International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 23 (2018) pp. 16468-16477

A New Three-Parameter Poisson-Lindley Distribution for Modelling Over-dispersed Count Data

Kishore K Das[1],Inzamul Ahmed[2] and Sahana Bhattacharjee[3]

[1,2,3,]*Department of Statistics, Gauhati University, Guwahati, India.*

Abstract

In this paper, a new Three-parameter Poisson-Lindley distribution (NTPPLD), of which Shanker and Mishra's [17] Two-parameter Poisson-Lindley distribution and Sankaran's [10] One-parameter Poisson-Lindley distribution are particular cases, is proposed. Distributional properties and common descriptive measures pertaining to this mixed distribution are derived. The behaviour of the probability mass function with variations in the parameters is also studied. Estimation of parameters by the method of maximum likelihood and method of moments are discussed and a simulation study is carried out to check the consistency of the maximum likelihood estimates. Finally, the proposed distribution is applied to real-data sets and it is seen that this distribution is a flexible model that may be a useful alternative to known distributions like Poisson, Poisson Lindley, Two-parameter Poisson Lindley and many others for count data analysis.

AMS Subject Classification: 65Cxx,62Exx,62Fxx,62Pxx.

Key Words and Phrases: Count data, Compound distribution, Poisson-Lindley distribution, Two-Parameter Poisson Lindley distribution, Simulation study, Root Mean Square Error.

1 Introduction

Count data are used to construe occurrences in many potential branches such as biology, clinical trials, engineering, insurance etc. [5]. Count data analysis may use a Poisson distribution to describe the data if its variance to mean ratio, called the dispersion index, is unity (equi-dispersion) [14]. However, in many practical situations, this assumption is often not valid and the Poisson distribution is therefore an inflexible model to give an account of these phenomena [9, 4]. An inequality of variance and mean results in over dispersion if the variance exceeds the mean. Contrary to it, when the variance is less than the mean, data reflects under dispersion [23]. Many researchers are looking at this over-dispersion impediment which can be addressed by the use of mixed Poisson distributions [9, 7, 4]. A mixed Poisson distribution arises when a random variable, say X follows Poisson distribution with some parameter, and the parameter, sayλ, itself behaves as a random variable with some specified distribution known as the mixing distribution [1, 9]. The negative binomial (NB) distribution is such a traditional mixed Poisson distribution where the mean of the Poisson variable is distributed as a gamma random variable [12]. It is an increasingly popular alternative in modelling count data. However, the NB distribution may not be an appropriate model for delineating some over-dispersed incidents.

Other mixed Poisson distribution arises from alternative mixing distributions. It is found out that the general characteristics of the mixed Poisson distribution follow some characteristics of its mixing distribution. In this cognition, Sankaran [10] introduced the one-parameter discrete Poisson-Lindley distribution (PLD) by compounding Poisson distribution with the Lindley Distribution (LD) of Lindley [3]. This model showed better fits to over-dispersed count data. Later, Shanker and Mishra [17] generalised this mixed distribution by proposing a Two-parameter Poisson-Lindley (TPPL) distribution by assuming the Poisson rate to follow a Two-parameter LD [16]. Theoretical study and empirical observation have justified the selection of Lindley and its generalised distributions as mixing densities to expound over dispersed data.

In this paper, an alternative distribution for over-dispersed count data, namely a New Three-parameter Poisson-Lindley Distribution (NTPPLD) is presented which is obtained by mixing the Poisson distribution with a Three-parameter Lindley distribution [18]. The probability density function of a Three-parameter LD is a generalisation of the Lindley and a Two-parameter Lindley distribution and thus, reflects more flexibility as a mixing model. Contents of the paper are as follows: Section 1 gist's out a review of literature on the current subject together with a discussion on the need for Mixed Poisson distributions. In section 2, the NTPPLD distribution, is introduced. Some special cases of the distribution are also considered in this section. Its basic distributional properties including distribution function, generating functions, moments are derived in Section 3. Section 4 discusses the methods of parameter estimation and in section 5, a simulation study is carried out. Finally, applications of the NTPPLD to real datasets are illustrated in Section 6.

2 A New Three-Parameter Poisson-Lindley Distribution

In this section, a new mixed Poisson distribution is proposed, which is obtained by mixing the Poisson distribution with a Three-parameter Lindley distribution (ATPLD) [18]. A general definition of this distribution is provided which subsequently introduces its pmf.

