axioms

Article

A Bimodal Extension of the Beta-Binomial Distribution with Applications

Jimmy Reyes [1], Josu Najera-Zuloaga [2], Dae-Jin Lee [3], Jaime Arrué [1] and Yuri A. Iriarte [1,*]

1 Departamento de Estadística y Ciencia de Datos, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1270300, Chile; jimmy.reyes@uantof.cl (J.R.); jaime.arrue@uantof.cl (J.A.)
2 Department of Mathematics, University of the Basque Country UPV/EHU, 48940 Leioa, Spain; josu.najera@ehu.eus
3 School of Science and Technology, IE University, 28046 Madrid, Spain; dae-jin.lee@ie.edu
* Correspondence: yuri.iriarte@uantof.cl

Abstract: In this paper, we propose an alternative distribution to model count data exhibiting uni/bimodality. It arises as a weighted version of the beta-binomial distribution, which is defined by a parametric weight function that admits up to two modes for the resulting probability mass function. Like the baseline beta-binomial distribution, the proposed distribution performs well in modeling overdispersed binomial data. Structural properties of the new distribution are studied. Raw moments are derived, which are used to describe the dispersion behavior relative to the mean and the skewness behavior. Parameter estimation is carried out using the maximum likelihood method. A simulation study is conducted in order to illustrate the behavior of the estimators. Finally, two applications illustrating the usefulness of the proposal are presented.

Keywords: beta-binomial distribution; bimodality; count data; maximum likelihood; moments; overdispersion

MSC: 62E10; 62F10

check for updates

Citation: Reyes, J.; Najera-Zuloaga, J.; Lee, D.-J.; Arrue, J.; Iriarte, Y.A. A Bimodal Extension of the Beta-Binomial Distribution with Applications. *Axioms* **2024**, *13*, 662. https://doi.org/10.3390/axioms13100662

Academic Editor: Simeon Reich

Received: 24 August 2024
Revised: 17 September 2024
Accepted: 19 September 2024
Published: 25 September 2024

1. Introduction

Count data represents the number of times a particular event occurs in an interval of time, space, or other unit of measurement. This type of data is commonly found in various areas, such as medicine, economics, and engineering, to name a few. For example, Böhning et al. [1] analyzed count data from a dental epidemiological study under the situation of additional zeros. Salman et al. [2] analyzed bankruptcy count data from Swedish small manufacturing firms with the aim of investigating the business failure risk factors of small manufacturing firms. Calabria et al. [3] analyzed the reliability of repairable systems from in-service failure count data.

There are many real-world scenarios where the probability of success in binomial experiments cannot be considered constant. For example, the probability of consuming alcohol across the 7 days of a particular week varies from one individual to another (see Alanko and Lemmens [4]). Considering a beta distribution for the probability of success in a binomial distribution (which gives rise to the beta-binomial distribution) is not overly restrictive since the beta distribution is very flexible in terms of the shapes of its probability density function.

A random variable X follows the beta-binomial distribution, denoted $X \sim \mathrm{BB}(n, \alpha, \beta)$, if its probability mass function (p.m.f.) is given by

$$P(X = x) = \binom{n}{x} \frac{B(x + \alpha, n - x + \beta)}{B(\alpha, \beta)}, \; x = 0, 1, 2, \ldots, n, \; \alpha, \beta > 0, \tag{1}$$

where $B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\,du$, $a,b > 0$, is the beta function.

In Bayesian inference, the beta-binomial distribution is used to make predictions about the number of successes in future trials, taking into account the uncertainty in the estimate of the probability of success. In classical inference, the beta-binomial distribution can be used to model data with overdispersion in binomial experiments, i.e., when the observed variability is greater than that expected under a standard binomial distribution.

A review of the applicability and extensions of the beta-binomial distribution can be found in Wilcox [5]. The use of the beta-binomial distribution in the context of regression is discussed in Crowder [6]. Details on the estimation of the parameters of the beta-binomial distribution can be found in Tripathi et al. [7].

Regarding more recent applications of the beta-binomial distribution, several studies can be found in the literature. To name a few, Palm et al. [8] use the beta-binomial distribution in the formulation of the BBARMA (Beta-Binomial Autoregressive Moving Average) model, which can capture the temporal dynamics and autoregressive structure in count data. Chen et al. [9] use the beta-binomial distribution to propose a GARCH model that captures the variation in the number of new cases of cryptosporidiosis infection, obtaining a useful model for time series data that present bounded counts and high volatility. Jansen and Holling [10], under a Bayesian approach, use the beta-binomial distribution in the meta-analysis of rare events.

Although the beta-binomial distribution is applied in various real-world settings, its performance is not good when empirical distributions exhibit bimodality, i.e., when there are two modes or peaks in the empirical distributions. The presence of bimodality can be explained by the existence of two groups or subpopulations with unique characteristics or by the existence of latent variables that significantly influence the distribution of the population.

A very popular methodology in the literature to incorporate flexibility in terms of asymmetry and multimodality is related to the definition of weighted distributions proposed by Fisher [11] and Rao [12]. Suppose that X is a random variable with probability function $f(x)$. The weighted random variable X_w has PDF

$$f_{X_w}(x) = \frac{w(x)f(x)}{\mu_w}, \tag{2}$$

where $w(\cdot)$ is a nonnegative weight function and $\mu_w = \mathbb{E}[W(X)] < \infty$.

A particularly salient case of (2) is obtained when $w(x) = x$, which defines a length-biased distribution. These distributions arise naturally in applied fields, such as reliability and survival analysis, when individuals or mechanical units are sampled with unequal probability due to the experimental design or the existing unequal probability of detection.

On the other hand, it is possible to find in the literature weight functions that can lead to multimodality for the weighted distributions resulting from (2). For example, if $w(x) = 1 + \left(1 - \frac{\alpha(x-\mu)}{\sigma}\right)^2$,$\alpha \in \mathbb{R}$, and $f(x)$ is the pdf of the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, then (2) reduces to the family of bimodal distributions called the alpha-skew-normal distribution, see Elal-Olivero [13]. Based on the same weight function, Gómez-Déniz et al. [14] introduces a bimodal version of the Poisson distribution. Cortés et al. [15] propose a parametric weight function that involves a power function of exponent 4, which can lead to a probability function with up to three modes.

In this paper, we propose an extension of the beta-binomial distribution appropriate to fit overdispersed binomial data that may exhibit both unimodality and bimodality. The proposal arises from (2), using the weight function proposed by Elal-Olivero [13] under a beta-binomial baseline distribution. In this way, the new distribution is aimed at expanding the use of beta-binomial distributions to real-world scenarios where empirical distributions exhibit bimodality.

The remainder of the paper is organized as follows. In Section 2, we define the bimodal beta-binomial random variable and study some of its properties, such as the probability mass function, cumulative distribution function, and the raw moments. The latter are used to describe the behavior of the relative dispersion with respect to the mean and the skewness behavior of the distribution. In Section 3, parameter estimation for the new distribution using the maximum likelihood method is discussed. A simulation study is carried out to evaluate the behavior of the estimators. In Section 4, two application examples with real data are presented to illustrate the usefulness of the proposed distribution. Finally, concluding remarks are presented in Section 5.

2. Bimodal Beta-Binomial Distribution

In this section, we derive the new distribution and study some of its main properties.

2.1. Bimodal Beta-Binomial Random Variable

The following proposition presents the p.m.f. of the new distribution.

