King Fahd University of Petroleum & Minerals

**DEPARTMENT OF MATHEMATICAL SCIENCES**

Technical Report Series

TR256

May 2000

# Random Ranked Set Samples

I. Rahimov, H.A. Muttlak

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

# RANDOM RANKED SET SAMPLES

[1] I. Rahimov and H. A. Muttlak

Ranked set sampling (RSS) as suggested by McIntyre (1952) uses fixed set size and number of cycles (or replications). In real life however, we may encounter problems that requiring random set size or number of cycles or both. In dealing with such problems we suggeste several unbiased estimators of the population mean using random ranked set sampling (RRSS) method. The efficiencies of the estimators of the population mean under RRSS and RSS are compared. The results show, under certain conditions, the efficiency of estimators is improved by using RRSS. The asymptotic properties of the newly suggested estimators are also considered.

KEY WORDS: Asymptotic properties; discrete uniform distribution; efficiency; random number of replications; random set size; ranked set sampling.

## 1. INTRODUCTION

The ranked set sampling (RSS) has attracted number of authors as an efficient sampling method, particularly in the area of environmental and ecological investigations. The RSS proposed by McIntyre (1952) is a sampling method proven to be more efficient when units are difficult and costly to measure, but are easy and cheap to rank with respect to a variable of interest without actual measurement. One can often tell which tree is the

---

[1] I. Rahimov is Professor of Probability and Statistics and H. A. Muttlak is Associate Professor of Statistics, Department of Mathematical Sciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia. The authors are indebted to King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia, for excellent research facilities.

tallest without measuring all the trees. The RSS method can be summarized as follows: From a population of interest, k random sets each of size *k* are selected. The members of each random set are ranked with respect to the variable of interest by a cost free method e.g. eye inspection. From the first ordered set, the smallest unit is selected for measurement. From the second ordered set, the second smallest unit is selected for measurement. This continues until the largest element from the last ordered set is measured. This process may be repeated r times (i.e. *r* cycles or replications) to yield a sample of size rk.

The RSS is based on fixed set sizes and number of replications. But in some applications we might be faced with problems where k, r, or both cannot be fixed. The example considered by Muttlak and McDonald (1992) demonstrates the need for at least r to be random. The line intercept (transect) is a widely used sampling method in ecological and environmental studies. Units are plants, animal species and etc. The RSS method will apply after the units are sampled using the line transect in the first phase. We know that the number of units $n_1$ (say) that are intercepted by the line can not be fixed. Muttlak and McDonald (1992) assumed that $n_1 \geq k^2 r$. Obviously, this assumption will be violated in most applications and hence we cannot use RSS. As we can see, this situation can be handled easily using the random ranked set sampling (RRSS) method by letting either r or k to be random.

Takahasi and Wakimoto (1968) supplied mathematical theory to support McIntyre's (1952) suggestion. Stokes and Sager (1988) developed the properties of the empirical distribution function based on RSS and compared these properties to the usual empirical distribution function using simple random sample (SRS) data. Bohn and Wolfe (1992, 1994) developed the Mann-Whitney-Wilcoxon statistic using RSS for both perfect ranking and ranking with errors. Kvam and Samaniego (1993, 1994) developed the estimation of the population distribution function and population mean using unbalanced RSS data i.e. the size of the $i^{th}$ set need not be the same for all sets and the various order statistics need not be represented an equal number of time. Koti and Babu (1996) derived the exact null distribution of the RSS sign test. Huang (1997) considers the asymptotic properties of the nonparametric maximum likelihood estimator (NPMLE) of a distribution function using RSS. Kim and Arnold (1999a, 1999b) considered estimating the distribution function F and Bayesian parameter estimation for specified parameter under both balanced and unbalanced RSS. Hartlaub and Wolfe (1999) generalized the one- and two-sample location problems considered in the previous nonparametric work in the area of RSS to m-sample location problem. Presnell and Bohn (1999) developed the U-statistics using RSS data for one and two sample cases. Özturk (2000) investigated the effect of different RSS protocols on sign test statistic.

Several authors considered some modifications of the RSS method either to improve the efficiency of the estimators or (and) to make the RSS method easier to implement in the field. Samawi et al. (1996) studied the properties of the extreme ranked set sampling (ERSS) in estimating the population mean. Muttlak (1997) suggested the use of median ranked ret sampling (MRSS) to estimate the population mean.

Li et al. (1999) introduced the notion of random selection of m sets out of k sets, $m < k$, where k is the set size in the usual RSS method. They studied the properties of the estimators of the population mean and variance based on the new randomly selected sample.

