Brazilian Journal of Biometrics, **44**, 1–20 (2026)
doi:10.28951/bjb.v44i2.939

BRAZILIAN JOURNAL OF BIOMΣTRICS
ISSN:2764-5290

ARTICLE

Zero truncated hyper-negative Binomial distribution and its applications

C. Satheesh Kumar* and Sandeep S.

Department of Statistics, University of Kerala, Thiruvanathapuram, India
*Corresponding author. Email: drcsatheeshkumar@gmail.com

(Received: July 30, 2025; Revised: December 18, 2025; Accepted: February 13, 2026; Published: May 14, 2026)
Section Editor: Camilla Marques Barroso

Abstract

Count data represent the number of occurrences of an event within a fixed time or space and arise frequently in areas such as epidemiology, insurance, demography, and reliability analysis. In many practical situations, zero counts are structurally absent or unobserved, resulting in zero-truncated count data. Standard zero-truncated models may be inadequate when such data exhibit substantial over-dispersion. In this paper, we introduce and study a zero-truncated hyper-negative binomial distribution (ZTHNBD) to address this limitation. This work constitutes the first systematic investigation of the ZTHNBD. Several important statistical properties of the proposed distribution are derived, including the probability mass function, cumulative distribution function, mode, log-concavity, survival function, and hazard function. Recurrence relations for probabilities, raw moments, and factorial moments are also obtained. Parameter estimation is carried out using the method of maximum likelihood, and a generalized likelihood ratio test is developed to assess the significance of the additional parameter. The practical usefulness of the proposed model is demonstrated using multiple real-life data sets, where it provides an improved fit compared to existing zero-truncated models based on goodness of fit measures and information criteria. A brief simulation study is conducted to examine the finite sample performance of the maximum likelihood estimators.

Keywords: Count data modeling; Probability generating function; Maximum likelihood estimation; Negative binomial distribution; Survival function; Simulation.

1. Introduction

The hyper-negative binomial distribution (Yousry and Srivastava 1987) is an important discrete probability distribution that is used to model over-dispersed count data, where the variance is greater than the mean. It is a generalization of the negative binomial distribution and can be used in many fields such as biology, epidemiology, insurance, and quality control. However, in some real-world situations, the value zero is not observed or is not possible. For example, when recording the number of failures before a success in a system where at least one failure must happen, or when studying the number of hospital visits by patients who have already visited at least once. In such cases, using the standard hyper-negative binomial distribution may not be appropriate because it includes the probability of zero. The statistical properties and applications of zero-truncated models have been extensively studied in the literature. The foundational work on the truncated negative binomial distribution can be traced back to early contributions by (Sampford 1955) and (Brass 1958), who developed simplified estimation procedures and theoretical properties for truncated count distributions. Subsequently, the zero truncated negative binomial distribution (ZTNBD) has been extensively applied and extended in various applied contexts. In transportation studies, (Liu *et al.* 2013) employed zero-truncated negative binomial regression and quantile regression to analyze U.S. freight-train derailment severity, demonstrating its superiority over Poisson-based models in handling overdispersion. In the field of population studies, (Cruyff and Heijden 2008) proposed point and interval estimation of population size using a zero-truncated negative binomial regression model, highlighting its effectiveness in capture–recapture frameworks. In recent years, the relevance of ZTNBD has expanded significantly in epidemiology, particularly in modeling infectious disease transmission, (Zhao *et al.* 2021) utilized a zero-truncated negative binomial model to infer superspreading potential in COVID-19, emphasizing its capability in modeling heterogeneity in transmission counts. Similarly, methodological advancements include (Brass 1958), who proposed simplified fitting techniques, and later developments introducing flexible generalizations of the ZTNBD. Several authors have proposed generalized and compound forms of the ZTNBD to enhance model flexibility, (Sitho *et al.* 2021) introduced a zero-truncated negative binomial weighted Weibull distribution, while (Arrabal *et al.* 2014) applied ZTNBD-based models to nonlinear data structures. More recently, (Cheng and Huang 202) employed finite-mixture zero-truncated negative binomial models for predicting wet-road crashes, illustrating the importance of mixture structures in accident data. Further extensions include the zero-truncated negative binomial–Erlang distribution proposed by (Bodhisuwan *et al.* 2017) and the Lagrange-based generalization with associated regression models developed by (Mohanan *et al.* 2025). However, in applications where zero counts are absent or unobservable, the direct use of the standard hyper-negative binomial distribution is inappropriate. To deal with this, we introduce a zero-truncated version of the hyper negative binomial distribution named it (ZTHNBD), where the probability of zero is removed and the remaining probabilities are adjusted properly. This new distribution can better fit the data where zero counts are not possible. As stated by Yousry and Srivastava (1987), the probability mass function (p.m.f) of the HNBD is the following, for $x = 0, 1, 2, ..., \theta > 0$, $r > 0$ and $0 < q < 1$:

$$f(x) = P(Z = x) = \frac{\Delta_0 \ (r)_x}{(\theta)_x} q^x. \tag{1}$$

where

$$\Delta_0 = [{}_2F_1(1, r; \theta; q)]^{-1},$$

and

$${}_2F_1(a, b; c; u) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n u^n}{(c)_n n!}.$$

Here ${}_2F_1(a, b; c; u)$ is the Gauss hypergeometric function. For further details regarding Gauss hypergeometric function, refer Mathai and Haubold (2008). Here $(a)_n$ denote the Pochhammer symbol: $(a)_0$=1, $(a)_n = a(a + 1)...(a + n - 1)$, for n= 1,2,3,... and $(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)}$, for $a > 0$, where $\Gamma(.)$ denotes the gamma function. So, through the present paper we study the ZTHNBD in detail through exploring more properties of the distribution and thereby illustrating its usefulness to real life data sets.

