Investigating the Estimation of the Population Mean Using Random Ranked Set Samples

I. Rahimov and H. A. Muttlak

Department of Mathematical Sciences,
King Fahd University of Petroleum & Minerals,
Dhahran 31261, Saudi Arabia

ABSTRACT

Ranked set sampling (RSS) 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 suggest several unbiased estimators of the population mean using random ranked set sampling (RRSS) method in case of perfect ranking and imperfect ranking. The efficiencies of the estimators of the population mean under RRSS and RSS are compared in the case of perfect ranking. The results show, under certain conditions, the efficiency of estimators is improved by using RRSS in the case of perfect ranking. Finally the asymptotic properties of the proposed estimators of the population mean are discussed for prefect and imperfect ranking.

Key words: Asymptotic properties, discrete uniform distribution, efficiency, errors in ranking, random number of replications, random set size, ranked set sampling, and relative precision.

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 that 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 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 results of this study will some value in practice in ecological, rang management and/or environmental sampling. For example if we are doing sampling for species diversity index, we may not have a fixed size of a given region. Also the example considered by Muttlak and McDonald (1992) demonstrates the need for at least r to be random.

Takahasi and Wakimoto (1968) supplied mathematical theory to support McIntyre's (1952) suggestion. Dell and Clutter (1972) showed that errors in ranking reduce the efficiency of the RSS mean relative to the simple random sampling (SRS) mean. However, the RSS mean remains unbiased and more efficient than the SRS mean unless the ranking is so poor as to yield a random sample.

Li et al. (1999) considered the random selection in the process of using the RSS method, and discussed several results from both nonparametric and parametric view. The method is consisting of randomly selecting n < k order statistics with n distinct indices from (1, …, k) sets of size k units each, where k is the set size in our notations.

For classified and extensively reviewed work in the area of RSS from 1952 to 1994 see Patil et al (1994). 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). Sections 2-6 are devoted for the case of perfect ranking. In Section 2 the idea of RRSS is introduced in the case of one replication i.e. single cycle. The general case of RRSS with random set sizes and replications is considered in Section 3. The asymptotic properties of the estimator of the population mean suggested for the general case of RRSS are established in Section 4. In Section 5 we calculate the efficiency of the newly suggested estimators of the population mean for specific probability distributions and compare these to the RSS estimators. Some concluding remarks are given in this section for the case of perfect ranking.

The general case of RRSS with errors in ranking with random set size and number of replications is considered in Section 6. In last Section the asymptotic properties of the estimator of the population mean based on the RRSS with errors in ranking are established.

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 μ and variance σ^2. Let ν be a random variable taking values from $\Lambda = \{2, 3, ...\}$. For i = 1, 2, .., ℓ, let $X_{i(1)}^{(\nu)} \le X_{i(2)}^{(\nu)} \le ... \le X_{i(\nu)}^{(\nu)}$ 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

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

as an estimator of the population mean μ. Assume from now on that the random variable ν 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

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

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

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

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

We can easily proof (a) by using the total probability formula and equation (1). For any ℓ 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

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

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 (b). Again by using the total probability formula we can write

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

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

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

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 the last equation 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 5 we will consider different concrete distributions to obtain the value τ.

3. Random Number of Replications with Random Set Size

Let $\nu_1, \ \nu_2, \ ...$ be independent and identically distributed random variables taking values from Λ, θ be a random variable taking values from Λ independent of the families $\{\nu_i, i \ge 1\}$ and $\{X_{ij}, i, j \in \mathbb{N}\}$. If the process of Section 2 is repeated θ times, i.e. we replicate the cycle θ times with set sizes ν_i, where i = 1, 2,…, θ, we will obtain a sequence of estimators $\bar{y}_{\nu_1 1}, \ \bar{y}_{\nu_2 2}, \ ..., \ \bar{y}_{\nu_\theta \theta}$. It is clear that $\bar{y}_{\nu_i i}, i \ge 1$ are independent and identically distributed with the following mean and variance respectively

