




Article

On the Conflation of Negative Binomial and Logarithmic Distributions

Anfal A. Alqefari ¹, Abdulhamid A. Alzaid ¹ and Najla Qarmalah ^{2,*}

¹ Department of Statistics and Operations Research, College of Sciences, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; 438203970@student.ksu.edu.sa (A.A.A.); alzaid@ksu.edu.sa (A.A.A.)

² Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

* Correspondence: nmbinqarmalah@pnu.edu.sa; Tel.: +966-18236238

Abstract: In recent decades, the study of discrete distributions has received increasing attention in the field of statistics, mainly because discrete distributions can model a wide range of count data. One common distribution used for modeling count data, for instance, is the negative binomial distribution (NBD), which performs well with over-dispersed data. In this paper, a new count distribution is introduced, called the conflation of negative binomial and logarithmic distributions, which is formed by conflating the negative binomial and logarithmic distributions, resulting in a distribution that possesses some of the properties of negative binomial and logarithmic distributions. The distribution has two parameters and is verified by a positive integer. Two modifications are proposed to the distribution, which includes zero as a support point. The new distribution is valuable from a theoretical perspective since it is a member of the weighted negative binomial distribution family. In addition, the distribution differs from the NBD in the sense that the probability of lower counts is inflated. This study discusses the characteristics of the proposed distribution and its modified versions, such as moments, probability generating functions, likelihood stochastic ordering, log-concavity, and unimodality properties. Real-world data are used to evaluate the performance of the proposed models against other models. All computations shown in this paper were produced using the R programming language.



Citation: Alqefari, A.A.; Alzaid, A.A.; Qarmalah, N. On the Conflation of Negative Binomial and Logarithmic Distributions. *Axioms* **2024**, *13*, 707. <https://doi.org/10.3390/axioms13100707>

Keywords: conflation distributions; weighted distribution; shifted distribution; negative binomial distribution; logarithmic distribution; discrete distributions; dispersion index; likelihood ratio stochastic order; log-concavity; unimodality

MSC: 60E05; 62F10; 65C20

Academic Editors: Yolanda Gómez and Inmaculada Barranco-Chamorro

Received: 19 August 2024
Revised: 7 October 2024
Accepted: 11 October 2024
Published: 13 October 2024



Copyright: © 2024 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/>).

1. Introduction

The analysis and modeling of count data have received significant attention in recent decades, with a particular emphasis being placed on the development of discrete distributions. A widely used model for the analysis and modeling of count data is the Poisson distribution (PD). The important condition for the PD is an equal dispersion of count data. The equal dispersion can be evaluated using a statistical measure known as an index of dispersion (ID). The ID defines the quantity of variability in a distribution. The definition of the ID is given below:

Definition 1. *The index of dispersion of a distribution, denoted as ID, can be defined as follows:*

$$ID = \frac{\text{Var}(Y)}{\mu},$$

where $\text{Var}(Y)$ and μ are the variance and the mean of Y , respectively. The ID implies the following:

- If $ID > 1$, then it is over-dispersion.
- If $ID < 1$, then it is under-dispersion.
- If $ID = 1$, then it is equal dispersion.

For more details, see [1].

However, this condition is rarely observed in practical scenarios. Count data often show over-dispersion, thereby requiring an investigation of modeling alternatives, which provide greater flexibility than the PD. The negative binomial distribution (NBD) is a frequently employed alternative distribution for modeling count data, particularly for data that show over-dispersion. The NBD is extensively employed in the modeling of diverse datasets, including biological, medical sciences, accident statistics, social sciences, economics, quality control, ecology, and so forth. The study [2] outlines the NBD as a mixture of PDs, and the mean of the PD is a random variable following a gamma distribution. The probability mass function (*pmf*) of the NBD is denoted as follows:

$$f(x) = \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} p^x (1-p)^r, \quad 0 < p < 1, r > 0, x = 0, 1, 2, \dots$$

For more details about the NBD and its properties, see [1].

Furthermore, count data frequently display an excess of zeros and heterogeneity in variance, making traditional statistical distributions insufficient for modeling purposes. Nevertheless, the NBD is frequently favored over the PD because of its ability to provide increased flexibility in modeling data that appear over-dispersed, as previously indicated. Many studies have developed alternative models to deal with the presence of excess zeros and variability in the dataset. For example, zero-inflated models, hurdle models, or finite mixture models have been proposed in order to more efficiently deal with this issue. In addition, the mixing of PD or NBD with a lifetime distribution is frequently employed for the same issue. For example, numerous research studies have demonstrated that the mixed negative binomial distribution offers a superior fit for count data in comparison to the PD and NBD. Moreover, weighted distributions are used to solve the problem by multiplying count distributions with weight functions, as developed in [3,4]. Since then, the concept of weighted distributions has established itself in the literature as a powerful tool for modeling. By allowing us to adjust probabilities based on specific weights assigned to each outcome, weighted distributions provide a flexible framework that enhances our ability to accurately represent and analyze complex real-world phenomena. This adaptability has made weighted distributions invaluable in various fields such as statistics, biostatistics, biomedicine, ecology, survival data analysis, meta-analysis, and intervention data analysis. For discrete distributions, the weighted distribution is defined as follows:

Definition 2. Let X be a random variable with pmf $f(x)$ and let $w(x)$ be a non-negative weighting function such that $0 < E[w(X)] < \infty$ exists and is finite. Then, the pmf $f_w(x)$ is defined as follows:

$$f_w(x) = \frac{w(x)f(x)}{E[w(X)]}, \quad x \in \mathbf{N}.$$

is a weighted distribution of $f(x)$.

For more information on weighted distributions for discrete random variables, refer to the work [1] on the subject. For example, the modified negative binomial distribution in [5] can be viewed as a weighted geometric distribution.

Although the standard distributions possess attractive characteristics, they do not provide the best fit for real-world data that have deviations. The source of deviations can be either inflation in low counts or high dispersion. Hence, there is a need to develop new distributions that demonstrate superior performance. In this study, we employ the idea of conflation as a tool to cope with the presence of excess zeros or generally low counts and

over-dispersed data. The concept of the conflation of probability distributions is presented by [6] and defined as follows:

Definition 3. If f_1, f_2, \dots, f_n are pmfs, then the corresponding conflated distribution f_C is

$$f_C(x) = \frac{\prod_{i=1}^n f_i(x)}{\sum_{y \in A} \prod_{i=1}^n f_i(y)}, \quad x \in \mathbf{N}. \quad (1)$$

where A is the intersection of the supports of all the distributions. In terms of random variables, if X_1, \dots, X_n are independent with pmfs f_1, \dots, f_n , respectively, then

$$f_C(x) = P(X_1 = x \mid X_1 = X_2 = \dots = X_n).$$

Ref. [6] presented conflation as a method for consolidating data from several independent experiments, all of which were designed to measure the same unknown quantity. In other words, distribution conflation is a distribution that inherits some properties from its components. For $n = 2$, Equation (1) can be viewed as a weighted distribution, where one mass function is the parent distribution and the other one is a weight function. In this sense, the conflation distributions are weighted distributions. Hence, one may use conflation methods to model data with a high excess of low counts and over-dispersion by conflating a distribution with a decreasing mass function with an over-dispersed distribution.

As a result, we introduce a new distribution by combining the NBD and the logarithmic distribution (LD) into a single distribution that reflects the common information between them with minimal loss of information. The pmf of the LD is given by:

$$f_{LD}(x) = \frac{-1}{\log(1-p)} \frac{p^x}{x}, \quad 0 < p < 1, \quad x = 1, 2, 3, \dots$$

For more details about the logarithmic distribution and its properties, see [1].

The new distribution is called the conflation of negative binomial and logarithmic distributions (CNBLD). The LD has a decreasing pmf, hence it is capable of modeling data with a high frequency for low counts while the NBD is over-dispersed; therefore, their conflation is expected to handle data expressing high excesses of low counts and over-dispersion. The LD does not support zero, hence the CNBLD inherits this property, which limits the applications of the CNBLD to positive count data. To overcome this problem, the study also presents two modifications of the CNBLD. The first modification shifts the CNBLD one position to the left, resulting in the shifted CNBLD that is denoted as SCNBLD. The SCNBLD retains the flexibility of the CNBLD but extends its support to zero values. The second modification conflates a shifted logarithmic distribution with the NBD, resulting in the conflation of a negative binomial shift logarithmic distribution (CNBSLD). The CNBSLD also aims to combine the features of both distributions to provide flexibility and the ability to model a wider range of data.

The structure of this paper is organized as follows: Section 2 presents the definitions and discusses the graphical representations of the proposed models; Section 3 describes some of the statistical properties of the proposed models, such as moments, log-concavity, index of dispersion, and likelihood ratio stochastic order; Section 4 discusses the estimation of the parameters using the method of moments and the maximum likelihood method and evaluates the accuracy of these estimates by a simulation study; Section 5 outlines the usefulness of the new distribution across several fields, showing its superior performance compared to the existing modified negative binomial distributions employed to fit similar data; finally, Section 6 presents a conclusion.

2. Conflation of Negative Binomial and Logarithmic Distributions

In this section, the conflation of negative binomial logarithmic distributions (CNBLD), and the developed versions of the CNBLD are introduced. The developed versions of

the CNBLD are named as follows: shifted conflation of negative binomial logarithmic distributions (SCNBLD) and conflation of negative binomial weighted by shift logarithmic distribution (CNBSLD). This section outlines the *pmfs*, the cumulative distribution functions (*cdf*), which are denoted as $F(\cdot)$, the survival functions (*sf*), which are denoted as $(\bar{F}(\cdot))$, and the hazard rate functions (*h*) of the CNBLD, SCNBLD, and CNBSLD.

Definition 4. The random variable Y is said to follow the CNBLD with parameters $r > 0$ and $0 < \theta < 1$ if its *pmf* is given as follows:

$$f(y) = P(Y = y) = C^{-1}(r, \theta) \frac{\Gamma(y + r)\theta^y}{\Gamma(y + 1)\Gamma(r)y}, \quad y = 1, 2, \dots$$

Here, $C^{-1}(r, \theta)$ is the normalizing constant that can be expressed as follows:

$$\begin{aligned} C(r, \theta) &= \sum_{k=1}^{\infty} \frac{\Gamma(k + r)}{\Gamma(k + 1)\Gamma(r)} \frac{\theta^k}{k} \\ &= \sum_{z=0}^{\infty} \frac{\Gamma(z + r + 1)}{\Gamma(z + 2)\Gamma(r)} \frac{\theta^{z+1}}{z + 1} \\ &= \theta r \sum_{z=0}^{\infty} \frac{(1)_z(1)_z(r + 1)_z}{(2)_z(2)_z} \frac{\theta^z}{z!} \\ &= \theta r {}_3F_2(1, 1, r + 1; 2, 2; \theta). \end{aligned}$$

where $(a)_u = \frac{\Gamma(a+u)}{\Gamma(a)}$ is the Pochhammer symbol, and ${}_3F_2$ is a generalized hypergeometric function; for more details, see [1,7].

