Environ Ecol Stat (2006) 13:25–40
DOI 10.1007/s10651-005-5688-9

ORIGINAL ARTICLE

Nonparametric ranked-set sampling confidence intervals for quantiles of a finite population

Jayant V. Deshpande · Jesse Frey · Omer Ozturk

Received: March 2004 / Revised: December 2004

Abstract Ranked-set sampling from a finite population is considered in this paper. Three sampling protocols are described, and procedures for constructing nonparametric confidence intervals for a population quantile are developed. Algorithms for computing coverage probabilities for these confidence intervals are presented, and the use of interpolated confidence intervals is recommended as a means to approximately achieve coverage probabilities that cannot be achieved exactly. A simulation study based on finite populations of sizes 20, 30, 40, and 50 shows that the three sampling protocols follow a strict ordering in terms of the average lengths of the confidence intervals they produce. This study also shows that all three ranked-set sampling protocols tend to produce confidence intervals shorter than those produced by simple random sampling, with the difference being substantial for two of the protocols. The interpolated confidence intervals are shown to achieve coverage probabilities quite close to their nominal levels. Rankings done according to a highly correlated concomitant variable are shown to reduce the level of the confidence intervals only minimally. An example to illustrate the construction of confidence intervals according to this methodology is provided.

Keywords Order statistics · Interpolated confidence intervals · Median · Efficiency · Sampling designs

1. Introduction

In settings where the cost of identifying experimental units for inclusion in a sample and ranking them according to the attribute of interest is small compared to the cost of actually measuring the value of the attribute of interest, a ranked-set sample (RSS) provides improved efficiency over a simple random sample (SRS) of the same size. This improved efficiency results from the additional information provided by units that are ranked, but not actually measured. Ranked-set sampling was originally proposed by McIntyre (1952) for

J.V. Deshpande · J. Frey · O. Ozturk (✉)
Department of Statistics, The Ohio State University, 1958 Neil Ave, Columbus, OH 43210, USA
e-mail: ozturk.4@osu.edu

use in estimating average yields in agriculture. Actually measuring a yield requires that one harvest the crops, but an expert may be able to produce a very accurate ranking of the yields in a sample of fields based solely on a visual inspection.

Ranked-set sampling has been applied to a variety of problems in ecology and the environmental sciences beyond the problem of estimating average yields that motivated McIntyre. For example, Chen et al. (2004) describe cases in which RSS was applied to the assessment of contamination at hazardous waste sites, the estimation of shrub photomass in an Appalachian oak forest, and the assessment of volatility of gasoline samples. While a large part of the RSS literature focuses on the problem of making inference on a population mean, it is clear that making inference on population quantiles is also of interest in ecology and the environmental sciences. For example, making inference on a population median may be an attractive alternative to making inference on a population mean when the underlying distribution is highly skewed, and making inference on extreme quantiles of a distribution is a way to assess the prevalence of dangerously high or low values in a population. These considerations motivated us to look at the problem of using RSS to produce confidence intervals for quantiles of a population.

The populations of interest in ecological studies are often quite small. For example, a population of interest might consist of a few sheep, a few trees, or a few insect colonies. Despite the small size of the population, measuring the characteristic of interest for every single unit might be too costly and time-consuming to be practical. However, it might be possible through visual inspection to rank small collections of units according to the attribute of interest, and this motivates one to use ranked-set sampling. We show in this paper that if the number of measurements to be made is fixed, procedures based on RSS yield shorter confidence intervals for quantiles of a finite population than do procedures based on simple random sampling.

In the infinite-population setting, a single cycle of the standard RSS protocol begins with the selection of n independent random samples of size n. Units in each of these n independent samples are then separately ranked from smallest to largest according to the attribute of interest. One selects the first order statistic from the first random sample, the second order statistic from the second random sample, and so on, until at last the nth order statistic from the nth sample is selected. Thus far, only ranking and selection has taken place, and no actual measurements have been carried out. Now, however, a full measurement is made on the attribute of interest for the n selected units. These measurements may be denoted $X_{(11)}, \ldots, X_{(nn)}$, where $X_{(ii)}$ is the value for the ith order statistic from the ith random sample. Because each of the n units on which measurement was carried out came from a separate random sample of size n, we have a collection of independent, nonidentically distributed random variables. Each measurement $X_{(ii)}$ has the same marginal distribution as the corresponding order statistic $X_{i:n}$, and the joint distribution of the n measurements is given by the product of the marginals. That is, the measurements $X_{(11)}, \ldots, X_{(nn)}$, which constitute one cycle of observations, have the joint distribution

$$\prod_{i=1}^{n} f_{i:n}(x_{(ii)}),$$

where $f_{i:n}(x)$ is the probability density function for the ith order statistic from a random sample of size n. To achieve a larger total sample size, one may draw a sample consisting of k cycles. Such a sample yields nk independent observations $X_{(ii)j}$, $i = 1, \ldots, n$, $j = 1, \ldots, k$, where $X_{(ii)j}$ is the ith order statistic from the ith random sample in the jth cycle. The expression for the joint density of the nk measured observations is then

$$\prod_{j=1}^{k}\prod_{i=1}^{n} f_{i:n}(x_{(ii)j}).$$

When the population of interest is finite rather than infinite, there is not just a single way to generalize the infinite-population sampling protocol we have described. Instead, a number of finite-population protocols can be imagined, with the exact joint distribution of the resulting sample of measurements depending on the details of the sampling process. To see why this is so, we first review the problem of finding nonparametric confidence intervals for quantiles of a finite population using simple random sampling. Suppose that we have a finite population consisting of N units, with the unknown ordered values of the attribute of interest for these units being

$$x_1 < x_2 < \cdots < x_N.$$

Suppose further that we wish to obtain a confidence interval for q_p, the quantile of order p of this population. To avoid ambiguity, we will assume throughout this paper that p and N are chosen so that $t = pN$ is an integer. The problem then reduces to that of obtaining a confidence interval for the value $x_t = q_p$.