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

Also, they have the common characteristic function

$$\varphi(t) \equiv E[e^{it\bar{y}_{\nu_i i}}] = E[\prod_{j=1}^{\nu_i}\varphi_j^{(\nu_i)}(t)] \tag{3}$$

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 ν_i. We propose

$$\bar{\bar{y}}_{(\theta)} = \frac{1}{\theta}\sum_{i=1}^{\theta}\bar{y}_{\nu_i i}$$

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

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

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

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

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

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

Thus we conclude that

I. $\bar{\bar{y}}_{(\theta)}$ is an unbiased estimator for the population mean μ;

II. $\bar{\bar{y}}_{(\theta)}$ has the variance

$$V(\bar{\bar{y}}_{(\theta)}) = E[\frac{1}{\theta}]E[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}]. \tag{4}$$

Now we consider some particular cases.

Example 1. Let θ has a discrete uniform distribution on the set {2, 3,…, m}. In this case we have

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

If in addition, the random variables $\nu_i, i \geq 1$ also have a common uniform distribution on the set {2, 3,…, N}, we can express $V(\overline{\overline{y}}_{(\theta)})$ as

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

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, $2 \leq k \leq N$ and number of replication r, $2 \leq r \leq m$. 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 $r > \frac{m^2 + m}{2m+1}$;

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

Prove of the proposition is not difficult.

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

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

Evaluation of function $\tau(k,r)$ for different probability distributions will be considered in Section 5.

Let now the random variable θ has a degenerate distribution with the atom of the mass 1 at point $m \in \Lambda$ i.e. has the jump of size 1 at point $m \in \Lambda$. It is clear that in this case we have a scheme with a fixed number of cycles with random set size. Obviously it follows from (I) and (II) that $\overline{\overline{y}}_{(m)}$ is an unbiased estimator for μ and has a variance

$$V(\overline{\overline{y}}_{(m)}) = \frac{1}{m} E\,[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}]. \tag{5}$$

To compare the proposed estimator $\overline{\overline{y}}_{(m)}$ with a similar estimator in the usual RSS case where $\nu_1 = \nu_2 = ... = \nu_m = k$, 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.

Now let the random variables $\nu_1, \nu_2, ...$ have a degenerate distribution at a point $k \in \Lambda$. This corresponds to the scheme of random number of cycles with fixed set size k. It follows from (I) and (II) that $\overline{\overline{y}}_{(\theta)}$ is an unbiased estimator with variance

$$V(\overline{\overline{y}}_{(\theta)}) = \frac{\sigma^2_{(k)}}{k} E[\frac{1}{\theta}] . \qquad (6)$$

Example 2. Let us assume that θ has a discrete uniform distribution on the set {2, 3,…, m}. Then the variance of $V(\overline{\overline{y}}_{(m)})$ is given by $V(\overline{\overline{y}}_{(m)}) = \frac{\sigma^2_{(k)}}{k \ (m-1)} \sum_{j=2}^{m} \frac{1}{j}$.

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

4. Asymptotic Properties

In this section we will prove that under the very natural assumptions the estimator $\overline{\overline{y}}_{(\theta)}$ is asymptotically normal. Recall that $\overline{\overline{y}}_{(\theta)} = \frac{1}{\theta}\sum_{i=1}^{\theta} \overline{y}_{\nu_i i}$ and $\overline{y}_{\nu_i i} = \frac{1}{\nu_i}\sum_{j=1}^{\nu_i} y_j^{(\nu_i)}$. As mentioned before $\overline{y}_{\nu_i i}, i \geq 1$ are independent and identically distributed random variables such that $E[\overline{y}_{\nu_i i}] = \mu$, $B^2 \equiv V(\overline{y}_{\nu_i i}) = E[\frac{1}{\nu_i}\sigma^2_{(\nu_i)}]$ and have characteristic function given in equation (3). Since $\sigma^2_{(n)} > \sigma^2_{(n+1)}$, n ≥ 1 and ν_i are random variables taking values from Λ, we find that $B^2 \le E[\sigma^2_{(\nu_i)}] < \sigma^2_{(1)} = \sigma^2$. Thus, if the initial distribution of $X^{(\ell)}_{ij}$ has a finite variance, then $B^2 < \infty$.

