

INFERENCE ON OVERLAP COEFFICIENTS UNDER THE WEIBULL DISTRIBUTION: EQUAL SHAPE PARAMETER

OBAID AL-SAIDY¹, HANI M. SAMAWI² AND MOHAMMAD F. AL-SALEH²

Abstract. In this paper we consider three measures of overlap, namely Matusia's measure ρ , Morisita's measure λ and Weitzman's measure Δ . These measures are usually used in quantitative ecology and stress-strength models of reliability analysis. Herein we consider two Weibull distributions having the same shape parameter and different scale parameters. This distribution is known to be the most flexible life distribution model with two parameters. Monte Carlo evaluations are used to study the bias and precision of some estimators of these overlap measures. Confidence intervals for the measures are also constructed via bootstrap methods and Taylor series approximation.

Mathematics Subject Classification. 62F10, 62F40.

Received July 5, 2004.

1. INTRODUCTION

It is essential in reliability analysis to estimate the proportion of machines or electronic devices, from two different sources or under different levels of stress that have similar range of failure time. Overlap coefficients (OVL) (Matusia's measure ρ , Morisita's measure λ and Weitzman's measure Δ) are defined to be the common area of graphing two probability density functions. In the literature, overlap coefficients are mostly used in ecology. However, other applications of are the lowest bound for the probability of failure in the stress-strength models of reliability analysis (Ichikawa [13]), an estimate of the proportion of genetic deviates in segregating populations (Federer *et al.* [9]), and a measure of distinctness of clusters (Sneath [35]). For more applications on OVL coefficients see Mulekar and Mishra [26,27]. Moreover, the history of such procedures is summarized by Inman and Bradley [14].

Let $f_1(x)$ and $f_2(x)$ be two continuous probability density functions. The overlap measures are defined as follows:

$$\text{Matusia's Measure [22]: } \rho = \int \sqrt{f_1(x) f_2(x)} dx.$$

$$\text{Morisita's Measure [25]: } \lambda = \frac{2 \int f_1(x) f_2(x) dx}{\int [f_1(x)]^2 dx + \int [f_2(x)]^2 dx}.$$

$$\text{Weitzman's Measure [39]: } \Delta = \int \min \{f_1(x), f_2(x)\} dx.$$

Keywords and phrases. Bootstrap method, Matusia's measure, Morisita's measure, Overlap coefficients, Taylor expansion, Weitzman's measure.

¹ Department of Mathematics and Statistics, Sultan Qaboos University, PO Box 36, P.C. 123 Al-Khod, Sultanate of Oman; obiad@squ.edu.om

² Department of Statistics, Yarmouk University, Irbid-Jordan 211-63, Jordan; hsamawi@yu.edu.jo; m-saleh@yu.edu.jo

For discrete type of distribution, one can replace the integrals with summations. Also, these forms can be generalized to multivariate distributions. All three overlap measures of two densities are measured on the scale of 0 to 1. An overlap value close to 0 indicates no common area between the two density functions. However, an overlap value of 1 indicates that the common area between the two densities is 1.

The mathematical structure of these measures is complicated; there are no results available on the exact sampling distributions of the commonly used OVL estimators. Researcher such as Smith [34] derived formulas for estimating the mean and the variance of discrete version of Weitzman's measure using the delta method. Mishra *et al.* [24] gave small and large sample properties of the sampling distribution for a function of $\hat{\Delta}$ under the assumption of homogeneity of variances. Lu *et al.* [17] investigated some of these measures using simulation. Mulekar and Mishra [26] derived the approximate expressions for the bias and variance of the OVL estimators. They also used simulation to study the sampling distribution of the overlap measures for normal densities with equal means. Dixon [6] used bootstrap and jackknife techniques for Gini coefficient of size hierarchy and Jaccard index of community similarity. Mulekar and Mishra [27] addressed the problem of making inferences about the overlap coefficients based on normal densities with equal means using jackknife, bootstrap, transformation method and Taylor series approximation. Reiser and Faraggi [30] 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 equal variances, based on the non-central t - and F -distributions. The sampling behavior of a nonparametric estimator of Δ is examined by Clemons and Bradley [4] by using Monte Carlo and bootstrap techniques.

In this paper we consider the three measures of overlap (ρ , λ and Δ) for two Weibull distributions having the same shape parameter and different scale parameters. Section 2 derives the measures. The estimators and their approximate biases and confidence intervals using the delta method and bootstrap techniques are introduced in Section 3. Simulation results are provided with discussion in Section 4. Section 5 presents a real data example with final comments.

The Weibull is a very flexible life distribution model named for Professor Waloddi Weibull of Sweden who suggested it as a distribution for a variety of applications (see Weibull [37, 38]). He considered the problems of yield strength of a Bofors steel, fiber strength of Indian cotton, length of syrtoideas, fatigue life of an St-37 steel, statures of adult male born in British Isles, and breadth of beans of *Phaseolus vulgaris*. The distribution was used as early as 1933 by Rosin and Rammler [31] in describing the "laws governing the fineness of powdered coal". The Weibull distribution began to be seriously considered as a competing model in the 1960's (Bain and Antle [1], Harter and Moore [11], Mann [18], Thoman *et al.* [36]), especially in problems in which the time to failure was the response of interest, see for example Mann [18], Bain and Engelhardt [2] and Murthy *et al.* [28].

A random variable X follows the two-parameter Weibull distribution (denoted by WE2(α, β)) has cdf and pdf given respectively, by:

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

and

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

where $\alpha, \beta > 0$ are respectively the scale and the shape parameters.

2. OVERLAP MEASURES (OVL) FOR WEIBULL DISTRIBUTION

Suppose $f_1(x)$ and $f_2(x)$ represent the Weibull densities with common shape parameter β and scale parameters α_1 and α_2 respectively. Figure 1 shows overlap of WE2(2, 2) and WE2(4, 2). Let $R = \frac{\alpha_1}{\alpha_2}$, then the

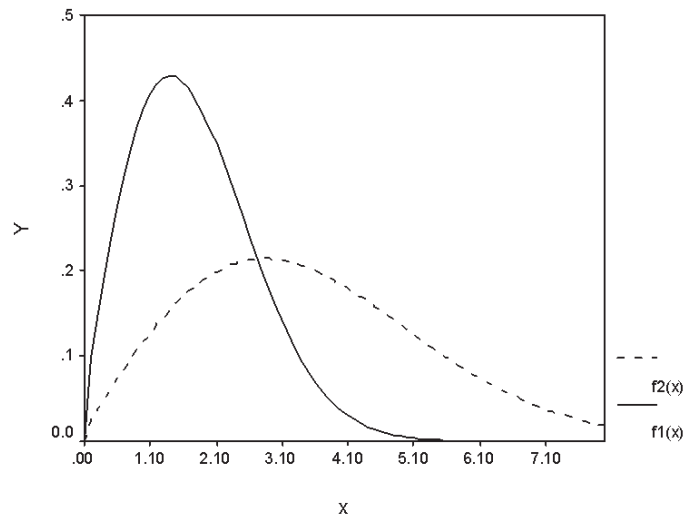


FIGURE 1. The overlap densities of two Weibull pdf $f_1(x)$ and $f_2(x)$.

