International Journal of Theoretical and Applied Mathematics
2024, Vol. 10, No. 2, pp. 14-22
https://doi.org/10.11648/j.ijtam.20241002.11

Research Article

On Inference of Weitzman Overlapping Coefficient Δ(X,Y) in the Case of Two Normal Distributions

Omar Mohammad Eidous*, Salam Khaled Daradkeh

Department of Statistics, Faculty of Science, Yarmouk University, Irbid, Jordan

Abstract

The Weitzman overlapping coefficient Δ(X,Y) is the most important and widely used overlapping coefficient, which represents the intersection area between two probability distributions. This research proposes a new general technique to estimate Δ(X,Y) assuming the existence of two independent random samples following normal distributions. In contrast to some studies in this scope that place some restrictions on the parameters of the two populations such as the equality of their means or the equality of their variances, this study did not assume any restrictions on the parameters of normal distributions. Two new estimators for Δ(X,Y) were derived based on the proposed new technique, and then the properties of the estimator resulting from taking their arithmetic mean was studied and compared with some corresponding estimators available in the literature based on the simulation method. An extensive simulation study was performed by assuming two normal distributions with different parameter values to cover most possible cases in practice. The parameter values were chosen taking into account the exact value of Δ(X,Y), which taken to be small (close to zero), medium (close to 0.5) and large (close to 1). The simulation results showed the effectiveness of the proposed technique in estimating Δ(X,Y). By comparing the proposed estimator of Δ(X,Y) with some existing corresponding estimators, its performance was better than the performances of the other estimators in almost all considered cases.

Keywords

Overlapping Weitzman Coefficient, Maximum Likelihood Method, Parametric Method, Normal Distribution, Expected Value, Relative Bias, Relative Mean Square Error

1. Introduction

The well-known Weitzman overlapping (OVL) coefficient $\Delta(X,Y)$ is a measure of similarity between two probability distributions. Although the Weitzman coefficient is the most important interference coefficient, there are some other overlapping coefficients that have been discussed and studied in the literature, see [2, 4-7, 11-14, 17]. $\Delta(X,Y)$ represents the common area under two probability density functions ($pdfs$). The OVL coefficients are widely used in the literature in many applications such as; comparison of income distributions [14]; distinctness clusters [18], reliability analysis [2] and goodness of fit test [1].

Let X and Y be two independent continuous random variables follow $f_1(x)$ and $f_2(x)$ respectively. The Weitzman [19] OVL coefficient between X and Y is defined by,

$$\Delta(X,Y) = \int min\{f_1(x), f_2(x)\}dx.$$

*Corresponding author: omarm@yu.edu.jo (Omar Mohammad Eidous)

Received: 9 July 2024; **Accepted:** 5 August 2024; **Published:** 20 August 2024

There are two main methods used to estimate the OVL coefficients; the parametric method and the non-parametric method. The parametric method assumes that the formulas of the probability density functions are known but depend on unknown parameter(s). To estimate the OVL coefficient, these parameters can be estimated by using one of the well-known statistical methods, such as the method of moments or the maximum likelihood method (see, [15, 6, 3]). If it is not possible to determine the probability density function models for the data or if there are some doubts about the validity of the assumption of a particular data model, the alternative is to use the non-parametric method instead of the parametric method. The non-parametric method does not require any assumptions about the formulas of probability density functions, as this method is used to estimate the formulas of the probability density functions themselves (see, [14, 9, 10]).

If X and Y follow $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$ respectively, where $N(\mu,\sigma^2)$ is the pdf of a normal distribution with mean μ and variance σ^2 then

$$\Delta(X,Y)=\int_{-\infty}^{\infty} min\{N(\mu_1,\sigma_1^2),N(\mu_2,\sigma_2^2)\}dx.$$

Under the assumption, $\sigma_1^2=\sigma_2^2=\sigma^2$, Inman and Bradley [15] derived the value of $\Delta(X,Y)$, which is given by,

$$\Delta(X,Y)=2\Phi\left(-\frac{|\mu_1-\mu_2|}{2\sigma}\right)$$

$$=2\Phi\left(-|\delta|/2\right)$$

where $\delta=(\mu_1-\mu_2)/\sigma$ and $\Phi(t)$ is the standard normal cumulative distribution function at a point t. Let $X_1,X_2,\dots,X_{n_1}$ and $Y_1,Y_2,\dots,Y_{n_2}$ be two independent random samples drawn from the two normal densities $N(\mu_1,\sigma^2)$ and $N(\mu_2,\sigma^2)$ respectively. The maximum-likelihood estimator of $\Delta(X,Y)$ is [15],