In section 2, we provide some important properties of the ZTHNBD through deriving expressions for its cumulative distribution function, index of dispersion, mean, variance, mode, survival function and hazard function. In section 3, we discuss the maximum likelihood estimators of the parameters of the distribution along with generalized likelihood ratio test procedures for testing the additional parameter of the distribution. Further, in section 4, we provide certain data illustrations for highlighting the usefulness of the model and section 5 contains a brief simulation study. Section 6 provides a summary and conclusion of the present work.

2. THE ZTHNBD

In this section we present the definition of the ZTHNBD and derive some of its important properties.

Definition 2.1: A non negative valued random variable X is said to follow ZTHNBD if its p.m.f has the following form in which $r > 0$, $\theta > 0$, $0 < q < 1$ for x=1,2,... is given by

$$g(x) = \frac{g[X = x]}{g[X > 0]} = W_0^{-1} \frac{(r)_x\, q^x}{(\theta)_x}. \tag{2}$$

Where

$$W_i = {}_2F_1(1 + i, r + i; \theta + i; q) - 1$$

For i= 0,1,2,.. . Clearly, when $\theta = 1$, the p.m.f (2) reduces to the p.m.f of Zero truncated negative binomial distribution (Sampford 1955).

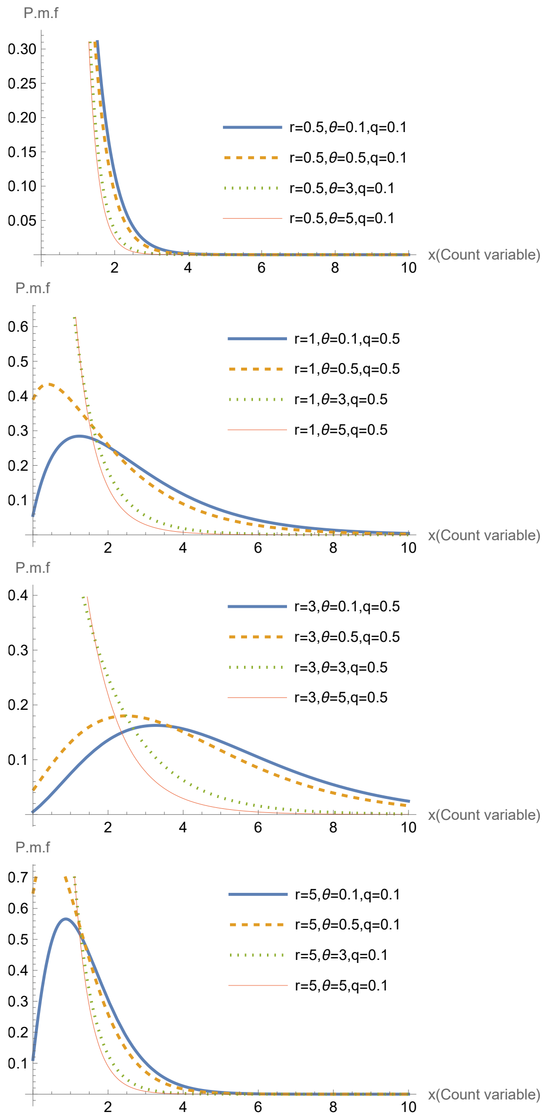

Figure 1. Illustration of the p.m.f of ZTHNBD for different values of r, θ and q.

Result 2.1 *The cumulative distribution function (c.d.f) of the ZTHNBD has the following form, for any* $x \in \Re = (-\infty, \infty)$.

$$G(x) = 1 - W_0^{-1}\, q^{x+1} \frac{(r)_{x+1}}{(\theta)_{x+1}}\, {}_2F_1(1, r+x+1; \theta+x+1; q) \tag{3}$$

Proof. By definition, the c.d.f of the ZTHNBD with p.m.f (2) is

$$\begin{aligned} G(x) &= P(X \leq x) \\ &= W_0^{-1} \sum_{k=0}^{x} \frac{(r)_k}{(\theta)_k} q^k \\ &= 1 - W_0^{-1} \sum_{k=x+1}^{\infty} \frac{(r)_k}{(\theta)_k} q^k \\ &= 1 - W_0^{-1}\, q^{x+1} \frac{(r)_{x+1}}{(\theta)_{x+1}} \sum_{k=0}^{\infty} \frac{(r+x+1)_k}{(\theta+x+1)_k}\, q^k, \end{aligned}$$

which leads to (3). ∎

Result 2.2 *The probability generating function (p.g.f) of the ZTHNBD is given by*

$$G(t) = W_0^{-1}\left[{}_2F_1(1, r; \theta; qt) - 1\right]. \tag{4}$$

Proof. By definition, the p.g.f of the ZTHNBD with p.m.f (2) is ∎

$$\begin{aligned} G(t) &= \sum_{x=1}^{\infty} P_x t^x \\ &= \frac{(1+W_0)}{W_0} \sum_{x=1}^{\infty} t^x \frac{(r)_x}{(\theta)_x} \frac{q^x}{1+W_0} \\ &= \frac{(1+W_0)}{W_0} \left[\sum_{x=0}^{\infty} \frac{(r)_x}{(\theta)_x} \frac{(qt)^x}{1+W_0}\right] - \frac{1}{W_0}, \end{aligned}$$

which on simplification gives (4)

