Ann. Inst. Statist. Math.
Vol. 54, No. 3, 626–640 (2002)
©2002 The Institute of Statistical Mathematics

ON THE RANKED-SET SAMPLING M-ESTIMATES FOR SYMMETRIC LOCATION FAMILIES

XIAOYUE ZHAO[1]* AND ZEHUA CHEN[2]

[1] *Department of Statistics, University of California, 367 Evans Hall, Berkeley, CA 94720-3860, U.S.A.*
[2] *Department of Statistics & Applied Probability, National University of Singapore, 3 Science Drive 2, Singapore 117543, Republic of Singapore,* e-mail: stachenz@nus.edu.sg

(Received April 6, 2000; revised June 11, 2001)

Abstract. The ranked-set sampling (RSS) is applicable in practical problems where the variable of interest for an observed item is costly or time-consuming but the ranking of a set of items according to the variable can be easily done without actual measurement. In this article, the M-estimates of location parameters using RSS data are studied. We deal mainly with symmetric location families. The asymptotic properties of M-estimates based on ranked-set samples are established. The properties of unbalanced ranked-set sample M-estimates are employed to develop the methodology for determining optimal ranked-set sampling schemes. The asymptotic relative efficiencies of ranked-set sample M-estimates are investigated. Some simulation studies are reported.

Key words and phrases: Asymptotic normality, asymptotic relative efficiency, M-estimates, optimal sampling design, ranked-set sampling, robustness.

1. Introduction

The ranked-set sampling (RSS) proposed by McIntyre (1952) is a sampling scheme that can be utilized for gaining more information when actual measurement of the variable of interest for an observed item is costly or time-consuming while the ranking of a set of items according to the variable can be relatively easily done without actual measurement. A ranked set sample is obtained by first drawing simple random samples, each of size k, then ranking the items of each sample by judgment and measuring, in each ranked sample, only one ranked order statistic with a pre-specified rank. If each ranked order statistic is measured the same number of times the RSS is said to be balanced, otherwise, it is said to be unbalanced. For details of a general RSS scheme, we refer the reader to Bai and Chen (2001).

Many statistical procedures including parametric and nonparametric procedures based on the balanced RSS have been investigated in the literature. The reader is referred to, among others, McIntyre (1952), Takahasi and Wakimoto (1968), Dell and Clutter (1972), Stokes (1980*a*, 1980*b*), Stokes and Sager (1988), Bohn and Wolfe (1992), Shen (1994), Sinha *et al.* (1994), Stokes (1995), Hettmansperger (1995), Koti and Babu (1996), Chen (1999, 2000*a*, 2000*b*) and Bai and Chen (2001). Recently, there is an increasing interest in research on unbalanced RSS to seek further improvement over the

*The research was done when she was a Master student of the National University of Singapore.

balanced RSS. See, e.g., Stokes (1995), Kaur *et al.* (1997), Chen (2001), and Chen and Bai (2000).

In this article, we consider the M-estimates using balanced and unbalanced ranked-set sampling data. The idea of M-estimates arises out of concern on the robustness of statistical procedures. For example, the usual estimate of the population mean, the sample mean, is not robust if the underlying distribution is heavily tailed. In RSS, the ranked set sample mean will be even more erratic when the tail of the underlying distribution is heavy. This motivates our research on the M-estimates in the context of RSS which, for convenience, are referred to as the RSS M-estimates hereafter. In this article, we investigate the properties of the RSS M-estimates including their asymptotic distribution and their relative efficiencies. We illustrate how unbalanced RSS can be designed to gain more efficiency over balanced RSS. The article is arranged as follows. In Section 2, we give some definitions and notations to be used in the article. The asymptotic properties of the RSS M-estimates are established in Section 3. The method for the design of optimal RSS schemes is discussed in Section 4. The relative efficiencies of the RSS M-estimates are investigated in Section 5. Some simulation results are reported in Section 6. Some concluding remarks are given in Section 7. Some technical details are provided in the Appendix.

2. Notation and definition

Let the cumulative distribution function (CDF) and the probability density function (PDF) of the underlying distribution be denoted by F and f respectively. The CDF and PDF of the r-th order statistic of a sample of size k from F are denoted by $F_{(r)}$ and $f_{(r)}$ respectively. Let a ranked-set sample be represented by

$$\begin{array}{llll} X_{(1)1}, & X_{(1)2}, & \ldots, & X_{(1)n_1}, \\ X_{(2)1}, & X_{(2)2}, & \ldots, & X_{(2)n_2}, \\ \ldots, & \ldots, & \ldots, & \ldots, \\ X_{(k)1}, & X_{(k)2}, & \ldots, & X_{(k)n_k}, \end{array} \tag{2.1}$$

where, in the subscripts, indices within parenthesis indicate pre-specified ranks and indices outside parenthesis represent repetitions of the same ranked order statistic. Define the empirical distribution function of the ranked-set sample (2.1) as follows:

$$\hat{F}_n(x) = \sum_{r=1}^{k} \frac{n_r}{n} \hat{F}_{(r)n_r}(x),$$

where

$$\hat{F}_{(r)n_r}(x) = \frac{1}{n_r} \sum_{i=1}^{n_r} I\{X_{(r)i} \le x\}, \qquad \text{and} \qquad n = \sum_{r=1}^{k} n_r.$$

If the n_r's are all equal, i.e., the RSS is balanced, we denote $\hat{F}_n$ by $\tilde{F}_n$ to distinguish between balanced and unbalanced RSS.

Suppose that $n_r/n \to p_r$ as $n \to \infty$. Let $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)'$. We shall refer to $\boldsymbol{p}$ as an allocation vector. Denote

$$F_{\boldsymbol{p}}(x) = \sum_{r=1}^{k} p_r F_{(r)}(x).$$

It is clear that $\hat{F}_n$ converges uniformly to $F_{\boldsymbol{p}}$ with probability 1. In the balanced case, $F_{\boldsymbol{p}} = F$.

