Statistica Sinica **11**(2001), 23-37

THE OPTIMAL RANKED-SET SAMPLING SCHEME FOR INFERENCE ON POPULATION QUANTILES

Zehua Chen

National University of Singapore

Abstract: In this article, we consider the design of unbalanced ranked-set sampling in order to achieve certain optimality for inference on quantiles. We first derive the asymptotic properties of the unbalanced ranked-set sample quantiles for any unbalanced ranked-set sampling scheme. Then these properties are employed to develop a methodology for determining optimal ranked-set sampling schemes. In the case of inference on a single quantile, the optimal scheme results in an estimator of the quantile which is asymptotically unbiased and with minimum variance among all ranked-set sample (balanced or unbalanced) quantiles. The striking feature of the methodology is that it is distribution-free. The optimal schemes for inference on certain quantiles are computed. Some simulation studies are reported.

Key words and phrases: Asymptotic normality, Bahadur representation, optimal sampling design, quantile, ranked-set sampling.

1. Introduction

In areas such as agriculture, environment, ecology, sociology and others, one often encounter problems where measurement of 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 by judgment without actual measurement. The notion of ranked-set sampling (RSS) introduced by McIntyre (1952) provides, in such circumstances, an applicable scheme which incorporates judgment ranking into sampling so as to gain more information than simple random sampling (SRS) without additional expense. The original form of an RSS scheme can be described as follows. A set of k items is drawn from the population, the items of the set are ranked by judgment, and only the item ranked the smallest is quantified. Then another set of size k is drawn and ranked, and only the item ranked the second smallest is quantified. The procedure continues until the item ranked the largest in the kth set is quantified. This completes a cycle of the sampling. The cycle is then repeated for as many times as desired. This original form of RSS is referred to as balanced in the sense that each order statistic in the ranked sets is quantified the same number of times. A ranked-set sample consists of the measurements on the quantified items.

Since a ranked-set sample contains not only the measurements but also the ranks of the quantified items in the ranked-sets, it usually carries more information than a simple random sample of the same size. The research in the literature has been focusing on the comparison between procedures based on RSS and their counterparts based on SRS. The earliest such research is on the relative precision of ranked-set sample means to simple random sample means as estimators of the population mean. This is found in, e.g., McIntyre (1952), Takahasi and Wakimoto (1968), Dell and Clutter (1972). The estimation of variance is considered in Stokes (1980). The ranked-set empirical distribution as an estimator of the population distribution and the related RSS version of Kolmogorov-Smirnov test are studied in Stokes and Sager (1988). The RSS version of the Mann-Whitney-Wilcoxon test is dealt with in Bohn and Wolfe (1992). The RSS version of the sign test is treated in Hettmansperger (1995) and Koti and Babu (1996). The procedures based on ranked-set sample quantiles are tackled in Chen (2000). Density estimation using RSS data is considered in Chen (1999a).

In this article, we focus on the issue of how unbalanced RSS schemes can result in optimal schemes for certain statistical problems. Unbalanced RSS schemes were mentioned in McIntyre (1952) and Takahasi and Wakimoto (1968). It was observed that the Neyman allocation is optimal for estimating the population mean. However, the Neyman allocation is not a practical scheme since the allocation proportions depend on unknown parameters. The first serious work that touches on this issue is, to our knowledge, Stokes (1995). Stokes (1995) considered the following sampling scheme: a cycle of k sets of size k is drawn and ranked, and the order statistics with orders $r_1, \ldots, r_k$ are quantified in these k ranked sets, where the r_j's can be any integer from 1 to k and are not necessarily different. The cycle is then repeated m times to produce the data

$$\begin{array}{c}
X_{(r_1)1}, X_{(r_1)2}, \ldots, X_{(r_1)m};\\
X_{(r_2)1}, X_{(r_2)2}, \ldots, X_{(r_2)m};\\
\ldots, \quad \ldots, \quad \ldots, \quad \ldots\\
X_{(r_k)1}, X_{(r_k)2}, \ldots, X_{(r_k)m}.
\end{array}$$

Under the framework of location-scale families, Stokes (1995) tackled the choice of the r_j's in order to obtain the most efficient best linear unbiased estimates for the parameters of location-scale families. However, the method given by Stokes (1995) has some limitations. The most serious limitation is that the method is based on asymptotic properties when k (in our notation) is large, which is impractical. A more reasonable and practical approach is proposed by Chen and Bai (1998) in parametric settings. They considered unbalanced RSS schemes described as follows. Let n sets of size k items be drawn from the population and each of them be ranked by judgment. Then, for $r = 1, \ldots, k$, n_r sets are randomly chosen and the rth order statistic is quantified in each. Here $0 \leq n_r \leq n$ and $\sum n_r = n$. An unbalanced ranked-set sample is represented by

$$\begin{array}{l} X_{(1)1}, X_{(1)2}, \ldots, X_{(1)n_1}; \\ X_{(2)1}, X_{(2)2}, \ldots, X_{(2)n_2}; \\ \ \ldots, \quad \ldots, \quad \ldots, \quad \ldots; \\ X_{(k)1}, X_{(k)2}, \ldots, X_{(k)n_k}. \end{array} \tag{1}$$

Chen and Bai (1998) considered both maximum likelihood estimates and best linear unbiased estimates for the parameters of the underlying parametric family. They developed methodology for determining optimal unbalanced schemes according to certain optimality criteria based on the asymptotic variance-covariance matrix of the estimates when n (not k) is large. Other inferences on unbalanced RSS schemes include Kaur, Patil and Taillie (1997, 1998a,b), and Muttlak (1998).