$$\hat{\Delta}_{IN}(X,Y)=2\Phi\left(-\frac{|\bar{X}-\bar{Y}|}{2S}\right),$$

where $\bar{X}$ and $\bar{Y}$ are the maximum likelihood (ML) estimators of μ_1 and μ_2 respectively and S^2 is the pooled ML estimator of σ^2, which is given by,

$$S^2=\frac{\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2}{n_1+n_2}.$$

Mulekar and Mishra [16] derived the formula of $\Delta(X,Y)$ under the assumption that the two means are equal, i.e. $\mu_1=\mu_2=\mu$. Now, define $C=\sigma_1/\sigma_2$, then they gave the following formula for $\Delta(X,Y)$,

$$\Delta(X,Y)=\begin{cases}1-2\Phi(b)+2\Phi(Cb), \text{if } 0<C<1\\ 1+2\Phi(b)-2\Phi(Cb), \text{if } C\geq 1\end{cases},$$

where $=\sqrt{-ln(C^2)/(1-C^2)}$. The corresponding estimator of $\Delta(X,Y)$ that suggested by [16] is,

$$\hat{\Delta}_{MM}(X,Y)=\begin{cases}1-2\Phi(\hat{b})+2\Phi(\hat{C}\hat{b}), \text{if } 0<\hat{\sigma}_1<\hat{\sigma}_2\\ 1+2\Phi(\hat{b})-2\Phi(\hat{C}\hat{b}), \text{if } \hat{\sigma}_1\geq\hat{\sigma}_2\end{cases}$$

where, $\hat{b}=\sqrt{-ln(\hat{C}^2)/(1-\hat{C}^2)}$, $\hat{C}=\hat{\sigma}_1/\hat{\sigma}_2$, $\hat{\sigma}_1^2=\frac{\sum_{i=1}^{n_1}(X_i-\hat{\mu})^2}{n_1}$, $\hat{\sigma}_2^2=\frac{\sum_{i=1}^{n_2}(Y_i-\hat{\mu})^2}{n_2}$ and $\hat{\mu}=\frac{\sum_{i=1}^{n_1}X_i+\sum_{i=1}^{n_2}Y_i}{n_1+n_2}$.

It is important to note that the value of b is undefined if $\sigma_1=\sigma_2$ and the corresponding estimator $\hat{b}$ is also undefined if $\hat{\sigma}_1=\hat{\sigma}_2$. In this case, the value of the parameter $\Delta(X,Y)=1$ (because the two densities are identical), and then the value of the corresponding estimator $\hat{\Delta}_{MM}(X,Y)$ must be 1.

The previous two studies placed some restrictions on the parameters of the distributions. The first assumed that the two variances were equal, while the second assumed that the two means were equal. To overcome this problem, Eidous and Al-Shourman [8] estimated $\Delta(X,Y)$ under a pair of normal distributions without using any assumptions on their parameters. Their proposal was based on approximating the integral of $\Delta(X,Y)$ and then the resulting approximation was estimated instead of estimating the exact value of $\Delta(X,Y)$.

The aim and main idea of this paper parallel the work of [8]. Without using any assumptions about the parameters of normal distribution, this paper proposed a new technique to deal with the integral of $\Delta(X,Y)$ by writing it as an expected value for a function or some functions and then estimating the resulting expected value instead of estimating $\Delta(X,Y)$ itself.

2. Main Results

Let X and Y be two random variables from $f_1(x;\mu_1,\sigma_1^2)=N(\mu_1,\sigma_1^2)$ and $f_2(y;\mu_2,\sigma_2^2)=N(\mu_2,\sigma_2^2)$ respectively. In this section, a new technique for estimating overlapping Weitzman coefficient is suggested,

$$\Delta(X,Y)=\int_{-\infty}^{\infty} min\{f_1(x;\mu_1,\sigma_1^2),f_2(x;\mu_2,\sigma_2^2)\}dx.$$

This proposed technique consists of two stages. In the first stage, the coefficient was written as an expected value of some function. In the second stage, a new estimator was proposed for the resulting expectation. Accordingly, the estimator was derived as follows.

