Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/050220

On Negative Binomial-Two Parameter Lindley Distribution

Adil Rashid[*], *Zahoor Ahmad and T. R Jan*

P.G Department of Statistics, University of Kashmir, Srinagar, India

Received: 27 May 2019, Revised: 2 Jan. 2020, Accepted: 27 Feb. 2020.
Published online: 1 Jul. 2021

Abstract: In this paper, we will explore the applications of the compound of negative binomial-two parameter Lindley distribution (NBTPLD) in traffic data where the observed data is over dispersed and heavy tailed. Furthermore, the comparative study of some existing classical distributions with respect to the (NBTPLD) will be made statistically. At the end, we will show how this compound distribution can be used as a device to accommodate any overdispersion caused due to high concentration of zeros.

Keywords: Negative binomial distribution, two parameter Lindley distribution, compound distribution, factorial moments, over-dispersion, count data.

1 Introduction

In probability distribution theory, there are vast number of discrete distributions that have specified applications but sometimes the observable data have distinct features that are not exhibited by these classical discrete distributions. So, to overcome these limitations, researchers often develop new probability distributions so that these new distributions can be employed in those situations where classical distributions are not providing any adequate fit. There are so many techniques by which we can obtain new distributions such as transmutation, compounding, etc., but compounding of distributions has received maximum attention and has been exploited by a good corpus of researchers due to its simplified approach. Some new compound distributions were obtained in the recent past by numerous researchers, for instance; Adil et.al constructed numerous discrte distributions to address various issues pertaining to countdata analysis and overdispesion phenomenon (2014a, 2014b, 2016, 2020, 2021). Intrestingly, these newly established discrete compound distributions provided optimally improved fit in comparison to some famous classical distribution like Poisson and nagative binomial. Some versatile continuous lifetime models were also proposed by Adil et al (2018a, 2018b) with applications in engineering and biological sciences. The authors like Sankaran et al (1970), Gomez et al (2008), Kulgman et al (2008), Zamni et al (2010) and Shanker et al (2013) worked rigorously on discrete compound distribution to hightlight their wide range of flexibility. Here, we shall explore some new applications of the already proposed compound distribution due to Adil and Jan (2015 proceedings) called negative binomial-two parameter Lindley distribution (NBTPLD).

2 Probability Mass Function of NBTPLD

The probability mass function of the compound of NBTPLD is given by

$$f(x;r,\alpha,\theta)=\binom{r+x-1}{x}\frac{\theta^2}{\theta+\alpha}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left(\frac{r+\theta+j+\alpha}{(r+\theta+j)^2}\right)$$

where $x=0,1,2....,$` $r,\alpha,\theta>0$

Special cases of NB-TPL (r,α,θ) model :

Case (i): For $\alpha=1,$ TPLD (α,θ) reduces to one parameter Lindley distribution L (θ); therefore a compound of NBD

[*]Corresponding author e-mail: adilstat@gmail.com

with one parameter Lindley distribution will be obtained by simply substituting $\alpha=1$ in (4)

$$f(x:r,\theta)=\frac{\theta^2}{\theta+1}\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left\{\frac{r+\theta+j+1}{(r+\theta+j)^2}\right\},\qquad x=0,1,2....\quad r,\theta>0$$

The above compound probability mass function was obtained by Zamani et al.(2010)

Case(ii): ***F***or $\alpha=0,$ TPLD (α,θ) reduces exponential distribution; therefore a compound of NBD with exponential distribution is followed from (4) when we put $\alpha=0$ in it.

$$f(x:r,\theta)=\theta\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left\{\frac{1}{(r+\theta+j)}\right\},\quad x=0,1,2....\quad r,\theta>0$$

The probability function displayed above was obtained by Willmot et al. (1981)

Case (iii): *For* $r=1,$ *NBD reduces to geometric distribution and a compound of geometric distribution with TPLD* (α,θ) *is followed from (4) when we put* $r=1$ *in it.*

$$f(x:\alpha,\theta)=\frac{\theta^2}{\theta+\alpha}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left\{\frac{1+\theta+j+\alpha}{(\theta+j+1)^2}\right\},\ x=0,1,2....\quad \theta,\alpha>0$$

