International Journal of Computational and Theoretical Statistics
ISSN (2210-1519)
Int. J. Comp. Theo. Stat. 3, No. 1 (May-2016)

New Nonparametric Ranked Set Sampling

Dinesh S. Bhoj

Department of Mathematical Sciences, Rutgers University, Camden, USA

Received 7 Mar. 2016 Revised 27 Mar. 2016. Accepted 5 Apr. 2016, Published 1 May 2016

Abstract: A new nonparametric ranked set sampling (NNRSS) procedure is proposed to estimate the mean of a population. The estimator based on NNRSS is compared with the estimators based on ranked set sampling (RSS) and median ranked set sampling (MRSS) procedures. It is shown that the relative precisions of the estimators based on NNRSS are higher than those based on RSS and MRSS schemes for unimodal symmetric distributions around the mean and the distributions with low skewness.

Keywords: Laplace distribution; Logistic distribution; Mean square error; Median ranked set sampling; Normal distribution; Ranked set sampling; Relative precision; Skew distributions; Symmetric distribution.

1. INTRODUCTION

The method of ranked set sampling (RSS) was introduced by McIntyre (1952) to estimate mean pasture yields with greater efficiency than simple random sampling (SRS). The RSS is a cost-efficient alternative to SRS for these situations where small sets of sampling units can be ranked easily by visual inspection or other gauging methods without actually quantifying the units. McIntyre (1952) indicated that the use of RSS is superior to SRS procedure to estimate the population mean. However, Dell and Clutter (1972) and Takahasi and Wakimoto provided mathematical foundation for RSS. They showed that the sample mean of RSS is an unbiased estimator for the population mean with smaller variance than the sample mean based on SRS with the same number of observations. Dell and Clutter also showed that the estimator of population mean based on RSS is at least as efficient as the SRS estimator even when there are ranking errors.

The procedure of selection of RSS involves drawing *n* random samples with *n* units in each sample. The *n* units in each sample are ranked with respect to a variable of interest without actually measuring them. Then the unit with the lowest rank is measured from the first sample, the unit with second lowest rank is measured from the second sample, and this procedure is continued until the unit with the highest rank is measured from the n^{th} sample. The n^2 ordered observations in the *n* samples can be displayed as:

$$
\begin{array}{lll}
x_{(11)} & x_{(12)} & \ldots\ldots x_{(1n)} \\
x_{(21)} & x_{(22)} & \ldots\ldots x_{(2n)} \\
\vdots & & \\
x_{(n1)} & x_{(22)} & \ldots\ldots x_{(nn)}
\end{array}
$$

The observations $x_{(11)}, x_{(22)}, \ldots, x_{(nn)}$ are only accurately measured and they constitute the RSS data. If *n* is small, the cycle may be repeated *r* times to increase the sample size. For convenience, we assume that r=1.

2. VARIATIONS OF RANKED SET SAMPLING:

There are various modifications of ranked set sampling to get better estimator for the population mean. One of the popular variations is to use the median ranked set sampling (MRSS) (see Bhoj 1997b, Muttlak 1997, Öztürk and Wolf, 2000). In the MRSS procedure, we use the n^2 ranked observations as in RSS. However, we measure the observation

E-mail address: dbhoj@camden.rutgers.edu

with rank $(n+1)/2$ from each sample when n is odd. If $n=2m$ is even, we measure the m^{th} order statistics from the first m samples and the $(m+1)^{th}$ order statistics from the last m samples. In recent years, the investigators have considered the varied set size ranked set sampling and ranked set sampling with random sub samples (see Samawi 2011; Amiri, Modarres and Bhoj 2015).

The other variation of RSS is its use in parametric setting, where the distribution is known. Several investigators (Bhoj and Ahsanullah (1996), Bhoj (1997a, 1997b, 1997c, 2000); Lam, Sinha and Wu (1994, 1995), Stokes (1995)) considered the distributions belong to the family of random variables with cumulative distribution function (CDF) of the form $F((x-\mu)/\sigma)$, where μ and σ are, in general, the location and scale parameters. Minimum variance linear unbiased estimators of the location and scale parameters were obtained by the above authors by using RSS for various distributions. Most of the above authors also suggested modified ranked set sampling to derive better estimators for the two parameters.