For classified and extensively reviewed work in the area of RSS from 1952 to 1994 see Patil et al (1994) and Kaur et al (1995). Finally for bibliography in the area of RSS see Patil et al. (1999).

In this paper we provided a new direction of RSS via the notion of random ranked set sampling (RRSS). In Section 2 the idea of RRSS is introduced in the case of one replication i.e. single cycle. The cases of r replications with random set size and fixed set size with random number of replications are discussed in Sections 3 and 4 respectively. The general case of RRSS with random set sizes and replications is considered in Section 5. The asymptotic properties of the estimator of the population mean suggested for the general case of RRSS is established in Section 6. In Section 7 we calculate the efficiency of the newly suggested estimators for specific probability distributions and compare these to the RSS. Some concluding remarks are given in the last Section.

## 2. SINGLE CYCLE WITH RANDOM SET SIZE

We consider the following family of random variables

$X_{11}, X_{12}, ..., X_{1\ell}; X_{21}, X_{22}, ..., X_{2\ell}; ....; X_{i1}, X_{i2}, ..., X_{i\ell}; ...; X_{\ell 1}, X_{\ell 2}, ..., X_{\ell\ell}$ where $X_{ij}, i, j = 1,2,...; \ell \in \Lambda = \{2,3,...\}$ are independent and identically distributed random variables with cdf F(x), pdf f(x), mean $\mu$ and variance $\sigma^2$. Let $\nu$ be a random variable taking values from $\Lambda = \{2, 3, ...\}$. Let $X_{i(1)}^{(\nu)}, X_{i(2)}^{(\nu)}, ..., X_{i(\nu)}^{(\nu)}$ be the order statistics of $X_{i1}, X_{i2}, ..., X_{i\nu}, i = 1,2,...,\ell$. To simplify the notations for any $\ell \in \Lambda$, we will use $y_i^{(\ell)} = X_{i(i)}^{(\ell)}, i = 1,2,...,\ell$. It is easy to see that $y_i^{(\ell)}, i = 1,2,...,\ell$ are independent but not identically distributed random variables. We propose

$$\bar{y}_{(\nu)} = \frac{1}{\nu}\sum_{i=1}^{\nu} y_i^{(\nu)} \qquad (1)$$

as an estimator of the population mean $\mu$. Assume from now on that the random variable $\nu$ and the family of random variables $X_{ij}^{(\ell)}$ are independent. We denote also the cdf, pdf, mean and variance of $y_i^{(\ell)}$ by $F_{\ell i}(x)$, $f_{\ell i}(x)$, $\mu_{\ell i}$, and $\sigma_{\ell i}^2$ respectively. It follows from the definition of $y_i^{(\ell)}$ for any $\ell \in \Lambda$ that

$$f(x) = \frac{1}{\ell}\sum_{i=1}^{\ell} f_{\ell i}(x). \tag{2}$$

The properties of the estimator $\bar{y}_{(\nu)}$ are:

(i) $\bar{y}_{(\nu)}$ is an unbiased estimator of population mean μ with a variance

(ii) $V(\bar{y}_{(\nu)}) = E\left[\frac{1}{\nu}\sigma_{(\nu)}^2\right]$, where $\sigma_{(\ell)}^2 = \frac{1}{\ell}\sum_{i=1}^{\ell}\sigma_{\ell i}^2$.

We can easily proof (i) by using the total probability formula and equation (2). For any $\ell$ we have

$$E[\frac{1}{\ell}\sum_{i=1}^{\ell} y_i^{(\ell)}] = \frac{1}{\ell}\sum_{i=1}^{\ell}\int_{-\infty}^{\infty} x f_{\ell i}(x)dx = \int_{-\infty}^{\infty} x \frac{1}{\ell}\sum_{i=1}^{\ell} f_{\ell i}(x)dx = \mu,$$

then

$$E(\bar{y}_{(\nu)}) = E\{E[\frac{1}{\nu}\sum_{i=1}^{\nu} y_i^{(\nu)} | \nu]\} = \sum_{\ell=1}^{\infty} E[\frac{1}{\ell}\sum_{i=1}^{\ell} y_i^{(\ell)}]P(\nu = \ell) = \mu. \tag{3}$$

