ONLINE FIRST ARTICLE
doi.org/10.36253/ijas-16758

Statistica Applicata - Italian Journal of Applied Statistics

Received: 20.02.24/ **Accepted:** 09.06.25
Published online: 02.09.25

THE DISCRETE NEW XLINDLEY DISTRIBUTION: A STATISTICAL FRAMEWORK FOR MODELLING MEDICAL AND BIOLOGICAL SCIENCE DATA

Na Elah Shah, Peer Bilal Ahmad[1]
Department of Mathematical Sciences, Islamic University of Science and Technology, Kashmir, India

Abstract *Modelling the frequency of events is a significant problem that has received a lot of attention in recent years. Discrete probability distributions such as the Poisson, Negative Binomial, Geometric, and Poisson-Lindley are commonly used for this purpose. However, these traditional distributions often exhibit limited flexibility in capturing the complexity of real-world count data. In this regard, we study the New Discrete XLindley distribution introduced by (Maya et al., 2024) and discussed its various structural properties. A Bayesian analysis is conducted to enhance the inferential understanding of the model. To address the presence of excess zeros in count data, we propose a zero-inflated extension of the New Discrete XLindley model. Parameters are estimated using the Maximum Likelihood Estimation method, and the performance of the estimators is assessed via simulation studies. The practical relevance of the proposed model is demonstrated through its application to a real-life dataset. Finally, a Likelihood Ratio Test is employed to test the significance of the zero-inflation parameter, providing strong evidence in support of the extended model. Overall, the zero-inflated New Discrete XLindley model offers a flexible and effective tool for modeling zero-inflated count data.*

Keywords: *Bayesian COVID-19 Simulation XLindley Zero-Inflation.*

[1]Peer Bilal Ahmad bilalahmadpz@gmail.com

1. INTRODUCTION

Statistical distributions play a vital role in decision-making and in describing the probabilistic behavior of random events. Across various fields of research, probability distributions have proven highly effective for modeling data. Researchers have shown significant interest in the discrete version of some classical distributions such as Weibull, Lindley, Burr, Gompertz, etc.

In many practical situations, data arise in discrete form due to inherent properties of the phenomena being measured or limitations of the measuring instruments. For instance, in reliability engineering, one may record the number of successful cycles before a device fails, or count how many times a device is switched on or off. In survival analysis, durations such as the number of days or weeks until an event-like remission or relapse in cancer patients; are often treated as discrete. Other examples include the number of deaths or daily cases during the COVID-19 pandemic, the annual number of earthquakes, hospital visits made by individuals, road accidents, counts of ecological species, and the number of insurance claims filed among others.

Given the discrete nature of such data, it is appropriate and often necessary to model these scenarios using suitable discrete probability distributions. One effective approach to obtaining such models is by discretizing continuous distributions. Among various techniques available, the survival discretization method is one of the most popular. A key advantage of this method is that the resulting discrete distribution retains the same functional form of the survival function as its continuous counterpart. This preservation ensures that many reliability properties of the original distribution remain intact. According to this technique, if X is a continuous random variable with survival function $S_X(x) = P(X > x)$, then the probability mass function of its discrete random variable X is given as:

$$P(X = x) = S(x) - S(x+1) \quad ; x = 0, 1, 2, 3... \tag{1}$$

Several authors have used equation (1) for creating the discrete counterpart of continuous distributions, such as the discrete Rayleigh distribution (Roy, 2004), discrete Burr and Pareto distributions (Krishna and Pundir, 2009), discrete Gamma distribution (Chakraborty and Chakravarty, 2012), discrete Modified Weibull distribution (Almalki and Nadarajah, 2014), discrete Generalised Exponential and Exponentiated discrete Weibull distributions (Nekoukhou and Bidram, 2015, 2020), the discrete extended Weibull distribution (Jia et al., 2019), the Geometric zero truncated Poisson distribution (Akdogan et al., 2019), the discrete Poisson quasi-Lindley regression model, the discrete Poisson-Bilal distribution (Altun, 2019, 2020), the discrete Burr-Hatke distribution (El-Morshedy et al., 2020), the discrete inverted Nadarajah-Haghighi distributions (Singh et al., 2022b), the discrete Teissier distribution (Singh et al., 2022a), etc.

Recently, (Al-Babtain et al., 2020) used the natural discrete Lindley distribution to analyse the remission time of Leukaemia patients and fatalities due to COVID-19 infection. Discrete Marshall-Olkin generalised exponential distribution was suggested by (Almetwally et al., 2020) to address the current Egyptian COVID cases on a regular basis. (Elbatal et al., 2022) developed discrete odd Perks-G distributions. They also presented the discrete Marshall-Olkin inverse Toppe-Leone distribution with COVID-19 data as its application. Discrete extended odd Weibull exponential was studied in detail along with various applications by (Nagy et al., 2021). (El-Morshedy et al., 2021) conducted an analysis using a discrete generalised Lindley to look at the counts of fresh daily fatalities in Iran as well as daily coronavirus cases in Hong Kong. Recent work on discrete count data can be seen from (Wani et al., 2023), (Skinder et al., 2023) and (Ahmad and Wani, 2024) where regression analysis and zero inflation of count data is analyzed in different fields.

In light of the existing literature, it is evident that numerous new discrete distributions have been introduced over the past few decades. Despite this progress, there remains considerable scope for developing novel discrete models that can more effectively capture the complexity and variability of real-world data. Motivated by this gap, we aim to explore a flexible discrete probability distribution proposed by (Maya et al., 2024) which shows promise in modeling a wide range of discrete datasets. In addition to its classical properties, we will also examine the Bayesian estimation framework and investigate its performance under zero-inflated conditions, an important aspect in count data modeling. The recurring patterns and trends observed in real-world count data serve as the primary motivation for adopting and studying such versatile models.

Recently, (Nawel et al., 2023) introduced a continuous distribution namely New XLindley distribution (NXLD). The probability density function (pdf) and survival function of NXLD are provided, respectively, as:

$$f_{NXL}(y;\theta) = \frac{\theta}{2}(1+\theta y)e^{-\theta y} \quad ; \, y>0\,, \theta>0$$

$$S_{NXL}(y;\theta) = \left(\frac{1}{2}\theta y + 1\right) e^{-\theta y} \tag{2}$$

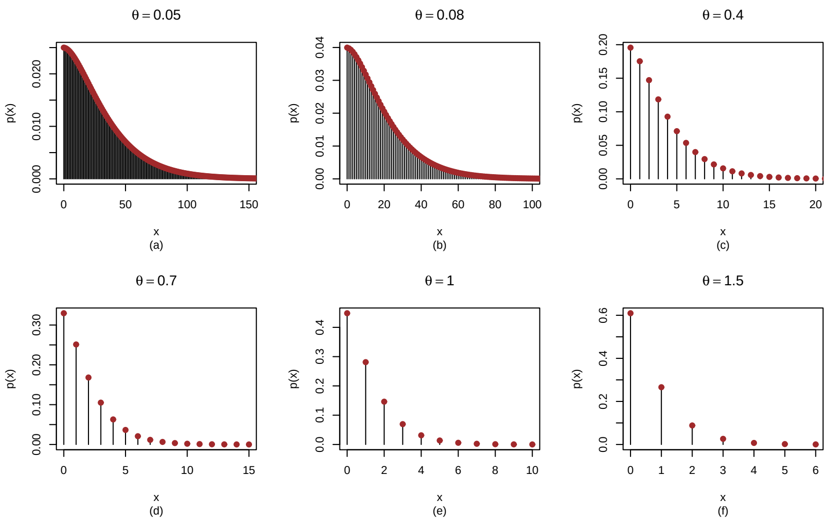

Figure 1: pmf plots of NDXL model with different parameter values

Using the survival discretization technique given in equation (1), the pmf of the discrete verison of NXLD is given as:

$$P_{NDXL}(x;\theta) \quad = \quad \left[\frac{\theta}{2}(1-e^{-\theta})x-\left(\frac{\theta}{2}+1\right)e^{-\theta}+1\right]e^{-\theta x} \qquad ;x=0,1,2,3,\ldots \quad ;\theta>0 \tag{3}$$

The cumulative distribution function (cdf) of the discrete New XLindley distribution (NDXL) model is given as:

$$F_{NDXL}(x;\theta)=1-\frac{\theta(x+1)+2}{2e^{\theta(1+x)}} \qquad ;x=0,1,2,3\ldots \quad ;\theta>0 \tag{4}$$

The pmf and cdf plots of the model are given in figures (1) and (2) respectively. From figure (1), it can be seen that as θ increases, the distribution becomes more concentrated near zero, i.e., the probability of small values of x increases.

