Statistica Sinica **10**(2000), 247-263

THE EFFICIENCY OF RANKED-SET SAMPLING RELATIVE TO SIMPLE RANDOM SAMPLING UNDER MULTI-PARAMETER FAMILIES

Zehua Chen

National University of Singapore

Abstract: When observations are costly or time-consuming but the ranking of the observations without actual measurement can be done relatively easily, ranked-set sampling (RSS) can be employed instead of simple random sampling (SRS) to gain more information. In this article, we deal with RSS under multi-parameter parametric families. It is proved that the Fisher information matrix of an RSS sample is the sum of its counterpart of an SRS sample and an additional positive definite matrix. This insures that the maximum likelihood estimates (MLE) based on RSS are always more efficient than their counterparts based on SRS. The effect of certain features of the underlying distribution such as skewness and kurtosis on the relative efficiency of the MLE is also investigated. Some other aspects of RSS are discussed as well.

Key words and phrases: Asymptotic relative efficiency, confidence interval, Fisher information, maximum likelihood estimation, ranked-set sampling.

1. Introduction

In certain practical problems, actual measurements of a variable of interest are costly or time-consuming, but the ranking of items according to the variable is relatively easy without actual measurement. Under such circumstances a sampling scheme called ranked-set sampling (RSS) can be employed to gain more information than simple random sampling (SRS) while keeping the cost of, or the time constraint on, the sampling about the same. The RSS scheme goes as follows. A sample of size k is drawn from the population under investigation. Then the items of the sample are ranked by judgment based on certain knowledge of subject, without actual measurement. After the ranking only one item is actually measured (or quantified), say, the rth smallest one. The procedure is repeated $N = \sum_{r=1}^{k} n_r$ times, where $n_r \geq 1$ is the number of samples for which the rth smallest item is quantified. In this article, we deal with the case of equal n_r's. The case of non-equal n_r's will be considered elsewhere. When all the n_r's are equal to n, say, the procedure can be described as n cycles. In each cycle, k samples of size k are drawn and one and only one of the items with the same order in these k samples is quantified.

The RSS was first applied by McIntyre (1952) in his study on estimation of mean pasture yields. Measuring yield of pasture plots requires mowing and weighing the hay, but an experienced person can fairly accurately rank a small number of plots without actual measurement. Later, more applications of RSS were made in agriculture, e.g., Halls and Dell (1966) and Cobby, Ridout, Bassett and Large (1985). Recently, the interest in RSS has been found in environmental studies, see Johnson, Patil and Sinha (1993), Patil and Taillie (1993), Patil, Sinha and Taillie (1993a, b), and Gore, Patil and Sinha (1994). Applications have also been suggested in situations where judgment ranking can be done through personal interviews, by use of photographs plus supplementary data, etc.

The properties of RSS have been investigated by several authors. McIntyre (1952) stated without rigor that $\hat{\mu}^*$, the sample mean of an RSS sample, is an unbiased estimate of the population mean, and the relative precision $RP(\hat{\mu}^*, \hat{\mu})$ of $\hat{\mu}^*$ to the sample mean of an SRS sample $\hat{\mu}$, the ratio of the variance of $\hat{\mu}$ and the variance of $\hat{\mu}^*$, is slightly less than $(k+1)/2$. To make the comparison between RSS and SRS meaningful, the number of quantified sampled items is set equal for both samples. Dell and Clutter (1972) investigated RP further and derived an explicit formula for it. They computed the RP for a variety of distributions for $k = 2$ to 5, and provided supporting evidence for McIntyre's statement. The above authors only considered the problem of estimating population means. Stokes (1980) considered the method-of-moment estimation of variance and showed that improved estimates of variance can be produced from RSS samples as well. Stokes and Sager (1988) characterized an RSS sample as a sample from a conditional distribution, conditioning on a multinomial random vector, and applied RSS to the estimation of the cumulative distribution functions. They showed that $RP(\hat{F}^*(t), \hat{F}(t)) \geq 1$ for all t, where $\hat{F}^*$ and $\hat{F}$ are the empirical distribution functions of the RSS sample and the SRS sample, respectively. Bohn and Wolfe (1992) used the RSS empirical distribution function to construct distribution-free competitors to the standard Mann-Whitney-Wilcoxon estimation and testing procedures. Hettmasperger (1995) investigated properties of the sign test along with the median and corresponding confidence interval for RSS. In all the problems investigated in the literature, RSS results in improved procedures over SRS.

The studies mentioned above focus on non-parametric settings, i.e., no assumption on the distribution of the observations is made. Studies on RSS in parametric settings have also attracted the attentions of researchers, e.g., Shen (1994) and Stokes (1995). Specifically, Stokes (1995) considered estimation of μ and σ for location-scale families with cumulative distribution functions of the form $F((x-\mu)/\sigma)$. She investigated the asymptotic relative efficiency (ARE) of the maximum likelihood estimates (MLE) from an RSS sample to the MLE from an SRS sample and showed the following. When σ is known, the Fisher information on μ from an RSS sample is

$$I^*_{nk}(\mu) = I_{nk}(\mu) + \frac{nk(k-1)}{\sigma^2} E\left\{\frac{[f(Z)]^2}{F(Z)[1-F(Z)]}\right\},$$

where $f(z)$ is the density function of F, $Z \sim F$ and $I_{nk}(\mu)$ denotes the Fisher information on μ from an SRS sample of size nk. Hence, the ARE of the MLE $\hat{\mu}^*_{ML}$ from an RSS sample to the MLE $\hat{\mu}_{ML}$ from an SRS sample is given by

$$ARE(\hat{\mu}^*_{ML}, \hat{\mu}_{ML}) = 1 + (k-1)E\left\{\frac{[f(Z)]^2}{F(Z)[1-F(Z)]}\right\} \bigg/ E\left\{\frac{f'(Z)}{f(Z)}\right\}^2.$$

