

## NEGATIVE BINOMIAL-IMPROVED SECOND DEGREE LINDLEY DISTRIBUTION AND ITS APPLICATION

R. ASHLY AND C. S. RAJITHA<sup>1</sup>

**ABSTRACT.** The objective of this paper is to introduce a new two parameter mixed negative binomial distribution, namely negative binomial-improved second degree Lindley(NB-ISL) distribution. This distribution is obtained by mixing the negative binomial distribution with the improved second degree Lindley distribution. Many mixed distributions have been used in the literature for modeling the over dispersed count data, which provide a better fit compared to the Poisson and negative binomial distribution. In addition, we present the basic statistical properties of the new distribution such as factorial moments, mean and variance and the behavior of mean, variance and coefficient of variation are also discussed. Parameter estimation is implemented by using maximum likelihood estimation method. The performance of the NB-ISL distribution is shown in practice by applying it on real data set and compare it with some well-known count distributions. The result shows that the negative binomial-improved second degree Lindley distribution provides a better fit compared to Poisson, negative binomial and negative binomial-Lindley distributions.

### 1. INTRODUCTION

One of the major issues faced in many research activities is the thorough monitoring and flexible modeling of count data. There are many examples of

---

<sup>1</sup>*corresponding author*

2010 *Mathematics Subject Classification.* 46N30, 47N30.

*Key words and phrases.* count data, mixed negative binomial distribution, overdispersion, negative binomial-improved second degree Lindley distribution.

count data such as the number of insurance claims, the number of accidents on some particular streets, number of crimes on a campus and so on. Poisson distribution is the most popular distribution used for modeling the count data. It is a discrete distribution for counting the number of occurrence of the event that may occur randomly in a given period of time. Equality of mean and variance called equal dispersion is a characteristic of Poisson distribution, but in practice, the observed data shows overdispersion with variance greater than mean. One of the main reason for overdispersion is due to the incidence of more zeros which is greater than expected for the Poisson distribution. As the number of occurrence of zeros increases the variance of the random variable also increase and become greater than the mean. In this case we need to adopt a more flexible form to deal with the overdispersion in the data which is an extension of the Poisson distribution.

The negative binomial distribution (NB) is a popular alternative to Poisson distribution for modeling the overdispersed count data. The NB distribution is a generalization of the Poisson distribution in which the distribution's parameter itself considered as a random variable which follows a gamma distribution. The variation in this parameter can account for the overdispersion in the data. Even though the NB distribution allows overdispersion it does not care for excess number of zeros. For such cases the traditional statistical distributions such as Poisson and NB distribution cannot be used efficiently.

The problem of overdispersion and excess number of zeros is solved by using mixed Poisson or mixed NB distribution. Several examples of mixed Poisson and mixed negative binomial distribution can be found in statistical literature including the discrete Poisson-Lindley Distribution proposed by Sankaran, see [6], which is an extended version of the compound Poisson distribution obtained by compounding the Poisson distribution with Generalized Lindley distribution. Recently Zamani et al., see [8] introduced a new discrete distribution which is a weighted version of the Poisson-Lindley distribution. These are some Poisson mixture distributions used for modeling count data. It is already shown that the NB mixed distribution provides better fit compared to Poisson mixture distribution. The negative binomial-Lindley was proposed by Zamani and Ismail [7], which provide a better fit compared to Poisson and the negative binomial for count data where the probability at zero has a large value

and applied in many count data analysis. The Lindley distribution was proposed by Ghitany. Ghitany, Atich and Nadarajah [2] shows that in many ways the Lindley distribution is a better model compared to exponential distribution. Saengthong and Bodhisuwan [5] proposed a mixture of negative binomial and crack distribution which contains special cases namely, negative binomial-inverse Gaussian (NB-IG), negative binomial-Birnbaum-Saunders (NB-BS) and negative binomial-length biased inverse Gaussian (NB-LBIG). Aryuyuen and Bodhisuwan [1] introduced the negative binomial-generalized exponential distribution (NB-GE) which provide the advantage of handling the overdispersed count data. Lord and Geedipally [3] analyze crash data containing large number of zeros using negative binomial-Lindley distribution. They showed that the NB-L performs better than NB for data with large number of zeros.