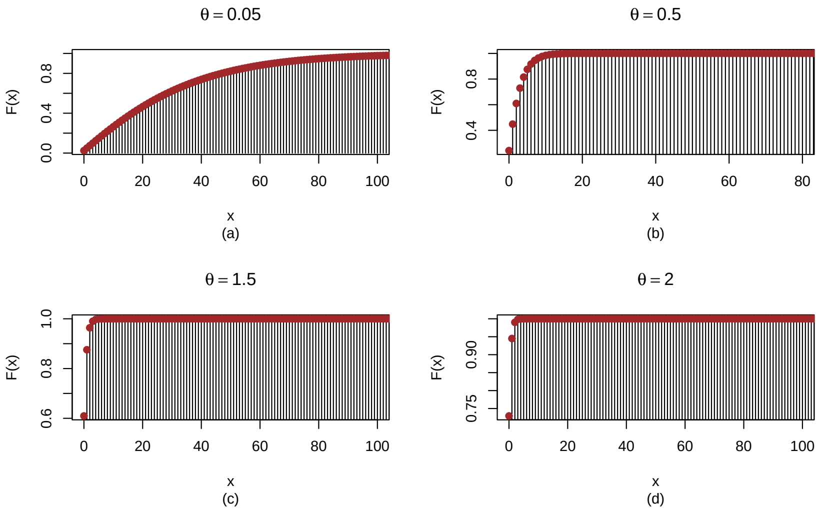

Figure 2: cdf plots of NDXL model

2. MOMENTS AND ASSOCIATED MEASURES

2.1. FACTORIAL MOMENTS

The factorial moments of NDXL the distribution are given as:

$$\mu_{(1)} = \frac{(\theta+2)e^{\theta}-2}{2(e^{\theta}-1)^2}$$

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

$$\mu_{(3)} = \frac{3(3\theta e^{\theta}+2e^{\theta}-2)}{(e^{\theta}-1)^4}$$

$$\mu_{(4)} = \frac{24(2\theta e^{\theta}+e^{\theta}-1)}{(e^{\theta}-1)^5}$$

2.2. MOMENTS ABOUT ORIGIN

The raw moments of NDXL the distribution are given as:

$$\begin{aligned}
\mu_1' &= \frac{(\theta+2)e^{\theta}-2}{2(e^{\theta}-1)^2} \\
\mu_2' &= \frac{(\theta+2)e^{2\theta}+3\theta e^{\theta}-2}{2(e^{\theta}-1)^3} \\
\mu_3' &= \frac{(\theta+2)e^{3\theta}+2(5\theta+3)e^{2\theta}+(7\theta-6)e^{\theta}-2}{2(e^{\theta}-1)^4} \\
\mu_4' &= \frac{(\theta+2)e^{4\theta}+5(5\theta+4)e^{3\theta}+55\theta e^{2\theta}+5(3\theta-4)e^{\theta}-2}{2(e^{\theta}-1)^5}
\end{aligned}$$

Mean and Variance of the NDXL model are:

$$\begin{aligned}
Mean_{(NDXL)} &= \frac{(\theta+2)e^{\theta}-2}{2(e^{\theta}-1)^2} \\
Variance_{(NDXL)} &= \frac{2(\theta+2)e^{3\theta}-(\theta^2+8)e^{2\theta}-2(\theta-2)e^{\theta}}{4(e^{\theta}-1)^4}
\end{aligned}$$

Skewness and Kurtosis can be obtained from the standard formulas.

2.3. INDEX OF DISPERSION AND COEFFICIENT OF VARIATION

A statistic used to assess the degree of over-dispersion or under-dispersion in data is the index of dispersion (iod). Over-dispersion is indicated by an iod more than one, and under-dispersion is indicated by an iod less than one. When the iod is equal to 1, the model is equi-dispersed. The iod for NDXL distribution is:

$$iod = \frac{Variance}{Mean} = \frac{2(\theta+2)e^{3\theta}-(\theta^2+8)e^{2\theta}-2(\theta-2)e^{\theta}}{2(\theta+2)e^{3\theta}-4(3+\theta)e^{2\theta}+2(\theta+6)e^{\theta}-4}$$

The coefficient of variation (C.V) is another indicator of data variability. The variability of independent samples is frequently compared using this measure. The C.V of a random variable X with the NDXL distribution is:

$$C.V = \frac{\sqrt{2(\theta+2)e^{3\theta}-(\theta^2+8)e^{2\theta}-2(\theta-2)e^{\theta}}}{(\theta+2)e^{\theta}-2} \tag{5}$$

From table (1) and figures (3 & 4), it is clear that:

1. The mean and variance of the model decreases as the value of parameter θ

Table 1: Few descriptive statistics for the NDXL distribution

Measure ↓ θ →	0.08	0.5	1.0	1.5	2.0	2.5	3.0	3.5
Mean	18.2533	2.5209	1.0423	0.5645	0.3375	0.2112	0.1351	0.08733
Variance	273.3958	6.9574	1.7049	0.7294	0.3859	0.2262	0.1397	0.0886
C.V	0.9058	1.0463	1.2527	1.5129	1.8405	2.2516	2.7662	3.4093

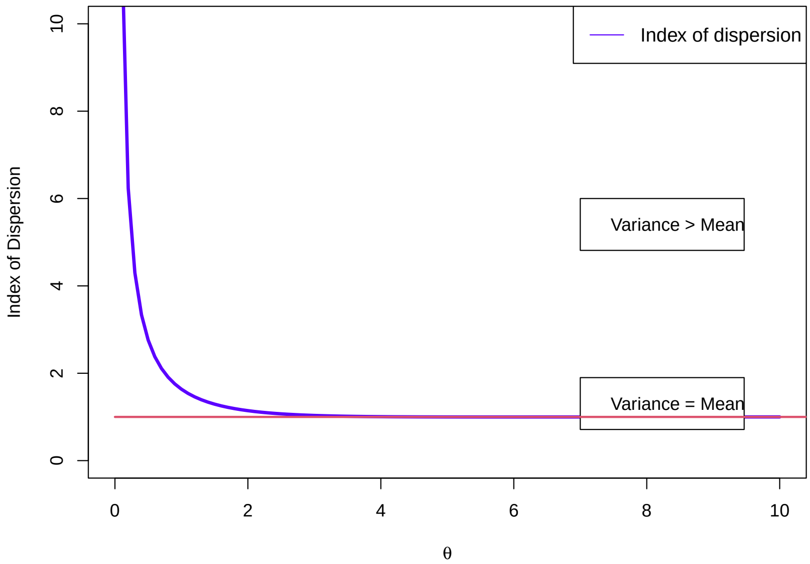

Figure 3: Index of dispersion plot of NDXL model

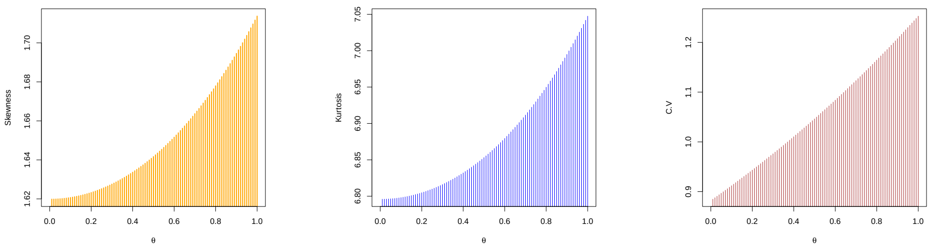

Figure 4: Skewness, Kurtosis and C.V plots of the NDXL model

increases.
2. The model is appropriate for modelling equi-dispersed, and over dispersed data.
3. The model is positively skewed.
4. The proposed model is leptokurtic in nature.

2.4. GENERATING FUNCTIONS

2.4.1. PROBABILITY GENERATING FUNCTION

When working with discrete random variables that have values 0, 1, 2,..., the probability generating function (PGF) is a helpful tool. Its main benefit is that it makes it simple for us to describe how X + Y will distribute when X and Y are independent. In general, using the traditional probability function to determine the distribution of a sum is challenging. The PGF converts a sum into a product, making it considerably easier to manage. The PGF of NDXL distribution is given as:

$$P_X(s) \quad = \frac{2e^{2\theta}-[(\theta+2)-(\theta-2)s]e^{\theta}+2s}{2(e^{\theta}-s)^2}$$

2.4.2. MOMENT GENERATING FUNCTION

The Moment Generating Function (MGF) is used to obtain the moments of a distribution. It has an essential property known as the uniqueness property in addition to aiding in the discovery of moments. According to the uniqueness property, if the MGF exists for a random variable, only one distribution may be linked to that MGF. Therefore, the distribution of a random variable is determined only by the MGF.

$$M_X(t) \quad = \frac{2e^{2\theta}-[(\theta+2)-(\theta-2)e^{t}]e^{\theta}+2e^{t}}{2(e^{\theta}-e^{t})^2}$$

Remark1: Recurrence relation between probabilities is given by:

$$P(x+1) = \frac{\frac{\theta}{2}(1-e^{-\theta})x+(\frac{\theta}{2}+1)-e^{-\theta}(1+\theta)}{e^{\theta}[\frac{\theta}{2}(1-e^{-\theta})x-(\frac{\theta}{2}+1)e^{-\theta}+1]}P(x)$$

with $\qquad P(0) = 1-\left(\frac{\theta}{2}+1\right)e^{-\theta}$

Remark2: $\qquad h(0) = \frac{e^{\theta}}{1+\frac{\theta}{2}}-1 = \frac{P(0)}{1-P(0)}$

3. RELIABILITY ANALYSIS

3.1. RELIABILITY FUNCTION/SURVIVAL FUNCTION

The likelihood that a component/device/equipment or system performs its intended function satisfactorily for a certain amount of time under specific operating circumstances. The reliability function for the model is given by

$$R(x,\theta) = P(X > x) = 1 - F(X = x)$$

$$R(x;\theta) = \frac{\theta(1+x)+2}{2e^{\theta(1+x)}}$$

3.2. HAZARD RATE FUNCTION (HRF)

The failure rate or hazard rate, or force of mortality is the ratio of the probability mass function P(x) to the survival function R(x), which is denoted by the expression:

$$\begin{aligned} h(x) &= \frac{P(x;\theta)}{R(x;\theta)} \\ \Rightarrow h(x) &= \frac{2\left[\frac{\theta}{2}(1-e^{-\theta})x - (\frac{\theta}{2}+1)e^{-\theta}+1\right]e^{\theta}}{\theta(1+x)+2} \end{aligned}$$

The HRF plots for different parameter values are shown in figure (5).

