Jackknife empirical likelihood inferences for the population mean with ranked set samples

Zhengjia Zhang^a, Tianqing Liu^{b,*}, Baoxue Zhang^{a,*}

^a*Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics,
Northeast Normal University, Changchun 130024, China*

^b*School of Mathematics, Jilin University, Changchun 130012, China*

Abstract

Without requiring any easily violated assumptions needed by existing rank-based nonparametric methods for ranked set samples, we propose a nonparametric approach for interval estimation and hypothesis testing for the population mean and the difference between two population means with balanced and unbalanced ranked set samples using jackknife empirical likelihood.

Keywords: Jackknife, Empirical likelihood, Ranked set samples, Population mean, Difference of means

1. Introduction

Ranked set sampling (RSS) was first introduced by McIntyre (1952) when studying the yields of pasture in Australia. He argued that RSS often provides more efficient inference than simple random sampling (SRS) of the same size. Since then, RSS has been widely used when the measurement of an observed item is costly and/or time-consuming but the ranking of a set of items can be easily done without actual measurement. McIntyre (1952) and Takahasi and Wakimoto (1968) proposed unbalanced RSS (UBRSS) as follows:

$$X_{[1]1}, \dots, X_{[1]n_1}, \dots, X_{[k]1}, \dots, X_{[k]n_k}. \quad (1)$$

*Corresponding author.

Email addresses: zhengjiazhang813@126.com (Zhengjia Zhang^a),
tianqingliu@outlook.com (Tianqing Liu^{b,*}), bxzhang@nenu.edu.cn (Baoxue Zhang^{a,*})

When $n_1 = \cdots = n_k$, it is the balanced RSS (BRSS). Throughout the paper, we will use notations with subscript $[a]$ to represent the values for the common distribution of the a -th judgement statistics $X_{[a]1}, \dots, X_{[a]n_a}$ corresponding to those of the population distribution. For example, if F is the cumulative distribution function (c.d.f.) for the population X , then $F_{[a]}$ denotes the c.d.f. of the common distribution of $X_{[a]1}, \dots, X_{[a]n_a}$. It is worth noting that, when the judgement ranking is perfect, $F_{[a]}$ is just the c.d.f. of the a -th-order statistic of a simple random sample (SRS) of size k from F , $F_{(a)}$; and on the other hand, when the judgement ranking is made at random, $F_{[a]}$ is the population c.d.f. F . We assume that the ranking mechanism is consistent in the sense that the fundamental equality

$$F(x) = \frac{1}{k} \sum_{a=1}^k F_{[a]}(x) \quad (2)$$

holds for all x (Chen et al., 2004, p.12-14). The crucial problem with the UBRSS is how to allocate the n_a 's. McIntyre (1952) and Kaur et al. (1997) showed that the estimator of the population mean based on UBRSS (1) with $n_a = \langle n \cdot \sqrt{\text{Var}(X_{[a]})} / \sum_{b=1}^k \sqrt{\text{Var}(X_{[b]})} \rangle$, $a = 1, \dots, k-1$, $n_k = n - \sum_{b=1}^{k-1} n_b$ has the smallest asymptotic variance among all RSS-based estimators with k fixed, where $\langle x \rangle$ denotes the integer closest to x , and n is the number of items to be actually measured. One can refer to Chen et al. (2004) and Wolfe (2012) for a comprehensive survey and references on this topic.

In recent years, some authors have made nonparametric inferences for parameters of interest with RSS-data using empirical likelihood (EL) method (Owen, 1990, 2001; Qin and Lawless, 1994). Baklizi (2009) used EL method to obtain intervals for the population mean and quantiles with BRSS-data, Baklizi (2011) generalized the method of Baklizi (2009) to UBRSS-data (1) for population quantiles, but the asymptotic null distributions of the log-empirical likelihood ratio statistics in Baklizi (2009, 2011) are weighted chi-square distributions. Liu et al. (2009) proposed a method called RSS-EL to obtain a nonparametric version of Wilk's theorem for parameters defined by general estimating equations with BRSS-data. However, it may not be an easy task to extend the RSS-EL method to UBRSS-data (1).

