

Negative Binomial-Lindley Distribution and Its Application

Hossein Zamani and Noriszura Ismail
Actuarial Science Program, School of Mathematical Sciences,
University Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Abstract: Problem statement: The modeling of claims count is one of the most important topics in actuarial theory and practice. Many attempts were implemented in expanding the classes of mixed and compound distributions, especially in the distribution of exponential family, resulting in a better fit on count data. In some cases, it is proven that mixed distributions, in particular mixed Poisson and mixed negative binomial, provided better fit compared to other distributions. **Approach:** In this study, we introduce a new mixed negative binomial distribution by mixing the distributions of negative binomial (r, p) and Lindley (θ) , where the reparameterization of $p = \exp(-\lambda)$ is considered. **Results:** The closed form and the factorial moment of the new distribution, i.e., the negative binomial-Lindley distribution, are derived. In addition, the parameters estimation for negative binomial-Lindley via the method of moments (MME) and the Maximum Likelihood Estimation (MLE) are provided. **Conclusion:** The application of negative binomial-Lindley distribution is carried out on two samples of insurance data. Based on the results, it is shown that the negative binomial-Lindley provides a better fit compared to the Poisson and the negative binomial for count data where the probability at zero has a large value.

Key words: Mixed negative binomial, negative binomial-Lindley, count data, insurance

INTRODUCTION

The modeling of claims count is one of the most important topics in actuarial theory and practice. In some cases, it is proven that mixed distributions, in particular mixed Poisson and mixed negative binomial, provided better fit on count data compared to other distributions. Examples of mixed Poisson and mixed negative binomial distributions are the negative binomial obtained as a mixture of Poisson and gamma (Klugman *et al.*, 2008; Lemaire, 1979; Simon, 1961), the negative binomial-Pareto (Klugman *et al.*, 2008; Meng *et al.*, 1999) and the Poisson-inverse Gaussian (Klugman *et al.*, 2008; Tremblay, 1992; Willmot, 1987).

The mixing of negative binomial with inverse Gaussian which considers the reparameterization of $p = \exp(-\lambda)$ was introduced by Gómez *et al.* (2008) who considered the distribution in univariate and multivariate versions and estimated the parameters using the method of moment and the maximum likelihood estimation. The distribution of Poisson-Lindley for modeling count data was introduced by Sankaran (1970). Ghitany *et al.* (2008a) showed that in

many ways, Lindley is a better distribution compared to exponential and hence, one should expect that the Poisson-Lindley provides a better fit compared to the Poisson-exponential. The distribution of zero-truncated Poisson-Lindley was introduced by Ghitany *et al.* (2008a; 2008b), who used the distribution for modeling count data in the case where the distribution has to be adjusted for the count of missing zeros.

In this study, we introduce a new mixed negative binomial distribution by mixing the distributions of negative binomial (r, p) and Lindley (θ) , where the reparameterization of $p = \exp(-\lambda)$ is considered. This new mixed distribution has a thick tail and may be considered as an alternative for modeling count data of insurance claims which has a thick tail and a large value at zero.

MATERIALS AND METHODS

Closed form and factorial moment: In this part, the closed form and the factorial moment of the negative binomial-Lindley distribution are given.

A classical negative binomial distribution is denoted with probability mass function (pmf):

Corresponding Author: Noriszura Ismail, Actuarial Science Program, School of Mathematical Sciences,
University Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
Tel: +603 89215754 Fax: +603 89254519

$$\Pr(x) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \dots \quad (1)$$

where, $r > p$ and $0 < p < 1$. The first two moments about zero and the factorial moment of order k of a negative binomial distribution are given respectively by Gómez *et al.* (2008):

$$E(x) = \frac{r(1-p)}{p}$$

$$E(X^2) = \frac{r(1-p)[1+r(1-p)]}{p^2}$$

and

$$\mu_{[k]}(X) = E[X(X-1)\dots(X-k+1)]$$

$$= \frac{\Gamma(r+k)(1-p)^k}{\Gamma(r)p^k}, \quad k = 1, 2, \dots \quad (2)$$

The Lindley distribution function which is specified by the probability density function (pdf):

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

was introduced by Lindley (1958). It can easily be shown that the Lindley distribution is a mixture of exponential and gamma with $p = \frac{1}{1+\theta}$, where $f_1(x) = \text{Gamma}(2, \theta)$, $f_2 = \exp(\theta)$ and $f(x) = pf_1(x) + (1-p)f_2(x)$. The moment generating function of Lindley distribution is given by:

$$M_x(z) = \frac{\theta^2}{\theta+1} \frac{\theta-z+1}{(\theta-z)^2} \quad (4)$$

Definition: A random variable X has a negative binomial-Lindley (r, θ) distribution if it satisfies the stochastic representation:

$$X | \lambda \sim \text{NB}(r, p = e^{-\lambda}) \quad (5)$$

and

$$\lambda \sim \text{Lin}(\theta) \quad (6)$$

where, $r > 0$ and $\theta > 0$.

