A COMPARATIVE STUDY ON ZERO-TRUNCATED GENERALIZED POISSON-LINDLEY AND ZERO-TRUNCATED POISSON-LINDLEY DISTRIBUTIONS

SAMUEL ADEWALE ADEROJU*1
1Department of Statistics and Mathematical Sciences, Kwara State University, Malete, P. M. B. 1530, Ilorin, Kwara State, Nigeria.

EMMANUEL TEJU JOLAYEMI2
2Department of Statistics, University of Ilorin, P. M. B. 1515, Ilorin, Nigeria.

ABDULWASIU OPEYEMI IBRAHIM3
3Department of Statistics and Mathematical Sciences, Kwara State University, Malete, P. M. B. 1530, Ilorin, Kwara State, Nigeria.

(Received On: 13-01-17; Revised & Accepted On: 23-03-17)

ABSTRACT

In this paper, the zero-truncated generalized Poisson-Lindley (ZTGPL) distribution has been proposed and its properties studied. The maximum likelihood method is used to obtain the estimators of its parameters through R-software. The comparative study with zero-truncated Poisson (ZTP) and zero-truncated Poisson-Lindley (ZTPL) distributions is done with two datasets. The proposed distribution is characterized by two parameters and flexible to account for both over- and under-dispersion in structurally non-zero count data. The statistic (chi square) is used to check its goodness-of-fit.

Key words: Generalized Poisson-Lindley distribution; goodness of fit; Poisson distribution; Poisson-Lindley distribution; Zero-Truncated distribution.

1. INTRODUCTION

In probability theory, zero-truncated discrete distributions are distribution that can handle the set of positive integers that are structurally non-zero counts.


Zero-truncated Poisson (ZTP) and zero-truncated Poisson-Lindley (ZTPL) distributions have been compared and studied using graphs for different values of their parameter by Shanker et al. (2015). A general expression for the rth factorial moment of ZTPL distribution has been obtained and the first four moments about origin has been given. A very simple and easy method for finding moments of ZTPL distribution has been suggested. Both ZTP and ZTPL distributions have been fitted to a number of data sets from different fields to study their goodness of fits and superiority of one over the other.
2. REVIEW ON OTHER DISTRIBUTIONS UNDER CONSIDERATION

2.1 Zero-truncated Poisson (ZTP) distribution

The probability mass function (pmf) of a ZTP variable as given independently by Plackett (1953) and Johnson et al. (1969) as:

\[ P(Y = y) = \frac{\lambda^y}{y!(e^\lambda - 1)}, \quad \text{if } y = 1, 2, \ldots \]

(1)

The difference with the standard Poisson distribution lies in the correction factor \( \frac{1}{e^\lambda - 1} \), which reflects the fact that a value of zero (0) cannot occur.

The moment generating function is obtained to be:

\[ M_y(t) = \frac{e^{\lambda t}}{e^\lambda - 1} \]

Moreover, the basic parameters such as the mean and variance respectively are:

\[ \mu = \frac{\lambda e^\lambda}{e^\lambda - 1} \quad \text{and} \quad \sigma^2 = \frac{\lambda e^\lambda}{e^\lambda - 1} \left[ 1 - \frac{\lambda}{e^\lambda - 1} \right] \]

2.2 Zero-truncated Poisson-Lindley (ZTPL) distribution

The ZTPL was studied by Shanker et al. (2015) and the pmf defined as:

\[ P(Y = y) = \frac{\theta^y}{\theta^2 + 3\theta + 1} \left( \theta + 1 \right)^y, \quad \text{if } y = 1, 2, \ldots ; \theta > 0 \]

(3)

Moreover, the basic parameters such as the mean and variance respectively are as given by Ghitany et al. (2008):

\[ \mu = \frac{(\theta + 1)^2(\theta + 2)}{\theta(3\theta + 1)} \quad \text{and} \quad \sigma^2 = \frac{(\theta + 1)^2(\theta^2 + 6\theta^2 + 10\theta + 2)}{\theta(3\theta + 1)^2} \]

Ghitany et al. (2008) showed that the MLE \( \hat{\theta} \) of \( \theta \) is consistent and asymptotically normal.

