Proceeding of ICSA 2019, p: 40-50
ISBN 978-979-19256-3-1 (PDF)

Zero-truncated Poisson-Lindley distribution

Afida Nurul Hilma[1], Dian Lestari[2*], Sindy Devila[3].

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, 16424, Indonesia*

[*)]Corresponding author: dian.Lestari@sci.ui.ac.id.

Abstract

In order to find a counting distribution that can handle the condition when the data has no zero-count. Distribution named Zero-truncated Poisson-Lindley distribution is developed. It can handle the condition when the data has no zero-count both in over-dispersion and under-dispersion. In this paper, characteristics of Zero-truncated Poisson-Lindley distribution are obtained and estimate distribution parameters using the maximum likelihood method. Then, the application of the model to real data is given.

***Keywords*:** Maximum Likelihood Method, Over-dispersion, Under-dispersion, Zero-count.

Introduction

Count data can be modeled by a counting distribution. Counting distribution that commonly used to model count data is Poisson distribution. It has been introduced by Poisson, S.D. But Poisson distribution has one characteristic that is the value of mean is equal to variance, named the condition of equal-dispersion. That characteristic of distribution is rarely found in real count data. In general, real count data are having over-dispersion and under-dispersion. As a result, Poisson distribution is no longer appropriate to be used for count data modeling.

ISBN 978-979-19256-3-1 (PDF)

Several distributions have been introduced as an alternative to Poisson distribution on handling the over-dispersion on data (Shanker & Fesshaye, 2015). However, the alternative distributions have higher complexity in the number of parameters. Modification needs to be done in Poisson distribution so that the distribution can represent the condition of the over-dispersion or under-dispersion on data.

In the last few years, there are some modifications to handle dispersion in data. One of the modifications is by mixing Poisson distribution to other alternative distribution. By doing mixing Poisson and Lindley distribution, a new alternative Poisson distribution called Poisson-Lindley has been introduced by Sankaran to model count data (Munuswamy, 1970). Especially count data that have over-dispersed on data. It has been found that Poisson-Lindley distribution is more flexible for analyzing different types of count data than Poisson distribution (Shanker & Fesshaye, 2015). However, Poisson-Lindley distribution cannot handle data that exhibits under-dispersion. On the other hand, there is real data that has no zero-count. Therefore, in order to obtain a more flexible distribution to fit count data that has no zero-count, a modification needs to be done on Poisson-Lindley distribution.

The modification to Poisson-Lindley distribution is by using zero-truncated method. The used of zero-truncated method on modeling count data, especially count data that has no zero-count. It was introduced by (Ghitany et al. 2007) and become famous in the last years. Zero-truncated distributions are certain distributions having support the set of positive integers, (Shanker, R et al. 2015). Its distribution modeling count data that has no zero-count, that is the probability of zero observation from Zero-truncated distribution is equal to zero. Therefore the minimum number of observation from Zero-truncated Poisson-Lindley distribution is one. Its distribution can handle the condition when the data has no zero-count both in over-dispersion and under-dispersion, (Ghitany et al. 2007). Then parameter estimation used maximum likelihood estimation. Finally, from the parameter estimation, we estimate the mean and variance from count data.

Statistical Properties

Let $X|\Lambda$ be a discrete random variable of Poisson distribution having probability mass function (p.m.f), denoted by $p_{X|\Lambda}(x|\lambda)$, where λ is a parameter of $X|\Lambda$. The Poisson distribution with its p.m.f:

$$p_{X|\Lambda}(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \; ; x = 0,1,2, \ldots , \; \lambda > 0. \qquad (1.1)$$
$$= 0 \qquad ; x \text{ elsewhere}$$

