Journal of Data Science and Modeling, Vol. 3, No. 1, 123-142, December 2024

Research Manuscript

Zero-Inflated Two-Parameter Distribution for Modeling Overdispersed Count data

Zahra Karimiezmareh*[1], Behdad Mostafaiy[2]

[1]Department of Statistics, Allameh Tabataba'i University, Tehran, Iran

[2]Department of Statistics, University of Mohaghegh Ardabili, Ardabil, Iran.

Recieved: 25/05/2025 Accepted: 20/02/2026

Abstract:

In this paper, we propose a new two-parameter discrete distribution based on the central Bell expansion, which is zero-inflated and designed to effectively model overdispersed count data. We study several structural properties of the proposed distribution and demonstrate that it is infinitely divisible, which adds theoretical strength and potential for wider applicability. The paper also discusses parameter estimation techniques for the distribution, focusing on two common approaches: the method of moments and the maximum likelihood estimation method. Both methods are developed and explained in detail. To evaluate the accuracy and reliability of these estimators, a simulation study is conducted across different sample sizes, allowing us to assess their performance under various conditions. To illustrate the practical importance and usefulness of the new distribution, we apply it to two real data sets and show how well it fits the observed data, reinforcing its value as a flexible tool for analyzing count data.

Keywords: Central Bell distribution; Count data; Overdispersion; Zero-inflated

Mathematics Subject Classification (2010): 62Exx, 62P99.

*Corresponding Author: zahra.karimiez@atu.ac.ir

1. Introduction

Discrete distributions have been extensively studied recently. Most of them are based on discretizing a continuous distribution. This approach allows researchers to leverage the well-established properties and flexibility of continuous models and adapt them for discrete outcomes commonly encountered in fields such as medicine, insurance, and quality control. By creating these discrete analogues, sophisticated models can be developed that effectively handle complex data characteristics like overdispersion, zero-inflation, and heavy tails, which are often poorly served by traditional count models such as the Poisson. The following recent studies exemplify this innovative trend.

Ascari and *et al.* (2024) proposed the Flexible Beta-Negative Binomial distribution, a novel model designed to capture extreme overdispersion and a high frequency of zeros more effectively than its predecessors. El-Alosey and *et al.* (2025) developed a zero-inflated regression model using a Poisson-modification of the Quasi Lindley distribution, further enhanced with a ridge estimator to handle the common issue of multicollinearity among predictors. Chesneau and *et al.* (2024) contributed to the field by creating a novel family of discrete trigonometric distributions, such as the discrete sin-Weibull, offering unique shapes for capturing patterns in diverse count datasets. Barbiero and *et al.* (2024) explored discrete analogues of the half-logistic distribution, resulting in simple yet effective models for overdispersed counts with a mode at zero, which are useful in fields like ecology and insurance. Sultan and Para (2025) presented the Poisson EGamma model, a versatile distribution that integrates the Poisson with a two-parameter EGamma distribution to adeptly model overdispersed healthcare data, such as infected cell counts. Maya and *et al.* (2024) defined a discrete analogue of the continuous new XLindley distribution, a flexible one-parameter model capable of handling both overdispersed and underdispersed data, and extended it for use in time series analysis and statistical quality control.

Emerging from the elegant realm of combinatorial mathematics, the Bell distribution captures the intricate patterns of count data through a single, powerful parameter. Unlike the Poisson distribution, which often struggles with real-world variability, the Bell distribution naturally accommodates overdispersion, making it an adept model for datasets where the variance exceeds the mean. Its mathematical foundation, linked to the Bell numbers that count the ways a set can be partitioned, grants it unique flexibility. This connection to fundamental combinatorics allows it to model complex, clustered phenomena in fields such as biology and sociology with remarkable parsimony, offering a sophisticated yet streamlined tool for modern statistical analysis.

The one-parameter Bell distribution is derived from the following expansion, originally introduced in the work of Bell (1934):

$$\exp\left(e^{x}-1\right)=\sum_{n=0}^{\infty}\frac{B_n}{n!}x^n, \quad x\in\mathbb{R}, \tag{1.1}$$

where the sequence B_n, known as the Bell numbers, is defined by the infinite sum

$$B_n=\frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{k!}. \tag{1.2}$$

Beginning with $B_0 = B_1 = 1$, the initial terms of this sequence are $B_2 = 2, B_3 = 5, B_4 = 15, B_5 = 52, B_6 = 203, B_7 = 877, B_8 = 4140, B_9 = 21147, B_{10} = 115975, B_{11} = 678570, B_{12} = 4213597, B_{13} = 27644437$.

Corollary 1.1. *It is worth noting that the Bell number B_n corresponds to the n^{th} moment of a Poisson distribution with a mean parameter equal to 1.*

Definition 1.2. *A discrete random variable Y is said to follow a Bell distribution with parameter $\theta > 0$, denoted by $Y \sim Bell(\theta)$, if its probability mass function is expressed as*

$$\Pr(Y=y)=\frac{\theta^{y}e^{-e^{\theta}+1}B_y}{y!}, \quad y=0,1,2,\dots, \tag{1.3}$$

where B_y represents the y^{th} Bell number as defined in Equation (1.2).

For more details on the Bell distribution, see Castellares and *et al.* (2018).

Recent literature has witnessed significant advancements in the generalization and application of the Bell distribution. Kim and Kim (2025) introduced probabilistic bivariate and r-Bell polynomials, deriving recurrence relations that extend classical results. Simultaneously, Xue and *et al.* (2025) developed probabilistic degenerate poly-Bell polynomials from degenerate polyexponential functions, obtaining explicit expressions and identities for special cases involving Bernoulli and gamma random variables. In a pivotal contribution, Soni and *et al.* (2024) established a comprehensive probabilistic framework for Bell polynomials connected to various random variables, deriving generating functions, recurrence relations, and demonstrating applications in stochastic modeling. Most recently, Santos and *et al.* (2025) transcended theoretical developments by proposing a practical Bell mixed-effects regression model, demonstrating through simulations and real-data applications its superiority over traditional Poisson-based models for handling overdispersed count data in fields including health sciences. These works collectively highlight the expanding utility and theoretical richness of Bell-type distributions in statistical modeling.

