Ann. Inst. Statist. Math.
Vol. 46, No. 4, 723–736 (1994)

ESTIMATION OF PARAMETERS IN A TWO-PARAMETER EXPONENTIAL DISTRIBUTION USING RANKED SET SAMPLE

KIN LAM[1], BIMAL K. SINHA[2] AND ZHONG WU[2]

[1]*Department of Statistics, University of Hong Kong, Pokfulam Road, Hong Kong*
[2]*Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21228-5398, U.S.A.*

(Received June 30, 1993; revised April 28, 1994)

Abstract. In situations where the experimental or sampling units in a study can be easily ranked than quantified, McIntyre (1952, *Aust. J. Agric. Res.*, **3**, 385–390) proposed that the mean of n units based on a *ranked set sample* (*RSS*) be used to estimate the population mean, and observed that it provides an unbiased estimator with a smaller variance compared to a simple random sample (*SRS*) of the same size n. McIntyre's concept of *RSS* is essentially nonparametric in nature in that the underlying population distribution is assumed to be completely unknown. In this paper we further explore the concept of *RSS* when the population is partially known and the parameter of interest is not necessarily the mean. To be specific, we address the problem of estimation of the parameters of a two-parameter exponential distribution. It turns out that the use of *RSS* and its suitable modifications results in much improved estimators compared to the use of a *SRS*.

Key words and phrases: Best linear unbiased estimator, exponential distribution, order statistics, ranked set sample, uniformly minimum variance unbiased estimator.

1. Introduction

In many sampling situations when the variable of interest from the experimental units can be more easily ranked than quantified, it turns out that the use of McIntyre's (1952) notion of '*Ranked Set Sampling*' (*RSS*) is highly beneficial and much superior to the standard simple random sampling (*SRS*) for estimation of the population mean. Fortunately, in many agricultural and environmental studies, it is indeed possible to rank the experimental or sampling units without actually measuring them rather cheaply. We refer to Halls and Dell (1966), Martin *et al.* (1980), Cobby *et al.* (1985) and Sinha *et al.* (1992) for some applications.

The basic concept behind *RSS* can be briefly described as follows. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $F(x)$ with a mean μ and a finite variance σ^2. Then a standard nonparametric estimator of μ is $\bar{X} = \sum_1^n X_i/n$ with $\mathrm{var}(\bar{X}) =$

σ^2/n. In contrast to *SRS*, *RSS* uses only one observation, namely, $X_{1:n} \equiv X_{(11)}$, the lowest observation, from this set, then $X_{2:n} \equiv X_{(22)}$, the second lowest from another independent set of n observations, and finally $X_{n:n} \equiv X_{(nn)}$, the largest observation from a last set of n observations. This process can be described in Table 1.

Table 1. Display of n^2 observations in n sets of n each.

$X_{(11)}$	$X_{(12)}$	$\cdots$	$X_{(1(n-1))}$	$X_{(1n)}$
$X_{(21)}$	$X_{(22)}$	$\cdots$	$X_{(2(n-1))}$	$X_{(2n)}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
$X_{(n1)}$	$X_{(n2)}$	$\cdots$	$X_{(n(n-1))}$	$X_{(nn)}$

The important point to emphasize is that although *RSS* requires identification of as many as n^2 sampling units, only n of them, namely, $\{X_{(11)}, \ldots, X_{(nn)}\}$, are actually measured, thus making a comparison of this sampling strategy with *SRS* of the same size n meaningful. Obviously, *RSS* would be a serious contender to *SRS* in situations where the task of assembly of the sampling units is easy and their relative rankings in terms of the characteristic under study can be done with negligible cost. It is obvious that the new sample $X_{(11)}, X_{(22)}, \ldots, X_{(nn)}$, termed by McIntyre (1952) a *Ranked Set Sample* (*RSS*), are independent but not identically distributed. Moreover, marginally, $X_{(ii)}$ is distributed as $X_{i:n}$, the i-th order statistic in a sample of size n from $F(x)$. In certain situations, the whole procedure to generate a *RSS* of size n is repeated m times. Throughout this paper, we consider the case $m = 1$.

McIntyre (1952) proposed

$$\hat{\mu}_{rss} = \sum_{1}^{n} X_{(ii)}/n \tag{1.1}$$

as a rival estimator of μ as opposed to $\bar{X}$. It should be mentioned that McIntyre (1952) gave no supporting mathematical theory to prefer $\hat{\mu}_{rss}$ over $\bar{X}$. It was provided much later by Takahasi and Wakimoto (1968). It is easy to verify that $E(\hat{\mu}_{rss}) = \mu$, and a somewhat surprising result which makes *RSS* preferable to *SRS* is that

$$\operatorname{var}(\hat{\mu}_{rss}) < \operatorname{var}(\bar{X}) \ ! \tag{1.2}$$

A direct proof of this variance inequality follows from the well-known positively associated property of the order statistics (Tukey (1958), Bickel (1967)). Dell (1969) and Dell and Clutter (1972) provided the following explicit expression for the variance of $\hat{\mu}_{rss}$ where $\mu_{(i)}$ is the mean of $X_{i:n}$.

$$\operatorname{var}(\hat{\mu}_{rss}) = \sigma^2/n - \sum^{n} (\mu_{(i)} - \mu)^2/n^2. \tag{1.3}$$

