Communications of the Korean Statistical Society Vol. 15, No. 2, 2008, pp. 147–159

Inference on Overlapping Coefficients in Two Exponential Populations Using Ranked Set Sampling

Hani M. Samawi[1], Mohammad F. Al-Saleh[2]

Abstract

We consider using ranked set sampling methods to draw inference about the three well-known measures of overlap, namely Matusita's measure ρ, Morisita's measure λ and Weitzman's measure Δ. Two exponential populations with different means are considered. Due to the difficulties of calculating the precision or the bias of the resulting estimators of overlap measures, because there are no closed-form exact formulas for their variances and their exact sampling distributions, Monte Carlo evaluations are used. Confidence intervals for those measures are also constructed via the bootstrap method and Taylor series approximation.

Keywords: Bootstrap method; Matusita's measure; Morisita's measure; overlap coefficients; Taylor expansion; Weitzman's measure; ranked set sampling.

1. Introduction

Overlap measures are widely used in reliability theory, ecology and other fields. Three overlap coefficients (OVL), (Matusita's measure ρ, Morisita's measure λ and Weitzman's measure Δ) were found in the literature. However, Weitzman's measure Δ is the most commonly used overlap coefficient. This OVL measure is defined as the intersection area of two probability density functions. It measures the similarity, the agreement or the closeness of the two probability distributions. It introduced first by Weitzman (1970) and then many other authors considered it (see for example, Bradley and Piantadosi, 1982; Inman and Bradley, 1989; Clemons, 1996; Reiser and Faraggi, 1999; Clemons and Bradley, 2000; Mulekar and Mishra, 2000).

Applications of Δ can be found in Ichikawa (1993) (for the probability of failure in the stress-strength models of reliability analysis), Federer *et al.* (1963) (for estimating of the proportion of genetic deviates in segregating populations and Sneath (1977) (as a measure of distinctness of clusters). For additional references of such methodology

1) Associate Professor, Jiann-Ping Hsu College of Public Health Biostatistics Center, Georgia Southern University, PO Box 8148, Statesboro, GA 30460, USA.
Correspondence: hsamawi@georgiasouthern.edu

2) Professor, Department of Statistics, Yarmouk University, Irbid-Jordan 211-63.
E-mail: m-saleh@yu.edu.jo

applications in ecology and other fields, see Mulekar and Mishra (1994, 2000). Also, the history of such procedures is summarized by Inman and Bradley (1989).

In many agricultural and environmental studies and recently in human populations and reliability analysis, quantification (the actual measurement) of a sampling unit can be more costly than the physical acquisition of the unit (see for example, Samawi and Al-Sakeer, 2001).

In many sampling survey and experimental studies considerable cost savings can be achieved if the number of measured sampling units (experimental units) is only a small fraction of the number of available units but all available units contribute to the information content of the measured units. Ranked set sampling(RSS) is a method of sampling that can achieve the goal of reducing the sampling cost. RSS was first introduced by McIntyre (1952). The use of RSS in testing hypotheses procedure is highly powerful and superior to the standard simple random sampling(SRS) and also more efficient than using simple random sample for estimating some of the population parameters (see, Kaur *et al.*, 1995; Patil *et al.*, 1999).

The RSS procedure can be summarized as follows: Select r random samples, each of size r units from the population and rank the units within each sample with respect to a variable of interest visually or by any inexpensive judgmental method. Then an actual measurement is taken from the unit with the smallest rank from the first sample. From the second sample, an actual measurement is taken from the unit with the second smallest rank and the procedure is continued until the unit with the largest rank is chosen for actual measurement from the r^{th} sample. In this way, we obtain a total of r measured units, one from each sample. The cycle may be repeated m times until $n = mr$ units have been measured.

Variations of RSS such as extreme ranked set sampling(ERSS) and median ranked set sampling(MRSS) were investigated by Samawi *et al.* (1996a) and Muttlak (1997) respectively. Samawi and Muttlak (1996, 2001) used RSS and MRSS to improve the efficiency of the ratio estimator comparing with simple random sample procedure. Moreover, Al-Saleh and Al-Kadiri (2000) showed that the efficiency of estimating the population mean can be significantly improved more when double ranked set sampling scheme(DRSS) is considered. They proved that ranking in the second stage is easier than in the first stage. Samawi (2001) suggested double extreme ranked set sampling scheme(DERSS) for estimating the population mean using naïve and regression estimators. Also, Al-Saleh and Al-Omari (2002) introduced the multistage ranked set sampling(MIRSS). More details about RSS, are available in Kaur *et al.* (1995) and Patil *et al.* (1999).

1.1. General setting and definitions of OVL measures

Let $f_1(x)$ and $f_2(x)$ be two probability density functions. Assuming samples of observations are drawn from continuous distributions (Slobdchikoff and Schulz, 1980; Harner and Whitmorte, 1977; MacArthur, 1972). The overlap measures are defined as follows:

$$\text{Matusita's Measure (1955)}: \quad \rho = \int \sqrt{f_1(x)\ f_2(x)}dx,$$

$$\text{Morisita's Measure (1959)}: \quad \lambda = \frac{2\int f_1(x) f_2(x) dx}{\int \{f_1(x)\}^2 \{f_2(x)\}^2 dx}$$

$$\text{Weitzman's Measure (1970)}: \Delta = \int \min\{f_1(x) f_2(x)\} dx.$$

These measures can be directly applied to discrete distributions by replacing the integrals with summations and also can be generalized to multivariate distributions. All three overlap measures of two densities are measured on the scale of 0 to 1. Note that, the overlap value close to 0 indicates extreme inequality of the two density functions and the overlap value of 1 indicates exact equality.

