

New Closed Form Estimators for the Beta Distribution

Victor Mooto Nawa ¹  and Saralees Nadarajah ^{2,*}¹ Department of Mathematics and Statistics, University of Zambia, Lusaka 10101, Zambia; vnawa@yahoo.com² Department of Mathematics, University of Manchester, Manchester M13 9PL, UK

* Correspondence: mbbssn2@manchester.ac.uk

Abstract: In this paper, we detail closed form estimators for beta distribution that are simpler than those proposed by Tamae, Irie and Kubokawa. The proposed estimators are shown to have smaller asymptotic variances and smaller asymptotic covariances compared to Tamae estimators and maximum likelihood estimators. The proposed estimators are also shown to perform better in real data applications.

Keywords: covariance; digamma function; trigamma function; variance

MSC: 62E99

1. Introduction

The most popular model for data in a finite interval is that based on the beta distribution. It is a two parameter distribution with probability density function given by

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

for $0 < x < 1$, $\alpha > 0$ and $\beta > 0$, where $B(\alpha, \beta)$ denotes the beta function defined by

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt,$$

which can be equivalently expressed as $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, where $\Gamma(\alpha)$ denotes the gamma function. We shall write $X \sim \text{Beta}(\alpha, \beta)$ to mean that a random variable X has the beta distribution. The mean and variance of $X \sim \text{Beta}(\alpha, \beta)$ are

$$E(X) = \frac{\alpha}{\alpha + \beta} \text{ and } \text{Var}(X) = \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}, \quad (1)$$

respectively. There are hundreds if not thousands of papers on the theory and applications of the beta distribution. It is impossible to cite all of these papers. Comprehensive accounts of the theory and applications of the beta distribution can be found in [1–5]. See also [6].

For a long time, estimating parameters of the beta distribution and other related distributions, such as the gamma distribution, was only possible through iteration methods. For instance, if x is an observation on $X \sim \text{Beta}(\alpha, \beta)$ then maximum likelihood estimators of α and β , say $\hat{\alpha}$ and $\hat{\beta}$, respectively, can be obtained by solving

$$\begin{aligned} \log x &= A, \\ \log(1-x) &= B, \end{aligned}$$



Citation: Nawa, V.M.; Nadarajah, S. New Closed Form Estimators for the Beta Distribution. *Mathematics* **2023**, *11*, 2799. <https://doi.org/10.3390/math11132799>

Academic Editor: Stelios Psarakis

Received: 24 May 2023

Revised: 15 June 2023

Accepted: 19 June 2023

Published: 21 June 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/>).

where $A = \psi(\alpha) - \psi(\alpha + \beta)$, $B = \psi(\beta) - \psi(\alpha + \beta)$ and $\psi(\alpha) = \frac{\partial \log \Gamma(\alpha)}{\partial \alpha}$ denotes the digamma function. Furthermore, the corresponding asymptotic variances and asymptotic covariance are given by

$$\sqrt{n}(\hat{\alpha} - \alpha, \hat{\beta} - \beta) \rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{G} \begin{pmatrix} D & \psi_1(C) \\ \psi_1(C) & E \end{pmatrix}\right)$$

as $n \rightarrow \infty$, where $\psi_1(\alpha) = \frac{\partial \psi(\alpha)}{\partial \alpha}$ denotes the trigamma function, $C = \alpha + \beta$, $D = \psi_1(\beta) - \psi_1(C)$, $E = \psi_1(\alpha) - \psi_1(C)$ and $G = \psi_1(\alpha)\psi_1(\beta) - \psi_1(\alpha)\psi_1(C) - \psi_1(\beta)\psi_1(C)$. Furthermore, let $F = \psi_1(\alpha) + \psi_1(\beta)$.

However, recently closed form estimators for the gamma and beta distributions have been proposed. Ref. [7] proposed closed form estimators for the gamma distribution by considering the likelihood equations of the generalized gamma distribution and taking the gamma distribution as a special case. Ref. [8] proposed closed form estimators for the gamma and beta distributions using the “score adjusted approach”. The estimators for the gamma distribution in [8] turned out to be the same as those obtained by [7].

Following the method applied by [8], two closed form and simpler estimators for the beta distribution are proposed in this paper. The two estimators appear to have smaller variances and smaller covariances for certain values of the parameters of the beta distribution.

The remainder of this paper is organized as follows. Section 2 re-derives the closed form estimators for the beta distribution due to [8]. Section 3 derives two new closed form estimators for the beta distribution. Section 4 establishes asymptotic normality of the new estimators. It also derives expressions for asymptotic variances and asymptotic covariances. Section 5 conducts a numerical comparison to show that the new estimators can be better than the estimators due to [8] as well as maximum likelihood estimators. A simulation study is conducted in Section 6 to check finite sample performance of all the estimators. Section 7 shows that the new estimators can provide better fit to a real dataset. Finally, some conclusions are given in Section 8. All computations in the paper were performed using the R software [9]. Sample code are given in Appendix A.

2. Tamae et al.’s [8] Closed Form Estimators

Let $X \sim \text{Beta}(\alpha, \beta)$. Using the facts

$$E(\log X) = \psi(\alpha) - \psi(C) \quad (2)$$

and

$$E(\log(1 - X)) = \psi(\beta) - \psi(C), \quad (3)$$

we can show that

$$E(X \log X) = \frac{\alpha}{C} \left[\psi(\alpha) - \psi(C) + \frac{1}{\alpha} - \frac{1}{C} \right] \quad (4)$$

and

$$E(X \log(1 - X)) = \frac{\alpha}{C} \left[\psi(\beta) - \psi(C) - \frac{1}{C} \right]. \quad (5)$$

Subtracting (5) from (4), using (1), (2) and (3) and solving for α gives

$$E(X \log X) - E(X \log(1 - X)) = E(X) \left[E(\log X) - E(\log(1 - X)) + \frac{1}{\alpha} \right], \quad (6)$$

which implies

$$\alpha = \frac{E(X)}{E(X \log X) - E(X \log(1 - X)) - E(X)E(\log X) + E(X)E(\log(1 - X))}.$$

Another way of expressing (6) is

$$E(X \log X) - E(X \log(1 - X)) = \frac{\alpha}{C} \left[E(\log X) - E(\log(1 - X)) + \frac{1}{\alpha} \right]. \quad (7)$$

Substituting α from (6) into (7) and solving for β yields

$$\beta = \frac{1 - E(X)}{E(X \log X) - E(X \log(1 - X)) - E(X)E(\log X) + E(X)E(\log(1 - X))}. \quad (8)$$

By the weak law of large numbers, we can replace the expectations in (6) and (8) by their sample versions yielding the closed form estimators proposed by [8] as

$$\begin{aligned} \hat{\alpha} &= \frac{\bar{X}}{\bar{X} \log \bar{X} - \bar{X} \log(1 - \bar{X}) - \bar{X} \log \bar{X} + \bar{X} \log(1 - \bar{X})} = \frac{\bar{X}}{\bar{X} \log\left(\frac{\bar{X}}{1 - \bar{X}}\right) - \bar{X} \log\left(\frac{\bar{X}}{1 - \bar{X}}\right)}, \\ \hat{\beta} &= \frac{1 - \bar{X}}{\bar{X} \log \bar{X} - \bar{X} \log(1 - \bar{X}) - \bar{X} \log \bar{X} + \bar{X} \log(1 - \bar{X})} = \frac{1 - \bar{X}}{\bar{X} \log\left(\frac{\bar{X}}{1 - \bar{X}}\right) - \bar{X} \log\left(\frac{\bar{X}}{1 - \bar{X}}\right)}. \end{aligned}$$

The corresponding asymptotic variances and asymptotic covariance are given by

$$\sqrt{n}(\hat{\alpha} - \alpha, \hat{\beta} - \beta) \rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \left[\sigma^2 C^2 F + 1\right] \begin{pmatrix} \alpha^2 & \alpha\beta \\ \alpha\beta & \beta^2 \end{pmatrix} - \frac{1}{C+1} \begin{pmatrix} \alpha\beta & C^2 - \alpha\beta \\ C^2 - \alpha\beta & \alpha\beta \end{pmatrix}\right)$$

as $n \rightarrow \infty$.

3. New Closed Form Estimators

In this section, we propose two new estimators for α and β . Throughout, we suppose that $X \sim \text{Beta}(\alpha, \beta)$.