continuous version of the three overlap measures, can be expressed as a function of R as follows (the derivation of the three overlap measures are straight forward and is omitted from the content of this paper):

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

$$\lambda = \frac{2^{C+1}R^\beta}{(1+R)(1+R^\beta)^C}, \text{ where } C = \frac{2\beta-1}{\beta} \quad (2.2)$$

and

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

Figure 2 shows curves of the three overlap measure (OVL) considered for WE2($\alpha, 2$). All three measures are not monotone for all $R > 0$. Similar to Mulekar and Mishra [27], ρ, λ and Δ have nice properties, such as, symmetry in R , *i.e.* $\text{OVL}(R) = \text{OVL}(1/R)$ and invariance under the linear transformation $Y = aX + b$, $a \neq 0$. These properties are true for these three measures in a general setting.

3. STATISTICAL INFERENCE

The OVL measures ρ, λ , and Δ are functions of β, α_1 and α_2 . In order to draw any inference about the OVL measures, we need first to get estimators of β, α_1 and α_2 .

3.1. Statistical inference for the two parameters Weibull

Early users of the two-parameter Weibull model were mainly engineers (McCool [23] and Weibull [38]); they used graphical methods for estimation of the parameters. Other types of estimation are considered by many authors, such as Lieblein and Zelen [16], White [40], and Mann [19–21] among others.

Maximum likelihood estimation of the parameters has been considered by Hater and Moore [11], Cohen [5] and Leone *et al.* [15], among others. Suppose that $(X_{11}, X_{12}, \dots, X_{1n_1})$ and $(X_{21}, X_{22}, \dots, X_{2n_2})$ are two independent random samples drawn from the two Weibull population $f_1(x)$ and $f_2(x)$ respectively. Assume further that $f_1(x)$ and $f_2(x)$ have the same shape parameter β and different scale parameters α_1 and α_2 .

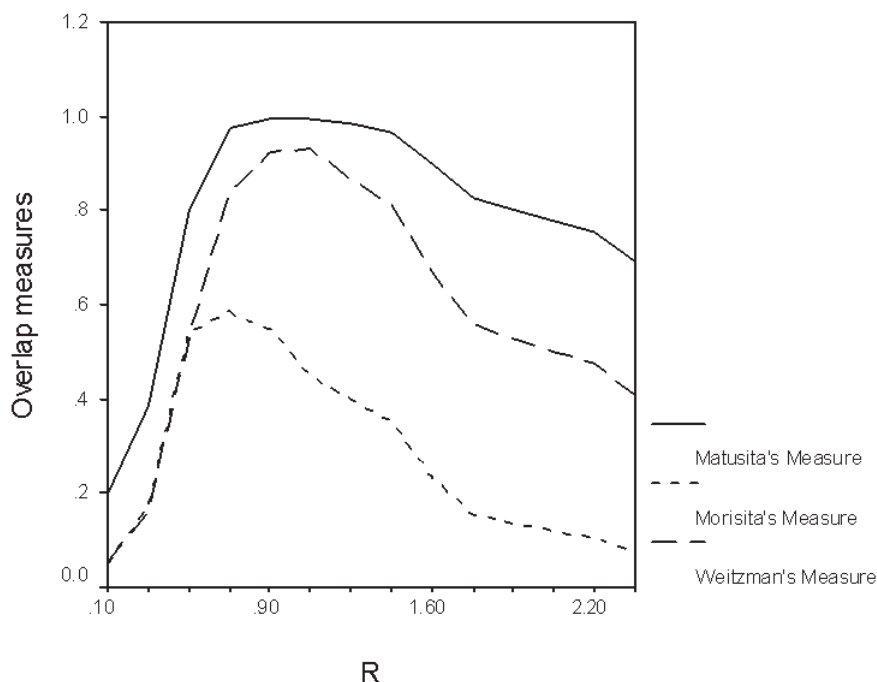


FIGURE 2. The three overlap measures for WE2 ($\alpha, 2$).

Then the maximum likelihood estimators (MLEs) based upon the two samples are given by the solution to the equations:

1) From the first sample

$$\frac{1}{\hat{\beta}} - \frac{\sum_{i=1}^{n_1} X_{1i}^{\hat{\beta}} \text{Ln}(X_{1i})}{\sum_{i=1}^{n_1} X_{1i}^{\hat{\beta}}} + \frac{\sum_{i=1}^{n_1} \text{Ln}(X_{1i})}{n_1} = 0, \tag{3.1}$$

and

$$\hat{\alpha}_1 = \left(\frac{\sum_{i=1}^{n_1} X_{1i}^{\hat{\beta}}}{n_1} \right)^{\frac{1}{\hat{\beta}}}. \tag{3.2}$$

2) From the second sample

$$\frac{1}{\tilde{\beta}} - \frac{\sum_{i=1}^{n_2} X_{2i}^{\tilde{\beta}} \text{Ln}(X_{2i})}{\sum_{i=1}^{n_2} X_{2i}^{\tilde{\beta}}} + \frac{\sum_{i=1}^{n_2} \text{Ln}(X_{2i})}{n_2} = 0, \tag{3.3}$$

and

$$\hat{\alpha}_2 = \left(\frac{\sum_{i=1}^{n_2} X_{2i}^{\tilde{\beta}}}{n_2} \right)^{\frac{1}{\tilde{\beta}}}. \tag{3.4}$$

The common shape parameter can be estimated using the weighted average of the solutions of (3.1) and (3.3) as follows:

$$\beta^* = \frac{n_1}{n_1 + n_2} \hat{\beta} + \frac{n_2}{n_1 + n_2} \tilde{\beta}. \quad (3.5)$$

The existence and uniqueness of the solution to these equations were shown by Pike [29] and McCool [23]. The MLEs are easily found using Newton's method to solve first for $\hat{\beta}$ and $\tilde{\beta}$ and then using these to get $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Moreover, for the two-parameter Weibull distribution, the MLEs are asymptotically efficient and they are asymptotically normally distributed. Therefore,

$$\left\{ \sqrt{n_1} (\hat{\alpha}_1 - \alpha_1), \sqrt{n_1} (\hat{\beta} - \beta) \right\} \sim N_2(0, 0, V_1),$$

and

$$\left\{ \sqrt{n_2} (\hat{\alpha}_2 - \alpha_2), \sqrt{n_2} (\tilde{\beta} - \beta) \right\} \sim N_2(0, 0, V_2)$$

where

$$V_1 = \begin{bmatrix} \frac{1.109\alpha_1^2}{\beta^2} & 0.257\alpha_1 \\ 0.257\alpha_1 & 0.608\beta^2 \end{bmatrix}, \text{ and } V_2 = \begin{bmatrix} \frac{1.109\alpha_2^2}{\beta^2} & 0.257\alpha_2 \\ 0.257\alpha_2 & 0.608\beta^2 \end{bmatrix}.$$