Proposition 1. *Let* $X \sim \mathrm{BB}(n, \alpha, \beta)$ *and* $w(\cdot)$ *be a parametric function given by*

$$w(x) = 1 + \left[1 - \frac{q(x-\mu)}{\sigma}\right]^2, \; x = 0, 1, \ldots, n,$$

where

$$\mu = \frac{n\alpha}{\alpha+\beta} \quad \text{and} \quad \sigma^2 = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)},$$

are the mean and variance of X*, respectively. Then, the p.m.f. of the weighted random variable* X_w *is*

$$\begin{aligned} f_{X_w}(x; \alpha, \beta, q) &= P(X_w = x) \\ &= \frac{1}{2+q^2}\left\{1 + \left[1 - \frac{q(x-\mu)}{\sigma}\right]^2\right\}\binom{n}{x}\frac{B(x+\alpha, n-x+\beta)}{B(\alpha,\beta)}, \; x = 0, 1, \ldots, n, \end{aligned} \tag{3}$$

such that $\alpha, \beta > 0$*,* $q \in \mathbb{R}$ *and* $B(\cdot, \cdot)$ *is the beta function.*

Proof. First, we observe that $f_{X_w}(x) > 0$ for all $x = 0, 1, 2, \ldots, n$ when $\alpha, \beta > 0$ and $q \in \mathbb{R}$. Second, it can be seen that

$$\begin{aligned} \sum_{x=0}^{n} f_{X_w}(x) &= \frac{1}{(2+q^2)}\sum_{x=0}^{n}\left(1 + \left[1 - \frac{q(x-\mu)}{\sigma}\right]^2\right) f_X(x) \\ &= \frac{1}{2+q^2}\left(2 - \frac{2q}{\sigma}\sum_{x=0}^{n}(x-\mu) f_X(x) + \frac{q^2}{\sigma^2}\sum_{x=0}^{n}(x-\mu)^2 f_X(x)\right) \\ &= 1. \end{aligned}$$

In consequence, it is concluded that (3) is a valid p.m.f. □

Definition 1. *Let* X_w *be a random variable with p.m.f. given in (3), then we say that* X_w *follows a bimodal beta-binomial distribution. We denote this as* $X_w \sim \mathrm{BBB}(n, \alpha, \beta, q)$*.*

The name given in Definition 1 to refer to the new distribution is based on the bimodal behavior that the p.m.f. can present. Figure 1 shows some plots of the p.m.f. of the bimodal beta-binomial distribution for different values of its parameters. In the figure, it can be seen that the BBB p.m.f can present a great variety of shapes depending on its parameters: monotonic shape, symmetric/asymmetric unimodal shape, bathtub shape, or asymmetric bimodal shape.

A function in the R programming language [16] for computing (3) is provided in Appendix A.

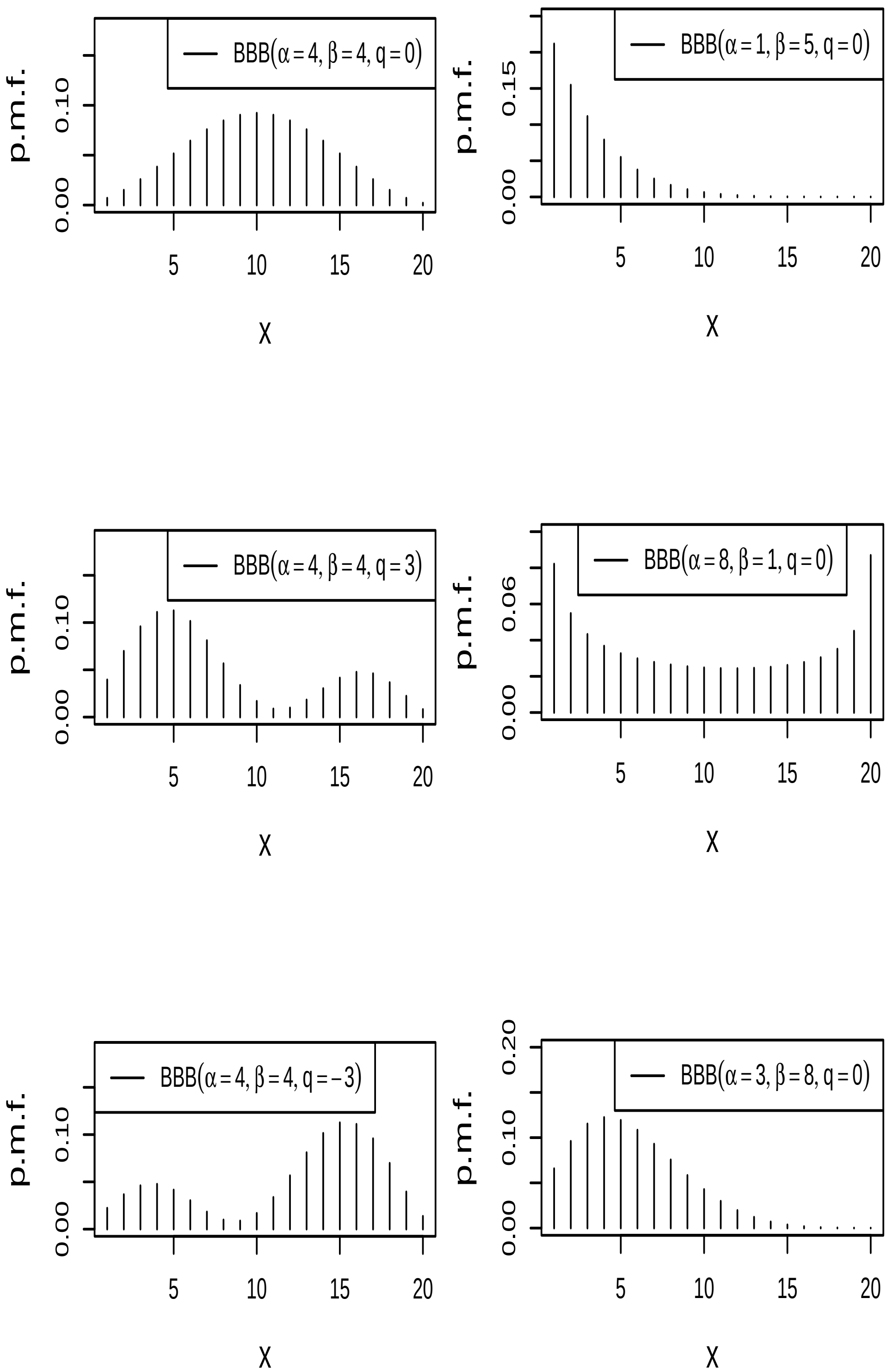

Figure 1. Plots of the p.m.f. of the bimodal beta-binomial distribution with $n = 20$ and different values of α, β and q.

2.2. Two Related Distributions

Corollary 1. *Let $X_w \sim \text{BBB}(n, \alpha, \beta, q)$. Then,*

1. $f_{X_w}(x; \alpha, \beta, q = 0) = \binom{n}{x} \dfrac{B(x + \alpha, n - x + \beta)}{B(\alpha, \beta)}$, $x = 0, 1, 2, \ldots, n$, $\alpha, \beta > 0$, *which is the p.m.f. of the beta-binomial distribution.*

2. *If $n = 1$, then $f_{X_w}(x; \alpha, \beta, q) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$, such that*

$$\theta = \begin{cases} \frac{\beta}{(2+q^2)(\alpha+\beta)} \left[1 + \left(1 + q\sqrt{\alpha/\beta}\right)^2\right], & \textit{if } x = 0, \\ \frac{\alpha}{(2+q^2)(\alpha+\beta)} \left[1 + \left(1 + q\sqrt{\beta/\alpha}\right)^2\right], & \textit{if } x = 1. \end{cases}$$

Corollary 1 is a direct consequence of (3) considering fixed values for q and n. Part 1 shows that the beta-binomial distribution is a special case of the bimodal beta-binomial distribution obtained when $q = 0$. The second part shows that the bimodal beta-binomial distribution reduces to the Bernoulli distribution with parameter θ, where θ is a function of the parameters α, β, and q.