In this article, we develop a method for the design of optimal unbalanced RSS schemes for inference on quantiles. The attracting feature of our method is that it is distribution-free. The article is arranged as follows. In Section 2, we give the definition of ranked-set sample quantiles and describe the notation. The asymptotic properties of the ranked-set sample quantiles when n is large are presented in Section 3. Section 4 is devoted to the development of the method for the design of optimal RSS schemes. The asymptotic relative efficiencies of optimal RSS schemes are discussed in Section 5. A simulation study is reported in Section 6. Some discussion on imperfect ranking and other issues is given in Section 7.

2. Notation and Definitions

The cumulative distribution function (CDF) and probability density function (PDF) of the underlying distribution are denoted by F and f. Although in most part we assume the judgment ranking in RSS is perfect, we also allow the possibility of imperfect ranking in some of the results discussed in this article. To distinguish perfect ranking from imperfect ranking, we denote by $X_{(r)}$ the ranked order statistic when ranking is perfect and by $X_{[r]}$ when there is a possibility of imperfect ranking. The CDF and PDF of the rth judgment-ranked order statistic in a set of size k are denoted by $F_{(r)}$ and $f_{(r)}$ if ranking is perfect, and by $F_{[r]}$ and $f_{[r]}$ if ranking might be imperfect. By a probability vector is meant a vector whose components are non-negative and sum to 1. Let $\mathbf{q} = (q_1, \ldots, q_k)'$ denote a probability vector. Define $f_{\mathbf{q}} = \sum_{r=1}^{k} q_r f_{[r]}$ and $F_{\mathbf{q}} = \sum_{r=1}^{k} q_r F_{[r]}$. Note that $f_{\mathbf{q}}$ is a PDF and $F_{\mathbf{q}}$ is the corresponding CDF. The distribution with CDF $F_{\mathbf{q}}$ can be regarded as the sampling distribution of the unbalanced RSS scheme with $n_r = [[nq_r]]$, $[[x]]$ denoting the integer nearest to x. Denote by ξ_p the pth quantile of F and by $\xi_{\mathbf{q},p}$ the pth quantile of $F_{\mathbf{q}}$.

Define the unbalanced ranked-set empirical distribution function by $\hat{F}_{\mathbf{q}_n}(x) = \sum_{r=1}^{k} \frac{n_r}{n} \hat{F}_{[r]n_r}(x)$, where $\hat{F}_{[r]n_r}(x) = \frac{1}{n_r} \sum_{i=1}^{n_r} I\{X_{[r]i} \leq x\}$, and $n = \sum n_r$. The pth quantile of $\hat{F}_{\mathbf{q}_n}$ is $\hat{\xi}_{\mathbf{q}_n,p} = \inf\{x : \hat{F}_{\mathbf{q}_n}(x) \geq p\}$. The quantile $\hat{\xi}_{\mathbf{q}_n,p}$ is referred to as the pth unbalanced ranked-set sample quantile.

Let the $X_{[r]i}$'s be ordered from smallest to largest and denote the ordered quantities by $Y_{(1:n)} \leq \cdots \leq Y_{(j:n)} \leq \cdots \leq Y_{(n:n)}$. The $Y_{(j:n)}$'s are then referred to as the unbalanced ranked-set order statistics.

3. The Asymptotic Properties of the Unbalanced Ranked-set Sample Quantiles

We state the asymptotic properties of the unbalanced ranked-set sample quantiles in this section. The properties include strong consistency, Bahadur representation and asymptotic normality. These results apply whether ranking is perfect or not.

Theorem 1. *Suppose $n_r/n = q_r + O(n^{-1})$ for $r = 1, \ldots, k$. If $0 < p < 1$ and $f_{\mathbf{q}}(\xi_{\mathbf{q},p}) > 0$, then, with probability 1,*

$$|\hat{\xi}_{\mathbf{q}_n,p} - \xi_{\mathbf{q},p}| \leq \frac{2(\log n)^2}{f_{\mathbf{q}}(\xi_{\mathbf{q},p}) n^{1/2}}$$

for all sufficiently large n.

The next result is the Bahadur representation of the unbalanced ranked-set sample quantile.

Theorem 2. *Suppose $n_r/n = q_r + O(n^{-1})$. If $f_{\mathbf{q}}$ is positive in a neighborhood of $\xi_{\mathbf{q},p}$ and is continuous at $\xi_{\mathbf{q},p}$, then*

$$\hat{\xi}_{\mathbf{q}_n,p} = \xi_{\mathbf{q},p} + \frac{p - \hat{F}_{\mathbf{q}_n}(\xi_{\mathbf{q},p})}{f_{\mathbf{q}}(\xi_{\mathbf{q},p})} + R_n,$$

where, with probability one, $R_n = O(n^{-3/4}(\log n)^{3/4})$ as $n \to \infty$.

The asymptotic normality of the unbalanced ranked-set sample quantiles follows as an immediate consequence of the Bahadur representation.

Corollary 1. *Under the same assumptions as in Theorem 2,*

$$\sqrt{n}(\hat{\xi}_{\mathbf{q}_n,p} - \xi_{\mathbf{q},p}) \to N\left(0, \frac{\sigma_{k,p}^2(\mathbf{q})}{f_{\mathbf{q}}^2(\xi_{\mathbf{q},p})}\right),$$

in distribution, where $\sigma_{k,p}^2(\mathbf{q}) = \sum_{r=1}^{k} q_r F_{[r]}(\xi_{\mathbf{q},p})[1 - F_{[r]}(\xi_{\mathbf{q},p})]$.

