CrossMark

ORIGINAL PAPER

Interval estimation of $P(X < Y)$ in ranked set sampling

M. Mahdizadeh[1] · **Ehsan Zamanzade[2]**

Abstract This article deals with constructing a confidence interval for the reliability parameter using ranked set sampling. Some asymptotic and resampling-based intervals are suggested, and compared with their simple random sampling counterparts using Monte Carlo simulations. Finally, the methods are applied on a real data set in the context of agriculture.

Keywords Bootstrap · Jackknife · Judgment ranking

1 Introduction

Ranked set sampling (RSS) is a data collection method introduced by McIntyre (1952) in an agricultural context involving pasture yields. It serves as an alternative to the usual simple random sampling (SRS) in situations in which exact measurements of sample units are difficult or expensive to obtain but judgment ranking of them according to the variable of interest is relatively easy and cheap. The judgment ranking is usually performed visually (by a field expert, say), or using one or more concomitant variables, but it cannot necessitate actual measurements on the selected units.

✉ M. Mahdizadeh
mahdizadeh.m@live.com

Ehsan Zamanzade
e.zamanzade@sci.ui.ac.ir; ehsanzamanzadeh@yahoo.com

[1] Department of Statistics, Hakim Sabzevari University, P.O. Box 397, Sabzevar, Iran

[2] Department of Statistics, University of Isfahan, Isfahan 81746-73441, Iran

The RSS design can be explained as follows:

1. Draw k random samples, each of size k, from the target population.
2. Apply the judgement ordering, by any cheap method without the actual measurement of the variable of interest, on the elements of the rth ($r = 1, \ldots, k$) sample and identify the rth smallest unit.
3. Actually measure the k identified units in step 2.
4. Repeat steps 1–3, m times (cycles), if needed, to obtain a ranked set sample of size mk.

Let $X_{[r]i}$ be the rth judgement order statistic from the ith cycle. This is a standard notation in the RSS literature (see Chapter 1 in Chen et al. 2004, for example). Then, the resulting ranked set sample is denoted by $\{X_{[r]i} : r = 1, \ldots, k; i = 1, \ldots, m\}$. The design parameter k is called set size. To facilitate the judgment ranking, the set size should be kept small in practice, say 2–8. Nonetheless, larger set sizes can be used as long as the ranking process is not hampered.

A ranked set sample comprising m cycles and with set size k exploits information about far more units than a simple random sample of size mk. In the RSS, the judgment ranking information about $mk(k-1)$ unmeasured units contributes to drawing a more representative sample. The SRS, however, has no mechanism for incorporating the judgment ranking information. Thus, the RSS-based procedures are usually more efficient than their SRS competitors. The extent of improvement hinges on the accuracy of the judgment ranking. The RSS has been applied in a variety of fields, including forestry (Halls and Dell 1966), entomology (Howard et al. 1982), environmental monitoring (Kvam 2003), clinical trials and genetic quantitative trait loci mappings (Chen 2007), segmentation of Terahertz images (Ayech and Ziou 2015), and medicine (Zamanzade and Mahdizadeh 2017).

In reliability theory, the probability $\theta = P(X < Y)$ represents the reliability of a stress–strength model, where X and Y represent the stress and strength variables, respectively. This probability also quantifies steady state availability of a repairable system with X and Y denoting repair time and lifetime of the system, respectively. In fact, θ provides a general measure of the difference between two populations, that has found applications in diverse areas (Kotz et al. 2003). For example, it is a measure of household financial fragility in economics when X and Y are disposable household income and consumption, respectively. In medicine, it is interpreted as a measure of treatment's effectiveness if X and Y are the response variables from control and treatment groups, respectively. The latter situation is illustrated using a real data set in Sect. 5.

It is well known that a point estimate is generally different from the true parameter value, say ϑ. Moreover, it does not convey any measure of reliability. Interval estimation is another type of estimation which contains more information about the data used to obtain the point estimate. It allows us to have some degree of confidence for securing ϑ. Interval estimators are called confidence intervals (CIs). Let L and U be two statistics such that $P(L < \vartheta < U) = 1 - \alpha$, for some $\alpha \in (0, 1)$. Then, (L, U) is a CI for ϑ with coverage probability (confidence level) $1 - \alpha$. The width of the CI reflects the amount of variability inherent in the point estimate. A good interval should be relatively narrow on the average, with high probability of enclosing the true parameter.

This article deals with constructing some CIs for θ in the RSS design. It is worth noting that point estimation of different population attributes have been comprehensively studied in the RSS literature, while hypothesis testing and interval estimation problems have received little attention (see Chen et al. 2004; Wolfe 2012; Chapter 15 in Hollander et al. 2014 for a good review of the RSS and its applications). Yin et al. (2016) proposed a CI for θ based on kernel density estimation. A simpler approach is to use empirical distribution function, which has not been investigated yet. We set out to fill this gap in this work. It emerges that the resulting intervals have an edge over the existing one.

In Sect. 2, our point estimator is introduced and its theoretical properties are studied. Some estimators for variance of this estimator also are presented. In Sect. 3, six types of intervals are developed. Section 4 contains results of Monte Carlo simulations assessing performances of the suggested intervals in terms of coverage probability and expected length. An agricultural data set is analyzed in Sect. 5. Final conclusions appear in Sect. 6. Proofs are put off to an appendix.

2 Nonparametric estimation

Let $\{X_{[r]i} : r = 1, \ldots, k ; i = 1, \ldots, m\}$ and $\{Y_{[s]j} : s = 1, \ldots, \ell ; j = 1, \ldots, n\}$ be independent ranked set samples from two populations with the distribution functions F and G, respectively. Also, the survival function associated with G is denoted by $\bar{G}$. The standard estimator of θ is given by

$$\hat{\theta}_{\text{RSS}} = \frac{1}{mkn\ell} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{k} \sum_{s=1}^{\ell} I(X_{[r]i} < Y_{[s]j}), \tag{1}$$

where $I(.)$ is the indicator function.

The properties of $\hat{\theta}_{\text{RSS}}$ in the especial case of $m = n = 1$ was investigated by Sengupta and Mukhuti (2008). They showed that the this estimator is unbiased and more efficient than its SRS counterpart, even in the presence of ranking errors. The following result shows asymptotic normality of $\hat{\theta}_{\text{RSS}}$.

Proposition 1 *Let $\hat{\theta}_{RSS}$ be as in (1), and $N = mk + n\ell$. If $m, n \to \infty$ and $(mk)/N \to \lambda \in (0, 1)$, then*

$$\sqrt{N}(\hat{\theta}_{RSS} - \theta) \xrightarrow{d} N\left(0, \frac{\sigma_1^2}{\lambda} + \frac{\sigma_2^2}{1 - \lambda}\right),$$

where

$$\sigma_1^2 = Var\left(\bar{G}(X)\right) - \frac{1}{k} \sum_{r=1}^{k} \left[E\left(\bar{G}(X_{[r]})\right) - \theta\right]^2,$$

and

$$\sigma_2^2 = Var\left(F(Y)\right) - \frac{1}{\ell}\sum_{s=1}^{\ell}\left[E\left(F(Y_{[s]})\right) - \theta\right]^2.$$

Suppose $\hat{\theta}_{SRS}$ is the counterpart of (1) based on two independent simple random samples of sizes mk and $n\ell$ from F and G, respectively. By virtue of the next result, $\hat{\theta}_{RSS}$ is asymptotically more efficient than $\hat{\theta}_{SRS}$. This statement is valid regardless of the accuracy of the judgment ranking process.

