Pak J. Statist.
2003 Vol. 19(3) pp 315 – 323

MODIFIED RANKED SET SAMPLING METHODS

Hassen A. Muttlak
Department Mathematical Sciences
KFUPM, Dhahran, 31261, Saudi Arabia

ABSTRACT

The ranked set sampling method (RSS) as suggested by McIntyre (1952) may be modified to yield new sampling methods with improved. Several modifications for the RSS are introduced by several authors such as extreme ranked set sampling (ERSS), suggested by Samawi et al. (1996), median ranked set sampling (MRSS), suggested by Muttlak (1997), etc. In this study a few other modifications for the RSS are introduced and compared to the RSS, ERSS and MRSS. It turns out that for probability distributions considered in this study, we can always improve upon the efficiency of RSS by using some sort of modification for the RSS method.

KEY WORDS

Extreme ranked set sampling, median ranked set sampling, simple random sampling, percentile ranked sampling and relative precision.

1. INTRODUCTION

Ranked set sampling (RSS) was first suggested by McIntyre (1952) without the mathematical theory to support his suggestion. Takahasi and Wakimoto (1968) supplied the necessary mathematical theory. They proved that the sample mean of the ranked set sample (RSS) is an unbiased estimator of the population mean with smaller variance than the sample mean of a simple random sample (SRS) with the same sample size. Dell and Clutter (1972) studied the case in which the ranking may not be perfect i.e., there are errors in ranking the units. Muttlak (1996) suggested using pair ranked set sampling instead of RSS. This can be used when it is difficult to select a large number of units from the population of interest. Samawi et al. (1996) suggested using extreme ranked set sampling (ERSS) to estimate the population mean. They showed that the ERSS estimator is an unbiased estimator of the population mean if the underlying distribution is symmetric and it is more efficient than the SRS estimator. Muttlak (1997) suggested using median ranked set sampling (MRSS) to estimate the population mean more efficiently than the usual RSS method. For review and more bibliography on the RSS see Patil et al. (1999).

In this paper, a further modification of the RSS method is considered, namely, percentile ranked sampling (PRSS) with different values of $0 \le p \le 1$. The newly suggested sampling method is compared with RSS, ERSS and MRSS. It is shown that for the probability distributions considered in this study, we can always improve the relative precision and reduce the errors in ranking by using the modified sampling method instead of the usual RSS method.

2. NOTIONS AND SOME USEFUL RESULTS

Let $X_1, X_2, ..., X_n$ be a random sample with probability density function f (x) with a finite mean μ and variance σ^2. Let $X_{11}, X_{12}, ..., X_{1n}; X_{21}, X_{22}, ..., X_{2n}; ...; X_{n1}, X_{n2}, ..., X_{nn}$ be independent random variables all with the same cumulative distribution function F(x). Let $X_{(i:n)}$ denotes the i^{th} order statistic from the i^{th} sample of size n (i = 1, 2, ..., n). The unbiased estimator of the population mean using RSS is defined as

$$\overline{X}_{rss} = \frac{1}{n}\sum_{i=1}^{n} X_{(i:n)} .$$

The variance of $\overline{X}_{rss}$ is given by

$$\operatorname{var}(\overline{X}_{rss}) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2_{(i:n)} ,$$

where $\sigma^2_{(i:n)} = E[X_{(i:n)} - E(X_{(i:n)})]^2$.

Let $X_{(i:e)}$, denote the smallest of the i^{th} sample (i = 1, 2, ..., L= n/2) and the largest of the i^{th} sample (i = L+1, L+2, ..., n) if the sample size n is even. Also denote the smallest of the i^{th} sample (i = 1, 2, ..., L_1= (n-1)/2), the median of the i^{th} sample (i = (n+1)/2) and the largest of the i^{th} sample (i = L_1+2, L_1+3, ..., n) if the sample size n is odd. The estimator of the population mean based on ERSS with one cycle can be written as

$$\overline{X}_{errs} = \frac{1}{n}\sum_{i=1}^{n} X_{(i:e)} .$$

The variance of $\overline{X}_{errs}$ can be written as

$$\operatorname{var}(\overline{X}_{errs}) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2_{(i:e)} ,$$

where $\sigma^2_{(i:n)} = E[X_{(i:e)} - E(X_{(i:e)})]^2$. For more details, see Samawi et al (1996).

