Kernel Method for Estimating Matusita Overlapping Coefficient Using Numerical Approximations

Omar (omarm@yu.edu.jo)
Department of Statistics - Yarmouk University - Irbid - Jordan
Enas
Department of Statistics, Yarmouk University

Research Article

Keywords: Matusita Overlapping Coefficient, Kernel Density Estimation, Reflection Kernel, Smoothing Parameter, Trapezoidal Rule, Simpson Rule, Bias and Relative Root Mean Square Error

Posted Date: June 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1731983/v1

Kernel Method for Estimating Matusita Overlapping Coefficient Using Numerical Approximations

Omar M. Eidous Enas A. Ananbeh

omarm@yu.edu.jo 2019107001@ses.yu.edu.jo

Department of Statistics – Faculty of Science

Yarmouk University

Irbid - Jordan

Abstract

In this paper, the nonparametric kernel method is used to estimate the well known overlapping coefficient known as Matusita ρ. Due to the complexity of finding the formula expression of this coefficient when using the kernel estimators, we suggest to use the numerical integration method to approximate its integral as a first step. Then the kernel estimators were combined with the new approximation to formulate the proposed estimators. Two numerical integration rules known as trapezoidal and Simpson rules were used to approximate the interesting integral. The proposed technique produced two new estimators for ρ. The proposed estimators are studied and compared with existing estimator developed by Eidous and Al-Talafheh (2020) via Monte-Carlo simulation technique. The simulation results clearly demonstrated the usefulness and efficiency of the new technique for estimating ρ.

Keywords: Matusita Overlapping Coefficient; Kernel Density Estimation; Reflection Kernel; Smoothing Parameter; Trapezoidal Rule; Simpson Rule; Bias and Relative Root Mean Square Error.

*** Correspondence:** omarm@yu.edu.jo

Omar Eidous, Department of Statistics, Yarmouk University, Jordan.

1. Introduction

Let $f_X(x)$ and $f_Y(x)$ be two continuous probability density functions, the formula for the overlapping (OVL) coefficient of Matusita (1955) ρ is defined by,

$$\rho = \int \sqrt{f_X(x) f_Y(x)}\, dx$$

As pointed out by some authors, the integrals in the above formulas should be replaced by summations (Inman and Bradly, 1989) for discrete distributions. The value of ρ is ranging from 0 to 1. If ρ value is 0 then it indicates that the two populations are entirely different and there is no common support between the two densities. On the other hand, if ρ value is 1 then it indicates a perfect agreement between two population densities. In general, the OVL measures became an important tool in goodness of fit test. Samawi *et al*. (2011) developed a new test of symmetry based on OVL using kernel density estimation. Recently, Alodat *et al.* (2021) showed the importance of OVL Matusita measure ρ in goodness-of-fit test for two independent populations by deriving the asymptotic sampling distribution of the kernel estimator of ρ. Also, they obtained an approximate confidence interval for ρ. For the applications importance of the OVL measures, and for parametric and nonparametric methods of estimation the different OVL measure, see Eidous and Al-Tafhah (2020) and the references therein.

Our main interest in this work is to suggest and study a new technique to estimate the OVL measure ρ.

Let $\hat{\rho} = \int \sqrt{\hat{f}_X(x) \hat{f}_Y(x)}\, dx$ be the kernel estimator of ρ, where $\hat{f}_X(x)$ and $\hat{f}_Y(x)$ are the kernel estimators (or reflection kernel estimators) of $f_X(x)$ and $f_Y(x)$ respectively (See Section 2.1 below). Without evaluating the closed form of the above integral, Alodat *et al.* (2021) studied the asymptotic properties of $\hat{\rho}$ and derived its asymptotic sampling distribution. Because the complexity of finding a closed form for $\hat{\rho}$, we suggest to use numerical integration methods to approximate $\rho = \int \sqrt{f_X(x) f_Y(x)}\, dx$ and then to derive new estimators based on the resulting approximations.

2. Kernel estimator and Numerical Integration Methods

2.1 Kernel estimators

Let $X_1, X_2, \ldots, X_{n_1}$ be a random sample of size n_1 from a continuous pdf $f_X(x)$ and let $Y_1, Y_2, \ldots, Y_{n_2}$ be another random sample of size n_2 from a continuous $f_Y(y)$, where the two samples are independent. The kernel density estimators of $f_X(x)$ and $f_Y(y)$ are respectively (Silverman, 1986),

$$\hat{f}_X(x) = \frac{1}{n_1 h_1} \sum_{i=1}^{n_1} K\left(\frac{x - X_i}{h_1}\right), \quad -\infty < x < \infty,$$

$$\hat{f}_Y(y) = \frac{1}{n_2 h_2} \sum_{i=1}^{n_2} K\left(\frac{y - Y_i}{h_2}\right), \quad -\infty < y < \infty,$$

where $h_i, i = 1, 2$ are called a smoothing parameters (or a bandwidths) and K is called a kernel function assumed to be symmetric density function. These quantities are under user control. In this study, the kernel function is taken to be Gaussian kernel, $K(u) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}u^2}$ and the smoothing parameters are computed by using the following rule, which recommended by Silverman (1986),

$$h_i = 0.9\, A\, n_i^{-1/5}, \; i = 1, 2$$

where $A = min\{\text{standard deviation}, \text{interquartile range}/1.34\}$ is computed based on the first sample $X_1, X_2, \dots, X_{n_1}$ to obtain h_1 and is computed based on the second sample $Y_1, Y_2, \dots, Y_{n_2}$ to obtain h_2.

