Math. Model. Nat. Phenom. 20 (2025) 9
https://doi.org/10.1051/mmnp/2025008

Mathematical Modelling of Natural Phenomena
www.mmnp-journal.org

GENERALIZATION OF THE OVERLAPPING COEFFICIENT FOR $k \geq 2$ NORMAL DISTRIBUTIONS

Omar M. Eidous* and Majd M. Alsheyyab

Abstract. The overlapping coefficients are the measures of homogeneity or closeness between two distributions. There are several coefficients in the literature including the Weitzman coefficient Δ, which represents the area of intersection between two distributions. In this paper, the Weitzman coefficient is generalized for $k \geq 2$ distributions and then several estimators for Δ were developed assuming k normal distributions. The derivation of the new estimators was done by adopting a new general approach, which relies on modeling the overlapping coefficient as the expectation of some functions, and then the resulting expectation is estimated by using the moments method. The behavior of the resulting estimators has been investigated and compared by Monte-Carlo simulation technique. Simulation results indicated the good properties of the resulting estimators.

Mathematics Subject Classification. 62E17.

Received November 4, 2024. Accepted February 13, 2025.

1. Introduction

Referring to the literature related to the overlapping (OVL) coefficients, we find that there are three well-known coefficients, which are the Matusita coefficient ρ, Morisita coefficient λ and Weitzman coefficient Δ (see [1]). These coefficients are defined as the degree of homogeneity or similarity between two statistical distributions. The Weitzman coefficient is the most obvious and widely used among these coefficients, as it is clearly defined as the common or the intersection area between two probability density functions (*pdf*). Weitzman [2] introduced a way of measuring this overlap mathematically, which can be especially useful when studying species interactions and biodiversity conservation. Assuming that $f_1(x)$ and $f_2(x)$ are two continuous *pdfs*, the Weitzman coefficient is given by the following formula,

$$\Delta_2 = \int min\{f_1(x), f_2(x)\}dx \tag{1.1}$$

The extra subscript 2 in Δ_2 indicates the number of densities used in calculating the Weitzman coefficient. The possible values of Δ_2 lie between 0 and 1. It is 1 if $f_1(x)$ and $f_2(x)$ are identical and 0 if they are completely different and have no common support. As [4] point out, if $f_1(x)$ and $f_2(x)$ are discrete probability density functions, the above integral is replaced by summation. The Weitzman overlapping coefficient is increasingly

Keywords and phrases: Overlaping coefficients, normal distribution, estimation, expectation, weitzman coefficient, relative bias, relative mean square error.

Department of Statistics, Yarmouk University, Irbid, Jordan.

* Corresponding author: omarm@yu.edu.jo

employed in the study of natural phenomena as it provides a robust method to quantify similarity and overlap between two probability distributions, which is crucial when dealing with uncertain or complex systems. For instance, in pharmacy studies, it can offer valuable insight to compare between different drugs in terms of their molecular structures, mechanisms of action, or side effects. This can lead to better personalized medicine and optimized drug therapies, where patients may be on multiple medications, and understanding overlap can help prevent adverse drug interactions and improve safety [5]. In income, the overlapping coefficient allows for better cross-country and cross-group comparisons, revealing hidden disparities across demographics such as age, gender, or ethnicity. This provides a more robust and flexible framework for addressing the multifaceted nature of income inequality [6]. In addition to pharmacy (see also, [7]) and income [8], the OVL coefficients have many applications in different fields, including, law [9], photography [10] and goodness-of-fit test [11, 12].

Several studies have been conducted in the past thirty years to estimate Δ_2 using parametric method, assuming two independent random samples from a pair of statistical distributions. Inman and Bradly [4] considered the case of normal distributions with equal variance. Reiser and Faraggi [13] constructed and investigated the confidence intervals for Δ_2 under normal distributions with equal variances. The estimation case of Δ_2 for two normal distributions with equal means is considered by [14]. Under the same distributions with equal means, [15] compared the confidence intervals of Δ_2 obtained by using re-sampling different technique, named, Jacknife, bootstrap and transformation methods. Wang and Tian [16] proposed a generalized inference approach and a parametric bootstrapping method to construct confidence interval of Δ_2 under the assumption of normality. To get rid of the assumptions used about the parameters of the normal distributions, [17] and [18] adopted two different techniques to estimate Δ_2 without using any assumptions about the means equality or equality of the variances for normal distributions. Al-Saidy *et al.* [19] considered the case of two Weibull distributions with equal shape parameters to estimate Δ_2, while [20] considered the same problem without using any assumptions on the shape parameters. Chaubey *et al.* [21] studied the case of two inverse Gaussian distributions with equal means. Under different settings and distributions, [22] derived point and confidence intervals estimators for Δ_2. Dhaker *et al.* [23] interested the case of two inverse Lomax distributions to estimate Δ_2 under these pair of distributions. Inác and Guillén [24] developed a Bayesian nonparametric approach based on Dirichlet process mixtures of additive normal models for estimating Δ_2. Finally, when it is difficult to determine the shape of the distribution for a given data set, the nonparametric kernel method can be used to estimate the OVL coefficients, which has also received attention from some researchers (see [25], [8], [1], [26] and [27]). Now, Let $f_i(x),\ i = 1, 2, \ldots,\ k \geq 2$ are continuous *pdfs*, then the common area between these k densities can be represented by generalizing the Weitzmann coefficient Δ_2 in 1.1 as follows,

$$\Delta_k = \int min\{f_1(x), f_2(x), ..., f_k(x)\}dx \tag{1.2}$$

The subscript k in Δ_k means that the number of densities used in the above integral is k. If $f_i(x),\ i = 1, 2, \ldots, k$ are identical then it is obvious from 1.2 that $\Delta_k = 1$. Also, if we assume that there is no common support between at least two densities (completely different) then $\Delta_k = 0$.

