Journal of Modern Applied Statistical Methods

Volume 6 | Issue 2 Article 15

11-1-2007

Interference on Overlapping Coefficients in Two Exponential Populations

Mohammad Fraiwan Al-Saleh
Yarmouk University, Jordan, m-saleh@yu.edu.jo

Hani M. Samawi
Georgia Southern University

Journal of Modern Applied Statistical Methods
November, 2007, Vol. 6, No. 2, 503- 516 1538 – 9472/05/$95.00

Inference on Overlapping Coefficients in Two Exponential Populations

Mohammad F. Al-Saleh
Yarmouk University

Hani M. Samawi
Georgia Southern University

Three measures of overlap, namely Matusita's measure ρ, Morisita's measure λ and Weitzman's measure Δ are investigated in this article for two exponential populations with different means. It is well that the estimators of those measures of overlap are biased. The bias is of these estimators depends on the unknown overlap parameters. There are no closed-form, exact formulas, for those estimators variances or their exact sampling distributions. Monte Carlo evaluations are used to study the bias and precision of the proposed overlap measures. Bootstrap method and Taylor series approximation are used to construct confidence intervals for the overlap measures.

Key words: Bootstrap method; Matusita's measure; Morisita's measure; overlap coefficients; Taylor expansion; Weitzman's measure.

Introduction

Overlap measure are commonly used in reliability analysis to estimate the proportion of machines or electronic devices that have similar range of failure time. The machines may come from two different sources or may be under different stress, which implies different probability densities of failure time. This proportion can be measured by the overlap coefficients of the two densities.

There are three overlap coefficients (OVL), (Matusita's measure ρ, Morisita's measure λ and Weitzman's measure Δ). However, the most commonly used overlap coefficient is the Weitzman's measure Δ. OVL measure is defined to be the area intersected by the graphs of two probability density functions. It measures the similarity, the agreement or the closeness of the two probability distributions.

Mohammad Fraiwan Al-Saleh is Associate Professor, College of Science, Department of Statistics, Yarmouk University, Irbid-Jordan. Email: m-saleh@yu.edu.jo. Hani M. Samawi is Associate Professor of Biostatistics at Georgia Southern University. Email: hsamawi@georgiasouthern.edu.

The OVL measure Δ was originally introduced by Weitzman (1970). Recently, many authors considered this measure, see Bradley and Piantadosi (1982), Inman and Bradley(1989), Clemons (1996), Reiser and Faraggi (1999), Clemons and Bradley (2000) and Mulekar and Mishra (2000).

For other applications of Δ, see Ichikawa (1993) (for the probability of failure in the stress-strength models of reliability analysis), Fedeer 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 applications in ecology and other fields, see Mulekar and Mishra (1994 and 2000). Inman and Bradley (1989) summarized the history of such procedures.

Let $f_1(x)$ and $f_2(x)$ be two probability density functions. Assume samples of observations are drawn from continuous distributions (Slobdchikoff and Schulz, 1980; Harner and Whiytmorte, 1997; MacArthur, 1972). The overlap measures are defined as follows:

Matusita's Measure (1955):

$$\rho = \int \sqrt{f_1(x) f_2(x)}\, dx,$$

Morisita's Measure (1959):

$$\lambda = \frac{2\int f_1(x) f_2(x)\, dx}{\int [f_1(x)]^2\, dx + \int [f_2(x)]^2\, dx} \ ,$$

and

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.

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 the Δ estimator, 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) described 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 Δ, 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 Δ was examined by Clemons and Bradley (2000), using Monte Carlo and bootstrap techniques. Finally, AL-Saidy et al. (2005) consider the problem of drawing inference about the three overlap measures under the Weibul distribution function with equal shape parameter.

Although, the exponential distribution is a special case of the Weibul distribution, this article considers the three proposed measures of overlap (ρ, λ and Δ) for two exponential distributions with different means. This special case provides some neat and closed form results. Exponential distributions are primarily used in reliability applications. They are used to model data with a constant failure rate (indicated by the hazard plot which is simply equal to a constant). Exponential distributions are the most commonly used life distribution models (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:

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

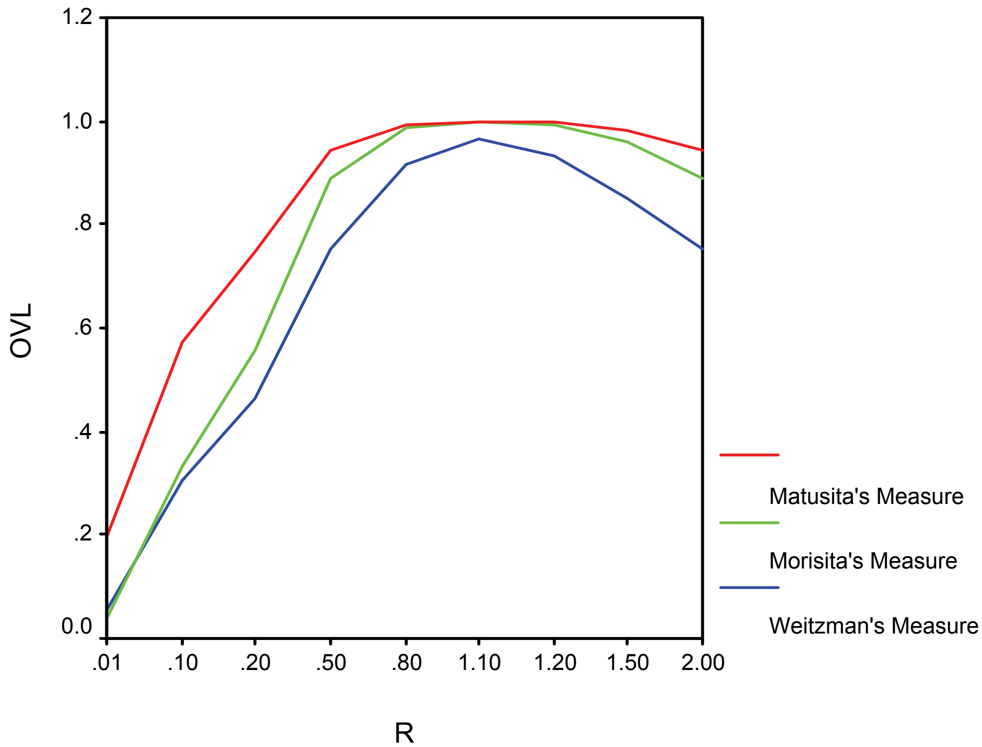
and

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

respectively, where $\theta > 0$.

Overlap measures (OVL) for Exponential Distribution

Suppose $f_1(x)$ and $f_2(x)$ represent the exponential densities with means θ_1 and θ_2 respectively. Letting $R = \frac{\theta_1}{\theta_2}$, 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 is straightforward and it is omitted from the content of this article):

$$\rho = \frac{2\sqrt{R}}{1+R}, \qquad (2.1)$$

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

and

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

Figure 1 shows curves of the three overlap measures. 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. OVL(*R*)=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.

Statistical Inference

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, one first needs to get estimates of θ_1 and θ_2. Suppose that $(X_{11}, X_{12}, ..., X_{1n_1})$ and $(X_{21}, X_{22}, ..., X_{2n_1})$ are two independent random samples drawn from $f_1(x)$ and $f_2(x)$ respectively, where

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

and

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

The maximum likelihood estimators (MLEs) based on the two samples are given by:

1) From the first sample:

$$\hat{\theta}_1 = \overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_{1i}}{n_1}. \tag{3.1}$$

2) From the second sample

$$\hat{\theta}_2 = \overline{X}_2 = \frac{\sum_{i=1}^{n_2} X_{2i}}{n_2}. \tag{3.2}$$

Note that, it is easy to show that $\hat{\theta}_1 \sim \mathrm{G}(n_1, \frac{\theta_1}{n_1})$ and $\hat{\theta}_2 \sim \mathrm{G}(n_2, \frac{\theta_2}{n_2})$, where G(., .) stands for the gamma distribution function. Hence, the variances of those MLE's are respectively $Var(\hat{\theta}_1) = \frac{\theta_1^2}{n_1}$ and $\mathrm{Var}(\hat{\theta}_2) = \frac{\theta_2^2}{n_2}$. Also, the MLE of R is $\hat{R} = \frac{\hat{\theta}_1}{\hat{\theta}_2}$. Therefore, using the relationship between Gamma distribution and Chi-square distribution and the fact that the two samples are independent, it is easy to show that $\frac{\theta_2}{\theta_1}\hat{R}$ has F-distribution{i.e., $F(2n_1, 2n_2)$ }. Hence, the variance of $\hat{R}$ is $Var(\hat{R}) = R^2 \frac{n_2^2(n_1+n_2-1)}{n_1(n_2-1)^2(n_2-2)}$. Also, an unbiased estimate of R is given by $\hat{R}^* = \frac{\hat{\theta}_1}{\hat{\theta}_2}\frac{(n_2-1)}{n_2} = \frac{(n_2-1)}{n_2}\hat{R}$ with $Var(\hat{R}^*) = R^2 \frac{(n_1+n_2-1)}{n_1(n_2-2)}$. Clearly, $\hat{R}^*$ has less variance than $\hat{R}$.