Result 2.3 *An expression for factorial moments of the ZTHNBD is given by*

$$\mu_{[n]} = \frac{q^n\, n!\, (r)_n}{(\theta)_n} \delta_1, \tag{5}$$

where $\delta_i = \dfrac{W_i}{W_0}$; $i = 1, 2.$

Proof. The factorial moment generating function of the ZTHNBD with p.g.f (4) is

$$F(t) = G(1+t) = W_0^{-1}\left[{}_2F_1(1, r; \theta; qt) - 1\right] \tag{6}$$

On differentiating (6) n times with respect to t and putting t=1, we get (5). ∎

Result 2.4 *The Mean and Variance of ZTHNBD with p.g.f (4) is given by*

$$Mean = \frac{rq}{\theta}\delta_1 = \nu \tag{7}$$

and

$$Variance = \nu\left[\frac{2q(r+1)}{(\theta+1)}\frac{\delta_2}{\delta_1} + 1 - \frac{rq}{\theta}\right], \tag{8}$$

Proof follows from the fact that Mean= E(X)= $\frac{\partial G(t)}{\partial t}|_{t=1}$ and variance =$E[X(X-1)+E(X)]-[E(X)]^2$ $\frac{\partial^2 G(t)}{\partial t^2}|_{t=1}$.

Result 2.5 *The mode x_0 of the ZTHNBD is the following, in which $\rho = \frac{rq-\theta}{1-q}$, Z is the set of all non-negative integers and [a] denote the integer part of a, for any $a \in \Re$.*

$$x_0 = \begin{cases} [\rho+1] & if \;\; \rho \notin Z \\ \rho \; and \; \rho+1 & if \;\; \rho \in Z \end{cases} \tag{9}$$

Proof. The mode x_0 of the ZTHNBD with p.m.f $g(x)$ given in (2) is the value of x satisfying

$$g(x) \geq g(x-1) \; and \; g(x) \geq g(x+1).$$

Now,
$g(x) \geq$ g(x-1) implies

$$x \leq \frac{rq-\theta}{1-q}+1 \tag{10}$$

and
$g(x) \geq$ g(x+1) implies

$$x \geq \frac{rq-\theta}{1-q}. \tag{11}$$

Inequalities (10) and (11) gives the mode of the ZTHNBD as given in (9). ■

Result 2.6 *The ZTHNBD with p.m.f g(x) as given in (2) is log concave only when $r \geq \theta$.*

Proof. Proof is straightforward in the light of the inequality:

$$g^2(x+1) \geq g(x)\, g(x+2).$$

■

Result 2.7 *The index of dispersion (I.D) of the ZTHNBD are the following, in which ν is as given in (7).*

$$I.D = \frac{2q^2 r(r+1)}{\theta(\theta+1)\nu}\,\delta_2 - \nu + 1, \tag{12}$$

Proof is straightforward and hence omitted.

Result 2.8 *The ZTHNBD becomes over dispersed when $2\delta_1(\nu+\delta_1 q) > \delta_1^2\nu(\theta+1)$, under dispersed when $2\delta_1(\nu+\delta_1 q) < \delta_1^2\nu(\theta+1)$ and equi dispersed when $2\delta_1(\nu+\delta_1 q) = \delta_1^2\nu(\theta+1)$.*

Result 2.9 *For any $x \in \Re$, the survival function S(x) and hazard function h(x) of the ZTHNBD are respectively*

$$S(x) = W_0^{-1}\, q^{x+1} \frac{(r)_{x+1}}{(\theta)_{x+1}}\; {}_2F_1(1, r+x+1; \theta+x+1; q)$$

and

$$h(x) = \frac{\theta+x}{(r+x)q}\,[{}_2F_1(1, r+x+1; \theta+x+1; q)]^{-1}. \tag{13}$$

Proof. The proof follows from(2), (3) and the definition of the survival function $S(x) = g(X > x)$ and hazard function $h(x) = \frac{g(x)}{S(x)}$. ■

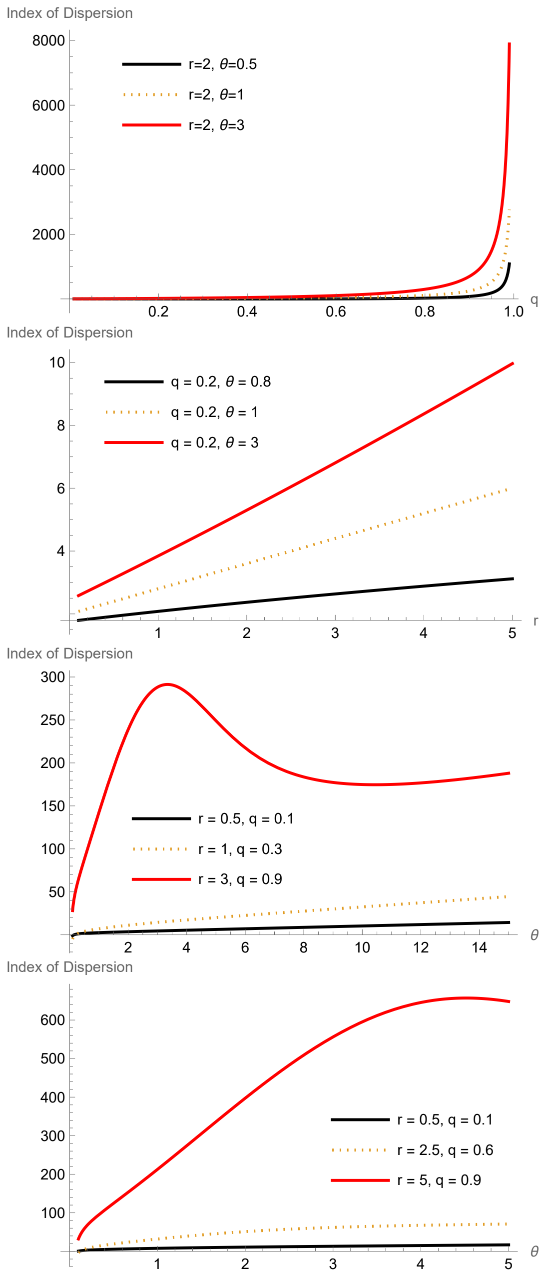

Figure 2. Illustration of the index of dispersion of ZTHNBD for different values of r, θ and q.