The mathematical structure of these measures is complicated; there are no results available on the exact sampling distributions of their estimators. Researcher such as Smith (1982) derived formulas for estimating the mean and the variance of the discrete version of Weizman's measure using delta method. Mishra *et al.* (1986) gave some properties of the sampling distributions for a function of $\widehat{\Delta}$, under the assumption of homogeneity of variances for the case of two normal distributions. Mulekar and Mishra (1994) simulated the sampling distribution of estimators of the overlap measures for normal densities with equal means and obtained the approximate expressions for the bias and variance of their estimators. Lu *et al.* (1989) investigated the sampling variability of some estimators of these measures using simulation Dixon (1993) describes the use of the bootstrap and jackknife techniques for Gini coefficient of size hierarchy and Jaccard index of community similarity. Mulekar and Mishra (2000) addressed the problem of making inferences about the overlap coefficients for two normal densities with equal means using jackknife, bootstrap, transformation and Taylor series approximation. Reiser and Faraggi (1999) considered the problem of making inference about the overlap coefficient, as a measure of bioequivalence under the name proportion of similar responses, for normal densities with the equal variances, based on the non-central t- and F-distributions. The sampling behavior of a nonparametric estimator of was examined by Clemons and Bradley (2000), using Monte Carlo and bootstrap techniques. Moreover, Al-Saidy *et al.* (2005) investigated drawing inference about OVL measures for two weibul distributions with equal shape parameter.

In this paper all above three overlap measures (ρ, λ and Δ) are considered for two exponential distributions with different means using RSS. The exponential distribution has been used in reliability applications. It is used to model data with a constant failure rate (indicated by the hazard plot which is simply equal to a constant, see Mann *et al.*, 1974.)

A random variable X follows the exponential (denotes by EXP(θ)) if it has the *cdf* and *pdf* given by:

$$F(x) = 1 - \exp\left(-\frac{x}{\theta}\right), \quad \text{for } x > 0 \tag{1.1}$$

and

$$f(x) = \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right), \quad \text{for } x > 0, \tag{1.2}$$

respectively, where $\theta > 0$.

2. Overlap Measures (OVL) for Exponential Distribution

Suppose $f_1(x)$ and $f_2(x)$ represent the exponential densities with θ_1 and θ_2 means respectively. Let $R = \theta_1/\theta_2$, as in Mulekar *et al.* (2001) and Al-Saleh and Samawi (2006), then the continuous version of the three proposed overlap measures, can be expressed as a function of R as follows (the derivation of the three overlap measures are straight forward and it is omitted from the content of this paper):

$$\rho = \frac{2\sqrt{R}}{1+R}, \tag{2.1}$$

$$\lambda = \frac{4R}{(1+R)^2}, \tag{2.2}$$

$$\Delta = 1 - R^{\frac{1}{1-R}}\left|1 - \frac{1}{R}\right|, \quad R \neq 1. \tag{2.3}$$

Note that, all three measures are not monotone for all $R > 0$. Similar to Mulekar and Mishra (2000), ρ, λ and Δ have nice properties, such as, symmetry in R, *i.e.*, $\mathrm{OVL}(R) = \mathrm{OVL}(1/R)$ and invariance under linear transformation, $Y = aX + b$, $a \neq 0$. They all attain the maximum value of 1 at $R = 1$.

3. Statistical Inference Using RSS

3.1. Estimation

The OVL measures ρ, λ and Δ are functions of θ_1 and θ_2. In order to draw any inference about the OVL measures, we need first to get estimates for θ_1 and θ_2. Suppose $(X_{1(1)k}, X_{1(2)k}, \ldots, X_{1(r_1)k})$ and $(X_{2(1)k}, X_{2(2)k}, \ldots, X_{2(r_2)k})$, $k = 1, 2, \ldots, m$ are two independent RSS samples drawn from $f_1(x)$ and $f_2(x)$ respectively, where

$$f_1(x) = \frac{1}{\theta_1}\exp\left(-\frac{x}{\theta_1}\right), \quad \text{for } x > 0,$$

$$f_2(x) = \frac{1}{\theta_2}\exp\left(-\frac{x}{\theta_2}\right), \quad \text{for } x > 0.$$

The empirical estimators based on the two RSS samples are given by:

1. From the first sample

$$\widehat{\theta}_1 = \overline{X}_{(1)} = \frac{\sum_{i=1}^{r_1}\sum_{k=1}^{m} X_{1(i)k}}{n_1}, \quad \text{where } n_1 = r_1 m. \tag{3.1}$$

2. From the second sample

$$\widehat{\theta}_2 = \overline{X}_{(2)} = \frac{\sum_{i=1}^{r_2}\sum_{k=1}^{m} X_{2(i)k}}{n_2}, \quad \text{where } n_2 = r_2 m. \tag{3.2}$$

Note that, it is easy to show

$$E(\widehat{\theta}_1) = \theta_1, \qquad E(\widehat{\theta}_2) = \theta_2,$$

$$\mathrm{Var}(\widehat{\theta}_1) = \frac{\theta_1^2}{mr_1^2}\sum_{i=1}^{r_1}\frac{1}{r_1 - i + 1} \quad \text{and} \quad \mathrm{Var}(\widehat{\theta}_2) = \frac{\theta_2^2}{mr_2^2}\sum_{i=1}^{r_2}\frac{1}{r_2 - i + 1}.$$