3.3. INVERSE HAZARD RATE

The reverse hazard rate is given by

$$r(x) = \frac{P_{NDXL}(x;\theta)}{F_{NDXL}(x;\theta)}$$

$$\Rightarrow r(x) = \frac{2[\frac{\theta}{2}(1-e^{-\theta})x - (\frac{\theta}{2}+1)e^{-\theta}+1]e^{\theta}}{2e^{\theta(1+x)} - \theta(1+x)+2} \tag{6}$$

4. PARAMETRIC ESTIMATION

Two estimation techniques viz., Maximum Likelihood Estimation (MLE) and Moment Method are used in this section to estimate the unknown parameter of the NDXL distribution.

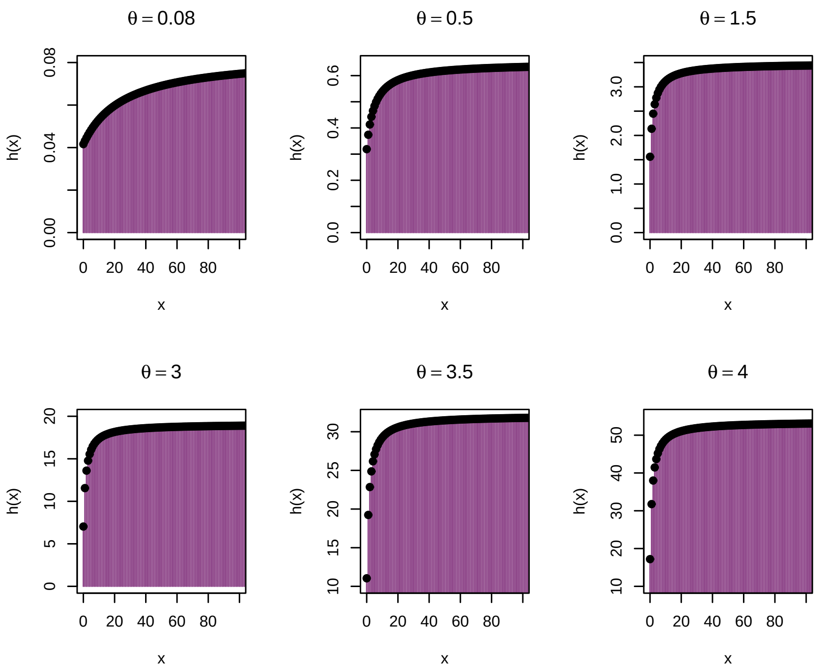

Figure 5: The hrf plots of NDXL model

4.1. MAXIMUM LIKELIHOOD ESTIMATION

Let the random variables $X_1, X_2, ..., X_n$ have a NDXL distribution. The likelihood function is given by

$$L(x;\theta) = \prod_{i=1}^{n} \left[\frac{\theta}{2}(1-e^{-\theta})x_i - (\frac{\theta}{2}e^{-\theta}) + 1 \right] e^{-\theta \Sigma x_i} \tag{7}$$

Equation (7) is differentiated with respect to the parameter θ in order to provide the following non-linear likelihood equation.

$$\sum_{i=1}^{n} x_i - \frac{1}{2}\sum_{i=1}^{n} \left[\frac{(\theta e^{-\theta} - e^{-\theta} + 1)x_i + (1+\theta)e^{-\theta}}{\frac{\theta}{2}(1-e^{-\theta})x_i - (\frac{\theta}{2}+1)e^{-\theta} + 1} \right] = 0 \tag{8}$$

The maximum likelihood estimator of θ is provided by the solution of equation (8). However, the solution to equation (8) has no explicit form. Therefore, iterative methods like Newton-Raphson, Nelder-Mead, etc. must be used to solve equation (8). The alternative option is to directly minimize the negative log-likelihood function. This may be accomplished by using the fitdistr or nlm commands of R software (see (R Core Team et al., 2013)).

4.2. MOMENT ESTIMATION

Equating the population moment (5) of NDXL distribution to the corresponding sample moment (m_1) we get:

$$m_1 = \frac{(\theta+2)e^{\theta} - 2}{2(e^{\theta}-1)^2}$$

$$\Rightarrow e^{\theta}[(\theta - 2e^{\theta}m_1) + 2(1+m_1)] = 2(m_1+1) \tag{9}$$

The **nleqslv** command in R software can be used to solve the above nonlinear equation.

5. ORDER STATISTICS

Let's suppose that the random variables $X_1, X_2, ..., X_n$ have a NDXL distribution and their associated order statistics are $X_{1:n}, X_{2:n}, ..., X_{n:n}$. The cdf of ith order statistic is

$$F_{(i:n)}(x;\theta) = \sum_{r=i}^{n}(^{n}C_{r})[F(x;\theta)]^{r}[1-F(x;\theta)]^{n-r}$$

$$= \sum_{r=i}^{n}\sum_{j=0}^{n-r}\zeta_{(m)}^{(n,r)}\left[1-\frac{\theta(x+1)+2}{2e^{\theta(1+x)}}\right]^{r+j}$$

Where $\zeta_{(m)}^{(n,r)} = (-1)^{j}(^{n}C_{r})(^{n-r}C_{j})$

Also, the pmf of ith order statistic is given by:

$$f_{i:n}(x;\theta) = \sum_{r=i}^{n}\sum_{j=0}^{n-r}\zeta_{(m)}^{(n,r)}[f(x;\theta)]^{r+j}$$

$$\Rightarrow f_{i:n}(x;\theta) = \sum_{r=i}^{n}\sum_{j=0}^{n-r}\zeta_{(m)}^{(n,r)}\left[\left(\frac{\theta}{2}(1-e^{-\theta})x-\left(\frac{\theta}{2}+1\right)e^{-\theta}+1\right)e^{-\theta x}\right]^{r+j}$$

The u^{th} moments of $X_{i:n}$ can therefore be written as:

$$E(X_{i:n}^{u}) = \sum_{x=0}^{\infty}\sum_{r=i}^{n}\sum_{j=0}^{n-r}\zeta_{(m)}^{(n,r)}x^{u}[f_{i}(x;\theta)]^{r+j} \tag{10}$$

To summarise the theoretical distribution, one can use the L-moment statistics given by (Hosking, 1990). L-moment statistics is an expectation of linear combination of order statistics.
The L-moment statistics of the random variable X is given by:

$$\Upsilon_{\alpha} = \frac{1}{\alpha}\sum_{j=0}^{\alpha-1}(-1)^{j}(^{\alpha-1}C_{j})E(X_{\alpha-j:\alpha}) \tag{11}$$

Some descriptive measures based on L-moment statistics can be obtained from equation (11), such as:
Mean=Υ_1, C.V=$\frac{\Upsilon_2}{\Upsilon_1}$, C.S=$\frac{\Upsilon_3}{\Upsilon_2}$, and C.K=$\frac{\Upsilon_4}{\Upsilon_2}$

6. BAYESIAN ANALYSIS OF NDXL MODEL

In Bayesian statistics, an unknown parameter is treated as a random variable with a specified prior probability distribution that incorporates existing knowledge or beliefs about the parameter. Estimating the lifetime of future samples based on an informative dataset is of significant interest to researchers, engineers, statisticians, and other applied scientists who employ predictive techniques for various purposes. In this section, we present a Bayesian analysis of the NDXL distribution.

