The negative binomial-weighted Lindley distribution

Sunthree Denthet[*] and Pramoch Promin

College of Industrial Technology King Mongkut's University of Technology North , Thailand

CHRONICLE	ABSTRACT
	This paper proposes a new distribution named the negative binomial-weighted Lindley. The study uses the maximum likelihood estimation to estimate the parameters of the proposed distribution and compares the performance of the new method with other distributions. The study finds that the negative binomial-weighted Lindley distribution, obtained by the mixing the negative binomial distribution with the weighted Lindley distribution is another mixed negative binomial distribution and may provide an appropriate fit for data estimation with overdispersion. Some characteristics of the proposed distribution, such as mean and variance are also derived.

1. Introduction

A count data distribution is only non-negative integers in its domain. We typically use the count data distribution to model the number of occurrences of a certain event. The Poisson and negative binomial (NB) distributions are the count data distribution for examples. The standard distribution for modeling count data has been the Poisson distribution, which is a proper model for counting the number of occurrences over a time interval at random when not many occurrences are observed within a short period of time. They occur at a constant rate through time, and one occurrence of the phenomenon does not alter the probability of any future occurrence (Rainer, 2008; Team, 2015). Let $X \sim Poisson(\lambda)$ be a Poisson distributed with parameter λ. The probability mass function (pmf) of X is given by

$$f(x) = \frac{\exp(-\lambda)\lambda^x}{x!}, \qquad x = 0, 1, 2, ..., \lambda > 0. \tag{1}$$

Then, the mean and variance are given by $E(X) = \lambda$ and $Var(X) = \lambda$. Equality of mean and variance, called equal dispersion, is a classic characteristic of the Poisson distribution. Moreover, there are other categories of dispersion which are overdisperssion when the variance is greater than the mean and underdispersion where the variance is smaller than the mean (Haight, 1967). The NB distribution is a

* Corresponding author.
E-mail address: srd_kmutnb@hotmail.com (S. Denthet)

popular alternative distribution for modelling overdispersed count data because it is more flexible in accommodating overdispersion in comparison with the Poisson model. The NB distribution is a mixture of Poisson distribution by mixing the Poisson and gamma distribution. Applications using the NB distribution can be found in many areas, for instance, economics, accident statistics, biostatistics and actuarial science. The problem of overdispersion is usually solved by introducing mixed NB distribution. In several studies, it is shown that mixed NB distribution provides better fit on count data compared with the Poisson and the NB distribution. These include the Poisson-inverse Gaussian (Klugman et al., 2008), negative binomial-inverse Gaussian (Gómez-Déniz et al., 2008), negative binomial-Lindley (Zamani & Ismail, 2010), negative binomial-Beta Exponential (Pudprommarat et al., 2012), and negative binomial-Erlang (Kongrod et al., 2014). The Lindley distribution has been generalized by many researchers in recent years. The Lindley distribution is the mixture of exponential (θ) and Gamma $(2,\theta)$ distributions (Lindley, 1958). Subsequently Ghitany et al. (2008) investigated Lindley distribution in the context of reliability analysis. Subsequently, a weighted Lindley (WL) distribution is proposed for modelling survival data. A random variable X follows the WL distribution with parameters $\theta > 0$ and $\theta > \phi$ and the probability density function (pdf) is follows,

$$f(x) = \frac{(1-\phi)^2}{(\theta-\phi+1)}(1+x)\exp(-x(\theta-\phi)), \quad \text{for } x > 0. \tag{2}$$

Let $X \sim \text{WL}(\theta,\phi)$, then its moment generating function (mgf) of X is given by

$$M_X(t) = \frac{(\theta-\phi)(\theta-\phi-t+1)}{(\theta-\phi+1)(\theta-\phi-t)^2}. \tag{3}$$

Some plots of the WL pdf with some specified values of θ and ϕ are shown in Fig. 1.
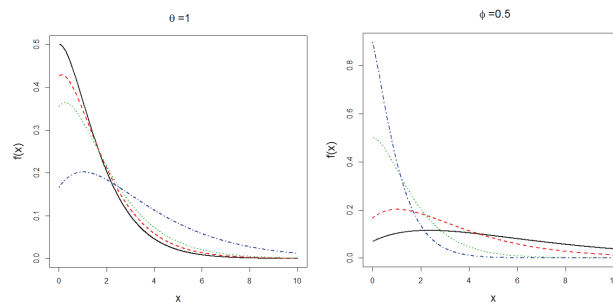
Fig. 1. Some pdf plots of the WL distribution