The generalized hypergeometric function is available in popular programming packages such as R, Mathematica, MATLAB, Python, and others. In this paper, we used the `genhypergeo(.)` function from the `hypergeo` package in R.

In comparison with (1), in terms of random variables, if X and Z are independent random variables with X following an NBD and Z following an LD, then $Y \stackrel{d}{=} X \mid X = Z$. In the special case when $r = 1$, the CNBLD reduces to the LD.

Remark 1. It should be noted that according to Definition 4, the CNBLD is a weighted negative binomial distribution with an LD as the weight function. In addition, the CNBLD can be considered as a weighted logarithmic distribution with an NBD as the weighting function.

Next, the *pmfs* of the CNBLD are visualized for different values of parameters $\theta = 0.2, 0.5, \text{ and } 0.8$ and $r = 1, 4, \text{ and } 8$ in Figure 1. The parameter θ has an impact on the dispersion of the distribution, while the parameter r is mainly responsible for the shape of the CNBLD. In general, the shape of the *pmfs* of the CNBLD is skewed to the right; however, with an increase in the value of r and θ , the distribution becomes less skewed and displays more symmetry. On the other hand, for smaller values of θ and r , the *pmf* of the CNBLD is a decreasing function with a high probability for low y values. However, as θ and r increase, the function’s behavior shifts, initially rising to a peak before decreasing. This change illustrates how larger parameters introduce greater variability into the distribution.

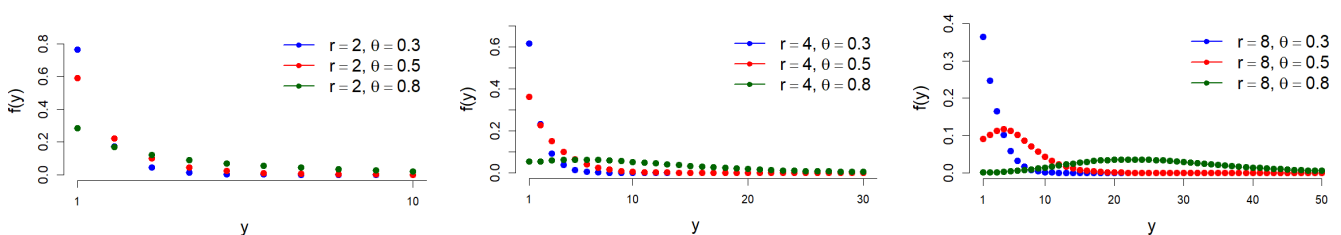


Figure 1. The *pmfs* of the CNBLD for different values of θ and r .

The *cdf*, *sf*, and *h* of the CNBLD, respectively, are as follows:

$$F(y) = 1 - \frac{\theta^y \Gamma(y+r+1) {}_3F_2(1, y+1, y+r+1; y+2, y+2; \theta)}{\Gamma(y+2)\Gamma(r+1)(y+1) {}_3F_2(1, 1, r+1; 2, 2; \theta)},$$

$$\bar{F}(y) = \frac{\theta^y \Gamma(y+r+1) {}_3F_2(1, y+1, y+r+1; y+2, y+2; \theta)}{\Gamma(y+2)\Gamma(r+1)(y+1) {}_3F_2(1, 1, r+1; 2, 2; \theta)},$$

and

$$h(y) = \frac{(y+1)^2}{\theta y(y+r) {}_3F_2(1, y+1, y+r+1; y+2, y+2; \theta)}.$$

We have

$$\begin{aligned} \bar{F}(y) &= \sum_{t=y+1}^{\infty} C^{-1}(r, \theta) \frac{\Gamma(t+r)}{\Gamma(t+1)\Gamma(r)} \frac{\theta^t}{t} \\ &= C^{-1}(r, \theta) A(y) \end{aligned}$$

where

$$\begin{aligned} A(y) &= \sum_{t=y+1}^{\infty} \frac{\Gamma(t+r)}{\Gamma(t+1)\Gamma(r)} \frac{\theta^t}{t} \\ &= \sum_{z=0}^{\infty} \frac{\Gamma(z+y+r+1)}{\Gamma(z+y+2)\Gamma(r)} \frac{\theta^{z+y+1}}{z+y+1} \\ &= \theta^{y+1} \sum_{z=0}^{\infty} \frac{(1)_z (y+1)_z (y+r+1)_z \Gamma(y+r+1)}{(y+2)_z (y+2)_z (y+1)\Gamma(y+2)(r-1)!} \frac{\theta^z}{z!} \\ &= \frac{\theta^{y+1} \Gamma(y+r)}{\Gamma(y+1)\Gamma(r-1)(y+1)} \sum_{z=0}^{\infty} \frac{(1)_z (y+1)_z (y+r+1)_z}{(y+2)_z (y+2)_z} \frac{\theta^z}{z!} \\ &= \frac{\theta^{y+1} \Gamma(y+r)}{\Gamma(y+1)\Gamma(r-1)(y+1)} {}_3F_2(1, y+1, y+r+1; y+2, y+2; \theta) \end{aligned}$$

Thus, according to *A(y)*, the *sf* can be given as follows:

$$\bar{F}(y) = \frac{\theta^y \Gamma(y+r+1) {}_3F_2(1, y+1, y+r+1; y+2, y+2; \theta)}{\Gamma(y+2)\Gamma(r+1)(y+1) {}_3F_2(1, 1, r+1; 2, 2; \theta)}.$$

Using the definition of the *sf*, the *cdf* can be defined as follows:

$$F(y) = 1 - \bar{F}(y)$$

Further, the *h* of the CNBLD can be calculated as follows:

$$h(y) = \frac{f(y)}{\bar{F}(y)}.$$

Most real-life count data have zero as a possible value. Therefore, the current study developed the CNBLD using two methods for this purpose. The first method was the obvious one which shifted the CNBLD by one to the left. The second method shifted the LD conflated with the NBD. Therefore, we obtained the following two definitions:

Definition 5. The random variable *Y* is said to follow the SCNBLD with parameters $r > 0$ and $0 < \theta < 1$, if its pmf is given by the following:

$$f(y) = \frac{\Gamma(y+r+1)\theta^y}{\Gamma(y+2)\Gamma(r+1) {}_3F_2(1, 1, r+1; 2, 2; \theta)(y+1)}, \quad y = 0, 1, 2, \dots$$

Consequently, we obtain the *cdf*, the *sf*, and the *h* functions of the SCNBLD, respectively, as follows:

$$F(y) = 1 - \frac{\theta^{y+1}\Gamma(y+r+2) {}_3F_2(1, y+2, y+r+2; y+3, y+3; \theta)}{\Gamma(y+3)\Gamma(r+1)(y+2) {}_3F_2(1, 1, r+1; 2, 2; \theta)},$$

$$\bar{F}(y) = \frac{\theta^{y+1}\Gamma(y+r+2) {}_3F_2(1, y+2, y+r+2; y+3, y+3; \theta)}{\Gamma(y+3)\Gamma(r+1)(y+2) {}_3F_2(1, 1, r+1; 2, 2; \theta)},$$

and

$$h(y) = \frac{(y+2)^2}{\theta(y+1)(y+r+1) {}_3F_2(1, y+2, y+r+2; y+3, y+3; \theta)}.$$

Definition 6. The random variable Y is said to follow the CNBSLD with parameters $r > 0$ and $0 < \theta < 1$ if its pmf is given as follows:

$$f(y) = \frac{\theta(r-1)\Gamma(y+r)\theta^y}{\Gamma(y+1)\Gamma(r)((1-\theta)^{-(r-1)} - 1)(y+1)}, \quad y = 0, 1, \dots$$

Remark 2. Note that the CNBSLD is a shifted LD for $r = 1$ and a geometric distribution for $r = 2$, and hence the CNBSLD can be considered as an extension of the two distributions.

The following theorem can be used to derive the CNBSLD.

Theorem 1. If X and Z are independent random variables following an NBD with parameters $r > 0$ and $0 < p_1 < 1$ and a shifted logarithmic distribution with parameter $0 < p_2 < 1$, then

$$P(Y = y) = P(X = y \mid X = Z)$$

where $\theta = p_1 p_2$.

Proof. The proof can be obtained directly by calculating the conditional probability. \square

The *cdf*, *sf*, and the *h* of the CNBSLD are, respectively:

$$F(y) = 1 - \frac{\theta^{y+2}(r-1)\Gamma(y+r+1) {}_2F_1(1, y+r+1; y+3; \theta)}{\Gamma(y+2)\Gamma(r)(y+2)[(1-\theta)^{-(r-1)} - 1]},$$

$$\bar{F}(y) = \frac{\theta^{y+2}(r-1)\Gamma(y+r+1) {}_2F_1(1, y+r+1; y+3; \theta)}{\Gamma(y+2)\Gamma(r)(y+2)[(1-\theta)^{-(r-1)} - 1]},$$

and

$$h(y) = \frac{y+2}{\theta(y+r) {}_2F_1(1, y+r+1; y+3; \theta)}.$$

Here, ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$ is the Gaussian hypergeometric function (see [1,7] for more information). It is possible to calculate the *sf* of the CNBSLD as follows:

$$\begin{aligned} \bar{F}(y) &= \sum_{t=y+1}^{\infty} \frac{\theta(r-1)\Gamma(t+r)\theta^t}{\Gamma(t+1)\Gamma(r)((1-\theta)^{-(r-1)} - 1)(t+1)} \\ &= \frac{\theta(r-1)}{((1-\theta)^{-(r-1)} - 1)(t+1)} B(y) \end{aligned}$$

where

$$\begin{aligned}
 B(y) &= \sum_{t=y+1}^{\infty} \frac{\Gamma(t+r)\theta^t}{\Gamma(t+1)\Gamma(r)(t+1)} \\
 &= \sum_{z=0}^{\infty} \frac{\Gamma(z+y+r+1)\theta^{z+y+1}}{\Gamma(z+y+2)\Gamma(r)(z+y+2)} \\
 &= \frac{\theta^{y+1}\Gamma(y+r+1)}{\Gamma(y+2)\Gamma(r)(y+2)} \sum_{z=0}^{\infty} \frac{z!\Gamma(y+r+1+z)\Gamma(y+3)}{\Gamma(y+r+1)\Gamma(y+3+z)} \frac{\theta^z}{z!} \\
 &= \frac{\theta^{y+1}\Gamma(y+r+1)}{\Gamma(y+2)\Gamma(r)(y+2)} \sum_{z=0}^{\infty} \frac{(1)_z(y+r+1)_z}{(y+3)_z} \frac{\theta^z}{z!} \\
 &= \frac{\theta^{y+1}\Gamma(y+r)_2F_1(1, y+r+1; y+3; \theta)}{\Gamma(y+1)\Gamma(r-1)(y+2)}.
 \end{aligned}$$