This shows that $\bar{y}_{(\nu)}$ is an unbiased estimator of population mean μ for any random variable ν. Now we consider the proof of (ii). Again by using the total probability formula we can write

$$V(\bar{y}_{(\nu)}) = \sum_{\ell=1}^{\infty} E\{[\frac{1}{\nu}\sum_{i=1}^{\nu} y_i^{(\nu)} - \mu]^2 | \nu = \ell\}P(\nu = \ell) = \sum_{\ell=1}^{\infty} E\{\frac{1}{\ell}\sum_{i=1}^{\ell}[y_i^{(\ell)} - E(y_i^{(\ell)})]\}^2 P(\nu = \ell),$$

then $V(\bar{y}_{(\nu)}) = E[\frac{1}{\nu^2}\sum_{i=1}^{\nu}\sigma_{\ell i}^2]$. If denote $\sigma_{(\ell)}^2 = \frac{1}{\ell}\sum_{i=1}^{\ell}\sigma_{\ell i}^2$, then we get

$$V(\bar{y}_{(\nu)}) = E\left[\frac{1}{\nu}\sigma_{(\nu)}^2\right]. \tag{4}$$

Let us consider particular cases of the formula (4) for given distributions of ν.

**Example 1.** Assume that ν has geometric distribution truncated at one with a parameter p. In this case we obtain $V(\bar{y}_{(\nu)})$ in equation (4) as

$$V(\bar{y}_{(\nu)}) = p\sum_{\ell=2}^{\infty}\frac{1}{\ell}q^{\ell-2}\sigma_{(\ell)}^2, \tag{5}$$

where q = 1- p.

**Example 2.** Let ν have a binomial distribution truncated at zero and one with parameters n and p i.e. $P(\nu = k) = P(\xi = k|\xi > 1), k = 2,3,\ldots,n$ where ξ is a binomial random variable of the same parameters. In this case the variance of $V(\bar{y}_{(\nu)})$ is

$$V(\bar{y}_{(\nu)}) = \frac{q^n n!}{1-q^n-npq^{n-1}}\sum_{\ell=2}^{n}\frac{1}{\ell!(n-\ell)!\ell}\left(\frac{p}{q}\right)^{\ell}\sigma_{(\ell)}^2. \tag{6}$$

**Example 3.** Let us assume that ν has a uniform distribution on the set {2, 3,..., N}. In this case the variance $V(\bar{y}_{(\nu)})$ is given by

$$V(\bar{y}_{(\nu)}) = \frac{1}{N-1}\sum_{\ell=2}^{N}\frac{1}{\ell}\sigma_{(\ell)}^2. \tag{7}$$

To compare the proposed estimator $\bar{y}_{(\nu)}$ to the RSS estimator, $\bar{y}_{(k)} = \frac{1}{k}\sum_{i=1}^{k}y_{(i)}$, where $y_{(i)}$ is the i[th] order statistic from the i[th] set of fixed size k, k =2, 3,..., N, it is easy to see that $V(\bar{y}_{(k)}) = \frac{1}{k}\sigma_{(k)}^2$, where $\sigma_{(k)}^2 = \frac{1}{k}\sum_{i=1}^{k}\sigma_{ki}^2$. As shown by Takahasi and Wakimoto (1968), $\sigma_{(k)}^2 > \sigma_{(k+1)}^2$ and consequently $\frac{1}{k}\sigma_{(k)}^2$ is also decreasing on k. Using these results and equation (7) we may state the following proposition

**Proposition 1.** There exist $2 \le \tau \le N$ such that $V(\bar{y}_{(k)}) > V(\bar{y}_{(\nu)})$ for $k \le \tau$ and $V(\bar{y}_{(k)}) < V(\bar{y}_{(\nu)})$ for $\tau < k \le N$.

It is clear that the number τ depends on the form of the initial density function f(x). In Section 7 we will consider different concrete distributions to obtain the value τ.

## 3. FIXED NUMBER OF CYCLES WITH RANDOM SET SIZE

Let $\nu_1, \nu_2, ..., \nu_m$ are independent and identically distributed random variables taking values from $\Lambda$. If the process of Section 2 is repeated m times, i.e. we replicate the cycle m times with set sizes $\nu_i$, where i = 1, 2,..., m, we will obtain a sequence of estimators $\bar{y}_{\nu_1 1}, \bar{y}_{\nu_2 2}, ..., \bar{y}_{\nu_m m}$. It is clear that $\bar{y}_{\nu_i i}, i \geq 1$ are independent and identically distributed with the following mean and variance respectively

$$E(\bar{y}_{\nu_i i}) = \mu, \; V(\bar{y}_{\nu_i i}) = E[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}], \; i = 1,2,...,m \tag{8}$$