Consider $min\{f_1(X;\mu_1,\sigma_1^2),f_2(X;\mu_2,\sigma_2^2)\}/f_1(X;\mu_1,\sigma_1^2)$ as a function of X and $min\{f_1(Y;\mu_1,\sigma_1^2),f_2(Y;\mu_2,\sigma_2^2)\}/f_2(Y;\mu_2,\sigma_2^2)$ as a function of Y. Then,

$$E\left(\frac{min\{f_1(X;\mu_1,\sigma_1^2),f_2(X;\mu_2,\sigma_2^2)\}}{f_1(X;\mu_1,\sigma_1^2)}\right) = \int_0^\infty \frac{min\{f_1(x;\mu_1,\sigma_1^2),f_2(x;\mu_2,\sigma_2^2)\}}{f_1(x;\mu_1,\sigma_1^2)} f_1(x;\mu_1,\sigma_1^2)\,dx$$

$$= \int_{-\infty}^{\infty} min\{f_1(x;\mu_1,\sigma_1^2), f_2(x;\mu_2,\sigma_2^2)\}dx$$

$$= \Delta(X,Y)$$

and

$$E\left(\frac{min\{f_1(Y;\mu_1,\sigma_1^2),f_2(Y;\mu_2,\sigma_2^2)\}}{f_2(Y;\mu_2,\sigma_2^2)}\right) = \int_0^\infty \frac{min\{f_1(y;\mu_1,\sigma_1^2),f_2(y;\mu_2,\sigma_2^2)\}}{f_2(y;\mu_2,\sigma_2^2)} f_2(y;\mu_2,\sigma_2^2)\,dy$$

$$= \int_{-\infty}^{\infty} min\{f_1(y;\mu_1,\sigma_1^2), f_2(y;\mu_2,\sigma_2^2)\}dy$$

$$= \int_{-\infty}^{\infty} min\{f_1(x;\mu_1,\sigma_1^2), f_2(x;\mu_2,\sigma_2^2)\}dx$$

$$= \Delta(X,Y).$$

From the last two formulas, $\Delta(X,Y)$ can also be expressed as the average of the above two formulas as follows,

$$\Delta(X,Y) = \frac{1}{2}\left[E\left(\frac{min\{f_1(X;\mu_1,\sigma_1^2),f_2(X;\mu_2,\sigma_2^2)\}}{f_1(X;\mu_1,\sigma_1^2)}\right) + E\left(\frac{min\{f_1(Y;\mu_1,\sigma_1^2),f_2(Y;\mu_2,\sigma_2^2)\}}{f_2(Y;\mu_2,\sigma_2^2)}\right)\right].$$

Based on the last three formulas for $\Delta(X,Y)$ and based on the two independent random samples $X_1, X_2, \dots, X_{n_1}$ and $Y_1, Y_2, \dots, Y_{n_2}$, the ML estimators of μ_1, μ_2, $\sigma_1^2$and σ_2^2 are $\hat{\mu}_1 = \bar{X}$, $\hat{\mu}_2 = \bar{Y}$, $\hat{\sigma}_1^2 = S_1^2$ and $\hat{\sigma}_2^2 = S_2^2$ respectively. Therefore, the ML estimators for $f_1(x;\mu_1,\sigma_1^2)$ and $f_2(y;\mu_2,\sigma_2^2)$ are $f_1(x;\hat{\mu}_1,\hat{\sigma}_1^2)$ and $f_2(y;\,\hat{\mu}_2,\hat{\sigma}_2^2)$ respectively. Now, $\Delta(X,Y)$ can be estimated using any of the following two estimators that are consistent with the first two formulas of $\Delta(X,Y)$,

$$\hat{\Delta}(X,Y) = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{min\{f_1(X_i;\hat{\mu}_1,\hat{\sigma}_1^2),f_2(X_i;\hat{\mu}_2,\hat{\sigma}_2^2)\}}{f_1(X_i;\hat{\mu}_1,\hat{\sigma}_1^2)}\right)$$

or,

$$\hat{\Delta}(X,Y) = \frac{1}{n_2}\sum_{i=1}^{n_2}\left(\frac{min\{f_1(Y_i;\hat{\mu}_1,\hat{\sigma}_1^2),f_2(Y_i;\hat{\mu}_2,\hat{\sigma}_2^2)\}}{f_2(Y_i;\hat{\mu}_2,\hat{\sigma}_2^2)}\right).$$