The main objective of this paper is to estimate the generalized Δ_k assuming k independent random samples each following a normal distribution. Throughout this paper it is assumed that $k \geq 2$.

2. Formulation of the main problem associated with $\Delta_k,\ k > 2$

Let $X_j,\ j = 1, 2, \ldots, k$ be k independent normal random variables with mean μ_j and variance σ_j^2. That is, the *pdf* of X_j is,

$$f_j(x; \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{\frac{-(x-\mu_j)^2}{2\sigma_j^2}},\ \ -\infty < x < \infty.$$

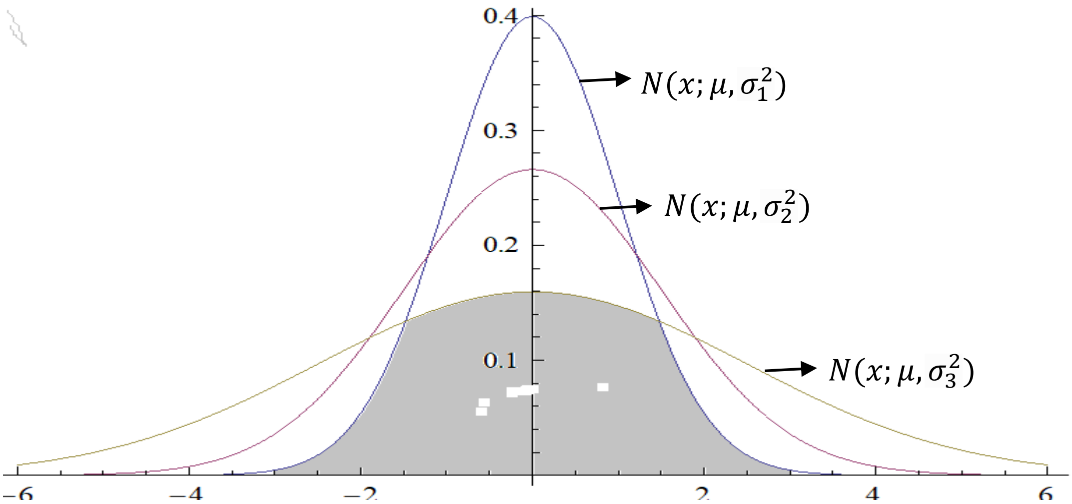

FIGURE 1. The shaded area represents Δ_3 for three normal distributions with equal means and different variances such that $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$.

The coefficient Δ_k becomes,

$$\Delta_k = \int_{-\infty}^{\infty} min\{f_1(x, \mu_1, \sigma_1^2), f_2(x, \mu_2, \sigma_2^2), ..., f_k(x, \mu_k, \sigma_k^2)\}dx \tag{2.1}$$

It is not simple task to deal with the last integral and then to compute Δ_k. To calculate the above integral in 2.1, we need to determine the intersection point(s) between any two functions of $f_j(x, \mu_j, \sigma_j^2),\ j = 1, 2, .., k$ and then study the sub-intervals of the entire interval $(-\infty, \infty)$ to determine the smallest $f_j(x, \mu_j, \sigma_j^2)$ for each sub-interval. To be more clearer, let

$$f_j(x, \mu_j, \sigma_j^2) = N_j(x; \mu_j, \sigma_j^2),\ j = 1, 2, \ldots, k.$$

If $k = 3$ then we consider the following three cases:

Case 1: Assume that $\mu_1 = \mu_2 = \mu_3 = (\mu, say)$ and consider the plot of Δ_3 as given in Figure 1. For $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$, the coefficient Δ_3 can be computed as follows,

$$\begin{aligned}
\Delta_3 &= \int_{-\infty}^{\infty} min\{N(x; \mu, \sigma_1^2), N(x; \mu, \sigma_2^2), N(x; \mu, \sigma_3^2)\}dx \\
&= \int_{-\infty}^{\infty} min\{N(x; \mu, \sigma_1^2), N(x; \mu, \sigma_3^2)\}dx \\
&= \int_{-\infty}^{a} N(x; \mu, \sigma_1^2)dx + \int_{a}^{b} N(x; \mu, \sigma_3^2)dx + \int_{b}^{\infty} N(x; \mu, \sigma_1^2)dx \\
&= \Phi\left(\frac{a-\mu}{\sigma_1}\right) + \Phi\left(\frac{b-\mu}{\sigma_3}\right) - \Phi\left(\frac{a-\mu}{\sigma_3}\right) + 1 - \Phi\left(\frac{b-\mu}{\sigma_1}\right),
\end{aligned}$$