Theorem 1. If $\sigma^2 < \infty$ and $\theta \to \infty$ in probability then

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

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

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

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

$$\overline{\varphi}(t) = 1 + itE(\overline{y}_{\nu_i i} - \mu) - \frac{t^2B^2}{2} + t^2\varepsilon(t),$$

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 δ >0 there is a t_o > 0 such that $|\varepsilon(t)| < \delta$ when $|t| < t_o$. Thus

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

It is easy to see that the first probability on the right side of last equation 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 $\overline{\varphi}(t)$ and the simple formula $\ln(\alpha) = \alpha - 1 + o(\alpha - 1)$, $\alpha \to 1$, we obtain that

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

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

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

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

$E[\overline{\varphi}(\frac{t}{B\sqrt{\theta}})]^\theta \to e^{-t^2/2}$ i.e. $B^{-1}\sqrt{\theta}\,(\overline{\overline{y}}_\theta - \mu)$ is asymptotically normal as $\theta \xrightarrow{P} \infty$.

5. Examples and Concluding Remarks

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 discrete uniform random variable defined on the set [2, 3,…, N] and the number of replications θ is following a discrete uniform distribution on the set [2, 3,…, m].

We calculate the value of τ, $2 \le \tau \le N$ as suggested by Proposition 1, which will give $V(\bar{y}_{(k)}) > V(\bar{y}_{(\nu)})$ for different parent distributions. Table 1 shows the values of τ with the corresponding $V(\bar{y}_{(\nu)})$ as if τ = N with the values of $V(\bar{y}_{(k)})$ 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 τ = 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

In the previous we have considered the case of random set size and/or random number of replications in case of perfect ranking. 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. 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 ν and number of cycles θ, the efficiency is substantially increased if the underlying for both ν and θ is discrete uniform distribution.

6. Random Number of Replications with Random Set Size In Case of Imperfect Judgment Ranking

In the previous sections we considered properties of estimators of population mean μ based on random ranked set sampling with perfect ranking i.e. there are no errors in ranking the units. But some times we cannot rank the unit without errors in ranking and we have to use our judgment to rank the units. In the remaining sections we will consider properties of estimators of population mean based on random ranked set sampling with imperfect ranking.

Let $X_{11}, X_{12}, ..., X_{1n}; X_{21}, X_{22}, ..., X_{2n}; ....; X_{i1}, X_{i2}, ..., X_{in_i};...; X_{n1}, X_{n2}$ ..., X_{nn} are independent and identically distributed random variables with cdf F (x), pdf f(x), mean, μ and variance σ^2. Let $\theta, v_1, v_2, ...$ be independent random variables taking values from $\Lambda = \{2, 3, ...\}$ and independent of the random variables $X_{ij}, i, j \geq 1$. First we randomly select v_1 sets of size θ units, where $v_1 \in \Lambda$. We order the units within each set by judgment order i.e. there are errors in ordering the units within each set. Let $X_{i[1]}, X_{i[2]}, ..., X_{i[\theta]}$, i = 1, 2,…, v_1 be the judgment order statistics of $X_{i1}, X_{i2}, ..., X_{i\theta}$ i = 1, 2,…, v_1, which are written as $X_{i[i]}$ to distinguish from the actual order statistics $X_{i(i)}$. We select for measurement the first judgment order statistics $X_{1[1]}^{(1)}, X_{2[1]}^{(1)}, ..., X_{v_1[1]}^{(1)}$ and let $\overline{y}_{v_1}^{(\theta)} = \frac{1}{v_1}\sum_{i=1}^{v_1} X_{i[1]}^{(1)}$. Now we randomly select another v_2 sets of size θ units each and order them by judgment order $X_{i[1]}, X_{i[2]}, ..., X_{i[\theta]}$, i = 1, 2,…,