After conducting a preliminary simulation study, this study shows that the average of the last two estimators for $\Delta(X,Y)$ (corresponding to the last formula of $\Delta(X,Y)$) is more stable than each of them individually. Therefore, the finite properties of the following proposed estimator is investigated in our simulation study in the next section,

$$\hat{\Delta}_{Prop}(X,Y) = \frac{1}{2}\left[\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{min\{f_1(X_i;\hat{\mu}_1,\hat{\sigma}_1^2),f_2(X_i;\hat{\mu}_2,\hat{\sigma}_2^2)\}}{f_1(X_i;\hat{\mu}_1,\hat{\sigma}_1^2)}\right) + \frac{1}{n_2}\sum_{i=1}^{n_2}\left(\frac{min\{f_1(Y_i;\hat{\mu}_1,\hat{\sigma}_1^2),f_2(Y_i;\hat{\mu}_2,\hat{\sigma}_2^2)\}}{f_2(Y_i;\hat{\mu}_2,\hat{\sigma}_2^2)}\right)\right].$$

3. Simulation Study

In this section, a simulation study was conducted to investigate the performance of the proposed estimator of $\Delta(X,Y)$ compared to some estimators found in the literature under pair of normal distributions. One of the general estimators that was taken into consideration in this study is the non-parametric kernel estimator developed by [9], which we will denote it by $\hat{\Delta}_k(X,Y)$ (see also [12] for the selection of bandwidth).

We considered the following three cases at which four pairs of distributions were chosen to simulate the data for each case. The basis of the selection process for these distributions is to ensure small, medium (less than 0.5), medium (greater than 0.5) and large values for the true values of $\Delta(X,Y)$. The $3 \times 4 = 12$ pairs are given in Table 1.

Four pairs of normal distributions with equal variances are selected (See Table 1). In this case, the estimators $\hat{\Delta}_k(X,Y)$, $\hat{\Delta}_{IN}(X,Y)$ and $\hat{\Delta}_{Prop}(X,Y)$ are considered and their performances were compared.

Four pairs of normal distributions with equal means are selected (See Table 1). The estimators $\hat{\Delta}_k(X,Y)$, $\hat{\Delta}_{MM}(X,Y)$ and $\hat{\Delta}_{Prop}(X,Y)$ are considered in this case.

Four pairs of normal distributions with different variances and different means are selected (See Table 1). In this case, only the two estimators $\hat{\Delta}_k(X,Y)$ and $\hat{\Delta}_{Prop}(X,Y)$ are investigated.

It should be noted here that, on the first hand, the estimator $\hat{\Delta}_{IN}(X,Y)$ was developed assuming that the variances are equal, while $\hat{\Delta}_{MM}(X,Y)$ was derived assuming that the means are equal. On the other hand, the estimator $\hat{\Delta}_k(X,Y)$ was developed without using any assumptions on the parameters of the distributions or even on the shape of the distributions themselves. Finally, the proposed estimator $\hat{\Delta}_{Prop}(X,Y)$ was derived assuming that the two distributions are normal but without using any assumptions on their parameters. Therefore, $\hat{\Delta}_k(X,Y)$ and $\hat{\Delta}_{Prop}(X,Y)$ can be used for all three cases mentioned above, while the estimator $\hat{\Delta}_{IN}(X,Y)$ can be used in the first case only, and the estimator $\hat{\Delta}_{IN}(X,Y)$ can only be used in the second case.

Let $x_1, x_2, \dots, x_{n_1}$ and $y_1, y_2, \dots, y_{n_2}$ are two independent simulated samples from $f_1(x)$ and $f_2(y)$ respectively, then to study the behavior of the various estimators for different sample sizes, $(n_1, n_2) = (10, 10), (50, 50), (100, 200)$ are chosen. For each sample size, $R = 1000$ replications are used. The Relative Bias (RB), Relative Mean Square Error (RMSE) and Efficiency (EFF) were computed for each estimator under study. These measures were computed according to the following rules.

If $\hat{\eta}$ is an estimator of η then,

$$RB = \frac{\hat{E}(\hat{\eta})-\eta}{\eta},$$

$$RMSE = \frac{\sqrt{\widehat{MSE}(\hat{\eta})}}{\eta},$$

Where

$$\hat{E}(\hat{\eta}) = \frac{\sum_{j=1}^{R}\hat{\eta}_{(j)}}{R}$$

and

$$\widehat{MSE}(\hat{\eta}) = \sum_{j=1}^{R}\left(\hat{\eta}_{(j)} - \hat{E}(\hat{\eta})\right)^2 / R.$$

The efficiency of each considered estimator is computed with respect to the nonparametric estimator, $\hat{\Delta}_k(X,Y)$. The efficiency of $\hat{\eta}$ with respect to $\hat{\Delta}_k(X,Y)$ is,

$$EFF = \frac{\widehat{MSE}(\hat{\Delta}_k(X,Y))}{\widehat{MSE}\ (\hat{\eta})}.$$

The simulation results are reported in Tables 2 to 4.