Also, they have the common characteristic function

$$\varphi(t) \equiv E[e^{it\bar{y}_{\nu_i i}}] = E[\prod_{j=1}^{\nu_i} \varphi_j^{(\nu_i)}(t)], \tag{9}$$

for $i \geq 1$, where $\varphi_j^{(\nu_i)}(t)$ is the characteristic function of $y_j^{(\nu_i)}$, the j$^{th}$ order statistic with set size $\nu_i$. We proposed as an estimator for the population mean μ as

$$\bar{\bar{y}}_{(m)} = \frac{1}{m}\sum_{i=1}^{m} \bar{y}_{\nu_i i}. \tag{10}$$

It is not difficult to show that $\bar{\bar{y}}_{(m)}$ is an unbiased estimator for the population mean μ with a variance

$$V(\bar{\bar{y}}_{(m)}) = \frac{1}{m}E[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}]. \tag{11}$$

To compare the proposed estimator $\overline{\overline{y}}_{(m)}$ with a similar estimator in the usual RSS case where $\nu_1 = \nu_2 = \ldots = \nu_m = k$, we have again to make comparison between $E[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}]$ and $\sigma^2_{(k)}$. For example, this comparison may use Proposition 1 in the case when the random variables $\nu_i, i \geq 1$ have common discrete uniform distribution.

## 4. RANDOM NUMBER OF CYCLES WITH FIXED SET SIZE

Assume now that the set size is fixed and equal to k. From the usual (fixed set size) RSS, we have that $\overline{y}_{(k)}$ is an unbiased estimator of the population mean μ. Let the number of replications θ be a random variable taking values from Λ and independent of $\overline{y}_{(k)}$. We consider

$$\overline{\overline{y}}_{(k)} = \frac{1}{\theta}\sum_{i=1}^{\theta}\overline{y}_{(k)i} \tag{12}$$

as an estimator for μ. Since θ and $\overline{y}_{(k)i}$ are independent it is easy to show that $\overline{\overline{y}}_{(k)}$ is an unbiased estimator for μ with variance

$$V(\overline{\overline{y}}_{(k)}) = \frac{\sigma^2_{(k)}}{k}E[\frac{1}{\theta}]. \tag{13}$$

**Example 4.** Let us assume that θ has a uniform distribution on the set {2, 3,…, m}. Then the variance of $V(\overline{\overline{y}}_{(k)})$ is given by

$$V(\overline{\overline{y}}_{(k)}) = \frac{\sigma^2_{(k)}}{k(m-1)}\sum_{j=2}^{m}\frac{1}{j}. \tag{14}$$

Let $\bar{\bar{y}}_{(r)}$ denote the estimator of the RSS method with r replications. Then the variance of $\bar{\bar{y}}_{(r)}$ is given $V(\bar{\bar{y}}_{(r)}) = \frac{\sigma^2_{(k)}}{kr}$. We can compare the variance of $\bar{\bar{y}}_{(k)}$, which is given in equation (14) with the variance of $\bar{\bar{y}}_{(r)}$. We can see that the proposed estimator has smaller variance if $\frac{1}{m-1}\sum_{j=2}^{m}\frac{1}{j} < \frac{1}{r}$.

## 5. RANDOM NUMBER OF CYCLES WITH RANDOM SET SIZE

Let θ be a random variable independent of $\bar{y}_{v_i i}, i \geq 1$ and taking values from $\Lambda$.

Then we propose

$$\bar{\bar{y}}_{(\theta)} = \frac{1}{\theta}\sum_{i=1}^{\theta}\bar{y}_{v_i i}, \tag{15}$$

as an estimator for μ. Since θ and $\bar{y}_{v_i i}$ are independent we can show that

$$E[\bar{\bar{y}}_{(\theta)}] = E\{E[\frac{1}{\theta}\sum_{i=1}^{\theta}\bar{y}_{v_i i}|\theta]\} = E\{\frac{1}{\theta}\sum_{i=1}^{\theta}E[\bar{y}_{v_i i}]\} = \mu,$$

i.e. $\bar{\bar{y}}_{(\theta)}$ is an unbiased estimator for μ. To find the variance of $\bar{\bar{y}}_{(\theta)}$ we again use the total probability formula and obtain

$$V(\bar{\bar{y}}_{(\theta)}) = E\{E[[\frac{1}{\theta}\sum_{i=1}^{\theta}(\bar{y}_{v_i i} - \mu)]^2|\theta]\} = E\{\frac{1}{\theta^2}\sum_{i,j=1}^{\theta}E[(\bar{y}_{v_i i} - \mu)(\bar{y}_{v_j j} - \mu)]\}.$$

Since $E[\bar{y}_{v_i i}] = E[\bar{y}_{v_j j}] = \mu$, we have

$$V(\bar{\bar{y}}_{(\theta)}) = E[\frac{1}{\theta}]E[\frac{1}{v_i}\sigma^2_{(v_i)}] \tag{16}$$