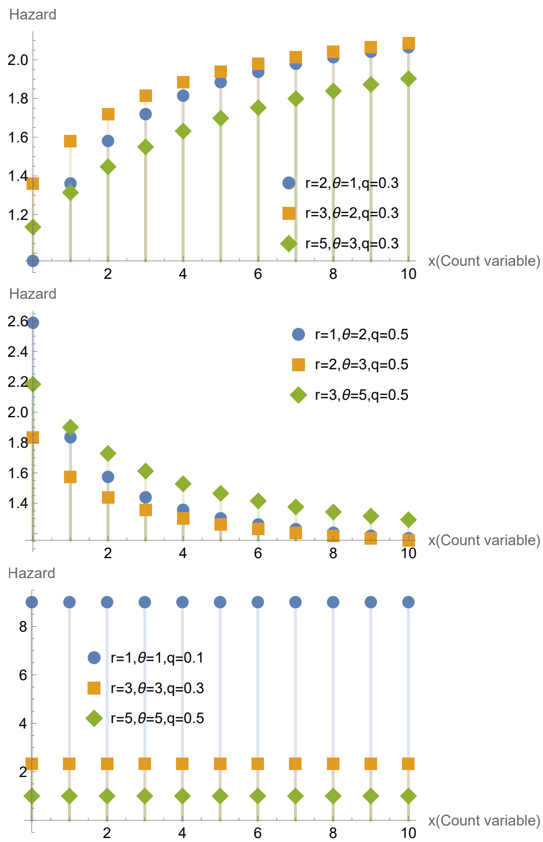

Figure 3. Illustration of the hazard of ZTHNBD for different values of r, θ and q.

Figure 2 illustrates the hazard rate function of the ZTHNBD for different values of r, θ and q. The Figure shows that the hazard rate function is non-decreasing when $r > \theta$, non-increasing when $r < \theta$ and constant when $r = \theta$, for all $0 < q < 1$.

Result 2.10 *The following is a simple recurrence relation for probabilities $g(x;\theta^*) = g(x)$ of the ZTHNBD with p.g.f (4) in which $\theta^* = (1,r;\theta)$*

$$g(x+1;\theta^*) = \frac{g(x;\theta^*+1)}{(x+1)}\,\nu\,\delta_1. \tag{14}$$

Proof.

From (4) we have

$$g(x) = \sum_{x=1}^{\infty} g(x;\theta^*)t^x = W_0^{-1}\,[{}_2F_1(1,r;\theta;qt)-1], \tag{15}$$

and

$$\sum_{x=0}^{\infty} g(x;\theta^*+1)t^x = \delta_1\ {}_2F_1(2,r+1;\theta+1;qt), \tag{16}$$

On differentiating (15) with respect to t, we get

$$\sum_{x=1}^{\infty} g(x;\theta^*)\ x\ t^{x-1} = W_0^{-1}\frac{rq}{\theta}\ {}_2F_1(2,r+1;\theta+1;qt). \tag{17}$$

Expressions (15) and (17) together lead to the following.

$$\sum_{x=1}^{\infty} g(x+1;\theta^*)\ (x+1)\ t^x = \delta_1\ \frac{rq}{\theta}\sum_{x=1}^{\infty} g(x;\theta^*+1)\ t^x \tag{18}$$

On equating coefficient of t^x on both side of (18) we get (14). ■

Result 2.11 *A recurrence relation for raw moments of the ZTHNBD is given by*

$$\mu_{x+1}(\theta^*) = \nu\left[1 + W_1\left\{\sum_{k-0}^{\infty}\binom{x}{k}\mu_{x-k}(\theta^*+1) - 1\right\}\right] \tag{19}$$

Proof. The characteristic function of ZTHNBD with p.g.f (4) has the following series representation.

$$\phi(t) = H(e^{it}) = W_0^{-1}\ \left[{}_2F_1(1,r;\theta;qe^{it})-1\right] = \sum_{x=1}^{\infty}\mu_x(\theta^*)\frac{(it)^x}{x!} \tag{20}$$

From (20) we have

$$\delta_1\ {}_2F_1(2,r+1;\theta+1;qe^{it}) = \sum_{x=1}^{\infty}\mu_x(\theta^*+1)\frac{(it)^x}{x!}. \tag{21}$$

Differentiate (20) with respect to t to get

$$\frac{rq}{\theta}\delta_1\ e^{it}\sum_{x=1}^{\infty}\mu_x(\theta^*+1)\frac{(it)^x}{x!} = \sum_{x=1}^{\infty}\mu_x(\theta^*)\ \frac{(it)^{x-1}}{(x-1)!}. \tag{22}$$

By using (20) and (21), equation (22) become

$$\sum_{x=1}^{\infty} \mu_{x+1}(\theta^*)\frac{(it)^x}{x!} = \frac{rq}{\theta}\delta_1 \sum_{k=0}^{\infty}\frac{(it)^k}{k!}\sum_{x=0}^{\infty}\mu_x(\theta^*+1)\frac{(it)^x}{x!} \tag{23}$$