Lets suppose the parameter θ in equation (3) is a random variable following Gamma distribution with parameters ϕ, β, i.e.,

$$\theta \sim G(\phi, \beta)$$

$$\therefore \pi(\theta) = \frac{\beta^{\phi}}{\Gamma(\phi)} e^{-\beta\theta} \theta^{\phi-1} \; ; \phi, \beta > 0 \tag{12}$$

Now, using equation (3) and the prior distribution (12), the posterior distribution is obtained as:

$$P(\theta|x) = \frac{P(x,\theta)\pi(\theta)}{\int_0^{\infty} P(x,\theta)\pi(\theta) d\theta} \tag{13}$$

where

$$P(x,\theta)\pi(\theta) = \frac{\phi^{\beta}}{\Gamma(\beta)} \theta^{\beta-1} e^{-\theta(x+\phi)} \left(1 + \frac{x\theta}{2}\right) - \frac{\phi^{\beta}}{\Gamma(\beta)} \theta^{\beta-1} e^{-\theta(x+\phi+1)} \left(1 + \frac{x\theta}{2} + \frac{\theta}{2}\right)$$

$$\therefore P(\theta|x) = \frac{\frac{\phi^{\beta}}{\Gamma(\beta)} \theta^{\beta-1} e^{-\theta(x+\phi)} \left(1 + \frac{x\theta}{2}\right) - \frac{\phi^{\beta}}{\Gamma(\beta)} \theta^{\beta-1} e^{-\theta(x+\phi+1)} \left(1 + \frac{x\theta}{2} + \frac{\theta}{2}\right)}{\int_0^{\infty} \frac{\phi^{\beta}}{\Gamma(\beta)} \theta^{\beta-1} e^{-\theta(x+\phi)} \left(1 + \frac{x\theta}{2}\right) - \frac{\phi^{\beta}}{\Gamma(\beta)} \theta^{\beta-1} e^{-\theta(x+\phi+1)} \left(1 + \frac{x\theta}{2} + \frac{\theta}{2}\right) d\theta}$$

$$\Rightarrow P(\theta|x) = \frac{e^{-(\phi+x+1)\theta} (\phi+1)^{\beta+1} (\phi+x+1)^{\beta+1} \theta^{\beta-1} \left[2 + \theta + x\theta - (2 + x\theta) e^{\theta}\right]}{(\phi+x)^{\beta+1} \left[2 + 2\phi + \beta + x(2+\beta)\right] - (\phi+x+1)^{\beta+1} \left[2\phi + x(2+\beta)\right]} \tag{14}$$

Bayes estimate under squared error loss function (SELF) is the posterior mean of equation (14) i.e.,

$$\hat{\theta} = \int_0^{\infty} \theta P(\theta|x) d\theta$$

Table 2: The competitive models of the NDXL distribution.

Distribution	Abbreviation	Author
Poisson distribution	PD	(Poisson, 1837)
Zero Inflated Poisson	ZIP	(Lambert, 1992)
Geometric Distribution	GD	(Gómez-Déniz, 2010)
Negative Binomial	NB	(Fermat)
Poisson Lindley	PL	(Sankaran, 1970)
Discrete Lindley	DL	(Gómez-Déniz and Calderín-Ojeda, 2011)
Discrete Burr	DB	(Krishna and Pundir, 2009)
Discrete Burr Hatke	DBH	(El-Morshedy et al., 2020)
Discrete Rayleigh	DR	(Roy, 2004)
Discrete Weibull	DW	(Nakagawa and Osaki, 1975)

$$\Rightarrow \hat{\theta} = \frac{(\phi+1)^{\beta+1)}(\phi+x+1)^{\beta+1}}{\Gamma(\beta)\left[(\phi+x)^{\beta+1}[2+2\phi+\beta+x(2+\beta)]-(\phi+x+1)^{\beta+1}[2\phi+x(2+\beta)]\right]} \times \int_0^\infty e^{-(\phi+x+1)\theta}\theta^\beta\left[2+\theta+x\theta-(2+x\theta)e^\theta\right]d\theta$$

$$\Rightarrow \hat{\theta} = \frac{\beta(\phi+1)^{\beta+1}(\phi+x+1)^{\beta+1}}{(\phi+x)^{\beta+1}[2+2\phi+\beta+(2+\beta)x]-(\phi+x+1)^{\beta+1}[2\phi+(2+\beta)x]} \times \left[\frac{2(\phi+x+1)+(1+x)(\beta+1)}{(\phi+x+1)^{\beta+2}} - \frac{2(\phi+x)+(\beta+1)x}{(\phi+x)^{\beta+2}}\right]$$

7. EMPIRICAL INVESTIGATION

On the basis of applications to real data sets, the significance of NDXL distribution is illustrated in this section. Several metrics are used to compare the fitted models, including the negative log-likelihood ($-logl$), Akaike information criterion (AIC), Bayesian information criterion (BIC), and the Chi-square (χ^2) with its related p-values. Comparing the NDXL distribution's fits to other competitive distributions with one or two parameters from table (2), allows us to see how well NDXL distribution fits the data.

Table 3: Expected frequencies of fitted distributions for Data set I

X	Frequency	Expected Frequency										
		NDXL	PD	ZIP	GD	NB	PL	DL	DB	DBH	DR	DW
0	126.00	122.37	74.94	126.00	137.96	120.22	128.68	118.78	128.74	198.27	46.89	120.12
1	80.00	90.69	115.71	65.08	83.74	92.99	87.14	92.78	109.10	64.12	106.32	92.88
2	59.00	58.51	89.34	68.97	50.82	59.17	55.26	60.26	45.36	30.77	101.22	59.04
3	42.00	35.11	45.98	48.73	30.85	34.94	33.63	35.82	21.88	17.57	61.17	35.13
4	24.00	20.15	17.75	25.82	18.72	19.84	19.89	20.21	12.26	11.06	25.65	20.07
5	8.00	11.21	5.48	10.95	11.36	10.99	11.52	11.02	7.63	7.42	7.72	11.13
6	5.00	6.11	1.41	3.87	6.90	5.99	6.57	5.86	5.12	5.20	1.70	6.03
7	4.00	3.27	0.31	1.17	4.19	3.22	3.70	3.06	3.63	3.76	0.28	3.20
8	3.00	1.73	0.06	0.31	2.54	1.72	2.06	1.58	2.68	2.78	0.03	1.67

7.1. Data Set I: *Epileptic Seizure*

The first data set shows the daily count of epileptic seizures that a patient had during the observation period after receiving consistent anticonvulsant medication (see (Chakraborty, 2010)). The data is given in table (3) with the expected frequencies of each competing models. The fitting resluts are shown in table (4).

7.2. Data Set II: *Vaccine Adverse Events*

The second data set, which is taken from (Rose et al., 2006) and is given in table (5), shows the adverse reactions to the Anthrax Vaccine Absorbed (AVA) vaccine that were documented during clinical trials. A total of 4020 observations were obtained after 1005 research participants had 4 study injections separated by intervals of 0 weeks, 2 weeks, 4 weeks, and 6 months. The fitting resluts are shown in table (6).

7.3. Data Set III: *COVID-19 in South Korea*

The daily new COVID-19 deaths in South Korea from 15 Feb 2020 to 12 Dec 2020 are included in this data set, which can be accessed at `https://www.worldometers.info/coronavirus/country/south-korea/`(South Korea). This data set has been recently used by (El-Morshedy et al., 2022) and is given in table (7) with the expected frequencies of each competing model. The fitting results are shown in table (8).

Table 4: Goodness of fit for dataset I

Model	$-logl$	AIC	BIC	χ^2	df	p-value	Est. Parameters(SE's)
NDXL	594.485	1190.969	1194.830	5.38	6	0.496	$\hat{\theta}$=0.74548(0.03524)
PD	636.045	1274.091	1277.952	81.02	4	<0.001	$\hat{\lambda}$=1.54416(0.06633)
ZIP	599.637	1203.275	1210.996	14.98	4	0.005	$\hat{\phi}$=2.11964(0.10799)
							$\hat{\lambda}$=0.27150(0.03102)
GD	598.396	1198.791	1202.652	9.57	6	0.140	$\hat{p}$= 0.39306(0.01634)
NB	594.942	1193.884	1201.605	5.31	4	0.252	$\hat{p}$= 0.50095(0.04728)
							$\hat{r}$=1.55005(0.27756)
PL	595.181	1192.362	1196.223	174.55	4	<0.001	$\hat{\theta}$=0.97344(0.05308)
DL	594.611	1191.222	1195.083	5.04	5	0.411	$\hat{\lambda}$=0.46047(0.01437)
DB	622.319	1248.638	1256.360	41.76	5	<0.001	$\hat{\alpha}$= 2.19169(0.19156)
							$\hat{\theta}$=0.51726(0.02920)
DBH	644.904	1291.809	1295.669	105.37	6	<0.001	$\hat{\lambda}$=0.87024(0.02195)
DR	672.298	1346.596	1350.457	5.55	6	0.478	$\hat{p}$=0.86641(0.00678)
DW	594.749	1193.499	1201.225	5.18	4	0.269	$\hat{q}$=0.65778(0.02405)
							$\hat{\beta}$=1.15609(0.05966)

Table 5: Expected frequencies of fitted distributions for Data set II

V.A.E	Frequency	Expected Frequency										
		NDXL	PD	ZIP	GD	NB	PL	DL	DB	DBH	DR	DW
0	1437.00	1425.00	890.76	1437.00	1603.53	1409.08	1500.13	1386.93	1468.30	2278.60	545.84	1410.52
1	1010.00	1046.80	1342.34	787.29	963.90	1068.65	1003.49	1069.87	1286.67	735.60	1231.72	1065.56
2	660.00	667.57	1011.43	803.25	579.41	670.65	629.24	686.09	522.44	352.25	1161.41	667.85
3	428.00	395.51	508.06	546.35	348.29	391.63	378.74	402.69	246.93	200.58	691.80	393.16
4	236.00	223.95	191.41	278.71	209.36	220.16	221.62	224.34	136.34	125.94	284.56	222.65
5	122.00	122.98	57.69	113.75	125.85	120.88	127.03	120.75	83.94	84.18	83.65	122.58
6	62.00	66.05	14.49	38.68	75.65	65.32	71.67	63.42	55.80	58.75	17.87	66.00
7	34.00	34.89	3.12	11.28	45.47	34.89	39.94	32.71	39.25	42.32	2.80	34.88
8	14.00	18.19	0.59	2.88	27.33	18.47	22.03	16.63	28.82	31.23	0.32	18.15
9	8.00	9.38	0.10	0.65	16.43	9.71	12.05	8.36	21.89	23.48	0.03	9.31
10	4.00	4.80	0.01	0.13	9.88	5.08	6.55	4.17	17.08	17.92	0.00	4.72
11	4.00	2.43	0.00	0.02	5.94	2.64	3.54	2.06	13.63	13.85	0.00	2.36
12	1.00	1.23	0.00	0.00	3.57	1.37	1.90	1.01	11.09	10.81	0.00	1.17

