Original

Parameters Estimation Methods for the Negative Binomial-Crack Distribution and Its Application

Pornpop Saengthong[1*], Winai Bodhisuwan[2]

Received: 29 March 2013 Accepted: 15 May 2013

Abstract

In this paper we study four parameters negative binomial-Crack (NB-CR) distribution. This new formulation distribution contains as special cases three parameters distribution, namely, negative binomial-inverse Gaussian (NB-IG), negative binomial-Birnbaum-Saunders (NB-BS) and negative binomial-length biased inverse Gaussian (NB-LBIG). The objective of our research is to estimate the parameters for NB-CR distribution by using maximum likelihood estimation and the method of moments. These methods are illustrated with an application to accident data.

Keywords: negative binomial-Crack distribution, parameter estimation, maximum likelihood estimation, method of moments count data

Introduction

With the huge growth in the collection and storage capabilities of data due to technological advances, count data have become widely available in many disciplines. 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. Also, they occur at a constant rate through time, and one occurrence of the phenomenon does not alter the probability of any future occurrence. Equality of mean and variance, called equal dispersion, is a quintessential characteristic of the Poisson distribution[1].

However, many count data often exhibit overdispersion, with a variance larger than the mean; in this case, an extension to the Poisson model is more appropriate. The negative binomial (NB) distribution is a popular alternative distribution for modeling 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 and gamma distribution. Applications using the NB distribution can be found in many areas, for instance, economics[2], accident statistics[3], biostatistics[4] and actuarial science[5]. Although, the NB distribution allows for over-dispersion, it does not take care of excess zeros in the data.[6] studied on a tool for analyzing crash data characterized by a large amount of zeros. They pointed out that traditional statistical distributions or models, such as the Poisson and the NB distributions, cannot be used efficiently in models for count data with many zeros. The Poisson distribution tends to under-estimate the number of zeros given the mean of the data, while the NB may over-estimate zeros, but under-estimate observations with a count.

The problem of overdispersion and excess zeros is usually solved by introducing mixed Poisson or mixed NB distribution. Several studies show that mixed Poisson and mixed NB distribution provided a better fit on count data compared to the Poisson and the NB distribution. These include the Poisson-inverse Gaussian[7], negative binomial-inverse Gaussian[8], negative binomial-Lindley[5] and negative binomial-Beta Exponential[9]. Therefore, in order to provide another competitive alternative to the models above, a new mixed model is considered. We

[1] *Department of Statistics, Faculty of Science, Kasetsart University, Bangkok, Thailand, 10900 E-mail: g5517400451@ku.ac.th*

[2] *Department of Statistics, Faculty of Science, Kasetsart University, Bangkok, Thailand, 10900 E-mail: fsciwnb@ku.ac.th*

[*] *Corresponding authour: E-mail: g5517400451@ku.ac.th*

propose the negative binomial-Crack (NB-CR) distribution which is a new mixed NB distribution obtained by mixing the distribution of NB(r,p) where, p = exp(-a) with a distribution of CR(λ , θ , γ). This last distribution has recently been studied[10,11]. The new mixed distribution has a heavy tail, seems to be skewed positively and may be considered as a competitive alternative for modeling overdispersed count data.

The purpose of this paper is to investigate the properties of the NB-CR distribution and its application. Parameters estimation is implemented using maximum likelihood estimation (MLE), method of moments (MoM) and present the comparison analysis between the Poisson, NB and NB-CR distributions based on a real data set using a goodness of fit test.

NB-CR distribution

This section describes the characteristics and some special cases of the NB-CR distribution.

Characteristics of the NB – CR distribution

As mentioned earlier, the NB-CR distribution is a mixture of the NB and Crack distributions. First we present the NB distribution and some of its properties. The probability mass function (pmf) of NB distribution is given by

$$f(x)=\binom{r+x-1}{x}p^r(1-p)^x,\ x=0,1,2,... \tag{1}$$

where $r>0$ and $0<p<1$. The mean and variance of this distribution can be given as

$$E(X)=\frac{r(1-p)}{p} \quad \text{and} \quad Var(X)=\frac{r(1-p)}{p^2} \quad . \tag{2}$$

