Study on Properties and Statistical Inference of Natural Discrete New Polynomial Exponential Distribution

A project submitted in fulfillment of the requirements for the degree of Master of Science in Statistics

by

Papu Mahanata
Roll No. - M.Sc- Sem (IV) Stat-06

Department of Statistics
Visva-Bharati, Santiniketan

Under the supervision of

Prof. Sudhansu S. Maiti
Visva-Bharati, Santiniketan

June, 2025

CERTIFICATION

This is to certify that the project work entitled **"Study on Properties and Statistical Inference of Natural Discrete New Polynomial Exponential Distribution"** has been prepared by **Papu Mahanata** under my supervision and guidance. The dissertation is his original work, completed after careful research and investigation. The work of the dissertation is of the standard expected of a candidate for the M.Sc. Degree (Statistics) Program, and I recommend that it be sent for evaluation.

Date:

Name and Signature of the Guide

Forwarded By:

Name and Signature of the Head of the Department

Acknowledgments

I would like to express my heartfelt gratitude to my project supervisor, Prof. Sudhansu S. Maiti, for his unwavering support, expert guidance, and continuous encouragement throughout the course of this project. His deep knowledge of the subject, insightful feedback, and valuable suggestions played a pivotal role in shaping the direction and quality of this dissertation.

His patience, availability, and meticulous attention to detail inspired me to approach each aspect of the research with critical thinking and dedication. His constructive criticism and persistent motivation enabled me to overcome challenges and grow both academically and personally.

I consider it a great privilege to have worked under his supervision. This project would not have reached its present form without his constant involvement, and I am truly indebted to him for his mentorship and confidence in my abilities.

Abstract

The project deals with the Natural Discrete New Polynomial Exponential distribution (NDNPED), a new discrete probability distribution with a mixture of geometric and negative binomial distributions. This distribution effectively models overdispersed count data, a common feature in various fields such as biology and reliability analysis. Also, different statistical properties of the distribution, such as moments, entropy, characterization, and the MGF, are discussed in this chapter. A detailed study on Natural Discrete xgamma (NDxgamma), a particular case of NDNPED, has been conducted. The uniqueness and existence of the MLE, as well as the computation of the bias and MSE, are used to evaluate the consistency of the estimator. Some biological data sets fit and compare the suggested distribution to other standard discrete distributions. The flexibility of the proposed distribution has been explored and contrasted with those of alternative discrete distributions.

Table of Contents

1 Introduction

Count data modeling is challenging in many domains, including health care, medicine, epidemiology, applied science, sociology, and agriculture. Because the life duration of a device cannot be calculated on a continuous scale, it is commonly assumed that the survival function is a function of a count random variable rather than a continuous time random variable. Hence, discrete distributions may be used to describe lifespan data when the output is discrete. Particularly for count data with over-dispersion when the variance is higher than the mean, the classic discrete distributions have limited application as models for dependability, failure duration, aggregate loss, etc.

Beghriche et. al (2022)[1] introduced the New Polynomial Exponential Distribution (NPED), a continuous distribution built as a mixture of gamma distributions, and the PDF is given by

$$f_T(t,\theta) = h(\theta)p(t,\theta)e^{-\theta t}, \qquad t,\ \theta > 0, \tag{1.1}$$

where, $h(\theta) = \frac{1}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}$, $p(t,\theta) = \sum_{k=0}^{r} a_{k,\theta} t^k$, $a_{k,\theta}$'s depend on θ and r is known non-negative integer.

The PDF can also be written as

$$\begin{aligned} f_T(t,\theta) &= h(\theta)\sum_{k=0}^{r} a_{k,\theta} t^k e^{-\theta t} \\ &= \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} f_{GA}(t; k+1, \theta)}{\sum_{k=0}^{r} a_{k,\ \theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}, \end{aligned} \tag{1.2}$$

where $f_{GA}(t; k+1, \theta)$ is the PDF of a gamma distribution with shape parameter $(k+1)$ and scale parameter θ. The distribution is a finite mixture of $(r+1)$ gamma distributions.

The CDF is given by

$$F_T(t,\theta) = 1 - \left(\frac{\sum_{k=0}^{r} \frac{a_{k,\theta}\Gamma(k+1)\Gamma(k+1,\theta t)}{\theta^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta} \frac{k!}{\theta^{k+1}}} \right); \quad t,\ \theta > 0, \tag{1.3}$$

where, $\Gamma(m,t) = \frac{1}{\Gamma(m)} \int_t^{\infty} e^{-u} u^{m-1} du$.

Inspired by its flexibility and the need for discrete analogues in practical applications, this work proposes a new discrete family, namely the Natural Discrete New Polynomial Exponential Distribution (NDNPED), which includes mixtures of geometric and negative binomial distributions. The resulting distribution is particularly suited for counting data with heavy tails, zero-inflated, or overdispersed, commonly observed in biological and reliability studies. A random variable X is said to have a Natural Discrete New Polynomial Exponential (NDNPED) distribution with success probability θ if its PMF is given by

$$p(w;\theta) = \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} f_{NB}(w; k+1,\ \theta)}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}; \qquad w = 0,\ 1,\ 2,\ ...,\ 0 < \theta < 1$$

$$= \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} \binom{w+k}{w} (1-\theta)^w \theta^{k+1}}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}, \tag{1.4}$$

where $f_{NB}(w; k+1, \theta)$ is the PMF of a negative binomial distribution with $(k+1)$ success and success probability θ, $a_{k,\theta}$'s are known non-negative constants depends on k and θ and r is a known positive integer. The distribution is the combinations of $(r+1)$ negative binomial distributions.
The PMF can also be expressed as

$$p(w; \theta) = h(\theta) p_1(w, \theta)(1-\theta)^w, \ 0 < \theta < 1, \ w = 0, \ 1, \ 2, \ \dots \tag{1.5}$$

where, $h(\theta) = \frac{1}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}$, $p_1(w, \theta) = \sum_{k=0}^{r} a_{k,\theta} \Gamma(k+1) \binom{w+k}{w}$.
The CDF of the random variable T is given by

$$F(t; \theta) = \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} I_{\bar{\theta}}(k+1, t+1)}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}, \ \bar{\theta} = 1 - \theta, \tag{1.6}$$

where, $I_p(m, \ n) = \frac{1}{B(m, \ n)} \int_0^p w^{m-1}(1-w)^{n-1} dw$.
If $a_{k,\theta} = a_k, \ \forall \ k$ and are non-negative constants, then (1.2) reduces to the NDOPPE distributions [2].
Remark 1.1. *It is to be noted that* $p(0; \theta) = \frac{\sum_{k=0}^{r} a_{k,\theta} \Gamma(k+1)}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}$, *and the other probabilities can be calculated recursively with the relationship* $p(w+1; \theta) = \frac{1-\theta}{1+w} \frac{\sum_{k=0}^{r} a_{k,\theta} (w+k+1)!}{\sum_{k=0}^{r} a_{k,\theta} (w+k)!} p(w; \theta)$.

The motivation for this development stems from the limitations of traditional discrete distributions in handling real-world datasets that exhibit complex features such as high variance, right skewness, or clustered zeros. The proposed NDNPED overcomes these limitations by introducing a more general framework with increased shape flexibility. A notable special case of this family, the NDxgamma distribution, retains analytical simplicity while providing strong empirical performance, as demonstrated through its superior fit to multiple biological datasets compared to classical models like Poisson, Zero-Inflated Poisson (ZIP), and Negative Binomial.

In addition to its modeling power, the NDNPED family possesses desirable theoretical properties such as closed-form expressions for moments, reliability functions, and entropy measures. Estimation can be efficiently handled through maximum likelihood and method of moments approaches, with simulation results supporting the consistency and precision of the estimators. These features make the NDNPED and its special cases practical and powerful tools for modern applied statistical modeling.

2 Moment and shape properties

The PGF is given by

$$P_X(s) = E(s^X)$$

$$= \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{(1-\bar{\theta}s)^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}, \qquad \mid s \mid < \frac{1}{1-\theta}.$$

The characteristic function is given by

$$\begin{aligned}\phi(t) &= E(e^{itX}) \\ &= \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{(1-\bar{\theta}e^{it})^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}}.\end{aligned}$$

The l^{th} factorial moment of the NDNPED is given by

$$\begin{aligned}\mu^{'}_{(l)} &= E[X(X-1)(X-2)...(X-l+1)] \\ &= h(\theta)\left(\frac{\bar{\theta}}{\theta}\right)^{l} \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(l+k+1)}{\theta^{k+1}}, \qquad l = 1,\ 2,\ ....\end{aligned}$$

The first four raw moments are given by

$$\begin{aligned}E(X) = \mu_1' &= \left(\frac{\bar{\theta}}{\theta}\right) h(\theta) \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+2)}{\theta^{k+1}}, \\ \mu_2' &= h(\theta)\left[\left(\frac{\bar{\theta}}{\theta}\right)^2 \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+3)}{\theta^{k+1}} + \frac{\bar{\theta}}{\theta} \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+2)}{\theta^{k+1}}\right],\end{aligned}$$

$$\begin{aligned}\mu_3' = h(\theta)\Bigg[&\left(\frac{\bar{\theta}}{\theta}\right)^3 \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+4)}{\theta^{k+1}} + 3\left(\frac{\bar{\theta}}{\theta}\right)^2 \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+3)}{\theta^{k+1}} + \\ &\frac{\bar{\theta}}{\theta} \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+2)}{\theta^{k+1}}\Bigg]\end{aligned}$$

$$\begin{aligned}\mu_4' = h(\theta)\Bigg[&\left(\frac{\bar{\theta}}{\theta}\right)^4 \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+5)}{\theta^{k+1}} + 6\left(\frac{\bar{\theta}}{\theta}\right)^3 \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+4)}{\theta^{k+1}} + \\ &7\left(\frac{\bar{\theta}}{\theta}\right)^2 \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+3)}{\theta^{k+1}} + \frac{\bar{\theta}}{\theta} \sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+2)}{\theta^{k+1}}\Bigg].\end{aligned}$$

The central moments can be calculated using the recursive relation between the raw moments and central moments.