v_2. We select the second judgment order statistics for measurement $X^{(2)}_{1[2]}$, $X^{(2)}_{2[2]}$, ..., $X^{(2)}_{v_1[2]}$ and denote $\overline{y}^{(\theta)}_{v_2} = \frac{1}{v_2}\sum_{i=1}^{v_2} X^{(2)}_{i[2]}$. We repeat this process θ time to get $\overline{y}^{(\theta)}_{v_1}$, $\overline{y}^{(\theta)}_{v_2}$, ..., $\overline{y}^{(\theta)}_{v_n}$. We propose the following as an estimator for the population mean μ

$$\bar{\tilde{y}}_{\theta} = \frac{1}{\theta}\sum_{i=1}^{\theta} \overline{y}^{(\theta)}_{v_i}$$

The basic properties of the estimator $\bar{\tilde{y}}_{\theta}$ are:

(i) $\bar{\tilde{y}}_{\theta}$ is unbiased estimator for of population mean μ with variance

(ii) $Var(\bar{\tilde{y}}_{\theta}) = E[\frac{1}{\theta^2}\sum_{i=1}^{\theta}\sigma^2_{[i]\theta}E(\frac{1}{v_i})]$, where $\sigma^2_{[i]\theta} = Var(X^{(\theta)}_{[i]})$.

To prove the above two properties, we need the following results. If $X^{(n)}_{[r]}$ is the r^th^ judgment order statistic with sample size n, then it is clear that $X^{(n)}_{[r]} \overset{d}{=} X^{(n)}_{i[r]}$. If $f^{(n)}_{[r]}(x)$ is the pdf of $X^{(n)}_{[r]}$, then as noted by Dell and Clutter (1972)

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} f^{(n)}_{[i]}(x). \tag{7}$$

It follows from (7) that, if we denote $\mu = E(X)$, $Var(X) = \sigma^2$, $\mu^{(n)}_{[i]} = E(X^{(n)}_{[i]})$ and $\sigma^2_{[i]n} = Var(X^{(n)}_{[i]})$, then $\sum_{i=1}^{n}\mu^{(n)}_{[i]} = n\mu$. Since $X^{(n)}_{i[r]}$, i = 1, 2,…, v_r are i i d and $E(X^{(n)}_{i[r]}) = \mu_{[r]}$, then

$$E(\overline{y}^{(n)}_{v_i}) = E[E[\frac{1}{v_i}\sum_{j=1}^{v_i} X^{(n)}_{j[i]}|v_i]] = E[\frac{1}{v_i}\sum_{j=1}^{v_i}\mu^{(n)}_{[i]}] = \mu^{(n)}_{[i]}. \tag{8}$$

Using the above results and the total probability formula we have

$$Var(\overline{y}^{(n)}_{v_i}) = \sum_{\ell=2}^{\infty}\frac{1}{\ell^2}E[\sum_{j=1}^{\ell}(X^{(n)}_{j[i]} - \mu^{(n)}_{[i]})]^2 P(v_i = \ell)$$

$$Var(\overline{y}^{(n)}_{v_i}) = \sum_{\ell=2}^{\infty}\frac{1}{\ell^2}\sum_{j=1,m=1}^{\ell}E[X^{(n)}_{j[i]} - \mu^{(n)}_{[i]}][X^{(n)}_{m[i]} - \mu^{(n)}_{[i]}]P(v_i = \ell)$$

$$Var(\overline{y}_{\nu_i}^{(n)}) = \sum_{\ell=2}^{\infty} \frac{1}{\ell^2} \sum_{j=1}^{\ell} \sigma_{[i]n}^2 P(\nu_i = \ell) = \sigma_{[i]n}^2 E(\frac{1}{\nu_i}) \qquad (9)$$