The Crack distribution is a mixture of inverse Gaussian distribution and length biased inverse Gaussian which has the density function[10,11]:

$$g(x)=\frac{1}{\theta\sqrt{2\pi}}\left[\gamma\lambda\left(\frac{\theta}{x}\right)^{3/2}+(1-\gamma)\left(\frac{\theta}{x}\right)^{1/2}\right] \times\exp\left[-\frac{1}{2}\left(\sqrt{\frac{x}{\theta}}-\lambda\sqrt{\frac{\theta}{x}}\right)^2\right],\quad x>0 \tag{3}$$

where $\lambda>0$, $\theta>0$, and $0<\gamma<1$. (3)

A random variable X is assumed to follow a NB-CR (r,λ,θ,γ) distribution, when X has a NB distribution with parameter $r>0$ and $p=\exp(-a)$, where a is distributed as CR with positive parameters λ, θ and γ, i.e., $X\mid a\sim NB(r,p=\exp(-a))$ and $a\sim CR(\lambda,\theta,\gamma)$ The pmf of X is given by[12]

$$h(x)=\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\exp\left[\lambda\left(1-\sqrt{1+2\theta(r+j)}\right)\right]}{\sqrt{1+2\theta(r+j)}} \times\left[1-\gamma\left(1-\sqrt{1+2\theta(r+j)}\right)\right],\quad x=0,1,2,... \tag{4}$$

where $r>0$, $\lambda>0$, $\theta>0$, and $0<\gamma<1$.

The first moment (i.e., the mean) of the NB-CR (r,λ,θ,γ) is given by

$$E(X)=r\left(\frac{\left(1-\gamma(1-\delta)\right)\exp\left(\lambda(1-\delta)\right)}{\delta}\right)-r \tag{5}$$

The second moment of the NB-CR $(r,\lambda,\theta,\gamma)$ is given as

$$E\left(X^2\right)=\exp(\lambda)\left(\frac{\left(r^2+r\right)\left(1-\gamma(1-\varsigma)\right)\exp(-\lambda\varsigma)}{\varsigma}-\frac{\left(2r^2+r\right)\left(1-\gamma(1-\delta)\right)\exp(-\lambda\delta)}{\delta}\right)+r^2, \tag{6}$$

where $\delta=\sqrt{1-2\theta}$ and $\varsigma=\sqrt{1-4\theta}$.

The variance of the NB-CR $(r,\lambda,\theta,\gamma)$ is calculated as

$$\begin{aligned}\operatorname{var}(X)&=E\left(X^2\right)-\left(E(X)\right)^2\\&=\exp(\lambda)\left(\frac{\left(r^2+r\right)\left(1-\gamma(1-\varsigma)\right)\exp(-\lambda\varsigma)}{\varsigma}-\frac{r\left(1-\gamma(1-\delta)\right)\exp(-\lambda\delta)}{\delta}-r^2\left(\frac{\left(1-\gamma(1-\delta)\right)\exp(-\lambda\delta)}{\delta}\right)^2\exp(\lambda)\right).\end{aligned} \tag{7}$$

The special case

Here, we consider some special cases of the NB-CR distribution. Let $X\sim NB-CR(r,\lambda,\theta,\gamma)$. Then, the pmf of

X when $\gamma = 0$ [negative binomial-length biased inverse Gaussian (NB-LBIG) distribution] is given by

$$h(x) = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \times \frac{\exp\left[\lambda\left(1-\sqrt{1+2\theta(r+j)}\right)\right]}{\sqrt{1+2\theta(r+j)}}, \quad x = 0,1,2,.. \tag{8}$$

where r, λ, and $\theta > 0$.

