

## Original Article

## Weighted negative binomial-Poisson Lindley with application to genetic data

Hossein Zamani<sup>1\*</sup>, Noriszura Ismail<sup>2</sup>, Marzieh Shekari<sup>1</sup><sup>1</sup>Department of Mathematics and Statistics, Faculty of Science, University of Hormozgan, Bandarabbas, Iran<sup>2</sup>School of Mathematical Sciences, Universiti Kebangsaan Malaysia, Selangor, Malaysia

## ARTICLE INFO

Received 18.12.2017

Revised 27.12.2017

Accepted 12.03.2018

## Key words:

Weighted distribution;  
Poisson distribution;  
Discrete distribution;  
Mixed distribution,  
Mixed Poisson

## ABSTRACT

**Background & Aim:** Mixed Poisson and mixed negative binomial distributions have been considered as alternatives for fitting count data with over-dispersion. This study introduces a new discrete distribution which is a weighted version of Poisson-Lindley distribution.**Methods & Materials:** The weighted distribution is obtained using the negative binomial weight function and can be fitted to count data with over-dispersion. The p.m.f., p.g.f. and simulation procedure of the new weighted distribution, namely weighted negative binomial-Poisson-Lindley (WNBPL), are provided. The maximum likelihood method for parameters estimation is also presented.**Results:** The WNBPL distribution is fitted to several datasets, related to genetics and compared with the Poisson distribution. The goodness of fit test shows that the WNBPL can be a useful tool for modeling genetics datasets.**Conclusion:** This paper introduces a new weighted Poisson-Lindley distribution which is obtained using negative binomial weight function and can be used for fitting over-dispersed count data. The p.m.f., p.g.f. and simulation procedure are provided for the new weighted distribution, namely the weighted negative binomial-Poisson Lindley (WNBPL) to better inform parents from possible time of occurrence reflux and treatment strategies.

## Introduction

Mixed Poisson and mixed negative binomial distributions have been considered as alternatives for fitting count data with over-dispersion (1-5). Several examples of mixed Poisson and mixed negative binomial distributions can be found in several statistical literatures (6-23), such as negative binomial which is obtained as a mixture of Poisson and gamma, Poisson-Lindley (6, 18), Poisson-lognormal (1), Poisson-inverse Gaussian (24, 25), negative binomial-Pareto (12), negative binomial-inverse Gaussian (7), negative binomial-Lindley (26, 11), Poisson-exponential (2), Poisson-weighted exponential (27), two parameter Poisson-Lindley (19) and Poisson-Janardan distributions (20).

Besides mixed distributions, weighted distributions have also been considered as

alternatives for fitting count data with over-dispersion, and can be generally obtained by multiplying a count distribution with a weight function. To derive a new weighted distribution, let  $X$  be a count random variable with p.m.f.  $P(X = k)$ , where  $k \in N_0 = \{0, 1, 2, \dots\}$ . Let  $\omega(k)$  be a non-negative function on  $N_0$  having a finite expectation

$$E(\omega(X)) = \sum_{k=0}^{\infty} \omega(k)P(X = k) < \infty,$$

where the weight function  $\omega(k)$  can be used to adjust the probability when  $X = k$  occur. Thus, the weighted version of r.v.  $X$ , which is the realization of count r.v.  $Y$ , has the following p.m.f:

$$P(Y = k) = p(k; \theta) = \frac{\omega(k)P(X = k)}{E(\omega(X))}, k \in N_0. \quad (1)$$

The most popular weighted count distributions are the weighted Poisson (WP) distributions which are obtained when the initial count r.v.,  $X$ , follows a Poisson distribution. The initial concept of WP distribution was introduced in (16), which lead to several more

\* Corresponding Author: Hossein Zamani, Postal Address: Department of Mathematics and Statistics, Faculty of Science, University of Hormozgan, Bandarabbas, Iran.

Email: zamani.huni@hormozgan.ac.ir

recent and different types of WP distributions derived and analyzed in other studies. Examples of a more recent WP distributions can be found in (3,17, 22).