Throughout this study, we will use the notation NB-L(r, θ) as a reference for the negative binomial-Lindley distribution. Figure 1-4 show that the pmf of

NB-L(r, θ) has the highest mass at zero for several values of r and θ . Hence, the NB-L(r, θ) may have a good fit for claims count data in the case where the probability at zero has a large value.

Theorem: Let $X \sim \text{NB-L}(r, \theta)$ be a negative binomial-Lindley distribution defined in (5) and (6). The probability mass function (pmf) and the factorial moment of order k of the distribution are given as (7) and (8) respectively:

$$\Pr(x) = \frac{\theta^2}{\theta+1} \binom{r+x-1}{x} \sum_{j=0}^k \binom{k}{j} (-1)^j \frac{\theta+r+j+1}{(\theta+r+j)^2}, \quad x = 0, 1, 2, \dots \quad (7)$$

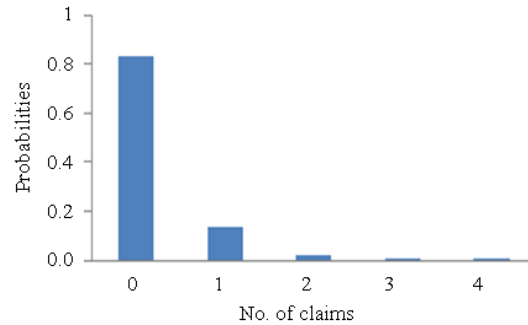


Fig. 1: Plot of p.m.f for NB-L ($r = 20, \theta = 100$)

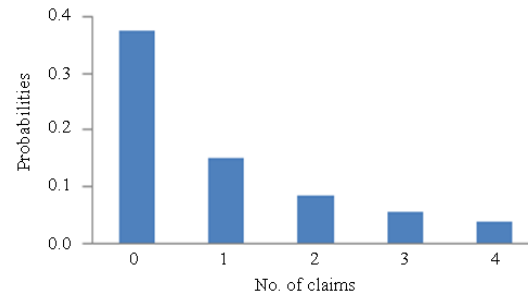


Fig. 2: Plot of p.m.f for NB-L ($r = 5, \theta = 20$)

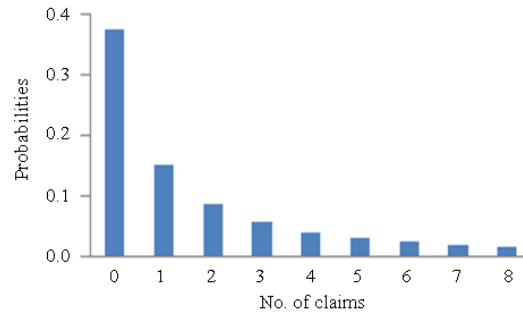


Fig. 3: Plot of p.m.f for NB-L ($r = 100, \theta = 100$)

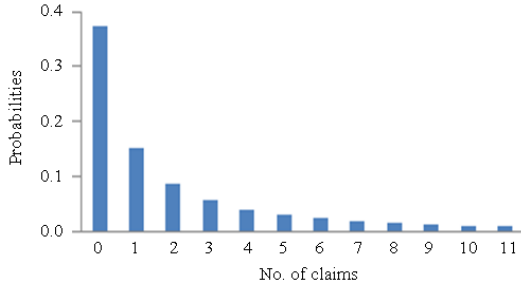


Fig. 4: Plot of p.m.f for NB-L (r = 1, θ = 1)

and

$$\mu_{[k]}(X) = \frac{\Gamma(r+k)}{\Gamma(r)} \frac{\theta^2}{\theta+1} \sum_{j=0}^k \binom{k}{j} (-1)^j \frac{\theta^{r+j+1}}{(\theta+r+j)^2}, \quad k = 0, 1, \dots \quad (8)$$