In this paper we introduced a new mixed NB distribution, namely negative binomial-improved second degree Lindley (NB-ISL) distribution obtained by mixing the NB distribution with improved second degree Lindley distribution (Sadasivan, Vinoth and Keerthana, see [4]). Properties of the NB-ISL distribution including the factorial moment, mean, variance and parameter of the NB-ISL is estimated by using maximum likelihood estimation method and provides a comparison to show that the NB-ISL provide a better fit compared to the existing traditional distributions. In section 2 we discuss about NB distribution. Section 3 is concerned with the improved second degree Lindley distribution and we present the new distribution. Section 4 discuss about the properties of the NB-ISL including factorial moments, mean and variance. Parameter estimation is given in section 5. In section 6, the usefulness of the proposed distribution is discussed. Finally conclusion is given in section 7.

## 2. NEGATIVE BINOMIAL DISTRIBUTION

The NB distribution is suitable in cases where the data is over dispersed, that is variance is greater than mean. If  $X$  is a random variable which follows NB distribution with parameters  $r$  and  $p$ , then its probability mass function (PMF) is

$$p(x) = \binom{r+x-1}{x} p^r (1-p)^x; \quad x = 0, 1, 2, \dots$$

The first two moments about zero and the factorial moment of order  $k$  of NB distribution are:

$$E(X) = \frac{r(1-p)}{p},$$

$$E(X^2) = \frac{r(1-p)[1+r(1-p)]}{p^2},$$

$$\mu_{[k]}(x) = \frac{\Gamma(r+k)}{\Gamma(r)} \frac{(1-p)^k}{p^k} \quad (k = 0, 1, 2, \dots).$$

### 3. IMPROVED SECOND DEGREE LINDLEY DISTRIBUTION

Improved second degree Lindley distribution (ISLD) is a one parameter continuous distribution. The probability density function PDF of ISLD is

$$f(x) = \frac{\lambda^3}{\lambda^2 + 2\lambda + 2} (1+x)^2 e^{-\lambda x}; \quad x > 0, \quad \lambda > 0.$$

The moment generating function MGF of ISLD is

$$M_X(t) = \frac{\lambda^3}{\lambda^2 + 2\lambda + 2} \left[ \frac{1}{\lambda - t} + \frac{2}{(\lambda - t)^2} + \frac{2}{(\lambda - t)^3} \right].$$

**Definition 3.1.** Let  $X$  be a random variable which follows the negative binomial-improved second degree Lindley distribution with parameters  $r$  and  $\lambda$ ,  $X \sim NB - ISL(r, \lambda)$ , when the NB distribution have parameters  $r > 0$  and  $p = e^{-a}$  where  $a$  is distributed as ISLD with parameter  $\lambda$ , which can be expressed as  $X|a \sim NB(r, p = e^{-a})$  and  $a \sim ISL(\lambda)$ .

**Theorem 3.1.** Let  $X \sim NB - ISL(r, \lambda)$  be a negative binomial-improved second degree Lindley distribution as defined in definition 3.1, then the PMF of  $X$  is:

$$h(x) = \left\{ \frac{\lambda^3}{\lambda^2 + 2\lambda + 2} \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \right. \\ \left. \times \left[ \frac{1}{\lambda + r + j} + \frac{2}{(\lambda + r + j)^2} + \frac{2}{(\lambda + r + j)^3} \right] \right\},$$

$$x = 0, 1, 2, \dots, r, \lambda > 0.$$

*Proof.* If  $X|a \sim NB(r, p = e^{-a})$  and  $a \sim ISL(\lambda)$  then PMF of  $X$  can be obtained by

$$(3.1) \quad h(x) = \int_0^\infty p(x|a) f(a; \lambda) da,$$

where

$$p(x|a) = \binom{r+x-1}{x} e^{-ar} (1 - e^{-a})^x.$$

That is,

$$(3.2) \quad p(x|a) = \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j e^{-a(r+j)},$$

and  $f(a; \lambda)$  is the probability density function of improved degree Lindley distribution.

