

Article

Estimation of Pianka Overlapping Coefficient for Two Exponential Distributions

Suad Alhihi ^{1,†}  and Maalee Almheidat ^{2,*,†} ¹ Department of Mathematics, Al-Balqa Applied University, Alsalt 19117, Jordan; suad.hihi@bau.edu.jo² Department of Mathematics, University of Petra, Amman 11196, Jordan

* Correspondence: malmheidat@uop.edu.jo

† These authors contributed equally to this work.

Abstract: Overlapping coefficients (OVL) are commonly used to estimate the similarity between populations in terms of their density functions. In this paper, we consider Pianka's overlap coefficient for two exponential populations. The methods for statistical inference of Pianka's coefficient are presented. The bias and mean square error (MSE) of the maximum likelihood estimator (MLE) and the Bayes estimator of Pianka's overlap coefficient are investigated by simulation. Confidence intervals for Pianka's overlap measure are constructed.

Keywords: overlapping measure; maximum likelihood estimator; limiting distribution; posterior distribution; Bayes estimator; multivariate delta method; exponential distribution

MSC: 62F10; 62F12

1. Introduction

Overlapping coefficients (OVL) are measures of how similar two populations are; this similarity is a function that assigns a real number between 0 and 1, where a value of zero indicates that the distributions are completely different and a value of one indicates that they are identical. There are many overlapping coefficients in the literature, including measures of overlap that determine the percentage of area that the two distributions have in common [1]. Gini and Livada [2] first introduced the idea of overlapping in 1943. Matusita's coefficient [3] was introduced to calculate the significant distance between two probability density functions, and has applications in several practical areas, including reliability analysis and clinical research [4,5]. Matusita developed a discrete version known as the Freeman–Tukey (FT) measure, which is related to the Hellinger distance [6,7] and the delta method [8]. The Chi-Squared measure [9] and Hellinger measure [10] play key roles in information theory, statistics, learning, signal processing, and other theoretical and applied branches of mathematics [11,12]. Morisita's coefficient [13] was proposed as an index of similarity between two communities. Weitzman's coefficient [14], primarily used to compare income distributions, was defined as the region where the curves of two probability distributions intersect. Kullback and Leibler [15] introduced the Kullback–Leibler measure, which measures the gain in information between two distributions and has been widely used in the literature on data mining. Jeffreys [16] introduced and studied a divergence measure called the Jeffreys distance, which is regarded as a symmetrization of the Kullback–Leibler measure. For a comprehensive review of various divergence measures, see [17–19].

The OVL coefficients are used in various fields, such as ecological processes [20], statistical ecology [21], clinical trials [5], data fusion [22], information processing [23], applied statistics [24], economics [25], and others.

Inference for OVL measures has been investigated by several researchers under normal, Weibull, and exponential distributions. In 2005, al-Saidy et al. [26] presented the inference



Citation: Alhihi, S.; Almheidat, M. Estimation of Pianka Overlapping Coefficient for Two Exponential Distributions. *Mathematics* **2023**, *11*, 4152. <https://doi.org/10.3390/math11194152>

Academic Editor: Velizar Pavlov

Received: 29 August 2023

Revised: 24 September 2023

Accepted: 27 September 2023

Published: 2 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

of three OVL coefficients for two Weibull distributions with the same shape parameter and different scale parameters.

Al-Saleh and Samawi [27] used bootstrap and Taylor series approximation to investigate the interval estimation of three OVL coefficients for two exponential distributions with different means. Samawi and Al-Saleh [28] studied three OVL coefficients for two exponential distributions and estimated them using ranked set sampling. Hamza et al. [29] proposed a new OVL coefficient based on the Kullback–Leibler measure for two exponential distributions. Sibil et al. [30] investigated both interval estimation and hypothesis testing for the OVL coefficients for one- and two-parameter exponential distributions using the concept of a generalized pivotal quantity.

Pianka’s overlap coefficient is used to assess the similarity of resource use by two species [31], they used Pianka’s overlap coefficient as a summary measure and to make inferences, typically about competition for resources.

Pianka’s overlap is used in mechanisms that favour morphological co-occurrence; Vieira and Port [32] evaluated the Pianka’s overlap between two species based on three main niche dimensions: habitat, food, and time. Jacqueline et al. [33] calculated dietary overlap between foxes and dingoes using Pianka’s index. Sa Oliveira et al. [34] investigated diet and niche breadth in fish communities, for which they estimated niche breadth using the liven index and Pianka’s measure.

In this paper, we consider the Pianka’s OVL coefficient (ρ) between two exponential distributions. We determine both the limiting and exact distributions for the maximum likelihood estimator (MLE) of ρ . We study the MLE and Bayesian estimators and compare their efficiency with each other. In addition, we consider interval estimation of ρ using the asymptotic technique and the transformation technique, and compare the effectiveness of both techniques.

2. General Setting and Definition of the Pianka Overlap Measure

Let $f_1(x)$ and $f_2(x)$ be two continuous probability density functions. Pianka’s overlap measure is defined as follows [31]:

$$\rho(f_1, f_2) = \frac{\int_{-\infty}^{\infty} f_1(x)f_2(x)dx}{\sqrt{\int_{-\infty}^{\infty} f_1^2(x)dx}\sqrt{\int_{-\infty}^{\infty} f_2^2(x)dx}}. \tag{1}$$

If a random variable X follows the exponential distribution, then the respective cdf and pdf of X are provided by

$$F(x) = 1 - e^{-\frac{x}{\theta}}, x > 0; \theta > 0 \quad \text{and} \quad f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, x > 0; \theta > 0,$$

and X is denoted by $Exp(\theta)$.

Now, let (X_1, \dots, X_n) and (Y_1, \dots, Y_m) be two independent random samples taken from $Exp(\theta_1)$ and $Exp(\theta_2)$, respectively. Then, the Pianka’s overlap coefficient ρ between the two exponential distributions, as defined in Equation (1), is provided by

$$\rho = \rho(\theta_1, \theta_2) = \frac{\int_0^{\infty} \frac{1}{\theta_1}e^{-\frac{x}{\theta_1}} \frac{1}{\theta_2}e^{-\frac{x}{\theta_2}} dx}{\sqrt{\int_0^{\infty} \frac{1}{\theta_1^2}e^{-\frac{2x}{\theta_1}} dx}\sqrt{\int_0^{\infty} \frac{1}{\theta_2^2}e^{-\frac{2x}{\theta_2}} dx}} = \frac{2\sqrt{\theta_1\theta_2}}{\theta_1 + \theta_2}, \theta_1 > 0, \theta_2 > 0. \tag{2}$$

Let $k = \frac{\theta_1}{\theta_2}$. Then, the Pianka’s OVL coefficient in (2) can be written as a function of k , as follows:

$$\rho = \rho(k) = \frac{2\sqrt{k}}{k + 1}, k > 0. \tag{3}$$

Several properties of $\rho(k)$ are provided in the following lemma.