Proposition 2 *The asymptotic relative efficiency of $\hat{\theta}_{RSS}$ to $\hat{\theta}_{SRS}$ is*

$$ARE(\hat{\theta}_{RSS}, \hat{\theta}_{SRS}) = 1 + \frac{\frac{1-\lambda}{k}\sum_{r=1}^{k}\left[E\left(\bar{G}(X_{[r]})\right) - \theta\right]^2 + \frac{\lambda}{\ell}\sum_{s=1}^{\ell}\left[E\left(F(Y_{[s]})\right) - \theta\right]^2}{\frac{1-\lambda}{k}\sum_{r=1}^{k}Var\left(\bar{G}(X_{[r]})\right) + \frac{\lambda}{\ell}\sum_{s=1}^{\ell}Var\left(F(Y_{[s]})\right)}.$$

In light of the above result, interval estimation of θ based on the RSS is expected to be more efficient than that based on the SRS. An estimate of $Var(\hat{\theta}_{RSS})$ is needed to propose an interval based on Proposition 1. To the best of our knowledge, this has not been studied yet. In the sequel, we introduce an estimator for σ_1^2. Similar arguments yield an estimate of σ_2^2. These are combined to arrive at the final estimator.

Let $\{X_{[r]i} : r = 1, \ldots, k; \ i = 1, \ldots, m\}$ be a ranked set sample from a population with finite mean μ and variance σ^2. If $\mu_{[r]}$ and $\sigma_{[r]}^2$ denote the mean and variance of $X_{[r]1}$, respectively, then Stokes (1980) showed that

$$\sigma^2 = \frac{1}{k}\sum_{r=1}^{k}\sigma_{[r]}^2 + \frac{1}{k}\sum_{r=1}^{k}\left(\mu_{[r]} - \mu\right)^2. \tag{2}$$

Suppose the random variable W is defined as $W = \bar{G}(X)$. Then using (2), we get

$$Var(W) = \frac{1}{k}\sum_{r=1}^{k}Var(W_{[r]}) + \frac{1}{k}\sum_{r=1}^{k}\left[E(W_{[r]}) - E(W)\right]^2,$$

where $W_{[r]} = \bar{G}(X_{[r]1})$. That is to say that

$$\sigma_1^2 = \frac{1}{k}\sum_{r=1}^{k}Var(W_{[r]}).$$

Now, from Equation 3 in MacEachern et al. (2002), one can construct an estimator for σ_1^2 as

$$\hat{\sigma}_1^2 = \frac{1}{2km(m-1)}\sum_{r=1}^{k}\sum_{i=1}^{m}\sum_{i'=1}^{m}\left(\mathcal{W}_{[r]i} - \mathcal{W}_{[r]i'}\right)^2, \tag{3}$$

where

$$\mathcal{W}_{[r]i} = \frac{1}{n\ell} \sum_{s=1}^{\ell} \sum_{j=1}^{n} I(X_{[r]i} < Y_{[s]j}).$$

Similarly, an estimator of σ_2^2 is obtained as

$$\hat{\sigma}_2^2 = \frac{1}{2\ell n(n-1)} \sum_{s=1}^{\ell} \sum_{j=1}^{n} \sum_{j'=1}^{n} \left(\mathcal{Z}_{[s]j} - \mathcal{Z}_{[s]j'} \right)^2, \tag{4}$$

where

$$\mathcal{Z}_{[s]j} = \frac{1}{mk} \sum_{r=1}^{k} \sum_{i=1}^{m} I(X_{[r]i} < Y_{[s]j}).$$

Combining (3) and (4), we conclude that

$$\widehat{Var}(\hat{\theta}_{\mathrm{RSS}}) = \frac{1}{N} \left(\frac{\hat{\sigma}_1^2}{\hat{\lambda}} + \frac{\hat{\sigma}_2^2}{1-\hat{\lambda}} \right), \tag{5}$$

where $\hat{\lambda} = (mk)/N$. The above estimator is expected to work well for moderate to large values of m and n, but not for small choices of them. In the sequel, three alternatives are suggested.

The jackknife methodology has been proposed to serve two purposes, namely, to reduce a possible bias of an estimator, and to yield an approximation for its variance (see Quenouille 1956; Tukey 1958). Let $\hat{\theta}(X_1, \ldots, X_n)$ be a statistic of interest, where X_i's are iid random variables, and $\hat{\theta}$ is invariant under permutation of the arguments. If $\hat{\theta}^{(i)}$ denotes the value of $\hat{\theta}$ based on $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, then the jackknife estimate of $Var(\hat{\theta})$ is given by

$$\widehat{Var}(\hat{\theta}) = \frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}^{(i)} - \hat{\theta}^{(0)} \right)^2,$$

where $\hat{\theta}^{(0)} = \sum_{i=1}^{n} \hat{\theta}^{(i)}/n$.

A ranked set sample consists of independent but not identically distributed random variables. Therefore, one should adapt the above technique to estimate $Var(\hat{\theta}_{\mathrm{RSS}})$. The first method is to treat data as $m+n$ iid random variables $\mathbf{X}_1, \ldots, \mathbf{X}_m, \mathbf{Y}_1, \ldots, \mathbf{Y}_n$, where $\mathbf{X}_i = (X_{[1]i}, \ldots, X_{[k]i})$ $(i = 1, \ldots, m)$ and $\mathbf{Y}_j = (Y_{[1]j}, \ldots, Y_{[\ell]j})$ $(j = 1, \ldots, n)$. This is to say that $\mathbf{X}_i$ ($\mathbf{Y}_j$) contains the elements of the X (Y) sample drawn in the ith (jth) cycle. Suppose $\tilde{\theta}_{\mathrm{RSS}}^{(t)}$ is value of the reliability estimator when $\mathbf{Z}_t$, $t = 1, \ldots, m+n$, is omitted from the data, where

$$Z_t = \begin{cases} \mathbf{X}_t & t = 1, \ldots, m \\ \mathbf{Y}_{t-m} & t = m+1, \ldots, m+n \end{cases}.$$

Now, the jackknife estimate of the variance is

$$\widetilde{Var}_1(\hat{\theta}_{\text{RSS}}) = \frac{m+n-1}{m+n} \sum_{t=1}^{m+n} \left(\tilde{\theta}_{\text{RSS}}^{(t)} - \tilde{\theta}^{(0)} \right)^2, \tag{6}$$

where $\tilde{\theta}^{(0)} = \sum_{t=1}^{m+n} \tilde{\theta}_{\text{RSS}}^{(t)} / (m+n)$.

It is possible to obtain another jackknife-type estimate of the variance by excluding the cycles from the two samples simultaneously. Let $\breve{\theta}_{\text{RSS}}^{(u,v)}$ denote value of the estimator when $\mathbf{X}_u$ and $\mathbf{Y}_v$ are removed from the data. The second estimator is then

$$\widetilde{Var}_2(\hat{\theta}_{\text{RSS}}) = \frac{mn-1}{mn} \sum_{u=1}^{m} \sum_{v=1}^{n} \left(\breve{\theta}_{\text{RSS}}^{(u,v)} - \breve{\theta}^{(0)} \right)^2, \tag{7}$$

where $\breve{\theta}^{(0)} = \sum_{u=1}^{m} \sum_{v=1}^{n} \breve{\theta}_{\text{RSS}}^{(u,v)} / (mn)$.

The bootstrap method, introduced by Efron (1979), can also be used to estimate the variance. The method involves drawing samples repeatedly from the empirical distribution function. Suppose $X_1, \ldots, X_n$ is a random sample from the target population, and $\hat{\theta}$ is an estimator of interest. First we draw a sample of size n, with replacement, from the data points (called a bootstrap sample). This sampling procedure is repeated B times, and the estimator is computed from each bootstrap sample. The sample variance of these B values is then the bootstrap estimate of $Var(\hat{\theta})$.