In recent studies, some authors used particular weights for deriving new versions of weighted distributions. Such examples can be found in (13) who used the Poisson weight function  $\omega(k; \varphi) = \varphi^k e^{-\varphi} (k!)^{-1}$ , Kokonendji (10) who utilized the binomial weight function  $\omega(k; \varphi) = 1 - (1 - \varphi)^k$ , and the negative binomial weight function  $\omega(k; \varphi) = \binom{\varphi + k - 1}{k}$  which was applied by (8). A more detailed study of weighted distributions and weight functions can be found in (15).

The objective of this study is to introduce a new discrete weighted distribution based on the Poisson-Lindley distribution. The weighted distribution, namely the weighted negative binomial-Poisson Lindley (WNBPL), is weighted with the negative binomial weight function and can be used as an alternative for fitting count data with over-dispersion. The rest of this paper is organized as follows. Section 2 provides the p.m.f., p.g.f. and simulation procedure for the WNBPL. Maximum likelihood method for parameters estimation is provided in Section 3. Several numerical illustrations are provided in Section 4, where the Poisson, and WNBPL are fitted to a few datasets.

### Methods

#### Weighted Poisson-Lindley Negative Binomial (WPLN)

**P.m.f., p.g.f., mean, and variance:** Assume r.v.  $Y | \lambda$  follows Poisson distribution with p.m.f:

$$p(y | \lambda) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots, \quad (2)$$

and parameter  $\lambda$  is distributed as Lindley with parameter  $\theta$ :

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

The Poisson-Lindley (PL) distribution is obtained by mixing the Poisson and Lindley distributions, and the p.m.f. is:

$$p(y) = \frac{\theta^2 (y + \theta + 2)}{(1 + \theta)^{y+3}}, \quad y = 0, 1, 2, 3, \dots, \quad (4)$$

with mean and variance:

$$E(Y) = \frac{\theta + 2}{\theta(\theta + 1)}, \quad Var(Y) = \frac{\theta^3 + 4\theta^2 + 6\theta + 2}{\theta^2(\theta + 1)^2}.$$

Using  $\theta + 1 = \frac{1}{p}$  for re-parameterization, the

PL p.m.f. in (4) can be re-written as:

$$p(y) = (1 - p)^2 p^y (1 + p + py) \quad y = 0, 1, 2, 3, \dots \quad (5)$$

A new discrete distribution can be easily obtained by inserting the negative binomial weight function  $\omega(k; r) = \binom{r + k - 1}{k}$  and the

PL p.m.f. (5) into the weighted equation in (1). The new distribution, namely the WNBPL, has the following p.m.f:

$$P(Y = k) = p(k) = \binom{r + k - 1}{k} \frac{(1 - p)^{r+1} p^k (1 + p + pk)}{(1 - p^2 + rp^2)}, \quad k = 0, 1, 2, \dots, \quad 0 < p < 1 \quad (6)$$

with mean and variance:

$$\mu = \frac{rp^2 - p^2 - p}{1 + rp^2 - p^2} + \frac{p(r + 1)}{1 - p} \quad (7)$$

$$\sigma^2 = \frac{(rp^2 - p^2 - p)(p + 1)}{(1 + rp^2 - p^2)^2} + \frac{p(r + 1)}{(1 - p)^2}$$

The p.g.f. can be obtained in a closed form, and is given by:

$$G_Y(t) = E(t^Y) = \frac{(1 - p^2 t + rp^2 t) + p(1 - t) \left( \frac{1 - p}{1 - pt} \right)^{r+1}}{(1 - p^2 + rp^2)} \quad (8)$$

#### Over-dispersion

In statistics, cases of over-dispersion can be determined by comparing the mean and variance, where a distribution is known to be over-dispersed if the variance is greater than the mean. For WNBPL, the variance and mean can be written as:

$$\sigma^2 - \mu = \frac{p^2(r + 1)}{(1 - p)^2} - \frac{(rp^2 - p^2 - p)^2}{(1 + rp^2 - p^2)^2},$$

so that we can determine whether the term  $\frac{(rp^2 - p^2 - p)^2}{(1 + rp^2 - p^2)^2}$  is less than one for all values of

$p$  and  $r$ . If  $\frac{(rp^2 - p^2 - p)^2}{(1 + rp^2 - p^2)^2}$  is less than one, then  $\frac{p^2(r+1)}{(1-p)^2}$  is greater than one, indicating that  $\sigma^2 - \mu$  is greater than zero. Therefore, the variance of WNBPL is greater than the mean, and the distribution can be used to handle over-dispersed count data.

