Sains Malaysiana 40(4)(2011): 397-401

On Comparison of Some Variation of Ranked Set Sampling (Tentang Perbandingan Beberapa Variasi Pensampelan Set Terpangkat)

KAMARULZAMAN IBRAHIM*

ABSTRACT

Many sampling methods have been suggested for estimating the population median. In the situation when the sampling units in a study can be easily ranked than quantified, the ranked set sampling methods are found to be more efficient and cost effective as compared to the simple random sampling. In this paper, the superiority of several ranked set sampling methods over the simple random sampling are illustrated through some simulation study. In addition, some new research topics under ranked set sampling are suggested.

Keywords: Ranked set sampling; simple random sampling; simulation study

ABSTRAK

Pelbagai kaedah pensampelan dicadangkan untuk menganggar median populasi. Dalam keadaan bila unit pensampelan senang untuk dipangkatkan berbanding dengan disukat, pensampelan set terpangkat didapati lebih cekap dan kos efektif berbanding dengan pensampelan rawak ringkas. Dalam kertas ini, kelebihan beberapa kaedah pensampelan set terpangkat ditunjukkan menerusi kajian simulasi. Di samping itu, beberapa tajuk baru dalam pensampelan set terpangkat dicadangkan.

Kata kunci: Kajian simulasi; pensampelan rawak ringkas; pensampelan set terpangkat

INTRODUCTION

Ranked set sampling (RSS) is a cost effective sampling procedure when compared to the commonly used simple random sampling (SRS) in the situations where visual ordering of units can be done easily but the exact measurement of the units is difficult and expensive. McIntyre (1952) proposed the sample mean based on RSS as an estimator of the population mean. He found that the estimator based on RSS is more efficient than SRS. Many modifications on RSS have been done since McIntyre (1952). Takahasi and Wakimoto (1968) provided the necessary mathematical theory of RSS. The modifications on RSS involve methods of estimation of the population mean and median which include median ranked set sampling (Muttlak 1997), extreme ranked set sampling (Samawi et al. 1996), two-stage median ranked set sampling (Jemain et al. 2007a) and balanced group ranked set sampling (Jemain et al. 2009). In addition, various modification of RSS have been suggested for the estimation of population ratio. Samawi and Muttlak (2001), for example, used median ranked set sampling (MRSS) to estimate the population ratio. Samawi and Tawalbeh (2002) introduced double median ranked set sampling (DMRSS) method for estimating the population mean and ratio. Muttlak (2003) proposed the use of quartile RSS for estimating the population mean.

In this paper, we discuss some variation of RSS, particularly regarding the efficiency of those methods when compared to SRS for estimating the population median.

The performance of the estimators is studied through a simulation study. In addition, some new research topics are suggested, focusing on the goodness of fit tests under ranked set sampling.

SAMPLING METHODS

RANKED SET SAMPLING

The RSS involves randomly selecting m^2 units from the population. These units are randomly allocated into m sets, each of size m. The m units of each sample are ranked visually or by any inexpensive method with respect to the variable of interest. From the first set of m units, the smallest unit is measured. From the second set of m units, the second smallest unit is measured. The process is continued until from the mth set of m units the largest unit is measured. Repeat the process n times to obtain a set of size mn from initial m^2n units.

MEDIAN RANKED SET SAMPLING

In median ranked set sampling (MRSS) method select m random samples each of size m units from the population and rank the units within each sample with respect to the variable of interest. If the sample size m is odd, then from each sample select for measurement the ((m+1)/2) th smallest rank (the median of the sample). If the sample size m is even, then select for measurement the (m/2)th smallest rank from the first m/2 samples, and the ((m+1)/2)

2)/2)th smallest rank from the second m/2 samples. The cycle can be repeated *n* times if needed to obtain a sample of size *nm* (Muttlak 1997).

EXTREME RANKED SET SAMPLING (ERSS)

To select a sample of size m under ERSS, first m^2 units are randomly selected from the population. In the usual RSS, we identify and measure the *i*th smallest ranked unit of the *i*th sample (i = 1, 2, ..., m). In the case when *m* is odd, for ERSS, we select the smallest ranked unit from the first (m-1)/2 sets, the largest ranked unit from the other (m-1)/2sets, and the median from the remaining set. In the case when *m* is even, we select the smallest ranked unit from the first m/2 sets and the largest ranked unit from the other m/2 sets. This cycle can be repeated *n* times to increase the sample size to *nm* units.