Modarres et al. (2006) suggested two bootstrap algorithms in the RSS design. The bootstrap ranked set sampling (BRSS) method, which is the most efficient one, is now delineated. Let F_{mk} be the empirical distribution function based on the ranked set sample $\{X_{[r]i} : r = 1, \ldots, k; i = 1, \ldots, m\}$, i.e.

$$F_{mk}(x) = \frac{1}{mk} \sum_{r=1}^{k} \sum_{i=1}^{m} I(X_{[r]i} \le x).$$

According to the BRSS algorithm, a bootstrap sample is drawn as follows:

1. Assign to each element of the ranked set sample a probability of $(mk)^{-1}$.
2. Randomly draw k elements $\mathcal{X}_1, \ldots, \mathcal{X}_k \overset{iid}{\sim} F_{mk}$, sort them in ascending order $\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(k)}$, and retain $X_{[r]1}^* = \mathcal{X}_{(r)}$.
3. Perform step 2 for $r = 1, \ldots, k$.
4. Repeat steps 2 and 3 m times to obtain $\{X_{[r]i}^*\}$.

Following similar steps, a bootstrap copy of $\{Y_{[s]j} : s = 1, \ldots, \ell; j = 1, \ldots, n\}$ is generated. Suppose B pairs of bootstrap samples are drawn as described above, and let $\hat{\theta}_{\text{RSS}}^b$ be the value of the reliability estimator based on data in the bth ($b = 1, \ldots, B$) replication. Then bootstrap variance estimator is given by

$$\widehat{Var}_{\text{boot}}(\hat{\theta}_{\text{RSS}}) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}_{\text{RSS}}^{b} - \bar{\theta}^{*} \right)^{2}, \tag{8}$$

where $\bar{\theta}^{*} = \sum_{b=1}^{B} \hat{\theta}_{\text{RSS}}^{b} / B$.

3 Proposed CIs

In this section, we construct several CIs for θ using asymptotic and resampling methods. Based on Proposition 1, one can employ the pivotal quantity

$$T = \frac{\hat{\theta}_{\text{RSS}} - \theta}{\sqrt{\widehat{Var}(\hat{\theta}_{\text{RSS}})}} \approx N(0, 1),$$

where $\widehat{Var}(\hat{\theta}_{\text{RSS}})$ is defined in (5). The corresponding approximate $(1 - \alpha)$-CI is

$$\left(\hat{\theta}_{\text{RSS}} - z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\theta}_{\text{RSS}})}, \; \hat{\theta}_{\text{RSS}} + z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\theta}_{\text{RSS}})} \right), \tag{9}$$

where $z_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution. The pivotal quantity T can be altered if $\widehat{Var}(\hat{\theta}_{\text{RSS}})$ is replaced by one of the estimates presented in (6), (7) and (8). Accordingly, natural modifications of (9) would be

$$\left(\hat{\theta}_{\text{RSS}} - z_{\alpha/2} \sqrt{\widetilde{Var}_1(\hat{\theta}_{\text{RSS}})}, \; \hat{\theta}_{\text{RSS}} + z_{\alpha/2} \sqrt{\widetilde{Var}_1(\hat{\theta}_{\text{RSS}})} \right), \tag{10}$$

$$\left(\hat{\theta}_{\text{RSS}} - z_{\alpha/2} \sqrt{\widetilde{Var}_2(\hat{\theta}_{\text{RSS}})}, \; \hat{\theta}_{\text{RSS}} + z_{\alpha/2} \sqrt{\widetilde{Var}_2(\hat{\theta}_{\text{RSS}})} \right), \tag{11}$$

and

$$\left(\hat{\theta}_{\text{RSS}} - z_{\alpha/2} \sqrt{\widehat{Var}_{\text{boot}}(\hat{\theta}_{\text{RSS}})}, \; \hat{\theta}_{\text{RSS}} + z_{\alpha/2} \sqrt{\widehat{Var}_{\text{boot}}(\hat{\theta}_{\text{RSS}})} \right). \tag{12}$$

We can construct a two-sided equal-tailed $(1 - \alpha)$-CI for θ from the empirical distribution function of a series of bootstrap replications of $\hat{\theta}_{\text{RSS}}$. The $\alpha/2$ and the $1 - \alpha/2$ quantiles of the bootstrap replications are used as lower and upper confidence bounds. This procedure is called percentile bootstrap, and the corresponding interval is given by

$$\left(\hat{\theta}_{\text{RSS}}^{\alpha/2}, \hat{\theta}_{\text{RSS}}^{1-\alpha/2} \right), \tag{13}$$

where $\hat{\theta}_{\text{RSS}}^{\beta}$ is the β quantile of $\hat{\theta}_{\text{RSS}}^{1}, \ldots, \hat{\theta}_{\text{RSS}}^{B}$.

The bootstrap-t method approximates quantiles of the distribution of T from sample quantiles of the quantities

$$T_b = \frac{\hat{\theta}_{\text{RSS}}^{b} - \hat{\theta}_{\text{RSS}}}{\sqrt{\widehat{Var}(\hat{\theta}_{\text{RSS}}^{b})}} \quad (b = 1, \ldots, B),$$

where $\hat{\theta}^b_{\text{RSS}}$ and $\widehat{Var}(\hat{\theta}^b_{\text{RSS}})$ are computed from the bth bootstrap sample. The bootstrap-t interval is defined as

$$\left(\hat{\theta}_{\text{RSS}} - t_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{\text{RSS}})}, \, \hat{\theta}_{\text{RSS}} - t_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{\text{RSS}})} \right), \qquad (14)$$

where t_β is the β quantile of $T_1, \ldots, T_B$.

The intervals (9), (10), (11), (12), (13) and (14) will be referred to as Normal, Normal-J1, Normal-J2, Normal-B, Boot-p and Boot-t, respectively. It should be mentioned that all the above intervals, except Boot-p, may have endpoints outside the interval (0,1). Therefore, we correct the original interval (L, U) as $(\max\{0, L\}, \min\{1, U\})$.

4 Simulation results

This section contains results of simulation studies conducted to compare the performances of the different intervals suggested in the previous section. We consider the cases where both X and Y follow either a normal or exponential distribution. If $X - \mu$ ($\mu \in \mathbb{R}$) and Y are independent standard normal random variables, then

$$\theta = \Phi\left(\frac{-\mu}{\sqrt{2}}\right),$$

where $\Phi(.)$ is the distribution function of Y. Similarly, for independent standard exponential random variables X/β ($\beta > 0$) and Y, it can be shown that

$$\theta = \frac{1}{1 + \beta}.$$

Under each parent distribution, three values were assigned to the associated parameter so as to produce $\theta = 0.25, 0.5, 0.75$ which are referred to as case A, B and C, respectively. The appropriate parameter values are given in Table 1. If the total sample sizes are denoted by $N_1 = mk$ and $N_2 = n\ell$, then we select $(N_1, N_2) \in \{(10, 10), (10, 20), (10, 30), (20, 20)\}$. Also, ranked set samples are drawn from the two populations using common set sizes $k = \ell = 1, 2, 5$, where the set size one simply represents the SRS design.