If the support of X and Y is $(0, \infty)$ then the reflection kernel estimator (Silverman, 1986) of $f_X(x)$ and $f_Y(y)$ is used to correct the kernel estimators near the boundaries $x = 0$ and $y = 0$. The reflection kernel estimators of $f_X(x)$ and $f_Y(y)$ are respectively,

$$\hat{f}_{XR}(x) = \frac{1}{n_1 h_1} \sum_{i=1}^{n_1} \left(K\left(\frac{x - X_i}{h_1}\right) + K\left(\frac{x + X_i}{h_1}\right) \right), \quad 0 < x < \infty$$

$$\hat{f}_{YR}(y) = \frac{1}{n_2 h_2} \sum_{i=1}^{n_2} \left(K\left(\frac{y - Y_i}{h_2}\right) + K\left(\frac{y + Y_i}{h_2}\right) \right), \quad 0 < y < \infty\, .$$

2.2 Numerical Integrals

Numerical integrals are used to approximate the value of a definite integral when a closed form of the integral is difficult to find or when an approximation of the definite integral is needed. The two methods of numerical integration, known as trapezoidal and Simpson rules (Chapra and Canale, 2014) are adopted in this work.

Let $u(x)$ be a continuous on $[a, b]$ where $[a, b]$ is partitioned into r subintervals of equal length (Δt, say) as follows:

$$a = t_0 < t_1 < t_2 < \cdots < t_{k-1} < t_r = b\,,$$

where $\Delta t = \frac{b-a}{r}$ and $t_i = a + i(\Delta t),\ i = 0, 1, 2, \dots, r.$

The Trapezoidal and Simpson rules for approximation $\int_a^b u(x)dx$ are, respectively,

$$\int_a^b u(x)dx \cong \frac{\Delta t}{2}[u(a) + 2u(t_1) + 2u(t_2) \dots + 2u(t_{r-1}) + u(b)]$$
$$= \frac{\Delta t}{2}\sum_{j=0}^{r} u(t_j).$$

$$\int_a^b u(x)dx \cong \frac{3\Delta t}{8}[u(a) + 3u(t_1) + 3u(t_2) + 2u(t_3) + 3u(t_4) + 3u(t_5)$$
$$+2u(t_6) + \cdots + 2u(t_{r-3}) + 3u(t_{r-2}) + 3u(t_{r-1}) + u(b)]$$
$$= \frac{3\Delta t}{8}\left(u(a) + 3\sum_{\substack{j=1, j\neq 3k \\ k=1,2,3,\dots}}^{r-1} u(t_j) + 2\sum_{j=1}^{r/3-1} u(t_{3j}) + u(b) \right).$$

where a and b are finite real numbers and r be a positive integer. The value of r for Simpson rule should be a multiple of 3.

3. Approximation of ρ

Assume that X and Y be two independent random variables with $pdfs$ $f_X(x)$ and $f_Y(y)$ respectively, where the two random variables X and Y are defined on $(-\infty, \infty)$ or on $(0, \infty)$. If we consider the first case where X and Y are defined on $(-\infty, \infty)$ (The second case $(0, \infty)$ can be treated similarly). The Matusita overlapping measure ρ is defined by,

$$\rho = \int_{-\infty}^{\infty} \sqrt{f_X(x) f_Y(x)}\, dx.$$

Evaluate the above integral for general $f_X(x)$ and $f_Y(x)$ is not possible. Also, it is not easy job to evaluate it by taking the kernel estimators of $f_X(x)$ and $f_Y(x)$ instead of $f_X(x)$ and $f_Y(x)$ themselves. Therefore, our main interest is to approximate the above integral by using the trapezoidal and Simpson numerical integration methods. The estimation process will be achieved after the approximation process.

Since the above definite integral is over the unbounded interval $(-\infty, \infty)$ then at first, we need to determine a proper transformation that enables us to apply the trapezoidal and Simpson rules. These numerical methods require integral limits (bounds) to be finite.