In this paper, using jackknife empirical likelihood (JEL) (Jing et al., 2009), we propose a systematic nonparametric approach, RSS-JEL, for interval estimation and hypothesis testing with balanced and unbalanced RSS-

data (1). The asymptotic null distributions of the log-empirical likelihood ratio statistics for the population mean and the difference between two population means are derived. Interval estimates for the population mean and the difference between two population means are obtained by inverting the log-likelihood ratio statistics. Moreover, the RSS-JEL method does not require any easily violated assumptions needed by existing rank-based nonparametric methods for RSS-data, such as perfect ranking, identical ranking scheme in two groups, and location shift between two population distributions. The merit of the RSS-JEL method is also demonstrated through numerical studies.

The rest of this paper is organized as follows. In Sections 2-3, we introduce JEL-based inference for the population mean and the difference between two population means with RSS-data. Numerical studies are reported in Sections 4-5.

2. JEL-based inference for the population mean with RSS-data

In this section we use the JEL method to make inference for the population mean

$$\mu = E(X) = \int x dF(x)$$

with RSS-data (1). We make the following assumption.

(A1) The ranking mechanism is consistent, F has a finite second moment, and the total sample size $n = n_1 + \cdots + n_k$ tends to infinity in such a way that n_a/n converges to a positive number q_a for each $a = 1, \dots, k$.

Since F has a finite second moment, so does each $F_{[a]}$ and this lets us define

$$\sigma_*^2 = \frac{1}{k^2} \sum_{a=1}^k \frac{\sigma_a^2}{q_a}$$

with σ_a the standard deviation of $F_{[a]}$. Next, set

$$T_n = \frac{1}{k} \sum_{a=1}^k \bar{X}_{[a]} \quad \text{with} \quad \bar{X}_{[a]} = \frac{1}{n_a} \sum_{b=1}^{n_a} X_{[a]b}.$$

For each $i = 1, \dots, n$, there are unique $r_i \in \{1, \dots, k\}$ and $j_i \in \{1, \dots, n_{r_i}\}$ such that i can be expressed as the sum $\sum_{a=0}^{r_i-1} n_a + j_i$ with $n_0 = 0$. This lets

us introduce the leave-one-out versions of T_n as

$$T_{n-1}^{(-i)} = \frac{1}{k} \sum_{a=1, a \neq r_i}^k \bar{X}_{[a]} + \frac{1}{k(n_{r_i} - 1)} \sum_{b=1, b \neq j_i}^{n_{r_i}} X_{[r_i]b}, \quad i = 1, \dots, n.$$

Then we can define the jackknife pseudo-values by $V_i = nT_n - (n-1)T_{n-1}^{(-i)}$. Simple calculations show that

$$V_i = T_n + \frac{n-1}{k(n_{r_i} - 1)} (X_{[r_i]j_i} - \bar{X}_{[r_i]}), \quad i = 1, \dots, n. \quad (3)$$

From this we easily conclude that the average of these pseudo-values equals T_n ,

$$\bar{V}_n = \frac{1}{n} \sum_{i=1}^n V_i = T_n.$$

The JEL for the mean is given by

$$L_n(t) = \sup \left\{ \prod_{i=1}^n np_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i V_i = t \right\}, \quad t \in \mathbb{R}.$$

Theorem 2.1. *Suppose (A1) holds. If μ is the true value of the population mean, then the following three statements are true.*

$$n^{1/2}(\bar{V}_n - \mu) \xrightarrow{d} N(0, \sigma_*^2), \quad (a1)$$

$$\frac{1}{n} \sum_{i=1}^n (V_i - \mu)^2 = \sigma_*^2 + o_p(1), \quad (a2)$$

$$\max_{1 \leq i \leq n} |V_i - \mu| = o_p(n^{1/2}). \quad (a3)$$

Moreover, if σ_*^2 is positive, then $-2 \log L_n(\mu)$ converges in distribution to a chi-square random variable with one degree of freedom.