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

$$\hat{\rho} = \frac{2\sqrt{\hat{R}^*}}{1+\hat{R}^*}, \tag{3.3}$$

$$\hat{\lambda} = \frac{4\hat{R}^*}{(1+\hat{R}^*)^2}, \tag{3.4}$$

and

$$\hat{\Delta} = 1 - (\hat{R}^*)^{\frac{1}{1-\hat{R}^*}} \left| 1 - \frac{1}{\hat{R}^*} \right|. \tag{3.5}$$

Asymptotic Properties

Let OVL= $g(R)$, then $O\hat{V}L = g(\hat{R}^*)$. Thus using the well-known delta method (Taylor series expansion) the approximate sampling variance of the OVL measures can be obtained as follows:

$$\mathrm{Var}(\hat{\rho}) = \sigma_{\hat{\rho}}^2 \approx \frac{R(1-R)^2(n_1+n_2-1)}{n_1(n_2-2)(1+R)^4}, \tag{3.6}$$

$$\mathrm{Var}(\hat{\lambda}) = \sigma_{\hat{\lambda}}^2 \approx \frac{16R^2(1-R)^2(n_1+n_2-1)}{n_1(n_2-2)(1+R)^6}, \tag{3.7}$$

and

$$\mathrm{Var}(\hat{\Delta}) = \sigma_{\hat{\Delta}}^2 \approx \frac{(n_1+n_2-1)(R)^{\frac{2}{1-R}}(\ln R)^2}{n_1(n_2-2)(1-R)^2}. \tag{3.8}$$

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. $Bias(\hat{\rho}^*) =$
$$\frac{(n_1+n_2-1)\sqrt{R}}{n_1(n_2-2)} \; \frac{3R(R-2)-1}{2(R+1)^3}$$

2. $Bias(\hat{\lambda}^*) =$

$$\frac{(n_1+n_2-1)}{n_1(n_2-2)}\ \frac{8R^2(R-2)}{(R+1)^4}$$

3. $Bias(\hat{\Delta}^*) =$

$$\left\{\begin{array}{ll} \dfrac{(n_1+n_2-1)R^2}{n_1(n_2-2)} & \\ \dfrac{R^{\frac{2R-1}{1-R}}[R(2R-Ln(R)-2)Ln(R)-(R-1)^2]}{(R-1)^3} & \text{if } R>1 \\ \dfrac{(n_1+n_2-1)R^2}{n_1(n_2-2)} & \\ \dfrac{R^{\frac{2R-1}{1-R}}[R(2R-Ln(R)-2)Ln(R)-(R-1)^2]}{(1-R)^3} & \text{if } R<1 \end{array}\right\}.$$

Reasonable estimates for the above variances and the biases can be obtained by substituting R by $\hat{R}^*$ in the above formulas.

Interval estimation

Transformation Technique

From Section 3.1, $\frac{\theta_2}{\theta_1}\hat{R} \sim F(2n_1, 2n_2)$, then $\frac{\theta_2(n_2)}{\theta_1(n_2-1)}\hat{R}^* \sim F(2n_1, 2n_2)$. Let L and U be the lower and upper confidence limits respectively of R, corresponding to the inclusion probability $1-\alpha$. Thus L and U can be determined by solving for R the equation $P(F^{\alpha/2}_{(2n_1,2n_2)} < \frac{\theta_2}{\theta_1}\hat{R} < F^{1-\alpha/2}_{(2n_1,2n_2)}) = 1-\alpha$, where $F^{\alpha/2}_{(2n_1,2n_2)}$ and $F^{1-\alpha/2}_{(2n_1,2n_2)}$ are the lower and the upper $\alpha/2$ quantile of the $F(2n_1, 2n_2)$ distribution respectively. Thus $L = \frac{\hat{R}}{F^{1-\alpha/2}_{(2n_1,2n_2)}}$ and $U = \frac{\hat{R}}{F^{\alpha/2}_{(2n_1,2n_2)}}$. However, the OVL coefficients are not monotone functions of R therefore, the $100(1-\alpha)$%confidence intervals for the OVL coefficients can be obtained using the transformation technique as follows:

1. $$\{Min(\frac{2\sqrt{L}}{(L+1)}, \frac{2\sqrt{U}}{(U+1)}) \le \rho \le Max(\frac{2\sqrt{L}}{(L+1)}, \frac{2\sqrt{U}}{(U+1)})\}$$

2. $$\{Min(\frac{4L}{(L+1)^2}, \frac{4U}{(U+1)^2}) \le \lambda \le Max(\frac{4L}{(L+1)^2}, \frac{4U}{(U+1)^2})\}$$