The ID is given by

$$
\begin{aligned}
ID(X) &= \frac{\mu_2' - (\mu_1')^2}{E(X)} \\
&= \frac{\bar{\theta}}{\theta}\left[\frac{\sum_{k=0}^{r} a_{k,\theta}\frac{\Gamma(k+3)}{\theta^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta}\frac{\Gamma(k+2)}{\theta^{k+1}}} - \frac{\sum_{k=0}^{r} a_{k,\theta}\frac{\Gamma(k+2)}{\theta^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta}\frac{\Gamma(k+1)}{\theta^{k+1}}}\right] + 1.
\end{aligned}
$$

Ecologists used this ID to identify the spatial pattern for measuring clustering in Ecology.

Note that

1. As ID $\geqslant 1$, the NDNPED is over-dispersed $\forall\ \theta \in (0,1)$;
2. ID decreases monotonically in θ i.e. ID $\to 1$ as $\theta \to 1$
3. Also, ID $\to \infty$ as $\theta \to 0$.

So, Statisticians can use the NDNPED to look at count data that is overdispersed.

3 Reliability Properties

This section will discuss some reliability properties, like reliability function, failure rate, MRL, and stress-strength reliability.

3.1 Reliability function and failure rate

The reliability function of the NDNPED is given by

$$
\begin{aligned}
S(t) &= P(X \geq t) \\
&= \frac{\sum_{k=0}^{r} a_{k,\theta}\frac{\Gamma(k+1)}{\theta^{k+1}} I_\theta(t, k+1)}{\sum_{k=0}^{r} a_{k,\theta}\frac{\Gamma(k+1)}{\theta^{k+1}}},
\end{aligned}
$$

and its failure rate function is given by

$$
\begin{aligned}
\lambda(t) &= \frac{p(t;\theta)}{R(t;\theta)} \\
&= \frac{\sum_{k=0}^{r} a_{k,\theta}\frac{\Gamma(k+1)}{\theta^{k+1}}\binom{t+k}{t}(1-\theta)^t\theta^{k+1}}{\sum_{k=0}^{r} a_k\frac{\Gamma(k+1)}{\theta^{k+1}} I_\theta(t, k+1)}.
\end{aligned}
$$

Note that: $\frac{\sum_{k=0}^{r} a_{k,\theta}\Gamma(k+1)}{\sum_{k=0}^{r} a_{k,\theta}\frac{\Gamma(k+1)}{\theta^{k+1}}} < r(t;\theta) < \theta$ for $\forall t$.

3.2 Mean Residual Lifetime

The MRL is given by

$$\begin{aligned} MRL(t) &= \frac{\sum_{x=t+1}^{\infty} \bar{F}(x)}{\bar{F}(t)} \\ &= \frac{\sum_{x=t+1}^{\infty} \sum_{k=0}^{r} a_{k,\theta} \dfrac{\Gamma(k+1)}{\theta^{k+1}} I_{\theta}(x, k+1)}{\sum_{k=0}^{r} a_{k,\theta} \dfrac{\Gamma(k+1)}{\theta^{k+1}} I_{\theta}(t, k+1)}. \end{aligned}$$

Note that, $MRL(0) = \mu$.

3.3 Stress Strength Reliability

Let $Y \sim NDNPED(\theta_1)$ denote the strength of a system subject to stress X, $X \sim NDNPED(\theta_2)$, X and Y are independent of each other. The stress-strength (R) for NDxgamma distribution is given by

$$\begin{aligned} R &= \sum_{y=0}^{\infty} P(X \leq Y | Y = y) p_Y(y) \\ &= \sum_{y=0}^{\infty} F_X(y) p_Y(y) \\ &= \sum_{k=0}^{r} \sum_{l=0}^{r} a_{k,\theta_2} a_{l,\theta_1} \frac{\Gamma(k+1)}{\theta_2^{k+1}} \frac{\Gamma(l+1)}{\theta_1^{l+1}} \sum_{y=0}^{\infty} I_{\bar{\theta_2}}(k+1, y+1) \binom{y+l}{y} \theta_1^{l+1} (1-\theta_1)^y. \end{aligned}$$

4 Order Statistics, Lorenz Curve and Entropies

Let $X_i, i = 1, 2, .., n$ be random variables iid of the NDNPED $().Then, the CDF of minimum,X_{(1)} = Min(X_1, X_2, ..X_n)$ and maximum, $X_{(n)} = Max(X_1, X_2, ..X_n)$ are given by

$$\begin{aligned} F_{X_{(1)}}(x) &= 1 - \left[\frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} I_{\theta}(x, k+1)}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}} \right]^n, \\ F_{X_{(n)}}(x) &= \left[\frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} I_{\bar{\theta}}(k+1, x+1)}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}} \right]^n. \end{aligned}$$

where $X_{(1)}$ is the lifetime of an n-component series system, and $X_{(n)}$ for an n-component parallel system, which plays an important role in reliability analysis.
The Lorenz curve for a positive random variable X is defined as

$$L(F(x)) = \frac{E(X | X \leq x) F(x)}{E(X)}.$$

For the NDNPED distribution, we get

$$E(X|X \leq x)F(x) = \frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\bar{\theta}(k+1)\Gamma(k+1)}{\theta^{k+2}}}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}} \left(\frac{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} I_{\bar{\theta}}(k+1, x+1)}{\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}}} \right).$$

The Rényi entropy is defined as

$$J(\gamma) = \frac{1}{1-\gamma} log \left[\int f^{\gamma}(x) dx \right],$$

. For the NDNPED the Renyi entropy is given by

$$J(\gamma) = \frac{1}{1-\gamma} log \left[\frac{\int p^{\gamma}(x,\theta)(1-\theta)^{\gamma x} dx}{\left(\sum_{k=0}^{r} a_{k,\theta} \frac{\Gamma(k+1)}{\theta^{k+1}} \right)^{\gamma}} \right]$$

where

$$p(x,\theta) = \sum_{k=0}^{r} a_{k,\theta} \Gamma(k+1) \binom{x+k}{x}$$

.

5 The NDxgamma Distribution: A particular case of NDNPED

Definition 1. *Let X be a nonnegative random variable that follows the NDxgamma distribution which is a finite mixture of $Geo(\theta)$ and $NB(3,\theta)$ with mixing probabilities $\frac{\theta}{1+\theta}$ and $\frac{1}{1+\theta}$, respectively. The PMF and CDF of NDxgamma distribution is given by*

$$\begin{aligned} p(x,\theta) &= \frac{\theta}{1+\theta} Geo(x,\theta) + \frac{1}{1+\theta} NB(x,3,\theta), \qquad x = 0,1,2,..,\infty; \qquad \theta \in (0,1) \\ &= \frac{\theta^2}{1+\theta}(1-\theta)^x \left(1 + (x+2)(x+1)\frac{\theta}{2} \right), \end{aligned} \tag{5.7}$$

and

$$F(x,\theta) = 1 - \frac{(1-\theta)^{x+1}(\theta(x+2)(\theta+\theta x+2)+2)}{2(\theta+1)}. \tag{5.8}$$

A single parameter discrete NDxgamma distribution is created to increase the flexibility of the modeling.

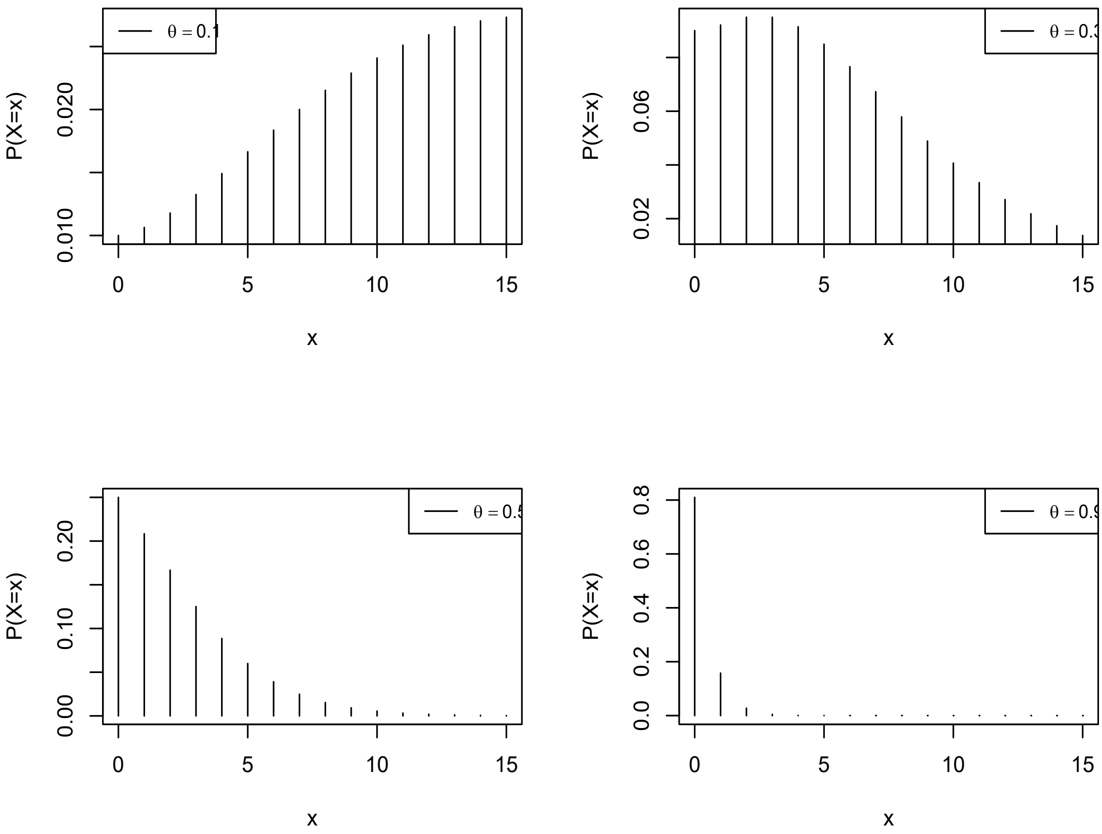

Fig. 1 The PMF of NDxgamma for θ=0.1, 0.3, 0.5, 0.9

Putting the value $r = 2,\ a_{0,\theta} = 1,\ a_{1,\theta} = 0,\ a_{2,\theta} = \frac{\theta}{2}$, we get m^{th} the factorial moment for the NDxgamma distribution.