PROOF. Let μ_a be the mean of $F_{[a]}$. From (3) it is immediate that

$$n^{1/2}(\bar{V}_n - \mu) = n^{1/2}(T_n - \mu) = \frac{1}{k} \sum_{a=1}^k \frac{\sqrt{n}}{\sqrt{n_a}} \frac{1}{\sqrt{n_a}} \sum_{b=1}^{n_a} (X_{[a]b} - \mu_a) \xrightarrow{d} N(0, \sigma_*^2),$$

$$\frac{1}{n} \sum_{i=1}^n (V_i - \mu)^2 = (T_n - \mu)^2 + \sum_{a=1}^k \frac{(n-1)^2}{nk^2(n_a-1)^2} \sum_{b=1}^{n_a} (X_{[a]b} - \bar{X}_{[a]})^2 = \sigma_*^2 + o_p(1).$$

By Lemma 11.2 in Owen (2001), we have

$$\max_{1 \leq i \leq n} |V_i - \mu| \leq |T_n - \mu| + \max_{1 \leq a \leq k} \frac{n-1}{k(n_a-1)} \max_{1 \leq b \leq n_a} 2|X_{[a]b}| = o_p(n^{1/2}).$$

Moreover, if σ_*^2 is positive, properties (a1)-(a3) and Theorem 6.1 in Peng and Schick (2013) then yield that $-2 \log L_n(\mu)$ is asymptotically chi-square with one degree of freedom.

Let $\chi_1^2(1-\alpha)$ be the $1-\alpha$ quantile of a chi-square random variable with one degree of freedom, where $0 < \alpha < 1$. Using Theorem 2.1, we obtain an approximate $1-\alpha$ log-empirical likelihood confidence interval for μ , defined by $\{t \mid -2 \log L_n(t) \leq \chi_1^2(1-\alpha)\}$. Theorem 2.1 can also be used to test the hypothesis $H_0 : t = \mu$. One could reject H_0 at level α if $-2 \log L_n(t) > \chi_1^2(1-\alpha)$.

3. JEL-based inference for the difference between two population means with RSS-data

Suppose that we have in addition to the ranked set sample in (1) an independent ranked set sample

$$Y_{[1]1}, \dots, Y_{[1]m_1}, \dots, Y_{[l]1}, \dots, Y_{[l]m_l}$$

from a distribution G with mean ν . We are now interested in inference about the difference

$$\Delta = \nu - \mu = \int x dG(x) - \int x dF(x)$$

between two population means. We impose the analogue conditions to (A1) on this sample.

(A2) The ranking mechanism is consistent, G has a finite second moment, and the total sample size $m = m_1 + \dots + m_l$ tends to infinity such that m_b/m converges to a positive number d_b for each $b = 1, \dots, l$.

In addition, we require

$$(A3) \quad n/m \rightarrow \rho > 0.$$

We write τ_b for the standard deviation of $G_{[b]}$ and set

$$\tau_*^2 = \frac{1}{l^2} \sum_{b=1}^l \frac{\tau_b^2}{d_b}.$$

Analogous to T_n define

$$S_m = \frac{1}{l} \sum_{a=1}^l \bar{Y}_{[a]} \quad \text{with} \quad \bar{Y}_{[a]} = \frac{1}{m_a} \sum_{j=1}^{m_a} Y_{[a]j}.$$

Let $W_j = mS_m - (m-1)S_{m-1}^{(-j)}$, $j = 1, \dots, m$ denote the jackknife pseudo-values for the second sample, where $S_{m-1}^{(-j)}$, $j = 1, \dots, m$ are the leave-one-out versions of S_m . The JEL for the mean and the second sample is

$$K_n(t) = \sup \left\{ \prod_{j=1}^m m p_j : p_j \geq 0, \sum_{j=1}^m p_j = 1, \sum_{j=1}^m p_j W_j = t \right\}, \quad t \in \mathbb{R}.$$

The JEL for Δ is

$$R(s) = L_n(\hat{\mu}) K_m(\hat{\mu} + s), \quad s \in \mathbb{R},$$

where $\hat{\mu}$ is the pooled estimator of μ defined by

$$\hat{\mu} = \frac{nS_W T_n + mS_V(S_m - \Delta)}{mS_V + nS_W}$$

with

$$S_V = \frac{1}{n} \sum_{i=1}^n (V_i - T_n)^2 \quad \text{and} \quad S_W = \frac{1}{m} \sum_{j=1}^m (W_j - S_m)^2,$$

and this estimator has the property that

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N \left(0, \frac{\rho \sigma_*^2 \tau_*^2}{\sigma_*^2 + \rho \tau_*^2} \right).$$