**Example 5.** Let $\theta$ be a random variable having a Poisson distribution with parameter $\lambda > 0$ truncated at zero and one, i.e. $P(\theta = m) = P(\xi = m|\xi > 1), m \geq 2$, where $\xi$ is a Poisson random variable of the parameter $\lambda$. In this case we obtain that

$$E[\frac{1}{\theta}] = \frac{1}{e^{\lambda} - 1 - \lambda} T(\lambda),\ T(\lambda) = \sum_{j=2}^{\infty} \frac{\lambda^j}{j!j}.$$

Using the fact that $T'(\lambda) = \frac{1}{\lambda}(e^{\lambda} - 1 - \lambda)$, we obtain the variance as

$$V(\overline{\overline{y}}_{(\theta)}) = E[\frac{1}{v_i}\sigma^2_{(v_i)}] \frac{1}{e^{\lambda} - 1 - \lambda} \int_0^{\lambda} \frac{e^u - 1 - u}{u} du.$$

**Example 6.** Let $\theta$ have a uniform distribution on the set {2, 3,..., m}. In this case we have

$$V(\overline{\overline{y}}_{(\theta)}) = \frac{1}{m-1} \sum_{j=2}^{m} \frac{1}{j} E[\frac{1}{v_i}\sigma^2_{(v_i)}].$$

If in addition, the random variables $v_i, i \geq 1$ also have a common uniform distribution on the set {2, 3, ..., N} as in example 3, we obtain

$$V(\overline{\overline{y}}_{(\theta)}) = A_m B_N, \tag{17}$$

where $A_m = \frac{1}{m-1} \sum_{j=2}^{m} \frac{1}{j}$ and $B_N = \frac{1}{N-1} \sum_{\ell=2}^{N} \frac{1}{\ell} \sigma^2_{(\ell)}$.

For comparison we consider the RSS unbiased estimator $\overline{\overline{y}}_{(r)}$ with fixed set size k, 2 ≤ k ≤ N and number of replication r, 2 ≤ r ≤ m. The variance of $\overline{\overline{y}}_{(r)}$ is given by

$V(\overline{\overline{y}}_{(r)}) = \frac{\sigma^2_{(k)}}{kr}$. Thus we have to compare $A_m$ with 1/r and $B_N$ with $\frac{\sigma^2_{(k)}}{k}$. In the latter case the comparison is based on Proposition 1. The following proposition is helpful in comparing the RRSS with the usual RSS method

**Proposition 2.** Let $\varepsilon_r = A_m - r^{-1}, r = 2,3,...,m$, then

(i) $\varepsilon_r > 0$ for $r > \dfrac{m^2+m}{2m+1}$;

(ii) $\varepsilon_r < 0$ for $r < \dfrac{\sqrt{2m^3+2m^2+1}-2m-1}{m-2}$.

To prove the proposition we consider the sum

$$(m-1)\varepsilon_r = \sum_{j=2}^{r}\frac{r-j}{jr} + \sum_{j=r+1}^{m}\frac{r-j}{jr} = I_1 + I_2 .$$

It is not difficult to see that

$$I_1 > \frac{1}{r^2}\sum_{j=1}^{r}(r-j) = \frac{1}{2}, \text{ and } I_2 > \frac{r-m-1}{2r(r+1)}.$$

If we use the opposite inequalities, then we obtain

$$I_1 < \frac{r}{4}, \text{ and } I_2 < \frac{r-m-1}{2mr}(m-r).$$

Using the bound for $I_1$ and $I_2$ in the previous equality we can obtain bounds for r.

The efficiency of the random ranked set sampling (RRSS) with random set size and number of replications with respect to RSS with set size k and r replications may be defined as

$$\tau(k,r) = \frac{\sigma_{(k)}^2 / rk - A_m B_N}{\sigma_{(k)}^2 / rk} = \frac{\sigma_{(k)}^2 - rkA_m B_N}{\sigma_{(k)}^2}.$$

Evaluation of function $\tau(k,r)$ for different probability distributions will be considered in Section 7.