Definition:2.1 A random variable X is said to follow a NTPPLD if it follows the stochastic representation X$|\lambda \sim$ P(λ) where $\lambda|\theta, \alpha, \beta \sim$ ATPLD(θ, α, β).[18]

Proposition:2.1 *Let X be a random variable according to the New Three-parameter Poisson-Lindley probability function, denoted by X∼NTPPLD(α, β, θ), then the pmf of X is*

$$p(x;\alpha,\beta,\theta) = \frac{\theta^2}{(\theta+1)^{x+2}}\left(1+\frac{\alpha+\beta x}{\theta\alpha+\beta}\right), \qquad (2.1)$$

$\forall\, x = 0, 1, ...;\ \theta > 0;\ \beta > 0;\ \theta\alpha + \beta > 0$

Proof Let X$|\lambda$ denote a random variable X following Poisson distribution with parameter λ. Accordingly, its probability mass function is given by

$$p(x) = \frac{e^{-\lambda}\lambda^x}{\Gamma(x+1)};\ x = 0, 1, 2, ...,\ \text{and}\ \lambda > 0. \qquad (2.2)$$

Again, the pdf of $\lambda|\theta, \alpha, \beta \sim ATPLD(\theta, \alpha, \beta)$ is given by

$$g(\lambda) = \frac{\theta^2}{\theta\alpha+\beta}(\alpha+\beta\lambda)e^{-\theta\lambda}; \lambda > 0,\ \theta > 0,\ \beta > 0,\ \theta\alpha+\beta > 0 \qquad (2.3)$$

The marginal pmf of $X \sim NTPPLD(\alpha, \beta, \theta)$, therefore, can be obtained by

$$p(x;\alpha,\beta,\theta) = \int_0^\infty p(x)g(\lambda)d\lambda \qquad (2.4)$$

Substituting (2.2) and (2.3) into (2.4), the marginal pmf of the New Three-Parameter Poisson-Lindley distribution is derived as:

$$p(x;\alpha,\beta,\theta) = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{\Gamma(x+1)}\frac{\theta^2}{(\theta\alpha+\beta)}(\alpha+\beta\lambda)e^{-\lambda\theta}d\lambda$$

$$= \frac{\theta^2}{(\theta+1)^{x+2}}\left(1+\frac{\alpha+\beta x}{\theta\alpha+\beta}\right)$$

The pmf of X satisfies the following properties:

I. $P(X = x) \geq 0, \forall x = 0, 1, 2, ...$

II. $\sum_x p(x;\alpha,\beta,\theta) = 1, \forall x = 0, 1, 2, ...$

2.1 Pmf Plots

Figure 1 illustrates the pmf plots of the NTPPLD (α, β, θ) for some selected values of the parameters. It is found out that the new distribution has a tendency to accommodate right tail and for particular values of the parameter, the tail tends to zero at a faster rate. This indicates that our proposed model fits appropriately to those data sets where there is an extended right tail or the tail approaches to zero at a faster rate. Such data sets are quite prevalent in the field of biology and insurance.

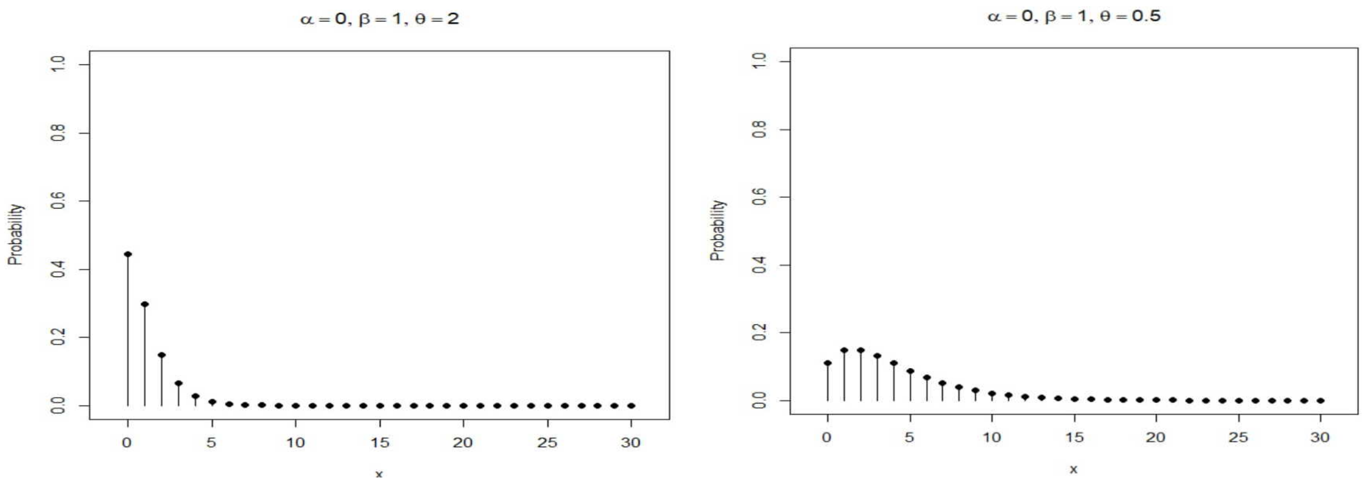

Figure 1: Some pmf plots of the NTPPL distribution with specified parameter values