In this paper, we propose a two-parameter discrete distribution based on the central Bell (CB) expansion, which is useful for modeling count data with overdispersion. The CB distribution offers significant advantages over the standard Bell distribution, particularly through its enhanced flexibility in modeling varying degrees of zero-inflation and improved control over dispersion patterns. Unlike the standard Bell distribution, where the mean and variance are intrinsically linked, the two-parameter structure of the CB distribution allows for more precise modeling of real-world data where dispersion characteristics may vary independently. Additionally, the property of infinite divisibility makes the CB distribution suitable for more complex statistical applications, including compound Poisson processes, thereby extending its utility beyond the capabilities of the standard Bell distribution.

2. The CB Expansion

For $n \geq 0$, the falling factorial sequence is defined by

$$(\theta)_0 = 1, \qquad (\theta)_n = (\theta)(\theta-1)\ldots(\theta-n+1), \quad n \geq 1. \tag{2.4}$$

The central factorial $\theta^{[n]}$ is

$$\theta^{[0]} = 1, \qquad \theta^{[n]} = \theta\left(\theta + \frac{n}{2} - 1\right)_{n-1}, \quad n \geq 1. \tag{2.5}$$

The central factorial numbers of the second kind, $T(n,k)$, define the connection between the coefficients of sequences θ^n and $\theta^{[k]}$ via the relation

$$\theta^n = \sum_{k=0}^{n} T(n,k)\theta^{[k]}, \qquad n \geq 0. \tag{2.6}$$

From (2.6), we have

$$T(n,k) = \frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}(-1)^i\left(\frac{k}{2} - i\right)^n, \qquad n,k \geq 0. \tag{2.7}$$

The CB polynomials $B_n^{(c)}(\theta)$ are defined by

$$B_n^{(c)}(\theta) = \sum_{k=0}^{n} T(n,k)\theta^k, \qquad n \geq 0. \tag{2.8}$$

See Kim and Kim (2020). For example, $B_0^{(c)}(\theta) = 1$, $B_1^{(c)}(\theta) = \theta$, $B_2^{(c)}(\theta) = \theta^2$, $B_3^{(c)}(\theta) = \theta^3 + \theta/4$, $B_4^{(c)}(\theta) = \theta^4 + \theta^2$, and so on. For the CB polynomials, we have the expansion

$$e^{2\theta \sinh(t/2)} = \sum_{n=0}^{\infty} B_n^{(c)}(\theta)\frac{t^n}{n!}. \tag{2.9}$$

The rest of the paper is organized as follows. In Section 3, we define the CB distribution and study its structural properties. The parameters of the CB distribution are estimated using the maximum likelihood and the method of moments in

Section 4. A simulation study is conducted in Section 5. In Section 6, we illustrate the usefulness of the CB distribution using two real datasets.

3. The CB Distribution

In this section, the CB distribution is introduced and its properties are examined.

Definition 3.1. *A random variable X has a CB distribution with parameters $\alpha > 0$ and $\theta > 0$, denoted by $X \sim CB(\alpha, \theta)$, if its probability mass function is*

$$\Pr(X = x) = \frac{e^{-2\theta \sinh(\alpha/2)} \alpha^x B_x^{(c)}(\theta)}{x!}, \qquad x = 0, 1, 2, \dots .$$

Some values of the $CB(\alpha, \theta)$ probability mass function are

$$\begin{aligned} \Pr(X = 0) &= e^{-2\theta \sinh(\alpha/2)}, \\ \Pr(X = 1) &= \alpha\theta e^{-2\theta \sinh(\alpha/2)}, \\ \Pr(X = 2) &= \alpha^2\theta^2 e^{-2\theta \sinh(\alpha/2)}, \\ \Pr(X = 3) &= \alpha^3(\theta/4 + \theta^3) e^{-2\theta \sinh(\alpha/2)}. \end{aligned}$$

Theorem 3.2. *Let $X \sim CB(\alpha, \theta)$. Then the moment generating function of X is*

$$M_X(s) = e^{2\theta(\sinh(\alpha e^s/2) - \sinh(\alpha/2))}.$$

Proof. By definition, we have

$$M_X(s) = E(e^{sX}) = \sum_{x=0}^{\infty} \frac{e^{-2\theta \sinh(\alpha/2)} (\alpha e^s)^x B_x^{(c)}(\theta)}{x!} = e^{2\theta(\sinh(\alpha e^s/2) - \sinh(\alpha/2))}.$$

□

Corollary 3.3. *Let $Y \sim CB(\alpha, \theta)$. Then, using the moment generating function derived in Theorem 3.2, the expectation and variance of Y are given by:*

$$\begin{aligned} E(X) &= \alpha\theta \cosh(\alpha/2), \\ Var(X) &= \alpha\theta \cosh(\alpha/2) + \frac{\alpha^2\theta}{2} \sinh(\alpha/2). \end{aligned}$$

We observe that $Var(X) > E(X)$. The index of dispersion is $ID = \frac{Var(X)}{E(X)} = 1 + \frac{\alpha}{2}\tanh(\alpha/2)$. It follows that $ID > 1$ for every $\alpha > 0$. Thus, count data with overdispersion may be modeled by the CB distribution. In addition, $ID \longrightarrow 1$ as $\alpha \longrightarrow 0$.