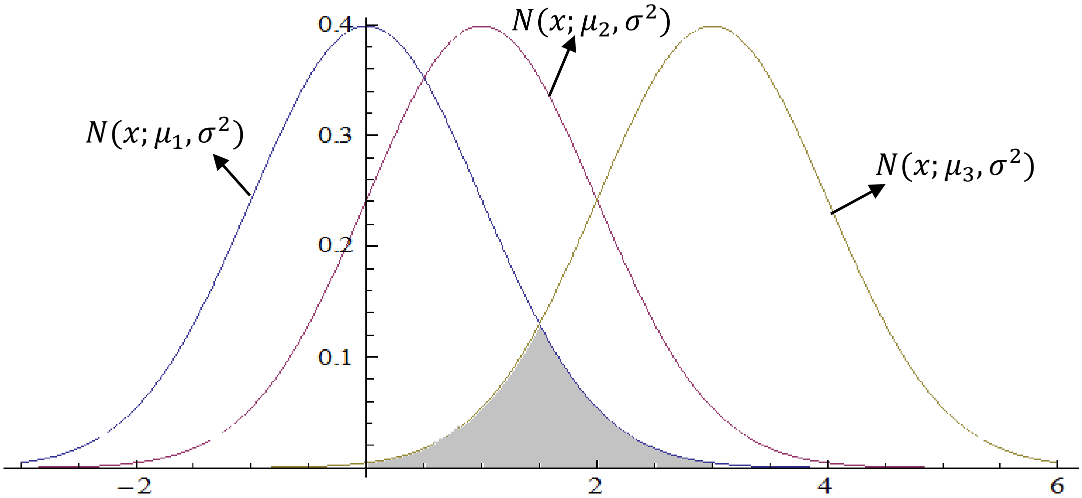

FIGURE 2. The shaded area represents Δ_3 for three normal distributions with equal variances and different means such that $\mu_1 < \mu_2 < \mu_3$.

where a and b are the intersection points between $N(x;\mu,\sigma_1^2)$ and $N(x;\mu,\sigma_3^2)$ and $\Phi(.)$ is the standard normal cumulative distribution function. Let $\sigma_{(1)}^2 \leq \sigma_{(2)}^2 \leq ... \leq \sigma_{(k)}^2$ be the order values of $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$, in the same way and under the assumptions $\mu_1 = \mu_2 = ... = \mu_k = (\mu, say)$, the coefficient Δ_k, $k \geq 2$ is computed as follows,

$$\Delta_k = \int_{-\infty}^{a} N(x;\mu,\sigma_{(1)}^2)dx + \int_{a}^{b} N(x;\mu,\sigma_{(k)}^2)dx + \int_{b}^{\infty} N(x;\mu,\sigma_{(1)}^2)dx$$
$$= \Phi\left(\frac{a-\mu}{\sigma_{(1)}}\right) + \Phi\left(\frac{b-\mu}{\sigma_{(k)}}\right) - \Phi\left(\frac{a-\mu}{\sigma_{(k)}}\right) + 1 - \Phi\left(\frac{b-\mu}{\sigma_{(1)}}\right)$$

The two quantities a and b are now the intersection points between $N(x;\mu,\sigma_{(1)}^2)$ and $N(x;\mu,\sigma_{(k)}^2)$

Case 2: Assume that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = (\sigma^2, say)$ and consider the plot of Δ_3 as given in Figure 2 below. For $\mu_1 < \mu_2 < \mu_3$, the coefficient Δ_3 can be computed as follows,

$$\Delta_3 = \int_{-\infty}^{\infty} min\{N(x;\mu_1,\sigma^2), N(x;\mu_2,\sigma^2), N(x;\mu_3,\sigma^2)\}dx$$
$$= \int_{-\infty}^{\infty} min\{N(x;\mu_1,\sigma^2), N(x;\mu_3,\sigma^2)\}dx$$
$$= \int_{-\infty}^{c} N(x;\mu_3,\sigma^2)dx + \int_{c}^{\infty} N(x;\mu_1,\sigma^2)dx$$
$$= \Phi\left(\frac{c-\mu_3}{\sigma}\right) + 1 - \Phi\left(\frac{c-\mu_1}{\sigma}\right).$$

where c is the intersection point between $N(x;\mu_1,\sigma^2)$ and $N(x;\mu_3,\sigma^2)$. Let $\mu_{(1)} \leq \mu_{(2)} \leq ... \leq \mu_{(k)}$ be the order values of $\mu_1, \mu_2, ..., \mu_k$, then under the assumptions $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = (\sigma^2, say)$, the

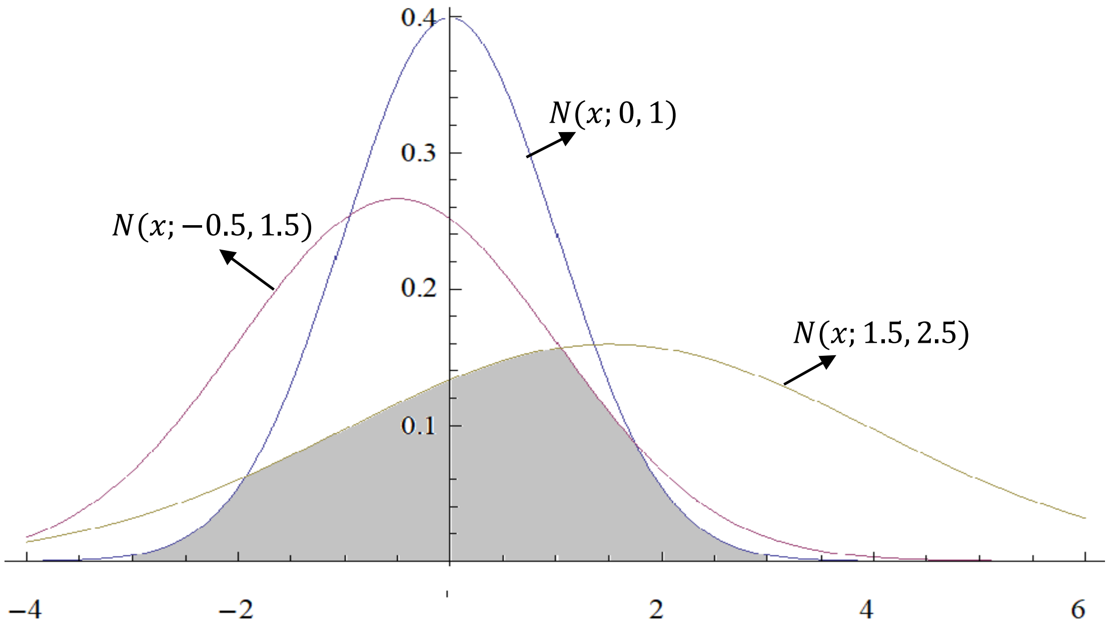

FIGURE 3. The shaded area represents Δ_3 for three normal distributions with unequal means and unequal variances.