$$\begin{aligned} \mu'_{(m)} &= E\left(X(X-1)..(X-m+1)\right) \\ &= \left(\frac{1-\theta}{\theta}\right)^m \left(\frac{m!}{\theta+1}\right)\left(\theta + \frac{(m+1)(m+2)}{2}\right). \end{aligned} \tag{5.9}$$

Similarly, the m^{th} central moment of NDxgamma is given by

$$\mu_m = \frac{\theta^3(\Phi(1-\theta,-m-2,0) + 3\Phi(1-\theta,-m-1,0))}{2(\theta+1)} + \theta^2\Phi(1-\theta,-m,0), \tag{5.10}$$

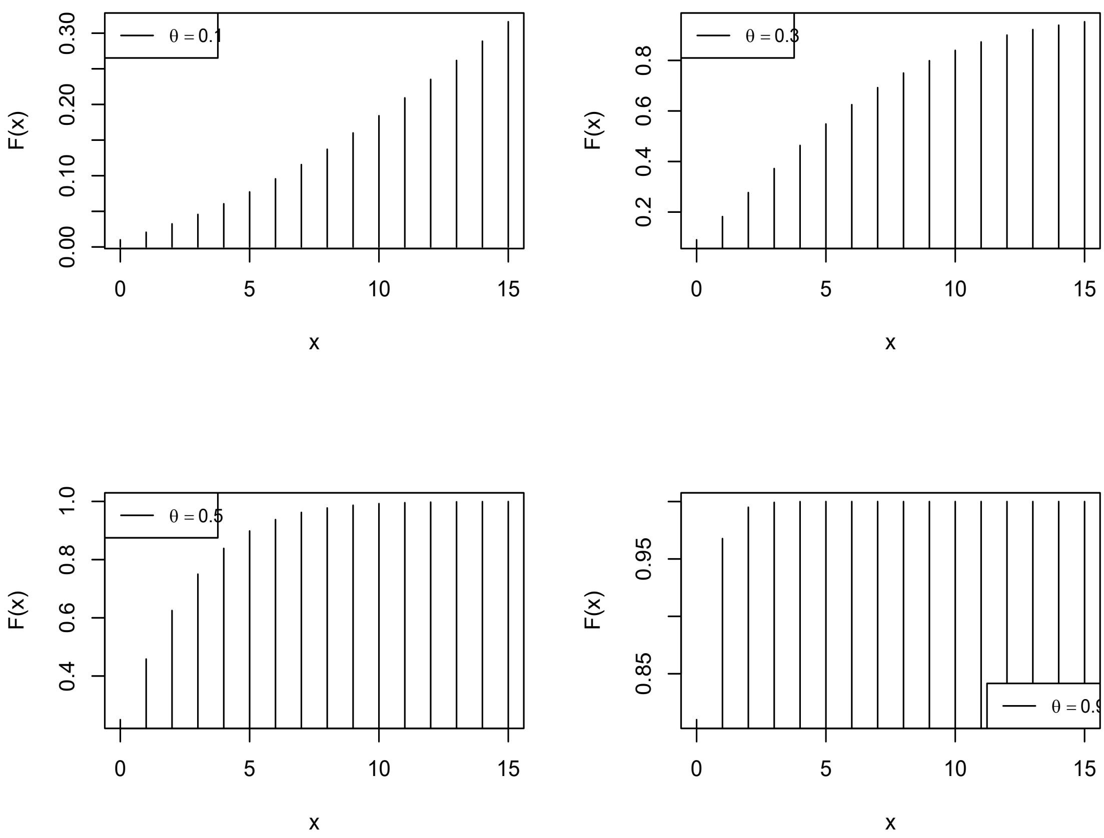

Fig. 2 The CDF of NDxgamma for θ=0.1, 0.3, 0.5, 0.9

where, $\Phi(z,s,a) = \sum_{k=0}^{\infty} \dfrac{z^k}{(a+k)^s}$ is the Lerchi transcendent.
Putting m = 1 and m = 2 in equation (5.9) and (5.10), respectively, we get mean and variance.

$$\mu = \frac{-\theta^2 - 2\theta + 3}{\theta(\theta+1)},$$

and

$$\mu_2 = \frac{(\theta-3)(\theta-1)(3\theta+1)}{\theta^2(\theta+1)^2},$$

respectively.

From Figure 3, we see that the mean and variance of the NDxgamma distribution decrease as the θ increases. In addition, the distribution is leptokurtic and positively skewed for different values of θ. The median can be calculated using the formula

$$\begin{aligned}\sum_{x=0}^{m}\frac{\theta^2}{1+\theta}(1-\theta)^x\left(1+(x+2)(x+1)\frac{\theta}{2}\right) &= \frac{1}{2}\\ or,\ \frac{(1-\theta)^{m+1}(\theta(m+2)(\theta+\theta m+2)+2)}{2(\theta+1)} &= \frac{1}{2}.\end{aligned}$$

Solving the above equation, we get the median for different values of θ shown in table 1.

Table 1 The value of median for different θ

median	1	2	3	4	5	6	7	8
θ	0.5271	0.4277	0.3626	0.3159	0.2805	0.2525	0.2299	0.2111
median	9	10	11	12	13	14	15	16
θ	0.1952	0.1816	0.1698	0.1594	0.1503	0.1422	0.1349	0.1283

In particular, the ID of NDxgamma distribution is given in the table 2.

$$ID=\frac{(\theta-3)(\theta-1)(3\theta+1)}{\theta(\theta+1)(3-\theta^2-2\theta)}$$

The survival function is given by

$$\begin{aligned}S(x) &= P(X\geq x)\\ &= \frac{(1-\theta)^{x+1}(\theta(x+2)(\theta+\theta x+2)+2)}{2(\theta+1)}\end{aligned}$$

Also, the HRF reduces to

$$\lambda(x) = \frac{p(x)}{S(x)}$$

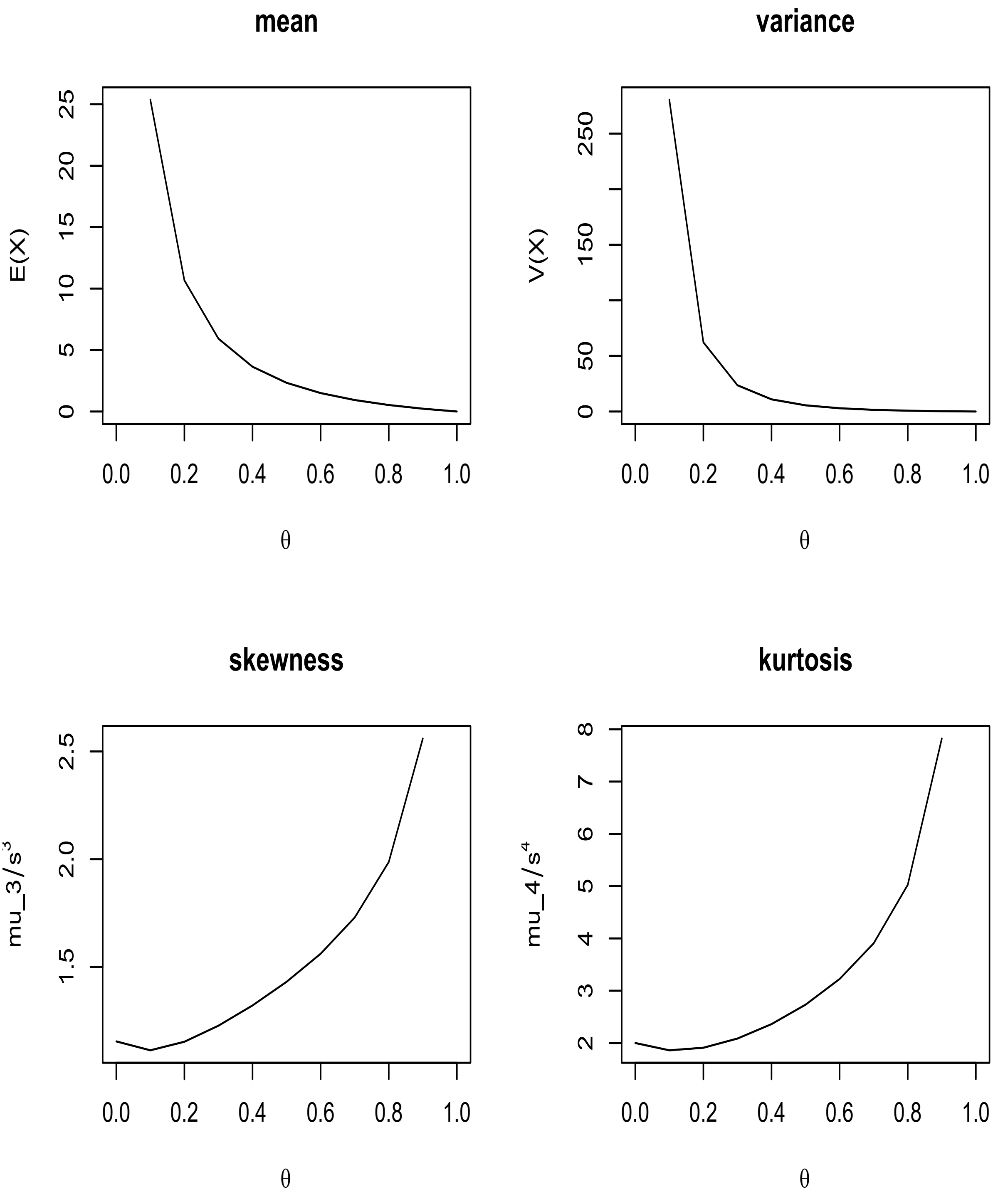

Fig. 3 Mean, variance, skewness, and Kurtosis plots for different values of θ

Table 2 The values of ID for different θ

θ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ID	11.056	5.833	3.986	3	2.380	1.944	1.619	1.366

$$= \frac{\theta^2\left[2+(x+1)(x+2)\theta\right]}{(1-\theta)\left[\theta(x+2)(\theta+\theta x+2)+2\right]}.$$

Also, $\lim_{x\to\infty}\lambda(x) = \dfrac{\theta}{1-\theta}$ which shows increasing hazard rate function.

