

# On Poisson-Lindley Distribution and Its Applications to Biological Sciences

Research Article

Volume 2 Issue 4 - 2015

**Rama Shanker<sup>1\*</sup> and Hagos Fesshaye<sup>2</sup>**

<sup>1</sup>Department of Statistics, Eritrea Institute of Technology, Eritrea

<sup>2</sup>Department of Economics, College of Business and Economics, Eritrea

\*Corresponding author: Rama Shanker, Department of Statistics, Eritrea Institute of Technology, Eritrea, Email: shankerrama2009@gmail.com

Received: April 13, 2015 | Published: April 27, 2015

## Abstract

A general expression for the  $r$  th factorial moment of Poisson-Lindley distribution has been obtained and hence its first four moments about origin has been obtained. The distribution has been fitted to some data-sets relating to ecology and genetics to test its goodness of fit and the fit shows that it can be an important tool for modeling biological science data.

**Keywords:** Lindley distribution, Poisson-Lindley distribution, moments, compounding, estimation of parameters, goodness of fit.

## Introduction

The Poisson-Lindley distribution (PLD) given by its probability mass function

$$P(X = x) = \frac{\theta^2(x + \theta + 2)}{(\theta + 1)^{x+3}}; \quad x = 0, 1, 2, \dots, \theta > 0. \quad (1.1)$$

has been introduced by Sankaran (1970) to model count data. The distribution arises from the Poisson distribution when its parameter  $\lambda$  follows Lindley (1958) distribution with its probability density function

$$f(\lambda, \theta) = \frac{\theta^2}{\theta + 1} (1 + \lambda) e^{-\theta\lambda}; \quad x > 0, \theta > 0 \quad (1.2)$$

We have  $P(X = x) =$

$$\int_0^{\infty} P(x | \lambda) f(\lambda; \theta) d\lambda = \int_0^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \cdot \frac{\theta^2}{\theta + 1} (1 + \lambda) e^{-\theta\lambda} d\lambda \quad (1.3)$$

$$= \frac{\theta^2}{(\theta + 1)^{x+3}} \int_0^{\infty} e^{-(\theta+1)\lambda} (\lambda^x + \lambda^{x+1}) d\lambda$$

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

$$= \frac{\theta^2(x + \theta + 2)}{(\theta + 1)^{x+3}}; \quad \theta > 0, \quad x = 0, 1, 2, 3, \dots$$

which is the Poisson-Lindley distribution (PLD).

The PLD has been extensively studied by Sankaran [1] and Ghitany and Mutairi [2] and they have discussed its various properties. The PLD has been generalized by many researchers. Shanker & Mishra [3] obtained a two parameter Poisson-Lindley distribution by compounding Poisson distribution with a two parameter Lindley distribution introduced by Shanker & Mishra [4]. A quasi Poisson-Lindley distribution has been introduced by

Shanker & Mishra [5] by compounding Poisson distribution with a quasi Lindley distribution introduced by Shanker & Mishra [6]. Shanker et al. [7] obtained a discrete two parameter Poisson-Lindley distribution by mixing Poisson distribution with a two parameter Lindley distribution for modeling waiting and survival time's data introduced by Shanker et al. [8]. Further, Shanker & Tekie [9] obtained a new quasi Poisson-Lindley distribution by compounding Poisson distribution with a new quasi Lindley distribution introduced by Shanker & Amanuel [10].

In this paper, a general expression for the  $r$  th factorial moment of PLD has been obtained and hence its first four moments about origin has also been obtained. It seems that not much work has been done on the applications of PLD. The PLD has been fitted to some data sets in ecology and genetics along with Poisson distribution and it has been found that PLD is more flexible for analyzing different types of count data than Poisson distribution.