When μ is known, the Fisher information on σ from an RSS sample is

$$I^*_{nk}(\sigma) = I_{nk}(\sigma) + \frac{nk(k-1)}{\sigma^2} E\left\{\frac{[Zf(Z)]^2}{F(Z)[1-F(Z)]}\right\},$$

and hence, the ARE of the MLE $\hat{\sigma}^*_{ML}$ from an RSS sample to the MLE $\hat{\sigma}_{ML}$ from an SRS sample is given by

$$ARE(\hat{\sigma}^*_{ML}, \hat{\sigma}_{ML}) = 1 + (k-1)E\left\{\frac{[Zf(Z)]^2}{F(Z)[1-F(Z)]}\right\} \bigg/ \left[E\left\{\frac{Zf'(Z)}{f(Z)}\right\}^2 - 1\right].$$

When both μ and σ are unknown, Stokes pointed out that if f is symmetric then the off-diagonal elements in the information matrix about (μ, σ) are zero so that $|I_{nk}(\mu, \sigma)| < |I^*_{nk}(\mu, \sigma)|$ and the ARE in estimating μ and σ remain the same though the MLE's might differ. However, in general, it is not obvious that $|I_{nk}(\mu, \sigma)| < |I^*_{nk}(\mu, \sigma)|$, and hence it can not be determined whether or not the MLE from an RSS sample is more efficient than its counterpart from an SRS sample. Stokes stated that it does not appear to be possible to determine that $|I_{nk}(\mu, \sigma)| < |I^*_{nk}(\mu, \sigma)|$ generally for non-symmetric distributions unless k is sufficiently large, (see Stokes (1995), p.472).

From a statistical point of view, an RSS sample should always contain more information than an SRS sample, since an RSS sample contains not only the information carried by the quantified observations but also the information provided by the judgment ranking. It is this belief that motivated our study. In this article, we consider a multi-parameter family of distributions $F(x, \theta)$, where θ is a vector of unknown parameters. In this general setting, we find a nice structure of the information matrix of the RSS sample about θ and prove that the information matrix from an RSS sample is the sum of the information matrix of the corresponding SRS sample and an additional non-negative definite matrix, whether the ranking is perfect or not. Thus the MLE of any function of θ from an RSS sample is always more efficient than its counterpart from an SRS sample, at least, asymptotically. We first consider the case of perfect ranking and deal with the ARE of the MLE of θ from an RSS sample relative to the MLE from an SRS sample. We also investigate, in the case of perfect ranking, how certain characteristics of $F(x, \theta)$ such as skewness and kurtosis affect the ARE of the MLE's. Then we consider the case of imperfect ranking. In Section 2, we derive the Fisher information matrix on θ from an RSS sample in the case of perfect ranking. The ARE of the MLE's in the case of perfect ranking is dealt with in Section 3. In Section 4, we consider a particular model for the case of imperfect ranking and derive the Fisher information matrix of the RSS sample in this case. Some remarks are given in Section 5.

2. The Fisher Information Matrix from an RSS Sample

We consider a population whose cumulative distribution function and density function are given by, respectively, $F(x;\theta)$ and $f(x;\theta)$, where $\theta = (\theta_1, \ldots, \theta_p)'$ is a vector of unknown parameters. Assume that $f(x;\theta)$ satisfies the following regularity conditions.

(C1) For all x, all the first, second and third partial derivatives of $f(x;\theta)$ with respect to the components of θ exist.

(C2) For each θ_0 in the range of θ, there is a neighborhood $N(\theta_0)$ such that every first or second partial derivative in (C1) is bounded by an integrable function and every third derivative is bounded by a function which has finite expectation when $\theta \in N(\theta_0)$.

(C3) For each θ, $0 < E\{(\partial \log f(X, \theta)/\partial\theta_i)^2\} < \infty$ for $i = 1, \ldots, p$.

These regularity conditions insure the consistency and asymptotic normality of the MLE of θ. In addition, we assume that the judgment ranking is perfect. Let the k quantified variables in the ith cycle be denoted by $X_{(1)i}, \ldots, X_{(k)i}$. Under the assumption of perfect judgment ranking, they are indeed order statistics but independent, differing from the order statistics of an SRS sample. It follows that $X_{(r)i}, i = 1, \ldots, n$, are independent identically distributed (i.i.d.) with density function given by

$$f_{(r)}(x;\theta) = \frac{k!}{(r-1)!(k-r)!} F^{r-1}(x;\theta)[1 - F(x;\theta)]^{k-r} f(x;\theta).$$

Let

$$l_1(\theta) = \sum_{r=1}^{k} [(r-1)\ln F(X_{(r)};\theta) + (k-r)\ln(1 - F(X_{(r)};\theta))],$$

$$l_2(\theta) = \sum_{r=1}^{k} \ln f(X_{(r)};\theta).$$

The log-likelihood function of an RSS sample with $n = 1$ can then be written as

$$l(\theta) = \sum_{r=1}^{k} \ln f_{(r)}(X_{(r)}; \theta) = l_1(\theta) + l_2(\theta) + C,$$

where C is a constant. Under the regularity conditions, the Fisher information matrix on θ from the sample is given by

$$I_k^*(\theta) = -E\left[\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta^T}\right] = -E\left[\frac{\partial^2 l_1(\theta)}{\partial\theta\partial\theta^T}\right] - E\left[\frac{\partial^2 l_2(\theta)}{\partial\theta\partial\theta^T}\right].$$