A simple random sample may be obtained from a finite population either through sampling with replacement or through sampling without replacement. A sample taken without replacement tends to be more informative, but properties of a sample taken with replacement may be more tractable analytically because of the independence of the observations. Confidence intervals for quantiles based on simple random sampling without replacement (SRSWOR) from a finite population have been discussed by Sedransk and Meyer (1978), Smith and Sedransk (1983), and others. Suppose that $X_{1:n} < \cdots < X_{n:n}$ are the ordered values of a SRSWOR of size n from the population. Then the probability $P(X_{r:n} \le x_t \le X_{s:n})$ can be calculated for any $r < s$, providing a nonparametric confidence interval for x_t, the tth ordered value from the population. The desired confidence coefficient is achieved or approximately achieved by adjusting the choice of the ranks r and s. We have that

$$\begin{aligned} P(X_{r:n} \le x_t \le X_{s:n}) &= P(X_{r:n} \le x_t) - P(X_{s:n} \le x_{t-1}) \\ &= \sum_{j=r}^{n} \frac{\binom{t}{j}\binom{N-t}{n-j}}{\binom{N}{n}} - \sum_{j=s}^{n} \frac{\binom{t-1}{j}\binom{N-t+1}{n-j}}{\binom{N}{n}} \\ &\equiv p_{r,t} - p_{s,t-1}. \end{aligned}$$

To get a confidence interval with coverage probability $1-\alpha$, we would select $r < s$ so that the probability $p_{r,t} - p_{s,t-1}$ of covering x_t is as close to $1-\alpha$ as possible.

When one takes a ranked-set sample from a finite population, the design of the sampling protocol involves more than a simple choice between sampling with replacement and sampling without replacement. This added complexity arises because there is both a final sample of units that are measured and a collection of samples that are used for ranking purposes only. If we require that the entire set of units used for ranking purposes be drawn without replacement, then we surely produce the most informative sample. If we sample without replacement only within and not across the samples that we use for ranking purposes, however, then our final sample of measured values consists of independent random variables and is thus easier to work with analytically. One can also imagine compromise protocols in which we do not require that the entire set of units used for ranking purposes be drawn without replacement, but we do insist that the final sample of units that are measured have no repeated units. In this paper, we consider three sampling protocols that feature varying degrees of insistence upon sampling without replacement. For each sampling protocol, we order the sample of $K = nk$ measured values as $X_{(1:K)} \leq X_{(2:K)} \leq \cdots \leq X_{(K:K)}$ and produce nonparametric confidence intervals of the form $[X_{(r:K)}, X_{(s:K)}]$ for quantiles.

For our confidence interval procedure, as for many nonparametric confidence interval procedures, the set of confidence levels that can be achieved exactly is a discrete set. One can often achieve a reasonable approximation to other confidence levels, however, by interpolating between two intervals whose exact confidence levels bracket the level of interest. A procedure for constructing interpolated confidence intervals for the population median in the simple random sampling setting was proposed by Hettmansperger and Sheather (1986). The Hettmansperger-Sheather procedure was later generalized to arbitrary quantiles by Nyblom (1992). For illustration, suppose that we seek a $100 \times (1-\alpha)\%$ confidence interval for the population quantile q_p. Let $[X_{(r:K)}, X_{(s:K)}]$ and $[X_{(r+1:K)}, X_{(s-1:K)}]$ be $100 \times (1 - \alpha_{r,s})\%$ and $100 \times (1 - \alpha_{r+1,s-1})\%$ confidence intervals for q_p, where $1 - \alpha_{r,s} > 1 - \alpha > 1 - \alpha_{r+1,s-1}$. Suppose further that the order statistics $X_{(r:K)}$ and $X_{(s:K)}$ have been chosen to give an approximately symmetric confidence interval in the sense that

$$P(X_{(r:K)} > q_p) \leq \alpha/2 \leq P(X_{(r+1:K)} > q_p)$$

and

$$P(X_{(s:K)} < q_p) \leq \alpha/2 \leq P(X_{(s-1:K)} < q_p).$$

Then, for $\epsilon_1, \epsilon_2 \in (0, 1)$, the interval

$$[L, U] = [(1 - \epsilon_1)X_{(r:K)} + \epsilon_1 X_{(r+1:K)}, (1 - \epsilon_2)X_{(s:K)} + \epsilon_2 X_{(s-1:K)}],$$

has a coverage probability between $1-\alpha_{r,s}$ and $1-\alpha_{r+1,s-1}$ whose exact value depends on the underlying distribution. Choices for ϵ_1 and ϵ_2 which approximately give coverage probability $1 - \alpha$ in the setting of ranked-set sampling from an infinite population are given by Ozturk and Deshpande (2005). Adjusting their formulas slightly to account for the discreteness of a finite population, we propose using the approximations

$$\epsilon_1 \approx \left[1 + \frac{r(1-p)(\pi_{r+1} - \alpha/2)}{(K-r)p(\alpha/2 - \pi_r)}\right]^{-1} \tag{1}$$

and

$$\epsilon_2 \approx \left[1 + \frac{(K-(s-1))p(\alpha/2 - \pi_{s-1})}{(s-1)(1-p)(\pi_s - \alpha/2)}\right]^{-1}, \tag{2}$$

where $\pi_r = P(X_{(r:K)} > q_p)$, $\pi_{r+1} = P(X_{(r+1:K)} > q_p)$, $\pi_s = P(X_{(s:K)} < q_p)$, and $\pi_{s+1} = P(X_{(s+1:K)} < q_p)$.

In either the finite-population setting or the infinite-population setting, it may be difficult to rank units according to the attribute of interest. In such situations, one may choose to rank the units according to a readily-available concomitant variable. If the relationship between the attribute of interest and the chosen concomitant variable is strong enough, and if the set size n is not too large, the good properties of ranked-set sampling procedures that assume perfect rankings may be largely maintained. In our simulation study, we find that the levels of our confidence intervals and their interpolated counterparts do not decline excessively when rankings are done according to a concomitant variable whose correlation with the attribute of interest is of the order of 0.9.

In Section 2, we describe three protocols for drawing a ranked-set sample from a finite population, and we describe methods for constructing confidence intervals for quantiles using these protocols. In Section 3, we give a numerical algorithm for computing the coverage probabilities of the confidence intervals produced by the various sampling protocols. In Section 4, we conduct a simulation study to compare the average lengths of confidence intervals for the median produced by the various sampling protocols. In Section 5, finally, we conclude with a summary and discussion.

2. Sampling protocols

In this section, we describe three protocols for drawing a ranked-set sample from a finite population. We will refer to these protocols as Level 0, Level 1, and Level 2 sampling, where higher levels correspond to greater insistence on sampling without replacement.