Hence, because the two random samples are independent,

$$\sqrt{n_1 + n_2} (\beta^* - \beta) \sim N(0, 0.608\beta^2).$$

3.2. Statistical inference for OVL measures

The OVL measures considered here are also functions of R and β , therefore, their MLE estimators are given by

$$\hat{\rho} = \frac{2\sqrt{\hat{R}^{\beta^*}}}{1 + \hat{R}^{\beta^*}}, \quad (3.6)$$

$$\hat{\lambda} = \frac{2^{\hat{C}+1} \hat{R}^{\beta^*}}{(1 + \hat{R})(1 + \hat{R}^{\beta^*})^{\hat{C}}}, \text{ where } \hat{C} = \frac{2\beta^* - 1}{\beta^*} \quad (3.7)$$

and

$$\hat{\Delta} = 1 - \left(\hat{R}^{\beta^*} \right)^{\frac{1}{1 - \hat{R}^{\beta^*}}} \left| 1 - \frac{1}{\hat{R}^{\beta^*}} \right|, \hat{R} \neq 1, \quad (3.8)$$

where $\hat{R} = \frac{\hat{\alpha}_1}{\hat{\alpha}_2}$. Let $g(\alpha_1, \alpha_2) = R = \frac{\alpha_1}{\alpha_2}$, then $\hat{R} = \frac{\hat{\alpha}_1}{\hat{\alpha}_2} = g(\hat{\alpha}_1, \hat{\alpha}_2)$. Thus using the well-known delta method (Taylor series expansion), the approximate sampling variance of \hat{R} can be obtained as follows: using Taylor series expansion \hat{R} can be written as

$$\hat{R} = \frac{\hat{\alpha}_1}{\hat{\alpha}_2} = g(\hat{\alpha}_1, \hat{\alpha}_2) = \frac{\alpha_1}{\alpha_2} + [(\hat{\alpha}_1 - \alpha_1), (\hat{\alpha}_2 - \alpha_2)] \left[\begin{array}{c} \frac{1}{\alpha_2} \\ \frac{-\alpha_1}{\alpha_2^2} \end{array} \right] + R_{n_1, n_2}.$$

Since \hat{R} is a function of the MLE's of α_1 and α_2 , then for large n_1 and n_2 , $E(\hat{R}) \approx \frac{\alpha_1}{\alpha_2}$,

$$\text{Var}(\hat{R}) \approx \frac{1.109R^2}{\beta^2} \left(\frac{n_1 + n_2}{n_1 n_2} \right), \quad (3.9)$$

and $(\hat{R} - R) \sim N\left(0, \frac{1.109R^2}{\beta^2} \left(\frac{n_1 + n_2}{n_1 n_2} \right)\right)$.

Theorem 1. Under the assumptions of the MLEs for Weibull distribution, the asymptotic covariance between \hat{R} and β^* {i.e., $Cov(\hat{R}, \beta^*)$ } converges to zero.

Proof.

$$\begin{aligned} Cov(\hat{R}, \beta^*) &= Cov\left(\frac{\hat{\alpha}_1}{\hat{\alpha}_2}, \frac{n_1}{n_1 + n_2}\hat{\beta} + \frac{n_2}{n_1 + n_2}\tilde{\beta}\right) \\ &= Cov\left(\frac{\hat{\alpha}_1}{\hat{\alpha}_2}, \frac{n_1}{n_1 + n_2}\hat{\beta}\right) + Cov\left(\frac{\hat{\alpha}_1}{\hat{\alpha}_2}, \frac{n_2}{n_1 + n_2}\tilde{\beta}\right). \end{aligned} \tag{3.10}$$

Now using the first term of the Taylor series expansion for \hat{R} , $\left(\hat{R} \approx \frac{\alpha_1}{\alpha_2} + \frac{(\hat{\alpha}_1 - \alpha_1)}{\alpha_2} - \frac{(\hat{\alpha}_2 - \alpha_2)\alpha_1}{\alpha_2^2}\right)$ in (3.10) we get $Cov(\hat{R}, \beta^*) \approx \frac{0.257\alpha_1}{\alpha_2(n_1+n_2)} - \frac{\alpha_1}{\alpha_2} \frac{0.257\alpha_2}{\alpha_2(n_1+n_2)} = 0$ and this concludes the proof. \square

Using the above results the approximate sampling variances of the overlap measures estimators, can be obtained using Taylor series expansion as follows:

$$\begin{aligned} Var(\ln \hat{\rho}) &= \sigma_{\ln \hat{\rho}}^2 \approx \frac{(1 - R^\beta)^2 \left[1.109 (n_1 + n_2)^2 + 0.608 (n_1 n_2) (\ln R)^2 \beta^2\right]}{4n_1 n_2 (n_1 + n_2) (1 + R^\beta)^2}, \\ Var(\ln \hat{\lambda}) &= \sigma_{\ln \hat{\lambda}}^2 \approx \frac{1.109 R^2 (n_1 + n_2)^2 g_1^2 + 0.608 (n_1 n_2) \beta^4 g_2^2}{n_1 n_2 (n_1 + n_2) \beta^2}, \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} g_1 &= \frac{\beta (1 - R^{\beta+1}) + (\beta - 1) (R - R^\beta)}{R (1 + R) (1 + R^\beta)} \text{ and} \\ g_2 &= \frac{\beta^2 \ln R + (1 + R^\beta) \ln 2 - \beta (\beta - 1) R^\beta \ln R - (1 + R^\beta) \ln (1 + R^\beta)}{\beta^2 (1 + R^\beta)}, \end{aligned} \tag{3.12}$$

$$Var(\hat{\Delta}) = \sigma_{\hat{\Delta}}^2 \approx \frac{(R^\beta)^{\frac{2}{1-R^\beta}} \beta^2 (\ln R)^2 \left[1.109 (n_1 + n_2)^2 + 0.608 (n_1 n_2) (\ln R)^2 \beta^2\right]}{n_1 n_2 (n_1 + n_2) (1 - R^\beta)^2}. \tag{3.13}$$

The MLEs for the two-parameter Weibull distribution are asymptotically efficient and they are asymptotically normally distributed (see, Thoman *et al.* [36]). However, the OVL measures are functions of the Weibull distribution parameters. Therefore, by using the Delta-method, the OVL measures estimators are asymptotically normally distributed. Thus, the $100(1 - \alpha)\%$ approximate confidence intervals are given by

$$\hat{\rho} e^{\pm Z_{\frac{\alpha}{2}} \hat{\sigma}_{\ln \hat{\rho}}}, \tag{3.14}$$

$$\hat{\lambda} e^{\pm Z_{\frac{\alpha}{2}} \hat{\sigma}_{\ln \hat{\lambda}}}, \tag{3.15}$$

and

$$\hat{\Delta} \pm Z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\Delta}}, \tag{3.16}$$

where $Z_{\frac{\alpha}{2}}$ is $100(\frac{\alpha}{2})\%$ upper quantile of the standard normal distribution.