## Moments of Poisson-Lindley Distribution

The  $r$  th factorial moment about origin of the PLD (1.1) can be obtained as

$$\mu'_{(r)} = E \left[ E \left( X^{(r)} | \lambda \right) \right] \quad (2.1)$$

where

$$X^{(r)} = X(X-1)(X-2)\dots(X-r+1).$$

From (1.3), we thus have

$$\begin{aligned} \mu'_{(r)} &= \int_0^{\infty} \left[ \sum_{x=0}^{\infty} x^{(r)} \frac{e^{-\lambda} \lambda^x}{x!} \right] \frac{\theta^2}{\theta + 1} (1 + \lambda) e^{-\theta\lambda} d\lambda \\ &= \int_0^{\infty} \left[ \lambda^r \sum_{x=r}^{\infty} \frac{e^{-\lambda} \lambda^{x-r}}{(x-r)!} \right] \frac{\theta^2}{\theta + 1} (1 + \lambda) e^{-\theta\lambda} d\lambda \end{aligned}$$

Taking  $x + r$  in place of  $x$ , we get

$$\mu'_{(r)} = \int_0^{\infty} \lambda^r \left[ \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \right] \frac{\theta^2}{\theta + 1} (1 + \lambda) e^{-\theta\lambda} d\lambda$$

The expression within the bracket is clearly unity and hence we have

$$\mu'_{(r)} = \frac{\theta^2}{\theta + 1} \int_0^\infty \lambda^r (1 + \lambda) e^{-\theta\lambda} d\lambda$$

Using gamma integral and a little algebraic simplification, we get finally a general expression for the  $r$  th factorial moment of PLD as

$$\mu'_{(r)} = \frac{r!(\theta + r + 1)}{\theta^r (\theta + 1)}; r = 1, 2, 3, \dots \quad (2.2)$$

Substituting  $r = 1, 2, 3,$  and  $4$  in (2.2), first four factorial moments can be obtained and using the relationship between factorial moments and moments about origin, the first four moment about origin of the PLD (1.1) are given by

$$\mu'_1 = \frac{\theta + 2}{\theta(\theta + 1)} \quad (2.3)$$

$$\mu'_2 = \frac{\theta + 2}{\theta(\theta + 1)} + \frac{2(\theta + 3)}{\theta^2(\theta + 1)} \quad (2.4)$$

$$\mu'_3 = \frac{\theta + 2}{\theta(\theta + 1)} + \frac{6(\theta + 3)}{\theta^2(\theta + 1)} + \frac{6(\theta + 4)}{\theta^3(\theta + 1)} \quad (2.5)$$

$$\mu'_4 = \frac{\theta + 2}{\theta(\theta + 1)} + \frac{14(\theta + 3)}{\theta^2(\theta + 1)} + \frac{36(\theta + 4)}{\theta^3(\theta + 1)} + \frac{24(\theta + 5)}{\theta^4(\theta + 1)} \quad (2.6)$$

Ghitany et al. [2] discussed the estimation methods for the PLD (1.1) and its applications.

### Estimation Of Parameters

**Maximum Likelihood (ML) Estimates:** Let  $x_1, x_2, \dots, x_n$  be a random sample of size  $n$  from the PLD (1.1). Let  $f_x$  be the observed frequency in the sample corresponding to  $X = x$  ( $x = 1, 2, 3, \dots, k$ ) such that

$$\sum_{x=1}^k f_x = n,$$

where  $k$  is the largest observed value having non-zero frequency. The likelihood function,  $L$ , of the PLD (1.1) is given by

$$L = \theta^{2n} \frac{1}{(\theta + 1)^{\sum_{x=1}^k f_x(x+3)}} \prod_{x=1}^k (x + \theta + 2)^{f_x}$$

The log likelihood function is given by

$$\text{Log}L = 2n \log \theta - \sum_{x=1}^k f_x (x + 3) \log(\theta + 1) + \sum_{x=1}^k f_x \log(x + \theta + 2)$$

The maximum likelihood estimate,  $\hat{\theta}$  of  $\theta$  is the solution of the equation

$$\frac{d \log L}{d \theta} = 0$$

and is given by solution of the following non-linear equation

$$\frac{2n}{\theta} - \frac{n(\bar{x} + 3)}{\theta + 1} + \sum_{x=1}^k \frac{f_x}{x + \theta + 2} = 0 \quad (3.1.1)$$

Where  $\bar{x}$  is the sample mean. It has been shown by Ghitany & Mutairi [2] that the ML estimator  $\hat{\theta}$  of  $\theta$  is consistent and

asymptotically normal.