Proof: If $X|\lambda \sim \text{NB}(r, p = e^{-\lambda})$ and $\lambda \sim \text{Lin}(\theta)$, then the pmf of X can be obtained by:

$$\Pr(x) = \int_0^\infty \Pr(X = x | \lambda) f(\lambda; \theta) d\lambda \quad (9)$$

We know that:

$$\begin{aligned} \Pr(X = x | \lambda) &= \binom{r+x-1}{x} e^{-\lambda r} (1 - e^{-\lambda})^x \\ &= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j e^{-\lambda(r+j)} \end{aligned} \quad (10)$$

By inserting (10) in (9) we have:

$$\begin{aligned} \Pr(X = x | \lambda) &= \binom{r+x-1}{x} \sum_{j=0}^x (-1)^j \int_0^\infty e^{-\lambda(r+j)} f(\lambda; \theta) d\lambda \\ &= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_\lambda(-r+j) \end{aligned} \quad (11)$$

Then, (7) is obtained by using (4) in (11).

Gómez *et al.* (2008) showed that the factorial moment of order k of a mixed negative binomial distribution where $p = \exp(-\lambda)$ can be obtained by using:

$$\begin{aligned} \mu_{[k]}(X) &= \frac{\Gamma(r+k)}{\Gamma(r)} E_\lambda (e^\lambda - 1)^k \\ &= \frac{\Gamma(r+k)}{\Gamma(r)} \sum_{j=0}^k \binom{k}{j} (-1)^j E(e^{\lambda(k-j)}) \\ &= \frac{\Gamma(r+k)}{\Gamma(r)} \sum_{j=0}^k \binom{k}{j} (-1)^j M_\lambda(k-j) \end{aligned} \quad (12)$$

Therefore, (8) is obtained by inserting (4) in (12). From (12), the first two moments of the NB-L(r,θ) are given by:

$$E(X) = r \left[\frac{\theta^3}{(\theta+1)(\theta-1)^2} - 1 \right]$$

and

$$E(X^2) = (r+r^2) \frac{\theta^2(\theta-1)}{(\theta+1)(\theta-2)^2} - (r+2r^2) \frac{\theta^3}{(\theta+1)(\theta-1)^2} + r^2 \quad (13)$$

Parameter estimation: In this segment, the estimation of parameters for NB-L(r,θ) via the method of moments and the maximum likelihood procedure are provided.

Method of moments (MME): For the method of moments, the parameters may be obtained by equating the sample and the theoretical moments. In the case of NB-L(r,θ), the first two moments are required for estimating r and θ.

The *i*th moment for the sample, *m_i*, is equated as:

$$m_i = \frac{1}{n} \sum_{x=0}^k x^i \quad (14)$$

Then, from (13) and (14) we have:

$$m_1 = r \left[\frac{\theta^3}{(\theta+1)(\theta-1)^2} - 1 \right] \quad (15)$$

and

$$m_2 = (r+r^2) \frac{\theta^2(\theta-1)}{(\theta+1)(\theta-2)^2} - (r+2r^2) \frac{\theta^3}{(\theta+1)(\theta-1)^2} + r^2 \quad (16)$$

The solution of r and θ may be obtained by solving both (15) and (16). It should be noted that $r > 0$ and $\theta > 0$.

Maximum Likelihood Estimation (MLE): The log likelihood function of the NB-L(r,θ) is given by:

$$\begin{aligned} \log L(r, \theta) &= \mathcal{L}(r, \theta) = \sum_{x=0}^k n_x \log p_x \\ &= \sum_{x=0}^k n_x \log \left[\binom{r+x-1}{x} \frac{\theta^2}{\theta+1} \sum_{j=0}^x \binom{x}{j} (-1)^j \frac{\theta^{r+j+1}}{(\theta+r+j)^2} \right] \end{aligned} \quad (17)$$