A more general result is the joint asymptotic normality of several unbalanced ranked-set sample quantiles, as given below.

Corollary 2. *Let $0 < p_1 < \cdots < p_l < 1$ be l probabilities. Let $\hat{\boldsymbol{\xi}} = (\hat{\xi}_{\mathbf{q}_n,p_1}, \ldots, \hat{\xi}_{\mathbf{q}_n,p_l})'$ and $\boldsymbol{\xi} = (\xi_{\mathbf{q},p_1}, \ldots, \xi_{\mathbf{q},p_l})'$. Then $\sqrt{n}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \to N(0, \Sigma)$, where Σ is a positive definite matrix and, for $i < j$, the (i,j)th entry of Σ is*

$$\sigma_{ij} = \frac{\sum_{r=1}^{k} q_r F_{[r]}(\xi_{\mathbf{q},p_i})[1 - F_{[r]}(\xi_{\mathbf{q},p_j})]}{f_{\mathbf{q}}(\xi_{\mathbf{q},p_i}) f_{\mathbf{q}}(\xi_{\mathbf{q},p_j})}.$$

The results in this section are just extensions of the results for balanced RSS found in Chen (2000). What is important is the implication of these results in the design of optimal RSS schemes for inference on quantiles, as is seen in the next section.

4. Method for the Determination of Optimal RSS Schemes for Inference on Quantiles When Ranking Is Perfect

In this section, we assume that ranking in RSS is perfect. Then

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

and hence $F_{(r)}(x) = B(r, k-r+1, F(x))$, where $B(r,s,t)$ denotes the CDF of the beta distribution with parameters r and s.

(i) *The optimal RSS scheme for inference on a single quantile when ranking is perfect.* First we consider inference on a single quantile, say the pth. We have

$$F_{\mathbf{q}}(\xi_p) = \sum_{r=1}^{k} q_r F_{(r)}(\xi_p) = \sum_{r=1}^{k} q_r B(r, k-r+1, p), \tag{2}$$

which is completely determined by p and the probability vector $\mathbf{q}$. Let the rightmost sum in (2) be denoted by $s(\mathbf{q}, p)$, that is, $s(\mathbf{q}, p) = F_{\mathbf{q}}(\xi_p)$. When no confusion is caused, $s(\mathbf{q}, p)$ is abbreviated as s. Equality (2) indicates that the pth quantile of F is the sth quantile of $F_{\mathbf{q}}$. This is a crucial fact for our development of optimal schemes. Note that

$$f_{\mathbf{q}}(\xi_p) = \sum_{r=1}^{k} q_r \frac{k!}{(r-1)!(k-r)!} p^{r-1}(1-p)^{k-r} f(\xi_p).$$

Then Corollary 1 implies that $\sqrt{n}(\hat{\xi}_{\mathbf{q}_n,s} - \xi_p) \to N\left(0, \frac{V(\mathbf{q},p)}{f^2(\xi_p)}\right)$ in distribution, where

$$V(\mathbf{q}, p) = \frac{\sum_{r=1}^{k} q_r B(r, k-r+1, p)[1 - B(r, k-r+1, p)]}{[\sum_{r=1}^{k} q_r \frac{k!}{(r-1)!(k-r)!} p^{r-1}(1-p)^{k-r}]^2}. \tag{3}$$

The argument above implies that the sth unbalanced ranked-set sample quantile with sampling distribution $F_{\mathbf{q}}$ provides an asymptotically unbiased estimator of ξ_p with asymptotic variance $V(\mathbf{q}, p)/(nf^2(\xi_p))$. Hence, we can find, by minimizing $V(\mathbf{q}, p)$ with respect to $\mathbf{q}$, a sampling distribution which results in an asymptotically unbiased minimum variance estimator among all ranked-set sample quantiles. The algorithm for determining the optimal sampling distribution and the corresponding s is as follows.

Algorithm 1. The following steps determine an optimal RSS scheme for the inference on ξ_p when ranking is perfect.

Step 1. Determination of the optimal sampling distribution. Minimize $V(\mathbf{q}, p)$ with respect to $\mathbf{q}$ and derive the minimizer $\mathbf{q}^* = (q_1^*, \ldots, q_k^*)'$. Then the optimal sampling distribution is determined as $F_{\mathbf{q}^*}$.

Step 2. Determination of rank s^*. The rank s^* is determined as $s^* = F_{\mathbf{q}^*}(\xi_p) = \sum q_r^* B(r, k-r+1, p)$.

The implementation of the algorithm does not pose any computational difficulty. The main computation is the minimization of $V(\mathbf{q}, p)$. According to a result of Chen and Bai (1998), the minimum of $V(\mathbf{q}, p)$ can be attained at a $\mathbf{q}$ that has at most two non-zero elements. Therefore, the minimum of $V(\mathbf{q}, p)$ can be searched on $k!/[(k-2)!2!]$ unit line segments. We wrote a trivial program in Splus to carry out this search. There are many other ways to implement the minimization as well. As a referee pointed out, even elementary optimization software such as the Solver add-in in Excel can settle the problem.

In the following, we consider some properties of $V(\mathbf{q}, p)$ and $s(\mathbf{q}, p)$ that give rise to some desirable properties of the optimal schemes.