Also, R can be estimated by $\widehat{R}_{RSS} = \widehat{\theta}_1/\widehat{\theta}_2$. Hence, by using
Delta method of approximation, the variance of $\widehat{R}$ can be approximated by

$$\mathrm{Var}(\widehat{R}) = R^2\left[\frac{\sum_{i=1}^{r_1}\frac{1}{r_1 - i + 1}}{mr_1^2} + \frac{\sum_{i=1}^{r_2}\frac{1}{r_2 - i + 1}}{mr_2^2}\right]. \tag{3.3}$$

The OVL measures considered here are functions of R, therefore, based on our estimate of R, the OVL coefficients can be estimated by

$$\widehat{\rho}_{RSS} = \frac{2\sqrt{\widehat{R}_{RSS}}}{1 + \widehat{R}_{RSS}}, \tag{3.4}$$

$$\widehat{\lambda}_{RSS} = \frac{4\widehat{R}_{RSS}}{(1 + \widehat{R}_{RSS})^2}, \tag{3.5}$$

$$\widehat{\Delta}_{RSS} = 1 - \widehat{R}_{RSS}^{\frac{1}{1-\widehat{R}_{RSS}}}\left|1 - \frac{1}{\widehat{R}_{RSS}}\right|. \tag{3.6}$$

3.2. Asymptotic properties

Let OVL$= g(R)$, then $\widehat{\mathrm{OVL}} = g(\widehat{R}_{RSS})$. Again by using the well-known Delta method (Taylor series expansion) the approximate sampling variance of the OVL measures can be obtained as follows:

$$\mathrm{Var}(\widehat{\rho}_{RSS}) \cong \frac{R(1-R)^2}{(1+R)^4}\left[\frac{\sum_{i=1}^{r_1}\frac{1}{r_1 - i + 1}}{mr_1^2} + \frac{\sum_{i=1}^{r_2}\frac{1}{r_2 - i + 1}}{mr_2^2}\right], \tag{3.7}$$

$$\mathrm{Var}(\widehat{\lambda}_{RSS}) \cong \frac{16R^2(1-R)^2}{(1+R)^6}\left[\frac{\sum_{i=1}^{r_1}\frac{1}{r_1-i+1}}{mr_1^2}+\frac{\sum_{i=1}^{r_2}\frac{1}{r_2-i+1}}{mr_2^2}\right], \tag{3.8}$$

$$\mathrm{Var}(\widehat{\Delta}_{RSS}) \cong \frac{R^{\frac{2}{1-R}}(\ln R)^2}{(1-R)^2}\left[\frac{\sum_{i=1}^{r_1}\frac{1}{r_1-i+1}}{mr_1^2}+\frac{\sum_{i=1}^{r_2}\frac{1}{r_2-i+1}}{mr_2^2}\right]. \tag{3.9}$$

It is known that the estimators of those OVL coefficients are biased. Approximations for the biases of the OVL coefficients estimates, using Taylor series expansion, are as follow:

1. $\mathrm{Bias}(\widehat{\rho}_{RSS}) \cong \frac{\sqrt{R}(3R(R-2)-1)}{2(1+R)^3}\left[\frac{\sum_{i=1}^{r_1}\frac{1}{r_1-i+1}}{mr_1^2}+\frac{\sum_{i=1}^{r_2}\frac{1}{r_2-i+1}}{mr_2^2}\right],$

2. $\mathrm{Bias}(\widehat{\lambda}_{RSS}) \cong \frac{8R^2(R-2)}{(1+R)^4}\left[\frac{\sum_{i=1}^{r_1}\frac{1}{r_1-i+1}}{mr_1^2}+\frac{\sum_{i=1}^{r_2}\frac{1}{r_2-i+1}}{mr_2^2}\right],$

3. $\mathrm{Bias}(\widehat{\Delta}_{RSS}) \cong \begin{cases} H(R)\left[\frac{\sum_{i=1}^{r_1}\frac{1}{r_1-i+1}}{mr_1^2}+\frac{\sum_{i=1}^{r_2}\frac{1}{r_2-i+1}}{mr_2^2}\right], & \text{if } R>1, \\ -H(R)\left[\frac{\sum_{i=1}^{r_1}\frac{1}{r_1-i+1}}{mr_1^2}+\frac{\sum_{i=1}^{r_2}\frac{1}{r_2-i+1}}{mr_2^2}\right], & \text{if } R<1, \end{cases}$

where $H(R) = R^2[R^{(2R-1)/(1-R)}R\{2R-\ln(R)-2\}\ln(R)-(R-1)^2]/(R-1)^3$.

Reasonable estimates for the above variances and the biases can be obtained by substituting R by $\widehat{R}_{RSS}$ in the above formulas.