2.3. Cumulative Distribution Function

The cumulative distribution function (c.d.f.) of $X_w \sim \mathrm{BBB}(n, \alpha, \beta, q)$ can be obtained straightforwardly from Proposition 1.

Corollary 2. *Let $X_w \sim \mathrm{BBB}(n, \alpha, \beta, q)$. Then, the cumulative distribution function (c.d.f.) of X_w is given by*

$$\begin{aligned} F_{X_w}(x) &= P(X_w \le x) \\ &= \begin{cases} 0, & \textit{if } \lfloor x \rfloor < 0, \\ \dfrac{1}{2+q^2} \displaystyle\sum_{t=0}^{\lfloor x \rfloor} \left\{1 + \left[1 - \dfrac{q(t-\mu)}{\sigma}\right]^2\right\} \binom{n}{t} \dfrac{B(t+\alpha, n-t+\beta)}{B(\alpha, \beta)}, & \textit{if } 0 \le \lfloor x \rfloor < n, \\ 1, & \textit{if } \lfloor x \rfloor \ge n, \end{cases} \end{aligned} \tag{4}$$

where $\lfloor x \rfloor = \max\{k \in \mathbb{Z} \,|\, k < x\}$, $x \in \mathbb{R}$.

Figure 2 shows some plots of the c.d.f. of the bimodal beta-binomial distribution for different values of α, β, and q. As expected, the figure shows that the frequencies are not decreasing as x increases. However, two sharp increases in frequency can be observed in two different intervals of x, which is explained by the bimodal behavior of the corresponding p.m.f.

A function in the R programming language for computing (4) is provided in Appendix A.

2.4. Moments

The following proposition derives the raw moments of the beta-binomial distribution. Essentially, these moments are expressed as a function of the raw moments of the beta-binomial distribution.

Proposition 2. *Let $X_w \sim \mathrm{BBB}(n, \alpha, \beta, q)$. Then, the rth raw moment of X_w is given by*

$$\mathbb{E}({X_w}^r) = a\mu_r - b\mu_{r+1} + c\mu_{r+2}, \quad r = 1, 2, \ldots, \tag{5}$$

where

$$a = \frac{1}{2+q^2}\left(2 + \frac{2q\mu}{\sigma} + \frac{q^2\mu^2}{\sigma^2}\right), \quad b = \frac{1}{2+q^2}\left(\frac{2q}{\sigma} + \frac{2q^2\mu}{\sigma^2}\right), \quad c = \frac{q^2}{(2+q^2)\sigma^2},$$

such that

$$\mu_j = \mathbb{E}\left(X^j\right) = \sum_{x=0}^{n} x^j \binom{n}{x} \frac{B(x+\alpha, n-x+\beta)}{B(\alpha, \beta)}, \quad j = 1, 2, \ldots$$

is the jth raw moment of the beta-binomial distribution.

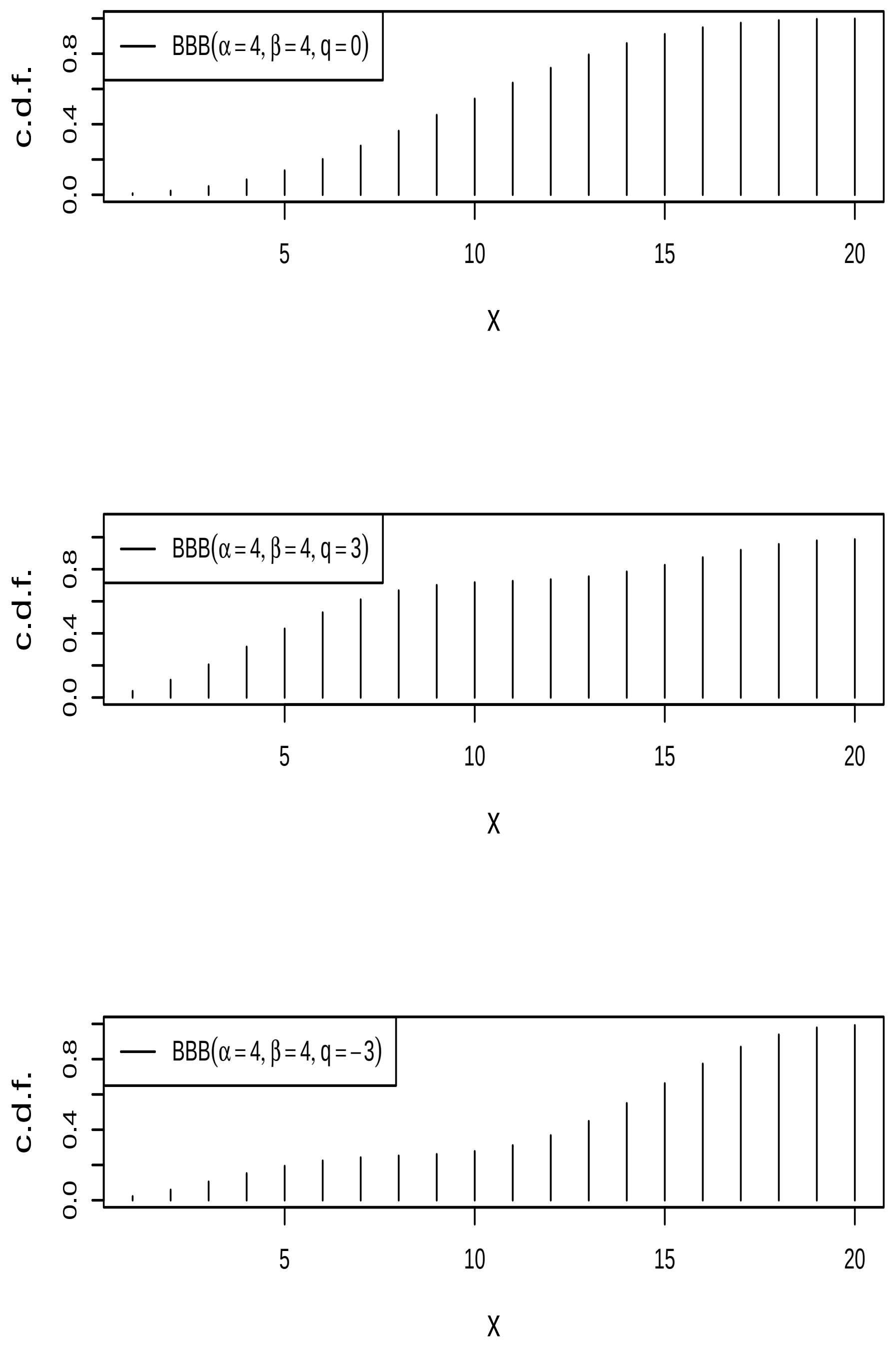

Figure 2. Plots of the c.d.f. of the bimodal beta-binomial distribution with $n = 20$, $\alpha = 4$, $\beta = 4$ and different values of q.