We assume that the ranking the variables of interest X and Y are done based on the concomitant variables $\mathcal{X}$ and $\mathcal{Y}$ which are related according to equations

$$\mathcal{X} = \rho_1 \left(\frac{X - \mu_x}{\sigma_x}\right) + \sqrt{1 - \rho_1^2}Z_1,$$

and

$$\mathcal{Y} = \rho_2 \left(\frac{Y - \mu_y}{\sigma_y}\right) + \sqrt{1 - \rho_2^2}Z_2,$$

Table 1 Parameter values corresponding to case A, B and C

Parameter	A	B	C
μ	0.95387	0	-0.95387
β	3	1	1/3

where $\rho_i \in [0, 1]$ $(i = 1, 2)$, and Z_1 (Z_2) is a standard normal random variable independent from X (Y). Moreover, Z_1 and Z_2 are independent. The quality of rankings are controlled by the parameter ρ_i's. It is easy to see that $Corr(X, \mathcal{X}) = \rho_1$ and $Corr(Y, \mathcal{Y}) = \rho_2$. The chosen values of (ρ_1, ρ_2) are $(1, 1)$ for perfect rankings of X and Y, $(1, 0.8)$ for perfect ranking of X and fairly accurate ranking of Y, and $(0.8, 0.8)$ for fairly accurate rankings of X and Y.

For each combination of distribution, sample sizes and correlations, 5000 pairs of samples were generated in the RSS design (with the aforesaid set sizes). The six intervals were constructed from each pair of samples for $\alpha = 0.05$. In doing so, number of the bootstrap replications is chosen to be 500. Then, coverage rate and expected length of any interval is estimated by fraction of the intervals containing true θ, and mean of the intervals' lengths, respectively. The results with the perfect ranking are reported in Tables 2, 3, 4 and 5, where the lengths of intervals appear in parentheses.

It can be seen generally that the higher length of interval, the better coverage probability. Normal-J2 and Boot-t CIs have the best coverage rates, and the latter is always shorter. Also, Normal-B and Boot-p are the shortest CIs, and their coverage rates are more or less the same. For a fixed $N_1 + N_2$, performances of the CIs generally improve with equal sample sizes setup. Compare similar intervals for sample sizes (10, 30) and (20, 20) under different parent distributions.

Given a pair of total sample sizes, the lengths of all intervals are decreasing in the set size regardless of the case (A, B or C). However, changes in the coverage probabilities are not regular. The above trends are consistent with some results in the RSS literature. For example, Terpstra and Miller (2006) studied exact inference for a population proportion based on the RSS. According to their findings, expected length of the RSS-based CI is uniformly (as a function of the true population proportion) smaller than that of the SRS-based CI. However, there is not a uniform superiority for the coverage probability. See Figure 3 in Terpstra and Miller (2006). In our problem, the situation is more complex because the intervals are based on asymptotic and/or resampling methods.

If the perfect rankings are assumed, the performances of each interval for cases A and C are in close agreement when the parent distribution is normal. This statement is true about the exponential distribution if $N_1 = N_2$. These properties can also be observed in the imperfect ranking setup (see Tables 1–8 in the supplementary material), but the additional assumption $\rho_1 = \rho_2$ is needed for the exponential distribution. In the presence of ranking errors, lengths of the CIs increase (as compared with the perfect ranking case) but the coverage probabilities do not behave regularly. Overall, Normal-J2 and Boot-t CIs have satisfactory coverage rates (which are close to the nominal level or higher than it) in this situation, although they are longer than the other CIs.

Mahdizadeh and Zamanzade (2016) used kernel density estimation to estimate θ in the RSS. Let h_1 and h_2 be bandwidth of the kernel density estimator based on $\{X_{[r]i} :$

Table 2 Estimated coverage rates and lengths of 95% intervals under normal distribution with the perfect ranking when $(N_1, N_2) = (10, 10), (10, 20)$

(N_1, N_2)	k, ℓ	Normal			Normal-J1			Normal-J2		
		A	B	C	A	B	C	A	B	C
(10, 10)	1	90.5	92.9	90.2	91.2	93.5	91.2	99.2	99.9	99.3
		(0.423)	(0.528)	(0.426)	(0.433)	(0.542)	(0.436)	(0.880)	(0.995)	(0.886)
	2	91.5	93.8	91.5	92.0	93.3	91.6	99.1	99.9	99.3
		(0.366)	(0.433)	(0.366)	(0.383)	(0.454)	(0.381)	(0.699)	(0.890)	(0.697)
	5	90.4	92.1	90.8	85.6	86.8	85.1	94.4	94.6	93.8
		(0.273)	(0.310)	(0.272)	(0.300)	(0.344)	(0.297)	(0.420)	(0.505)	(0.418)

(N_1, N_2)	k, ℓ	Normal-B			Boot-p			Boot-t		
		A	B	C	A	B	C	A	B	C
(10, 10)	1	90.0	92.4	90.0	92.1	93.6	91.9	99.6	100	99.7
		(0.413)	(0.511)	(0.416)	(0.414)	(0.503)	(0.416)	(0.740)	(0.637)	(0.737)
	2	92.2	94.2	91.9	92.9	94.4	92.7	98.1	98.7	98.2
		(0.367)	(0.438)	(0.367)	(0.365)	(0.433)	(0.365)	(0.568)	(0.487)	(0.571)
	5	94.6	95.6	94.1	94.6	95.2	94.0	98.1	97.10	97.7
		(0.290)	(0.333)	(0.291)	(0.288)	(0.330)	(0.288)	(0.446)	(0.392)	(0.439)

Table 2 continued

(N_1, N_2)	k, ℓ	Normal			Normal-J1			Normal-J2		
		A	B	C	A	B	C	A	B	C
(10, 20)	1	91.3	93.2	91.1	91.8	93.8	91.6	99.7	100	99.8
		(0.370)	(0.451)	(0.369)	(0.380)	(0.463)	(0.378)	(0.912)	(0.998)	(0.910)
	2	91.9	93.5	91.8	92.1	93.3	92.1	99.8	100	99.8
		(0.315)	(0.367)	(0.314)	(0.329)	(0.383)	(0.328)	(0.728)	(0.911)	(0.726)
	5	91.2	92.7	91.7	88.2	88.7	87.7	97.2	97.3	97.0
		(0.232)	(0.262)	(0.233)	(0.255)	(0.290)	(0.255)	(0.433)	(0.523)	(0.434)

(N_1, N_2)	k, ℓ	Normal-B			Boot-p			Boot-t		
		A	B	C	A	B	C	A	B	C
(10, 20)	1	90.7	92.6	90.7	92.1	93.6	92.3	99.1	98.8	99.1
		(0.361)	(0.438)	(0.360)	(0.357)	(0.432)	(0.356)	(0.518)	(0.505)	(0.521)
	2	92.4	93.7	91.8	93.1	94.0	92.9	97.7	97.3	97.9
		(0.315)	(0.371)	(0.314)	(0.311)	(0.368)	(0.310)	(0.401)	(0.399)	(0.400)
	5	93.9	95.6	94.1	93.9	95.4	94.0	96.3	95.7	96.4
		(0.244)	(0.279)	(0.244)	(0.242)	(0.277)	(0.241)	(0.291)	(0.302)	(0.293)

Table 3 Estimated coverage rates and lengths of 95% intervals under normal distribution with the the perfect ranking when $(N_1, N_2) = (10, 30), (20, 20)$

(N_1, N_2)	k, ℓ	Normal			Normal-J1			Normal-J2		
		A	B	C	A	B	C	A	B	C
(10, 30)	1	90.8	93.1	91.1	91.6	93.8	91.6	99.9	100	100
		(0.345)	(0.420)	(0.346)	(0.355)	(0.433)	(0.356)	(0.936)	(0.999)	(0.938)
	2	91.1	92.9	91.0	91.5	93.1	91.2	99.8	100	99.9
		(0.293)	(0.343)	(0.293)	(0.308)	(0.360)	(0.308)	(0.779)	(0.937)	(0.777)
	5	91.3	92.4	90.9	88.1	87.8	88.0	97.9	98.0	97.5
		(0.216)	(0.244)	(0.216)	(0.240)	(0.273)	(0.244)	(0.465)	(0.564)	(0.473)