Thus, according to $B(y)$, the sf can be given as follows:

$$\bar{F}(y) = \frac{\theta^{y+2}(r-1)\Gamma(y+r+1) {}_2F_1(1, y+r+1; y+3; \theta)}{\Gamma(y+2)\Gamma(r)(y+2)[(1-\theta)^{-(r-1)} - 1]}$$

Using the definition of the sf , the cdf can be defined as follows:

$$F(y) = 1 - \bar{F}(y)$$

Further, the h of the CNBSLD can be calculated as follows:

$$h(y) = \frac{f(y)}{\bar{F}(y)}.$$

A comparison between the SCNBLD, the CNBSLD, and the NBD can be made by looking at the $pmfs$ for different values of θ and r . Figure 2 shows the $pmfs$ for $r = 1, 4,$ and 8 with $\theta = 0.2, 0.5,$ and 0.8 . The shape of all distributions is skewed to the right but tends to be symmetric for large θ and r . The difference between the distributions decreases significantly as θ and r increase, and the distributions behave more similarly. The $pmfs$ appears identical for relatively large y , depending on the value of r in all plots with small probabilities of all distributions. For example, the $pmfs$ of all distributions are the same after $y = 4$ when $\theta = 0.25$ and $r = 1$. In general, as the value of r increases, the value of y at which the pmf is constant increases. The plots show that both r and θ have a clear influence on the behavior of the different distributions.

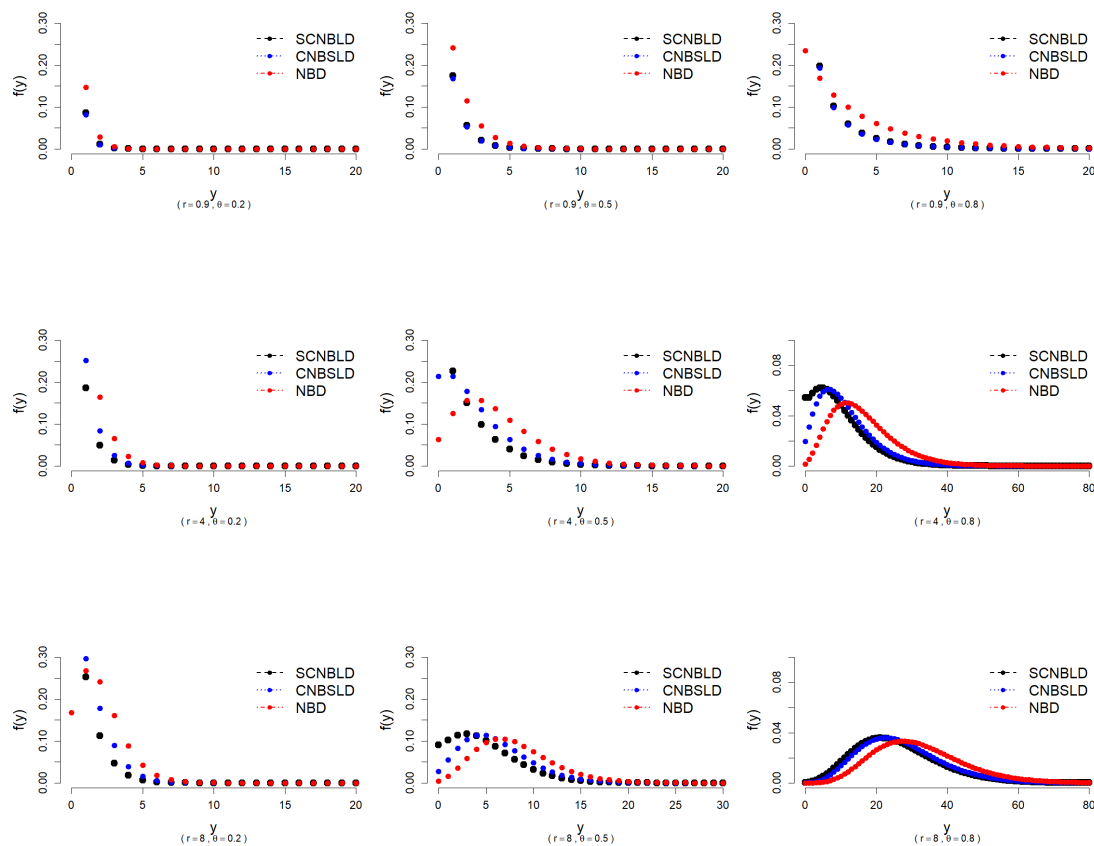


Figure 2. The pmfs of the SCNBLD, the CNBSLD, and the NBD for different values of θ and r .

3. Some Statistical Properties

In this section, we examine several useful statistical properties of the CNBLD, SCNBLD, and CNBSLD. These include deriving the mean, variance, and probability generating functions for each distribution. In addition, we calculate the index of dispersion (ID) for the CNBLD, SCNBLD, and CNBSLD, which provides information about the variability relative to their means. Furthermore, we discuss the likelihood ratio stochastic order and log-concavity property for the new distributions. The likelihood ratio stochastic order study is extended to the NBD to provide a more comprehensive understanding of the relative behaviors and properties of these distributions.

3.1. Moments and Probability Generating Functions

The statistical results for the moment and probability generating functions associated with the CNBLD are reviewed below.

The mean, the variance, and the probability generating function for the CNBLD are given as follows:

$$\mu = C^{-1}(r, \theta) [(1 - \theta)^{-r} - 1],$$

$$Var(Y) = \frac{r^2 \theta^2 (1 - \theta)^{-(r+1)} {}_3F_2(1, 1, r + 1; 2, 2; \theta) - [(1 - \theta)^{-r} - 1]^2}{[\theta r {}_3F_2(1, 1, r + 1; 2, 2; \theta)]^2},$$

and

$$G(s) = \frac{s {}_3F_2(1, 1, r + 1; 2, 2; \theta s)}{{}_3F_2(1, 1, r + 1; 2, 2; \theta)}.$$

The formulas above can be given as follows:

$$\begin{aligned} \mu &= \sum_{y=1}^{\infty} \frac{y C^{-1}(r, \theta) \theta^y \Gamma(y+r)}{y \Gamma(y+1) \Gamma(r)} \\ &= C^{-1}(r, \theta) \sum_{y=1}^{\infty} \frac{\theta^y \Gamma(y+r)}{\Gamma(y+1) \Gamma(r)} \\ &= C^{-1}(r, \theta) [(1-\theta)^{-r} - 1]. \end{aligned}$$

For the variance, the second moment can be expressed as follows:

$$\begin{aligned} E(Y^2) &= \sum_{y=1}^{\infty} \frac{y^2 C^{-1}(r, \theta) \theta^y \Gamma(y+r)}{y \Gamma(y+1) \Gamma(r)} \\ &= C^{-1}(r, \theta) \sum_{y=1}^{\infty} \frac{y \theta^y \Gamma(y+r)}{\Gamma(y+1) \Gamma(r)} \\ &= C^{-1}(r, \theta) r \theta \sum_{z=0}^{\infty} \frac{\Gamma(z+r+1) \theta^z}{\Gamma(z+1) \Gamma(r+1)} \\ &= \frac{(1-\theta)^{-(r+1)}}{{}_3F_2(1, 1, r+1; 2, 2; \theta)} \end{aligned}$$

Hence, the variance can be shown to be as follows:

$$\begin{aligned} Var(Y) &= E(Y^2) - \mu^2 \\ &= \frac{(1-\theta)^{-(r+1)}}{{}_3F_2(1, 1, r+1; 2, 2; \theta)} - [C^{-1}(r, \theta) [(1-\theta)^{-r} - 1]]^2 \\ &= \frac{r^2 \theta^2 (1-\theta)^{-(r+1)} {}_3F_2(1, 1, r+1; 2, 2; \theta) - [(1-\theta)^{-r} - 1]^2}{[\theta r {}_3F_2(1, 1, r+1; 2, 2; \theta)]^2} \end{aligned}$$

The form of the probability generating function becomes obvious from the pmf of the CNBLD.

The following introduces the moments and probability generating function related to the SCNBLD.

$$\begin{aligned} \mu &= C^{-1}(r, \theta) [(1-\theta)^{-r} - 1] - 1, \\ Var(Y) &= \frac{r^2 \theta^2 (1-\theta)^{-(r+1)} {}_3F_2(1, 1, r+1; 2, 2; \theta) - [(1-\theta)^{-r} - 1]^2}{[\theta r {}_3F_2(1, 1, r+1; 2, 2; \theta)]^2}, \end{aligned}$$

and

$$G(s) = \frac{{}_3F_2(1, 1, r+1; 2, 2; \theta s)}{{}_3F_2(1, 1, r+1; 2, 2; \theta)}.$$

The results can be obtained from the fact that $Z = Y - 1$, where Y follows a CNBLD. Finally, the moments and probability generating function of the CNBSLD are as follows:

The mean, the variance, and the probability generating function for the CNBSLD can be given as follows:

$$\begin{aligned} \mu &= \frac{(1 - \theta)^r + r\theta - 1}{1 - (1 - \theta)^r - \theta}, \\ \text{Var}(Y) &= \frac{\theta(r - 1)[(1 - \theta)^r[\theta(r - 1) + 1] + \theta - 1]}{(\theta - 1)[(1 - \theta)^r + \theta - 1]^2}, \\ \text{and} \\ G(s) &= \frac{(1 - s\theta)^{-(r-1)} - 1}{s[(1 - \theta)^{-(r-1)} - 1]}. \end{aligned}$$