It is easy to see that $-E[\partial^2 l_2(\theta)/\partial\theta\partial\theta^T]$ is the same as the Fisher information matrix on θ from an SRS sample of size k. Write

$$\begin{aligned} F_{(r)} &= F(X_{(r)}; \theta), \\ F'_{(r)i} &= \frac{\partial F(X_{(r)}; \theta)}{\partial \theta_i}, \\ F''_{(r)ij} &= \frac{\partial^2 F(X_{(r)}; \theta)}{\partial\theta_i \partial\theta_j}. \end{aligned}$$

After some straightforward manipulation, we find the (i,j)th element of $\partial^2 l_1(\theta)/\partial\theta\partial\theta^T$ to be

$$\begin{aligned} \frac{\partial^2 l_1(\theta)}{\partial\theta_i\partial\theta_j} = &\sum_{r=1}^{k}(r-1)\left[\frac{F''_{(r)ij}}{F_{(r)}(1-F_{(r)})} + \frac{F'_{(r)i}F'_{(r)j}}{F_{(r)}(1-F_{(r)})^2}\right] \\ &-(k-1)\sum_{r=1}^{k}\left[\frac{F''_{(r)ij}}{(1-F_{(r)})} + \frac{F'_{(r)i}F'_{(r)j}}{(1-F_{(r)})^2}\right] \\ &-\sum_{r=1}^{k}(r-1)\left[\frac{F'_{(r)i}F'_{(r)j}}{F^2_{(r)}(1-F_{(r)})}\right]. \end{aligned} \tag{1}$$

Now let us state and prove a lemma.

Lemma 1. *Let $Y_{k,r}$ denote the rth order statistic of a random sample of size k from a distribution with cumulative distribution function $F(x)$. Then for any function $G(\cdot)$,*

$$E\left\{\sum_{r=1}^{k}(r-1)\frac{G(Y_{k,r})}{F(Y_{k,r})}\right\} = k(k-1)EG(X),$$

provided $EG(X)$ exists, where X is a random variable with cumulative distribution function $F(x)$.

Proof.

$$
\begin{aligned}
& E\left\{\sum_{r=1}^{k}(r-1)\frac{G(Y_{k,r})}{F(Y_{k,r})}\right\} \\
&= \sum_{r=1}^{k}(r-1)\int \frac{k!}{(r-1)!(k-r)!}\frac{G(y)}{F(y)}F^{r-1}(y)(1-F(y))^{k-r}dF(y) \\
&= k\sum_{r=2}^{k}\int \frac{(k-1)!}{(r-2)!(k-r)!}G(y)F^{r-2}(y)(1-F(y))^{k-r}dF(y) \\
&= k\sum_{r=1}^{k-1}\int \frac{(k-1)!}{(r-1)!(k-1-r)!}G(y)F^{r-1}(y)(1-F(y))^{k-1-r}dF(y) \\
&= k\sum_{r=1}^{k-1} EG(Y_{k-1,r}) \\
&= k(k-1)EG(X).
\end{aligned}
$$

By applying the lemma to the sums in (1), the expectation of the first sum and the second sum cancel each other, and the expectation of the third sum becomes $-k(k-1)\Delta_{ij}(\theta)$ with

$$
\Delta_{ij}(\theta) = E\left\{\frac{\partial F(X;\theta)/\partial\theta_i \ \partial F(X;\theta)/\partial\theta_j}{F(X;\theta)[1-F(X;\theta)]}\right\},
$$

where expectation is taken with respect to X with distribution function $F(x;\theta)$. Let $I(\theta)$ denote the Fisher information matrix of a single random observation from distribution $F(x;\theta)$. Let $\Delta(\theta)$ denote the matrix with (i,j)th element $\Delta_{ij}(\theta)$. Then we have

Theorem 1. *Under regularity conditions* (C1)—(C3),

$$
I_k^*(\theta) = kI(\theta) + k(k-1)\Delta(\theta).
$$

Note that the matrix $\Delta(\theta)$ can be written as

$$
\Delta(\theta) = E\left\{\frac{1}{F(X;\theta)[1-F(X;\theta)]}\left(\frac{\partial F(X;\theta)}{\partial\theta}\right)\left(\frac{\partial F(X;\theta)}{\partial\theta}\right)^T\right\}.
$$

Hence $\Delta(\theta)$ is non-negative definite. It will be referred to as the information gain matrix.

It follows from the i.i.d. structure of the cycles in the RSS scheme that the Fisher information matrix on θ from an RSS sample with n cycles is given by

$$
I_{nk}^*(\theta) = nI_k^*(\theta).
$$

We consider some special cases.

(i) *Location-scale families.* If $F(x;\theta) = F((x-\mu)/\lambda)$, where $\theta = (\mu,\lambda)^T$, the family is called a location-scale family. Then

$$\frac{\partial F((X-\mu)/\lambda)}{\partial \mu} = -\frac{1}{\lambda} f(\frac{X-\mu}{\lambda}),$$
$$\frac{\partial F((X-\mu)/\lambda)}{\partial \lambda} = -\frac{1}{\lambda}\frac{X-\mu}{\lambda} f(\frac{X-\mu}{\lambda}).$$

Hence the information gain matrix is given by

$$\Delta_{11} = \frac{1}{\lambda^2} E\left\{\frac{[f(X)]^2}{F(X)[1-F(X)]}\right\},$$
$$\Delta_{22} = \frac{1}{\lambda^2} E\left\{\frac{[Xf(X)]^2}{F(X)[1-F(X)]}\right\},$$
$$\Delta_{12} = \frac{1}{\lambda^2} E\left\{\frac{X[f(X)]^2}{F(X)[1-F(X)]}\right\},$$

where the expectation is taken with respect to X with distribution function $F(x)$ and density $f(x)$. If $f(x)$ is symmetric about zero, then $\Delta_{12} = 0$. The information gain matrix is independent of the location parameter μ and is inversely proportional to the square of the scale parameter λ. This is the case considered by Stokes (1995). The normal and exponential families fall into this class. We have, for the normal family $N(\mu,\lambda^2)$, $\Delta_{11} = 0.4805/\lambda^2, \Delta_{22} = 0.0675/\lambda^2$ and $\Delta_{12} = 0$. For the exponential family $\mathcal{E}(\theta)$, $\Delta(\theta) = 0.4041/\theta^2$.