Let $u = q(x)$, where $q(x)$ is a continuous increasing function in x defined on the support of x, ($i.e. q(x)$ is defined on $(-\infty, \infty)$) so that $q(-\infty) = a$ and $q(\infty) = b$, where $a < b$ and a and b are finite numbers. Then,

$$x = q^{-1}(u) \text{ and } dx = \frac{dq^{-1}(u)}{du}.$$

Therefore,

$$\rho = \int_a^b \sqrt{f_X(q^{-1}(u)) f_Y(q^{-1}(u))} \frac{dq^{-1}(u)}{du}. \quad (1)$$

Let Z be a continuous random variable with cumulative distribution function (CDF) $F_Z(z)$, $-\infty < z < \infty$ then the case $q(x) = F_Z(x)$ is of our special interest. In this case, the transformation $u = F_Z(x)$ gives $= F_Z^{-1}(u)$, $dx = \frac{d}{du} F_Z^{-1}(u)$ and

$$\rho = \int_0^1 \sqrt{f_X(F_Z^{-1}(u)) f_Y(F_Z^{-1}(u))} \frac{d}{du} F_Z^{-1}(u). \quad (2)$$

Let $\rho\big(F_Z(X), F_Z(Y)\big)$ be the Matusita overlapping measure between $F_Z(X)$ and $F_Z(Y)$ and $\rho(X, Y)$ is the Matusita overlapping measure between X and Y then we state the following theorem.

Theorem (1): The coefficient ρ in (2) is $\rho\big(F_Z(X), F_Z(Y)\big)$. That is,

$$\rho\big(F_Z(X), F_Z(Y)\big) = \int_0^1 \sqrt{f_X(F_Z^{-1}(u)) f_Y(F_Z^{-1}(u))} \frac{d}{du} F_Z^{-1}(u).$$

Proof: Let X be a random variables with pdf $f_X(x)$ and let $W_1 = F_Z(X)$ then $x = F_Z^{-1}(w_1)$ and $dx = \frac{d}{dw_1} F_Z^{-1}(w_1)$. Therefore and by using the transformation method, the pdf of W_1 is,

$$f_{W_1}(w_1) = f_X(F_Z^{-1}(w_1)) \frac{d}{dw_1} F_Z^{-1}(w_1)$$

If Y is a random variables with pdf $f_Y(y)$ and if $W_2 = F_Z(Y)$ then $y = F_Z^{-1}(w_2)$ and $dy = \frac{d}{dw_2} F_Z^{-1}(w_2)$. Therefore, the pdf of W_2 is,

$$f_{W_2}(w_2) = f_Y(F_Z^{-1}(w_2)) \frac{d}{dw_2} F_Z^{-1}(w_2)$$

The Matusita overlapping coefficient between $F_Z(X)$ and $F_Z(Y)$ is,

$$\begin{aligned}
\rho\big(F_Z(X)\,,F_Z(Y)\big) &= \rho(W_1\,,W_2) \\
&= \int_0^1 \sqrt{f_{W_1}(w_1) f_{W_2}(w_1)}\,dw_1 \\
&= \int_0^1 \sqrt{f_X(F_Z^{-1}(w_1))\frac{d}{dt}F_Z^{-1}(w_1) f_Y(F_Z^{-1}(w_1))\frac{d}{dt}F_Z^{-1}(w_1)}\,dw_1 \\
&= \int_0^1 \sqrt{f_X(F_Z^{-1}(w_1)) f_Y(F_Z^{-1}(w_1))}\,\frac{d}{dt}F_Z^{-1}(w_1) \\
&= \int_0^1 \sqrt{f_X(F_Z^{-1}(u)) f_Y(F_Z^{-1}(u))}\,\frac{d}{du}F_Z^{-1}(u)
\end{aligned}$$

This completes the proof.

Because the formula of ρ in (2) is derived based on the formula of ρ in (1) then the main message of Theorem (1) is:

The Matusita overlapping measure ρ is invariant with respect to any increasing function. That is, $\rho(X\,,Y) = \rho\big(F_Z(X)\,,F_Z(Y)\big)$.

It is worth to note here the following *important* issues:

1. Despite that F_Z represents a CDF for a continuous random variable Z, the result of Theorem (1) is still correct for any continuous increasing function that is not necessarily a CDF.
2. Analogous to the above discussion, it is easy to verify that ρ is invariant with respect to a decreasing function.
3. It is a straightforward technique to show that ρ is invariant with respect to increasing and decreasing functions when the support of X and Y is $(0, \infty)$.

Now, the coefficient ρ that is given by (2) can be approximated by using trapezoidal and Simpson rules as follows:

Let $g(u) = \sqrt{f_X(F_Z^{-1}(u)) f_Y(F_Z^{-1}(u))}\,\frac{d}{du}F_Z^{-1}(u)$ then,

$$\rho = \int_0^1 \sqrt{f_X\big(F_Z^{-1}(u)\big) f_Y\big(F_Z^{-1}(u)\big)}\,\frac{d}{du}F_Z^{-1}(u) = \int_0^1 g(u)du.$$

We will denote the approximate value of ρ used the trapezoidal rule by ρ_1 and that used the Simpson rule by ρ_2, which are given as follows:

- The approximation of $\rho = \int_0^1 g(u)du$ by using trapezoidal rule is given by,

$$\rho_1 \cong \frac{1}{2r}\sum_{j=0}^{r} g\left(\frac{j}{r}\right).$$

- The approximation of $\rho = \int_0^1 g(u)du$ by using Simpson rule is,

$$\rho_2 \cong \frac{3\Delta t}{8}\left(g(0) + 3 \sum_{\substack{j=1, j\neq 3k \\ k=1,2,3,\ldots}}^{r-1} g\left(\frac{j}{r}\right) + 2 \sum_{j=1}^{r/3-1} g\left(\frac{3j}{r}\right) + g(1) \right)$$

In all of the above approximations, $g(0) = g(1) = 0$.

