# Editorial Note on Research & Reviews: Journal of Statistics and Mathmetical Science Editoral

Received date: 10/05/2018 Accepted date: 02/07/2018 Published date: 06/07/2018

\*For Correspondence

Department of Statistics, Faculty of SocialScience, University of Botswana, Gaborone, Botswana.

E-mail: tlhaloganyangs@gmail.com

#### ABSTRACT

The objective of this paper is to provide an alternative distribution to the varieties of discrete distributions to be used to fit count data. We propose a compound of Generalized Negative Binomial and Shanker distribution, namely, the Generalized Negative Binomial-Shanker (GNB-SH) distribution. GNB-SH distribution can be used to fit count data while still maintaining similar characteristics as the traditional negative binomial. This new formulation is a generalization of new mixture distributions namely the Negative Binomial-Shanker (NB-SH) distribution and the Binomial-Shanker (BI-SH) distribution. Some mathematical properties of this distribution and that of its special cases are studied. Parameter estimation for GNB-SH and NB-SH distributions is also implemented using maximum likelihood.

## INTRODUCTION

Most often, existing discrete distributions sometimes fail to fit count data well due to various reasons such variations within the data, the shape of the distribution and assumptions related to these distributions <sup>[1]</sup>. As a result, experiencing poor fit of existing discrete models in analysis of count data is a major concern in fields such as medicine, transport, engineering and agriculture. Therefore, researchers are striving to come up with new discrete distributions which could provide a better fit to the observed count data when compared to other existing models. For that reason, we propose the new distribution namely the Generalized Negative Binomial-Shanker (GNB-SH) distribution which is obtained by compounding the distribution of GNB

(m,p, $\beta$ ), where  $p = \exp(-\lambda)$  with distribution of SH( $\theta$ ). The expectation is that GNB-SH distribution should provide a better fit to observed count

data when compared to other competing distributions such as the traditional NB distribution.

Various researchers used the concept of mixing distributions to explore new flexible distributions that performs better than standard well known models. In many cases, mixed Poisson and Negative Binomial (NB) distributions usually provide better fit when compared to other existing distributions. When the data is over-dispersed, stands out to be the best the when compared to Poisson due to its assumptions flexibility. Based on this strategy of mixing distributions, various researchers were able to explore new flexible distributions.

Looking at previous researches, <sup>[2]</sup> mixed NB and Lindely distribution which have been extended and applied in many count data analysis <sup>[3-6]</sup>. Saengthong et al. obtained a mixture of NB and Crack distribution which contains three special cases namely Negative Binomial-Inverse Gaussian (NB-IG), Negative Binomial-Birnbaum-Saunders (NB-BS) and Negative Binomial-Length Biased Inverse Gaussian (NB-LBIG) <sup>[7]</sup>. These results were extended to make obtain a distribution suitable for zero inflated count data <sup>[8]</sup>.

Gerstenkorn compounded the Generalized Negative Binomial (GNB) distribution and the Generalized Beta (GB) distribution which was later modified by Rashid et al. to study zero truncated count data <sup>[9,10]</sup>. Rashid et al. studied a mixture of Generalized Negative Binomial with Generalized Exponential (GNB-GE) distribution which entailed a mixture of Negative Binomial with Generalized Exponential (NB-GE) <sup>[11,12]</sup>. For its appealing performance, a zero inflation parameter was added to NB-GE distribution to make it more suitable for count data with excess number of zeros <sup>[13]</sup>.

In mixtures related to Poisson distribution <sup>[14]</sup> introduced a mixture of one parameter Lindely distribution <sup>[15]</sup> with Poisson distribution. Some extensions and modification related to this formulation can be found <sup>[16-20]</sup>. Other Poisson mixtures include the Poisson-Shanker mixture <sup>[1]</sup>, the Poisson-Amarenda mixture <sup>[21]</sup> and the Poisson-Sujatha mixture <sup>[22-25]</sup>.

In this work, we present the concept of compounding distributions and the distributions involved in formulating GNB-SH distribution in Section 2. This section ends with mixing GNB distribution and Shanker distribution and provides its special cases. Section 3 entails mathematical properties related to this distribution including that of special cases. Section 4 deals with parameter estimation of NB-SH and GNB-SH using maximum likelihood. Section 5 presents the conclusion of this paper that includes our future plans.