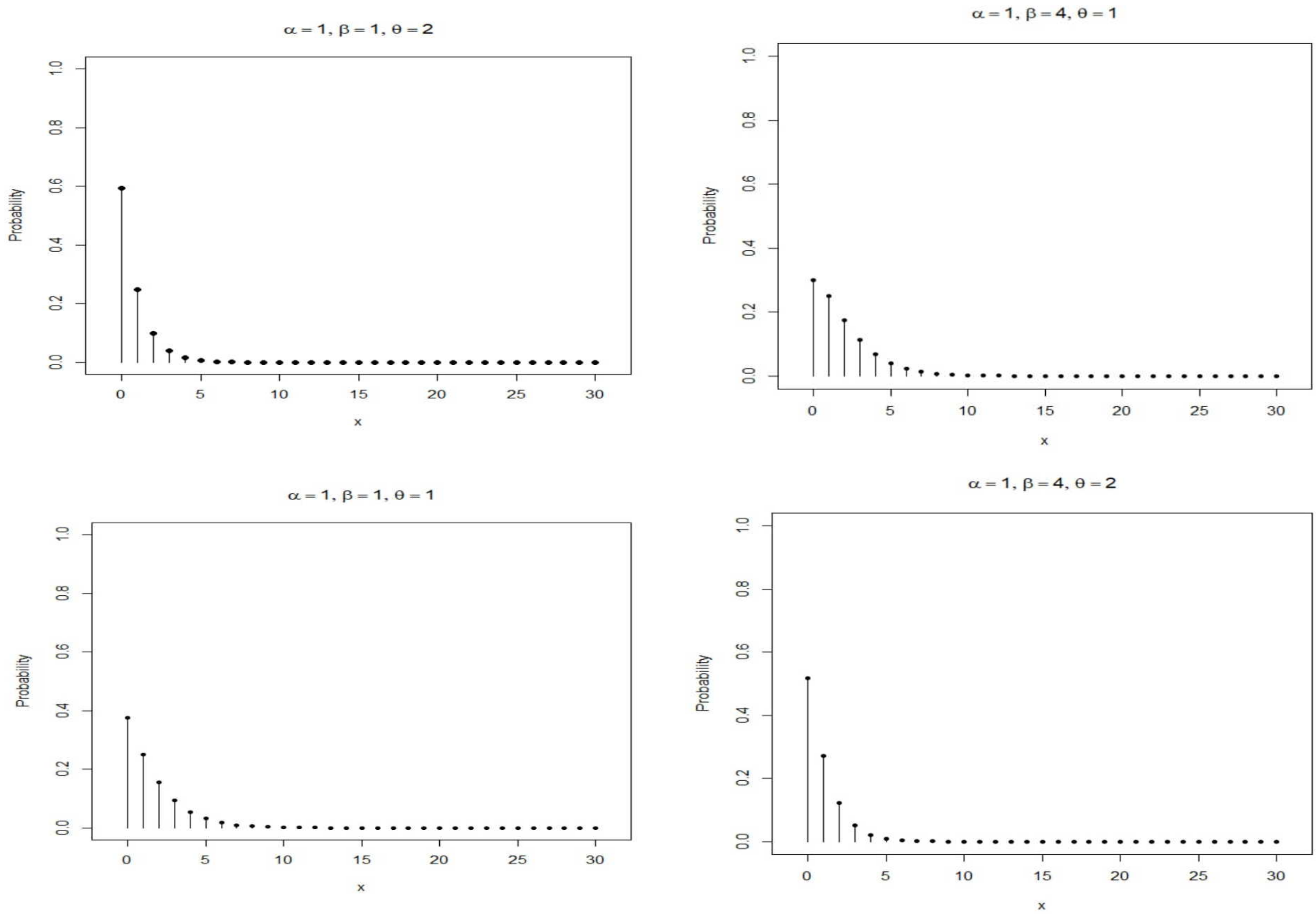

Figure 1: Some pmf plots of the NTPPL distribution with specified parameter values

2.2 Special Cases:

I. For $\beta = 1$, pmf (2.1) reduces to a Two-Parameter PLD[17].

II. For $\alpha = \beta = 1$, pmf (2.1) becomes the discrete Poisson Lindley distribution[10].

III. For $\alpha = 0$ and $\beta = 1$, pmf (2.1) reduces to NBD with parameters r=2 and $p = \frac{\theta}{\theta+1}$[12].