coefficient $\Delta_k,\ k \geq 2$ is computed as follows,

$$\begin{aligned}\Delta_k &= \int_{-\infty}^{c} N(x;\mu_{(k)},\sigma^2)dx + \int_{c}^{\infty} N(x;\mu_{(1)},\sigma^2)dx \\ &= \Phi\left(\frac{c-\mu_{(k)}}{\sigma}\right) + 1 - \Phi\left(\frac{c-\mu_{(1)}}{\sigma}\right).\end{aligned}$$

The quantity c is now the intersection point between $N(x;\mu_{(1)},\sigma^2)$ and $N(x;\mu_{(k)},\sigma^2)$.

Case 3: As have been shown in Case (1) and Case (2), under the assumption of equal means or equal variances, the problem of calculating $\Delta_k,\ k \geq 2$ is reduced to be the same as of calculating Δ_2. However, if there is no assumptions on the means and variances then the problem becomes different. It is more complicated as shown in Figure 3, which plots three normal distributions for specific values of their parameters as given on the same plot. Based on Graph (3), the coefficient Δ_3 can be computed as follows,

$$\begin{aligned}\Delta_3 &= \int_{-\infty}^{\infty} min\{N(x;0,1), N(x;-0.5,1.5), N(x;1.5,2.5)\}dx \\ &= \int_{-\infty}^{a_1} N(x;0,1)dx + \int_{a_1}^{a_2} N(x;1.5,2.5)dx + \int_{a_2}^{a_3} N(x;-0.5,1.5)dx + \int_{a_3}^{\infty} N(x;0,1)dx \\ &= \Phi(a_1) + \Phi\left(\frac{a_2-1.5}{\sqrt{2.5}}\right) - \Phi\left(\frac{a_1-1.5}{\sqrt{2.5}}\right) + \Phi\left(\frac{a_3+0.5}{\sqrt{1.5}}\right) - \Phi\left(\frac{a_2+0.5}{\sqrt{1.5}}\right) + 1 - \Phi(a_3).\end{aligned}$$

where a_1 is the first intersection point between $N(x;0,1)$ and $N(x;1.5,2.5)$, a_2 is the second intersection point between $N(x;-0.5,1.5)$ and $N(x;1.5,2.5)$ and, a_3 is the second intersection point between $N(x;0,1)$ and $N(x;-0.5,1.5)$. For the general case where $k \geq 2$ and the values of the unequal means and unequal variances, it is not easy to determine the values of the required intersections and accordingly the integral in the formula of Δ_k cannot be computed. Therefore, solving such a problem requires unconventional thinking. In the next two sections of this paper, some ideas are presented through which we can perform the inference process, specifically estimating Δ_k for any $k \geq 2$ and whether or not there are restrictions on the equality of means or variances of the normal distributions under study.

3. Modelling Δ_k as expected value

For simplicity, let $f_j(x)$ be a normal *pdf* with mean μ_j and variance σ_j^2 $(i,e, f_j(x) = N(x;\mu_j,\sigma_j^2))$ and let $X_j \sim f_j(x)$. The coefficient, $\Delta_k = \int_{\infty}^{\infty} min\{f_1(x), f_2(x), \ldots, f_k(x)\}dx$ can be expressed in various ways as an expected value of some functions as given below. Define

$$\psi_j(X_j) = \frac{min\{f_1(X_j), f_2(X_j), ..., f_k(X_j)\}}{f_j(X_j)}, \ j = 1, 2, ..., k$$

then

$$\begin{aligned} E\psi_j(X_j) &= \int_{-\infty}^{\infty} \psi_j(x_j)f_j(x_j)dx_j \ , \ j = 1, 2, ..., k \\ &= \int_{-\infty}^{\infty} \psi_j(x)f_j(x)dx \\ &= \int_{-\infty}^{\infty} min\{f_1(x), f_2(x), ..., f_k(x)\}dx \\ &= \Delta_k. \end{aligned}$$

Therefore,

$$\Delta_k = E\psi_1(X_1) = E\psi_2(X_2) = ... = E\psi_k(X_k).$$

Also, we can rewrite Δ_k as the average of any two of the above functions, the average of any three, the average of any four functions, and so on. In the end, we can rewrite Δ_k as the average of all these k functions, which is given by,

$$\Delta_k = \frac{1}{k}(E\psi_1(X_1) + E\psi_2(X_2) + ... + E\psi_k(X_k)).$$

Notice that the number of possible ways to rewrite Δ_k in terms of $E\psi_j(X_j)$, $j = 1, 2, \ldots, k$ is,

$$\binom{k}{1} + \binom{k}{2} + ... + \binom{k}{k} = 2^k - 1,$$

which means we can derive $2^k - 1$ corresponding estimators for Δ_k.

