Folded Ranked Set Sampling for Asymmetric Distributions

Ahmed Bani-Mustafa[1,a], Amjad D. Al-Nasser[b], Muhammad Aslam[c]

[a]Prince Sultan College Jeddah, Alfisel University
[b]Division of Area Planning & Residents Relations, WRM
[c]Department of Statistics, Forman Christian College University

Abstract

In this paper a new sampling procedure for estimating the population mean is introduced. The performance of the new population mean estimator is discussed, along with its properties, and it is shown that the proposed method generates an unbiased estimator. The relative efficiency of the suggested estimator is computed, in regards to the simple random sample(SRS), and comparisons are made to the ranked set sampling(RSS) and extreme ranked set sampling(ERSS) estimators used for asymmetric distributions. The results indicate that the proposed estimator is more efficient than the estimators based on the ERSS. In addition, the folded ranked set sampling(FRSS) procedure has an advantage over the RSS and ERSS in that it reduces the number of unused sampling units.

Keywords: Ranked set sampling(RSS), extreme ranked set sampling(ERSS), asymmetric distributions.

1. Introduction

Ranked set sampling(RSS) is a two-stage scheme that can be used to increase the precision of estimators (Chen *et al.*, 2004). At the first stage, simple random samples are drawn and a free cost ranking mechanism is employed to rank the units in each simple random sample. At the second stage, measurements of the variable of interest are made on the units selected based on the ranking information obtained at the first stage. The selected observations are a RSS of size n denoted by $X_{i:n}$, $i = 1, 2, \ldots, n$, which represents the ith ordered statistic obtained from the ith SRS of size n. Note that we actually need n^2 observations selected via SRS to obtain n RSS units which mean that we have to, unfortunately, discard $n(n-1)/2$ observations. The RSS procedure may be repeated m times(cycles) to obtain a RSS of size $n \times m$. The majority of research regarding RSS has been concerned with estimating the population mean, where the unbiased estimate of the population mean μ based on RSS is

$$\bar{X}_{RSS} = \frac{1}{m+n} \sum_{j=1}^{m} \sum_{i=1}^{n} X_{(i:n),j},$$

where $X_{(i:n)}$ is the ith ordered statistic from the ith SRS in the jth cycle. The associated variance is given by:

$$\text{Var}\left(\bar{X}_{RSS}\right) = \frac{1}{m+n^2} \sum_{i=1}^{n} \sigma_{(i:n)}^2.$$

[1] Corresponding address: Associate Professor, Prince Sultan College Jeddah, Alfaisel University, P. O. Box 7307 Jeddah 21462, KSA. E-mail: a.ajjour@pscj.edu.sa

Where $\sigma^2_{(i:n)}$ is the variance of the ith order statistic. Since the time when McIntyre (1952) first published a paper advocating the use of ranked sets for unbiased selective sampling, several authors have introduced improvements to the RSS scheme. Sinha and Purkayastha (1996) and Muttlak (1997) suggested using the median ranked set sampling(MRSS) and then, in 2003, Muttlak introduced percentiles ranked set sampling(PRSS). Recently, Al-Nasser (2007) proposed a generalized robust sampling scheme for the RSS, MRSS and PRSS called L ranked set sampling(LRSS) and most recently, Al-Nasser and Bani-Mustafa (2009) suggested a robust ERSS as an alternative sampling scheme. In this paper, we introduce a new sampling scheme(FRSS) which reduces the number of wasted sampling units by at least 50% as well as provides more efficient estimates for skewed distributions. The remainder of the paper is organized as follows. In Section 2, the sampling methods considered in this article for the estimation population mean, ERSS and the proposed FRSS are introduced and then, in Section 3 the performance of the population mean estimator is discussed, along with its properties, and a comparison is made between RSS, ERSS and the proposed FRSS. The final Section briefly summarizes our findings with some concluding remarks.

2. Sampling Schemes

2.1. Extreme ranked set sampling

Samawi *et al.* (1996) suggested using the extreme ranked set sampling(ERSS). This sampling scheme can be summarized in the following steps:

Step 1. Select n SRS's each of size n.

Step 2. Rank the units within each sample with respect to a variable of interest by a visual inspection.

Step 3. Select the 1st unit (minimum) from the first $[n/2]$ SRS for actual measurement.

Step 4. Select the nth unit (maximum) of the last $[n/2]$ SRS for actual measurement.

Step 5. If n is odd then select the median of the last SRS.

Step 6. The cycle may be repeated m times to obtain the desired sample size $n \times m$.