Let $\psi(x)$ be an appropriate function. Define the functional $T(F)$ over all distribution functions as the solution of $\lambda_F(t) = \int \psi(x-t)dF(x) = 0$, if exists. Denote $\hat{T}_n = T(\hat{F}_n)$ and $\tilde{T}_n = T(\tilde{F}_n)$, respectively, when the RSS is unbalanced and balanced. In generic notation, if F is an unknown population distribution function and $\hat{F}$ is an appropriate estimate of F, then $T(\hat{F})$ is called an M-estimate of $T(F)$. Thus, $\hat{T}_n$ is an M-estimate of $T(F_{\boldsymbol{p}})$ in general, and $\tilde{T}_n$ is an M-estimate of $T(F)$ in particular. Our goal is to estimate $T(F)$. We shall find conditions such that $T(F) = T(F_{\boldsymbol{p}})$ and hence both $\tilde{T}_n$ and $\hat{T}_n$ are M-estimates of $T(F)$.

We also introduce the following notation:

$$A_{(r)}(t) = \int \psi^2(x-t)dF_{(r)}(x) - \left[\int \psi(x-t)dF_{(r)}(x)\right]^2,$$
$$\lambda'_F(T(F)) = \left.\frac{d\int \psi(x-t)dF(x)}{dt}\right|_{t=T(F)}.$$

Throughout the article, we assume that the underlying distribution is symmetric and that $\psi(x)$ is odd. Under this assumption, the median, the mean of the distribution F and $T(F)$ are all the same.

3. The asymptotic properties of RSS M-estimates

We present the asymptotic properties of the RSS M-estimates in this section. The properties include the strong consistency and the asymptotic normality. First, we give conditions on the allocation vector $\boldsymbol{p}$ such that $T(F_{\boldsymbol{p}}) = T(F)$. The conditions follow from the lemma below.

LEMMA 1. *Suppose that $\psi(x)$ is an odd function and F is a symmetric location distribution, then the population mean μ is a solution of $\lambda_F(t) = 0$, i.e., $\mu = T(F)$, and, further, μ satisfies*

$$\int_{-\infty}^{+\infty} \psi(x-\mu)dF_{(r)}(x) + \int_{-\infty}^{+\infty} \psi(x-\mu)dF_{(k-r+1)}(x) = 0.$$

Hence, if the component of the allocation vector $\boldsymbol{p}$ satisfies that $p_r = p_{k-r+1}$ for all r, then μ is also a solution of $\lambda_{F_{\boldsymbol{p}}}(t) = 0$, i.e., $\mu = T(F_{\boldsymbol{p}})$.

An allocation vector $\boldsymbol{p}$ satisfying the conditions given in Lemma 1 will be referred to as being symmetric.

THEOREM 1. *Suppose that $\psi(x)$ is odd, continuous and either monotone or bounded, and that F is a symmetric location distribution. Let the allocation vector $\boldsymbol{p}$ be symmetric. Then there is a solution sequence $\{\hat{T}_n\}$ of $\lambda_{\hat{F}_n}(t) = 0$ such that $\{\hat{T}_n\}$ converges to μ with probability 1.*

There are other conditions on ψ such that the above theorem holds. However, since the ψ's in practical applications satisfy the conditions in Theorem 1, we will concentrate on the ψ's satisfying these conditions.

In the following, we give three sets of conditions each of which, together with the conditions on F given in Theorem 1, guarantees the asymptotic normality of the sequence $\hat{T}_n$.

A1. $\psi(x)$ is odd and monotone; $\lambda_F(t)$ is differentiable at $t = \mu$, with $\lambda'_F(\mu) \neq 0$; $\int \psi^2(x-t)dF(x)$ is finite for t in a neighborhood of μ and is continuous at $t = \mu$.

A2. $\psi(x)$ is odd and continuous and satisfies

$$\lim_{t\to\mu} \| \psi(\cdot,t) - \psi(\cdot,\mu) \|_V = 0;$$

$\lambda_F(t)$ is differentiable at $t = \mu$, with $\lambda'_F(\mu) \neq 0$ and $\int \psi^2(x-t)dF(x) < \infty$.

A3. $\psi(x)$ is odd and uniformally continuous; $\int \partial\psi(x-t)/\partial t|_{t=\mu} dF(x)$ is finite and nonzero; $\int \psi^2(x-\mu)dF(x) < \infty$.

THEOREM 2. *Assume that F is a symmetric location distribution and that the allocation vector $\boldsymbol{p}$ is symmetric. Then, under either* (A1), (A2) *or* (A3),

$$\sqrt{n}(\hat{T}_n - \mu) \to N(0, \sigma^2(\boldsymbol{p}, F)), \tag{3.1}$$

in distribution, where under (A1) *and* (A2),

$$\sigma^2(\boldsymbol{p}, F) = \sum_{r=1}^{k} p_r A_{(r)}(\mu) \Bigg/ \left[\sum_{r=1}^{k} p_r \lambda'_{F_{(r)}}(\mu)\right]^2,$$

and under (A3),

$$\sigma^2(\boldsymbol{p}, F) = \sum_{r=1}^{k} p_r A_{(r)}(\mu) \Bigg/ \left[\sum_{r=1}^{k} p_r \int \partial\psi(x-t)/\partial t|_{t=\mu} dF_{(r)}(x)\right]^2.$$

The results in this section are straightforward extensions of the corresponding results in SRS. A sketch of the proof of the theorems is given in the Appendix. The results can be used to compare the efficiency between RSS and SRS. Furthermore, they can be used to determine optimal RSS schemes for M-estimates, as will be seen in the next section.

4. Optimal RSS schemes for M-estimates

In this section, we describe a method for determining an allocation vector such that the resultant RSS M-estimate is an asymptotically unbiased minimum variance estimator among all RSS M-estimates.

First we present two lemmas.

LEMMA 2. *Assume $\psi(x)$ is an odd function and F is a symmetric location distribution. Then for any r,*

(i) $A_{(r)}(\mu) = A_{(k-r+1)}(\mu)$.