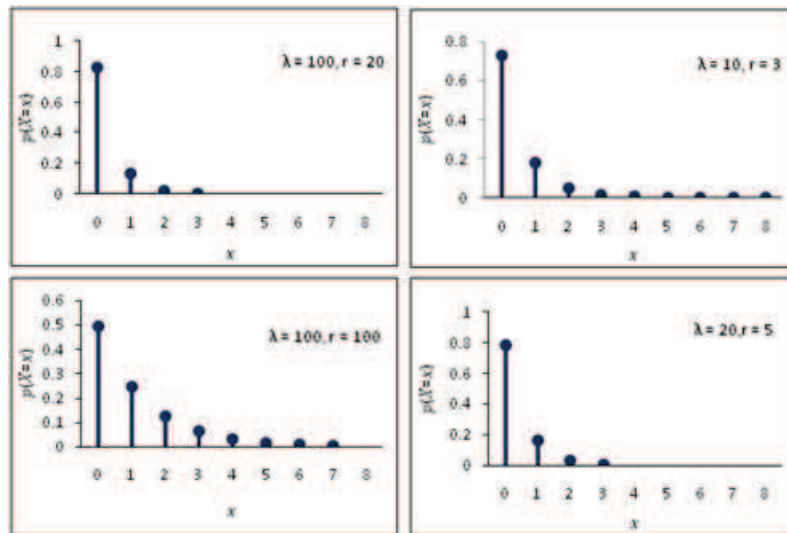


FIGURE 1. The PMF of the NB-ISL distribution for various values of parameters

Substitute (3.2) in (3.1) we get

$$\begin{aligned}
 h(x) &= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \int_0^\infty e^{-a(r+j)} f(a; \lambda) da \\
 &= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_a(-(r+j)) \\
 &= \frac{\lambda^3}{\lambda^2 + 2\lambda + 2} \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \\
 &\quad \times \left[ \frac{1}{\lambda + r + j} + \frac{2}{(\lambda + r + j)^2} + \frac{2}{(\lambda + r + j)^3} \right]
 \end{aligned}$$

□

#### 4. PROPERTIES OF NB-ISL DISTRIBUTION

This section gives some properties of the new proposed mixed distribution such as factorial moments, mean and variance.

**Theorem 4.1.** *If  $X \sim NB - ISL(r, \lambda)$ , then the factorial moment of order  $k$  of  $X$  is given by*

$$\begin{aligned}
 \mu_{[k]}(X) &= \left( \frac{\lambda^3}{\lambda^2 + 2\lambda + 2} \frac{\Gamma(r+k)}{\Gamma(r)} \sum_{j=0}^k \binom{k}{j} (-1)^j \right. \\
 (4.1) \quad &\quad \left. \times \left[ \frac{1}{\lambda - k + j} + \frac{2}{(\lambda - k + j)^2} + \frac{2}{(\lambda - k + j)^3} \right] \right).
 \end{aligned}$$

*Proof.* If  $X|a \sim NB(r, p = e^{-a})$  and  $a \sim ISL(\lambda)$ , then the factorial moment of the order  $k$  of  $X$  can be obtained by:

$$\mu_{[k]}(X) = E_a[\mu_k(X|a)].$$

The factorial moments of order  $k$  of a mixed NB distribution where  $p = e^{-a}$  is

$$\begin{aligned}
 \mu_{[k]}(X) &= E_a \left[ \frac{\Gamma(r+k)}{\Gamma(r)} \frac{(1 - e^{-a})^k}{e^{-ak}} \right] \\
 &= \frac{\Gamma(r+k)}{\Gamma(r)} E_a(e^a - 1)^k.
 \end{aligned}$$