3. $$\{Min(1-L^{\frac{1}{1-L}}|1-\frac{1}{\mathrm{L}}|, 1-U^{\frac{1}{1-U}}|1-\frac{1}{\mathrm{U}}|) \le \Delta \le Max(1-L^{\frac{1}{1-L}}|1-\frac{1}{\mathrm{L}}|, 1-U^{\frac{1}{1-U}}|1-\frac{1}{\mathrm{U}}|)\}$$

Asymptotic technique

Normal approximation to the sampling distribution, using Delta-method, work fairly well for large sample because of the nice asymptotic properties of the MLE estimates of the exponential distribution. Therefore, the $100(1-\alpha)$% confidence intervals for the OVL coefficients can be computed easily as $\{O\hat{V}L - Z_{1-\alpha/2}\hat{\sigma}_{O\hat{V}L}, O\hat{V}L + Z_{1-\alpha/2}\hat{\sigma}_{O\hat{V}L}\}$, where $Z_{1-\alpha/2}$ is the $\alpha/2$ 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 Section 3.2, approximate the bias of those OVL coefficients. Using these approximations, the bias corrected interval can be computed as

$$\{(O\hat{V}L - Bias(O\hat{V}L)) - Z_{1-\alpha/2}\hat{\sigma}_{O\hat{V}L}, (O\hat{V}L - Bias(O\hat{V}L)) + Z_{1-\alpha/2}\hat{\sigma}_{O\hat{V}L}\}.$$

Bootstrap Interference

Bootstrap methods are computer intensive which involves simulated data sets. 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. Uniform bootstrap resampling as described by Efron(1979) and others is an assumption-free method that can be used for some inferential problems. However, it is designed for complete and continuous set of observations. For two-sample case the uniform resampling rules will apply to each sample separately and independently (see Ibrahim, 1991; Samawi et al., 1996; Samawi et al., 1998).

Suppose $\aleph_1 = (X_{11}, X_{12}, ..., X_{1n_1})$ and $\aleph_2^* = (X_{21}^*, X_{22}^*, ..., X_{2n_2}^*)$ are two independent random samples drawn from $f_1(x)$ and $f_2(x)$ respectively. Assume that the parameter of interest is the OVL coefficient, say θ. Let S be an estimate based on the random samples $\aleph_1$ and $\aleph_2$ i.e., $S=S(\aleph_1, \aleph_2)$. Furthermore, assume S is a smooth function of the samples. Assume that U is a function of S i.e., $U = U(S)$. Write U^* for the same function of the data but in resamples $\aleph_1^* = (X_{11}^*, X_{12}^*, ..., X_{1n_1}^*)$ and $\aleph_2^* = (X_{21}^*, X_{22}^*, ..., X_{2n_2}^*)$ which are drawn from $\aleph_1$ and $\aleph_2$ according to the rules which places probability $\frac{1}{n_1}$ on each sample value of $\aleph_1$ and probability $\frac{1}{n_2}$ on each sample value of $\aleph_2$. Let $u = E(U)$ then the bootstrap estimate (say $\hat{u}$) of u is given by

$$\hat{u} = E(U^* \mid \aleph_1, \aleph_2) \qquad (3.17)$$

This expected value is often not computable.

Uniform Resampling Approximation for Bootstrap Estimate

Assume that the probability of selecting X_{1i} in a resample is

$$P(X_1^* = X_{1i} \mid \aleph_1) = \frac{1}{n_1} \qquad (3.18)$$

and probability of selecting X_{2i} in a resample is

$$P(X_2^* = X_{2i} \mid \aleph_2) = \frac{1}{n_2} \qquad (3.19)$$

Let $\aleph_{11}^*, \aleph_{12}^*, ..., \aleph_{1B}^*$ and $\aleph_{21}^*, \aleph_{22}^*, ..., \aleph_{2B}^*$ denote two independent resamples sets of size B each drawn from $\aleph_1$ and $\aleph_2$ respectively. To obtain a Monte Carlo approximation to $\hat{u}$ using uniform resampling, let U_b^* denote U computed from $\aleph_{1b}^*$ and $\aleph_{2b}^*$. Then, the uniform resampling approximation to the bootstrap estimate $\hat{u}$ is given by

$$\hat{u}_B^* = B^{-1} \sum_{b=1}^{B} \left(U_b^* \right) \qquad (3.20)$$

Do and Hall (1991) showed that $\hat{u}_B^*$ is an unbiased approximation to $\hat{u}$, in the sense that $E(\hat{u}_B^* \mid \aleph_1, \aleph_2) = \hat{u}$. Moreover, an approximation of the bootstrap bias of u can be obtained by $\hat{bias}^* = |\hat{u}_B^* - \hat{u}|$, and an approximation of the bootstrap MSE can be obtained by $M\hat{S}E^* = B^{-1} \sum_{b=1}^{B} \left(U_b^* - \hat{u} \right)^2$.