Substituting (3) into (5) and rearranging the terms gives

$$E(X \log(1 - X)) = E(X) \left[E(\log(1 - X)) - \frac{1}{C} \right],$$

which implies

$$\frac{1}{C} = E(\log(1 - X)) - \frac{E(X \log(1 - X))}{E(X)}. \quad (9)$$

Multiplying (9) by α , using (1) and solving for α gives

$$\alpha = \frac{[E(X)]^2}{E(X)E(\log(1 - X)) - E(X \log(1 - X))}. \quad (10)$$

Substituting (10) in (9) and solving for β gives

$$\beta = \frac{E(X) - [E(X)]^2}{E(X)E(\log(1 - X)) - E(X \log(1 - X))}. \quad (11)$$

By the weak law of large numbers, we can replace the expectations in (10) and (11) by sample versions to obtain the estimators:

$$\hat{\alpha} = \frac{(\bar{X})^2}{\bar{X} \log(1-X) - \bar{X} \log(1-\bar{X})}, \text{ and } \hat{\beta} = \frac{\bar{X} - (\bar{X})^2}{\bar{X} \log(1-X) - \bar{X} \log(1-\bar{X})}. \quad (12)$$

Substituting (1) and (2) in (4) and rearranging yields

$$E(X \log X) = E(X) \left[E(\log X) + \frac{1}{\alpha} - \frac{1}{C} \right],$$

which implies

$$\frac{1}{\alpha} - \frac{1}{C} = \frac{E(X \log X) - E(X)E(\log X)}{E(X)}. \quad (13)$$

Multiplying (13) by α , using (1) and solving for α gives

$$\alpha = \frac{E(X)[1 - E(X)]}{E(X \log X) - E(X)E(\log X)}. \quad (14)$$

Substituting (14) in (13) and solving for β gives

$$\beta = \frac{[1 - E(X)]^2}{E(X \log X) - E(X)E(\log X)}. \quad (15)$$

By the weak law of large numbers, we can replace the expectations in (14) and (15) by their sample versions to obtain

$$\hat{\alpha} = \frac{\bar{X}(1 - \bar{X})}{\bar{X} \log \bar{X} - \bar{X} \log X} \text{ and } \hat{\beta} = \frac{(1 - \bar{X})^2}{\bar{X} \log \bar{X} - \bar{X} \log X}. \quad (16)$$

Note that $\frac{\hat{\beta}}{\hat{\alpha}} = \frac{1}{\bar{X}} - 1$. This relationship can be useful in deriving large sample properties given in Section 4.

4. Large Sample Properties

Large sample properties of the estimators given by (12) and (16) are derived in this section. Theorem 1 proves asymptotic normality of the estimators given by (12). Theorem 2 proves asymptotic normality of the estimators given by (16).

Theorem 1. *The estimators given in (12) satisfy*

$$\sqrt{n}(\hat{\alpha} - \alpha, \hat{\beta} - \beta) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}\right)$$

as $n \rightarrow \infty$, where

$$\Sigma_{11} = \frac{\alpha\beta}{C+1} (1 + C^2 D) + \frac{\alpha C(\alpha+1)(C^2+1)}{(C+1)^3},$$

$$\Sigma_{12} = \frac{\beta^2 C^2 D - \alpha(\beta+C)}{C+1} + \frac{\beta C}{(C+1)^3} [(\alpha+2)C^2 + C + \alpha + 1]$$

and

$$\Sigma_{22} = \frac{\alpha\beta}{C+1} + \frac{\beta^2 C}{\alpha(C+1)}(\beta C D - \alpha + 1) + \frac{2\beta C(\alpha+1)}{\alpha(C+1)^3}(\beta C^2 - \alpha C - \alpha).$$

Proof. Let the empirical means of X , $\log(1-X)$, and $X \log(1-X)$ be denoted by $\bar{X}$, $\bar{Y}$, and $\bar{Z}$, respectively. We can easily show that

$$E(\bar{X}) = \frac{\alpha}{C},$$

$$E(\bar{Y}) = B$$

and

$$E(\bar{Z}) = \frac{\alpha}{C} \left(B - \frac{1}{C} \right).$$

We can also show that

$$\text{Var}(\sqrt{n}\bar{Y}) = D. \quad (17)$$

Using the fact that

$$E[X^2 \log^2(1-X)] = \frac{\alpha(\alpha+1)}{C(C+1)} \{ [\psi(\beta) - \psi(C+2)]^2 + \psi_1(\beta) - \psi_1(C+2) \},$$

we can show that

$$\text{Var}(\sqrt{n}\bar{Z}) = \sigma^2 \left(B - \frac{1}{C} \right)^2 + \frac{\alpha(\alpha+1)}{C(C+1)} \left[\frac{2}{(C+1)^2} - \frac{2}{C+1} \left(B - \frac{1}{C} \right) + \frac{1}{C^2} + D \right]. \quad (18)$$

Similarly, we can show that

$$\text{Cov}(\sqrt{n}\bar{X}, \sqrt{n}\bar{Y}) = -\frac{\alpha}{C^2}, \quad (19)$$

$$\text{Cov}(\sqrt{n}\bar{X}, \sqrt{n}\bar{Z}) = \sigma^2 \left(B - \frac{1}{C} \right) - \frac{\alpha(\alpha+1)}{C(C+1)^2} \quad (20)$$

and

$$\text{Cov}(\sqrt{n}\bar{Y}, \sqrt{n}\bar{Z}) = \frac{\alpha}{C} \left(\frac{2}{C^2} - \frac{B}{C} + D \right). \quad (21)$$

By the central limit theorem,

$$\sqrt{n} \left[(\bar{X}, \bar{Y}, \bar{Z}) - \left\{ \frac{\alpha}{C}, B, \frac{\alpha}{C} [\psi(\beta) - \psi(C+1)] \right\} \right] \xrightarrow{d} N(\mathbf{0}_3, \Sigma)$$

as $n \rightarrow \infty$, where

$$\mathbf{0}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

and

$$\Sigma = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, Y) & \text{Cov}(X, Z) \\ \text{Cov}(X, Y) & \text{Var}(Y) & \text{Cov}(Y, Z) \\ \text{Cov}(X, Z) & \text{Cov}(Y, Z) & \text{Var}(Z) \end{pmatrix}.$$

The entries of this matrix are given by (1) and (17)–(21) as

$$\text{Var}(X) = \frac{\alpha\beta}{C^2(C+1)} = \sigma^2,$$

$$\text{Cov}(X, Y) = -\frac{\alpha}{C^2},$$

$$\text{Cov}(X, Z) = \sigma^2 \left(B - \frac{1}{C} \right) - \frac{\alpha(\alpha+1)}{C(C+1)^2},$$

$$\text{Var}(Y) = D,$$

$$\text{Cov}(Y, Z) = \frac{\alpha}{C} \left(\frac{2}{C^2} - \frac{B}{C} + D \right)$$

and

$$\text{Var}(Z) = \sigma^2 \left(B - \frac{1}{C} \right)^2 + \frac{\alpha(\alpha+1)}{C(C+1)} \left[\frac{2}{(C+1)^2} - \frac{2}{C+1} \left(B - \frac{1}{C} \right) + \frac{1}{C^2} + D \right].$$

Let $g_1(x, y, z) = \frac{x^2}{xy-z}$, $g_2(x, y, z) = \frac{x-x^2}{xy-z}$ and $(x_0, y_0, z_0) = \left[\frac{\alpha}{C}, B, \frac{\alpha}{C} [\psi(\beta) - \psi(C+1)] \right]$.
Then

$$\left. \frac{\partial g_1}{\partial x} \right|_{(x_0, y_0, z_0)} = C(2 - BC),$$

$$\left. \frac{\partial g_1}{\partial y} \right|_{(x_0, y_0, z_0)} = -\alpha C,$$

$$\left. \frac{\partial g_1}{\partial z} \right|_{(x_0, y_0, z_0)} = C^2,$$

$$\left. \frac{\partial g_2}{\partial x} \right|_{(x_0, y_0, z_0)} = C^2 \left(\frac{\beta - \alpha}{\alpha C} - \frac{\beta B}{\alpha} \right),$$

$$\left. \frac{\partial g_2}{\partial y} \right|_{(x_0, y_0, z_0)} = -\beta C$$

and

$$\left. \frac{\partial g_2}{\partial z} \right|_{(x_0, y_0, z_0)} = \frac{\beta C^2}{\alpha}.$$

Let

$$\mathbf{M} = \begin{pmatrix} C(2 - BC) & -\alpha C & C^2 \\ C^2 \left(\frac{\beta - \alpha}{\alpha C} - \frac{\beta B}{\alpha} \right) & -\beta C & \frac{\beta C^2}{\alpha} \end{pmatrix}.$$