For large samples these confidence intervals work fairly well. However, for small sample sizes more refined versions of the above confidence intervals can be obtained by

$$\left\{ \left(\text{OVL} - \text{Bias}(\text{OVL}) \right) - Z_{\frac{\alpha}{2}} \hat{\sigma}_{\text{OVL}}, \left(\text{OVL} - \text{Bias}(\text{OVL}) \right) + Z_{\frac{\alpha}{2}} \hat{\sigma}_{\text{OVL}} \right\}$$

or

$$\left\{ \begin{array}{l} \text{Ln} \left(\widehat{\text{OVL}} - \widehat{\text{Bias}}(\text{ln OVL}) \right) - Z_{\frac{\alpha}{2}} \hat{\sigma}_{\widehat{\text{OVL}}}, \\ \text{Ln} \left(\widehat{\text{OVL}} - \widehat{\text{Bias}}(\text{ln OVL}) \right) + Z_{\frac{\alpha}{2}} \hat{\sigma}_{\widehat{\text{OVL}}} \end{array} \right\},$$

where $\widehat{\text{Bias}}(\widehat{\text{OVL}})$ is an estimate of $\text{Bias}(\widehat{\text{OVL}}) = E(\widehat{\text{OVL}}) - \text{OVL}$.

In the case of using the logarithmic transformation to the OVL estimates. The bias for the OVL coefficient estimates can be approximated using Taylor expansion as follows:

$$\begin{aligned} \text{Bias}(\hat{\rho}) &= \frac{\sqrt{R^\beta} (\beta - 2 - 6\beta R^\beta + R^{2\beta} (\beta + 2))}{2(1 + R^\beta)^3} \frac{1.109 (n_1 + n_2)}{(n_1 n_2) \beta} \\ &+ \frac{\sqrt{R^\beta} (1 + R^\beta (R^\beta - 6)) [\text{Ln}(R)]^2}{2(1 + R^\beta)^3} \frac{0.608\beta^2}{(n_1 + n_2)}, \end{aligned} \quad (3.17)$$

$$\text{Bias}(\hat{\lambda}) = \frac{1.109 (n_1 + n_2) R^2}{n_1 n_2 \beta^2} K_1(R, \beta) + \frac{0.608\beta^2}{(n_1 + n_2)} K_2(R, \beta), \quad (3.18)$$

where

$$K_1(R, \beta) = \frac{1}{(1 + R)^3} \left\{ \begin{array}{l} 2^{\frac{3\beta-1}{\beta}} R^{\beta-2} (1 + R^\beta)^{\frac{1-4\beta}{\beta}} \\ \left[\begin{array}{l} 2R^2 - \beta - 4R\beta - 3R^2\beta + \beta^2 + 2R\beta^2 + R^2\beta^2 + R^\beta \\ (R^2 - 1 - 4R + 3(1 + R)^2\beta - 4(1 + R)^2\beta^2) \\ + R^{2\beta} (\beta^2 (1 + R)^2 + \beta (R^2 - 1) - 2R) \end{array} \right] \end{array} \right\}$$

and

$$K_2(R, \beta) = \frac{1}{(1 + R)\beta^4} \left\{ \begin{array}{l} 2^{\frac{3\beta-1}{\beta}} R^\beta (1 + R^\beta)^{\frac{1-4\beta}{\beta}} \\ \left[\begin{array}{l} (\text{Ln}(2))^2 - 2R^\beta (2\beta - \text{Ln}(2)) \text{Ln}(2) \\ + R^{2\beta} \text{Ln}(2) (\text{Ln}(2) - 2\beta) - \beta \text{Ln}(4) + \beta^2 \\ (R^{2\beta} (\beta - 1)^2 + \beta^2 - R^\beta \beta (4\beta - 3)) (\text{Ln}(R))^2 \\ + (1 + R^\beta)^2 \text{Ln}(1 + R^\beta) (2\beta - \text{Ln}(4) + \text{Ln}(1 + R^\beta)) \\ - \beta (R^\beta + 1) \text{Ln}(R) \\ (2R^\beta (\beta (\beta - 1) \text{Ln}(2)) \\ - \beta \text{Ln}(4) + 2(R^\beta + \beta - R^\beta \beta) \text{Ln}(1 + R^\beta)) \end{array} \right] \end{array} \right\}$$

and

$$\text{Bias}(\hat{\Delta}) = \begin{cases} \frac{1.109 (n_1 + n_2) R^2}{n_1 n_2 \beta^2} D_1(R, \beta) + \frac{0.608\beta^2}{(n_1 + n_2)} D_2(R, \beta), & \text{for } R > 1, \\ \frac{-1.109 (n_1 + n_2) R^2}{n_1 n_2 \beta^2} D_1(R, \beta) - \frac{0.608\beta^2}{(n_1 + n_2)} D_2(R, \beta), & \text{for } R < 1, \end{cases} \quad (3.19)$$

where

$$D_1(R, \beta) = \frac{1}{(R^\beta - 1)^4 R^2} \left\{ \beta \left[\begin{array}{l} R^{4\beta} (R^\beta)^{\frac{1}{R-\beta-1}} (1 + \beta) + (R^\beta)^{\frac{1}{1-R^\beta}} \\ (\beta + (\beta - 1) \text{Ln}(R^\beta)) - (R^\beta)^{\frac{2-R^\beta}{1-R^\beta}} \\ (3 + (3\beta - 1) \text{Ln}(R^\beta)) - R^{3\beta} \\ \left[(R^\beta)^{\frac{1}{1-R^\beta}} (1 + 2\beta - (1 + \beta) \text{Ln}(R^\beta)) \right] + R^{2\beta} \\ + (R^\beta)^{\frac{1}{R-\beta-1}} (3 + (1 + 3\beta) \text{Ln}(R^\beta)) \end{array} \right] + \left[\begin{array}{l} (R^\beta)^{\frac{1}{1-R^\beta}} (3(1 + \beta) - \text{Ln}(R^\beta) (2 - 2\beta + \beta \text{Ln}(R^\beta))) \\ (R^\beta)^{\frac{1}{R-\beta-1}} (3 - 3\beta + \text{Ln}(R^\beta) (2 + 2\beta + \beta \text{Ln}(R^\beta))) \end{array} \right] \right\}$$