So, the MRL function of NDxgamma distribution is given by placing suitable values of r and $a_{k,\theta}$

$$MRL(x,\theta) = \frac{(1-\theta)\left(2\theta^2+6\theta+\theta^2x^2+3\theta^2x+4\theta x+6\right)}{\theta(\theta(x+2)(\theta+\theta x+2)+2)}$$

5.1 Log-Concavity

Definition 2. *A discrete random variable X with PMF $P(X=x)$ is said to be increasing failure rate (IFR)/ unimodal if $P(X=x)$ is log-concave (see, Keilson and Gerber (1971) [3]), i.e., if*

$$P(X=x)P(X=x+2) \le P(X=x+1)^2, \qquad x=0,1,2,..,.$$

.

Theorem 5.1. *The PMF of the NDxgamma distribution is log-concave for all $\theta \in$ (0, 1).*

Proof. From equation (5.7), We get,

$$\begin{aligned}
p_{x+1}^2 &= \frac{\theta^4}{(1-\theta)^2}(1+\theta)^{2x+2}\left[1+(x+2)(x+3)\frac{\theta}{2}\right]^2\\
p_xp_{x+2} &= \frac{\theta^4}{(1-\theta)^2}(1-\theta)^{2x+2}\left[1+(x+1)(x+2)\frac{\theta}{2}\right]\left[1+(x+3)(x+4)\frac{\theta}{2}\right]\\
\therefore\ p_{x+1}^2-p_xp_{x+2} &= \frac{\theta^5}{2(1-\theta)^2}(1-\theta)^{2x+2}\left[\theta(x+2)(x+3)-2\right]>0
\end{aligned}$$

$\forall$ x and $\forall\theta \in (0,1)$.
So, NDxgamma is unimodal.

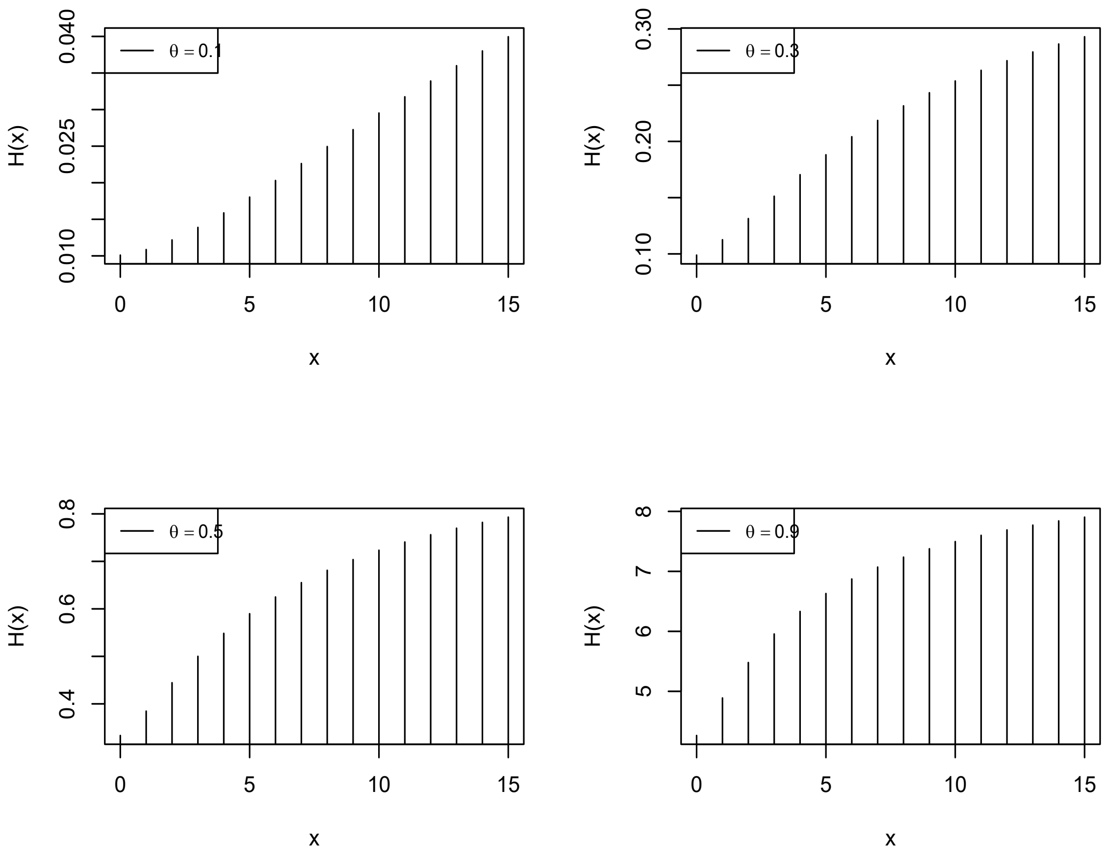

Fig. 4 The hazard rate plot of NDxgamma for θ=0.1, 0.3, 0.5, 0.9

$\square$

Corollary 5.1. *If $X \sim NDxgamma(\theta)$, then it has an IFR and DMRL.*

As we find in section 5.1 that the NDxgamma distribution is log-concave, so, according to Gupta et. al (1997) [4], the NDxgamma distribution has an IFR property. According to Kemp (2004) [5], chain is as follows: $IFR \Rightarrow IFRA \Rightarrow NBU \Rightarrow NBUE \Rightarrow DMRL$ i.e. the distribution has

1. increasing failure rate (IFR)
2. increasing failure rate average (IFRA)
3. new better than used (NBU)
4. new better than used in expectation (NBUE)

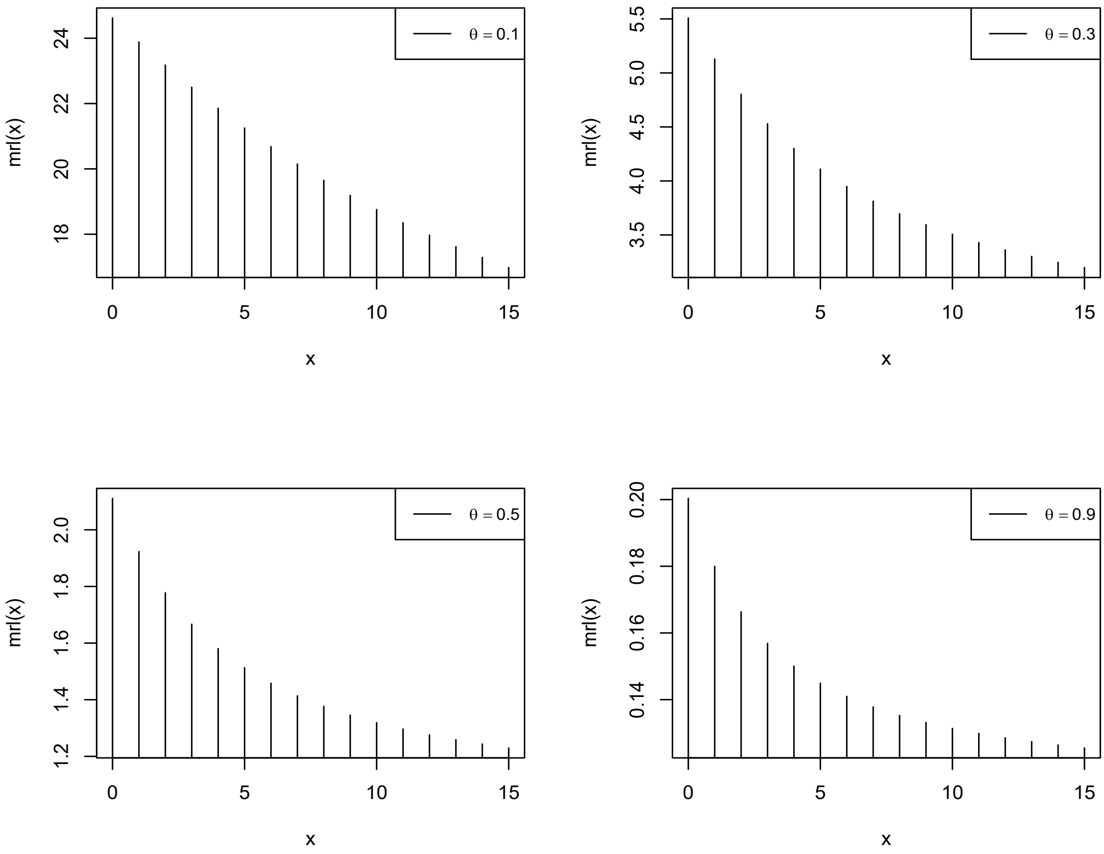

Fig. 5 The MRL plot of NDxgamma for θ=0.1, 0.3, 0.5, 0.9

5. decreasing mean residual lifetime (DMRL)

6 Stochastic ordering

Theorem 6.1. *Let $X \sim NDxgamma(\theta_1)$ and $Y \sim NDxgamma(\theta_2)$. Then $X \leq_{lr} Y \Rightarrow X \leq_{hr} Y \Rightarrow X \leq_{st} Y \Rightarrow X \leq_{mrl} Y \leq_{lr} Y \Rightarrow X \leq_{rh} Y, \;\; \forall\; \theta_1 < \theta_2$.*
Proof 6.1. *We have*

$$L(x) = \frac{p_X(x)}{p_Y(x)} = \frac{(1-\theta_1)[1+(x+2)(x+3)\theta_1/2]}{(1-\theta_2)[1+(x+2)(x+3)\theta_2/2]}$$

Clearly, we can see that $L(x+1) \leq L(x), \;\forall \theta_1 < \theta_2$. Hence, the theorem follows.

7 Estimation and Simulation

In this section, we estimate the parameter θ by the MLE and MME methods.