Using the delta method,

$$\sqrt{n}(\hat{\alpha} - \alpha, \hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}_2, \mathbf{M}\mathbf{\Sigma}\mathbf{M}^T)$$

as $n \rightarrow \infty$, where

$$\mathbf{0}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$\mathbf{M}\Sigma\mathbf{M}^T = \begin{pmatrix} M\Sigma M_{11}^T & M\Sigma M_{12}^T \\ M\Sigma M_{12}^T & M\Sigma M_{22}^T \end{pmatrix},$$

where

$$M\Sigma M_{11}^T = \sigma^2 C^2 (2 - BC)^2 + 2\text{Cov}(X, Z) C^3 (2 - BC) + \text{Var}(Z) C^4 - \alpha^2 C^2 D,$$

$$M\Sigma M_{12}^T = \sigma^2 C^3 (2 - BC) \left(\frac{\beta - \alpha}{\alpha C} - \frac{\beta B}{\alpha} \right) + \text{Cov}(X, Z) \frac{C^3}{\alpha} (3\beta - 2\beta BC - \alpha) \\ + \text{Var}(Z) \frac{\beta C^4}{\alpha} - \alpha (C + \beta C^2 D)$$

and

$$M\Sigma M_{22}^T = \sigma^2 C^4 \left(\frac{\beta - \alpha}{\alpha C} - \frac{\beta B}{\alpha} \right)^2 + 2\text{Cov}(X, Z) \frac{\beta C^4}{\alpha} \left(\frac{\beta - \alpha}{\alpha C} - \frac{\beta B}{\alpha} \right) \\ + \text{Var}(Z) \frac{\beta^2 C^4}{\alpha^2} - \beta C (2 + \beta CD).$$

The theorem follows by simplification of these expressions. $\square$

Theorem 2. The estimators given in (16) satisfy

$$\sqrt{n}(\hat{\alpha} - \alpha, \hat{\beta} - \beta) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}\right)$$

as $n \rightarrow \infty$, where

$$\Sigma_{11} = \frac{\sigma^2 \alpha^2 C^2}{\beta^2} \left[1 + \frac{2C^2}{C+1} + \frac{\beta C^3}{(\alpha+1)(C+1)^2} \right] + \frac{\alpha^2 C}{\beta} \left[2 + \frac{\alpha CE}{C+1} \right] \\ + \frac{\alpha^3 C^3 (\alpha+1)}{\beta^2 (C+1)} \left[\frac{1}{C^2} + \frac{1}{(C+1)^2} - \frac{1}{\alpha^2} - \frac{1}{(\alpha+1)^2} \right],$$

$$\Sigma_{12} = \frac{\sigma^2 \alpha C^2}{\beta} \left[\frac{\beta}{\alpha} \left(1 + \frac{C^2}{C+1} \right) + \frac{\beta C^3}{(\alpha+1)(C+1)^2} + 2 + \frac{3C^2}{C+1} \right] + \alpha C \left[1 + \frac{\alpha CE}{C+1} \right] \\ + \frac{\alpha^2 C^3 (\alpha+1)}{\beta (C+1)} \left[\frac{1}{C^2} + \frac{1}{(C+1)^2} - \frac{1}{\alpha^2} - \frac{1}{(\alpha+1)^2} \right]$$

and

$$\Sigma_{22} = \sigma^2 C^2 \left[4 \left(1 + \frac{\beta}{\alpha} + \frac{C^2}{C+1} \right) + \frac{\beta}{\alpha} \left(\frac{\beta}{\alpha} + \frac{2C^2}{C+1} \right) + \frac{\beta C^3}{(\alpha+1)(C+1)^2} \right] + \frac{\alpha \beta C^2 E}{C+1} \\ + \frac{\alpha C^3 (\alpha+1)}{C+1} \left[\frac{1}{C^2} + \frac{1}{(C+1)^2} - \frac{1}{\alpha^2} - \frac{1}{(\alpha+1)^2} \right].$$

Proof. Let the empirical means of X , $\log(X)$ and $X \log(X)$ be denoted by $\bar{X}$, $\bar{U}$ and $\bar{V}$, respectively. We can easily show that

$$E(\bar{U}) = A$$

and

$$E(\bar{V}) = \frac{\alpha}{C} \left[A + \frac{1}{\alpha} - \frac{1}{C} \right].$$

We can also show that

$$\text{Var}(\sqrt{nU}) = E. \quad (22)$$

Using the fact that

$$E[X^2 \log^2(X)] = \frac{\alpha(\alpha+1)}{C(C+1)} \left\{ [\psi(\alpha+2) - \psi(C+2)]^2 + \psi_1(\alpha+2) - \psi_1(C+2) \right\},$$

we have

$$\begin{aligned} \text{Var}(\sqrt{nV}) &= \sigma^2 \left(A + \frac{\beta}{\alpha C} \right)^2 + \frac{\sigma^2 C}{C+1} \left[\frac{\beta}{(\alpha+1)(C+1)} + 2A + \frac{2\beta}{\alpha C} \right] \\ &\quad + \frac{\alpha(\alpha+1)}{C(C+1)} \left[\frac{1}{C^2} + \frac{1}{(C+1)^2} - \frac{1}{\alpha^2} - \frac{1}{(\alpha+1)^2} + E \right]. \end{aligned} \quad (23)$$

Similarly, we can show that

$$\text{Cov}(\sqrt{nX}, \sqrt{nU}) = \frac{\beta}{C^2}, \quad (24)$$

$$\text{Cov}(\sqrt{nX}, \sqrt{nV}) = \sigma^2 \left[A + \frac{\beta}{\alpha C} + \frac{C}{C+1} \right] \quad (25)$$

and

$$\text{Cov}(\sqrt{nU}, \sqrt{nV}) = \frac{\alpha}{C} \left[\frac{A\beta}{\alpha C} + E - \frac{2\beta}{\alpha C^2} \right]. \quad (26)$$

By the central limit theorem,

$$\sqrt{n} \left[(\bar{X}, \bar{U}, \bar{V}) - \left\{ \frac{\alpha}{C}, A, \frac{\alpha}{C} [\psi(\alpha+1) - \psi(C+1)] \right\} \right] \xrightarrow{d} N(\mathbf{0}_3, \Sigma)$$

as $n \rightarrow \infty$, where

$$\Sigma = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, U) & \text{Cov}(X, V) \\ \text{Cov}(X, U) & \text{Var}(U) & \text{Cov}(U, V) \\ \text{Cov}(X, V) & \text{Cov}(U, V) & \text{Var}(V) \end{pmatrix}.$$

The entries of this matrix are given by (1) and (22)–(26) as

$$\text{Var}(X) = \frac{\alpha\beta}{C^2(C+1)} = \sigma^2,$$

$$\text{Cov}(X, U) = \frac{\beta}{C^2},$$

$$\text{Cov}(X, V) = \sigma^2 \left[A + \frac{\beta}{\alpha C} + \frac{C}{C+1} \right],$$

$$\text{Var}(U) = E,$$

$$\text{Cov}(U, V) = \frac{\alpha}{C} \left[\frac{A\beta}{\alpha C} + E - \frac{2\beta}{\alpha C^2} \right]$$

and

$$\begin{aligned} \text{Var}(V) = & \sigma^2 \left(A + \frac{\beta}{\alpha C} \right)^2 + \frac{\sigma^2 C}{C+1} \left[\frac{\beta}{(\alpha+1)(C+1)} + 2A + \frac{2\beta}{\alpha C} \right] \\ & + \frac{\alpha(\alpha+1)}{C(C+1)} \left[\frac{1}{C^2} + \frac{1}{(C+1)^2} - \frac{1}{\alpha^2} - \frac{1}{(\alpha+1)^2} + E \right]. \end{aligned}$$

Let $h_1(x, y, z) = \frac{x-x^2}{xz-y}$, $h_2(x, y, z) = \frac{1-2x+x^2}{z-xy}$ and $(x_0, y_0, z_0) = [\frac{\alpha}{C}, A, \frac{\alpha}{C}[\psi(\alpha+1) - \psi(C+1)]]$. Then,

$$\frac{\partial h_1}{\partial x} \Big|_{(x_0, y_0, z_0)} = \frac{C}{\beta} (\alpha AC + \beta - \alpha),$$

$$\frac{\partial h_1}{\partial y} \Big|_{(x_0, y_0, z_0)} = \frac{\alpha^2 C}{\beta},$$

$$\frac{\partial h_1}{\partial z} \Big|_{(x_0, y_0, z_0)} = -\frac{\alpha C^2}{\beta},$$

$$\frac{\partial h_2}{\partial x} \Big|_{(x_0, y_0, z_0)} = AC^2 - 2C,$$

$$\frac{\partial h_2}{\partial y} \Big|_{(x_0, y_0, z_0)} = \alpha C$$

and

$$\frac{\partial h_2}{\partial z} \Big|_{(x_0, y_0, z_0)} = C^2.$$

Let

$$\mathbf{W} = \begin{pmatrix} \frac{C}{\beta} (\alpha AC + \beta - \alpha) & \frac{\alpha^2 C}{\beta} & -\frac{\alpha C^2}{\beta} \\ AC^2 - 2C & \alpha C & -C^2 \end{pmatrix}.$$