Theorem 3.1. *Suppose (A1)-(A3) hold, and σ_*^2 and τ_*^2 are positive. If Δ is the true value of the difference between two population means, then $-2 \log R(\Delta)$ converges in distribution to a chi-square random variable with one degree of freedom.*

PROOF. Similar to the proof of theorem 6.1 in Peng and Schick (2013), one can show that

$$-2 \log L_n(\hat{\mu}) = \frac{n(T_n - \hat{\mu})^2}{\sigma_*^2} + o_p(1).$$

In fact, let $T_{ni} = V_i - \hat{\mu}$, then the conditions of theorem 6.1 in Peng and Schick (2013) are satisfied by noting that

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n T_{ni} &= \sqrt{n}(T_n - \mu) - \sqrt{n}(\hat{\mu} - \mu) \\ &= \frac{mS_V \sqrt{n}[(T_n - \mu) - (S_m - \Delta - \mu)]}{mS_V + nS_W} \xrightarrow{d} N(0, \sigma_*^4/(\sigma_*^2 + \rho\tau_*^2)), \end{aligned}$$

$$\frac{1}{n} \sum_{i=1}^n T_{ni}^2 = \frac{1}{n} \sum_{i=1}^n (V_i - \mu)^2 - 2(T_n - \mu)(\hat{\mu} - \mu) + (\hat{\mu} - \mu)^2 = \sigma_*^2 + o_p(1),$$

and

$$\max_{1 \leq i \leq n} |T_{ni}| \leq \max_{1 \leq i \leq n} |V_i - \mu| + |\hat{\mu} - \mu| = o_p(n^{1/2}).$$

Similarly, one can verify that

$$-2 \log K_m(\hat{\mu} + \Delta) = \frac{m(S_m - \Delta - \hat{\mu})^2}{\tau_*^2} + o_p(1).$$

Combining the above results, we obtain

$$\begin{aligned} -2 \log R(\Delta) &= \frac{n(T_n - \hat{\mu})^2}{\sigma_*^2} + \frac{m(S_m - \Delta - \hat{\mu})^2}{\tau_*^2} + o_p(1) \\ &= \frac{m^2 S_V^2 \tau_*^2 + mn S_W^2 \sigma_*^2}{\sigma_*^2 \tau_*^2 (mS_V + nS_W)^2} n[(T_n - \mu) - (S_m - \Delta - \mu)]^2 + o_p(1) \\ &= \frac{n[(T_n - \mu) - (S_m - \Delta - \mu)]^2}{\sigma_*^2 + \rho\tau_*^2} + o_p(1). \end{aligned}$$

Moreover, by the central limit theorem, we have

$$n^{1/2}[(T_n - \mu) - (S_m - \Delta - \mu)] \xrightarrow{d} N(0, \sigma_*^2 + \rho\tau_*^2),$$

Based on these two facts, it follows that $-2\log R(\Delta)$ is asymptotically chi-square with one degree of freedom.

Let $\chi_1^2(1 - \alpha)$ be the $1 - \alpha$ quantile of a chi-square random variable with one degree of freedom, where $0 < \alpha < 1$. Using Theorem 3.1, we obtain an approximate $1 - \alpha$ log-empirical likelihood confidence interval for Δ , defined by $\{s \mid -2\log R(s) \leq \chi_1^2(1 - \alpha)\}$. Theorem 3.1 can also be used to test the hypothesis $\tilde{H}_0 : s = \Delta$. One could reject $\tilde{H}_0$ at level α if $-2\log R(s) > \chi_1^2(1 - \alpha)$.

4. Simulation study

In this section, we use simulation studies to show the efficiency of the RSS-JEL approach over the EL approach based on SRSs (SRS-EL). In each simulation, we generate 5,000 Monte Carlo samples to compare the coverage rate and average length of the 95% confidence intervals based on SRS and RSS of the same size. For RSSs, we consider perfect ranking (RSS), imperfect ranking (RSS*). The imperfect ranking is introduced by adding random perturbations $\epsilon \sim \text{Uniform}(-0.5, 0.5)$.