Lemma 1. For ρ defined in (3):

1. $0 \leq \rho(k) \leq 1$ for all $k > 0$
2. $\rho(k) = 1$ iff $k = 1$, i.e., $\theta_1 = \theta_2$
3. $\rho(k) > 0$, since $\theta_1, \theta_2 > 0$
4. $\rho(k) = \rho(\frac{1}{k})$
5. $\rho(k)$ is monotonically increasing for $k < 1$ and decreasing for $k > 1$, with a maximum of $\rho(k)$ at $k = 1$.

Proof. It is easy to derive the above results from the formula of the Pianka’s overlap coefficient formula in (3). \square

Figure 1 shows the plot of the Pianka’s overlap coefficient between two exponential distributions as a function of k , where $k < 1$.

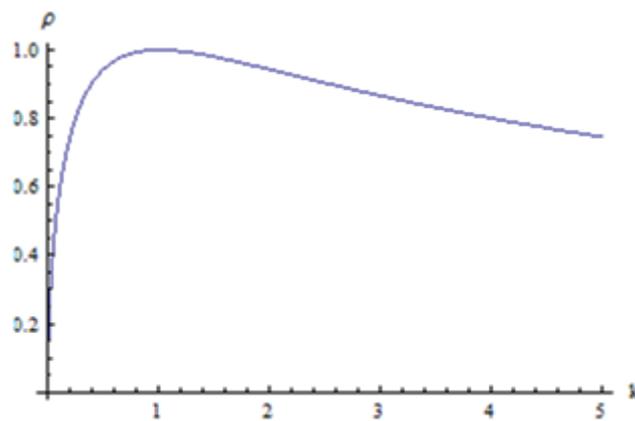


Figure 1. Pianka overlap coefficient as a function of k .

In the following section, we find the maximum likelihood estimator of Pianka’s overlap coefficient ρ , namely, $\hat{\rho}_{MLE}$, along with its distribution. In addition, we investigate the limiting distribution of $\hat{\rho}_{MLE}$.

3. Maximum Likelihood Estimator of $\rho(\theta_1, \theta_2)$

It is known that the MLEs for θ_1 and θ_2 based on two samples taken from $Exp(\theta_1)$ and $Exp(\theta_2)$ are provided by $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = \bar{Y}$, respectively. From the basic properties of the exponential distribution, we have $\hat{\theta}_1 \sim Gamma\left(n, \frac{\theta_1}{n}\right)$ with $Var\left(\hat{\theta}_1\right) = \frac{\theta_1^2}{n}$ and $\hat{\theta}_2 \sim Gamma\left(m, \frac{\theta_2}{m}\right)$, and with $Var\left(\hat{\theta}_2\right) = \frac{\theta_2^2}{m}$, where $Gamma(\alpha, \beta)$ stands for the gamma distribution with shape parameter α and scale parameter β . It follows that the estimates $(\hat{\theta}_1, \hat{\theta}_2)$ represent a complete minimal sufficient statistic for (θ_1, θ_2) . Thus, from the invariance property of the MLE, the MLE of $\rho(\theta_1, \theta_2)$ is

$$\hat{\rho}_{MLE} = \rho(\hat{\theta}_1, \hat{\theta}_2) = \frac{2\sqrt{\bar{x}\bar{y}}}{\bar{x} + \bar{y}}. \tag{4}$$

3.1. Limiting Distribution of $\hat{\rho}_{MLE}$

The following theorem concludes that the limiting distribution for the MLE of Pianka’s overlap coefficient for two exponential distributions with different scale parameters is the normal distribution, using $N(\mu, \sigma^2)$ to denote the normal distribution with location parameter μ and scale parameter σ .

Theorem 1. Let (X_1, \dots, X_n) and (Y_1, \dots, Y_m) be two independent random samples from $Exp(\theta_1)$ and $Exp(\theta_2)$, respectively, with $\theta_1 \neq \theta_2$. Then, the asymptotic distribution for $\hat{\rho}_{MLE}$ is

$$\sqrt{\frac{nm}{n+m}}(\hat{\rho}_{MLE} - \rho(\theta_1, \theta_2)) \xrightarrow{d} N\left(0, \frac{\theta_1\theta_2(\theta_1 - \theta_2)^2}{(\theta_1 + \theta_2)^4}\right), \quad n, m \rightarrow \infty.$$

Proof. Using the asymptotic property of MLE and the multivariate delta method (δ -method), we have

$$\sqrt{n}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} N(0, I_1^{-1}(\theta_1)), \quad n \rightarrow \infty.$$

That is,

$$\sqrt{n}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} N(0, \theta_1^2), \quad n \rightarrow \infty \quad \text{and} \quad \sqrt{m}(\hat{\theta}_2 - \theta_2) \xrightarrow{d} N(0, \theta_2^2), \quad m \rightarrow \infty,$$

where $I_1^{-1}(\theta)$ is the Fisher information.

We want to find the asymptotic distribution of $\hat{\rho}_{MLE} = \rho(\hat{\theta}_1, \hat{\theta}_2)$, as $n, m \rightarrow \infty$.

Using the fact that $\hat{\theta}_1 \xrightarrow{P} \theta_1$ and $\hat{\theta}_2 \xrightarrow{P} \theta_2$ and the continuous mapping theorem, we obtain

$$\hat{\rho}_{MLE} = \frac{2\sqrt{\hat{\theta}_1\hat{\theta}_2}}{\hat{\theta}_1 + \hat{\theta}_2} \xrightarrow{P} \frac{2\sqrt{\theta_1\theta_2}}{\theta_1 + \theta_2} = \rho(\theta_1, \theta_2).$$

Now, we are interested in the asymptotic distribution of $\hat{\rho}_{MLE} = \rho(\hat{\theta}_1, \hat{\theta}_2)$.

Because $\hat{\rho}_{MLE}$ is a function of $\hat{\theta}_1$ and $\hat{\theta}_2$, using an alternative form of the multivariate δ -method [35] we obtain $E(\hat{\rho}_{MLE}) \approx \rho(\theta_1, \theta_2)$ and

$$\begin{aligned} Var(\hat{\rho}_{MLE}) &\simeq Var(\hat{\theta}_1) \left(\frac{\partial \rho(\theta_1, \theta_2)}{\partial \theta_1}\right)^2 + Var(\hat{\theta}_2) \left(\frac{\partial \rho(\theta_1, \theta_2)}{\partial \theta_2}\right)^2, \\ &= \frac{\theta_1^2}{n} \left(\frac{\theta_2(\theta_2 - \theta_1)}{\sqrt{\theta_1\theta_2}(\theta_1 + \theta_2)^2}\right)^2 + \frac{\theta_2^2}{m} \left(\frac{\theta_1(\theta_1 - \theta_2)}{\sqrt{\theta_1\theta_2}(\theta_1 + \theta_2)^2}\right)^2, \\ &= \frac{(n+m)\theta_1\theta_2(\theta_1 - \theta_2)^2}{nm(\theta_1 + \theta_2)^4}. \end{aligned}$$