Figure 1 shows the p.m.f. of WNBPL for different values of  $(r, p)$ . The graphs indicate that the distribution can be considered as an alternative for over-dispersed count data.

**Random data generation**

P.m.f. (6) indicates that  $WNBPL(r, p)$  is a mixture of negative binomial distributions, which can be written as:

$$p(k) = \frac{1-p^2}{1-p^2+rp^2} NB(r, 1-p) + \frac{rp^2}{1-p^2+rp^2} NB(r+1, 1-p)$$

Therefore, the  $WNBPL(r, p)$  random samples can be generated via the weighted negative binomial approach.

We analyze the performance of ML estimates of  $WNBPL(r, p)$  based on 1000 simulations. The average estimators, average mean square

errors and average standard errors of the ML estimates for several sample sizes,  $n$ , and several initial values  $(r, p)$ , are provided in Table 1. The results show that increasing the sample size is an effective way of decreasing the standard errors of parameters. As shown in this table, the MSEs decrease when the sample size increase, and thus, suggesting the consistency of the proposed model.

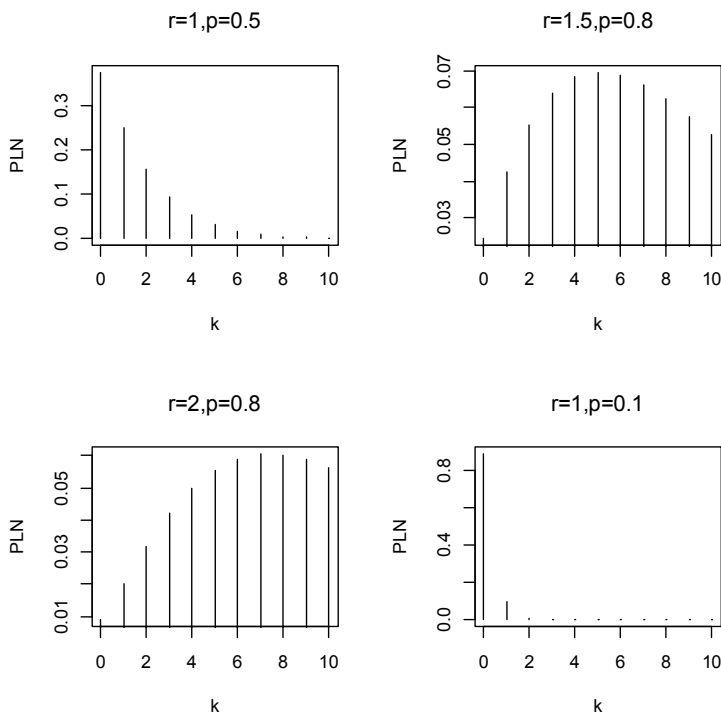
**Parameter Estimation**

Let  $Y_1, Y_2, \dots, Y_n$  be an i.i.d. random sample drawn from WNBPL distribution, with observed values  $k_1, k_2, \dots, k_n$ . The log-likelihood is:

$$\ln L(r, p) = \ell(r, p) = n(r+1)\ln(1-p) + \sum_{i=1}^n k_i \ln p + \sum_{i=1}^n \ln(1+p+pk_i) - n \ln(1-p^2+rp^2) + \sum_{i=1}^n \binom{r+k_i-1}{k_i}$$