Using the delta method,

$$\sqrt{n}(\hat{\alpha} - \alpha, \hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}_2, \mathbf{W}\Sigma\mathbf{W}^T)$$

as $n \rightarrow \infty$, where

$$\mathbf{W}\Sigma\mathbf{W}^T = \begin{pmatrix} \mathbf{W}\Sigma\mathbf{W}_{11}^T & \mathbf{W}\Sigma\mathbf{W}_{12}^T \\ \mathbf{W}\Sigma\mathbf{W}_{12}^T & \mathbf{W}\Sigma\mathbf{W}_{22}^T \end{pmatrix},$$

where

$$\mathbf{W}\Sigma\mathbf{W}_{11}^T = \frac{2\alpha^2 C}{\beta} - \frac{\sigma^2 \alpha C^2}{\beta^2} (\alpha AC + \beta - \alpha) \left(1 + AC + \frac{\beta}{\alpha} + \frac{2C^2}{C+1} \right) - \frac{\alpha^4 C^2 E}{\beta^2} + \frac{\alpha^2 C^4}{\beta^2} \text{Var}(V),$$

$$\begin{aligned} \mathbf{W}\Sigma\mathbf{W}_{12}^T = & -\frac{\sigma^2 \alpha C^2}{\beta} \left[AC \left(AC + \frac{2\beta}{\alpha} + \frac{2C^2}{C+1} \right) + \frac{\beta}{\alpha} \left(-1 + \frac{\beta}{\alpha} + \frac{C^2}{C+1} \right) - 2 - \frac{3C^2}{C+1} \right] \\ & + \alpha C - \frac{\alpha^3 C^2 E}{\beta} + \frac{\alpha C^4}{\beta} \text{Var}(V) \end{aligned}$$

and

$$W\Sigma W_{22}^T = -\sigma^2 C^2 (AC - 2) \left(2 + \frac{2\beta}{\alpha} + \frac{2C^2}{C+1} + AC \right) - \alpha^2 C^2 E + C^4 \text{Var}(V).$$

The theorem follows by simplification of these expressions. $\square$

5. Numerical Comparison

In this section, we compare the asymptotic variances and asymptotic covariances of the estimators given by Theorems 1 and 2, Tamae et al.'s estimators and maximum likelihood estimators. Figures 1–3 show how the asymptotic variances and asymptotic covariances of these estimators vary versus $\alpha = 0.1, 0.2, \dots, 10$ and $\beta = 0.1, 0.2, \dots, 10$. Figures 4–6 show the differences in asymptotic variances and the differences in asymptotic covariances for two of the estimators at a time as $\alpha = 0.1, 0.2, \dots, 10$ and $\beta = 0.1, 0.2, \dots, 10$. We shall refer to Tamae et al.'s estimators as TIK estimators.

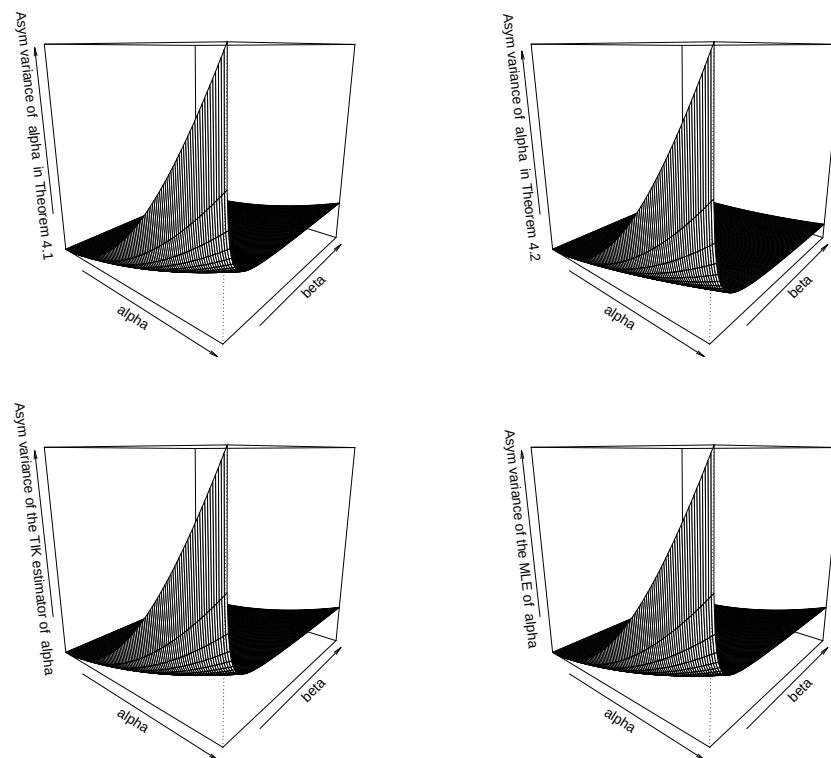


Figure 1. Asymptotic variance of the estimator of α versus α and β for the estimator given by Theorem 1 (**top left**), the estimator given by Theorem 2 (**top right**), TIK estimator (**bottom left**), and maximum likelihood estimator (**bottom right**).

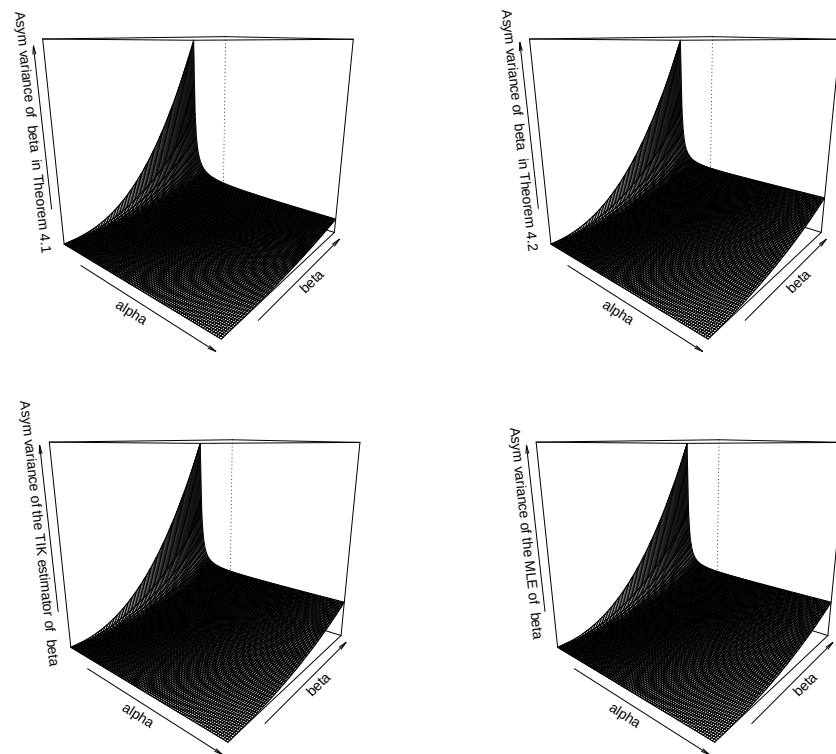


Figure 2. Asymptotic variance of the estimator of β versus α and β for the estimator given by Theorem 1 (top left), the estimator given by Theorem 2 (top right), TIK estimator (bottom left), and maximum likelihood estimator (bottom right).