Therefore, the asymptotic distribution of $\hat{\rho}_{MLE}$ is

$$\sqrt{\frac{nm}{n+m}}(\hat{\rho}_{MLE} - \rho(\theta_1, \theta_2)) \xrightarrow{d} N\left(0, \frac{\theta_1\theta_2(\theta_1 - \theta_2)^2}{(\theta_1 + \theta_2)^4}\right).$$

□

3.2. The Exact Distribution of $\hat{\rho}_{MLE}$

To ease the derivation of the distribution of $\hat{\rho}_{MLE}$, we can rewrite Equation (4) as follows:

$$\hat{\rho}_{MLE} = \frac{2}{\sqrt{nm} \left(\frac{1}{n}\sqrt{\frac{V}{W}} + \frac{1}{m}\sqrt{\frac{W}{V}}\right)}, \tag{5}$$

where $V = \sum_{i=1}^n X_i \sim Gamma(n, \theta_1)$ and $W = \sum_{i=1}^m Y_i \sim Gamma(m, \theta_2)$. Now, we apply the following steps.

Step 1. Find the pdf of $H = \sqrt{\frac{V}{W}}$ by considering the following transformations.

Let $H_1 = \sqrt{\frac{V}{W}}$ and $H_2 = \sqrt{W}$; then, $V = H_1^2 H_2^2$ and $W = H_2^2$. The absolute value of the Jacobian of this transform is $|J| = 4h_1 h_2^3$.

Thus, the joint pdf of H_1 and H_2 is

$$\begin{aligned} f_{H_1, H_2}(h_1, h_2) &= f_{V, W}(v = h_1^2 h_2^2, w = h_2^2) |J|, \\ &= \frac{(h_1^2 h_2^2)^{n-1} e^{-\frac{h_1^2 h_2^2}{\theta_1}}}{\Gamma(n) \theta_1^n} \frac{(h_2^2)^{m-1} e^{-\frac{h_2^2}{\theta_2}}}{\Gamma(m) \theta_2^m} (4h_1 h_2^3), \\ &= \frac{4h_1^{2n-1} h_2^{2n+2m-1} e^{-h_2^2 \left(\frac{h_1^2}{\theta_1} + \frac{1}{\theta_2}\right)}}{\Gamma(n) \Gamma(m) \theta_1^n \theta_2^m}, \quad h_1 > 0, h_2 > 0; \theta_1 > 0, \theta_2 > 0. \end{aligned}$$

By integrating h_2 out, the pdf of H_1 is

$$f_{H_1}(h_1) = \int_0^\infty \frac{4h_1^{2n-1} h_2^{2n+2m-1} e^{-h_2^2 \left(\frac{h_1^2}{\theta_1} + \frac{1}{\theta_2}\right)}}{\Gamma(n) \Gamma(m) \theta_1^n \theta_2^m} dh_2 = c \frac{h_1^{2n-1}}{(\theta_1 + \theta_2 h_1^2)^{n+m}}, \quad h_1 > 0; \theta_1 > 0, \theta_2 > 0,$$

Consequently, the pdf of H is

$$f_H(h) = \frac{ch^{2n-1}}{(\theta_1 + \theta_2 h^2)^{n+m}}, \quad h > 0, \theta_1 > 0, \theta_2 > 0,$$

where $c = \frac{2\Gamma(n+m)\theta_1^m \theta_2^n}{\Gamma(n)\Gamma(m)}$.

Step 2. Solve $\hat{\rho}_{MLE}$ for h .

From Equation (5) and the transformation $H = \sqrt{\frac{V}{W}}$, we have

$$\hat{\rho}_{MLE} = \frac{2}{\sqrt{nm} \left(\frac{h}{n} + \frac{1}{mh}\right)} = \frac{2h}{\sqrt{nm} \left(\frac{h^2}{n} + \frac{1}{m}\right)}.$$

Now, let $\hat{\rho}_{MLE} = R$, allowing $R = \frac{2h}{\sqrt{nm} \left(\frac{h^2}{n} + \frac{1}{m}\right)}$ to be rewritten as the quadratic equation

$$m\sqrt{nm}Rh^2 - 2nmh + m\sqrt{nm}R = 0. \tag{6}$$

The two solutions of Equation (6) are

$$U_1 = \frac{\sqrt{nm}}{mR} (1 - \sqrt{1 - R^2}), U_1 > 0 \text{ and } U_2 = \frac{\sqrt{nm}}{mR} (1 + \sqrt{1 - R^2}), U_2 > 0.$$

Step 3. The pdf of $R = \hat{\rho}_{MLE}$ is

$$\begin{aligned} f_R(r) &= f_H(u_1) \left| \frac{\partial u_1}{\partial r} \right| + f_H(u_2) \left| \frac{\partial u_2}{\partial r} \right| \\ &= \frac{c}{r\sqrt{1-r^2}} \left[\left(\frac{\sqrt{\frac{n}{m}}(-1 + \sqrt{1-r^2})}{r} \right)^{2n} \left(\theta_1 + \frac{n\theta_2(-1 + \sqrt{1-r^2})^2}{mr^2} \right)^{-n-m} \right. \\ &\quad \left. + \left(\frac{\sqrt{\frac{n}{m}}(1 + \sqrt{1-r^2})}{r} \right)^{2n} \left(\theta_1 + \frac{n\theta_2(1 + \sqrt{1-r^2})^2}{mr^2} \right)^{-n-m} \right], \quad 0 < r < 1; \theta_1 > 0, \theta_2 > 0. \end{aligned}$$

Figure 2 shows different plots of the density of $\hat{\rho}_{MLE}$ for $(n, m) = (30, 50)$. Based on the figure, the pdf of $\hat{\rho}_{MLE}$ can be bell-shape, bimodal, or J-shaped.

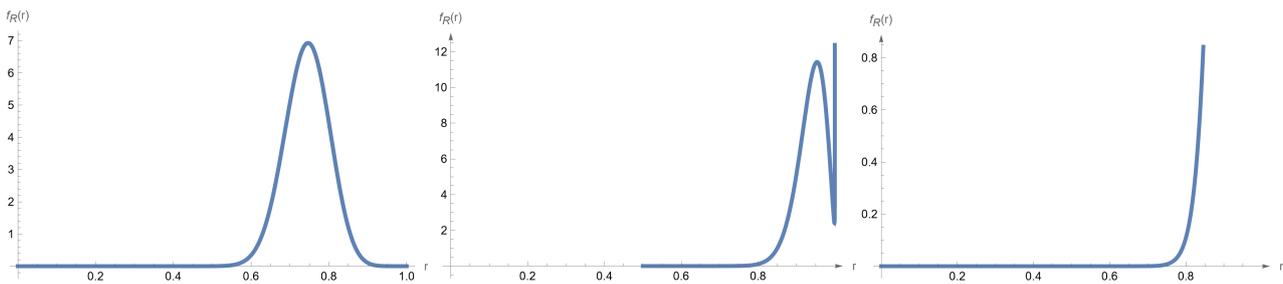


Figure 2. The pdf of $\hat{\rho}_{MLE}$ for $(\theta_1, \theta_2) = (2, 10), (5, 10),$ and $(8, 10)$.