We have from (8) that $E(\overline{y}_{\nu_i}^{(\theta)}) = E[\mu_{[i]}^{(\theta)}]$. Using this fact and the total probability formula we can show

$$E(\tilde{\overline{y}}_\theta) = \sum_{\ell=2}^{\infty} \frac{1}{\ell} \sum_{i=1}^{\ell} \mu_{[i]}^{(\ell)} P(\theta = \ell) = \mu ,$$

i.e. $\tilde{\overline{y}}_\theta$ is unbiased estimator for μ. Now we evaluate the variance of $\tilde{\overline{y}}_\theta$. Since θ, $\nu_1, \nu_2, \ldots$ are independent, using the fact that $\sum_{i=1}^{\ell} \mu_{[i]}^{(\ell)} = \ell\mu$, and the total probability formula, we have

$$Var(\tilde{\overline{y}}_\theta) = E[\tilde{\overline{y}}_\theta - \mu]^2 = E[\frac{1}{\theta} \sum_{i=1}^{\theta} (\overline{y}_{\nu_i}^{(\theta)} - \mu_{[i]}^{(\theta)})]^2$$

$$Var(\tilde{\overline{y}}_\theta) = \sum_{\ell=2}^{\infty} \frac{1}{\ell^2} \sum_{i,j=1}^{\ell} E[(\overline{y}_{\nu_i}^{(\ell)} - \mu_{[i]}^{(\ell)})(\overline{y}_{\nu_j}^{(\ell)} - \mu_{[j]}^{(\ell)})] P(\theta = \ell)$$

$$Var(\tilde{\overline{y}}_\theta) = \sum_{\ell=2}^{\infty} \frac{1}{\ell^2} \sum_{i=1}^{\ell} Var(\overline{y}_{\nu_i}^{(\ell)}) P(\theta = \ell)$$

Using equation (9), we can write the variance of $\tilde{\overline{y}}_\theta$ as in (ii).

Now we consider some particular cases. Let first θ has a degenerate distribution at point $n \in \Lambda$. Then we have the scheme with fixed set size. It is clear that $\tilde{\overline{y}}_n$ is an unbiased estimator of μ and we obtain from (ii) that

$$Var(\tilde{\overline{y}}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_{[i]n}^2 E(\frac{1}{\nu_i}) \qquad (10)$$

Example 3. Let $P(\nu_i = k) = \dfrac{1}{n-1}$, k = 2, 3,…, n. In this case we find that

$$Var(\tilde{\overline{y}}_n) = \frac{1}{n^2(n-1)} \sum_{\ell=2}^{n} \frac{1}{\ell} \sum_{i=1}^{n} \sigma_{[i]n}^2 ,$$

Since from equation (7) it follows that

$$\sum_{i=1}^{n} \sigma_{[i]n}^2 = n\sigma^2 - \sum_{i=1}^{n} (\mu_{[i]}^{(n)} - \mu)^2 , \qquad (11)$$

we have an alternative representation of

$$Var(\overline{\overline{y}}_n) = \frac{1}{n(n-1)} \sum_{\ell=2}^{n} \frac{1}{\ell} \; [\sigma^2 - \; \frac{1}{n} \; \sum_{i=1}^{n} (\mu_{[i]}^{(n)} - \mu)^2] . \tag{12}$$

To compare $\overline{\overline{y}}_n$ as an estimator for the population mean to the estimator proposed by Dell and Clutter (1972) which is denoted by $\hat{\mu}_{rss} = \frac{1}{n} \sum_{i=1}^{n} \overline{X}_{[i]}$, $\overline{X}_{[i]} = \frac{1}{k_i} \sum_{j=1}^{k_i} X_{[i]j}$ we need only to compare $Var(\overline{\overline{y}}_n)$ as given above to $Var(\hat{\mu}_{rss}) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_{[i]n}^2 / _{k_i}$. Clearly we need to know the distribution(s) of v_i. As a special case let assume that k_i = k for i=, 1, 2,…, n. In Dell and Clutter scheme the variance of $\hat{\mu}_{rss}$ can be written as $Var(\hat{\mu}_{rss}) = \frac{1}{n^2 k} \sum_{i=1}^{n} \sigma_{[i]}^2$. It follows from the above results that to compare variances of $\overline{\overline{y}}_n$ and of $\hat{\mu}_{rss}$ we need to compare $A_n = \frac{1}{n-1} \sum_{\ell=2}^{n} \frac{1}{\ell}$ with 1/k. Here we can use again proposition 2.