4. Estimation of ρ

Let the common support of the two independent random variables X and Y is $(-\infty, \infty)$ and let $\hat{g}(u) = \sqrt{\hat{f}_X(F_Z^{-1}(u))\hat{f}_Y(F_Z^{-1}(u))\frac{d}{du}F_Z^{-1}(u)}$ be a kernel estimator of $g(u)$, where $\hat{f}_X(F_Z^{-1}(u))$ is a kernel estimator of f_X at a point $F_Z^{-1}(u)$ and $\hat{f}_Y(F_Z^{-1}(u))$ is a kernel estimator of f_Y at a point $F_Z^{-1}(u)$. The proposed estimators of ρ are:

- The proposed estimator of ρ_1 is (r is a positive integer number),

$$\hat{\rho}_1 = \frac{1}{r}\sum_{j=1}^{r-1} \hat{g}\left(\frac{j}{r}\right).$$

- The proposed estimator of ρ_2 is (r is a positive integer and multiple of 3),

$$\hat{\rho}_2 = \frac{3}{8r}\left(2 \sum_{j=1}^{r/3-1} \hat{g}\left(\frac{3j}{r}\right) + 3 \sum_{j=1, j\neq 3k}^{r-1} \hat{g}\left(\frac{j}{r}\right) \right), \qquad k = 1,2,3, \ldots$$

If the support of X and Y is $(0, \infty)$ then the reflection kernel estimators (See Subsection 2.1) of f_X and f_Y is used to obtain $\hat{g}(u)$ then $\hat{\rho}_1$ and $\hat{\rho}_1$.

5. Simulation study design and results

In this section, a simulation study is designed and performed to investigate the properties of the two proposed estimators of ρ. For sake of comparison, we include the estimator of ρ that suggested by Eidous and Al-Talafheh (2020). Since all of these estimators are nonparametric, we simulate the two samples of data from 4 families of distributions at which two families are defined on the positive real line and the other two families are defined on the entire real line. For each family of distributions, we considered 4 cases depending on the selection of the corresponding parameter(s). Despite that the selection of distribution parameters seem to be arbitrary, we keep in our mind that they should be selected to vary the values of ρ between 0 and 1. For each of the 16 considered cases that stated in Table (1), the two independent samples $x_1, x_2, \dots, x_{n_1}$ and $y_1, y_2, \dots, y_{n_2}$ with $(n_1, n_2) = (12,12)$, $(36,54)$, $(96,180)$ were generated 1000 times.

The simulated distributions and their selected parameters together with the exact values of ρ in each case are given in Table (1). Figure (1) shows the shape of pair distributions for each case of simulations, where the shaded area represents the intersection area between the two distributions.

- Weibull distribution, ($Weib(\alpha, \beta)$):

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-(x/\alpha)^{\beta}}, \qquad x > 0, \qquad \alpha, \beta > 0$$

- Lomax distribution ($Lom(\alpha, \lambda)$):

$$f(x) = \frac{\alpha}{\lambda}\left(1 + \frac{x}{\lambda}\right)^{-(\alpha+1)}, \quad x > 0, \ \alpha, \lambda > 0$$

- Gumbel distribution ($Gum(\alpha, \beta)$):

$$f(x) = \frac{e^{-e^{\frac{x-\alpha}{\beta}} + \frac{x-\alpha}{\beta}}}{\beta}, \quad x \in \mathcal{R}, \quad \alpha \in \mathcal{R}, \quad \beta > 0$$

- Normal distribution ($N(\mu, \sigma^2)$):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathcal{R}, \ \mu \in \mathcal{R}, \ \sigma^2 > 0$$

Note that the support of the first two distributions is $(0, \infty)$, while the support of the last two distributions is $(-\infty, \infty)$.

The interested estimators are the two proposed estimators $\hat{\rho}_{Rtrap}$ and $\hat{\rho}_{Rsimp}$ (or $\hat{\rho}_{trap}$ and $\hat{\rho}_{simp}$), which are given in Section (2.6) and the estimator $\hat{\rho}_{Rk}$ (or $\hat{\rho}_k$) of Eidous and Al-Talafheh, which are given by (See Eidous and Al-Talafheh, 2020),

$$\hat{\rho}_{Rk} = \frac{1}{2}\left[\frac{1}{n_1}\sum_{k=1}^{n_1}\left(\frac{\hat{f}_{YR}(x_k)}{\hat{f}_{XR}(x_k)}\right)^{\frac{1}{2}} + \frac{1}{n_2}\sum_{k=1}^{n_2}\left(\frac{\hat{f}_{XR}(y_k)}{\hat{f}_{YR}(y_k)}\right)^{\frac{1}{2}}\right]$$

and

$$\hat{\rho}_k = \frac{1}{2}\left[\frac{1}{n_1}\sum_{k=1}^{n_1}\left(\frac{\hat{f}_Y(x_k)}{\hat{f}_X(x_k)}\right)^{\frac{1}{2}} + \frac{1}{n_2}\sum_{k=1}^{n_2}\left(\frac{\hat{f}_X(y_k)}{\hat{f}_Y(y_k)}\right)^{\frac{1}{2}}\right]$$