(ii) *If, in addition, $\int \psi(x)\partial f(x+t)/\partial t dx$ is finite, then $\lambda'_{F_{(r)}}(\mu) = \lambda'_{F_{(k-r+1)}}(\mu)$.*

(iii) *If, in addition, $\int \partial\psi(x-t)/\partial t dF(x)$ is finite,*

$$\int \partial\psi(x-t)/\partial t|_{t=\mu} dF_{(r)}(x) = \int \partial\psi(x-t)/\partial t|_{t=\mu} dF_{(k-r+1)}(x).$$

LEMMA 3. (Kaur *et al.* (1996)) *Assume, for a fixed t,*

$$\frac{A_t}{C_t^2} \leq \frac{A_r}{C_r^2} \quad \textit{for all} \quad r \neq t,$$

and all $A_r > 0$, *then*

$$\frac{A_t}{C_t^2} \leq \frac{\sum_{r=1}^k p_r A_r}{\{\sum_{r=1}^k p_r C_r\}^2},$$

for all allocation vectors.

In order to obtain an asymptotically unbiased minimum variance RSS M-estimate, we need to minimize $\sigma^2(\boldsymbol{p}, F)$ with respect to the allocation vector $\boldsymbol{p}$. According to Lemma 3, an optimal allocation can be found by allocating all the quantifications to the r-th order statistic such that $A_{(r)}/C_r^2$ is the smallest where $C_r = \lambda'_{F_{(r)}}(\mu)$ under (A1) or (A2) and $C_r = \int \partial\psi(x-t)/\partial t|_{t=\mu} dF_{(r)}(x)$ under (A3). By Lemma 2, we only need to compare $k/2$ or $(k+1)/2$ such ratios according to whether k is even or odd to obtain the minimum.

In the remaining of this section, we apply the method above to two ψ functions and a variety of distribution families. The first ψ is given by

$$\psi_1 = \begin{cases} -1.5, & x < -1.5, \\ x, & |x| \leq 1.5, \\ 1.5, & x > 1.5. \end{cases}$$

The corresponding M-estimator T_n is a type of Winsorized mean. The other ψ is a smoothed "Hampel" given by

$$\psi_2 = \begin{cases} \sin(x/2.1), & |x| < 2.1\pi, \\ 0, & |x| \geq 2.1\pi. \end{cases}$$

The distribution families under consideration are: (i) standard normal distribution $N(0,1)$, (ii) $Cauchy(0,1)$ distribution and (iii) symmetrically contaminated normal distributions of forms $(1-\epsilon)N(0,1) + \epsilon N(0,9)$ and $(1-\epsilon)N(0,1) + \epsilon Cauchy(0,1)$, for $\epsilon = 0.05(0.05)0.5$. Here, $Cauchy(0,1)$ denotes the Cauchy distribution with location parameter 0 and scale parameter 1. We also examine the performance of the above RSS M-estimators with the RSS mean which is obtained with $\psi = x$ from the consideration of robustness.

For each combination of the above ψ functions and distribution families, we minimized $A_{(r)}/C_r^2$ with respect to r for set size $k = 3(1)10$ and found that the minimum is always attained at the middle rank, i.e., $r = k/2$ or $r = (k+1)/2$ according to whether k is even or odd. In other words, the optimal schemes prescribe to quantify the median. Therefore, if k is even, the order statistic $X_{(k/2)}$ should be quantified for half of the ranked sets and the order statistic $X_{((k/2)+1)}$ should be quantified for the other half. If k is odd, the order statistic $X_{((k+1)/2)}$ should be quantified for all the ranked sets.

We have tried to minimize $A_{(r)}/C_r^2$ for general ψ and F by treating r as a continuous variable and found that $r = (k+1)/2$ is a stationary point of $A_{(r)}/C_r^2$ but failed to establish whether or not it is a minimum point. Nevertheless, we believe it must be true that the quantification of only the median in all ranked sets is the optimal RSS scheme. It is in line with the principle, like that of importance sampling (see, e.g., Jones (1988)), that sampling should be made at the data points such that it is more likely for the estimator of a parameter to assume a value in the vicinity of the parameter. Therefore, we suggest the use of the scheme of quantifying only the median in practice for the M-estimates of the mean of a symmetric distribution. In any case, if the rule is not beyond doubt, the ratio $A_{(r)}/C_r^2$ can be checked for those suspected underlying distribution families. It should be noted that if the family is a location family, $A_{(r)}/C_r^2$ does not depend on the location parameter of the family.

5. The asymptotic relative efficiency of RSS M-estimates

In this section, we deal with the asymptotic relative efficiency (ARE) of the RSS M-estimates. We compare the RSS M-estimates (balanced or optimal) with the corresponding SRS M-estimates. We compare the optimal RSS M-estimates with the balanced RSS M-estimates. We also compare the RSS M-estimates with the RSS mean. The SRS M-estimate of μ is given by $T_n = T(F_n)$ where F_n is the empirical distribution of a simple random sample of size n. The SRS M-estimate has asymptotically a normal distribution with mean μ and variance $\sigma^2(F)$ given by either

$$\int \psi^2(x-\mu)dF(x)/[\lambda_F'(\mu)]^2$$

or

$$\int \psi^2(x-\mu)dF(x)\Big/\left[\int \partial\psi(x-t)/\partial t\,|_{t=\mu}dF(x)\right]^2$$

depending on the assumptions on $\psi(x)$. Hence the relative efficiency of the balanced RSS M-estimate to the SRS M-estimate is given by

$$ARE(\tilde{T}_n, T_n) = \int \psi^2(x-\mu)dF(x)/d,$$

where

$$d = \int \psi^2(x-\mu)dF(x) - \frac{1}{k}\sum_{r=1}^{k}\left[\int \psi(x-\mu)dF_{(r)}(x)\right]^2.$$