We know that

$$\sum_{x=0}^{\infty}\sum_{s=0}^{\infty} B(x,s) = \sum_{x=0}^{\infty}\sum_{s=0}^{x} B(s, s-x) \tag{24}$$

In the light of(24) Equating the coefficients of $\frac{(it)^x}{x!}$ on both sides of (23), we get (20). ■

Result 2.12 *Recurrence relation for factorial moments of the ZTHNBD with p.g.f (4) is given by*

$$\mu_{[x+1]}(\theta^*) = \nu\ \mu_x(\theta^*+1). \tag{25}$$

Proof. The factorial moment generating function $F_X(t)$ of the ZTHNBD with p.g.f (4) has the following series representation.

$$F_X(t) = H(1+t) = W_0^{-1}\ {}_2F_1(1, r; \theta; q+qt) = \sum_{x=1}^{\infty}\mu_x(\theta^*)\ \frac{t^x}{x!} \tag{26}$$

From (26) we have

$$\sum_{x=1}^{\infty}\mu_x(\theta^*+1)\ \frac{t^x}{x!} = \delta_1\ {}_2F_1(2, r+1; \theta+1; q+qt). \tag{27}$$

On differentiating (26) with respect to t, we have

$$\sum_{x=1}^{\infty}\mu_x(\theta^*)\ \frac{t^{x-1}}{(x-1)!} = \frac{rq}{\theta}\delta_1\ \sum_{x=0}^{\infty}\mu_x(\theta^*+1)\ \frac{t^x}{x!}. \tag{28}$$

By using (26) and (27) we obtain the following from (28).

$$\sum_{x=1}^{\infty}\mu_{x+1}(\theta^*)\ \frac{t^x}{x!} = \frac{rq}{\theta}\delta_1\sum_{x=1}^{\infty}\mu_x(\theta^*+1)\ \frac{t^x}{x!} \tag{29}$$

Equating the coefficients of $\frac{t^x}{x!}$ in (29), we get (25). ■

3. Estimation and Testing

Here we consider the method of maximum likelihood for estimating the parameters r, θ and q of ZTHNBD. Let x be a non-negative integer a(x) be the observed frequency of x events and y be the highest value of x observed. Then the likelihood function of the sample is

$$L = \prod_{x=0}^{y}[g(x)]^{a(x)},$$

which implies

$$\log L = \sum_{x=0}^{y} a(x)\ log\ g(x).$$

That is,

$$l = \log L = \sum_{x=0}^{y} a(x) \left[log\, \Gamma(r+x) + log\Gamma(\theta) + x\, log\, q - log\, W_0^{-1} - log\Gamma(r) - log\Gamma(\theta+x) \right]. \tag{30}$$

Now, on differentiating (31) with respect to the parameters r, θ and q, we obtain the following likelihood equations, in which $\psi(w) = \frac{\partial}{\partial w} log\Gamma(w)$. That is,

$$\frac{\partial l}{\partial r} = 0$$

implies

$$\sum_{x=0}^{y} a(x) \left[\psi(r+x) - \triangle_0 \sum_{x=0}^{\infty} \frac{q^x}{(\theta)_x} \frac{\Gamma(r+x)}{\Gamma(r)} [\psi(r+x) - \psi(r)] - \psi(r) \right] = 0, \tag{31}$$

$$\frac{\partial l}{\partial \theta} = 0$$

implies

$$\sum_{x=0}^{y} a(x) \left[\psi(\theta) - \triangle_0 \sum_{x=0}^{\infty} q^x (r)_x \frac{\Gamma(\theta)}{\Gamma(\theta+x)} [\psi(\theta) - \psi(\theta+x)] - \psi(\theta+x) \right] = 0. \tag{32}$$

and

$$\frac{\partial l}{\partial q} = 0$$

implies

$$\sum_{x=0}^{y} a(x) \left[\frac{x}{q} - \frac{r}{\theta}\, \delta_1 \right] = 0. \tag{33}$$

Now, on solving the likelihood equations (31), (32) and (33) by using some mathematical softwares like MATHEMATICA, one can obtain the maximum likelihood estimators of the parameters of the ZTHNBD.

Hypothesis Testing

To test the significance of the parameter θ of the ZTHNBD, we adopt the generalized likelihood ratio test (GLRT) procedure. The null hypothesis is

$$H_0 : \theta = 1$$

against the alternative hypothesis

$$H_1 : \theta \neq 1.$$

Here, the test statistic is

$$-2ln\, \Psi = 2(\Lambda_1 - \Lambda_2), \tag{34}$$

in which $\Lambda_1 = lnL(\hat{\Theta}; x)$, where $\hat{\Theta}$ is the maximum likelihood estimator for $\lambda = (r, \theta, q)$ with no restrictions, and $\Lambda_2 = lnL(\hat{\Theta}^*; x)$, were $\hat{\Theta}^*$ is the maximum likelihood estimator for λ under the null hypothesis H_0. The test statistic defined in (34) is asymptotically distributed as chi-square with one degree of freedom(df). For further details regarding GLRT, see Rao(1973).

4. Applications

In this section, we consider certain real-life data sets for illustrating the methods discussed in Section 3. To illustrate the applicability of the proposed model, six real data sets from different fields are analyzed. The first data set relates to the number of households having at least one migrant according to the number of migrants, as reported by Singh and Yadav (1981). The second data set consists of the number of mothers in a rural area with at least two live births classified by the number of infant and child deaths, reported by Shanker (2015). The third data set represents the number of free-forming small groups according to group size, reported by Coleman and James (1961). The fourth data set is concerned with the number of literate mothers with at least one live birth classified by the number of infant deaths, also reported by Shanker (2015). The fifth data set, taken from Shanker (2015), provides additional count observations related to demographic events. The sixth data set consists of snowshoe hare counts captured over seven consecutive days, as reported by Keith and Meslow (1968). All six data sets involve positive count data and are therefore suitable for demonstrating the performance and flexibility of the proposed zero-truncated model.