We note here that $\hat{\rho}_{Rtrap}$, $\hat{\rho}_{Rsimp}$ and $\hat{\rho}_{Rk}$ are used when the support of simulated distribution is $(0, \infty)$, while $\hat{\rho}_{trap}$, $\hat{\rho}_{simp}$ and $\hat{\rho}_k$ were used when its support is $(-\infty, \infty)$. For all estimators included $\hat{\rho}_k$ and $\hat{\rho}_{Rk}$, the kernel function is taken to be the Gaussian function and the corresponding smoothing parameters are obtained by using the formulas given by Equations (2.1) and (2.2).
The practical implementation of the proposed estimators $\hat{\rho}_1$ and $\hat{\rho}_2$ requires the determination of the transformation function. We suggest to use the exponential transformation if the support of X and Y is $(0, \infty)$. That is,

$$F_Z(x) = 1 - e^{-x}, \;\; 0 < x < \infty \,.$$

In this case, the proposed two estimators $\hat{\rho}_1$ and $\hat{\rho}_2$ will be denoted by $\hat{\rho}_{Rtrap}$ and $\hat{\rho}_{RSimp}$ respectively.

If the support of X and Y is $(-\infty, \infty)$ then we suggest taking $F_Z(x)$ to be the CDF of a generalized Logistic distribution, which is given by,

$$F_Z(x) = 1 - \frac{1}{(1 + e^x)}\,, \quad -\infty < x < \infty\,,$$

Again, in this case, the two estimators $\hat{\rho}_1$ and $\hat{\rho}_2$ will be denoted by $\hat{\rho}_{trap}$ and $\hat{\rho}_{Simp}$ respectively.

For each case of simulations and for each considered estimators, the following statistical properties are computed based on 1000 samples (replication = 1000):
The relative bias (RB),

$$RB = \frac{\hat{E}(estimator) - exact}{exact}.$$

The relative root of mean squared error ($RMSE$),

$$RMSE = \frac{\sqrt{\widehat{MSE}(estimator)}}{exact}$$

and the efficiency (EFF), which is computed with respect to Eidous and Al-Talafhah (2020) estimator,

$$EFF = \frac{\widehat{MSE}(\hat{\rho}_k)}{\widehat{MSE}(proposed\ estimator)}.$$

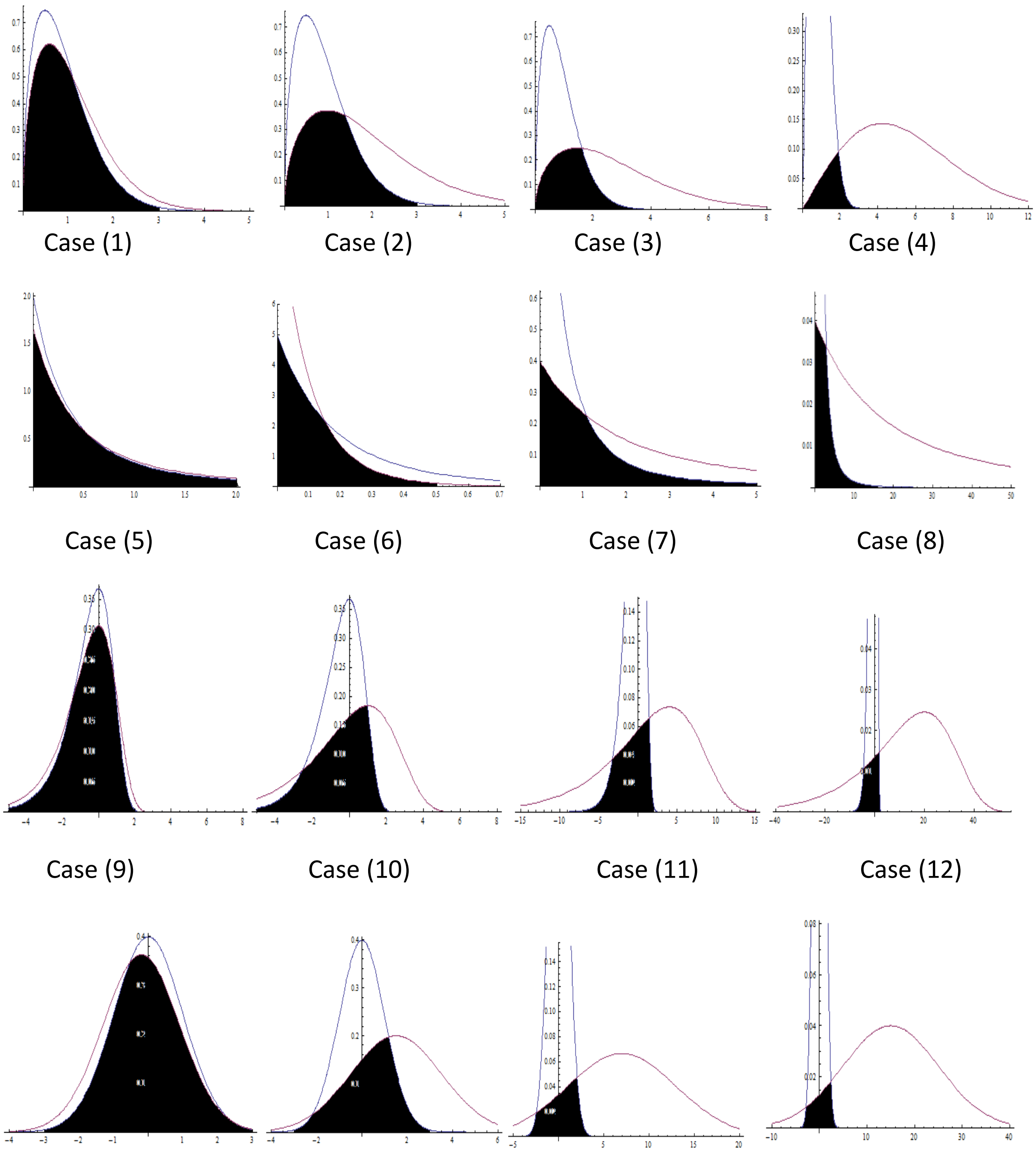
Based on the results of the simulation study shown in Table (2.2), we can summarize the general results as follows:

1. In general, and by considering the values of RB, RMSE, and thus EFF, it is cleer that the three proposed estimates of ρ ($\hat{\rho}_{trap}$, $\hat{\rho}_{simp1}$and $\hat{\rho}_{simp2}$) give very good and reasonable results for all considered cases with different sample sizes.
2. The results of all proposed estimators are satisfactory even in the cases of small sample sizes with different exact values of ρ, where the values of RB and RMSE were small especially in the cases of large exact values of ρ. In addition, the values of RB and RMSE decrease with increase of the sample sizes; this is a good sign that the proposed estimators are consistent.
3. The EFF values are generally very good and were the best when the exact values are large, except for the Case (13) where the generated data followed the Gumbel distribution with exact value of ρ = 0.9925, but scored the smallest value of EFF=0.658 for $\hat{\rho}_{trap}$, as well as the Case (16) for the same distribution recorded the largest value of EFF = 4.82, although the exact values of ρ = 0.2822 for $\widehat{\rho}_{simp2}$, this may be due to properties related to gumbel distribution in particular.
4. All results for RB, RMSE, and EFF in particular were better than the results mentioned for Eidous and Al-Talafheh (2020), which indicates that the estimates proposed in this thesis are more acceptable, especially that the simulation study was conducted on samples of small sizes, and the results were very satisfactory.
5. It was noticed that the values of RB were mostly negative(RB<0), which indicates that the kernal estimates underestimate the true value of the corresponding overlapping coefficient ρ.

Table (1). The simulated pair distributions $f_X(x)$ and $f_Y(y)$ and the corresponding exact values of the overlapping coefficient ρ.

Name of distribution	$f_X(x)$	$f_Y(y)$	ρ
Case(1): Weibull	Weib(1.5, 1)	Weib(1.5, 1.2)	0.9907
Case(2): Weibull	Weib(1.5, 1)	Weib(1.5, 2)	0.8785
Case(3): Weibull	Weib(1.5, 1)	Weib(1.5, 3)	0.7357
Case(4): Weibull	Weib(2, 1)	Weib(2, 6)	0.3243
Case(5): Lomax	Lom(1, 2)	Lom(1.2, 2)	0.9978
Case(6): Lomax	Lom(1, 5)	Lom(1, 10)	0.9428
Case(7): Lomax	Lom(1, 2)	Lom(5, 2)	0.8536
Case(8): Lomax	Lom(1, 2)	Lom(50, 2)	0.4309
Case(9): Gumbel	Gum(0, 1)	Gum(0, 1.2)	0.9925
Case(10): Gumbel	Gum(0, 1)	Gum(1, 2)	0.8412
Case(11): Gumbel	Gum(0, 1)	Gum(4, 5)	0.5541
Case(12): Gumbel	Gum(0, 1)	Gum(20, 15)	0.2822
Case (13): Normal	N(0, 1)	N(-0.2, 1.1)	0.9932
Case (14): Normal	N(0, 1)	N(1.5, 2)	0.7992
Case (15): Normal	N(0, 1)	N(7, 6)	0.4089
Case (16): Normal	N(0, 1)	N(15, 10)	0.2549

Figure (1). The plots of 16 cases of OVL between two distributions $f_X(x)$ and $f_Y(y)$ as given in Table (1).



Case (1) Case (2) Case (3) Case (4)

Case (5) Case (6) Case (7) Case (8)

Case (9) Case (10) Case (11) Case (12)

Case (13) Case (14) Case (15) Case (16)

Table (2). Relative bias (RB), relative root of mean square error (RMSE), and efficiency (EFF) of the different estimators of ρ.