Let $X_{(i:m)}$, denote the median of the i^{th} sample if the sample size is odd, and the $(n/2)^{th}$ order statistic of the i^{th} sample (i = 1, 2, ..., L= n/2) and the $((n+2)/2)^{th}$ order statistic of the i^{th} sample (i = L+1, L+2, ..., n) if the sample size is even. The estimator of the population mean using MRSS then can be written as

$$\overline{X}_{mrss} = \frac{1}{n}\sum_{i=1}^{n} X_{(i:m)} .$$

The variance of $\overline{X}_{mrss}$ can be written as

$$\mathrm{var}(\overline{X}_{mrss}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2_{(i:m)}$$

where $\sigma^2_{(i:m)} = E[X_{(i:m)} - E(X_{(i:m)})]^2$. For more details, see Muttlak (1997).

3. PERCENTILE RANKED SET SAMPLING

In the percentile ranked set sampling (PRSS) procedure, select n random samples of size n units from the population and rank the units within each sample with respect to a variable of interest. If the sample size is even, select for measurement from the first n/2 samples the $(p(n+1))^{th}$ smallest rank and from the second n/2 samples the $(q(n+1))^{th}$ smallest rank, where $0 \le p \le 1$ and q = 1-p. If the sample size is odd, select from the first (n-1)/2 samples the $(p(n+1))^{th}$ smallest rank and from the other (n-1)/2 samples the $(q(n+1))^{th}$ smallest rank, and from one sample the median for that sample for actual measurement. The cycle may be repeated r times to get nr units. These nr units form the PRSS data.

Let $X_{11}, X_{12}, ..., X_{1n}; X_{21}, X_{22}, ..., X_{2n}; ...; X_{n1}, X_{n2}, ..., X_{nn}$ be independent random variables all with the same cumulative distribution function F(x). Let $X_{i\,(p(n+1))}$ and $X_{i\,(q(n+1))}$ denote the $(p(n+1))^{th}$ order statistic $(q(n+1))^{th}$ order statistic of the i^{th} sample respectively (i = 1, 2, ..., n), where $0 \le p \le 1$ and q = 1-p. The estimator of the population mean using percentile ranked set sample (PRSS) with one cycle can be defined in the case of an even sample size as

$$\overline{X}_{prss1} = \frac{1}{n}\left(\sum_{i=1}^{L_1} X_{i(p(n+1))} + \sum_{i=L_1+1}^{n} X_{i(q(n+1))}\right),$$

where $L_1 = n/2$. In the case of an odd sample size, the estimator of the population mean can be defined as

$$\overline{X}_{prss2} = \frac{1}{n}\left(\sum_{i=1}^{L_2} X_{i(p(n+1))} + \sum_{i=L_2+2}^{n} X_{i(q(n+1))} + X_{i((n+1)/2)}\right),$$

where $L_2 = (n-1)/2$ and $X_{i\,((n+1)/2)}$ is the median of sample i = (n+1)/2

The variance of $\overline{X}_{prss}$ can be written as

$$\mathrm{var}(\overline{X}_{prss}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2_{(i:p)}$$

where $s^2_{(i:p)} = E[X_{(i:p)} - E(X_{(i:p)})]^2$. Here $X_{(i:p)}$ is the $p_{(n+1)}^{th}$ order statistic of the i^{th} sample.

Let $\overline{X}_{srs}$ denote the sample mean of simple random sample (SRS) of size n. The properties of $\overline{X}_{prss}$ are

1. $\overline{X}_{prss}$ is an unbiased estimator of the population mean μ if the underling distribution is symmetric about the population μ and
2. Var ($\overline{X}_{prss}$) is less than Var ($\overline{X}_{srs}$).
3. If the distribution is not symmetric about μ than the mean square error (MSE) of $\overline{X}_{prss}$ is less than the variance of $\overline{X}_{srs}$.

It is not difficult to prove (1)-(3) using the results by Takahasi and Wakimoto (1968), Samawi et al (1996) and Muttlak (1997).