Theorem 3.4. *The CB distribution with parameters $\alpha > 0$ and $\theta > 0$ is identifiable.*

Proof. Suppose that for all $x = 0, 1, 2, \dots,$

$$\frac{e^{-2\theta_1 \sinh(\alpha_1/2)} \alpha_1^x B_x^{(c)}(\theta_1)}{x!} = \frac{e^{-2\theta_2 \sinh(\alpha_2/2)} \alpha_2^x B_x^{(c)}(\theta_2)}{x!}.$$

Then

$$e^{-2\theta_1 \sinh(\alpha_1/2)} \alpha_1^x B_x^{(c)}(\theta_1) = e^{-2\theta_2 \sinh(\alpha_2/2)} \alpha_2^x B_x^{(c)}(\theta_2). \tag{3.10}$$

For $x = 0$, since $B_0^{(c)}(\theta) = 1$, we have

$$e^{-2\theta_1 \sinh(\alpha_1/2)} = e^{-2\theta_2 \sinh(\alpha_2/2)}. \tag{3.11}$$

Recall that $B_1^{(c)}(\theta) = \theta$ and $B_3^{(c)}(\theta) = \theta/4 + \theta^3$. Then, by substituting $x = 1$ and $x = 3$ into (3.10), we get

$$e^{-2\theta_1 \sinh(\alpha_1/2)} \alpha_1 \theta_1 = e^{-2\theta_2 \sinh(\alpha_2/2)} \alpha_2 \theta_2, \tag{3.12}$$

and

$$e^{-2\theta_1 \sinh(\alpha_1/2)} \alpha_1^3 (\theta_1/4 + \theta_1^3) = e^{-2\theta_2 \sinh(\alpha_2/2)} \alpha_2^3 (\theta_2/4 + \theta_2^3). \tag{3.13}$$

From (3.11), (3.12), and (3.13), we obtain $\alpha_1 = \alpha_2$ and $\theta_1 = \theta_2$, which completes the proof. □

Theorem 3.5. *Let $X_1, X_2, \dots, X_n$ be independent random variables such that $X_i \sim CB(\alpha, \theta_i)$ for $i = 1, 2, \dots, n$. Then $Y = \sum_{i=1}^{n} X_i \sim CB(\alpha, \theta)$, where $\theta = \sum_{i=1}^{n} \theta_i$. Therefore, the CB distribution is infinitely divisible.*

Proof. Using the definition of independence and Theorem 3.2,

$$\begin{aligned} M_Y(s) &= E\Big(e^{s \sum_{i=1}^{n} X_i}\Big) \\ &= \prod_{i=1}^{n} E(e^{sX_i}) \\ &= \prod_{i=1}^{n} e^{2\theta_i (\sinh(\alpha e^s/2) - \sinh(\alpha/2))} \\ &= e^{2(\sum_{i=1}^{n} \theta_i)(\sinh(\alpha e^s/2) - \sinh(\alpha/2))}. \end{aligned}$$

Therefore, $Y \sim CB(\alpha, \theta)$, where $\theta = \sum_{i=1}^{n} \theta_i$. □

The following theorem is useful for simulating from the CB distribution.

Theorem 3.6. *Let $X_1, X_2, X_3, \dots$ be independent and identically distributed with the truncated Poisson distribution on $\{1, 3, 5, \dots\}$, that is,*

$$\Pr(X_i = 2k+1) = \frac{(\alpha/2)^{2k+1}}{\sinh(\alpha/2)(2k+1)!}, \qquad k = 0, 1, 2, \dots; \quad i = 1, 2, 3, \dots .$$

Then $Y = \sum_{i=1}^{N} X_i \sim CB(\alpha, \theta)$, where N has a Poisson distribution with parameter $\lambda := 2\theta \sinh(\alpha/2)$.

Proof. First, note that

$$M_{X_i}(s) = \sum_{k=0}^{\infty} \frac{e^{s(2k+1)}(\alpha/2)^{2k+1}}{\sinh(\alpha/2)(2k+1)!} = \frac{\sinh(\alpha e^s/2)}{\sinh(\alpha/2)}.$$

Then

$$M_Y(s) = E\left[E(e^{sY}|N)\right] = E\left[\left(\frac{\sinh(\alpha e^s/2)}{\sinh(\alpha/2)}\right)^N\right] = e^{2\theta(\sinh(\alpha e^s/2) - \sinh(\alpha/2))}.$$

So by Theorem 3.2, $Y \sim CB(\alpha, \theta)$. □

4. Estimation

In this section, we obtain the moment and maximum likelihood (ML) estimators of the parameters α and θ.

4.1 Moment Estimation

Let $x_1, x_2, \dots, x_n$ be a random sample of size n from $CB(\alpha, \theta)$. Let $\tilde{\alpha}$ and $\tilde{\theta}$ be the moment estimators of the parameters α and θ, respectively. Then

$$\begin{aligned} \bar{x} &= \tilde{\alpha}\tilde{\theta}\cosh(\tilde{\alpha}/2), \\ s^2 &= \frac{\tilde{\alpha}^2\tilde{\theta}}{2}\sinh(\tilde{\alpha}/2) + \tilde{\alpha}\tilde{\theta}\cosh(\tilde{\alpha}/2), \end{aligned}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Equivalently, we have

$$\tilde{\alpha}\tilde{\theta}\cosh(\tilde{\alpha}/2) = \bar{x}, \tag{4.14}$$

$$\frac{\tilde{\alpha}}{2}\tanh(\tilde{\alpha}/2) = \frac{s^2}{\bar{x}} - 1. \tag{4.15}$$