In this research, a count distribution, which is represented as an alternative distribution for overdispersed count data, namely the negative binomial- weighted Lindley (NB-WL) distribution is developed. The NB-WL distribution is a mixture of the NB and WL distributions. The method is more flexible alternative to the Poisson and NB distribution. Some of the characteristics of the proposed distribution can be studied through factorial moments, e.g., mean and variance. The parameters of the proposed distributions are estimated by using the maximum likelihood estimation (MLE). The MLE is a popular technique for estimating parameter of a given function which makes that likelihood function a maximum and it is also a powerful and unbiased estimation in estimating parameters (Hamid, 2014). The proposed distribution is compared with the performance of Poisson and NB distributions.

2. Methodology

2.1 Research objectives

The objectives of this research are to propose a new mixed distributions, to derive the parameter estimation of the proposed distributions by using the MLE method and compares the efficiencies of the proposed distribution with other distributions for count data analysis.

2.2 The materials

The materials of this research are as high performance personal computer for running the coded program. The maximum likelihood estimates $\hat{r}, \hat{\theta}$ and $\hat{\phi}$ for the parameters r, θ and ϕ respectively, are taken by solving iteratively differential equations to zero. These differential equations are not in closed form and a numerical method can be employed to obtain the expectations of them. The MLE solution of $\hat{r}, \hat{\theta}$ and $\hat{\phi}$ can be obtained by solving the resulting equations simultaneously using optim function in R language.

2.3 The methods

The methods of the research are to investigate pmf and some properties of the NB-WL distribution. To estimate the parameters of the NB-WL distribution, a MLE method is implemented. Random variate generation of the NB-WL distribution is derived and application of the NB-WL distribution to real data set has been studied by comparing with the Poisson and NB distributions using the Kolmogorov-Smirnov (K-S) from the dgof package of R language (Arnold & Emerson, 2011).

3. Results

This section presents the results of the research and provides the probability mass function (pmf) of the proposed distribution. Moreover, some characteristics including the plots of the pmf with various values of parameters, parameter estimation, random variate generation, and application of the proposed distribution to real dataset are included in each part.

3.1 The propose Distribution

We propose a new mixed NB distribution which is an NB-WL distribution obtained by mixing the NB distribution with a WL distribution. The distribution has three parameters, namely, r, θ and ϕ. We begin with a general definition of the NB-WL distribution which will consequently reveal its the probability mass function (pmf). Fig.2. displays the NB-WL pmf plots with some specified parameter values of r, θ and ϕ.

Definition 1. Let $X|\lambda$ be a random variable following a NB distribution with parameters r and $p = \exp(-\lambda),\ X|\lambda \sim \mathrm{NB}(r, p = \exp(-\lambda))$. If λ is distributed as the WL distribution with positive parameters θ and ϕ, denoted by $\lambda \sim \mathrm{WL}(\theta, \phi)$, then X is called a NB-WL random variable.

Theorem 1. Let $X \sim \mathrm{NB\text{-}WL}(r, \theta, \phi)$. The pmf of X is given by

$$f(x; r, \theta, \phi) = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \frac{(\theta-\phi)(\theta-\phi+r+j+1)}{(\theta-\phi+1)(\theta-\phi+r+j)^2}, \quad x = 0, 1, 2, \ldots \tag{4}$$

where $\theta > 0$ and $\theta > \phi$.

Proof. If $X|\lambda \sim \mathrm{NB}(r, p = \exp(-\lambda))$ and $\lambda \sim \mathrm{WL}(\theta, \phi)$, then the pmf of X can be obtained by

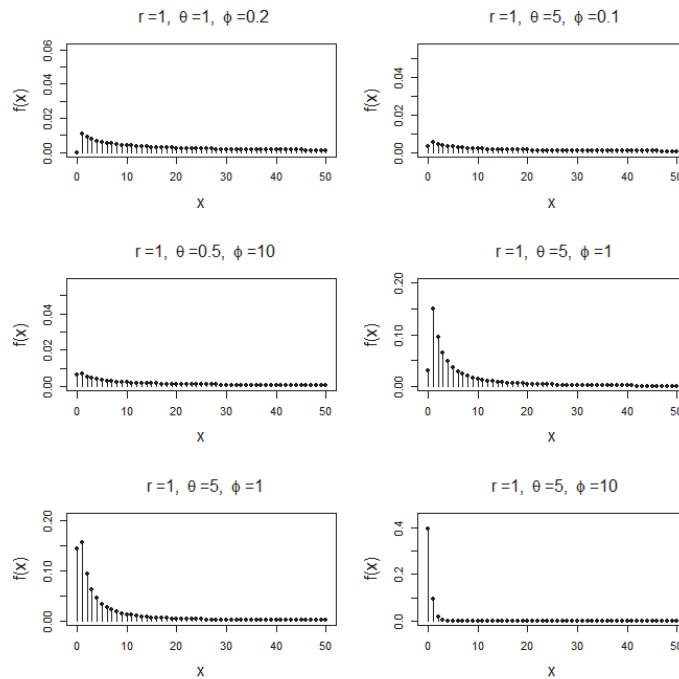$f(x) = \int_0^\infty f_1(x|\lambda) g(\lambda; \theta, \phi) d\lambda$, where $f_1(x|\lambda)$ is express as