Many aspects of RSS have been studied in the literature. Takahasi and Wakimoto (1968) have shown that the relative precision (RP) of $\hat{\mu}_{rss}$ relative to $\bar{X}$, defined as $RP = \operatorname{var}(\bar{X})/\operatorname{var}(\hat{\mu}_{rss})$, satisfies: $1 \leq RP \leq (n+1)/2$, with $RP = (n+1)/2$ in case the population is uniform. Patil *et al.* (1992a) computed the expressions for RP for many discrete and continuous distributions. David and Levine (1972) and Ridout and Cobby (1987) discussed the consequences of presence of errors in ranking. Muttlak and McDonald (1990a, 1990b) developed ranked set sampling theory when the experimental units are selected with size-biased probability with respect to a concomitant variable. Stokes (1980) and Sinha *et al.* (1992) discussed the estimation of variance based on a ranked set sample. Yanagawa and Shirahata (1976) developed RSS theory with selective probability matrix, and Yanagawa and Shan-Huo (1980) discussed the MG-procedure in RSS. For some other aspects of RSS, we refer to Takahashi (1969, 1970), Stokes (1977, 1986), Takahashi and Futatsuya (1988), Kvam and Samaniego (1991) and Patil *et al.* (1992b).

Admittedly, the concept of RSS is nonparametric in nature, and $\hat{\mu}_{rss}$ is a natural candidate for unbiased estimation of μ on the basis of RSS as described above when $F(x)$ is completely unknown. The object of this paper is to further explore the concept of RSS and its suitable modifications for estimation of the parameters in a two-parameter exponential distribution in the same spirit as in Sinha *et al.* (1992) where *normal* and *exponential* distributions are considered. However, one main point of difference is that unlike in Sinha *et al.* (1992) where mostly estimation of the mean is discussed, here the parameters of interest are not the means. We note that the pdf of a two-parameter exponential distribution can be written as

$$f(x \mid \theta, \sigma) = \frac{1}{\sigma} \exp[-(x-\theta)/\sigma], \qquad x \geq \theta, \quad \sigma > 0. \tag{1.4}$$

Section 2 is devoted to the estimation of θ, Section 3 to σ, and Section 4 to $\delta_p = \theta + \eta_p \sigma$, the p-th quantile of (1.4). Here $\eta_p = \log \frac{1}{1-p}$ stands for the p-th quantile of a standard exponential distribution, i.e., (1.4) with $\theta = 0$ and $\sigma = 1$. It may be noted that $p = 1 - e^{-1}$ corresponds to the mean of (1.4) so that McIntyre's (1952) general result is directly applicable in this case.

We note in passing that if $X_1, \ldots, X_n$ is a SRS of size n from (1.4), then

$$\hat{\theta} = X_{(1)} - \frac{\hat{\sigma}}{n}, \qquad \hat{\sigma} = \frac{\sum_1^n (X_i - X_{(1)})}{n-1} \tag{1.5}$$

are the uniformly minimum variance unbiased estimators (UMVUEs) of θ and σ respectively. Also, $\bar{X}$ is the UMVUE of $E(X) = \theta + \sigma$, and $\hat{\delta}_p = \hat{\theta} + \eta_p \hat{\sigma}$ is the UMVUE of δ_p. Moreover,

$$\begin{aligned} &\operatorname{var}(\hat{\theta}) = \frac{\sigma^2}{n(n-1)}, \qquad \operatorname{var}(\hat{\sigma}) = \frac{\sigma^2}{n-1}, \\ &\operatorname{var}(\hat{\delta}_p) = \sigma^2 \left[\frac{1}{n^2} + \frac{\left(\eta_p - \frac{1}{n}\right)^2}{n-1} \right]. \end{aligned} \tag{1.6}$$

2. Estimation of θ

In this section we discuss the problem of estimation of the parameter θ in (1.4), and point out that the use of *RSS* and its suitable variations results in much improved estimators compared to the use of a *SRS*.

2.1 *Best linear unbiased estimators*

We first address the issue of how best to use the *RSS*, namely, $X_{(11)}, \ldots, X_{(nn)}$, for estimation of θ. Recall that $E(X_{(ii)}) = \theta + c_{i:n}\sigma$ and $\mathrm{var}(X_{(ii)}) = d_{i:n}\sigma^2$, where

$$c_{i:n} = \sum_{j=1}^{i} \frac{1}{n-j+1}, \qquad d_{i:n} = \sum_{j=1}^{i} \frac{1}{(n-j+1)^2} \tag{2.1}$$

are respectively the expected value and the variance of the i-th order statistic in a sample of size n from a standard exponential distribution (David (1981), Arnold and Balakrishnan (1989), Balakrishnan and Cohen (1991)). Starting with $\sum_1^n c_i X_{(ii)}$ and minimizing $\mathrm{var}(\sum_1^n c_i X_{(ii)})$ subject to the unbiasedness conditions: $\sum_1^n c_i = 1$, $\sum_1^n c_i c_{i:n} = 0$, leads to the best linear unbiased estimator (*BLUE*) of θ as

$$\tilde{\theta}_{blue} = \frac{(\sum_1^n X_{(ii)}/d_{i:n})(\sum_1^n c_{i:n}^2/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})(\sum_1^n c_{i:n}X_{(ii)}/d_{i:n})}{(\sum_1^n 1/d_{i:n})(\sum_1^n c_{i:n}^2/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})^2} \tag{2.2}$$

with

$$\mathrm{var}(\tilde{\theta}_{blue}) = \sigma^2 \frac{(\sum_1^n c_{i:n}^2/d_{i:n})}{(\sum_1^n 1/d_{i:n})(\sum_1^n c_{i:n}^2/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})^2}. \tag{2.3}$$

The above formulae can be simplified a little bit using the fact that

$$\sum_{i=1}^{n} 1/d_{i:n} = \sum_{i=1}^{n} 1/(n-i+1). \tag{2.4}$$

Table 2 provides a comparison of $\mathrm{var}(\hat{\theta})$ and $\mathrm{var}(\tilde{\theta}_{blue})$ through $RP = \mathrm{var}(\hat{\theta})/\mathrm{var}(\tilde{\theta}_{blue})$ for $n = 6, 7, 8, 9, 10$ and clearly reveals the superiority of *RSS* over the use of *SRS*.