We compared the fits of the zero truncated negative binomial distribution (ZTNBD) to that of the ZTHNBD in case of all the three data sets considered in the paper . We calculated the values of the chi-square statistic, AIC, BIC, and AICc for the purpose of comparing the model. The numerical results are shown in Tables 1, 2, 3, 4, 5 and 6. Comparing the ZTHNBD to other models based on the data sets considered here, one can see that the ZTHNBD provides a superior fit.

It can be shown that ZTNBD is not given best fit to the data sets, while the ZTHNBD only gives a better fit, based on the P-value and chi-square values. And also the values of information measures like AIC, BIC and AICc support the fact that the ZTHNBD can be considered as suitable model compared to other existing models considered in the paper.

Table 1. Observed frequencies and computed values of expected frequencies of the ZTNBD and the ZTHNBD by the method of maximum likelihood for the first dataset

X	Observed frequency	ZTNBD	ZTHNBD
1	375	375	374
2	143	144	143
3	49	46	48
4	12	13	12
5	2	4	2
6	1	0.9	2
7	1	0.1	1
Total	583	583	583
df		2	1
Log L		-560.82	-558.82
Estimates		$r = 3.02$	$r = 0.97$
		$p = 0.19$	$\theta = 0.16$
			$q = 0.22$
χ^2-value		9.39	0.52
P-value		0.0091	0.46
AIC		1125.64	1123.64
BIC		11127.16	1123.68
AICc		1125.66	1123.68

Table 2. Observed frequencies and computed values of expected frequencies of the ZTNBD and the ZTHNBD by the method of maximum likelihood for the second dataset

X	Observed frequency	ZTNBD	ZTHNBD
1	745	747	746
2	212	206	209
3	50	60	52
4	21	19	20
5	7	5	7
6	3	1	4
Total	1038	1038	1038
df		2	1
Log L		-873.49	-871.46
Estimates		$r = 0.60$	$r = 2.83$
		$p = 0.34$	$\theta = 4.56$
			$q = 0.39$
χ^2-value		6.85	0.42
P-value		0.03	0.51
AIC		1750.98	1748.92
BIC		1753.00	1751.95
AICc		1750.99	1748.94

Table 3. Observed frequencies and computed values of expected frequencies of the ZTNBD and the ZTHNBD by the method of maximum likelihood for the third dataset

X	Observed frequency	ZTNBD	ZTHNBD
1	1486	2100	1486
2	694	250	694
3	195	60	194
4	37	10	39
5	10	2.5	9
6	1	0.5	1
Total	2423	2423	2423
df		1	1
Log L		-6882.19	-2303
Estimates		$r = 24.57$	$r = 4.40$
		$p = 0.34$	$\theta = 0.0007$
			$q = 0.086$
χ^2-value		1367.7	0.218
P-value		0.0001	0.63
AIC		13768.4	4162.00
BIC		13771.1	4616.14
AICc		13768.4	4612.01

Table 4. Observed frequencies and computed values of expected frequencies of the ZTNBD and the ZTHNBD by the method of maximum likelihood for the fourth dataset

X	Observed frequency	ZTNBD	ZTHNBD
1	683	686	683
2	145	140	144
3	29	35	30
4	11	10	10
5	5	2	6
Total	873	873	873
df		1	1
Log L		-604.24	-602.23
Estimates		$r = 0.24$	$r = 0.76$
		$p = 0.33$	$\theta = 2.13$
			$q = 0.36$
χ^2-value		5.82	0.306
P-value		0.015	0.57
AIC		1212.48	1210.46
BIC		1214.36	1213.28
AICc		1212.49	1210.47

Table 5. Observed frequencies and computed values of expected frequencies of the ZTNBD and the ZTHNBD by the method of maximum likelihood for the fifth dataset

X	Observed frequency	ZTNBD	ZTHNBD
1	567	569	567
2	135	131	133
3	28	33	30
4	11	9	10
5	5	2	6
Total	746	746	746
df		1	1
Log L		-553.34	-551.34
Estimates		$r = 0.445$	$r = 1.13$
		$p = 0.319$	$\theta = 2.19$
			$q = 0.34$
χ^2-value		5.83	0.34
P-value		0.015	0.511
AIC		1110.68	1108.68
BIC		1112.42	1111.29
AICc		1110.70	1108.01

Table 6. Observed frequencies and computed values of expected frequencies of the ZTNBD and the ZTHNBD by the method of maximum likelihood for the sixth dataset

X	Observed frequency	ZTNBD	ZTHNBD
1	184	183	183
2	55	53	53
3	14	17	15
4	4	6	5
5	4	1	5
Total	260	260	260
df		1	1
Log L		-227.67	-225.67
Estimates		r = 0.708	r = 1.78
		p = 0.336	θ = 2.42
			q = 0.55
χ^2-value		10.27	0.54
P-value		0.001	0.45
AIC		459.34	457.34
BIC		460.16	458.57
AICc		459.38	457.43

Table 7. Value of the test statistic from the generalized likelihood ratio test

	$\ln L(\hat{\Theta}^*; x)$	$\ln L(\hat{\Theta}; x)$	Test statistic
Data set1	-560.82	-558.82	4
Data set 2	-873.49	-871.46	4.06
Data set 3	-6882.19	-2303.00	9158.3
Data set 4	-604.24	-602.23	4.02
Data set 5	-553.34	-551.34	4
Data set 6	-227.67	-225.67	4

We evaluated the test statistic given in (34) as shown in Table 7. The null hypothesis is rejected in each case since the critical value for the test at the 5% threshold of significance and one df is 3.84. Then one can observe that the parameter θ of ZTHNBD is significant in case of all the data sets considered here.

