Modified Sign Test Using Reverse Ranked Ordering-Set Samples

Hyun Gee Kim¹⁾ and Dong Hee Kim²⁾

Abstract

The method of Reverse Ranked Ordering-Set Sampling(RROSS) as an opposed Ranked Ordering-Set Sampling(ROSS) and Ranked-Set Sampling(RSS) is discussed. We propose the test statistic using sign test on RROSS. This method is effective when observations are expensive and measurement is perhaps destructive or invasive. This method obtains more informations than ROSS and RSS. The asymptotic relative efficiencies of RROSS with respect to ROSS and RSS are always greater than 1 for all sample sizes. We consider a simple model to describe the effect of imperfect judgment errors.

Keywords: Reverse ranked ordering-set sample(RROSS); Ranked ordering-set samples(ROSS); Ranked-setsamples(RSS); Asymptotic relative efficiency; Imperfect judgment error.

1. Introduction

Many authors discussed the sampling method, the Ranked-Set Sampling(RSS) method. This sampling method is useful when measurements are destructive or expensive or invasive, while ranking of the observation is relatively easy. This sampling was introduced first by McIntyre (1952). He suggested the RSS method for assessing yields of pasture plots without actually mowing and weighing the hay for a large number of plots. Dell and Clutter (1972) considered a useful technique for improving estimates of the mean in the situation that the measurements of observations were imperfectly judged. Stokes (1980) investigated the estimation of a population variance and asymptotic relative efficiency(ARE) on the RSS. In one sample location problem, Hettmansperger (1995) studied the sign test statistic on RSS. Koti and Babu (1996) calculated the exact null distribution of the ranked-set sample sign test statistic. In two sample location problem, Bohn and Wolfe (1992, 1994) proposed the Mann-Whitney-Wilcoxon test statistic and

¹⁾ Instructor, Department of Statistics, Pusan National University, Pusan, 609-735, Korea. Correspondence: hgeekim@pusan.ac.kr

²⁾ Professor, Department of Statistics, Statistical research institute, Pusan National University, Pusan, 609-735, Korea.

investigated the properties of the test procedures based on RSS for perfect and imperfect judgments. Öztürk (1999a, b) studied the one- and two-sample sign tests for RSS and extended one sample sign test statistic to two sample sign test statistic on RSS. Kim and Kim (2003) considered the sign test on ROSS and the imperfect judgements errors.

In this paper, we discuss the reverse ranked ordering-set sampling(RROSS) method as an alternative to ROSS and RSS. The RROSS method is similar to ROSS, but RROSS is different from ROSS in the sense of using the reverse ordering on ROSS. In one sample location problem, we will propose sign test statistic using RROSS which have more informations than ROSS and RSS for the sign test. In section 2, we introduce RROSS method and propose the sign test statistic. The asymptotic properties of the proposed test statistic are shown in section 3. In section 4, we consider the sampling for the imperfect judgment errors. Section 5 has conclusions.

2. Test Statistic

2.1 Reverse Ranked Ordering-Set Sampling

In order to use more informations than ROSS and RSS, we introduce a Reverse Ranked Ordering-Set Sampling(RROSS). In the first place, we introduce the ROSS that take a sample of size k from a specified population and an expert ranks the observations using a visual ordering of the observations or using the ordering of concomitant variable in the samples. This process is repeated until we have k different samples each having k ordered units; $X_{(11)}, \dots, X_{(1k)}, X_{(21)}, \dots, X_{(2k)}, \dots, X_{(k1)}, \dots, X_{(kk)}$, where $X_{(ij)}$ is the j-th order statistic in i-th sample. At this stage, we perform another ordering among the units having the same ranking from all of the k samples, i. e. $X_{(1j)}, \dots, X_{(kj)}$, for $j = 1, \dots, k$ and measure the j-th ordered units for quantification. The quantified observations will be denoted by $X_{(j)1}^{(j)}$, $j = 1, \dots, k$. This sampling method is concretely explained in <Figure 1>. But, the quantified observations for ROSS in each cycle are ordered, i. e. $X_{(1j)}^{(1)} \leq X_{(2)i}^{(2)} \leq \dots \leq X_{(k)i}^{(k)}$. Kim and Kim (2003) have shown that the ARE of ROSS with respect to RSS is greater than 1, but for the case of k = 2 only, less than 1.