4. Interval Estimation of $\rho(\theta_1, \theta_2)$

In this section, we find interval estimation of Pianka’s overlap coefficient ρ by considering both asymptotic and transformation techniques; later, in Section 6, we perform a Monte Carlo analysis to compare the effectiveness of these two different approaches.

4.1. Asymptotic Technique

A large sample confidence interval for $\rho(\theta_1, \theta_2)$ can be easily calculated. From Theorem (2.1) and the continuous mapping theorem, we have

$$\frac{\hat{\theta}_1 \hat{\theta}_2 (\hat{\theta}_1 - \hat{\theta}_2)^2}{(\hat{\theta}_1 + \hat{\theta}_2)^4} \xrightarrow{p} \frac{\theta_1 \theta_2 (\theta_1 - \theta_2)^2}{(\theta_1 + \theta_2)^4}.$$

Hence, a $100(1 - \alpha)\%$ large sample confidence interval for $\rho(\theta_1, \theta_2)$ is

$$\left(\hat{\rho}_{MLE} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{(n+m)\hat{\theta}_1\hat{\theta}_2(\hat{\theta}_1-\hat{\theta}_2)^2}{nm(\hat{\theta}_1+\hat{\theta}_2)^4}}, \hat{\rho}_{MLE} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{(n+m)\hat{\theta}_1\hat{\theta}_2(\hat{\theta}_1-\hat{\theta}_2)^2}{nm(\hat{\theta}_1+\hat{\theta}_2)^4}} \right),$$

where Z_γ is the γ th percentile of the standard normal distribution.

4.2. Transformation Technique

From the assumption in Equation (3), $\rho(k) = \frac{2\sqrt{k}}{k+1}$, where the MLE of k is $\hat{k} = \frac{\hat{\theta}_1}{\hat{\theta}_2}$. From Section 3 and the relationship between the gamma distribution and the chi-square distribution, it is easy to conclude that $\frac{2n\hat{\theta}_1}{\theta_1} \sim \chi^2_{(2n)}$ and $\frac{2m\hat{\theta}_2}{\theta_2} \sim \chi^2_{(2m)}$; thus, $\frac{\hat{\theta}_2}{\hat{\theta}_1} \frac{\theta_1}{\theta_2}$ has an F -distribution with $(2m, 2n)$ degrees of freedom.

Let L and U be the lower and upper confidence limits, respectively; from the concept of the confidence interval, we have

$$1 - \alpha = \Pr(F_{\frac{\alpha}{2}, 2m, 2n} < \frac{\hat{\theta}_2}{\hat{\theta}_1} k < F_{1-\frac{\alpha}{2}, 2m, 2n}). \tag{7}$$

By solving (7) for k , we obtain the values of L and U as $L = \frac{\hat{\theta}_1}{\hat{\theta}_2} F_{\frac{\alpha}{2}, 2m, 2n}$ and $U = \frac{\hat{\theta}_1}{\hat{\theta}_2} F_{1-\frac{\alpha}{2}, 2m, 2n}$.

However, the overlap coefficient $\rho(k)$ is not a monotone function of k . Therefore, using the transformation technique, we can obtain a $100(1 - \alpha)\%$ confidence interval for ρ , as follows:

$$\left(\text{Min} \left(\frac{2\sqrt{L}}{L+1}, \frac{2\sqrt{U}}{U+1} \right), \text{Max} \left(\frac{2\sqrt{L}}{L+1}, \frac{2\sqrt{U}}{U+1} \right) \right),$$

where F_{γ, r_1, r_2} is the γ^{th} percentile of the F -distribution with (r_1, r_2) degrees of freedom.

5. Bayes Estimator of $\rho(\theta_1, \theta_2)$

Let (X_1, \dots, X_n) and (Y_1, \dots, Y_m) be two independent random samples taken from $Exp(\theta_1)$ and $Exp(\theta_2)$, respectively. Let $V = \sum X_i, W = \sum Y_i, \theta_1 \sim InvGamma(a, b)$, and $\theta_2 \sim InvGamma(c, d)$, where $InvGamma(\cdot, \cdot)$ is the inverse gamma distribution.

Using the fact that $V \sim Gamma(n, \theta_1)$ and $W \sim Gamma(m, \theta_2)$, the posterior distribution of $\theta = (\theta_1, \theta_2)$ given V, W is

$$\begin{aligned} \pi(\theta_1, \theta_2 | v, w) &= \frac{f_V(v | \theta_1) p_1(\theta_1) f_W(w | \theta_2) p_2(\theta_2)}{\int_0^\infty \int_0^\infty f_V(v | \theta_1) p_1(\theta_1) f_W(w | \theta_2) p_2(\theta_2) dv dw} \\ &= \frac{\frac{v^{n-1} e^{-\frac{v}{\theta_1}}}{\Gamma(n)\theta_1^n} \frac{\theta_1^{-a-1} b^a e^{-\frac{b}{\theta_1}}}{\Gamma(a)} \frac{w^{m-1} e^{-\frac{w}{\theta_2}}}{\Gamma(m)\theta_2^m} \frac{\theta_2^{-c-1} d^c e^{-\frac{d}{\theta_2}}}{\Gamma(c)}}{\int_0^\infty \int_0^\infty \frac{v^{n-1} e^{-\frac{v}{\theta_1}}}{\Gamma(n)\theta_1^n} \frac{\theta_1^{-a-1} b^a e^{-\frac{b}{\theta_1}}}{\Gamma(a)} \frac{w^{m-1} e^{-\frac{w}{\theta_2}}}{\Gamma(m)\theta_2^m} \frac{\theta_2^{-c-1} d^c e^{-\frac{d}{\theta_2}}}{\Gamma(c)} dv dw} \\ &\propto \theta_1^{-(n+a)-1} e^{-\frac{1}{\theta_1}(v+b)} \theta_2^{-(m+c)-1} e^{-\frac{1}{\theta_2}(w+d)}, \quad \theta_1 > 0, \theta_2 > 0; a > 0, b > 0, c > 0, d > 0, \end{aligned}$$

where $p_1(\theta_1)$ and $p_2(\theta_2)$ are prior probability distributions for θ_1 and θ_2 , respectively.

Then,

$$\pi(\theta | v, w) = \pi_1(\theta_1 | v) \pi_2(\theta_2 | w),$$

where $\theta_1 | v \sim InvGamma(n + a, v + b)$ and $\theta_2 | w \sim InvGamma(m + c, w + d)$.