Obviously, the RSS M-estimate is always more efficient than its SRS counterpart. The relative efficiencies of the optimal RSS M-estimate to the SRS M-estimate and to the balanced RSS M-estimate are given, respectively, by

$$ARE(\hat{T}_n, T_n) = \frac{\sigma^2(F)}{\sigma^2(\boldsymbol{p}, F)}, \qquad ARE(\hat{T}_n, \tilde{T}_n) = \frac{\sigma_1^2(F)}{\sigma^2(\boldsymbol{p}, F)},$$

where

$$\sigma_1^2(F) = d/[\lambda_F'(\mu)]^2 \quad \text{or} \quad d\Big/\left[\int \partial\psi(x-t)/\partial t\,|_{t=\mu}dF(x)\right]^2,$$

depending on the assumptions on $\psi(x)$, and

$$\sigma^2(\boldsymbol{p}, F) = \begin{cases} A_{((k+1)/2)}(\mu)/d_1, & k \text{ odd}, \\ \frac{1}{2}[A_{(k/2)}(\mu) + A_{((k/2)+1)}(\mu)]/d_2, & k \text{ even}, \end{cases}$$

with

$$d_1 = [\lambda'_{F_{((k+1)/2)}}(\mu)]^2 \quad \text{or} \quad \left[\int \partial\psi(x-t)/\partial t|_{t=\mu} dF_{((k+1)/2)}(x)\right]^2,$$

$$d_2 = \frac{1}{4}[\lambda'_{F_{(k/2)}}(\mu) + \lambda'_{F_{((k/2)+1)}}(\mu)]^2$$

$$\text{or} \quad \frac{1}{4}\left[\int \partial\psi(x-t)/\partial t\,|_{t=\mu} dF_{(k/2)}(x) + \int \partial\psi(x-t)/\partial t\,|_{t=\mu} dF_{((k/2)+1)}(x)\right]^2,$$

depending on the assumptions on $\psi(x)$. The ARE's of the balanced RSS M-estimates to the RSS mean is given by

$$ARE(\tilde{T}_n, \tilde{\tilde{T}}_n) = \frac{\bar{\sigma}^2(F)}{\sigma_1^2(F)},$$

where $\tilde{\tilde{T}}_n$ is the balanced RSS mean and $\bar{\sigma}^2(F)$ is its asymptotic variance.

Table 1. The ARE's of the balanced RSS M-estimate w.r.t. SRS mean with ψ_1 for Cauchy, Normal and some contaminated Normal distributions.

Dist. k	$Cauchy(0,1)$	$N(0,1)$	$0.9N(0,1)$ $+0.1N(0,9)$	$0.7N(0,1)$ $+0.3N(0,9)$	$0.5N(0,1)$ $+0.5N(0,9)$
$k=2$	1.49	1.49	1.50	1.49	1.49
$k=3$	1.97	1.99	1.99	1.98	1.96
$k=4$	2.44	2.48	2.48	2.47	2.43
$k=5$	2.90	2.96	2.97	2.95	2.90
$k=6$	3.36	3.45	3.45	3.43	3.36
$k=7$	3.82	3.93	3.94	3.91	3.82
$k=8$	4.27	4.41	4.42	4.38	4.27
$k=9$	4.71	4.90	4.90	4.85	4.72
$k=10$	5.16	5.38	5.38	5.32	5.17

Table 2. The ARE's of the balanced RSS M-estimate w.r.t. SRS mean with ψ_2 for Cauchy, Normal and some contaminated Normal distributions.

Dist. k	$Cauchy(0,1)$	$N(0,1)$	$0.9N(0,1)$ $+0.1N(0,9)$	$0.7N(0,1)$ $+0.3N(0,9)$	$0.5N(0,1)$ $+0.5N(0,9)$
$k=2$	1.30	1.49	1.48	1.46	1.44
$k=3$	1.54	1.97	1.94	1.89	1.84
$k=4$	1.73	2.44	2.39	2.30	2.22
$k=5$	1.90	2.91	2.83	2.69	2.57
$k=6$	2.06	3.38	3.26	3.06	2.90
$k=7$	2.20	3.85	3.67	3.42	3.21
$k=8$	2.33	4.31	4.08	3.76	3.51
$k=9$	2.46	4.77	4.48	4.08	3.79
$k=10$	2.58	5.24	4.88	4.39	4.07

Table 3. The ARE's of the optimal RSS M-estimate w.r.t. SRS mean with ψ_1 for Cauchy, Normal and some contaminated Normal distributions.

Dist. k	$Cauchy(0,1)$	$N(0,1)$	$0.9N(0,1)$ $+0.1N(0,9)$	$0.7N(0,1)$ $+0.3N(0,9)$	$0.5N(0,1)$ $+0.5N(0,9)$
$k=3$	2.55	2.30	2.37	2.45	2.47
$k=4$	3.18	2.87	2.96	3.06	3.08
$k=5$	4.32	3.61	3.79	3.99	4.06
$k=6$	5.03	4.21	4.41	4.65	4.73
$k=7$	6.24	4.93	5.22	5.57	5.71
$k=8$	7.01	5.54	5.86	6.27	6.41
$k=9$	8.28	6.24	6.65	7.18	7.39
$k=10$	9.10	6.87	7.31	7.90	8.13

Table 4. The ARE's of the optimal RSS M-estimate w.r.t. SRS mean with ψ_2 for Cauchy, Normal and some contaminated Normal distributions.

Dist. k	$Cauchy(0,1)$	$N(0,1)$	$0.9N(0,1)$ $+0.1N(0,9)$	$0.7N(0,1)$ $+0.3N(0,9)$	$0.5N(0,1)$ $+0.5N(0,9)$
$k=3$	2.30	2.25	2.40	2.56	2.67
$k=4$	2.81	2.80	2.99	3.19	3.32
$k=5$	3.91	3.52	3.86	4.29	4.50
$k=6$	4.53	4.10	4.50	5.00	5.25
$k=7$	5.76	4.79	5.33	6.07	6.45
$k=8$	6.46	5.39	5.99	6.82	7.25
$k=9$	7.79	6.07	6.79	7.86	8.44
$k=10$	8.55	6.68	7.47	8.65	9.28

Table 5. The ARE's of the optimal RSS M-estimate w.r.t. balanced RSS M-estimate with ψ_1 for Cauchy, Normal and some contaminated Normal distributions.