2.1. Level 0 sampling

In the Level 0 sampling protocol, we sample without replacement only within the samples used for ranking purposes only. We thus allow the same unit to appear in more than one of the samples used just for ranking purposes. As a result, a single unit may also appear more than once in the final sample of values to be measured. One advantage of this protocol, however, is that the final sample consists of independent observations. Specifically, each observation $X_{(ii)j}$ has a marginal distribution determined by the cumulative distribution function

$$P(X_{(ii)j} \leq x_t) = \frac{\sum_{r=i}^{n} \binom{t}{r} \binom{N-t}{n-r}}{\binom{N}{n}} = p_{i,t}, \tag{3}$$

and the joint distribution is the product of the marginals. In algorithmic form, the sampling protocol is as follows.

Step I: *Draw a SRSWOR of size n from the population.*
Step II: *Rank the sampled units and select the* ith *order statistic for measurement.*
Step III: *Return all n units back to the population.*
Step IV: *Repeat steps I to III for* $i = 1, \ldots, n$.
Step V: *Repeat steps I to IV for* $j = 1, \ldots, k$.

2.2. Level 1 sampling

In the Level 1 sampling protocol, we sample without replacement within the samples used for ranking purposes only, and we allow the same unit to appear in more than one of the samples used just for ranking purposes. Unlike in Level 0 sampling, however, we do not allow the same unit to appear more than once in the final sample of values to be measured. This exclusion of repeat units is accomplished by returning to the population only those units that were not selected for measurement. Because of this restriction on having units appear in the final sample multiple times, the units in the final sample are negatively dependent, with the exact structure of that dependence depending on the sequence in which the final sample was drawn. Two natural orderings for drawing a Level 1 sample are given below in algorithmic form.

2.2.1. Ascending order

Step I: *Draw a SRSWOR of size n from the population.*
Step II: *Rank the sampled units and select the* ith *order statistic for measurement.*
Step III: *Return the* $n - 1$ *units not selected back to the population.*

Step IV: *Repeat steps I to III for* $i = 1, \ldots, n$.
Step V: *Repeat steps I to IV for* $j = 1, \ldots, k$.

2.2.2. Descending order

Step I: *Draw a SRSWOR of size n from the population.*
Step II: *Rank the sampled units and select the* $(n+1-i)$th *order statistic for measurement.*
Step III: *Return the* $n - 1$ *units not selected back to the population.*
Step IV: *Repeat steps I to III for* $i = 1, \ldots, n$.
Step V: *Repeat steps I to IV for* $j = 1, \ldots, k$.

2.3. Level 2 sampling

In the Level 2 sampling protocol, we sample without replacement both within and across the samples used for rankings purposes only. As a result, there are no repeated units in the final sample of values to be measured, and the units in that final sample are more strongly negatively dependent than under Level 1 sampling. This stronger negative dependence between the units in the final sample means that a Level 2 sample is more informative than either a Level 0 sample or a Level 1 sample. However, because of the strict insistence on sampling without replacement both within and across the samples used for ranking purposes only, a Level 2 sample can only be drawn for sampling fractions nk/N less than or equal to $1/n$. Restrictions on the size of the sampling fraction are much looser (Level 1) or nonexistent (Level 0) for the other two sampling protocols. The properties of Level 2 sampling for estimation of means were studied by Patil et al. (1995). The Level 2 sampling protocol is given in algorithmic form below.

Step I: *Draw a SRSWOR of size n from the population.*
Step II: *Rank the sampled units and select the* ith *order statistic for measurement.*
Step III: *Do not return any of the n units back to the population.*
Step IV: *Repeat steps I to III for* $i = 1, \ldots, n$.
Step V: *Repeat steps I to IV for* $j = 1, \ldots, k$.

Given a sample of size $K = nk$ drawn using one of the sampling protocols described above, we order the measured values as $X_{(1:K)} \leq X_{(2:K)} \leq \cdots \leq X_{(K:K)}$. We then consider confidence intervals of the type $[X_{(r:K)}, X_{(s:K)}]$, for $r < s$. The coverage probability of such an interval as an interval for x_t is given by

$$\begin{aligned} P(X_{(r:K)} \leq x_t \leq X_{(s:K)}) &= P(X_{(r:K)} \leq x_t) - P(X_{(s:K)} \leq x_{t-1}) \\ &\equiv q_{r,t} - q_{s,t-1} \end{aligned}$$

since we automatically have that $X_{(r:K)} \leq X_{(s:K)}$. The values $q_{r,t}$, which depend on the particular sampling protocol used, can be computed using the algorithms given in Section 3. For the case of Level 0 sampling, in which the measured values are independent, a fairly compact analytic expression is also available. We have for Level 0 sampling that

$$\begin{aligned} q_{r,t} = P(X_{(r:K)} \leq x_t) &= \sum_{j=r}^{K} P(\text{exactly } j \text{ of } X_{(11)1}, \ldots, X_{(nn)k} \text{ are less than or equal to } x_t) \\ &= \sum_{j=r}^{K} \sum_{(u_1,\ldots,u_n)\in S_j} \prod_{l=1}^{n} \binom{k}{u_l} p_{l,t}^{u_l} (1 - p_{l,t})^{k-u_l}, \end{aligned}$$

where

$$S_j = \{(u_1, \ldots, u_n) : u_1 + \cdots + u_n = j, 0 \leq u_i \leq k; i = 1, \ldots, n\}$$

and $p_{i,t}$ is as given in (3).

To highlight the differences between these sampling protocols, we consider the simple case in which $N = 4$, $n = 2$, and $k = 1$. That is, we consider the case in which a single cycle of size 2 is drawn from the finite population whose ordered values are $x_1 < x_2 < x_3 < x_4$. The possible ranked-set samples and their probabilities under the various sampling protocols we have described are then as given in Table 1.

As can be seen in Table 1, the three sampling protocols may produce quite different sampling distributions for small populations. It may be verified that under Level 0 and Level 2 sampling, the sample means are unbiased estimators of the true mean, while under Level 1 sampling, the sample mean, regardless of whether sampling is done in ascending or descending order, is in general biased. The effect of the strong insistence on sampling without replacement in the Level 2 protocol can be seen in the fact that only 4 distinct ranked-set samples are possible under Level 2 sampling. Samples such as x_1, x_2 and x_3, x_4, which are unrepresentative of the population in the sense that they consist entirely of low population values or high population values, have no probability of being chosen under Level 2 sampling. The Level 1 protocol also renders certain unrepresentative samples impossible, but it does this to a lesser extent than does Level 2 sampling. The number of possible samples is greatest under Level 0 sampling.