**Estimates from Moments:** Let  $x_1, x_2, \dots, x_n$  be a random sample of size  $n$  from the PLD (1.1). Equating the first moment about origin to the sample mean, the method of moment (MOM) estimate,  $\tilde{\theta}$ , of  $\theta$  is given by

$$\tilde{\theta} = \frac{-(\bar{x} - 1) + \sqrt{(\bar{x} - 1)^2 + 8\bar{x}}}{2\bar{x}}; \bar{x} > 0 \quad (3.2.1)$$

Where  $\bar{x}$  is the sample mean? It has been shown by Ghitany & Mutairi [2] that the MOM estimator  $\hat{\theta}$  of  $\theta$  is positively biased, consistent and asymptotically normal.

### Applications of Poisson-Lindley Distribution

The Poisson distribution is a suitable model for the situations where events seem to occur at random such as the number of customers arriving at a service point, the number of telephone calls arriving at an exchange, the number of fatal traffic accidents per week in a given state, the number of radioactive particle emissions per unit of time, the number of meteorites that collide with a test satellite during a single orbit, the number of organisms per unit volume of some fluid, the number of defects per unit of some materials, the number of flaws per unit length of some wire, etc. However, the Poisson distribution requires events to be independent- a condition which is rarely satisfied completely. In biological science and medical science, the occurrence of successive events is dependent. The negative binomial distribution is a possible alternative to the Poisson distribution when successive events are possibly dependent [11]. Further, for fitting Poisson distribution to the count data equality of mean and variance should be satisfied. Similarly, for fitting negative binomial distribution (NBD) to the count data, mean should be less than the variance. In biological and medical sciences, these conditions are not fully satisfied.

The theoretical and empirical justification for the selection of the PLD to describe biological science and medical science data is that PLD is over dispersed ( $\mu < \sigma^2$ )

### Application in Ecology

The organisms and their environment in the nature are not only complex and dynamic but also interdependent, mutually reactive and interrelated. Ecology deals with the various principles which govern such relationship between organisms and their environment. Fisher et al. [12] has discussed the applications of Logarithmic series distribution (LSD) to model count data in the science of ecology. It was Kempton [13] who fitted the generalized form of Fisher's Logarithmic series distribution (LSD) to model insect data and concluded that it gives a superior fit as compared to ordinary Logarithmic series distribution (LSD). He also concluded that it gives better explanation for the data having exceptionally long tail. Tripathi & Gupta [14] proposed another generalization of the Logarithmic series distribution (LSD) and fitted it to insect data and found that it gives better fit as compared to ordinary Logarithmic series distribution. They concluded that the distribution is flexible to describe short-tailed as well as long-tailed data. Mishra & Shanker [15] have discussed applications of generalized logarithmic series distributions (GLSD) to models data in ecology.

In this section we have tried to fit Poisson distribution and Poisson-Lindley distribution to many biological data using maximum likelihood estimates. The data were on haemocytometer yeast cell counts per square, on European red mites on apple leaves and European corn borers per plant.

It is obvious from above (Table 1-3) that PLD gives much closer fit than Poisson distribution and thus it can be considered as an important tool for modeling data in ecology.

**Table 1:** Observed and expected number of Haemocytometer yeast cell counts per square observed by 'Student' 1907.

Number of cells per square	Observed frequency	Expected frequency	
		Poisson distribution	Poisson-Lindley distribution
0	128	118.1	127.4
1	37	54.3	41.1
2	18	12.5	12.9
3	3	1.9	3.9
4	1	0.2	1.2
5+	0	0.0	0.5
Total	187	187.0	187.0
Estimate of parameter		$\hat{\theta}=0.459893$	$\hat{\theta}=2.751579$
$\chi^2$		9.903	1.431
d.f.		1	1
p-value		0.0016	0.2316

**Table 2:** Observed and expected number of red mites on Apple leaves.

Number mites per leaf	Observed frequency	Expected frequency	
		Poisson distribution	Poisson-Lindley distribution
0	38	25.3	35.8
1	17	29.1	20.7
2	10	16.7	11.4
3	9	6.4	6.0
4	3	1.8	3.1
5	2	0.4	1.6
6	1	0.2	0.8
7+	0	0.1	0.6
Total	80	80.0	80.0
Estimate of parameter		$\hat{\theta}=1.15$	$\hat{\theta}=1.255891$
$\chi^2$		18.275	2.469
d.f.		2	3
p-value		0.0001	0.4809