They can be obtained as follows:

$$\begin{aligned} E(Y + 1) &= \sum_{y=0}^{\infty} \frac{(y + 1)\theta(r - 1)\theta^y\Gamma(y + r)}{((1 - \theta)^{-(r-1)} - 1)(y + 1)\Gamma(y + 1)\Gamma(r)} \\ &= \frac{\theta(r - 1)}{(1 - \theta)^{-(r-1)} - 1} \sum_{y=0}^{\infty} \frac{\Gamma(y + r)\theta^y}{\Gamma(y + 1)\Gamma(r)} \\ &= \frac{\theta(r - 1)(1 - \theta)^{-r}}{(1 - \theta)^{-(r-1)} - 1} \\ &= \frac{\theta(r - 1)}{1 - (1 - \theta)^r - \theta}. \end{aligned}$$

Hence,

$$\begin{aligned} \mu &= E(Y) \\ &= E(Y + 1) - 1 \\ &= \frac{\theta(r - 1)}{1 - (1 - \theta)^r - \theta} - 1 \\ &= \frac{(1 - \theta)^r + r\theta - 1}{1 - (1 - \theta)^r - \theta}. \end{aligned}$$

For the variance, we obtain the following:

$$\begin{aligned} E[Y(Y + 1)] &= \sum_{y=0}^{\infty} \frac{y(y + 1)\theta(r - 1)\Gamma(y + r)\theta^y}{((1 - \theta)^{-(r-1)} - 1)(y + 1)\Gamma(y + 1)\Gamma(r)} \\ &= \frac{\theta(r - 1)}{(1 - \theta)^{-(r-1)} - 1} \sum_{y=0}^{\infty} \frac{y\theta^y\Gamma(y + r)}{\Gamma(y + 1)\Gamma(r)} \\ &= \frac{\theta(r - 1)\theta r(1 - \theta)^{-(r+1)}}{(1 - \theta)^{-(r-1)} - 1} \\ &= \frac{\theta^2 r(r - 1)}{(\theta - 1)[(1 - \theta)^r + \theta - 1]}. \end{aligned}$$

Hence, the variance is

$$\begin{aligned} \text{Var}(Y) &= E[y(y + 1)] - E(y) - \mu^2 \\ &= \frac{\theta^2 r(r - 1)}{(\theta - 1)[(1 - \theta)^r + \theta - 1]} - \frac{(1 - \theta)^r + r\theta - 1}{1 - (1 - \theta)^r - \theta} - \left[\frac{(1 - \theta)^r + r\theta - 1}{1 - (1 - \theta)^r - \theta} \right]^2 \\ &= \frac{\theta(r - 1)[(1 - \theta)^r(\theta(r - 1) + 1) + \theta - 1]}{(\theta - 1)[(1 - \theta)^r + \theta - 1]^2}. \end{aligned}$$

The form of the probability generating function becomes obvious from the shape of the pmf of the CNBSLD.

3.2. Index of Dispersion

In this subsection, we introduce the ID of the NBD, SCNBLD, and CNBSLD that are denoted as ID_{NBD} , ID_{SCNBLD} , and ID_{CNBSLD} , respectively, for different values of r and θ . The ID_{NBD} , ID_{SCNBLD} , and ID_{CNBSLD} for different values of θ and r are calculated in Table 1. Since ID_{SCNBLD} and ID_{CNBSLD} have complicated mathematical formulas, the IDs are calculated for selected values of r and θ .

- ID_{NBD} is given by $\frac{1}{1-\theta}$. This implies that as θ increases, ID_{NBD} increases, indicating higher dispersion with higher θ .
- For SCNBLD and CNBSLD, as θ increases, ID_{SCNBLD} and ID_{CNBSLD} increase. This means that for a fixed r , the dispersion increases as θ increases.
- For SCNBLD and CNBSLD, as r increases, ID_{SCNBLD} and ID_{CNBSLD} also increase. This suggests that for a fixed θ , the dispersion increases as r increases.
- If $r > 1$, then $ID_{CNBSLD} < ID_{SCNBLD}$; as a result, the SCNBLD is more suitable for data with greater dispersion. This suggests that the value of r determines the interchangeability of the two distributions.

Table 1. The index of dispersion for the NBD, SCNBLD, and CNBSLD for different values of r and θ .

r	θ	SCNBLD			CNBSLD			NBD		
		Mean	Variance	ID	Mean	Variance	ID	Mean	Variance	ID
0.5	0.2	0.0883	0.1042	1.1807	0.0590	0.0699	1.1840	0.1250	0.1562	1.25
	0.5	0.3079	0.5220	1.6955	0.2071	0.3536	1.7071	0.5	0.5	2
	0.8	0.9104	3.2622	3.5833	0.6180	2.2361	3.6180	2	10	5
2	0.2	0.1889	0.2378	1.2592	0.25	0.3125	1.25	0.5	0.625	1.25
	0.5	0.7718	1.5855	2.0541	1	2	2	2	4	2
	0.8	3.2785	17.3486	5.2916	4	20	5	8	40	5
5	0.2	0.4323	0.6115	1.4143	0.6937	0.9421	1.3579	1.25	1.5625	1.25
	0.5	2.3418	6.0804	2.5965	3.2667	7.3956	2.2639	5	10	2
	0.8	13.5341	79.5349	5.8766	15.0256	79.7172	5.3054	20	100	5
8	0.2	0.7403	1.1536	1.5581	1.2143	1.7396	1.4325	2	2.5	1.25
	0.5	4.7797	13.0137	2.7226	6.0551	13.7213	2.2660	8	16	2
	0.8	25.7471	140.5006	5.4569	27.0003	139.9917	5.1848	32	160	5

3.3. Log-Concavity Property

Log-concave probability distributions are essential in various areas, including reliability theory, labor economics, monopoly theory, mechanism design theory, political science, and law. Refer to [8] for additional information.

Definition 7. A discrete random variable X is log-concave if $f^2(x + 1) \geq f(x)f(x + 2)$ for all x .

Theorem 2. The pmf of the CNBLD is log-concave for $r \geq 7$ and log-convex for $r \leq 2$.

Proof. For $y = 1, 2, 3, \dots$, we have

$$A(y, r) = \frac{(f(y + 1))^2}{f(y)f(y + 2)} = \frac{\left(\frac{\Gamma(y+r+1)}{\Gamma(y+2)}\right)^2}{\frac{\Gamma(y+r)\Gamma(y+r+2)}{\Gamma(y+1)\Gamma(y+3)}} \cdot \frac{y(y + 2)}{(y + 1)^2}$$

Using the property that $\Gamma(y + 1) = y\Gamma(y)$, we can obtain:

$$\begin{aligned} A(y, r) &= \frac{y(y + 2)^2}{(y + 1)^3} \frac{(y + r)}{(y + 1 + r)} \\ &= \frac{y(y + 2)^2}{(y + 1)^3} \left(1 - \frac{1}{y + 1 + r}\right). \end{aligned}$$

As a result, we observe that $A(y, r)$ shows an increase in r for $y = 1, 2, 3, \dots$. Therefore, $A(1, r) \geq 1$ implies $A(y, r) \geq 1$ for any y , but $A(1, r) \geq 1$ if and only if $\frac{9}{8} \cdot \frac{1+r}{2+r} \geq 1$. Equivalently, this is true if and only if $r \geq 7$. Thus, $A(y, r) \geq 1$ indicates that $f(y)$ is log-concave when $r \geq 7$.

For $r \leq 2$, $A(y, r) \leq A(y, 2)$, resulting in

$$A(y, r) \leq \frac{y(y + 2)^2(y + 2)}{(y + 1)^3(y + 3)} \leq \frac{(y + 1)(y + 2)^2(y + 2)}{(y + 2)(y + 1)^2(y + 3)} = \frac{(y + 2)^2}{(y + 1)(y + 3)} \leq 1.$$

This completes the proof. \square

The CNBLD offers a flexible alternative to the NBD, with properties that depend on the parameter r . While the NBD is log-concave and unimodal for $r \geq 1$ and log-convex for $r \leq 1$, the CNBLD has similar properties, but with different transition points: it is log-concave and unimodal when $r \geq 7$ and log-convex when $r \leq 2$. Moreover, the transition from log-convex to log-concave in the CNBLD is gradual as r increases from 2 to 7, which improves its ability to model more diverse and precise datasets. This flexibility makes the CNBLD particularly well suited for developing precise statistical models that better fit the unique characteristics of the data and allow for a more effective analysis and interpretation compared to the NBD.

Remark 3. Using a similar argument, we can conclude that $A(2, r) \geq 1$ if and only if $r \geq 3.4$. Thus, for $3.4 \leq r < 7$, the pmf of the CNBLD is log-concave on the set $\{2, 3, 4, \dots\}$. The transition from log-convexity to log-concavity occurs gradually as r rises from 2 to 7.

Remark 4. The SCNBLD is log-convex for $r \leq 2$ and log-concave for $r \geq 7$, in contrast to the NBD, which is log-convex for $r \leq 1$ and log-concave for $r \geq 1$. This is because log-concavity does not change with shifting.

Theorem 3. The CNBSLD is log-convex for $r \leq 1$ and log-concave for $r \geq 1$.

Proof. The conclusion is derived from the log-concavity of the NBD, which remains unchanged by truncation and shifting. \square

3.4. Likelihood Ratio Stochastic Order

The likelihood ratio stochastic ordering provides a powerful method for comparing distributions, regardless of whether they belong to the same family with different parameters or are of completely different types. We can determine the ordering relationship between random variables by analyzing the likelihood ratio, which gives us insights into their probabilistic behavior and trends. In this section, we discuss the likelihood ratio stochastic ordering for our new distributions. We also extend this discussion to compare the likelihood ratio stochastic ordering for our new distributions with the NBD.

First, we introduce the definition of the likelihood ratio stochastic order used in this subsection.

Definition 8. Let Y_1 and Y_2 be two discrete random variables with pmfs $f(y)$ and $g(y)$, respectively. We say that Y_1 is smaller than Y_2 in the likelihood ratio stochastic order (denoted by $Y_1 \leq_{lr} Y_2$) if the ratio $\frac{g(y)}{f(y)}$ is non-decreasing in y over the union of the supports of Y_1 and Y_2 .

The likelihood ratio stochastic order is very strong; it implies the hazard stochastic order and other stochastic orders. For more details on the implications and applications of stochastic ordering, see [9]. In this subsection, we refer to the CNBLD with parameters θ and r as CNBLD(r, θ).

Theorem 4. Let Y_1 and Y_2 be two random variables following CNBLD(θ_1, r) and CNBLD(θ_2, r), respectively. If $\theta_1 \leq \theta_2$, then $Y_1 \leq_{lr} Y_2$.