4. Estimation of Δ_k

To formulate the present problem, let $X_{i1}, X_{i2}, \ldots, X_{in_i}$ be a random sample of size n_i from $f_i(x) = N(x;\mu_i,\sigma_i^2)$, $i = 1, 2, \ldots, k$, where all parameters are assumed to be unknown. The k random samples are assumed to be independent. Define

$$\overline{X}_i = \sum_{j=1}^{n_i} \frac{X_{ij}}{n_i},\ S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2,\ \hat{f}_i(x) = N(\overline{X}_i, \frac{S_i^2}{n_i}),\ i = 1, 2, \ldots, k.$$

In addition, define $\overline{X}_{(i)}$ and $S_{(i)}^2$ to be the corresponding estimators for $\mu_{(i)}$ and $\sigma_{(i)}^2$ respectively with sample size $n_{(i)}$. For $k = 2$ and under the two normal distributions, we consider the following estimators of Δ_2:

a) If $\sigma_1^2 = \sigma_2^2 (= \sigma^2, say)$, [4] derived the exact value of Δ_2 and then gave its maximum likelihood (ML) estimator, which is,

$$\hat{\Delta}_2 = 2\Phi\left(-\frac{|\overline{X}_1 - \overline{X}_2|}{2S}\right),$$

where $S^2 = \frac{(S_1^2 + S_2^2)}{(n_1 + n_2)}$. If $k \geq 2$ and with equal variances for the k normal distributions (see Case 2 in the previous section), we can directly derive the ML estimator of Δ_k, which is given by,

$$\hat{\Delta}_k = 2\Phi\left(-\frac{|\overline{X}_{(1)} - \overline{X}_{(k)}|}{2\delta}\right),$$

where $\delta^2 = \frac{(S_{(1)}^2 + S_{(k)}^2)}{(n_{(1)} + n_{(k)})}$.

b) If $\mu_1 = \mu_2 (= \mu, say)$ then the exact value of Δ_2 was derived by [14], who gave the following ML estimator of Δ_2,

$$\hat{\Delta}_2 = \begin{cases} 1 - 2\Phi(\hat{b}) + 2\Phi(\hat{C}\hat{b}) & \text{if } 0 < \hat{C} < 1 \\ 1 + 2\Phi(\hat{b}) - 2\Phi(\hat{C}\hat{b}) & \text{if } \hat{C} \geq 1 \end{cases}$$

where $\hat{b} = \sqrt{\frac{-\ln(\hat{C}^2)}{(1-\hat{C}^2)}}$, $\hat{C} = \frac{\hat{\sigma}_1}{\hat{\sigma}_2}$, $\hat{\sigma}_1^2 = \sum_{j=1}^{n_1} \frac{(X_{1j} - \hat{\mu})^2}{n_1}$, $\hat{\sigma}_2^2 = \sum_{j=1}^{n_2} \frac{(X_{2j} - \hat{\mu})^2}{n_2}$, and $\hat{\mu} = \frac{(n_1\overline{X}_1 + n_2\overline{X}_2)}{(n_1 + n_2)}$.

For $k \geq 2$ and with equal means for the k normal distributions (see Case 1 in the previous section), it can be directly derived the ML estimator of Δ_k, which is given by,

$$\hat{\Delta}_k = \begin{cases} 1 - 2\Phi(\hat{a}) + 2\Phi(\hat{D}\hat{a}) & \text{if } 0 < \hat{D} < 1 \\ 1 + 2\Phi(\hat{a}) - 2\Phi(\hat{D}\hat{a}) & \text{if } \hat{D} \geq 1 \end{cases}$$

where

$$\hat{a} = \sqrt{\frac{-\ln(\hat{D}^2)}{(1 - \hat{D}^2)}},\ \hat{D} = \frac{\hat{\delta_{(1)}}}{\hat{\delta_{(k)}}},\ \hat{\delta}_{(1)}^2 = \sum_{j=1}^{n_{(1)}} \frac{(X_{(1)j} - \hat{a})^2}{n_{(1)}},$$

$$\hat{\delta}_{(k)}^2 = \sum_{j=1}^{n_{(k)}} \frac{(X_{(k)j} - \hat{\mu})^2}{n_{(2)}} \text{ and } \hat{a} = \frac{(n_{(1)}\overline{X}_{(1)} + n_{(2)}\overline{X}_{(2)})}{(n_{(1)} + n_{(2)})}$$

The symbol $X_{(i)j}$ represents the observed value obtained from the corresponding distribution with variance $\sigma_{(i)}^2$, $i = 1, k$.

c) Without using any restrictions or assumptions on the equality of the location parameters or the equality of scale parameters of normal distributions, [18] developed an estimator for Δ_2 where they first expressed the coefficient Δ_2 as expected value for some functions and then gave the following estimator for Δ_2,

$$\hat{\Delta}_2 = \frac{1}{2n_1}\sum_{j=1}^{n_1}\frac{min\{\hat{f}_1(X_{1j}), \hat{f}_2(X_{1j})\}}{\hat{f}_1(X_{1j})} + \frac{1}{2n_2}\sum_{j=1}^{n_2}\frac{min\{\hat{f}_1(X_{2j}), \hat{f}_2(X_{2j})\}}{\hat{f}_2(X_{2j})},$$

where

$$\hat{f}_i(X_{ij}) = \frac{1}{\sqrt{2\pi\frac{S_i^2}{n_i}}}e^{\frac{-n_i(X_{ij}-\bar{X}_i)^2}{2S_i^2}}, \ i = 1, 2.$$

In order to estimate Δ_k based on the k independent samples we have to remember that we can write Δ_k as the expected value of some function as explained in Section 4. That is,

$$\Delta_k = E\psi_1(X_1) = E\psi_2(X_2) = ... = E\psi_k(X_k).$$

The estimation of the parameter Δ_k can be achieved by individually estimating each expectation $\psi_j(X_j)$, $j = 1, 2, \ldots, k$. Specifically, the estimation of $E\psi_j(X_j)$ relies on its corresponding random sample $X_{j1}, X_{j2}, \ldots, X_{jn_j}$. Thus, the core of the estimation process involves estimating these expectations indivitually based on the respective random sample. Since Δ_k is defined as the mean of $\psi_j(X_j)$ (or equivalently, the mean of

$$\frac{min\{f_1(X_j), f_2(X_j), \ldots, f_k(X_j)\}}{f_j(X_j)},$$

we propose the following approach to estimate $E\psi_j(X_j)$ (thereby providing an estimator of Δ_k),

$$\hat{\Delta}_k^{(i)} = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{min\{\hat{f}_1(X_{ij}), \hat{f}_2(X_{ij}), ..., \hat{f}_k(X_{ij})\}}{\hat{f}_i(X_{ij})}, \ i = 1, 2, ..., k.$$