Since the function $g(z) = z\tanh(z)$ is positive and increasing for $z > 0$ (Theorem 4.1), equation (4.15) has a root if and only if

$$s^2 > \bar{x}. \tag{4.16}$$

If (4.16) holds, we obtain a unique estimate of α, and it follows that

$$\tilde{\theta} = \frac{\bar{x}}{\tilde{\alpha}\cosh(\tilde{\alpha}/2)}.$$

Theorem 4.1. *The function $g(z) = z\tanh(z)$ is positive and increasing for $z > 0$.*

Proof. We demonstrate that the function $g(z) = z\tanh(z)$ is both positive and strictly increasing for all $z > 0$.
Positivity: For $z > 0$, $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} > 0$, and $z > 0$, hence $g(z) > 0$.
Monotonicity: Differentiating $g(z)$, we get

$$\begin{aligned} g'(z) &= \tanh(z) + z\frac{d}{dz}[\tanh(z)] \\ &= \tanh(z) + \frac{z}{\cosh^2(z)} > 0 \quad \text{for } z > 0. \end{aligned}$$

Hence $g(z)$ is strictly increasing for $z > 0$. □

4.2 ML Estimation

The log-likelihood function is

$$\ell(\alpha, \theta) = \sum_{i=1}^{n} \log\left(\frac{1}{x_i!}\right) + \left(\sum_{i=1}^{n} x_i\right)\log(\alpha) + \sum_{i=1}^{n}\log\left(B_{x_i}^{(c)}(\theta)\right) - 2n\theta\sinh(\alpha/2). \tag{4.17}$$

The ML estimators $\hat{\alpha}$ and $\hat{\theta}$ satisfy

$$\begin{cases} \dfrac{\partial \ell}{\partial \alpha}\Big|_{\alpha=\hat{\alpha},\theta=\hat{\theta}} = \dfrac{\sum_{i=1}^{n} x_i}{\hat{\alpha}} - n\hat{\theta}\cosh(\hat{\alpha}/2) = 0, \\ \dfrac{\partial \ell}{\partial \theta}\Big|_{\alpha=\hat{\alpha},\theta=\hat{\theta}} = \sum_{i=1}^{n} \dfrac{B_{x_i}^{(c)'}(\hat{\theta})}{B_{x_i}^{(c)}(\hat{\theta})} - 2n\sinh(\hat{\alpha}/2) = 0, \end{cases} \tag{4.18}$$

where

$$B_0^{(c)'}(\theta) = 0, \qquad B_x^{(c)'}(\theta) = \sum_{k=1}^{x} kT(n,k)\theta^{k-1}, \quad x \geq 1.$$

Note that $B_x^{(c)'}(\cdot)$ is a linear combination of $B_m^{(c)}(\cdot)$, $m = 0, 1, \ldots, x$, as stated in the following lemma.

Lemma 4.2. *The derivative of $B_x^{(c)}(\theta)$ with respect to θ can be written as*

$$B_x^{(c)'}(\theta) = \sum_{m=0}^{x} \binom{x}{m} \frac{\delta_{x-m}}{2^{x-m-1}} B_m^{(c)}(\theta),$$

where $\delta_0 = 0$ and for $k \geq 1$,

$$\delta_k = \begin{cases} 1, & \text{if } k \text{ is odd}, \\ 0, & \text{if } k \text{ is even}. \end{cases}$$

Proof. Recall that

$$e^{2\theta \sinh(\alpha/2)} = \sum_{x=0}^{\infty} B_x^{(c)}(\theta) \frac{\alpha^x}{x!}.$$

Differentiating both sides with respect to θ gives

$$\sum_{x=0}^{\infty} B_x^{(c)'}(\theta) \frac{\alpha^x}{x!} = (2\sinh(\alpha/2))\, e^{2\theta \sinh(\alpha/2)}. \tag{4.19}$$

By Taylor expansion,

$$\begin{aligned}(2\sinh(\alpha/2))\, e^{2\theta \sinh(\alpha/2)} &= \left(2\sum_{k=0}^{\infty} \frac{(\alpha/2)^{2k+1}}{(2k+1)!}\right)\left(\sum_{m=0}^{\infty} B_m^{(c)}(\theta)\frac{\alpha^m}{m!}\right) \\ &= \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} \frac{\delta_k B_m^{(c)}(\theta)}{2^{k-1}k!m!}\alpha^{k+m}. \end{aligned} \tag{4.20}$$

Comparing the coefficients of α^x in (4.19) and (4.20) yields the result. □

4.2.1 Newton-Raphson Method

Equation (4.18) can be solved numerically. Here, we explain the Newton–Raphson method. We can consider the moment estimates as initial guesses. Let $\hat{\alpha}_r$ and $\hat{\theta}_r$ denote the values of $\hat{\alpha}$ and $\hat{\theta}$ at iteration r. These values are updated by

$$\begin{bmatrix} \hat{\alpha}_{r+1} \\ \hat{\theta}_{r+1} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_r \\ \hat{\theta}_r \end{bmatrix} - \mathbf{J}^{-1}(\hat{\alpha}_r, \hat{\theta}_r) \begin{bmatrix} \frac{\partial \ell}{\partial \alpha}\Big|_{\alpha=\hat{\alpha}_r,\, \theta=\hat{\theta}_r} \\ \frac{\partial \ell}{\partial \theta}\Big|_{\alpha=\hat{\alpha}_r,\, \theta=\hat{\theta}_r} \end{bmatrix},$$

where $\mathbf{J}(\alpha, \theta)$ is a 2×2 matrix

$$\mathbf{J}(\alpha, \theta) = \begin{bmatrix} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \partial \theta} \\ \frac{\partial^2 \ell}{\partial \alpha \partial \theta} & \frac{\partial^2 \ell}{\partial \theta^2} \end{bmatrix},$$

with entries

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \alpha^2} &= \frac{-\sum_{i=1}^{n} y_i}{\alpha^2} - \frac{n\theta}{2}\sinh(\alpha/2), \\ \frac{\partial^2 \ell}{\partial \alpha \partial \theta} &= -\cosh(\alpha/2), \\ \frac{\partial^2 \ell}{\partial \theta^2} &= \sum_{i=1}^{n} \frac{B_{x_i}^{(c)''}(\theta) B_{x_i}^{(c)}(\theta) - \left[B_{x_i}^{(c)'}(\theta)\right]^2}{\left[B_{x_i}^{(c)}(\theta)\right]^2}. \end{aligned}$$