(ii) *Shape-scale families.* By a shape-scale family we refer to a family with cumulative distribution functions of the form $F(x/\lambda,\alpha)$, where $\theta = (\lambda,\alpha)^T$, λ is called the scale parameter and α the shape parameter. Let $F_\alpha(x)$ denote the cumulative distribution with $\lambda = 1$ and $f_\alpha(x)$ the corresponding density function. The information gain matrix for a shape-scale family is given by

$$\begin{pmatrix} 1 & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \begin{pmatrix} \delta_{11}(\alpha) & \delta_{12}(\alpha) \\ \delta_{21}(\alpha) & \delta_{22}(\alpha) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$$

where

$$\delta_{11}(\alpha) = E\left\{\frac{(\partial F_\alpha(X)/\partial\alpha)^2}{F_\alpha(X)[1-F_\alpha(X)]}\right\},$$
$$\delta_{12}(\alpha) = -E\left\{\frac{X(\partial F_\alpha(X)/\partial\alpha)(\partial f_\alpha(X)/\partial\alpha)}{F_\alpha(X)[1-F_\alpha(X)]}\right\},$$
$$\delta_{22}(\alpha) = E\left\{\frac{(X\partial f_\alpha(X)/\partial\alpha)^2}{F_\alpha(X)[1-F_\alpha(X)]}\right\},$$

where the expectation is taken with respect to $X \sim F_\alpha(x)$.

We consider two families of this type: the gamma family and the Weibull family. The gamma family $Gamma(\alpha, \lambda)$ is a shape-scale family with density function given by $f_\alpha(x) = x^{\alpha-1}e^{-x}/\Gamma(\alpha)$. The values of $\delta_{11}(\alpha)$, $\delta_{12}(\alpha)$ and $\delta_{22}(\alpha)$ for $\alpha = 1.5(0.5)10$ are given in Table 1. The Weibull family $Weibull(\lambda, \alpha)$ is a shape-scale family with density function given by $f_\alpha(x) = \alpha x^{\alpha-1}e^{-x^\alpha}$. The values of $\delta_{11}(\alpha)$, $\delta_{12}(\alpha)$ and $\delta_{22}(\alpha)$ for $\alpha = 1.5(0.5)10$ are given in Table 2. Tables 1 and 2 are to be used in the computation of the asymptotic relative efficiencies for estimating mean and variance for the two families in the next section.

Table 1. The information gain matrix for Gamma distributions.

α	$\delta_{11}(\alpha)$	$\delta_{22}(\alpha)$	$\delta_{12}(\alpha)$	α	$\delta_{11}(\alpha)$	$\delta_{22}(\alpha)$	$\delta_{12}(\alpha)$
1.5	0.4174	0.6403	0.4872	6.5	0.0790	3.0356	0.4845
2	0.2995	0.8785	0.4944	7	0.0730	3.2759	0.4842
2.5	0.2268	1.1175	0.4872	7.5	0.0679	3.5168	0.4840
3	0.1855	1.3570	0.4897	8	0.0634	3.7582	0.4839
3.5	0.1557	1.5967	0.4884	8.5	0.0595	3.9995	0.4839
4	0.1341	1.8366	0.4874	9	0.0560	4.2397	0.4838
4.5	0.1177	2.0765	0.4866	9.5	0.0530	4.4773	0.4835
5	0.1049	2.3163	0.4860	10	0.0502	4.7113	0.4829
5.5	0.0918	2.5560	0.4729				

Table 2. The information gain matrix for Weibull distributions.

α	$\delta_{11}(\alpha)$	$\delta_{22}(\alpha)$	$\delta_{12}(\alpha)$	α	$\delta_{11}(\alpha)$	$\delta_{22}(\alpha)$	$\delta_{12}(\alpha)$
1.5	0.11195	0.9092	0.02334	6.5	0.00596	17.0738	0.02334
2	0.06297	1.6164	0.02334	7	0.00514	19.8015	0.02334
2.5	0.04030	2.5257	0.02334	7.5	0.00447	22.7314	0.02334
3	0.02798	3.6370	0.02334	8	0.00393	25.8632	0.02334
3.5	0.0205	4.9503	0.02334	8.5	0.00348	29.1972	0.02334
4	0.01574	6.4658	0.02334	9	0.00310	32.7332	0.02334
4.5	0.01243	8.1833	0.02334	9.5	0.00279	36.4712	0.02334
5	0.01007	10.1028	0.02334	10	0.00251	40.4113	0.02334
5.5	0.00832	12.2244	0.02334				

3. The Asymptotic Relative Efficiency of the Maximum Likelihood Estimates

In this section, we deal with the asymptotic relative efficiency of the MLE from RSS samples with respect to the MLE from SRS samples in the case of perfect ranking. We also investigate how the asymptotic relative efficiency is affected by the skewness and kurtosis of the underlying distribution.