Proof. By definition of expectation, we have that

$$\begin{aligned} \mathbb{E}(X_w^r) &= \sum_{x=0}^{n} x^r f_{X_w}(x;\alpha,\beta,q) \\ &= \frac{1}{2+q^2}\sum_{x=0}^{n} x^r \left\{1+\left[1-\frac{q(x-\mu)}{\sigma}\right]^2\right\} f_X(x;\alpha,\beta), \end{aligned} \tag{6}$$

where $f_X(x;\alpha,\beta)$ is the p.m.f. of the beta-binomial distribution. Therefore, after some algebra, we see that

$$\begin{aligned}\mathbb{E}(X_w^r) &= \frac{1}{2+q^2}\left(2+\frac{2q\mu}{\sigma}+\frac{q^2\mu^2}{\sigma^2}\right)\sum_{x=0}^{n} x f_X(x;\alpha,\beta)-\frac{1}{2+q^2}\left(\frac{2q}{\sigma}+\frac{2q^2\mu}{\sigma^2}\right)\sum_{x=0}^{n} x^{r+1} f_X(x;\alpha,\beta)\\ &+\frac{q^2}{(2+q^2)\sigma^2}\sum_{x=0}^{n} x^{r+2} f_X(x;\alpha,\beta),\end{aligned}$$

and the result is obtained by recognizing the raw moments of the beta-binomial distribution in the above expression. □

Alternatively, in (6) we can write $\left[1-\frac{q(x-\mu)}{\sigma}\right]^2 = c^2\left(1-\frac{q}{\sigma c}x\right)^2$, where $c = 1+\frac{q\mu}{\sigma}$. Then, using the binomial theorem, we have

$$\left[1-\frac{q(x-\mu)}{\sigma}\right]^2 = c^2\sum_{k=0}^{2} v_k x^k, \quad \text{with} \quad v_k = \frac{q^k}{\sigma^k c^k}.$$

Thus, we can write (7) as

$$\begin{aligned}\mathbb{E}({X_w}^r) &= \frac{1}{2+q^2}\left[\sum_{x=0}^{n} x^r f_X(x;\alpha,\beta)+c^2\sum_{k=0}^{2} v_k\sum_{x=0} x^{r+k} f_X(x;\alpha,\beta)\right]\\ &= \frac{1}{2+q^2}\left(\mu_r+c^2\sum_{k=0}^{2} v_k\mu_{r+k}\right), \quad r=1,2,\dots,\end{aligned}$$

where μ_r is the rth raw moment of the beta-binomial distribution.

Corollary 3. *Let $X_w \sim \mathrm{BBB}(n,\alpha,\beta,q)$. Then, the coefficient of variation (c.v.(X_w)) and the Fisher's skewness coefficient ($\sqrt{\beta}$) of X_w are given by*

$$\begin{aligned}c.v.(X_w) &= \frac{\sqrt{\mu_2-b\mu_3+c\mu_4-(a\mu_1-b\mu_2+c\mu_3)^2}}{a\mu_1-b\mu_2+c\mu_4} \quad \textit{and}\\ \sqrt{\beta} &= \frac{a\mu_3-b\mu_4+c\mu_5-3(a\mu_1-b\mu_2+c\mu_3)(a\mu_2-b\mu_3+c\mu_4)+2(a\mu_1-b\mu_2+c\mu_3)^3}{\left[a\mu_2-b\mu_3+c\mu_4-(a\mu_1-b\mu_2+c\mu_3)^2\right]^{3/2}},\end{aligned}$$

where

$$\begin{aligned}\mu_1 &= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)n}{\Gamma(\alpha+\beta+1)\Gamma(\alpha)},\\ \mu_2 &= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)n(n\alpha+n+\beta)}{(\alpha+\beta+1)\Gamma(\alpha+\beta+1)\Gamma(\alpha)},\\ \mu_3 &= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)n\left(3\alpha\beta n+\beta^2+3n\beta+3n^2\alpha+n^2\alpha^2+2n^2-\alpha\beta\right)}{(2+\alpha^2+3\alpha+2\alpha\beta+\beta^2+3\beta)\Gamma(\alpha+\beta+1)\Gamma(\alpha)},\\ \mu_4 &= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)A\,n}{(6+\alpha^3+12\alpha\beta+6\beta^2+6\alpha^2+3\alpha\beta^2+3\beta\alpha^2+11\alpha+11\beta+\beta^3)\Gamma(\alpha+\beta+1)\Gamma(\alpha)},\\ \mu_5 &= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)B\,n}{(\alpha+\beta+3)(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta+4)\Gamma(\alpha+\beta+1)\Gamma(\alpha)}.\end{aligned}$$

such that a, b and c are as in Proposition 2 and

$$
\begin{aligned}
A \quad &= \quad -\alpha\beta + 6n^3 - 4\alpha\beta^2 + 7n\beta^2 + n^3\alpha^3 + 12n^2\beta + \beta^3 + 18\alpha n^2\beta + 7\beta^2 n\alpha - 4\beta n\alpha^2 - 5\beta n\alpha + \beta\alpha^2 \\
&\quad +6\beta n^2\alpha^2 - n\beta - \beta^2 + 6n^3\alpha^2 + 11\alpha n^3, \\
B \quad &= \quad 15\beta^3 n + 50n^2\beta^2 + 10n^4\alpha^3 + 60n^3\beta + 50n^4\alpha - 15n\beta^2 + 35n^4\alpha^2 + n^4\alpha^4 - 30\beta^2\alpha^2 n - 35\alpha n^2\beta \\
&\quad +25\beta^2 n^2\alpha^2 + 15\beta^3 n\alpha - 10\beta\alpha^3 n^2 + 5\beta\alpha^3 n + 10\beta n^3\alpha^3 - 45\beta^2 n\alpha - 11\alpha\beta^3 - 5\beta n\alpha + 5\beta\alpha^2 \\
&\quad -35\beta n^2\alpha^2 + 60n^3\alpha^2\beta - \alpha^3\beta - 10n^2\beta + 75n^2\alpha\beta^2 + 110n^3\alpha\beta + \beta^4 + 11\beta^2\alpha^2 + 24n^4 - 5\beta^3.
\end{aligned}
$$

Figure 3 shows some curves of the coefficient of variation and the coefficient of skewness of the bimodal beta-binomial distribution as a function of q under fixed values for α and β. In the figure, it can be seen that the bimodal beta-binomial distribution (depending on q) can present a greater or lesser relative dispersion (and a greater or lesser skewness level) than the beta-binomial distribution (special case $q = 0$).

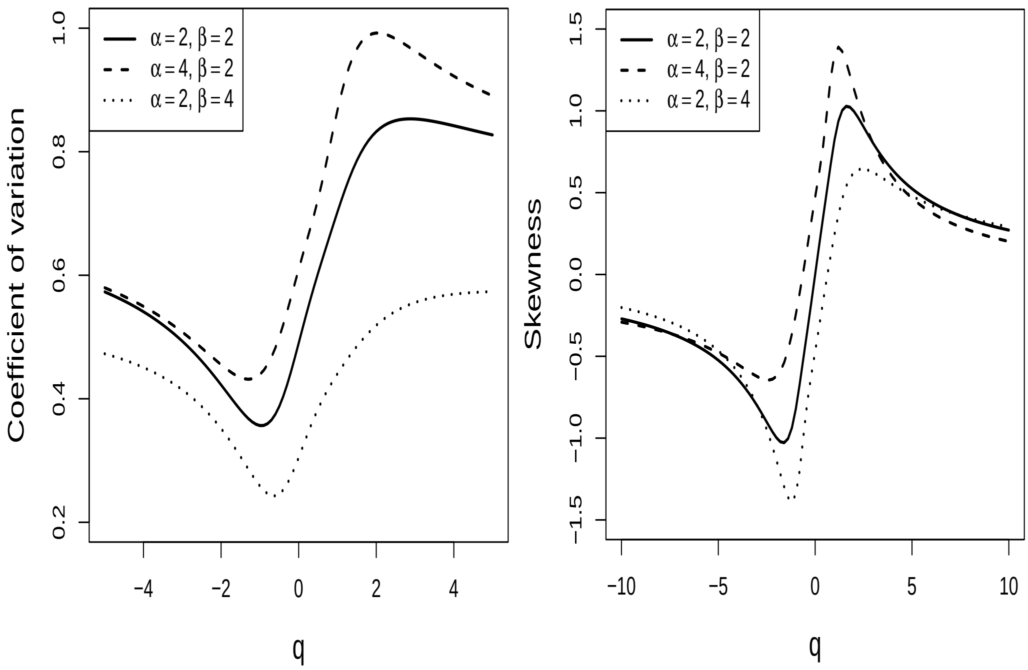

Figure 3. Plots of the coefficient of variation and the skewness coefficient (as a function of q) of the bimodal beta-binomial distribution with $n = 20$ and different values of α and β.