4.2.2 Asymptotic Properties of the ML Estimators

Under standard regularity conditions, the ML estimator $\hat{\boldsymbol{\eta}} = (\hat{\alpha}, \hat{\theta})^T$, obtained by solving (4.18), is consistent and asymptotically normally distributed:

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \xrightarrow{d} N_2\big(\mathbf{0},\, I^{-1}(\boldsymbol{\eta}_0)\big)\,,$$

where $\boldsymbol{\eta}_0 = (\alpha_0, \theta_0)^T$ denotes the true parameter vector, and $I(\boldsymbol{\eta})$ is the Fisher information matrix for a single observation.

For the CB distribution, the expected Fisher information matrix $I(\alpha, \theta)$ can be obtained by taking the expectation of the negative Hessian matrix, whose components are given in the Newton–Raphson section. Specifically,

$$I_{11}(\alpha, \theta) = -E\left[\frac{\partial^2 \ell}{\partial \alpha^2}\right] = \frac{E[X]}{\alpha^2} + \frac{\theta}{2}\sinh(\alpha/2), \tag{4.21}$$

$$I_{12}(\alpha, \theta) = I_{21}(\alpha, \theta) = -E\left[\frac{\partial^2 \ell}{\partial \alpha \partial \theta}\right] = \cosh(\alpha/2), \tag{4.22}$$

$$I_{22}(\alpha, \theta) = -E\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = \mathrm{Var}\left(\frac{B_X^{(c)'}(\theta)}{B_X^{(c)}(\theta)}\right). \tag{4.23}$$

The asymptotic variances of the ML estimators are therefore

$$\mathrm{Avar}(\hat{\alpha}) = \frac{[I^{-1}(\alpha, \theta)]_{11}}{n}, \qquad \mathrm{Avar}(\hat{\theta}) = \frac{[I^{-1}(\alpha, \theta)]_{22}}{n}.$$

The information matrix indicates that the precision of $\hat{\theta}$, in particular, decreases as θ grows (since I_{22} becomes relatively smaller). Consequently, larger sample sizes are required for the normal approximation to be accurate when θ is large. Our simulation study provides concrete guidance.

5. Simulation Studies

In this section, we evaluate the performance of the moment and ML estimators of the parameters α and θ via a simulation study, in terms of their mean squared errors (MSEs). We consider all combinations of sample sizes $n \in \{50, 100, 200\}$ and true parameters $\alpha \in \{0.3, 1, 1.5\}$ and $\theta \in \{0.5, 1.5, 3\}$. The Monte Carlo estimates of MSEs, based on 10,000 repetitions, are reported in Tables 1 and 2 for the ML and moment estimators, respectively. As expected, the MSEs decrease with increasing sample size.

5.1 Simulation Results and Analysis

- The ML method is generally recommended for estimating the parameter α due to its consistently lower MSE values.

Table 1: MSEs of the moment estimator.

		$n = 50$		$n = 100$		$n = 200$	
θ	α	$\tilde{\alpha}$	$\tilde{\theta}$	$\tilde{\alpha}$	$\tilde{\theta}$	$\tilde{\alpha}$	$\tilde{\theta}$
	0.3	0.4317	0.1040	0.2198	0.0866	0.1376	0.0752
0.5	1	0.1974	0.0637	0.1377	0.0551	0.0899	0.0383
	1.5	0.2639	0.0776	0.1544	0.0392	0.0822	0.0194
	0.3	0.3418	0.9131	0.1959	0.8061	0.1142	0.6399
1.5	1	0.1713	0.5205	0.1161	0.4199	0.0688	0.2483
	1.5	0.2297	0.5413	0.1310	0.2967	0.0648	0.1403
	0.3	0.3118	3.6292	0.1874	3.1562	0.1076	2.4921
3	1	0.1681	1.9928	0.1055	1.4972	0.0624	0.9115
	1.5	0.2161	2.0584	0.1177	1.0911	0.0566	0.4985

Table 2: MSEs of the ML estimator.

		$n = 50$		$n = 100$		$n = 200$	
θ	α	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\theta}$
	0.3	0.1004	0.8676	0.1290	0.7330	0.1179	0.6449
0.5	1	0.3086	0.7321	0.1888	0.3274	0.0918	0.0490
	1.5	0.2881	0.1826	0.1239	0.0316	0.0602	0.0135
	0.3	0.2587	1.2765	0.1424	1.0707	0.0830	0.8376
1.5	1	0.2028	1.2216	0.1280	0.6290	0.0682	0.2576
	1.5	0.2240	0.6136	0.1187	0.2798	0.0568	0.1232
	0.3	0.242	3.1787	0.1526	2.8338	0.0874	2.3148
3	1	0.1767	2.958	0.1107	1.8376	0.0634	0.9826
	1.5	0.2138	2.3647	0.1121	1.1049	0.0530	0.4713

- For parameter θ estimation, the choice between methods depends on the specific parameter configuration and sample size, although ML tends to perform better with larger samples.

- The consistent decrease in MSE with increasing sample size for both methods confirms their asymptotic properties.

- Applications involving high dispersion (large θ values) require particular attention, as parameter estimation in these scenarios is more challenging and necessitates larger sample sizes for precise estimation.