Incidentally, we can also derive the *BLUE* of θ based on a partial *RSS*, namely, $X_{(11)}, \ldots, X_{(ll)}$, for $l < n$. Starting with $\sum_1^l c_i X_{(ii)}$ and minimizing $\mathrm{var}(\sum_1^l c_i X_{(ii)})$ subject to the unbiasedness conditions: $\sum_1^l c_i = 1$, $\sum_1^l c_i c_{i:n} = 0$, leads to

$$\tilde{\theta}_{blue}(l) = \frac{\left(\sum_1^l \frac{X_{(ii)}}{d_{i:n}}\right)\left(\sum_1^l \frac{c_{i:n}^2}{d_{i:n}}\right) - \left(\sum_1^l \frac{c_{i:n}}{d_{i:n}}\right)\left(\sum_1^l \frac{c_{i:n}X_{(ii)}}{d_{i:n}}\right)}{\left(\sum_1^l \frac{1}{d_{i:n}}\right)\left(\sum_1^l \frac{c_{i:n}^2}{d_{i:n}}\right) - \left(\sum_1^l \frac{c_{i:n}}{d_{i:n}}\right)^2} \tag{2.5}$$

Table 2. Comparison of $\mathrm{var}(\hat{\theta})$ and $\mathrm{var}(\tilde{\theta}_{blue})$ through RP.

n	6	7	8	9	10
RP	1.0439	1.1333	1.2177	1.2870	1.3537

with

$$(2.6) \qquad \mathrm{var}(\tilde{\theta}_{blue}(l)) = \sigma^2 \frac{(\sum_1^l c_{i:n}^2/d_{i:n})}{(\sum_1^l 1/d_{i:n})(\sum_1^l c_{i:n}^2/d_{i:n}) - (\sum_1^l c_{i:n}/d_{i:n})^2}.$$

Table 3 provides, for $n = 8, 9$ and 10 minimum values of l for which $RP = \mathrm{var}(\hat{\theta})/\mathrm{var}(\tilde{\theta}_{blue}(l)) > 1$, and shows that often a partial RSS combined with optimum weights does better than a SRS of size n. Thus, for example, $\tilde{\theta}_{blue}(7)$ based on a partial RSS of size $l = 7$ is as efficient as the UMVUE of θ based on a SRS of size $n = 10$. As an extreme example, one can verify that $\tilde{\theta}_{blue}(9)$ based on only 9 observations is as efficient as $\hat{\theta}$ based on $n = 100$.

Table 3. Values of l and n for which $RP > 1$.

n	l	RP
8	7	1.1049
9	7	1.0692
10	7	1.0374

2.2 *Which order statistic?*

We next address the issue of the right selection of order statistics in the context of RSS, given that we must select one from each set of n observations, there are n such sets, and that the resultant estimator of θ is unbiased. Recall that McIntyre's scheme is based upon selecting the diagonal elements $(X_{(11)}, \ldots, X_{(nn)})$ in Table 1, where $X_{(ij)}$ refers to the j-th order statistic in the i-th row of this table. Unfortunately, unlike in the case of normal and exponential distributions, there is no *obvious* clear-cut choice of any 'optimum' order statistic in the present problem.

We first discuss the case of the minimum order statistic and examine the performance of the use of $(X_{(11)}, \ldots, X_{(l1)})$ for estimation of θ for various choices of $l = 1, \ldots, n$ in an attempt to determine the *minimum* value of l for which dominance over $\hat{\theta}$ holds. Noting that the pdf of $X_{(i1)}$ is of the same form as (1.4) with σ replaced by $\sigma^* = \sigma/n$, and that $X_{(11)}, \ldots, X_{(l1)}$ are iid, we can readily use (1.5) to conclude that

$$(2.7) \qquad \tilde{\theta}_{\min}(l) = Y_{(1)} - \frac{\sum_1^l (X_{(i1)} - Y_{(1)})}{l(l-1)}$$

is the UMVUE of θ based on $(X_{(11)}, \ldots, X_{(l1)})$, where $Y_{(1)} = \min\{X_{(11)}, \ldots, X_{(l1)}\}$, with the resultant variance given by

$$\operatorname{var}(\tilde{\theta}_{\min}(l)) = \frac{\sigma^2}{n^2 l(l-1)}. \tag{2.8}$$

A direct comparison of $\operatorname{var}(\hat{\theta})$ and $\operatorname{var}(\tilde{\theta}_{\min}(l))$ immediately shows that $l = 2$ does the job. In other words, appropriate use of *only two* smallest order statistics from Table 1 results in an unbiased estimator of θ better than the use of a *SRS* of size n, whatever be n! This is an extremely powerful and highly interesting result, and similar to what Sinha *et al.* (1992) observed in connection with estimation of normal and exponential means. It should be noted that the optimality of the minimum order statistic for estimation of θ is essentially due to its (partial) sufficiency under the model (1.4), and cannot be expected to hold for other distributions. Thus, for example, as noted in Sinha *et al.* (1992), a similar result holds for estimation of the normal mean only for the sample median, and for the exponential mean for a very special order statistic different from the sample median and the smallest order statistic.