Based on this sampling scheme, the estimator of the population mean is given by:

$$\bar{X}_{ERSS} = \begin{cases} \dfrac{1}{m \times n} \left\{ \displaystyle\sum_{j=1}^{m} \left(\dfrac{n+1}{2}\right) \left(X_{(1:n),j} + X_{(n:n),j}\right) + X_{\left(\frac{n+1}{2}\right),j} \right\}, & \text{if } n \text{ is odd,} \\[4mm] \dfrac{1}{m \times n} \left\{ \displaystyle\sum_{j=1}^{m} \left(\dfrac{n}{2}\right) \left(X_{(1:n),j} + X_{(n:n),j}\right) \right\}, & \text{if } n \text{ is even.} \end{cases} \tag{2.1}$$

This estimator is unbiased only if the underlying distribution is symmetric around zero. The associated variance of this estimator is given by:

$$\text{Var}\left(\bar{X}_{ERSS}\right) = \begin{cases} \left(\dfrac{1}{n}\right)^2 \left\{ \left(\dfrac{n+1}{2}\right) \left(\sigma^2_{(1:n)} + \sigma^2_{(n:n)}\right) + \sigma^2_{\left(\frac{n+1}{2}\right):n} \right\}, & \text{if } n \text{ is odd,} \\[4mm] \left(\dfrac{1}{n}\right)^2 \left\{ \left(\dfrac{n}{2}\right) \left(\sigma^2_{(1:n)} + \sigma^2_{(n:n)}\right) \right\}, & \text{if } n \text{ is even.} \end{cases} \tag{2.2}$$

Table 1: Relative Efficiency of ERSS versus SRS for different skewed distributions

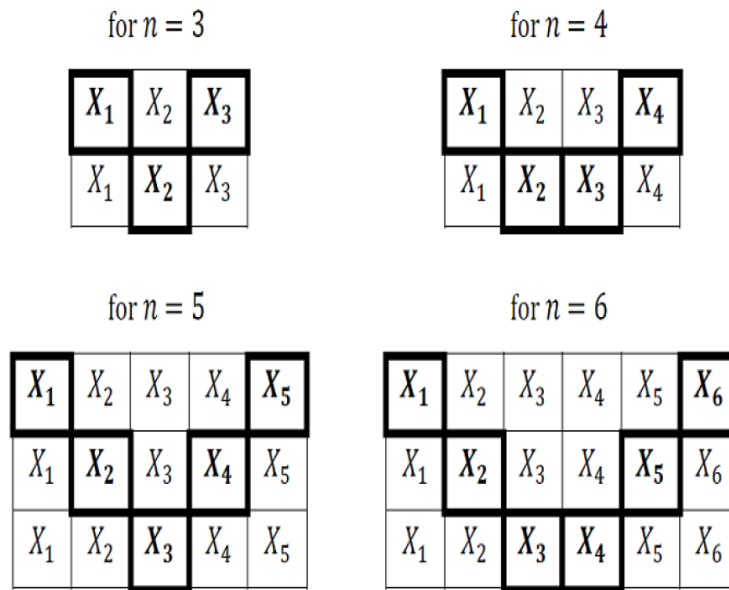| Distribution	Sample size			
	4	5	6	7
Exponential(1)	1.3457	1.3216	0.7518	0.8125
Inverse Gaussian(2, 1)	**0.8696**	**0.8233**	0.7247	0.4646
Log $N(0, 1)$	**0.7515**	**0.7057**	0.4734	0.4541
Pareto(1, 3)	**0.7236**	0.6997	0.5483	0.4183
Gamma(0.5, 1)	1.0760	1.0077	**0.9805**	0.5603
Weibull(0.5, 1)	**0.7279**	0.6853	0.5707	0.3851

Figure 1: *Selection Scheme of FRSS for different sample sizes*

However, as illustrated in Table 1, the simple random sample(SRS) is actually more efficient than ERSS with some skewed populations.

The results suggest a need for a more efficient sampling scheme. Furthermore, all previously discussed sampling schemes waste $n(n - 1)/2$ sampling units. To reduce the costs involved with this wasting of sampling units we have developed a new sampling procedure based on the folded range(quasi range) of the data. This new procedure improves the efficiency of the estimator for skewed distributions and reduces the wasted number of sampling units.

2.2. Proposed sampling scheme: Folded ranked set sampling(FRSS)

In order to plan a FRSS design, n random samples should be selected each of size n, where n is typically small to reduce ranking error. For the sake of convenience it is assumed that the judgment ranking is as good as the actual ranking. FRSS scheme consists of following steps:

Step 1. Select $[(n + 1)/2]$ random samples each of size n.

Step 2. Rank the units within each sample with respect to a variable of interest by a visual inspection.

Step 3. Select the 1st and the nth units from the first sample for actual measurement.

Step 4. Select the first 2nd and the $(n-1)^{th}$ units of the second sample for the actual measurement.