**Table 3:** Observed and expected number of European corn-borer of Mc. Guire et al [18].

Number of bores per plant	Observed frequency	Expected frequency	
		Poisson distribution	Poisson-Lindley distribution
0	83	78.9	87.2
1	36	42.9	31.8
2	14	11.7	11.2
3	2	2.1	3.8
4+	1	0.4	2.0
Total	136	136.0	136.0
Estimate of parameter		$\hat{\theta}=0.544118$	$\hat{\theta}=2.372252$
$\chi^2$		1.885	0.757
d.f.		1	1
p-value		0.1698	0.3843

### Application in Genetics

**Table 4:** Distribution of number of Chromatid aberrations (0.2 g chinon 1, 24 hours).

Number of aberrations	Observed frequency	Expected frequency	
		Poisson distribution	Poisson-Lindley distribution
0	268	231.3	257.0
1	87	126.1	93.4
2	26	34.7	32.8
3	9	6.3	11.2
4	4	0.8	3.8
5	2	0.1	1.2
6	1	0.1	0.4
7+	3	0.1	0.2
Total	400	400.0	400.0
Estimate of parameter		$\hat{\theta}=0.5475$	$\hat{\theta}=2.380442$
$\chi^2$		38.208	6.208
d.f.		2	3
p-value		0.0000	0.1019

Genetics is the branch of biological science which deals with heredity and variation. Heredity includes those traits or characteristics which are transmitted from generation to generation, and is therefore fixed for a particular individual. Variation, on the other hand, is mainly of two types, namely hereditary and environmental. Hereditary variation refers to differences in inherited traits whereas environmental variations are those which are mainly due to environment. In the field of

genetics much quantitative studies seem to have been done. The segregation of chromosomes has been studied using statistical tool, mainly chi-square ( $\chi^2$ ). In the analysis of data observed on chemically induced chromosome aberrations in cultures of human leukocytes, Loeschke & Kohler [16] suggested the negative binomial distribution while Janardan & Schaeffer [17] suggested modified Poisson distribution. Mishra & Shanker [15] have discussed applications of generalized Logarithmic series distributions (GLSD) to model data in mortality, ecology and genetics. In this section an attempt has been made to fit to data relating to genetics using PLD and Poisson distribution using maximum likelihood estimate. Also an attempt has been made to fit Poisson distribution and PLD to the data of Catcheside et al. [19,20] in (Tables 3-7) [21].

**Table 5:** Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by streptonigrin (NSC-45383), Exposure -60  $\mu\text{g} | \text{kg}$

Class/Exposure ( $\mu\text{g}   \text{kg}$ )	Observed frequency	Expected frequency	
		Poisson distribution	Poisson-Lindley distribution
0	413	374.0	405.7
1	124	177.4	133.6
2	42	42.1	42.6
3	15	6.6	13.3
4	5	0.8	4.1
5	0	0.1	1.2
6	2	0.0	0.5
Total	601	601.0	601.0
Estimate of parameter		$\hat{\theta}=0.47421$	$\hat{\theta}=2.685373$
$\chi^2$		48.169	1.336
d.f.		2	3
p-value		0.0000	0.7206

**Table 6:** Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by streptonigrin (NSC-45383), Exposure -70  $\mu\text{g} | \text{kg}$

Class/Exposure ( $\mu\text{g}   \text{kg}$ )	Observed frequency	Expected frequency	
		Poisson distribution	Poisson-Lindley distribution
0	200	172.5	191.8
1	57	95.4	70.3
2	30	26.4	24.9
3	7	4.9	8.6
4	4	0.7	2.9
5	0	0.1	1.0
6	2	0.0	0.5
Total	300	300.0	300.0
Estimate of parameter		$\hat{\theta}=0.55333$	$\hat{\theta}=2.353339$
$\chi^2$		29.68	3.91
d.f.		2	2
p-value		0.0000	0.1415

**Table 7:** Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by streptonigrin (NSC-45383), Exposure -90  $\mu\text{g} | \text{kg}$