BALANCED GROUP RANKED SET SAMPLING (BGRSS)

The balanced groups ranked set sampling (BGRSS) can be described as follows: randomly select m = 3k (k = 1, 2, ...) sets each of size m from the target population, and rank the units within each set with respect to the variable of interest. Then allocate the 3k selected sets randomly into three groups, each of size k sets. For each group, select for measurement the lowest ranked unit from each set in the first group, and the median unit from each set in the second group, and the largest ranked unit from each set in the second group. One cycle is done after completing all these steps; thus, a sample of size m = 3k units is obtained. If the cycle is repeated n times, the sample size obtained by using BGRSS is increased to 3kn, based on $9k^2n$ units from a SRS.

TWO-STAGE RANKED SET SAMPLING (TSRSS)

The TSRSS procedure can be summarized as the followings. First, randomly select $m^3 = 27k^3$ (k = 1, 2, ...) units from the target population and divide these units randomly into $m^2 = 9k^2$ sets each of size *m*. Then, allocate these $9k^2$ sets into three groups, each of $3k^2$ sets. From each set in the first group select the smallest rank unit, from each set in the second group select the median rank unit, and from each set in the third group select the largest rank unit. This step yields k sets in each group. Finally, without doing any actual quantification, from the k sets in the first group select the smallest rank unit, from the k sets in the second group select the median rank unit, and from the k sets in the third group select the largest rank unit. This step yields one set of size m = 3k. If the procedure is repeated *n* times, a sample of size *nm* is obtained. If more than two stages are required in the sampling process, the work by Jemain et al. (2007b) on multistage median ranked set sampling may be referred.

SELECTIVE ORDER RANKED SET SAMPLING (SORSS)

Consider a random sample of size *m*. Rank the units in the sample. Select the *i*th ranked unit as the unit of interest.

Repeat this r times to obtain a selective order ranked set sample of size r. Essentially, this is a random sample found based on the ranked set sampling involving a set size m via the *i*th order statistic. Al-Subh et al. (2009) has shown that the empirical distribution function based tests for testing the logistic distribution perform better under selective order ranked set sampling (SORSS) as compared to the simple random sampling.

ESTIMATION OF THE POPULATION MEDIAN

Simple random sampling Let $X_1, X_2, ..., X_m$ be a random sample with probability density function f(x) and cumulative distribution function F(x), with finite mean μ and variance σ^1 . The estimator for the population median based on SRS, denoted as η_{RSS} is defined as:

$$\hat{\eta}_{SRS} = \begin{cases} X_{\left(\frac{m+1}{2},m\right)}, & \text{if } m \text{ is odd} \\ X_{\left(\frac{m}{2},m\right)} + X_{\left(\frac{m+2}{2},m\right)} \\ \hline 2, & \text{if } m \text{ is even} \end{cases}$$
(1)

Ranked set sampling Let $X_1, X_2, ..., X_m$ be a random sample from the distribution with probability density function f(x)and cumulative distribution function F(x) having finite mean μ and variance σ^1 . Assume that m independent simple random samples each of size m for the hth cycle be denoted by $X_{11h}, X_{12h}, ..., X_{1mh}, X_{21h}, X_{22h}, ..., X_{m1h}, X_{m2h},$ $..., X_{mmh}$ (h = 1, 2, ..., n). Let $X_{i(1:m)h}, X_{i(2:m)h}, ..., X_{i(m:m)h}$ be the order statistics of the *i*th sample (i = 1, 2, ..., m). Then $X_{1(1:m)h}, X_{2(2:m)h}, ..., X_{m(m:m)h}$ denote the measured RSS. The estimator for the population median based on RSS, denoted as η_{RSS} , is defined as:

$$\hat{\eta}_{RSS} = \text{median} \{ X_{i(i:m)h}, i = 1, 2, \dots, m; h = 1, 2, \dots, n \}.$$
 (2)