Case (iv): *For* $r=1$ *and* $\alpha=1$ *in (4), we get a compound of geometric distribution with one parameter Lindley distribution*

$$f(x,\theta)=\frac{\theta^2}{\theta+1}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left\{\frac{\theta+j+2}{(\theta+j+1)^2}\right\},\ x=0,1,2....\quad \theta>0$$

Case (v): *For* $r=1$ *and* $\alpha=0$ *in (4), we get a compound of geometric distribution with exponential distribution*

$$f(x,\theta)=\theta\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left\{\frac{1}{(\theta+j+1)}\right\},\ x=0,1,2....\ ,\theta>0$$

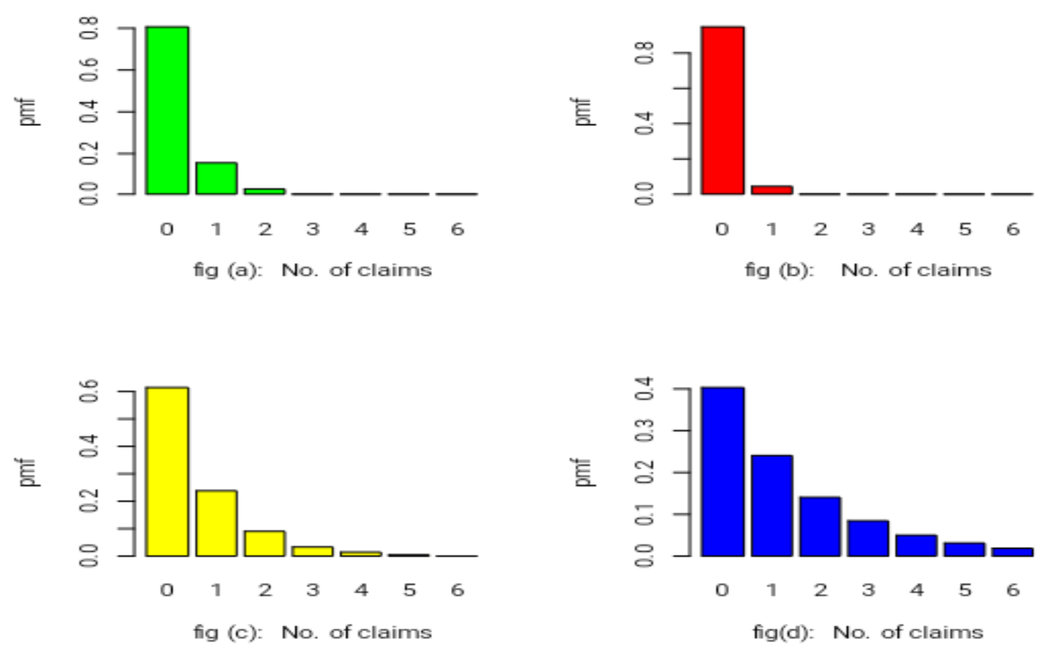

Fig.1(a): Plot of pmf for NB-TPLD $(r=20,\alpha=20,\theta=100)$, fig. 1(b): Plot of pmf for NB-TPLD

$(r=7,\alpha=20,\theta=150)$, fig. 1(c): Plot of pmf for NB-TPLD $(r=22,\alpha=30,\theta=50)$, fig. 1(d): Plot of pmf for NB-TPLD $(r=9,\alpha=10,\theta=10)$.

3 Properties of NB-TPL (r,α,β) Model

The mean of NB-TPL (r,α,θ) model

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

$$\mu_{[1]}(X)=r\delta_1$$

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

$$\mu_{[2]}(X)=r(r+1)\delta_2$$

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

$$\mu_{[3]}(X)=r(r+1)(r+2)\delta_3$$

$$E(X^2)=\mu_{[2]}(X)+\mu_{[1]}(X)=r(r+1)\delta_2+r\delta_1$$

$$E(X^3)=\mu_{[3]}(X)+3\mu_2'-2\mu_1'=(r+1)(r+2)\delta_3+3r(r+1)\delta_2+r\delta_1$$