X when $\gamma = \frac{1}{2}$ [negative binomial-Birnbaum-Saunders (NB-BS) distribution] is given by

$$h(x) = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \frac{\exp\left[\lambda\left(1-\sqrt{1+2\theta(r+j)}\right)\right]}{\sqrt{1+2\theta(r+j)}} \times \frac{1}{2}\left[1+\sqrt{1+2\theta(r+j)}\right], \quad x = 0,1,2,... \tag{9}$$

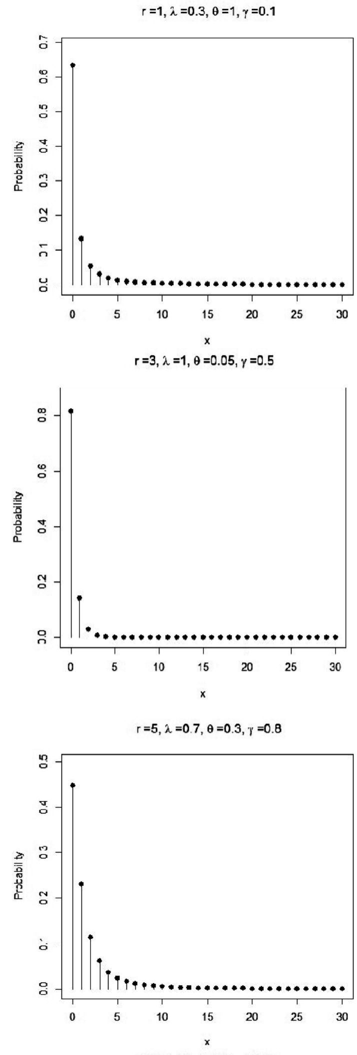
where r, λ, and $\theta > 0$.

X when $\gamma = 1$, $\lambda = \frac{\psi}{\mu}$, and $\theta = \frac{\mu^2}{\psi}$ [negative binomial-inverse Gaussian (NB-IG) distribution] is given by[8]

$$h(x) = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \times \exp\left[\frac{\psi}{\mu}\left(1-\sqrt{1+\frac{2\mu^2(r+j)}{\psi}}\right)\right], \quad x = 0,1,2,... \tag{10}$$

where r, μ, and $\psi > 0$.

In order to study the behavior of the distribution for different value of $r, \lambda, \theta, \gamma$, the pmf is calculated. We show the graphs of the pmf of the NB-CR random variable of some values of parameters in Fig. 1.



The pmf of a NB-CR random variable (X) for some specified values of $(r, \lambda, \theta, \gamma)$

Parameter Estimation

In this section, the estimation of parameters for NB-CR $(r, \lambda, \theta, \gamma)$ via the maximum likelihood estimation and the method of moments are provided. The R program [10] is used to obtain the solutions of $\hat{r}, \hat{\lambda}, \hat{\theta}$ and $\hat{\gamma}$.

Maximum Likelihood Estimation (MLE)

The log-likelihood function of the NB-CR $(r, \lambda, \theta, \gamma)$ is given by

$$\log L(r,\lambda,\theta,\gamma) = \sum_{i=1}^{n}\left(\log(r+x_i-1)! - \log x_i! - \log(r-1)!\right) + \sum_{i=1}^{n}\left(\log \sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j \frac{\exp\left[\lambda\left(1-\sqrt{1+2\theta(r+j)}\right)\right]}{\sqrt{1+2\theta(r+j)}} \times\left[1-\gamma\left(1-\sqrt{1+2\theta(r+j)}\right)\right]\right). \tag{11}$$

It can be verified that the first partial derivatives equation (11) with respect to r, λ, θ and γ, we then obtain the following differential equations;

$$\frac{\partial}{\partial r}\log L(r,\lambda,\theta,\gamma)$$
$$=\sum_{i=1}^{n}\left(\frac{\frac{\partial}{\partial r}\Gamma(r+x_i)}{\Gamma(r+x_i)}-\frac{\frac{\partial}{\partial r}\Gamma(r)}{\Gamma(r)}\right)$$
$$+\sum_{i=1}^{n}\frac{\partial}{\partial r}\log\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]$$
$$=\sum_{i=1}^{n}\psi(r+x_i)-n\psi(r)$$
$$+\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\partial}{\partial r}\left(\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]\right)}{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]}\right),$$

where $\psi(k)=\frac{\Gamma'(k)}{\Gamma(k)}$ is the digamma function.