3. ZERO-TRUNCATED GENERALIZED POISSON-LINDLEY DISTRIBUTION

Baghestani et al. (2014) derived a generalized Poisson-Lindley (GPL) distribution by compounding Poisson distribution with the generalized Lindley distribution introduced by Zakerzadeh and Dolati (2010). For details on its various properties, see Baghestani et al. (2014). However, the probability mass function of the GPL distribution is

\[ P(y; \alpha, \beta) = \left\{ \begin{array}{ll}
\frac{\Gamma(y_i + \alpha)\beta^{\alpha+1}}{y_i!\Gamma(\alpha+1)(\beta+1)^{\alpha+1}} \left[ \frac{\alpha + \alpha + y_i}{\beta + 1} \right], & \text{for } y_i = 1, 2, \ldots ; \alpha, \beta > 0 \\
0, & \text{e.w}
\end{array} \right. \]

(4)

Hence, the Zero-truncated version of the distribution, which we refer to as Zero-truncated Generalized Poisson-Lindley (ZTGPL) distribution can be derived as

\[ f(y; \alpha, \beta) = \frac{P(y; \alpha, \beta)}{1 - P(0; \alpha, \beta)}, \quad y_j = 1, 2, \ldots , n \]

(5)
Where, \( P(0; \alpha, \beta) = \frac{(\alpha - 1)!}{\alpha!} \frac{\beta^{\alpha+1}}{(1 + \beta)^{\alpha+1}} \left( \alpha + \alpha + 1 \beta \right) = \frac{2 + \beta}{(1 + \beta)^{\alpha+2}} \) and \( f(y; \alpha, \beta) \) is the ZTGPL.

Therefore, the pmf of the ZTGPL distribution is derived as
\[
f(y; \alpha, \beta) = \frac{\Gamma(y_i + \alpha)(\beta^{\alpha+1} \left( \alpha + \alpha + \beta \right)}{y_i! \Gamma(\alpha + 1)(\beta + 1)} \left( 1 - \frac{1}{1 + \beta} \right) \left( \alpha - 1 \right)! \left( 1 + \beta \right)^{-\alpha} \left( \alpha + \alpha + \beta \right)
\]

\[
= \frac{\beta^{\alpha+1}}{y_i!} \left( 1 + \beta \right)^{\alpha+1} \left( \alpha + \alpha + \beta \right) \left( \alpha - 1 \right)! \left( 1 + \beta \right)^{-\alpha} \left( \alpha + \alpha + \beta \right)
\]

\[
y_i = 1, 2, \ldots, \infty, \quad \beta > 0, \quad \alpha > 0
\]

Hence, this is the pmf of ZTGPL with parameters \( \alpha \) and \( \beta \)

3.1 Generating Functions of ZTGPL distribution

The Probability Generating function is
\[
P(t) = E(t^y) = \sum_{y=1}^{\infty} t^y f(y; \alpha, \beta)
\]
\[
= \sum_{y=1}^{\infty} \frac{\beta^{\alpha+1}}{y_i!} \left( 1 + \beta \right)^{\alpha+1} \left( \alpha + \alpha + \beta \right) \left( \alpha - 1 \right)! \left( 1 + \beta \right)^{-\alpha} \left( \alpha + \alpha + \beta \right)
\]

\[
= \frac{\beta^{\alpha+1}}{ \left( 1 + \beta \right)^{\alpha+1} \left( \alpha + \alpha + \beta \right) \left( \alpha - 1 \right)! \left( 1 + \beta \right)^{-\alpha} \left( \alpha + \alpha + \beta \right)
\]

\[
= \frac{1}{\left( 1 + \beta \right)^{\alpha+1} \left( \alpha + \alpha + \beta \right) \left( \alpha - 1 \right)! \left( 1 + \beta \right)^{-\alpha} \left( \alpha + \alpha + \beta \right)
\]