The relative precision of $\overline{\overline{y}}_n$ with respect to $\hat{\mu}_{rss}$ is

$$RP(\overline{\overline{y}}_n , \; \hat{\mu}_{rss}) = \frac{Var(\hat{\mu}_{rss})}{Var(\overline{\overline{y}}_n)} = [\frac{k}{n-1} \sum_{\ell=2}^{n} \frac{1}{\ell}]^{-1} .$$

Now let $v_1, \; v_2, \; ..., v_\theta$ have degenerate distributions on points $k_1, k_2, ..., k_\theta$ respectively. Then we have the scheme with fixed number of replications and with random set size. In this case $\overline{\overline{y}}_\theta$ is still unbiased estimator of μ and we obtain form (ii) its variance to be

$$Var(\overline{\overline{y}}_\theta) = E \;\; [\frac{1}{\theta^2} \;\; \sum_{i=1}^{\theta} \frac{1}{k_i} \sigma_{[i]\theta}^2] . \tag{13}$$

We now compare $\overline{\overline{y}}_n$ to $\hat{\mu}_{srs}$ the simple random sampling (SRS) estimator for μ with sample size nk. Let θ = n be non-random and $v_1, v_2, ...$ be independent and identically distributed random variables taking values from $\Lambda = \{2,3,...\}$. In this case we obtain from (13) that $Var(\overline{\overline{y}}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_{[i]n}^2 E(\frac{1}{v_i})$. Also the variance of $\hat{\mu}_{srs}$ is

$Var(\hat{\mu}_{srs}) = \dfrac{\sigma^2}{nk}$. Using this result and the result of equation (11), we find the relative precision of $\tilde{\bar{\bar{y}}}_n$ with respect to $\hat{\mu}_{srs}$:

$$RP(\tilde{\bar{\bar{y}}}_n,\ \hat{\mu}_{srs}) = \frac{Var(\hat{\mu}_{srs})}{Var(\tilde{\bar{\bar{y}}}_n)} = \{kE(\frac{1}{\nu_i})[1 - \frac{1}{n\sigma^2}\ \sum_{i=1}^{n}(\mu_{[i]}^{(n)} - \mu)^2]\}^{-1}\ . \qquad (14)$$

If ν_i is distributed on {2, 3,…, n} as a discrete uniform random variable, then (14) will become

$$RP(\tilde{\bar{\bar{y}}}_n,\ \hat{\mu}_{srs}) = \{\frac{k}{n-1}\ \sum_{\ell=2}^{n}\frac{1}{\ell}[1 - \frac{1}{n\sigma^2}\ \sum_{i=1}^{n}(\mu_{[i]}^{(n)} - \mu)^2]\}^{-1}\ .$$

Let $\bar{\bar{y}}_{nk}$ be estimator with fixed number k quantification in each cycle, then the relative precision of $\tilde{\bar{\bar{y}}}_n$ with respect to $\bar{\bar{y}}_{nk}$ is $RP = \{kE(\dfrac{1}{\nu_i})\}^{-1}$.

7. Asymptotic Properties In case of Imperfect Judgment Ranking

We now consider the asymptotic properties of the estimator $\tilde{\bar{\bar{y}}}_n$ as $n \to \infty$. We will prove that under some natural conditions $\tilde{\bar{\bar{y}}}_n$ is asymptotically normal. We assume that $\sigma_{[i]n}^2 = Var(X_{[i]}^{(n)})$ and $\beta_{[i]n}^3 = E\left|X_{[i]}^{(n)} - \mu_{[i]}^{(n)}\right|^3$, where $\mu_{[i]}^{(n)} = E[X_{[i]}^{(n)}]$, are finite for each $i, n \geq 1$. Note that in the case of perfect ranking the above assumptions are satisfied if $\sigma^2 = Var(X)$ and $B^3 = E\left|X\right|^3$ are finite. We also need the following conditions

C1. $\lim_{n\to\infty}(nB_n)^{-3}\sum_{i=1}^{n}E[\dfrac{1}{\nu_i^2}]\beta_{[i]n}^3 = 0$, where $B_n^2 = Var(\bar{\bar{y}}_n) = n^{-2}\sum_{i=1}^{n}E[\dfrac{1}{\nu_i}]\sigma_{[i]n}^2$.