7.1 Maximum likelihood estimation

Let $X_1, X_2, .., X_n$ be the observed values from the NDxgamma distribution with parameter θ, then the likelihood and log-likelihood function are given by

$$L(\theta) = \prod_{i=1}^{n} \frac{\theta^2}{1+\theta}(1-\theta)^{x_i} \left(1 + (x_i+2)(x_i+1)\frac{\theta}{2}\right)$$

and after taking "*log*" on both sides, we get

$$l(\theta) = n \log\left(\frac{\theta^2}{1+\theta}\right) + \sum_{i=1}^{n} \log(1-\theta) + \sum_{i=1}^{n} [1 + (1+x_i)(2+x_i)\theta/2]$$

respectively.
The following equation can be solved numerically to get the MLE of the parameter θ.

$$\frac{dl}{=} \frac{(2+\theta)n}{\theta(1+\theta)} - \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} \left[\frac{(1+x_i)(2+x_i)}{2+(1+x_i)(2+x_i)}\right] = 0$$

7.2 Method of Moment Estimation

Let $X_1, X_2, .., X_n$ be the observed values from the NDxgamma distribution with parameter θ, then the moment estimate (MME) of the parameter θ can be obtained by solving the following equation:

$$\frac{-\theta^2 - 2\theta + 3}{\theta(\theta+1)} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

So, the value of θ are:

$$\left\{\left\{\hat{\theta} \to \frac{-\sqrt{\bar{X}^2+16\bar{X}+16}-\bar{X}-2}{2(\bar{X}+1)}\right\}, \left\{\hat{\theta} \to \frac{\sqrt{\bar{X}^2+16\bar{X}+16}-\bar{X}-2}{2(\bar{X}+1)}\right\}\right\}$$

As, $\theta > 0$ so the required estimate of θ is

$$\hat{\theta} = \frac{\sqrt{\bar{X}^2 + 16\bar{X} + 16} - \bar{X} - 2}{2(\bar{X}+1)}$$

8 Simulation: Bias and MSE

Let $X_1,\ X_2,\ \ldots,\ X_n$ be random sample drawn from NDNPED Distribution.

Put $F(x) = u$, and then proceed the following in R software as follows:

1. Generate $U_i \sim Uniform(0,1), i = 1(1)n$.
2. If $\frac{\sum_{k=0}^{j-1} a_{k,\theta} \frac{k!}{\theta^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta} \frac{k!}{\theta^{k+1}}} < U_i \leq \frac{\sum_{k=0}^{j} a_{k,\theta} \frac{k!}{\theta^{k+1}}}{\sum_{k=0}^{r} a_{k,\theta} \frac{k!}{\theta^{k+1}}}$, $j = 2, ..., r$, then set $X_i = V_i$, where $V_i \sim NB(j+1, \theta)$ and
3. If $U_i \leq \frac{a_{0,\theta}(\frac{1}{\theta})}{\sum_{k=0}^{r} a_{k,\theta} \frac{k!}{\theta^{k+1}}}$, then set $X_i = W_i$, where $W_i \sim Geo(\theta)$.

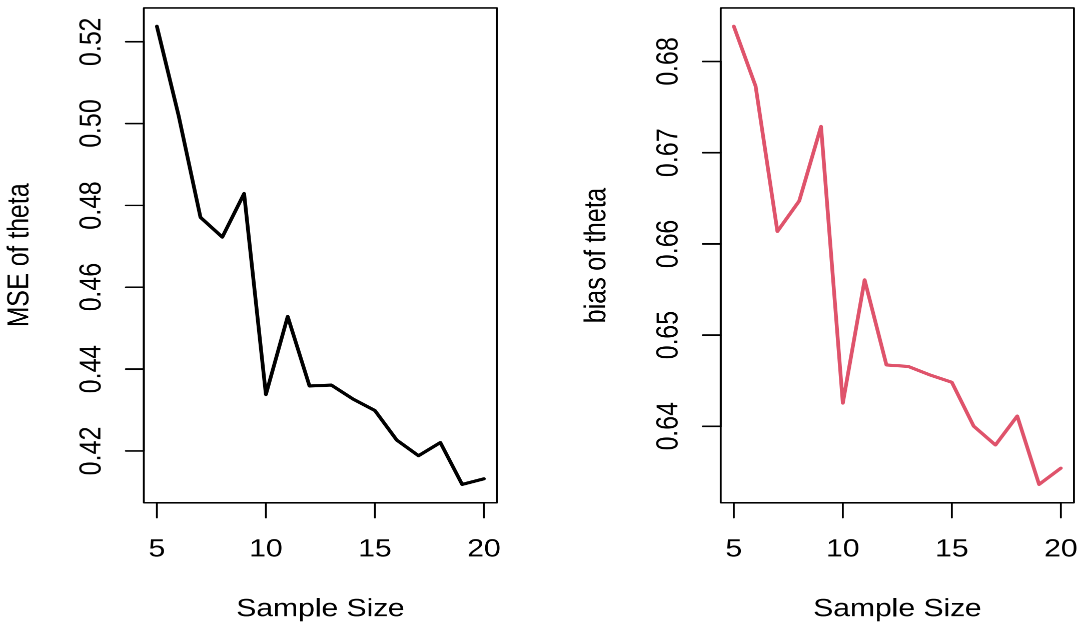

Fig. 6 Mse and Bias of NDxgamma for θ=0.4

The above operation is carried out with $1,000(= N)$ repetitions choosing $a_{0,\theta} = 1$, $a_{1,\theta} = 0$, $a_{2,\theta} = \dfrac{1}{\theta}$ that indicates the NDxgamma distribution. For generation of random observations, take $\theta = 0.4$. The following two measures are also computed:

- Bias of the simulated estimates $\hat{\theta}$, for i=1, 2, 3, ...,N: $\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)$,
- Mean Square Error (MSE) of the simulated estimates $\hat{\theta}$, for i=1, 2, 3, ...,N: $\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2$.

Table 3 **Summary of the Datasets**

Dataset	Mean	Median	Mode	Std. dev	Min	Max	Q_1	Q_3
I	107.83	31.5	0	173	2	447	8	109.5
II	18.75	9.5	0	23.95	1	70	2.75	22.25
III	54	25	0	72.43	1	188	5	71.25

As n increases, it is found from Figure 8 that both Bias and MSE decrease. The MLE is positively skewed and gets closer to zero as n increases. This shows how precise the MLE is. We found that the MSE is moving down as n increases to ensure that the estimate is valid.

9 Goodness of fit

We chose three examples of actual datasets observed in the biological sciences, such as insurance claim data and ecological data, which consists of an excess number of zeroes. The first dataset presented in Table 7.10.1 is the accidents of 647 women who were working at H.E. Shells for 5 weeks. The datasets are available in Consul and Jain (1973)[6]. The second dataset is taken from Garman (1951) [7], which comprises the number of European red mites on apple leaves, which is overdispersed, and the details are shown in Table 3. The third data set is taken from McGuire et al. (1957) [8], which contains information on the number of European corn borers. A full summary of the datasets is analyzed and presented below.

For each data set, we have displayed the fitted probabilities for four different probability distributions, namely Possion, Zero Inflated Poisson (VIP), Dxgamma - I, and Dxgamma - II, including the proposed NDxgamma distribution, and also compared them using the sense of negative log-likelihood and AIC as ZIP has more than one parameter. A summary of the findings is presented in Table 4, 5 and 6 for each of the three datasets. There is a significant percentage of zero values in all data sets, which is noticeable. For each of the datasets, we have calculated the negative likelihood, AIC, and χ^2 statistic. In terms of AIC and χ^2 statistic, it is evident that the NDxgamma distribution gives a better fit than the other distributions. The entire data analysis was carried out using the R software.

Table 4 Frequency table of the accidents of 647 women who are working on H. E. Shells for 5 weeks

X	**Observed frequencies**	**Fitted Poisson(θ)**	**Fitted ZIP(θ, π)**	**Fitted dxgamma-I**	**Fitted dxgamma-II**	**Fitted NDxgamma**
0	447	406.33	427.95	421.02	420.69	433.22
1	132	189.01	157.96	165.8	166.28	149.58
2	42	43.95	45.56	46.4	46.4	46.49
3	21	6.81	17.51	10.9	10.81	13.12
4	3	0.79	1.65	2.3	2.26	3.44
≥ 5	2	0.07	0.29	0.45	0.44	0.85
$\hat{\theta}$	-	0.4651	0.8819	0.8596	0.8618	0.8182
$\hat{\pi}$	-	-	0.4725	-	-	-
Negative log-likelihood	-	617.1843	593.2722	599.6271	599.8784	593.9928
AIC	-	1236.3686	1190.5444	1201.2542	1201.7568	**1189.9856**
χ^2 **statistic**	-	110.31	6.42	23.81,	24.50	**6.27**

Table 5 Frequency table of the number of European red mites on apple leaves which is over dispersed

X	**Observed frequencies**	**Fitted Poisson(θ)**	**Fitted ZIP(θ, π)**	**Fitted dxgamma-I**	**Fitted dxgamma-II**	**Fitted NDxgamma**
0	70	47.65	60.26	59.54	59.41	64.75
1	38	54.64	32.87	42.92	42.56	39.82
2	17	31.32	29.45	25.14	25.32	22.71
3	10	11.97	17.58	12.63	12.82	11.93
4	9	3.43	7.87	5.73	5.83	5.86
5	3	0.78	2.82	2.42	2.46	2.74
6	2	0.15	0.84	0.98	0.99	1.23
7	1	0.02	0.21	0.38	0.38	0.54
$\hat{\theta}$	-	1.1465	1.7916	0.701	0.705	0.657
$\hat{\pi}$	-	-	0.359	-	-	-
Negative log-likelihood	-	242.8099	226.4332	225.1087	225.1177	223.0043
AIC	-	487.6198	456.8664	452.2174	452.2354	**448.0086**
χ^2 **statistic**	-	108.539	14.441	9.646	9.599	**4.823**

Table 6 Frequency table of the amount of European corn borers

X	**Observed frequencies**	**Fitted Poisson(θ)**	**Fitted ZIP(θ, π)**	**Fitted dxgamma-I**	**Fitted dxgamma-II**	**Fitted NDxgamma**
0	188	169.46	179.85	182.54	182.25	190.66
1	83	109.83	109.04	92.84	92.68	82.95
2	36	35.59	24.99	34.27	34.38	32.79
3	14	7.68	7.85	10.71	10.69	11.81
4	2	1.24	1.84	3.02	0.77	1.26
5	1	0.16	0.34	0.8	0.77	1.26
$\hat{\theta}$	-	0.6481	0.9423	0.8135	0.8167	0.7671
$\hat{\pi}$	-	-	0.3122	-	-	-
Negative log-likelihood	-	362.2451	355.0395	355.9639	355.9881	355.9439
AIC	-	726.4902	714.079	713.9278	713.9762	**713.8878**
χ^2 **statistic**	-	18.66	1.24	2.70	4.27	**1.20**