We next address the problem of using a subset of the r-th order statistics, namely, $X_{(1r)}, \ldots, X_{(lr)}$ from the r-th column of Table 1 for some fixed $r > 1$ and for some $1 \le l \le n$ to efficiently estimate θ, and in the sequel determine the minimum l for every value of r for which the desired dominance holds. Note that for an arbitrary $r > 1$, $X_{(1r)}, \ldots, X_{(lr)}$ are iid with a common pdf of the form $\frac{1}{\sigma} g_{r,n}(\frac{x-\theta}{\sigma})$, $x > \theta$, for some $g_{r,n}(\cdot)$. It may be noted that, unless $r = 1$, an exact optimum inference in the form of a UMVUE of θ is extremely difficult in the general situation. Following the idea given in (1.5), we therefore propose to use

$$Z_{(l:r)} = \min\{X_{(1r)}, \ldots, X_{(lr)}\} \tag{2.9}$$

and

$$Z^*_{(l:r)} = \frac{\sum_1^l (X_{(ir)} - Z_{(l:r)})}{l-1}. \tag{2.10}$$

Noting that

$$E(Z_{(l:r)}) = \theta + \sigma\delta_{(l,r,n)}, \qquad E(Z^*_{(l:r)}) = \sigma\delta^*_{(l,r,n)} \tag{2.11}$$

where $\delta_{(l,r,n)}$ and $\delta^*_{(l.r,n)}$ are two absolute constants, we may use

$$\tilde{\theta}_{(l,r)} = Z_{(l:r)} - \frac{\delta_{(l,r,n)} Z^*_{(l:r)}}{\delta^*_{(l,r,n)}} \tag{2.12}$$

as an estimator of θ based on $X_{(1r)}, \ldots, X_{(lr)}$. We have numerically computed the values of $\delta_{(l,r,n)}$, $\delta^*_{(l,r,n)}$, and performed extensive simulation to evaluate the variance of $\tilde{\theta}_{(l,r)}$ for various values of l, r and $n = 5, 10$. It turns out that for

Table 4.1a. Values of $\delta_{(l,2,n)}$ and $\delta^*_{(l,2,n)}$ for $n = 5$.

l	2	3	4	5
$\delta^*_{(l,2,n)}$	.3389	.3511	.3595	.3659
$\delta_{(l,2,n)}$	.2806	.2159	.1804	.1573

Table 4.1b. Values of $\delta_{(l,2,n)}$ and $\delta^*_{(l,2,n)}$ for $n = 10$.

l	2	3	4	5	6	7	8	9	10
$\delta^*_{(l,2,n)}$	.1584	.1643	.1683	.1713	.1737	.1756	.1773	.1787	.1799
$\delta_{(l,2,n)}$	.1318	.1015	.0848	.0740	.0663	.0605	.0559	.0522	.0491

Table 4.2. Simulated values of $\mathrm{var}(\tilde{\theta}_{(l,2)})$ for $n = 5$.

l	2	3	4	5
$\mathrm{var}(\hat{\theta}_{(l,2)})$	$0.1078\sigma^2$	$0.0368\sigma^{2*}$	$0.0191\sigma^{2*}$	$0.0128\sigma^{2*}$

Table 4.3. Simulated values of $\mathrm{var}(\tilde{\theta}_{(l,2)})$ for $n = 10$.

l	2	3	4	5
$\mathrm{var}(\hat{\theta}_{(l,2)})$	$0.1096\sigma^2$	$0.0372\sigma^2$	$0.0192\sigma^2$	$0.0129\sigma^2$

6	7	8	9	10
$0.0092\sigma^{2*}$	$0.0064\sigma^{2*}$	$0.0049\sigma^{2*}$	$0.0042\sigma^{2*}$	$0.0037\sigma^{2*}$

Table 4.4. Simulated standard errors of simulated $\mathrm{var}(\tilde{\theta}_{(l,2)})$ for $n = 5$.

l	2	3	4	5
s.e.	6.1074×10^{-06}	3.3474×10^{-07}	5.1550×10^{-08}	1.9139×10^{-08}

these values of n, only $r = 2$ works. Values of $\delta_{(l,2,n)}$, and $\delta^*_{(l,2,n)}$ for $n = 5$ and 10 respectively are given in Tables 4.1a and 4.1b.

In Tables 4.2 and 4.3 we present the simulated values of $\mathrm{var}(\tilde{\theta}_{(l,r)})$ for all the combinations of l when $r = 2$ for $n = 5$ and 10 respectively. The combinations for which dominance over $\hat{\theta}$ holds, namely, $\mathrm{var}(\tilde{\theta}_{(l,2)}) < \frac{\sigma^2}{n(n-1)}$, are denoted by * (see (1.6) for $\mathrm{var}(\hat{\theta})$). Thus, for $n = 5$, the unbiased estimator $\tilde{\theta}_{(3,2)}$ based on three second order statistics $\{X_{(1,2)}, X_{(2,2)}, X_{(3,2)}\}$ performs better than $\hat{\theta}$, while for $n = 10$, the unbiased estimator $\tilde{\theta}_{(6,2)}$ based on $\{X_{(1,2)}, X_{(2,2)}, X_{(3,2)}, X_{(4,2)}, X_{(5,2)}, X_{(6,2)}\}$ is better than $\hat{\theta}$. The simulated values of the variances of $\tilde{\theta}_{(l,r)}$ are based on 10,000 replications of the values of $\tilde{\theta}_{(l,r)}$, each $\tilde{\theta}_{(l,r)}$ in turn being generated from n^2 simulated standard exponential variables.