Next, let λ be a realization of the random variable Λ have probability density function (p.d.f) of continuous type denoted by $f_\Lambda(\lambda;\theta)$. The modification from Poisson distribution to Poisson-Lindley distribution used mixing method. It arises from the Poisson distribution when the parameter λ follows Lindley with its probability density function (p.d.f):

$$f_\Lambda(\lambda;\theta) = \frac{\theta^2}{1+\theta}(1+\lambda)e^{-\theta\lambda} \; ; \; \lambda > 0, \qquad \theta > 0 \qquad (1.2)$$

(Ghitany et al. 2007)

Then, with mixture method, the unconditional probability mass function of X is given by:

$$p_X(x;\theta) = \int_0^\infty p(x|\lambda)\, f_\Lambda(\lambda;\theta) d\lambda$$

Therefore, the Poisson-Lindley distribution with its probability mass function (p.m.f):

$$p_X(x;\theta) = \frac{\theta^2(x+\theta+2)}{(1+\theta)^{x+3}} \; ; \; x = 0,1,2,3,..; \; \theta > 0 \qquad (1.3)$$

Detailed study of Poisson-Lindley distribution (1.3) has been done by Shanker et al. (2015) and shown that (1.3) is a better model than Poisson distribution on modeling over-dispersion data. In this paper we consider the Zero-truncated Poisson-Lindley distribution with its probability mass function (p.m.f):

$$p_1(x;\theta) = \frac{p_X(x;\theta)}{1 - p_X(0;\theta)}; \quad x = 1,2,3,\dots, \qquad \theta > 0 \tag{1.4}$$

$p_X(x;\theta)$ is the probability mass function from Poisson-Lindley distribution.

Then, using (1.3) and (1.4), the p.m.f of ZTPLD with parameter $\theta > 0$ can be obtained by solving this equation, as a result:

$$p_1(x;\theta) = \frac{\dfrac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}}}{1 - \dfrac{\theta^2(\theta+2)}{(\theta+1)^3}}$$

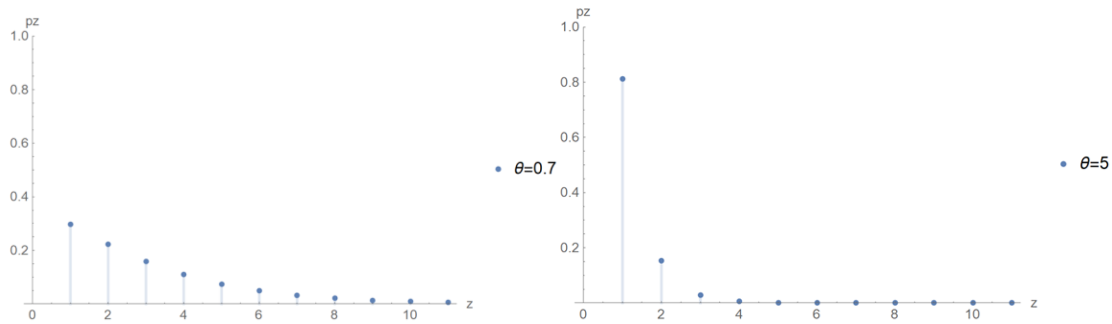
$$= \frac{\theta^2}{\theta^2 + 3\theta + 1} \frac{x+\theta+2}{(\theta+1)^x}$$

From equation above, the probability mass function of ZTPLD distribution is given by:

$$p_1(x;\theta) = \frac{\theta^2}{\theta^2 + 3\theta + 1} \frac{x+\theta+2}{(\theta+1)^x}; \quad x = 1,2,3,\dots, \qquad \theta > 0 \tag{1.5}$$

$$= 0 \qquad\qquad ;\text{elsewhere}$$

It has been proved that the p.m.f of ZTPLD satisfied the properties of discrete probability mass function. Next, figure 1.1 contains the graphic p.m.f of ZTPLD for some value of $\theta > 0$:



ISBN 978-979-19256-3-1 (PDF)