$$V(X)=r(r+1)\delta_2+r\delta_1(1-r\delta_1)$$

$$\sigma(X)=\sqrt{r(r+1)\delta_2+r\delta_1(1-r\delta_1)}$$

$$CV(X)=\frac{\sqrt{r(r+1)\delta_2+r\delta_1(1-r\delta_1)}}{r\delta_1}$$

where

$$\delta_1=\frac{\theta^2+2\alpha\theta-\alpha-\theta}{(\theta+\alpha)(\theta-1)^2},\ \delta_2=\frac{\theta^2(\theta-1)^2(\alpha+\theta-2)-2\theta^2(\theta-2)^2(\alpha+\theta-1)}{(\theta+\alpha)(\theta-1)^2(\theta-2)^2}+1$$

and

$$\delta_3=\frac{\theta^2(\theta-1)^2(\theta-2)^2(\alpha+\theta-3)-3\theta^2(\theta-1)^2(\theta-3)^2(\alpha+\theta-2)+3\theta^2(\theta-2)^2(\theta-3)^2(\alpha+\theta-1)}{(\theta+\alpha)(\theta-1)^2(\theta-2)^2(\theta-3)^2}-1$$

4 Results and Discussion

Here, we illustrate the application of the proposed model by fitting it to the real data set taken from Klugman et al. (2008). The data which appears in the first two columns of table 1, 2 and 3 provides information of observed counts of the number of accidents and crashes on automobile insurance policies. It is clear that given data is over-dispersed since sample variance

(0.288) is greater than the sample mean (0.214). Therefore, the given data must be fitted by some over-dispersed model. For this reason, we strongly believe that proposed model is suitable for them. By comparing the fitted distributions in Table 1,2 and 3; based on p value and log likelihood value it is quite clear and evident that NBTPLD model is a better and highly flexible when it comes to model the over-dispersed count data with large number of zeroes.

Table 1: Number of accidents on automobile insurance policies (Klugman et al. 2008).

<table>
<tr><th rowspan="2">Number of accidents</th><th rowspan="2">Observed Frequency</th><th colspan="4">Fitted Distribution</th></tr>
<tr><th>Poisson</th><th>NBD</th><th>NBD-L</th><th>NBD-TPL</th></tr>
<tr><td>0</td><td>7840</td><td>7638.30</td><td>7843.70</td><td>7853.6</td><td>7878.08</td></tr>
<tr><td>1</td><td>1317</td><td>1634.61</td><td>1290.20</td><td>1287.4</td><td>1271.95</td></tr>
<tr><td>2</td><td>239</td><td>174.90</td><td>257.30</td><td>247.6</td><td>241.22</td></tr>
<tr><td>3</td><td>42</td><td rowspan="6">12.5
0.7
0
0
0
0
1 3.2</td><td>54.50</td><td>54.2</td><td>49.64</td></tr>
<tr><td>4</td><td>14</td><td rowspan="5">11.8
2.6
0.6
0.2
0.1
15.3</td><td>13.2</td><td>12.5</td></tr>
<tr><td>5</td><td>4</td><td rowspan="4">3.5
1.0
0.3
0.2
5</td><td rowspan="4">3.30
3.94
0.28
0.09
7.6</td></tr>
<tr><td>`6</td><td>4</td></tr>
<tr><td>7</td><td>1</td></tr>
<tr><td>8+</td><td>0</td></tr>
<tr><td>Total</td><td>9461</td><td>9461</td><td>9461</td><td>9461</td><td>9461</td></tr>
<tr><td>Parameter Estimation</td><td></td><td>$\lambda = 0.214$</td><td>$r = 0.70$
$\theta = 0.76$</td><td>$r = 4.63$
$\theta = 23.55$</td><td>$r = 4.63$
$\theta = 23.55$
$\alpha = 0.52$</td></tr>
<tr><td>Chi-squareEstimates and DF P-value</td><td></td><td>293.80, DF=2 P-value=0</td><td>865,DF=2 P-value<0.01</td><td>6.99,DF=3 P-value=0.07</td><td>4.21, DF=2 P-value=0.12</td></tr>
</table>

Table 2: Single Vehicle Fatal Crashes on Divided Multilane Rural Highways in Texas between 1997 and 2001.