3 Distributional Properties of NTPPLD (θ, α, β)

In this section, expressions for the cumulative distribution function, generating functions, characteristic function, moments and other measures like coefficient of skewness, kurtosis, coefficient of variation and index of dispersion are derived.

3.1 Cumulative distribution function:

Proposition:3.1 *Let X be a random variable which follows New Three-parameter Poisson-Lindley probability function given by (2.1), then the cumulative distribution function of X is given by*

$$F_X(x;\alpha,\beta,\theta) = \frac{[\theta\{\alpha(\theta+1)+\beta\}\{(\theta+1)^{x+1}-1\}+\beta(\theta+1)\{(\theta+1)^x-1\}-x\beta\theta]}{(\theta+1)^{x+2}(\theta\alpha+\beta)}$$

Proof The cdf of NTPPLD (α, β, θ) can be obtained as

$$F_X(x;\alpha,\beta,\theta) = \sum_{n=0}^{x} \frac{\theta^2}{(\theta+1)^{n+2}}\left(1+\frac{\alpha+\beta n}{\theta\alpha+\beta}\right)$$

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

3.2 Generating functions:

Proposition 3.2 Let X be a New Three-parameter Poisson-Lindley (NTPPL) variable with parameters α, β and θ, then the Probability Generating Function of X denoted by $P_x(t)$ is given by

$$P_X(t) = \frac{\alpha(\theta+1-t)\theta^2+\theta^2\beta}{(\alpha\theta+\beta)(\theta+1-t)^2}$$

Proof The p.g.f of NTPPLD (α, β, θ) of (2.1) is obtained as

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

$$= \sum_{x=0}^{\infty} t^x \frac{\theta^2}{(\theta+1)^{x+2}}\left[1+\frac{\alpha+\beta x}{\theta\alpha+\beta}\right]$$

$$= \frac{\theta^2}{(\theta+1)^2}\sum_{x=0}^{\infty}\left(\frac{t}{\theta+1}\right)^x + \frac{\alpha\theta^2}{(\theta+1)^2(\theta\alpha+\beta)}\sum_{x=0}^{\infty}\left(\frac{t}{\theta+1}\right)^x$$

$$+ \frac{\beta\theta^2}{(\theta+1)^2(\theta\alpha+\beta)}\sum_{x=0}^{\infty} x\left(\frac{t}{\theta+1}\right)^x$$

$$= \frac{\alpha(\theta+1-t)\theta^2+\theta^2\beta}{(\alpha\theta+\beta)(\theta+1-t)^2}$$

Proposition:3.3 *Let X $\sim$ NTPPLD (α, β, θ).The moment generating function of X is given by*

$$M_X(t) = \frac{\alpha(\theta+1-e^t)\theta^2+\theta^2\beta}{(\alpha\theta+\beta)(\theta+1-e^t)^2}$$

Proof The moment generating function of the NTPPLD can be obtained by setting $t = e^t$ in the expression for the p.g.f.

3.3 Characteristic function

Proposition:3.4 *The Characteristic Function of X$\sim$ NTPPLD (θ, α, β) is given by*

$$\phi_X(t) = \frac{\alpha(\theta+1-e^{it})\theta^2+\theta^2\beta}{(\alpha\theta+\beta)(\theta+1-e^{it})^2}$$

Proof The characteristic function of the NTPPLD can be obtained by replacing $t = e^{it}$ in the expression for the p.g.f.

3.4 Raw Moments

Proposition:3.5 The first four raw moments about origin of the NTPPLD (2.1) are

$$\mu_1^{'} = \frac{\theta\alpha + 2\beta}{\theta(\theta\alpha + \beta)}$$

$$\mu_2^{'} = \frac{\theta\alpha + 2\beta}{\theta(\theta\alpha + \beta)} + \frac{2(\theta\alpha + 3\beta)}{\theta^2(\theta\alpha + \beta)}$$

$$\mu_3^{'} = \frac{\theta\alpha + 2\beta}{\theta(\theta\alpha + \beta)} + \frac{6(\theta\alpha + 3\beta)}{\theta^2(\theta\alpha + \beta)} + \frac{6(\theta\alpha + 4\beta)}{\theta^3(\theta\alpha + \beta)}$$

$$\mu_4^{'} = \frac{\theta\alpha + 2\beta}{\theta(\theta\alpha + \beta)} + \frac{14(\theta\alpha + 3\beta)}{\theta^2(\theta\alpha + \beta)} + \frac{36(\theta\alpha + 4\beta)}{\theta^3(\theta\alpha + \beta)} + \frac{24(\theta\alpha + 5\beta)}{\theta^4(\theta\alpha + \beta)}$$