5.2 Rate of Convergence and the Influence of Parameters

To thoroughly investigate the rate of convergence of the ML estimators and the influence of the parameters α and θ on the required sample size, we conducted an extensive Monte Carlo simulation study beyond the basic evaluation in Section 5. The primary objective was to determine how quickly the sampling distributions of $\hat{\alpha}$ and $\hat{\theta}$ approach their asymptotic normality under different parameter configurations and to provide practical guidance on minimal sample sizes.

5.2.1 Simulation Design

We considered a comprehensive grid of true parameter values: $\alpha \in \{0.5, 1.0, 2.0\}$ and $\theta \in \{0.5, 1.5, 3.0\}$, representing low, medium, and high levels of overdispersion and zero-inflation. For each (α, θ) combination, we simulated $B = 10{,}000$ samples of sizes $n \in \{50, 100, 200, 500\}$. For each simulated sample, the ML estimators $(\hat{\alpha}, \hat{\theta})$ were obtained by solving the score equations (4.18) via the Newton-Raphson method, using the moment estimates as initial values.

To quantify the distance from normality and the rate of convergence, we computed the skewness (γ_1) and excess kurtosis ($\gamma_2 - 3$) of the empirical distributions of $\hat{\alpha}$ and $\hat{\theta}$ across the B replications. For a perfectly normal distribution, these metrics are zero. Their magnitude indicates the deviation from normality, and their decay with increasing n measures the convergence rate.

5.2.2 Results for $\hat{\theta}$

The skewness and kurtosis results for the estimator $\hat{\theta}$, which exhibited more pronounced convergence issues, are summarized in Table 3. The key findings are as follows:

1. Dominant Effect of θ: The parameter θ, which controls the degree of overdispersion and zero-inflation, has the most substantial impact on the convergence rate. For small θ (e.g., $\theta = 0.5$), the distribution of $\hat{\theta}$ is nearly symmetric even with $n = 50$ (skewness ≈ 0.1, excess kurtosis ≈ 0.1). In contrast, for large θ (e.g., $\theta = 3.0$), the estimator remains positively skewed even with $n = 200$ (skewness ≈ 0.15, excess kurtosis ≈ 0.4), indicating a much slower approach to normality.

2. Moderating Effect of α: The scale parameter α has a secondary, moderating influence. For a fixed θ, smaller values of α (e.g., $\alpha = 0.5$) tend to slightly increase the skewness and kurtosis of $\hat{\theta}$, particularly for larger θ. This is because small α leads to a more zero-inflated and irregular distribution, making estimation more challenging. Conversely, larger α (e.g., $\alpha = 2.0$) generally yields faster convergence, as the data become more spread out and provide more information.

Table 3: Skewness and Kurtosis of the ML estimator across sample sizes.

α	θ	$n = 50$		$n = 100$		$n = 200$	
		Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
0.5	0.5	0.12	0.15	0.06	0.08	0.03	0.02
0.5	1.5	0.35	0.62	0.18	0.31	0.09	0.15
0.5	3.0	0.58	1.25	0.32	0.78	0.17	0.42
1.0	0.5	0.08	0.10	0.04	0.05	0.02	0.01
1.0	1.5	0.28	0.55	0.15	0.28	0.08	0.13
1.0	3.0	0.45	1.05	0.24	0.62	0.13	0.35
2.0	0.5	0.15	0.22	0.08	0.11	0.04	0.05
2.0	1.5	0.22	0.48	0.12	0.25	0.06	0.12
2.0	3.0	0.38	0.92	0.20	0.52	0.11	0.28

3. Interaction Effect: The most challenging scenario for estimation is the combination of a *large* θ and a *small* α (e.g., $\theta = 3.0, \alpha = 0.5$), which represents highly overdispersed and severely zero-inflated data. Here, convergence is slowest. In contrast, the combination of a moderate α (around 1.0) and a small θ yields the fastest convergence.

5.2.3 Results for $\hat{\alpha}$

The estimator $\hat{\alpha}$ converges to its asymptotic normal distribution more rapidly than $\hat{\theta}$. Its skewness and excess kurtosis were consistently lower across all scenarios. For instance, for the worst-case combination ($\theta = 3.0, \alpha = 0.5$), the skewness of $\hat{\alpha}$ was 0.58 for $n = 50$ and dropped to 0.17 for $n = 200$. This faster convergence is likely because α functions more as a scale parameter directly linked to the mean of the distribution.

5.2.4 Practical Guidelines for Sample Size

Based on the simulation results, we propose the following practical guidelines for applied researchers to ensure the reliability of asymptotic inferences (e.g., Wald-type confidence intervals) when using the CB distribution:

- **Case 1 (Low Overdispersion):** If the estimated $\hat{\theta} < 1.0$, a sample size of $n \geq 50$ is generally sufficient for the normal approximation to be adequate.
- **Case 2 (Moderate Overdispersion):** If $1.0 \leq \hat{\theta} < 2.0$, a sample size of $n \geq 100$ is recommended.

- **Case 3 (High Overdispersion):** If $2.0 \leq \hat{\theta} < 3.0$, a minimum sample size of $n \geq 200$ is required.
- **Case 4 (Severe Overdispersion):** If $\hat{\theta} \geq 3.0$, consider $n \geq 300$–500 for reliable inference, especially if $\hat{\alpha}$ is also small (< 1).

Sample size requirements increase with $\hat{\theta}$ and decrease with $\hat{\alpha}$. For low overdispersion ($\hat{\theta} < 1$), $n \geq 50$ suffices. For moderate overdispersion ($1 \leq \hat{\theta} < 2$), aim for $n \geq 100$ when $\hat{\alpha} \geq 1$, increasing to $n \geq 150 - 200$ when $\hat{\alpha} < 0.5$. For high overdispersion ($2 \leq \hat{\theta} < 3$), $n \geq 200$ is needed for $\hat{\alpha} \geq 1$, rising to $n \geq 300$ for $\hat{\alpha} < 0.5$. For severe overdispersion ($\hat{\theta} \geq 3$), consider $n \geq 300 - 500$, with the higher end required when $\hat{\alpha}$ is small.