Estimation of distributions function and quantiles

Bootstrap method for calculating confidence limits, distribution function or a problem in testing hypothesis involves estimation of probabilities of the form

$$p = P(S \le d_p) \qquad (3.21)$$

Using the bootstrap estimation conditioned on the original samples one can estimate p by $\hat{p}$ where

$$\hat{p} = P(S^* \le d_p) \qquad (3.21)$$

and $S^* = S(\aleph_1^*, \aleph_2^*)$. Note that (3.22) can be approximated by using empirical frequencies such as the proportion of B simulated samples for which $S^* \le d_p$. In the literature, the problem is solved by defining a smooth transformation h of S viz., $T=h(S)$ with the property that the distribution of T is approximately normal, see Hall (1992).

Adopting the notation of Section 3.3.1, the definitions of U and U^* for this problem become $U = I(S \le d_p)$ and $U^* = I(S^* \le d_p)$ respectively, where I is the indicator function. Let $S_1^*, S_2^*, ..., S_B^*$ be the resampling realization of S. Then, the uniform resampling approximation to the bootstrap estimate $\hat{p}$ is

$$\hat{p}_B^* = B^{-1} \sum_{b=1}^{B} \left(I(S_b^* \le d_b) \right) \qquad (3.22)$$

To find the uniform resample approximation of the *p-th* quantile of the bootstrap distribution of S, say $\hat{q}_b$, let $S_{(1)}^*, S_{(2)}^*, ..., S_{(B)}^*$ be the order statistics of $S_1^*, S_2^*, ..., S_B^*$. Define K=Int(B*P).
Then uniform resampling approximation of the lower limit is $\hat{q}_p^* = \dfrac{S_{(K)}^* + S_{(K+1)}^*}{2}$.

Simulation Study

A Monte Carlo simulation study was conducted for R=0.2, 0.5 and 0.8, (n_1, n_2)=(20,20), (20,30), (50, 50), (50, 100) and α=0.05. All the 1000 simulated sets of observations were generated under the assumption that both densities have exponential distribution with the different means. A bootstrap approximation, based on 1000 resamples, was used.

Tables 1-3 indicate that the bias of the proposed OVL estimators are negligible and |bias| decreases as the sample sizes are increased. With respect to the coverage probability (1-α), Taylor series approximation method seem to work well, except for R close to one and very small sample sizes.

The coverage probability for all three OVL coefficients are getting closer to the nominal value when the sample sizes are increased. Bootstrap methods coverage probability work fairly good and increases when R increases close to one. However, Transformation method, which is the easiest to be used, works very well when R <0.5 and for small sample sizes. Also, transformation method is the best for all three OVL coefficients, with respect to the length of the confidence interval, except when the sample sizes are (50, 50).

Illustration: Survival Time from Dinse (1982)

In most of medical studies the progress of the patients is often monitored for a limited time after treatment. Dinse (1982) gives data for survival times in weeks for 10 patients with symptomatic lymphocytic non-Hodgkin's lymphoma and 28 asymptomatic patients. The precise survival time is not known for one patient in the symptomatic group and 12 patients in the asymptomatic group. They were alive when the study was terminated. Therefore, those patients were excluded from our illustration. Table 4 contains the survival time of the symptomatic and the asymptomatic group. The aim of this illustration to estimate the percentage of similarity in the range of survival time in the two groups.

Figure 2 and 3 indicate that the data for both groups (symptomatic and asymptomatic) can be accepted as exponential data. The MLE estimates for the scale parameters are respectively $\hat{\theta}_1$=138.22 and $\hat{\theta}_2 = 207.13$.

From Table 5, all three methods gave reasonable point and confidence interval estimates for the proposed OVL coefficients. However, Δ have the lowest asymptotic bias but the largest asymptotic variance. The confidence interval based on Taylor series

Table 1: Bias, length of interval (L.), and the coverage probability (Cov.) for R=0.20. Exact OVL coefficients: ρ=0.745, λ=0.556 and Δ=0.465.