3. NEW NONPARAMETRIC RANKED SET SAMPLING:

In this paper, we will concentrate on new ranked set sampling (NRSS) proposed by Bhoj (1997c, 2000). Bhoj (1997c) proposed NRSS to estimate mean and standard deviation of the distribution when the sample size is even, while Bhoj (2000) proposed NRSS to estimate the mean of the distribution when n is odd. First we assume that $n=2m$ is even. The NRSS procedure involves drawing *m* random samples of size *2n* from the population. The *2n* members from each sample are ranked among themselves by visual inspection or other inexpensive method. The n^2 ordered observations can be displayed as:

$x_{(11)}$ $x_{(12)}$ …..$x_{(1\,2n)}$

$x_{(21)}$ $x_{(22)}$ …..$x_{(2\,2n)}$

.
.
.

$x_{(m1)}$ $x_{(m2)}$ …..$x_{(m\,2n)}$

From each sample we select appropriate j^{th} and k^{th} order statistics. The choices of these two order statistics depend on the distribution under consideration and the parameter to be estimated. The n measured observations $(x_{(ij)}, x_{(ik)})$ $i=1,2,\dots,m$ constitute NRSS sample. We note that $x_{(ij)}$ and $x_{(ik)}$ are not independently distributed. However, $(x_{(ij)}, x_{(ik)})$ and $(x_{(i'j)}, x_{(i'k)})$ for $i \neq i'$ are independently distributed.

Now we consider the variations of NRSS which can be used for odd values of $n=2m^*+1$. The NRSS procedure involves drawing $m^*=(n-1)/2$ samples, each of size *2n*. In addition, we draw $(m^*+1)^{th}$ sample of size *n*. The members of each sample are ranked among themselves by visual inspection or other inexpensive methods. The n^2 ranked set sampling units in NRSS scheme can be displayed as:

$x_{(11)}$ $x_{(12)}$ …..$x_{(12n)}$

$x_{(21)}$ $x_{(22)}$ …..$x_{(22n)}$

.
.
.

$x_{(m^*2)}$ …..$x_{(m^*2n)} x_{(m^*1)}$

$x_{(m^*+1\,1)}$ $x_{(m^*+1\,2)}$ …..$x_{(m^*+1\,n)}$

We select the j^{th} and k^{th} order statistics from each of the m^* samples and p^{th} order statistic from the last sample. Thus $(x_{(ij)}, x_{(ik)})$ $i=1,2,\dots,m^*$ and $x_{(m*+1\,p)}$ would constitute the NRSS sample. The choices of the order statistics depend on the distribution under investigation.

In practice, the distribution of the variable under consideration is unknown. Therefore, the estimation of the population mean under the nonparametric framework is very important. In this section we propose a new nonparametric ranked set sampling scheme to estimate the population mean. We always select n^{th} and $(n+1)^{th}$ order statistics from each ordered sample of size $2n$ when n is even. When n is odd we select n^{th} and $(n+1)^{th}$ order statistics from the first m^* samples of size $2n$, and $((n+1)/2)^{th}$ order statistics from last sample of size n. We measure n observations in this new nonparametric ranked set sampling (NNRSS) from the n^2 ordered observations as in the case of RSS and MRSS procedures. Therefore, the comparison of NNRSS scheme is meaningful with RSS and MRSS procedures.

4. ESTIMATION OF THE POPULATION MEAN:

McIntyre (1952) proposed the estimator for population mean, μ, based on RSS as

$$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_{(ii)}.$$

This is an unbiased estimator of μ with the property that $Var(\bar{\mu}) < Var(\bar{x})$, where $\bar{x}$ is the sample mean based on simple random sample of size n . The estimator, μ^*, based on MRSS, defined in Section 2, is:

$$\mu^* = \frac{1}{n}\left[\sum_{i=1}^{m} x_{(i\,m)} + \sum_{i=m+1}^{n} x_{(im+1)}\right] \text{ for even } n$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_{(ik)} \text{, where } k = (n+1)/2 \text{ for odd } n .$$

μ^* is an unbiased estimator for μ when the distribution is symmetric around μ.