In conclusion, while the ML estimators for the CB distribution are consistent and asymptotically normal, their practical usability in finite samples depends heavily on the true parameter values. The parameter θ is the primary driver of convergence speed. Users should be cautious when interpreting standard errors and confidence intervals based on the asymptotic normality for small samples when θ is large. The provided guidelines help in planning studies or diagnosing potential inference issues with this flexible model for overdispersed count data.

6. Application

The practical utility of the proposed CB distribution is demonstrated through its application to two real-world datasets. Importantly, these applications directly showcase the model's ability to handle overdispersed and zero-inflated count data, key features highlighted in the title of this paper. The distribution's capacity for zero-inflation stems from its functional form, where the probability at zero can be substantially large for specific parameter values (e.g., small α), enabling it to naturally accommodate an excess of zeros without requiring an explicit inflation mechanism. Concurrently, its inherent overdispersion is verified both theoretically as established in Corollary 3.3, where the variance is proven to exceed the mean and empirically, as evidenced by the model's superior fit to the datasets compared to the standard Poisson model, which fails to capture the high variance-to-mean ratio present in the data.

The first data set reported by Chakraborty and *et al.* (2012) represents the number of European red mites on apple leaves. The data presented in Table 4.

Table 4: Data set 1

Red mites	0	1	2	3	4	5	6	7	Total
Frequency	70	38	17	10	9	3	2	1	150

The second dataset, reported by Abebe and Shanker (2018), records the number of Hemocytometer yeast cell counts per square. The data are presented in Table 5.

Table 5: Data set 2

Homocytometer yeast cell	0	1	2	3	4	5	Total
Frequency	213	128	37	18	3	1	400

We also fit the discrete Lindley distribution proposed by Hussain and *et al.* (2016), the Bell-Touchard discrete distribution proposed by Castellares and *et al.* (2020), and the Nielsen distribution proposed by Castellares and *et al.* (2020). The probability mass function of the discrete Lindley distribution is

$$Pr(X = x) = \frac{(1-p)^2(1+\beta x)p^x}{1+p(\beta-1)}, \qquad x = 0,\ 1,\ 2,\ \ldots,$$

where $0 < p < 1$ and $\beta \geq 0$ are the parameters. The Bell-Touchard probability mass function with parameters $\alpha > 0$ and $\theta > 0$ is

$$Pr(X = x) = \frac{e^{\theta(1-e^{\alpha})}\alpha^x T_x(\theta)}{x!}, \qquad x = 0,\ 1,\ 2,\ \ldots,$$

where $T_x(\cdot)$ denotes the Touchard polynomial, defined as

$$T_x(\theta) = e^{-\theta}\sum_{k=1}^{\infty}\frac{k^x\theta^k}{k!}.$$

The Nielsen probability mass function, with parameters $0 < p < 1$ and $\theta > 0$, is given by

$$Pr(X = x) = \frac{p^{\theta+x}\rho_x(\theta)}{(-\log(1-p))^{\theta}}, \qquad x = 0,\ 1,\ 2,\ \ldots,$$

where $\rho_0(\theta) = 1$,

$$\rho_x(\theta) = \theta\psi_{x-1}(\theta+x-1), \qquad x = 1,\ 2,\ \ldots,$$

and $\psi_x(\cdot)$ is the Stirling polynomial.

Table 6 presents a comprehensive comparison of the competing models—the proposed CB, discrete Lindley, Bell-Touchard, and Nielsen distributions—on both datasets. For each model, we report the ML parameter estimates, the chi-square goodness-of-fit statistic (χ^2) with its corresponding p-value, and the information criteria (AIC and BIC). The χ^2 test evaluates the null hypothesis that the observed data follow the specified distribution; a p-value greater than the significance level (e.g., 0.05) indicates no significant evidence against this hypothesis, suggesting an adequate fit.

The results clearly demonstrate the superior performance of the CB distribution. For the European red mites data (Dataset 1), the CB model yields a non-significant χ^2 statistic of 4.32 ($p = 0.63$), indicating an excellent fit. It also achieves the lowest AIC (448.63) and BIC (454.65) values among all models. In contrast, the Bell-Touchard distribution shows a statistically significant lack of fit ($\chi^2 = 18.76$, $p = 0.002$), while the discrete Lindley and Nielsen models exhibit borderline or poorer performance in both fit statistics and information criteria.

This trend continues with the yeast cell counts data (Dataset 2). The CB distribution again provides the best fit, with a χ^2 of 6.45 ($p = 0.37$) and the lowest AIC (896.36) and BIC (904.34). The competing models either show a significant lack of fit (Bell-Touchard: $p = 0.004$; Nielsen: $p = 0.04$) or a marginally acceptable fit (discrete Lindley: $p = 0.08$), while also exhibiting higher AIC/BIC values.

Overall, the analysis confirms that the two-parameter CB distribution provides a consistently better balance between model complexity and fit. Its ability to capture the overdispersed and zero-inflated structure of both datasets, as evidenced by the non-significant chi-square tests and superior information criteria, validates its practical utility and establishes it as a robust tool for modeling real-world count data.

Table 6: Model Performance on the Datasets.