$$\frac{\partial}{\partial\lambda}\log L(r,\lambda,\theta,\gamma)$$
$$=\sum_{i=1}^{n}\frac{\partial}{\partial\lambda}\log\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]$$
$$=\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\left[1-\gamma\left(1-Z_j\right)\right]}{Z_j}\exp\left[\lambda\left(1-Z_j\right)\right]\left(1-Z_j\right)}{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]}\right)$$

$$\frac{\partial}{\partial\theta}\log L(r,\lambda,\theta,\gamma)$$
$$=\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]$$
$$=\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\partial}{\partial\theta}\left(\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]\right)}{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]}\right),$$

$$\frac{\partial}{\partial\gamma}\log L(r,\lambda,\theta,\gamma)$$
$$=\sum_{i=1}^{n}\frac{\partial}{\partial\gamma}\log\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]$$
$$=\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^{j+1}\left(\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left(1-Z_j\right)\right)}{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\exp\left[\lambda\left(1-Z_j\right)\right]}{Z_j}\left[1-\gamma\left(1-Z_j\right)\right]}\right),$$

where $Z_j=\sqrt{1+2\theta(r+j)}$.

These four derivation equations cannot be solved analytically, as they need to rely on Newton-Raphson: a simple iterative numerical method to approximate MLE. In this paper, we obtained the MLE solutions of $\hat{r},\hat{\lambda},\hat{\theta}$ and $\hat{\gamma}$ by solving the resulting equations simultaneously using nlm function in R package, namely stats[13].

Method of moments (MoM)

For the method of moments, the parameters can be obtained by equating the sample and population moments. Because we have four parameters, we need the 3 rd-moment and 4 th-moment of (4), which are given by

$$E(X^3)=\exp(\lambda)\left(\frac{\left(r^3+3r^2+2r\right)\left(1-\gamma(1-\vartheta)\right)\exp(-\lambda\vartheta)}{\vartheta}-\frac{\left(3r^3+6r^2+3r\right)\left(1-\gamma(1-\varsigma)\right)\exp(-\lambda\varsigma)}{\varsigma}+\frac{\left(3r^3+3r^2+r\right)\left(1-\gamma(1-\delta)\right)\exp(-\lambda\delta)}{\delta}\right)-r^3, \quad (12)$$

$$E(X^4)=\exp(\lambda)\left(\frac{\left(r^4+6r^3+11r^2+6r\right)\left(1-\gamma(1-\kappa)\right)\exp(-\lambda\kappa)}{\kappa}-\frac{\left(4r^4+18r^3+26r^2+12r\right)\left(1-\gamma(1-\vartheta)\right)\exp(-\lambda\vartheta)}{\vartheta}+\frac{\left(6r^4+18r^3+19r^2+7r\right)\left(1-\gamma(1-\varsigma)\right)\exp(-\lambda\varsigma)}{\varsigma}-\frac{\left(4r^4+6r^3+4r^2+r\right)\left(1-\gamma(1-\delta)\right)\exp(-\lambda\delta)}{\delta}\right)+r^4, \quad (13)$$

where $\vartheta=\sqrt{1-6\theta}$ and $\kappa=\sqrt{1-8\theta}$.

The ith moment for the sample, m_i, is equated as

$$m_i=\frac{1}{n}\sum\nolimits_{j=1}^{n}x_j^i \quad (14)$$

Then, the method of moments estimator is derived by solving equation $m_1=E(X)$, $m_2=E(X^2)$, $m_3=E(X^3)$, and $m_4=E(X^4)$ using gmm function in R package, namely gmm[3].

Results and Conclusion

The number of injured from accidents on major roads in Bangkok of Thailand in 2007[9] was used to estimate the parameters of the Poisson, NB and NB-CR distribution based on the MLE and MoM. Table 1 shows the observed and expected frequencies, grouped in classes of expected frequency greater than five for the chi-square goodness-of-fit test as criteria of comparison, computed as $\chi^2=\sum_{i=1}^{k}\left(O_i-E_i\right)^2/E_i$. Based on the p-value, the MLE provides very poor fit for the Poisson distribution and the NB and acceptable fits for the NB-CR.