Table 8. Value of the index of the dispersion of the data sets

	Data ID	ZTNBD ID	ZTHNBD ID
Data 1	230.27	230.34	229.31
Data 2	489.47	491.16	489.93
Data 3	867.11	1732.6	866.98
Data 4	481.2	485.53	480.86
Data 5	384.28	388.75	383.56
Data 6	112.47	111.07	109.9

The Table 8 compares the index of dispersion (ID) values of the observed data with those obtained under the ZTNBD and ZTHNBD for six data sets. For all cases, the ID values from the ZTHNBD are closer to the empirical IDs than those from the ZTNBD. This indicates that the ZTHNBD more effectively captures the dispersion structure of zero-truncated count data, explaining its improved performance in applications.

5. Simulation

Since the maximum likelihood estimators of the parameters of the ZTHNBD are not in explicit form it is quite difficult for examining the performance of the estimators. So we conducted a brief simulation study in this section using simulated data sets. calculated the absolute bias and standard errors for each of the following set of parameters and provide in Table 8 corresponding to sample sizes 100, 300, and 500.

(i). $r = 2$, $\theta = 1$, $q = 0.05$ (Over-dispersed case)

(ii). $r = 1$, $\theta = 0.1$, $q = 0.05$ (Under-dispersed case)

For comparison purposes, a parallel simulation experiment is also carried out for the zero-truncated negative binomial distribution (ZTNBD). The corresponding bias and standard error results for the ZTNBD estimators are given in Table 9, allowing a direct comparison of the estimation efficiency under the two zero-truncated models.
The simulation study is performed under the following two parameter configurations

(i). $r = 2$, $p = 0.05$ (Over-dispersed case)

(ii). $r = 1$, $p = 0.05$ (Under-dispersed case)

Table 9. Bias and standard errors in the parenthesis of the estimators of the parameters r, θ and q of the ZTHNBD for the simulated data sets

Parameter set	Sample size	MLE		
		$\hat{r}$	$\hat{\theta}$	$\hat{q}$
(i)	n = 100	-1.99	0.14	0.12
		(3.99)	(2.01)	(0.02)
	n = 300	-2.00	-2.8	0.04
		(2.99)	(0.69)	(0.01)
	n = 500	-3.2	-0.82	0.003
		(2.54)	(0.53)	(0.004)
(ii)	n = 100	-0.99	1.33	0.4
		(0.98)	(4.98)	(0.40)
	n = 300	-1.29	0.96	0.09
		(0.87)	(1.09)	(0.01)
	n = 500	-1.51	-0.88	0.01
		(0.71)	(0.007)	(0.001)

Table 10. Bias and standard errors in the parenthesis of the estimators of the parameters r and p of the ZTNBD for the simulated data sets

Parameter set	Sample size	MLE	
		$\hat{r}$	$\hat{p}$
(i)	n = 100	-0.003	-0.003
		(0.024)	(0.0034)
	n = 300	-0.51	0.07
		(0.32)	(0.008)
	n = 500	0.20	-0.02
		(0.90)	(0.009)
(ii)	n = 100	3.53	-0.3
		(16.82)	(0.09)
	n = 300	4.4	-0.33
		(21.09)	(0.10)
	n = 500	2.9	-0.27
		(9.08)	(0.07)

From Table 8, it can be seen that as the sample size increases, both absolute bias and standard errors corresponding to the parameter sets are in decreasing order in both the cases. From Table 9 the results indicate that the estimation performance of the ZTNBD parameters is unstable across sample sizes. In particular, for both parameter sets, the absolute bias and standard errors do not exhibit a consistent decreasing trend as the sample size increases.Overall, the estimators under the ZTHNBD exhibit smaller bias and reduced standard errors relative to those of the ZTNBD, particularly for moderate and large sample sizes. This improvement can be attributed to the additional flexibility introduced by the extra parameter in the ZTHNBD, which allows the model to better capture over-dispersion in zero-truncated count data. These findings indicate that, while the ZTNBD performs reasonably well, the ZTHNBD provides more efficient parameter estimation and constitutes a superior alternative in practical applications involving zero-truncated over-dispersed data.

6. Conclusion

In this paper, we studied several properties of the ZTHNBD, including the derivative of its cumulative distribution function, mean, variance, mode, and recursive formulas. We estimated the parameters of the distribution using the maximum likelihood method. We also developed a generalized likelihood ratio test to check the importance of the extra parameter in the model. To show the practical utility of the proposed model is we applied it to six real-life data sets and compared its performance with the existing ZTNBD model. The results show that while the ZTNBD does not provide the best fit, the ZTHNBD gives a better fit based on Chi-square values and P-values. The information criteria values such as AIC, BIC, and AICC also suggest that the ZTHNBD is a more suitable model than the others discussed in the paper. Finally, we conducted a small simulation study to check how well the estimation methods perform.

Acknowledgments

The authors are highly thankful to the Editor in Chief, the Associate Editor and the anonymous Referees for their fruitful comments on an earlier version of the paper that greatly improved the quality and presentation of the paper.

Conflicts of Interest

The authors declare that they have no competing interests.