and

$$D_2(R, \beta) = \frac{1}{(1 + R^\beta)^4} \left\{ (\text{Ln}(R))^2 \left[\begin{array}{l} (R^{4\beta}) (R^\beta)^{\frac{1}{R-\beta-1}} - 3 (R^\beta)^{\frac{2-R^\beta}{1-R^\beta}} \text{Ln}(R^\beta) \\ + (R^\beta)^{\frac{1}{1-R^\beta}} (1 + \text{Ln}(R^\beta)) + R^{3\beta} \\ \left[(R^\beta)^{\frac{1}{1-R^\beta}} (\text{Ln}(R^\beta) - 2) - 3 (R^\beta)^{\frac{1}{R-\beta-1}} \text{Ln}(R^\beta) \right] \\ + R^{2\beta} \left[\begin{array}{l} (R^\beta)^{\frac{1}{R-\beta-1}} (\text{Ln}(R^\beta) - 1) (\text{Ln}(R^\beta) + 3) \\ - (R^\beta)^{\frac{1}{1-R^\beta}} (\text{Ln}(R^\beta) + 1) (\text{Ln}(R^\beta) - 3) \end{array} \right] \end{array} \right] \right\}$$

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

3.3. Bootstrap inference

Bootstrap methods are computer intensive and involve simulated data sets. Uniform (ordinary) bootstrap resampling by Efron [8] is based on resampling with replacement from the observed sample according to a rule which places equal probabilities on sample values. Uniform bootstrap resampling is an assumption-free method that can be used for some inferential problems. However, it is designed for complete and continuous set of observations (Hall [10]). In the two-sample case the uniform resampling rules applies to each sample separately and independently (see Ibrahim [12], Samawi *et al.* [32], Samawi *et al.* [33]).

Suppose that $\aleph_1 = (X_{11}, X_{12}, \dots, X_{1n_1})$ and $\aleph_2 = (X_{21}, X_{22}, \dots, X_{2n_2})$ 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. Let S be an estimate of OVL based on the random samples \aleph_1 and \aleph_2 *i.e.*, $S = S(\aleph_1, \aleph_2)$. Assume that U is a function of S *i.e.*, $U = U(S)$. Write U^* for the same function of the data but in the resamples $\aleph_1^* = (X_{11}^*, X_{12}^*, \dots, X_{1n_1}^*)$ and $\aleph_2^* = (X_{21}^*, X_{22}^*, \dots, X_{2n_2}^*)$ which are drawn from \aleph_1 and \aleph_2 according to a rule 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

$$\hat{u} = E(U^* | \aleph_1, \aleph_2). \tag{3.20}$$

This expected value is often not computable.

3.3.1. Uniform resampling approximation for bootstrap estimate

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

$$P(X_1^* = X_{1i} | \aleph_1) = \frac{1}{n_1}, \quad (3.21)$$

and probability of selecting X_2 in a resample is

$$P(X_2^* = X_{2i} | \aleph_2) = \frac{1}{n_2}. \quad (3.22)$$

Let $\aleph_{11}^*, \aleph_{12}^*, \dots, \aleph_{1B}^*$ and $\aleph_{21}^*, \aleph_{22}^*, \dots, \aleph_{2B}^*$ 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

$$\hat{u}_B^* = B^{-1} \sum_{b=1}^B (U_b^*). \quad (3.23)$$

Do and Hall [7] showed that \hat{u}_B^* is an unbiased approximation to \hat{u} , in the sense that $E(\hat{u}_B^* | \aleph_1, \aleph_2) = \hat{u}$. Moreover, an approximation of the bootstrap bias of u can be obtained by $\text{bias}^* = |\hat{u}_B^* - \hat{u}|$, and an approximation of the bootstrap MSE can be obtained by $M\hat{S}E^* = B^{-1} \sum_{b=1}^B (U_b^* - \hat{u})^2$.

3.3.2. Bootstrap confidence intervals of OVL

Adopting the notations of Section 3.3.1, let $S_1^*, S_2^*, \dots, S_B^*$ be the resampling realization of S . Then, the uniform resampling approximation to the bootstrap $100(1-d)\%$ confidence limits can be obtained as follows: Let $S_{(1)}^*, S_{(2)}^*, \dots, S_{(B)}^*$ be the order statistics of $S_1^*, S_2^*, \dots, S_B^*$. Define $K_1 = \text{int}\{Bd\}$ and $K_2 = \text{int}\{B(1-d)\}$. Then the uniform resampling approximation of the $100(1-d)\%$ confidence interval is $\left\{ \frac{S_{(K_1)}^* + S_{(K_1+1)}^*}{2}, \frac{S_{(K_2)}^* + S_{(K_2+1)}^*}{2} \right\}$.

4. SIMULATION STUDY

A Monte Carlo simulation study was conducted for $R = 0.2, 0.5$, and 0.80 , $(n_1, n_2) = (20, 20), (20, 30), (50, 50), (100, 100)$ and $(200, 200)$ and $d = 0.05$. All the 1000 simulated sets of observations were generated under the assumption that both variables have Weibull distribution with the same shape parameter ($\beta = 2$) but with different scale parameters. For bootstrap approximation 1000 resamples were used. Note that the results presented in Table 1–3 are based on (3.14)–(3.16).

Tables 1–3, indicate that the $|\text{bias}|$ in all cases are less than 0.04 and decreases as the sample size increases. It seems that the Taylor series approximation works well. Also, bootstrap method, which does not need complicated formula for the variances seems to work fairly well except for small R .

The asymptotic method gives a good estimate for the coverage probability of the 95% confidence intervals for ρ , λ and Δ . However, the coverage probability of those 95% confidence intervals does not converge to the nominal values when the bootstrap method is used for small R .

5. APPLICATION ON BODY MASS INDEX DATA FROM 65+ RHS STUDY

The 65+ Rural Health Study (RHS) is a prospective longitudinal cohort study of 3673 individuals (1420 men and 2253 women) aged 65 or older living in Washington and Iowa counties of Iowa in 1982. This study is one of four studies supported by the National Institute on Aging and collectively referred to as EPESE, Established Populations for Epidemiological Studies of the Elderly, see Brock *et al.* [3].

TABLE 1. Estimates (Est.), Bias, MSE, length of interval (L.), and the coverage probability (Cover.) for $R = 0.20$. Exact OVL coefficients: $\rho = 0.385, \lambda = 0.178$ and $\Delta = 0.161$.