Crashes	Observed Frequency	Fitted Distribution		
		Poisson	NB	NB-L
0	1532	1509.2	534.4	1532.9
1	162	198.2	154.7	158.3
2	19	13.0	25.8	23.7
3	6	0.6	4.9	4.6
4+	2	0.0	1.2	1.4
Parameter Estimation		$\mu=0.131$	$\mu=0.131$ $\varphi=0.434$	$\theta=15.984$ $r = 1.851$
Chi-square		102.99	2.73	1.68
Log-Likelihood		-715.1	-696.1	-695.6

Table 3: Single Vehicle Roadway Departure Crashes on Rural Two Lane Horizontal curves in Texas between 2003 and 2008.

Crashes	Observed Frequency	Fitted Distribution		
		Poisson	NB	NB-L
0	29087	28471.6	29204.8	29133.6
1	2952	39180.0	2706.0	2855.5
2	464	269.6	567.4	503.1
3	108	12.4	141.1	120.9
4	40	0.4	37.8	35.9
5	9	0.0	10.6	13.1
6	5	0.0	3.0	3.3
7	2	0.0	0.9	3.3
8	3	0.0	0.3	0.0
9	1	0.0	0.1	0.0
10+	1	0.0	0.0	3.3
Parameter Estimation		μ=0.138	μ=0.138 φ=0.284	θ=9.212 r = 1.018
Chi-square		2297.31	57.47	11.68
Log-Likelihood		-14208.1	-13557.7	-13529.8

5 Conclusions

It has been shown statistically NBTPLD performs excellently well when it comes to model heavy tailed over-dispersion; hence, we recommend practitioners to use NBTPLD model while dealing with heavy tailed over-dispersion phenomenon.

Conflict of interest

There is no conflict of interest in this work.

References

[1] Gomez, D.E., Sarabia, J.M and Ojeda, E.C (2008) Univariate and Multivariate Versions of the Negative Binomial-Inverse Gaussian Distributions with Applications. *Insurance Mathematics and Economics.,* **42**, 39–49(2008).
[2] Kulgman, S.A., Panger, H.H and Willmot, G.E. (2008) Loss Models from data to decision.3rdEdn. John Wiley and sons, USA., 101-159(2008).
[3] Adil, R and Jan, T.R. (2013) A compound of zero truncated generalized negative binomial distribution with the generalized beta distribution. Journal of Reliability and Statistical studies., **6(1)**, 11-19(2013).
[4] Adil, R and Jan, T.R. (2014a) A compound of Geeta distribution with the Generalized Beta Distribution.Journal of Modern Applied Statistical Methods.,**13(1)**, 278-286(2014a).
[5] Adil, R and Jan, T.R. (2014b) A compound of Consul distribution with the Generalized Beta Distribution. Journal of Applied probability and Statistics Methods., **9(2)**, 17-24(2014b).

[6] Adil, R and Jan, T.R. (2015) A compound of negative binomial distribution with two parameter Lindley distribution as a tool for handling oversiapersion, Proceedings of International Conference held at Department of Electronics, University of Kashmir, (SEEDS)
[7] Adil, R and Jan, T.R. (2016) A new count data model with application in genetics and ecology. Electronic Journal of Applied Statistical Analysis., **9(1)**, 213-226(2016).
[8] Adil, R and Jan, T.R. (2018a) A new lifetime distribution with application for series system: Model, properties and application., Journal of Modern Applied Statistical Methods., **17(1)**, 1-25(2018a).
[9] Zahoor, A., Adil R., and Jan T. R. Complementary compound Lindley power series distribution with application. Pakistan Journal of Statistics and Operations Research., **14(1)**, 139-155.
[10] Anwar, H., Gulzar A. S, Sheikh B. A and Adil, R (2020) A new flexible discrete dstribution with application to count data, Journal of Statistical Theory and Application., 19(1), 102-108(2020).
[11] Zahoor, A., Adil, R and Arshid, I (2021) Generalized Negative binomial distribution: Compound negative binomial distribution as a sum of Laplace variates, Journal of Statistics Applications and Probability., **10(1)**, 77-84(2021).
[12] Sankaran, M. (1970) The Discrete Poisson-Lindley Distribution. Biometrics., **26**, 145-149(1970).
[13] Shanker, R. Mishra, A. (2013) A two parameter Lindley distribution.Statistics in Transition new Series.,**14(1)**, 45-56(2013).
[14] Zamani, H. and Ismail, N. (2010) Negative Binomial–Lindley Distribution And Its Application. J. Mathematics And Statistics.,**1**, 4–9(2010).