Step 5. The procedure is continued until the nth unit is selected from the sample.

Step 6. The cycle may be repeated m times to obtain the desired sample size.

Without loss of generality, suppose that the cycle is repeated once, Figure 1 clarifies the selected FRSS for different sample sizes (Bold $X's$ are selected).

   Note that the observed data are judgmental order statistics, where these values are dependent if they are observed from the same row while they are independent from the other random samples of different rows. We will refer to this random sample as partial judgment order statistics.

3. Estimating of the Population Mean

Using this sampling plan the estimator of the population mean μ based on a sample of size $n \times m$ is given by:

$$\bar{X}_{FRSS} = \frac{1}{nm} \sum_{j=1}^{m} \sum_{\substack{i=1 \\ i<n-i+1}}^{\left[\frac{n+1}{2}\right]} \left(X_{(i:n),j} + X_{(n-i+1),j} \right),$$

$$\bar{X}_{FRSS} = \frac{1}{nm} \sum_{j=1}^{m} \sum_{1}^{n} X_{(i:n),j}.$$

The expected value of the proposed estimator is given by:

$$E\left(\bar{X}_{FRSS}\right) = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} x_{(i:n)} dF\left(x_{(i:n)}\right).$$

Where the probability density functions of $X_{i:n}$ is given by:

$$f\left(x_{(i:n)}\right) = \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1-F(x)]^{n-i} f(x).$$

Hence,

$$E\left(\bar{X}_{FRSS}\right) = \frac{1}{n} \sum_{i=1}^{n} \mu_{(i:n)} = \mu.$$

Therefore, the proposed estimator is an unbiased estimator of the population mean.

   The variance of the proposed estimator is given by:

$$\text{Var}\left(\bar{X}_{FRSS}\right) = \frac{1}{m \times n^2} \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} \text{Var}\, X_{(i:n),j} + 2 \sum_{\substack{i=1 \\ i<n-i+1}}^{\left[\frac{n+1}{2}\right]} \text{Cov}\left(X_{(i:n),j}, X_{(n-i+1:n),j} \right) \right\} \right]. \qquad (3.1)$$

   Moreover, the relative efficiency of the estimator of the population mean based on ranked samples schemes with respect to the traditional simple random sample(SRS) can be defined by:

$$\text{RE} = \frac{\text{MSE}\left(\bar{X}_{SRS}\right)}{\text{MSE}\left(\bar{X}_{*RSS}\right)}. \qquad (3.2)$$

Where MSE stands for the mean squared error and *RSS could be RSS, ERSS or FRSS; noting that when the given estimator; θ; is unbiased then the MSE(θ) is equivalent to Var(θ).

Regarding to McIntyre (1952), the relative efficiency of the classical RSS with respect to SRS, in estimation of the population mean, is between 1 and $(n + 1)/2$. The RE is not much less than $(n + 1)/2$ when the underlying distribution is symmetric; however, as the underlying distribution became asymmetric the RE dropped down.

Lemma 1. *The maximum value of the RE of FRSS with respect to SRS for estimating the population mean is obtained when the underlying distribution is standard uniform and it is equal to $(n + 1)/3$ if n is even and $n/3$ if n is odd.*

Proof: Based on the results of Balakrishnan and Cohen (1991), assuming $m = 1$, then the mean and the variance of the ith order statistics from a standard uniform distribution are given respectively as

$$E\left(X_{(i)}\right) = p_i = \frac{i}{n + 1} \qquad \text{and} \qquad \text{Var}\left(X_{(i)}\right) = \frac{p_i(1 - p_i)}{n + 2}.$$

Moreover, the covariance between the ith and the jth order statistics is given as:

$$\text{Cov}\left(X_{(i)}, X_{(j)}\right) = \frac{p_i\left(1 - p_j\right)}{n + 2}.$$

Then by substituting the variance and the covariance values in Equation 1, we obtain that

$$\text{Var}\left(\bar{X}_{FRSS}\right) = \frac{1}{n^2}\left[\left\{\sum_{i=1}^{n}\frac{p_i(1 - p_i)}{n + 2} + 2\sum_{\substack{i=1 \\ i<n-i+1}}^{\left[\frac{n+1}{2}\right]}\frac{p_i(1 - p_{n-i+1})}{n + 2}\right\}\right].$$

This formula can be simplified easily to:

$$\text{Var}\left(\bar{X}_{ERSS}\right) = \begin{cases} \dfrac{1}{4n(n + 1)}, & \text{if } n \text{ is even,} \\ \dfrac{1}{4n^2}, & \text{if } n \text{ is odd.} \end{cases} \tag{3.3}$$

Therefore the RE can be computing by dividing the $\text{Var}(\bar{X}_{FRSS}) = 1/12n$ with Equation 2, to be

$$\text{RE}\left(\bar{X}_{SRS}, \bar{X}_{ERSS}\right) = \frac{\bar{X}_{SRS}}{\bar{X}_{RSS}} = \begin{cases} \dfrac{n + 1}{3}, & \text{if } n \text{ is even,} \\ \dfrac{n}{3}, & \text{if } n \text{ is odd.} \end{cases} \tag{3.4}$$

Hence, the Lemma 1.   □

Without loss of generality, we assumed $m = 1$. Then the exact relative efficiency (RE) is computed for some asymmetric distributions in Table 2.