We perform two simulations. The first one is for one-sample problem: $X \sim F \equiv N(0, 1)$ and $X^* = X + \epsilon$. Here F denotes the population distributions, and X^* is concomitant variable used for ranking sampling units under imperfect ranking (Chen et al., 2004, p.12-14). The parameter of interest is the population mean, and the sampling scheme used for X is $n_r = \langle n \cdot \sqrt{\text{Var}(X_{(r)})} / \sum_{a=1}^3 \sqrt{\text{Var}(X_{(a)})} \rangle$, $r = 1, 2$, $n_3 = n - n_1 - n_2$. Simulation results are given in Table 1. The other one is for two-sample problem with unequal size of ranked sets in two groups: $X \sim F \equiv N(0, 1)$, $Y \sim G \equiv \chi_1^2$, $X^* = X + \epsilon$ and $Y^* = Y + \epsilon$. Here F and G denote the population distributions, and X^* and Y^* are concomitant variables used for ranking sampling units under imperfect ranking (Chen et al., 2004, p.12-14). The sampling schemes used for X and Y are $n_r = \langle n \cdot \sqrt{\text{Var}(X_{(r)})} / \sum_{a=1}^3 \sqrt{\text{Var}(X_{(a)})} \rangle$, $r = 1, 2$, $n_3 = n - n_1 - n_2$, and $m_1 = \langle m \cdot \sqrt{\text{Var}(Y_{(1)})} / \sum_{b=1}^2 \sqrt{\text{Var}(Y_{(b)})} \rangle$, $m_2 = m - m_1$ respectively, and the parameter of interest is the difference between two population means. Simulation results are given in Table 2.

The results given in Tables 1-2 clearly shows that the advantages of the RSS-JEL and RSS*-JEL over the SRS-EL, since the RSS-based confidence intervals are always shorter than the SRS-based intervals and achieve the

correct coverage level. Moreover, these scenarios violate the assumptions required by the rank-based nonparametric methods in Bohn and Wolfe (1992, 1994), and Fligner and MacEachern (2006), yet our RSS-JEL method still achieves correct coverage probabilities and outperforms the SRS-based method.

Table 1 This table reports the coverage rates and average lengths of the 95% confidence intervals of the population mean of variable $X \sim F \equiv N(0, 1)$ with ranked set samples in the first simulation. n_a is the number of measurements on units with rank a , $a = 1, 2, 3$; $n = n_1 + n_2 + n_3$.

Method	$(n_1, n_2, n_3; n)$	Coverage	Length		$(n_1, n_2, n_3; n)$	Coverage	Length
RSS-JEL	(3, 3, 3; 9)	0.922	0.9759		(9, 8, 9; 26)	0.947	0.5729
RSS*-JEL	(3, 3, 3; 9)	0.924	1.0110		(9, 8, 9; 26)	0.945	0.5934
SRS-EL	(*, *, *, 9)	0.899	1.1892		(*, *, *, 26)	0.936	0.7630
RSS-JEL	(6, 5, 6; 17)	0.936	0.7140		(24, 22, 24; 70)	0.951	0.3443
RSS*-JEL	(6, 5, 6; 17)	0.938	0.7403		(24, 22, 24; 70)	0.950	0.3564
SRS-EL	(*, *, *, 17)	0.931	0.9251		(*, *, *, 70)	0.945	0.4684

Table 2 This table reports the coverage rates and average lengths of the 95% confidence intervals of the difference between two population means of variables $X \sim F \equiv N(0, 1)$ and $Y \sim G \equiv \chi_1^2$ with ranked set samples in the second simulation. For X , n_a is the number of measurements on units with rank a , $a = 1, 2, 3$; $n = n_1 + n_2 + n_3$. For Y , m_b is the number of measurements on units with rank b , $b = 1, 2$; $m = m_1 + m_2$.