Table 1 Goodness-of-fit test from MLE for the accident data

No. of injured	No. of accident	Fitting distribution *Poisson*	*NB*	*NB-CR*
0	1273	1187.6	1278.7	1278.9
1	300	410.5	278.1	286.1
2	71	70.9	81.8	74.1
3	18	} 8.9	26.1	23.3
4	9		} 13.2	8.6
5	4			} 6.9
6+	3			
Estimated parameters		$\hat{\lambda} = 0.345$	$\hat{r} = 0.586$, $\hat{p} = 0.629$	$\hat{r} = 2.275$, $\hat{\lambda} = 1.099$, $\hat{\theta} = 0.074$, $\hat{\gamma} = 0.296$
Chi-squares		106.236	6.301	2.058
Degree of freedom		2	2	1
p-value		<0.001	0.043	0.151

The results for the goodness-of-fit test based on the MoM of the Poisson, NB and NB-CR are listed in Table II. Similar results to the previous table, the NB-CR provides a better fit compared to the Poisson and NB distribution. Furthermore, the MLE was a superior method to estimate parameters of the NB and NB-CR for the accident data. The further work, we may uses the Bayesian approach to estimate the parameter of the NB-CR distribution and conduct a simulation study in order to compare results under estimating parameters of the NB-CR distribution by applying the MLE, MoM and Bayesian approach.

Table 2 Goodness-of-fit test from MoM for the accident data

No. of injured	No. of accident	Fitting distribution *Poisson*	*NB*	*NB-CR*
0	1273	1187.6	1310.4	1270.9
1	300	410.5	239.2	283.5
2	71	70.9	78.4	77.1
3	18	} 8.9	29.5	26.1
4	9		11.8	10.4
5	4		} 8.6	} 10.0
6+	3			
Estimated parameters		$\hat{\lambda} = 0.345$	$\hat{r} = 0.386$, $\hat{p} = 0.527$	$\hat{r} = 2.116$, $\hat{\lambda} = 1.013$, $\hat{\theta} = 0.093$, $\hat{\gamma} = 0.407$
Chi-squares		106.236	22.646	5.040
Degree of freedom		2	3	1
p-value		<0.001	<0.001	0.025

Acknowledgment

The authors would like to thank Assumption University for their financial assistance as a research fellowship.

References

1. R Development Core Team, "A language and environment for statistical computing," R foundation for statistical computing, Austria, http://www.R-project.org, 2012.
2. E. Gomez-Deniz, J.M. Sarabia and E.Caldern-Ojeda, "Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications," Insurance Mathematics and Economics, 42(1), 2008, 39-49.
3. D. Lord and R.S. Geedipally, "The negative binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros," Accident Analysis & Prevention, 43(5), 2011, 1738-1742.
4. N. Alexander, R. Moyeed and J. Stander, "Spatial modelling of individual-level parasite counts using the negative binomial distribution," Biostatistics, 453-463.
5. P. Saengthong and W. Bodhisuwan, "Negative binomial-crack (NB-CR) distribution," International Journal of Pure and Applied Mathematics, 84(3), 213-230, 2013.

6. S. Kumnadee, A. Volodin and S. Lisawadi, “Crack distribution parameters estimation by the method of moments,” Proceedings of the 11th Statistics and Applied Statistics Conferences, Chiang Mai, Thailand, 2010.
7. J. Hausman, B.H. Hall and Z. Griliches, “Econometric models for count data with an application to the patents-R&D relationship,” Econometrica, 52, 909-938, 1984.
8. P. Chausse, “Computing generalized method of moments and generalized empirical likelihood with R,” Journal of Statistical Software, 34(11), 1-35, 2010.
9. M. Poch and F. Mannering, “Negative binomial analysis of intersection accident frequency,” J. Transportation Engineering, 105-113, 1996.
10. P. Bowonrattanaset and K. Budsaba, “Some properties of the three parameter Crack distribution,” Thailand Statistician, vol. 9, no. 2, 195-203, 2011.
11. S.A. Klugman, H.H. Panjer and G.E. Willmot, Loss Models: From Data to Decision, John Wiley and Sons, USA, 2008.
12. W. Rainer, Econometric Analysis of Count Data, 3rd edn, Springer-Verlag, Berlin, Germany, 2000.
13. C. Pudprommarat, W. Bodhisuwan and P. Zeephongsekul, “A new mixed negative binomial distribution,” J. Applied Sciences,12(17), 1853-1858, 2012.
14. H. Zamani and N. Ismail, “Negaive binomial-Lindley distribution and its application,” J. Mathematics and Statistics, 6(1), 2010, 4-9.