The Bayes estimator $\hat{\rho}_{Bayes}$ is

$$\begin{aligned} \hat{\rho}_{Bayes} &= \iint \rho(\theta_1, \theta_2) \pi(\theta_1, \theta_2 | v, w) d\theta_1 d\theta_2 \\ &= \int_0^\infty \int_0^\infty \frac{2\sqrt{\theta_1 \theta_2}}{\theta_1 + \theta_2} \frac{\theta_1^{-(n+a)-1} e^{-\frac{1}{\theta_1}(v+b)} (v+b)^{n+a}}{\Gamma(n+a)} \frac{\theta_2^{-(m+c)-1} e^{-\frac{1}{\theta_2}(w+d)} (w+d)^{m+c}}{\Gamma(m+c)} d\theta_1 d\theta_2 \\ &= \frac{2(v+b)^{n+a} (w+d)^{m+c}}{\Gamma(n+a)\Gamma(m+c)} \int_0^\infty \int_0^\infty \frac{\theta_1^{-(n+a)-\frac{1}{2}} \theta_2^{-(m+c)-\frac{1}{2}} e^{-\frac{1}{\theta_2}(w+d)} e^{-\frac{1}{\theta_1}(v+b)}}{\theta_1 + \theta_2} d\theta_1 d\theta_2. \end{aligned}$$

The above estimate does not have a simple closed form; thus, we obtain it numerically. For the asymptotic distribution of the Bayes estimator $\hat{\rho}_{Bayes}$, the Bernstein–von Mises theorem [36] concludes that the Bayesian estimator and the maximum likelihood estimator are asymptotically equivalent for large sample sizes.

In the next section, we present a simulation study to compare the two approaches for finding the interval estimator of Pianka’s overlap coefficient, as described earlier in Section 4. Additionally, we investigate the performance of the maximum likelihood estimator ($\hat{\rho}_{MLE}$) and Bayes estimator ($\hat{\rho}_{Bayes}$) of the Pianka’s overlap coefficient detailed in Sections 3 and 5.

6. Simulation Study

To compare the two approaches of interval estimation of Pianka’s overlap coefficient, we consider two criteria:

1. The term “valid confidence level” can be applied to an interval estimation process when, in repeated sampling, the actual coverage of the true but unmeasured statistic is close to the nominal confidence level;
2. If the expected length of the simulated period is short, a method for estimating intervals can be described as “valid length-efficient”.

To compare the estimators, we use the bias, mean square error (MSE), and efficiency for each estimator. In order to use the above criteria, we conducted a simulation study, as follows:

1. A random sample of size n is generated from $Exp(\theta_1)$. This random sample is used to calculate $\hat{\theta}_1$.
2. A random sample of size m is generated from $Exp(\theta_2)$. This random sample is used to calculate $\hat{\theta}_2$.
3. The lower limit L_i , upper limit U_i , and width W_i , are calculated with a nominal confidence level of 95%.
4. The MLE $(\hat{\rho}_{MLE})_i$ and the Bayes $(\hat{\rho}_{Bayes})_i$ estimators are calculated.
5. Steps 1–4 above are repeated 10,000 times.
6. The average of the lower limits (AL), median of the lower limits (ML), average of the upper limits (AU), median of the upper limits (MU), average width (AW), and median width (MW) are calculated for each interval.
7. The percentage of ρ out of the 10,000 samples generated in Step 3 is called the “coverage probability” and is denoted by AP .
8. Histogram Plots for $\hat{\rho}_{MLE}$ and $\hat{\rho}_{Bayes}$ are generated.
9. Bias and MSE are calculated for $\hat{\rho}_{MLE}$ and $\hat{\rho}_{Bayes}$, then efficiency is calculated (i.e., efficiency $(\hat{\rho}_{MLE} \cdot \hat{\rho}_{Bayes}) = \frac{MSE(\hat{\rho}_{MLE})}{MSE(\hat{\rho}_{Bayes})}$).
10. Steps 1–9 above are repeated for $(n, m) = (20, 20), (20, 30), (30, 50), (50, 50), (50, 100), (100, 100)$, and $(\theta_1, \theta_2) = (2, 10), (5, 10), (8, 10)$ for each value of $k = 0.2, 0.5, 0.8$.

Mathematica was used to simulate each of the interval estimation and point estimation methods for the Pianka’s overlap measure ρ .

Tables 1–3 show the simulated interval estimators using the asymptotic and transformation techniques based on exponential random samples with a nominal confidence level of 95%. These results show that the average width (AW) is almost the same as the median width (MW) and that the transformation method consistently performs better in terms of the confidence interval width. Moreover, the transformation method appears to be effective in terms of the coverage probability except for values of k around one and very small sample sizes.

As the sample size increases, the coverage probability of the two techniques approaches the nominal value. The coverage probability of the asymptotic technique works very well, and increases as k approaches one; however, when $k < 0.5$ and for small sample sizes the transformation technique performs exceptionally well.

Table 1. Simulation results for the two approaches of interval estimation of Pianka’s OVL coefficient for $k = 0.2, \rho = 0.7454$.

(n, m)	Technique	AL	AU	AW	ML	MU	MW	AP
(20, 20)	Asymptotic	0.594	0.894	0.299	0.592	0.890	0.305	0.931
	Transformation	0.592	0.884	0.292	0.591	0.891	0.299	0.951
(20, 30)	Asymptotic	0.606	0.880	0.275	0.604	0.885	0.279	0.935
	Transformation	0.608	0.876	0.269	0.607	0.882	0.274	0.952
(30, 30)	Asymptotic	0.622	0.869	0.247	0.620	0.872	0.250	0.937
	Transformation	0.621	0.863	0.243	0.619	0.867	0.247	0.941
(30, 50)	Asymptotic	0.632	0.854	0.221	0.632	0.857	0.224	0.935
	Transformation	0.634	0.853	0.218	0.634	0.856	0.221	0.945
(50, 50)	Asymptotic	0.649	0.841	0.192	0.648	0.843	0.194	0.942
	Transformation	0.648	0.839	0.191	0.647	0.831	0.193	0.951
(50, 100)	Asymptotic	0.651	0.827	0.167	0.651	0.829	0.169	0.946
	Transformation	0.662	0.828	0.166	0.662	0.829	0.167	0.951
(100, 100)	Asymptotic	0.676	0.813	0.137	0.676	0.814	0.138	0.942
	Transformation	0.676	0.812	0.136	0.676	0.813	0.137	0.944

Table 2. Simulation results for the two approaches of interval estimation of Pianka’s OVL coefficient for $k = 0.5, \rho = 0.9428$.