3. An algorithm for computing confidence coefficients

The determination of the confidence coefficients for our confidence intervals can be computationally intensive even for Level 0 sampling, the simplest case. With the negative correlations between the sample observations that occur in Level 1 and Level 2 sampling, the amount of computation needed increases. This negative correlation is a consequence, as described in Section 2, of insisting upon sampling without replacement at one or more levels. In this section, we describe an approach that can be used to compute confidence coefficients for confidence intervals produced by any of the RSS sampling protocols described in Section 2. While our focus in this paper is on balanced RSS, these algorithms can also be applied to samples obtained either from unbalanced RSS or from any scheme that yields a sample consisting of order statistics. Applying the algorithms in these more general cases simply involves using appropriate values for r and n at each step. The FORTRAN subroutines we provide implement the algorithms in full generality.

Table 1 Probabilities of the possible samples under various sampling protocols

Sample	Level 0	Level 1		Level 2
		Ascending	Descending	
x_1, x_2	1/12	0	1/9	0
x_1, x_3	1/6	1/6	2/9	1/6
x_1, x_4	1/4	1/3	1/3	1/3
x_2, x_2	1/18	0	0	0
x_2, x_3	5/36	1/6	1/6	1/3
x_2, x_4	1/6	2/9	1/6	1/6
x_3, x_3	1/18	0	0	0
x_3, x_4	1/12	1/9	0	0

Our approach to computing confidence coefficients consists of computing the values

$q_{r,t} = P(\text{at least } r \text{ elements of the sample do not exceed } x_t)$

by computing the constituent probabilities

$E_{s,t} = P(\text{exactly } s \text{ elements of the sample do not exceed } x_t),\ s = 0, \ldots, K,$

and summing them in the obvious way. We compute the $E_{s,t}$ in a sequential fashion, thinking of drawing our final sample of units to be measured one unit at a time. Some of the building blocks needed in the computational process are $p^{\star}_{j:n,t:N}$, the probability that exactly j elements of a sample of size n taken without replacement from a population of size N are less than or equal to the tth smallest value in the population, and $p_{j:n,t:N} = \sum_{r=j}^{n} p^{\star}_{r:n,t:N}$, the probability that at least j elements of a sample of size n taken without replacement from a population of size N are less than or equal to the tth smallest value in the population. Note that $p_{j:n,t:N}$ is the same as the $p_{j,t}$ introduced in Section 2. We are simply modifying the notation slightly so that both the population and sample sizes are explicitly denoted. Note also that we can compute $p^{\star}_{j:n,t:N}$ via the equality

$$p^{\star}_{j:n,t:N} = \frac{\binom{t}{j}\binom{N-t}{n-j}}{\binom{N}{n}}.$$

We discuss the computation of $E_{s,t}$ separately below for each sampling protocol.

3.1. Level 0 sampling

We define $E^{l}_{i,t}$ to be the probability that exactly i units less than or equal to x_t have been obtained among the first l units selected for measurement. Then clearly $E^{0}_{0,t} = 1$. Assume that for some fixed $l \geq 0$, $E^{l}_{i,t}$ is known for each i, and suppose that the next unit to be added to the final sample will be an rth order statistic from a sample of size n. With probability $p_{r:n,t:N}$, that next unit will be less than or equal to x_t, and with probability $1 - p_{r:n,t:N}$, that next unit will be greater than x_t. Thus, we can compute each value $E^{l+1}_{i,t}$ via the recursive equation

$$E^{l+1}_{i,t} = p_{r:n,t:N} E^{l}_{i-1,t} + (1 - p_{r:n,t:N}) E^{l}_{i,t}.$$

When the entire sample of size $K = nk$ has been accounted for in this way, we simply set $E_{s,t} = E^{K}_{s,t}$. We can then compute the value $q_{r,t}$ for any selected r.

As an example, consider the case, already discussed at the end of Section 2, in which $N = 4$, $n = 2$, and $k = 1$. We compute the values $E_{0,2}$, $E_{1,2}$, and $E_{2,2}$ for the situation in which $t = 2$. Our starting value is $E^{0}_{0,2} = 1$, and the first unit to be added to the sample may be taken to be a first order statistic. Thus, we have that

$$E^{1}_{0,2} = (1 - p_{1:2,2:4}) E^{0}_{0,2} = (1 - 5/6)1 = 1/6, \text{ and}$$
$$E^{1}_{1,2} = p_{1:2,2:4} E^{0}_{0,2} = (5/6)1 = 5/6.$$

The next unit to be added is then a second order statistic, giving

$$E^{2}_{0,2} = (1 - p_{2:2,2:4}) E^{1}_{0,2} = (1 - 1/6)(1/6) = 5/36,$$
$$E^{2}_{1,2} = p_{2:2,2:4} E^{1}_{0,2} + (1 - p_{2:2,2:4}) E^{1}_{1,2} = (1/6)(1/6) + (1 - 1/6)(5/6) = 26/36, \text{ and}$$
$$E^{2}_{2,2} = p_{2:2,2:4} E^{1}_{1,2} = (1/6)(5/6) = 5/36.$$

Thus, the values for $E_{0,2}$, $E_{1,2}$, and $E_{2,2}$ are $5/36$, $13/18$, and $5/36$, respectively. These values may be confirmed by consulting Table 1 in Section 2.

3.2. Level 1 sampling

Let $E^l_{i,t}$ be defined as for Level 0 sampling. The difference between the sampling procedure under this sampling protocol and under the Level 0 sampling protocol is that units already added to the final sample are not returned to the population. If i elements less than or equal to x_t have been added to the final sample among the first l elements sampled, then the remaining population contains $t - i$ elements less than or equal to x_t and $N - l$ total elements. Thus, assuming that an rth order statistic from a sample of size n is the next unit to be added to the final sample, we can compute each value $E^{l+1}_{i,t}$ via the recursive equation

$$E^{l+1}_{i,t} = p_{r:n,t-i+1:N-l} E^l_{i-1,t} + (1 - p_{r:n,t-i:N-l}) E^l_{i,t}.$$

The procedure is otherwise identical to the procedure for the Level 0 sampling protocol.

As an example, we again consider the case in which $N = 4$, $n = 2$, $k = 1$, and $t = 2$. The starting value is $E^0_{0,2} = 1$. Adding the first unit, a first order statistic, gives that

$$E^1_{0,2} = (1 - p_{1:2,2:4}) E^0_{0,2} = (1 - 5/6)1 = 1/6, \text{ and}$$
$$E^1_{1,2} = p_{1:2,2:4} E^0_{0,2} = (5/6)1 = 5/6.$$

Adding the second unit, a second order statistic, then gives that

$$E^2_{0,2} = (1 - p_{2:2,2:3}) E^1_{0,2} = (1 - 1/3)(1/6) = 1/9, \text{ and}$$
$$E^2_{1,2} = p_{2:2,2:3} E^1_{0,2} + (1 - p_{2:2,1:3}) E^1_{1,2} = (1/3)(1/6) + (1 - 0)(5/6) = 8/9.$$

Thus, the values for $E_{0,2}$, $E_{1,2}$, and $E_{2,2}$ are $1/9$, $8/9$, and 0, respectively. These values may be confirmed by consulting Table 1 in Section 2.