#### MATERIALS AND METHODS

Generalized Negative-Binomial Distribution

A discrete random variable X is said to be a Generalized Negative-Binomial (GNB) distribution if its pmf is given as:

$$f x m p(; , \overline{\beta}) = m \qquad \Box \Box m + \beta x \Box \Box p^{x} (1-p)_{m+-\beta x x}$$
(1)  

$$m + \beta x \Box x \Box$$
for x = 0,1, 2, 3,....., and zero otherwise, where  

$$0 \le p < 1, m > 0, p\beta \beta < 1, \ge 1$$
(a)

When  $\beta = 0 \& m \in \mathbb{N}$ , the pmf of equation (1) reduces to Binomial distribution and when  $\beta = 1$ , equation (1) reduces to the pmf of NB distribution which its mean and factorial moment respectively are given as:

$$m(1-p) \qquad \Gamma(m \ r+)(1-p)^{r}$$

$$E X() = ----- and \mu_{r}(x) = ------r$$

$$(2) p \qquad \Gamma(m \ p)$$

where  $\Gamma$ (.) is the Gamma function, see <sup>[23-25]</sup>. In GNB distribution, the parameters m, pand  $\beta$  are constants but here it is assumed that where  $\lambda$  is a random variable following the Shanker distribution.

#### Shanker distribution

 $0 \leq \leq \theta \ 1, m N \in \beta = 0$ 

As an extension of Lindely distribution <sup>[15]</sup> Shanker (SH) distribution was proposed by Shanker <sup>[26]</sup> who also provided its mathematical properties. This distribution is a mixture of Exponential distribution with scale parameter  $\theta$  and a Gamma distribution with shape parameter 2 and scale parameter  $\theta$ . This distribution has shown a better fit when it was compared with Exponential and Lindely distribution in modelling of lifetime data. The density function of this distribution is given as:

$$\theta^{2}$$

$$g(\lambda\theta;) = \underline{\qquad}_{2}(\theta \ \lambda + )e$$

$$\theta + 1$$
(3)

for  $\lambda > 0$  and zero otherwise, where  $\theta > 0$ . Its moment generating function (mgf) is given as:

$$(4) \stackrel{-2}{=}$$

$$M_{\lambda}(t) = \Box_{1-2} \underbrace{\Box \Box_{1-}}_{\Box \Box_{1-}} t = \int_{-\infty} \underbrace{d_{1-1}}_{\Box \Box_{1-}} d_{1-1} d$$

#### Compound Distribution

+100000

According to the definitions provided by Gurland <sup>[27]</sup>, Compounding of distributions occurs when all or some parameters of a certain distribution (Parent probability distribution) is treated as a random variable of another probability distribution called Compounding distribution. In compounding, the support of the Parent distribution determines the support of the compound distribution <sup>[27,28]</sup>. If the parent distribution is discrete (continuous), then the compound distribution will become discrete (continuous).

(b)

Compounding played an important role in revival of NB distribution which is a compound of Poisson distribution with its parameter λ treated as a Gamma variable. Considering the case of one discrete variable, the definitions and relations provided by Gurland <sup>[27]</sup> to compound a distribution are as follows:

Let X be a discrete random variable with  $f_X(\lambda)$  where parameter  $\lambda$  is a random variable with probability density function  $g(\lambda)$ , then a compound distribution h(x) is defined as:

Compounding of Generalized Negative-Binomial distribution with the Shanker Distribution

Definition: Let X be a random variable of a GNB-SH(m,  $\beta$ ,  $\theta$ ) distribution denoted by X GNB SH  $-(m,\beta\theta,)$  when X has a GNB distribution with parameters m,  $\beta$  and  $p = e^{-\lambda}$  where  $\lambda$  is distributed as SH with parameter  $\theta > 0$ , i.e  $X \lambda \sim GNB m(\beta, p e^{-\lambda})$ and  $\lambda \Box SH(\theta)$ .