Proof. Let $f_{(\theta, r)}(y)$ be the pmf of the CNBLD(θ, r). Then, we obtain the following:

$$\frac{f_{(\theta_2, r)}(y)}{f_{(\theta_1, r)}(y)} = \Phi_r(\theta_1, \theta_2) \left(\frac{\theta_2}{\theta_1}\right)^y.$$

Here, $\Phi_r(\theta_1, \theta_2) = \frac{C(r, \theta_2)}{C(r, \theta_1)}$. Since the ratio $\left(\frac{\theta_2}{\theta_1}\right)^y$ increases in y if and only if $\theta_1 \leq \theta_2$, this implies $Y_1 \leq_{lr} Y_2$. \square

Remark 5. For the SCNBLD and the CNBSLD, the following implications hold:

- If Y_1 and Y_2 are two random variables following SCNBLD(θ_1, r) and SCNBLD(θ_2, r), respectively, such that $\theta_1 \leq \theta_2$, then $Y_1 \leq_{lr} Y_2$.
- If Y_1 and Y_2 are two random variables following CNBSLD(θ_1, r) and CNBSLD(θ_2, r), respectively, such that $\theta_1 \leq \theta_2$, then $Y_1 \leq_{lr} Y_2$.

Proof. The proof is similar to the proof of Theorem 4. \square

Theorem 5. Let Y_1 and Y_2 be two random variables following CNBLD(θ, r_1) and CNBLD(θ, r_2), respectively. If $r_1 \leq r_2$, then $Y_1 \leq_{lr} Y_2$.

Proof. Let $f_{(\theta, r)}(y)$ be the pmf of CNBLD(θ, r). Then, we obtain the following:

$$\begin{aligned} \frac{f_{(\theta, r_2)}(y)}{f_{(\theta, r_1)}(y)} &= \Psi_\theta(r_1, r_2) \frac{(y + r_2 - 1)!}{(y + r_1 - 1)!} \\ &= \Psi_\theta(r_1, r_2) (y + r_2 - 1)(y + r_2 - 2) \dots (y + r_1). \end{aligned}$$

Here $\Psi_\theta(r_1, r_2) = \frac{C(r_2, \theta)}{C(r_1, \theta)(r_2 - 1)(r_2 - 2) \dots r_1!}$. Since, the ratio $\frac{f_{(\theta, r_2)}(y)}{f_{(\theta, r_1)}(y)}$ increases in y if and only if $r_1 \leq r_2$, this implies $Y_1 \leq_{lr} Y_2$. \square

Remark 6. For the SCNBLD and the CNBSLD, the following implications hold:

- If Y_1 and Y_2 are two random variables following SCNBLD(θ, r_1) and SCNBLD(θ, r_2), respectively, such that $r_1 \leq r_2$, then $Y_1 \leq_{lr} Y_2$.
- If Y_1 and Y_2 are two random variables following CNBSLD(θ, r_1) and CNBSLD(θ, r_2), respectively, such that $r_1 \leq r_2$, then $Y_1 \leq_{lr} Y_2$.

Proof. The proof is similar to the proof of Theorem 5. \square

Corollary 1. Let $Y_{(\theta_1, r_1)}$ be a random variable from CNBLD(θ_2, r_2). If $r_1 \leq r_2$, and $\theta_1 \leq \theta_2$, then we conclude from Theorems 4 and 5 the following:

$$Y_{(\theta_1, r_1)} \leq_{lr} Y_{(\theta_2, r_1)} \leq_{lr} Y_{(\theta_2, r_2)}.$$

Hence, the following is given:

$$Y_{(\theta_1, r_1)} \leq_{lr} Y_{(\theta_2, r_2)}.$$

Theorem 6. Let $Y_1, Y_2,$ and Y_3 be three random variables following $SCNBLD(\theta, r), CSNBLD(\theta, r)$ and $NBD(\theta, r),$ respectively. Then, $Y_1 \leq_{lr} Y_2 \leq_{lr} Y_3.$

Proof. To prove that $Y_1 \leq_{lr} Y_2,$ we examine the following ratio:

$$\frac{f_{SCNBLD}(y)}{f_{CNBSLD}(y)} = \Lambda_1(r, \theta) \frac{y+r}{y+1}.$$

where $\Lambda_1(r, \theta) = \frac{(1-\theta)^{-(r-1)} - 1}{\theta r(r-1) {}_3F_2(1, 1, r+1; 2, 2; \theta)}$. We observe that the term $\frac{y+r}{y+1}$ is an increasing function of y when $r > 1.$ Therefore, $Y_1 \leq_{lr} Y_2.$

Similarly, we need to examine the following ratio:

$$\frac{f_{NBD}(y)}{f_{SCNBLD}(y)} = \Lambda_2(r, \theta) \frac{(y+1)^2}{y+r}.$$

Here, $\Lambda_2(r, \theta) = r(1-\theta)^r {}_3F_2(1, 1, r+1; 2, 2; \theta).$ We observe that the term $\frac{(y+1)^2}{y+r}$ is an increasing function of $y.$ Hence, $Y_2 \leq_{lr} Y_3.$

Since we have showed that $Y_1 \leq_{lr} Y_2$ and $Y_2 \leq_{lr} Y_3,$ we conclude that $Y_1 \leq_{lr} Y_2 \leq_{lr} Y_3.$ This means that in the likelihood ratio stochastic order, Y_3 is stochastically larger than $Y_2,$ and Y_2 is stochastically larger than $Y_1.$

In various fields such as economics, insurance, and risk management, the stochastic order of the likelihood ratio is concerned with risk analysis and decision-making by identifying which distributions are more or less likely to produce large values. □

4. Estimation and Simulation Study

This section examines the estimation of CNBLD and CNBSLD parameters using the method of moment (MM) and the maximum likelihood (ML) method. Two scenarios were investigated: one where r was known, and another where r was unknown. In all scenarios, it was assumed that Y_1, Y_2, \dots, Y_n represented a random sample drawn from the distribution under study. Furthermore, simulation studies were employed to assess the efficiency of the estimates delivered by the proposed methods.

4.1. Parameter Estimation of the CNBLD

4.1.1. Case 1: r Is Known

Here, we had only one parameter θ to estimate using the MM as follows:

$$E(Y) = \bar{y},$$

Or equivalently, this can be achieved as follows:

$$C^{-1}(r, \theta) [(1-\theta)^{-r} - 1] - \bar{y} = 0. \tag{2}$$

Using the ML method, the likelihood function can be given as follows:

$$L(\theta, r | y_1, y_2, \dots, y_n) = \frac{\theta^{\sum_{i=1}^n y_i - n}}{[r! {}_3F_2(1, 1, r+1; 2, 2; \theta)]^n} \prod_{i=1}^n \frac{\Gamma(y_i + r)}{y_i \Gamma(y_i + 1)}, \tag{3}$$

Further, the log-likelihood function $l(\theta)$ from Equation (3) above is given as follows:

$$\begin{aligned} \ell(\theta) &= \left(\sum_{i=1}^n y_i - n \right) \ln \theta - n [\ln \Gamma(r+1) + \ln {}_3F_2(1, 1, r+1; 2, 2; \theta)] \\ &\quad + \sum_{i=1}^n [\ln \Gamma(y_i + r) - \ln \Gamma(y_i + 1) - \ln y_i] \end{aligned}$$

The ML estimate of θ is the solution of the equation $\frac{\partial l(\theta)}{\partial \theta} = 0$. Now, since $\frac{\partial}{\partial \theta} {}_3F_2(1, 1, r+1; 2, 2; \theta) = \frac{(r+1)}{4} {}_3F_2(2, 2, r+2; 3, 3; \theta)$, we have:

$$\begin{aligned} & \left(\sum_{i=1}^n y_i - n \right) \frac{1}{\theta} - \frac{n(r+1)}{4} \frac{{}_3F_2(2, 2, r+2; 3, 3; \theta)}{{}_3F_2(1, 1, r+1; 2, 2; \theta)} = 0, \\ & \left(\sum_{i=1}^n y_i - n \right) \frac{1}{\theta} - \frac{n(r+1)}{4} \frac{4(1-\theta)^{-r} - 4\theta r {}_3F_2(1, 1, r+1; 2, 2; \theta) - 4}{\theta^2 r(1+r)} = 0, \\ & \left(\sum_{i=1}^n y_i - n \right) \frac{1}{\theta} - \frac{n}{\theta^2 r} \frac{(1-\theta)^{-r} - \theta r {}_3F_2(1, 1, r+1; 2, 2; \theta) - 1}{{}_3F_2(1, 1, r+1; 2, 2; \theta)} = 0, \\ & \frac{(1-\theta)^{-r} - 1}{\theta r {}_3F_2(1, 1, r+1; 2, 2; \theta)} - \bar{y} = 0. \\ & \iff C^{-1}(r, \theta) [(1-\theta)^{-r} - 1] - \bar{y} = 0. \end{aligned} \tag{4}$$

which is the same as Equation (2). Thus, if r is known, the ML estimate of θ is the same as that of the MM estimate.

4.1.2. Case 2: r Is Unknown

To obtain the MM estimates of θ and r , the following equations were used:

$$E(Y) = \bar{y},$$

and

$$E(Y^2) = \frac{1}{n} \sum_{i=1}^n y_i^2,$$

Or equivalently, the following could be used:

$$\frac{(1-\theta)^{-r} - 1}{\theta r {}_3F_2(1, 1, r+1; 2, 2; \theta)} - \bar{y} = 0 \tag{5}$$

and

$$\frac{(1-\theta)^{-(r+1)}}{{}_3F_2(1, 1, r+1; 2, 2; \theta)} - \frac{1}{n} \sum_{i=1}^n y_i^2 = 0. \tag{6}$$

Numerical solutions for Equations (5) and (6) are not achievable using algebraic methods. Therefore, the equations were solved using the “nleqslv” function from the “nleqslv” package in the R programming language, which gave the MM estimate of r and θ .

Using the ML method, the likelihood function as defined in Equation (3) were used to determine the ML estimates for the unknown parameters r and θ as the solutions of the likelihood equations as the following:

$$\begin{aligned} \frac{\partial \ell(\theta, r)}{\partial \theta} &= \frac{\partial [(\sum_{i=1}^n y_i - n) \ln \theta - n(\ln {}_3F_2(1, 1, r+1; 2, 2; \theta))]}{\partial \theta} = 0, \\ \frac{\partial \ell(\theta, r)}{\partial r} &= \frac{\partial [-n(\ln \Gamma(r+1) + \ln {}_3F_2(1, 1, r+1; 2, 2; \theta)) + \sum_{i=1}^n \ln \Gamma(y_i + r)]}{\partial r} = 0, \end{aligned}$$

These two equations must be solved numerically to determine the MLEs of the parameters θ and r . Since the generalized hypergeometric function ${}_3F_2$ makes the system of equations complex, numerical methods are generally used to obtain the solutions.