Now we propose the estimator for μ based on NNRSS procedure as defined in Section 3.

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{m}(x_{(in)} + x_{(i\,n+1)}) \text{ for even } n$$

$$= \frac{1}{n}\left[\sum_{i=1}^{m^*}(x_{(in)} + x_{(i\,n+1)}) + x_{(m*+1\,k)}\right] \text{ for odd } n\text{, where } k = (n+1)/2.$$

$\hat{\mu}$ is also an unbiased estimator for μ when the distribution is symmetric around μ. We note that both μ^* and $\hat{\mu}$ are biased estimators for μ when the distribution is skew.

5. COMPARISON OF ESTIMATORS:

In this section, we compare the three estimators for μ based on RSS, MRSS and NNRSS procedures. For this purpose, we define the following two nonparametric relative precisions (NRPs):

$$NRP_1 = \frac{Var(\bar{\mu})}{Var(\hat{\mu})} \quad \text{, if } \hat{\mu} \text{ is an unbiased estimator,}$$

$$= \frac{Var(\bar{\mu})}{MSE(\hat{\mu})} \quad \text{, if } \hat{\mu} \text{ is a biased estimator,}$$

$$NRP_2 = \frac{Var(\bar{\mu})}{Var(\mu^*)} \quad \text{, if } \mu^* \text{ is an unbiased estimator,}$$

$$= \frac{Var(\bar{\mu})}{MSE(\mu^*)} \quad \text{, if } \mu^* \text{ is a biased estimator,}$$

where $MSE = Variance + (bias)^2$.

The values of these relative precisions, the variances of $\bar{\mu}$, μ^* and $\hat{\mu}$ and the $(bias)^2$ of $\mu^*$and $\hat{\mu}$ are presented in Table 1 for eleven distributions. We have included three symmetric distributions and eight distributions with low to moderately large skewness. For these eleven distributions the means, variances and covariances of the order statistics are readily available in the literatures; see Balakrishnan and Chen (1997) and Harter and Balakrishnan (1996). Table 1 shows that $Var(\bar{\mu}) \geq Var(\mu^*) > Var(\hat{\mu})$ for all distributions and the three sample sizes. $Var(\bar{\mu}) = Var(\mu^*)$ only for $= 2$, since RSS and MRSS procedures are the same. We note that the values of NRP_1 increases with n for unimodal symmetric distributions around μ and distributions with low skewness such as Gamma (5) and Weibull (4). However, this property may not hold for distributions with relative larger skewness. The relative precisions for such distributions decrease slightly when n increases from three to four.

Our estimator based on NNRSS procedure is better than the estimator based on MRSS for all distributions and sample sizes except for Gamma (3) and sample size 4. We therefore tried Log Gamma (3) and found that the estimator $\hat{\mu}$ is better than μ^* for all three sample sizes. In the case of skew distribution, the bias in $\hat{\mu}$ increases as n increases and therefore NRP_1 decreases when n increases from three to four. We note that the bias in μ^* remains the same as n increases from three to four for all eight skew distributions.

Table 1. Variances, biases and relative precisions of the estimators.