## 6. ASYMPTOTIC PROPERTIES

In this section we will prove that under the very natural assumptions the estimator $\bar{\bar{y}}_{(\theta)}$ is asymptotically normal. Recall that $\bar{\bar{y}}_{(\theta)} = \frac{1}{\theta}\sum_{i=1}^{\theta}\bar{y}_{v_i i}$ and $\bar{y}_{v_i i} = \frac{1}{v_i}\sum_{j=1}^{v_i} y_j^{(v_i)}$. As mentioned before $\bar{y}_{v_i i}, i \geq 1$ are independent and identically distributed random variables such that $E[\bar{y}_{v_i i}] = \mu$, $B^2 \equiv V(\bar{y}_{v_i i}) = E[\frac{1}{v_i}\sigma^2_{(v_i)}]$ and have characteristic function given in equation (9). Since $\sigma^2_{(n)} > \sigma^2_{(n+1)}$, $n \geq 1$ and $v_i$ are random variables taking values from $\Lambda$, we find that $B^2 \leq E[\sigma^2_{(v_i)}] < \sigma^2_{(1)} = \sigma^2$. Thus, if the initial distribution of $X^{(\ell)}_{ij}$ has a finite variance, then $B^2 < \infty$.

**Theorem.** If $\sigma^2 < \infty$ and $\theta \to \infty$ in probability then

$$P(B^{-1}\sqrt{\theta}\,(\bar{\bar{y}}_{(\theta)} - \mu) \leq x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

**Proof.** Let $W_\theta = B^{-1}\sqrt{\theta}\,(\bar{\bar{y}}_{(\theta)} - \mu)$, then by total probability formula we have

$$E[e^{itw_\theta}] = \sum_{\ell=1}^{\infty} E\{\exp[\frac{it}{B\sqrt{\ell}}\sum_{i=1}^{\ell}(\bar{y}_{v_i i} - \mu)]\}P(\theta = \ell) = E\{[\bar{\varphi}(\frac{t}{B\sqrt{\theta}})]^{\theta}\},$$

where $\bar{\varphi}(t) = E\{\exp[it(\bar{y}_{v_i i} - \mu)]\}$. Since $B^2 < \infty$, we can write following representation for $\bar{\varphi}(t)$:

$$\bar{\varphi}(t) = 1 + itE(\bar{y}_{v_i i} - \mu) - \frac{t^2 B^2}{2} + t^2 \varepsilon(t), \qquad (18)$$

where $\varepsilon(t) \to 0$ as $t \to 0$. Then if $\theta \to \infty$ in probability, then $\varepsilon(\theta^{-1/2}) \to 0$ in probability. In fact for any $\delta > 0$ there is a $t_o > 0$ such that $|\varepsilon(t)| < \delta$ when $|t| < t_o$. Thus

$$P\{|\varepsilon(1/\sqrt{\theta})| > \delta\} = P\{|\varepsilon(1/\sqrt{\theta})| > \delta, 1/\sqrt{\theta} \leq t_o\} + P\{|\varepsilon(1/\sqrt{\theta})| > \delta, 1/\sqrt{\theta} > t_o\}. \quad (19)$$

It is easy to see that the first probability on the right side of equation (19) is equal to zero and the second is less than $p(\sqrt{\theta} < 1/t_o)$ which tends to zero when $\theta \xrightarrow{p} \infty$. Using equation (18) and the simple formula $\ln(\alpha) = \alpha - 1 + o(\alpha - 1)$, $\alpha \to 1$, we obtain that

$$\theta \ln \bar{\varphi}(\frac{t}{B\sqrt{\theta}}) = -\frac{t^2}{2} + \gamma(\theta)$$

where $\gamma(\theta) \xrightarrow{P} 0$ as $\theta \xrightarrow{P} \infty$. Consequently

$$[\overline{\varphi}(\frac{t}{B\sqrt{\theta}})]^{\theta} \xrightarrow{P} e^{-t^2/2}, \tag{20}$$

as $\theta \xrightarrow{P} \infty$. Since $\left|\overline{\varphi}(\frac{t}{B\sqrt{\theta}})\right| \leq 1$, by the dominated convergence theorem (see Shiryaev, (1996), Theorem 3, p 187and remark on p 258). We conclude from (20) that

$$E[\overline{\varphi}(\frac{t}{B\sqrt{\theta}})]^{\theta} \rightarrow e^{-t^2/2}$$

i.e. $B^{-1}\sqrt{\theta}\,(\overline{\overline{y}}_{\theta} - \mu)$ is asymptotically normal as $\theta \xrightarrow{P} \infty$.

## 7. EXAMPLES

In this section we will consider comparing the RRSS with the RSS for estimating the population mean μ if the parent distribution is known to be normal, exponential, double exponential or logistic. Also, we are assuming that the set size ν is a uniform random variable defined on the set [2, 3, …, N] and the number of replications θ is following a uniform distribution on the set [2, 3, …, m].