Author Contributions

Conceptualization: KUMAR, C. S.; SANDEEP, S. **Data curation**: KUMAR, C. S.; SANDEEP, S. **Formal analysis**: KUMAR, C. S.; SANDEEP, S. **Funding acquisition**: **Investigation**: KUMAR, C. S.; SANDEEP, S. **Methodology**: KUMAR, C. S.; SANDEEP, S. **Project administration**: KUMAR, C. S.; SANDEEP, S. **Software**: KUMAR, C. S.; SANDEEP, S.**Resources**: KUMAR, C. S.; SANDEEP, S. **Supervision**: KUMAR, C. S.; SANDEEP, S. **Validation**: KUMAR, C. S.; SANDEEP, S. **Visualization**: KUMAR, C. S.; SANDEEP, S. **Writing – original draft**: KUMAR, C. S.; SANDEEP, S. **Writing – review and editing**: KUMAR, C. S.; SANDEEP, S.

References

[1] Abramowitz, M. & Stegun, I.A . *Handbook of Mathematical Functions*. Dover, New York (1965).

[2] Anscomb, F.J. Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika* **37**, 358–382 (1950).

[3] Arrabal, C. T., dos Santos Silva, K. P., & Bandeira, L. N. Zero-truncated negative binomial applied to nonlinear data. *JP Journal of Biostatistics*, **11** (1), 55-67 (2014).

[4] Bodhisuwan, W., Pudprommarat, C., Bodhisuwan, R., & Saothayanun, L. Zero-truncated negative binomial-Erlang distribution. *In AIP Conference Proceedings* (Vol. **1905** (1), p. 050011). AIP Publishing LLC (2017).

[5] Brass, W. Simplified methods of fitting the truncated negative binomial distribution. *Biometrika*, **45** (1/2), 59-68 (1958).

[6] Chen, Y., & Huang, Z. Predicting Wet-Road Crashes Using the Finite-Mixture Zero-Truncated Negative Binomial Model. Journal of advanced transportation, 8828939 (2020).

[7] Coleman J.S & James J . The equilibrium size distribution of freely–forming groups. *Sociometry*, **24** (1):36–45 (1961).

[8] Cruyff, M. J., & Van Der Heijden, P. G. Point and interval estimation of the population size using a zero-truncated negative binomial regression model. *Journal of Mathematical Methods in Biosciences*, **50** (6), 1035-1050 (2008).

[9] Jain, G. C., & Consul, P. C. . A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics*, **21** (4), 501-513 (1971).

[10] Feller, W. *An Introduction to Probability Theory and its Applications*, Wiley, New York. 15 (1957).

[11] Keith L.B & Meslow E.C . Trap response by snowshoe hares. *Journal of Wildlife Management*, **32** ,795–801 (1968).

[12] Khatri, C.G. (1959). On certain properties of power series distributions. *Biometrika* **46**, 486-480 (1968).

[13] Kumar, C.S. A new class of discrete distributions, *Brazilian Journal of Probability and Statistics* **23**, 49-56 (2009)

[14] Lal D.N . A Demographic Sample Survey. Demographic Research Center, Department of Statistics, Patna University, Patna, India (1955) .

[15] Liu, X., Saat, M. R., Qin, X., & Barkan, C. P. Analysis of US freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. *Accident Analysis & Prevention*, **59**, 87-93 (2013).

[16] Mathai, A. M., & Haubold, H. J. . Special functions for Applied Scientists. *Springer* **4** (2008).

[17] Meegama S.A . Socio–economic determinants of infant and child mortality in Sri Lanka, an analysis of postwar experience. International Statistical Institute (World Fertility Survey), Netherland (1980) .

[18] Mishra A. Generalizations of Some Discrete Distributions. Unpublished Ph.D thesis, Patna University, Patna, India. (1979)

[19] Monisha, M., Maya, R., Irshad, M. R., Chesneau, C., & Shibu, D. S. A new generalization of the zero-truncated negative binomial distribution by a lagrange expansion with associated regression model and applications. *International Journal of Data Science and Analytics*, **20** (2), 637-651 (2025).

[20] Patil, G.P. Certain properties of the generalized power series distribution. *Annals of the Institute of Statistical Mathematics*. **14**, 179-182 (1962).

[21] Rao, C.R. . *Linear Statistical Inference and its Applications* . New York: Wiley (1973).

[22] Sampford, M. R . The truncated negative binomial distribution. *Biometrika*, **42** (1/2), 58-69 (1955).

[23] Shanker, R., Hagos, F., Sujatha, S., & Abrehe, Y. On zero-truncation of Poisson and Poisson-Lindley distributions and their applications. *Biometrics & Biostatistics International Journal*, **2** (6), 1-14 (2015).

[24] Singh S.N, Yadava K.N . Trends in rural out–migration at household level. Rural Demogr, **8** (1),53–61 (1981).

[25] Sitho, S., Denthet, S., & Nadeem, H. Zero truncated negative binomial weighted weibull distribution and its application. Lobachevskii Journal of Mathematics, **42** (13), 3241-3252 (2021).

[26] Slater, L.J . *Confluent Hypergeometric Functions*. Cambridge University Press, New York (1960).

[27] Yousry, M.A. & Srivastava, R.C. The hyper-negative binomial distribution. *Biometrical Journal* **29**, 875-884 (1987).

[28] Zhao, S., Shen, M., Musa, S. S., Guo, Z., Ran, J., Peng, Z., ...& Wang, M. H. Inferencing superspreading potential using zero-truncated negative binomial model: exemplification with COVID-19. *BMC Medical Research Methodology*, **21** (1), 30 (2021).