		RSS	*MRSS*		*NNRSS*		NRP_1	NRP_2
Distribution	*n*	*Variance*	*Variance*	*(Bias)*^{2}	*Variance*	*(Bias)*^{2}		
Normal	2	0.34084	0.34084	0.00000	0.29820	0.00000	1.143	1.000
	3	0.17418	0.14956	0.00000	0.14529	0.00000	1.199	1.165
	4	0.10652	0.09012	0.00000	0.0841	0.00000	1.267	1.182
Logistic	2	0.34802	0.34802	0.00000	0.25991	0.00000	1.339	1.000
	3	0.18135	0.13069	0.00000	0.12334	0.00000	1.470	1.388
	4	0.11279	0.07902	0.00000	0.06889	0.00000	1.637	1.427
Laplace	2	0.71875	0.71875	0.00000	0.4201	0.00000	1.711	1.000
	3	0.38544	0.21296	0.00000	0.18696	0.00000	2.062	1.810
	4	0.24531	0.13018	0.00000	0.09366	0.00000	2.619	1.884
Weibull (2)	2	0.07361	0.07361	0.00000	0.06608	0.00088	1.100	1.000
	3	0.03810	0.03316	0.00088	0.03081	0.00088	1.202	1.119
	4	0.02308	0.02002	0.00088	0.01897	0.00180	1.111	1.104
Weibull (4)	2	0.02193	0.02193	0.00000	0.02003	0.00001	1.095	1.000
	3	0.01115	0.01004	0.00001	0.00982	0.00001	1.134	1.110
	4	0.00680	0.00605	0.00001	0.00573	0.00002	1.182	1.122
Gamma (3)	2	1.06055	1.06055	0.00000	0.64944	0.03196	1.556	1.000
	3	0.55522	0.41765	0.03196	0.39881	0.04581	1.249	1.235
	4	0.34571	0.25398	0.03196	0.22647	0.06589	1.182	1.209
Gamma (5)	2	1.74298	1.74298	0.00000	1.42426	0.03272	1.196	1.000
	3	0.90419	0.71648	0.03272	0.68907	0.04685	1.229	1.207
	4	0.55923	0.43407	0.03272	0.39444	0.05486	1.245	1.198
Inverse Gaussian (.5)	2	0.34354	0.34354	0.00000	0.29240	0.00204	1.167	1.000
	3	0.17570	0.14777	0.00204	0.09309	0.00292	1.830	1.173
	4	0.10842	0.08861	0.00204	0.08181	0.00419	1.261	1.196
Inverse Gaussian (1)	2	0.35087	0.35087	0.00000	0.27653	0.00748	1.235	1.000
	3	0.18295	0.13923	0.00748	0.08665	0.01068	1.880	1.247
	4	0.11363	0.08447	0.00748	0.07559	0.01532	1.250	1.236
Log Gamma (3)	2	0.13643	0.13643	0.00000	0.11280	0.0.0014	1.197	1.000
	3	0.07042	0.05667	0.00114	0.05457	0.00163	1.253	1.218
	4	0.04340	0.03424	0.00114	0.03126	0.00232	1.292	1.227
Extreme Value	2	0.58134	0.58134	0.00000	0.43544	0.01387	1.294	1.000
	3	0.30577	0.21951	0.01387	0.20782	0.01970	1.344	1.310
	4	0.19102	0.13342	0.01387	0.11671	0.02806	1.319	1.297

6. CONCLUSIONS AND DISCUSSION:

In this paper, we proposed nonparametric ranked set sampling (NNRSS) procedure for the symmetric and low to moderate skew distributions. The estimator for population mean is suggested based on NNRSS procedure. This estimator is then compared with the estimators based on ranked set sampling (RSS) and median ranked set sampling (MRSS) procedures. We computed the relative precisions of the estimators for eleven distributions and sample sizes 2, 3 and 4. The computations given in Table 1 show that estimator based on NNRSS is uniformly superior to the estimator based on RSS and the gains in precisions are substantial. The relative precisions of the estimators based on NNRSS are higher than the estimator based on MRSS for symmetric distributions and the distributions with lower skewness for all three sample sizes. The gains in precisions are very good for *n*=2 and 3. However, the gains in precisions may be marginal for *n*=4 for the skew distributions. As skewness of the distribution increases the relative precision of $\hat{\mu}$ over μ^* might get smaller than one for *n*=4.