The last formula yields $\binom{k}{1} = k$ estimators for Δ_k. Additionally, these estimators can be combined in various ways to generate new ones. For instance, there are $\binom{k}{2} = \frac{k(k-1)}{2}$ possible estimators formed by taking the pairwise average of any two estimators. Denoting these pairwise averages as $\hat{\Delta}_k^{(p,q)}$, each of these represents a new estimator for Δ_k. Specifically, $\hat{\Delta}_k^{(p,q)}$ is obtained by averaging the estimators $\hat{\Delta}_k^{(p)}$ and $\hat{\Delta}_k^{(q)}$, where, $1 \le p < q \le k$, *e.g.*,

$$\hat{\Delta}_k^{(p,q)} = \frac{1}{2}(\hat{\Delta}_k^{(p)} + \hat{\Delta}_k^{(q)})$$

This process can be continued until the final proposed estimator is the average of all k original estimators, expressed as,

$$\hat{\Delta}_k^{(aver)} = \frac{1}{k}\sum_{i=1}^{k}\hat{\Delta}_k^{(i)}$$

This is equivalent to,

$$\hat{\Delta}_k^{(aver)} = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{1}{n_i}\frac{min\{\hat{f}_1(X_{ij}), \hat{f}_2(X_{ij}), ..., \hat{f}_k(X_{ij})\}}{\hat{f}_i(X_{ij})}$$

If $k = 2$ then $\hat{\Delta}_k^{(aver)}$ is precisely the estimator $\hat{\Delta}_2$, as defined by [18].

In the case of $k = 3$ as used in the simulation study in the next section, there are 7 distinct proposed estimators for Δ_3, calculated as $2^3 - 1 = 7$. The first three estimators are $\hat{\Delta}_3^{(1)}$, $\hat{\Delta}_3^{(2)}$ and $\hat{\Delta}_3^{(3)}$, which correspond to the three parameters $E\psi_1(X_1)$, $E\psi_2(X_2)$ and $E\psi_3(X_3)$, respectively. Next, there are three pairwise average estimators,

$$\hat{\Delta}_3^{(1,2)} = \frac{1}{2}(\hat{\Delta}_3^{(1)} + \hat{\Delta}_3^{(2)}),\ \hat{\Delta}_3^{(1,3)} = \frac{1}{2}(\hat{\Delta}_3^{(1)} + \hat{\Delta}_3^{(3)}),\ \hat{\Delta}_3^{(2,3)} = \frac{1}{2}(\hat{\Delta}_3^{(2)} + \hat{\Delta}_3^{(3)})$$

The seventh and last estimator is the average of all individual estimators:

$$\hat{\Delta}_3^{(aver)} = \frac{1}{3}(\hat{\Delta}_3^{(1)} + \hat{\Delta}_3^{(2)} + \hat{\Delta}_3^{(3)})$$

It is crucial to note that for $k \geq 2$, the aforementioned estimators can be utilized regardless of whether the equality condition for the location parameters (or scale parameters) holds. In other words, these estimators remain valid even if the condition $\mu_1 = \mu_2 = ... = \mu_k$ (or $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$) is not satisfied.

5. Simulation study and results

There are many choices resulting from the many possible values for choosing the value of k and the number of estimators associated with this choice, along with the many other options for choosing the parameters of the associated distributions in addition to choosing the sample sizes. Therefore, in this section, we conducted a limited simulation study where we chose the value of $k = 3$ and with it we chose to study the behavior of four estimators out of the 7 resulting estimators, which are the estimators $\hat{\Delta}_3^{(1)}$, $\hat{\Delta}_3^{(1,2)}$, $\hat{\Delta}_3^{(2,3)}$ and $\hat{\Delta}_3^{(aver)}$ (see Sect. 5).

The data was generated according to four scenarios, in each of which 3 normal distributions were chosen. In the first scenario, three distributions with equal means were chosen so that they give a large value for Δ_3. In the second scenario, 3 normal distributions with equal variances were chosen so that they give a value around 0.5 for Δ_3. For the last two scenarios, 3 distributions were chosen in each of them with unequal means and also unequal variances so that the first case gives a value for Δ_3 close to 0.5 while the other gives a small value for Δ_3. These four scenarios along with the exact value of Δ_3 for each of them are as follows:

$$S1 : N(x; 0, 0.95), N(x; 0, 1) \text{ and } N(x; 0, 1.1) \text{ with exact } \Delta_3 = 0.929$$
$$S2 : N(x; -0.1, 1), N(x; 0, 1) \text{ and } N(x; 0.1, 1) \text{ with exact } \Delta_3 = 0.689$$
$$S3 : N(x; -0.5, 1), N(x; 0, 0.5) \text{ and } N(x; 0.75, 1) \text{ with exact } \Delta_3 = 0.469$$
$$S4 : N(x; -1, 1.5), N(x; 0, 0.8) \text{ and } N(x; 2, 0.4) \text{ with exact } \Delta_3 = 0.074.$$

The sample sizes were taken to be $(n_1, n_2, n_3) = (50, 50, 50), (100, 150, 200)$. The empirical measures; Average (AV), Relative Bias (RB) and Relative Root Mean Square Error (RRMSE) were computed for each of the 4 selected estimators based on 1000 replications. The rules of these empirical measures are as follows:

TABLE 1. The AV, RB and RMSE of the estimators $\hat{\Delta}_3^{(1)}$, $\hat{\Delta}_3^{(1,2)}$, $\hat{\Delta}_3^{(2,3)}$ and $\hat{\Delta}_3^{(aver)}$.