		Taylor Series			Bootstrap			Transformation		
(n_1, n_2)		Bias	L.	Cov.	Bias	L.	Cov	Bias*	L.	Cov.
(20,20)	ρ	-0.028	0.314	0.933	-0.016	0.279	0.936	-0.015	0.296	0.959
	λ	-0.029	0.457	0.928	-0.014	0.404	0.936	-0.017	0.428	0.959
	Δ	-0.018	0.337	0.949	-0.006	0.311	0.936	-0.004	0.330	0.959
(20,30)	ρ	-0.023	0.283	0.949	-0.013	0.259	0.941	0.003	0.270	0.957
	λ	-0.024	0.415	0.947	-0.010	0.376	0.941	-0.011	0.396	0.957
	Δ	-0.015	0.305	0.950	-0.005	0.286	0.941	-0.002	0.303	0.957
(50, 50)	ρ	-0.011	0.196	0.937	-0.005	0.185	0.934	-0.005	0.086	0.036
	λ	-0.014	0.290	0.937	-0.002	0.274	0.934	-0.004	0.109	0.036
	Δ	-0.007	0.212	0.944	-0.000	0.204	0.934	-0.002	0.078	0.036
(50, 100)	ρ	-0.008	0.169	0.945	-0.006	0.162	0.937	-0.004	0.125	0.868
	λ	-0.008	0.250	0.945	-0.005	0.240	0.937	-0.004	0.185	0.868
	Δ	-0.005	0.182	0.949	-0.002	0.177	0.937	-0.003	0.137	0.868

* Estimated bias using Monte Carlo simulation methods

Table 2. Bias, length of interval (L.), and the coverage probability (Cov.) for R=0.50. Exact OVL coefficients: ρ =0.943, λ =0.0.889 and Δ =0.75.

		Taylor Series			Bootstrap			Transformation		
(n_1, n_2)		Bias	L.	Cov.	Bias	L.	Cov	Bias*	L.	Cov.
(20,20)	ρ	-0.035	0.203	0.921	-0.024	0.179	0.944	-0.016	0.186	0.957
	λ	-0.059	0.369	0.919	-0.039	0.316	0.944	-0.027	0.327	0.957
	Δ	-0.029	0.430	0.917	-0.016	0.369	0.944	-0.013	0.365	0.957
(20,30)	ρ	-0.029	0.186	0.915	-0.025	0.171	0.930	-0.017	0.171	0.943
	λ	-0.048	0.339	0.915	-0.042	0.304	0.930	-0.030	0.303	0.943
	Δ	-0.024	0.395	0.926	-0.021	0.318	0.930	0.003	0.347	0.943
(50, 50)	ρ	-0.014	0.125	0.930	-0.010	0.118	0.931	-0.007	0.453	0.034
	λ	-0.024	0.232	0.930	-0.017	0.215	0.931	-0.014	0.583	0.034
	Δ	-0.012	0.271	0.925	-0.005	0.260	0.931	-0.005	0.443	0.034
(50, 100)	ρ	-0.010	0.107	0.951	-0.008	0.104	0.946	-0.005	0.078	0.850
	λ	-0.018	0.200	0.948	-0.014	0.191	0.946	-0.008	0.144	0.850
	Δ	-0.009	0.234	0.942	-0.004	0.228	0.946	0.0009	0.175	0.850

* Estimated bias using Monte Carlo simulation methods

Table 3. Bias, length of interval (L.), and the coverage probability (Cov.) for R=0.80. Exact OVL coefficients: ρ=0.994, λ=0.988 and Δ=0.918.

		Taylor Series			Bootstrap			Transformation		
(n_1, n_2)		Bias	L.	Cov.	Bias	L.	Cov	Bias*	L.	Cov.
(20,20)	ρ	-0.031	0.106	0.712	-0.026	0.108	0.955	-0.016	0.096	0.337
	λ	-0.059	0.204	0.714	-0.049	0.201	0.955	-0.030	0.178	0.337
	Δ	-0.020	0.500	0.953	-0.071	0.320	0.955	-0.046	0.226	0.337
(20,30)	ρ	-0.025	0.087	0.720	-0.022	0.094	0.952	-0.012	0.078	0.333
	λ	-0.048	0.168	0.721	-0.041	0.176	0.952	-0.023	0.145	0.333
	Δ	-0.018	0.539	0.940	-0.059	0.297	0.952	-0.033	0.199	0.333
(50, 50)	ρ	-0.012	0.051	0.894	-0.011	0.053	0.958	-0.006	0.616	0.023
	λ	-0.024	0.10	0.892	-0.020	0.103	0.958	-0.011	0.794	0.023
	Δ	-0.011	0.320	0.881	-0.028	0.220	0.958	-0.015	0.652	0.023
(50, 100)	ρ	-0.009	0.043	0.943	-0.009	0.045	0.945	-0.005	0.030	0.844
	λ	-0.017	0.083	0.943	-0.017	0.088	0.945	-0.009	0.058	0.844
	Δ	-0.008	0.269	0.858	-0.023	0.020	0.945	-0.012	0.135	0.844