(n_1, n_2)	Asymptotic inference					Bootstrap inference			
		Est.	MSE(Bias)	Cover	L.	Est.	MSE*(bias*)	Cover	L.
(20, 20)	ρ	0.352	0.0092(0.033)	0.930	0.352	0.317	0.0120(0.068)	0.810	0.320
	λ	0.158	0.0057(0.020)	0.928	0.305	0.138	0.0059(0.040)	0.810	0.236
	Δ	0.144	0.0036(0.017)	0.985	0.378	0.126	0.0040(0.035)	0.810	0.191
(20, 30)	ρ	0.358	0.0070(0.026)	0.947	0.320	0.327	0.0094(0.058)	0.805	0.290
	λ	0.162	0.0045(0.016)	0.946	0.275	0.275	0.0049(0.035)	0.808	0.218
	Δ	0.147	0.0028(0.014)	0.995	0.365	0.365	0.0324(0.030)	0.805	0.175
(50, 50)	ρ	0.373	0.0034(0.020)	0.950	0.226	0.351	0.0042(0.034)	0.878	0.211
	λ	0.171	0.0023(0.007)	0.951	0.192	0.156	0.0025(0.022)	0.878	0.168
	Δ	0.154	0.0014(0.006)	0.996	0.311	0.142	0.0553(0.019)	0.878	0.133
(100, 100)	ρ	0.379	0.0016(0.006)	0.961	0.161	0.362	0.0020(0.023)	0.893	0.151
	λ	0.174	0.0011(0.004)	0.959	0.136	0.162	0.0013(0.016)	0.893	0.123
	Δ	0.158	0.0007(0.003)	0.999	0.227	0.147	0.0067(0.013)	0.893	0.097
(200, 200)	ρ	0.380	0.0008(0.004)	0.972	0.114	0.367	0.0011(0.018)	0.898	0.108
	λ	0.175	0.0005(0.003)	0.095	0.095	0.164	0.0007(0.014)	0.899	0.089
	Δ	0.158	0.0003(0.003)	0.999	0.161	0.149	0.0723(0.011)	0.898	0.070

TABLE 2. Estimates (Est.), Bias, MSE, length of interval (L.), and the coverage probability (Cover.) for $R = 0.50$. Exact OVL coefficients: $\rho = 0.800, \lambda = 0.675$ and $\Delta = 0.528$.

(n_1, n_2)	Asymptotic inference					Bootstrap inference			
		Est.	MSE(Bias)	Cover	L.	Est.	MSE*(bias*)	Cover	L.
(20, 20)	ρ	0.774	0.0091(0.026)	0.945	0.354	0.746	0.0120(0.054)	0.905	0.366
	λ	0.642	0.0183(0.033)	0.974	0.598	0.608	0.0218(0.067)	0.903	0.500
	Δ	0.508	0.0182(0.020)	0.992	0.643	0.487	0.0175(0.041)	0.905	0.418
(20, 30)	ρ	0.774	0.0077(0.026)	0.947	0.323	0.749	0.0101(0.051)	0.886	0.328
	λ	0.641	0.0156(0.034)	0.962	0.546	0.610	0.0186(0.065)	0.886	0.452
	Δ	0.506	0.0098(0.022)	0.990	0.589	0.486	0.0437(0.042)	0.886	0.370
(50, 50)	ρ	0.786	0.0036(0.014)	0.954	0.224	0.774	0.0043(0.026)	0.907	0.222
	λ	0.656	0.0077(0.019)	0.976	0.382	0.640	0.0086(0.035)	0.906	0.319
	Δ	0.515	0.0049(0.013)	0.998	0.410	0.505	0.0606(0.022)	0.907	0.259
(100, 100)	ρ	0.797	0.0015(0.003)	0.957	0.157	0.790	0.0017(0.010)	0.936	0.152
	λ	0.672	0.0034(0.003)	0.981	0.272	0.662	0.0036(0.012)	0.936	0.224
	Δ	0.526	0.0023(0.001)	0.998	0.290	0.520	0.0691(0.007)	0.936	0.183
(200, 200)	ρ	0.798	0.0007(0.002)	0.961	0.11	0.794	0.0008(0.006)	0.947	0.108
	λ	0.672	0.0016(0.003)	0.980	0.193	0.666	0.0017(0.009)	0.947	0.160
	Δ	0.526	0.0011(0.001)	0.998	0.205	0.522	0.0734(0.011)	0.947	0.130

TABLE 3. Estimates (Est.), Bias, MSE, length of interval (L.), and the coverage probability (Cover.) for $R = 0.80$. Exact OVL coefficients: $\rho = 0.976, \lambda = 0.958$ and $\Delta = 0.837$.

(n_1, n_2)	Asymptotic inference					Bootstrap inference				
		Est.	MSE(Bias)	Cover	L.	Est.	MSE*(bias*)	Cover	L.	
(20, 20)	ρ	0.958	0.0021(0.017)	0.890	0.156	0.940	0.0034(0.036)	0.945	0.183	
	λ	0.929	0.0058(0.029)	0.940	0.398	0.900	0.0088(0.058)	0.945	0.294	
	Δ	0.815	0.0121(0.022)	0.995	0.528	0.786	0.0292(0.051)	0.945	0.404	
(20, 30)	ρ	0.962	0.0018(0.013)	0.853	0.135	0.947	0.0026(0.028)	0.938	0.155	
	λ	0.936	0.0049(0.022)	0.922	0.347	0.911	0.0070(0.046)	0.938	0.253	
	Δ	0.826	0.0110(0.012)	0.989	0.479	0.800	0.0537(0.038)	0.938	0.367	
(50, 50)	ρ	0.968	0.0008(0.008)	0.900	0.093	0.961	0.0010(0.014)	0.938	0.095	
	λ	0.945	0.0021(0.013)	0.949	0.247	0.934	0.0026(0.024)	0.938	0.159	
	Δ	0.829	0.0058(0.009)	0.981	0.329	0.820	0.0660(0.017)	0.938	0.271	
(100, 100)	ρ	0.973	0.0003(0.003)	0.921	0.063	0.970	0.0003(0.006)	0.942	0.062	
	λ	0.954	0.0008(0.004)	0.977	0.172	0.948	0.0009(0.010)	0.942	0.105	
	Δ	0.837	0.0028(0.000)	0.976	0.230	0.834	0.0719(0.003)	0.942	0.200	
(200, 200)	ρ	0.974	0.0001(0.002)	0.932	0.045	0.972	0.0001(0.003)	0.945	0.043	
	λ	0.955	0.0004(0.003)	0.992	0.124	0.952	0.0004(0.005)	0.945	0.074	
	Δ	0.836	0.0014(0.000)	0.970	0.163	0.835	0.0748(0.002)	0.945	0.144	

TABLE 4. Body mass index (kg/m²) of diabetic women aged 80 to 85 years in the Iowa 65+ Rural Health Study by urinary incontinence.