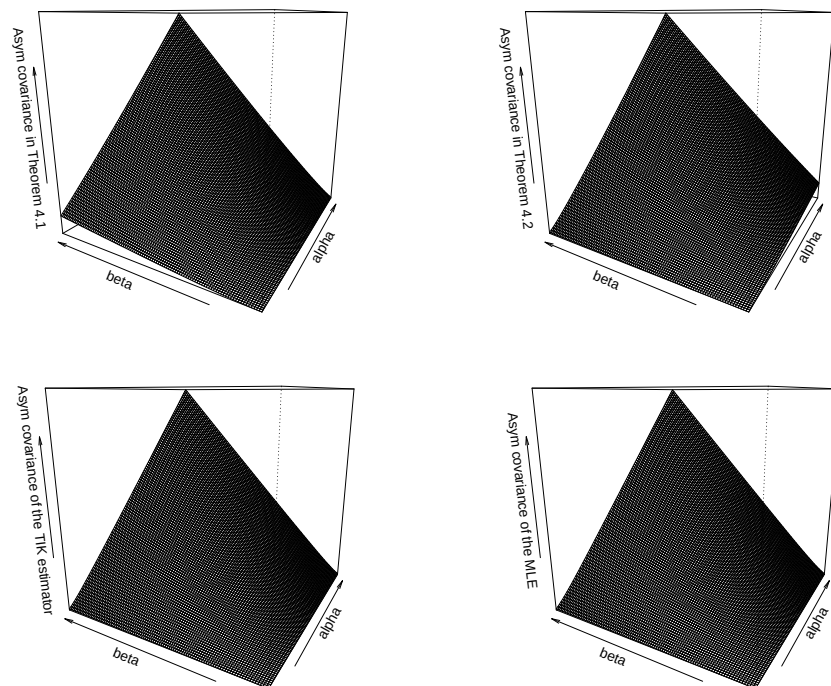


Figure 3. Asymptotic covariance between the estimators of α and β versus α and β for the estimators given by Theorem 1 (top left), the estimators given by Theorem 2 (top right), TIK estimators (bottom left), and maximum likelihood estimators (bottom right).

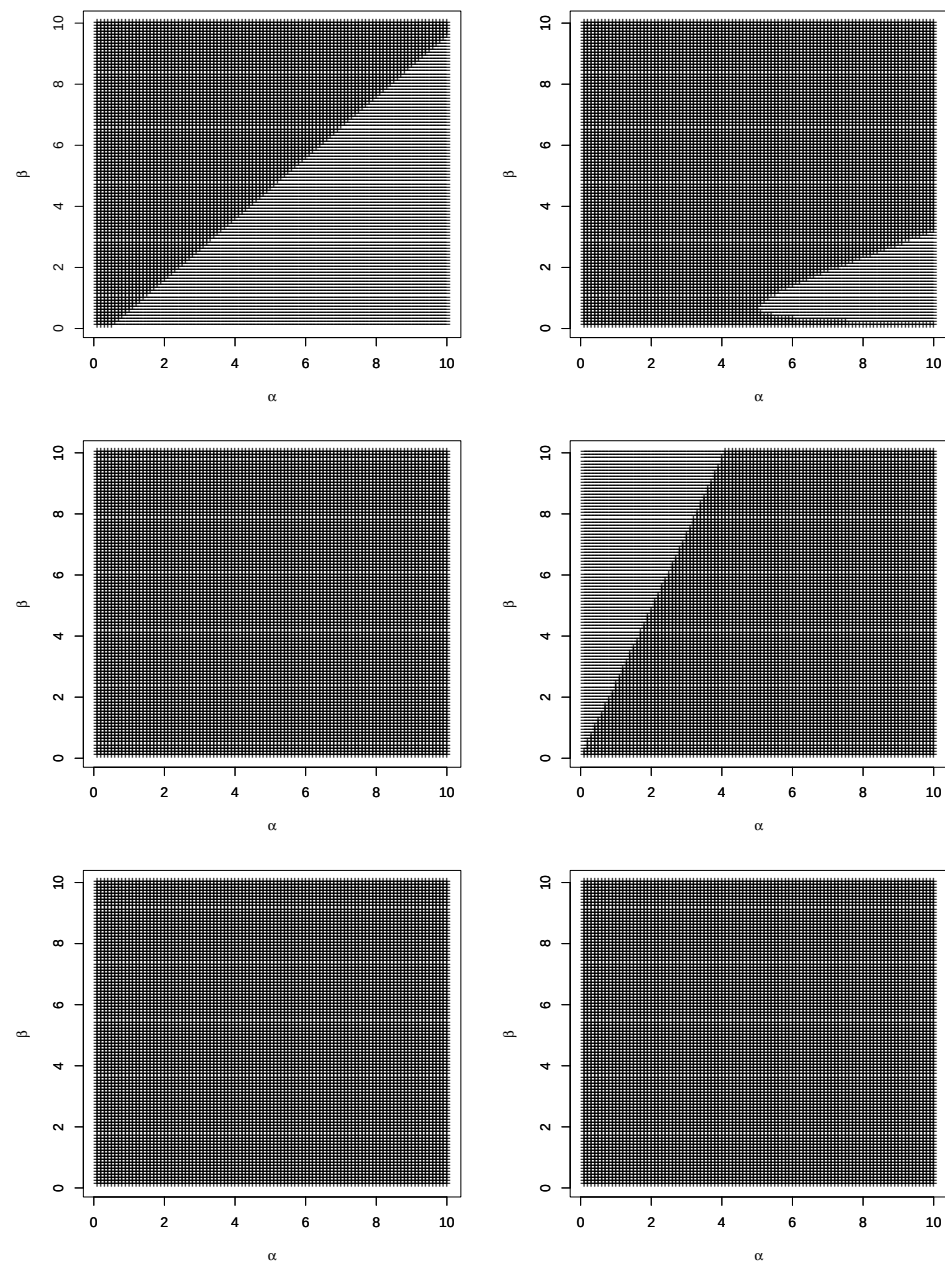


Figure 4. Sign of the difference between the asymptotic variances of the estimators of α given by Theorems 1 and 2 (**top left**); sign of the difference between the asymptotic variances of the estimator of α given by Theorem 1 and TIK estimator (**top right**); sign of the difference between the asymptotic variances of the estimator of α given by Theorem 1 and maximum likelihood estimator (**middle left**); sign of the difference between the asymptotic variances of the estimator of α given by Theorem 2 and TIK estimator (**middle right**); sign of the difference between the asymptotic variances of the estimator of α given by Theorem 2 and maximum likelihood estimator (**bottom left**); sign of the difference between the asymptotic variances of TIK estimator of α and maximum likelihood estimator (**bottom right**).