Class/Exposure ( $\mu\text{g}   \text{kg}$ )	Observed frequency	Expected frequency	
		Poisson distribution	Poisson-Lindley distribution
0	155	127.8	158.3
1	83	109.0	77.2
2	33	46.5	35.9
3	14	13.2	16.1
4	11	2.8	7.1
5	3	0.5	3.1
6	1	0.2	2.3
Total	300	300.0	300.0
Estimate of parameter		$\hat{\theta}=0.853333$	$\hat{\theta}=1.617611$
$\chi^2$		24.969	1.51
d.f.		2	3
p-value		0.0000	0.6799

It is obvious from above tables that PLD gives much closer fit than Poisson distribution and thus it can be considered as an important tool for modeling data in genetics.

**Acknowledgement**

None.

**Conflict of Interest**

No conflict of interest

**References**

1. Sankaran M (1970) The discrete Poisson-Lindley distribution. *Biometrics* 26(1): 145-149.
2. Ghitany ME, Al-Mutairi DK (2009) Estimation Methods for the discrete Poisson-Lindley distribution. *Journal of Statistical Computation and Simulation* 79(1): 1-9.
3. Shanker R, Mishra A (2014) A two-parameter Poisson-Lindley distribution. *International Journal of Statistics and Systems* 9(1): 79-85.
4. Shanker R, Mishra A (2013 a) A two-parameter Lindley distribution. *Statistics in Transition new Series* 14(1): 45-56.
5. Shanker R, Mishra A (2015) A quasi Poisson-Lindley distribution (submitted).
6. Shanker R, Mishra A (2013 b) A quasi Lindley distribution. *African journal of Mathematics and Computer Science Research* 6(4): 64-71.
7. Shanker R, Sharma S, Shanker R (2012) A Discrete two-Parameter Poisson Lindley Distribution. *Journal of Ethiopian Statistical Association* 21: 15-22.
8. Shanker R, Sharma S, Shanker, R (2013) A two-parameter Lindley distribution for modeling waiting and survival times data, *Applied Mathematics* 4: 363-368.
9. Shanker R, Tekie AL (2014) A new quasi Poisson-Lindley distribution. *International Journal of Statistics and Systems* 9(1): 87-94.
10. Shanker R, Amanuel AG (2013) A new quasi Lindley distribution, *International Journal of Statistics and Systems* 8(2): 143-156.

11. Johnson NL, Kotz S, Kemp AW (1992) Univariate Discrete Distributions, 2<sup>nd</sup> edition, John Wiley & sons Inc.
12. Fisher RA, Corpet AS, Williams CB (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* 12(1): 42-58.
13. Kempton RA (1975) A generalized form of Fisher's logarithmic series. *Biometrika* 62 (1): 29-38.
14. Tripathi RC, Gupta RC (1985) A generalization of the log-series distribution. *Comm. in Stat. (Theory and Methods)* 14(8): 1779-1799.
15. Mishra A, Shanker R (2002) Generalized logarithmic series distribution-Its nature and applications, *Proceedings of the Vth International Symposium on Optimization and Statistics* 28-30: 155-168.
16. Loeschke V, Kohler W (1976) Deterministic and Stochastic models of the negative binomial distribution and the analysis of chromosomal aberrations in human leukocytes. *Biometrische Zeitschrift* 18: 427-451.
17. Janardan KG, Schaeffer DJ (1977) Models for the analysis of chromosomal aberrations in human leukocytes. *Biometrical Journal* (8)599-612.
18. Mc Guire JU, Brindley TA, Bancroft TA (1957) The distribution of European corn-borer larvae *pyrausta* in field corn. *Biometrics* 13: 65-78.
19. Catcheside DG, Lea DE, Thoday JM (1946) Types of chromosome structural change induced by the irradiation on *Tradescantia* microspores. *Journal of Genetics* 47: 113-136.
20. Catcheside DG, Lea DE, Thoday JM (1946) The production of chromosome structural changes in *Tradescantia* microspores in relation to dosage, intensity and temperature. *Journal of Genetics* 1.47: 137-149.
21. Lindley DV (1958) Fiducial distributions and Bayes theorem. *Journal of Royal Statistical Society Ser. B* Vol.20: 102-107.