Functions in the R programming language for computing the rth moment (7) and for the coefficients of variation and the coefficient of skewness of Corollary 3 are provided in Appendix A.

3. Parameter Estimation

In this section, we discuss the maximum likelihood estimator and conduct a simulation study to evaluate the performance of the estimators.

3.1. Maximum Likelihood Estimation

Given a random sample $X_1, \ldots, X_m$ of the random variable $X_w \sim \mathrm{BBB}(n, \alpha, \beta, q)$, the log-likelihood function for $\theta = (\alpha, \beta, q)$ can be written as

$$
\begin{aligned}
\ell(\theta;x_i) &= \log\prod_{i=1}^{m} f_{X_w}(x_i;\alpha,\beta,q) \\
&= c+\sum_{i=1}^{m}\log\binom{n}{x_i}+\sum_{i=1}^{m}\log(x_{1i})+\sum_{i=1}^{m}\log\Gamma(x_i+\alpha)+\sum_{i=1}^{m}\log\Gamma(n-x_i+\beta),
\end{aligned} \tag{7}
$$

where $c=-m\log(2+q^2)+m\log\Gamma(\alpha+\beta)-m\log\Gamma(\alpha)-m\log\Gamma(\beta)-m\log\Gamma(\alpha+\beta+n)$, $x_{1i}=1+[1-q(x_i-\mu)/\sigma]^2$ and $\Gamma(a)=\int_0^\infty u^{a-1}e^{-u}\,du$, $a>0$, is the gamma function.

Then, the score functions are given by

$$
\begin{aligned}
\frac{\partial\ell(\theta;x_i)}{\partial\alpha} &= c_1-2q\Big(1+\frac{q\mu}{\sigma}\Big)\sum_{i=1}^{m}\frac{x_{2i}}{x_{1i}}+\frac{2q^2}{\sigma}\sum_{i=1}^{m}\frac{x_ix_{2i}}{x_{1i}}+\sum_{i=1}^{m}\Psi(x_i+\alpha), && (8)\\
\frac{\partial\ell(\theta;x_i)}{\partial\beta} &= c_2-2q\Big(1+\frac{q\mu}{\sigma}\Big)\sum_{i=1}^{m}\frac{x_{3i}}{x_{1i}}+\frac{2q^2}{\sigma}\sum_{i=1}^{m}\frac{x_ix_{3i}}{x_{1i}}+\sum_{i=1}^{m}\Psi(n-x_i+\beta), && (9)\\
\frac{\partial\ell(\theta;x_i)}{\partial q} &= c_3+\frac{2\mu}{\sigma^2}\Big(1-\frac{q}{\sigma}\Big)\sum_{i=1}^{m}\frac{1}{x_{1i}}-\frac{2}{\sigma}\Big(1-\frac{q\mu}{\sigma}+\frac{q}{\sigma}\Big)\sum_{i=1}^{m}\frac{x_i}{x_{1i}}+\frac{2q}{\sigma^2}\sum_{i=1}^{m}\frac{x_i^2}{x_{1i}}, && (10)
\end{aligned}
$$

where $c_1=-m\Psi(\alpha)+m\Psi(\alpha+\beta)-m\Psi(\alpha+\beta+n)$, $c_2=-m\Psi(\beta)+m\Psi(\alpha+\beta)-m\Psi(\alpha+\beta+n)$, $c_3=-2mq/(2+q^2)$, $\Psi(a)=\partial\log\Gamma(a)/\partial a$, with $a>0$, is the digamma function,

$$
\begin{aligned}
x_{2i} &= \frac{\partial}{\partial\alpha}\Big(\frac{x_i-\mu}{\sigma}\Big)=-\frac{k_{1i}}{2}\Big[n\alpha^3\beta(\alpha+\beta+1)(\alpha+\beta+n)^3\Big]^{-1/2} \quad\text{and}\\
x_{2i} &= \frac{\partial}{\partial\beta}\Big(\frac{x_i-\mu}{\sigma}\Big)=\frac{k_{2i}}{2}\Big[n\alpha^3\beta(\alpha+\beta+1)(\alpha+\beta+n)^3\Big]^{-1/2},
\end{aligned}
$$

such that $k_{1i}=-\alpha^3x_i+\alpha^3n-2x_i\alpha^2n-x_i\alpha^2\beta+2n^2\alpha^2+2n\alpha^2\beta+n^2\alpha\beta+n\alpha\beta^2+\alpha n^2+n\alpha\beta-\alpha x_in-x_i\alpha n\beta+x_i\alpha\beta^2+\alpha x_i\beta+x_in\beta+x_i\beta^2+x_i\beta^3+x_i\beta^2n$ and $k_{2i}=-\alpha^3x_i+\alpha^3n+n^2\alpha^2-x_i\alpha^2n-\alpha^2x_i+n\alpha^2+2n\alpha^2\beta-x_i\alpha^2\beta+n\alpha\beta^2-\alpha x_i\beta-\alpha x_in+2n\alpha\beta+\alpha n^2+x_i\alpha n\beta+x_i\alpha\beta^2+x_in\beta+2x_i\beta^2n+x_i\beta^3$.

Maximum likelihood (ML) estimator $\hat\theta=(\hat\alpha,\hat\beta,\hat q)$ of $\theta=(\alpha,\beta,q)$ can be obtained by setting (8)–(10) equal to zero and solving the resulting system of equations. However, due to the analytical complexity of these equations, estimates must be obtained using numerical methods.

The standard errors of the ML estimators can be obtained as the square roots of the elements of the diagonal of the matrix

$$
\mathrm{K}^{-1}(\hat\theta)=\left\{-\frac{\partial^2\ell(\theta;x_i)}{\partial\theta\partial\theta^{\mathrm{T}}}\bigg|_{\theta=\hat\theta}\right\}^{-1},
$$

where $\partial\ell(\theta;x_i)/\partial\theta\partial\theta^{\mathrm{T}}$ is the hessian matrix.

Alternatively, ML estimates can be obtained by solving the optimization problem $\max_\theta\ell(\theta;x_i)$, subject to $\alpha>0$, $\beta>0$ and $q\in\mathbb{R}$, where $\ell(\theta;x_i)$ is as in (7). For this, we recommend the use of the function `stat:optim()` of the R programming language, which also returns the numeric Hessian function. In particular, we consider the L-BFGS-B method [17], which allows the imposition of box constraints on the parameters. This means that it is possible to specify lower and upper bounds for each parameter, which is very valuable in optimization problems with high dimensions and specific constraints.

An R function for computing (7) is provided in Appendix A.

