International Journal of Theoretical and Applied Mathematics
2024, Vol. 10, No. 1, pp. 1-5
https://doi.org/10.11648/j.ijtam.20241001.11

Research Article

Negative Binomial Three Parameter Lindley Distribution and Its Properties

Wanjala Anjela[1, *], George Muhua[1, 2]

[1]Department of Mathematics, Faculty of Science and Technology, University Nairobi, Nairobi, Kenya
[2]Department of Mathematics, University Nairobi, Nairobi, Kenya

Abstract

Many researchers have proposed mixed distributions as one of the most important methods for obtaining new probability distributions. Several studies have shown that mixed Negative Binomial distributions fits count data better than Poisson and Negative Binomial distribution itself. In this paper, we introduce a mixed distribution by mixing the distributions of negative binomial and three Parameter Lindley distribution. This new distribution has a thick tail and may be considered as an alternative for fitting count data with over dispersion. The parameters of the new distribution are estimated using MLE method and properties studied. Special cases of the new distribution and also identified. A simulation study carried out shows that the ML estimators give the parameter estimates close to the parameter when the sample is large, that is, the bias and variance of the parameter estimates decrease with increase in sample size showing the consistent nature of the new compound distribution. The study also compares the performance of the new distribution over distributions of Poisson, Negative Binomial, Negative Binomial oneParameter Lindley Distribution, Negative Binomial two Parameter distribution, three parameter Lindley distribution using a real count over dispersed dataset and the results shows that Negative Binomial three parameter Lindley distribution gave the smallest Kolmogorov Smirnov test statistic, AIC and BIC as compared to other distributions, hence the new distribution provided a better fit compared to other distributions under study for fitting over dispersed count data.

Keywords

Overdispersion, Negative Binomial Distribution, Lindley Distribution

1. Introduction

Since 1894, a large number of writers have studied the concept of mixed distribution. For example, Blischke defined the mixed distribution as the sum of the positive-weighted probability distributions, or the mixing distributions that add up to one. The Poisson distribution is typically used to fit the count data when the number of occurrences is randomly distributed over the time and/or space over which the event counts occur. In reality, though, observed count data often exhibit traits like under dispersion and overdispersion that are common in applied data analysis [15]. [5] proposed a model with a Poisson mean gamma, or negative binomial (NB), distribution. There is growing popularity for the NB distribution as a more flexible option [12].

*Corresponding author: Wanjalaanjela@gmail.com (Wanjala Anjela)

Received: 17 October 2023; **Accepted:** 2 January 2024; **Published:** 14 June 2024

2. Literature Review

A one-parameter distribution known as the Lindley distribution was initially presented by Lindley [3, 11]. It is a finite mixture of the exponential θ and γ(2, θ) distributions. where the scaling parameter is denoted by θ. By combining the inverse Gaussian distribution with the p negative binomial parameter, [4] created a novel compound negative binomial distribution. The new distribution's fundamental characteristics were described, along with three methods for estimating parameters: the maximum likelihood technique, the zero proportion method, and the method of moments [14]. Lastly, both univariate and bivariate case application examples were provided [1]. For count data when the probability at zero has a big value, [18] developed the negative binomial-Lindley distribution, which offers a better fit than the Poisson and the negative binomial. They provided estimates of the parameters and basic attributes.

3. Negative Binomial Three Parameter Lindley Distribution

3.1. Definition

If the random variable X complies with the following stochastic formulation, it has a negative binomial-3 parameter Lindley distribution.

$$X \mid r, \lambda \sim NB(r, p = \exp(-\lambda))$$

And

$$\lambda \sim L_3(\lambda \mid a, b, c)$$

Where x>0, r>0, a>0, b>0, c>0

$$\lambda \sim f(\lambda \mid a, b, c) = \frac{c^2}{ca+b}(\mathrm{a}+\lambda\,\mathrm{b})^{\exp(-cx)}$$

distribution refers to 3 parameter lindley distribution.