Scenario	(n_1, n_2, n_3)		$\hat{\Delta}_3^{(1)}$	$\hat{\Delta}_3^{(1,2)}$	$\hat{\Delta}_3^{(2,3)}$	$\hat{\Delta}_3^{(aver)}$
$S1$ Δ_3=0.929	(50, 50, 50)	AV	0.8492	0.8495	0.8499	0.8497
		RB	–0.0860	–0.0857	–0.0854	–0.0856
		RRMSE	0.1033	0.1029	0.1026	0.1027
	(100, 150, 200)	AV	0.9041	0.9042	0.9044	0.9043
		RB	–0.0270	–0.0269	–0.0267	–0.0268
		RRMSE	0.0416	0.0415	0.0414	0.0414
$S2$ Δ_3=0.689	(50, 50, 50)	AV	0.6774	0.6770	0.6771	0.6772
		RB	–0.0171	–0.0176	–0.0175	–0.0174
		RRMSE	0.1084	0.1078	0.1079	0.1074
	(100, 150, 200)	AV	0.6865	0.6865	0.6865	0.6865
		RB	–0.0039	–0.0038	–0.0038	–0.0038
		RRMSE	0.0570	0.0563	0.0563	0.0563
$S3$ Δ_3=0.469	(50, 50, 50)	AV	0.4560	0.4548	0.4555	0.4557
		RB	–0.0274	–0.0299	–0.0283	–0.0280
		RRMSE	0.1289	0.1182	0.1180	0.1166
	(100, 150, 200)	AV	0.4655	0.4650	0.4647	0.4650
		RB	–0.0069	–0.0080	–0.0086	–0.0081
		RRMSE	0.0646	0.0579	0.0571	0.0571
$S4$ Δ_3=0.074	(50, 50, 50)	AV	0.0677	0.0671	0.0663	0.0668
		RB	–0.0782	–0.0863	–0.0967	–0.0906
		RRMSE	0.4186	0.3376	0.3126	0.3151
	(100, 150, 200)	AV	0.0726	0.0723	0.0721	0.0722
		RB	–0.0112	–0.0148	–0.0186	–0.0161
		RRMSE	0.2142	0.1599	0.1457	0.1467

Let $\hat{\Delta}$ be the estimator of Δ_3 and let $\hat{\Delta}_{(i)}$ be the value of $\hat{\Delta}$ based on iteration number i, where $i = 1, 2, ..., 1000$, then,

$$AV(\hat{\Delta}) = \sum_{i=1}^{1000} \frac{\hat{\Delta}_{(i)}}{1000},$$

$$RB(\hat{\Delta}) = \frac{(AV(\hat{\Delta}) - \Delta_3)}{\Delta_3},$$

and

$$RRMSE(\hat{\Delta}) = \frac{\sqrt{\sum_{i=1}^{1000} \frac{(\hat{\Delta}_{(i)} - \Delta_3)^2}{1000}}}{\Delta_3}.$$

All computations were performed using Mathematica, Version 7 and presented in Table 1. Based on the results of Table 1, we can draw the following general conclusions:

1. The simulation results clearly show that the RRMSE value of each estimator decreases as the sample sizes increase. This is a good sign that each of these estimators fulfills the important property of a good estimator, which is the “consistency property”
2. The relative bias values associated with all estimators are negative, indicating that these estimators, on average, give values less than (underestimate) the exact value of Δ_3 for all considered scenarious.

3. By comparing the RRMSE values of the different estimators, the results clearly and surprisingly show that their performance is very close, especially when the exact value of Δ_3 is large towards 1 and for all sample sizes. Their performance starts to differ slightly for small values of Δ_3.
4. The results show that all the proposed estimators perform well to estimate Δ_3. This is evident whether the data are simulated from normal distributions with some constraints on their parameters (such as distributions with equal means, *i.e.* scenario $S1$ or distributions with equal variances, *i.e.* scenario 4) or from distributions that do not assume any constraints on their parameters (such as scenarios $S3$ and $S4$). We conclude that using the proposed estimators, it does not matter whether there are constraints on the parameters of the distributions or not to obtain good results.
5. Finally, although the performance of all the proposed estimators is good and satisfactory, and at the same time their performance is close to each other, we recommend using the estimator $\hat{\Delta}_3^{(aver)}$ as a general estimator for Δ_3 due to its stable performance and non-volatile from one case to another.

6. Discussion and conclusion

Although the technique discussed in Section 3 focuses primarily on normal distributions, it is important to note that it can be applied to a wide variety of parametric distributions beyond just the normal. The key is to accurately estimate the parameters of the distribution of interest using a good point estimation method, such as the maximum likelihood (ML) method. This highlights the broader applicability of the technique presented in this paper. As mentioned in the previous section, the relative biases for the proposed estimators were negative for all considered cases with $k = 3$. As a reviewer pointed out, this systematic bias does not appear in [18] numerical examples (for $k = 2$), even though most values of the relative bias in that study were negative, with those that were nonnegative being very small. We believe that conducting additional simulation studies for different values of k, along with using bias correction methods (see for instance, [28] and [3]) could help further improve the method discussed in this study. This topic can be addressed as future work.

Conflicts of interest

The authors state that there is no conflict of interest

Data availability statement

No new data/codes were created or analyzed in this study.