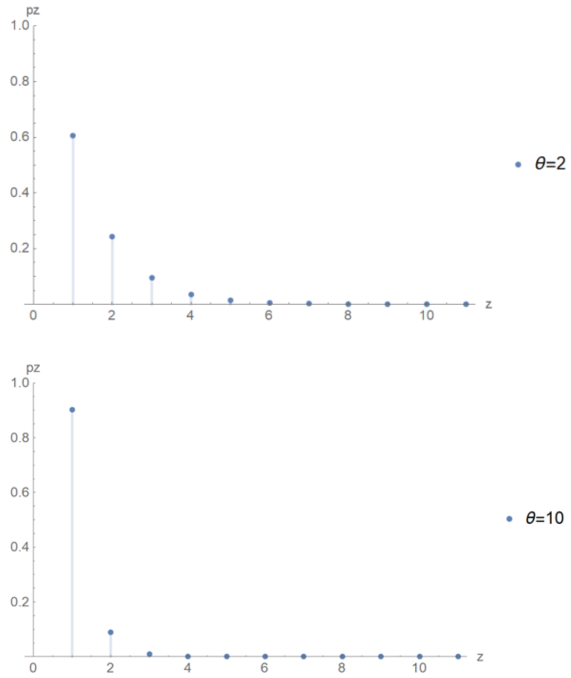


Figure 1.1. Plot *pmf* ZTPLD for some value of $\theta > 0$

The mean and variance of the ZTPLD with the p.m.f (1.5) are given by:

$$\mu = \frac{(\theta+1)^2(\theta+2)}{\theta(\theta^2+3\theta+1)} \tag{1.6}$$

and

$$\sigma^2 = \frac{(\theta+1)^2(\theta^3+6\theta^2+10\theta+2)}{\theta^2(\theta^2+3\theta+1)^2} \tag{1.7}$$

Next, the value of θ has been found based on the value from mean and variance of ZTPLD by setting the condition of equal-dispersion, as a result:

$$\mu = \sigma^2$$

$$\leftrightarrow \frac{(\theta+1)^2(\theta+2)}{\theta(\theta^2+3\theta+1)} = \frac{(\theta+1)^2(\theta^3+6\theta^2+10\theta+2)}{\theta^2(\theta^2+3\theta+1)^2}$$

$$\leftrightarrow \frac{(\theta+1)^2(\theta+2)}{\theta(\theta^2+3\theta+1)} \frac{\theta^2(\theta^2+3\theta+1)^2}{(\theta+1)^2(\theta^3+6\theta^2+10\theta+2)} = 1$$

$$\leftrightarrow (\theta+2).\theta(\theta^2+3\theta+1) = \theta^3+6\theta^2+10\theta+2$$

$$\leftrightarrow \theta^4+5\theta^3+7\theta^2+2\theta = \theta^3+6\theta^2+10\theta+2$$

$$\leftrightarrow \theta^4+4\theta^3+\theta^2-8\theta-2 = 0$$

$$\leftrightarrow \theta(\theta^3 + 4\theta^2 + \theta - 8) = 2 \tag{1.8}$$

By solving equation (1.8) we got the value of $\theta \approx 1.258627$. When it has equal-dispersion, which has the same value of its mean and variance, the value of θ is equal to θ_0=1.258627. Therefore the connection between mean and variance of ZTPLD can be written as follows:

1. $\mu < \sigma^2$ (Over-dispersion) when the value of $\theta < \theta_0$.
2. $\mu = \sigma^2$ (Eqial-dispersion) when the value of $\theta = \theta_0$.
3. $\mu > \sigma^2$ (Under-dispersion) when the value of $\theta > \theta_0$.