Now we introduce the new sampling method, Reverse Ranked Ordering-Set Sampling(RROSS). The RROSS method is similar to ROSS, but the quantified observations for RROSS are not ordered and independent in each cycle. The sampling method is as follows;

 \langle Figure 1 \rangle Display of k measurement observations of the first cycle(ROSS).

We take a sample of size k from a specified population and an expert ranks the observations using a visual ordering of the observations or using the ordering of a concomitant variable in the samples. This process is repeated until we have k different samples each having k ordered units; $X_{(11)}, \dots, X_{(1k)}, X_{(21)}, \dots, X_{(2k)}, \dots, X_{(k1)}, \dots, X_{(kk)}$ where $X_{(ij)}$ is the j-th order statistic in the i-th sample. At this stage, we perform another ordering among the units having the same ranking from different samples, i. e. $X_{(1j)}, \dots, X_{(kj)}$, for $j=1,\dots,k$ and measure the (k-j+1)-th ordered units for quantification. The quantified observations will be denoted by $X_{(j)1}^{(k-j+1)}$, $j=1,\dots,k$. For example, if j is 1, we judge the largest observation among $X_{(11)}, \dots, X_{(k1)}$, which will be denoted as $X_{(1)1}^{(k)}$. This sampling method is explained in $\langle Figure 2 \rangle$.

<Figure 2> Display of k measurement observations of the first cycle(RROSS).

2.2 The Proposed Test Statistic

We consider one sample location problem on RROSS. Let $X_{(1)1}^{(k)}, \dots, X_{(1)n}^{(k)}, \dots, X_{(k)1}^{(n)}, \dots, X_{(k)n}^{(n)}$ be a RROSS of size nk as shown in <Figure 2>. We repeat the process for n cycles to get nk measurements.

$$egin{array}{lll} X_{(1)1}^{(k)}, \cdots, X_{(1)n}^{(k)} &\sim \ iid \, h_{(1)}^{(k)}(t) \ &dots &dots &dots &dots \ X_{(j)1}^{(k-j+1)}, \cdots, X_{(j)n}^{(k-j+1)} &\sim \ iid \, h_{(j)}^{(k-j+1)}(t) \ &dots &dots &dots &dots \ X_{(k)1}^{(1)}, \cdots, X_{(k)n}^{(1)} &\sim \ iid \, h_{(k)}^{(1)}(t) \end{array}$$

The density function $h_{(j)}^{(k-j+1)}(t)$ represents the pdf of the (k-j+1)-th order statistic from a distribution, $H_{(j)}$, where $H_{(j)}$ is the cdf of the j-th order statistic from a set of size k, i.e. the density is of the form ;

$$\begin{split} h_{(j)}^{(k-j+1)}(t) &= \frac{k!}{(j-1)!(k-j)!} [H_{(j)}(t)]^{k-j} [1 - H_{(j)}(t)]^{j-1} h_{(j)}(t) \\ &= [\frac{k!}{(j-1)!(k-j)!}]^2 [H_{(j)}(t)]^{k-j} [1 - H_{(j)}(t)]^{j-1} \\ &\times H(t)^{j-1} [1 - H(t)]^{k-j} h(t), \end{split}$$

where $H_{(j)}(t) = \int_{-\infty}^{t} h_{(j)}(x) dx$, $h_{(j)}(t)$ is the pdf of j-th order statistic from the distribution having pdf h(t).

We suppose the population is distributed as $H(x) = F(x - \theta)$ and wish to test for H_0 : $\theta = 0$ against H_A : $\theta \neq 0$. Sign test is a very useful and easy test for one sample location problem. So, we propose a sign test statistic as a test statistic on RROSS, which is given by

$$S_{RROSS}^{+} = \sum_{j=1}^{k} \sum_{i=1}^{n} I(X_{(j)i}^{(k-j+1)} > 0)$$

where $I(\cdot)$ is an indicator function.