Median ranked set sampling If the sample size is odd, let $X_{i\left(\frac{m+1}{2},m\right)^{h}}$ be the median of the *i*th sample (i = 1, 2, ..., m) for the *h*th cycle, i.e. the ((m+1)/2)th order statistics, denotes the MRSSO. If the sample size is even, let $X_{i\left(\frac{m}{2},m\right)}$ be the (m/2) th order statistics of the ith sample (i = 1, 2, ..., L = m/2) and $X_{i\left(\frac{m+2}{2},m\right)}$ be the ((m+2)/2)th order statistics of the *i*th sample (i = L, L+1, ..., m) denote the MRSSE. The estimators for the population median are given by:

$$\hat{\eta}_{MRSSO} = \text{median} \left\{ X_{i\left(\frac{m+1}{2}:m\right)h}, i = 1, 2, \dots, m; h = 1, 2, \dots, n \right\},$$
(3)

and

$$\hat{\eta}_{MRSSE} = \text{median} \begin{cases} X_{i\left(\frac{m+1}{2},m\right)h}, i = 1, 2, \dots, \frac{m}{2}; h = 1, 2, \dots, n; \\ X_{i(m:m)h}, i = \frac{m+2}{2}, \frac{m+4}{2}, \dots, m; h = 1, 2, \dots, n \end{cases},$$
(4)

based on MRSSO and MRSSE, respectively.

Extreme ranked set sampling Let $X_{i(1:m)h}$ and $X_{i(m:m)h}$ denote the first and the *m*th order statistics of the *i*th sample (i=1,2,...,m) for the *h*th cycle respectively. In the case of even sample, the estimator of the population median based on ERSS can be given by:

$$\hat{\eta}_{ERSSE} = \text{median} \begin{cases} X_{i(1:m)h}, i = 1, 2, \dots, \frac{m}{2}; h = 1, 2, \dots, n; \\ X_{i(m:m)h}, i = \frac{m+2}{2}, \frac{m+4}{2}, \dots, m; h = 1, 2, \dots, n \end{cases}$$
(5)

In the case of odd sample, the estimator for the population mean can be defined by:

$$\hat{\eta}_{ERSSO} = \text{median} \begin{cases} X_{i(1:m)h}, i = 1, 2, \dots, \frac{m-1}{2}; h = 1, 2, \dots, n; \\ X_{i\left(\frac{m+1}{2}:m\right)h}, i = \frac{m+1}{2}: h = 1, 2, \dots, n: \\ X_{i(m:m)h}, i = \frac{m+3}{2}, \frac{m+5}{2}, \dots, m; h = 1, 2, \dots, n \end{cases}$$
(6)

Balanced group ranked set sampling If *m* is odd, let $X_{i(1:m)h}$ be the first order statistics of the *i*th sample (i=1, 2, ..., k) for the *h*th cycle, $X_{i\left(\frac{m+1}{2}:m\right)h}$ be the median of the *i*th sample (i=k + 1, k + 2, ..., 2k) for the *h*th cycle, and $X_{i(m:m)h}$ be the maximum order statistics of the *i*th sample (i=2k + 1, 2k + 2, ..., 3k) for the *h*th cycle. Corresponding to the three groups, we have the measured units consisting of $X_{1(1:m)h}, X_{2(1:m)h}, ..., X_{k(1:m)h}, X_{k+1\left(\frac{m+1}{2}:m\right)h}, X_{k+2\left(\frac{m+1}{2}:m\right)}, ..., X_{2k\left(\frac{m+1}{2}:m\right)h}, X_{2k+1(m:m)h}, X_{2k+2(m:m)h}, X_{3k(m:m)h}$ which is the BGRSS of size m=3k for the odd sample of the *h*th cycle, denoted as BGRSSO. The estimator of the population median based on BGRSSO, $\hat{\eta}_{BGRSSO}$, can be defined as:

$$\hat{\eta}_{BGRSSO} = \text{median} \begin{cases} X_{i(1:m)h}, i = 1, 2, \dots, k; h = 1, 2, \dots, n; \\ X_{i\left(\frac{m+1}{2}:m\right)h}, i = k+1, k+2, \dots, 2k; h = 1, 2, \dots, n; \\ X_{i(m:m)h}, i = 2k+1, 2k+2, \dots, m; h = 1, 2, \dots, n \end{cases}$$
(7)