Lemma 1. *Let $\mathbf{q}$ be any probability vector and let $\tilde{\mathbf{q}}$ be the probability vector whose rth element, $\tilde{q}_r$, equals the $(k-r+1)$st element, q_{k-r+1}, of $\mathbf{q}$. Then we have, for $p = 0.5$, $V(\mathbf{q}, p) = V(\tilde{\mathbf{q}}, p) = V(\frac{1}{2}(\mathbf{q} + \tilde{\mathbf{q}}), p)$.*

The lemma relies on the following observation. Let $c_r(p) = B(r, k-r+1, p)$ and $d_r(p) = \frac{k!}{(r-1)!(k-r)!} p^{r-1}(1-p)^{k-r}$. If $p = 0.5$ then we have $c_r(p) = c_{k-r+1}(p)$, $d_r(p) = d_{k-r+1}(p)$ for all r.

Thus for any probability vector $\mathbf{q}$ there is a probability vector given by $\mathbf{q}^* = \frac{1}{2}(\mathbf{q} + \tilde{\mathbf{q}})$ such that $V(\mathbf{q}, p) = V(\mathbf{q}^*, p)$. Hence the optimal probability vector $\mathbf{q}^*$ can be made symmetric, i.e, the elements of $\mathbf{q}^*$ satisfy $q_r^* = q_{k-r+1}^*$.

Lemma 2. *If $p = 0.5$ and $\mathbf{q}$ is symmetric, then $s(\mathbf{q}, 0.5) = \sum_{r=1}^k q_r F_{(r)}(\xi_{0.5}) = 0.5$, where $\xi_{0.5}$ is the median of F.*

To verify the lemma, first consider the case that F is symmetric. If F is symmetric about μ then, for any symmetric $\mathbf{q}$, $F_{\mathbf{q}}$ is also symmetric about μ. Note that, in this case, $\mu = \xi_{0.5}$. Hence we have $s(\mathbf{q}, 0.5) = 0.5$. But, since the quantities $F_{(r)}(\xi_{0.5})$ do not depend on F, the lemma follows.

Lemma 3. *Let* $\mathbf{q}$ *be any probability vector. Let* $\tilde{\mathbf{q}}$ *be the probability vector obtained by reversing the components of* $\mathbf{q}$. *Then, for any* $0 < p < 1$, $V(\mathbf{q}, p) = V(\tilde{\mathbf{q}}, 1-p)$, $s(\mathbf{q}, p) = 1 - s(\tilde{\mathbf{q}}, 1-p)$.

This lemma follows from the equalities: $c_r(p) = 1 - c_{k-r+1}(1-p)$,$d_r(p) = d_{k-r+1}(1-p)$. These can be easily verified by the definition of $c_r(p)$and $d_r(p)$.

Lemma 3 implies that if $\mathbf{q}^*$ is an optimal probability vector and $s(\mathbf{q}^*, p)$ is the corresponding rank for the inference on the pth quantile, then $\tilde{\mathbf{q}}^*$ is an optimal probability vector and $1 - s(\mathbf{q}^*, p)$ is the corresponding rank for the inference on the $(1-p)$th quantile. In other words, the optimal schemes have a symmetric structure. Hence, we only need to compute the optimal schemes for $0 < p \leq 0.5$.

We now report computed optimal schemes for $p = 0.05$ to 0.5 in steps of 0.05, $k = 2, \ldots, 10$. It turned out that in all cases except for $p = 0.5$, optimal probability vectors have only one non-zero component. In the case of $p = 0.5$, optimal probabilities are equal on the medians of the sets of size k. That is, if k is odd then $q^*_{(k+1)/2} = 1$, and if k is even then $q^*_{k/2} = q^*_{k/2+1} = 0.5$. The index $r(\mathbf{q}^*)$ of the non-zero component of the probability vector and the corresponding $s(\mathbf{q}^*, p)$ of the optimal designs for $p = 0.05$ to 0.45 in steps of 0.05, $k = 2, \ldots, 10$, are given in Table 1.

Table 1. Optimal unbalanced RSS designs for estimating a single quantile ξ_p, for selected p and set size k, based on minimizing asymptotic variance and assuming perfect ranking.

$p\backslash k$		2	3	4	5	6	7	8	9	10
0.05	$r(\mathbf{q}^*)$	1	1	1	1	1	1	1	1	1
	$s(\mathbf{q}^*, p)$	0.10	0.14	0.19	0.23	0.26	0.30	0.34	0.37	0.40
0.10	$r(\mathbf{q}^*)$	1	1	1	1	1	1	1	2	2
	$s(\mathbf{q}^*, p)$	0.19	0.27	0.34	0.41	0.47	0.52	0.57	0.23	0.26
0.15	$r(\mathbf{q}^*)$	1	1	1	1	2	2	2	2	2
	$s(\mathbf{q}^*, p)$	0.28	0.39	0.48	0.56	0.22	0.28	0.34	0.40	0.46
0.20	$r(\mathbf{q}^*)$	1	1	1	2	2	2	2	2	3
	$s(\mathbf{q}^*, p)$	0.36	0.49	0.59	0.26	0.34	0.42	0.50	0.56	0.32
0.25	$r(\mathbf{q}^*)$	1	1	2	2	2	2	3	3	3
	$s(\mathbf{q}^*, p)$	0.44	0.58	0.26	0.37	0.47	0.56	0.32	0.40	0.47
0.30	$r(\mathbf{q}^*)$	1	1	2	2	2	3	3	3	4
	$s(\mathbf{q}^*, p)$	0.51	0.66	0.35	0.47	0.58	0.35	0.45	0.54	0.35
0.35	$r(\mathbf{q}^*)$	1	2	2	2	3	3	3	4	4
	$s(\mathbf{q}^*, p)$	0.58	0.28	0.44	0.57	0.35	0.47	0.57	0.39	0.49
0.40	$r(\mathbf{q}^*)$	1	2	2	3	3	3	4	4	5
	$s(\mathbf{q}^*, p)$	0.64	0.35	0.52	0.32	0.46	0.58	0.41	0.52	0.37
0.45	$r(\mathbf{q}^*)$	1	2	2	3	3	4	4	5	5
	$s(\mathbf{q}^*, p)$	0.70	0.43	0.61	0.41	0.56	0.39	0.52	0.38	0.50