For variance and bias for the OVL measures under simple random sampling (SRS) see Mulekar *et al.* (2001) or Al-Saleh and Samawi (2006). Now, the asymptotic relative efficiency of OVL measures estimates is defined by

$$Eff(\widehat{\mathrm{OVL}}_{SRS}, \widehat{\mathrm{OVL}}_{RSS}) = \frac{\mathrm{MSE}(\widehat{\mathrm{OVL}}_{SRS})}{\mathrm{MSE}(\widehat{\mathrm{OVL}}_{RSS})},$$

where

$$\mathrm{MSE}(\widehat{\mathrm{OVL}}) = \mathrm{Var}(\widehat{\mathrm{OVL}}) + \mathrm{Bias}(\widehat{\mathrm{OVL}})^2$$

Table 3.1 and 3.2 show the asymptotic relative efficiencies for OVL measures using RSS relative to using SRS.

Table 3.1: Asymptotic relative efficiency of OVL estimates using RSS relative to using SRS (m is 8).

		ρ				λ				Δ			
R	r_1/r_2	2	3	4	5	2	3	4	5	2	3	4	5
0.10	2	1.52	1.57	1.58	1.58	1.49	1.54	1.56	1.56	1.49	1.55	1.56	1.56
	3	1.65	1.79	1.87	1.91	1.62	1.76	1.84	1.88	1.62	1.76	1.85	1.89
	4	1.70	1.91	2.06	2.15	1.67	1.88	2.03	2.12	1.68	1.89	2.03	2.12
	5	1.73	1.93	2.18	2.32	1.70	1.95	2.15	2.29	1.70	1.96	2.15	2.29
0.50	2	1.69	1.74	1.75	1.74	1.67	1.71	1.72	1.71	1.52	1.57	1.59	1.59
	3	1.85	2.00	2.08	2.11	1.82	1.97	2.05	2.08	1.65	1.79	1.87	1.91
	4	1.91	2.14	2.28	2.38	1.88	2.11	2.25	2.34	1.71	1.92	2.06	2.15
	5	1.93	2.21	2.42	2.56	1.90	2.18	2.38	2.52	1.73	1.99	2.18	2.32
1.01	2	2.18	2.35	2.40	2.41	2.18	2.35	2.40	2.41	1.54	1.59	1.60	1.60
	3	2.57	3.05	3.35	3.50	2.57	3.05	3.35	3.50	1.67	1.81	1.89	1.93
	4	2.76	3.50	4.06	4.44	2.76	3.50	4.06	4.44	1.73	1.94	2.08	2.17
	5	2.84	3.76	4.56	5.17	2.84	3.76	4.56	5.17	1.75	2.01	2.20	2.34
1.50	2	1.57	1.62	1.63	1.63	1.56	1.61	1.62	1.62	1.54	1.60	1.61	1.61
	3	1.71	1.85	1.93	1.97	1.70	1.84	1.92	1.96	1.68	1.82	1.90	1.94
	4	1.77	1.98	2.12	2.21	1.75	1.96	2.11	2.20	1.74	1.95	2.09	2.18
	5	1.79	2.05	2.25	2.38	1.77	2.03	2.23	2.37	1.76	2.02	2.22	2.35
2.0	2	1.48	1.53	1.55	1.55	1.48	1.53	1.55	1.55	1.56	1.61	1.62	1.62
	3	1.61	1.75	1.83	1.87	1.61	1.75	1.83	1.87	1.69	1.83	1.91	1.95
	4	1.66	1.87	2.02	2.11	1.66	1.87	2.02	2.11	1.75	1.96	2.10	2.19
	5	1.69	1.94	2.14	2.28	1.69	1.94	2.14	2.28	1.77	2.03	2.23	2.36

Table 3.1 and 3.2 show that, using RSS for estimating all three overlap measure is more efficient that using SRS. The efficiency increases as the set size r_1 and/or r_2 increases. Increasing the number of cycles m slightly decreases the efficiency. This may be to the fact that this relative efficiency is based on large sample approximation. Therefore, the larger is the sample size is the closer is the relative efficiency to the exact one.

3.3. Interval estimation

Normal approximation to the sampling distribution, using Delta-method, work fairly well for large sample. Therefore, the $100(1-\alpha)\%$ confidence intervals for the OVL coefficients can be computed easily as $\{\widehat{\text{OVL}}_{RSS} \pm Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\widehat{\text{OVL}}_{RSS})}\}$, where $Z_{1-\alpha/2}$ is $\alpha/2$ the upper quantile of the standard normal distribution.

These confidence intervals are not the best because of the bias involved in OVL coefficients estimates, however, for large samples they work fairly well. In previous Sections we approximate the bias of those OVL coefficients. Using these approximations, the bias corrected interval can be computed as $[\{\widehat{\text{OVL}}_{RSS} - \text{Bias}(\widehat{\text{OVL}}_{RSS})\} \pm Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\widehat{\text{OVL}}_{RSS})}]$.

Table 3.2: Asymptotic relative efficiency of OVL estimates using RSS relative to using SRS (m is 40).