In the case of even sample size, let $X_{i(1:m)h}$ be the first order statistics of the *i*th sample (i = 1, 2, ..., k) for the *h*th cycle, $\frac{1}{2} \left(X_{i(\frac{m}{2}m)h} + X_{i(\frac{m+2}{2}m)h} \right)$ be the median of the *i*th sample (i = k + 1, k + 2, ..., 2k) for the *h*th cycle, and $X_{i(m:m)h}$ be the maximum order statistics of the *i*th sample (i = 2k+1, 2k + 2, ..., 3k) for the *h*th cycle. The estimator of the population median, denoted as $\hat{\eta}_{BGRSSE}$, is defined as:

$$\hat{\eta}_{BGRSSE} = \text{median} \begin{cases} X_{i(1:m)h}, i = 1, 2, \dots, k; h = 1, 2, \dots, n; \\ \frac{1}{2} \left(X_{i\left(\frac{m}{2}:m\right)h}, + X_{i\left(\frac{m+2}{2}:m\right)h} \right), i = k + 1, k + 2, \dots, 2k; h = 1, 2, \dots, n; \\ X_{i(m:m)h}, i = 2k + 1, 2k + 2, \dots, m; h = 1, 2, \dots, n \end{cases}$$
(8)

Two-stage ranked set sampling Assume that we are interested in a sample of size m = 3k based on TSRSS. This sample can be obtained based on a simple random sample of size m^3 . In the first stage, the samples found based on SRS are divided into 3 groups, each having $3k^2$ sets of size m. For the *h*th cycle of the first stage of the sampling process, let $X_{i(1m)h}^{(1)}$ be the lowest rank unit of the *i*th sample (i = 1, 2, ..., k), $X_{i(\frac{m+1}{2},m)h}^{(1)}$ be the median of the *i*th sample (i = k + 1, k + 2, ..., 2k) and $X_{i(1m)h}^{(1)}$ be the largest rank unit of the sample (i = 2k + 1, 2k + 2, ..., 3k). For the second stage, the units in each group are ranked ordered, where the smallest rank unit is selected from the first group, the median rank unit from the second group and the largest rank unit from the third group. This results in TRSS of size m, consisting of $\left\{X_{i(1m)h}^{(2)}, X_{2(2m)h}^{(2)}, X_{m(mm)h}^{(2)}\right\}$. The estimator for the population median based on TRSS, $\hat{\eta}_{TRSS}$, is given by:

$$\hat{\eta}_{TRSS} = \text{median} \left\{ X_{1(1:m)h}^{(2)}, X_{2(2:m)h}^{(2)}, \dots, X_{m(m:m)h}^{(2)} \right\}.$$
(9)

Selective order ranked set sampling Consider a sample of size *r* based on SORSS. Let $X_{(i:m)1}, X_{(i:m)2}, ..., X_{(i:m)r}$ denotes this sample. The estimator for the population median based on SORSS, $\hat{\eta}_{SORSS}$, is given by:

$$\hat{\eta}_{SORSS} = \text{median} \left\{ X_{(im)1}, X_{(im)2}, \dots, X_{(im)r} \right\}.$$
 (10)

Assume that we are interested in SORSS via minimum, median and maximum order statistics. If m = 3, the estimators via minimum, median and maximum order statistics are

$$\hat{\eta}_{SORSS(\min)} = \text{median} \left\{ X_{(1:3)1}, X_{(1:3)2}, \dots, X_{(1:3)r} \right\},$$
(11)

$$\hat{\eta}_{SORSS(\text{median})} = \text{median} \left\{ X_{(2:3)1}, X_{(2:3)2}, \dots, X_{(2:3)r} \right\},$$
(12)

and

$$\hat{\eta}_{SORSS(max)} = \text{median} \left\{ X_{(3:3)}, X_{(3:3)}, \dots, X_{(3:3)r} \right\},$$
(13)

respectively.