Using binomial expansion for the term  $(e^a - 1)^k$  we can write

$$\begin{aligned}\mu_{[k]}(X) &= \frac{\Gamma(r+k)}{\Gamma(r)} \sum_{j=0}^k \binom{k}{j} (-1)^j E(e^{a(k-j)}) \\ &= \frac{\Gamma(r+k)}{\Gamma(r)} \sum_{j=0}^k \binom{k}{j} (-1)^j M_a(k-j) \\ &= \frac{\lambda^3}{\lambda^2 + 2\lambda + 2} \frac{\Gamma(r+k)}{\Gamma(r)} \sum_{j=0}^k \binom{k}{j} (-1)^j \\ &\quad \times \left[ \frac{1}{\lambda - k + j} + \frac{2}{(\lambda - k + j)^2} + \frac{2}{(\lambda - k + j)^3} \right]\end{aligned}$$

□

The mean and second order moment can be derived from (4.1) which are given by

$$\begin{aligned}E(X) &= r \left[ \frac{\lambda^3(\lambda^2 + 1)}{(\lambda - 1)^3(\lambda^2 + 2\lambda + 2)} - 1 \right] \\ E(X^2) &= \left[ r(r+1) \frac{\lambda^3(\lambda^2 - 2\lambda + 2)}{(\lambda - 2)^3(\lambda^2 + 2\lambda + 2)} \right. \\ &\quad \left. - (2r^2 + r) \frac{\lambda^3(\lambda^2 + 1)}{(\lambda - 1)^3(\lambda^2 + 2\lambda + 2)} + r^2 \right].\end{aligned}$$

Table 1 summarizes the behavior of mean, variance and coefficient of variation (CV) of NB-ISL distribution for varying values of the parameters  $r$  and  $\lambda$ . It is clear from table 1 and figure 2 that mean and variance are decreasing for increasing values of the parameter  $\lambda$  and mean and variance are increasing for increasing value of the parameter  $r$ . Since the coefficient of variation is greater than 1, the distribution is suitable for over dispersed count data. C.V initially decreases for increasing value of the parameter  $\lambda$  and afterwards increases when  $\lambda$  reaches a certain value. For increasing value of  $r$ , C.V decreases.

## 5. PARAMETER ESTIMATION

In this section we estimated the parameters of the negative binomial-improved second degree Lindley distribution by using the method of moments and the method of maximum likelihood estimation.

TABLE 1. Mean, variance and CV of NB-ISL distribution for various parameter values

Mean						
$r \backslash \lambda$	3	4	5	6	7	8
3	2.9558	1.6495	1.1173	0.8361	0.6645	0.5497
4	3.9411	2.1994	1.4898	1.1148	0.886	0.7329
5	4.9264	2.7492	1.8623	1.3936	1.1075	0.9162
6	5.9117	3.2991	2.2347	1.6723	1.329	1.0994
7	6.8970	3.849	2.6072	1.9510	1.5505	1.2827
8	7.8823	4.3988	2.9797	2.2297	1.772	1.4659
Variance						
$r \backslash \lambda$	3	4	5	6	7	8
3	53.866	10.655	4.455	2.508	1.650	1.194
4	87.820	16.906	6.914	3.823	2.48	1.773
5	129.774	24.507	9.86	5.378	3.449	2.442
6	179.728	33.458	13.293	7.173	4.559	3.202
7	237.681	43.758	17.212	9.208	5.807	4.053
8	303.633	55.408	21.619	11.482	7.196	4.994
Coefficient of Variation						
$r \backslash \lambda$	3	4	5	6	7	8
3	2.4829	1.9788	1.8889	1.8938	1.9331	1.9874
4	2.3777	1.8694	1.7648	1.7538	1.7773	1.8164
5	2.3123	1.8006	1.686	1.6641	1.6768	1.7056
6	2.2677	1.7532	1.6314	1.6015	1.6064	1.6276
7	2.2352	1.7186	1.5912	1.5552	1.5541	1.5695
8	2.2106	1.6921	1.5604	1.5196	1.5137	1.5244