		ρ				λ				Δ			
R	r_1/r_2	2	3	4	5	2	3	4	5	2	3	4	5
0.10	2	1.37	1.46	1.50	1.52	1.36	1.46	1.50	1.51	1.36	1.46	1.50	1.51
	3	1.48	1.66	1.77	1.83	1.47	1.66	1.76	1.82	1.47	1.66	1.77	1.82
	4	1.52	1.78	1.95	2.05	1.52	1.77	1.94	2.05	1.52	1.77	1.94	2.05
	5	1.54	1.84	2.06	2.21	1.54	1.83	3.05	3.21	1.54	1.84	2.05	2.21
0.50	2	1.40	1.50	1.54	1.55	1.40	1.49	1.53	1.55	1.37	1.46	1.50	1.52
	3	1.51	1.71	1.81	1.87	1.51	1.70	1.81	1.86	1.48	1.65	1.77	1.83
	4	1.56	1.81	1.99	2.10	1.55	1.80	1.98	2.10	1.52	1.78	1.95	2.06
	5	1.58	1.88	2.11	2.26	1.57	1.88	2.10	2.26	1.54	1.84	2.06	2.22
1.01	2	1.84	2.12	2.24	2.28	1.84	2.12	2.24	2.28	1.37	1.47	1.51	1.52
	3	2.16	2.74	3.09	3.29	2.16	2.74	3.09	3.29	1.48	1.67	1.77	1.83
	4	2.29	3.12	3.73	4.15	2.29	3.12	3.73	4.15	1.53	1.78	1.95	2.06
	5	2.35	3.34	4.17	4.81	2.35	3.34	4.17	4.81	1.54	1.84	2.06	2.22
1.50	2	1.37	1.47	1.51	1.53	1.37	1.47	1.51	1.53	1.37	1.47	1.51	1.52
	3	1.49	1.68	1.78	1.84	1.48	1.67	1.78	1.84	1.48	1.67	1.78	1.83
	4	1.53	1.79	1.96	2.07	1.53	1.79	1.96	2.06	1.53	1.78	1.95	2.06
	5	1.55	1.85	2.07	2.23	1.55	1.85	2.07	2.22	1.55	1.85	2.07	2.22
2.0	2	1.36	1.46	1.50	1.51	1.36	1.46	1.50	1.51	1.37	1.47	1.51	1.52
	3	1.47	1.66	1.76	1.82	1.47	1.66	1.76	1.82	1.48	1.67	1.78	1.84
	4	1.52	1.77	1.94	2.05	1.52	1.77	1.94	2.05	1.53	1.79	1.95	2.06
	5	1.53	1.83	2.05	2.21	1.53	1.83	2.05	2.21	1.55	1.85	2.07	2.22

3.4. Bootstrap inference

The uniform (ordinary) bootstrap resampling by Efron (1979) is based on resampling with replacement from the observed sample according to a rule which places equal probabilities on sample values. For two-sample case the uniform resampling rules will apply to each sample separately and independently (see, Ibrahim, 1991; Samawi *et al.*, 1996b; Samawi *et al.*, 1998). In case of RSS, we will adopt the stratified bootstrap algorithm (see, Hui *et al.*, 2005), where each stratum contains only one type of order statistics. The method is described as follows:

1. For the first sample, divide the sample into r_1 mutually exclusive strata each contains m *i.i.d* observations (one type of order statistics).

2. Independently from each stratum generate a resample with replacement of size m by placing a mass probability $(1/m)$ on each original observation in that stratum.

3. Combine all r_1 resamples to form a RSS resample of size $n_1 = mr_1$.

4. For the second sample, divide the sample into r_2 mutually exclusive strata each contains m *i.i.d* observations (one type of order statistics).

5. Again independently from each stratum generate a resample with replacement of size m by placing a mass probability $(1/m)$ on each original observations in that stratum.

6. Using the resamples from Step 1 to 5, compute the OVL measures

7. Repeat Step 1–6 B times to have B different resampled OVL measures (say, $\text{OVl}_1, \ldots, \text{OVl}_B$)

8. Ranked them from the smallest to the largest $(\text{OVl}_{(1)}, \ldots, \text{OVl}_{(B)})$

9. An approximate 95% bootstrap confidence interval is $(\text{OVl}_{(k)}, \text{OVl}_{(B-k)})$, where $k = \text{Int}[0.025(B+1)]$

4. Simulation study

In our simulation study we include the following parameters: $R = 0.2, 0.5, 0.8$; $r_1 = 2, 3$; $r_2 = 2, 3$; $m = 10, 40$ and $\alpha = 0.05$. For each combinations of R, r_1, r_2 and m, α, 1,000 simulated sets of SRS and RSS observations were generated under the assumption that both densities have exponential distribution with different means. All three OVL measures were computed from the SRS simulated samples and from the RSS simulated Samples. Then the associated approximate 95% confidence intervals bias were computed for the SRS and RSS samples using Taylor and bootstrap approximation. The bootstrap approximation was based on $B = 1000$ resamples.

Tables 4.1–4.3 indicate that the bias of the proposed OVL estimators is negligible and |bias| decreases as the sample sizes are increased for both SRS and RSS. However, the asymptotic bias when using RSS is smaller than when using SRS. The bootstrap bias using SRS is smaller than when using RSS. With respect to the coverage probability $(1-\alpha)$, Taylor series approximation method seem to work well when SRS is used except for R close to one and very small sample sizes. The coverage probabilities for all three OVL coefficients are getting closer to the nominal value when the sample sizes are increased for both SRS and RSS. Bootstrap methods coverage probability work fairly good in all cases.

In conclusion, it seems that there is no best method in all situations. If computers are available, bootstrap method can be used. Taylor series approximation is recommended for larger sample sizes and $R < 0.8$.

Table 4.1: Absolute value of Bias = |Bias|, Length of the Interval (L.) and the Coverage Probability (Cov.), when R is 0.20 and m is 10 or 40, using RSS. Results for SRS in (**Bold**).