(ii) *The optimal RSS schemes for simultaneous inference on several quantiles when ranking is perfect.* Corollary 2 can be applied to determine optimal RSS schemes for simultaneous inference on several quantiles, as Corollary 1 is applied in the single quantile case. Without loss of generality, we discuss how this can be done for inference involving two quantiles.

Let $s_1 = s(\mathbf{q}, p_1) = F_{\mathbf{q}}(\xi_{p_1})$ and $s_2 = s(\mathbf{q}, p_2) = F_{\mathbf{q}}(\xi_{p_2})$. Then the p_1th and p_2th quantiles, ξ_{p_1} and ξ_{p_2}, of F are the s_1th and s_2th quantiles of $F_{\mathbf{q}}$ respectively. Note that $f_{\mathbf{q}}(\xi_{p_1}) = \sum_{r=1}^{k} q_r d_r(p_1) f(\xi_{p_1})$, $f_{\mathbf{q}}(\xi_{p_2}) = \sum_{r=1}^{k} q_r d_r(p_2) f(\xi_{p_2})$. Then it follows from Corollary 2 that

$$\sqrt{n}\left[\begin{pmatrix}\hat{\xi}_{\mathbf{q}n,s_1}\\ \hat{\xi}_{\mathbf{q}n,s_2}\end{pmatrix} - \begin{pmatrix}\xi_{p_1}\\ \xi_{p_2}\end{pmatrix}\right] \to N\left(0, \Sigma(\mathbf{q})\right),$$

in distribution, where $\Sigma(\mathbf{q}) = C^{-1}B^{-1}(\mathbf{q})A(\mathbf{q})B^{-1}(\mathbf{q})C^{-1}$, and

$$C = \begin{pmatrix} f(\xi_{p_1}) & 0 \\ 0 & f(\xi_{p_2}) \end{pmatrix},$$

$$B(\mathbf{q}) = \begin{pmatrix} \sum_{r=1}^{k} q_r d_r(p_1) & 0 \\ 0 & \sum_{r=1}^{k} q_r d_r(p_2) \end{pmatrix},$$

$$A(\mathbf{q}) = \begin{pmatrix} \sum_{r=1}^{k} q_r c_r(p_1)[1-c_r(p_1)] & \sum_{r=1}^{k} q_r c_r(p_1)[1-c_r(p_2)] \\ \sum_{r=1}^{k} q_r c_r(p_1)[1-c_r(p_2)] & \sum_{r=1}^{k} q_r c_r(p_2)[1-c_r(p_2)] \end{pmatrix}.$$

Let $V(\mathbf{q}) = B^{-1}(\mathbf{q})A(\mathbf{q})B^{-1}(\mathbf{q})$. Optimal RSS schemes can be determined based on $V(\mathbf{q})$. However, unlike the case of a single quantile, we are faced with the choice of optimality criteria. Various criteria can be considered, such as D-optimality, A-optimality, E-optimality, etc. Suppose that the choice of criteria has been made and that the optimality criterion entails the minimization of a function of $V(\mathbf{q})$, say, $G(V(\mathbf{q}))$. Then the algorithm for determining optimal RSS schemes for simultaneous inference on several quantiles can be described as follows.

Algorithm 2. The following steps determine the optimal scheme for the simultaneous inference on $\xi_{p_1}, \ldots, \xi_{p_l}$ when ranking is perfect.

Step 1. Minimize $G(V(\mathbf{q}))$ with respect to $\mathbf{q}$ to determine the optimal probability vector $\mathbf{q}^* = (q_1^*, \ldots, q_k^*)'$.

Step 2. Compute $s_j^* = \sum q_r^* B(r, k-r+1, p_j)$ for $j = 1, \ldots, l$.

For illustration, consider the criterion of D-optimality in what follows. This entails the minimization of $|V(\mathbf{q})|$, the determinant of $V(\mathbf{q})$. As important examples, we computed the D-optimal RSS schemes for inference on the pairs (ξ_p, ξ_{1-p}) for $p = 0.01, 0.05, 0.1, 0.15, 0.2$ and 0.25. The optimal probability vectors and the corresponding vector $s(\mathbf{q}^*, \boldsymbol{p}\ ) = (s(\mathbf{q}^*, p), s(\mathbf{q}^*, 1-p))$ are given in Table 2.

Table 2. Optimal unbalanced RSS designs for estimating a pair of quantiles (ξ_p, ξ_{1-p}), for selected p and set size k, based on minimizing asymptotic generalized variance and assuming perfect ranking.