4.2. Simulation Study for the CNBLD

To evaluate the methods of estimation, we performed the following simulations as outlined below:

4.2.1. Case 1: r Is Known

In this simulation study, we considered the values $r = 1, 5$ and 10 and $\theta = 0.2, 0.5$ and 0.8 . The simulation algorithm consisted of the following steps:

Algorithm 1: Simulation algorithm wher r is known

1. Choose the values r, θ and the sample size $n = 10, 20, 50, 100$, and 500 .
2. Generate a total of 1000 random samples of size n from a standard uniform distribution, namely, $\{U_1, U_2, \dots, U_n\}$ such that $U_i \sim Unif(0, 1)$ and $i = 1, \dots, n$.
3. Divide the unit interval $[0, 1]$ into intervals: $I_j = [F(Y_{j-1}), F(Y_j)]$, where $F(Y_j)$ is the *cdf* of the CNBLD for $Y_j, j = 1, 2, \dots$.
4. Find j such that $U_i \in I_j$.
5. Return Y_j .
6. Repeat steps 4 and 5 for $i = 1, 2, \dots, n$.
7. Use the likelihood function in Equation (3) that takes these parameters as input and returns a negative log-likelihood function.
8. Find the ML for θ .
9. Repeat steps 1 to 8 for $N = 1000$ times to calculate the following:
 - (a) The standardized bias of the simulated estimates is defined as follows:

$$SBias(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^N \frac{(\hat{\theta} - \theta)}{\theta}$$

- (b) The average of the mean squared error of the simulated estimates is defined as follows:

$$MSE(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^N (\hat{\theta} - \theta)^2$$

Simulation and concluding results:

The simulation results are shown in Tables 2–4 below.

Table 2. The ML results of the CNBLD for different values of θ when $r = 1$.

n	θ	$\hat{\theta}$	SBias	MSE
10	0.2	0.1874	0.0627	0.0153
20		0.1912	0.0437	0.0075
50		0.1931	0.0343	0.0055
100		0.1954	0.0229	0.0037
500		0.1996	0.0018	0.0018
10	0.5	0.4789	0.0421	0.0106
20		0.4883	0.0233	0.0059
50		0.4962	0.0074	0.0048
100		0.4971	0.0058	0.0016
500		0.5006	0.0012	0.0009
10	0.8	0.7841	0.0199	0.0049
20		0.7919	0.0101	0.0021
50		0.7974	0.0032	0.0009
100		0.7984	0.0018	0.0004
500		0.7987	0.0015	0.0002

Table 3. The ML results of the CNBLD for different values of θ when $r = 5$.

n	θ	$\hat{\theta}$	 SBias 	MSE
10	0.2	0.1879	0.0604	0.0063
20		0.1909	0.0452	0.0032
50		0.1982	0.0091	0.0013
100		0.1985	0.0077	0.0006
500		0.2005	0.0025	0.0001
10	0.5	0.4892	0.0215	0.0044
20		0.4961	0.0078	0.0021
50		0.4980	0.0039	0.0008
100		0.5004	0.0008	0.0004
500		0.5003	0.0006	0.0001
10	0.8	0.7961	0.0049	0.0009
20		0.7981	0.0024	0.0004
50		0.7994	0.0008	0.0002
100		0.7998	−0.0003	0.0001
500		0.8001	0.0002	0.00002

Table 4. The ML results of the CNBLD for different values of θ when $r = 10$.

n	θ	$\hat{\theta}$	 SBias 	MSE
10	0.2	0.1920	0.0396	0.0025
20		0.1977	0.0114	0.0013
50		0.1983	0.0083	0.0005
100		0.1997	0.0011	0.0002
500		0.1999	0.0003	0.0001
10	0.5	0.4981	0.0037	0.0014
20		0.4987	0.0025	0.0007
50		0.4991	0.0019	0.0002
100		0.4995	0.0009	0.0001
500		0.4996	0.0007	0.000004
10	0.8	0.7984	0.0019	0.0004
20		0.7992	0.0009	0.0002
50		0.7995	0.0006	0.00007
100		0.8002	0.0003	0.000004
500		0.8000	0.00004	0.000002

From Tables 2–4, it is possible to conclude the following:

- The MSE of θ decreases along with an increase in n , and thus the estimator of θ is consistent.
- The |SBias| decreases as θ increases.

4.2.2. Case 2: r Is Unknown

Numerical solutions were used to obtain the parameter estimates of θ and r , when r was unknown. In that scenario, we considered $r = 1, 4$ and 8 and $\theta = 0.2, 0.5$, and 0.8 . The simulation algorithm involved the following steps:

Algorithm 2: Simulation algorithm wher r is unknown

1. Generate 1000 random samples of size n from the CNBLD following steps 1–6 in Algorithm 1.
 2. Solve the system of nonlinear equations in Equations (5) and (6) and find the MM estimates for θ and r .
 3. Find the ML estimates for θ and r using Equation (3).
 4. Repeat steps 1 to 3 for $N = 1000$ times to calculate the Bias, SBias, MSE, and SMSE for both ML and MM methods.
-

Simulation and concluding results:

The results are reported in Table 5. From Table 5, one can conclude the following:

- For MM and ML estimates, the MSE of r and θ decrease as n increases.
- For a large n , both the ML and MM estimates are good, but for a small n , the ML estimate is better than the MM estimate according to small values of the SBias for ML estimates.

Table 5. The MM and ML results for the CNBLD for different values of θ and r .

n	θ		Method	$\hat{\theta}$		SBias	MSE
	r			\hat{r}			
10	0.2		MM	0.3651	0.8258	0.0705	
	1			1.4973	0.4973	0.9397	
	0.2		ML	0.1220	0.3895	0.0122	
	1			1.4926	0.4926	0.8077	
20	0.2		MM	0.2217	0.1088	0.0090	
	1			1.2123	0.2123	0.7119	
	0.2		ML	0.2390	0.1951	0.0086	
	1			1.1533	0.1533	0.6424	
50	0.2		MM	0.1869	0.0655	0.0071	
	1			1.1565	0.1565	0.5158	
	0.2		ML	0.2038	0.0192	0.0038	
	1			1.1286	0.1286	0.4967	
100	0.2		MM	0.2095	0.0479	0.0043	
	1			0.9843	0.0156	0.2600	
	0.2		ML	0.1969	0.0175	0.0022	
	1			1.0454	0.0454	0.2554	
500	0.2		MM	0.2047	0.0239	0.0039	
	1			0.9971	0.0028	0.0038	
	0.2		ML	0.1967	0.0161	0.0019	
	1			0.9935	0.0064	0.0026	
10	0.5		MM	0.4710	0.0579	0.0022	
	4			4.3024	0.0756	0.7953	
	0.5		ML	0.5091	0.0182	0.0015	
	4			4.1335	0.0333	0.1841	
20	0.5		MM	0.5041	0.0083	0.0019	
	4			4.2085	0.0521	0.2568	
	0.5		ML	0.4932	0.0134	0.0005	
	4			4.0683	0.0171	0.0733	
50	0.5		MM	0.4980	0.0039	0.0003	
	4			4.1167	0.0292	0.2368	
	0.5		ML	0.5046	0.0093	0.0001	
	4			3.9921	0.0019	0.0584	
100	0.5		MM	0.5009	0.0019	0.0002	
	4			3.9884	0.0028	0.1885	
	0.5		ML	0.4960	0.0078	0.00009	
	4			3.9982	0.0004	0.0049	

Table 5. Cont.

n	θ	Method	$\hat{\theta}$	SBias	MSE
	r		\hat{r}		
500	0.5	MM	0.4975	0.0048	0.00005
	4		3.9954	0.0011	0.0041
	0.5	ML	0.4994	0.0011	0.00002
	4		3.9993	0.0002	0.0007
10	0.8	MM	0.8015	0.0019	0.0001
	8		7.9200	0.0799	0.2022
	0.8	ML	0.7969	0.0038	0.00008
	8		8.1446	0.0181	0.1093
20	0.8	MM	0.7992	0.0009	0.00002
	8		8.0291	0.0036	0.0668
	0.8	ML	0.7994	0.0008	0.00001
	8		8.0045	0.0006	0.0569
50	0.8	MM	0.7994	0.0007	0.00001
	8		8.0232	0.0029	0.0211
	0.8	ML	0.7999	0.0001	0.000009
	8		8.0031	0.0004	0.0199
100	0.8	MM	0.7997	0.0004	0.000001
	8		8.0196	0.0024	0.0107
	0.8	ML	0.7998	0.0003	0.0000005
	8		8.0029	0.0003	0.0014
500	0.8	MM	0.7999	0.0001	0.0000008
	8		8.0122	0.0015	0.0010
	0.8	ML	0.7998	0.00003	0.0000002
	8		7.9999	0.000009	0.0003

4.3. Parameter Estimation of the CNBSLD

In this subsection, estimation methods and simulation study were conducted to assess the performance of the CNBSLD as follows:

4.3.1. Case 1: r Is Known

Here, we had only one parameter θ to estimate using the MM. It was obtained by the following equation:

$$E(Y) = \bar{y},$$

Or equivalently, this could be achieved with the following:

$$\frac{(1 - \theta)^r + r\theta - 1}{1 - (1 - \theta)^r - \theta} - \bar{y} = 0.$$

Using the ML method, the likelihood function can be given as follows:

$$L(\theta, r | y_1, y_2, \dots, y_n) = \prod_{i=1}^n \left(\frac{\theta(r - 1)\theta^{y_i}\Gamma(y_i + r)}{((1 - \theta)^{-(r-1)} - 1)(y_i + 1)\Gamma(y_i + 1)\Gamma(r)} \right)$$

Simplified, the likelihood function becomes:

$$L(\theta, r | y_1, y_2, \dots, y_n) = \left(\frac{\theta(r - 1)}{(1 - \theta)^{-(r-1)} - 1} \right)^n \theta^{\sum_{i=1}^n y_i} \prod_{i=1}^n \left(\frac{\Gamma(y_i + r)}{(y_i + 1)\Gamma(y_i + 1)\Gamma(r)} \right). \tag{7}$$