(n, m)	Technique	AL	AU	AW	ML	MU	MW	AP
(20, 20)	Asymptotic	0.841	1.027	0.185	0.845	1.035	0.195	0.879
	Transformation	0.812	0.987	0.176	0.815	0.994	0.184	0.957
(20, 30)	Asymptotic	0.848	1.011	0.172	0.852	1.028	0.179	0.891
	Transformation	0.827	0.988	0.162	0.831	0.995	0.168	0.951
(30, 30)	Asymptotic	0.859	1.013	0.154	0.861	1.020	0.161	0.896
	Transformation	0.839	0.987	0.149	0.841	0.994	0.155	0.949
(30, 50)	Asymptotic	0.867	1.006	0.131	0.869	1.013	0.144	0.913
	Transformation	0.853	0.987	0.134	0.856	0.993	0.138	0.952
(50, 50)	Asymptotic	0.879	0.999	0.120	0.881	1.004	0.123	0.916
	Transformation	0.866	0.984	0.118	0.869	0.989	0.120	0.952
(50, 100)	Asymptotic	0.877	0.992	0.105	0.889	0.996	0.107	0.927
	Transformation	0.879	0.982	0.102	0.881	0.985	0.104	0.953
(100, 100)	Asymptotic	0.898	0.984	0.086	0.899	0.986	0.087	0.932
	Transformation	0.891	0.976	0.085	0.892	0.979	0.086	0.947

Table 3. Simulation results fir the two approaches of interval estimation of Pianka’s OVL coefficient for $k = 0.8, \rho = 0.9938$.

(n, m)	Technique	AL	AU	AW	ML	MU	MW	AP
(20, 20)	Asymptotic	0.935	1.028	0.093	0.949	1.031	0.084	0.924
	Transformation	0.908	0.969	0.087	0.915	0.971	0.078	0.266
(20, 30)	Asymptotic	0.943	1.024	0.081	0.955	1.026	0.072	0.914
	Transformation	0.911	0.975	0.074	0.928	0.983	0.066	0.290
(30, 30)	Asymptotic	0.952	1.011	0.068	0.962	1.022	0.061	0.898
	Transformation	0.930	0.982	0.065	0.936	0.989	0.059	0.375
(30, 50)	Asymptotic	0.958	1.016	0.058	0.967	1.018	0.053	0.884
	Transformation	0.941	0.986	0.054	0.947	0.991	0.049	0.437
(50, 50)	Asymptotic	0.965	1.013	0.047	0.971	1.015	0.045	0.873
	Transformation	0.950	0.992	0.046	0.954	0.996	0.043	0.582
(50, 100)	Asymptotic	0.961	1.001	0.040	0.974	1.011	0.039	0.861
	Transformation	0.959	0.994	0.038	0.962	0.997	0.363	0.691
(100, 100)	Asymptotic	0.976	1.007	0.031	0.979	1.008	0.031	0.856
	Transformation	0.967	0.997	0.031	0.969	0.999	0.030	0.853

Figures 3–5 plot the MLE and Bayes estimators of ρ for $k = 0.2, 0.5, 0.8$, and $(n, m) = (50, 50)$.

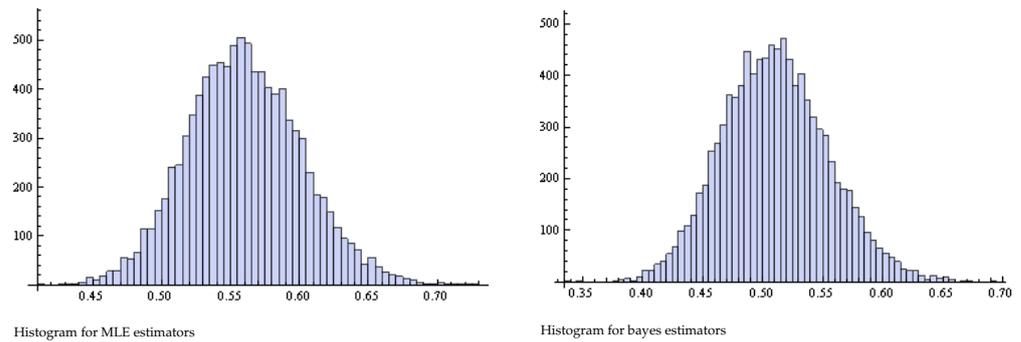


Figure 3. Histogram of Pianka estimators for $(n, m) = (50, 50)$ and $k = 0.2$.

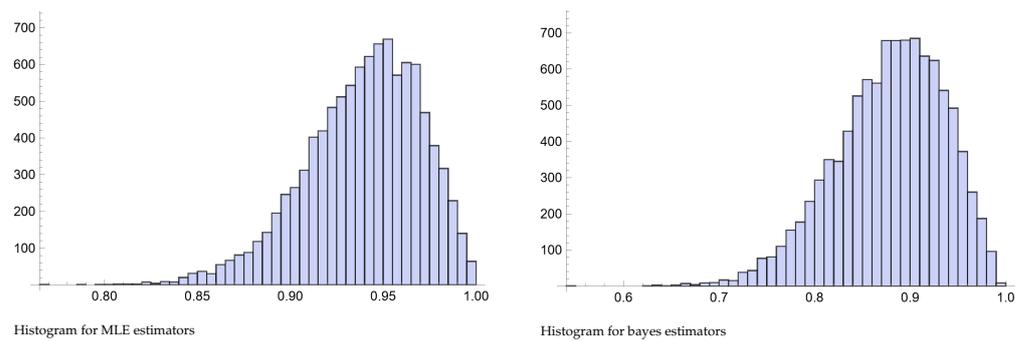


Figure 4. Histogram of Pianka estimators for $(n, m) = (50, 50)$ and $k = 0.5$.

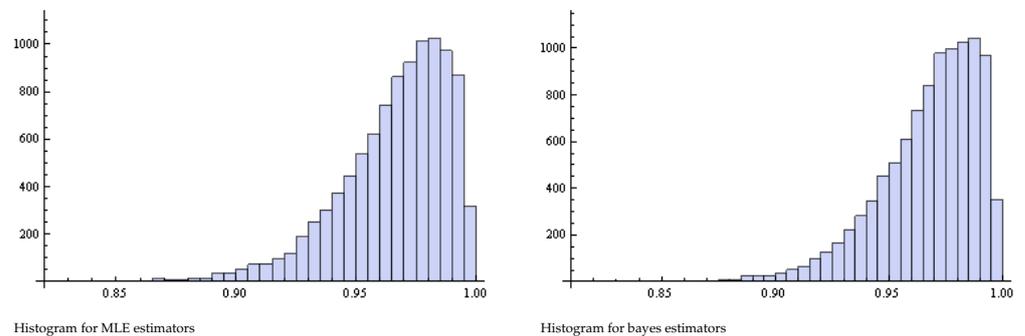


Figure 5. Histogram of Pianka estimators for $(n, m) = (50, 50)$ and $k = 0.8$.

Tables 4–6 present the results of the simulation study carried out to compare the MLE and Bayes estimators for the Pianka’s overlap coefficient. Based on these results, which only consider the values of $k < 1$, the absolute values of the bias are in all cases less than 0.05 and decrease as the sample size increases. It appears that the MLE estimator works well, and the Bayes estimator seems to work quite well at $k = 0.5$. However, for $k > 1$ the calculations are provided in terms of $\frac{1}{k}$ for the Pianka’s overlap measure. For sample sizes larger than 30, the bias and MSE are quite close to zero.