(N_1, N_2)	k, ℓ	Normal-B			Boot-p			Boot-t		
		A	B	C	A	B	C	A	B	C
(10, 30)	1	90.7	92.2	90.7	92.0	93.3	92.0	99.1	97.8	98.7
		(0.335)	(0.407)	(0.337)	(0.331)	(0.402)	(0.332)	(0.457)	(0.470)	(0.454)
	2	91.9	93.5	91.3	92.4	93.7	91.9	96.9	96.6	97.0
		(0.292)	(0.345)	(0.291)	(0.288)	(0.342)	(0.287)	(0.359)	(0.375)	(0.362)
	5	93.7	95.3	93.8	93.4	95.2	93.8	96.0	95.7	95.8
		(0.225)	(0.258)	(0.226)	(0.223)	(0.256)	(0.224)	(0.273)	(0.285)	(0.275)

Table 3 continued

(N_1, N_2)	k, ℓ	Normal			Normal-J1			Normal-J2		
		A	B	C	A	B	C	A	B	C
(20, 20)	1	92.6	93.9	93.2	92.8	94.2	93.4	100	100	99.9
		(0.303)	(0.367)	(0.304)	(0.307)	(0.371)	(0.308)	(0.902)	(0.999)	(0.904)
	2	93.8	94.3	93.2	94.0	94.0	93.3	100	100	100
		(0.258)	(0.299)	(0.257)	(0.263)	(0.305)	(0.263)	(0.677)	(0.905)	(0.672)
	5	93.4	94.2	94.2	92.4	92.4	93.0	99.6	99.7	99.7
		(0.191)	(0.214)	(0.192)	(0.202)	(0.225)	(0.202)	(0.408)	(0.469)	(0.407)

(N_1, N_2)	k, ℓ	Normal-B			Boot-p			Boot-t		
		A	B	C	A	B	C	A	B	C
(20, 20)	1	92.2	93.7	93.0	93.0	94.2	93.8	98.3	97.8	98.3
		(0.300)	(0.361)	(0.301)	(0.296)	(0.357)	(0.297)	(0.348)	(0.385)	(0.347)
	2	94.2	94.8	93.5	94.4	94.9	93.9	97.2	96.4	97.3
		(0.259)	(0.303)	(0.257)	(0.256)	(0.300)	(0.254)	(0.280)	(0.308)	(0.281)
	5	94.6	95.6	95.4	94.3	95.3	95.2	95.6	95.2	96.3
		(0.197)	(0.223)	(0.197)	(0.195)	(0.221)	(0.195)	(0.205)	(0.221)	(0.205)

Table 4 Estimated coverage rates and lengths of 95% intervals under exponential distribution with the perfect ranking when $(N_1, N_2) = (10, 10), (10, 20)$

(N_1, N_2)	k, ℓ	Normal			Normal-J1			Normal-J2		
		A	B	C	A	B	C	A	B	C
(10, 10)	1	89.8	92.3	90.4	90.5	92.8	91.1	98.9	100	99.0
		(0.426)	(0.528)	(0.429)	(0.436)	(0.543)	(0.439)	(0.882)	(0.994)	(0.887)
	2	91.2	93.7	91.1	91.5	93.2	91.0	99.2	99.9	99.3
		(0.369)	(0.433)	(0.369)	(0.384)	(0.451)	(0.384)	(0.702)	(0.887)	(0.700)
	5	89.6	92.3	90.6	84.6	87.4	86.0	93.7	94.5	94.2
		(0.273)	(0.310)	(0.273)	(0.297)	(0.347)	(0.299)	(0.416)	(0.508)	(0.420)

(N_1, N_2)	k, ℓ	Normal-B			Boot-p			Boot-t		
		A	B	C	A	B	C	A	B	C
(10, 10)	1	89.4	91.7	90.0	91.1	92.9	91.7	99.4	99.9	99.6
		(0.415)	(0.511)	(0.418)	(0.416)	(0.503)	(0.419)	(0.745)	(0.638)	(0.746)
	2	91.8	93.9	91.6	92.8	94.2	92.6	97.8	98.9	98.0
		(0.371)	(0.439)	(0.370)	(0.369)	(0.434)	(0.368)	(0.578)	(0.489)	(0.584)
	5	93.5	95.7	94.3	93.3	95.2	94.2	98.2	97.5	97.8
		(0.292)	(0.333)	(0.293)	(0.290)	(0.330)	(0.291)	(0.458)	(0.393)	(0.460)

Table 4 continued

(N_1, N_2)	k, ℓ	Normal			Normal-J1			Normal-J2		
		A	B	C	A	B	C	A	B	C
(10, 20)	1	90.5	93.2	92.1	91.1	93.6	92.6	99.6	100	99.9
		(0.391)	(0.452)	(0.354)	(0.402)	(0.464)	(0.362)	(0.929)	(0.998)	(0.891)
	2	90.6	93.0	92.2	90.6	92.9	92.6	99.5	100	99.9
		(0.333)	(0.368)	(0.300)	(0.349)	(0.384)	(0.311)	(0.772)	(0.912)	(0.683)
	5	90.5	92.9	92.4	87.0	89.5	90.7	96.5	97.6	98.0
		(0.244)	(0.262)	(0.224)	(0.266)	(0.291)	(0.247)	(0.457)	(0.525)	(0.414)

(N_1, N_2)	k, ℓ	Normal-B			Boot-p			Boot-t		
		A	B	C	A	B	C	A	B	C
(10, 20)	1	90.1	92.5	91.7	91.4	93.4	93.2	99.3	99.0	98.9
		(0.380)	(0.439)	(0.347)	(0.377)	(0.433)	(0.343)	(0.603)	(0.507)	(0.462)
	2	91.4	93.3	92.5	92.4	93.4	93.0	97.5	96.4	97.8
		(0.332)	(0.372)	(0.301)	(0.328)	(0.368)	(0.297)	(0.469)	(0.400)	(0.359)
	5	94.1	96.1	94.5	94.0	95.8	94.6	96.4	96.1	96.5
		(0.260)	(0.280)	(0.233)	(0.257)	(0.277)	(0.231)	(0.344)	(0.302)	(0.262)

Table 5 Estimated coverage rates and lengths of 95% intervals under exponential distribution with the perfect ranking when $(N_1, N_2) = (10, 30), (20, 20)$

(N_1, N_2)	k, ℓ	Normal			Normal-J1			Normal-J2		
		A	B	C	A	B	C	A	B	C
(10, 30)	1	90.0	92.8	92.3	90.7	93.5	92.8	99.9	100	100
		(0.374)	(0.421)	(0.321)	(0.386)	(0.434)	(0.329)	(0.953)	(0.999)	(0.914)
	2	89.6	93.1	93.3	89.9	93.2	93.5	99.8	100	99.9
		(0.318)	(0.342)	(0.271)	(0.336)	(0.359)	(0.282)	(0.832)	(0.937)	(0.714)
	5	89.2	92.6	92.6	85.3	88.2	90.6	96.2	97.8	98.5
		(0.233)	(0.243)	(0.199)	(0.260)	(0.274)	(0.221)	(0.508)	(0.565)	(0.423)

(N_1, N_2)	k, ℓ	Normal-B			Boot-p			Boot-t		
		A	B	C	A	B	C	A	B	C
(10, 30)	1	89.4	92.3	92.0	91.0	93.1	93.0	98.8	97.8	98.6
		(0.361)	(0.408)	(0.314)	(0.357)	(0.403)	(0.310)	(0.557)	(0.472)	(0.385)
	2	90.3	93.6	93.2	91.2	94.0	93.8	96.3	96.6	97.5
		(0.316)	(0.345)	(0.271)	(0.312)	(0.342)	(0.268)	(0.435)	(0.374)	(0.307)
	5	94.0	95.3	94.4	93.8	95.2	94.6	95.4	96.0	96.0
		(0.248)	(0.259)	(0.208)	(0.245)	(0.257)	(0.206)	(0.338)	(0.284)	(0.229)