To compare the proposed estimators for the population mean using PRSS with RSS, ERSS, MRSS and SRS methods, eight probability distribution functions were considered: rectangular, normal, exponential, gamma, weibull, double exponential, inverse Gaussian and lognormal. The variance or the mean square error of the sample means for the RSS, ERSS, MRSS and PRSS with different values of p were calculated for the above distributions using the moments of the order statistics, see Harter and Balakrishnan (1996) and Balakrishnan and Chen (1997). The relative precision (RP) of estimating the population mean using any of the RSS based methods with respect to the usual estimator using SRS is defined as following

$$RP(\overline{X}_{srs}, \overline{X}_{rss}) = \frac{Var(\overline{X}_{srs})}{Var(\overline{X}_{rss})},$$

if the distribution is symmetric and

$$RP(\overline{X}_{srs}, \overline{X}_{rss}) = \frac{Var(\overline{X}_{srs})}{MSE(\overline{X}_{rss})},$$

if the distribution is not symmetric.

Results are summarized by the relative precision (RP) and the bias in Tables I-III for RSS ERSS, MRSS and PRSS with p = 20%, 30% and 40%. For each population calculations were done with sample size n = 8 in Table I, n =9 in Table II and n=10 inn Table III. Considering the results in Tables I-III, a gain in efficiency is obtained by using PRSS for different values of n and for all the distributions considered in this study. For example, for n = 10 in Table III and p = 0.3 the relative precision (RP) of the PRSS is

5.329 for estimating the population mean of a weibull distribution with shape parameter 2.5.

4. PERCENTILE RANKED SET SAMPLING WITH ERRORS IN RANKING

Dell and Clutter (1972) considered the case in which there are errors in ranking; that is, the quantified observation from the i^{th} sample in the j^{th} cycle may be not the i^{th} order statistic but rather the i^{th} judgment order statistic. They showed that the sample mean of RSS with errors in ranking is an unbiased estimator of the population mean μ, regardless of the errors in ranking and has a smaller variance than the usual estimator based on SRS with the same sample size.

Let $X_{i\,[p(n+1)]}$ and $X_{i\,[q(n+1)]}$ denote the $[p(n+1)]^{th}$ and $[q(n+1)]^{th}$ judgment order statistics respectively, of the i^{th} sample (i = 1, 2, ..., n), where $0 \le p \le 1$ and $q = 1-p$. If the cycle is repeated once, the estimator of the population mean using percentile ranked set sample (PRSS) with errors in ranking can be defined in the case of an even sample size as

$$\tilde{X}_{prsse1} = \frac{1}{n}\left(\sum_{i=1}^{L_1} X_{i[p(n+1)]} + \sum_{i=L_1+1}^{n} X_{i[q(n+1)]} \right),$$

where $L_1 = n/2$. In the case of an odd sample size, the estimator of the population mean can be defined as

$$\tilde{X}_{prsse2} = \frac{1}{n}\left(\sum_{i=1}^{L_2} X_{i[p(n+1)]} + \sum_{i=L_2+2}^{n} X_{i[q(n+1)]} + X_{i[(n+1)/2]} \right),$$

where $L_2 = (n-1)/2$ and $X_{i\,[(n+1)/2]}$ is the judgment median of sample i = (n+1)/2.
The variance of $\tilde{X}_{prsse}$ can be written as

$$\operatorname{var}(\tilde{X}_{prsse}) = \frac{1}{n^2}\sum_{i=1}^{n} s^2_{[i:p]}$$

where $s^2_{[i:p]} = E[X_{[i:p]} - E(X_{[i:p]})]^2$. Here $X_{[i:p]}$ is the $p_{(n+1)}{}^{th}$ judgment order statistic of the i^{th} sample.

Let $\overline{X}_{srs}$ denote the sample means of simple random sample (SRS) of size n. The properties of $\tilde{X}_{prsse}$ are

1. $\tilde{X}_{prsse}$ is an unbiased estimator of the population mean μ if the underling distribution is symmetric about μ and
2. Var ($\tilde{X}_{prsse}$) is less than Var ($\overline{X}_{srs}$).

3. If the distribution is not symmetric about μ than the mean square error (MSE) of $\tilde{X}_{prsse}$ is less than the variance of $\overline{X}_{srs}$.

It is not difficult to prove a-c, using the results by Takahasi and Wakimoto (1968), Dell and Clutter (1972) Samawi et al (1996) and Muttlak (1997).

5. ACKNOWLEDGMENTS

The author is indebted to King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia for excellent research facilities. I am thankful for the referee for healthy comments.

REFERENCES

1. Balakrishnan, N., and Chen, W.W. S. (1997). *CRC Handbook of Tables for Order Statistics from Inverse Gussian Distributions with Aplications*, CRC Press, Boca Raton.