$p\backslash k$		3	4	5
0.01	$\mathbf{q}^*$	(0.5, 0, 0.5)	(0.5,0,0,0.5)	(0.5,0,0,0,0.5)
	$s(\mathbf{q}^*, \boldsymbol{p}\ )$	(0.015,0.985)	(0.020,0.980)	(0.025, 0.975)
0.05	$\mathbf{q}^*$	(0.5, 0, 0.5)	(0.5,0,0,0.5)	(0.5,0,0,0,0.5)
	$s(\mathbf{q}^*, \boldsymbol{p}\ )$	(0.071,0.929)	(0.093,0.907)	(0.113, 0.887)
0.10	$\mathbf{q}^*$	(0.5, 0, 0.5)	(0.5,0,0,0.5)	(0.5,0,0,0,0.5)
	$s(\mathbf{q}^*, \boldsymbol{p}\ )$	(0.136,0.864)	(0.172,0.828)	(0.205, 0.795)
0.15	$\mathbf{q}^*$	(0.5, 0, 0.5)	(0.5,0,0,0.5)	(0.5,0,0,0,0.5)
	$s(\mathbf{q}^*, \boldsymbol{p}\ )$	(0.195,0.805)	(0.239,0.761)	(0.278, 0.722)
0.20	$\mathbf{q}^*$	(0, 1, 0)	(0,0.5,0.5,0)	(0,0.5,0,0.5,0)
	$s(\mathbf{q}^*, \boldsymbol{p}\ )$	(0.104,0.896)	(0.104,0.896)	(0.135, 0.865)
0.25	$\mathbf{q}^*$	(0, 1, 0)	(0,0.5,0.5,0)	(0, 0, 1, 0, 0)
	$s(\mathbf{q}^*, \boldsymbol{p}\ )$	(0.156,0.844)	(0.156,0.844)	(0.104, 0.896)

5. Asymptotic Relative Efficiency of Optimal RSS Schemes When Ranking Is Perfect

In this section, we discuss the asymptotic relative efficiency (ARE) of optimal RSS schemes with respect to SRS schemes and also compare them with balanced RSS schemes. The SRS counterpart of the estimator of ξ_p is the pth sample quantile $\hat{\xi}_p$, asymptotically normal with mean ξ_p and variance $p(1-p)/[nf^2(\xi_p)]$. (See, e.g., Serfling (1980, Chapter 2)). The balanced RSS counterpart of the estimator of ξ_p is given by the pth balanced ranked-set sample quantile $\tilde{\xi}_{mk,p}$, asymptotically normal with mean ξ_p and variance $(1/k)\sum c_r(p)[1-c_r(p)]/[mkf^2(\xi_p)]$, see Chen (2000). Hence the relative efficiencies of the optimal unbalanced RSS scheme and the balanced RSS scheme with respect to the SRS scheme for estimating ξ_p are given, respectively, by

$$ARE(\tilde{\xi}_{mk,p}, \hat{\xi}_p) = \frac{p(1-p)}{(1/k)\sum_{r=1}^{k} c_r(p)[1-c_r(p)]},$$

$$ARE(\hat{\xi}_{\mathbf{q}^*,s(p)}, \hat{\xi}_p) = \frac{p(1-p)}{\sum_{r=1}^{k} q_r^* c_r(p)[1-c_r(p)]/[\sum_{r=1}^{k} q_r^* d_r(p)]^2}.$$

The ARE of the optimal unbalanced RSS schemes with respect to the SRS schemes, for $k = 2, \ldots, 10$ and $p = 0.05$ to 0.5 in steps of 0.05, are given in Table 3. It can be seen that the gain in efficiency by using the optimal unbalanced RSS schemes is large, the n quantified order statistics do almost as well as a simple random sample of size kn. It is also interesting to compare the ARE of the optimal unbalanced RSS schemes with the ARE of the balanced RSS schemes.

As a function of p, the ARE of the balanced RSS schemes is a bow shaped curve with its maximum at $p = 0.5$ and reduces to 1 at both ends, see Chen (2000). Though the efficiency gain by using balanced RSS for the inference on medians is quite significant, the efficiency gain for the inference on extreme quantiles is almost negligible. However, the optimal unbalanced RSS schemes achieve about the same efficiency gain for all quantiles. This is due to the fact that every quantile of the underlying distribution is made a central quantile of the sampling distribution of the optimal unbalanced RSS scheme. This draws a similarity to other statistical procedures, such as importance sampling and saddlepoint approximation. They share the idea that if data values are sampled in a way which makes it more likely for a statistic to assume a value in the vicinity of a given point of interest, then that point may be estimated or approximated with greater accuracy.

Table 3. The ARE of optimal unbalanced RSS schemes with respect to SRS schemes, for selected p and set size k, assuming perfect ranking.

p \ k	2	3	4	5	6	7	8	9	10
0.05	1.949	2.848	3.698	4.501	5.258	5.970	6.639	7.267	7.853
0.10	1.895	2.690	3.392	4.005	4.537	4.991	5.375	6.118	6.954
0.15	1.838	2.528	3.084	3.519	4.054	4.906	5.680	6.365	6.956
0.20	1.778	2.361	2.775	3.365	4.279	4.966	5.517	5.933	6.682
0.25	1.714	2.189	2.763	3.590	4.243	4.714	5.337	6.142	6.783
0.30	1.647	2.014	2.879	3.569	4.025	4.734	5.483	6.014	6.569
0.35	1.576	2.095	2.912	3.433	4.059	4.817	5.299	6.005	6.719
0.40	1.500	2.182	2.874	3.307	4.161	4.671	5.368	6.047	6.500
0.45	1.419	2.233	2.773	3.463	4.102	4.680	5.408	5.879	6.695
0.50	1.333	2.250	2.618	3.516	3.896	4.785	5.172	6.056	6.447