Table 5 continued

(N_1, N_2)	k, ℓ	Normal			Normal-J1			Normal-J2		
		A	B	C	A	B	C	A	B	C
(20, 20)	1	92.4	93.7	92.9	92.6	93.9	93.2	99.9	100	100
		(0.306)	(0.366)	(0.306)	(0.310)	(0.371)	(0.310)	(0.903)	(0.999)	(0.905)
	2	93.1	94.5	93.1	93.3	94.1	93.2	100	100	100
		(0.260)	(0.300)	(0.260)	(0.265)	(0.306)	(0.266)	(0.679)	(0.906)	(0.680)
	5	93.3	94.0	93.3	92.4	92.3	91.9	99.7	99.7	99.7
		(0.194)	(0.214)	(0.193)	(0.203)	(0.225)	(0.202)	(0.408)	(0.469)	(0.407)

(N_1, N_2)	k, ℓ	Normal-B			Boot-p			Boot-t		
		A	B	C	A	B	C	A	B	C
(20, 20)	1	92.1	93.2	92.7	93.3	93.7	93.6	98.2	97.7	98.2
		(0.302)	(0.361)	(0.302)	(0.298)	(0.357)	(0.298)	(0.352)	(0.385)	(0.353)
	2	93.2	94.5	93.2	93.8	94.8	93.9	96.9	95.9	96.9
		(0.260)	(0.302)	(0.260)	(0.257)	(0.300)	(0.258)	(0.285)	(0.308)	(0.286)
	5	95.0	95.5	95.0	94.8	95.1	94.7	95.9	94.9	95.9
		(0.199)	(0.223)	(0.199)	(0.197)	(0.221)	(0.197)	(0.208)	(0.221)	(0.207)

$r = 1, \ldots, k \, ; i = 1, \ldots, m\}$ and $\{Y_{[s]j} : s = 1, \ldots, \ell \, ; j = 1, \ldots, n\}$, respectively (see Chen 1999 for the kernel density estimation in the RSS). If $t = \sqrt{h_1^2 + h_2^2}$, then kernel-based estimator is given by

$$\tilde{\theta}_{RSS} = \frac{1}{mkn\ell} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{k} \sum_{s=1}^{\ell} \Phi\left(\frac{Y_{[s]j} - X_{[r]i}}{t}\right),$$

where $\Phi(.)$ is the distribution function of standard normal random variable. Yin et al. (2016) established asymptotic normality of the above estimator, and employed this result in developing a CI for θ. The corresponding interval is defined as

$$\left(\tilde{\theta}_{RSS} - z_{\alpha/2}\sqrt{\widehat{Var}(\tilde{\theta}_{RSS})}, \tilde{\theta}_{RSS} + z_{\alpha/2}\sqrt{\widehat{Var}(\tilde{\theta}_{RSS})}\right), \qquad (15)$$

where $\widehat{Var}(\tilde{\theta}_{RSS})$ is computed similar to (5) based on

$$\mathcal{W}_{[r]i} = \frac{1}{n\ell} \sum_{s=1}^{\ell} \sum_{j=1}^{n} \Phi\left(\frac{Y_{[s]j} - X_{[r]i}}{t}\right),$$

and

$$\mathcal{Z}_{[s]j} = \frac{1}{mk} \sum_{r=1}^{k} \sum_{i=1}^{m} \Phi\left(\frac{Y_{[s]j} - X_{[r]i}}{t}\right).$$

We conducted a partial simulation study to compare CIs (9) and (15) in terms of the coverage probability and length, based on 10,000 pairs of samples. In determining h_1 and h_2, the following three methods of the bandwidth selection were utilized: normal reference (NR) rule, unbiased cross-validation (UCV), and plug-in (PI). Although these techniques are developed for the SRS (see Sheather 2004 for more details), they can be applied in the RSS setup by considering data as if collected by the SRS.

Figures 1 and 2 display the results for $(N_1, N_2) = (10, 10)$ with $k = \ell = 2, 5$, when the perfect rankings are assumed. Here, black/solid curve is corresponding to the interval (9). Also, blue/dashed, red/dotted and orange/longdash curves are associated with the interval (15) using NR, UCV and PI methods, respectively. It is observed that the interval (9) has better coverage rate, while the interval (15) is always shorter. Hence, there is not a single interval preferred from both aspects. Among the kernel-based intervals, overall performance of the CI using PI method is satisfactory.

5 Illustration

We now apply the proposed procedures to an agricultural data set. Murray et al. (2000) conducted an experiment in which apple trees are sprayed with chemical containing fluorescent tracer, Tinopal CBS-X, at 2% concentration level in water. Two nine-tree
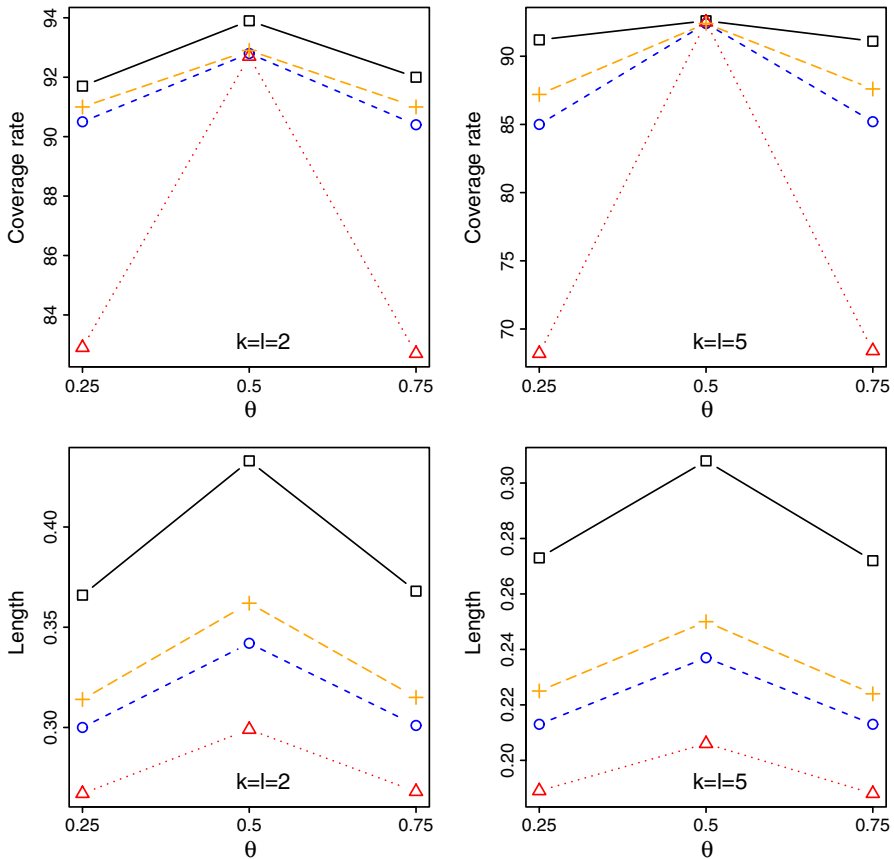
Fig. 1 Estimated coverage rates and lengths of 95% intervals under normal distribution with the perfect ranking when $(N_1, N_2) = (10, 10)$

plots were chosen for spraying. One plot was sprayed at high volume, using coarse nozzles on the sprayer to give a large average droplet size. The other plot was sprayed at low volume, using fine nozzles to give a small average droplet size. Fifty sets of five leaves were identified from the central five trees of each plot, and used to draw 10 copies a ranked set sample of size five, from each plot. The variable of interest is the percentage of area covered by the spray on the surface of the leaves. The formal measurement entails chemical analysis of the solution collected from the surface of the leaves, and thereby is a time-consuming and expensive process. The judgment ranking within each set is based on the visual appearance of the spray deposits on the leaf surfaces when viewed under ultraviolet light. Clearly, the latter method is cheap, and fairly accurate if implemented by an expert observer.