No Urinary Incontinence	18.88	19.35	19.52	19.77	20.57	21.65	21.67	22.90	23.45	23.45
Urinary Incontinence	23.46	23.62	26.27	25.66	27.50	28.95	30.10	30.17	36.65	
Urinary Incontinence	19.61	22.88	24.07	24.38	25.24	25.66	26.01	27.31	27.49	29.29
	31.31	31.95	33.52	37.66						

Table 4 presents the body mass index data (BMI) collected at the baseline for the Iowa 65+ RHS. In this study there were 33 diabetic women aged 80 to 85 of whom 14 reported urinary incontinence. The question of interest is to draw inference about the percentage of similarity in the BMI of the diabetic women with or without urinary incontinence. The BMI is the ratio of the subject’s weight (kg) divided by height square (m²). Figures 3 and 4 (including the Anderson-Darling goodness of fit test) indicate that the underlying distribution for this data follows Weibull distributions with P -values > 0.99 .

Moreover, using the asymptotic normality of the MLEs of the shape parameters, our test of the two shape parameters equality using two samples Z -test ($Z = 0.55; P$ -value = 0.291) indicates that the two distributions shape parameters are statistically equal. Also, the combined MLE from the two samples for β is $\beta^* = 5.72$. Table 5, summarizes the asymptotic inference and the bootstrap inference on the OVL measures for the BMI data of urinary incontinence and non- urinary incontinence women aged 80-85 at the baseline for the Iowa 65+ RHS.

Table 5 shows an illustration of estimating the OVL coefficients for two weibull distributions. For example MLE of $\Delta = 0.686$ indicates that about 68% of diabetic women aged 80–85 years with or without urinary incontinence have the same range of BMI.

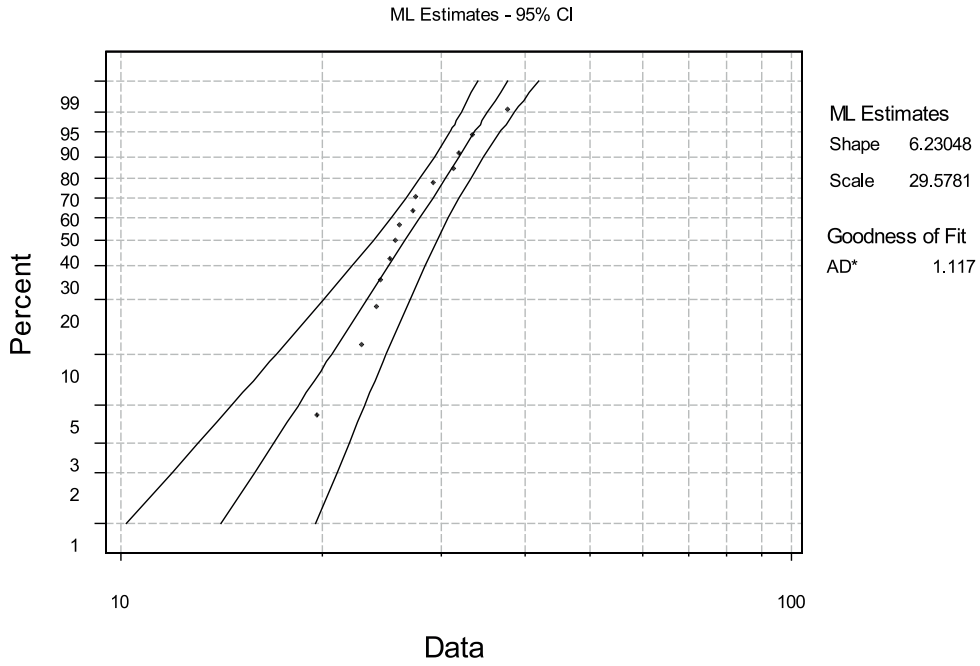


FIGURE 3. Weibull probability plot for body mass index of diabetic women aged 80–85 with urinary incontinence.

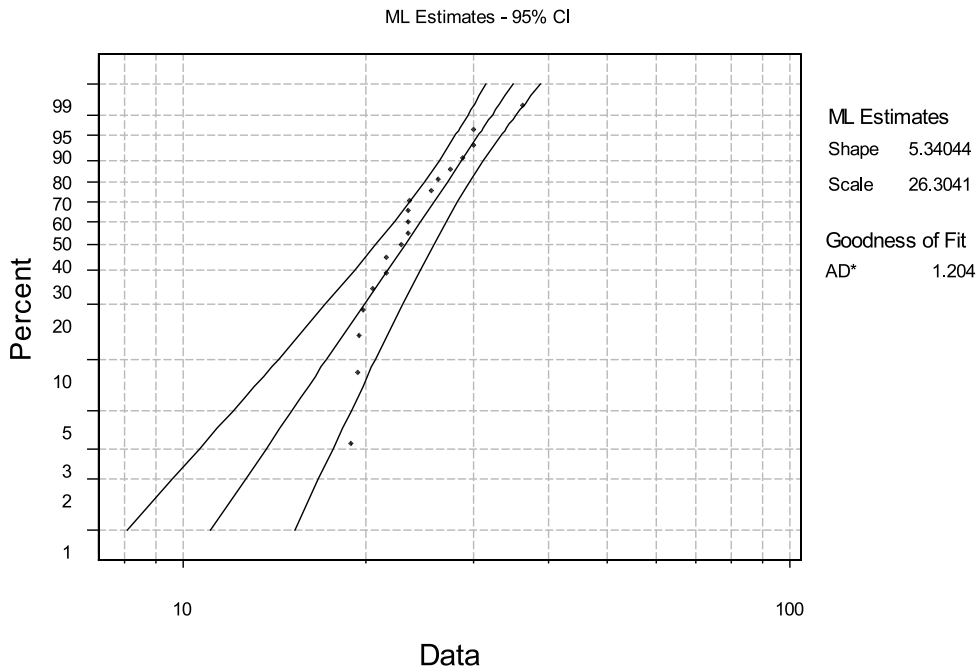


FIGURE 4. Weibull probability plot for body mass index of diabetic women aged 80–85 with no urinary incontinence.

TABLE 5. Results based on the real data of BMI in Table 4.