6. Result of Simulation Studies

In this section, we report some simulation results that are part of a larger study. For a given distribution and a give sample size n, we generate 1000 simple random samples and then, for each k, we generate 1000 unbalanced ranked-set samples with set size k and sample size n according to the optimal RSS schemes. For each simple random sample, the pth sample quantile is computed. For each ranked-set sample, the corresponding $s^*(p)$th quantile is computed. The computed sample quantiles are used to compute an approximation to the MSE, $M\hat{S}E = \frac{1}{N}\sum_{i=1}^{N}(\hat{\xi}_j - \xi_p)^2$, where $\hat{\xi}_j$ is the quantile of the jth sample, ξ_p is the theoretical pth quantile of the underlying distribution and N is the simulation size ($N = 1000$). The ratio of the approximated MSEs of the simple random sample quantiles and the unbalanced ranked-set sample quantiles is then computed. In

Tables 4 and 5, we report these ratios for two underlying distributions: the Extreme Value distribution with location parameter 0 and scale parameter 1 and the log-normal distribution with $\mu = 0$ and $\sigma = 1$. Only the results for $n = 20, 40, 120, 240$ and $k = 3, 5$ are reported.

Table 4. Simulated ratios of MSEs of the quantile estimates based, respectively, on optimal RSS schemes and SRS schemes for the Extreme Value distribution.

		Ratio of Estimated MSE's				
n	$k\backslash p$	0.05	0.25	0.5	0.75	0.95
20	3	6.095	2.465	2.564	2.389	2.977
	5	8.293	4.121	3.875	3.539	5.009
40	3	4.132	2.587	2.373	2.355	3.157
	5	6.271	3.528	3.550	3.823	4.803
120	3	3.059	2.125	2.067	2.255	2.979
	5	4.811	3.731	3.241	3.495	4.490
240	3	3.212	2.310	2.066	2.234	2.976
	5	4.767	3.438	3.217	3.689	4.727

Table 5. Simulated ratios of MSEs of the quantile estimates based, respectively, on optimal RSS schemes and SRS schemes for the Log-normal distribution.

		Ratio of Estimated MSE's				
n	$k\backslash p$	0.05	0.25	0.5	0.75	0.95
20	3	1.988	1.305	2.866	2.080	1.199
	5	2.795	3.135	4.372	2.685	2.649
40	3	2.229	1.954	2.299	1.914	2.162
	5	4.181	2.959	3.662	3.395	2.900
120	3	2.474	2.382	2.178	2.141	2.408
	5	4.164	3.743	3.870	3.082	4.061
240	3	2.815	2.124	2.208	2.156	2.897
	5	4.537	3.330	3.756	3.363	4.128

For a large sample size, say, $n = 240$, the simulation results are much in line with the theoretical results given in Table 3. For small sample sizes, it seems that, for the estimation of quantiles in the tails, the optimal RSS is even more efficient when the tails are heavy. This can be seen from the simulation results for the Extreme Value distribution and also in the simulation studies for other heavy tail distributions. On the other hand, it appears that for the log-normal distribution the efficiency for small samples is less than anticipated from the asymptotic results.

7. Imperfect Ranking and Other Issues

In this section, we discuss the case of imperfect ranking and some other issues.

(i) *Imperfect ranking.* Except for Sections 2 and 3, we have assumed that judgment ranking in RSS is perfect. But in most practical cases, there are errors in judgment ranking. When judgment ranking is imperfect, the unbalanced schemes derived in Section 4 are not necessarily optimal. Here we consider imperfect ranking under the following model. Suppose that in judgment ranking, each (numerical) order statistic is ranked as an other order statistic with a certain probability. Let p_{sr} denote the probability with which the sth (numerical) order statistic is wrongly ranked as the rth order statistic. Let $F_{[r]}$ denote the distribution function of the rth judgment-ranked order statistic and $F_{(s)}$ denote the cumulative distribution function of the sth (numerical) order statistic. We have $F_{[r]}(t) = \sum_{s=1}^{k} p_{sr} F_{(s)}(t)$. The component $V(\mathbf{q}, p)$ in the asymptotic variance of the unbalanced RSS with allocation probability vector $\mathbf{q}$ is of the form

$$V(\mathbf{q}, p) = \frac{\sum_{r=1}^{k} q_r F_{[r]}(\xi_p)[1 - F_{[r]}(\xi_p)]}{[\sum_{r=1}^{k} q_r f_{[r]}(\xi_p)/f(\xi_p)]^2}.$$

Since

$$\begin{aligned} F_{[r]}(\xi_p) &= \sum_{s=1}^{k} p_{sr} B(s, k-s+1, p), \\ \frac{f_{[r]}(\xi_p)}{f(\xi_p)} &= \sum_{s=1}^{k} p_{sr} \frac{k!}{(s-1)!(k-s)!} p^{s-1}(1-p)^{k-s}, \end{aligned}$$

The component $V(\mathbf{q}, p)$ still depends only on $\mathbf{q}$ if the probabilities of ranking errors are known. Hence an optimal scheme can be obtained in the same way as in the perfect ranking case. In practice, the probabilities of ranking errors cannot be exactly known, there might be only certain estimates available.