We easily evaluate the mean and variance of the test statistic S_{RROSS}^+

$$E(S_{RROSS}^{+}) = n \sum_{j=1}^{k} (1 - H_{(j)}^{(k-j+1)}(0)) = nk - n \sum_{j=1}^{k} H_{(j)}^{(k-j+1)}(0),$$

$$Var(S_{RROSS}^{+}) = n [\sum_{j=1}^{k} (1 - H_{(j)}^{(k-j+1)}(0)) H_{(j)}^{(k-j+1)}(0)],$$

where $H_{(j)}^{(k-j+1)}(0) = \int_{-\infty}^{0} h_{(j)}^{(k-j+1)}(x) dx$, $h_{(j)}^{(k-j+1)}(t)$ is the pdf of the (k-j+1) -th order statistic from a distribution, $H_{(j)}$.

3. The Asymptotic Properties

To obtain the asymptotic properties of S^+_{RROSS} , we consider the results of the mean and variance for the sign test statistic. The testing problem of our interest is $H_0: \theta = 0$ in favor of $H_A: \theta \neq 0$. The following theorem provides the mean and variance of the RROSS using sign test statistic under the null hypothesis.

Theorem. Assume that H_0 : $\theta = 0$ is true, then H(0) = F(0) = 1/2,

$$E_0(S_{RROSS}^+) = nk/2, \ Var_0(S_{RROSS}^+) = (nk/4)\delta_{Roo}^2,$$

where
$$\delta_{Roo}^2 = 1 - \frac{4}{k} \sum_{j=1}^k [F_{(j)}^{(k-j+1)}(0) - \frac{1}{2}]^2$$
 with $F_{(j)}^{(k-j+1)}(0) = \int_{-\infty}^0 \frac{k!}{(j-1)!(k-j)!} [F_{(j)}(t)]^{k-j} [1 - F_{(j)}(t)]^{j-1} dF_{(j)}(t).$

Proof. Using the fact $\frac{1}{k} \sum_{j=1}^{k} H_{(j)}^{(k-j+1)}(0) = H(0) = 1/2$ and from Mathematica, the expectation follows at once. The variance follows after a bit of algebra.

We reject the null hypothesis if $|S_{RROSS}^+ - nk/2| > z_{\alpha/2} (nk\delta_{Roo}^2/4)^{1/2}$ where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile from the standard normal distribution. To decide the efficiency of S_{RROSS}^+ , as a competitor of S_{RROSS}^+ , we consider two statistic, Kim and Kim's (2003) statistic given by

$$S_{ROSS}^{+} = \sum_{j=1}^{k} \sum_{i=1}^{n} I(X_{(j)i}^{(j)} > 0)$$

and Hettmansperger's (1995) statistic given by

$$S_{RSS}^+ = \sum_{j=1}^k \sum_{i=1}^n I(X_{(j)i} > 0).$$

Corollary 1. (Theorem of Kim and Kim (2003))

Under the assumption that H_0 : $\theta = 0$ is true, H(0) = 1/2,

$$E_0(S_{ROSS}^+) = nk/2, \ Var_0(S_{ROSS}^+) = (nk/4)\delta_{oo}^2,$$

where
$$\delta_{oo}^2 = (4/k) \left[\sum_{j=1}^k (1 - F_{(j)}^{(j)}(0)) F_{(j)}^{(j)}(0) + 2 \sum_{j=1}^{k-1} \sum_{j'=j+1}^k (1 - F_{(j)}^{(j)}(0)) F_{(j')}^{(j')}(0) \right]$$
 with

$$F_{(j)}^{(j)}(0) = \int_{-\infty}^{0} \frac{k!}{(j-1)!(k-j)!} F_{(t)(1-F_{(j)}(t))^{k-j}dF_{(j)}(t).(j)}^{j-1}$$

In <Table 1>, we provide the values of $F_{(j)}^{(k-j+1)}(0)$ for each k under H_{0} .