Let $\hat{\theta}^*_{nk}$ denote the MLE of θ from an RSS sample with cycle size k and number of cycles n. First, we have the following

Proposition 1. *Under regularity conditions* (C1)—(C3),

(i) $\hat{\theta}^*_{nk}$ *is strongly consistent as* n *goes to infinity;*

(ii) *As* n *goes to infinity,* $\hat{\theta}^*_{nk}$ *converges in distribution to the normal distribution with mean* θ *and variance-covariance matrix* $[kI(\theta) + k(k-1)\Delta(\theta)]^{-1}$*;*

(iii) *If* ϕ *is a function of* θ*, the asymptotic efficiency of the MLE* $\hat{\phi}^*$ *of* ϕ *from RSS samples, relative to its counterpart* $\hat{\phi}$ *from SRS samples, is given by*

$$ARE(\hat{\phi}^*, \hat{\phi}) = \frac{(\phi'(\theta))^T I^{-1}(\theta)\phi'(\theta)}{(\phi'(\theta))^T[I(\theta) + (k-1)\Delta(\theta)]^{-1}\phi'(\theta)},$$

where $\phi'(\theta)$ *is the vector of the derivatives of* ϕ *with respect to the components of* θ*. In particular, if* θ *itself is a scalar, the asymptotic relative efficiency is*

$$ARE(\hat{\phi}^*, \hat{\phi}) = 1 + (k-1)\frac{\Delta(\theta)}{I(\theta)},$$

for any differentiable function $\phi(\theta)$*.*

Note that since the matrix $\Delta(\theta)$ is non-negative definite the relative efficiency $ARE(\hat{\phi}^*, \hat{\phi})$ is always greater than or equal to 1.

In what follows, we examine the ARE in estimating the mean and variance for some particular families. We investigate how the skewness and kurtosis of a distribution affect the AREs. To this end, we work with the Gamma, Weibull and non-central t-distribution families.

(a) For the $Gamma(\alpha, \lambda)$ family, the information matrix about (α, λ) is as follows:

$$\begin{pmatrix} [\Gamma(\alpha)\Gamma''(\alpha) - [\Gamma'(\alpha)]^2][\Gamma(\alpha)]^{-2} & \lambda^{-1} \\ \lambda^{-1} & \alpha\lambda^{-2} \end{pmatrix}.$$

The mean and the variance of the gamma distribution are, respectively, $\mu = \phi_1 = \alpha\lambda$ and $\sigma^2 = \phi_2 = \alpha\lambda^2$. We have $\phi_1' = (\lambda, \alpha)^T$ and $\phi_2' = (\lambda^2, 2\alpha\lambda)^T$. We computed the ARE, the skewness and the kurtosis for $\alpha = 1.5(0.5)10, \lambda = 1$ and $k = 2$. The results are given in Table 3.

(b) For the $Weibull(\lambda, \alpha)$ family, the information matrix about (α, λ) is as follows:

$$\begin{pmatrix} \alpha^{-2} + \tau_2 & \lambda^{-1}(1 - \alpha\tau_1 - \tau_0) \\ \lambda^{-1}(1 - \alpha\tau_1 - \tau_0) & \alpha\lambda^{-2}[(\alpha+1)\tau_0 - 1] \end{pmatrix},$$

where $\tau_i = E[X^\alpha(lnX)^i], i = 0, 1, 2,$ with $X \sim Weibull(\alpha, 1)$. The mean and the variance of the Weibull distribution are, respectively, $\mu = \phi_1 =$

$\lambda\Gamma(1+1/\alpha)$ and $\sigma^2 = \phi_2 = \lambda^2[\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)]$. We have

$$\phi_1^{'} = [-\frac{\lambda}{\alpha^2}\Gamma^{'}(1+\frac{1}{\alpha}), \Gamma(1+\frac{1}{\alpha})]^T,$$
$$\phi_2^{'} = [-\frac{2\lambda^2}{\alpha^2}[\Gamma^{'}(1+\frac{2}{\alpha}) - \Gamma(1+\frac{1}{\alpha})\Gamma^{'}(1+\frac{1}{\alpha})], 2\lambda[\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})]]^T.$$

The ARE, the skewness and the kurtosis for $\alpha = 1.5(0.5)10, \lambda = 1$ and $k = 2$ are given in Table 4.

(c) The relative efficiency in estimating the mean, and the kurtosis of the non-central t_n-distribution for $k = 2$ and $n = 7(1)10(5)30, 40, 50, 75, 100, 150, 200$ are given in Table 5. The non-central t-distributions are artificial but they provide us with an example for examining how the kurtosis alone will affect the ARE in estimating the mean.

Table 3. The relative efficiency of the RSS under Gamma distributions.

α	ARE(μ)	ARE(σ^2)	Skewness	Kurtosis
1.5	1.4192	1.2321	1.633	7
2	1.4302	1.1979	1.4142	6
2.5	1.4409	1.2072	1.2649	5.4
3	1.4458	1.1854	1.1547	5
3.5	1.4504	1.1807	1.069	4.7143
4	1.4540	1.1767	1	4.5
4.5	1.4568	1.1733	0.9428	4.3333
5	1.4590	1.1702	0.8944	4.2
5.5	1.4641	1.1700	0.8528	4.0909
6	1.4623	1.1659	0.8165	4
6.5	1.4637	1.1641	0.7845	3.9231
7	1.4649	1.1624	0.7559	3.8571
7.5	1.4660	1.1613	0.7303	3.8
8	1.4670	1.1595	0.7071	3.75
8.5	1.4679	1.1584	0.686	3.7059
9	1.4686	1.1556	0.6667	3.6667
9.5	1.4688	1.1545	0.6489	3.6316
10	1.4687	1.1528	0.6325	3.6

The AREs of the MLEs of the means of the Gamma and Weibull distributions given in Tables 3 and 4 can be compared with Table 1 of Dell and Clutter (1972). The ARE of the MLE is always larger than the ARE of the sample mean for these two families. This is also true for normal and exponential distributions. We guess that this is a general phenomenon for all distributions, and hence we might conclude that, in general, more gain can be obtained in parametric RSS than in nonparametric RSS.