Dist. k	$Cauchy(0,1)$	$N(0,1)$	$0.9N(0,1)$ $+0.1N(0,9)$	$0.7N(0,1)$ $+0.3N(0,9)$	$0.5N(0,1)$ $+0.5N(0,9)$
$k=3$	2.81	1.16	1.19	1.23	1.26
$k=4$	2.83	1.16	1.19	1.24	1.27
$k=5$	3.22	1.22	1.28	1.35	1.40
$k=6$	3.24	1.22	1.28	1.35	1.41
$k=7$	3.54	1.25	1.32	1.43	1.50
$k=8$	3.56	1.26	1.33	1.43	1.50
$k=9$	3.81	1.28	1.36	1.48	1.57
$k=10$	3.82	1.28	1.36	1.48	1.57

The ARE's of the balanced RSS M-estimates with ψ_1 and ψ_2 for the $Cauchy(0,1)$, $N(0,1)$ and some contaminated normal distributions are given, respectively, in Tables 1 and 2. The ARE's of the corresponding optimal RSS M-estimates with respect to SRS

Table 6. The ARE's of the optimal RSS M-estimate w.r.t. balanced RSS M-estimate with ψ_2 for Cauchy, Normal and some contaminated Normal distributions.

Dist. k	$Cauchy(0,1)$	$N(0,1)$	$0.9N(0,1)$ $+0.1N(0,9)$	$0.7N(0,1)$ $+0.3N(0,9)$	$0.5N(0,1)$ $+0.5N(0,9)$
$k=3$	1.50	1.14	1.24	1.36	1.45
$k=4$	1.62	1.15	1.25	1.39	1.50
$k=5$	2.05	1.21	1.37	1.59	1.76
$k=6$	2.20	1.21	1.38	1.63	1.81
$k=7$	2.62	1.25	1.45	1.78	2.01
$k=8$	2.77	1.25	1.47	1.82	2.07
$k=9$	3.17	1.27	1.51	1.93	2.23
$k=10$	3.31	1.28	1.53	1.97	2.28

Table 7. The ARE's of the balanced RSS M-estimate w.r.t. RSS mean for some contaminated Normal distributions.

Dist.	$0.9N(0,1)+0.1N(0,9)$		$0.7N(0,1)+0.3N(0,9)$		$0.5N(0,1)+0.5N(0,9)$	
k	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2
$k=2$	1.51	1.46	1.85	1.65	1.69	1.42
$k=3$	1.63	1.56	2.00	1.74	1.79	1.46
$k=4$	1.74	1.65	2.13	1.82	1.87	1.48
$k=5$	1.85	1.72	2.25	1.88	1.95	1.49
$k=6$	1.94	1.79	2.36	1.92	2.01	1.50
$k=7$	2.03	1.86	2.45	1.96	2.06	1.50
$k=8$	2.11	1.91	2.54	1.99	2.11	1.50
$k=9$	2.19	1.96	2.61	2.01	2.15	1.50
$k=10$	2.26	2.01	2.68	2.02	2.19	1.49

and balanced RSS ones are given in Tables 3, 4, 5 and 6. Table 7 gives the ARE's of the balanced RSS M-estimates to the RSS mean.

It can be seen from Tables 1–6 that, as expected, the balanced RSS M-estimates are much more efficient than their SRS counterparts and the efficiency increases as set size k increases, and that the optimal RSS M-estimates improve the balanced RSS M-estimates significantly further. Table 7 shows that the RSS M-estimates with ψ_1 and ψ_2 are more efficient than the RSS mean when the underlying distribution is not normal. It can also be noticed that the efficiency with ψ_1 is uniformly higher that the efficiency with ψ_2. This might be attributable to the fact that ψ_2 throws away entirely the information of the extreme observations while ψ_1 still makes use of those information but with adjustment. We do not elaborate this aspect further since it is beyond the scope of this article.

6. Results of simulation studies

In this section, we present some results of our simulation studies. The protocol of our simulation study is as follows. For each distribution and a given $n = mk$, we generate 1000 simple random samples of size n and 1000 ranked-set samples of size n and set size k (either balanced or optimal) by using the random number generating functions in Splus 4.5. For each sample, the M-estimate T_n, $\tilde{T}_n$ or $\hat{T}_n$ is computed according to whether the sample is simple random, balanced RSS or optimal RSS. Then the computed M-estimates are used to compute an approximation to the MSE by using the formula:

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(\hat{\mu}_j - \mu)^2,$$

where $\hat{\mu}_j$ is the M-estimate from the j-th sample, μ is the theoretical mean of the underlying distribution and N is the simulation size ($N = 1000$). The ratio of the approximated MSEs of the simple random sample M-estimate and the ranked-set sample M-estimate is then computed.

We report the results for four distributions: $N(0,1)$, $Cauchy(0,1)$, $0.9N(0,1)+0.1N(0,9)$ and $0.7N(0,1)+0.3N(0,9)$, with $\psi = \psi_1$, $n = 20, 60, 120, 240$, and $k = 2,4,5$ for balanced case and $k = 3,4,5$ for unbalanced case. The balanced RSS case is reported in Table 8 and the optimal RSS case is reported in Table 9. The simulation results show that, for small sample sizes, the balanced RSS M-estimates and optimal RSS M-estimates can achieve relative efficiencies comparable with the asymptotic relative efficiencies. It can also be seen that when n is large, say $n = 240$, the simulated relative efficiencies are quite in line with the theoretical asymptotic relative efficiencies. A noticeable feature is that the simulated relative efficiencies for small sample sizes are larger than their corresponding asymptotic relative efficiencies. This is not accidental. In fact, we can express the variances of the estimators in the form:

$$\frac{1}{n}\Delta_1(\boldsymbol{X}_n) + \frac{1}{n^2}\Delta_2(\boldsymbol{X}_n) + \cdots,$$

where $\boldsymbol{X}_n$ represents a sample of size n. Each of the SRS versions of the Δ terms can be improved by its RSS version. The asymptotic relative efficiency only accounts for the improvement on Δ_1 involved in the order $O(1/n)$ while the relative efficiency for small

Table 8. Simulated relative efficiencies of balanced RSS M-estimates with $\psi = \psi_1$.