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

3. Harter, H. L. and Balakrishnan, N. (1996). *CRC Handbook of Tables for the use of Order Statistics in Estimation*, CRC Press, Boca Raton.

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

5. Muttlak, H. A. (1996). Pair ranked set sampling, *the Biometrical Journal*, ***38***, 879-885.

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

7. Patil, G. P., Sinha, A. K., and Taillie, C (1999). Ranked set sampling: A bibliography,

8. *Environmental and Ecological Statistics*, 6, 91-98.

9. Samawi, H, Abu-Dayyeh, W and Ahmed, S (1996). Extreme ranked set sampling. *The Biometrical Journal* **30**, 577-586.

10. 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.

Table I. Summary of the relative precision (RP) values for estimating the population mean using RSS, ERSS, MRSS, and PRSS with values of p = 0.2, 0.3 and 0.4, with sample size n = 8.

Distribution	RSS	ERSS	MRSS	PRSS	20%	30%	40%
Uniform (0, 1)	RP	4.50	8.348	3.375	4.821	3.750	3.375
	Bias						
Normal (0, 1)	RP	3.999	2.682	5.342	4.177	4.981	5.342
	Bias						
Exponential (1)	RP	2.943	0.457	1.673	3.545	2.426	1.673
	Bias		0.421	0.241	0.007	0.174	0.241
Gamma (2)	RP	3.354	0.725	2.404	3.812	3.126	2.404
	Bias		0.345	0.253	0.009	0.184	0.253
Gamma (3)	RP	3.535	0.939	2.903	3.921	3.533	2.903
	Bias		0.453	0.257	0.009	0.182	0.257
Gamma (5)	RP	3.702	1.253	3.524	4.018	4.044	3.524
	Bias		0.459	0.259	0.011	0.187	0.259
Lognormal (0, 1)	RP	1.891	0.279	1.814	4.068	2.586	1.814
	Bias		1.037	0.538	0.083	0.416	0.538
Double Exponential (0, 1)	RP	3.124	1.309	9.509	3.768	6.863	9.509
	Bias						
Inverse Gaussian (0.5)	RP	3.885	1.951	4.633	4.124	4.641	4.633
	Bias		0.115	0.065	0.003	0.047	0.065
Inverse Gaussian (1)	RP	3.603	1.127	3.432	3.989	3.932	3.432
	Bias		0.219	0.124	0.006	0.090	0.124
Inverse Gaussian (1.5)	RP	3.262	0.715	2.563	3.821	3.276	2.563
	Bias		0.308	0.173	0.008	0.127	0.173
Inverse Gaussian (2.5)	RP	2.657	0.404	1.747	3.546	2.495	1.747
	Bias		0.434	0.241	0.015	0.178	0.241
Weibull (0.5)	RP	1.665	0.222	1.478	3.750	2.164	1.478
	Bias		2.472	1.272	0.207	0.993	1.272
Weibull (1.5)	RP	3.647	0.972	2.718	3.934	3.370	2.718
	Bias		0.163	0.094	0.002	0.067	0.094
Weibull (2)	RP	3.962	1.750	3.788	4.122	4.092	3.788
	Bias		0.075	0.042	0.002	0.031	0.042
Weibull(2.5)	RP	4.088	2.534	4.524	4.187	4.502	4.524
	Bias		0.035	.0190	0.001	0.014	.0190

Table II: Summary of the relative precision (RP) values for estimating the population mean using RSS, ERSS, MRSS, and PRSS with values of p = 0.2, 0.3 and 0.4, with sample size n = 9.