3.2. Simulation Study

In this section, we perform the simulation study by the acceptance–rejection sampling procedure; see Neumann [18]. We assume that X_w is a random variable that follows a bimodal beta-binomial distribution, i.e., $X_w \sim \text{BBB}(n, \alpha, \beta, q)$, and we generate 1000 random replications of samples of X_w with sample sizes $m = 50$, 100, 150, 200 and 300, respectively. We fix the maximum score number n equal to 30, and we define different scenarios setting different values for the parameters α, β, and q. For each scenario and sample size, Table 1 shows the behavior of the ML estimates, which are computed numerically using the optim function of the R programming language. Table 1 reports the mean and the standard deviation (sd) of the ML estimates in each scenario for each sample size. As expected, it can be observed that the average estimates move closer to the true values of the parameters as the sample size increases. Furthermore, it can be seen that the standard deviation decreases towards 0 as the sample size increases.

Table 1. Simulation of 1000 replications for the BBB distribution.

m	n	α	β	q	$\widehat{\alpha}$	$sd(\widehat{\alpha})$	$\widehat{\beta}$	$sd(\widehat{\beta})$	$\widehat{q}$	$sd(\widehat{q})$
50	30	0.5	1	2	0.517	0.117	1.043	0.292	2.184	0.681
100	30	0.5	1	2	0.506	0.081	1.012	0.199	2.072	0.404
150	30	0.5	1	2	0.505	0.066	1.004	0.160	2.062	0.323
200	30	0.5	1	2	0.502	0.057	0.999	0.138	2.034	0.273
300	30	0.5	1	2	0.501	0.046	0.999	0.112	2.020	0.219
50	30	0.5	2	1	0.533	0.146	2.322	0.879	1.032	0.322
100	30	0.5	2	1	0.512	0.100	2.132	0.594	1.010	0.220
150	30	0.5	2	1	0.507	0.081	2.065	0.473	1.005	0.176
200	30	0.5	2	1	0.504	0.070	2.048	0.413	1.005	0.150
300	30	0.5	2	1	0.502	0.057	2.037	0.338	1.003	0.121
50	30	3.0	5	2	3.167	0.717	5.340	1.388	2.155	0.620
100	30	3.0	5	2	3.068	0.484	5.141	0.931	2.075	0.394
150	30	3.0	5	2	3.046	0.389	5.084	0.743	2.062	0.315
200	30	3.0	5	2	3.023	0.334	5.045	0.638	2.039	0.268
300	30	3.0	5	2	3.015	0.271	5.030	0.517	2.026	0.215
50	30	5.0	3	2	5.347	1.261	3.317	1.049	2.197	0.723
100	30	5.0	3	2	5.112	0.829	3.103	0.680	2.085	0.442
150	30	5.0	3	2	5.072	0.668	3.057	0.547	2.066	0.349
200	30	5.0	3	2	5.045	0.573	3.038	0.469	2.048	0.296
300	30	5.0	3	2	5.039	0.466	3.034	0.381	2.036	0.237
50	30	1.0	2	3	1.015	0.163	2.029	0.323	3.217	0.871
100	30	1.0	2	3	1.008	0.145	2.013	0.286	3.159	0.741
150	30	1.0	2	3	1.005	0.118	2.002	0.231	3.110	0.575
200	30	1.0	2	3	1.001	0.102	2.001	0.200	3.064	0.481
300	30	1.0	2	3	1.001	0.083	2.000	0.163	3.032	0.382
50	30	1.0	3	2	1.039	0.231	3.127	0.724	2.135	0.574
100	30	1.0	3	2	1.012	0.159	3.040	0.492	2.053	0.368
150	30	1.0	3	2	1.008	0.129	3.011	0.395	2.042	0.295
200	30	1.0	3	2	1.003	0.111	3.004	0.341	2.022	0.251
300	30	1.0	3	2	1.002	0.090	3.003	0.278	2.016	0.203
50	30	2.0	3	1	2.233	0.631	3.647	1.344	1.060	0.463
100	30	2.0	3	1	2.099	0.421	3.284	0.846	1.018	0.288
150	30	2.0	3	1	2.060	0.342	3.177	0.681	1.007	0.229
200	30	2.0	3	1	2.039	0.291	3.122	0.577	1.005	0.192
300	30	2.0	3	1	2.032	0.239	3.093	0.474	1.007	0.153

sd corresponds to the standard deviation for the ML estimates.

4. Applications

In this section, two applications are presented to illustrate the utility of the bimodal beta-binomial (BBB) distribution in modeling count data. In each application, the beta-binomial (BB) and McDonald generalized beta-binomial (McGBB) [19] distributions are incorporated into the analysis. The p.m.fs of the McGBB distribution is given by

$$P(X=x)=\binom{n}{x}\frac{1}{B(\alpha,\beta)}\sum_{j=0}^{n-x}(-1)^j\binom{n-x}{j}B\left(\frac{x}{q}+\alpha+\frac{j}{q},\beta\right),\ x=0,1,\ldots,n,\ \alpha,\beta,q>0,$$

where $B(\cdot,\cdot)$ is the beta function.

Like the BB distribution, the McGBB distribution performs well in modeling overdispersed binomial data. However, due to a larger parameter dimension, the McGBB distribution may outperform the BB distribution in modeling overdispersed binomial data. Furthermore, it is important to note that the McGBB distribution can model bimodality when the empirical frequency distribution exhibits a bathtub shape, thus making the BBB distribution a natural alternative to the McGBB distribution.

We assessed the quality of the fits using the chi-square goodness-of-fit test and evaluated the comparative performance using the Akaike information criterion (AIC) [20] and the Bayesian information criterion (BIC) [21]. R codes used in this section are provided in Appendix A.

Furthermore, we use the excess mass test proposed in Ameijeiras-Alonso et al. [22] to show the bimodality of the data considered in the first application and the unimodality of the data considered in the second application. For this, we used the modetest function [23] in the R programming language.

4.1. Alcohol Consumption Data

The first data set consists of observations on the number of days on the $n=7$ days of two reference weeks (week 1 and week 2), in which 399 individuals consume alcohol (See Table 2). Although there may be an attempt to use the binomial distribution to fit these data, it must be taken into account that the probability of consuming alcohol on a randomly chosen day in a week is variable from one individual to another. Based on the latter, Alanko and Lemmens [4] use the beta-binomial distribution to fit these data. On the other hand, Manoj et al. [19] illustrate that these data present an overdispersion with respect to the binomial distribution and that the McGBB distribution performs better than the BB and Kumaraswamy binomial [24] distributions in fitting these data.

For these data, we test hypothesis H_0: the data have exactly two modes versus the alternative hypothesis H_1: the data have more than two modes. For the data from week 1, we obtain an observed statistic equal to 0.021 with a p-value equal to 0.644. For the data from week 2, we obtain an observed statistic value equal to 0.023, with a p-value equal to 0.406. Therefore, with a significance level equal to 0.05, H_0 is not rejected in both weeks; that is, the frequency distributions of the data corresponding to weeks 1 and 2 exhibit bimodal behavior.

Other previous studies with these data can be found in Rodríguez-Avi et al. [25].

Table 2 reports the results obtained when fitting the alcohol consumption data using the BB, McGBB, and BBB distributions. The table shows that the BBB distribution presents the highest p-values in the chi-square goodness-of-fit test and the lowest AIC and BIC values, suggesting that the BBB distribution should be selected among the fitted distributions for modeling the alcohol consumption data.

Figure 4 shows the frequency distribution of the alcohol consumption data (Weeks 1 and 2) and the fitted BB, McGBB, and BBB distributions. In the figure, it can be seen that the frequency distributions of the number of drinking days present two frequency peaks and that the mass values of the BBB distribution are the closest to the empirical frequency values.