Further, the log-likelihood function $l(\theta)$ from Equation (7) above is given as follows:

$$\ell(\theta) = n \log\left(\frac{\theta(r-1)}{(1-\theta)^{-(r-1)} - 1}\right) + \sum_{i=1}^n y_i \log(\theta) + \sum_{i=1}^n \log\left[\frac{\Gamma(y_i+r)}{(y_i+1)\Gamma(y_i+1)\Gamma(r)}\right]$$

The ML estimate of θ is the solution of the equation $\frac{\partial l(\theta)}{\partial \theta} = 0$. Then,

$$\frac{n}{\theta} + \frac{(1-\theta)^{-r}n(1-r)}{(1-\theta)^{1-r} - 1} + \frac{\sum_{i=1}^n y_i}{\theta} = 0$$

$$\frac{(1-\theta)^{-r}(\theta r - 1) + 1}{(1-\theta)^{1-r} - 1} - \bar{y} = 0.$$

4.3.2. Case 2: r Is Unknown

To obtain the MM estimates of θ and r , the following equations were used:

$$E(Y) = \bar{y},$$

and

$$E(Y^2) = \frac{1}{n} \sum_{i=1}^n y_i^2,$$

Or equivalently, the following could be used:

$$\frac{(1-\theta)^r + r\theta - 1}{1 - (1-\theta)^r - \theta} - \bar{y} = 0 \tag{8}$$

and

$$\frac{\theta r(\theta r - 1) - (1-\theta)^{r+1} - \theta + 1}{(\theta - 1)[(1-\theta)^r + \theta - 1]} - \frac{1}{n} \sum_{i=1}^n y_i^2 = 0. \tag{9}$$

Numerical solutions for Equations (8) and (9) are not achievable using algebraic methods. Therefore, the equations were solved numerically, which gave the MM estimate of r and θ .

Using the ML method, the likelihood function as defined in Equation (7) were used to determine the ML estimates for the unknown parameters r and θ as the solutions of the likelihood equations as the following:

$$\frac{\partial \ell(\theta, r)}{\partial \theta} = \frac{\partial \left[n \log\left(\frac{\theta(r-1)}{(1-\theta)^{-(r-1)} - 1}\right) + \sum_{i=1}^n y_i \log(\theta) \right]}{\partial \theta} = 0,$$

$$\frac{\partial \ell(\theta, r)}{\partial r} = \frac{\partial \left[n \log\left(\frac{\theta(r-1)}{(1-\theta)^{-(r-1)} - 1}\right) + \sum_{i=1}^n \log\left(\frac{\Gamma(y_i+r)}{\Gamma(r)}\right) \right]}{\partial r} = 0,$$

These two equations must be solved simultaneously to determine the MLEs of the parameters θ and r .

4.4. Simulation Study for the CNBSLD

To evaluate the methods of estimation, we performed the following simulations as outlined below:

4.4.1. Case 1: r Is Known

In this simulation study, Algorithm 1 of the simulation outlined in Section 4.2.1 was implemented with the exception of the following:

- In step 3, $F(Y_j)$ was the *cdf* of the CNBSLD.
- In step 2, the likelihood function employed was represented by Equation (7).

Simulation and concluding results:

The simulation results are shown in Tables 6–8 below.

From Tables 6–8, it is possible to conclude the following:

- The MSE decreases along with an increase in n .
- The $|SBias|$ decrease as n increases.
- Both $|SBias|$ and MSE show that $n = 10$ or more is required to accurately estimate θ .

Table 6. The ML estimates of the CNBSLD for different values of θ when $r = 0.9$.

n	θ	$\hat{\theta}$	$ SBias $	MSE
10	0.2	0.1805	0.0971	0.0258
20		0.1859	0.0700	0.0139
50		0.1951	0.0244	0.0059
100		0.1966	0.0169	0.0029
500		0.1997	0.0012	0.0006
10	0.5	0.4696	0.0606	0.0201
20		0.4769	0.0461	0.0129
50		0.4895	0.0209	0.0076
100		0.4964	0.0072	0.0034
500		0.4998	0.0004	0.0005
10	0.8	0.7821	0.0222	0.0049
20		0.7919	0.0101	0.0027
50		0.7922	0.0097	0.0017
100		0.7964	0.0044	0.0006
500		0.7998	0.00013	0.00012

Table 7. The ML estimates of the CNBSLD for different values of θ when $r = 5$.

n	θ	$\hat{\theta}$	$ SBias $	MSE
10	0.2	0.1937	0.0314	0.0043
20		0.1971	0.0145	0.0013
50		0.1981	0.0095	0.0007
100		0.1985	0.0071	0.0004
500		0.1995	0.0025	0.00008
10	0.5	0.4914	0.0173	0.0036
20		0.4945	0.0109	0.0019
50		0.4989	0.0022	0.0006
100		0.5003	0.0005	0.0003
500		0.50004	0.00007	0.00006
10	0.8	0.7970	0.0037	0.0025
20		0.7980	0.0024	0.0004
50		0.7991	0.0012	0.0002
100		0.8001	0.0002	0.00008
500		0.80002	0.00003	0.00001

Table 8. The ML estimates of the CNBSLD for different values of θ when $r = 10$.

n	θ	$\hat{\theta}$	 SBias 	MSE
10	0.2	0.1962	0.0191	0.0019
20		0.1972	0.0139	0.0009
50		0.1984	0.0076	0.0004
100		0.1995	0.0025	0.0002
500		0.2004	0.0019	0.00003
10	0.5	0.4978	0.0044	0.0014
20		0.4992	0.0016	0.0007
50		0.4998	0.0002	0.0003
100		0.4999	0.0001	0.000006
500		0.49996	0.00006	0.000002
10	0.8	0.7979	0.0026	0.0004
20		0.7992	0.0009	0.0002
50		0.7999	0.0001	0.00006
100		0.7999	0.00006	0.000007
500		0.800008	0.00001	0.000003

4.4.2. Case 2: r Is Unknown

In this case, a similar strategy and the steps of Algorithm 2 of the simulation outlined in Section 4.2.2 were followed, with the exception of the following:

- In step 1, random samples were generated from the CNBSLD.
- In step 2, the nonlinear equations represented by Equations (8) and (9) were solved to find the MM estimates of θ and r .
- In step 3, Equation (7) was solved numerically to obtain the ML estimates of θ and r .

Simulation and concluding results:

The simulation results are shown in Table 9 below.

Table 9. The MM and ML for the CNBSLD for different values of θ and r .

n	θ		Method	$\hat{\theta}$		 SBias 	MSE
	r			\hat{r}			
10	0.2	MM	0.1804	0.0979	0.0155		
	0.9		1.3265	0.4739	0.5303		
	0.2	ML	0.1881	0.0591	0.0131		
	0.9		1.2100	0.3444	0.2422		
20	0.2	MM	0.1936	0.0315	0.0102		
	0.9		1.1537	0.2819	0.1378		
	0.2	ML	0.1912	0.0439	0.0094		
	0.9		1.1431	0.2701	0.1245		
50	0.2	MM	0.1967	0.0160	0.0071		
	0.9		1.1131	0.2368	0.0901		
	0.2	ML	0.1922	0.0392	0.0062		
	0.9		1.1037	0.2263	0.0789		
100	0.2	MM	0.1984	0.0077	0.0027		
	0.9		1.0856	0.2062	0.0600		
	0.2	ML	0.1980	0.0096	0.0018		
	0.9		1.0726	0.1918	0.0509		
500	0.2	MM	0.2004	0.0021	0.0012		
	0.9		0.8898	0.0112	0.0439		
	0.2	ML	0.1995	0.0023	0.0009		
	1		0.8989	0.0011	0.0261		

Table 9. Cont.

n	θ	Method	$\hat{\theta}$	SBias	MSE
	r		\hat{r}		
10	0.5	MM	0.4532	0.0935	0.0126
	4		3.7684	0.0579	0.6007
	0.5	ML	0.5192	0.0384	0.0031
	4		3.8734	0.0316	0.5942
20	0.5	MM	0.4906	0.0187	0.0028
	4		4.1546	0.0387	0.3911
	0.5	ML	0.4933	0.0133	0.0015
	4		4.0958	0.0239	0.3027
50	0.5	MM	0.4977	0.0045	0.0019
	4		4.0633	0.0158	0.2131
	0.5	ML	0.4982	0.0034	0.0013
	4		4.0153	0.0038	0.1142
100	0.5	MM	0.5011	0.0022	0.0007
	4		4.0362	0.0091	0.0179
	0.5	ML	0.5014	0.0029	0.0002
	4		4.0041	0.0010	0.0155
500	0.5	MM	0.5001	0.0003	0.00009
	4		4.0212	0.0053	0.0099
	0.5	ML	0.5009	0.0018	0.00003
	4		4.0009	0.0002	0.0033
10	0.8	MM	0.7970	0.0036	0.0001
	8		8.1414	0.0177	0.2299
	0.8	ML	0.7972	0.0034	0.00009
	8		8.1306	0.0163	0.1737
20	0.8	MM	0.7989	0.0013	0.00003
	8		7.9725	0.0034	0.0471
	0.8	ML	0.7999	0.0002	0.00002
	8		8.1433	0.0179	0.0337
50	0.8	MM	0.7992	0.0009	0.00001
	8		8.0094	0.0012	0.0175
	0.8	ML	0.80009	0.0001	0.000002
	8		7.9884	0.0014	0.0033
100	0.8	MM	0.7995	0.0006	0.000007
	8		7.9958	0.0005	0.0034
	0.8	ML	0.80001	0.00002	0.000001
	8		7.9992	0.00009	0.0025
500	0.8	MM	0.8002	0.0003	0.000002
	8		8.0039	0.0004	0.0029
	0.8	ML	0.800001	0.000002	0.0000008
	8		8.00001	0.000001	0.0016

From Table 9, one can conclude the following:

- For the MM estimate, the MSE of r and θ decreases as n increases.
- For the ML estimate, the MSE of r and θ decreases as n increases.
- For a large n , both the ML and MM estimates are good, but for a small n , the ML estimate is better than the MM estimate according to the |SBias|.

5. Applications

In this section, the effectiveness of the CNBLD, SCNBLD, and CNBSLD is evaluated using real datasets and compared with other existing distributions. The parameter estimates for these distributions were obtained using the ML method. The tools of comparison used were the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). In general, the smaller the values of these statistics, the better the fit to the data. The calculations were performed using the R programming language.

5.1. The Number of Eggs per Flower Head

According to [10], the zero-truncated Poisson–Lindley distribution (ZTPLD) provides a better fit for data relating to the number of eggs per flower head, shown in Table 10, compared to the zero-truncated PD (ZTPD). The study [11] proposes the application of the ZTPD to the dataset.