C2. $\lim_{n\to\infty}(nB_n)^{-2}\sum_{i=1}^{n}E\left|\dfrac{1}{\nu_i} - E(\dfrac{1}{\nu_i})\right|\sigma_{[i]n}^2 = 0$

C3. $\lim_{n\to\infty}(nB_n)^{-2}\max_{1\leq i\leq n}\sigma_{[i]n}^2 = 0$.

Theorem 2. If conditions C1 – C3 are satisfied, then $B_n^{-1}(\tilde{\bar{\bar{y}}}_n - \mu)$ is asymptotically normal i.e. as $n \to \infty$

$$\sup_x \left| P\{B_n^{-1}(\tilde{\bar{y}}_n - \mu) \le x\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \right| \to 0 .$$

Note that condition (C1) is a version of so called Lyapuhov condition for asymptotic normality of a sum of independent random variables. Condition (C2) may be considered as a condition for the variance of $1/\nu_i$. In fact it is hard to expect asymptotic normality of $\tilde{\bar{y}}_n$ without a condition of "smallity" of deviation of $1/\nu_i$ from their expectations.

Proof. We denote $S_n = B_n^{-1}(\tilde{\bar{y}}_n - \mu)$. Using the independence of the random variables ν_i , $X_{[i]j}^{(n)}$, i , j , $n \ge 1$, by the total probability formula we obtain the following formula for the characteristic function of S_n :

$$E[e^{itS_n}] = E[\prod_{i=1}^{n} f_{in}^{\nu_i}(\frac{t}{nB_n\nu_i})] , \tag{15}$$

where $f_{in}(t) = E[e^{itv_{i1}}]$, $v_{ij}^{(n)} = X_{[i]j}^{(n)} - \mu_{[i]}^{(n)}$, $j \ge 1$. We now consider

$$\varphi_n(t) = \prod_{i=1}^{n} f_{in}^{\nu_i}(\frac{t}{nB_n\nu_i})$$

for arbitrary but fixed $t \in R$. First we prove that as $n \to \infty$

$$I_1 = \sum_{i=1}^{n} \nu_i [f_{in}(\frac{t}{nB_n\nu_i}) - 1] \xrightarrow{P} -\frac{t^2}{2} , \tag{16}$$

where P means convergence in probability. To do it we use the following representation for $f_{in}(t)$

$$f_{in}(t) = 1 - \frac{t^2}{2}\sigma_{[i]n}^2 - i\frac{t^3}{6}[E[v_{ij}^{(n)3}] + \varepsilon_{in}(t)] \tag{17}$$

for $j \ge 1$, where $|\varepsilon_{in}(t)| \le 3B_{[i]n}^3$ and $\varepsilon_{ij}(t) \to 0$ as $t \to 0$. If we use (17) it is not difficult to see that

$$I_1 = -\frac{t^2}{2} + \frac{t^2}{2n^2B_n^2}\sum_{i=1}^{n}(E[\frac{1}{\nu_i}] - \frac{1}{\nu_i})\sigma_{[i]n}^2 + \frac{it^3}{6} I_2 , \tag{18}$$

where $I_2 = \frac{1}{n^3B_n^3}\sum_{i=1}^{n}\frac{1}{\nu_i^2}[E(v_{ij}^{(n)3}) + \varepsilon_{in}(\frac{t}{nB_n\nu_i})]$. Using the Chebyshev inequality and condition (C2) we obtain that as $n \to \infty$ the second term in (18) converges in probability to zero. Now we consider I_2, taking into the account the estimate for $\varepsilon_{in}(t)$ and using the Chebyshev inequality again we have for any δ

$$P\{|I_2| > \delta\} \le \frac{4}{\delta \; n^3 B_n^3} \sum_{i=1}^{n} E[\frac{1}{\nu_i^2}] \beta_{[i]n}^3 .$$

This together with condition (C1) yields that the last term of (18) converges in probability to zero as $n \to \infty$ for each fixed t. Thus relation (16) is proved.