The estimates of the bias are plotted in Figure 6 for the MLE and Bayes estimators. From these results, it can be seen that the bias decreases significantly as the sample size increases. Figure 6a shows that the actual Pianka’s overlap is underestimated; however, for very small values of k and small sample sizes the true Pianka’s overlap is overestimated. Furthermore, the bias increases as k increases for the MLE estimator.

Table 4. Bias, MSE, and efficiency of the two estimators of Pianka’s OVL coefficient for $k = 0.2$. Exact Pianka’s coefficient $\rho = 0.7454$.

(n, m)	MLE Estimator		Bayes Estimator		Efficiency
	Bias	MSE	Bias	MSE	
(20, 20)	0.0017	0.0061	0.0386	0.0055	1.0752
(20, 30)	0.0038	0.0051	0.0433	0.0052	0.9698
(30, 30)	0.0005	0.0040	0.0274	0.0039	1.0188
(30, 50)	0.0026	0.0032	0.0317	0.0035	0.9261
(50, 50)	0.0004	0.0024	0.0172	0.0024	1.0111
(50, 100)	0.0011	0.0018	0.0201	0.0019	0.9529
(100, 100)	0.0001	0.0012	0.0086	0.0012	1.0145

Table 5. Bias, MSE, and efficiency of the two estimators of Pianka’s OVL coefficient for $k = 0.5$. Exact Pianka’s coefficient $\rho = 0.9428$.

(n, m)	MLE Estimator		Bayes Estimator		Efficiency
	Bias	MSE	Bias	MSE	
(20, 20)	0.0095	0.0025	0.0102	0.0018	1.3915
(20, 30)	0.0093	0.0022	0.0051	0.0014	1.5146
(30, 30)	0.0061	0.0017	0.0069	0.0013	1.2805
(30, 50)	0.0059	0.0014	0.0024	0.0010	1.3485
(50, 50)	0.0033	0.0001	0.0031	0.0009	1.1096
(50, 100)	0.0038	0.0008	0.0012	0.0006	1.2178
(100, 100)	0.0014	0.0005	0.0022	0.0005	1.0559

Table 6. Bias, MSE, and efficiency of the two estimators of Pianka’s OVL coefficient for $k = 0.8$. Exact Pianka’s coefficient $\rho = 0.9938$.

(n, m)	MLE Estimator		Bayes Estimator		Efficiency
	Bias	MSE	Bias	MSE	
(20, 20)	0.0111	0.0007	0.0213	0.0009	0.8145
(20, 30)	0.0103	0.0006	0.0175	0.0006	0.8923
(30, 30)	0.0079	0.0004	0.0146	0.0005	0.7897
(30, 50)	0.0067	0.0003	0.0111	0.0003	0.9122
(50, 50)	0.0049	0.0002	0.0091	0.0002	0.8199
(50, 100)	0.0038	0.0001	0.0064	0.0001	0.9160
(100, 100)	0.0023	0.0001	0.0046	0.0001	0.8554

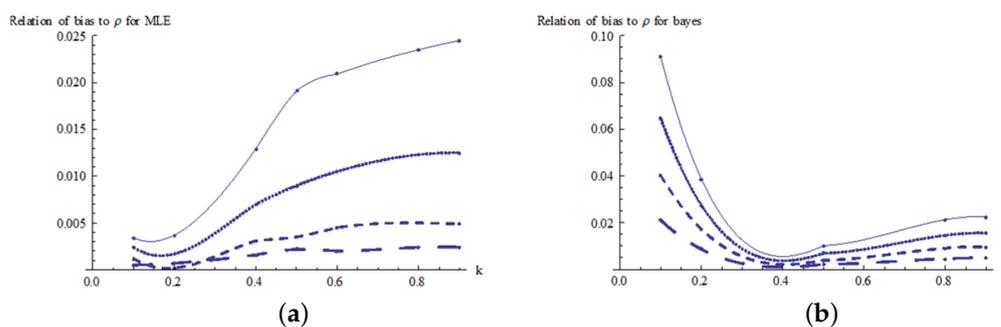


Figure 6. Relationship of bias to k for Pianka’s coefficient: (a) relation of bias to ρ for the MLE estimator and (b) relation of bias to ρ for the Bayes estimator.

The estimates of MSE are plotted in Figure 7 for the MLE and Bayes estimators. From these results, it can be seen that the MSE decreases significantly as the sample size increases. Figure 7a shows that for small k values and small sample sizes, there is a significant increase in the MSE for the MLE estimator. For the Bayes estimator, Figures 6b and 7b show that both the bias and the MSE decrease as the value of k increases.

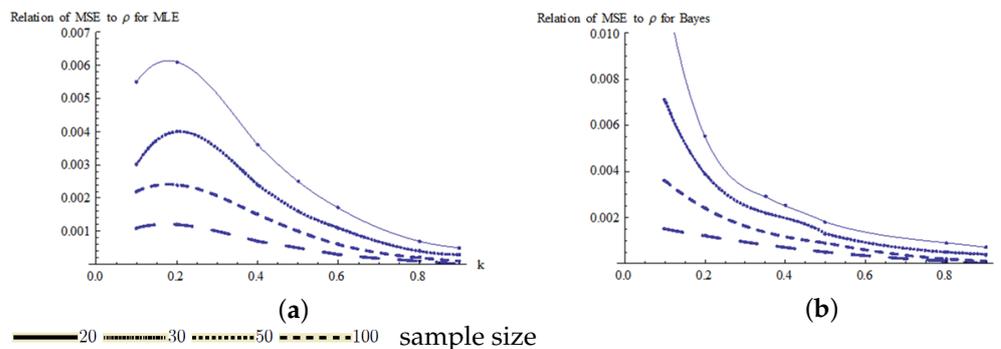


Figure 7. Relationship of MSE to k for Pianka’s coefficient: (a) relation of MSE to ρ for the MLE estimator and (b) relation of bias to ρ for Bayes estimator.

7. Conclusions

We have estimated Pianka’s overlap coefficient for two exponential populations with different scale parameters using the MLE and Bayes estimators, then compared these estimators by calculating the bias and MSE in a simulation study. In addition, we have constructed confidence intervals for the Pianka’s overlap measure using asymptotic and transformation techniques, then compared them using the “valid confidence level” and “valid length-efficiency”.

We investigated the accuracy of the Pianka’s overlap coefficient through a Monte Carlo analysis. In conclusion, it appears that there is no ideal approach. Therefore, a transformation procedure is recommended when $k < 0.5$ and the sample size is small. The asymptotic approach can be used if computers are available. For larger sample sizes and $k < 0.8$, the transformation approach is recommended.