Table 6: Goodness of fit for dataset II

Model	$-logl$	AIC	BIC	$Chi-sq$	df	p-value	Est. Parameters(SE's)
NDXL	6738.497	13478.995	13485.294	6.28	9	0.712	$\hat{\theta}$=0.75966(0.01062)
PD	7231.13	14464.27	14470.57	1279.25	5	<0.001	$\hat{\lambda}$=1.50697(0.01936)
ZIP	6868.79	13741.59	13754.19	1101.35	5	<0.001	$\hat{\lambda}$=2.04054(0.03144)
							$\hat{\phi}$=0.26149(0.00935)
GD	6778.04	13558.09	13564.39	74.28	10	<0.001	$\hat{p}$=0.39889(0.00488)
NB	6740.60	13485.21	13497.81	10.04	8	0.255	$\hat{p}$=0.50327(0.01384)
							$\hat{r}$=1.52678(0.07989)
PL	6745.99	13493.99	13500.29	19.24	10	0.037	$\hat{\theta}$=0.99415(0.01612)
DL	6739.750	13481.500	13487.800	9.3	9	0.410	$\hat{\lambda}$=0.45459(0.00427)
DB	6993.27	13990.54	14003.14	363.09	10	<0.001	$\hat{\alpha}$=2.27232(0.05941)
							$\hat{\theta}$=0.51906(0.00868)
DBH	7286.33	14574.66	14580.96	2372.12	11	<0.001	$\hat{\lambda}$= 0.86637(0.00659)
DR	7694.39	15390.79	15397.08	302.77	5	<0.001	$\hat{p}$=0.86429(0.00203)
DW	6739.679	13483.359	13495.957	8.84	9	0.452	$\hat{q}$=0.64913(0.00710)
							$\hat{\beta}$=1.14699 (0.01727)

Table 7: Expected frequencies of fitted distributions for Data set III

		Expected frequencies										
X	Frequency	NDXL	PD	ZIP	GD	NB	PL	DL	DB	DBH	DR	DW
0	89.00	91.36	45.41	89.00	104.78	89.41	95.02	86.33	92.89	166.60	29.50	89.08
1	79.00	72.48	86.34	49.47	68.67	74.51	31.43	74.62	97.79	54.60	72.40	74.42
2	49.00	51.24	82.08	60.80	45.00	51.98	9.68	53.71	42.67	26.67	80.78	51.86
3	29.00	33.99	52.02	49.82	29.49	33.90	2.85	35.40	21.18	15.54	61.94	34.09
4	19.00	21.66	24.73	30.62	19.32	21.34	0.82	22.15	12.17	10.02	35.68	21.62
5	17.00	13.42	9.40	15.05	12.66	13.15	0.23	13.39	7.75	6.89	15.99	13.36
6	9.00	8.15	2.98	6.17	8.30	7.98	0.06	7.90	5.31	4.95	5.66	8.10
7	6.00	4.87	0.81	2.17	5.44	4.79	0.02	4.58	3.83	3.68	1.60	4.83
8	6.00	2.88	0.19	0.67	3.56	2.85	0.00	2.62	2.88	2.81	0.36	2.84
9	1.00	1.68	0.04	0.18	2.34	1.69	0.00	1.48	2.23	2.18	0.07	1.65

Table 8: Goodness of fit for dataset III

Model	$-logl$	AIC	BIC	$Chi-sq$	df	p-value	Est. Parameters(SE's)
NDXL	564.82	1131.64	1135.35	4.20	6	0.65	$\hat{\theta}$=0.63204(0.03203)
PD	621.10	1244.20	1247.91	116.07	8	<0.001	$\hat{\lambda}$=1.90132(0.07908)
ZIP	587.70	1179.39	1186.83	5.96	6	0.43	$\hat{\lambda}$=2.45830(0.11653)
							$\hat{\phi}$=0.22657(0.02976)
GD	568.08	1138.17	1141.88	6.09	5	0.29	$\hat{p}$=0.34467(0.01600)
NB	564.88	1133.76	1141.19	4.11	5	0.53	$\hat{p}$=0.43833(0.04329)
							$\hat{r}$=1.48382(0.24370)
PL	565.13	1132.25	1135.97	13.82	6	0.32	$\hat{\theta}$= 0.81458(0.04544)
DL	565.05	1132.10	1135.81	4.06	5	0.54	$\hat{\lambda}$=0.51078(0.01466)
DB	587.65	1179.30	1186.73	5.63	6	0.47	$\hat{\alpha}$=2.46633 (0.24846)
							$\hat{\theta}$=0.59093(0.03142)
DBH	620.47	1242.93	1246.65	105.44	5	<0.001	$\hat{\theta}$= 0.90395(0.02032)
DR	638.90	1279.81	1283.53	72.98	7	<0.001	$\hat{p}$=0.90297(0.00537)
DW	564.62	1133.25	1140.68	4.06	5	0.54	$\hat{q}$=0.70696(0.02408)
							$\hat{\beta}$=1.15432(0.06059)

Table 9: Expected frequencies of fitted distributions for Data set IV

X	Frequency	Expected frequencies										
		NDXL	PD	ZIP	GD	NB	PL	DL	DB	DBH	DR	DW
0	188.00	185.89	169.46	188.00	196.58	185.89	194.05	188.13	188.00	217.82	122.56	185.90
1	83.00	89.02	109.83	81.84	77.31	89.20	79.53	86.38	93.21	59.78	153.03	89.15
2	36.00	33.10	35.59	38.56	30.40	32.98	31.32	32.74	24.95	23.59	43.91	32.99
3	14.00	11.04	7.69	12.11	11.96	10.98	11.99	11.33	8.71	10.85	4.34	11.05
4	2.00	3.47	1.25	2.85	4.70	3.46	4.50	3.72	3.79	5.43	0.16	3.47
5	1.00	1.05	0.16	0.54	1.85	1.05	1.66	1.18	1.92	2.86	0.00	1.04

7.4. Data Set IV: *Larvae Pyrausta*

This data set comes from a biological experiment that quantified the amount of larvae Pyrausta (the European corn borer), that were present in the field (see (Holt et al., 2002)). The data is given in table (9) with the expected frequencies of each competing model. The fitting results are shown in table (10).

All the results were obtained from R software using the package "fitdistr". The ML technique is used to estimate the parameters of all the distributions. Table (4,6,8,10) shows the parameter estimates for each data set together with their standard errors (SE's) and goodness of fit metrics. The results in these tables demonstrate that, in comparison to other models, the NDXL model offers a good fit to the data. These examples of real datasets demonstrate that the NDXL model not only offers strong p-values but also the lowest AIC and BIC values when compared to the distributions displayed in table (2). As a result, it reveals that the NDXL model experiences the least degree of information loss when compared to the competing distributions that were taken into consideration.

8. ZERO INFLATED NDXL DISTRIBUTION

Excess zeros in count data are frequently encountered across various disciplines. Traditional models, such as the Poisson and Negative Binomial distributions, often underestimate the probability of zero counts. To address this issue, several methods have been proposed, including the Zero-Inflated Poisson (ZIP) model by (Lambert, 1992) and the Zero-Inflated Negative Binomial (ZINB) model by (Greene, 1994). In this subsection, we introduce a novel zero-inflated count

Table 10: Goodness of fit for dataset IV

Model	$-logl$	AIC	BIC	χ^2	df	p-value	Est. Parameters(SE's)
NDXL	355.91	713.82	717.61	0.82	2	0.664	$\hat{\theta}$=1.37633(0.06942)
PD	362.25	726.49	730.27	15.45	2	<0.01	$\hat{\lambda}$=0.64815(0.04473)
ZIP	355.04	714.08	721.64	0.33	1	0.566	$\hat{\lambda}$=0.94237(0.10305)
							$\hat{\phi}$=0.31222(0.06377)
GD	357.88	717.76	721.54	4.10	3	0.25	$\hat{p}$=0.60674(0.02114)
NB	355.95	715.91	723.47	0.88	1	0.348	$\hat{p}$=0.74039(0.07280)
							$\hat{r}$=1.84852(0.68430)
PL	357.05	716.10	719.88	3	3	0.39	$\hat{\theta}$=2.04325(0.15134)
DL	355.94	713.87	717.65	1.82	3	0.611	$\hat{\lambda}$=0.26731(0.01580)
DB	362.69	729.39	736.95	10.51	2	0.005	$\hat{\alpha}$=2.01275(0.18919)
							$\hat{\theta}$=0.28582(0.02669)
DBH	369.70	741.40	745.18	23.92	3	<0.01	$\hat{\lambda}$=0.65546(0.03455)
DR	404.49	810.97	814.75	67.42	1	<0.01	$\hat{p}$=0.62172(0.01752)
DW	355.91	715.82	723.38	0.86	1	0.354	$\hat{q}$=0.42623(0.02721)
							$\hat{\beta}$=1.14806(0.07678)

model based on the NDXL distribution as an alternative to the existing ZIP and ZINB models.