By partially differentiating the log-likelihood with respect to  $p$  and  $r$ , we obtained:

$$\frac{\partial \ell(r, p)}{\partial p} = -\frac{n(r+1)}{1-p} + \frac{n\bar{k}}{p} + \sum_{i=1}^n \frac{1+k_i}{1+p+pk_i} - \frac{2np(r-1)}{1-p^2+rp^2}$$



**Figure 1.** P.m.f. of WNBPLN distribution for different values of  $(r, p)$

**Table 1.** Average estimates, average MSE and average standard error (1000 simulation)

n	Initial values		Average estimates		Average MSE		Average Std	
	r	p	$\hat{r}$	$\hat{p}$	mse( $\hat{r}$ )	mse( $\hat{p}$ )	se( $\hat{r}$ )	se( $\hat{p}$ )
50	0.3	0.1	1.227	0.130	1.773	0.046	0.955	0.213
	0.6	0.6	0.629	0.567	0.105	0.010	0.323	0.094
	0.2	0.8	5.924	0.179	105.247	0.404	8.513	0.147
75	0.3	0.1	2.051	0.161	8.413	0.048	2.312	0.211
	0.6	0.6	0.594	0.572	0.062	0.006	0.249	0.076
	0.2	0.8	4.081	0.189	58.593	0.389	6.597	0.131
100	0.3	0.1	1.948	0.177	7.543	0.050	2.197	0.210
	0.6	0.6	0.581	0.575	0.041	0.005	0.203	0.066
	0.2	0.8	3.490	0.184	43.061	0.389	5.677	0.118
125	0.3	0.1	1.577	0.196	5.541	0.094	.970	0.200
	0.6	0.6	0.567	0.577	0.027	0.004	0.163	0.058
	0.2	0.8	2.864	0.191	31.381	0.382	4.927	0.112
150	0.3	0.1	1.402	0.198	4.614	0.044	1.844	0.186
	0.6	0.6	0.555	0.581	0.021	0.003	0.140	0.050
	0.2	0.8	2.694	0.190	27.342	0.382	4.595	0.107

$$\frac{\partial \ell(r, p)}{\partial r} = n \ln(1-p) - \frac{np^2}{1-p^2+rp^2} + \frac{\partial}{\partial r} \sum_{i=1}^n \ln \binom{r+k_i-1}{k_i}$$

Klugman et.al (2012) showed that the term  $\frac{\partial}{\partial r} \sum_{x=0}^k \ln \binom{r+x-1}{x}$  can be simplified into:

$$\frac{\partial}{\partial r} \sum_{x=0}^k \ln \binom{r+x-1}{x} = \sum_{x=0}^k \sum_{m=0}^{x-1} \ln(r+m).$$

Therefore, the partial differentiation  $\frac{\partial \ell(r, p)}{\partial p}$  can be written in a simpler form, which is:

$$\frac{\partial \ell(r, p)}{\partial r} = n \ln(1-p) - \frac{np^2}{1-p^2+rp^2} + \sum_{i=1}^n \sum_{m=0}^{k_i-1} \ln(r+m)$$

ML estimates  $(\hat{r}, \hat{p})$  can be obtained numerically using statistical packages such as R 3.3.1 with *nlminb* command. Under regularity conditions, the ML estimates  $(\hat{r}, \hat{p})$  for

WNBPL has a bivariate normal distribution with mean  $(r, p)$  and variance-covariance matrix  $[I(r, p)]^{-1}$ , where  $I(r, p)$  is the Fisher information matrix, which is given as:

$$I(\hat{r}, \hat{p}) = \begin{bmatrix} E \left[ -\frac{\partial^2 \ell(r, p)}{\partial p^2} \right] & E \left[ -\frac{\partial^2 \ell(r, p)}{\partial r \partial p} \right] \\ E \left[ -\frac{\partial^2 \ell(r, p)}{\partial r \partial p} \right] & E \left[ -\frac{\partial^2 \ell(r, p)}{\partial r^2} \right] \end{bmatrix}$$