Author Contributions: Conceptualization, S.A. and M.A.; Methodology, S.A. and M.A.; Software, S.A. and M.A.; Validation, S.A. and M.A.; Formal analysis, S.A. and M.A.; Resources, S.A. and M.A.; Writing—original draft, S.A.; Writing—review & editing, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tilton, J.W. The measurement of overlapping. *J. Educ. Psychol.* **1937**, *28*, 656–662. [[CrossRef](#)]
2. Gini, C.; Livada, G. *Nuovi Contributi Alla Teoria Della Transvariazione*; Atti della VI Riunione della Società Italiana di Statistica: Rome, Italy, 1943.
3. Matusita, K. Decision rules based on distance, for problems of fit, two samples and applications. *Ann. Inst. Math. Stat.* **1955**, *19*, 181–192. [[CrossRef](#)]
4. Anderson, G. Toward an empirical analysis of polarization. *J. Econom.* **2004**, *122*, 1–26. [[CrossRef](#)]
5. Mizuno, S.; Yamauchi, T.; Fukushima, A.; Matsuyama, Y.; Ohashi, Y. Overlap coefficient for assessing the similarity of pharmacokinetic data between ethnically different populations. *Clin. Trials* **2005**, *2*, 174–181. [[CrossRef](#)]
6. Beran, R. Minimum Hellinger distance estimates for parametric models. *Ann. Stat.* **1977**, *5*, 455–463. [[CrossRef](#)]
7. Rao, K.J.N.; Tintner, G. On the variate difference method. *Aust. J. Stat.* **1963**, *5*, 106–116. [[CrossRef](#)]
8. Smith, E.P. Niche breadth, resource availability, and inference. *Ecology* **1982**, *63*, 1675–1681. [[CrossRef](#)]
9. Pearson, K. On the Criterion that a Given System of Deviations From the Probable in the Case of a Correlated System of Variables is such that it Can be Reasonably Supposed to have a Risen From Random Sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1991**, *50*, 157–172. [[CrossRef](#)]
10. Hellinger, E. Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *J. Reine Angew. Math.* **1909**, *136*, 210–271. [[CrossRef](#)]
11. Nishiyama, T. A tight lower bound for the Hellinger distance with given means and variances. *arXiv* **2020**, arXiv:2010.13548.
12. Nishiyama, T.; Sason, I. On relations between the relative entropy and χ^2 -divergence, generalizations and applications. *Entropy* **2020**, *22*, 563. [[CrossRef](#)]
13. Morisita, M. Measuring of the dispersion and analysis of distribution patterns, Memoires of the Faculty of Science, Series E. *Biol. Kyushu Univ.* **1959**, *2*, 215–235.
14. Weitzman, M.S. Measures of overlap of income distributions of white and Negro families in the United States. In *US Bureau of the Census*; U.S. Department of Commerce: Washington, DC, USA, 1970; Volume 22.
15. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
16. Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **1946**, *186*, 453–461.
17. Abu, A.H.; Hassanat, A.; Lasassmeh, O.; Tarawneh, A.; Alhasanat, M.; Eyal, S.H.; Prasath, V. Effects of distance measure choice on k-nearest neighbor classifier performance: A review. *Big Data* **2019**, *7*, 221–248.
18. Cha, S. Comprehensive survey on distance/similarity measures between probability density functions. *City* **2007**, *1*, 1.
19. Taneja, I. On symmetric and nonsymmetric divergence measures and their generalizations. *Adv. Imaging Electron Phys.* **2005**, *138*, 177–250.
20. Abele, L.G. The community structure of coral-associated decapod crustaceans in a variable environment. *Ecol. Process. Coast. Mar. Syst. Mar. Sci.* **1979**, *10*, 265–287.
21. Chao, A.; Hwang, W.; Chen, Y.; Kuo, C. Estimating the number of shared species in two communities. *Stat. Sin.* **2000**, *10*, 227–246.
22. Moravec, H. *Mind Children: The Future of Robot and Human Intelligence*; Harvard University Press: Cambridge, MA, USA, 1988.
23. Viola, P.; Wells, W., III. Alignment by maximization of mutual information. *Int. J. Comput. Vis.* **1997**, *24*, 137–154. [[CrossRef](#)]
24. Inman, H.F.; Bradley, E.L. The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun. Stat. Theory Methods* **1989**, *18*, 3851–3874. [[CrossRef](#)]
25. Milanovic, B.; Shlomo, Y. Decomposing world income distribution: Does the world have a middle class? *Rev. Income Wealth* **2002**, *48*, 155–178. [[CrossRef](#)]
26. Al-Saidy, O.; Samawi, H.M.; Al-Saleh, M.F. Inference on overlap coefficients under the Weibul distribution: Equal Shape Parameter. *ESAIM Probab. Stat.* **2005**, *9*, 206–219. [[CrossRef](#)]
27. Al-Saleh, M.F.O.; Samawi, H. Interference on Overlapping Coefficients in Two Exponential Populations. *J. Mod. Appl. Stat. Methods* **2007**, *6*, 503–516. [[CrossRef](#)]
28. Samawi, H.; Al-Saleh, M.F.O. Inference on Overlapping Coefficients in Two Exponential Populations Using Ranked Set Sample. *Commun. Korean Stat. Soc.* **2008**, *15*, 147–159.
29. Hamza, D.; Papa, N.; Malick, M. Overlap Coefficients Based on Kullback-Leibler Divergence: Exponential Populations Case. *Int. J. Appl. Math. Res.* **2017**, *6*, 135–140.
30. Sibil, J.; Seemon, T.; Thomas, M. Interval Estimation of the Overlapping Coefficient of Two Exponential Distributions. *J. Stat. Theory Appl.* **2019**, *18*, 26–32.
31. Pianka, E. Niche Overlap and Diffuse Competition. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 2141–2145. [[CrossRef](#)]
32. Vieira, E.M.; Port, D. Niche overlap and resource partitioning between two sympatric fox species in southern Brazil. *J. Zool.* **2006**, *272*, 57–63. [[CrossRef](#)]
33. Jacqueline, B.C.; Mathew, S.C.; Georgeanna, S.; Mike, L. Dietary overlap and prey selectivity among sympatric carnivores: could dingoes suppress foxes through competition for prey? *J. Mammal.* **2011**, *92*, 590–600.
34. Sa-Oliveira, J.C.; Ronaldo, A.; Victoria, J.I.N. Diet and niche breadth and overlap in fish communities within the area affected by an Amazonian reservoir (Amapá, Brazil). *Ann. Braz. Acad. Sci.* **2014**, *86*, 383–405. [[CrossRef](#)] [[PubMed](#)]

-
35. Bodkin, R.G.; Klein, L.R.; Marwah, K. *A History of Macroeconometric Model-Building*; Edward Elgar Publishing: Cheltenham, UK, 1991.
 36. Doob, J. Application of the theory of martingales. *Calc. Des Probab. Ses Appl.* **1949**, *13*, 23–27.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.