3.3. Level 2 sampling

To compute the values $E_{s,t}$ for this procedure, we must keep track at each stage not only of the number of units less than or equal to x_t that have been added to the final sample among the first l elements sampled, but also of the number of units less than or equal to x_t left in the rest of the population. This extra level of record-keeping is necessary since these two quantities are no longer in a one-to-one correspondence. To this end, we define $F^l_{i,j,t}$ to be the probability that exactly i units with measured values less than or equal to x_t have been sampled among the first l items sampled and that j total items less than or equal to x_t have been used up in the process of sampling the first l items. Then clearly $F^0_{0,0,t} = 1$. Assume that for some fixed $l \geq 0$, the $F^l_{i,j,t}$ are known for all i and j, and suppose that the next unit to be sampled is an rth order statistic from a sample of size n. Since the entire remaining population at this point consists of $N - nl$ elements, we can compute the value of $F^{l+1}_{i,j,t}$ via the recursive equation

$$F^{l+1}_{i,j,t} = \sum_{k=0}^{r-1} p^\star_{k:n,t-j+k:N-nl} F^l_{i,j-k,t} + \sum_{k=r}^{n} p^\star_{k:n,t-j+k,N-nl} F^l_{i-1,j-k,t}.$$

Once we have obtained the entire sample of $K = nk$ elements, we can compute the $E_{s,t}$ by setting

$$E_{s,t} = \sum_j F^K_{s,j,t}. \tag{4}$$

As an example, we again consider the case in which $N = 4$, $n = 2$, $k = 1$, and $t = 2$. As before, we compute the values $E_{0,2}$, $E_{1,2}$, and $E_{2,2}$. Our starting value is $F^0_{0,0,2} = 1$. Adding the first unit, a first order statistic, gives that

$$F^1_{0,0,2} = p^{\star}_{0:2,2:4} F^0_{0,0,2} = (1/6)1 = 1/6,$$
$$F^1_{1,1,2} = p^{\star}_{1:2,2:4} F^0_{0,0,2} = (2/3)1 = 2/3, \text{ and}$$
$$F^1_{1,2,2} = p^{\star}_{2:2,2:4} F^0_{0,0,2} = (1/6)1 = 1/6.$$

Adding the second unit, a second order statistic, then gives that

$$F^2_{1,2,2} = p^{\star}_{2:2,2:2} F^1_{0,0,2} + p^{\star}_{0:2,0:2} F^1_{1,2,2} + p^{\star}_{1:2,1:2} F^1_{1,1,2} = 1/6 + 2/3 + 1/6 = 1.$$

Summing appropriate F values as given in (4), we find that the values for $E_{0,2}$, $E_{1,2}$, and $E_{2,2}$ are 0, 1 and 0, respectively. These values may be confirmed by consulting Table 1 in Section 2.

Using the algorithms given in this section, coverage probabilities may be determined for any confidence interval $[X_{(r:K)}, X_{(s:K)}]$ arising under any of the three sampling protocols. Also, FORTRAN subroutines for implementing these algorithms are available at "www19.homepage.villanova.edu/jesse.frey/software/cifinder.txt". As an additional resource, however, we provide Tables 2 and 3, computed using these algorithms, which identify, for population sizes 20, 30, 40, and 50 and selected sampling designs, the order statistics and interpolation coefficients appropriate for producing approximately symmetric 95% confidence intervals for the median of a finite population.

Table 2 Order statistics and interpolation coefficients for constructing approximately symmetric 95% confidence intervals for the population median

Population size	(n, k)	Sampling protocol	$r(\epsilon_1)$	$s(\epsilon_2)$	Upper coverage probability	Lower coverage probability
20	(3,2)	Level 0	1 (0.851)	5 (0.051)	0.977	0.767
		Level 1	1 (0.978)	5 (0.309)	0.995	0.855
		Level 2	2 (0.152)	5 (0.360)	0.992	0.664
		SRSWOR	1 (0.683)	6 (0.898)	0.992	0.892
20	(2,3)	Level 0	1 (0.687)	6 (0.906)	0.992	0.895
		Level 1	1 (0.877)	5 (0.150)	0.985	0.802
		Level 2	1 (0.985)	5 (0.199)	0.990	0.821
		SRSWOR	1 (0.683)	6 (0.898)	0.992	0.892
20	(2,4)	Level 0	2 (0.261)	7 (0.673)	0.977	0.848
		Level 1	2 (0.685)	6 (0.132)	0.983	0.800
		Level 2	2 (0.958)	6 (0.228)	0.993	0.842
		SRSWOR	2 (0.431)	7 (0.814)	0.987	0.875
30	(3,2)	Level 0	1 (0.835)	6 (0.968)	0.996	0.924
		Level 1	1 (0.916)	5 (0.130)	0.985	0.787
		Level 2	2 (0.031)	5 (0.160)	0.969	0.562
		SRSWOR	1 (0.583)	6 (0.746)	0.987	0.857
30	(2,3)	Level 0	1 (0.677)	6 (0.831)	0.990	0.883
		Level 1	1 (0.803)	6 (0.975)	0.996	0.920
		Level 2	1 (0.867)	6 (0.998)	0.997	0.932
		SRSWOR	1 (0.583)	6 (0.746)	0.987	0.857
30	(2,4)	Level 0	2 (0.245)	7 (0.529)	0.974	0.831
		Level 1	2 (0.526)	7 (0.853)	0.990	0.890
		Level 2	2 (0.664)	7 (0.926)	0.995	0.912
		SRSWOR	2 (0.212)	7 (0.494)	0.973	0.821

Table 3 Order statistics and interpolation coefficients for constructing approximately symmetric 95% confidence intervals for the population median