Table 4. The relative efficiency of the RSS under Weibull distributions.

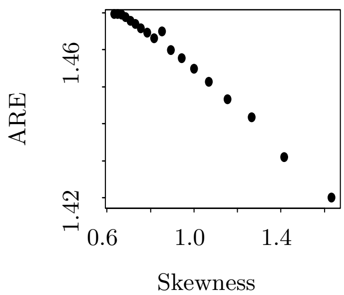
α	ARE(μ)	ARE(σ^2)	Skewness	Kurtosis
1.5	1.44281	1.15844	1.072	4.3904
2	1.46693	1.14179	0.6311	3.2451
2.5	1.47771	1.13122	0.3586	2.8568
3	1.48277	1.12695	0.1681	2.7295
3.5	1.48517	1.12676	0.0251	2.7127
4	1.48626	1.12876	-0.0872	2.7478
4.5	1.48667	1.13175	-0.1784	2.8081
5	1.4867	1.13509	-0.2541	2.8803
5.5	1.48651	1.13845	-0.3182	2.9574
6	1.48619	1.14166	-0.3733	3.0355
6.5	1.48581	1.14465	-0.4211	3.1125
7	1.48539	1.14739	-0.4632	3.1872
7.5	1.48498	1.14991	-0.5005	3.259
8	1.48458	1.15222	-0.5337	3.3277
8.5	1.48417	1.15432	-0.5636	3.3931
9	1.48375	1.15623	-0.5907	3.4552
9.5	1.48343	1.15809	-0.6152	3.5142
10	1.48316	1.15978	-0.6376	3.5702

Table 5. The relative efficiency of the RSS under t-distributions.

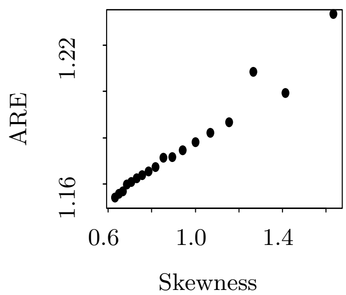
n	$\delta(\mu)$	$I(\mu)$	ARE	Kurtosis
7	0.3996	0.8	1.4995	9.8
8	0.4086	0.81818	1.4994	8
9	0.4158	0.83333	1.49897	6.94286
10	0.4217	0.84615	1.49837	6.25
15	0.44011	0.88889	1.49513	4.72028
20	0.44975	0.91304	1.49258	4.16667
25	0.45568	0.92857	1.49073	3.88199
30	0.45969	0.93939	1.48935	3.70879
40	0.46478	0.95349	1.48745	3.50877
50	0.46787	0.96226	1.48622	3.39674
75	0.47204	0.97436	1.48446	3.25584
100	0.47415	0.98058	1.48353	3.18878
150	0.47626	0.98693	1.48257	3.12384
200	0.47733	0.99015	1.48208	3.09215

To better understand how the ARE is affected by skewness and kurtosis, the AREs of the three examples above are depicted in Figures 1, 2 and 3, respectively. Since skewness and kurtosis are highly correlated in the first two examples (for Gamma distributions, the correlation is 0.9910, and for Weibull distributions the correlation is 0.9637), the AREs are only plotted against skewness. In each of Figures 1 and 2, Panel (a) depicts the AREs for the MLEs of the means, Panel (b) depicts the AREs for the MLEs of the variances and Panel (c) depicts the kurtosis against the skewness. In Figure 3, the ARE of the MLE of the mean is depicted against the kurtosis. Because of the high correlation between skewness and kurtosis in the first two examples, their effects on the ARE are aliased and can not be distinguished. However, in both Figure 1 and Figure 2, there is a common feature: the ARE of the MLE of the mean decreases as the skewness (or the kurtosis) increases and the ARE of the MLE of the variance increases as the skewness (or the kurtosis) increases. However, in the case of non-central t-distributions, the increment of kurtosis also increases the ARE of the MLE of the mean, as shown in Figure 3.

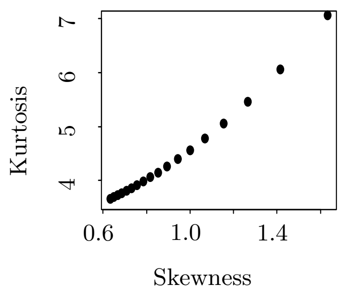
We make some general comments to end this section. Although the ARE is highly affected by skewness and kurtosis, it is not completely determined by them. Since we have confined ourselves to distributions satisfying the regularity conditions, whether or not the features of these distributions are shared by others is still unknown.



1(a). ARE for the means

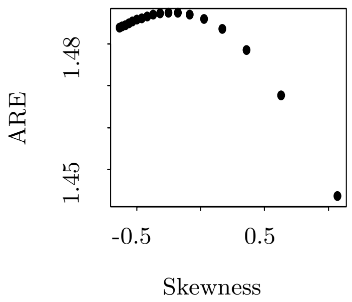


1(b). ARE for the variances

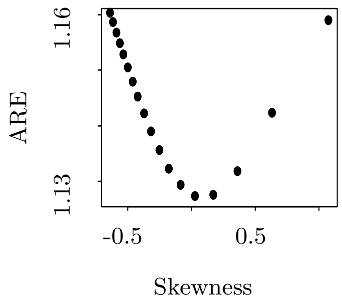


1(c). Kurtosis vs skewness

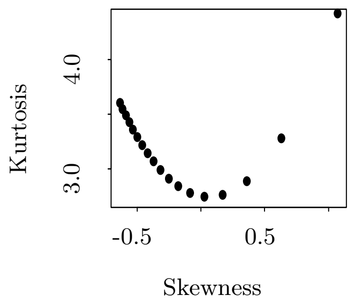
Figure 1. The effects of skewness and kurtosis of Gamma distributions on the ARE.