Proof Let X follow pmf (2.2) and λ follow pdf (2.3).
The r^{th} moment about origin of the NTPPLD (2.1) can then be obtained as

$$\mu_r^{'} = E[E(X^r|\lambda)] = \int_0^\infty E(X^r)g(\lambda)d\lambda \tag{3.1}$$

Denoting Stirling's number of second kind by S(r,j) and pgf of (2.2) by $G_X(t)$, (3.1) may be written as

$$\begin{aligned} \mu_r^{'} &= \int_0^\infty \left[\sum_{j=1}^{r} S(r,j)\frac{d^j}{dt^j}G_X(t)|_{t=1}\right] g(\lambda)d\lambda \\ &= \sum_{j=1}^{r} S(r,j)\frac{\theta^2}{\theta\alpha + \beta}\left[\alpha\frac{\Gamma(j+1)}{\theta^{j+1}} + \beta\frac{\Gamma(j+2)}{\theta^{j+2}}\right] \end{aligned} \tag{3.2}$$

Taking r=1, in (3.2),

$$\mu_1^{'} = \frac{\theta\alpha + 2\beta}{\theta(\theta\alpha + \beta)}$$

Similarly, taking r=2,3 and 4 in (3.2), completes the proof of this theorem.

Remark:3.1

It can be seen that for $\beta = 1$ these moments reduce to the respective moments of the TPPLD [17]and for $\alpha = \beta = 1$, we get the corresponding moments of one-parameter PLD.[10]

3.5 Skewness and Kurtosis

The expressions for skewness and kurtosis are large and complicated; however, their values for different values of the parameters are determined and presented in Table 1.

Remark:3.2

I. For fixed β and θ, as α increases, both skewness and kurtosis increase.
II. When α and θ is kept fixed, skewness and kurtosis decrease with an increase in β (except when $\alpha = 0$).
III. As θ increases, there is increase in skewness and kurtosis.

Table 1: Skewness and (Kurtosis)

	$\alpha=0$			$\alpha=1$			$\alpha=2$		
	$\theta=0.5$	$\theta=1$	$\theta=2$	$\theta=0.5$	$\theta=1$	$\theta=2$	$\theta=0.5$	$\theta=1$	$\theta=2$
$\beta=0.5$	1.4 (6.1)	1.5 (6.2)	1.6 (6.7)	1.7 (7.2)	1.9 (8.2)	2.2 (9.6)	1.8 (7.8)	2.0 (8.8)	2.2 (10.0)
$\beta=1$	1.4 (6.1)	1.5 (6.2)	1.6 (6.7)	1.6 (6.7)	1.8 (7.5)	2.1 (8.9)	1.7 (7.2)	1.9 (8.2)	2.2 (9.6)
$\beta=2$	1.4 (6.1)	1.5 (6.2)	1.6 (6.7)	1.5 (6.4)	1.7 (7.0)	2.0 (8.2)	1.6 (6.7)	1.8 (7.5)	2.1 (8.9)

3.6 Coefficient of variation and index for dispersion

Proposition:3.6 *The coefficient of variation of New Three-parameter PLD is given by*

$$C.V=\sqrt{1+\frac{\theta(\theta\alpha+\beta)}{\theta\alpha+2\beta}-\frac{2\beta^2}{(\theta\alpha+2\beta^2)}}$$

Proof The mean of the NTPPLD (θ,α,β) is

$$\mu_1^{'}=\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}$$

The variance is obtained using the relation

$$\mu_2=\mu_2^{'}-(\mu_1^{'})^2=\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}+\frac{2(\theta\alpha+3\beta)}{\theta^2(\theta\alpha+\beta)}-\left\{\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}\right\}^2$$

Therefore, the C.V is given by

$$C.V=\sqrt{\frac{\mu_2}{(\mu_1^{'})^2}}$$

$$=\frac{\sqrt{\mu_2}}{\mu_1^{'}}$$

$$=\sqrt{\frac{\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}+\frac{2(\theta\alpha+3\beta)}{\theta^2(\theta\alpha+\beta)}-\left\{\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}\right\}^2}{\left\{\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}\right\}^2}}$$

$$=\sqrt{1+\frac{\theta(\theta\alpha+\beta)}{\theta\alpha+2\beta}-\frac{2\beta^2}{(\theta\alpha+2\beta)^2}}$$

Proposition:3.7 *The index for dispersion is given by*

$$r=1+\frac{\theta^2+4\theta\alpha\beta+2\beta^2}{\theta(\theta\alpha+\beta)(\theta\alpha+2\beta)}$$

Proof The mean of the NTPPLD (θ,α,β) is

$$\mu_1^{'}=\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}$$

The variance is given by

$$\mu_2=\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}+\frac{2(\theta\alpha+3\beta)}{\theta^2(\theta\alpha+\beta)}-\left\{\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}\right\}^2$$

Therefore, the index of dispersion is given by

$$r=\frac{\mu_2}{\mu_1^{'}}$$

$$=1+\frac{\theta^2\alpha^2+4\theta\alpha\beta+2\beta^2}{\theta(\theta\alpha+\beta)(\theta\alpha+2\beta)}$$

Remark:3.3 The variance-to-mean ratio is greater than one. Therefore, the proposed distribution is over-dispersed.