* Estimated bias using Monte Carlo simulation methods

Table 4. Survival time of symptomatic and asymptomatic lymphocytic patients by Dinse (1982)

Symptomatic	49	58	75	110	112	132	151	276	281							
Asymptomatic	50	58	96	139	152	159	189	225	239	242	257	262	292	294	301	359

Table 5. Results based on the real data of Dinse (1982)

			Asymptotic Inference		**Transformation Technique**		**Bootstrap Inference based on 1000 resamples**	
			95% confidence Interval limits		95% confidence interval limits		95% confidence interval limits	
Coeff	MLEs (bias)	Asymptotic variance	Lower	Upper	Lower	upper	Lower	Upper
ρ	0.973(-0.063)	0.0018	0.815	1.000	0.860	0.990	0.904	0.999
λ	0.947(-0.117)	0.0070	0.643	1.000	0.740	0.981	0.817	0.999
Δ	0.829(-0.060)	0.0247	0.654	0.883	0.606	0.898	0.675	0.976

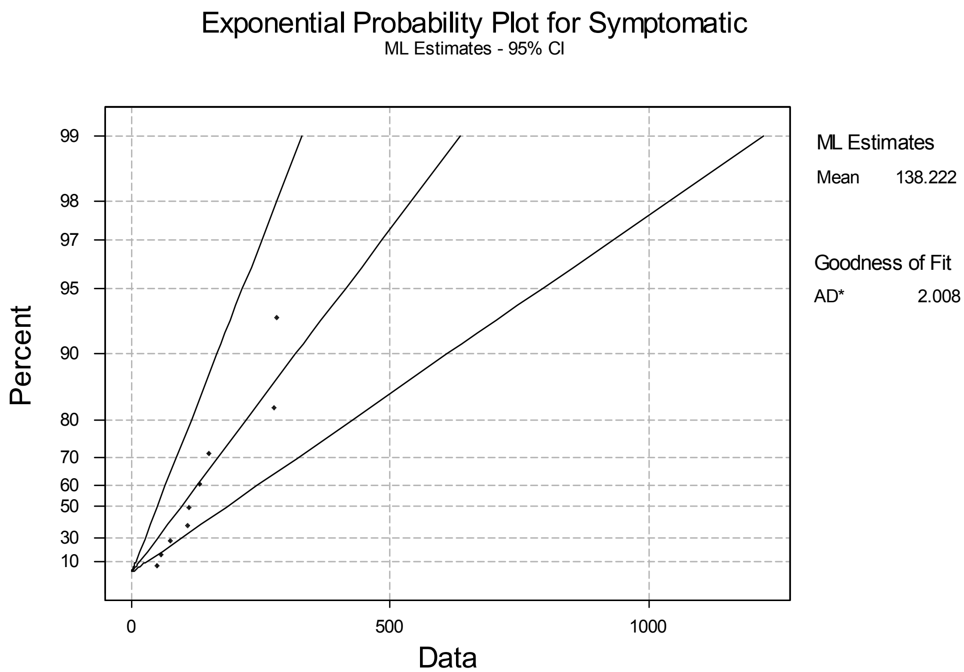


Figure 2. Exponential probability plot for symptomatic patients

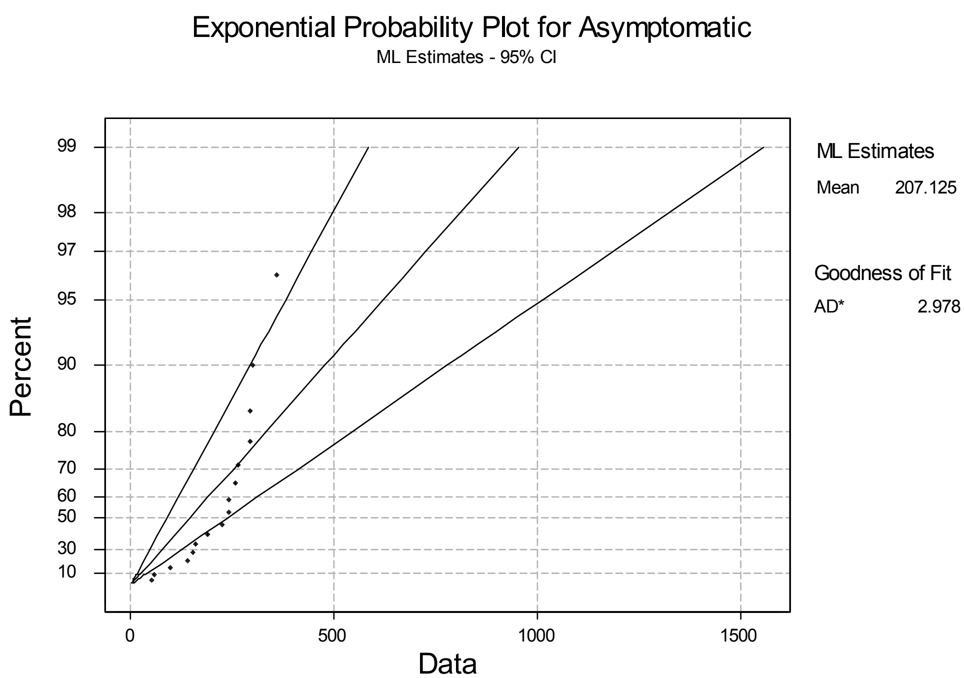


Figure 3: Exponential probability plot for asymptomatic patients.

approximation gave the shortest confidence for Δ.

In conclusion, it seems that there is no best method in all situations. Therefore, when the sample size is small and R<0.5, transformation method is recommended. If computers are available, bootstrap method can be used. Taylor series approximation is recommended for larger sample sizes and R<0.8.

Results

Al-Saidy, O., Samawi, H. M., & Al-Saleh, M. F. (2005). Inference on overlap coefficients under the Weibul distribution: Equal Shape Parameter. *ESAM*: *PS*, *9*, 206-219.

Bradley, E. L., & 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., & Bradley Jr. (2000). A nonparametric measure of the overlapping coefficient. *Comp. Statist. And Data Analysis*, *34*, 51-61.

Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and type of failure Data. *Biometrics*, *38*, 417-431.

Dixon, P. M., (1993). The Bootstrap and the Jackknife: describing the precision of ecological Indices. In: Scheiner, S.M., Gurevitch J. (Eds.), *Design and Analysis of Ecological Experiments.* NY: Chapman & Hall, p. 209-318.

Do, K. N., & Hall, P. (1991). On importance resampling for the bootstrap. *Biometrika*, *78*, 161-167.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Biometrika* 73, 555-566.