2(a). ARE for the means

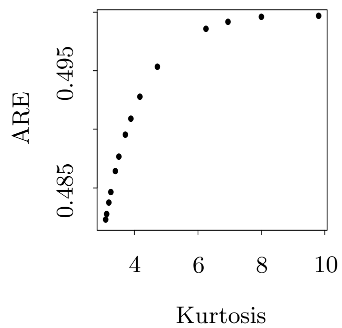


2(b). ARE for the variances



2(c). Kurtosis vs skewness

Figure 2. The effects of skewness and kurtosis of Weibull distributions on the ARE.



ARE for the means of t-distributions

Figure 3. The effects of kurtosis of non-central t-distributions on the ARE.

4. The Fisher Information Matrix When Ranking Is Imperfect

This section is devoted to the case of imperfect ranking. By imperfect ranking one means that the judgment-ranked order of the items do not match their true numerical orders. In practice, since the numerical orders of the items are unknown unless they are all measured, one can never be sure that the judgment ranking is perfect. In other words, imperfect ranking is inevitable. When the ranking is imperfect, does an RSS sample still contain more information than an SRS sample? We answer this question in this section. We work with a particular but reasonable model for imperfect ranking considered by Bohn and Wolfe (1994) and Hettmansperger (1995). Let p_{sr} denote the probability that the item that actually has numerical rank s is judgment-ranked as the rth order statistic. The model assumes $p_{sr} = p_{rs}$. In this section, we derive the Fisher information matrix of an RSS sample under the above model.

We denote the judgment-ranked rth order statistic by $X_{[r]}$ and let $f_{[r]}$ denote the density function of $X_{[r]}$. Then

$$\begin{aligned} f_{[r]} &= \sum_{s=1}^{k} p_{sr} \frac{k!}{(k-s)!(s-1)!} F^{s-1}(1-F)^{k-s} f \\ &= g_r f, \end{aligned}$$

say. Note that $\sum_s p_{sr} = \sum_r p_{sr} = 1$. It follows that

$$\sum_{r=1}^{k} f_{[r]} = kf, \tag{2}$$

$$\sum_{r=1}^{k} g_r = k. \tag{3}$$

Let $\tilde{I}^*_{nk}(\theta)$ denote the Fisher information matrix about θ in an RSS sample with set size k and number of cycles n under the above model for imperfect ranking. We have the following result.

Theorem 2. *Under regularity conditions* (C1)—(C3),

$$\tilde{I}^*_{nk}(\theta) = nkI(\theta) + \tilde{\Lambda}(\theta),$$

where $\tilde{\Lambda}(\theta)$ *is a non-negative definite matrix given by*

$$\tilde{\Lambda}(\theta) = \sum_{r=1}^{k} E\left[\frac{\frac{\partial g_r(X)}{\partial \theta}\frac{\partial g_r(X)}{\partial \theta^T}}{g_r(X)}\right].$$

Here the expectation is taken with respect to $X(\sim F)$.

Proof. It suffices to prove the theorem for $n = 1$. We have

$$\begin{aligned} \tilde{I}^*_{nk}(\theta) &= -E\left[\frac{\partial^2 \sum_{r=1}^{k} \log f_{[r]}(X_{[r]})}{\partial\theta\partial\theta^T}\right] \\ &= -E\left[\frac{\partial^2 \sum_{r=1}^{k} \log g_r(X_{[r]})}{\partial\theta\partial\theta^T}\right] - E\left[\frac{\partial^2 \sum_{r=1}^{k} \log f(X_{[r]})}{\partial\theta\partial\theta^T}\right]. \end{aligned} \tag{4}$$

It follows from (2) that

$$-E\left[\frac{\partial^2 \sum_{r=1}^k \log f(X_{[r]})}{\partial\theta\partial\theta^T}\right] = kI(\theta). \tag{5}$$

Write

$$-E\left[\frac{\partial^2 \sum_{r=1}^k \log g_r(X_{[r]})}{\partial\theta\partial\theta^T}\right] = \sum_{r=1}^k E\left[\frac{\frac{\partial g_r(X_{[r]})}{\partial\theta}\frac{\partial g_r(X_{[r]})}{\partial\theta^T}}{g_r(X_{[r]})^2}\right] - \sum_{r=1}^k E\left[\frac{\frac{\partial^2 g_r(X_{[r]})}{\partial\theta\partial\theta^T}}{g_r(X_{[r]})}\right]. \tag{6}$$

It follows from (3) that

$$\begin{aligned}\sum_{r=1}^k E\left[\frac{\frac{\partial^2 g_r(X_{[r]})}{\partial\theta\partial\theta^T}}{g_r(X_{[r]})}\right] &= \sum_{r=1}^k \int \frac{\partial^2 g_r(x)}{\partial\theta\partial\theta^T} f(x)dx \\ &= \int \left[\frac{\partial^2 \sum_{r=1}^k g_r(x)}{\partial\theta\partial\theta^T}\right] f(x)dx \\ &= 0. \end{aligned} \tag{7}$$

Finally, we have

$$\begin{aligned}\sum_{r=1}^k E\left[\frac{\frac{\partial g_r(X_{[r]})}{\partial\theta}\frac{\partial g_r(X_{[r]})}{\partial\theta^T}}{g_r(X_{[r]})^2}\right] &= \sum_{r=1}^k \int \frac{\partial g_r(x)}{\partial\theta}\frac{\partial g_r(x)}{\partial\theta^T}\frac{f(x)}{g_r(x)}dx \\ &= \sum_{r=1}^k E\left[\frac{\frac{\partial g_r(X)}{\partial\theta}\frac{\partial g_r(X)}{\partial\theta^T}}{g_r(X)}\right]. \end{aligned} \tag{8}$$

The theorem then follows from (4) — (8).