		Case (1) , ρ = 0.9907				Case (2) , ρ = 0.8786		
(n_1, n_2)		$\hat{\rho}_{Rker}$	$\hat{\rho}_{Rtrap}$	$\hat{\rho}_{RSimp}$		$\hat{\rho}_{Rker}$	$\hat{\rho}_{Rtrap}$	$\hat{\rho}_{RSimp}$
(12,12)	RB	-0.1568	-0.1194	-0.1122		-0.1587	-0.1341	-0.1291
	RMSE	0.1830	0.1460	0.1440		0.2127	0.1855	0.1836
	EFF	**1.000**	**1.570**	**1.615**		**1.000**	**1.315**	**1.342**
(36,54)	RB	-0.0578	-0.0424	-0.0412		-0.0731	-0.0625	-0.0607
	RMSE	0.0660	0.0508	0.0498		0.1001	0.0902	0.0890
	EFF	**1.000**	**1.686**	**1.754**		**1.000**	**1.231**	**1.264**
(96,180)	RB	-0.0278	-0.0196	-0.0191		-0.0409	-0.0338	-0.0331
	RMSE	0.0315	0.0236	0.0232		0.0559	0.0500	0.0495
	EFF	**1.000**	**1.771**	**1.834**		**1.000**	**1.249**	**1.272**
		Case (3) , ρ=0.7358				Case (4) , ρ= 0.3243		
(12,12)	RB	-0.1771	-0.1558	-0.1494		-0.2141	-0.0350	-0.0204
	RMSE	0.2603	0.2260	0.2217		0.4523	0.3142	0.3181
	EFF	**1.000**	**1.327**	**1.379**		**1.000**	**2.073**	**2.022**
(36,54)	RB	-0.0879	-0.0723	-0.0698		-0.0997	0.0444	0.0471
	RMSE	0.1301	0.1139	0.1125		0.2316	0.1709	0.1725
	EFF	**1.000**	**1.304**	**1.336**		**1.000**	**1.837**	**1.803**
(96,180)	RB	-0.0505	-0.0392	-0.0383		-0.0644	0.0378	0.0388
	RMSE	0.0755	0.0663	0.0658		0.1354	0.1075	0.1079
	EFF	**1.000**	**1.295**	**1.316**		**1.000**	**1.587**	**1.573**

Continue …

		Case (5) , ρ=0.9978				Case (6) , ρ= 0.9428		
(n_1, n_2)		$\hat{\rho}_{Rker}$	$\hat{\rho}_{Rtrap}$	$\hat{\rho}_{RSimp}$		$\hat{\rho}_{Rker}$	$\hat{\rho}_{Rtrap}$	$\hat{\rho}_{RSimp}$
(12,12)	RB	-0.2148	-0.2071	-0.1937		-0.4287	-0.3877	-0.3747
	RMSE	0.2406	0.2301	0.2197		0.4603	0.4337	0.4263
	EFF	**1.000**	**1.094**	**1.199**		**1.000**	**1.127**	**1.166**
(36,54)	RB	-0.1277	-0.1085	-0.1035		-0.1877	-0.1588	-0.1543
	RMSE	0.1362	0.1158	0.1114		0.2002	0.1731	0.1698
	EFF	**1.000**	**1.383**	**1.493**		**1.000**	**1.338**	**1.391**
(96,180)	RB	-0.0958	-0.0698	-0.0677		-0.0939	-0.0816	-0.0804
	RMSE	0.0990	0.0727	0.0708		0.0991	0.0877	0.0867
	EFF	**1.000**	**1.852**	**1.957**		**1.000**	**1.277**	**1.305**
		Case (7) , ρ=0.8536				Case (8) , ρ=0.4309		
(12,12)	RB	-0.8567	-0.7832	-0.7803		-0.9782	-0.9910	-0.9901
	RMSE	0.8652	0.7975	0.7951		0.9787	0.9917	0.9909
	EFF	**1.000**	**1.177**	**1.184**		**1.000**	**0.974**	**0.975**
(36,54)	RB	-0.8428	-0.7725	-0.7720		-0.9911	-0.9876	-0.9868
	RMSE	0.8472	0.7783	0.7778		0.9912	0.9878	0.9870
	EFF	**1.000**	**1.185**	**1.186**		**1.000**	**1.007**	**1.008**
(96,180)	RB	-0.8250	-0.7671	-0.7670		-0.9970	-0.9944	-0.9945
	RMSE	0.8273	0.7698	0.7697		0.9970	0.9945	0.9945
	EFF	**1.000**	**1.155**	**1.155**		**1.000**	**1.005**	**1.005**