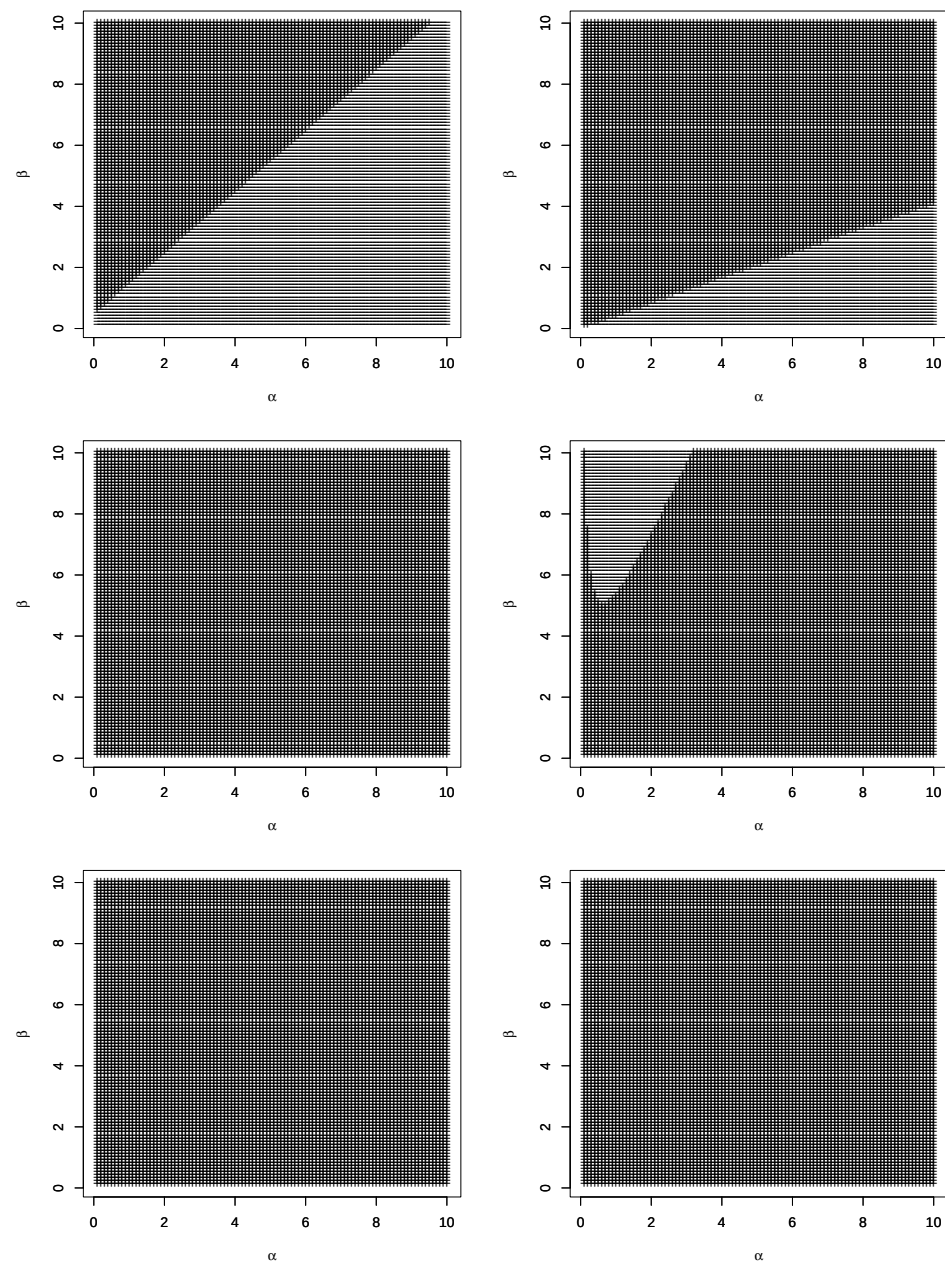


Figure 5. Sign of the difference between the asymptotic variances of the estimators of β given by Theorems 1 and 2 (**top left**); sign of the difference between the asymptotic variances of the estimator of β given by Theorem 1 and TIK estimator (**top right**); sign of the difference between the asymptotic variances of the estimator of β given by Theorem 1 and maximum likelihood estimator (**middle left**); sign of the difference between the asymptotic variances of the estimator of β given by Theorem 2 and TIK estimator (**middle right**); sign of the difference between the asymptotic variances of the estimator of β given by Theorem 2 and maximum likelihood estimator (**bottom left**); sign of the difference between the asymptotic variances of TIK estimator of β and maximum likelihood estimator (**bottom right**).

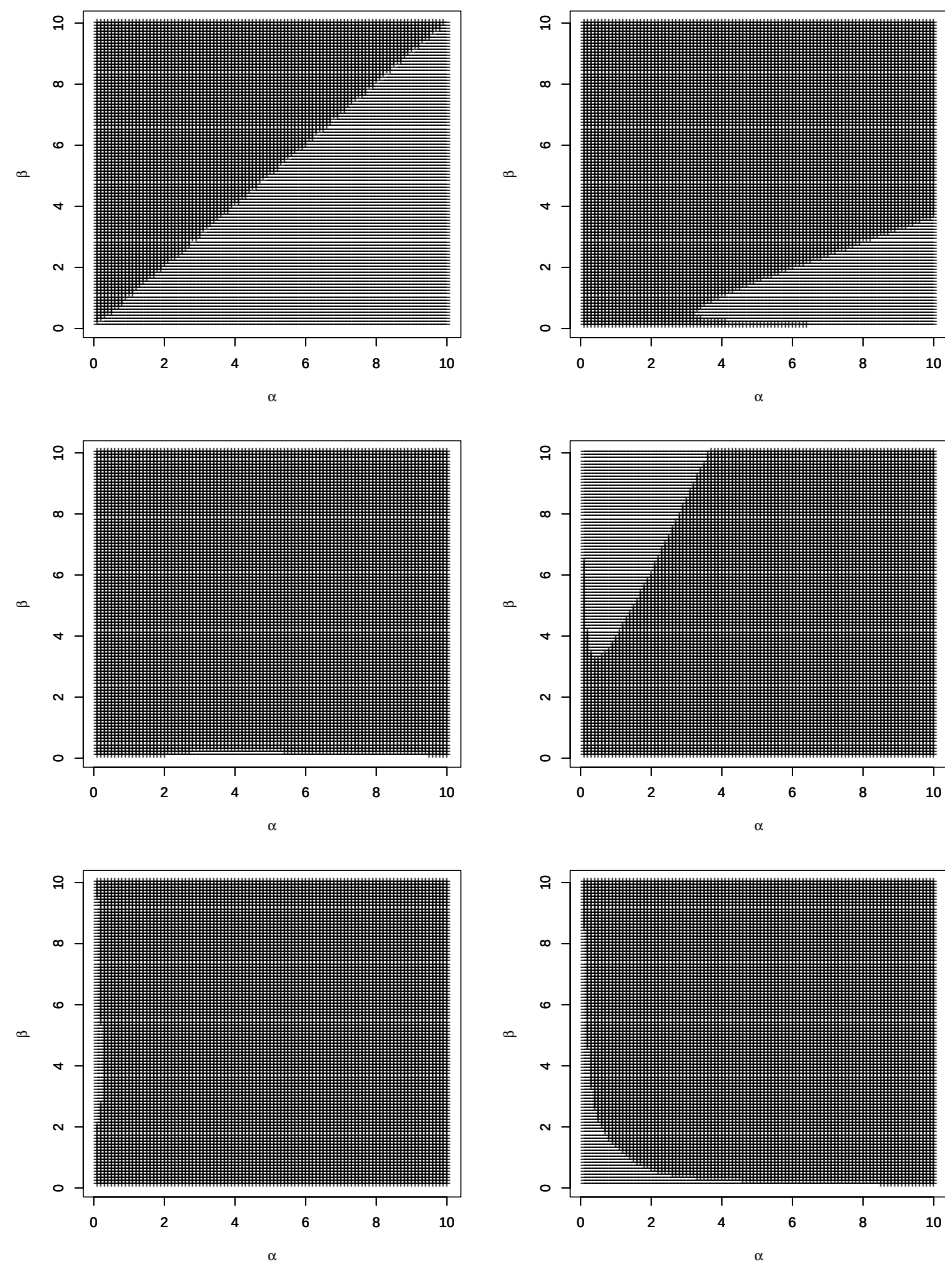


Figure 6. Sign of the difference between the asymptotic covariances of the estimators given by Theorems 1 and 2 (**top left**); sign of the difference between the asymptotic covariances of the estimators given by Theorem 1 and TIK estimator (**top right**); sign of the difference between the asymptotic covariances of the estimators given by Theorem 1 and maximum likelihood estimator (**middle left**); sign of the difference between the asymptotic covariances of the estimators given by Theorem 2 and TIK estimator (**middle right**); sign of the difference between the asymptotic covariances of the estimators given by Theorem 2 and maximum likelihood estimator (**bottom left**); sign of the difference between the asymptotic covariances of TIK estimator and maximum likelihood estimator (**bottom right**).

We can observe the following from Figures 1–3. Asymptotic variances of all of the estimators for α increase with respect to α ; asymptotic variances of all of the estimators for α decrease with respect to β ; asymptotic variances of all of the estimators for β decrease with respect to α ; asymptotic variances of all of the estimators for β increase with respect to β ; asymptotic covariances of all of the estimators increase with respect to α ; and asymptotic

covariances of all of the estimators increase with respect to β . All four estimators in each of the Figures 1–3 appear to behave similarly.

We can observe the following from Figures 4–6. With respect to asymptotic variances of the estimators for α , the estimator in Theorem 1 is more efficient than the corresponding estimator in Theorem 2 for all $\alpha > \beta$; with respect to asymptotic variances of the estimators for α , the estimator in Theorem 1 is more efficient than the corresponding TIK estimator in a parabolic region containing large values of α and small values of β ; with respect to asymptotic variances of the estimators for α , the maximum likelihood estimator is more efficient than the corresponding estimator in Theorem 1 for all values of α and β ; with respect to asymptotic variances of the estimators for α , the estimator in Theorem 2 is more efficient than the corresponding TIK estimator for all $\beta > 2\alpha$; with respect to asymptotic variances of the estimators for α , the maximum likelihood estimator is more efficient than the corresponding estimator in Theorem 2 and the corresponding TIK estimator for all values of α and β ; with respect to asymptotic variances of the estimators for β , the estimator in Theorem 1 is more efficient than the corresponding estimator in Theorem 2 for all $\beta < \alpha$; with respect to asymptotic variances of the estimators for β , the estimator in Theorem 1 is more efficient than the corresponding TIK estimator for all $2\beta < \alpha$; with respect to asymptotic variances of the estimators for β , the estimator in Theorem 2 is more efficient than the corresponding TIK estimator in a parabolic region containing small values of α and large values of β ; with respect to asymptotic variances of the estimators for β , the maximum likelihood estimator is more efficient than the three remaining estimators for all values of α and β ; with respect to asymptotic covariances, the estimators in Theorem 1 are more efficient than the corresponding estimators in Theorem 2 for all $\beta < \alpha$; with respect to asymptotic covariances, the estimators in Theorem 1 are more efficient than the corresponding TIK estimators in a parabolic region containing large values of α and small values of β ; with respect to asymptotic covariances, the estimators in Theorem 1 are more efficient than the corresponding maximum likelihood estimators for all small values of β ; with respect to asymptotic covariances, the estimators in Theorem 2 are more efficient than the corresponding TIK estimators in a parabolic region containing small values of α and large values of β ; with respect to asymptotic covariances, the estimators in Theorem 2 are more efficient than the corresponding maximum likelihood estimators for all small values of β ; and with respect to asymptotic covariances, the TIK estimators are more efficient than the corresponding maximum likelihood estimators in a hyperbolic region containing either small values of both α and β or small values of α and large values of β or large values of α and small values of β .