Theorem: Let  $X \square GNB - SH m(, \beta \theta, )$ , then the pmf of X is given as:

 $-2m \Box m + \beta x \Box^{m+} -\beta x x \Box m + \beta x$ 

$$x = \frac{x - \Box}{k \Box \Theta(x \ k + )\Box} (x \ k + )\Box (x \ k + )\Box (\theta) = \frac{x - \Box}{k \Box \Box} (\theta) = \frac{x - \Box}{k \Box \Box} x = \frac{x - \Box}{k \Box} = \frac{x - \Box}{k \Box}$$

for x = 0, 1, 2, 3, ..., and zero otherwise, where

$$0 \le p < 1, m > 0, \ p\beta\beta\theta < 1, \ge 1, > 0$$
 (a)

$$0 \le \le \theta \ 1, m \ N \in \beta \ \theta = 0, >0 \tag{b}$$

Proof: If *X* GNB SH  $m \sim -(\beta, \theta)$  defined in equation (1) and  $\lambda \square SH(\theta)$  defined in equation (3), then using equation (5), the pmf of X can be obtained by:

$$h x(\bigcup_{0}^{\infty} f x(\lambda \partial \lambda)g(z))d$$
(7)

Substituting  $p e = -\lambda$  in equation (1) we have  $fX(|\lambda)$  being defined as

$$fX(\lambda) = \overline{m} \Box \Box m + \beta x \Box \Box (e_{-\lambda})_x (1 - e_{-\lambda})_{m + -\beta x x}$$

 $m + \beta x \Box x$ 

using binomial expansion we obtain

 $\infty$ 

$$\begin{matrix} & & & \\ &$$

By substituting equation (3) and equation (8) into equation (7) we obtain

$$m \Box m + \beta x \Box^{m + -\beta_{x}x} \Box m + \beta x x - \Box \qquad k^{\infty} -\lambda(xk+) \Theta^{2} \qquad -\Theta\lambda$$

$$(9)$$

$$h x() = \underline{\qquad} \Box \Sigma \Box k \Box \Box (-1) \int 0 e \Theta \underline{\qquad} 2 + 1 (\Theta \lambda + ) e d\lambda m + \beta x \Box x \Box_{k=0} \Box$$

$$m \Box m + \beta x \Box^{m + -\beta_{x}x} \Box m + \beta x x - \Box \qquad k$$

$$=h x() = \underline{\qquad} m + \beta x \Box \Box \qquad x \quad \Box \Sigma_{k=0} \Box^{\Box} \qquad k \quad \Box^{\Box}(-1) M_{\lambda} \Box \Box - (xk+) \Box \Box$$

Substituting the moment generating function of SH distribution in equation (4) into equation (9) the pmf of *GNB SH*  $m-(\beta,\theta)$  the distribution is finally given as

-2

 $m + \beta x \Box$  x  $\Box_{k=0} \Box$ 

Next we provide the special cases of GNB-SH and their probability mass functions. Note that these special cases can simply be proven by substituting in the assumed values provided for each case.

Corollary: If  $\beta = 1$ , then the GNB-SH pmf in equation (6) reduces to a mixture of NB and Shanker distribution denoted as *X NB SH m*  $-(, \theta)$  with pmf

$$h x m(; \qquad , \theta) = \Box \Box m x + -1 \Box \Box \sum_{m} \Box \Box \Box \Box km (-1)_{k} \Box \Box \Box 1 + \theta \theta(x \\ k++1) \Box \Box \Box \Box \Box (x k\theta + ) \Box \Box (10) 1 + 2 \qquad \Box \qquad (10) 1 + 2 \qquad \Box$$

-2

Corollary: If  $\beta = 0$  and  $m N \in$ , then the GNB-SH pmf in equation (6) reduces to a mixture of Binomial and Shanker distribution denoted as  $X \square BI - SHm(, \theta)$  with pmf

$$h x m(; , \theta) = \Box \Box \Box \Box \Box m x^{m x} \sum_{k=0}^{\infty} \Box \Box \Box m x^{k-1} \Box (-1)_{k} \Box \Box \Box 1 + \theta \theta(x k_{2} + + 1)$$