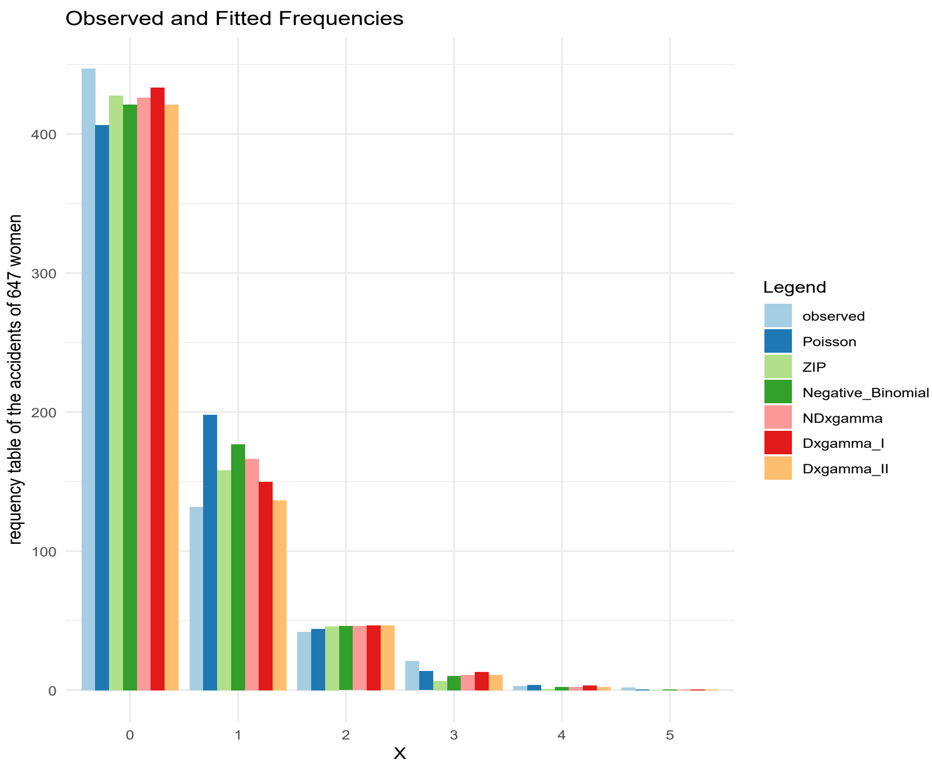

Fig. 7 Bar diagram for the observed and expected frequencies for the accidents of 647 women

10 Concluding Remarks

A discrete form of the NPED distribution has been derived, which combines geometric and negative binomial distributions and is called the NDNPED distribution. This discrete distribution is studied in terms of its characteristics, including the hazard rate function and the density function's shape properties, moments, stress strength reliability, the Lorenz curve, entropies, and measurements based on the maximum likelihood estimation. The detailed simulations provide findings that indicate that the estimation performance has been met as expected. We analyze the model on three real datasets, and the findings show that, in comparison to other discrete distributions, the proposed approach is suitable for the datasets since the distribution works well with over-dispersed data.

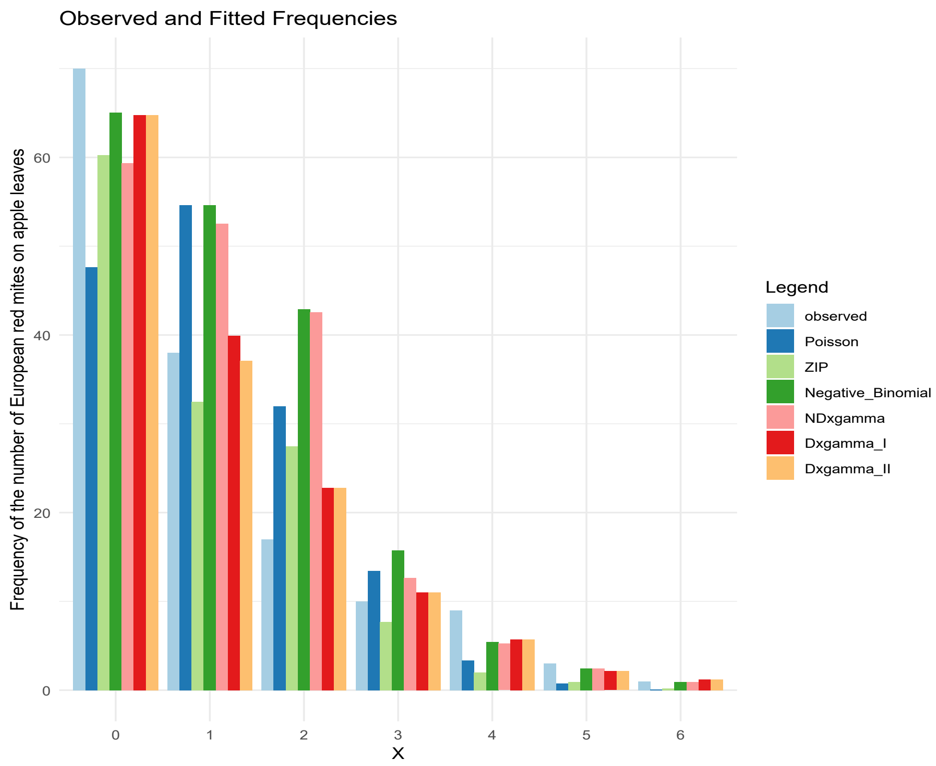

Fig. 8 Bar diagram for the observed and expected frequencies of the number of European red mites on apple leaves

References

[1] Beghriche, A., Zeghdoudi, H., Raman, V., Chouia, S.: New polynomial exponential distribution: properties and applications. Statistics in Transition new series **23**(3), 95–112 (2022)

[2] Maiti, S.S., Ruidas, M.K., Adhya, S.: Natural discrete one parameter polynomial exponential family of distributions and the application. Annals of Data Science **11**(3), 1051–1076 (2024)

[3] Keilson, J., Gerber, H.: Some results for discrete unimodality. Journal of the American Statistical Association **66**(334), 386–389 (1971)

[4] Gupta, P.L., Gupta, R.C., Tripathi, R.C.: On the monotonic properties of discrete failure rates. Journal of statistical planning and inference **65**(2), 255–268 (1997)

[5] Kemp, A.W.: Classes of discrete lifetime distributions (2004)

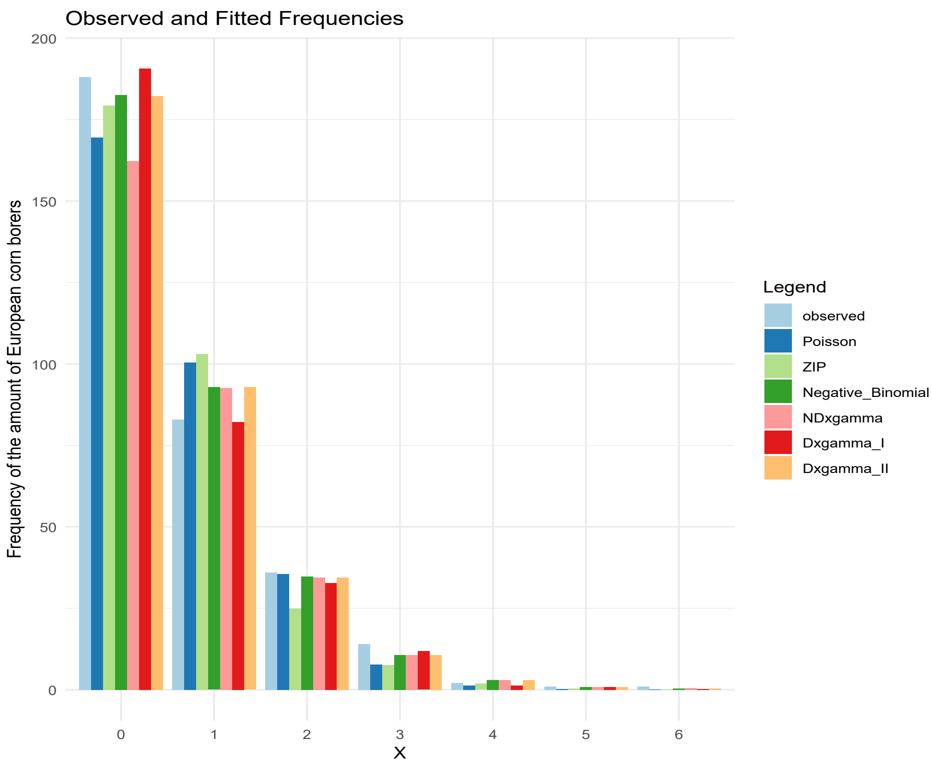

Fig. 9 Bar diagram for the observed and expected frequencies of the number of European corn borers

[6] Consul, P.C., Jain, G.C.: A generalization of the poisson distribution. Technometrics **15**(4), 791–799 (1973)

[7] Garman, P.: Original data on european red mite on apple leaves. Report. Connecticut (1951)

[8] McGuire, J.U., Brindley, T.A., Bancroft, T.A.: The distribution of european corn borer larvae pyrausta nubilalis (hbn.), in field corn. Biometrics **13**(1), 65–78 (1957)

Appendix: R Code for NDxgamma Distribution

1. PMF and CDF Definitions

```
# PMF
dndxgamma <- function(x, theta) {
  if (any(theta <= 0 | theta >= 1)) stop("theta must be in (0,1)")
  if (any(x < 0 | x != floor(x))) stop("x must be a non-negative integer")
  (theta^2 / (1 + theta)) * (1 - theta)^x * (1 + (x + 2)*(x + 1)*theta / 2)
}

# CDF
pndxgamma <- function(x, theta) {
  if (any(theta <= 0 | theta >= 1)) stop("theta must be in (0,1)")
  if (any(x < 0 | x != floor(x))) stop("x must be a non-negative integer")
  1 - ((1 - theta)^(x + 1) * (theta * (x + 2) * (theta + theta * x + 2) + 2)) / (2
      * (1 + theta))
}
```