To examine the stability of the above simulated values, we generated 20 sets of values of simulated $\mathrm{var}(\tilde{\theta}_{(l,2)})$ for $n = 5$ and $l = 2, 3, 4, 5$, each set in turn being based on 10,000 replications of standard exponential variables. The standard errors of these 20 values are given above in Table 4.4. It is clear that the amount of variation is very small, and dominance of $\tilde{\theta}_{(l,2)}$ over $\hat{\theta}$ for $n = 5$ and $l \geq 3$ always holds.

3. Estimation of σ

In this section we discuss the problem of estimation of the parameter σ in (1.4), and point out that the use of *RSS* and its suitable variations results in much improved estimators compared to the use of a *SRS*.

3.1 *BLUE*

To derive the *BLUE* of σ based on the entire McIntyre sample $X_{(11)}, \ldots, X_{(nn)}$, we minimize the variance of $\sum_1^n c_i X_{(ii)}$ subject to the unbiasedness conditions: $\sum_1^n c_i = 0$, $\sum_1^n c_i c_{i:n} = 1$. This results in

$$\tilde{\sigma}_{blue} = \frac{(\sum_1^n X_{(ii)} c_{i:n}/d_{i:n})(\sum_1^n 1/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})(\sum_1^n X_{(ii)}/d_{i:n})}{(\sum_1^n 1/d_{i:n})(\sum_1^n c_{i:n}^2/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})^2} \tag{3.1}$$

with

$$\mathrm{var}(\tilde{\sigma}_{blue}) = \sigma^2 \frac{(\sum_1^n 1/d_{i:n})}{(\sum_1^n 1/d_{i:n})(\sum_1^n c_{i:n}^2/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})^2}. \tag{3.2}$$

As before, the above expressions can be simplified a bit using (2.4). In Table 5 we have presented the values of $RP = \mathrm{var}(\hat{\sigma})/\mathrm{var}(\tilde{\sigma}_{blue})$ for $n = 4, 5, 6, 7, 8, 9, 10$. The overwhelming dominance of $\tilde{\sigma}_{blue}$ over $\hat{\sigma}$ in all the cases is obvious. For example, $\tilde{\sigma}_{blue}$ based on a *RSS* of size 7 is twice as efficient as $\hat{\sigma}$ based on $n = 7$.

Table 5. Comparison of $\mathrm{var}(\hat{\sigma})$ and $\mathrm{var}(\tilde{\sigma}_{blue})$ through RP.

n	RP
4	1.1865
5	1.4535
6	1.7241
7	1.9988
8	2.2755
9	2.5615
10	2.8414

As in the case of estimation of θ, here also we can use a partial RSS, namely, $X_{(11)}, \ldots, X_{(ll)}$ for some $l < n$. Starting with $\sum_1^l c_i X_{(ii)}$ and minimizing $\mathrm{var}(\sum_1^l c_i X_{(ii)})$ subject to the usual unbiasedness conditions: $\sum_1^l c_i = 0$, $\sum_1^l c_i c_{i:n} = 1$, we readily obtain the $BLUE$ of σ based on the partial RSS as

$$(3.3)\quad \tilde{\sigma}_{blue}(l) = \frac{(\sum_1^l X_{(ii)} c_{i:n}/d_{i:n})(\sum_1^l 1/d_{i:n}) - (\sum_1^l c_{i:n}/d_{i:n})(\sum_1^l X_{(ii)}/d_{i:n})}{(\sum_1^l 1/d_{i:n})(\sum_1^l c_{i:n}^2/d_{i:n}) - (\sum_1^l c_{i:n}/d_{i:n})^2}$$

with

$$(3.4)\qquad \mathrm{var}(\tilde{\sigma}_{blue}(l)) = \sigma^2 \frac{(\sum_1^l 1/d_{i:n})}{(\sum_1^l 1/d_{i:n})(\sum_1^l c_{i:n}^2/d_{i:n}) - (\sum_1^l c_{i:n}/d_{i:n})^2}.$$

Table 6 provides the minimum values of l for $n = 6, 7, 8, 9, 10$ for which dominance of $\tilde{\sigma}_{blue}(l)$ over $\hat{\sigma}$ holds, and clearly shows the superiority of this method. For example, $\tilde{\sigma}_{blue}(6)$ based on a RSS of size 6 is as efficient as $\hat{\sigma}$ based on $n = 9$.

Table 6. Values of l and n for which $RP > 1$.

n	l	RP
6	5	1.1422
7	6	1.4420
8	6	1.2059
9	6	1.0280
10	7	1.3086

3.2 *Which order statistic?*

In this subsection we explore the possibility of finding ‘optimum’ order statistics $X_{(1r)}, \ldots, X_{(nr)}$ corresponding to the r-th column in Table 1 so that a suitable combination of a subset of these order statistics, namely, $X_{(1r)}, \ldots, X_{(lr)}$ for some $l < n$ would produce an unbiased estimator of σ better than $\hat{\sigma}$. A close inspection of the formula given in (1.7) coupled with the observations made in Subsection 2.2 reveals that $r = 1$ will *not* do the job because

$$(3.5)\qquad \tilde{\sigma}_{\min}(l) = \frac{n\sum_1^l (Y_i - Y_{(1)})}{l-1},$$

the UMVUE of σ based on $X_{(11)}, \ldots, X_{(l1)}$, satisfies: $\mathrm{var}(\tilde{\sigma}_{\min}(l)) = \sigma^2/(l-1)$ which is always bigger than $\sigma^2/(n-1)$.