4 Estimation of Parameters

This section discusses two widely used methods of estimation of the parameters viz. Method of Moments and Maximum Likelihood Method for the NTPPLD.

4.1 Moments Estimates:

The NTPPLD (θ,α,β) has three parameters to be estimated and so, the first three moments are required to get their estimates.
From section **3.4**, we have,

$$\mu_1^{'}=\frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)}=K_1(say) \tag{4.1}$$

$$\frac{\mu_2^{'}-\mu_1^{'}}{\mu_1^{'}}=\frac{2(\theta\alpha+3\beta)}{\theta(\theta\alpha+2\beta)}=K_2(say) \tag{4.2}$$

$$\frac{\mu_3^{'}-3\mu_2{'}+2\mu_1^{'}}{\mu_2^{'}-\mu_1^{'}}=\frac{3(\theta\alpha+4\beta)}{\theta(\theta\alpha+3\beta)}=K_3(say) \tag{4.3}$$

Replacing the population moments by their respective sample moments in (4.1), (4.2) and (4.3), an estimate of K_1,K_2 and K_3 can be obtained. Using them, we can solve the system of equations for θ,α,β to obtain its moments estimates.

4.2 Maximum Likelihood Estimates:

Let $x_1,x_2,...,x_n$ be a random sample of size n from our proposed NTPPLD (θ,α,β) . The likelihood function for the vector of parameters ϑ= $(\theta,\alpha,\beta)^T$ of this sample is given by:

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

Accordingly, the log-likelihood function can be written as:

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

The maximum likelihood estimates (MLE) of the parameters are computed by solving the maximum likelihood equations

$$\frac{\partial}{\partial\theta}\log L = 0, \quad \frac{\partial}{\partial\alpha}\log L = 0, \quad \frac{\partial}{\partial\beta}\log L = 0$$

Since the maximum likelihood equations are non-linear in nature and difficult to solve analytically, they are to be solved using some suitable numerical technique.

A detailed simulation algorithm for generating data from (2.1) and obtaining the MLE for ϑ is enclosed in Section 5.

5 Simulation study

In this section, simulation is carried out to generate random variables from NTPPLD (θ, α, β). Thereafter, the MLE of the parameters are obtained from the generated sample. Finally, the Bias and mean squared error (MSE) of the MLE of the parameters are calculated to assess the consistency of the estimates.

Enumerated below, is the algorithm for the desired simulation study:

Step I: Generating a random sample from NTPPLD (θ, α, β).

Step 1: A random variable is generated from the U(0,1) distribution, say u.

Step 2: If the $x_i, i \geq 0$, are ordered so that $x_0 < x_1 < x_2 < ..$ and if we let F, as defined in section **3.1**, denote the distribution function of X.

Then, $X = x_i$ if $F(x_i - 1) \leq u < F(x_i)$,$i = 0, 1, 2, ...$

Steps 1 and *2* are repeated as many times as the desirable sample size is.

The above method is called the discrete inverse transform method for generating X.

Step II: Obtaining MLE of the parameters.

The MLE of θ, α and β is obtained by solving the maximum likelihood equations for the generated sample procured in the previous step.

Step III: Calculating Bias and MSE of the MLE's.

Suppose that the true value of the parameter θ is θ_0 and the MLE is θ^*. Then the Bias of θ^* in estimating θ_0 is given by

$$Bias(\theta^*) = E(|\theta^* - \theta_0|)$$

The expectation being with respect to the mass function of NTPPLD (θ, α, β).

Similarly, the MSE of θ_0 is obtained as

$$MSE(\theta^*) = E[(\theta^* - \theta_0)^2]$$

The Bias and MSE of the MLE of θ is approximated by the *Monte Carlo* approximation technique, by taking M=1000 replicates. In a similar manner, the Bias and MSE of the MLE of α and β are calculated. The MLE is said to be consistent if the Bias decreases (approaches to zero) with an increase in the sample size and so does the MSE. Table 2 shows the values of the Bias and MSE of the MLE of θ, α and β for the different sample sizes. From Table 2, it is seen that the Bias and MSE of θ, α and β approaches towards zero with an increase in the sample size. Thus, it can be concluded that they are consistent and precise in estimating the true value of the parameters. Calculations pertaining to the study are carried out using the **R** software, version *3.4.3*, with the help of self-programmed codes. The *maxLik package*(Henningsen, Arne and Toomet, Ott, 2011) in **R** software is used to obtain the maximum likelihood estimates of the parameters from NTPPLD (θ, α, β).