Listing 1 R Code for PMF and CDF of NDxgamma Distribution

2. Central Moments, Skewness, and Kurtosis

```
central_moment_ndxgamma <- function(m, theta) {
  if (m == 1) return(mean_ndxgamma(theta))
  if (m == 2) return(var_ndxgamma(theta))
  x <- 0:1000
  pmf <- pmf_ndxgamma(x, theta)
  mu <- mean_ndxgamma(theta)
  if (m == 3) return(sum((x - mu)^3 * pmf))
  if (m == 4) return(sum((x - mu)^4 * pmf))
  stop("Only moments up to order 4 are supported")
}

skewness_ndxgamma <- function(theta) {
  mu3 <- central_moment_ndxgamma(3, theta)
  sigma <- sqrt(var_ndxgamma(theta))
  mu3 / (sigma^3)
}

kurtosis_ndxgamma <- function(theta) {
  mu4 <- central_moment_ndxgamma(4, theta)
  sigma2 <- var_ndxgamma(theta)
  mu4 / (sigma2^2) - 3
}
```

Listing 2 R Code for Central Moments and Shape Statistics

3. Mean, Variance, Skewness, and Kurtosis Plot

```
theta_vals <- seq(0.05, 0.95, by = 0.01)
mean_vals <- sapply(theta_vals, mean_ndxgamma)
var_vals <- sapply(theta_vals, var_ndxgamma)
skew_vals <- sapply(theta_vals, skewness_ndxgamma)
kurt_vals <- sapply(theta_vals, kurtosis_ndxgamma)

par(mfrow = c(2, 2), mar = c(3, 3, 2, 1))
plot(theta_vals, mean_vals, type = "l", lwd = 2, col = "black", main = "mean")
plot(theta_vals, var_vals, type = "l", lwd = 2, col = "black", main = "variance")
plot(theta_vals, skew_vals, type = "l", lwd = 2, col = "black", main = "skewness")
plot(theta_vals, kurt_vals, type = "l", lwd = 2, col = "black", main = "kurtosis")
```

Listing 3 R Code for Plotting Mean, Variance, Skewness, and Kurtosis

4. Hazard Rate Function and Plot

```
ndxgammapmf <- function(x, theta) {
  (theta^2 / (1 - theta)) * (1 + theta)^x * (1 + (x + 1) * (x + 2) * (theta / 2))
}

compute_hazard <- function(x_max, theta) {
  x_large <- 0:50
  pmf <- sapply(x_large, ndxgammapmf, theta = theta)
  pmf <- pmf / sum(pmf)
  survival <- sapply(0:x_max, function(x) sum(pmf[(x+1):length(pmf)]))
  hazard <- pmf[1:(x_max+1)] / survival
  hazard
}

x <- 0:15
theta_values <- c(0.1, 0.3, 0.5, 0.9)
par(mfrow = c(2, 2), mar = c(4, 4, 2, 1))
for (theta in theta_values) {
  hazard <- compute_hazard(max(x), theta)
  hist_data <- rep(x, times = round(hazard * 10000))
  hist(hist_data, breaks = seq(-0.5, 15.5, 1), prob = TRUE, main = bquote(theta ==
      .(theta)))
}
```

Listing 4 R Code for Hazard Rate of NDxgamma Distribution

5. Mean Residual Life (MRL) Plot

```
dgl=function()
{
  xaxis=seq(0,15,1)
  par(mfrow=c(2,2))
  hrf=function(x,theta)
  {
    ((1-theta)*(2*theta^2+6*theta+theta^2*x^2+3*theta^2*x+4*theta*x+6))/(theta*(
        theta*(x+2)*(theta*x+theta+2)+2))
  }

  plot(xaxis, hrf(x=xaxis,theta=0.1), type = "h", xlab = "x", ylab = "mrl(x)", lwd
      =1)
  legend("topright", cex=0.85, legend=c(expression(paste(theta==0.1))), lty=1)

  plot(xaxis, hrf(x=xaxis,theta=0.3), type = "h", xlab = "x", ylab = "mrl(x)", lwd
      =1)
  legend("topright", cex=0.85, legend=c(expression(paste(theta==0.3))), lty=1)
```

```

  plot(xaxis, hrf(x=xaxis,theta=0.5), type = "h", xlab = "x", ylab = "mrl(x)", lwd
       =1)
  legend("topright", cex=0.85, legend=c(expression(paste(theta==0.5))), lty=1)

  plot(xaxis, hrf(x=xaxis,theta=0.9), type = "h", xlab = "x", ylab = "mrl(x)", lwd
       =1)
  legend("topright", cex=0.85, legend=c(expression(paste(theta==0.9))), lty=1)
}

dgl()
```

Listing 5 R Code for Mean Residual Life Plot

6. Simulation and Histogram using Mixture of Negative Binomial Distributions

```
f=function(theta,r,n)
{
  htheta=0
  for (k in 1:r) htheta=htheta+gamma(k)/theta^(k)
  htheta=1/htheta
  x=rep(1,n)
  t=rep(1,r+1)
  t[0]=0
  t[1]=htheta/theta
  for (j in 2:r) t[j]=t[j-1]+htheta*gamma(j)/theta^j
  u=runif(n,0,1)
  size=c(1:r)
  prob=theta
  for (i in 1:n)
  {
    if (u[i]<t[1])
    {
      x[i]=rnbinom(1,size[1],prob)
    }
    for (j in 2:length(t))
    {
      if(t[j-1]<u[i] && t[j]>u[i])
      {
        x[i]=rnbinom(1,size[j],prob)
      }
    }
  }
  cat(x,"\n")
  h <- hist(x, col=4, main="Histogram")
}

f(theta=0.4, r=3, n=50)
```

Listing 6 R Code for Mixture Simulation and Histogram

7. Simulation Study: MSE and Bias of MLE Estimator

```
res <- function(xz, r, theta, replication) { #====================[first]start
  ############ PDF CDF Simulation ##############
  ninitial = 5; nend = 20
  nrange = nend - ninitial
  xaxis <- seq(ninitial, nend, 1)
  #cat(xaxis,"\n")
  tt = length(xaxis)
  #-----------------------------------
```

```
  Ymse_mle <- vector("numeric", tt)
  Ybias_mle <- vector("numeric", tt)

  #------------PDF of GLD-------------
  pdf <- function(xz, theta) {
    sum1 = 0
    for (k in 0:r) {
      sum1 = sum1 + (gamma(k + 1) / theta^(k + 1))
    }
    valuesum1 = sum1
    sum2 = 0
    for (k in 0:r) {
      sum2 = sum2 + gamma(k + 1) * choose(xz + k, xz) * (1 - theta)^xz
    }
    valuesum2 = sum2
    pdfvalue = valuesum2 / valuesum1
    #cat("pdf=",pdfvalue,"\n")
  }
  #---------------CDF of GLD----------------

  #----------------------------------------

  for (ij in ninitial:nend) { #==============[1]start
    n = ij

    mse_mle = 0
    bias_mle = 0
    #---------------------
    for (j in 1:replication) { #================[2]start
      #-----------random sample generation-----------------
      htheta = 0
      for (k in 1:(r + 1)) htheta = htheta + gamma(k) / theta^(k)
      htheta = 1 / htheta
      x = rep(1, n)
      t = rep(1, r + 1)
      t[0] = 0
      t[1] = htheta / theta
      for (j in 2:(r + 1)) t[j] = t[j - 1] + htheta * gamma(j) / theta^j
      #cat(t,"\n")
      u = runif(n, 0, 1)
      #cat(u,"\n")
      shape = c(1:(r + 1))
      scale = theta
      for (i in 1:n) {
        if (u[i] < t[1]) {
          x[i] = rnbinom(1, shape[1], scale)
          #cat(shape[1],"\n")
        }
        for (j in 2:length(t)) {
          if (t[j - 1] < u[i] && t[j] > u[i]) {
            x[i] = rnbinom(1, shape[j], scale)
            #cat(shape[j],"\n")
          }
        }
      }
      #cat("x=",x,"n=",n,"\n")
      z = sum(x)
      #---------------------------------------------------
      #------------MLE of theta from negative log likelihood function
          ----------------------------

      msemle <- function(n) {
        fn <- function(y) {
          theta = y[1]
          sumpx = 0
          for (i in 1:n) {
            sum1 = 0
            for (k in 0:r) {
```

```
              sum1 = sum1 + x[i]^k
              #cat("xi=",x[i],"i=",i,"")
            }
            sumpx = sumpx + log(sum1)
          }
          value = sumpx
          #cat("value=",value,"\n")

          #----------------nlog(h(theta))-----------
          sumhtheta = 0
          for (k in 0:r) {
            sumhtheta = sumhtheta + (gamma(k + 1) / theta^(k + 1))
          }
          hthetavalue = n * log(1 / sumhtheta)
          #-----------------------------------

          sum3 = 0
          for (i in 1:n) {
            sum3 = sum3 + (-theta * x[i])
          }
          sum3value = sum3
          #cat("sum3value=",sum3value,"\n")

          nll = -(hthetavalue + value + sum3value)
          #cat("nll=",nll,"\n")
        }
        fit = optim(c(0.5), fn)
        thetahat = fit$par[1]
      }
      #---------------------------------------------------

      #----------------------------------------------------
      mse_mle = mse_mle + msemle(n)^2 / replication
      bias_mle = bias_mle + msemle(n) / replication
    } #============================================[2]end

    Ymse_mle[ij - ninitial + 1] = mse_mle
    Ybias_mle[ij - ninitial + 1] = bias_mle
  }
  #==========PDF plot for MSE========
  plot(xaxis, Ymse_mle, lty = 1, col = 1, lwd = 2, xlab = "Sample Size", ylab = "
      MSE of theta", type = "l")

  #==========PDF plot for bias========
  plot(xaxis, Ybias_mle, lty = 1, col = 2, lwd = 2, xlab = "Sample Size", ylab = "
      Bias of theta", type = "l")
} #====================[first]end

res(xz = 2, r = 2, theta = 0.4, replication = 1000)
```