In general, the derived optimal schemes are sensitive to the assumed probabilities of ranking errors. However, in the important special case of $p = 0.5$, optimal schemes are insensitive to ranking errors, as will be shown in the following. It is reasonable to assume that $p_{s,k-r+1} = p_{r,k-s+1}$ and $p_{sr} = p_{rs}$, the assumption of a symmetric ranking mechanism. Under this assumption, Lemmas 1 and 2 in Section 4 still hold. To see this, let

$$\begin{aligned} C_r(p) &= \sum_{s=1}^{k} p_{sr} B(s, k-s+1, p), \\ D_r(p) &= \sum_{s=1}^{k} p_{sr} \frac{k!}{(s-1)!(k-s)!} p^{s-1}(1-p)^{k-s}. \end{aligned}$$

If $p = 0.5$, we have

$$\begin{aligned} C_r(p) &= \sum_{s=1}^{k} p_{sr} c_s(p) = \sum_{s=1}^{k} p_{sr} c_{k-s+1}(p) \\ &= \sum_{s=1}^{k} p_{k-s+1,r} c_s(p) = \sum_{s=1}^{k} p_{k-r+1,s} c_s(p) = C_{k-r+1}(p). \end{aligned}$$

Similarly, we can verify that $D_r(p) = D_{k-r+1}(p)$. Thus Lemma 1 follows. Write

$$F_{\mathbf{q}}(\xi_{0.5}) = \sum_{r=1}^{k} q_r \sum_{s=1}^{k} p_{sr} F_{(s)}(\xi_{0.5}) = \sum_{s=1}^{k} (\sum_{r=1}^{k} q_r p_{sr}) F_{(s)}(\xi_{0.5}).$$

Denote $Q_s = \sum q_r p_{sr}$. It is easy to verify that if $q_r = q_{k-r+1}$ then $Q_s = Q_{k-s+1}$. Hence, Lemma 2 is valid.

The implication of the above results is that (a) for whatever error probabilities, as long as they satisfy the assumption given above, the optimal allocation probability vector can be taken symmetric, and (b) it is always the median of the resulting unbalanced ranked set sample that is to be used as the estimator of the population median. We have computed optimal schemes for a variety of error probabilities satisfying the given assumption. It turns out that if $p_{rr} > 0.5$ then the optimal scheme puts all the mass on the medians of the ranked sets, which is the same as for the perfect ranking case.

More research needs to be carried out on optimal design in general cases when ranking is imperfect.

(ii) *Inference procedures based on optimal unbalanced RSS schemes.* The procedures for confidence interval and hypothesis testing discussed in Chen (2000) can be adapted to the optimal unbalanced RSS schemes.

Suppose that for given p and k, the pth quantile ξ_p of the underlying distribution is the sth quantile of the sampling distribution of the optimal RSS scheme. Let $Y_{(1:n)} \leq \cdots \leq Y_{(n:n)}$ be the order statistics of the unbalanced ranked-set sample from the optimal RSS scheme. Then a confidence interval of confidence coefficient $1 - 2\alpha$ for ξ_p can be constructed as $[Y_{(l_1:n)},\ \ Y_{(l_2:n)}]$, where

$$\begin{aligned} l_1 &= ns - z_\alpha \sqrt{n \sum_{r=1}^{k} q_r^* c_r(p)[1 - c_r(p)]}, \\ l_2 &= ns + z_\alpha \sqrt{n \sum_{r=1}^{k} q_r^* c_r(p)[1 - c_r(p)]}, \end{aligned}$$

z_α denoting the $(1 - \alpha)$th quantile of the standard normal distribution.

To test $H_0 : \xi_p = \xi_0$, the test statistic can be constructed as

$$Z_n = \frac{\sqrt{n}\hat{f}_{\mathbf{q}^*}(\xi_0)[\hat{\xi}_{\mathbf{q}^*,s} - \xi_0]}{\sqrt{\sum_{r=1}^{k} q_r^* c_r(p)[1 - c_r(p)]}},$$

where $\hat{f}_{\mathbf{q}^*}$ is taken as a kernel estimate based on the unbalanced ranked-set sample. For the kernel estimate using balanced RSS data, see Chen (1999a). The results of Chen (1999a) can be extended to unbalanced RSS data by using the methodology developed in Chen (1999b). Under the null hypothesis, the test statistic has approximately a standard normal distribution.

(iii) *Feasible computation programs.* Some further research needs to be done to develop feasible computation programs for the minimization of general $G(V(\mathbf{q}))$ other than $|V(\mathbf{q})|$ when other optimality criteria are considered.

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.

Chen, Z. (1999a). Density estimation using ranked-set sampling data. *Environ. Ecological Statist.* **6**, 135-146.

Chen, Z. (1999b). Non-parametric inferences based on general unbalanced ranked-set samples. Manuscript.

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

Chen, Z. and Bai, Z. H. (1998). The optimal ranked-set sampling scheme for parametric families. *Sankya* A, accepted.

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

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

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

Kaur, A., Patil, G. P. and Taillie, C. (1998a). Optimal allocation for symmetric distributions in ranked set sampling. *Ann. Inst. Statist. Math.* Under revision.

Kaur, A., Patil, G. P. and Taillie, C. (1998b). Ranked set sample sign test under unequal allocation. *J. Statist. Plann. Inference.* Under revision.

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

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

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

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.

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.

Department of Statistics and Applied Probability, National University of Singapore, 3 Science Drive 2, Singapore 117543, Republic of Singapore.

E-mail: stachenz@nus.edu.sg

(Received April 1999; accepted June 2000)