Application and Concluding Results

The calculated values for the mean, variance, and dispersion index were 3.0340, 3.3436, and 1.1020, respectively. These values indicated the presence of over-dispersion. In Table 10, when comparing the observed frequencies of the ZTPLD and CNBLD, we noticed that the LD had an impact on the CNBLD by gradually raising the probability of small values of Y. On the other hand, the truncated zero effect on the ZTPLD just increased its probability at zero. Table 10 clearly shows that the CNBLD model provided a superior fit to the data with lower AIC and BIC values.

Table 10. Number of counts of flower heads as per the number of fly eggs.

Y	O.F ¹	ZTPD	ZTPLD	CNBLD
1	22	15.28	26.78	21.42
2	18	21.86	19.77	19.79
3	18	20.84	13.94	16.81
4	11	14.90	9.53	12.45
5	9	8.53	6.37	8.12
6	6	4.06	4.19	4.73
7	3	1.66	2.72	2.51
8	0	0.59	1.74	1.22
9	1	0.19	1.11	0.55
Total	88	88	88	88
ML		$\hat{\theta} = 2.8604$	$\hat{\theta} = 0.7186$	$\hat{\theta} = 0.1267$ $\hat{\tau} = 28.1719$
AIC		335.09	336.76	331.93
BIC		337.57	339.24	336.89

¹ O.F = observed frequency.

5.2. The Number of Hospital Stays by United States Residents Aged 66 and Over

These data cover the number of hospitalizations of United States residents aged 66 and up, as reported by [12]. This dataset had 80.37% zeros with a sample ID of 1.882. These characteristics showed over-dispersion and a high number of zero counts. The zero-inflated negative binomial-generalized exponential (ZINB-GE) distribution was used on the same dataset as in the analysis of [13]. The ZINB-GE distribution outperformed the zero-inflated Poisson distribution (ZIPD) and the zero-inflated negative binomial distribution (ZINBD) in terms of data fit.

Application and Concluding Results

Table 11 shows that the SCNBLD and CNBSLD models suited the data well. These models outperformed the ZINB-GE model based on the lower AIC and BIC values, indicating they could handle over-dispersion and significant numbers of zero counts in the

dataset. As a result, the SCNBLD and CNBSLD models are most suited for modeling hospitalizations of United States residents aged 66 and older.

Table 11. Number of hospital stays by United States residents aged 66 and over.

Y	O.F ¹	PD	ZIPD	NBD	ZINBD	ZINB-GE	SCNBLD	CNBSLD
0	3541	3277.3	1816.5	3544.4	3541.1	3541.1	3543.1	3543.7
1	599	969.9	1609.5	583.5	533.4	601.0	595.8	591.3
2	176	143.5	712.9	177.4	217.7	168.7	167.7	170.9
3	48	14.1	210.6	62.3	68.7	56.4	58.5	59.9
4	20	1.05	46.7	23.3	18.9	21.6	22.9	23.2
5	12	0.06	8.4	9	4.4	9.7	9.7	9.5
6	5	0	1.3	3.6	1.3	4.4	4.3	4.1
7	1	0	0.1	1.4	0.5	1.8	2	1.8
8	4	0	0	0.6	0	1.3	0.9	0.8
Total	4406	4406	4406	4406	4406	4406	4406	4406
ML		$\hat{\theta} = 0.2959$	$\hat{\phi} = 0.6659$ $\hat{\theta} = 0.8859$	$\hat{\theta} = 0.4437$ $\hat{r} = 0.3709$	$\hat{\phi} = 0.6040$ $\hat{r} = 3.9683$ $\hat{\theta} = 0.8415$	$\hat{\phi} = 0.1645$ $\hat{r} = 2.2040$ $\hat{\alpha} = 1.0331$ $\hat{\alpha} = 7.3573$	$\hat{\theta} = 0.5935$ $\hat{r} = 0.1333$	$\hat{\theta} = 0.5337$ $\hat{r} = 0.6252$
AIC		6611.01	6122.84	6023.24	6078	6022	6019.23	6020.16
BIC		6617.40	6135.62	6036	6079	6048	6032.02	6032.94

¹ O.F = observed frequency.

5.3. Accident Frequency Data among Machinists

In this subsection, the data concerning the frequency of accidents among 414 machinists originated from a study conducted towards the end of World War I. This study was performed by the Industrial Fatigue Research Board and was documented in a report published in [14]. The data covered a three-month period and were designed to assess the frequency of industrial accidents, particularly in environments with heavy machinery; they were used in [2].

Application and Concluding Results

The values for the mean, variance, and dispersion index were 0.4831, 1.0106, and 2.0919, respectively. Based on Table 12, the SCNBLD and CNBSLD models were considered the best fit for these data as they had the lowest AIC and BIC values.

Table 12. Accident frequency among 414 machinists: a statistical count over an undefined period.

Y	O.F ¹	PD	NBD	SCNBLD	CNBSLD
0	296	255.38	296.70	296.60	296.57
1	74	123.37	71.01	72.42	72.18
2	26	29.80	26.41	72.42	25.65
3	8	4.79	10.99	10.43	10.56
4	4	0.58	4.81	4.66	4.69
5	4	0.06	2.17	2.19	2.19
6	1	0.005	1.002	1.08	1.06
7	0	0.0003	0.46	0.54	0.52
8	1	0	0	0.27	0.27
Total	414	414	414	414	414
ML		$\theta = 0.4830$	$\theta = 0.5046$ $r = 0.4742$	$\theta = 0.6048$ $r = 0.6148$	$\theta = 0.5794$ $r = 0.8400$
AIC		855.82	768.06	767.63	767.68
BIC		859.85	776.11	775.68	775.73

¹ O.F = observed frequency.

6. Conclusions

This study presented a novel distribution by employing the concept of conflating probability distributions, specifically combining the NBD and the LD into a single distribution that captured their shared information with minimal loss. The newly introduced distribution was referred to as the conflation of negative binomial and logarithmic distributions (CNBLD). The CNBLD is capable of modeling positive count data from the LD, but it does not take into consideration zero values, unlike the NBD. In order to overcome this constraint, two novel modified models were introduced. The first model was the SCNBLD, which was obtained by shifting the CNBLD one position to the left. The second model was the combination of the negative binomial and shifted logarithmic distributions (CNBSLD), which merged a shifted logarithmic distribution with the NBD to incorporate the characteristics of both distributions. These two distributions provided increased flexibility and the capacity to model a wider range of data. An investigation was conducted to examine the valuable statistical characteristics of the CNBLD, SCNBLD, and CNBSLD. In addition, we studied the estimation of the CNBLD and CNBSLD parameters using the methods of MM and ML. The efficiency of the estimations given by these methods was assessed by simulation studies. The simulation results demonstrated that both the ML and MM estimates exhibited consistency. The efficacy of these models was assessed by testing them against diverse distributions using real data. The new distributions exhibited superior performance in accurately fitting the data when compared to other models. Finally, despite the fact that we used these new models on particular datasets, we were able to expand their suitability and employ them in various study domains, demonstrating their potential as a proficient substitute for modeling count data.

Author Contributions: Conceptualization, A.A.A. (Abdulhamid A. Alzaid) and A.A.A. (Anfal A. Alqefari); methodology, A.A.A. (Abdulhamid A. Alzaid), A.A.A. (Anfal A. Alqefari) and N.Q.; validation, A.A.A. (Abdulhamid A. Alzaid), A.A.A. (Anfal A. Alqefari), and N.Q.; writing—original draft preparation, A.A.A. (Anfal A. Alqefari); writing—review and editing, A.A.A. (Abdulhamid A. Alzaid), A.A.A. (Anfal A. Alqefari), and N.Q.; visualization, A.A.A. (Anfal A. Alqefari) and N.Q.; funding acquisition, N.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R376), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Data Availability Statement: We used of publicly available data.

Acknowledgments: The authors gratefully acknowledge Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R376), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia for the financial support for this project.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Johnson, N.L.; Kemp, A.W.; Kotz, S. *Univariate Discrete Distributions*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
2. Greenwood, M.; Yule, G.U. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *J. R. Stat. Soc.* **1920**, *83*, 255–279. [[CrossRef](#)]
3. Fisher, R.A. The effect of methods of ascertainment upon the estimation of frequencies. *Ann. Eugen.* **1934**, *6*, 13–25. [[CrossRef](#)]
4. Rao, C.R. On discrete distributions arising out of methods of ascertainment. *Sankhyā Indian J. Stat. Ser. A* **1965**, *27*, 311–324.
5. Barmalzan, G.; Saboori, H.; Kosari, S. A Modified Negative Binomial Distribution: Properties, Overdispersion and Underdispersion. *J. Stat. Theory Appl.* **2019**, *18*, 343–350. [[CrossRef](#)]
6. Hill, T. Conflations of probability distributions. *Trans. Am. Math. Soc.* **2011**, *363*, 3351–3372. [[CrossRef](#)]
7. Abramowitz, M.; A., S.I. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; US Government Printing Office: Washington, DC, USA, 1968.

8. Bagnoli, M.; Bergstrom, T. Log-concave probability and its applications. In *Rationality and Equilibrium: A Symposium in Honor of Marcel K. Richter*; Springer: New York, NY, USA, 2006; pp. 217–241.
9. Shaked, M.; Shanthikumar, J.G. *Stochastic Orders*; Series in Statistics; Springer: New York, NY, USA, 2007.
10. Shanker, R.; Hagos, F.; Sujatha, S.; Abrehe, Y. On zero-truncation of Poisson and Poisson-Lindley distributions and their applications. *Biom. Biostat. Int. J.* **2015**, *2*, 168–181. [[CrossRef](#)]
11. Finney, D.; Varley, G. An example of the truncated Poisson distribution. *Biometrics* **1955**, *11*, 387–394. [[CrossRef](#)]
12. Flynn, M.; Francis, L.A. More flexible GLMs zero-inflated models and hybrid models. *Casualty Actuar. Soc.* **2009**, *2009*, 148–224.
13. Aryuyuen, S.; Bodhisuwan, W.; Supapakorn, T. Zero inflated negative binomial-generalized exponential distribution and its applications. *Songklanakarin J. Sci. Technol.* **2014**, *36*, 483–491.
14. Greenwood, M.; Woods, H.M. *The Incidence of Industrial Accidents upon Individuals: With Special Reference to Multiple Accidents*; HM Stationery Office: London, UK, 1919.

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.