The pmf of ZINDXL distribution can be obtained as:

$$P(X=x)=\begin{cases}\alpha+(1-\alpha)P_{NDXL}(x;\theta) & ;x=0\\ (1-\alpha)P_{NDXL}(x;\theta) & ;x=1,2,3,...\end{cases}$$

$$\Rightarrow P(X=x)=\begin{cases}\alpha+(1-\alpha)[1-(\frac{\theta}{2}+1)e^{-\theta}] & ;x=0\\ (1-\alpha)\left[\frac{\theta}{2}(1-e^{-\theta})x-\left(\frac{\theta}{2}+1\right)e^{-\theta}+1\right]e^{-\theta x} & ;x=1,2,3,...\end{cases} \tag{15}$$

The pmf plots are given in Figure (6) for different values of the parameters θ and α. The parameter θ controls the shape and dispersion of the underlying NDXL distribution, while α introduces zero-inflation, accounting for excess zeros often observed in real-world data.

In subplot (a), with θ=0.05 and α=0.01, the distribution exhibits a long right tail, suggesting a high level of dispersion. The probability at zero is slightly inflated, reflecting minimal zero-inflation.

In subplot (b), with θ=0.2 and α=0.08, the distribution becomes more concentrated around lower values of x, and the spike at zero becomes more pronounced, indicating moderate zero-inflation.

In subplot (c), with θ=0.4 and α=0.2, the distribution is tightly concentrated near the origin, with a significant spike at x=0. This reflects strong zero-inflation and reduced dispersion due to a higher value of θ .

Overall, the plots illustrate that as α increases, the probability at zero becomes more dominant, while increasing θ leads to a steeper decline in the pmf, indicating a more concentrated distribution and can handle both zero-inflation as well as overdispersion. These features highlight the flexibility of the ZINDXL model in accommodating datasets with both heavy tails and excess zeros.

The mean and variance of ZINDXL distribution are given as:

$$\begin{aligned}Mean=m_1' &= (1-\alpha)\frac{(\theta+2)e^{\theta}-2}{2(e^{\theta}-1)^2}\\ m_2' &= (1-\alpha)\frac{(\theta+2)e^{2\theta}+3\theta e^{\theta}-2}{2(e^{\theta}-1)^3}\\ \therefore Variance(m2) &= m_2'-m_1'^2\\ &= \frac{1-\alpha}{4(e^{\theta}-1)^4}\Big[2(\theta+2)e^{3\theta}-\big((1-\alpha)\theta^2-4\alpha\theta-4(\alpha-2)\big)e^{2\theta}\\ &\qquad -(2\theta+4\alpha\theta+8\alpha-4)e^{\theta}+4\alpha\Big]\end{aligned}$$

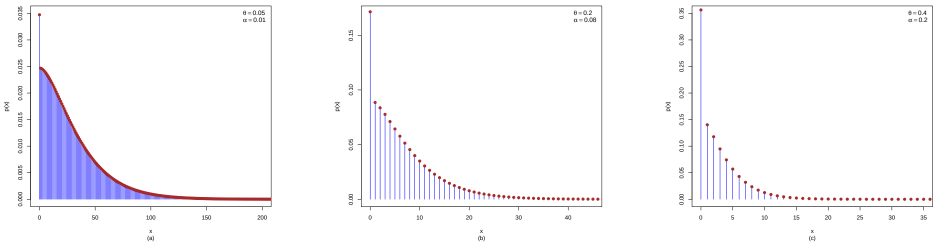

Figure 6: The pmf plots of the ZINDXL model

9. MAXIMUM LIKELIHOOD ESTIMATION OF ZINDXL

One of the popular technique to determine the parameters of the statistical model is MLE. In MLE, we want to maximise the chance of observing the data from the joint probability distribution given a certain probability distribution and its parameters.

Let's take a sample $X_1, X_2, X_3, ..., X_n$ from ZINDXL distribution of size n with parameters θ, α and Y be the number of $x_i' s$ taking value zero. The likelihood function is given by:

$$l(\alpha, \theta|x) = \prod_{i=1}^{n} P(x_i)$$

$$\begin{aligned} \Rightarrow l(\alpha, \theta|x) &= \left[\alpha + (1-\alpha)[1 - (\frac{\theta}{2}+1)e^{-\theta}]\right]^Y \times \\ & \prod_{i=1}^{n}(1-\alpha)\left[\frac{\theta}{2}(1-e^{-\theta})x_i - (\frac{\theta}{2}+1)e^{-\theta}+1\right]e^{-\theta x_i} \end{aligned}$$

$$\begin{aligned} &= \left[\alpha + (1-\alpha)[1 - (\frac{\theta}{2}+1)e^{-\theta}]\right]^Y (1-\alpha)^{n-Y} e^{-\theta x} \\ & \prod_{i=1}^{n}\left[\frac{\theta}{2}(1-e^{-\theta})x_i - (\frac{\theta}{2}+1)e^{-\theta}+1\right] \end{aligned}$$

$$\begin{aligned} \therefore \log l &= Y \log\left[\alpha + (1-\alpha)\left(1 - (\frac{\theta}{2}+1)\right)e^{-\theta}\right] + (n-Y)\log(1-\alpha) \\ & -\theta \sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\left[\frac{\theta}{2}(1-e^{-\theta})x_i - (\frac{\theta}{2}+1)e^{-\theta}+1\right] \end{aligned}$$

Differentiating $\log l$ w.r.t α and equating to zero we get

$$\hat{\alpha} = \frac{Y-n}{n(\frac{\hat{\theta}}{2}+1)e^{-\theta}}+1 \tag{16}$$

Since the equation for θ is not in closed form, thus we employ fitdistr function in R-software for obtaining the ML estimates of ZINDXL model.

10. SIMULATION STUDY

In this section, we carried out a Monte Carlo simulation study to examine the finite sample behaviour of the MLE's. The procedure is given below:

1. Using the inverse cdf technique and the arbitrary set of parameter $(\alpha, \theta) \in$ (0.2,0.5), (0.4,0.7), (0.5,0.8), (0.7,1.0)), generate 10,000 samples of size n $\in$ [25, 75, 100, 300, 600] from the ZINDXL distribution.

2. Calculate the MLE's for each of the 10,000 samples.

3. Compute the Bias, Mean-squared errors (MSE's), and Mean relative estimates (MRE's)

R software has been used for all the calculations. Simulation results are given in table (11). The following conclusions may be taken from the table.

- As the sample size increases, the magnitude of bias approaches to zero.

- The MSE's of the parameter approaches to zero as the sample size increases. This demonstrates the consistency of estimators.

- As sample size increases, the MRE's approaches to 1, which proves that estimators are asymptotically unbiased.

10.1. Data Set V: No. of still births

This dataset consists of frequencies of stillbirths in 402 litters of New Zealand white rabbits, originally used by (Morgan et al., 2007).

From the fitting results shown in the table (13), it is very clear that the ZINDXL distribution is the most effective of the models studied, exhibiting maximum p-value, minimum AIC and BIC values.

Table 11: Simulation results based on MLE's of ZINDXL

Sample Size(n)	Parameter	$\alpha=0.2,\theta=0.5$ Bias	MSE	MRE	$\alpha=0.4,\theta=0.7$ Bias	MSE	MRE
25	$\hat{\alpha}$	0.01409	0.01580	0.92057	-0.03732	0.02968	0.90671
	$\hat{\theta}$	0.00220	0.01330	0.86379	0.03970	0.05045	1.05672
75	$\hat{\alpha}$	-0.01233	0.00570	0.80697	-0.01169	0.00914	0.97077
	$\hat{\theta}$	0.02826	0.00417	0.90861	0.04597	0.01502	1.06567
100	$\hat{\alpha}$	0.00279	0.00425	0.87201	-0.01594	0.00850	0.96015
	$\hat{\theta}$	0.00324	0.00208	0.86557	0.03081	0.01006	1.04401
300	$\hat{\alpha}$	0.00785	0.00159	0.89375	0.00425	0.00261	1.01062
	$\hat{\theta}$	-0.00059	0.00088	0.85898	0.00027	0.00319	1.00039
600	$\hat{\alpha}$	-0.00188	0.00091	0.85191	0.00900	0.00121	1.02250
	$\hat{\theta}$	0.00751	0.00071	0.87291	-0.00061	0.00115	0.99913
Sample Size(n)	Parameter	$\alpha=0.5,\theta=0.8$ Bias	MSE	MRE	$\alpha=0.7,\theta=1.0$ Bias	MSE	MRE
25	$\hat{\alpha}$	-0.04422	0.03777	0.91156	-0.05500	0.04523	0.92142
	$\hat{\theta}$	0.06682	0.14441	1.08352	0.46009	6.95561	1.06009
75	$\hat{\alpha}$	-0.01391	0.00961	0.97217	-0.01486	0.01232	0.97877
	$\hat{\theta}$	0.07627	0.03804	1.09534	0.12016	0.08222	1.12016
100	$\hat{\alpha}$	-0.00496	0.00784	0.99009	-0.01555	0.00748	0.97778
	$\hat{\theta}$	0.03150	0.01421	1.03937	0.07496	0.04686	1.07496
300	$\hat{\alpha}$	0.00686	0.00265	1.01373	-0.00211	0.00213	0.99698
	$\hat{\theta}$	-0.00445	0.00492	0.99444	0.00340	0.01204	1.00340
600	$\hat{\alpha}$	0.00336	0.00127	1.00672	0.00470	0.00121	1.00672
	$\hat{\theta}$	0.00390	0.00239	1.00488	0.00209	0.00801	1.00209