**5.1. Method of Moments.** For the method of moments, parameters in the PMF are estimated by equating the sample and population moments. In the case of  $NB - ISL(r, \lambda)$ , the first two moments are required for estimating  $r$



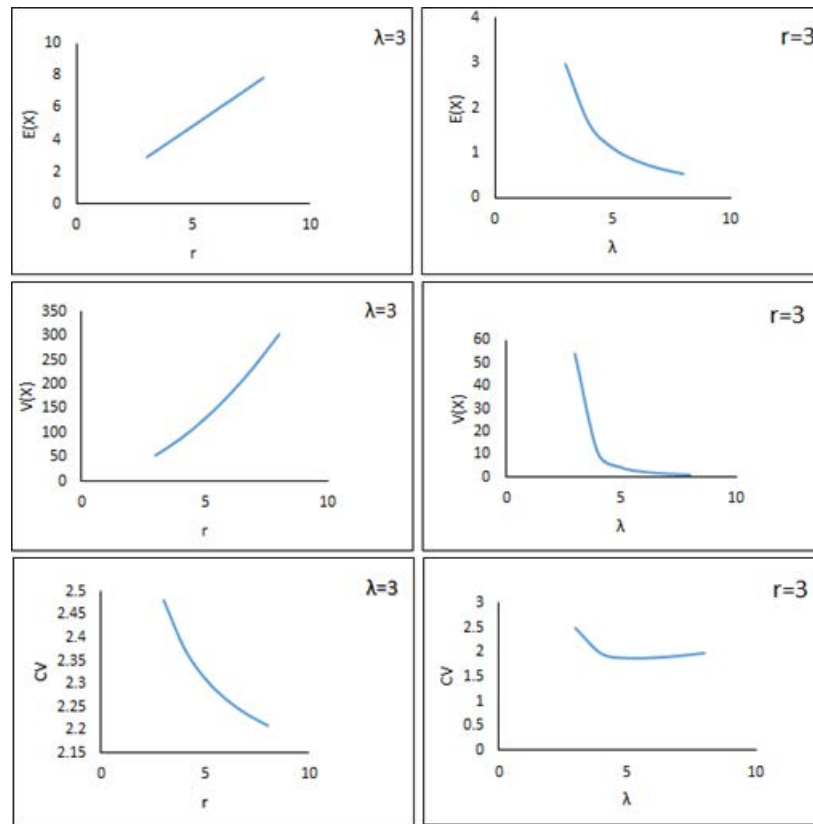


FIGURE 2. Behavior of mean, variance and CV for various values of parameters

and  $\lambda$

$$m_1 = r \left[ \frac{\lambda^3(\lambda^2 + 1)}{(\lambda - 1)^3(\lambda^2 + 2\lambda + 2)} - 1 \right]$$

and

$$m_2 = \left[ r(r+1) \frac{\lambda^3(\lambda^2 - 2\lambda + 2)}{(\lambda - 2)^3(\lambda^2 + 2\lambda + 2)} - (2r^2 + r) \frac{\lambda^3(\lambda^2 + 1)}{(\lambda - 1)^3(\lambda^2 + 2\lambda + 2)} + r^2 \right]$$

By solving equations (5.1) and (5.1), the moment estimates of the two parameters  $r$  and  $\lambda$  can be obtained.