Method	$(n_1, n_2, n_3; n)$	$(m_1, m_2; m)$	Coverage	Length
RSS-JEL	(3, 3, 3; 9)	(2, 6; 8)	0.905	1.7683
RSS*-JEL	(3, 3, 3; 9)	(2, 6; 8)	0.907	1.8047
SRS-EL	(*, *, *, 9)	(*, *, 8)	0.869	1.9891
RSS-JEL	(6, 5, 6; 17)	(4, 12; 16)	0.927	1.2868
RSS*-JEL	(6, 5, 6; 17)	(4, 12; 16)	0.929	1.3112
SRS-EL	(*, *, *, 17)	(*, *, 16)	0.909	1.5472
RSS-JEL	(9, 8, 9; 26)	(6, 18; 24)	0.937	1.0478
RSS*-JEL	(9, 8, 9; 26)	(6, 18; 24)	0.936	1.0665
SRS-EL	(*, *, *, 26)	(*, *, 24)	0.912	1.2981
RSS-JEL	(24, 22, 24; 70)	(17, 51; 68)	0.941	0.6361
RSS*-JEL	(24, 22, 24; 70)	(17, 51; 68)	0.944	0.6472
SRS-EL	(*, *, *, 70)	(*, *, 68)	0.932	0.8085

In Figure 1, we give the chi-square QQ-plots for the RSS-JEL statistics to verify their null distributions given in Theorems 2.1 and Theorem 3.1. No matter the ranking is perfect or not, the plots clearly show that the RSS-JEL statistics always follow the asymptotic chi-square distribution χ_1^2 closely.

Figure 1: Theoretical quantiles of χ_1^2 random variable versus empirical quantiles of log-empirical likelihood ratio evaluated at the true value of parameter of interest for the proposed method with ranked set samples. Fig.A and Fig.B are for the case $(n_1, n_2, n_3; n) = (9, 8, 9; 26)$ in Table 1 with ratio $-2\log L_n(\mu)$ in Theorem 2.1 under **perfect** and **imperfect** ranking respectively, while Fig.C and Fig.D are for the case $(n_1, n_2, n_3; n) = (9, 8, 9; 26)$ and $(m_1, m_2; m) = (6, 18; 24)$ in Table 2 with ratio $-2\log R(\Delta)$ in Theorem 3.1 under **perfect** and **imperfect** ranking respectively.

5. An Application

In this section, we apply our RSS-JEL method to the data collected from 396 conifer (*pinus palustris*) trees, which are given in the Appendix B of the book *Ranked Set Sampling : Theory and Applications* (Chen et al. 2004, p.208-211). The 396 data are records of X (the diameter in centimeters at breast height) and Y (the entire height in feet).

Here, the mean of Y is our interest, and X is the concomitant variable used for ranking sampling units under imperfect ranking (Chen et al. 2004, p.12-14). The sample mean and variance of Y are $\hat{\mu} = 52.6768$, $\hat{\sigma}^2 = 3253.44$, and the sample variances of the three order statistics $Y_{(1)}$, $Y_{(2)}$ and $Y_{(3)}$ are $\hat{\sigma}_{(1)}^2 = 156.37$, $\hat{\sigma}_{(2)}^2 = 1140.40$ and $\hat{\sigma}_{(3)}^2 = 4174.35$, respectively.

We sample 5,000 SRSs with size n and 5,000 RSSs with $n_a = \langle n \cdot \hat{\sigma}_{(a)} / \sum_{b=1}^3 \hat{\sigma}_{(b)} \rangle$, $a = 1, 2$ and $n_3 = n - n_1 - n_2$ from the 396 data to compute the average length of the 95% confidence intervals, where $\langle x \rangle$ denotes the integer closest to x . We consider different sample sizes $n = 15, 24, 33$ and 42 .

The results given in Table 3 clearly show the advantages of the RSS-JEL and RSS*-JEL over the SRS-EL as the RSS-JEL and RSS*-JEL-based confidence intervals are always shorter than the SRS-based intervals.

Table 3 This table reports the average lengths of the 95% confidence intervals of the population mean of variable Y with ranked set samples. n_a is the number of measurements on units with rank a , $a = 1, 2, 3$; $n = n_1 + n_2 + n_3$.