## Results

### Application to Genetic Data

The Poisson distribution is a tool which widely used in modeling count data in many areas such as ecology and genetics. But the Poisson model has a good fitting on the equi-dispersion datasets. For the case of the over dispersed data, i.e. the data in which the variance is greater than the mean, the alternatives distributions are used. In this case, the mixed

**Table 2.** Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by streptonigrin, exposure -60

Class/Exposure ( $\mu g   kg$ )	Observed frequency	Poisson	WNBPL
0	413	374	412.9
1	124	177.4	124.3
2	42	42.1	41.9
3	15	6.6	14.4
4	5	0.8	4.9
5	0	0.1	1.7
6	2	0.0	0.9
Parameters		$\hat{\lambda} = 0.4742$	$\hat{p} = 0.3207$ $\hat{r} = 0.7555$
-ln L		582.67	556.18
AIC		1167.34	1116.36
chi-square		48.169	0.06
p-value of chi-square		0.00	0.97

**Table 3.** Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by streptonigrin, exposure -70

Class/Exposure ( $\mu g   kg$ )	Observed frequency	Poisson	WNBPL
0	200	172.5	199.2
1	57	95.4	61.6
2	30	26.4	23.6
3	7	4.9	9.3
4	4	0.7	3.8
5	0	0.1	1.6
6	2	0.0	0.9
Parameters		$\hat{\lambda} = 0.553$	$\hat{p} = 0.384$
			$\hat{r} = 0.630$
-ln L		323.44	302.67
AIC		648.88	609.34
chi-square		29.68	3.82
<i>p</i> -value of chi-square		0.00	0.15

**Table 4.** Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by streptonigrin, exposure -70

Class/Exposure ( $\mu g   kg$ )	Observed frequency	Poisson	WNBPL
0	155	127.8	155.3
1	83	109	80.2
2	33	46.5	36.9
3	14	13.2	16.2
4	11	2.8	6.9
5	3	0.5	2.9
6	1	0.2	1.6
Parameters		$\hat{\lambda} = 0.853$	$\hat{p} = 1.154$
			$\hat{r} = 0.354$
-ln L		400.46	382.89
AIC		802.9	769.78
chi-square		24.97	2.18
<i>p</i> -value of chi-square		0.00	0.335

Poisson distribution or the sized biased distribution can be used for modeling the over dispersed datasets. In this section three genetic data sets which was used by (21) are given. The Poisson and the WNBPL are fitted to the datasets and compared using the AIC criteria and the goodness of fit test.

In Table 2-4 the Poisson and the WNBPL are fitted to the Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by streptonigrin for the exposure of (-60, -70, -90)  $\mu g | kg$  respectively which considered by Catcheside et al. It can be seen that based on the AIC and the goodness of fit test, the WNBPL has a better fit compared to the Poisson distribution.

### Discussion

This paper introduces a new weighted Poisson-Lindley distribution which is obtained using negative binomial weight function and can be used for fitting over-dispersed count data. The p.m.f., p.g.f. and simulation procedure are provided for the new weighted distribution, namely the weighted negative binomial-Poisson

Lindley (WNBPL). The WNBPL( $r, p$ ) can also be shown to be equivalent to a mixture of negative binomial distributions, and thus, allowing the random samples to be generated via weighted approach. The estimation procedures of WNBPL parameters via the maximum likelihood are also shown. For numerical illustrations, the WNBPL distribution is fitted to three sets of genetic count data, and the results are compared to the Poisson distribution. Based on chi-square and log likelihood of the fitted models, the WNBPL distributions provide significant improvements over the Poisson, and the WNBPL provide the largest log likelihood and the smallest chi-square. Considering the straightforward manner of obtaining its MLE estimators, the WNBPL can be considered as an alternative model for fitting over-dispersed count data.

### Conclusion

To summarize, the proposed method in this study is recommended for modeling over-dispersed count data as an alternative to the Poisson and negative binomial distributions.

### Conflicts of interests

The authors declare that there is no conflict of interest regarding the publication of this article.