Maximum Likelihood Estimator

Consider a random sample $X_1, X_2, \dots . X_n$ of size n from Zero-truncated Poisson-Lindley distribution having p.m.f $p_1(x;\theta)$ which depend to parameter $\theta \epsilon \Omega$, where Ω is the space parameter. The joint p.m.f of $X_1, X_2, \dots . X_n$ is:

$$p(x_1, x_2, \dots, x_n) = p(x_1;\theta).p(x_2;\theta)\dots..p(x_n;\theta)$$

Then, the joint p.m.f could be assumed as a function of θ, its called the joint p.m.f as likelihood function L of the random sample. The likelihood function of the ZTPLD is denoted by:

$$L(\theta; z_1, z_2, \dots, z_n) = \prod_{i=1}^{n} p_1(x;\theta)$$

$$= p(z_1;\theta).p(z_2;\theta)\dots..p(z_n;\theta)$$

$$= \left[\left(\frac{\theta^2}{\theta^2+3\theta+1}\right)\frac{(z_1+\theta+2)}{(\theta+1)^{z_1}}\right]\left[\left(\frac{\theta^2}{\theta^2+3\theta+1}\right)\frac{(z_2+\theta+2)}{(\theta+1)^{z_2}}\right]$$

$$\dots\left[\left(\frac{\theta^2}{\theta^2+3\theta+1}\right)\frac{(z_n+\theta+2)}{(\theta+1)^{z_n}}\right]$$

$$= \left(\frac{\theta^2}{\theta^2+3\theta+1}\right)^n \left(\frac{1}{\theta+1}\right)^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n}(x_i+\theta+2)$$

In maximum likelihood estimation, we find the value of θ that maximized the likelihood or log likelihood function. For ZTPLD, the form of the log likelihood function is given by:

$$\text{Log}L(\theta; x_1, x_2, \dots, x_n) = \ nlog\left(\frac{\theta^2}{\theta^2+3\theta+1}\right) - \sum_{i=1}^{n} x_i \log(\theta+1) + \sum_{i=1}^{n} \log(x_i+\theta+2)$$

The maximum likelihood estimate, $\hat{\theta}$ of θ is the solution of the equation $\frac{dLogL(\theta; x_1, x_2, \dots, x_n)}{d\theta} = 0$. Then, the solution of the non-linear solution from likelihood function is given by:

$$\frac{2n}{\theta} - \frac{(2\theta+3)n}{\theta^2+3\theta+1} - \frac{\sum_{i=1}^{n} x_i}{\theta+1} + \sum_{i=1}^{n} \frac{1}{x_i+\theta+2} = 0 \qquad (1.9)$$

This non-linear equation can be solved by any numerical iteration methods such as Newton-Raphson method, Bisection method, Regula-Falsi method etc or using software RStudio. Beside of that, Ghitany et al., (2007) showed that the maximum likelihood estimation $\hat{\theta}$ is consistent and asymptotically normal.

Applications

In this section, Zero-truncated Poisson-Lindley distribution has been fitted to a number of data-sets using maximum likelihood estimates. It's the number of egg-cells on a flower head. The eggs are easily seen and counted when the flower-head is split open (Finney & Varley, 1955). As part of his intensive study, he had records from sample of flower-heads in 1935 and 1936. In this research, the data based on data that has been recorded on 1936. Table 1.1 contains the results, each sample relating to different flower heads since the process of counting is destructive; the number of empty flower-heads is omitted because a multitude of causes not relevant to the investigation may secure that no eggs are laid (Finney & Varley, 1955).

Tabel 1.1 The number of egg-cells on a flower-head

	The numbers of egg-cells										
	1	2	3	4	5	6	7	8	9	>9	Total
The numbers of flower head	22	18	18	11	9	6	3	0	1	0	88

That characteristic from data Table 1.1 is given by:

Tabel 1.2 The characteristics of data the number of egg-cells on a flower-head

Sample (n)	Mean	Variance	Median	Modus	Std.Deviation
88	3.03	3.344	3.00	1.00	1.829