Method	$(n_1, n_2, n_3; n)$	Length	Method	$(n_1, n_2, n_3; n)$	Length
RSS-JEL	(2, 4, 9; 15)	41.2754	RSS-JEL	(4, 10, 19; 33)	27.5278
RSS*-JEL	(2, 4, 9; 15)	43.5718	RSS*-JEL	(4, 10, 19; 33)	29.1447
RSS-JEL	(5, 5, 5; 15)	48.2060	RSS-JEL	(11, 11, 11; 33)	31.9739
RSS*-JEL	(5, 5, 5; 15)	49.4479	RSS*-JEL	(11, 11, 11; 33)	32.7787
SRS-EL	(*, *, *, 15)	53.9854	SRS-EL	(*, *, *, 33)	38.2490
RSS-JEL	(3, 7, 14; 24)	32.5321	RSS-JEL	(5, 13, 24; 42)	24.3609
RSS*-JEL	(3, 7, 14; 24)	34.6099	RSS*-JEL	(5, 13, 24; 42)	25.8687
RSS-JEL	(8, 8, 8; 24)	37.7164	RSS-JEL	(14, 14, 14; 42)	27.9440
RSS*-JEL	(8, 8, 8; 24)	38.6412	RSS*-JEL	(14, 14, 14; 42)	28.7229
SRS-EL	(*, *, *, 24)	43.9453	SRS-EL	(*, *, *, 42)	33.7914

6. Conclusion

In this paper, we proposed an approach for making inference for the population mean and the difference between two population means with balanced and unbalanced ranked set samples using jackknife empirical likelihood. The approach does not require any easily violated assumptions needed by existing rank-based nonparametric methods for ranked set samples. Moreover, the approach can be used for making inference for other parameters of interest, such as population quantiles. We will present these developments in subsequent papers.

Acknowledgements

Zhengjia Zhang and Baoxue Zhang were partly supported by the Program for the National Science Foundation of China (No.11271064), Ph.D. Programs Foundation of Ministry of Education of China (No.20100043110002) and Fund of Jilin Provincial Science & Technology Department (No.20111804). Tianqing Liu was partly supported by the NSFC (No.11201174) and the Natural Science Foundation for Young Scientists of Jilin Province, China (No.20150520054JH). The authors are grateful to the editor, the associate editor and two referees for their valuable comments and suggestions which have been important in improving the paper.

References

- Baklizi, A., 2009. Empirical likelihood intervals for the population mean and quantiles based on balanced ranked set samples. *Stat. Methods Appl.* 18, 483-505.

- Baklizi, A., 2011. Empirical likelihood inference for population quantiles with unbalanced ranked set samples. *Communications in Statistics - Theory and Methods* 40, 4179-4188.
- Bohn, L.L., Wolfe, D.A., 1992. Nonparametric two-sample procedures for ranked-set samples data. *J. Amer. Statist. Assoc.* 87, 552-561.
- Bohn, L.L., Wolfe, D.A., 1994. The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitney-Wilcoxon statistics. *J. Amer. Statist. Assoc.* 89, 168-176.
- Chen, Z.H., Bai, Z.D., Sinha, B.K., 2004. *Ranked Set Sampling: Theory and Applications*. Springer-Verlag, New York.
- Fligner, M.A., MacEachern, S.N., 2006. Nonparametric two-sample methods for ranked-set sample data. *J. Amer. Statist. Assoc.* 101, 1107-1118.
- Jing, B.Y., Yuan, J.Q., Zhou, W., 2009. Jackknife empirical likelihood. *J. Amer. Statist. Assoc.* 104, 1224-1232.
- Kaur, A., Patil, G.P., Taillie, C., 1997. Unequal allocation models for ranked set sampling with skew distributions. *Biometrics* 53, 123-130.
- Liu, T.Q., Lin, N., Zhang, B.X., 2009. Empirical likelihood for balanced ranked-set sampled data. *Science in China Series A* 52, 1351-1364.
- McIntyre, G.A., 1952. A method of unbiased selective sampling, using ranked sets. *Australian Journal Agriculture Research* 3, 385-390.
- Owen, A.B., 1990. Empirical likelihood ratio confidence regions. *Ann. Statist.* 18, 90-120.
- Owen, A.B., 2001. *Empirical likelihood*. Chapman & Hall/CRC, New York.
- Peng, H., Schick A., 2013. Empirical likelihood approach to goodness of fit testing. *Bernoulli* 19, 954-981.
- Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. *Ann. Statist.* 22, 300-325.
- Takahasi, K., Wakimoto, K., 1968. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics* 30, 814-824.
- Wolfe, D.A., 2012. Ranked set sampling: Its relevance and impact on statistical inference. *ISRN Probab. Statist.* 2012, article ID 568385.