Table 12: Expected frequencies of ZI distributions for Data set V

X	Frequency	Expected frequencies			
		ZINDXL	NDXL	ZIP	ZINB
0	314	314.00	264.39	313.00	313.00
1	48	42.00	99.83	36.00	43.88
2	20	23.00	27.86	28.00	22.27
3	7	12.00	6.88	15.00	11.10
4	5	6.00	1.59	6.00	5.48
5	2	3.00	0.35	2.00	2.69
6	6	1.00	0.08	0.00	1.32

Table 13: Goodness of fit for dataset V

Model	$-logl$	AIC	BIC	χ^2	df	p-value	Est. Parameters(SE's)
ZINDXL	333.18	670.35	678.34	4.23	2	0.122	$\hat{\theta}$=0.92779(0.08736)
							$\hat{\phi}$=0.62089(0.04184)
NDXL	372.03	746.06	750.05	51.91	2	<0.001	$\hat{\theta}$=1.68906(0.07869)
ZIP	341.05	686.10	694.09	13.68	2	0.001	$\hat{\lambda}$=1.57835(0.15537)
							$\hat{\phi}$=0.72350(0.02833)
ZINB	333.44	670.88	682.87	3.43	1	0.06	$\hat{p}$=0.519316(0.27074)
							$\hat{r}$=1.11100(1.86877)
							$\hat{\phi}$=0.57562(0.26283)

Table 14: Calculated value of test statistic in case of LRT.

	l_1	l_2	Test Statistic
Data Set V	372.03	333.18	77.7

11. HYPOTHESIS TESTING

To assess the significance of the zero-inflation parameter in ZINDXL model, we use likelihood ratio test to evaluate the following null hypothesis:

$$H_0 : \alpha = 0 \quad v/s \quad H_1 : \alpha > 0$$

11.1. LIKELIHOOD RATIO TEST

The Likelihood Ratio Test (LRT) compares the null hypothesis H_0 with the alternative hypothesis H_1 using the ratio of two log-likelihood functions. The test statistic for the LRT is given by:

$$LRT_\alpha = -2(l_1 - l_2) \tag{17}$$

where l_1 and l_2 are the maximum log-likelihood under NDXL and ZINDXL distribution respectively. The test statistic (17) is asymptotically distributed as χ^2 with one degree of freedom.

Since the LRT statistic exceeds the critical value of 3.84, we reject the null hypothesis, indicating that the zero-inflation parameter plays a statistically significant role in our proposed model (i.e., ZINDXL).

12. CONCLUSION

In this study, we began by examining the properties of the New Discrete XLindley (NDXL) model and subsequently derived some of its key structural characteristics, including its order statistics. We then carried out a Bayesian analysis to further explore the model's inferential capabilities. Building upon the NDXL framework, we proposed a zero-inflated version to better handle datasets with excessive zeros; an extension particularly relevant for real-world applications in count data modeling. To estimate the parameters of the proposed model, we employed the Maximum Likelihood Estimation (MLE) method and assessed the performance of the estimators through a comprehensive simulation study. The results demonstrated the robustness and efficiency of the MLEs under various sample sizes and parameter settings.

Furthermore, the practical utility of the zero-inflated NDXL model was validated using a real-life dataset. The application confirmed the model's ability to effectively capture the underlying data structure, particularly in the presence of zero inflation. Finally, hypothesis testing was conducted to assess the significance of the zero-inflation parameter. The Likelihood Ratio Test (LRT) provided strong evidence in favour of the zero-inflated structure, highlighting its statistical importance and justifying its inclusion in the model. Overall, the proposed zero-inflated NDXL model presents a flexible and powerful tool for analyzing count data, with promising theoretical properties and practical relevance.

Future work may explore the extension of the zero-inflated NDXL model to regression frameworks and its application to a wider range of real-world datasets.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the anonymous referees and the Editor-in-Chief for their insightful and constructive comments. Their valuable feedback has significantly enhanced the quality and clarity of this manuscript.

References

Ahmad, P.B. and Wani, M.K. (2024). New discrete distributon for zero-inflated count data. In *Reliability: Theory & Applications*, 19 (1 (77)): 717–728.

Akdogan, Y., Coskun, K., BIDRAM, H., and KINACI, İ. (2019). Geometric-zero truncated poisson distribution: properties and applications. In *Gazi University Journal of Science*, 32 (4): 1339–1354.

Al-Babtain, A.A., Ahmed, A.H.N., and Afify, A.Z. (2020). A new discrete analog of the continuous lindley distribution, with reliability applications. In *Entropy*, 22 (6): 603.

Almalki, S.J. and Nadarajah, S. (2014). A new discrete modified weibull distribution. In *IEEE Transactions on Reliability*, 63 (1): 68–80.

Almetwally, E.M., Almongy, H.M., and Saleh, H.A. (2020). Managing risk of spreading âcovid-19â in egypt: Modelling using a discrete marshall-olkin generalized exponential distribution. In *Int. J. Probab. Stat*, 9 (2): 33–41.

Altun, E. (2019). A new model for over-dispersed count data: Poisson quasi-lindley regression model. In *Mathematical Sciences*, 13: 241–247.

Altun, E. (2020). A new one-parameter discrete distribution with associated regression and integer-valued autoregressive models. In *Mathematica slovaca*, 70 (4): 979–994.

Chakraborty, S. (2010). On some distributional properties of the family of weighted generalized poisson distribution. In *Communications in StatisticsâTheory and Methods*, 39 (15): 2767–2788.

Chakraborty, S. and Chakravarty, D. (2012). Discrete gamma distributions: Properties and parameter estimations. In *Communications in Statistics-Theory and Methods*, 41 (18): 3301–3324.

El-Morshedy, M., Altun, E., and Eliwa, M. (2021). A new statistical approach to model the counts of novel coronavirus cases. In *Mathematical Sciences*, 1–14.

El-Morshedy, M., Eliwa, M.S., and Altun, E. (2020). Discrete burr-hatke distribution with properties, estimation methods and regression model. In *IEEE access*, 8: 74359–74370.

El-Morshedy, M., Eliwa, M., and Tyagi, A. (2022). A discrete analogue of odd weibull-g family of distributions: properties, classical and bayesian estimation with applications to count data. In *Journal of Applied Statistics*, 49 (11): 2928–2952.

Elbatal, I., Alotaibi, N., Almetwally, E.M., Alyami, S.A., and Elgarhy, M. (2022). On odd perks-g class of distributions: properties, regression model, discretization, bayesian and non-bayesian estimation, and applications. In *Symmetry*, 14 (5): 883.

Fermat, P. (). Varia opera mathematica d. petri de fermat. In .

Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. In *Test*, 19: 399–415.

Gómez-Déniz, E. and Calderín-Ojeda, E. (2011). The discrete lindley distribution: properties and applications. In *Journal of statistical computation and simulation*, 81 (11): 1405–1416.

Greene, W.H. (1994). Accounting for excess zeros and sample selection in poisson and negative binomial regression models. In .

Holt, A.R., Gaston, K.J., and He, F. (2002). Occupancy-abundance relationships and spatial distribution: a review. In *Basic and Applied Ecology*, 3 (1): 1–13.

Hosking, J.R. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. In *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 52 (1): 105–124.

Jia, J.M., Yan, Z.Z., and Peng, X.Y. (2019). A new discrete extended weibull distribution. In *IEEE access*, 7: 175474–175486.

Krishna, H. and Pundir, P.S. (2009). Discrete burr and discrete pareto distributions. In *Statistical methodology*, 6 (2): 177–188.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. In *Technometrics*, 34 (1): 1–14.

Maya, R., Jodrá, P., Aswathy, S., and Irshad, M. (2024). The discrete new xlindley distribution and the associated autoregressive process. In *International Journal of Data Science and Analytics*, 1–27.

Morgan, B.T., Palmer, K.J., and Ridout, M.S. (2007). Negative score test statistic. In *The American Statistician*, 61 (4): 285–288.

Nagy, M., Almetwally, E.M., Gemeay, A.M., Mohammed, H.S., Jawa, T.M., Sayed-Ahmed, N., and Muse, A.H. (2021). The new novel discrete distribution with application on covid-19 mortality numbers in kingdom of saudi arabia and latvia. In *Complexity*, 2021: 1–20.

Nakagawa, T. and Osaki, S. (1975). The discrete weibull distribution. In *IEEE Transactions on Reliability*, R-24 (5): 300–301. doi:10.1109/TR.1975.5214915.

Nawel, K., Gemeay, A.M., Zeghdoudi, H., Karakaya, K., Alshangiti, A.M., Bakr, M., Balogun, O.S., Muse, A.H., and Hussam, E. (2023). Modeling voltage real data set by a new version of lindley distribution. In *IEEE Access*.

Nekoukhou, V. and Bidram, H. (2015). The exponentiated discrete weibull distribution. In *Sort*, 39: 127–146.