From Table 1.2, the value of variance is equal to 3.344 greater than 3.03 that is the value of mean, this condition is called over-dispersion on data. Then, the minimum number of observation start from one observation. It indicates that the data has no zero-count. Therefore, Zero-truncated Poisson-Lindley distribution can be used to model this data. By fitting data on Table 1.1 using maximum likelihood estimates, we got the value of $\hat{\theta}$ = 0.71856 used the value of θ_0 that is θ_0=1.258627 based on equation (1.8). The solution of its non-linear solution from likelihood function (1.7) has been calculated with software R Studio. Table 1.3 contains the result of Goodness-of-fit from data the numbers of egg-cells on a flower head. It has been compared to Zero-truncated Poisson distribution.

Table 1.3 The observed and expected number of data

Number of egg-cells	Observed frequency	Expected frequency	
		Zero-truncated Poisson-Lindley distribution	Zero-truncated Poisson distribution
1	22	26.77	51.21
2	18	19.77	25.61
3	18	13.94	17.07
4	11	9.53	12.80
5	9	6.37	10.24
6	6	4.19	8.53
7	3	2.72	7.32
8	0	1.74	6.40
9	1	1.11	5.70
9+	0	0	0

Estimate of parameter	$\hat{\theta}$	0.71856	1
α		0.05	
d.f		4	
χ^2		3.5086	30.5833

Zero-truncated Poisson-Lindley distribution is more suitable to model data the number of egg-cells on a flower head. From the value of chi-square with α=0.05 and d.f=4, the test is rejected when using Zero-truncated Poisson distribution. The other characteristic from this data, that is the estimate of pmf ZTPLD from data above with $\hat{\theta}$ = 0.71856 is given by:

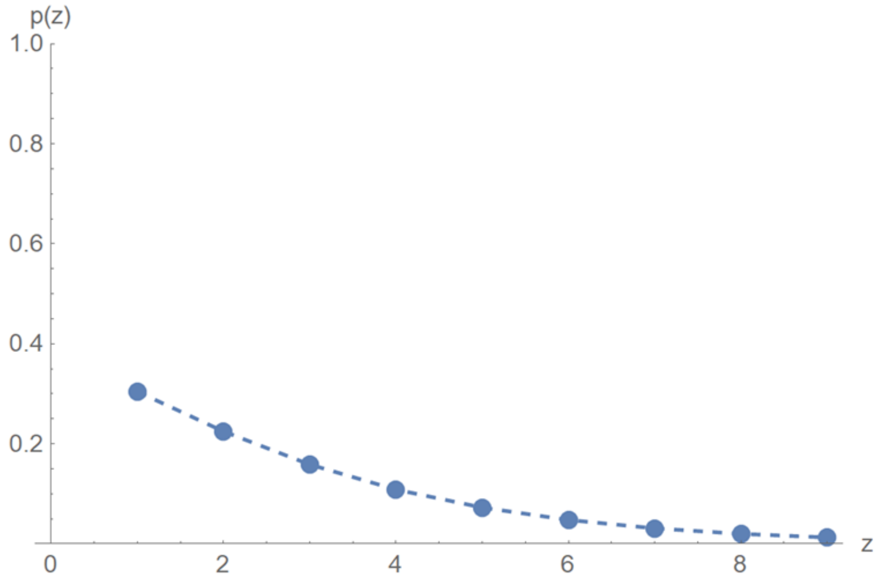


Figure 1.2 The estimate of p.m.f of ZTPLD from data

Therefore, the Zero-truncated Poisson distribution with its probability mass function (p.m.f):

$$p(w;\theta) = \frac{\theta^w}{(e^\theta - 1)w!} \ ; \ w = 1,2,3,..; \ \theta > 0 \tag{1.9}$$

Detailed study of Zero-truncated Poisson distribution (1.9) has been done by (Shanker et al. 2015). Next,using (1.9) the estimate of pmf ZTPD from data above with $\hat{\boldsymbol{\theta}}$ = 1 is given by:

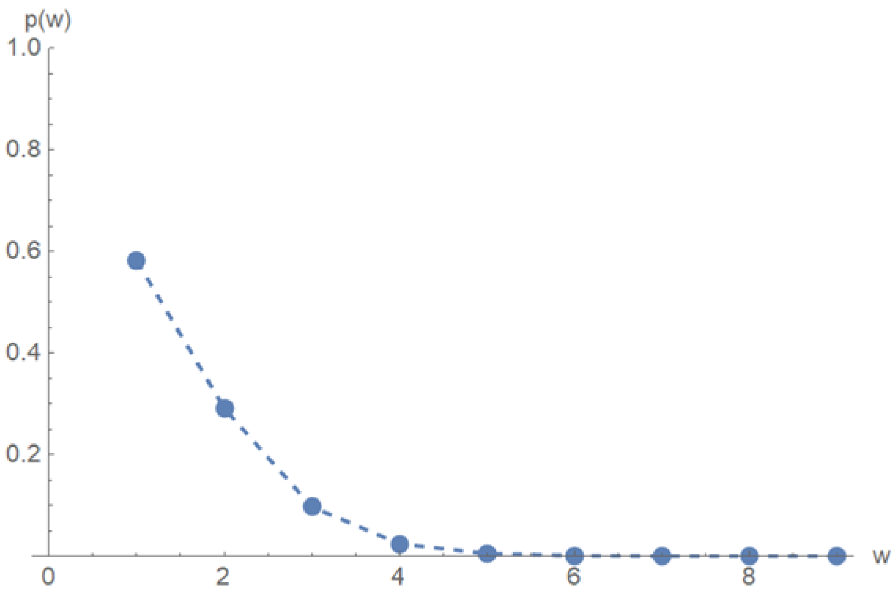


Figure 1.3 The estimate of p.m.f of ZTPD from data

Based on graphic pmf of Zero-truncated Poisson-Lindley distribution compared to Zero-truncated Poisson distribution, we can conclude that Zero-truncated Poisson-Lindley is more suitable to model data the number of egg-cells on a flower-head.

Conclusion

Zero-truncated Poisson-Lindley distribution arises first by doing mixing between Poisson distribution and Lindley distribution, used mixture method. That distribution is named Poisson-Lindley distribution. Then, next modification is by applying the probability at zero from Poisson-Lindley distribution having its probability zero. That modification called the zero-truncated method. Therefore, the newly formed distribution is named Zero-truncated Poisson-Lindley distribution.

References

E Bruce Brooks, University of Massachusetts at Amherst. (2000, March 22). Retrieved Aug 24, 2007, from https://www.umass.edu/wsp/resources/poisson/

Finney, D., & Varley, G. (1955, September). An Example of The Truncated Poisson Distribution. *11*. International Biometric Society.

Ghitany, M., Al-Mutiari, D., & Nadarajah, S. (2007). Zero-truncated Poisson-Lindley distribution. *Mathematics and Computers in Simulation*, 279-287.

Ghitany, M., Atieh, B., & Nadarajah, S. (2007). Lindley distribution and its application. *Mathematics and Computers in Simulation*, 493-506.

Klugman, S., Panjer, H., & Willmot, G. (2012). *Loss Models: from data to decision 4th Edition.* New Jersey: Wiley seried in probability and statistics.

Munuswamy, S. (1970, Mar). The Discrete Poisson-Lindley Distribution. *26*. International Biometric Society.

Shanker, R., & Fesshaye, H. (2015, April 27). On Poisson-Lindley Distribution and Its Applications to Biological Sciences. *Biometrics & Biostatistics International Journal*, 1-5.

Shanker, R., Fesshaye, H., Selveraj, S., & Yemane, A. (2015, July 22). On Zero-Truncatation of Poisson and Poisson-Lindley Distributions and Their Applications. *Biometrics & Biostatistics International Journal, 2*(6), 1-14.