Table 2: Average values of bias and MSE for θ^*,$\alpha^*$and β^*

Sample Size	θ=3		$\alpha = 4$		$\beta = 2$	
	Bias (θ)	MSE (θ)	Bias (α)	MSE (α)	Bias (β)	MSE (β)
50	0.64619	0.51673	1.87777	3.52602	1.20045	1.46335
100	0.58956	0.37721	1.60015	2.65715	1.19996	1.44537
200	0.238507	0.06977	1.03593	1.23732	0.69749	0.50399
500	0.21259	0.05419	0.90122	0.82120	0.49964	0.25864

6 Application to Real Data Sets

Some real data sets are considered in this section to fit with the proposed distribution (NTPPLD) along with One and Two-parameter Poisson Lindley distributions. The first data set is due to Beall [6] regarding the distribution of Pyrausta nublilalis in 1937. It is reproduced in Table 4. The second is due to Kemp and Kemp [2] which records the distribution of mistakes in copying of random digits and is displayed in table 5. Finally, the third data set, as illustrated in table 6, is the number of claims in automobile insurance observed by Klugman et al [22]. Descriptive summaries of these data are shown in Table 3 and it is seen that the index of dispersion for these data sets are greater than unity thereby indicating over-dispersion.

Table 3: Summary data

	Mean	Variance	Dispersion Index
Number of Pyrausta nublilalis	0.7500	1.3182	1.7576
Number of mistakes in copying groups	0.7833	1.2573	1.6051
Number of claims in automobile insurance	0.1941	0.2259	1.1638

The expected frequencies according to the One-parameter PLD and Two-parameter PLD are given in the following tables for ready comparison with those obtained by the New Three-parameter PLD. Parameters for the NTPPLD have been estimated by maximum likelihood method and for those of the other competing models by method of moments because of their existing availability.

Table 4: Distribution of Pyrausta nublilalis in 1937

No. of Insects	Observed Frequency	Expected Frequency		
		One-parameter PLD	Two-parameter PLD	New Three-parameter PLD
0	33	31.5	31.9	32.0
1	12	14.2	13.8	13.7
2	6	6.1	5.9	5.9
3	3	2.5	2.5	2.5
4	1	1.0	1.1	1.1
5	1	0.4	0.5	0.5
Parameter Estimates		$\hat{\theta} = 1.8081$	$\hat{\theta} = 1.5255$	$\hat{\theta} = 1.3970$
			$\hat{\alpha} = 3.8919$	$\hat{\alpha} = 8.1694$
				$\hat{\beta} = 0.5744$
RMSE		1.13	0.91	0.86

Table 5: Distribution of mistakes in copying groups of random digits

No. of errors per group	Observed Frequency	Expected Frequency		
		One-parameter PLD	Two-parameter PLD	New Three-parameter PLD
0	35	33.0	32.4	33.2
1	11	15.3	15.9	15.1
2	8	6.7	7.0	6.7
3	4	2.9	2.9	2.9
4	2	1.2	1.1	1.2
Parameter Estimates		$\hat{\theta} = 1.7421$	$\hat{\theta} = 2.0000$	$\hat{\theta} = 1.6322$
			$\hat{\alpha} = 0.3824$	$\hat{\alpha} = 4.5289$
				$\hat{\beta} = 2.8890$
RMSE		2.28	2.60	2.17

Table 6: Number of claims of policy holders in automobile insurance

Claim Count	Observed Frequency	Expected Frequency		
		One-parameter PLD	Two-parameter PLD	New Three-parameter PLD
0	1563	159.5	1566.5	1564.7
1	271	256.3	261.1	264.2
2	32	41.3	40.4	39.7
3	7	6.6	6.0	5.6
4	2	1.0	0.9	0.8
Parameter Estimates		$\hat{\alpha}$ =5.8979	$\hat{\theta}$ =7.1855	$\hat{\theta}$ =7.8580
			$\hat{\alpha}$ =0.2132	$\hat{\alpha}$ =1.1019
				$\hat{\beta}$ = 9.5877
RMSE		8.32	6.05	4.73

It is seen from these data sets that the Root Mean Square error (RMSE) statistic is least when modelled using the New Three-parameter PLD. This substantiates our claim that the proposed distribution gives much closer fit than the One and Two-parameter Poisson-Lindley distributions while describing count data.