SIMULATION STUDY

The relative efficiency of the proposed estimators of the population median based on RSS, ERSS, MRSS, BGRSS, TSRSS and SORSS are compared to SRS. Seven probability distribution functions that are considered include uniform, normal, beta, logistic, exponential, gamma and Weibull. Without loss of generality, for simulation, we assume that the cycle is repeated once except for the case of SORSS where r = m. If the distribution is symmetric, the efficiency of, say, RSS relative to SRS, is given by:

$$eff\left(\hat{\eta}_{SRS},\hat{\eta}_{RSS}\right) = \frac{Var\left(\hat{\eta}_{SRS}\right)}{Var\left(\hat{\eta}_{RSS}\right)}$$

and if the distribution is asymmetric, the efficiency is defined by:

$$eff\left(\hat{\eta}_{SRS},\hat{\eta}_{RSS}\right) = \frac{MSE\left(\hat{\eta}_{SRS}\right)}{MSE\left(\hat{\eta}_{RSS}\right)},$$

where MSE denotes the mean square error. The same formulation applies for finding the efficiency of the other estimators relative to SRS. The values of the efficiency and bias are given in Table 1 to 3 for m = 3, 6, 9 respectively.

From the results, TSRSS is more efficient than SRS whether the underlying is symmetric or not. When the underlying distribution is symmetric, the estimators based on TSRSS are unbiased, and the bias is found to be less than those of the other estimators when the distribution is asymmetric. In general, for a particular sample size, based on the relative efficiency of each estimator with respect to SRS, it is found that TSRSS is most efficient compared to RSS, ERSS, MRSS, BGRSS and SORSS.

CONCLUSIONS

When compared to other estimators, TSRSS is recommended for estimating the population median for all the distributions considered except for the uniform distribution. When the underlying distribution is uniform, SORSS based on minimum and maximum are suggested. The methods suggested in this study can be used to estimate other

TABLE 1. The efficiency and bias for estimating the population median based on RSS, ERSS, MRSS, BGRSS, TSRSS and SORSS for m = 3 assuming several different distributions

Distribution	Efficiency (bias) based on								
	RSS	ERSS	MRSS	BGRSS	TSRSS	SORSS (min)	SORSS (max)	SORSS (median)	
Uniform (0,1)	1.454	1.454	1.884	1.454	1.974	2.778	2.778	1.852	
Normal (0,1)	1.613	1.613	2.231	1.613	2.393	1.820	1.820	2.277	
Beta (4,4)	1.582	1.582	2.145	1.582	2.303	2.00	2	2.333	
Logistic (0,1)	1.718	1.718	2.409	1.718	2.507	1.580	1.580	2.437	
Exponential(1)	1.787 (0.086)	1.787 (0.086)	2.591 (0.062)	1.787 (0.086)	2.734 (0.062)	1.859 (-0.416)	0.257 (1.010)	2.731 (0.067)	
Gamma (2,1)	1.702 (0.087)	1.702 (0.087)	2.415 (0.064)	1.702 (0.087)	2.583 (0.060)	0.661 (-1.246)	2.428 (-0.538)	0.984 (-1.007)	
Weibull (1,3)	1.815 (0.258)	1.815 (0.258)	2.578 (0.190)	1.815 (0.258)	2.726 (0.174)	15.560 (-0.058)	0.254 (4.23)	1.788 (1.396)	

TABLE 2. The efficiency and bias for estimating the population median based on RSS, ERSS, MRSS, BGRSS, TSRSS and SORSS for m = 6 assuming several different distributions

Distribution	Efficiency (bias) based on									
	RSS	ERSS	MRSS	BGRSS	TSRSS	SORSS (min)	SORSS (max)	SORSS (median)		
Uniform (0,1)	2.413	3.171	3.331	2.131	8.245	9	9	4.50		
Normal (0,1)	2.730	2.294	3.919	2.352	9.370	2.4607	2.461	4.867		
Beta (4,4)	2.595	2.502	3.757	2.379	8.977	3.50	3.50	3.50		
Logistic (0,1)	2.786	2.017	4.195	2.363	9.726	1.786	1.786	4.861		
Exponential(1)	2.872 (0.054)	1.120 (0.229)	4.459 (0.025)	2.337 (0.074)	9.428 (0.042)	0.511 (-0.563)	0.058 (1.592)	4.970 (.045)		
Gamma (2,1)	2.807 (0.048)	1.556 (0.238)	4.192 (0.024)	2.447 (0.073)	9.394 (0.045)	0.589 (-1.321)	8.095 (-0.245)	0.985 (-1.018)		
Weibull (1,3)	2.841 (0.147)	1.117 (0.687)	4.445 (0.073)	2.464 (0.218)	9.311 (0.130)	12.266 (-0.496)	0.093 (5.960)	1.755 (1.312)		