Nekoukhou, V. and Bidram, H. (2020). A new discrete distribution based on geometric odds ratio. In *Journal of Statistical Modelling: Theory and Applications*, 1 (2): 153–166.

Poisson, S.D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile: précédées des règles générales du calcul des probabilités*. Bachelier.

R Core Team, R. et al. (2013). R: A language and environment for statistical computing. In .

Rose, C.E., Martin, S.W., Wannemuehler, K.A., and Plikaytis, B.D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. In *Journal of biopharmaceutical statistics*, 16 (4): 463–481.

Roy, D. (2004). Discrete rayleigh distribution. In *IEEE transactions on reliability*, 53 (2): 255–260.

Sankaran, M. (1970). 275. note: The discrete poisson-lindley distribution. In *Biometrics*, 145–149.

Singh, B., Agiwal, V., Nayal, A.S., and Tyagi, A. (2022a). A discrete analogue of teissier distribution: Properties and classical estimation with application to count data. In *Reliability: Theory & Applications*, 17 (1 (67)): 340–355.

Singh, B., Singh, R.P., Nayal, A.S., and Tyagi, A. (2022b). Discrete inverted nadarajah-haghighi distribution: Properties and classical estimation with application to complete and censored data. In *Statistics, Optimization & Information Computing*, 10 (4): 1293–1313.

Skinder, Z., Ahmad, P.B., and Elah, N. (2023). A new zero-inflated count model with applications in medical sciences. In *Reliability: Theory & Applications*, 18 (3 (74)): 841–855.

Wani, M.A., Ahmad, P.B., Para, B.A., and Elah, N. (2023). A new regression model for count data with applications to health care data. In *International Journal of Data Science and Analytics*, 1–15.

13. APPENDIX

1. **SOLUTION OF MOMENTS ABOUT ORIGIN**

$$
\begin{aligned}
E(X)=\mu_1' &= \sum_{x=0}^{\infty}\frac{\theta}{2}(1-e^{-\theta})x^2e^{-\theta x}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\sum_{x=0}^{\infty}xe^{-\theta x}\\
&= \frac{\theta}{2}(1-e^{-\theta})\sum_{x=0}^{\infty}x^2e^{-\theta x}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\sum_{x=0}^{\infty}xe^{-\theta x}\\
&= \frac{\theta}{2}(1-e^{-\theta})\frac{e^{\theta}(e^{\theta}+1)}{(e^{\theta}-1)^3}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\frac{e^{\theta}}{(e^{\theta}-1)^2}\\
&= \frac{\theta}{2}\frac{(e^{\theta}+1)}{(e^{\theta}-1)^2}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\frac{e^{\theta}}{(e^{\theta}-1)^2}\\
\therefore \mu_1' &= \frac{(\theta+2)e^{\theta}-2}{2(e^{\theta}-1)^2}
\end{aligned}
$$

$$
\begin{aligned}
E(X^2)=\mu_2' &= \frac{\theta}{2}(1-e^{-\theta})\sum_{x=0}^{\infty}x^3e^{-\theta x}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\sum_{x=0}^{\infty}x^2e^{-\theta x}\\
&= \frac{\theta}{2}\frac{(4e^{\theta}+e^{2\theta}+1)}{(e^{\theta}-1)^3}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\frac{e^{\theta}(e^{\theta}+1)}{(e^{\theta}-1)^3}\\
\therefore \mu_2' &= \frac{(\theta+2)e^{2\theta}+3\theta e^{\theta}-2}{2(e^{\theta}-1)^3}
\end{aligned}
$$

$$
\begin{aligned}
E(X^3)=\mu_3' &= \frac{\theta}{2}(1-e^{-\theta})\sum_{x=0}^{\infty}x^4e^{-\theta x}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\sum_{x=0}^{\infty}x^3e^{-\theta x}\\
&= \frac{\theta(11e^{\theta}+11e^{2\theta}+e^{3\theta}+1)}{2(e^{\theta}-1)^4}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\times\\
&\qquad \frac{e^{\theta}(4e^{\theta}+e^{2\theta}+1)}{(e^{\theta}-1)^4}\\
\therefore \mu_3' &= \frac{(\theta+2)e^{3\theta}+2(5\theta+3)e^{2\theta}+(7\theta-6)e^{\theta}-2}{2(e^{\theta}-1)^4}
\end{aligned}
$$

$$
\begin{aligned}
E(X^4)=\mu_4' &= \frac{\theta}{2}(1-e^{-\theta})\sum_{x=0}^{\infty}x^5e^{-\theta x}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\sum_{x=0}^{\infty}x^4e^{-\theta x}\\
&= \frac{\theta}{2}\frac{(26e^{\theta}+66e^{2\theta}+26e^{3\theta}+3^{4\theta}+1)}{(e^{\theta}-1)^5}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\times\\
&\qquad \frac{e^{\theta}(11e^{\theta}+11e^{2\theta}+e^{3\theta}+1)}{(e^{\theta}-1)^5}\\
\therefore \mu_4' &= \frac{(\theta+2)e^{4\theta}+5(5\theta+4)e^{3\theta}+55\theta e^{2\theta}+5(3\theta-4)e^{\theta}-2}{2(e^{\theta}-1)^5}
\end{aligned}
$$

2. SOLUTION OF FACTORIAL MOMENTS

Using relation between factorial moments and moments about origin

$$\mu_{(1)}=\mu_1'=\frac{(\theta+2)e^{\theta}-2}{2(e^{\theta}-1)^2}$$

$$\mu_{(2)}=\mu_2'-\mu_1'=\frac{2(\theta e^{\theta}+e^{\theta}-1)}{(e^{\theta}-1)^3}$$

$$\mu_{(3)}=\mu_3'-3\mu_2'+2\mu_1'=\frac{3(3\theta e^{\theta}+2e^{\theta}-2)}{(e^{\theta}-1)^4}$$

$$\mu_{(4)}=\mu_4'-6\mu_3'+11\mu_2'-6\mu_1'=\frac{24(2\theta e^{\theta}+e^{\theta}-1)}{(e^{\theta}-1)^5}$$

3. SOLUTION OF MOMENT GENERATING FUNCTION

$$
\begin{aligned}
M_X(t)=E(e^{tx}) &= \frac{\theta}{2}(1-e^{-\theta})\sum_{x=0}^{\infty}xe^{-(\theta-t)x}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\sum_{x=0}^{\infty}e^{-(\theta-t)x}\\
&= \frac{\theta}{2}(1-e^{-\theta})\frac{e^{\theta+t}}{(e^{\theta}-e^{t})^2}-\left[\left(\frac{\theta}{2}+1\right)e^{-\theta}-1\right]\frac{e^{\theta}}{e^{\theta}-e^{t}}\\
\Rightarrow M_X(t) &= \frac{2e^{2\theta}-[(\theta+2)-(\theta-2)e^{t}]e^{\theta}+2e^{t}}{2(e^{\theta}-e^{t})^2}
\end{aligned}
$$

4. SOLUTION OF RECURRENCE RELATION

$$\begin{aligned}
\frac{P(x+1)}{P(x)} &= \frac{\left[\frac{\theta}{2}(1-e^{-\theta})(x+1)-\left(\frac{\theta}{2}+1\right)e^{-\theta}+1\right]e^{-\theta(x+1)}}{\left[\frac{\theta}{2}(1-e^{-\theta})x-\left(\frac{\theta}{2}+1\right)e^{-\theta}+1\right]e^{-\theta x}} \\
&= \frac{\left[\frac{\theta}{2}(1-e^{-\theta})x+\frac{\theta}{2}(1-e^{-\theta})-\frac{\theta}{2}e^{-\theta}-e^{-\theta}+1\right]e^{-\theta}}{\left[\frac{\theta}{2}(1-e^{-\theta})x-\left(\frac{\theta}{2}+1\right)e^{-\theta}+1\right]} \\
&= \frac{\left[\frac{\theta}{2}(1-e^{-\theta})x+\left(\frac{\theta}{2}+1\right)-e^{-\theta}(1+\theta)\right]}{e^{\theta}\left[\frac{\theta}{2}(1-e^{-\theta})x-\left(\frac{\theta}{2}+1\right)e^{-\theta}+1\right]} \\
\Rightarrow P(x+1) &= \frac{\left[\frac{\theta}{2}(1-e^{-\theta})x+\left(\frac{\theta}{2}+1\right)-e^{-\theta}(1+\theta)\right]}{e^{\theta}\left[\frac{\theta}{2}(1-e^{-\theta})x-\left(\frac{\theta}{2}+1\right)e^{-\theta}+1\right]}P(x)
\end{aligned}$$

14. APPENDIX

```
ndxl = function(x,theta){
(((theta/2)*(1-exp(-theta))*x)-(((theta/2)+1)*exp(-theta)-1))*exp(-theta*
x)
}
obs = rep(seq(0,8),c(126,80,59,42,24,8,5,4,3))
values = seq(0,8)
x = seq(0,100)
library(MASS)
Est = fitdistr(obs,densfun = ndxl,start = list(theta = 0.5),lower = c(0.001),upper =
c(Inf))
Est
dxl = function(x,theta = Est$estimate[1]){
(((theta/2)*(1-exp(-theta))*x)-(((theta/2)+1)*exp(-theta)-1))*exp(-theta*
x)
}
dxl11 = c(Est$loglik,AIC(Est),BIC(Est))
dxl11
```