Theorem Let X be a random variable that has three parameters r, a, b, and c—and a negative binomial Lindley distribution. Next, X's pmf is provided by

$$\Pr(X = x \mid r, a, b, c) = \binom{x+r-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \frac{c^2}{ac+b} [\frac{b+a(c+r+j)}{(c+r+j)^2}$$

X=0,1; where a> 0, b > 0 and c> 0

and the corresponding cdf is

$$F_X(\mathrm{x}) = \sum_{s<=x} f_x(x) = \sum_{s<=x} \binom{r+s-1}{s} \sum_{j=0}^{s} \binom{s}{j} (-1)^j \left(\frac{c^b}{(c+r+j)^b} + \frac{c^{a+1}}{(c+r+j)^a}\right) \frac{1}{c+1}$$

3.2. Properties of Negative Binomial Three Parameter Lindley Distribution

3.2.1. Shape of the Distribution

The following graph displays the pdf of the three-parameter Lindley-negative binomial distribution for various values of c, b, a, and r:

According to the figure, when c, b, are large, the pmf of LNB3 has the lowest mass at zero and the probability is noticeably tiny at zero. However, the LNB3 pmf is right-tailed and has the maximum mass at zero when these parameters are minimal. In the event that the probability at zero has a small or big value, the proposed distribution provides an alternate distribution that can adequately match the proportional data [16]. Additionally, c is a shape parameter because it alters the pmf's overall shape, which determines the distribution's tail. r is the dispersion parameter from Negative Binomial distribution, and it indicates whether the distribution is wide or narrow. from the pdf diagram, the large values of r indiate a narrow distribution whereas small values of r results to a wider distribution [10].

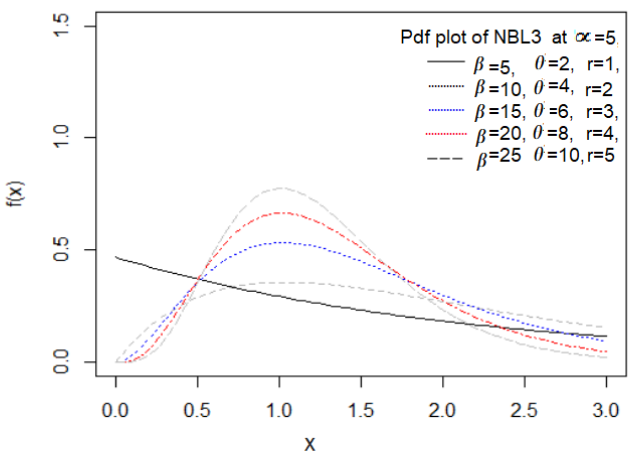

Figure 1. Shape of the NBL3 with varying values of the parameters.

3.2.2. Index of Dispersion

When comparing a set of observations to a conventional statistical model, the variance-to-mean ratio, also known as the index of dispersion, is a highly helpful tool for determining how distributed or clustered the data are. The LB3 distribution's index of dispersion is found to be

$$\delta = \frac{\delta^2}{\zeta}$$

$$= \frac{(r^2+r)\pi_2 - (2r^2+r)\pi_1 + r^2 - r(\pi_1 - 1)}{r(\pi_1 - 1)}$$

$$= \frac{(r^2+r)\pi_2 - r\pi_1(1+r\pi_1)}{r(\pi_1 - 1)}$$

Given that the value or r is greater than one, the dispersion index will be greater than one, implying that the distribution is over dispersed.

3.3. Special Cases of Three Parameter Lindley Distributio

Let $X \sim NBL_3(\mathrm{r,a,b,c})$.

Given that a = 1 and b = 1, With positive r and c parameters, the negative binomial Lindley (NB-L) distribution is obtained. Its PMF is

$$f(x:r,c) = \binom{r+x-1}{x}\frac{c^2}{c+1}\sum_{j=0}^{x}\binom{x}{j}(-1)^j \frac{a(c+r+j)+1}{(c+r+j)^2}, \quad x=0,1,2\ldots$$

It is the NB-L distribution that [18] suggested. If a=1, we get

$$f(x:r,b,c) = \frac{c^2}{c+b}\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j \frac{a(c+r+j)+b}{(c+r+j)^2}, \quad x=0,1,2\ldots$$

whichis the Negative Binomial Two Parameter Lindley Distribution

If a=1 and b=0, we get

$$f(x:r,c) = c\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j \frac{1}{(c+r+j)}, \quad x=0,1,2\ldots$$

This is the exponential distribution.

Simulation

This part simulates the generation of random variables from the NBL3 (r, a, b, c). The created sample is then used to determine the MLE of the parameters. To determine the consistency of the estimates, The parameters' MLE's bias and mean squared error (MSE) are calculated. The algorithm for the requested simulation research is shown below.