\[
3.2 Moments of ZTGPL distribution

The first four moments are derived as
\[
\mu_1 = \beta^a \frac{(\beta^{\alpha+1})}{\alpha! \left( 1 + \beta \right)^{\alpha+1} \left( \alpha + \alpha + \beta \right) \left( \alpha - 1 \right)! \left( 1 + \beta \right)^{-\alpha} \left( \alpha + \alpha + \beta \right)
\]

\[
\mu_2 = \frac{(1 + \alpha) \beta^{2+\alpha}}{\alpha! \left( 1 + \beta \right)^{\alpha+1} \left( \alpha + \alpha + \beta \right) \left( \alpha - 1 \right)! \left( 1 + \beta \right)^{-\alpha} \left( \alpha + \alpha + \beta \right)
\]
The mean and variance of the distribution can therefore be derived as:

\[
\mu = \frac{\beta^a \left( \frac{\beta}{1+\beta} \right)^{\alpha+a}}{(1+\beta)^{\alpha+a}} \left( 1 + \alpha + \alpha \beta \right)
\]

(12)

\[
\sigma^2 = \left( \frac{\beta^{-2} \left( \frac{\beta}{1+\beta} \right)^{2\alpha}}{(1+\beta)^{\alpha}} \left( 1 - \beta^a \right) \left( 1 + a + \alpha \beta \right)^2 a^{\alpha-1} \left( 1 + a \right) \left( \frac{1}{1+\beta} \right)^{\alpha} \right) \left( 2 + \alpha + \alpha \beta \right) \left( 2 + \beta \right) \left( 1+\beta \right)^{2\alpha} a!igr) - \left( \alpha \beta^{\alpha+1} (2+\beta)(\alpha-1)! \right) \right)
\]

(13)
In figure 1, as $\alpha \to 0$ while $\beta$ remain around 1 the distribution skewed to the right but as $\alpha \to \infty$ while $\beta$ remain around 1 the distribution skewed to the left. However, when $\alpha$ remains around 1 and $\beta \to \infty$, the distribution right tail disappears gradually. In figure 2, as $\alpha$ and $\beta$ is simultaneously increase, the distribution right tail disappears gradually.

4. ESTIMATION OF PARAMETERS OF ZTGPL DISTRIBUTION

The likelihood function of the ZTGPL distribution is given as:

$$L(\alpha, \beta; y_i) = \prod_{i=1}^{n} \frac{\beta^{\alpha y_i+1} (1 + \beta)^{-\alpha y_i} (y_i + \alpha(2 + \beta)) (y_i + \alpha - 1)!}{y_i! (1 + \beta)^{\alpha} \alpha^{\alpha y_i+1} (2 + \beta)(\alpha - 1)!}$$

(14)

The log-likelihood function therefore is:

$$\ell = n(\alpha + 1) \log \beta - \sum_{i=1}^{n} y_i \log(\beta + 1) + \sum_{i=1}^{n} \log(y_i + 2\alpha + 2\beta) + \sum_{i=1}^{n} \log(y_i + \alpha - 1)!$$

$$- \sum_{i=1}^{n} \log(y_i)! - n \log \left[ (1 + \beta)^{2a-\alpha} \alpha! - \alpha\beta^{\alpha}(2 + \beta)(\alpha - 1)! \right]$$

(15)

We are to find the first and second partial derivative of equation (15) with respect to each parameter and equate them to zero as:

$$\frac{\partial \ell}{\partial \alpha} = 0, \quad \frac{\partial^2 \ell}{\partial \alpha^2} = 0, \quad \frac{\partial \ell}{\partial \beta} = 0, \quad \text{and} \quad \frac{\partial^2 \ell}{\partial \beta^2} = 0$$

However, the equations do not have closed form. Therefore, the maximum likelihood estimates (MLEs) of ZTGPL distribution cannot be solved analytically, an iterative methods such as Fisher Score Algorithm, Bisecition method Regula-Falsi method or Newton-Raphson (NR) iterative method, as implemented by Jolayemi (1990).