We have done extensive simulation for all $r > 1$ when $n = 5$ and found that there is no value of l for which the dominance of

$$(3.6)\qquad \tilde{\sigma}_{(l,r)} = \frac{Z^*_{(l,r)}}{\delta^*_{(l,r,n)}}$$

over $\hat{\sigma}$ holds. We also noted that the performance of $\tilde{\sigma}_{(l,r)}$, which is always worse than $\hat{\sigma}$, quickly deteriorates as r increases.

4. Estimation of quantiles

In this section we discuss the problem of estimation of quantiles δ_p of (1.4) for various values of $0 < p < 1$, and point out that the use of *RSS* and its suitable variations results in much improved estimators compared to the use of a *SRS*. The representation of δ_p as a linear combination of θ and σ makes our task quite easy. It should be noted that Stokes and Sager (1988) discussed the nonparametric estimation of a distribution function based on a *RSS*. However, no such result for estimation of a quantile is available.

4.1 *BLUE*

The *BLUE* of δ_p on the basis of the complete *RSS* $X_{(11)}, \ldots, X_{(nn)}$ can be derived by minimizing $\mathrm{var}(\sum_1^n c_i X_{(ii)})$ subject to the unbiasedness conditions: $\sum_1^n c_i = 1$, $\sum_1^n c_i c_{i:n} = \eta_p$. This immediately results in

$$\tilde{\delta}_{p,blue} = \lambda_1 \sum_1^n \frac{X_{(ii)}}{d_{i:n}} + \lambda_2 \sum_1^n \frac{c_{i:n} X_{(ii)}}{d_{i:n}} \tag{4.1}$$

where

$$\lambda_1 = \frac{(\sum_1^n c_{i:n}^2/d_{i:n}) - \eta_p(\sum_1^n c_{i:n}/d_{i:n})}{(\sum_1^n 1/d_{i:n})(\sum_1^n c_{i:n}^2/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})^2} \tag{4.2}$$

and

$$\lambda_2 = \frac{\eta_p(\sum_1^n 1/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})}{(\sum_1^n 1/d_{i:n})(\sum_1^n c_{i:n}^2/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})^2}. \tag{4.3}$$

Moreover, straightforward computations yield

$$\mathrm{var}(\tilde{\delta}_{p,blue}) = \sigma^2 \frac{\sum_1^n (c_{i:n} - \eta_p)^2/d_{i:n}}{(\sum_1^n 1/d_{i:n})(\sum_1^n c_{i:n}^2/d_{i:n}) - (\sum_1^n c_{i:n}/d_{i:n})^2}. \tag{4.4}$$

As before, the above expressions can be simplified a bit using (2.4). Table 7 represents the values of $RP = \mathrm{var}(\hat{\delta}_p)/\mathrm{var}(\tilde{\delta}_{p,blue})$ for $p = 0.1, 0.3, 0.5, 0.7, 0.9$ and $n = 4, 5, 6, 7, 8, 9, 10$. The usefulness of this appropriate variation of *RSS* is obvious. For example, for $n = 10$, $\tilde{\delta}_{p,blue}$ is twice as efficient as $\hat{\delta}_p$ for $p = 0.1$ and 6 times as efficient as $\hat{\delta}_p$ for $p = 0.5$.

Table 7. Values of $\mathrm{var}(\hat{\delta}_p)$ and $\mathrm{var}(\tilde{\delta}_{p,blue})$ through RP.

$n \backslash p$	0.1	0.3	0.5	0.7	0.9
4	1.1554	2.9097	4.3638	2.8859	1.8595
5	1.3266	3.3802	4.5603	3.1290	2.1526
6	1.4895	3.8182	4.8284	3.4201	2.4572
7	1.6369	4.2453	5.1182	3.7380	2.7701
8	1.7636	4.6154	5.4403	4.0669	3.0876
9	1.8889	5.0508	5.7812	4.4140	3.4103
10	2.0164	5.3830	6.1429	4.7569	3.7349

Analogously, based on a partial RSS of size l, the $BLUE$ of δ_p turns out to be

$$\tilde{\delta}_{p,blue}(l) = \lambda_1(l)\sum_1^l \frac{X_{(ii)}}{d_{i:n}} + \lambda_2(l)\sum_1^l \frac{c_{i:n}X_{(ii)}}{d_{i:n}} \tag{4.5}$$

where

$$\lambda_1(l) = \frac{(\sum_1^l c_{i:n}^2/d_{i:n}) - \eta_p(\sum_1^l c_{i:n}/d_{i:n})}{(\sum_1^l 1/d_{i:n})(\sum_1^l c_{i:n}^2/d_{i:n}) - (\sum_1^l c_{i:n}/d_{i:n})^2} \tag{4.6}$$

and

$$\lambda_2(l) = \frac{\eta_p(\sum_1^l 1/d_{i:n}) - (\sum_1^l c_{i:n}/d_{i:n})}{(\sum_1^l 1/d_{i:n})(\sum_1^l c_{i:n}^2/d_{i:n}) - (\sum_1^l c_{i:n}/d_{i:n})^2}. \tag{4.7}$$

Also, direct computations lead to

$$\operatorname{var}(\tilde{\delta}_{p,blue}(l)) = \sigma^2 \frac{\sum_1^l (c_{i:n} - \eta_p)^2/d_{i:n}}{(\sum_1^l 1/d_{i:n})(\sum_1^l c_{i:n}^2/d_{i:n}) - (\sum_1^l c_{i:n}/d_{i:n})^2}. \tag{4.8}$$

In Table 8 we have given the values of minimum l for $p = 0.1, 0.3, 0.5, 0.7, 0.9$ and $n = 4, 5, 6, 7, 8, 9, 10$ for which $RP = \operatorname{var}(\hat{\delta}_p)/\operatorname{var}(\tilde{\delta}_{p,blue}(l)) > 1$. The savings in the number of observations is quite impressive. As an example, compared to a SRS of size 10, a partial RSS based on as few as 2 and a maximum of 6 will provide a better estimator of δ_p as p ranges from 0.1 to 0.9.