**5.2. Method of Maximum Likelihood.** The likelihood function of the  $NB - ISL(r, \lambda)$  is given by

$$\begin{aligned} \mathcal{L}(x; r, \lambda) &= \frac{\lambda^{3n}}{(\lambda^2 + 2\lambda + 2)^n} \prod_{i=1}^n \binom{r + x_i - 1}{x_i} \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \\ &\quad \times \left[ \frac{1}{\lambda + r + j} + \frac{2}{(\lambda + r + j)^2} + \frac{2}{(\lambda + r + j)^3} \right] \end{aligned}$$

From which we calculate the log-likelihood function

$$\begin{aligned} \ell(x; r, \lambda) &= 3n \log(\lambda) - n \log(\lambda^2 + 2\lambda + 2) \\ &\quad + \sum_{i=1}^n \log(r + x_i - 1)! - \log x_i! - \log(r - 1)! \\ &\quad + \sum_{i=1}^n \log \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left[ \frac{1}{\lambda + r + j} + \frac{2}{(\lambda + r + j)^2} + \frac{2}{(\lambda + r + j)^3} \right]. \end{aligned}$$

The first order conditions for finding the optimal values of the parameters are obtained by differentiating this equation with respect to  $r$  and  $\lambda$  give rise to the following differential equations:

$$\begin{aligned} \frac{\partial}{\partial r} \ell(x; r, \lambda) &= \sum_{i=1}^n \Psi(r + x_i) - n \Psi(r) \\ &\quad + \sum_{i=1}^n \frac{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left( \frac{1}{(\lambda + r + j)^2} + \frac{4}{(\lambda + r + j)^3} + \frac{6}{(\lambda + r + j)^5} \right)}{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left( \frac{1}{\lambda + r + j} + \frac{2}{(\lambda + r + j)^2} + \frac{2}{(\lambda + r + j)^3} \right)}, \end{aligned}$$

where  $\Psi(k) = \frac{\Gamma'(k)}{\Gamma(k)}$  is a digamma function,

$$\begin{aligned} \frac{\partial}{\partial \lambda} \ell(x; r, \lambda) &= \frac{3n}{\lambda} - \frac{2n(\lambda + 1)}{\lambda^2 + 2\lambda + 2} \\ &\quad + \sum_{i=1}^n \frac{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left( \frac{1}{(\lambda + r + j)^2} + \frac{4}{(\lambda + r + j)^3} + \frac{6}{(\lambda + r + j)^5} \right)}{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left( \frac{1}{\lambda + r + j} + \frac{2}{(\lambda + r + j)^2} + \frac{2}{(\lambda + r + j)^3} \right)}. \end{aligned}$$

Maximum likelihood estimates are obtained by equating both the partial derivatives to zero. But solving these equations is difficult and complicated. So these equations are solved numerically by using Newton Raphson method.

## 6. RESULT AND DISCUSSION

Usefulness of the proposed distribution has been showed by considering a data set from the paper Lord and Geedipally [3]. A real data set is used to establish the result. The data contain single vehicle roadway departure fatal crashes occurred on rural two lane horizontal curves between 2003 and 2008. The data has more number of zeros, hence the data is overdispersed and highly skewed to the right as shown in figure 3. Parameters are estimated using maximum likelihood estimation method and Poisson distribution, NB distribution, NB-L and NB-ISL distributions are fitted to the data. The observed and expected frequencies are summarize by grouping expected frequency greater than five. Based on chi-square test , p value and log likelihood, the table 2 shows that the NB-ISL provide a better fit compared to the Poisson distribution, NB distribution, negative binomial-Lindley distribution.

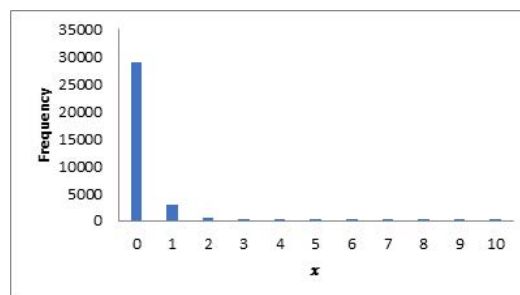


FIGURE 3. Frequency distribution of fatal crashes

## 7. CONCLUSION

In this paper we introduced a new two parameter negative binomial mixed distribution named as negative binomial-improved second degree Lindley distribution obtained by mixing the negative binomial and improved second degree Lindley distribution. We have obtained the moments of the NB-ISL distribution including the factorial moments, mean and variance. Parameter of the NB-ISL distribution is estimated by maximum likelihood estimation method and the usefulness of the NB-ISL distribution is illustrated by a real data set.