We note that NRP_1 increases as skewness increases for *n*=2 and 3, and decreases for *n*=4. In the case of extreme skew distributions NNRSS is better than MRSS for *n*=2. For example we calculated NRP_1 for Exponential, Log Normal and Pareto (5) distributions. The values of NRP_1 for these distributions for *n*=2 are, respectively, 1.421, 2.690

and 2.151. These gains in relative precisions are dramatic. The estimator based on NNRSS is better than the one based on RSS even for n=3 and 4. However, when $n \geq 3$, the estimator based on NNRSS is slightly inferior to the estimator based on MRSS procedure for these highly skew distributions.

Based on the numerical computations of relative precisions, we recommend the NNRSS procedure for n=2, 3 and 4 when the samples are drawn from symmetric distributions or distributions with low skewness. For moderate skew distributions, NNRSS may be recommended for n=2 and 3. Most importantly for extreme skew distributions we can recommend NNRSS procedure for n=2. Since we are selecting m^* (for odd n) and m (for even n) samples of size $2n$ the use of NNRSS procedure is recommended only for small values of n in order to minimize the error due to ranking. In order to increase the sample size, the procedure may be repeated $r \geq 2$ times.

ACKNOWLEDGMENT

The author is thankful to the editor and the referee for their suggestions.

REFERENCES

[1] Amiri, S., Modarres, R. and Bhoj, D.S. (2015). Ranked set sampling with random subsamples. Journal of Statistical Computation and Simulation 85, 935-946.

[2] Balakrishnan, N. and Chen, W.S. (1997). CRC Handbook of Tables for Order Statistics from Inverse Gaussian Distributions with Applications, CRC Press, Boca Raton, New York.

[3] Bhoj, D.S. (2001). Ranked set sampling with unequal samples. Biometrics 57, 957-962.

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

[5] Bhoj, D.S. (2000). Estimation of mean and standard deviation of Laplace distribution using ranked set sampling. Advances on Methodological and Applied Aspects and Probability and Statistics. Gordon and Breach Publishers, Newark, New Jersey, N. Balakrishnan editor, 169-182.

[6] Bhoj, D.S. (1997a). Estimation of parameters of the extreme value distribution using ranked set sampling. Communications in Statistics – Theory and Methods 26, 653-667.

[7] Bhoj, D.S. (1997b). Estimation of parameters using modified ranked set sampling. In Applied Statistical Science, Volume II, M. Ahsanullah (ed)., 145-163. New York: Nova Science.

[8] Bhoj, D.S. (1997c). New parametric ranked set sampling. Journal of Applied Statistical Science 6, 275-289.

[9] Bhoj, D.S. and Ahsanullah, M. (1996). Estimation of parameters of the generalized geometric distribution using ranked set sampling. Biometrics 52, 685-694.

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

[11] Lam, K., Sinha, B.K. and Zhong, W. (1994). Estimation of parameters in the two-parameter exponential distribution using ranked set sample. Annals of the Institute Statistical Mathematics 46, 723-736.

[12] Lam, K., Sinha, B.K. and Zhong, W. (1995). Estimation of location and scale parameters of a logistic distribution using a ranked set sample. In Collected Essays in Honor of Professor Hubert A. David, H.N. Nagaraja, P.K. Sen, and D.F. Morrison (eds), 189-197. New York: Springer.

[13] McIntyre, G.A. (1952). A method of unbiased selective sampling, using ranked sets. Australian Journal of Agriculture Research 3, 385-390.

[14] Muttlak, H.A. (1997). Median ranked set sampling. Journal of Applied Statistical Science 6, 245-255.

[15] Özturk, O. and Wolfe, D.A. (2000). Alternative ranked set sampling protocols for the sign test. Statistics and Probability Letters 47, 15-23.

[16] Samawi, H.M. (2011) Varied set size ranked set sampling with applications to mean and ratio estimation. International Journal of Modelling and Simulation 31, 6-13.

[17] Stokes, S.L. (1995). Parametric ranked set sampling. Annals of the Institute of Statistical Mathematics 47, 465-482.

[18] 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.