		n			
dist.	k	20	60	120	240
$N(0,1)$	2	1.62	1.43	1.48	1.49
	4	2.48	2.59	2.46	2.47
	5	3.09	2.97	3.01	2.95
$Cauchy(0,1)$	2	1.52	1.47	1.51	1.49
	4	2.56	2.47	2.46	2.42
	5	3.00	2.94	2.91	2.88
$0.9N(0,1)$	2	1.50	1.48	1.48	1.50
$+0.1N(0,9)$	4	2.50	2.51	2.51	2.46
	5	2.88	2.93	2.95	2.94
$0.7N(0,1)$	2	1.53	1.53	1.51	1.48
$+0.3N(0,9)$	4	2.46	2.47	2.47	2.47
	5	2.92	2.93	2.94	2.96

Table 9. Simulated relative efficiencies of optimal RSS M-estimates with $\psi = \psi_1$.

		n			
dist.	k	20	60	120	240
$N(0,1)$	3	2.39	2.37	2.34	2.32
	4	2.85	2.84	2.90	2.89
	5	3.68	3.67	3.66	3.59
$Cauchy(0,1)$	3	2.59	2.51	2.54	2.58
	4	3.15	3.14	3.19	3.19
	5	4.36	4.35	4.35	4.29
$0.9N(0,1)$	3	2.35	2.42	2.36	2.37
$+0.1N(0,9)$	4	2.94	2.98	2.98	2.95
	5	3.75	3.81	3.77	3.80
$0.7N(0,1)$	3	2.46	2.46	2.44	2.45
$+0.3N(0,9)$	4	3.00	3.07	3.09	3.06
	5	4.04	4.04	4.01	3.98

samples also reflects the improvement involved in higher orders of $1/n$. In general, we can expect that the relative efficiencies for small samples be larger than the asymptotic relative effienciencies.

7. Concluding remarks

The M-estimates arise from the concern on the robustness of statistical procedures. We have dealt with the M-estimates in the context of ranked set sampling in this article. In particular, we obtained the asymptotic properties of the M-estimates for the mean of symmetric distributions and developed the method for finding optimal RSS scheme for the M-estimates. We also investigated the relative efficiencies of balanced RSS M-estimates and optimal RSS M-estimates. We found that the scheme of quantifying only the median in all the ranked sets is optimal for all the distribution families which we have considered including *Normal*, *Cauchy* and contaminated *Normal* distributions. Though we are unable to establish the result in general for any symmetric distribution, we have a reasonable ground to believe that it must be true in general. In any case, the method we developed can be used to check whether the scheme is optimal for any suspected underlying distribution. Of course, further research in this respect for a general theoretical result will be interesting and worthy.

In our discussion we assumed tacitly that the ranking in RSS is perfect. However, in practice, ranking errors have to be taken into account. Here we briefly discuss the case of imperfect ranking to end this article. Let us denote the distribution function of the r-th judgment ranked order statistic by $F_{[r]}$ to indicate the existence of ranking errors. A basic assumption we need is that $F = \frac{1}{k}\sum_{r=1}^{k} F_{[r]}$ which is satisfied if the judgment ranking is made consistently. In the balanced RSS case, though ranking errors could reduce the efficiency of the M-estimate, the consistency and asymptotic normality of the M-estimate remain intact and the RSS M-estimate is still more efficient than its SRS counterpart. In the case of unbalanced RSS, if we further assume the symmetry property that $F_{[r]}(\mu + x) = F_{[k-r+1]}(\mu - x)$, the lemmas and theorems in the previous sections still hold if wherever $F_{(r)}$ is replaced by $F_{[r]}$. The prescription of measuring only the median still provides a consistent and asymptotically normally distributed M-estimate. Though it might not necessarily be optimal, it can still be expected to be more efficient than the balanced RSS by reasons we argued at the end of Section 4. In the following, we describe a particular situation such that the symmetry property holds. Let $p_{s,r}$ be the probability at which the actual s-th order statistic is ranked as the r-th order statistic. For symmetric underlying distributions it is reasonable to assume that $p_{s,r} = p_{k-s+1,k-r+1}$. Without loss of generality, assume that $\mu = 0$. Then we have

$$\begin{aligned} F_{[r]}(x) &= \sum_{s=1}^{k} p_{s,r} F_{(s)}(x) = \sum_{s=1}^{k} p_{k-s+1,k-r+1} F_{(s)}(x) \\ &= \sum_{s=1}^{k} p_{s,k-r+1} F_{(k-s+1)}(x) = \sum_{s=1}^{k} p_{s,k-r+1}(1 - F_{(s)}(-x)) \\ &= 1 - F_{[k-r+1]}(-x). \end{aligned}$$

Appendix

Here we sketch the proofs of Theorem 1, Theorem 2 and Lemma 2.

PROOF OF THEOREM 1. Assume that $\psi(x,t)$ is nonincreasing in t and thus $\lambda_{F_P}(t)$ and $\lambda_{\hat{F}_n}(t)$ are nonincreasing. Since μ is an isolated root of $\lambda_{F_P}(t) = 0$, μ is the unique root and thus we get

$$\lambda_{F_P}(\mu + \varepsilon) < 0 < \lambda_{F_P}(\mu - \varepsilon), \qquad \text{for any} \quad \varepsilon > 0.$$

By the SLLN, $\lambda_{\hat{F}_n}(t)$ converges to $\lambda_{F_P}(t)$ wp1, each t. Thus,

$$\lim_{n\to\infty} P(\lambda_{\hat{F}_m}(\mu + \varepsilon) < 0 < \lambda_{\hat{F}_m}(\mu - \varepsilon), \text{ all } m \geq n) = 1.$$

Since $\lambda_{\hat{F}_m}(t)$ is nonincreasing and $\lambda_{\hat{F}_m}(\hat{T}_m) = 0$, thus

$$\lim_{n\to\infty} P(\mu - \varepsilon < \hat{T}_m < \mu + \varepsilon, \text{ all } m \geq n) = 1.$$