The results indicate that the RE increases as the sample size increases. The new sampling scheme (FRSS) gives a better estimate than the ERSS in most asymmetric distributions when the set size is 4. However, when the set size is more than 4 the FRSS gives increasingly better estimates than ERSS. Furthermore, an advantage of the proposed sampling method(FRSS) over RSS and ERSS, is that the

Table 2: Exact relative efficiency for some asymmetric distributions

Distribution	n = 4			n = 5			n = 6		
	RSS	ERSS	FRSS	RSS	ERSS	FRSS	RSS	ERSS	FRSS
Exp(1)	1.92	1.3457	1.5652	2.1898	1.3216	1.9468	2.4489	1.3165	2.0678
Log N(0,1)	1.4711	0.7515	1.3518	1.5891	0.8057	1.5191	1.6971	0.4541	1.6017
Gamma(2,1)	2.0958	1.4534	1.6152	2.4244	1.6649	2.0653	2.7423	1.0896	2.1825
Gamma(0.5,1)	1.6963	1.0760	1.4873	1.8979	1.0077	1.7754	2.0908	0.9805	1.8974
Weibull(0.5,1)	1.3345	0.7236	1.2809	1.4249	0.6853	1.4002	1.5094	0.5707	1.4755
Weibull(2,1)	2.3251	2.0532	1.6694	2.7436	2.3327	2.2167	3.1551	2.4588	2.3177
Beta(2,9)	2.2667	1.8228	1.6577	3.6310	1.6205	2.1869	3.7731	2.2719	3.5800
CHI(1)	2.2393	1.9287	1.6489	2.6284	2.0482	2.1761	3.0100	2.2441	2.2722
Pareto(1,3)	1.3305	0.7236	1.2599	1.4072	0.6997	1.3682	1.4754	0.5483	1.4242
Inverse Gaussian (2,1)	1.5056	0.8696	1.3849	1.6482	0.8233	1.5744	1.7718	0.7247	1.6758

Table 3: Wasted sampling units in RSS, ERSS and FRSS and their ratio

Set Size	Wasted sampling units in RSS or ERSS	Wasted sampling units in FRSS	Ratio FRSS/RSS
3	6	3	0.5
4	12	4	0.333
5	20	10	0.5
6	30	12	0.4
7	42	21	0.5

number of discard sampling units in ERSS and RSS is $n^2 - n$, while in FRSS it is only $[(n+1)/2] \times n - n$, where $[\cdot]$ is the integer operator. Table 3 shows the number of wasted sampling units in RSS, ERSS and FRSS as well as their ratio.

The results from the table above indicate that we reduce the number of wasted measurements by at least a half $(1 - \text{Ratio})$ by using FRSS.

4. Concluding Remarks

In this paper a new ranked set sampling scheme based on the idea of folded range is introduced for estimating the population mean when the underlying distribution is asymmetric. The FRSS is compared with SRS, RSS and ERSS for different distributions. There are several advantages of the proposed sampling scheme, the first is that the proposed FRSS is more efficient than the SRS in estimating the population mean. Secondly, as given in Table 1, when the well-known ERSS gives biased and inefficient estimates; the proposed FRSS gives unbiased and efficient estimates. Moreover, FRSS is a cost-effective sampling scheme that has an advantage over the RSS as well as ERSS in reducing the wasted sampling units by at least 50% thus reducing sampling costs.

Acknowledgement

References

Al-Nasser, A. (2007). L ranked set sampling: A generalization procedure for robust visual sampling, *Communications in Statistics*, **36**, 33–43.

AL-Nasser, A. and Bani-Mustafa, A. S. (2009). Robust extreme ranked set sampling, *Journal of Statistical Computation and Simulation*, **79**, 859–867.

Balakrishnan, N. and Cohen, A. (1991). *Order Statistics and Inference: Estimation Methods*, Academic Press.

Chen, Z., Bai, Z. and Sinha, B. (2004). *Ranked Set Sampling: Theory and Applications*, Springer, New York.

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

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

Muttlak, H. (2003). Modified ranked set sampling methods, *Pakistan Journal of Statistics*, **19**, 315–323.

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

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