In practice, to carry out the maximum likelihood estimation of θ, one needs to determine the probabilities p_{sr}. One can specify the probabilities by experience or by a resampling procedure described as follows. Let samples of size k be resampled from the nk measured items and the re-samples ranked by judgment ranking. Suppose that the person who ranks the samples is blinded from the measurements of the items. Then judgment ranks are compared with numerical ranks. Estimates of the probabilities p_{sr} can then be obtained. Of course, there is inevitable misspecification of these probabilities. How the misspecification affects the MLE and its relative efficiency needs further investigation.

5. Some Remarks

(i) *Confidence intervals.* Proposition 1 can be used to construct confidence intervals for ϕ when n is large. A $100(1-\alpha)\%$ confidence interval of ϕ can be taken as

$$\hat{\phi}^* \pm \frac{z_{\alpha/2}}{\sqrt{nk}}\hat{se}(\hat{\phi}^*),$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution, and

$$\hat{\phi}^* = \phi(\hat{\theta}^*_{nk}),$$
$$\hat{se}(\hat{\phi}^*) = [(\phi'(\hat{\theta}^*_{nk}))^T[I(\hat{\theta}^*_{nk}) + (k-1)\Delta(\hat{\theta}^*_{nk})]^{-1}\phi'(\hat{\theta}^*_{nk})]^{1/2}.$$

The length of the confidence interval based on the MLE from an RSS sample is $\sqrt{ARE}$ times shorter than its counterpart based on the MLE from an SRS sample. For example, if the underlying distribution is normal and k is set to 7, the confidence interval based on the MLE from an RSS sample is only $1/\sqrt{ARE} = [1+(7-1)(0.4807)]^{-1/2} \approx 1/2$ as long as the confidence interval based on the MLE from an SRS sample.

(ii) *Choice of k.* The asymptotic relative efficiency increases as k increases but, in practice, we can not set k too large: as k increases, the difficulty of judgment ranking increases. However, it is sensible to take k as large as judgment ranking allows.

(iii) *Small sample efficiency.* The efficiency achieved by RSS comes not only in large sample theory but also in small sample situations. We conducted a simulation study to illustrate this. We generated RSS samples with $nk = 24, k = 2, 3, 4, 6, 8, 12, 24$ and SRS samples of size 24 from a standard normal distribution. The MLE of the mean was computed for each of these samples. This procedure was repeated 500 times. The approximations of the mean square errors were computed. The simulation results are reported in Table 6. The first column of the table lists the values of k with $k = 1$ corresponding to SRS samples. The second column gives the values of $(1/500)\sum_{i=1}^{500}(\hat{\theta}^*_{nk} - 0)^2$. The values in the third column are estimated relative efficiencies (ERE), ratios of the MSE of the MLE from SRS samples to the MSE of the MLE from RSS samples. The values in the last column are the asymptotic relative efficiencies which are given by $1 + 0.4805(k-1)$. It can be seen from the table that the estimated relative efficiencies closely match the theoretical asymptotic relative efficiencies.

Table 6. Comparison of RSS MLE and SRS MLE of the mean based on samples of size $nk = 24$ from a standard normal distribution.

k	MSE	RE	ARE
24	0.00313	12.46	12.04
12	0.00693	5.63	6.28
8	0.00988	3.95	4.36
6	0.01253	3.11	3.40
4	0.01623	2.40	2.44
3	0.02101	1.86	1.96
2	0.02603	1.50	1.48
1	0.03901	1	1

References

Bohn, L. L. and Wolfe, D. A. (1992). Nonparametric two-sample procedures for ranked-set samples data. *J. Amer. Statist. Assoc.* **87**, 552-561.

Bohn, L. L. and Wolfe, D. A. (1994). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitey-Wilcoxon statistic. *J. Amer. Statist. Assoc.* **89**, 168-176.

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.

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

Gore, S. D., Patil, G. P. and Sinha, A. K. (1994). Environmental chemistry, statistical modeling, and observational economy. *Environmental Statistics, Assessment, and Forecasting* (Edited by C. R. Cothern and N. P. Ross), 57-97. Lewis Publishing/CRC Press, Boca Raton.

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

Hettmansperger, T. P. (1995). The ranked-set sampling sign test. *Nonparametric Statistics* **4**, 263-270.

Johnson, G. D., Patil, G. P. and Sinha, A. K. (1993). Ranked set sampling for vegetation research. *Abstracta Botanica* **17**, 87-102.

McIntyre, G. A. (1952). A method of unbiased selective sampling, using ranked sets. *Austral. J. Agriculture Research* **3**, 385-390.

Patil, G. P. and Taillie, C. (1993). Environmental sampling, observational economy, and statistical inference with emphasis on ranked set sampling, encounter sampling, and composite sampling. *Bulletin ISI Proceedings of 49th Session*, 295-312. Firenze.

Patil, G. P., Sinha, A. K. and Taillie, C. (1993a). Ranked set sampling from a finite population in the presence of a trend on a site. *J. Appl. Statist. Science* **1**, 51-65.

Patil, G. P., Sinha, A. K. and Taillie, C. (1993b). Observational economy of ranked set sampling: comparison with the regression estimator. *Environmetrics* **4**, 399-412.

Shen, W. H. (1994). Use of ranked set sampling for test of a normal mean. *Calcutta Statist. Assoc. Bulletin* **44**, 183-193.

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

Stokes, S. L. (1995). Parametric ranked set sampling. *Ann. Inst. Statist. Math.* **47**, 465-482.

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

Department of Statistics and Applied Probability, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Republic of Singapore.

E-mail: stachenz@nus.edu.sg

(Received March 1998; accepted March 1999)