Table 1. *The 12 simulated pair normal distributions $f_1(x)$ and $f_2(y)$ together with the corresponding exact values of the overlapping coefficient $\Delta(X,Y)$.*

	Normal distributions	$\boldsymbol{f_1(x)}$	$\boldsymbol{f_2(y)}$	$\boldsymbol{\Delta(X,Y)}$
Case 1: Equal variances	A	$N(0,1)$	$N(-0.5,1)$	0.8025
	B	$N(0,1)$	$N(1,1)$	0.671
	C	$N(0,1)$	$N(1.5,1)$	0.4532
	D	$N(0,1)$	$N(3,1)$	0.1336
Case 2: Equal means	A	$N(0,1)$	$N(0,1.5)$	0.8064
	B	$N(0,1)$	$N(0,2.5)$	0.585
	C	$N(0,1)$	$N(0,5)$	0.3528
	D	$N(0,1)$	$N(0,10)$	0.2017
Case 3: Different means and different variances	A	$N(0,1)$	$N(-0.2,1.1)$	0.9151
	B	$N(0,1)$	$N(1,2)$	0.6099
	C	$N(0,1)$	$N(2.5,4)$	0.3577
	D	$N(0,1)$	$N(5,2)$	0.0891

4. Simulation Results

The RB, RMSE and EFF for each estimator mentioned in the previous sections are displayed in Table 2, Table 3 and Table 4. By examining these results, we conclude the following:

As a general conclusion, it is clear that the RMSE values of the different estimators decrease with increasing sample sizes. This a good sign for concluding that the different estimators are consistent estimators for $\Delta(X,Y)$.

By examining the results of Table 2, which concern the case of two normal distributions with equal variances, it is clear that the two estimators $\hat{\Delta}_{IN}(X,Y)$ and $\hat{\Delta}_{prop}(X,Y)$ perform similar to each other with some preferring for $\hat{\Delta}_{IN}(X,Y)$ over $\hat{\Delta}_{prop}(X,Y)$. These two estimators perform better than the general kernel estimator $\hat{\Delta}_k(X,Y)$.

From Table 3, the performance of the estimator $\hat{\Delta}_{MM}(X,Y)$ is better than the other two counterpart estimators $\hat{\Delta}_k(X,Y)$ and $\hat{\Delta}_{prop}(X,Y)$ when the exact values of overlapping coefficient are large. The opposite is true for small values of $\Delta(X,Y)$. This may be due to the estimation of the common mean by using the pooled mean as suggested by [16].

In all cases, the performance of the proposed estimator $\hat{\Delta}_{prop}(X,Y)$ is better than that of $\hat{\Delta}_k(X,Y)$. It is worthwhile to mention here that the estimator of [16] can only be used when the mean of the two normal distributions is assumed to be equal.

Based on the results of Table 4, the proposed estimator $\hat{\Delta}_{prop}(X,Y)$ achieves good performance in general. Its performance is better than that of the kernel estimator $\hat{\Delta}_k(X,Y)$

in all cases studies. This is very evident when examining the values RMSE and EFF that associated with the proposed and kernel estimators. It is important to note that this result is expected because the kernel estimator can be used more generally without looking to the distributions at which the data come from. It should also be noted here that the proposed estimator was derived without using the assumption of equal means or equal variances for the two normal distributions.

Finally, it should be noted that the Inman and Bradley [15] estimator $\hat{\Delta}_{IN}(X,Y)$ (or Mulekar and Mishra [16] estimator $\hat{\Delta}_{MM}(X,Y)$) can be used only if the two variances (or the two means) of the normal distributions are equal. Because of this drawback of each of them, we recommend using the proposed estimator $\hat{\Delta}_{prop}(X,Y)$ as a general estimator for $\Delta(X,Y)$ under a pair of normal distributions.