		Taylor App.			Bootstrap		
(n_1, n_2)		|Bias|	L.	Cov.	|Bias|	L.	Cov.
(20, 20)	ρ	0.020(**0.029**)	0.26(**0.31**)	0.89(**0.93**)	0.013(**0.009**)	0.38(**0.28**)	0.98(**0.93**)
$r_1 = 2$	λ	0.022(**0.031**)	0.38(**0.46**)	0.89(**0.93**)	0.012(**0.009**)	0.55(**0.41**)	0.98(**0.93**)
$r_2 = 2$	Δ	0.013(**0.019**)	0.29(**0.35**)	0.90(**0.94**)	0.016(**0.011**)	0.45(**0.32**)	0.98(**0.93**)
(20, 30)	ρ	0.015(**0.023**)	0.23(**0.28**)	0.89(**0.93**)	0.009(**0.003**)	0.35(**0.26**)	0.97(**0.93**)
$r_1 = 2$	λ	0.015(**0.025**)	0.33(**0.42**)	0.89(**0.94**)	0.019(**0.002**)	0.49(**0.38**)	0.97(**0.93**)
$r_2 = 3$	Δ	0.009(**0.015**)	0.25(**0.31**)	0.90(**0.94**)	0.010(**0.002**)	0.38(**0.29**)	0.97(**0.93**)
(30, 30)	ρ	0.011(**0.019**)	0.19(**0.25**)	0.91(**0.94**)	0.009(**0.007**)	0.30(**0.24**)	0.97(**0.94**)
$r_1 = 3$	λ	0.011(**0.020**)	0.29(**0.38**)	0.91(**0.94**)	0.009(**0.001**)	0.44(**0.35**)	0.97(**0.94**)
$r_2 = 3$	Δ	0.007(**0.012**)	0.21(**0.28**)	0.93(**0.94**)	0.010(**0.004**)	0.34(**0.26**)	0.97(**0.94**)
(80, 80)	ρ	0.005(**0.007**)	0.13(**0.15**)	0.91(**0.94**)	0.009(**0.007**)	0.21(**0.15**)	0.97(**0.94**)
$r_1 = 2$	λ	0.005(**0.007**)	0.20(**0.23**)	0.91(**0.94**)	0.002(**0.005**)	0.31(**0.22**)	0.97(**0.94**)
$r_2 = 2$	Δ	0.003(**0.004**)	0.14(**0.17**)	0.93(**0.94**)	0.003(**0.005**)	0.23(**0.17**)	0.97(**0.94**)
(120, 120)	ρ	0.003(**0.004**)	0.10(**0.13**)	0.93(**0.93**)	0.005(**0.003**)	0.16(**0.12**)	0.97(**0.93**)
$r_1 = 3$	λ	0.003(**0.005**)	0.15(**0.19**)	0.93(**0.94**)	0.002(**0.002**)	0.24(**0.18**)	0.97(**0.93**)
$r_2 = 3$	Δ	0.002(**0.003**)	0.11(**0.14**)	0.94(**0.95**)	0.002(**0.002**)	0.17(**0.13**)	0.97(**0.93**)

Table 4.2: Absolute value of Bias = |Bais|, Length of the Interval (L.) and the Coverage Probability (Cov.), when R is 0.50 and m is 10 or 40 using RSS. Results for SRS in (**Bold**).

		Taylor App.			Bootstrap		
(n_1, n_2)		|Bias|	L.	Cov.	|Bias|	L.	Cov.
(20, 20)	ρ	0.024(**0.035**)	0.16(**0.19**)	0.88(**0.90**)	0.011(**0.011**)	0.24(**0.17**)	0.98(**0.93**)
$r_1 = 2$	λ	0.042(**0.060**)	0.30(**0.35**)	0.88(**0.90**)	0.043(**0.031**)	0.42(**0.31**)	0.98(**0.93**)
$r_2 = 2$	Δ	0.021(**0.029**)	0.37(**0.44**)	0.91(**0.94**)	0.013(**0.006**)	0.48(**0.37**)	0.98(**0.93**)
(20, 30)	ρ	0.019(**0.029**)	0.15(**0.18**)	0.87(**0.93**)	0.011(**0.011**)	0.23(**0.16**)	0.97(**0.94**)
$r_1 = 2$	λ	0.032(**0.049**)	0.28(**0.33**)	0.87(**0.93**)	0.055(**0.027**)	0.41(**0.29**)	0.97(**0.94**)
$r_2 = 3$	Δ	0.016(**0.025**)	0.32(**0.40**)	0.92(**0.95**)	0.031(**0.006**)	0.45(**0.35**)	0.97(**0.94**)
(30, 30)	ρ	0.014(**0.023**)	0.12(**0.16**)	0.93(**0.95**)	0.021(**0.021**)	0.18(**0.14**)	0.97(**0.94**)
$r_1 = 3$	λ	0.024(**0.040**)	0.22(**0.29**)	0.93(**0.95**)	0.019(**0.017**)	0.33(**0.26**)	0.97(**0.94**)
$r_2 = 3$	Δ	0.011(**0.020**)	0.27(**0.36**)	0.93(**0.95**)	0.003(**0.002**)	0.40(**0.32**)	0.97(**0.94**)
(80, 80)	ρ	0.006(**0.009**)	0.08(**0.10**)	0.92(**0.93**)	0.011(**0.011**)	0.13(**0.09**)	0.97(**0.93**)
$r_1 = 2$	λ	0.011(**0.015**)	0.16(**0.18**)	0.92(**0.93**)	0.009(**0.008**)	0.24(**0.17**)	0.97(**0.93**)
$r_2 = 2$	Δ	0.005(**0.007**)	0.19(**0.22**)	0.93(**0.94**)	0.004(**0.001**)	0.29(**0.21**)	0.97(**0.93**)
(120, 120)	ρ	0.003(**0.006**)	0.06(**0.08**)	0.93(**0.94**)	0.020(**0.020**)	0.10(**0.08**)	0.97(**0.94**)
$r_1 = 3$	λ	0.006(**0.010**)	0.12(**0.15**)	0.93(**0.94**)	0.006(**0.007**)	0.18(**0.15**)	0.97(**0.94**)
$r_2 = 3$	Δ	0.003(**0.005**)	0.14(**0.18**)	0.94(**0.94**)	0.001(**0.001**)	0.22(**0.17**)	0.97(**0.94**)