## THE PROPERTIES OF THE GNB-SH DISTRIBUTION

This deals with provision of Factorial moments of the distributions. The ordinary (crude) moments of the GNB-SH distributions can be obtained by using the formula

$$m_l = \sum_{k=0}^{\infty} S m_{kl}[k]$$

l

where  $S_{k^{l}}$  stands for the Stirling numbers of the second kind <sup>[29]</sup>. Therefore, only the factorial moments of the mixture of NB

and Shanker distribution will be considered.

$$m_l = \sum_{k=0}^{\infty} S \ m_{kl}[k]$$

where  $S_{k^{l}}$  stands for the Stirling numbers of the second kind [29]. Therefore, only the factorial moments of the mixture of NB and Shanker distribution will be considered.

Definition: If  $^{X} \square NB - SH m(, \theta)$ , then the factorial moment polynomial

$$x_r = x x (-1)(x-2), \dots, (x-r+1)$$

$$m_r(x) = E\lambda \mu \lambda \Box \Box r(x) \Box \Box = E X(r)$$

is called the factorial moment of order r of a mixture of NB with Shanker distribution (NB-SH), where  $\mu \lambda_r(x)$  is the factorial moment of NB distribution.

Theorem: The factorial moment of order r of NB-SH distribution is given by

$$\Gamma(m+r) = \frac{\Gamma(m+r)}{\sum} \Box \Box (-1) \Box 1 - \frac{1}{2} \Box \Box 1 - \frac{1}{2} \Box 1 - \frac{1}{2} = \frac{r-k}{(12)} \frac{r-k}{\Gamma(m)} = \frac{r-k}{k=0} \Box k \Box \theta + 1 \Box \Box \theta \Box \theta$$
  
for r = 1,2,3,..., where m,  $\theta > 0$ .

Proof: From the factorial moment of NB distribution in equation (2), if we let  $p = -\lambda$  then the factorial moment of order r of NB-SH distribution is given as

$$m x_r () = E_{\lambda} \Box \Box \Gamma(m r+) (1 - \lambda_r e_{-\lambda})_r \Box \Box \Box \Gamma(m e) \Box$$

$$\Box \qquad \Box$$

$$\Gamma(m r+)$$

$$= - \Gamma(m) E_{\lambda} \Box^{\Box} \Box (e_{\lambda} - 1)_r$$

□□□using binomial expansion we

$$= \binom{m}{r} \sum_{r} \binom{r}{r} (-1)^{k} \frac{1}{2}$$

obtain

$$= \underline{\sum_{k=0} \Box_k (-1) M r k_{\lambda} (-)}$$

 $\Gamma(m)$ 

Substituting the moment generating function of SH distribution with t = r- k we get

$$\Gamma + (m r) \quad r \square \square r_k \square \Theta(r k -) \square \square r k - \square^{-2} m x_r () =$$

$$\sum_{k=0}^{\infty} \square k (-1) \square \square \square - \Theta_2 + 1 \square \square \square \square \square - \Theta \square \square$$

$$\Gamma(m) \square \square$$

From the factorial moments of NB-SH distribution in equation (12), for convenience we let

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

Then, the first four moments about zero are respectively given as

$$E x\left(\frac{m\theta_{2}}{\omega}\right) = 2$$

$$(\omega \delta -) \theta + 1$$

$$2 mm(+1)\theta$$

$$E x\left(\frac{mm(+1)\theta}{\omega}\right) = 2(\rho \omega)$$

$$\delta - 2 + \theta + 1$$

$$\theta^{2}\Gamma + (m3)$$

$$E x\left(\frac{\theta^{2}\Gamma + (m3)}{\omega}\right) = (\tau - t - 3\rho)$$

$$\omega \delta 3 + (\theta + \Gamma - 1) (m) = 2$$

$$4\theta^{2}\Gamma + (m4) = 0 = 0\theta(-t + 41)$$

$$E x\left(\theta + \tau - 1\right) = 0$$

$$(\theta + \tau - 1) (m) = 0$$

# **PARAMETER ESTIMATION**

Definition: In this section, maximum likelihood is used to provide parameter estimates of  $GNB SH m-(\beta, \theta)$  distribution and that of its special case of NB-SH distribution (m, $\theta$ ).