Table 2. *The RB, RMSE and EFF of the estimators $\hat{\Delta}_k(X,Y)$, $\hat{\Delta}_{IN}(X,Y)$ and $\hat{\Delta}_{Prop}(X,Y)$when the data are simulated from pair normal distributions with equal variances ($\sigma_1^2 = \sigma_2^2 = 1$) (Case, 1 of Table 1).*

$\Delta(X,Y)$	(n_1,n_2)		$\hat{\Delta}_k(X,Y)$	$\hat{\Delta}_{IN}(X,Y)$	$\hat{\Delta}_{Prop}(X,Y)$
0.8025	(10,10)	RB	0.0364	-0.033	-0.0995
		RMSE	0.246	0.1951	0.2037
		EFF	1	1.5899	1.459
	(50,50)	RB	-0.004	-0.0038	-0.0164
		RMSE	0.0914	0.0973	0.095
		EFF	1	0.8821	0.9255
	(100,200)	RB	0.0018	0.002	-0.0022
		RMSE	0.0615	0.0606	0.0601
		EFF	1	1.031	1.049
0.617	(10,10)	RB	-0.0024	-0.0008	-0.0484
		RMSE	0.2803	0.2784	0.268
		EFF	1	1.0141	1.094
	(50,50)	RB	0.002	0.0029	-0.0051
		RMSE	0.1342	0.1297	0.1286
		EFF	1	1.07	1.0894
	(100,200)	RB	-0.0012	-0.0004	-0.0031
		RMSE	0.0812	0.0726	0.0726
		EFF	1	1.2507	1.2501
0.4532	(10,10)	RB	0.0009	-0.0036	-0.041
		RMSE	0.3567	0.3461	0.3368
		EFF	1	1.0619	1.1214
	(50,50)	RB	0.0029	0.0011	-0.0049
		RMSE	0.1698	0.1565	0.1556
		EFF	1	1.1759	1.1899
	(100,200)	RB	-0.0004	-0.001	-0.0032
		RMSE	0.1061	0.089	0.0902
		EFF	1	1.4206	1.3833
0.1336	(10,10)	RB	0.0607	0.0166	-0.0229
		RMSE	0.7837	0.6801	0.6807

$\Delta(X,Y)$	(n_1,n_2)		$\hat{\Delta}_k(X,Y)$	$\hat{\Delta}_{IN}(X,Y)$	$\hat{\Delta}_{Prop}(X,Y)$
		EFF	1	1.328	1.3257
		RB	0.0295	0.0213	0.0158
	(50,50)	RMSE	0.3642	0.3073	0.3062
		EFF	1	1.4052	1.4148
		RB	0.0009	-0.0043	-0.0057
	(100,200)	RMSE	0.2093	0.1738	0.1768
		EFF	1	1.4494	1.4006

Table 3. *The RB, RMSE and EFF of the estimators $\hat{\Delta}_k(X,Y)$, $\hat{\Delta}_{MM}(X,Y)$ and $\hat{\Delta}_{Prop}(X,Y)$ when the data are simulated from pair normal distributions with equal means $(\mu_1 = \mu_2 = 0)$ (Case, 2 of Table 1).*

$\Delta(X,Y)$	(n_1,n_2)		$\hat{\Delta}_k(X,Y)$	$\hat{\Delta}_{MM}(X,Y)$	$\hat{\Delta}_{Prop}(X,Y)$
		RB	-0.1686	0.0014	-0.0093
	(10,10)	RMSE	-0.2447	0.1481	0.1606
		EFF	1	2.7332	2.3214
		RB	-0.042	0.0014	-0.0007
0.8064	(50,50)	RMSE	0.0954	0.0825	0.0844
		EFF	1	1.3367	1.2793
		RB	-0.01	0.0001	0.0008
	(100,200)	RMSE	0.0537	0.0463	0.047
		EFF	1	1.3505	1.3077
		RB	-0.1213	0.0688	0.0152
	(10,10)	RMSE	0.297	0.2353	0.2523
		EFF	1	1.5925	1.386
		RB	-0.0274	0.0083	-0.004
0.5850	(50,50)	RMSE	0.1266	0.0985	0.1074
		EFF	1	1.653	1.3906
		RB	-0.0029	0.0053	0.0029
	(100,200)	RMSE	0.0723	0.0593	0.0609
		EFF	1	1.4848	1.4057
		RB	-0.1349	0.2267	0.0105
	(10,10)	RMSE	0.3798	0.4077	0.3296
		EFF	1	0.8684	1.3284
		RB	-0.03	0.0536	0.0032
0.3528	(50,50)	RMSE	0.1613	0.1278	0.1323
		EFF	1	1.5931	1.4858
		RB	-0.013	0.0187	0.0014
	(100,200)	RMSE	0.0949	0.0754	0.0781