### References

1. Bulmer MG. On fitting the Poisson-lognormal distribution to species-abundance data. *Biometrics*. 1974;101-110.
2. Cancho VG, Louzada-Neto F, Barriga, GDC. The Poisson-exponential lifetime distribution. *Computa Stat Data Analys*. 2011;55(1):677-686.
3. Castillo J, Perez-Casany, M. Overdispersed and underdispersed Poisson generalizations. *J Stat Plan Infer*. 2005;134:486-500.
4. Catcheside DG, Lea DE, Thoday JM. Types of chromosome structural change induced by the irradiation on *Tradescantia* microspores. *J Genet*. 1946;47:113-136.
5. Denuit M. A new distribution of Poisson-type for the number of claims. *Astin Bull*. 1997;27(2): 229-242.
6. Ghitan ME, Atieh B, Nadarajah S. Lindley distribution and its application. *Math Comput Simulat*. 2008;78:493-506.
7. Gomez-Deniz E, Sarabia JM, Calderin-Ojeda E. Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance: Math Econ*. 2008; 42:39-49.
8. Hussain T, Aslam M, Ahmad M. A two parameter discrete Lindley Distribution. *Revista Colomb Estad*. 2016;39(1):45-61.
9. Klugman SA, Panjer HH, Willmot GE. *Loss models: from data to decisions*. John Wiley & Sons; 2012 Jan 25.
10. Kokonendji CC, Casany MP. A Note on Weighted Count Distributions. *J Stat Theory Appl*. 2012;11(4):337-352.
11. Lord D, Geedipally SR. The negative binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. *Accident Anal. Prevent*. 2011;43:1738-1742.
12. Meng S, Wei Y, Whitmore GA. Accounting for individual overdispersion in a bonus-malus system. *ASTIN Bull*. 1999;29:327-337.
13. Neel JV, Schull WJ. *Human Heredity*. University of Chicago Press, Chicago. 1966:211-227.
14. Shanker R, Fesshaye H. On Poisson Lindley Distribution and Applications to Biological Sciences. *Biometr Biostat Int J*. 2015;2(4).
15. Patil GP, Rao CR, Ratnaparkhi MV. On discrete weighted distribution and their use in model choice for observed data. *Commun Stat Theory Math*. 1986;15(3):907-918.
16. Rao CR. On discrete distributions arising out of methods of ascertainment. *Ind J Stat*. 1965 Dec 1:311-24.
17. Ridout MS, Besbeas P. An empirical model for underdispersed count data. *Stat Model*. 2004 Apr;4(1):77-89.
18. Sankaran M. 275. note: The discrete poisson-lindley distribution. *Biometrics*. 1970 Mar 1:145
19. Shanker R, Mishra A. A two-parameter Poisson-Lindley distribution. *Int J Stat Syst*. 2014;9(1):79-85.
20. Shanker R, Sharma S, Shanker U, Shanker R, Leonida TA. The discrete Poisson-Janardan distribution with applications. *Int J Soft Comput Eng*. 2014;4(2):31-3.
21. Shanker R, Hagos F. On Poisson-Lindley distribution and its applications to biological sciences. *Biometr Biostat Int J*. 2015;2(4):1-5.
22. Shmueli G, Minka TP, Kadane JB, Borle S, Boatwright P. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *J Royal Stat Soc*. 2005 Jan 1;54(1):127-42.
23. Simon LJ. Fitting negative binomial distributions by the method of maximum likelihood. *Proceed Casualty Actuar Soc*. 1961;48:45-53.
24. Tremblay L. Using the Poisson inverse Gaussian in bonus-malus systems. *ASTIN Bull: J IAA*. 1992 May;22(1):97-106.
25. Willmot GE. The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scand Actuar J*. 1987 Jul 1;1987(3-4):113-27.
26. Zamani H, Ismail N. Negative binomial-Lindley distribution and its application. *J Math Stat*. 2010;6(1):4-9.
27. Zamani H, Ismail N, Faroughi P. Poisson-weighted exponential univariate version and regression model with applications. *J Math Stat*. 2014;10(2):148-54.