Table 4.3: Absolute value of Bias = |Bais|, Length of the Interval (L.) and the Coverage Probability (Cov.), when R is 0.80 and m is 10 or 40, using RSS. Results for SRS in (**Bold**).

		Taylor App.			Bootstrap		
(n_1, n_2)		|Bias|	L.	Cov.	|Bias|	L.	Cov.
(20, 20)	ρ	0.021(**0.030**)	0.08(**0.10**)	0.75(**0.67**)	0.010(**0.010**)	0.09(**0.10**)	0.98(**0.96**)
$r_1 = 2$	λ	0.040(**0.058**)	0.15(**0.18**)	0.75(**0.67**)	0.061(**0.042**)	0.29(**0.19**)	0.98(**0.96**)
$r_2 = 2$	Δ	0.011(**0.017**)	0.34(**0.38**)	0.93(**0.95**)	0.085(**0.060**)	0.41(**0.31**)	0.98(**0.96**)
(20, 30)	ρ	0.017(**0.025**)	0.08(**0.08**)	0.80(**0.70**)	0.010(**0.010**)	0.14(**0.09**)	0.98(**0.96**)
$r_1 = 2$	λ	0.032(**0.048**)	0.14(**0.16**)	0.80(**0.71**)	0.060(**0.040**)	0.27(**0.17**)	0.98(**0.96**)
$r_2 = 3$	Δ	0.013(**0.017**)	0.31(**0.35**)	0.91(**0.95**)	0.083(**0.050**)	0.39(**0.29**)	0.98(**0.96**)
(30, 30)	ρ	0.012(**0.020**)	0.05(**0.07**)	0.88(**0.75**)	0.020(**0.020**)	0.10(**0.07**)	0.98(**0.96**)
$r_1 = 3$	λ	0.023(**0.038**)	0.09(**0.13**)	0.88(**0.75**)	0.034(**0.029**)	0.19(**0.14**)	0.98(**0.96**)
$r_2 = 3$	Δ	0.008(**0.014**)	0.25(**0.31**)	0.95(**0.95**)	0.049(**0.042**)	0.32(**0.26**)	0.98(**0.96**)
(80, 80)	ρ	0.006(**0.008**)	0.03(**0.03**)	0.93(**0.98**)	0.010(**0.010**)	0.06(**0.04**)	0.98(**0.96**)
$r_1 = 2$	λ	0.011(**0.015**)	0.06(**0.07**)	0.93(**0.98**)	0.017(**0.010**)	0.11(**0.07**)	0.98(**0.96**)
$r_2 = 2$	Δ	0.005(**0.007**)	0.19(**0.21**)	0.94(**0.96**)	0.023(**0.012**)	0.24(**0.18**)	0.98(**0.96**)
(120, 120)	ρ	0.003(**0.005**)	0.02(**0.03**)	0.93(**0.95**)	0.009(**0.009**)	0.04(**0.03**)	0.98(**0.96**)
$r_1 = 3$	λ	0.006(**0.010**)	0.04(**0.06**)	0.93(**0.95**)	0.010(**0.008**)	0.08(**0.06**)	0.98(**0.96**)
$r_2 = 3$	Δ	0.003(**0.005**)	0.14(**0.18**)	0.94(**0.97**)	0.010(**0.008**)	0.19(**0.16**)	0.98(**0.96**)

References

Al-Saleh, M. F. and Al-Kadiri, M. A. (2000). Double-ranked set sampling. *Statistics & Probability Letters*, **48**, 205–212.

Al-Saleh, M. F. and Al-Omari, A. I. (2002). Multistage ranked set sampling. *Journal of Statistical Planning and Inference*, **102**, 273–286.

Al-Saleh, M. F. and Samawi, H. M. (2006). Inference on overlapping coefficients in two exponential populations. Submitted for publication.

Al-Saidy, O., Samawi H. M. and Al-Saleh, M. F. (2005). Inference on overlap coefficients under the weibull distribution: Equal shape parameter. *ESAIM: Probability and Statistics*, **9**, 206–219.

Bradley, E. L. and Piantadosi, S. (1982). The overlapping coefficient as a measure of agreement between distributions. *Technical Report*, Department of Biostatistics and Biomathematics, University of Alabama at Birmingham, Birmingham, AL.

Clemons, T. E. (1996). The overlapping coefficient for two normal probability functions with unequal variances. Unpublished Thesis, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL.

Clemons, T. E. and Bradley, E. L. (2000). A nonparametric measure of the overlapping coefficient. *Computational Statistics & Data Analysis*, **34**, 51–61.

