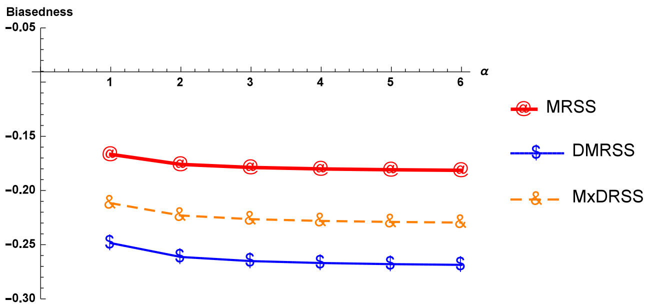


(a) Biasedness.

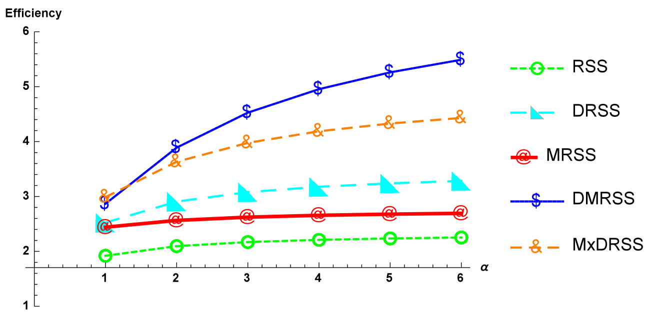


(b) Efficiency.

Figure 1: The effectiveness of the skewness parameter of the Gamma$(\alpha, 1)$ on the biasedness and efficiency of the estimates when $m = 3$ under the considered sampling schemes.



(a) Biasedness.



(b) Efficiency.

Figure 2: The effectiveness of the skewness parameter of the Gamma$(\alpha, 1)$ on the biasedness and efficiency of the estimates when $m = 4$ under the considered sampling schemes.

7. A REAL DATA EXAMPLE

In this section, a real data set is analyzed to illustrate the usefulness of our proposed methodology.

The body mass index (BMI) is a measure of relative size based on the mass and height of an individual. It is commonly employed among children and adults to predict health outcomes. Commonly accepted BMI ranges are underweight: under 18.5, normal weight: 18.5 to 25, overweight: 25 to 30, obese: over 30. A data set that has a BMI for 2107 people is contained in R-package `mixsmsn`. Six types of samples (obtained by using SRS, RSS, MRSS, DRSS, DMRSS, and MxRSS) of size 5 each are presented in Table 6 and the question of interest is to estimate the mean of the BMI. The estimated BMI mean and the standard error of the mean under SRS, RSS, MRSS, DRSS, DMRSS, and MxDRSS are obtained and reported in Table 6.

Table 6: Body mass index example.

	SRS	RSS	MRSS	DRSS	DMRSS	MxDRSS
	20.00	22.00	25.91	22.00	26.36	26.36
	22.62	20.25	21.97	26.89	23.08	28.30
	23.70	26.36	31.63	22.09	28.68	22.09
	32.79	31.96	26.51	30.78	24.86	26.30
	35.18	33.46	34.63	32.64	26.51	23.32
estimated mean	26.858	26.806	28.130	26.880	25.898	25.274
estimated standard error	2.9951	2.6184	2.2361	2.1811	0.9313	1.1257

As suggested by [23] the estimated variance of the sample mean obtained by RSS is given by

$$S^2_{\text{RSS}} = \frac{\sum_{i=1}^m (Y_i^{(1)} - \bar{Y}^{(1)})^2}{m-1}.$$

Accordingly, one may define the estimated variances of the sample means obtained by MRSS, DRSS, DMRSS, and MxDRSS in the same way. For example, in case of MxDRSS,

$$S^2_{\text{MxDRSS}} = \frac{\sum_{i=1}^m (Z_i - \bar{Z})^2}{m-1},$$

and hence the estimated standard error is given by

$$\text{SE}(\bar{Z}) = \sqrt{\frac{S^2_{\text{MxDRSS}}}{m}}.$$

8. CONCLUSION

Practically, given an RSS in stage 1, applying RSS or MRSS in stage 2 is the same because identifying the sample observations is done after the ranking process. But as discussed in Section 6 it is shown that efficiency is higher if we apply MRSS in stage 2. It is also found that efficiency decreases by increases in the kurtosis and skewness. To sum up, DRSS and MxDRSS will behave the same in practicality, but in terms of efficiency MxDRSS is better than DRSS (except for the uniform distribution, which is fatter tailed).