Distribution –	Efficiency (bias) based on							
	RSS	ERSS	MRSS	BGRSS	TSRSS	SORSS (min)	SORSS (max)	SORSS (median)
Uniform (0,1)	2.981	1.142	5.256	2.343	11.621	23	23	5.75
Normal (0,1)	2.942	1.194	6.062	2.482	13.591	2.879	2.879	6.185
Beta (4,4)	2.772	1.148	5.972	2.331	13.230	5	5	5
Logistic (0,1)	2.882	1.181	6.112	2.479	13.830	1.835	1.835	6.464
Exponential(1)	3.002 (0.018)	1.213 (0.044)	6.451 (0.009)	2.478 (0.022)	14.924 (0.004)	0.318 (-0.610)	0.03 (1.949)	6.611 (.008)
Gamma (2,1)	2.901 (0.019)	1.183 (0.047)	6.402 (0.006)	2.461 (0.022)	14.323 (0.004)	0.583 (-1.345)	18.49 (-0.068)	0.98 (-1.035)
Weibull (1,3)	3.046 (0.05)	1.210 (0.131)	6.437 (0.028)	2.503 (0.067)	14.767 (0.011)	7.021 (-0.638)	0.057 (7.039)	1.8 (1.217)

TABLE 3. The efficiency and bias for estimating the population median based on RSS, ERSS, MRSS, BGRSS, TSRSS and SORSS for m = 9 assuming several different distributions

parameters such as quartiles, ratio and variance. These methods could possibly be implemented with other sampling procedure such as stratified sampling. Since RSS is a cost effective technique, it is worthwhile investigating the performance of various goodness of fit tests under this technique since not much works have been done in this area.

AKNOWLEDGEMENT

This research is financially supported by the Fundamental Research Grant of Universiti Kebangsaan Malaysia (UKM-ST-06-FRGS0096-2009).

REFERENCES

- Al-Subh, S.A., Al-Odat, M.T., Ibrahim, K. & Jemain, A.A. 2009. EDF goodness of fit tests of logistic distribution under selective order statistics. *Pakistan Journal of Statistics* 25(3): 265-274.
- Jemain, A.A., Al-Omari, A.I. & Ibrahim, K. 2007a. Two-stage ranked set sampling for estimating the population median. *Sains Malaysiana* 37(1): 95-99.
- Jemain, A.A., Al-Omari, A.I. & Ibrahim, K. 2007b. Multistage ranked set sampling for estimating the population median. *Journal of Mathematics and Statistics* 3(2): 58-64.
- Jemain, A.A., Al-Omari, A.I. & Ibrahim, K. 2009. Balanced groups ranked set samples for estimating the population median. *Journal of Applied Statistical Science* 17: 39-46.
- McIntyre, G.A. 1952. A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research* 3: 385-390.
- Muttlak, H.A. 1997. Median ranked set sampling. *Journal of Applied Statistical Science* 6: 245-255.

- Muttlak, H.A. 2003. Investigating the use of quartile ranked set sampling for estimating the population mean. *Applied Mathematics and Computation* 146: 437-443.
- Samawi, H.M., Ahmed, M.S. & Abu Dayyeh, W. 1996. Estimation of ratio using rank set sampling. *The Biometrical Journal* 36(6): 753-764.
- Samawi, H.M. & Muttlak, H.A. 2001. On ratio estimation using median ranked set sampling. *Journal of Applied Statistical Science* 10(2): 89-98.
- Samawi, H.M. & Tawalbeh, E.M. 2002. Double median ranked set sample: Comparison to other double ranked samples for mean and ratio estimators. *Journal of Modern Applied Statistical Methods* 1(2): 428-442.
- Takahasi, K. & 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.

School of Mathematical Sciences Faculty of Science and Technology Universiti Kebangsaan Malaysia 43600 UKM Bangi, Selangor D.E. Malaysia

*Corresponding author; email: kamarulz@ukm.my

Received: 28 April 2010 Accepted: 1 November 2010