$$f_1(x|\lambda) = \binom{r+x-1}{x} \exp(-\lambda r)(1 - \exp(-\lambda))^x = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \exp(-\lambda(r+j)). \tag{5}$$

By substituting $f_1(x|\lambda)$ into $f(x) = \int_0^\infty f_1(x|\lambda)g(\lambda;\theta,\phi)d\lambda$, thus

$$f(x) = \binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left(\int_0^\infty \exp(-\lambda(r+j))g(\lambda;\theta,\phi)d\lambda\right) = \binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j M_\lambda(-(r+j)). \tag{6}$$

Substituting $M_\lambda(-(r+j))$ the mgf of the WL distribution in the equation above, the pmf of the NB-WL(r,θ,ϕ) is given as

$$f(x;r,\theta,\phi) = \binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{(\theta-\phi)(\theta-\phi+r+j+1)}{(\theta-\phi+1)(\theta-\phi+r+j)^2}, \tag{7}$$

Fig. 2. The pmf of the NB-WL distribution of some specified values of r, θ and ϕ

3.2 Characteristics of the NB-WL distribution

Some characteristics of the NB-WL distribution will be discussed as follows. The factorial moment of the NB-WL distribution is introduced. Some of the most important structures and characteristics of the NB-WL distribution can be studied through factorial moments.

Theorem 2. If $X \sim$ NB-WL(r,θ,ϕ). the factorial moment of order a of X is

$$\mu_a'(X) = \frac{\Gamma(r+a)}{\Gamma(r)}\sum_{j=0}^{a}\binom{a}{j}(-1)^j\frac{(\theta-\phi)(\theta-\phi+r+j+1)}{(\theta-\phi+1)(\theta-\phi+r+j)^2}, \quad x = 0,1,2,... \tag{8}$$

for $\theta > 0$ and $\theta > \phi$.

Proof. Gómez-Déniz et al. (2008) showed that the factorial moment of order a of mixed NB distribution can be expressed in the terms of elementary function by

$$\mu_a'(X) = E_\lambda\left(\frac{\Gamma(r+a)}{\Gamma(r)}\frac{(1-\exp(-\lambda))^a}{\exp(-\lambda a)}\right) = \frac{\Gamma(r+a)}{\Gamma(r)}E_\lambda(\exp(\lambda)-1)^a. \tag{9}$$

Using the binomial expansion of $(\exp(\lambda)-1)^a$, then $\mu'_a(X)$ can be written as

$$\mu'_a(X) = \frac{\Gamma(r+a)}{\Gamma(r)}\sum_{j=0}^{a}\binom{a}{j}(-1)^j E_\lambda(\exp(\lambda(a-j))) = \frac{\Gamma(r+a)}{\Gamma(r)}\sum_{j=0}^{a}\binom{a}{j}(-1)^j M_\lambda(a-j). \tag{10}$$

From the mgf of the NWL distribution with $t = a-j$, the $\mu'_a(X)$ is finally given as

$$\mu'_a(X) = \frac{\Gamma(r+a)}{\Gamma(r)}\sum_{j=0}^{a}\binom{a}{j}(-1)^j \frac{(\theta-\phi)(\theta-\phi+r+j+1)}{(\theta-\phi+1)(\theta-\phi+r+j)^2}. \tag{11}$$

Definition 2. Let $X \sim$ NB-WL(r,θ,ϕ). some properties of X are as follows

1) The first two moments about zero of X are

$$E(X) = r\,(\varpi-1), \tag{12}$$
$$E(X^2) = r(r+1)\varpi_2 - r(2r+1)\varpi_1 + r2, \tag{13}$$