In summary, we can see that the estimators in Theorem 1 can be more efficient than the TIK estimators if either $2\beta < \alpha$ or in a parabolic region containing large values of α and small values of β . The estimators in Theorem 1 can be more efficient than the maximum likelihood estimators for small values of β . The estimators in Theorem 2 can be more efficient than the TIK estimators if either $\beta > 2\alpha$ or in a parabolic region containing large values of β and small values of α . The estimators in Theorem 2 can be more efficient than the maximum likelihood estimators for small values of β . The estimators in Theorem 1 can be more efficient than the estimators in Theorem 2 if $\alpha > \beta$.

6. Simulation Study

Finite sample performances of the estimators given by Theorems 1 and 2, Tamae et al.'s estimators and maximum likelihood estimators are compared in this section. We use the following simulation scheme:

- (i) Simulate a random sample of size n from a beta distribution with parameters $a = 2$ and $b = 2$;
- (ii) Compute the estimators given by Theorems 1 and 2, Tamae et al.'s estimators and maximum likelihood estimators;
- (iii) Repeat steps (i) and (ii) 1000 times;
- (iv) Compute the bias and mean squared error of the estimators;

- (v) Compute also the p -value for [10]'s test of bivariate normality of the estimators;
- (vi) Repeat steps (i) to (v) for $n = 10, 20, \dots, 1000$.

Biases versus n for all four estimators are shown in Figure 7. Mean squared errors versus n for all four estimators are shown in Figure 8. p -values versus n for all four estimators are shown in Figure 9.

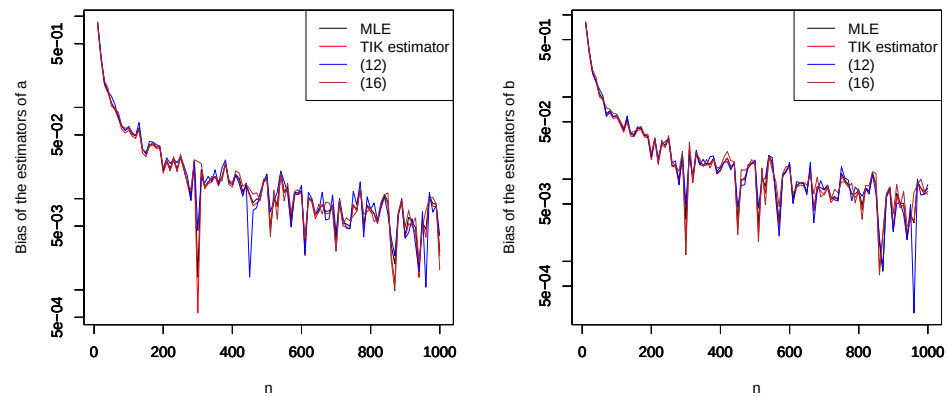


Figure 7. Biases of the estimators of a (left) and b (right) versus n .

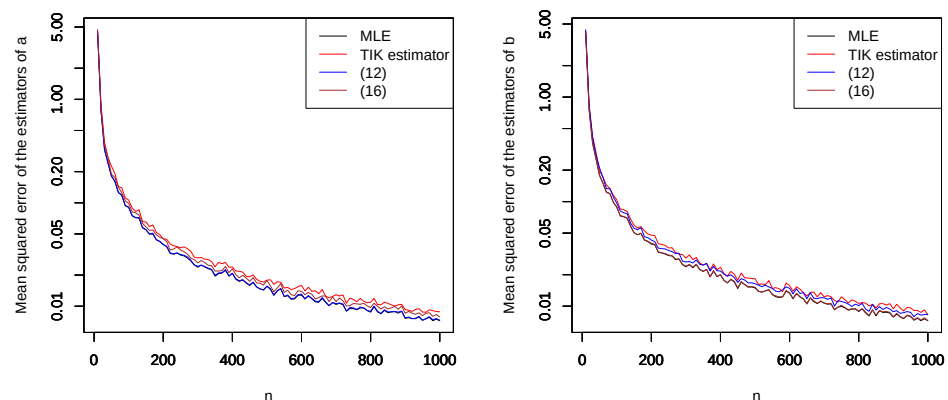


Figure 8. Mean squared errors of the estimators of a (left) and b (right) versus n .

We observe the following from the figures: increasing n leads the biases generally decreasing to zero; the relative performances of the estimators with respect to bias appear similar; the biases appear reasonably small for all $n \geq 200$; increasing n leads the mean squared errors generally decreasing to zero; the mean squared errors appear largest for Tamae et al.'s estimators and smallest for maximum likelihood estimators; the mean squared errors appear reasonably small for all $n \geq 200$; asymptotic normality of all estimators appears to have been achieved for all $n \geq 400$.

These observations are for $a = 2$ and $b = 2$. However, the same observations held for a wide range of other values of a and b . In particular, increasing n always led to the biases decreasing to zero, increasing n always led to the mean squared errors decreasing to zero, and the mean squared errors always appeared largest for Tamae et al.'s estimators and smallest for maximum likelihood estimators.

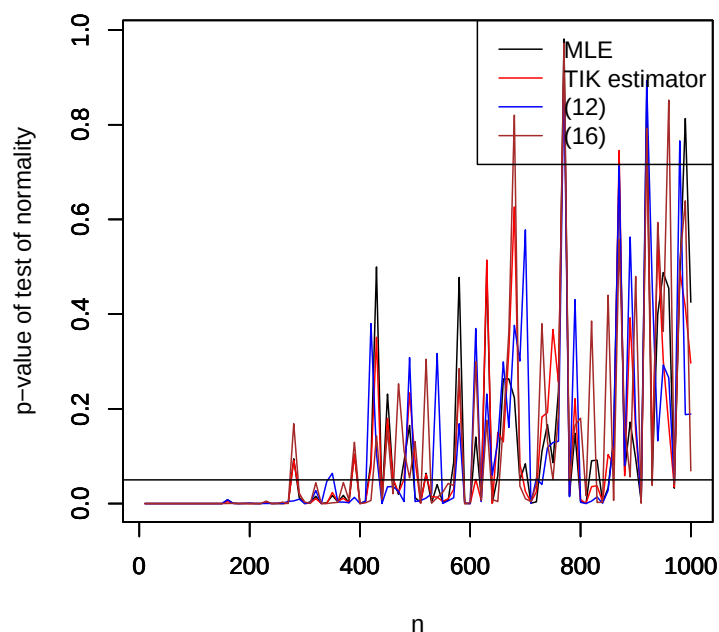


Figure 9. *P*-values for [10]’s test of bivariate normality of the estimators versus n . The horizontal line corresponds to 0.05.

7. Real Data Illustration

In this section, we compare the performance of the estimators given by Theorems 1 and 2, Tamae et al.’s estimators and maximum likelihood estimators using a real dataset. The data are the proportion voting Remain in the Brexit (EU referendum) poll outcomes for 127 polls from January 2016 to the referendum date on June 2016. The actual data values are 0.52, 0.55, 0.51, 0.49, 0.44, 0.54, 0.48, 0.41, 0.45, 0.42, 0.53, 0.45, 0.44, 0.44, 0.42, 0.42, 0.37, 0.46, 0.43, 0.39, 0.45, 0.44, 0.46, 0.40, 0.48, 0.42, 0.44, 0.45, 0.43, 0.43, 0.48, 0.41, 0.43, 0.40, 0.41, 0.42, 0.44, 0.51, 0.44, 0.44, 0.41, 0.41, 0.45, 0.55, 0.44, 0.44, 0.52, 0.55, 0.47, 0.43, 0.55, 0.38, 0.36, 0.38, 0.44, 0.42, 0.44, 0.43, 0.42, 0.49, 0.39, 0.41, 0.45, 0.43, 0.44, 0.51, 0.51, 0.49, 0.48, 0.43, 0.53, 0.38, 0.40, 0.39, 0.35, 0.45, 0.42, 0.40, 0.39, 0.44, 0.51, 0.39, 0.35, 0.41, 0.51, 0.45, 0.49, 0.40, 0.48, 0.41, 0.46, 0.47, 0.43, 0.45, 0.48, 0.49, 0.40, 0.40, 0.40, 0.39, 0.41, 0.39, 0.48, 0.48, 0.37, 0.38, 0.42, 0.51, 0.45, 0.40, 0.54, 0.36, 0.43, 0.49, 0.41, 0.36, 0.42, 0.38, 0.55, 0.44, 0.54, 0.41, 0.52, 0.42, 0.38, 0.42, 0.44.