We calculate the value of τ, $2 \leq \tau \leq N$ as suggested by Proposition 1, which will give $V(\overline{y}_{(k)}) > V(\overline{y}_{(\nu)})$ for different parent distributions. Table 1 shows the values of $\tau$ with the corresponding $V(\overline{y}_{(\nu)})$ as if $\tau = \text{N}$ with the values of $V(\overline{y}_{(k)})$ for set size k = 2, 3, 4, 5 for the above probability distributions. It is clear that the RRSS will do better than the RSS with set size k=3, for example if $\tau = 5$ for most of the distributions considered in this study.

**Table 1**

The value of the efficiency $\tau(k,r)$ of RRSS with respect to RSS is evaluated for k = 3, 5, r =3, 5, N =10, 15 and m = 10, 20. Table 2 shows different values $\tau(k,r)$ for the normal, exponential, double exponential and logistic distributions. We observe that the RRSS improves the efficiency of estimating the population mean if the values of N and m are moderately large. For example, if N = m =10 and r = k = 3, the RRSS is about 66% more efficient than the RSS.

**Table 2**

## 8. CONCLUDING COMMENTS

In this paper we have considered the case of random set size and/or random number of replications. The reason for considering such a method is to resolve the problem of unfixed number of units that we might come cross in real life problems like the line intercept (transect) example given in Section 1. It has been shown that under certain conditions the efficiency of the estimator of the population mean may be improved by using RRSS instead of RSS. The following conclusions are drawn:

1. In the case of single cycle with random set size we might be able to improve the efficiency of the estimator of the population mean by using RRSS instead of RSS by choosing the suitable distribution for the set size. The result of Table 1 confirms this fact in the case of choosing the discrete uniform distribution.
2. If the set size is fixed and the number of replications is random we can easily show that the RRSS is more efficient than RSS, if the number of replications are following the discrete uniform distribution.

3. The results of Table 2 show that in the case of random set size $\nu$ and number of cycles $\theta$, the efficiency is substantially increased if the underlying for both $\nu$ and $\theta$is discrete uniform.

The final recommendation is to use RRSS to handle some practical problems and/or to increase the efficiency of the estimator of the population mean.

## REFERENCES


Bohn, L. L. and Wolfe, D. A. (1992), "Nonparametric Two-Sample Procedures for Ranked Set Samples Data," *Journal of American Statistical Association*, 87, 552-561.

------ (1994), "The Effect of Imperfect Judgment Rankings on Properties of Procedures Based on Ranked-Set Samples analog of the Mann-Whitney-Wilcoxon Statistic," *Journal of American Statistical Association*, 89, 168-176.

Hartlaub, B. A. and Wolfe, D. A. (1999), "Distribution-Free Ranked Set Sample Procedures for Umbrella Alternatives in the m-Sample Setting," *Environmental and Ecological Statistics*, 6, 105-118.

Huang, J. (1997), "Asymptotic Properties of the NPMLE of a Distribution Function Based on Ranked Set Samples," *The Annals of Statistics*, 25(3), 1036-1049.

Kaur, A., Patil, G. P., Sinha, B. K. and Taillie, C. (1995), "Ranked Set Sampling: an Annotated Bibliography," *Environmental and Ecological Statistics*, 2, 25-54.

Kim, Y. H and Arnold, B. C. (1999a), "Estimation of a Distribution Function under Generalized Ranked Set Sampling," *Commun. Statist. –Simula,* 28(3), 657-666.

------ (1999b), "Parameter Estimation under Generalized Ranked Set Sampling," *Statistics & Probability Letters*, 42, 353-360.

Koti, K. M. and Babu, G. J. (1996), "Sign Test for Ranked-Set Sampling," *Commun. Statist. Theory Meth.* 25(7), 1617-1630.

Kvam, P. H. and Samaniego, F. J. (1993), "On the Inadmissibility of Empirical Averages as Estimators in Ranked Set Sampling," *Journal of Statistical Planning and Inference*, 36, 39-55.

------- (1994) "Nonparametric Maximum Likelihood Estimation based on Ranked Set Samples," *Journal of American Statistical Association*, 89, 526-537.

Li, D. Sinha, B. K. and Perron, F. (1999), "Random Selection in Ranked Set Sampling and its Applications," *Journal of Statistical Planning and Inference*, 76, 185-201.

McIntyre, G. A. (1952), "A Method of Unbiased Selective Sampling, using Ranked Sets," *Australian Journal of Agricultural Research*, 3, 385-390.

Muttlak, H. A. and McDonald, L. L. (1992), "Ranked Set Sampling and the Line Intercept Method: A more Efficient procedure," *The Biometrical Journal*, 34, 329-46.