2) The mean and variance of X respectively, are

$$E(X) = r\,(\varpi-1), \tag{14}$$
$$Var(X) = r(r+1)\varpi_2 - r\,(1+r\varpi)\varpi_1. \tag{15}$$

where $\varpi_k = \dfrac{(\theta-\phi)(\theta-\phi+k+1)}{(\theta-\phi+1)(\theta-\phi+k)^2}.$

3.3 Applications study of NB-WL distribution

We illustrated the NB-WL, NB and Poisson distributions by applying the number of hospitalized patients with diabetes at Ratchaburi hospital, Thailand. The log-likelihood values and the p-values of K-S test for the discrete goodness of fit test are summarized in Table 1. The expected frequencies of the NB-WL distribution are close to the observed frequencies, the values of K-S test of NB-WL distribution is smaller than the values of the K-S test of the Poisson and NB distributions and Also, based on the p-values of K-S test, the proposed distribution is appropriate to fit the data compared to the Poisson and NB distributions.

Table 1
Observed and expected frequencies for number of hospitalized patients with diabetes

No. of hospitalization	No. of cases	Expected value by fitting distribution		
		Poisson	NB	NB-WL
0	63	261.2574	73.5711	34.3315
1	29	449.3630	155.7058	189.0602
2	12	386.4520	205.2514	171.5407
3	15	221.5659	215.9518	147.2508
4	8	95.2733	198.4813	124.3442
5	9	32.7740	166.5815	104.7453
6	5	9.3952	130.9441	88.5186
7	4	2.3085	97.9537	75.2206
8	6	0.4963	70.4826	64.3310
9	2	0.0949	49.1530	55.3816
10	3	0.0163	33.4063	47.9862
11	3	0.0026	22.2195	41.8357
12	2	0.0004	14.5100	36.6865
13	2	0.0000	9.3271	32.3472
Total				
Parameter estimates		$\hat{\lambda} = 1.72$	$\hat{r} = 4.07$ $\hat{p} = 0.48$	$\hat{r} = 4.15$ $\hat{\alpha} = 0.52$ $\hat{\beta} = 2.01$
log-likelihood		-1140.449	-1014.642	-825.985
K-S test		0.319	0.013	0.018
p-value		<0.001	0.086	0.572

4. Conclusions

This work has proposed a new mixed negative binomial distribution called the negative binomial-new weighted Lindley distribution. In particular, some of the most important characteristics of the distribution can be studied through factorial moments, e.g., mean, variance, skewness, and kurtosis. In the application of the NB-WL distribution, we have compared the accuracy of the proposed distribution with the Poisson and NB distributions. The usefulness of the NB-WL distribution has been illustrated by the number of hospitalized patients with diabetes at Ratchaburi hospital, Thailand. We have used the log-likelihood and *p*-values of the K-S test for the goodness of fit for model selection purpose. Finally, the result of this study has shown that the NB-WL distribution provides a better fit compared with the Poisson and NB distributions. Obviously, the NB-WL distribution is an alternative distribution to the other for count data.

Acknowledgement

References

Arnold, T. B., & Emerson, J. W. (2011). Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions. *R Journal*, *3*(2), 34-39.

Ghitany, M. E., Atieh, B., & Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and Computers in Simulation*, *78*(4), 493-506.

Gómez-Déniz, E., Sarabia, J. M., & Calderín-Ojeda, E. (2008). Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance: Mathematics and Economics*, *42*(1), 39-49.

Hamid, H. (2014). *Integrated Smoothed Location Model and Data Reduction Approaches for Multi Variables Classification (Unpublished doctoral dissertation)*. Universiti Utara Malaysia, Kedah, Malaysia.

Haight, F. (1967). *Handbook of the Poisson distribution*. John Wiley and Sons, New York.

Klugman,S., Panjer, H. and Willmot, G. (2008). *Loss models: from data to decisions*. 3*rd*. John Wiley and Sons.

Kongrod, S., Bodhisuwan, W., & Payakkapong, P. (2014). The negative binomial-Erlang distribution with applications. *International Journal of Pure and Applied Mathematics*, *92*(3), 389-401.

Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, *20*(1), 102-107.

Pudprommarat, C., Bodhisuwan, W., & Zeephongsekul, P. (2012). A new mixed negative binomial distribution. *Journal of Applied Sciences(Faisalabad)*, *12*(17), 1853-1858.

Rainer,W. (2008). *Econometric analysis of count data*. Library of congress control, New York.

Team, R.C. (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Zamani, H., & Ismail, N. (2010). Negative binomial-Lindley distribution and its application. *Journal of Mathematics and Statistics*, *6*(1), 4-9.