Dixon, P. M. (1993). The bootstrap and the jackknife: Describing the precision of ecological indices. *Design and Analysis of Ecological Experiments* (Scheiner, S. M. and Gurevitch J., eds.), 290–318, Chapman & Hall, New York.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7**, 1–26.

Federer, W. T., Powers, L. R. and Payne, M. G. (1963). Studies on statistical procedures applied to chemical genetic data from sugar beets. Technical Bulletin, **77**, Agricultural Experimentation Station, Colorado State University.

Harner, E. J. and Whitmore, R. C. (1977). Multivariate measures of niche overlap using discriminant analysis. *Theoretical Population Biology*, **12**, 21–36.

Hui, T. P., Modarres, R. and Zheng, G. (2005). Bootstrap confidence interval estimation of mean via ranked set sampling linear regression. *Journal of Statistical Computation and Simulation*, **75**, 543–553.

Ibrahim, H. I. (1991). Evaluating the power of the Mann-Whitney test using the bootstrap method. *Communications in Statistics-Theory and Methods*, **20**, 2919–2931.

Ichikawa, M. (1993). A meaning of the overlapped area under probability density curves of stress and strength. *Reliability Engineering & System Safety*, **41**, 203–204.

Inman, H. F. and Bradley, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics-Theory and Methods*, **18**, 3851–3874.

Kaur, A., Patil, G. P., Sinha, A. K. and Tailie, C. (1995). Ranked set sampling: An annotated bibliography. *Environmental and Ecological Statistics*, **2**, 25–54.

Lu, R. P., Smith, E. P. and Good, I. J. (1989). Multivariate measures of similarity and niche overlap. *Theoretical Population Biology*, **35**, 1–21.

MacArthur, R. H. (1972). *Geographical Ecology.* Harper and Row, New York.

Mann, N. R., Schafer, R. E. and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data.* John Wiley & Sons, New York.

Matusita, K. (1955). Decision rules, based on the distance for problems of fir, two samples and estimation. *The Annals of Mathematical Statistics*, **26**, 631–640.

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

Mishra, S. N., Shah, A. K. and Lefante, J. J. Jr. (1986). Overlapping coefficient: The generalized t approach. *Communications in Statistics-Theory and Methods*, **15**, 123–128.

Morisita, M. (1959). Measuring interspecific association and similarity between communities. *Memoirs of the Faculty of Kyushu University*, Ser. E, Biology, **3**, 36–80.

Mulekar, M. S. and Mishra, S. N. (1994). Overlap Coefficient of two normal densities: Equal means case. *Journal of Japan Statistical Society*, **24**, 169–180.

Mulekar, M. S. and Mishra, S. N. (2000). Confidence interval estimation of overlap: Equal means case. *Computational Statistics and Data Analysis*, **34**, 121–137.

Mulekar, M. S., Gonzales, S. and Aryal, S. (2001). Estimation and inference for the overlap of two exponential distributions. In *Proceedings of American Statistical Association,* Joint Statistical Meetings.

Muttlak, H. A. (1997). Median ranked set sampling. *Journal of Applied Statistical Science*, **6**, 245–255.

Patil, G. P., Sinha, A. K. and Taillie, C. (1999). Ranked set sampling: A bibliography. *Environmental Ecological Statistics*, **6**, 91–98.

Reiser, B. and Faraggi, D. (1999). Confidence intervals for the overlapping coefficient: The normal equal variance case. *The Statistician*, **48**, 3, 413–418.

Samawi, H. M. (2002). On double extreme rank set sample with application to regression estimator. *Metron-International Journal of Statistics*, **LX**,1–2.

Samawi, H. M. Ahmed, M. S. and Abu Dayyeh, W. (1996a). Estimating the population mean using extreme ranked set sampling. *Biometrical Journal*, **38**, 577–586.

Samawi, H. M. and Al-Sageer, O. A. M. (2001). On the estimation of the distribution function using extreme and median ranked set sampling. *Biometrical Journal*, **43**, 357–373.

Samawi, H. M. and Muttlak, H. A. (1996). Estimation of ratio using ranked set sampling. *Biometrical Journal*, **38**, 753–764.

Samawi, H. M. and Muttlak, H. A. (2001). On ratio estimation using median ranked set sampling. *Journal of Applied Statistical Science*, **10**, 89–98.

Samawi H. M., Woodworth G. G and Al-Saleh M. F. (1996b). Two-sample importance resampling for the bootstrap. *Metron-International Journal of Statistics*, **LIV**, 3–4.

Samawi H. M., Woodworth G. G. and Lemke, J. H. (1998). Power estimation for two-sample tests using importance and antithetic resampling. *Biometrical Journal*, **40**, 341–354.

Slobodchikoff, C. N. and Schulz, W. C. (1980). Measures of niche overlap. *Ecology*, **61**, 1051–1055.

Smith, E. P. (1982). Niche breadth, resource availability, and inference. *Ecology*, **63**, 1675–1681.

Sneath, P. H. A. (1977). A method for testing the distinctness of clusters: a test of the disjunction of two clusters in euclidean space as measured by their overlap. *Mathematical Geology*, **9**, 123–143.

Weitzman, M. S. (1970). Measures of overlap of income distributions of white and Negro families in the United States. *Technical Paper*, **22**, Departement of Commerce, Bureau of Census, Washington. U.S.

[Received April 2007, Accepted July 2007]