j	sample size									
	2	3	4	5	6	7	8	9	10	
1	0.563	0.670	0.773	0.853	0.910	0.947	0.969	0.983	0.990	
2	0.437	0.500	0.630	0.763	0.867	0.934	0.969	0.987	0.995	
3		0.330	0.370	0.500	0.660	0.803	0.903	0.960	0.985	
4			0.227	0.237	0.340	0.500	0.678	0.827	0.923	
5				0.147	0.133	0.197	0.322	0.500	0.690	
6					0.090	0.066	0.097	0.173	0.310	
7						0.053	0.031	0.040	0.077	
8							0.031	0.013	0.015	
9								0.017	0.005	
10									0.010	
$\delta_{Roo}^{(k)^2}$	0.984	0.923	0.818	0.690	0.563	0.452	0.365	0.299	0.251	

<Table 1> Values of $F_{(j)}^{(k-j+1)}(0), \ \delta_{Roo}^{(k)2}$

Corollary 2. (Theorem of Hettmansperger (1995))

Under the assumption that H_0 : $\theta = 0$ is true, H(0) = 1/2,

$$E_0(S_{RSS}^+) = nk/2, \ Var_0(S_{RSS}^+) = (nk/4)\delta_o^2,$$

where
$$\delta_o^2 = 1 - (4/k) \sum_{j=1}^k [F_{(j)}(0) - 1/2]^2$$
 with $F_{(j)}(0) = \int_{-\infty}^0 \frac{k!}{(j-1)!(k-j)!} [F(t)]^{j-1} [1 - F(t)]^{k-j} dt$.

We next derive the asymptotic relative efficiency of S_{RROSS}^+ with respect to S_{ROSS}^+ and S_{RSS}^+ by the definition of Randles and Wolfe (1979). We must compute the efficacy

$$e = \lim_{n \to \infty} \mu'_{nk}(0) / \sqrt{nk} \, \sigma_{nk}$$

where $\mu'_{nk}(0)$ is the derivation of $E_{\theta}(S^+_{RROSS})$ evaluated at $\theta = 0$ and σ_{nk} is the null standard deviation of S^+_{RROSS} given in Theorem.

$$E_{\theta}(S_{RROSS}^{+}) = n \sum_{j=1}^{k} (1 - H_{(j)}^{(k-j+1)}(-\theta)),$$

The $\mu'_{nk}(0)$ is given by

$$\partial E_{\theta}(S^+_{RROSS})/\partial \theta|_{\theta=0} = n \sum_{j=1}^k h_{(j)}^{(k-j+1)}(-\,\theta)|_{\theta=0} = n \sum_{j=1}^k h_{(j)}^{(k-j+1)}(0).$$

The asymptotic relative efficiencies of S_{RROSS}^+ with respect to S_{ROSS}^+ and S_{RSS}^+ are given by the ratio of the square of the efficacy;

$$ARE(S_{RROSS}^{+}, S_{ROSS}^{+}) = \left(\sum_{j=1}^{k} h_{(j)}^{(k-j+1)}(0)\right)^{2} \delta_{oo}^{2} / \left(\sum_{j=1}^{k} h_{(j)}^{(j)}(0)\right)^{2} \delta_{Roo}^{2}$$
$$= d^{2} \delta_{oo}^{2} / c^{2} \delta_{Roo}^{2}$$

and

$$egin{align} ARE(S^+_{RROSS}, S^+_{RSS}) &= (\sum_{j=1}^k h^{(k-|j|+1)}_{(j)}(0))^2 \delta_o^2/(kh(0))^2 \delta_{Roo}^2 \ &= d^2 \delta_o^2/k^2 \delta_{Roo}^2, \end{split}$$

where

$$c = \sum_{j=1}^{k} \left[\frac{k!}{(j-1)!(k-j)!} \right]^{2} [F_{(j)}(0)]^{j-1} [1 - F_{(j)}(0)]^{k-j} F^{j-1}(0) (1 - F(0))^{k-j}$$

$$= \sum_{j=1}^{k} k^{2} {k-1 \choose j-1}^{2} 2^{-(k^{2}-1)} \left[\sum_{r=j}^{k} {k \choose r} \right]^{j-1} \left[\sum_{r=0}^{j-1} {k \choose r} \right]^{k-j}$$