Population size	(n, k)	Sampling protocol	$r(\epsilon_1)$	$S(\epsilon_1)$	Upper coverage probability	Lower coverage probability
40	(3,2)	Level 0	1 (0.827)	6 (0.930)	0.996	0.918
		Level 1	1 (0.887)	5 (0.045)	0.978	0.752
		Level 2	1 (0.969)	5 (0.068)	0.981	0.759
		SRSWOR	1 (0.531)	6 (0.662)	0.983	0.838
40	(2,3)	Level 0	1 (0.673)	6 (0.791)	0.990	0.877
		Level 1	1 (0.767)	6 (0.903)	0.994	0.905
		Level 2	1 (0.813)	6 (0.918)	0.996	0.914
		SRSWOR	1 (0.531)	6 (0.662)	0.983	0.838
40	(2,4)	Level 0	2 (0.237)	7 (0.454)	0.972	0.822
		Level 1	2 (0.450)	7 (0.698)	0.986	0.867
		Level 2	2 (0.548)	7 (0.743)	0.990	0.883
		SRSWOR	2 (0.094)	7 (0.325)	0.963	0.794
50	(3,2)	Level 0	1 (0.822)	6 (0.906)	0.996	0.914
		Level 1	1 (0.870)	6 (0.994)	0.997	0.932
		Level 2	1 (0.933)	5 (0.012)	0.975	0.736
		SRSWOR	1 (0.498)	6 (0.609)	0.980	0.827
50	(2,3)	Level 0	1 (0.670)	6 (0.766)	0.989	0.873
		Level 1	1 (0.745)	6 (0.857)	0.993	0.896
		Level 2	1 (0.782)	6 (0.869)	0.994	0.903
		SRSWOR	1 (0.498)	6 (0.609)	0.980	0.827
50	(2,4)	Level 0	2 (0.232)	7 (0.408)	0.971	0.816
		Level 1	2 (0.404)	7 (0.606)	0.982	0.852
		Level 2	2 (0.482)	7 (0.639)	0.986	0.864
		SRSWOR	2 (0.018)	7 (0.216)	0.957	0.777

The lack of symmetry evident in Tables 2 and 3 may seem somewhat counterintuitive, given that we seek a confidence interval for the median. For Level 0 and Level 2 sampling, this lack of symmetry arises solely because we are defining the median of an even-numbered finite population as the population value $x_{0.5N}$. For Level 1 sampling, however, asymmetry arises both from the fact that the median as we define it is not symmetrically located and from the asymmetry in the sampling protocol itself. The tabled values correspond to the ascending version of Level 1 sampling as described in Section 2, and they would be different if the order of the sampling were changed.

As an example of how to use the tables, consider the first row of Table 2. It tells us that for a finite population of size 20, set size $n = 3$, and number of cycles $k = 2$, the interval $[X_{(r:K)}, X_{(s:K)}] = [X_{(1:6)}, X_{(5:6)}]$ has exact coverage probability 0.977 under Level 0 sampling, while the shorter interval $[X_{(2:6)}, X_{(4:6)}]$ has exact coverage probability 0.767. Using formulas (1) and (2) from Section 1, we find that $\epsilon_1 = 0.851$ and $\epsilon_2 = 0.051$ are the interpolation coefficients appropriate for obtaining an approximate 95% confidence interval for the median, x_{10}. Thus, the interpolated confidence interval

$$[0.149X_{(1:6)} + 0.851X_{(2:6)}, 0.949X_{(5:6)} + 0.051X_{(4:6)}]$$

is an approximate 95% confidence interval for the median proposed in this paper.

In practice, one may not have an exact value for the size N of the finite population from which one is sampling. However, if an upper bound $N^\star$ for the population size is available, then one may treat $N^\star$ as the population size and be effectively assured that the confidence interval obtained in this way will be conservative. The validity of this conclusion for certain concrete cases may be verified both by consulting Tables 2 and 3 and by consulting Table 4. Table 4 shows that if one fixes the design as $(n, k) = (2, 5)$, then the interpolated order

Table 4 Order statistics and interpolation coefficients for contructing approximately symmetric 95% confidence intervals for the population median when the sampling design is $(n, k) = (2, 5)$

Population size	Sampling protocol	$r(\epsilon_1)$	$s(\epsilon_2)$	Upper coverage probability	Lower coverage probability
40	Level 0	2 (0.946)	8 (0.164)	0.977	0.876
80		2 (0.939)	8 (0.012)	0.971	0.857
160		2 (0.936)	9 (0.970)	0.992	0.942
240		2 (0.935)	9 (0.958)	0.992	0.940
320		2 (0.935)	9 (0.952)	0.992	0.940
400		2 (0.934)	9 (0.948)	0.992	0.939
∞		2 (0.933)	9 (0.933)	0.992	0.938
40	Level 1	3 (0.218)	8 (0.534)	0.977	0.838
80		3 (0.042)	8 (0.208)	0.960	0.787
160		2 (0.977)	8 (0.037)	0.973	0.860
240		2 (0.962)	9 (0.990)	0.993	0.946
320		2 (0.955)	9 (0.976)	0.993	0.944
400		2 (0.950)	9 (0.968)	0.993	0.943
∞		2 (0.933)	9 (0.933)	0.992	0.938
40	Level 2	3 (0.364)	8 (0.615)	0.984	0.862
80		3 (0.115)	8 (0.245)	0.964	0.798
160		2 (0.994)	8 (0.056)	0.974	0.862
240		2 (0.974)	9 (0.995)	0.994	0.947
320		2 (0.964)	9 (0.980)	0.993	0.945
400		2 (0.957)	9 (0.971)	0.993	0.944
∞		2 (0.933)	9 (0.933)	0.992	0.938

statistics appropriate for obtaining a 95% confidence interval for the median get farther apart as the population size increases. Table 4 also shows that as the population size increases, the coverage probabilities for Level 0 sampling converge most rapidly to the corresponding values in the infinite-population case, while the coverage probabilities for Level 2 sampling converge most slowly to the values in the infinite-population case.

A heuristic explanation of the general rule that overestimating the population size will lead to conservative confidence intervals can be obtained by considering what it is that makes these finite-population confidence intervals good. Put simply, the two factors making these finite-population confidence intervals better than the corresponding infinite-population confidence intervals are the discreteness of the population and the gradual exhaustion of the population when we sample without replacement. If the population size increases while the design is kept fixed, the influence of each of these factors is reduced, meaning that one obtains a conservative confidence interval when one overestimates the population size N. In the extreme situation in which nothing is known of the population size, one may safely use the infinite population confidence intervals developed by Ozturk and Deshpande (2005) and be assured of a conservative result.