Listing 7 R Function to Compute MSE and Bias of MLE of Theta

R Code: Fitting NDxgamma and Related Distributions

```
discrete_fitting=function(ini)
{
  f=c(70,38,17,10,9,3,2,1)
  cat("observed=",f,"\n")

  n=length(f)

  lldxg1=function(w)
  {
```

```
    p=w[1]
    pmfdxg1=function(y)
    {
      a_1=function(p) {(1-p+p*log(p)+(log(p))^2/2)/((1-p)*(1-log(p)))}
      b_1=function(p) {-((1-p)*log(p)+(3-p)*(log(p))^2/2)/((1-p)^2*(1-log(p)))}
      c_1=function(p) {((1-p)*(log(p))^2)/((1-p)^3*(1-log(p)))}
      value1=a_1(p)*dgeom(y,p)+b_1(p)*dnbinom(y,2,p)+c_1(p)*dnbinom(y,3,p)
      return(value1)
    }
    llsum=0
    for(j in 1:n) llsum=llsum+f[j]*log(pmfdxg1(j-1))
    return(-llsum)
  }
  z=optim(c(ini),lldxg1)
  par=z$par
  phat=par[1]
  cat("p,dxg-I=",phat,"\n")

  xgammapmf1=function(y,p)
  {
    a_1=function(p) {(1-p+p*log(p)+(log(p))^2/2)/((1-p)*(1-log(p)))}
    b_1=function(p) {-((1-p)*log(p)+(3-p)*(log(p))^2/2)/((1-p)^2*(1-log(p)))}
    c_1=function(p) {((1-p)*(log(p))^2)/((1-p)^3*(1-log(p)))}
    value1=a_1(p)*dgeom(y,p)+b_1(p)*dnbinom(y,2,p)+c_1(p)*dnbinom(y,3,p)
    return(value1)
  }
  x=seq(0,n-1,1)
  cat("dxgammaI=",xgammapmf1(x,phat)*sum(f),"-loglikelihood=",lldxg1(phat),"\n")

  lldxg2=function(w)
  {
    p=w[1]
    pmfdxg2=function(y)
    {
      a_2=function(p) {(2*(1-p)^2*(1-log(p)/2))/(2*(1-p)^2-p*(1+p)*log(p))}
      b_2=function(p) {(3*(1-p)*log(p))/(2*(1-p)^2-p*(1+p)*log(p))}
      c_2=function(p) {-(2*log(p))/(2*(1-p)^2-p*(1+p)*log(p))}
      value2=a_2(p)*dgeom(y,p)+b_2(p)*dnbinom(y,2,p)+c_2(p)*dnbinom(y,3,p)
      return(value2)
    }
    llsum=0
    for(j in 1:n) llsum=llsum+f[j]*log(pmfdxg2(j-1))
    return(-llsum)
  }
  z=optim(c(ini),lldxg2)
  par=z$par
  phat2=par[1]
  cat("p,dxg-II=",phat2,"\n")

  xgammapmf2=function(y,p)
  {
    a_2=function(p) {(2*(1-p)^2*(1-log(p)/2))/(2*(1-p)^2-p*(1+p)*log(p))}
    b_2=function(p) {(3*(1-p)*log(p))/(2*(1-p)^2-p*(1+p)*log(p))}
    c_2=function(p) {-(2*log(p))/(2*(1-p)^2-p*(1+p)*log(p))}
    value2=a_2(p)*dgeom(y,p)+b_2(p)*dnbinom(y,2,p)+c_2(p)*dnbinom(y,3,p)
    return(value2)
  }
  x=seq(0,n-1,1)
  cat("dxgammaII=",xgammapmf2(x,phat2)*sum(f),"-loglikelihood=",lldxg2(phat2),"\n")

  lldlind=function(w)
  {
    p=w[1]
    pmfndxgamma=function(y)
    {
      a_1=function(p) {p/(1+p)}
      b_1=function(p) {1/(1+p)}
      valuedlind=a_1(p)*dgeom(y,p)+b_1(p)*dnbinom(y,3,p)
```

```
      return(valuedlind)
    }
    llsum=0
    for(j in 1:n) llsum=llsum+f[j]*log(pmfndxgamma(j-1))
    return(-llsum)
  }
  z=optim(c(ini),lldlind)
  par=z$par
  phatlind=par[1]
  cat("p,Lind =",phatlind,"\n")

  dlindpmf=function(y,p)
  {
    a_1=function(p) {p/(1+p)}
    b_1=function(p) {1/(1+p)}
    valuedlind=a_1(p)*dgeom(y,p)+b_1(p)*dnbinom(y,3,p)
    return(valuedlind)
  }
  x=seq(0,n-1,1)
  cat("dLindley=",dlindpmf(x,phatlind)*sum(f),"-loglikelihood=",lldlind(phatlind),"
      \n")
}

discrete_fitting(ini=0.2)
```

R Code: Chi-Square Goodness of Fit Test

```
observed <- c(213, 128, 37, 18, 3, 1, 0)  # Observed frequencies
expected <- c(235.8941, 97.79981, 39.83747, 16.00578, 6.360385, 2.504825,
     0.9790591)  # Expected frequencies

# Calculate Chi-Square statistic manually
chi_sq_statistic <- sum((observed - expected)^2 / expected)
df <- length(observed) - 1
p_value <- 1 - pchisq(chi_sq_statistic, df)

# Print results
print(paste("Chi-Square Statistic:", chi_sq_statistic))
print(paste("Degrees of Freedom:", df))
print(paste("P-value:", p_value))
```

R Code: Fitting a Zero-Inflated Poisson (ZIP) Model

```
library(MASS)

# Data
obs1 <- c(rep(0,530642), rep(1,33495), rep(2,2585), rep(3,211), rep(4,25))

# ZIP density
dzip <- function(x, lambda, pi) {
  ifelse((x == 0),
         (pi + (1 - pi) * exp(-lambda)),
         ((1 - pi) * dpois(x, lambda)))
}

# Attempt fit
fit_zip1 <- tryCatch({
  fitdistr(obs1, dzip,
           start = list(lambda = mean(obs1), pi = 0.5),
           lower = list(lambda = 0.001, pi = 0.001),
           upper = list(lambda = 10, pi = 0.999))
```

```
}, error = function(e) {
  message("First attempt failed, trying different starting values")
  fitdistr(obs1, dzip,
           start = list(lambda = 0.5, pi = 0.7),
           lower = list(lambda = 0.001, pi = 0.001),
           upper = list(lambda = 10, pi = 0.999))
})

# Manual fit if needed
if(inherits(fit_zip1, "error")) {
  negloglik <- function(params) {
    lambda <- params[1]
    pi <- params[2]
    -sum(log(dzip(obs1, lambda, pi)))
  }
  fit <- optim(par = c(0.5, 0.5), fn = negloglik,
               lower = c(0.001, 0.001), upper = c(10, 0.999),
               method = "L-BFGS-B")
  fit_zip1 <- list(estimate = fit$par,
                   loglik = -fit$value,
                   convergence = fit$convergence)
  names(fit_zip1$estimate) <- c("lambda", "pi")
}

# Output and evaluation
if(!inherits(fit_zip1, "error")) {
  lambda_hat <- fit_zip1$estimate["lambda"]
  pi_hat <- fit_zip1$estimate["pi"]
  cat("Estimated lambda:", lambda_hat, "\n")
  cat("Estimated pi:", pi_hat, "\n")

  expected <- dzip(0:4, lambda_hat, pi_hat) * length(obs1)
  observed_counts <- c(sum(obs1 == 0), sum(obs1 == 1),
                       sum(obs1 == 2), sum(obs1 == 3), sum(obs1 == 4))

  chi_sq <- sum((observed_counts - expected)^2 / expected)
  df <- length(observed_counts) - 3
  p_value <- pchisq(chi_sq, df, lower.tail = FALSE)

  cat("\nGoodness-of-fit test:\n")
  cat("Chi-squared:", chi_sq, "\n")
  cat("df:", df, "\n")
  cat("p-value:", p_value, "\n")

  k <- 2
  nll <- -fit_zip1$loglik
  aic <- 2*k + 2*nll
  cat("\nAIC:", aic, "\n")

  fit_pois <- fitdistr(obs1, "poisson")
  lambda_pois <- fit_pois$estimate
  nll_pois <- -fit_pois$loglik
  aic_pois <- 2*1 + 2*nll_pois

  cat("\nRegular Poisson:\n")
  cat("Lambda:", lambda_pois, "\n")
  cat("AIC:", aic_pois, "\n")

  barplot(rbind(observed_counts/sum(observed_counts),
                expected/sum(expected)),
          beside = TRUE,
          names.arg = 0:4,
          xlab = "Count",
          ylab = "Proportion",
          main = "Observed vs Expected Frequencies",
          col = c("blue", "red"))
  legend("topright",
         legend = c("Observed", "ZIP Expected"),
```

```r
        fill = c("blue", "red"))
} else {
  message("Fitting failed with all attempted methods")
}
```

R Code: Plotting Observed and Fitted Frequencies

```r
# Define the observed and fitted frequencies
X <- c(0, 1, 2, 3, 4, 5)
observed <- c(447, 132, 42, 21, 3, 2)
fitted_poisson <- c(406.33, 198.01, 43.95, 13.65, 3.61, 0.45)
fitted_zip <- c(427.75, 157.96, 45.85, 6.81, 0.79, 0.07)
fitted_nb <- c(421.02, 176.82, 46.04, 10.31, 2.26, 0.45)
fitted_nb_gamma <- c(426.09, 166.28, 46.04, 10.93, 2.30, 0.45)
fitted_nb_deg_I <- c(433.22, 149.85, 46.49, 13.22, 3.44, 0.56)
fitted_nb_deg_II <- c(421.00, 136.54, 46.40, 10.93, 2.30, 0.45)

# Combine the data into a data frame
data <- data.frame(X, observed, fitted_poisson, fitted_zip, fitted_nb,
                   fitted_nb_gamma, fitted_nb_deg_I, fitted_nb_deg_II)

# Load the necessary libraries
library(ggplot2)
library(reshape2)

# Melt the data frame to long format for plotting
data_long <- melt(data, id="X")

# Plot the bar chart
ggplot(data_long, aes(x=factor(X), y=value, fill=variable)) +
  geom_bar(stat="identity", position="dodge") +
  labs(x="X", y="Frequency", title="Observed and Fitted Frequencies") +
  theme_minimal() +
  scale_fill_brewer(palette="Paired", name="Legend")
```