Making a random sample with NBL3 r, a, b, and c

The U(0,1) distribution is used to create a random variable.

If the xi, i >= 0, are arranged in the following order: x0, x1, x2.. and if we let

The CDF define the distribution of X,

If $F(x_i - 1) = uF(x_i), i = 0,1,2....$, then X=xi

As many times as the desired sample size is, above steps are performed. The discrete inverse transform method for generating X is what the aforementioned technique is called. Getting the parameters' MLE For the produced sample collected in the preceding phase, the maximum likelihood equations (MLE) of r, c, a, and b are solved. Bias and MSE of the MLEs are calculated. Assume that the MLE is c^* and that the parameter c true value is c_0. When predicting c_0, the bias of c^* is thus given by

$$\mathrm{Bias}(c^*) = E(c^* - c_0)$$

Regarding the NBL3 mass function, the expectation is (r, c, b, a).

Similar to this, MSE(c^*) = E[(c $*$ −c_0)2 is used to calculate the MSE of c_0.

Table 1. Average values of Bias and Variance of r, a, b, c for NBL3.

Sample Size	c=3		a=4		b=2		r=5	
	Bias (c)	Var (c)	Bias (a)	Var (a)	Bias (b)	Var (b)	Bias (r)	Var (r)
50	0.64619	0.51673	1.87777	3.52602	1.20045	1.46335	1.27721	1.56335
100	0.58956	0.37721	1.60015	2.65715	1.19996	1.44537	1.15996	1.44557
200	0.23850	0.06977	1.03593	1.23732	0.69749	0.50399	0.65745	0.50399
300	0.21259	0.05419	0.05419	0.82120	0.49964	0.25864	0.45964	0.25564

The Monte Carlo approximation method, using T=1000 repetitions, approximates the Bias and MSE of the MLE of c. The Bias and MSE of the MLE of r, a and b are determined in a similar way. If the Bias lowers (gets closer to zero) with an increase in sample size and the MSE also decreases, the MLE is considered to be consistent. The values of the Bias and MSE of the MLE of r, b, a, and c for the various sample sizes are displayed in Table 1. Table 1 shows that as the sample size is increased, the bias and MSE of r, beta, alpha, and theta decrease and eventually reach zero. The MLE estimations are therefore consistent and accurate in determining the true value of the parameters [8], it can be said. R software, version 3.4.4, is used to do calculations related to the study with the aid of self-programmed codes. The parameters from NBL3 (r, β, α, andθ) are estimated using the maximum likelihood method using the R software's maxLik package.

4. Application to Real Data

This is to show the importance of NBL3 for count data analysis. We consider a real data set. Distributions such the Poisson, NB, L3, NBL, NB-NWL, and NBL3 distributions are utilized to fit this data set. To determine the significance of the K-S statistics, we utilized the ks. test function in the dgof package of R. The optimal distribution is indicated by the KS test's lower statistic result [6]. The dataset includes all of the claims filed under the third-party liability auto insurance policy that [17] examined. We compare the fit of the NBL3 distribution to that of the Poisson, Negative Binomial, and three parameter Lindley distributions.

Table 2. Comparison of NBL3 distribution with other distributions.

Number of claims	Observed Values	Poisson	Negative Binomial	Negative Binomial one Parameter Lindley	Negative Binomial two Parameter Lindley	Three parameter Lindley	Negative Binomial three parameter Lindley
0	7,840	7,635.27	7,846.64	7,763.54	7,796.23	7,811.55	7,837.40
1	1,317	1,637.00	1,288.58	1,356.39	1,361.03	1,346.35	1,326.20
2	239	175.49	256.64	241.98	205.80	244.61	226.34
3	42	12.54	54.10	44.29	30.25	46.67	48.74
4	14	0.67	11.72	8.31	4.45	9.32	14.03
5	4	0.03	2.58	1.60	0.66	1.94	4.97
6	4	0.00	0.57	0.31	0.10	0.42	1.95
7	1	0.00	0.13	0.06	0.02	0.10	0.79
Estimated Values of parameters		$\lambda = 0.2144$	P=0.7659 r=0.7015	r=30.4775 c=143.966	r=21.7015 a= 1.4737 c=162.372	b= 21.7876 a=1.4737 c=162.372	r=4.8261 a=1.8355 b=12.1718 c=48.1251
-l		5,590.78	5,396.31	5,396.29	5,431.77	5,348.97	5,341.93
KS test		0.0216	0.0022	0.0081	0.0031	0.0066	0.0007
(p-value)		(0.0003)	(0.9999)	(0.5669)	(0.8039)	(0.9999)	(0.9999)
AIC		857.49	857.41	857.32	857.31	857.01	855.06
BIC		865.78	865.71	864.67	864.15	864.14	859.20