Assume that $\psi(x)$ is bounded. It's easy to get that $\lambda_{F_P}(t)$ and $\lambda_{\hat{F}_n}(t)$ are continuous. Since ψ is odd and F is symmetric location distribution, we can get $\lambda_{F_P}(\mu+\varepsilon) \times \lambda_{F_P}(\mu - \varepsilon) < 0$. Assume that $\lambda_{F_P}(\mu + \varepsilon) > 0$. By the SLLN, $\lambda_{\hat{F}_n}(t)$ converges to $\lambda_{F_P}(t)$ wp1, each t. Let $\varepsilon_1 < \min\{\frac{1}{2}|\lambda_{F_P}(\mu + \varepsilon)|, \frac{1}{2}|\lambda_{F_P}(\mu - \varepsilon)|\}$. Thus we have

$$P(\lambda_{\hat{F}_m}(\mu + \varepsilon) > \lambda_{F_P}(\mu + \varepsilon) - \varepsilon_1(> 0), \text{ all } m \geq n) \to 1,$$

and

$$P(\lambda_{\hat{F}_m}(\mu - \varepsilon) < \lambda_{F_P}(\mu - \varepsilon) + \varepsilon_1(< 0), \text{ all } m \geq n) \to 1.$$

Since $\lambda_{\hat{F}_n}(t)$ is continuous, there is a solution sequence $\{\hat{T}_n\}$ such that

$$\lim_{n\to\infty} P(\mu - \varepsilon < \hat{T}_m < \mu + \varepsilon, \text{ all } m \geq n) = 1. \qquad \square$$

PROOF OF THEOREM 2. Under A1, to obtain (3.1), it is equivalent to show that

$$\lim_{n\to\infty} P\left(\hat{Y}_{n\boldsymbol{p}} \le \frac{-n^{1/2}\sum_{r=1}^{k} p_r \lambda_{F_{(r)}}(\hat{t}_{z,n,\boldsymbol{p}})}{\hat{s}_{z,n,\boldsymbol{p}}}\right) = \Phi(z), \qquad \text{each} \quad z,$$

where

$$\hat{Y}_{n\boldsymbol{p}} = \left\{ n^{-1/2}\sum_{r=1}^{k}\sum_{i=1}^{n_r} \psi(X_{(r)i} - \hat{t}_{z,n,\boldsymbol{p}}) - n^{1/2}\sum_{r=1}^{k} p_r \lambda_{F_{(r)}}(\hat{t}_{z,n,\boldsymbol{p}}) \right\} \Big/ \hat{s}_{z,n,\boldsymbol{p}},$$

and

$$\hat{s}_{z,n,\boldsymbol{p}} = \left[\sum_{r=1}^{k} p_r A_{(r)}(\hat{t}_{z,n,\boldsymbol{p}})\right]^{1/2}, \qquad \hat{t}_{z,n,\boldsymbol{p}} = \mu + z\sigma(\boldsymbol{p}, F)n^{-1/2}.$$

It thus suffices to show that

$$\lim_{n\to\infty} P(\hat{Y}_{n\boldsymbol{p}} \le z) = \Phi(z), \qquad \text{each} \quad z,$$

and equivalently

$$P\left(n_r^{-1/2}\sum_{i=1}^{n_r} Y_{r\boldsymbol{p}ni} \le z\right) \quad \text{is} \quad AN(0, B_{r\boldsymbol{p}n}^2),$$

where

$$Y_{r\boldsymbol{p}ni} = \frac{\psi(X_{(r)i} - \hat{t}_{z,n,\boldsymbol{p}}) - \lambda_{F_{(r)}}(\hat{t}_{z,n,\boldsymbol{p}})}{\hat{s}_{z,n,\boldsymbol{p}}},$$

and

$$B_{r\boldsymbol{p}n}^2 = \left[\int \psi^2(x - \hat{t}_{z,n,\boldsymbol{p}})dF_{(r)}(x) - \left(\int \psi(x - \hat{t}_{z,n,\boldsymbol{p}})dF_{(r)}(x)\right)^2\right] \Big/ \hat{s}_{z,n,\boldsymbol{p}}^2.$$

Since $Y_{r\boldsymbol{p}ni}, 1 \le i \le n_r$, are I.I.D. with mean 0 and variance $B_{r\boldsymbol{p}n}^2$, each n, we may apply the double array CLT to get the result and it remains to verify the Lindberg condition

$$\lim_{n\to\infty} \int_{|y|>n_r^{1/2}\varepsilon} y^2 dF_{Y_{r\boldsymbol{p}n1}}(y) = 0, \qquad \text{every} \quad \varepsilon > 0,$$

or equivalently

$$\lim_{n\to\infty} \int_{|\psi(x-\hat{t}_{z,n,\boldsymbol{p}})-\lambda_{F_{(r)}}(\hat{t}_{z,n,\boldsymbol{p}})|>n_r^{1/2}\varepsilon} [\psi(x - \hat{t}_{z,n,\boldsymbol{p}}) - \lambda_{F_{(r)}}(\hat{t}_{z,n,\boldsymbol{p}})]^2 dF_{(r)}(x) = 0, \qquad \text{every} \quad \varepsilon > 0.$$

For any $\eta > 0$, we have for n sufficiently large that

$$\psi(x - (\mu + \eta)) - \lambda_{F_{(r)}}(\mu - \eta) \le \psi(x - \hat{t}_{z,n,\boldsymbol{p}}) - \lambda_{F_{(r)}}(\hat{t}_{z,n,\boldsymbol{p}}) \le \psi(x - (\mu - \eta)) - \lambda_{F_{(r)}}(\mu + \eta), \qquad \text{all} \quad x.$$

So putting $\mu(x) = \max\{|\psi(x-(\mu+\eta)) - \lambda_{F_{(r)}}(\mu-\eta)|, |\psi(x-(\mu-\eta)) - \lambda_{F_{(r)}}(\mu+\eta)|\}$, we get

$$\int_{|\psi(x-\hat{t}_{z,n,\boldsymbol{p}})-\lambda_{F_{(r)}}(\hat{t}_{z,n,\boldsymbol{p}})|>n_r^{1/2}\varepsilon} [\psi(x-\hat{t}_{z,n,\boldsymbol{p}}) - \lambda_{F_{(r)}}(\hat{t}_{z,n,\boldsymbol{p}})]^2 dF_{(r)}(x)$$
$$\leq \int_{\mu(x)>n_r^{1/2}\varepsilon} \mu^2(x) dF_{(r)}(x).$$

Hence the result follows.