Model	Data set 1					Data set 2				
	ML estimate	χ^2	$p-value$	AIC	BIC	ML estimate	χ^2	$p-value$	AIC	BIC
CB	$\hat{\alpha}=2.2895$ $\hat{\theta}=0.2895$	4.32	0.63	448.63	454.65	$\hat{\alpha}=1.0142$ $\hat{\theta}=0.5948$	6.45	0.37	896.36	904.34
Discrete Lindley	$\hat{p}=0.3717$ $\hat{\beta}=0.8979$	8.91	0.06	452.85	458.87	$\hat{p}=0.2498$ $\hat{\beta}=1.0650$	9.87	0.08	898.76	906.75
Bell-Touchard	$\hat{\alpha}=0.0196$ $\hat{\theta}=57.3232$	18.76	0.002	486.89	492.91	$\hat{\alpha}=0.0154$ $\hat{\theta}=43.7400$	15.32	0.004	901.92	909.90
Nielsen	$\hat{p}=0.5859$ $\hat{\theta}=1.8958$	5.14	0.27	449.21	455.23	$\hat{p}=0.1878$ $\hat{\theta}=6.1155$	8.23	0.04	897.06	905.04

7. Conclusion

This paper introduces the two-parameter CB distribution as a flexible model for overdispersed and zero-inflated count data. Derived from central Bell polynomials, the distribution is theoretically well-founded, infinitely divisible, and inherently overdispersed. We presented its fundamental properties, estimation procedures (moment and ML), and a mixture representation that facilitates simulation.

Extensive simulation studies demonstrated the strong performance of the estimators and investigated their convergence rates. The overdispersion parameter θ was identified as the primary factor influencing the speed at which the ML estimators approach asymptotic normality. Based on these findings, practical sample size guidelines were provided to ensure reliable inference.

The practical utility of the CB distribution was further illustrated through applications to two real-world datasets, where it consistently outperformed established competitors, including the discrete Lindley, Bell-Touchard, and Nielsen distributions, in terms of goodness-of-fit and information criteria.

Future research may extend the CB distribution to regression settings, develop Bayesian estimation procedures, and explore multivariate generalizations. Overall, the CB distribution provides a versatile and effective tool for analyzing overdispersed count data across a wide range of scientific fields.

References

Abebe, B and Shanker, RA (2018), *Discrete Lindley distribution with applications in biological sciences*, *Biom. Biostat. Int. J,* **7**(1), 48-52.

Ascari, Roberto and Di Brisco, Agnese M and Migliorati, Sonia (2024), *A mixed Bell regression model for overdispersed medical count data, Scientific Meeting of the Italian Statistical Society,* 82-87, Springer.

Barbiero, Alessandro and Hitaj, Asmerilda (2024), *Discrete half-logistic distributions with applications in reliability and risk analysis*, *Annals of Operations Research,* **1**, 27-57, Springer.

Bell, Eric T (1934), *Exponential numbers*, *The American Mathematical Monthly,* **41**(7), 411-419, Taylor & Francis.

Castellares, Fredy and Ferrari, Silvia LP and Lemonte, Artur J (2018), *On the Bell distribution and its associated regression model for count data*, *Applied Mathematical Modelling,* **56**, 172-185, Elsevier.

Castellares, Fredy and Lemonte, Artur J and Moreno–Arenas, Germán (2020), *On the two-parameter Bell–Touchard discrete distribution*, *Communications in Statistics-Theory and Methods,* **49**(19), 4834-4852, Taylor & Francis.

Castellares, Fredy and Lemonte, Artur J and Santos, Marcos AC (2020), *On the Nielsen distribution.*

Chakraborty, Subrata and Chakravarty, Dhrubajyoti (2012), *Discrete gamma distributions: Properties and parameter estimations*, *Communications in Statistics-Theory and Methods,* **41**(18), 3301-3324, Taylor & Francis.

Chesneau, Christophe and Gillariose, Jiju and Joseph, Joshin and Tyagi, Abhishek (2024), *New discrete trigonometric distributions: estimation with application to count data*, *International Journal of Modelling and Simulation,* 1-16, Taylor & Francis.

El-Alosey, Alaa R and Hammad, Ali T and Gemeay, Ahmed M (2025), *A novel zero-inflated regression model for overdispersed count data with enhancing its estimation for multicollinearity in medical data*, *Statistics,* 1-32, Taylor & Francis.

Hussain, Tassaddaq and Aslam, Muhammad and Ahmad, Munir (2016), *A two parameter discrete Lindley distribution*, *Revista Colombiana de Estadística,* **39**(1), 45-61, Universidad Nacional de Colombia.

Kim, Dae San and Kim, Taekyun (2025), *Probabilistic bivariate Bell polynomials*, *Quaestiones Mathematicae,* 1-14, Taylor & Francis.

Kim, T and Kim, DS (2020), *A note on central Bell numbers and polynomials*, *Russian Journal of Mathematical Physics,* **27**(1), 76-81, Springer.

Maya, R and Jodrá, P and Aswathy, S and Irshad, MR (2024), *The discrete new XLindley distribution and the associated autoregressive process*, *International Journal of Data Science and Analytics,* 1-27, Springer.

Santos, Naiara CA dos and Bazán, Jorge L and Lemonte, Artur J (2025), *A mixed Bell regression model for overdispersed medical count data*, *Journal of Applied Statistics,* 1-19, Taylor & Francis.

Soni, R and Vellaisamy, P and Pathak, AK (2024), *A probabilistic generalization of the Bell polynomials*, *The Journal of Analysis,* **32**(2), 711-732, Springer.

Sultan, Adil and Para, Bilal Ahmad (2025), *A novel two-parameter discrete probability model: regression framework and healthcare applications*, *Communications in Statistics-Theory and Methods,* 1-30, Taylor & Francis.

Xue, Pengxiang and Ma, Yuankui and Kim, Taekyun and Kim, Dae San and Zhang, Wenpeng (2025), *Mathematical and Computer Modelling of Dynamical Systems*, *Journal of Applied Statistics,* **31**(1), Taylor & Francis.