4. A simulation study

In order to compare the performance of the three sampling protocols both among themselves and also to simple random sampling without replacement, a simulation study was performed. Finite populations with sizes 20, 30, 40, and 50 were produced by separately drawing simple random samples without replacement from a larger dataset listed in Chen et al. (2004). These small finite populations were then fixed and used throughout the simulation study. The dataset listed in Chen et al. (2004), which consists of measurements of height and diameter

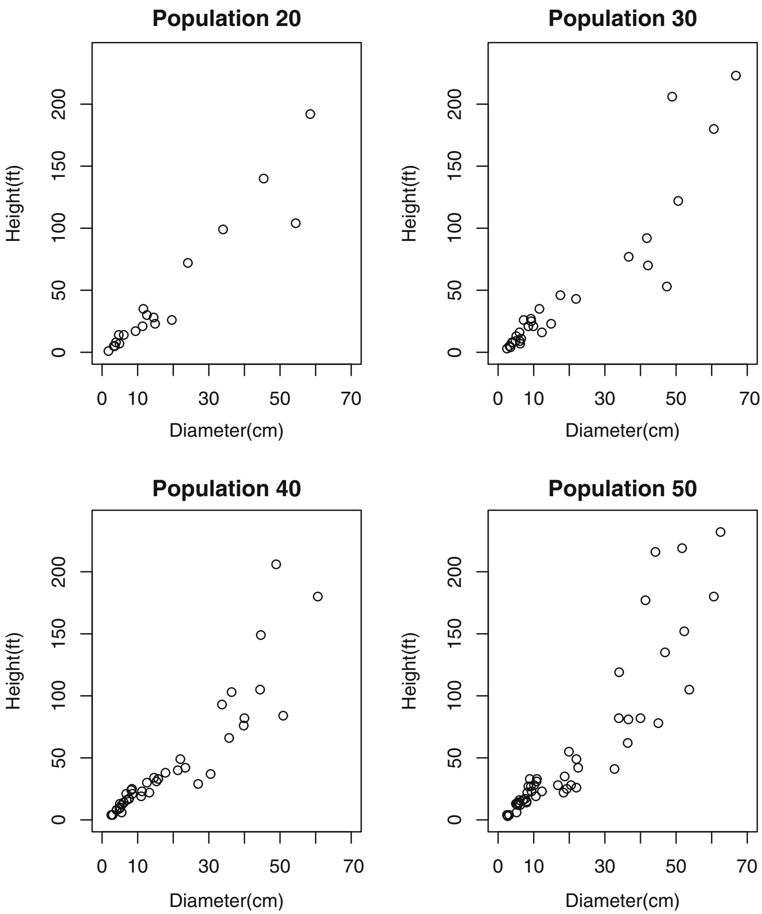

Fig. 1 Plots of height (in feet) versus diameter at chest height (in cm) for the four finite populations of long-leafed pines

at chest height for 396 long-leafed pines, is part of a more comprehensive dataset originally developed by Platt et al. (1988). The attribute of interest was taken to be tree height, which was measured in feet.

We considered two different methods for carrying out the ranking needed in the sampling process. Specifically, we considered the case in which ranking was done according to the concomitant variable diameter at chest height and the case in which ranking was done according to height itself. The relationship between height and diameter at chest height for the four finite populations is shown in Fig. 1. The correlations between height and diameter at chest height were 0.957 for the population of size 20, 0.935 for the population of size 30, 0.905 for the population of size 40, and 0.900 for the population of size 50.

The simulation study was carried out using a factorial design, with the sampling design (n, k) taking on the values (3, 2), (2, 3), and (2, 4), the population size taking on the values 20, 30, 40, and 50, and the ranking being done either according to the concomitant variable diameter at chest height or according to height itself. The sampling protocols considered were Level 0 sampling, Level 1 sampling (ascending), Level 2 sampling, and simple random sampling without replacement. For each combination of the factors, a 95% confidence interval

Table 5 Simulated average lengths for nominal 95% confidence intervals for the population median

Population size	(n, k)	Ranking variable	Sampling protocol			
			Level 0	Level 1	Level 2	SRSWOR
20	(3,2)	Diameter	54.2	46.2	39.8	60.1
	(2,3)	Diameter	61.6	50.5	46.7	60.1
	(2,4)	Diameter	50.4	33.0	27.9	42.3
20	(3,2)	Height	54.4	46.4	39.9	60.1
	(2,3)	Height	61.5	50.6	46.8	60.1
	(2,4)	Height	50.5	33.1	28.0	42.3
30	(3,2)	Diameter	75.7	68.3	62.8	92.4
	(2,3)	Diameter	86.6	75.4	71.4	92.5
	(2,4)	Diameter	72.7	54.4	48.4	72.5
30	(3,2)	Height	75.9	68.4	62.8	92.6
	(2,3)	Height	86.8	75.6	71.5	92.6
	(2,4)	Height	72.7	54.5	48.4	72.5
40	(3,2)	Diameter	54.4	49.4	46.9	68.8
	(2,3)	Diameter	62.1	55.9	54.1	68.8
	(2,4)	Diameter	52.3	43.9	41.1	55.1
40	(3,2)	Height	54.4	49.4	46.8	68.8
	(2,3)	Height	62.2	56.0	54.1	68.8
	(2,4)	Height	52.4	44.0	41.5	55.1
50	(3,2)	Diameter	72.1	66.4	63.6	92.4
	(2,3)	Diameter	81.8	75.8	74.1	92.3
	(2,4)	Diameter	69.8	60.7	58.0	77.0
50	(3,2)	Height	72.3	66.5	63.4	92.4
	(2,3)	Height	82.1	76.1	74.2	92.4
	(2,4)	Height	69.9	60.7	58.0	77.1

for the population median was constructed. Since the confidence level 0.95 could not be achieved exactly under any of the factor combinations, approximate 95% confidence intervals were constructed using the interpolation scheme described in Section 1. In the course of the study, 1000000 ranked-set samples were simulated for each combination of factors. The true coverage probability for the intervals was estimated using the proportion of the 1000000 intervals which contained the true median, and the average length of the confidence intervals was estimated using the mean length of the 1000000 intervals simulated for each combination of factors.

The results of the study indicate that for all combinations of factors considered, intervals produced by the Level 1 and Level 2 sampling protocols are substantially shorter than intervals produced by simple random sampling without replacement. The intervals produced by the Level 0 sampling protocol also tend to be shorter than the intervals produced by simple random sampling without replacement, but they are slightly longer for certain designs when the population size is only 20 or 30. These effects are clearly evident in Table 5. The Level 1 and Level 2 RSS sampling protocols yield intervals of comparable average length, but the average length is always smaller under Level 2 sampling. In fact, for each combination of factors considered, Level 2 sampling produced the shortest intervals among the three RSS protocols, and Level 0 sampling produced the longest. The difference in average lengths among the three protocols is, as might have been anticipated, larger for population size 20 than for population size 50. As expected, the average lengths of the confidence intervals decrease, for the fixed confidence level 95%, when the set size n is fixed at 2 and the number of cycles k increases from 3 to 4.