Muttlak, H. A. (1997), "Median Ranked Set Sampling, " *Journal of Applied Statistical Science*, 6(4), 245-55.

Özturk, Ö. (2000), "Alternative Ranked Set Sampling Protocols for Sign Test," *Statistics & Probability Letters*, 47, 15-23

Patil, G. P., Sinha, A. K. and Taillie, C. (1994), "Ranked Set Sampling," *A Handbook of Statistics*, vol. 12, pp167-200.

----- (1999), "Ranked Set Sampling: A Bibliography," *Environmental and Ecological Statistics*, 6, 91-98.

Presnell, B. and Bohn, L. L. (1999), "U-Statistics and Imperfect Ranking in Ranked Set Sampling," *Nonparametric Statistics,* 10, 111-126.

Samawi, H. M., Ahmed, M. S. and Abu-Dayyeh (1996), "Estimating the Population Mean using Extreme Raked Set Sampling," *The Biometrical Journal*, 38(5), 577-86.

Shiryaev, A. N. (1996), *Probability, 2end*: Springer, New York.

Stokes, S. L. and Sager, T. W. (1988), "Characterization of a Ranked-Set Sample with Application to Estimating Distribution Function," *Journal of American Statistical Association*, 83, 374-381.

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,* 21, 249-55.

Table 1. The value of the random set size $\tau$ along the corresponding variance $V(\bar{y}_{(v)})$ of RRSS as if $N = \tau$ and the RSS variance $V(\bar{y}_{(k)})$ for different set size k and different probability distributions.

| Distribution | | k = 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Normal | $V(\bar{y}_{(k)})$ | 0.3408 | 0.1742 | 0.1065 | 0.0722 |
| | $\tau$ | 3 | 5 | 9 | 14 |
| | $V(\bar{y}_{(v)})$ | 0.2575 | 0.1734 | 0.1053 | 0.0708 |
| Exponential | $V(\bar{y}_{(k)})$ | 0.3750 | 0.2037 | 0.1303 | 0.0913 |
| | $\tau$ | 3 | 5 | 9 | 13 |
| | $V(\bar{y}_{(v)})$ | 0.2894 | 0.2001 | 0.1248 | 0.0911 |
| Logistic | $V(\bar{y}_{(k)})$ | 0.3480 | 0.1814 | 0.1128 | 0.0776 |
| | $\tau$ | 3 | 4 | 9 | 14 |
| | $V(\bar{y}_{(v)})$ | 0.2647 | 0.2141 | 0.1104 | 0.0748 |
| Double Exponential | $V(\bar{y}_{(k)})$ | 0.7368 | 0.3854 | 0.2453 | 0.1719 |
| | $\tau$ | 3 | 5 | 9 | 14 |
| | $V(\bar{y}_{(v)})$ | 0.5611 | 0.3848 | 0.2385 | 0.1630 |

Table 2. The efficiency of the RRSS $\tau(k,r)$ with respect to RSS for different probability distributions.

| N | m | k = 3, r = 3 | k = 3, r = 5 | k = 5, r = 3 | k = 5, r = 5 |
|---|---|---|---|---|---|
| | | **Normal** | | | |
| 10 | 10 | 0.655 | 0.410 | 0.146 | - 0.423 |
| | 20 | 0.774 | 0.624 | 0.455 | 0.092 |
| 15 | 10 | 0.755 | 0.591 | 0.408 | 0.014 |
| | 20 | 0.844 | 0.739 | 0.623 | 0.371 |
| | | **Exponential** | | | |
| 10 | 10 | 0.640 | 0.399 | 0.196 | - 0.340 |
| | 20 | 0.770 | 0.618 | 0.487 | 0.145 |
| 15 | 10 | 0.746 | 0.577 | 0.435 | 0.058 |
| | 20 | 0.838 | 0.730 | 0.639 | 0.399 |
| | | **Logistic** | | | |
| 10 | 10 | 0.643 | 0.404 | 0.165 | - 0.392 |
| | 20 | 0.772 | 0.620 | 0.467 | 0.112 |
| 15 | 10 | 0.751 | 0.585 | 0.418 | 0.029 |
| | 20 | 0.841 | 0.735 | 0.628 | 0.381 |
| | | **Double Exponential** | | | |
| 10 | 10 | 0.636 | 0.394 | 0.184 | - 0.361 |
| | 20 | 0.768 | 0.613 | 0.479 | 0.132 |
| 15 | 10 | 0.744 | 0.574 | 0.426 | 0.044 |
| | 20 | 0.837 | 0.728 | 0.634 | 0.390 |