$\Delta(X,Y)$	(n_1,n_2)		$\hat{\Delta}_k(X,Y)$	$\hat{\Delta}_{MM}(X,Y)$	$\hat{\Delta}_{Prop}(X,Y)$
		EFF	1	1.5846	1.4756
		RB	-0.187	0.5953	-0.0112
	(10,10)	RMSE	0.4733	0.8444	0.4167
		EFF	1	0.3141	1.2901
		RB	-0.0524	0.1694	0.0012
0.2017	(50,50)	RMSE	0.2116	0.278	0.1732
		EFF	1	0.5795	1.4938
		RB	-0.0325	0.0833	0.0028
	(100,200)	RMSE	0.1226	0.1533	0.0983
		EFF	1	0.6402	1.5566

Table 4. *The RB, RMSE and EFF of the estimators $\hat{\Delta}_k(X,Y)$ and $\hat{\Delta}_{Prop}(X,Y)$ when the data are simulated from pair normal distributions with different parameters (Case, 3 of Table 1).*

$\Delta(X,Y)$	(n_1,n_2)		$\hat{\Delta}_k(X,Y)$	$\hat{\Delta}_{Prop}(X,Y)$
		RB	−0.2384	−0.1537
	(10,10)	RMSE	0.2801	0.2029
		EFF	1.0000	1.9056
		RB	−0.0793	−0.0404
0.9151	(50,50)	RMSE	0.1020	0.0777
		EFF	1.0000	1.722
		RB	−0.0359	−0.0132
	(100,200)	RMSE	0.0551	0.048
		EFF	1.0000	1.3183
		RB	−0.1117	−0.0682
	(10,10)	RMSE	0.2813	0.2452
		EFF	1.0000	1.316
		RB	−0.0281	−0.0192
0.6099	(50,50)	RMSE	0.1209	0.1037
		EFF	1.0000	1.3596
		RB	−0.0041	−0.0063
	(100,200)	RMSE	0.069	0.0598
		EFF	1.0000	1.3324
		RB	−0.1141	−0.0594
	(10,10)	RMSE	0.3804	0.3234
0.3577		EFF	1.0000	1.3833
		RB	−0.0353	−0.0170
	(50,50)	RMSE	0.1756	0.1438

$\Delta(X,Y)$	(n_1,n_2)		$\hat{\Delta}_k(X,Y)$	$\hat{\Delta}_{Prop}(X,Y)$
		EFF	1.0000	1.4914
		RB	−0.0064	−0.0023
	(100,200)	RMSE	0.0944	0.0791
		EFF	1.0000	1.4263
		RB	−0.0098	−0.0464
	(10,10)	RMSE	0.8987	0.8097
		EFF	1.0000	1.2317
		RB	0.0671	−0.0014
0.0891	(50,50)	RMSE	0.4347	0.3660
		EFF	1.0000	1.4105
		RB	0.0578	0.0118
	(100,200)	RMSE	0.2547	0.2103
		EFF	1.0000	1.4662

5. Conclusion

This study presented a new technique for estimating $\Delta(X,Y)$ under a pair of normal distributions by writing it as an expected value for some functions. One of the most important benefits of this technique is to estimate $\Delta(X,Y)$ without placing any conditions on the parameters of normal distributions. Based on the results of numerical simulations, these results demonstrated the effectiveness of the new technique and that the performance of the estimator resulting from the use of this technique is better than the performance of the nonparametric kernel estimator of $\Delta(X,Y)$ that developed by Eidous and AL-Talafha [9]. Accordingly, this technique can be used to estimate other OVL coefficients mentioned in the literature, such as the Matusita coefficient (see, Eidous and Ananbeh [11]) and Pianka and Kullback-Leibler coefficients (see, Eidous and Abu Al-Hayja`a [5]).