Table 8. Values of l for which $RP > 1$.

p	$\backslash n$	4	5	6	7	8	9	10
0.1	l	4	4	4	3	3	2	2
	RP	1.1554	1.1379	1.1672	1.0363	1.1412	1.1250	1.3510
0.3	l	2	2	3	3	3	4	4
	RP	2.6916	1.7860	2.7222	1.9824	1.3846	2.1912	1.7211
0.5	l	3	3	4	4	4	5	5
	RP	2.7176	1.5283	2.1821	1.5257	1.1205	1.6336	1.3083
0.7	l	3	4	4	5	5	5	6
	RP	1.3434	1.8353	1.2440	1.7013	1.3160	1.0575	1.4410
0.9	l	4	4	5	5	6	6	6
	RP	1.8595	1.2311	1.6115	1.2394	1.6058	1.3303	1.1273

4.2 *Which order statistic?*

As discussed previously, we can address the problem of determining which subset of order statistics $X_{(1r)}, \ldots, X_{(lr)}$ provides the best scenario for unbiased estimation of δ_p for a given value of p. Obviously, for $r = 1$, the arguments presented in Subsections 2.2 and 3.2 show that the UMVUE of δ_p is given by

$$\tilde{\delta}_{p,\min}(l) = \tilde{\theta}_{\min}(l) + \eta_p \tilde{\sigma}_{\min}(l) \tag{4.9}$$

where $\tilde{\theta}_{\min}(l)$ and $\tilde{\sigma}_{\min}(l)$ are given by (2.5) and (3.5) respectively, and consequently

$$\text{var}(\tilde{\delta}_{p,\min}(l)) = \frac{\sigma^2}{n^2 l^2} + \left(n\eta_p - \frac{1}{l}\right)^2 \frac{\sigma^2}{n^2(l-1)}. \tag{4.10}$$

Table 9 provides a comparison of the variances of $\hat{\delta}_p$ and $\tilde{\delta}_{p,\min}(l)$ through RP and provides the smallest values of l for which $RP > 1$ for $p = 0.1, 0.3$, $n = 5, 10$. It should be noted that the desired dominance does not hold for higher values of p.

Table 9. Values of l for which $RP > 1$.

n	p	l	RP	p	l	RP
5	0.1	2	5.2802	0.3	4	2.4201
10	0.1	2	2.1961	0.3	8	1.4804

As noted before, for $r > 1$, it is extremely difficult to get an 'optimum' estimator of δ_p, and we propose to use

$$\tilde{\delta}_{p,(l,r)} = \tilde{\theta}_{(l,r)} + \eta_p \tilde{\sigma}_{(l,r)} \tag{4.11}$$

where $\tilde{\theta}_{(l,r)}$ and $\tilde{\sigma}_{(l,r)}$ are given by (2.12) and (3.6) respectively. Simulation studies reported in Table 10 for $p = 0.1, 0.3, 0.5, 0.7, 0.9$ and $n = 5$ reveal the interesting fact that $\tilde{\delta}_{p,(l,r)}$ dominates $\hat{\delta}_p$ (denoted as *) and even the *BLUE* $\hat{\delta}_{p,blue}$ (denoted as **) for many combinations of l and r. The simulated values of the variances of $\tilde{\delta}_{p,(l,r)}$ are based on 10,000 replications of the values of $\tilde{\delta}_{p,(l,r)}$, each $\tilde{\delta}_{p,(l,r)}$ in turn being generated from n^2 simulated standard exponential variables.

Remark 1. The patterns of the values of RP in Table 8, Table 9, and $\text{var}(\tilde{\delta}_{p,(l,r)})$ in Table 10 can probably be explained by the asymmetry of the underlying distribution and also the dependence between $\hat{\theta}$ and $\hat{\sigma}$.

Remark 2. Throughout this paper, we have reported our results for n up to 10, the minimum value of n in each of the tables being the one for which the desired dominance holds.

Remark 3. Since the standard errors of all our proposed estimators of θ, σ and δ_p based on a *RSS* depend on σ, these can be easily estimated by plugging in the proposed *RSS* estimators of σ.