and

$$\begin{split} d &= \sum_{j=1}^{k} \left[\frac{\mathbf{k}!}{(\mathbf{j}-1)!(\mathbf{k}-\mathbf{j})!} \right]^{2} [\mathbf{F}_{(\mathbf{j})}(0)]^{\mathbf{k}-\mathbf{j}} [1-\mathbf{F}_{(\mathbf{j})}(0)]^{\mathbf{j}-1} [\mathbf{F}(0)]^{\mathbf{j}-1} [1-\mathbf{F}(0)]^{\mathbf{k}-\mathbf{j}} \\ &= \sum_{j=1}^{k} k^{2} {k-1 \choose j-1}^{2} 2^{-(k-1)} \sum_{s=0}^{j-1} {j-1 \choose s} (-1)^{s} \\ &\times \left[\frac{k!}{(k-j)!(j-1)!} \sum_{u=0}^{k-j} {k-j \choose u} (-1)^{u} \frac{2^{-(j+u)}}{j+u} \right]^{k-j+s} \end{split}$$

In order to obtain the asymptotic relative efficiencies of S^+_{RROSS} with respect to S^+_{ROSS} and S^+_{RSS} , we must see the relation between $\sum_{j=1}^k h^{\binom{k-j+1}{j}}(0)$ and h(0), i.e. $\sum_{j=1}^k h^{\binom{k-j+1}{j}}(0) = d \cdot h(0)$. The values of d vary as k does. It is very difficult to find the values of d analytically, so we find them from Mathematica. Some values of d and k are in <Table 3>.

Table 2> Values d of $\sum_{j=1}^k h_{(j)}^{(k-j+1)} = d \cdot h(0)$ and Values c of $\sum_{j=1}^k h_{(j)}^{(j)} = c \cdot h(0)$

\overline{k}	2	3	4	5	6	7	8	9	10
c	1.00	2.32	2.42	3.78	3.94	5.33		6.93	7.15
\overline{d}	3	5.70	8.61	11.30	13.46	15.07	16.24	17.16	17.99

< Table 3> Values of $ARE(S_{RROSS}^+, S_{RSS}^+)$ and $ARE(S_{RROSS}^+, S_{ROSS}^+)$.

\overline{k}	2	3	4	5	6	7	8	9	10
$ARE(S_{RROSS}^+, S_{RSS}^+)$	1.71	2.44	3.10	3.63	4.03	4.26	4.44	4.51	4.54
$\overline{ARE(S_{RROSS}^{+}, S_{ROSS}^{+})}$	2.29	2.20	2.89	2.79	3.07	2.85	2.89	2.67	2.65

The asymptotic relative efficiencies of S_{RROSS}^+ with respect to S_{ROSS}^+ and S_{RSS}^+ are in <Table 3> for several values of k. <Table 3> represents that the asymptotic relative efficiencies are greater than 1 for all sample size k.

4. Imperfect judgment

So far, we have assumed that the expert can make perfect judgment. Now, we consider the case of imperfect judgment errors with the RROSS. Let q_{ij} be the probability that we judge the item by $X_{(j)}^{(k-j+1)}$ in spite of $X_{(i)}^{(k-i+1)}$. The representation of q_{ij} is $\sum_{l=1}^k p_{il} r_{lj}$, where p_{il} is the probability that the item judged to have rank l actually has rank i in the first phase ranking and r_{lj} is the probability that the item judged to have rank (k-j+1) actually has rank (k-l+1) in the second phase ranking. We will let $x_{(j)}^{[k-j+1]}$ denote the measurements of the observation judged to have rank (k-j+1) in the second phase ranking among judged to have rank j in the first phase ranking and let $x_{(j)}^{(k-j+1)}$ denote the observation with the (k-j+1)-th rank in the second phase ranking among the j-th rank in the first phase ranking. The distribution function of $X_{(j)}^{(k-j+1)}$ is then given by

$$H_{[j]}^{[k-j+1]}(x) = \sum_{i=1}^{k} q_{ij} H_{(i)}^{(k-i+1)}(x)$$
 ,

where $H_{(i)}^{(k-i+1)}(x)$ is the cdf of the (k-i+1)-th order statistic of sample from a distribution $H_{(i)}(x)$. Note that the matrix of q_{ij} is doubly stochastic; hence $\sum_{i=1}^k q_{ij} = 1$ and we replace the variance of ROSS having imperfect judgment errors

by $(nk/4)\delta_{Rp}^2$, where

$$\delta_{Rp}^2 = 4/k \sum_{j=1}^k \left[\sum_{i=1}^k q_{ij} F_{(i)}^{(k-i+1)}(0) \left(1 - \sum_{i=1}^k q_{ij} F_{(i)}^{(k-i+1)}(0) \right) \right].$$