Federer, W. T., Powers, L. R., & 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., & Whitmore, R.C. (1977). Multivariate measures of niche overlap using discriminant analysis. *Theoret. Pobulation Biol.*, *12*, 21-36.

Ibrahim, H. I. (1991). Evaluating the power of the Mann-Whitney test using the bootstrap method. *Commun. Statist. Theory Meth.*, *20*, 2919- 2931.

Ichikawa, M. (1993). A meaning of the overlapped area under probability density curves of stress and strength. *Reliab. Eng. System Safety, 41*, 203-204.

Inman, H. F. , & 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. *Comm. Statist. Theory and Methods*, *18*, 3851-3874.

Lu, R., Smith, E. P., & Good, I. J. (1989). Multivariate measures of similarity and niche overlap. *Theoret. Population Ecol.*, *35*, 1-21.

MacArthur, R.H. (1972). *Geographical ecology*. NY: Harper and Row.

Mann, N. R., Schafer, R. E., & Singpurwalla, N. D. (1974), *Methods For Statistical Analysis Of Reliability & Life Data*, NY: John Wiley & Sons.

Matusita, K. (1955). Decision rules based on the distance for problem of fir, two samples, and Estimation. *Ann. Math. Statist*., *26*, 631-640.

Mishra, S. N., Shah, A. K., & Lefante, J. J. (1986). Overlapping coeffecient: the generalized *t* approach. *Commun. Statist.-Theory and Methods*, *15*, 123-128.

Morisita, M. (1959). *Measuring interspecific association and similarity between communities*. [Incomplete reference available]

Memoirs of the faculty of Kyushu University. *Series E. Biology , 3*, 36-80.

Mulekar, M. S., & Mishra, S. N. (1994). Overlap Coefficient of two normal densities: equal means case. *J. Japan Statist. Soc.*, *24*, 169-180.

Mulekar, M. S., & Mishra, S. N. (2000). Confidence interval estimation of overlap: equal means case. Comp. *Statist .and Data Analysis,* 34, 121-137.

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

Samawi, H. M., Woodworth, G. G, & Al-Saleh M. F. (1996). Two-Sample importance resampling for the bootstrap. *Metron*, *LIV*, n. 3-4.

Samawi, H. M. (1998). Power estimation for two-sample tests using importance and antithetic resampling. *Biometrical Journal*. *40*, 3, 341-354.

Slobodchikoff, C. N., & 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. *Math. Geol*., *9*, 123-143.

Weitzman, M. S. (1970). Measures of overlap of income distributions of white and Negro families in the United States. *Technical paper No*. 22, Departement of Commerce, Bureau of Census, Washington, D. C.