References

[1] O. Eidous and S. Al-Talafha, Kernel method for overlapping coefficients estimation. *Commun. Statist. Simul. Computat.* **51** (2022) 5139–5156.

[2] M.S. Weitzman, Measures of overlap of income distributions of white and negro families in the United States. Technical Report 22, US Department of Commerce (1970).

[3] Q. Xu, X. Wang, J. Yi and Y. Wang, Bias correction in species distribution models based on geographic and environmental characteristics. *Ecol. Inform.* **61** (2024) 1–14.

[4] H.F. Inman and E.L. Bradley Jr, The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun. Statist. Theory Methods* **18** (1989) 3851–3874.

[5] J. Wolff, G. Hefner, C. Normann, K. Kaier, H. Binder, C. Hiemke, S. Toto, K. Domschchke, M. Marschollek and A. Klimke, Polypharmacy and the risk of drug-drug interactions and potentially inappropriate medications in hospital psychiatry. *Pharmacoepidemiol. Drug Saf.* **30** (2021) 1258–1268.

[6] S.P. Jenkins and P. Van Kerm, The Measurement of Economic Inequality. The Oxford Handbook of Economic Inequality (2012). https://doi.org/10.1093/oxfordhb/9780199606061.013.0003

[7] S. Mizuno, T. Yamaguchi, A. Fukushima, Y. Matsuyama and Y. Ohashi, Overlap coefficient for assessing the similarity of pharmacokinetic data between ethnically different populations. *Clin. Trials* **2** (2005) 174–181.

[8] F. Schmid and A. Schmidt, Nonparametric estimation of the coefficient of overlapping—theory and empirical application. *Computat. Statist. Data Anal.* **50** (2006) 1583–1596.

[9] J.A. Ekstrom and G.T. Lau, Exploratory text mining of ocean law to measure overlapping agency and jurisdictional authority. *Proceedings of the 2008 international conference on Digital government research*, 53–62.

[10] M.S. Ridout and M. Linkie, Estimating overlap of daily activity patterns from camera trap data. *J. Agric. Biol. Environ. Statist.* **14** (2009) 322–337.

[11] T. Alodat, M. Al Fayez and O. Eidous, On the asymptotic distribution of Matusita's overlapping measure. *Commun. Statist. Theory Methods* **51** (2022) 6963–6977.

[12] G. Núñez-Antonio, M. Mendoza, A. Contreras-Cristán, E. Gutiérrez-Peña and E. Mendoza, Bayesian nonparametric inference for the overlap of daily animal activity patterns. *Environ. Ecol. Statist.* **25** (2018) 471–494.

[13] B. Reiser and D. Faraggi, Confidence intervals for the overlapping coefficient: the normal equal variance case. *J. Roy. Statist. Soc. Ser. D (tatistician)* **48** (1999) 413–418.

[14] M.S. Mulekar and S.N. Mishra, Overlap coefficient of two normal densities: equal means case. *J. Japan Statist. Soc.* **24** (1994) 169–180 (Japanese issue).

[15] M.S. Mulekar and S.N. Mishra, Confidence interval estimation of overlap: equal means case. *Computat. Statist. Data Anal.* **34** (2000) 121–137.

[16] D. Wang and L. Tian, Parametric methods for confidence interval estimation of overlap coefficients. *Computat. Statist. Data Anal.* **106** (2017) 12–26.

[17] O. Eidous and A. Alshorman, Estimating the overlapping coefficient in the case of normal distributions. *World J. Math.* **1** (2023) 1–13.

[18] O. Eidous and S. Daradkeh, On inference of Weitzman overlapping coefficient $\Delta(X, Y)$ in the case for two normal distributions. *Int. J. Theor. Appl. Math.* **10** (2024) 14–22.

[19] O. Al-Saidy, H.M. Samawi and M.F. Al-Saleh, Inference on overlap coefficients under the Weibull distribution: equal shape parameter. *ESAIM: Probab. Statist.* **9** (2005) 206–219.

[20] O. Eidous and M. Abu Al-hayja'a, Numerical integration approximations to estimate the Weitzman overlapping measure: Weibull distributions. *Yugoslav J. Oper. Res.* **33** (2023) 699–712.

[21] Y.P. Chaubey, D. Sen and S.N. Mishra, Inference on overlap for two inverse Gaussian populations: equal means case. *Commun. Statist. Theory Methods* **37** (2008) 1880–1894.

[22] J.A. Montoya, G.P. Figueroa and D. González-Sánchez, Statistical inference for the Weitzman overlapping coefficient in a family of distributions. *Appl. Math. Model.* **71** (2019) 558–568.

[23] H. Dhaker, E. Deme and S. El-Adlouni, On inference of overlapping coefficients in two inverse Lomax populations. *Statist. Theory Appl.* (2021) 61–75.

[24] V. Inácio and J.E.G. Guillén, Bayesian nonparametric inference for the overlap coefficient: with an application to disease diagnosis. *Statist. Med.* **41** (2022) 3879–3898.

[25] T.E. Clemons, Erratum to "A nonparametric measure of the overlapping coefficient". *Computat. Statist. Data Anal.* **36** (2001) 243.

[26] O. Eidous and E. Ananbeh, Kernel method for estimating matusita overlapping coefficient using numerical approximations. *Ann. Data Sci.* (2024) 1–19.

[27] O. Eidous and E. Ananbeh, Kernel method for estimating overlapping coefficient using numerical integration methods. *Appl. Math. Computat.* **462** (2024) 1–10.

[28] O. Eidous, Bias correction for histogram estimator using line transect sampling. *Environmetrics* **16** (2005) 61–69.