We obtained the MLEs of the parameters by direct maximization of the log-likelihood function using “optim” routine of R software (R Development Core Team, 2016) with “L-BFGS-B” method. This can as well be done by using PROC NLMIXED in SAS.

5. APPLICATION

We would want to compare the distributions with two real datasets so as to establish their goodness-of-fit.

Example 1: Immunogold assay data

The data is taking from Mathews et al (1993), who gave counts of sites with 1, 2, 3, 4 and 5 particles from immunogold assay data. The sample mean and variance are 1.576 and 0.7897, respectively.

<table>
<thead>
<tr>
<th>X</th>
<th>Obs.Freq</th>
<th>ZTP</th>
<th>ZTPL</th>
<th>ZTGPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>122</td>
<td>115.86</td>
<td>124.77</td>
<td>121.89</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>57.39</td>
<td>46.76</td>
<td>50.07</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>18.95</td>
<td>17.07</td>
<td>17.74</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4.69</td>
<td>6.11</td>
<td>5.78</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.93</td>
<td>2.15</td>
<td>1.78</td>
</tr>
<tr>
<td>Total</td>
<td>198</td>
<td>198</td>
<td>198</td>
<td>198</td>
</tr>
</tbody>
</table>
From the results in table 1 above, the three distributions fit the data well. However, ZTGPL distribution outperforms the ZTPL and ZTP distributions.

**Example 2: Mortality data**

The data is taken from Shanker et al. (2015), who gave counts of mothers of the rural area having at least one live birth and one neonatal death.

<table>
<thead>
<tr>
<th>X (# of neonatal deaths)</th>
<th>Obs. Freq</th>
<th>ZTP</th>
<th>ZTPL</th>
<th>ZTGPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>409</td>
<td>399.7</td>
<td>408.1</td>
<td>409.6</td>
</tr>
<tr>
<td>2</td>
<td>88</td>
<td>102.3</td>
<td>89.4</td>
<td>87.3</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>17.5</td>
<td>19.3</td>
<td>19.6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2.2</td>
<td>4.1</td>
<td>4.5</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.3</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Total</td>
<td>522</td>
<td>198</td>
<td>198</td>
<td>198</td>
</tr>
</tbody>
</table>

Considering the mortality data in table 2, the value of the Chi-square test of ZTGPL distribution is smaller than that of the ZTPL distribution but the p-value is otherwise. That is due to the difference in the number of parameters; the ZTPL distribution is one parameter distribution while the ZTGPL distribution is a two parameter distribution. However, the two distributions perform better than the ZTP distribution. Several other examples, which are not reported in this article show that the ZTGPL distribution is a good alternative to the ZTPL and zero-truncated Poisson-Gamma distributions.

### 6. CONCLUSION

In this paper the ZTGPL distribution has been proposed and its properties studied. The maximum likelihood method is used to obtain the estimators of its parameters through R-software. The two datasets that have been studied previously with ZTP and ZTPL distributions were used to study its goodness of fit to count data. The proposed distribution is characterized by two parameters and flexible to account for both over- and under-dispersion in structurally non-zero count data. The statistic (chi square) is used to check its goodness-of-fit.

**REFERENCE**

APPENDIX:

R function for the parameters estimation:

```r
mlefn=function(par,x,n)
{
  α=par[1]
  β=par[2]
  ll=(α+1)*log(β)-x*log(β+1)+log(x+2)*α*β+lgamma(x+1)-log(factorial(α)+(2+α)*β)-α*(2+β)*factorial(α-1)*β^(α+1))
  sum(ll)
}
optim(c(0.05,0.05),mlefn,method="L-BFGS-B",lower=c(0.05,0.05),upper=c(Inf,Inf),x=data,n)
```

Source of support: Nil, Conflict of interest: None Declared.

[Copy right © 2017. This is an Open Access article distributed under the terms of the International Journal of Mathematical Archive (IJMA), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.]