Table 10. Simulated values of $\mathrm{var}(\tilde{\delta}_{p,(l,r)})$ for $n = 5$ and $\sigma = 1$.

p	l	$r = 2$	$r = 3$	$r = 4$	$r = 5$
0.1	2	0.0623	0.5705	2.4781	11.4645
	3	0.0245**	0.1342	0.6546	3.8194
	4	0.0145**	0.0538	0.2990	2.1312
	5	0.0111**	0.0459*	0.2373	5.3539
0.3	2	0.0484*	0.0983	1.1775	8.0159
	3	0.0397*	0.0076**	0.2497	2.4617
	4	0.0326*	0.0033**	0.0933	1.2873
	5	0.0299*	0.0409*	0.1729	4.7862
0.5	2	0.2382	0.0774*	0.1855	4.3622
	3	0.1584*	0.0825*	0.0079**	1.1089
	4	0.1207*	0.0721*	0.0014**	0.4889
	5	0.1061*	0.0853*	0.1567*	4.0913
0.7	2	0.2382*	0.0774**	0.1855*	4.3622
	3	0.1584*	0.0825**	0.0079**	1.1089
	4	0.1207**	0.0721**	0.0014**	0.4889
	5	0.1061**	0.0853**	0.1567*	4.0913
0.9	2	4.4485	9.6596	7.3036	2.1428
	3	2.2808	4.3121	3.6085	2.0026
	4	1.5542	2.6843	2.4704	1.9167
	5	1.2769*	1.1064*	1.1866*	1.7985

Acknowledgements

Our sincere thanks are due to an Associate Editor and two referees who provided many critical comments and helpful suggestions for significant improvements over the original version of the paper.

References

Arnold, B. C. and Balakrishnan, N. (1989). Relations, bounds and approximations for order statistics, *Lecture Notes in Statist.*, **53**, Springer, New York.

Balakrishnan, N. and Cohen, A. C. (1991). *Order Statistics and Inference*, Academic Press, Boston.

Bickel, P. J. (1967). Some contributions to the theory of order statistics, *Proc. 5th Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 575–591, Univ. of California Press, Berkeley.

Cobby, J. M., Ridout, M. S., Bassett, P. J. and Large, R. V. (1985). An investigation into the use of ranked set sampling on grass and grass-clover swards, *Grass and Forage Science*, **40**, 257–263.

David, H. A. (1981). *Order Statistics*, 2nd ed., Wiley, New York.

David, H. A. and Levine, D. N. (1972). Ranked set sampling in the presence of judgement error, *Biometrics*, **28**, 553–555.

Dell, T. R. (1969). The theory of some applications of ranked set sampling, Ph.D. Thesis, University of Georgia, Athens.

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

Halls, L. S. and Dell, T. R. (1966). Trial of ranked set sampling for forage yields, *Forest Science*, **12**(1), 22–26.

Kvam, P. H. and Samaniego, F. J. (1991). On the inadmissibility of standard estimators based on ranked set sampling, *1991 Joint Statistical Meetings of ASA Abstracts*, 291–292.

Martin, W. L., Sharik, T. L., Oderwald, R. G. and Smith, D. W. (1980). Evaluation of ranked set sampling for estimating shrub phytomass in appalachian Oak forest, Publication No. FWS-4-80, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets, *Aust. J. Agric. Res.*, **3**, 385–390.

Muttlak, H. A. and McDonald, L. L. (1990*a*). Ranked set sampling with respect to concomitant variables and with size biased probability of selection, *Comm. Statist. Theory Methods*, **19**(1), 205–219.

Muttlak, H. A. and McDonald, L. L. (1990*b*). Ranked set sampling with size biased probability of selection, *Biometrics*, **46**, 435–445.

Patil, G. P., Sinha, A. K. and Taillie, C. (1992*a*). Ranked set sampling and ecological data analysis, Tech. Reports and Reprints Series, Department of Statistics, Pennsylvania State University.

Patil, G. P., Sinha, A. K. and Taillie, C. (1992*b*). Ranked set sampling in the presence of a trend on a site, Tech. Reports and Reprints Series, Department of Statistics, Pennsylvania State University.

Ridout, M. S. and Cobby, J. M. (1987). Ranked set sampling with non-random selection of sets and errors in ranking, *Appl. Statist.*, **36**(2), 145–152.

Sinha, B. K., Sinha, B. K. and Purkayastha, S. (1992). On some aspects of ranked set sampling for estimation of normal and exponential parameters (submitted for publication).

Stokes, S. L. (1977). Ranked set sampling with concomitant variables, *Comm. Statist. Tneory Methods*, **6**(12), 1207–1211.

Stokes, S. L. (1980). Estimation of variance using judgment ordered ranked set samples, *Biometrics*, **36**, 35–42.

Stokes, S. L. (1986). Ranked set sampling, *Encyclopedia of Statistical Sciences*, **7** (eds. S. Kotz, N. L. Johnson and C. B. Read), 585–588, Wiley, New York.

Stokes, L. S. and Sager, T. (1988). Characterization of a ranked set sample with application to estimating distribution functions, *J. Amer. Statist. Assoc.*, **83**, 374–381.

Takahasi, K. (1969). On the estimation of the population mean based on ordered samples from an equicorrelated multivariate distribution, *Ann. Inst. Statist. Math.*, **21**, 249–255.

Takahasi, K. (1970). Practical note on estimation of population means based on samples stratified by means of ordering, *Ann. Inst. Statist. Math.*, **22**, 421–428.

Takahasi, K. and Futatsuya, M. (1988). Ranked set sampling from a finite population, *Proc. Inst. Statist. Math.*, **36**(1), 55–68 (in Japanese).

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering, *Ann. Inst. Statist. Math.*, **20**, 1–31.

Tukey, J. W. (1958). A problem of Berkson, and minimum variance orderly estimators, *Ann. Math. Statist.*, **29**, 588–592.

Yanagawa, T. and Chen, S. H. (1980). The MG-procedure in rank set sampling, *J. Statist. Plann. Inference*, **4**, 33–44.

Yanagawa, T. and Shirahata, S. (1976). Ranked set sampling theory with selective probability matrix, *Austral. J. Statist.*, **18**(1, 2), 45–52.