## Estimation of GNB-SH distribution parameters using full likelihood function

Let X1,X2,X3,...,Xn a random sample of size n from the GNB-SH distribution with observed values x1, x2, x3,..., xn. We find the values of m,  $\beta$  and  $\theta$  that maximizes the likelihood function (joint pmf of the sample) of GNB-SH. Parameter estimates can easily be obtained by maximizing the logarithm of the likelihood function with respect to m,  $\beta$  and  $\theta$  as the product is replaced by the sums. Consider the likelihood function of GNB-SH distribution defined by

$$L x m(;,\beta\theta, \qquad b = 1 \ b = 0 \ b = 0 \ b = 0 \ b = 0 \ b = 0 \ b = 0 \ b = 0 \ b = 0 \ c = 0$$

with corresponding log-likelihood function given as:

$$n = m = n = \Gamma(m + \beta x_i + 1) =$$

Maximum likelihood estimators of m,  $\beta$  and  $\theta$  can be obtained by maximizing Log L (x; m,  $\beta$ ,  $\theta$ ) with respect to m,  $\beta$  and  $\theta$  respectively. That is

$$\frac{\partial L_{0gL,x,m}\left(\begin{array}{c} \beta & \theta \\ \end{array}\right)}{\sum_{n=1}^{n} \left[ \begin{array}{c} -\frac{1}{2} & -\frac{1}{2} \\ \end{array}\right]} \sum_{n=1}^{n} \sum_{i=1}^{n} \sum_{m=1}^{n} \sum_{m=1}^{n} \sum_{i=1}^{n} \sum_{m=1}^{n} \sum_{i=1}^{n} \sum_{m=1}^{n} \sum_{i=1}^{n} \sum_{m=1}^{n} \sum_{i=1}^{n} \sum_{m=1}^{n} \sum_{m=1$$

Consider the log-likelihood function of NB-SH distribution defined by

<sup>n</sup> 
$$\Box m x + -i1\Box$$
 <sup>n</sup>  $\Box m k \Box \Box m \Box \Theta(x k_i + )\Box \Box (x k_i + )\Box^{-2}$ 

$$LogL x m(:;\beta\theta,) = \sum_{i=1}^{n} \log \left[ x_{i} \Box_{i+1} \log \Box_{\Box} k^{\pm 0} (-1) \Box_{\Box} \Box_{i} \Box_{\Box} \Box_{i} \Box_{i+1} \Box_{\Box} \Box_{\Box} \Box_{i+1} \Box_{\Box} \Box_{\Box} \Box_{i+1} \Box_{\Box} \Box_{\Box} \Box_{i+1} \Box_{i+1} \Box_{\Box} \Box_{i+1} \Box_{i+1} \Box_{\Box} \Box_{i+1} \Box_{i$$

 $\partial LogLx \ m(\partial \Theta; \beta \Theta, ) = \sum_{i+1} \Box \Box \Box \Box \Box_{k=} \sum_{0 \le m=0} (-1) \Box \Box_k \Box \Box \Box \Box \Box \Box \Box m k \Box \Box \Box \Box \Box + \Theta \Theta(x \ k_{2i} + +1 + 1) \Box \Box \Box \Box \Box \Box \Box + (x \ k_i \Theta + \Theta) \Box = 0$ 

 $\Gamma'(k)$  [29,30] where  $\Psi(k) =$  \_\_\_\_\_ is the digamma function .  $\Gamma(k)$ 

The above derivative equations cannot be solved analytically, therefore we use Newton Raphson method which is a simple and powerful technique for solving equations numerically. Therefore, parameter estimates will be obtained by maximizing the loglikelihood function using a numerical iterative method.

#### CONCLUSION

This paper proposed a new distribution which was obtained by mixing GNB distribution with a shanker distribution. It was found that NB-SH and Binomial-Shanker distributions are its special cases. Some mathematical properties which relates to its special case was provided. Parameter estimation of GNB-SH and NB-SH using MLE. Finally, our future interest will be in comparing the efficiency of this distribution with that of Poisson and NB distributions using real data sets.