ACKNOWLEDGMENTS

The authors wish to thank the associated editor and referees for their constructive comments which improve the final version of this paper. The authors also would like to acknowledge excellent research support from KFUPM grant number SR151009.

REFERENCES

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

[2] Al-Saleh, M.F. and Diab, Y.A. (2009). Estimation of the parameters of Downton's bivariate exponential distribution using ranked set sampling scheme, *Journal of Statistical Planning and Inference*, **139**(2), 277–286.

[3] Arnold, B.; Balakrishnan, N. and Nagaraja, H. (2008). *A First Course in Order Statistics*, Society for Industrial and Applied Mathematics.

[4] 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**(6), 697–710.

[5] Chen, Z.; Bai, Z. and Sinha, B. (2004). *Ranked Set Sampling: Theory and Applications*, Vol. 176, Springer Science & Business Media.

[6] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*, 3rd ed., Wiley, New York.

[7] Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background, *Biometrics*, **28**(2), 545–555.

[8] Jemain, A.A.; Al-Omari, A. and Ibrahim, K. (2008). Some variations of ranked set sampling, *Electronic Journal of Applied Statistical Analysis*, **1**(1), 1–15.

[9] McIntyre, G. (1952). A method for unbiased selective sampling, using ranked sets, *Australian Journal of Agricultural Research*, **3**(4), 385–390.

[10] McIntyre, G. (2005). A method for unbiased selective sampling, using ranked sets, *The American Statistician*, **59**(3), 230–232.

[11] Mehmood, R.; Riaz, M. and Does, R.J.M.M. (2013). Control charts for location based on different sampling schemes, *Journal of Applied Statistics*, **40**(3), 483–494.

[12] Mehmood, R.; Riaz, M. and Does, R.J.M.M. (2014). Quality quandaries: On the application of different ranked set sampling schemes, *Quality Engineering*, **26**(3), 370–378.

[13] Murff, E.J.T. and Sager, T.W. (2006). The relative efficiency of ranked set sampling in ordinary least squares regression, *Environmental and Ecological Statistics*, **13**(1), 41–51.

[14] Murray, R.A.; Ridout, M.S. and Cross, J.V. (2000). The use of ranked set sampling in spray deposit assessment, *Aspects of Applied Biology*, **57**, 141–146.

[15] Muttlak, H. (1997). Median ranked set sampling, *Journal of Applied Statistical Science*, **6**, 245–255.

[16] Nikulin, M.S. (2001). Hellinger Distance, *Encyclopedia of Mathematics*.

[17] Patil, G.P. (1995). Editorial: ranked set sampling, *Environmental and Ecological Statistics*, **2**(4), 271–285.

[18] Philip, L.H. and Lam, K. (1997). Regression estimator in ranked set sampling, *Biometrics*, **53**(3), 1070–1080.

[19] Riaz, M.; Mahmood, T.; Abbasi, S.A. and Abbas, N. (2017). Linear profile monitoring using EWMA structure under ranked set schemes, *The International Journal of Advanced Manufacturing Technology*, **91**(3), 2751–2775.

[20] Samawi, H.M. and Tawalbeh, E.M. (2002). Double median ranked set sample: comparing to other double ranked samples for mean and ratio estimators, *Journal of Modern Applied Statistical Methods*, **1**(2), 428–442.

[21] Samuh, M.H. and Al-Saleh, M.F. (2011). The effectiveness of multistage ranked set sampling in stratifying the population, *Communications in Statistics — Theory and Methods*, **40**(6), 1063–1080.

[22] Sarikavanij, S.; Kasala, S.; Sinha, B.K. and Tiensuwan, M. (2014). Estimation of location and scale parameters in two-parameter exponential distribution based on ranked set sample, *Communications in Statistics – Simulation and Computation*, **43**(1), 132–141.

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

[24] Stokes, S.L. and Sager, T.W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions, *Journal of the American Statistical Association*, **83**(402), 374–381.

[25] 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), 1–31.

[26] Wolfe, D.A. (2010). Ranked set sampling, *Wiley Interdiscip. Rev. Comput. Stat.*, **2**, 460–466.