It follows from representation

$$\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - ..., |z| < 1$$

that $|\ln(1+z) - z| \le |z|^2$ for any complex z such that $|z| \le 1/2$. Note that here $\ln(z)$ denotes the principal value of the logarithm, i.e., $\ln(z) = \ln|z| + i \arg z$, $-\pi < \arg z \le \pi$. To use the last estimate for $\varphi_n(t)$ we need to prove that for sufficiently large n and each fixed t

$$\max_{1 \le i \le n} \left| f_{in}(\frac{t}{nB_n \nu_i}) - 1 \right| \le \frac{1}{2}. \tag{19}$$

Now we use the representation

$$f_{in}(t) = 1 - \frac{t^2}{2}[\sigma_{[i]n}^2 + \varepsilon_{in}^{(1)}(t)], \tag{20}$$

where $\left|\varepsilon_{in}^{(1)}(t)\right| \le 3\sigma_{[i]n}^2$. From (20) we get for any $i, n \ge 1$

$$\left| f_{in}(\frac{t}{nB_n \nu_i}) - 1 \right| \le 2t^2 \frac{\sigma_{[i]n}^2}{(nB_n \nu_i)^2} . \tag{21}$$

It follows from the last inequality that under the condition (C3) as $n \to \infty$ for each fixed t

$$\max_{1 \le i \le n} \left| f_{in}(\frac{t}{nB_n \nu_i}) - 1 \right| \to 0 \tag{22}$$

Consequently (19) holds for sufficiently large n and for each fixed t.

Now we will prove that as $n \to \infty$

$$I_3 = \sum_{i=1}^{n} \nu_i (f_{in}(\frac{t}{nB_n \nu_i}) - 1)^2 \xrightarrow{P} 0. \tag{23}$$

To show that we use (21) and obtain the following estimate $|I_3| \le \theta_n^{(1)} \theta_n^{(2)}$, where

$$\theta_n^{(1)} = \frac{2t^2}{n^2 B_n^2} \sum_{i=1}^{n} \frac{1}{v_i} \sigma_{[i]n}^2 \, , \; \theta_n^{(2)} = \frac{2t^2}{n^2 B_n^2} \max_{1 \le i \le n} \sigma_{[i]n}^2 \, .$$

Note that here $E[\theta_n^{(1)}] = 2t^2$. Thus for any $\delta > 0$

$$P\{|I_3| > \delta\} \le P\{\theta_n^{(1)} \theta_n^{(2)} > \delta\} \le 2t^2 \theta_n^{(2)}$$

As we can see that (23) follows from this and condition (C3). From (16) and (23) we conclude that

$$\varphi_n(t) = e^{\ln(\varphi_n(t))} \xrightarrow{P} e^{-t^2/2} \, ,$$

as $n \to \infty$. Since $|\varphi_n(t)| \le 1, n \ge 1$ by dominated convergence theorem (see Shiryaev (1996), Theorem 3, p187 and remark on p258), we obtain from the last that.

$$E[\varphi_n(t)] \to e^{-t^2/2}, n \to \infty \, .$$

This complete the prove of the Theorem

Acknowledgments

The authors are indebted to King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia for excellent research facilities.

REFERENCES

Dell and Clutter (1972). Ranked set sampling theory with order statistic background, *Biometrics*. 28, pp. 545-555.

Li, D., Sinha, B. K. and Perron, P. (1999). Random selection on 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.

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

Patil, G. P., Sinha, A. K., and Taillie, C (1999). Ranked set sampling: A bibliography, *Environmental and Ecological Statistics,* 6, 91-98.

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

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 τ along the corresponding variance $V(\bar{y}_{(v)})$ of RRSS as if N = τ 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
	τ	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
	τ	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
	τ	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
	τ	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