## REFERENCES

- 1. Shanker R. The discrete poisson-shanker distribution. J of Biostat. 2015a.
- 2. Zamani H, et al. Negative binomial-lindley distribution and its application. J Math Stat. 2010;6(1):4-9.
- 3. Lord D, et al. The negative binomial-lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. Accid Anal Prev. 2011;43(5):1738-1742.
- 4. Geedipally SR, et al. The negative binomial-lindley generalized linear model: Characteristics and application using crash data. Accid Anal Prev. 2012;45:258-265.

- 5. Shirazi M, et al. A methodology to design heuristics for model selection based on the characteristics of data: Application to investigate when the negative binomial lindley (nb-l) is preferred over the negative binomial (nb). Accid Anal Prev. 2017;107:186-194.
- 6. Pudprommarat C, et al. Stochastic orders comparisons of negative binomial distribution with negative binomial-lindley distribution. Open J Stat. 2012;2(02):208.
- 7. Saengthong P, et al. Negative binomial-crack (nb-cr) distribution. Int J Pure Appl Math. 2013;84(3):213-230.
- 8. Saengthong P, et al. The zero inflated negative binomial-crack distribution: some properties and parameter estimation. J Sci Technol. 2015;37(6):701-711.
- 9. Gerstenkorn T. A compound of the generalized negative binomial distribution with the generalized beta distribution. Open Math. 2004;2(4):527-537.
- 10. Rashid A, et al. A compound of zero truncated generalized negative binomial distribution with the generalized beta distribution. J Rel Stat Stud. 2013;6(1):11-19.
- 11. Rashid A, et al. A mixture of generalized negative binomial distribution with generalized exponential distribution. J Stat App Pro. 2014;3(3):451.
- 12. Aryuyuen S, et al. The negative binomial-generalized exponential (nb-ge) distribution. Appl Math Sci. 2013;7(22):1093-1105.
- 13. Aryuyuen S, et al. Zero inflated negative binomial-generalized exponential distribution and its applications. Songklanakarin J Sci Technol. 2014;36(4).
- 14. Sankaran M. The discrete poisson-lindley distribution. Biometrics. 1970;145-149.
- 15. Lindley DV. Fiducial distributions and bayes' theorem. J R Stat Soc Series B Stat Methodol. 1958;102-107.
- 16. Ghitany M, et al. Size-biased poisson-lindley distribution and its application. Met Inter J Stat. 2008a;66(3):299-311.
- Ghitany M, et al. Zero-truncated poisson-lindley distribution and its application. Math Comput Simul. 2008b;79(3):279-287.
- 18. Mahmoudi E, et al. Generalized poisson-lindley distribution. Commun Stat Simul Comput. 2010;39(10):1785-1798.
- 19. Asgharzadeh A, et al. Pareto poisson-lindley distribution with applications. J App Stat. 2013;40(8):1717-1734.
- 20. Shanker R, et al. A two-parameter poisson-lindley distribution. Int J Stat Sys. 2014;9(1):79-85.
- 21. Shanker R. The discrete poisson-amarendra distribution. IJSDA. 2016a.
- 22. Shanker R. The discrete poisson-sujatha distribution. Int J Pro Stat. 2016b;5(1):1-9.
- 23. Consul PC, et al. Lagrangian probability distributions. Springer. 2006.
- 24. Balakrishnan N, et al. A primer on statistical distributions. John Wiley & Sons. 2004.
- 25. Johnson N, et al. Univariate discrete distributions. 1992.
- 26. Shanker R. Shanker distribution and its applications. Int J Stat App. 2015b;5(6):338-348.
- 27. Gurland J. Some interrelations among compound and generalized distributions. Biometrika. 1957;44(1/2):265-268.
- 28. Ahmad Z, et al. A new discrete compound distribution with application. J Stat Appl Pro. 2017;6:233-241.
- 29. Gupta PL, et al. On the moments and factorial moments of a mpsd. In Statistical Distributions in Scientific Work. Springer. 1981;189-195.
- 30. Abramowitz M, et al. Handbook of mathematical functions: with formulas, graphs, and mathematical tables, volume 55. Courier Corporation. 1964.