Distribution	RSS	ERSS	MRSS	PRSS	20%	30%	40%
Uniform (0, 1)	RP	5.0	10.19	3.667	5.729	4.365	3.819
	Bias						
Normal (0, 1)	RP	4.394	2.798	6.020	4.431	5.365	5.863
	Bias						
Exponential (1)	RP	3.181	0.484	1.432	3.770	2.577	1.636
	Bias		0.389	0.254	0.000	0.158	0.232
Gamma (2)	RP	3.650	0.771	2.166	4.102	3.344	2.409
	Bias		0.412	0.267	0.001	0.167	0.244
Gamma(3)	RP	3.858	1.003	2.708	4.241	3.795	2.949
	Bias		0.419	0.271	0.001	0.170	0.248
Gamma (5)	RP	4.052	1.347	3.439	4.362	4.292	3.665
	Bias		0.425	0.274	0.002	0.172	0.251
Lognormal (0, 1)	RP	1.980	0.288	1.528	4.199	2.635	1.719
	Bias		0.975	0.562	0.064	0.387	0.523
Double Exponential (0, 1)	RP	3.374	1.299	11.42	3.695	6.735	9.900
	Bias						
Inverse Gaussian (0.5)	RP	4.265	2.139	4.914	4.495	5.036	4.993
	Bias		0.106	0.068	0.001	0.043	0.062
Inverse Gaussian (1)	RP	3.936	1.211	3.062	4.311	4.227	3.547
	Bias		0.203	0.141	0.006	0.082	0.119
Inverse Gaussian (1.5)	RP	3.542	0.760	2.320	4.100	3.487	2.569
	Bias		0.286	0.183	0.003	0.116	0.167
Inverse Gaussian (2.5)	RP	2.850	0.736	1.494	3.739	2.615	1.692
	Bias		0.228	0.253	0.007	0.163	0.233
Weibull (0.5)	RP	1.736	0.228	1.228	3.807	2.182	1.387
	Bias		2.327	1.327	0.162	0.926	1.239
Weibull (1.5)	RP	3.992	1.039	2.530	4.244	3.642	2.770
	Bias		0.151	0.098	0.008	0.061	0.090
Weibull (2)	RP	4.406	1.891	3.855	4.504	4.457	4.039
	Bias		0.069	0.045	0.003	0.028	0.041
Weibull (2.5)	RP	4.507	2.779	4.899	4.596	4.927	4.930
	Bias		0.032	0.020	0.001	0.013	0.018

Table III: Summary of the relative precision (RP) values for estimating the population mean using RSS, ERSS, MRSS, and PRSS with values of p = 0.2, 0.3 and 0.4, with sample size n =10.

Distribution	RSS	ERSS	MRSS	PRSS	20%	30%	40%
Uniform (0, 1)	RP	5.50	12.10	4.033	6.722	5.042	4.321
	Bias						
Normal (0, 1)	RP	4.795	2.904	6.620	4.662	5.714	6.332
	Bias						
Exponential (1)	RP	3.414	0.292	1.328	2.985	3.259	1.736
	Bias		0.514	0.254	0.007	0.117	0.213
Gamma (2)	RP	3.940	0.486	2.061	3.591	4.002	2.556
	Bias		0.544	0.267	0.072	0.125	0.224
Gamma (3)	RP	4.177	0.656	2.628	3.880	4.407	3.140
	Bias		0.554	0.271	0.072	0.127	0.228
Gamma (5)	RP	4.397	0.934	3.428	4.155	4.829	3.898
	Bias		0.561	0.274	0.072	0.129	0.230
Lognormal (0, 1)	RP	2.064	0.182	1.392	3.522	3.205	1.759
	Bias		1.288	0.562	0.076	0.313	0.489
Double Exponential (0, 1)	RP	3.617	1.291	12.63	3.633	6.593	9.940
	Bias						
Inverse Gaussian (0.5)	RP	4.641	1.727	5.165	4.479	5.422	5.330
	Bias		0.140	0.068	0.018	0.032	0.057
Inverse Gaussian (1)	RP	4.263	0.828	3.267	4.055	4.774	3.758
	Bias		0.269	0.131	0.034	0.062	0.110
Inverse Gaussian (1.5)	RP	3.820	0.485	2.203	3.602	4.121	2.707
	Bias		0.377	0.183	0.047	0.087	0.154
Inverse Gaussian (2.5)	RP	3.038	0.259	1.3740	2.997	3.244	1.7720
	Bias		0.533	.254	0.060	0.125	.215
Weibull (0.5)	RP	1.807	0.142	1.110	3.099	2.672	1.408
	Bias		3.074	1.328	0.169	0.754	1.162
Weibull (1.5)	RP	4.333	0.669	2.457	3.915	4.287	2.965
	Bias		0.199	0.099	0.003	0.045	0.083
Weibull (2)	RP	4.752	1.381	3.953	4.511	4.969	4.320
	Bias		0.092	0.045	0.012	0.021	0.038
Weibull(2.5)	RP	4.922	2.379	5.245	4.754	5.329	5.313
	Bias		0.042	0.020	0.004	0.010	0.017