The estimated parameters for each distribution are obtained using R's nlm and MLE functions. To compare the parameter estimate of each distribution, we consider the ˆℓ, which is the maximal value of the log-likelihood function under the studied distributions. The ratio of variance to mean, which measures the dispersion of the data and is 1.78, shows that the dataset is overdispersed.

We offer the values of the Kolmogorov-Smirnov (KS) statistic for further discussion. Goodness of fit test statistics, like the Komolgorov-Smirnov test statistics, are used to validate the superiority of the NBL3 distribution for the number of claims in the third liability auto insurance dataset [2] in order to model the number of claims of the third liability auto insurance data [9], the NBL3 performs better than the poisson [7], NB, NBL, L3, and NBL2 (see Table 2 above), with the NBL3 having the lowest value of −2log(L), AIC, BIC, and KS. The NBL3 distribution, which has the lowest AIC, BIC, and KS [13], is the best model to describe the volume of third-party liability insurance claims, according to the table's results.

$$\Pr(X = x | r,a,b,c) = \binom{x+r-1}{x} \frac{c^2}{ac+b} \sum_{j=0}^{x} \binom{x}{j} (-1)^j [\frac{a(c+r+j)+b}{(c+r+j)^2}$$

Where r=4.8261, a=1.8353, b=12.1718, c=48.1251

5. Conclusion

The proposed distribution, called the NBL3 distribution, generalizes the two- and three-parameter negative binomial-Lindley distributions. Due to its unique features, the proposed distribution has a large number of mixed negative binomial distributions, including negative binomial-generalized exponential distributions. Some statistical features for the proposed distribution have been provided to help understand the NBL3 distribution. Among the characteristics examined were the dispersion index and the kth factorial moment. Determining the skewness, kurtosis, and higher order moments of the NBL3 is made easier by these statistical characteristics: The dispersion index suggests that the data generated by the NBL3 distribution may exhibit overdispersion characteristics.

Conflicts of Interest

The authorsdeclareno conflicts of interest.

References

[1] Elbatal, I., Merovci, F. & Elgarhy, M. (2013) A new generalized Lindley distribution. Mathematical Theory and Modeling, 3(13), 30-47.

[2] Gencturk, Y. & Yi ¨ giter, A. (2016) Modelling claim number using a new mixture ˘ model: negative binomial gamma distribution. Journal of Statistical Computation and Simulation, 86(10), 1829-1839.

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

[4] Gomez-D ´ eniz, E., Sarabia, J. M. & Calder´ın-Ojeda, E. (2008). Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. Insurance: Mathematics and Economics, 42(1), 39-49.

[5] Greenwood, M. & Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. Journal of the Royal Statistical Society, 83(2), 255-279.

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

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

[8] Jambunathan M. B. (1954). Some properties of beta and gamma distributions. The annals of mathematical statistics. 25(2), 401-405.

[9] Klugman, S. A., Panjer H. H. & Willmot G. E. (2008). Loss Models: From Data to Decision. 3rd Edn., John Wiley and Sons, USA: 101-159.

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

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

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

[13] Marsaglia, G., Tsang, W. W. & Wang, J. (2003), Evaluating Kolmogorov's distribution. Journal of Statistical Software, 8(18), 1-4.

[14] Panjer, H. H. & Willmot, G. E. (1981). Finite sum evaluation of the negative binomial-exponential model. ASTIN Bulletin: The Journal of the IAA, 12(2), 133- 137.

[15] Rainer, W. (2000). Econometric Analysis of Count Data, 3rd edn, Springer Verlag, Berlin, Germany.

[16] Sankaran, M. (1970). The discrete Poisson-Lindley distribution, Biometrics, 26(1), 145-149.

[17] Wang, Z. (2011). On mixed negative binomial distribution with application. Journal of Statistical Planning and Inference, 141(3), 1153-1160.

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