TABLE 2. Observed and expected frequencies of crash data

No. of claims	No. of drivers	Fitting of distribution			
		Poisson	NB	NB-L	NB-ISL
0	29087	28471.6	29204.8	29134.1	29099.7
1	2952	3918	2706	2855.2	2904
2	464	269.6	567.4	503.6	500.3
3	108	12.4	141.1	121.7	116.6
4	40	0.4	37.8	36.2	33.3
5	9	0	10.6	12.4	11
6	5	0	3	4.8	4.1
7	2	0	0.9	2	1.6
8	3	0	0.3	0.9	0.7
9	1	0	0.1	0.4	0.3
10+	1	0	0	0.7	0.4
Total	32,672	32,672	32,672	32,672	32,672
Estimated parameter		$\hat{\mu} = 0.138$	$\hat{r} = 0.138$ $\hat{p} = 0.2584$	$\hat{r} = 1.018$ $\hat{\theta} = 9.212$	$\hat{r} = 1.115$ $\hat{\lambda} = 10.764$
df		2	3	4	4
Chi-square		2297.78	51.94	10.51	9.16
Log-likelihood		-14,208.1	-13,557.7	-13,529.8	-13,529.02
p-value		<0.0001	<0.0001	0.03	0.06

The performance of the proposed distribution is compared with some traditional count distributions based on various measures such as chi-square test-of-goodness of fit, p-value and log-likelihood. Based on the result it shows that the NB-ISL provides a better fit compared to the Poisson distribution, NB distribution and negative binomial- Lindley distribution.

#### ACKNOWLEDGMENT

We are very grateful to experts for their appropriate and constructive suggestions to improve this paper.

## REFERENCES

- [1] S. ARYUYUEN, W. BODHISUWAN: *The Negative Binomial-Generalized Exponential Distribution*, Applied Mathematical Sciences, **7**(22) (2013), 1093–1105.
- [2] M. E. GHITANY, B. ATIEH, S. NADARAJAH: *Lindley Distribution and its Application*, Mathematics and Computers in Simulation, **78** (2008), 493–506.
- [3] D. LORD, S. R. GEEDIPALLY: *The Negative Binomial-Lindley Bistribution as a tool for analyzing crash data characterized by a large amount of zeros*, Accident Analysis and Prevention, **43** (2011), 1738–1742.
- [4] K. SADASIVAN, B. VINOTH S. KEERTHANA: *Improved second degree Lindley distribution and its applications*, IOSR Journal of Mathematics, **13** (6) (2017), 01–10.
- [5] P. SAENGTHONG, W. BODHISUWAN: *Negative Binomial-Crack distribution*, International Journal Pure and Applied Mathematics, **84** (3) (2013), 213–230.
- [6] M. SANKARAN: *Discrete Poisson-Lindley distribution*, Biometric, **26**(1) (1970), 145–149.
- [7] H. ZAMANI, N. ISMAIL: *Negative binomial-Lindley distribution and its application*, Journal of Mathematics and Statistics, **6**(1) (2010), 4–9.
- [8] H. ZAMANI, N. ISMAIL, P. FAROUGHI: *Poisson-weighted exponential: Univariate version and regression model with applications*, Journal of Mathematics and Statistics, **10**(2) (2014), 148–154.

DEPARTMENT OF MATHEMATICS  
 AMRITA SCHOOL OF ENGINEERING, COIMBATORE  
 AMRITA VISHWA VIDYAPEETHAM, INDIA  
 E-mail address: ashlyregi@gmail.com

DEPARTMENT OF MATHEMATICS  
 AMRITA SCHOOL OF ENGINEERING, COIMBATORE  
 AMRITA VISHWA VIDYAPEETHAM, INDIA  
 E-mail address: cs\_rajitha@cb.amrita.edu