7 Conclusions

In this paper, a New Three-parameter Poisson-Lindley discrete distribution (NTPPLD) with an infinite and non-negative integer support is proposed. The behaviour of the probability mass function for varying values of the parameter is studied. Distributional properties and common descriptive measures pertaining to this mixed distribution are derived. Estimation of parameters by the method of maximum likelihood and method of moments are discussed and a simulation study is carried out to check the consistency of the maximum likelihood estimates. From the results obtained, they are found to be consistent and precise in estimating the true value of the parameters. Finally, the proposed distribution is applied to real-data sets and it is shown that this distribution is a flexible model that may be a useful alternative to known distributions like Poisson Lindley, Two-parameter Poisson Lindley etc. for over-dispersed count data analysis.

8 Acknowledgements

The authors acknowledge the financial support provided by the Department of Science and Technology (DST), Government of India, to the second author through the INSPIRE programme for carrying out this research work.

References

[1] B. Everitt, and D. Hand, *Finite mixture distributions. Monographs on Statistics and Applied Probability*, Chapman and Hall, U.K (1981).

[2] C.D. Kemp, and A.W. Kemp, Some properties of the Hermite distribution, *Biometrika*, **52** (1965), 381-394.

[3] D.V. Lindley ,Fiducial distributions and Bayes theorem, *Journal of Royal Statistical Society*, **Ser. B Vol.20** (1958), 102-107.

[4] D. Karlis, and E. Xekalaki ,Mixed Poisson distributions, *International Statistical Review / Revue Internationale de Statistique*, **73(1)** (2005), 35-58.

[5] D. Bhati ,D.V. Sastry ,P.Z.M. Qadri , A New Generalized Poisson-Lindley Distribution: Applications and Properties, *Austrian Journal of Statistics*, **44** (2015), 35-51.

[6] G. Beall, The fit and significance of contagious distributions when applied to observations on larval insects, *Ecology*, **21** (1940), 460-474.

[7] H. H. Panjer,Mixed Poisson Distributions, In: *Encyclopedia of Actuarial Science*, John Wiley and Sons, Ltd. Hoboken, New Jersey, U.S.A (2006).

[8] J. C. Nash, On best practice optimization methods in R, *Journal of Statistical Software*, **60(2)** (2014), 1-14.

[9] M. Raghavachari ,A. Srinivasan , and P. Sullo , Poisson mixture yield models for integrated circuits: A critical review., *Microelectronics Reliability*, **37(4)** (1997),565-580.

[10] M. Sankaran, The Discrete Poisson-Lindley Distribution, *Biometrics*, **26 (1)** (1970),145-149.

[11] M.E. Ghitany and D.K. Al-Mutairi ,Estimation Methods for the Discrete Poisson-Lindley Distribution, *Journal of Statistical Computation and Simulation*, **79 (1)** (2009), 1-9.

[12] M. Greenwood, and G. U. Yule, , An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, *Journal of the Royal Statistical Society*, **83(2)** (1920), 255-279.

[13] M.E. Ghitany ,B. Atieh and S. Nadarajah , Lindley Distribution and Its Application, *Mathematics and Computers in Simulation*, **78** (2008), 493-506.

[14] N.L. Johnson , S. Kotz and A.W. Kemp , *Univariate Discrete Distributions*, Second Edition. John Wiley and Sons (2005).

[15] R Core Team , *R: A Language and Environment for Statistical Computing*,R Foundation for Statistical Computing, Vienna, Austria (2014).

[16] R. Shanker and A. Mishra,A two-parameter Lindley distribution, *Statistics in Transition New Series*, **Vol. 14, No.1 pp** (2013), 45-56.

[17] R. Shanker and A. Mishra ,A two-parameter Poisson-Lindley distribution, *International Journal of Statistics and Systems* , **Vol. 9, No.1 pp.** (2014),79-85.

[18] R. Shanker , K. K. Shukla,R. Shanker ,T. A. Leonida,A Three-Parameter Lindley Distribution, *American Journal of Mathematics and Statistics*, **7(1)** (2017),15-26.

[19] S. Chakraborty, On Some Distributional Properties of the Family of Weighted Generalized Poisson Distribution, *Communications in Statistics - Theory and Methods*, **39 (15)** (2010), 2767-2788.

[20] S. Nadarajah and S. Kotz , Compound Mixed Poisson Distributions I, *Scandinavian Actuarial Journal*, **3** (2006a), 141-162.

[21] S. Nadarajah and S. Kotz, Compound Mixed Poisson Distributions II, *Scandinavian Actuarial Journal*, **3** (2006b), 163-181.

[22] S. Klugman,H. Panjer ,G. Willmot , *Loss Models. From Data to Decisions*, John Wiley and Sons, New York (2012).

[23] W. Wongrin , W. Bodhisuwan ,The Poisson-generalised Lindley distribution and its applications, *Songklankarin J. Sci. Technol*, **38(6)** (2016),645-66.