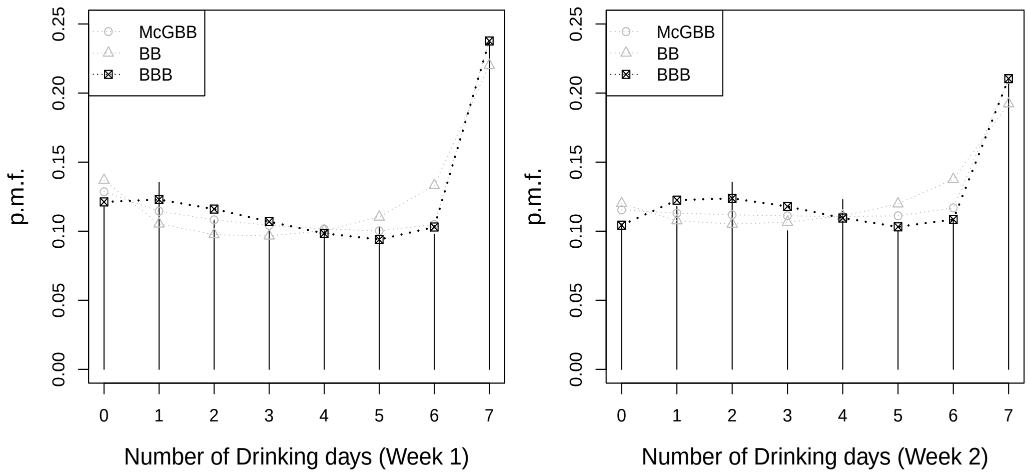

Figure 4. Frequency distributions of the number of drinking days in weeks 1 and 2 and p.m.f.s for the BB, McGBB, and BBB distributions provided with the ML estimates reported in Table 2.

Table 2. ML estimates, maximum log-likelihood value, AIC and BIC values, and observed statistic (χ^2), degrees of freedom (DF), and p-value obtained in the chi-square goodness-of-fit tests for the BB, McGBB, and BBB distributions fitted to the alcohol consumption data.

Number of Drinking Days	Observed Frequency (Week 1)	Expected BB Frequency (Week 1)	Expected McGBB Frequency (Week 1)	Expected BBB Frequency (Week 1)	Observed Frequency (Week 2)	Expected BB Frequency (Week 2)	Expected McGBB Frequency (Week 2)	Expected BBB Frequency (Week 2)
0	47	54.60	51.29	48.39	42	47.90	45.92	41.62
1	54	42.00	45.67	41.93	47	42.90	45.13	48.86
2	43	38.90	43.17	46.28	54	41.90	44.75	49.37
3	40	38.50	41.61	42.65	40	42.50	44.50	47.03
4	40	40.10	40.52	39.26	49	44.30	44.35	43.70
5	41	44.00	40.01	37.46	40	47.80	44.51	41.14
6	39	53.10	41.83	41.11	43	54.90	46.57	43.28
7	95	87.80	94.90	94.85	84	76.70	83.26	83.96
Total	399	399	399	399	399	399	399	399
χ^2		9.600	2.162	0.366		9.700	4.004	1.397
DF		5	4	4		5	4	4
p-value		0.086	0.706	0.833		0.082	0.406	0.986
ML estimates	$\hat{\alpha}$	0.722	0.028	1.147		0.858	0.027	1.354
	$\hat{\beta}$	0.581	0.155	0.325		0.701	0.215	0.392
	$\hat{q}$	-	32.345	0.701		-	36.075	0.732
Log-Likelihood		−813.457	−809.627	−809.329		−821.392	−818.402	−817.650
AIC		1630.9	1625.3	1624.7		1646.8	1642.8	1641.3
BIC		1638.9	1637.2	1636.6		1654.8	1654.8	1653.3

4.2. Candidate Assessment Data

In this section, we consider a dataset on candidate performance on an exam consisting of 9 questions. Each question is scored out of a total of 20 points, and when assessing a candidate's final score, special attention is paid to the total number of questions on which he or she has an "alpha" ("alpha"—scoring at least 15 points on the question). The number of alphas is a rough indication of the quality of the candidate's exam performance. Therefore, the distribution of alphas across candidates is of interest. A total of 209 candidates attempted questions from this 9-question section, and 326 alphas were awarded in total. Thus, we consider $n = 9$ (number of trials/questions), where the dichotomous variable on each trial is whether or not the candidate scored an alpha.

For these data, we test the hypothesis H_0: the data has exactly one mode versus the alternative hypothesis H_1: the data have more than one mode, obtaining an observed statistic equal to 0.031, with a p-value equal to 0.720. Consequently, with a significance level equal to 0.05, H_0 is not rejected; that is, the frequency distribution of the data exhibits unimodal behavior.

A previous study with these data can be found in Paul [26].

Table 3 reports the results obtained by fitting the number of alphas using the BB, McGBB, and BBB distributions. In the table, it can be seen that the BBB distribution presents the highest p-values in the chi-square goodness-of-fit test and the lowest AIC and BIC values, suggesting that the BBB distribution should be selected among the fitted distributions for modeling the number of alphas.

Table 3. ML estimates, maximum log-likelihood value, AIC and BIC values, and observed statistic (χ^2), degrees of freedom (DF), and p-value obtained in the chi-square goodness-of-fit tests for the BB, McGBB, and BBB distributions fitted to the number of alphas.

Number of Alphas	**Observed Frequency**	**Expected BB Frequency**	**Expected McGBB Frequency**	**Expected BBB Frequency**
0	63	63.67	63.49	63.41
1	67	49.23	50.37	66.31
2	34	35.84	35.89	38.93
3	18	24.81	24.35	17.05
4	11	16.24	15.77	8.29
5	8	9.89	9.64	6.18
6	4	5.47	5.43	4.80
7	3	2.63	2.70	2.79
8	1	1.00	1.09	1.05
9	0	0.23	0.28	0.20
Total	209	209	209	209
χ^2		11.121	9.669	2.457
DF		7	6	6
p-value		0.133	0.139	0.873
ML estimates	$\hat{\alpha}$	1.057	16.543	2.465
	$\hat{\beta}$	4.300	3.702	5.921
	$\hat{q}$	-	0.101	0.942
Log-Likelihood		−354.025	−353.414	−351.319
AIC		712.051	712.828	708.639
BIC		718.735	722.855	718.665

Figure 5 shows the frequency distribution of the number of alphas and the fitted BB, McGBB, and BBB distributions. In the figure, it can be seen that the frequency distribution of the number of alphas presents a single frequency peak and that the mass values of the BBB distribution are the closest to the empirical frequency values.

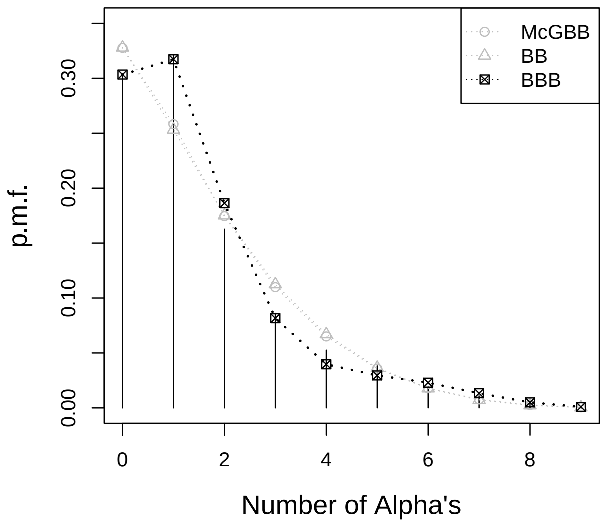

Figure 5. Frequency distributions of the number of alphas and p.m.f.s for the BB, McGBB, and BBB distributions provided with the ML estimates reported in Table 3.