By taking partial derivatives of the log likelihood function each with respect to r and θ and by equating both partial derivatives to zero, we obtain the equations:

$$\frac{\partial}{\partial \theta} \mathcal{L}(r, \theta) = n \left(\frac{2}{\theta} - \frac{1}{\theta + 1} \right) + \sum_{x=0}^k n_x \left\{ \frac{\sum_{j=0}^x (-1)^{j+1} \binom{x}{j} \frac{\theta+r+j+2}{(\theta+r+j)^2}}{\sum_{j=0}^x (-1)^j \binom{x}{j} \frac{\theta+r+j+1}{(\theta+r+j)^2}} \right\} = 0 \tag{18}$$

and

$$\frac{\partial}{\partial r} \mathcal{L}(r, \theta) = \frac{\partial}{\partial r} \sum_{x=0}^k n_x \log \left(\frac{x+r-1}{x} \right) + \sum_{x=0}^k n_x \left\{ \frac{\sum_{j=0}^x (-1)^{j+1} \binom{x}{j} \frac{\theta+r+j+2}{(\theta+r+j)^2}}{\sum_{j=0}^x (-1)^j \binom{x}{j} \frac{\theta+r+j+1}{(\theta+r+j)^2}} \right\} = 0 \tag{19}$$

where, $n = \sum_{x=0}^k n_x$.

Klugman *et al.* (2008) showed that:

$$\frac{\partial}{\partial r} \sum_{x=0}^k n_x \log \left(\frac{r+x-1}{x} \right) = \sum_{x=0}^k n_x \sum_{m=0}^{x-1} \ln(r+m) \tag{20}$$

Therefore, by using (18-20) we have:

$$n \left(\frac{2}{\theta} - \frac{1}{\theta + 1} \right) = \sum_{x=0}^k n_x \sum_{m=0}^{x-1} \ln(r+m) \tag{21}$$

Let us assume:

$$B = \sum_{x=0}^k n_x \sum_{m=0}^{x-1} \ln(r+m)$$

Equation 21 has two distinct real roots for θ . However, since $\theta > 0$, the only acceptable root is given by:

$$\theta(r) = \frac{n - B + \sqrt{(n + B)^2 + 4nB}}{2B} \tag{22}$$

Replacing (22) in (19) yields the following:

$$H(\hat{r}) = \sum_{x=0}^k n_x \sum_{m=0}^{x-1} \frac{1}{\hat{r} + m} + \sum_{x=0}^k n_x \left\{ \frac{\sum_{j=0}^x (-1)^{j+1} \binom{x}{j} \frac{\theta(\hat{r})+\hat{r}+2}{(\theta(\hat{r})+\hat{r}+j)^2}}{\sum_{j=0}^x (-1)^j \binom{x}{j} \frac{\theta(\hat{r})+\hat{r}+1}{(\theta(\hat{r})+\hat{r}+j)^2}} \right\} \tag{23}$$

The solution for \hat{r} in Eq. 23 may be solved numerically by using Newton-Raphson method. The required equation for the k th iteration is:

$$rk = r_{k-1} - \frac{H(r_{k-1})}{H'(r_{k-1})}$$

The solution for $\hat{\theta}$ can be obtained by inserting \hat{r} in (21).

RESULTS AND DISCUSSION

The NB-L(r, θ) distribution is applied on two samples of insurance count data and the results are presented in this results.

Example 1: The data for this example was taken from Klugman *et al.* (2008), whereby it was collected by Dropkin in 1956-1958 and analyzed in a paper in 1959. The distributions of NB-L, Poisson and negative binomial are fitted to the data using R programming and the results are provided in Table 1. Based on the log likelihood and p-value, the NB-L distribution provides the best fit for the data. The p-value of chi-square statistics for the NB-L is 99.9%.

Table 1: Estimated parameters, p-value and log likelihood for Example 1

Number of accidents	Number of claims	Fitting distribution			
		Poisson	Negative binomial	NB-L	
0	81,714	80,655.90	81,692.60	81,718.60	
1	11,306	13,146.90	11,294.80	11,294.70	
2	1,618	1,071.50	1,656.20	1,630.60	
3	250 } 40 } = 297 7 }	58.20 } 2.40 } = 60.70 0.10 }	247.50 } 37.30 } = 291.40 6.60 }	245.30 } 38.40 } = 291.10 7.40 }	
4					3+
5+					
Parameters		$\lambda = 0.163$	$r = 0.892$	$R = 19.20$	
Chi squares		1140.74	$\theta = 0.845$	$\theta = 119.70$	
Degrees of freedom	2	3	1.14	0.29	
p-value		<1%	97.9%	99.9%	
Log likelihood	-45,298.00	-44,764.60	-44,764.30		