We fitted the beta distribution using the four estimators. The estimates of (α, β) of the beta distribution were (45.315, 57.107), (45.196, 56.974), (44.061, 55.543) and (46.139, 58.163) for the maximum likelihood estimators, TIK estimators, the estimators given by Theorem 1 and the estimators given by Theorem 2, respectively.

The standard errors of the estimates of (α, β) can be obtained using Theorems 1 and 2. However, the theorems are asymptotic and do not account for variability of the estimators. We used the following bootstrapping procedure to check variability:

- (i) Remove the i th observation from the data;
- (ii) Refit the beta distribution to the modified data using the four estimators;
- (iii) Repeat steps (i) and (ii) for $i = 1, 2, \dots, 127$.

The histograms not shown here looked similar for all four estimators. However, the standard deviations of the histograms were 0.448, 0.448, 0.435 and 0.459 for the maximum likelihood estimator, TIK estimator, the estimator given by Theorem 1 and the estimator given by Theorem 2, respectively, of α . The standard deviations of the histograms were 0.600, 0.599, 0.580 and 0.617 for the maximum likelihood estimator, TIK estimator, the estimator given by Theorem 1 and the estimator given by Theorem 2, respectively, of β . Hence, the estimators given by Theorem 1 provide the best performance with respect to accuracy of fit.

The probability and quantile plots not shown here looked similar for all four estimators. However, the sums of the squares of the deviations between expected and observed

probabilities were 0.2970, 0.2932, 0.2884, and 0.2981 for the maximum likelihood estimators, TIK estimators, the estimators given by Theorem 1 and the estimators given by Theorem 2, respectively. The sums of the squares of the deviations between expected and observed quantiles were 0.0101, 0.0101, 0.0100, and 0.0102 for the maximum likelihood estimators, TIK estimators, the estimators given by Theorem 1 and the estimators given by Theorem 2, respectively. Hence, the estimators given by Theorem 1 provide the best performance with respect to goodness of fit.

8. Conclusions

Motivated by [8], we have proposed two closed form estimators for the parameters of the beta distribution. We have proved their asymptotic normality and derived expressions for asymptotic variances and asymptotic covariances. Through a numerical comparison and a real data application, we have shown that proposed estimators can be more efficient than Tamae et al.'s estimators as well as maximum likelihood estimators for certain parameter values. We have also performed a simulation study to check finite sample behavior of all the estimators.

Future work is theoretical comparison of the performances of proposed estimators, Tamae et al.'s estimators and maximum likelihood estimators for finite n . Other future work are to see if closed form estimators can be derived for known extensions of the beta distribution, bivariate beta distributions, multivariate beta distributions, matrix variate beta distributions and complex variate beta distributions [1,11].

Author Contributions: Methodology, V.M.N.; Software, S.N.; Formal analysis, V.M.N.; Writing—original draft, V.M.N.; Supervision, S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This paper has received no external funding.

Data Availability Statement: Data and code are given as part of the manuscript.

Acknowledgments: The authors would like to thank the Editor and the four referees for careful reading and comments which greatly improved the paper.

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

Appendix A. R Code

The following code computes the estimators given by Theorems 1 and 2, Tamae et al.'s estimators and maximum likelihood estimators.

```
f=function (p)
{tt=1.0e20
if (p[1]>0&p[2]>0) tt=-sum(dbeta(x,shape1=p[1],shape2=p[2],log=TRUE))
return(tt)}
est=optim(f,par=c(1,1))
a=est$par[1]
b=est$par[2]

a2=mean(x)/(mean(x*log(x/(1-x)))-mean(x)*mean(log(x/(1-x))))
b2=(1-mean(x))/(mean(x*log(x/(1-x)))-mean(x)*mean(log(x/(1-x))))

a3=(mean(x)**2/(mean(x)*mean(log(1-x))-mean(x*log(1-x))))
b3=(mean(x)-(mean(x)**2)/(mean(x)*mean(log(1-x))-mean(x*log(1-x))))

a4=(mean(x))*(1-mean(x))/(mean(x*log(x))-mean(x)*mean(log(x)))
b4=(1-mean(x)**2/(mean(x*log(x))-mean(x)*mean(log(x))))
```

The following code computes the asymptotic variances and asymptotic covariances of the estimators given by Theorems 1 and 2, Tamae et al.'s estimators and maximum likelihood estimators.

```

A=digamma(a)-digamma(a+b)
B=digamma(b)-digamma(a+b)
C=a+b
D=trigamma(b)-trigamma(a+b)
E=trigamma(a)-trigamma(a+b)
F=trigamma(a)+trigamma(b)
G=trigamma(a)*trigamma(b)-trigamma(a)*trigamma(a+b)-trigamma(b)*trigamma(a+b)
s2=a*b/((a+b)**2*(a+b+1))

vic111=a*b*(1+C*C*D)/(C+1)+a*(a+1)*C*(C*C+1)/(C+1)**3
vic112=(b*b*C*C*D-a*(b+C))/(C+1)+(b*C)*((a+2)*C*C+C+a+1)/(C+1)**3
vic122=a*b/(C+1)+b*b*C*(b*C*D-a+1)/(a*(C+1))+
2*(a+1)*b*C*(b*C*C-a*C-a)/(a*(C+1)**3)

vic211=s2*a*a*C*C*(1+2*C*C/(C+1)+b*C**3/((a+1)*(C+1)**2))/(b*b)+
a*a*C*(2+a*C*E/(C+1))/b+a**3*C**3*(a+1)*(1/C**2+1/(C+1)**2-
1/(a*a)-1/(a+1)**2)/(b*b*(C+1))
vic212=s2*a*C*C*((b/a)*(1+C*C/(C+1))+b*C**3/((a+1)*(C+1)**2)+
2+3*C*C/(C+1))/b+a*C*(1+a*C*E/(C+1))+a*a*C**3*(a+1)*(1/C**2+
1/(C+1)**2-1/a**2-1/(a+1)**2)/(b*(C+1))
vic222=s2*C*C*(4*(1+b/a+C**2/(C+1))+(b/a)*(b/a+2*C**2/(C+1))+
b*C**3/((a+1)*(C+1)**2))+a*b*C*C*E/(C+1)+a*(a+1)*C**3*(1/C**2+
1/(C+1)**2-1/a**2-1/(a+1)**2)/(C+1)

TIK11=(s2*C*C*F+1)*a**2-a*b/(C+1)
TIK22=(s2*C*C*F+1)*b**2-a*b/(C+1)
TIK12=(s2*C*C*F+1)*a*b-(C*C-a*b)/(C+1)

MLE11=D/G
MLE22=E/G
MLE12=(trigamma(a+b))/G

```

References

1. Gupta, A.K.; Nadarajah, S. *Handbook of Beta Distribution and Its Applications*; CRC Press: New York, NY, USA, 2004.
2. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; John Wiley and Sons: New York, NY, USA, 1995; Volume 2.
3. Kotz, S.; van Dorp, J.R. *Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications*; World Scientific Publishing Company: Hackensack, NJ, USA, 2004.
4. Larson, H.J. *Introduction to Probability Theory and Statistical Inference*; John Wiley and Sons: New York, NY, USA, 1982.
5. Seber, G.A.F. *The Linear Model and Hypothesis: A General Unifying Theory*; Springer: Cham, Switzerland, 2015.
6. Ferrari, S.L.P.; Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [\[CrossRef\]](#)
7. Ye, Z.-S.; Chen, N. Closed-form estimators for the gamma distribution derived from likelihood equations. *Am. Stat.* **2017**, *71*, 177–181. [\[CrossRef\]](#)
8. Tamae, H.; Irie, K.; Kubokawa, T. A score-adjusted approach to closed form estimators for the gamma and beta distributions. *Jpn. J. Stat. Data Sci.* **2020**, *3*, 543–561. [\[CrossRef\]](#)
9. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.

10. Henze, N.; Zirkler, B. A class of invariant consistent tests for multivariate normality. *Commun. Stat.-Theory Methods* **1990**, *19*, 3595–3617. [[CrossRef](#)]
11. Balakrishnan, N.; Lai, C.D. *Continuous Bivariate Distributions*; Springer Verlag: New York, NY, USA, 2009.

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.