5. Final Comments

The beta-binomial (BB) and McDonald's generalized beta-binomial (McGBB) distributions are discrete probability distributions used for modeling overdispersed binomial data. The McGBB distribution, presenting a larger parameter dimension than the BB distribution (three parameters), is capable of modeling even bimodality in cases where empirical frequency distributions present a bathtub shape. In this article, we propose the bimodal beta-binomial (BBB) distribution as an alternative for modeling overdispersed binomial data, both unimodal and bimodal. The new distribution arises from a weighted version of the BB distribution, where the weight function has the quadratic form proposed by Elal-Olivero [13]. Consequently, the BBB distribution is capable of presenting a flexible probability mass function in terms of shapes: monotonic, unimodal, and even bimodal. The bimodal shape of the BBB distribution is not limited to the bathtub shape (like the McGBB distribution), but the bimodality can be accompanied by various levels of skewness.

We derive the main structural functions of the BBB distribution, such as the p.m.f., the c.d.f., and the raw rth moment. We use the rth moment to describe the behavior of the coefficient of variation and the coefficient of skewness. We observe that the BBB distribution may exhibit a larger relative dispersion and a larger skewness than the BB distribution. We discuss parameter estimation via the maximum likelihood (ML) method. The estimators are not closed-form, so numerical methods are required to obtain the estimates. We develop a simulation study to evaluate the behavior of the ML estimators, in which we observe that the ML method provides acceptable estimates. Finally, we illustrate the utility of the BBB distribution by fitting real data sets. The illustrations show that the BBB distribution can outperform the BB and McGBB distributions in modeling count data that exhibit both unimodality and bimodality.

Author Contributions: Conceptualization, J.R., J.N.-Z. and D.-J.L.; Methodology, J.N.-Z., D.-J.L., J.A. and Y.A.I.; Software, J.A. and Y.A.I.; Validation, J.R., J.N.-Z., D.-J.L., J.A. and Y.A.I.; Formal analysis, J.R., J.N.-Z., D.-J.L., J.A. and Y.A.I.; Investigation, J.R., J.N.-Z., D.-J.L. and J.A.; Writing—review & editing, Y.A.I.; Supervision, J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

Appendix A. R Functions

The R codes used in this article are available at https://github.com/YuriIriarte/BBBdistribution (accessed on 18 September 2024) stored in the following files:

1. BBB-distribution;
   Includes R codes for computing the structural functions of the BBB distribution.
   p.m.f. → `dBBB()`; c.d.f. → `pBBB()`; *r*th moment → `momBBB()`
   coefficient of variation → `coef.var()`; skewness coefficient → `skewnessBBB()`.
2. BB-McGBB-distribution;
   Includes R codes for computing the p.m.f.s of the BB and McGBB distributions.
   BB p.m.f. → `dBB()`; McGBB p.m.f. → `dMcGBB()`.
3. Log-likelihood;
   Includes R codes for computing the log-likelihood functions to obtain the ML estimators using the `optim()` function.
   BB log-likelihood → `loglikBB()`
   McGBB log-likelihood → `loglikMcGBB()`
   BBB log-likelihood → `loglikBBB()`.
4. Application-alcohol-week1 (week2);
   Includes the data and results obtained in the analysis in Section 4.1.
5. Application-candidate;
   Includes the data and results obtained in the analysis in Section 4.2.

References

1. Böhning, D.; Dietz, E.; Schlattmann, P.; Mendonca, L.; Kirchner, U. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *J. R. Stat. Soc. Ser. A Stat. Soc.* **1999**, *162*, 195–209. [CrossRef]
2. Salman, A.K.; Fuchs, M.; Zampatti, D. Assessing risk factors of business failure in manufactoring sector: A count data approach from Sweden. *Int. J. Econ. Commer. Manag.* **2015**, *3*, 42–62.
3. Calabria, R.; Guida, M.; Pulcini, G. Reliability analysis of repairable systems from in–service failure count data. *Appl. Stoch. Model. Data Anal.* **1994**, *10*, 141–151. [CrossRef]
4. Alanko, T.; Lemmens, P.H. Response effects in consumption surveys: An application of the beta-binomial model to self-reported drinking frequencies. *J. Off. Stat.* **1996**, *12*, 253.
5. Wilcox, R.R. A review of the beta-binomial model and its extensions. *J. Educ. Stat.* **1981**, *6*, 3–32. [CrossRef]
6. Crowder, M.J. Beta-binomial ANOVA for proportions. *Appl. Stat.* **1978**, *27*, 34–37. [CrossRef]
7. Tripathi, R.C.; Gupta, R.C.; Gurland, J. Estimation of parameters in the beta binomial model. *Ann. Inst. Stat. Math.* **1994**, *46*, 317–331. [CrossRef]
8. Palm, B.G.; Bayer, F.M.; Cintra, R.J. Signal detection and inference based on the beta binomial autoregressive moving average model. *Digit. Signal Process.* **2021**, *109*, 102911. [CrossRef]
9. Chen, H.; Li, Q.; Zhu, F. A covariate-driven beta-binomial integer-valued GARCH model for bounded counts with an application. *Metrika* **2023**, *86*, 805–826. [CrossRef]
10. Jansen, K.; Holling, H. Rare events meta-analysis using the Bayesian beta-binomial model. *Res. Synth. Methods* **2023**, *14*, 853–873. [CrossRef]
11. Fisher, R.A. The effect of methods of ascertainment upon the estimation of frequencies. *Ann. Eugen.* **1934**, *6*, 13–25. [CrossRef]
12. Rao, C.R. On discrete distributions arising out of methods of ascertainment. *Sankhyā Indian J. Stat. Ser. A* **1965**, *27*, 311–324.
13. Elal-Olivero, D. Alpha-skew-normal distribution. *Proyecciones* **2010**, *29*, 224–240. [CrossRef]
14. Gómez-Déniz, E.; Pérez-Rodríguez, J.V.; Reyes, J.; Gómez, H.W. A bimodal discrete shifted Poisson distribution. a case study of tourists' length of stay. *Symmetry* **2020**, *12*, 442. [CrossRef]
15. Cortés, I.; Reyes, J.; Iriarte, Y.A. A Weighted Skew-Logistic Distribution with Applications to Environmental Data. *Mathematics* **2024**, *12*, 1287. [CrossRef]
16. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
17. Byrd, R.H.; Lu, P.; Nocedal, J.; Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208. [CrossRef]
18. Neumann, V. Various techniques used in connection with random digits. *Natl. Bur. Stand.* **1951**, *12*, 36–38.

19. Manoj, C.; Wijekoon, P.; Yapa, R.D. The McDonald generalized beta-binomial distribution: A new binomial mixture distribution and simulation based comparison with its nested distributions in handling overdispersion. *Int. J. Stat. Probab.* **2013**, *2*, 24. [CrossRef]
20. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
21. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
22. Ameijeiras-Alonso, J.; Crujeiras, R.M.; Rodríguez-Casal, A. Mode testing, critical bandwidth and excess mass. *Test* **2019**, *28*, 900–919. [CrossRef]
23. Ameijeiras-Alonso, J.; Crujeiras, R.M.; Rodriguez-Casal, A. Multimode: An R package for mode assessment. *arXiv* **2018**, arXiv:1803.00472. [CrossRef]
24. Xiaohu, L.; Yanyan, H.; Xueyan, Z. The Kumaraswamy binomial distribution. *Chin. J. Appl. Probab. Stat.* **2011**, *27*, 511–521.
25. Rodríguez-Avi, J.; Conde-Sánchez, A.; Sáez-Castillo, A.; Olmo-Jiménez, M. A generalization of the beta–binomial distribution. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2007**, *56*, 51–61. [CrossRef]
26. Paul, S. A three-parameter generalization of the binomial distribution. *Hist. Philos. Log.* **1985**, *14*, 1497–1506. [CrossRef]