On Poisson-Lindley Distribution and Its Applications to Biological Sciences

Abstract

A general expression for the r\textsuperscript{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

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

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

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

We have

\[
P(X = x) = \frac{\theta^2}{(\theta + 1)^x} \int_0^\infty P(x | \lambda) f(\lambda; \theta) d\lambda
\]

\[
= \frac{\theta^2}{(\theta + 1)^x} \int_0^\infty \frac{e^{-\lambda \theta} \lambda^x}{x!} \frac{\theta^2}{\theta + 1} (1 + \lambda) e^{-\theta \lambda} d\lambda
\]

\[
= \frac{\theta^2}{(\theta + 1)^x} \int_0^\infty \frac{\lambda^x (1 + \lambda)^{x+1}}{x!} d\lambda
\]

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

which is the Poisson-Lindley distribution (PLD).


In this paper, a general expression for the r\textsuperscript{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\textsuperscript{th} factorial moment about origin of the PLD (1.1) can be obtained as

\[
\mu'_r = E \left[ (X^{(r)})^{\theta} \right] \quad (2.1)
\]

where

\[
X^{(r)} = X(X - 1)(X - 2) \ldots (X - r + 1).
\]

From (1.3), we thus have

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

\[
= \int_0^\infty \frac{\lambda^r}{x!} (x + 1) e^{-\theta \lambda} d\lambda
\]

Taking \( x + r \) in place of \( x \), we get

\[
\mu'_r = \sum_{x=0}^{\infty} \frac{\lambda^r}{x!} (x + 1) e^{-\theta \lambda} d\lambda
\]

\[
= \frac{\theta^2}{\theta + 1} (1 + \lambda) e^{-\theta \lambda} d\lambda
\]
The expression within the bracket is clearly unity and hence we have
\[ \mu'_r = \frac{\theta^2}{\theta + 1} \frac{1}{\theta} e^{\frac{\theta}{\theta + 1} - 1} \]

Using gamma integral and a little algebraic simplification, we get finally a general expression for the \( r \) th factorial moment of PLD as
\[ \mu'_r = \frac{r!(\theta + r + 1)}{\theta (\theta + 1)} ; r = 1, 2, 3, \ldots \quad (2.2) \]

Substituting \( r = 1, 2, 3, \text{and} 4 \) 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
\[ \mu'_1 = \frac{\theta + 2}{\theta (\theta + 1)} \quad (2.3) \]
\[ \mu'_2 = \frac{\theta + 2}{\theta (\theta + 1)} + \frac{2(\theta + 3)}{\theta (\theta + 1)} \quad (2.4) \]
\[ \mu'_3 = \frac{\theta + 2}{\theta (\theta + 1)} + \frac{6(\theta + 3)}{\theta (\theta + 1)} + \frac{6(\theta + 4)}{\theta (\theta + 1)} \quad (2.5) \]
\[ \mu'_4 = \frac{\theta + 2}{\theta (\theta + 1)} + \frac{14(\theta + 3)}{\theta (\theta + 1)} + \frac{36(\theta + 4)}{\theta (\theta + 1)} + \frac{24(\theta + 5)}{\theta (\theta + 1)} \quad (2.6) \]

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

**Estimation Of Parameters**

**Maximum Likelihood (ML) Estimates:** Let \( x_1, x_2, \ldots, x_n \) be a random sample of size \( n \) from the PLD (1.1). Let \( f_k \) be the observed frequency in the sample corresponding to \( X = x \) (\( x = 1, 2, 3, \ldots, k \)) 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
\[ L = \theta^{\sum_{x=1}^{k} f_x (x+3)} (\theta + 1)^{-\sum_{x=1}^{k} f_x (x+2)} \]
The log likelihood function is given by
\[ \log L = 2n \log \theta - \sum_{x=1}^{k} f_x (x+3) \log (\theta + 1) + \sum_{x=1}^{k} f_x \log (x + \theta + 2) \]
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
\[ \frac{2n}{\theta} - \frac{n(x+3)}{\theta + 1} + \sum_{x=1}^{k} f_x (x+\theta+2) = 0 \quad (3.1.1) \]
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 \( x_1, x_2, \ldots, x_n \) 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, \( \hat{\theta} \), of \( \theta \) is given by
\[ \hat{\theta} = \frac{(\bar{x} - 1) + \sqrt{\bar{x}(\bar{x} - 1)^2 + 8\bar{x}}}{2\bar{x}} ; \bar{x} > 0 \quad (3.2.1) \]

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 meteores 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 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 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.

<table>
<thead>
<tr>
<th>Number of cells per square</th>
<th>Observed frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poisson distribution</td>
<td>Poisson-Lindley distribution</td>
</tr>
<tr>
<td>0</td>
<td>128</td>
<td>118.1</td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>54.3</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>12.5</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>5+</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Total</td>
<td>187</td>
<td>187.0</td>
</tr>
</tbody>
</table>

Estimate of parameter

\[ \hat{\lambda} = 0.459893 \]

\[ \hat{\lambda} = 2.751579 \]

\[ \chi^2 \]

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.

<table>
<thead>
<tr>
<th>Number of mites per leaf</th>
<th>Observed frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poisson distribution</td>
<td>Poisson-Lindley distribution</td>
</tr>
<tr>
<td>0</td>
<td>38</td>
<td>25.3</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>29.1</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>16.7</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>6.4</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1.8</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.4</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>7+</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>80.0</td>
</tr>
</tbody>
</table>

Estimate of parameter

\[ \hat{\lambda} = 1.15 \]

\[ \hat{\lambda} = 1.255891 \]

\[ \chi^2 \]

18.275

2.469

\[ d.f. \]

2

3

p-value

0.0001

0.4809

Application in Genetics

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

<table>
<thead>
<tr>
<th>Number of aberrations</th>
<th>Observed frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poisson distribution</td>
<td>Poisson-Lindley distribution</td>
</tr>
<tr>
<td>0</td>
<td>268</td>
<td>231.3</td>
</tr>
<tr>
<td>1</td>
<td>87</td>
<td>126.1</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>34.7</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>6.3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.8</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>7+</td>
<td>3</td>
<td>0.1</td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td>400.0</td>
</tr>
</tbody>
</table>

Estimate of parameter

\[ \hat{\lambda} = 0.5475 \]

\[ \hat{\lambda} = 2.380442 \]

\[ \chi^2 \]

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

On Poisson-Lindley Distribution and Its Applications to Biological Sciences

genetics much quantitative studies seem to have been done. The
segregation of chromosomes has been studied using statistical
tools, 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 μg | kg

| Class/Exposure (μg | 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
$\hat{\lambda} = 0.47421$  $\hat{\theta} = 2.685373$

$\chi^2$  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 μg | kg

| Class/Exposure (μg | 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
$\hat{\lambda} = 0.55333$  $\hat{\theta} = 2.35339$

$\chi^2$  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 μg | kg

| Class/Exposure (μg | kg ) | Observed frequency | Expected frequency |
|-------------------|---------------------|-------------------|
|                   | Poisson distribution | Poisson-Lindley distribution |
| 0                 | 155                 | 127.8              | 158.3 |
| 1                 | 83                  | 109.0              | 72.7  |
| 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
$\hat{\lambda} = 0.85333$  $\hat{\theta} = 1.617611$

$\chi^2$  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