In the following, we sketch the proof of normality under A3. Since $\psi(x-t)$ is differentiable in t, so is the function $\sum_{r=1}^{k}\sum_{i=1}^{n_r}\psi(X_{(r)i}-t)$, and we have

$$\sum_{r=1}^{k}\sum_{i=1}^{n_r}\psi(X_{(r)i}-\hat{T}_n) - \sum_{r=1}^{k}\sum_{i=1}^{n_r}\psi(X_{(r)i}-\mu) = (\hat{T}_n-\mu)\sum_{r=1}^{k}\sum_{i=1}^{n_r}\frac{\partial\psi(X_{(r)i}-t)}{\partial t}|_{t=\hat{T}_n^*},$$

where $|\hat{T}_n^* - \mu| \leq |\hat{T}_n - \mu|$. Since $\lambda_{\hat{F}_n}(\hat{T}_n) = 0$, we thus have

$$n^{1/2}(\hat{T}_n - \mu) = -\frac{\hat{A}_{n\boldsymbol{p}}}{\hat{B}_{n\boldsymbol{p}}},$$

where

$$\hat{A}_{n\boldsymbol{p}} = n^{-1/2}\sum_{r=1}^{k}\sum_{i=1}^{n_r}\psi(X_{(r)i}-\mu),$$

and

$$\hat{B}_{n\boldsymbol{p}} = n^{-1}\sum_{r=1}^{k}\sum_{i=1}^{n_r}\left.\frac{\partial\psi(X_{(r)i}-t)}{\partial t}\right|_{t=\hat{T}_n^*}.$$

Complete the proof using the CLT and Theorem 1.

To prove Theorem 2 under A3, a very important and critical step is to replace F by $F_{\boldsymbol{p}}$ and thus use the result $\lambda_{F_{\boldsymbol{p}}}(\mu) = 0$. Other steps are replacing F_n by $\hat{F}_n$, T_n by $\hat{T}_n$ and $\sigma^2(T,F)$ by $\sigma^2(\boldsymbol{p},F)$. □

PROOF OF LEMMA 2. Suppose $F(x) = F_0(x-\mu)$ and $f(x) = f_0(x-\mu)$. According to Lemma 1, to get (i), it suffices to show

$$\text{(A.1)} \qquad \int_{-\infty}^{+\infty}\psi^2(x-\mu)dF_{(r)}(x) - \int_{-\infty}^{+\infty}\psi^2(x-\mu)dF_{(k-r+1)}(x) = 0.$$

To obtain (ii) and (iii), it suffices to show that

$$\text{(A.2)} \qquad \left.\frac{\partial\int\psi(x-t)h(x-\mu)dx}{\partial t}\right|_{t=\mu} = 0,$$

and

$$\text{(A.3)} \qquad \int \partial\psi(x-t)/\partial t|_{t=\mu}h(x-\mu)dx = 0,$$

respectively, where

$$h(x-\mu) = f_0(x-\mu)(F_0^{r-1}(x-\mu)(1-F_0(x-\mu))^{k-r} \\ -F_0^{k-r}(x-\mu)(1-F_0(x-\mu))^{r-1}).$$

We apply transformation $y = x - \mu$ to (A.1) and (A.3), and $z = x - t$ to (A.3). By the odd property of ψ and symmetric property of F, it is not difficult to get the results. □

References

Bai, Z. and Chen, Z. (2001). On the theory of ranked-set sampling and its ramifications, *J. Statist. Plann. Inference*, Special issue in honor of C. R. Rao (in print).

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

Chen, Z. (1999). Density estimation using ranked-set sampling data, *Environmental and Ecological Statistics*, **6**, 135–146.

Chen, Z. (2000*a*). On ranked-set sample quantiles and their applications, *J. Statist. Plann. Inference*, **83**, 125–135.

Chen, Z. (2000*b*). The efficiency of ranked-set sampling relative to simple random sampling under multi-parameter families, *Statist. Sinica*, **10**, 247–263.

Chen, Z. (2001). The optimal ranked-set sampling scheme for inference on population quantiles, *Statist. Sinica*, **11**, 23–37.

Chen, Z. and Bai, Z. (2000). The optimal ranked-set sampling scheme for parametric families, *Sankhyā Ser. A*, **62**, 178–192.

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

Fei, H., Sinha, B. K. and Wu, Z. (1994). Estimation of parameters in two-parameter Weibull and extreme-value distributions using ranked set sample, *J. Statist. Res.*, **28**, 149–162.

Hettmansperger, T. P. (1995). The ranked-set sampling sign test, *J. Nonparametr. Statist.*, **4**, 263–270.

Jones, M. V. (1988). Importance sampling for bootstrap confidence intervals, *J. Amer. Statist. Assoc.*, **83**, 709–714.

Kaur, A., Patil, G. P. and Taillie, C. (1996). Ranked set sample sign test under unequal allocation, Tech. Report, 96–0692, Department of Statistics, Center for Statistical Ecology and Environmental Statistics, Penn State University.

Kaur, A., Patil, G. P. and Taillie, C. (1997). Unequal allocation models for ranked set sampling with skew distributions, *Biometrics*, **53**, 123–130.

Koti, K. M. and Babu, G. J. (1996). Sign test for ranked-set sampling, *Comm. Statist. Theory Method*, **25**(7), 1617–1630.

McIntyre, G. A. (1952). A method of unbiased selective sampling, using ranked sets, *Australian Journal of Agriculture Research*, **3**, 385–390.

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

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

Stokes, S. L. (1980*b*). Inference on the correlation coefficient in bivariate normal populations from ranked set samples, *J. Amer. Statist. Assoc.*, **75**, 989–995.

Stokes, S. L. (1995). Parametric ranked set sampling, *Ann. Inst. Statist. Math.*, **47**(3), 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.

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.*, **30**, 814–824.