Table 6 Simulated coverage probabilities of nominal 95% confidence intervals for the population median

Population size	(n, k)	Ranking variable	Sampling protocol			
			Level 0	Level 1	Level 2	SRSWOR
20	(3,2)	Diameter	0.930	0.949	0.949	0.947
	(2,3)	Diameter	0.949	0.934	0.945	0.947
	(2,4)	Diameter	0.935	0.931	0.934	0.940
20	(3,2)	Height	0.933	0.952	0.957	0.947
	(2,3)	Height	0.952	0.935	0.945	0.947
	(2,4)	Height	0.935	0.933	0.936	0.940
30	(3,2)	Diameter	0.936	0.916	0.913	0.940
	(2,3)	Diameter	0.948	0.943	0.932	0.941
	(2,4)	Diameter	0.940	0.944	0.946	0.943
30	(3,2)	Height	0.951	0.939	0.938	0.940
	(2,3)	Height	0.952	0.951	0.943	0.940
	(2,4)	Height	0.945	0.950	0.955	0.942
40	(3,2)	Diameter	0.936	0.927	0.941	0.938
	(2,3)	Diameter	0.938	0.936	0.935	0.938
	(2,4)	Diameter	0.946	0.941	0.943	0.945
40	(3,2)	Height	0.938	0.929	0.942	0.938
	(2,3)	Height	0.940	0.938	0.937	0.938
	(2,4)	Height	0.948	0.943	0.945	0.945
50	(3,2)	Diameter	0.954	0.944	0.922	0.945
	(2,3)	Diameter	0.951	0.955	0.957	0.945
	(2,4)	Diameter	0.937	0.946	0.949	0.930
50	(3,2)	Height	0.966	0.961	0.943	0.945
	(2,3)	Height	0.957	0.961	0.963	0.945
	(2,4)	Height	0.945	0.953	0.958	0.930

The interpolated confidence intervals have coverage probabilities quite close to their nominal levels when rankings are done according to height, as can be seen in Table 6. The interpolated confidence intervals also essentially maintain their level when rankings are done according to the concomitant variable diameter at chest height. We suspect, however, that the confidence levels would not hold as well either when rankings are done according to less highly correlated concomitant variables or when larger set sizes n are used.

5. Conclusions

We described three protocols for drawing a ranked-set sample from a finite population, and we proposed nonparametric confidence intervals for quantiles based on these sampling protocols. Though the sampling protocols lead to highly divergent sampling distributions when population sizes are small, the differences decrease with increasing population size. However, in our simulations study involving finite populations of sizes 20, 30, 40, and 50, we found noticeable differences among the three protocols in the average length of 95% confidence intervals for a population median. All three RSS protocols were found in the simulation study to be superior to simple random sampling without replacement, but the difference was greatest for Level 1 and Level 2 sampling. The performance of the protocols suggests that Level 2 sampling is to be preferred for small finite population sizes such as those we considered. Once larger populations are considered, however, convenience might also be an important consideration in deciding which protocol to use in a given situation.

We proposed that the same sort of interpolated confidence intervals that have been used for SRS and RSS in the infinite-population setting could be used for finite-sample RSS, and our simulation results showed close agreement between the desired confidence levels and the levels actually achieved using interpolated confidence intervals. We examined the impact of imperfect rankings on our confidence interval procedure by considering rankings according to a concomitant variable. Our simulation study suggests that for concomitant variables, such as diameter at chest height in the simulation study, that are highly correlated with the attribute of interest, the true confidence levels are only marginally lower than the nominal levels for small set sizes. However, for larger set sizes and for less highly correlated concomitant variables, we recommend proceeding with caution.

Acknowledgements The authors would like to thank three anonymous referees for helpful comments that have improved the paper. A partial support for this research was provided by the National Security Agency under award MSPF-04G-109.

References

Chen Z, Bai Z, Sinha BK (2004). Ranked set sampling: theory and applications, Springer, New York.

Hettmansperger TP, Sheather SJ (1986). Confidence intervals based on interpolated order statistics. Stat probabil lett 4:75–79.

McIntyre GA (1952). A method for unbiased selective sampling, using ranked sets. Aust J Agr Res 3:385–390.

Nyblom J (1992). Note on interpolated order statistics. Stat Probabil Lett 14:129–131.

Ozturk O, Deshpande JV (2005). Ranked-set sample nonparametric quantile confidence intervals. J Stat Plan Infer 136:570–577.

Patil GP, Sinha AK, Taillie C (1995). Finite population corrections for ranked set sampling. Ann I Stat Math, 47:621–636.

Platt WN, Evans GM, Rathbun SL (1988). The population dynamics of a long-lived conifer (Pinus palustris). Am Nat 131:491–525.

Sedransk J, Meyer J (1978). Confidence intervals for the quantiles of a finite population: simple random and stratified simple random sampling. J Roy Stat Soc B 40:239–252.

Smith PJ, Sedransk J (1983). Lower bounds for confidence coefficients for confidence intervals for finite population quantiles. Commun Stat – Theor M 12:1329–1344.

Biographical sketches

Professor Jayant Deshpande has been teaching in India since 1972, first at the Panjab University and later, since 1984, at the University of Pune. He has held visiting positions at Case Western Reserve University, the University of California at Santa Barbara, the University of Michigan, the Ohio State University, and the University of Sheffield (UK). The work presented in this paper was done while he was visiting the Ohio State University during 2002–2004. His research interests include nonparametric inference, ranking and selection procedures, survival analysis, and reliability theory. He is a Fellow of the Institute of Mathematical Statistics, an Elected Member of the International Statistical Institute, and a Past President of the Indian Society for Probability and Statistics.

Jesse Frey is a Ph.D. student in statistics at the Ohio State University. His research interests include order statistics, nonparametrics, and statistical computing. He spent a year as a research assistant in the Program in Spatial Statistics and Environmental Sciences at the Ohio State University.

Omer Ozturk is an Associate Professor of Statistics at the Ohio State University. His recent research interests include nonparametric and robust procedures for ranked-set sample data, and conditional inference and rank regression for simple random sample data. His research has direct relevance for the analysis of biological, environmental, and ecological data, which very often are not normally distributed. He is an active contributor to ranked-set sampling research.