			Asymptotic Inference		Bootstrap Inference based on 1000 resamples				
			95% confidence Interval limits				95% confidence Interval limits		
Coefficients	MLEs	MSE	Lower	Upper	Estimate	$\hat{MSE}^*(\hat{bias}^*)$	Lower	Upper	
ρ	0.910	0.0065	0.777	1.000	0.922	0.0080 (0.012)	0.703	1.000	
λ	0.840	0.0940	0.460	1.000	0.865	0.0211 (0.025)	0.516	1.000	
Δ	0.686	0.0273	0.496	0.948	0.751	0.0292 (0.065)	0.421	0.986	

REFERENCES

- [1] L.J. Bain and C.E. Antle, Estimation of parameters in Weibull the distribution. *Technometrics* **9** (1967) 621–627.
- [2] L.J. Bain and M. Engelhardt, *Statistical analysis of reliability and life-testing models*. Marcel Dekker (1991).
- [3] D.B. Brock, T. Wineland, D.H. Freeman, J.H. Lemke and P.A. Scherr, Demographic characteristics, in *Established Population for Epidemiologic Studies of the Elderly*, Resource Data Book, J. Cornoni-Huntley, D.B. Brock, A.M. Ostfeld, J.O. Taylor and R.B. Wallace Eds. National Institute on Aging, NIH Publication No. 86-2443. US Government Printing Office, Washington, DC (1986).
- [4] T.E. Clemons and Bradley Jr., A nonparametric measure of the overlapping coefficient. *Comp. Statist. Data Analysis* **34** (2000) 51–61.
- [5] A.C. Cohen, Multi-censored sampling in three-parameter Weibull distribution. *Technometrics* **17** (1974) 347–352.
- [6] P.M. Dixon, The Bootstrap and the Jackknife: describing the precision of ecological Indices, in *Design and Analysis of Ecological Experiments*, S.M. Scheiner and J. Gurevitch Eds. Chapman & Hall, New York (1993) 209–318.
- [7] K.N. Do and P. Hall, On importance resampling for the bootstrap. *Biometrika* **78** (1991) 161–167.
- [8] B. Efron, Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** (1979) 1–26.
- [9] W.T. Federer, L.R. Powers and M.G. Payne, Studies on statistical procedures applied to chemical genetic data from sugar beets. *Technical Bulletin*, Agricultural Experimentation Station, Colorado State University **77** (1963).
- [10] P. Hall, On the removal of Skewness by transformation. *J. R. Statist. Soc. B* **54** (1992) 221–228.
- [11] H.L. Harter and A.H. Moore, Asymptotic variances and covariances of maximum-likelihood estimators, from censored samples, of the parameters of the Weibull and gamma populations. *Ann. Math. Statist.* **38** (1967) 557–570.
- [12] H.I. Ibrahim, Evaluating the power of the Mann-Whitney test using the bootstrap method. *Commun. Statist. Theory Meth.* **20** (1991) 2919–2931.
- [13] M. Ichikawa, A meaning of the overlapped area under probability density curves of stress and strength. *Reliab. Eng. System Safety* **41** (1993) 203–204.
- [14] H.F. Inman and E.L. Bradley, The Overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Comm. Statist. Theory Methods* **18** (1989) 3851–3874.
- [15] F.C. Leone, Y.H. Rutenberg and C.W. Topp, Order statistics and estimators for the Weibull population. *Tech. Reps.* AFOSR TN 60-489 and AD 237042, Air Force Office of Scientific Research, Washington, DC (1960).
- [16] J. Lieblein and M. Zelen, Statistical investigations of the fatigue life of deep groove ball bearings. Research Paper 2719. *J. Res. Natl. Bur Stand.* **57** (1956) 273–316.
- [17] R. Lu, E.P. Smith and I.J. Good, Multivariate measures of similarity and niche overlap. *Theoret. Population Ecol.* **35** (1989) 1–21.
- [18] N. Mann, *Point and Interval Estimates for Reliability Parameters when Failure Times have the Two-Parameter Weibull Distribution*. Ph.D. dissertation, University of California at Los Angeles, Los Angeles, CA (1965).
- [19] N. Mann, Results on location and scale parameters estimation with application to Extreme-Value distribution. *Tech. Rep. ARL 670023*, Office of Aerospace Research, USAF, Wright-Patterson **AFB, OH** (1967a).
- [20] N. Mann, Tables for obtaining the best linear invariant estimates of parameters of the Weibull distribution. *Technometrics* **9** (1967b) 629–645.
- [21] N. Mann, Best linear invariant estimation for Weibull distribution. *Technometrics* **13** (1971) 521–533.
- [22] K. Matusita, Decision rules based on the distance for problem of fir, two samples, and Estimation. *Ann. Math. Statist.* **26** (1955) 631–640.
- [23] J.I. McCool, Inference on Weibull Percentiles and shape parameter from maximum likelihood estimates. *IEEE Trans. Rel.* **R-19** (1970) 2–9.
- [24] S.N. Mishra, A.K. Shah and J.J. Lefante, Overlapping coefficient: the generalized t approach. *Commun. Statist. Theory Methods* (1986) **15** 123–128.
- [25] M. Morisita, Measuring interspecific association and similarity between communities. *Memoirs of the faculty of Kyushu University. Series E. Biology* **3** (1959) 36–80.

- [26] M.S. Mulekar and S.N. Mishra, Overlap Coefficient of two normal densities: equal means case. *J. Japan Statist. Soc.* **24** (1994) 169–180.
- [27] M.S. Mulekar and S.N. Mishra, Confidence interval estimation of overlap: equal means case. *Comp. Statist. Data Analysis* **34** (2000) 121–137.
- [28] D.N.P. Murthy, M. Xie and R. Jiang, *Weibull Models*. John Wiley & Sons (2004).
- [29] M. Pike, A suggested method of analysis of a certain class of experiments in carcinogenesis. *Biometrics* **29** (1966) 142–161.
- [30] B. Reser and D. Faraggi, Confidence intervals for the overlapping coefficient: the normal equal variance case. *The statistician* **48** (1999) 413–418.
- [31] P. Rosen and B. Rammler, The laws governing the fineness of powdered coal. *J. Inst. Fuels* **6** (1933) 29–36.
- [32] H.M. Samawi, G.G. Woodworth and M.F. Al-Saleh, Two-Sample importance resampling for the bootstrap. *Metron* (1996) Vol. LIV No. 3–4.
- [33] H.M. Samawi, Power estimation for two-sample tests using importance and antithetic r resampling. *Biometrical J.* **40** (1998) 341–354.
- [34] E.P. Smith, Niche breadth, resource availability, and inference. *Ecology* **63** (1982) 1675–1681.
- [35] P.H.A. Sneath, 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** (1977) 123–143.
- [36] D.R. Thoman, L.J. Bain and C.E. Antle, Inference on the parameters of the Weibull distribution. *Technometrics* **11** (1969) 445–460.
- [37] W. Weibull, A statistical theory of the strength of materials. *Ing. Vetenskaps Akad. Handl.* **151** (1939) 1–45.
- [38] W. Weibull, A statistical distribution function of wide application. *J. Appl. Mech.* **18** (1951) 293–297.
- [39] M.S. Weitzman, Measures of overlap of income distributions of white and Negro families in the United States. *Technical paper No. 22*. Department of Commerce, Bureau of Census, Washington, US (1970).
- [40] J.S. White, The moments of log-Weibull Order Statistics. *General Motors Research Publication GMR-717*. General Motors Corporation, Warren, Michigan (1967).