Corollary 3. Let Q be the matrix with the ij-th element equal to $q_{ij}=1/k$ (the case of completely random ranking) and $F_{(j)}^{(k-j+1)}(0)$ is in <Table 1>. Then $\delta_{Rp}^2=1$.

When the ranking is completely random, the asymptotic relative efficiency of S^+_{RROSS} with respect to S^+_{RSS} is 1.

Example. Suppose k=5 and let the matrix Q be $P \cdot P$ where P is the same matrix in Hettmansperger (1995).

Risperger (1993).
$$Q = P \cdot P = \begin{pmatrix} 0.6250 & 0.3125 & 0.0625 & 0 & 0 \\ 0.3125 & 0.3750 & 0.2500 & 0.0625 & 0 \\ 0.0625 & 0.2500 & 0.3750 & 0.2500 & 0.06250 \\ 0 & 0.0625 & 0.2500 & 0.3750 & 0.3125 \\ 0 & 0 & 0.0625 & 0.3125 & 0.6250 \end{pmatrix}$$

In this case, δ_{Rp}^2 =0.7935 with δ_{Ro}^2 =0.690. The percent reduction in variance due to RROSS is 1-0.7935=0.2065

5. Conclusion

In one sample location problem with sign test, we propose the new sampling method, Reverse Ranked Ordering-Set Samples(RROSS), which contains more informations than ROSS and RSS method when there is no judgement error. Proposed statistic using the RROSS is superior to the ROSS and RSS for all of k.

But it is difficult to show analytically that $\sum_{j=1}^{k} H_{(j)}^{(k-j+1)}(0)$ is equal to k/2, so until now we can only show it by Mathematica. We consider the case of imperfect judgment errors with the RROSS, the percent reduction in variance due to RROSS is 0.2061 for the previous example of k=5.

References

[1] Bohn, L.L. and Wolfe, D.A. (1992). Nonparametric two-sample procedures for ranked-set samples data. *Journal of the American Statistical*

- Association, Vol. 87, 552-561.
- [2] Bohn, L.L. and Wolfe, D.A. (1994). The effect of imperfect judgement rankings on properties of procedures based on the ranked-set samples analog of Mann-Whitney-Wilcoxon statistic. *Journal of the American Statistical Association*, Vol. 89, 168-176.
- [3] Dell, T.R. and Clutter, J.L. (1972). Ranked-set sampling theory with order statistics background. *Biometrics*, Vol. 28, 545-553.
- [4] Hettmansperger, T.P. (1995). The ranked-set sample sign test. *Journal of the Nonparametric Statistics*, Vol. 4, 271–282.
- [5] Kim, D.H. and Kim, H.G. (2003). Sign Test Using Ranked Ordering-Set Sampling. *Journal of Nonparametric Statistics*, Vol. 15, 303-309.
- [6] Koti, K.M. and Babu, G.J. (1996). Sign test for ranked-set sampling. Communications in Statistics-Theory and Methods, Vol. 25, 1617-1630.
- [7] McIntyre, G.A. (1952). A method of unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, Vol. 3, 385–390.
- [8] Öztürk, O. (1999a). One- and two-sample sign tests for ranked set sample selective designs. *Communications in Statistics-Theory and Methods*, Vol. 28, 1231–1245.
- [9] Öztürk, O. (1999b). Two-sample inference based on one-sample ranked set sample sign statistics. *Journal of Nonparametric Statistics*, Vol. 10, 197-212.
- [10] Randles, R.H. and Wolfe, D.A. (1979). *Introduction to the Theory of Nonparametric Statistics*, John Wiley, New York.
- [11] Stokes, S.L. (1980). Estimation of variance using judgement ordered ranked-set samples. *Biometrics*, Vol. 36, 35-42.

[Received March 2006, Accepted June 2006]