Continue …

		Case (9) , ρ=0.9925				Case (10) , ρ=0.8412		
(n_1, n_2)		$\hat{\rho}_{ker}$	$\hat{\rho}_{trap}$	$\hat{\rho}_{Simp}$		$\hat{\rho}_{ker}$	$\hat{\rho}_{trap}$	$\hat{\rho}_{Simp}$
(12,12)	RB	-0.0998	-0.1349	-0.1161		-0.0827	-0.1517	-0.1282
	RMSE	0.1274	0.1540	0.1385		0.1557	0.1848	0.1674
	EFF	**1.000**	**0.685**	**0.847**		**1.000**	**0.710**	**0.865**
(36,54)	RB	-0.0400	-0.0364	-0.0313		-0.0262	-0.0278	-0.0183
	RMSE	0.0475	0.0440	0.0404		0.0682	0.0599	0.0578
	EFF	**1.000**	**1.169**	**1.386**		**1.000**	**1.296**	**1.393**
(96,180)	RB	-0.0197	-0.0142	-0.0131		-0.0153	0.0004	0.0031
	RMSE	0.0227	0.0177	0.0170		0.0383	0.0319	0.0327
	EFF	**1.000**	**1.648**	**1.784**		**1.000**	**1.437**	**1.370**
		Case (11) , ρ= 0.5541				Case (12) , ρ=0.2822		
(12,12)	RB	-0.0946	-0.1711	-0.1458		-0.1271	-0.1886	-0.1634
	RMSE	0.2615	0.2171	0.2015		0.5235	0.2653	0.2533
	EFF	**1.000**	**1.451**	**1.685**		**1.000**	**3.89**	**4.27**
(36,54)	RB	-0.0372	-0.0370	-0.0257		-0.0534	-0.0432	-0.0318
	RMSE	0.1219	0.0885	0.0867		0.2453	0.1126	0.1118
	EFF	**1.000**	**1.895**	**1.975**		**1.000**	**4.74**	**4.82**
(96,180)	RB	-0.0229	-0.0019	0.0020		-0.0459	-0.0129	-0.0086
	RMSE	0.0695	0.0519	0.0531		0.1384	0.0715	0.0721
	EFF	**1.000**	**1.790**	**1.711**		**1.000**	**3.74**	**3.68**

Continue …

		Case (13) , ρ = 0.9932				Case (14) , ρ = 0.7992		
(n_1, n_2)		$\hat{\rho}_{ker}$	$\hat{\rho}_{trap}$	$\hat{\rho}_{Simp}$		$\hat{\rho}_{ker}$	$\hat{\rho}_{trap}$	$\hat{\rho}_{Simp}$
(12,12)	RB	-0.1137	-0.0914	-0.0826		-0.1169	-0.1109	-0.0918
	RMSE	0.1406	0.1154	0.1107		0.1991	0.1744	0.1667
	EFF	**1.000**	**1.485**	**1.614**		**1.000**	**1.303**	**1.427**
(36,54)	RB	-0.0391	-0.0253	-0.0248		-0.0483	-0.0255	-0.0212
	RMSE	0.0464	0.0332	0.0329		0.0909	0.0747	0.0746
	EFF	**1.000**	**1.948**	**1.982**		**1.000**	**1.481**	**1.487**
(96,180)	RB	-0.0183	-0.0110	-0.0110		-0.0281	-0.0094	-0.0089
	RMSE	0.0216	0.0148	0.0148		0.0527	0.0436	0.0438
	EFF	**1.000**	**2.139**	**2.132**		**1.000**	**1.461**	**1.447**
		Case (15) , ρ = 0.4089				Case (16) , ρ = 0.2549		
(12,12)	RB	-0.1343	-0.1182	-0.0977		-0.1578	-0.0904	-0.0694
	RMSE	0.3870	0.2543	0.2526		0.5694	0.2992	0.3020
	EFF	**1.000**	**2.316**	**2.348**		**1.000**	**3.62**	**3.55**
(36,54)	RB	-0.0623	-0.0157	-0.0104		-0.0714	0.0119	0.0172
	RMSE	0.1713	0.1147	0.1158		0.2714	0.1551	0.1576
	EFF	**1.000**	**2.232**	**2.187**		**1.000**	**3.06**	**2.97**
(96,180)	RB	-0.0390	0.0007	0.0014		0.0348	0.0246	0.0254
	RMSE	0.0998	0.0731	0.0735		0.1509	0.1007	0.1013
	EFF	**1.000**	**1.865**	**1.842**		**1.000**	**2.243**	**2.218**

References

- Alodat, M. T. , Al Fayez, M. and Eidous, O. (2021). On the asymptotic distribution of Matusita's overlapping measure. *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2020.186926.
- Chapra C. C. and Canale. R. P. (2014). Numerical Methods for Engineers. McGraw-Hill Education; 7th edition.
- Eidous, O. and Al-Talafhah, S. (2020). Kernel method for overlapping coefficients estimation. *Communications in Statistics – Simulation and Computation*. DOI: 10.1080/03610918.2020.1757711.
- Inman, H. F., & Bradley Jr, 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*(10), 3851-3874.
- Matusita, K. 1955. Decision rules based on the distance for problem of fir, two samples, and estimation, Ann. Math. Statist.,26,631-640.
- Samawi, H. M, Helu, A. and Vogel, R. (2011). A Nonparametric test of symmetry based on the overlapping coefficient. *Journal of Applied Statistics,* 38(5): 885-898.
- Silverman, B, W. (1986). Density estimation of statistics and data analysis. Chapman and Hall. London.