The data are given in Table 6, where measurements obtained from the plot sprayed at high (low) volume constitute the control (treatment) group. The interest centers on knowing whether the sprayer settings affect the percentage area coverage. If $X(Y)$ denotes the response variable from the control (treatment) group, then $\hat{\theta}_{\text{RSS}}$
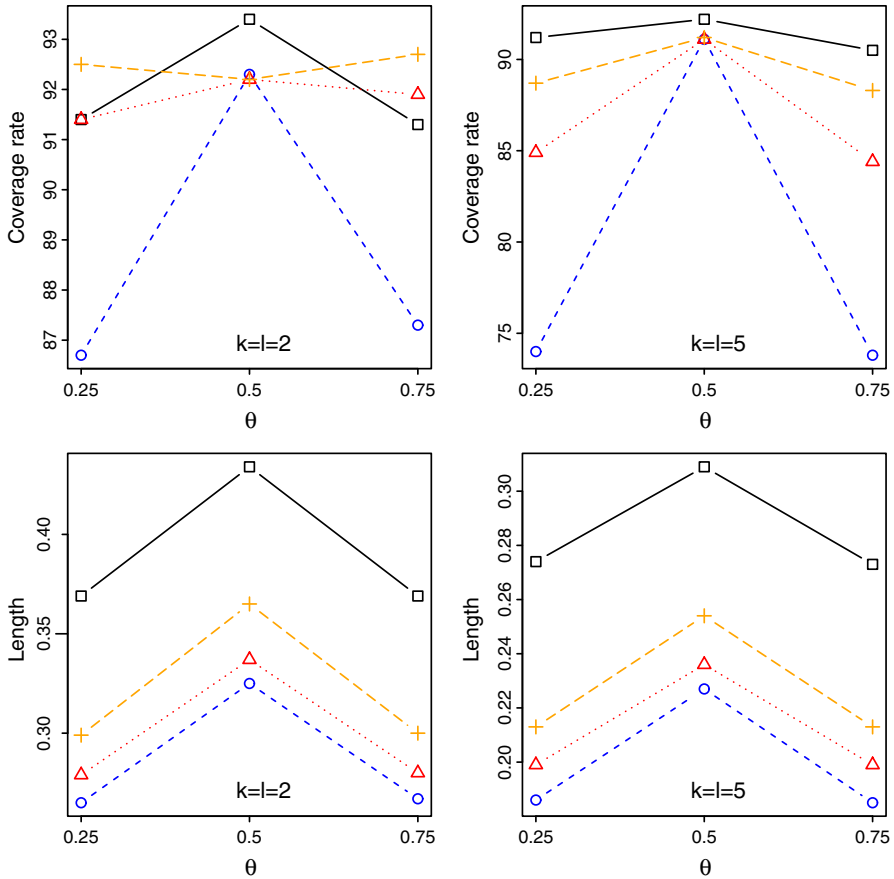
Fig. 2 Estimated coverage rates and lengths of 95% intervals under exponential distribution with the perfect ranking when $(N_1, N_2) = (10, 10)$

is a measure of the treatment effect. From the data in Table 6, $\hat{\theta}_{\text{RSS}} = 0.6184$ is obtained with estimated variances $\widetilde{Var}(\hat{\theta}_{\text{RSS}}) = 0.001344$, $\widetilde{Var}_1(\hat{\theta}_{\text{RSS}}) = 0.001825$, $\widetilde{Var}_2(\hat{\theta}_{\text{RSS}}) = 0.019038$, and $\widetilde{Var}_{\text{boot}}(\hat{\theta}_{\text{RSS}}) = 0.001169$. For the bootstrap-based estimate, $B = 5000$ is used. It is seen that with the exception of $\widetilde{Var}_2(\hat{\theta}_{\text{RSS}})$, all of the estimates are in good agreement. Table 7 displays 95% CIs for θ based on different methods. Apart from Normal-J2 interval, we may conclude that the treatment effect is significant at 0.05 level as none of the intervals contain 0.5. It should be mentioned that Normal-J2 is the longest interval in this example, and this is consistent with simulation results in Sect. 4.

As a reviewer pointed out, the accuracy of sampling and statistical inference largely hinges on properties of the population, conditions of the sample, and the method of estimation, so called sampling and statistical trinity (see Wang et al. 2012, for example). Here, the proposed procedures are illustrated using agricultural data. Spatial population may be dominated by spatial autocorrelation, spatial stratified heterogeneity, or both.

Table 6 Ranked set sample data for the percentage area covered on the surface of the leaves of apple trees

Group	Copy	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Control	1	0.003	0.028	0.244	0.057	0.143
	2	0.039	0.119	0.126	0.105	0.565
	3	0.034	0.118	0.130	0.218	0.296
	4	0.051	0.104	0.193	0.210	0.150
	5	0.032	0.141	0.130	0.250	0.229
	6	0.069	0.070	0.260	0.225	0.285
	7	0.100	0.091	0.244	0.130	0.347
	8	0.012	0.096	0.069	0.373	0.133
	9	0.046	0.117	0.126	0.223	0.273
	10	0.028	0.083	0.108	0.212	0.261
Treatment	1	0.036	0.137	0.183	0.270	0.487
	2	0.250	0.181	0.290	0.328	0.715
	3	0.089	0.032	0.269	0.419	0.315
	4	0.180	0.111	0.130	0.194	0.742
	5	0.100	0.009	0.184	0.277	0.122
	6	0.042	0.089	0.199	0.269	0.395
	7	0.044	0.083	0.227	0.177	0.742
	8	0.044	0.171	0.067	0.192	0.336
	9	0.009	0.017	0.217	0.438	0.544
	10	0.071	0.132	0.310	0.343	0.379

Table 7 95% CIs for θ based on the apple trees data

Normal	Normal-J1	Normal-J2
(0.547, 0.690)	(0.535, 0.702)	(0.348, 0.889)
Normal-B	Boot-p	Boot-t
(0.551, 0.685)	(0.551, 0.683)	(0.546, 0.691)

Also, there may be significant covariates. The properties of the population should be tested before making a choice of the most suitable one among numerous estimators. To justify the choice of a method, a table may be drawn to compare the assumptions of the mainstream models in the topic and the properties of the data under study (e.g. spatial autocorrelation, spatial stratified heterogeneity, and the significance of covariates). Unfortunately, we have not any information about the population from which our sample in Table 6 is drawn. Thus, it is not possible to check the aforesaid properties.

6 Conclusion

The RSS method combines measurement with the judgment ranking information for purpose of statistical inference. It is advantageous in settings where precise mea-

surement on the variable of interest is difficult (e.g., time-consuming, expensive or destructive), but small sets of units can be accurately ranked without actual quantification.

While point estimation of different population attributes have been exhaustively studied in the RSS literature, hypothesis testing and interval estimation problems have received little attention. This article aims to fill this gap in the context of estimating the reliability parameter. Several asymptotic and resampling-based intervals are developed, and compared with their SRS analogs through extensive simulation study. The results confirm the preference of the RSS-based CIs with respect to length, although their coverage rates are not uniformly superior. An agricultural data set is used to illustrate the suggested interval estimation procedures.