Table 2: Estimated parameters, p-value and log likelihood for Example 2

Number of claims	Number of drivers	Fitting distribution		
		Poisson	Negative binomial	NB-L
0	7,840	7,638.3	7,843.3	7853.6
1	1,317	1,634.6	1,290.2	1287.4
2	239	174.9	257.7	247.6
3	42	12.5	54.5	54.2
4	14	0.7	11.8	13.2
5	4	0	2.6	3.5
6	4	0	0.6	1.0
7	1	0	0.2	0.3
8+	0	0	0.1	0.2
Parameters		$\lambda = 0.214$	$r = 0.70$ $\theta = 0.765$	$r = 4.63$ $\theta = 23.55$
Chi squares	293.8	8.65	6.99	
Degrees of freedom	2	2	3	
p-value		<1%	7%	32%
Log likelihood	-5490.78	-5348.00	-5344.70	

Example 2: The data which was obtained from Klugman *et al.* (2008), provides information on 9,461 automobile insurance policies whereby the number of accidents of each policy is recorded. The distributions of Poisson, negative binomial and NB-L are fitted to the data using R programming. Based on the log likelihood and p-value, the negative binomial provides a better fit compared to the Poisson. However, the p-value of chi-square statistics for the negative binomial is still considered as small. The NB-L distribution however, provides quite a significant improvement over the Poisson and the negative binomial by providing a much better fit for the data (Table 2).

CONCLUSION

In this study, a two-parameter negative binomial-Lindley distribution, NB-L(r, θ), is introduced by mixing the distributions of negative binomial (r, p) and Lindley (θ), where the reparameterization of $p = \exp(-\lambda)$ is considered. In particular, the closed form and the factorial moment of NB-L(r, θ) are derived. In addition, the parameters estimation of NB-L(r, θ) via the method of moments (MME) and the Maximum Likelihood Estimation (MLE) are shown. The application of NB-L(r, θ) is carried out on two samples of insurance data. Based on the results, it is shown that NB-L(r, θ) provides a better fit compared to the Poisson and the negative binomial for count data where the probability at zero has a large value. In particular, Fig. 1-4 show that the pmf of NB-L(r, θ) has the highest mass at zero for several values of r and θ and the results of log-likelihood and p-value provided

in Table 1 and 2 proved that the NB-L(r, θ) provides a better fit for the sample data compared to the distributions of Poisson and Negative Binomial.

REFERENCES

Ghitany, M.E., D.K. Al-Mutairi and S. Nadarajah, 2008a. Zero-truncate Poisson-Lindley distribution and its application. *Math. Comput. Simulat.*, 79: 279-287. <http://portal.acm.org/citation.cfm?id=1461225>

Ghitany, M.E., B. Atieh and S. Nadarajah, 2008b. Lindley distribution and its application. *Math. Comput. Simulat.*, 78: 493-506. <http://portal.acm.org/citation.cfm?id=1377430>

Gómez, E., M. Sarabia and E. Calderin, 2008. Univariate and multivariate versions of the negative binomial-inverse Gaussian distribution with applications. *Insur. Math. Stat.*, 42: 39-49. DOI: 10.1016/j.insmatheco.2006.12.001

Klugman, S.A., H.H. Panjer and G.E. Willmot, 2008. *Loss Models: From Data to Decision*. 3rd Edn., John Wiley and Sons, USA, pp: 101-159.

Lemaire, J., 1979. How to define a bonus-malus system with an exponential utility function. *ASTIN Bull.*, 10: 274-282. <http://www.actuaries.org/LIBRARY/ASTIN/vol10no3/274.pdf>

Lindley, D.V., 1958. Fiducial distributions and Bayes' theorem. *J. R. Stat. Soc.*, 20: 102-107. <http://www.jstor.org/stable/2983909>

Meng, S., Y. Wei and G.A. Whitmore, 1999. Accounting for individual overdispersion in a bonus-malus system. *ASTIN Bull.*, 29: 327-337. <http://www.actuaries.org/LIBRARY/ASTIN/vol29no2/327.pdf>

- Sankaran, M., 1970. The discrete Poisson-Lindley distribution. *Biometrics*, 26: 145-149.
- Simon, L.J., 1961. Fitting negative binomial distributions by the method of maximum likelihood. *Proc. Casual. Actuar. Soc.*, 48: 45-53. <http://www.casact.net/pubs/proceed/proceed61/61045.pdf>
- Tremblay, L., 1992. Using the Poisson inverse Gaussian in bonus-malus systems. *ASTIN Bull.*, 22: 97-106. <http://www.actuaries.org/LIBRARY/ASTIN/vol22no1/97.pdf>
- Willmot, G.E., 1987. The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuar. J.*, 2:113-127.