Abbreviations

OVL	Overlapping
pdf	Probability Density Function
ML	Maximum Likelihood
RB	Relative Bias
MSE	Mean Square Error
RMSE	Relative Mean Square Error
EFF	Efficiency

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] 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*. 51(20), 6963-6977. https://doi.org/10.1080/03610926.2020.186926

[2] Dhaker, H., Deme, E. and El-Adlouni, S. (2021). On inference of overlapping coefficients in two inverse Lomax populations. *Statistical Theory and Applications*, 20 (1), 61-75. https://doi.org/10.2991/jsta.d.210107.002

[3] Eidous, O., and Abu Al-Hayja's, M. (2023a). Numerical integration approximations to estimate the Weitzman overlapping measure: Weibull distributions. *Yugoslav Journal of Operations Research*, 33 (4), 699-712. http://dx.doi.org/10.2298/YJOR221215021E

[4] Eidous, O. M. and Abu Al-Hayja,a, M. (2023b). Estimation of overlapping measures using numerical approximations: Weibull distributions. *Jordan Journal of Mathematics and Statistics* (JJMS), 16(4), 741–761. https://doi.org/10.47013/16.4.8

[5] Eidous, O. and Abu Al-Hayja'a, M. (2023c). Weibull distributions for estimation of Pianka and Kullback-Leibler overlapping measures. *Journal of Mathematics and Statistics Research*, 5 (1), 165. https://www.researchgate.net/publication/376681761

[6] Eidous, O., and Al-Daradkeh, S. (2022). Estimation of Matusita overlapping coefficient for pair normal distributions. *Jordan Journal of Mathematics and Statistics* (JJMS), 15(4B), 1137-1151. https://doi.org/10.47013/15.4.23

[7] Eidous, O. M., Al-Shourman, A. (2022). Numerical Integral Approximation to Estimate Matusita Overlapping Coefficient for Normal Distributions. *Journal of Mathematical Techniques and Computational Mathematics*, 1 (3), 264-270. https://api.semanticscholar.org/CorpusID:255699518

[8] Eidous, O. M. and Al-Shourman, A. (2023). Estimating the Weitzman Overlapping Coefficient Using Integral Approximation Method in the Case of Normal Distributions. *Applied Mathematics, Modeling and Computer Simulation*, 42, 1011-1020. https://doi.org/10.54647/wjm5071001

[9] Eidous, O. M., and AL-Talafha, S. A. (2020). Kernel method for overlapping coefficients estimation. *Communications in Statistics: Simulation and Computation*, 51(9), 5139–5156. https://doi.org/10.1080/03610918.2020.1757711

[10] Eidous, O. M., and Ananbeh, E. A. (2024a). Kernel method for estimating overlapping coefficient using numerical integration methods. *Applied Mathematics and Computation*, 462, 128339. https://doi.org/10.1016/j.amc.2023.128339

[11] Eidous, O. M., and Ananbeh, E. A. (2024b). Kernel method for estimating Matusita overlapping coefficient using numerical approximations. *Annals of Data Science.* https://doi.org/10.1007/s40745-024-00563-y

[12] Eidous, O. M, Marie, M. and Al-Haj Ibrahim, M. (2010). A comparative study for bandwidth selection in kernel density estimation. *Journal of Modern Applied Statistical Methods*, 9 (1), 263-273. https://doi.org/10.22237/jmasm/1272687900

[13] Ichikawa, M. (1993). A meaning of the overlap pedarea under probability density curves of stress and strength. *Reliability Engineering and System Safety*, 41(2), 203-204. https://doi.org/10.1016/0951-8320(93)90033-U

[14] Inácio V. and Garrido Guillén J. E. (2022). Bayesian nonparametric inference for the overlap coefficient: With an application to disease diagnosis. *Statistics in Medicine*, 41(20), 3879-3898. https://doi.org/10.1002/sim.9480

[15] Inman, H. F. and Bradley, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Comm. Statist.- Theory Methods*, 18, 3851-3874. https://doi.org/10.1080/03610928908830127

[16] Mulekar, M. S. and Mishra, S. N. (1994). Overlap Coefficient of two normal densities: equal means case. *J. Japan Statist. Soc.*, *24*, 169-180. https://www.jstage.jst.go.jp/article/jjss1970/24/2/24_2_169/_pdf

[17] Samawi, H. M. and Al-Saleh, M. F. (2008). Inference of overlapping coefficients in two exponential populations using ranked set sample. *Communication of Korean of Statistical Society*, 15 (2): 147-159.

[18] 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. *Journal of the International Association for Mathematical Geology*, 9, 123-143. https://doi.org/10.1007/BF02312508

[19] Weitzman, M. S. (1970). Measures of overlap of income distributions of white and Negro families in the United States. Washington: U.S. Bureau of the Census. https://babel.hathitrust.org/cgi/pt?id=mdp.39015085502204&seq=17