The intervals presented in this work utilize a point estimator constructed based on empirical distribution function. We have partly investigated performance of one of the CIs modified using kernel density estimation. The other intervals can be adapted similarly. This will be considered in a separate article.

Appendix

In this section, we first provide some results about the two-sample U-statistics, and then present proofs of Propositions 1 and 2.

Suppose that $h(x_1, \ldots, x_p; y_1, \ldots, y_q)$ is a symmetric kernel of degree (p, q) for the parameter $\theta = E(h)$. For independent simple random samples $X_1, \ldots, X_{mk}$ from F, and $Y_1, \ldots, Y_{n\ell}$ from G, the corresponding two-sample U-statistic for θ is given by

$$
\begin{aligned}
U_{\text{SRS}} &= U(X_1, \ldots, X_{mk}; Y_1, \ldots, Y_{n\ell}) \\
&= \frac{1}{\binom{mk}{p}\binom{n\ell}{q}} \sum_{\alpha \in \mathcal{A}} \sum_{\beta \in \mathcal{B}} h(X_{\alpha_1}, \ldots, X_{\alpha_p}; Y_{\beta_1}, \ldots, Y_{\beta_q}),
\end{aligned}
$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)$, and $\mathcal{A}$ ($\mathcal{B}$) is the collection of all subsets of size p (q) chosen from integers $1, \ldots, mk$ ($1, \ldots, n\ell$). Let

$$
h_{10}(x) = E\left(h(x, X_2, \ldots, X_p; Y_1, \ldots, Y_q)\right),
$$

and

$$
h_{01}(y) = E\left(h(X_1, \ldots, X_p; y, Y_2, \ldots, Y_q)\right).
$$

Also, in connection with the above functions, define $\zeta_{10} = Var\left(h_{10}(X_1)\right)$ and $\zeta_{01} = Var\left(h_{01}(Y_1)\right)$.

Let $\{X_{[r]i} : r = 1, \ldots, k \, ; i = 1, \ldots, m\}$ and $\{Y_{[s]j} : s = 1, \ldots, \ell \, ; j = 1, \ldots, n\}$ be independent ranked set samples from two populations with the distribution functions F and G, respectively. The two-sample U-statistic for θ in the RSS is given by

$$U_{\text{RSS}} = U(\mathbf{X}_1, \ldots, \mathbf{X}_m; \mathbf{Y}_1, \ldots, \mathbf{Y}_n),$$

where $\mathbf{X}_i = (X_{[1]i}, \ldots, X_{[k]i})$ $(i = 1, \ldots, m)$, and $\mathbf{Y}_j = (Y_{[1]j}, \ldots, Y_{[\ell]j})$ $(j = 1, \ldots, n)$. In addition, suppose $\gamma_{r0} = E\left(h_{10}(X_{[r]1})\right)$, $\gamma_{0s} = E\left(h_{01}(Y_{[s]1})\right)$, $\xi_{r0} = Var\left(h_{10}(X_{[r]1})\right)$, and $\xi_{0s} = Var\left(h_{01}(Y_{[s]1})\right)$.

Assume that $N = mk + n\ell$, and $(mk)/N \to \lambda \in (0, 1)$ as $m, n \to \infty$. Then, according to Theorem 2 in Presnell and Bohn (1999),

$$\sqrt{N}(U_{\text{RSS}} - \theta) \xrightarrow{d} N\left(0, \frac{p^2 \phi}{\lambda} + \frac{q^2 \varphi}{1 - \lambda}\right),$$

where

$$\phi = \zeta_{10} - \frac{1}{k} \sum_{r=1}^{k} (\gamma_{r0} - \theta)^2,$$

and

$$\varphi = \zeta_{01} - \frac{1}{\ell} \sum_{s=1}^{\ell} (\gamma_{0s} - \theta)^2.$$

If we set $h(x, y) = I(x < y)$, which is a kernel of degree $(1, 1)$, then $\theta = P(X < Y)$ and Proposition 1 follows.

Also, from Corollary 2 in Presnell and Bohn (1999), the asymptotic relative efficiency of U_{RSS} to U_{SRS} is

$$ARE(U_{\text{RSS}}, U_{\text{SRS}}) = 1 + \frac{\frac{(1-\lambda)p^2}{k} \sum_{r=1}^{k} (\gamma_{r0} - \theta)^2 + \frac{\lambda q^2}{\ell} \sum_{s=1}^{\ell} (\gamma_{0s} - \theta)^2}{\frac{(1-\lambda)p^2}{k} \sum_{r=1}^{k} \xi_{r0} + \frac{\lambda q^2}{\ell} \sum_{s=1}^{\ell} \xi_{0s}}.$$

Again, by the same choice of the kernel mentioned above, Proposition 2 is concluded.

References

Ayech MW, Ziou D (2015) Segmentation of Terahertz imaging using k-means clustering based on ranked set sampling. Expert Syst Appl 42:2959–2974

Chen Z (1999) Density estimation using ranked-set sampling data. Environ Ecol Stat 6:135–146

Chen Z (2007) Ranked set sampling: its essence and some new applications. Environ Ecol Stat 14:355–363

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

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Stat 7:1–26

Halls LK, Dell TR (1966) Trial of ranked-set sampling for forage yields. For Sci 12:22–26

Hollander M, Wolfe DA, Chicken E (2014) Nonparametric statistical methods, 3rd edn. Wiley, New York

Howard RW, Jones SC, Mauldin JK, Beal RH (1982) Abundance, distribution, and colony size estimates for *Reticulitermes* spp. (Isoptera: Rhinotermitidae) in Southern Mississippi. Environ Entomol 11:1290–1293

Kotz S, Lumelskii Y, Pensky M (2003) The stress-strength model and its generalizations. Theory and applications. World Scientific, Singapore

Kvam P (2003) Ranked set sampling based on binary water quality data with covariates. J Agric Biol Environ Stat 8:271–279

MacEachern SN, Ozturk O, Stark GV, Wolfe DA (2002) A new ranked set sample estimator of variance. J R Stat Soc Ser B 64:177–188

Mahdizadeh M, Zamanzade E (2016) Kernel-based estimation of $P(X>Y)$ in ranked set sampling. SORT 40:243–266

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

Modarres R, Hui TP, Zheng G (2006) Resampling methods for ranked set samples. Comput Stat Data Anal 51:1039–1050

Murray RA, Ridout MS, Cross JV (2000) The use of ranked set sampling in spray deposit assessment. Asp Appl Biol 57:141–146

Presnell B, Bohn L (1999) U-statistics and imperfect ranking in ranked set sampling. J Nonparametr Stat 10:111–126

Quenouille MH (1956) Notes on bias in estimation. Biometrika 43:353–360

Sengupta S, Mukhuti S (2008) Unbiased estimation of $P(X>Y)$ using ranked set sample data. Statistics 42:223–230

Sheather SJ (2004) Density estimation. Stat Sci 19:588–597

Stokes SL (1980) Estimation of variance using judgment ordered ranked set samples. Biometrics 36:35–42

Terpstra J, Miller ZA (2006) Exact inference for a population proportion based on a ranked set sample. Commun Stat Simul Comput 35:19–26

Tukey JW (1958) Bias and confidence in not quite large samples (abstract). Ann Math Stat 29:614

Wang JF, Stein A, Gao BB, Ge Y (2012) A review of spatial sampling. Spat Stat 2:1–14

Wolfe DA (2012) Ranked set sampling: its relevance and impact on statistical inference. ISRN Probability and Statistics, pp 1–32

Yin J, Hao Y, Samawi H, Rochani H (2016) Rank-based kernel estimation of the area under the ROC curve. Stat Methodol 32:91–106

Zamanzade E, Mahdizadeh M (2017) A more efficient proportion estimator in ranked set sampling. Stat Probab Lett 129:28–33