ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **12** (2023), no.3, 457–464
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.12.3.3

BETA TRANSFORMATION OF THE LINDLEY POISSON DISTRIBUTION FOR OVER-DISPERSED COUNT DATA

Chedly Gélin Louzayadio[1], Michel Koukouatikissa Diafouka, and Rodnellin Onesime Malouata

ABSTRACT. A new distribution for over-dispersed count data is proposed, and its properties are studied. This is a two-parameter distribution which is obtained by introducing an additional parameter beta into the Poisson-Lindley distribution. The goodness-of-fit of this distribution is compared with other distributions that have been proposed to model overdispersion. Two illustrative examples are presented to show the flexibility of the model.

1. INTRODUCTION

Even though the Poisson distribution is considered as a reference for modelling count data, the restriction to have the variance equal to the mean (equidispersion) is often too constraining in practice. For many observed count data, it is common for the sample variance to be greater or less than the sample mean, known as overdispersion or underdispersion with respect to the Poisson distribution. There is then a rich variety of alternative distributions that can be used to model the data (see, [3, 4, 6, 9, 10]).

To account for dispersion, several authors have proposed a mixed Poisson distribution or an extension of the Poisson distribution. Sankaran [11] proposed the

[1]*corresponding author*
2020 *Mathematics Subject Classification.* 60E05, 62H05.
Key words and phrases. Beta transformation, Poisson-Lindley distribution, over-dispersion, goodness-of-fit.
Submitted: 02.02.2023; *Accepted:* 17.02.2023; *Published:* 04.03.2023.

discrete Poisson-Lindley distribution by combining the Poisson distribution with one provided by Lindley. This distribution has probability mass function

$$P(X=k)=\frac{\theta^2(\theta+k+2)}{(\theta+1)^{k+3}},\;\; k=0,1,\ldots;\;\; \theta>0. \tag{1.1}$$

The author estimates its parameter and gives two examples of fitting this distribution to overdispersed data. He considers that this distribution can be used as an approximation to the negative binomial distribution (NBD). Mahmoudi [8] proposed an extension of the Lindley-Poisson distribution by combining the Poisson distribution with the generalized Lindley distribution. He calls it generalized Poisson-Lindley distribution (GPLD) and estimates its parameters using the method of moments and maximum likelihood. He gives examples of fitting this distribution of data and then compares it with other discrete distributions. Bhati [2] introduced a new generalized Poisson-Lindley distribution (NGPLD) through combination of the Poisson distribution with a two-parameter generalized Lindley distribution. They studied its properties and shown that it is better performing than other competing models through applications on real data sets. Aderoju [1] introduced a New Generalized Poisson-Sujatha distribution (NGPSD). He constructs it from a mixture of a Poisson distribution with a generalized two-parameter sujatha distribution and studies its properties and goodness of fit by comparing it with other distributions in statistical literature.

As one of the alternatives to the Poisson distribution, the authors in [5] proposed a new specific transformation for discrete distributions, called the beta transformation. They define it as follows: Let X be a discrete random variable with probability mass function $p_k = P(X=k)$. Its beta transformation Y is also a discrete random variable with probability mass function

$$p_k = P(Y=k)=\begin{cases}\dfrac{1-p_0}{\beta} & k=0\\ p_{k-1}-\dfrac{p_k}{\beta} & k=1,2,\ldots\end{cases}$$

where β satisfies the conditions $\beta > 1-p_0$ and $\beta \geq \max_{k\geq 1}\left(\dfrac{p_k}{p_{k-1}}\right)$. This is another way to introduce an additional parameter into the distribution of a given discrete random variable that can be used as a competitor for some common two-parameter distributions.

In this article, we propose the beta transformation of the Poisson-Lindley distribution (BTPLD) as one of the alternatives to the Poisson distribution for modeling count data. We show that our distribution provides enough flexibility for analyzing different types of count data. The paper is organized as follows: Section 2, we present some definitions and properties of BTPLD. In Section 3, we estimate parameters of BTPLD. Finally, in Section 4, we compare BTPLD with the POISSON, NB, NGPL, and NGPS distributions.

2. Proposed model and Some properties

2.1. Probability mass function. Let X be a Poisson-Lindley non-negative integer random variable with parameter $\theta > 0$ and probability mass function (pmf) given by (1.1)

Definition 2.1. *The beta transformation Y of X is a non-negative integer random variable which probability mass function is given by*

$$\begin{aligned} p_k &= P(Y=k) \\ &= \begin{cases} \dfrac{\theta^2+3\theta+1}{\beta(\theta+1)^3}, & k=0 \\ \dfrac{\theta^2(\theta+k+2)}{\beta(\theta+1)^{k+1}}\left(\dfrac{\beta(\theta+1)(\theta+k+1)}{\theta+k+2}-1\right), & k=1,2,\ldots \end{cases}, \end{aligned} \tag{2.1}$$

where the parameter β is presupposed that

$$\beta > 1-\frac{\theta^2(\theta+2)}{(\theta+1)^3} \quad \textit{and} \quad \beta \geq \max_{k\geq 1}\left(\frac{\theta+k+2}{(\theta+1)(\theta+k+1)}\right).$$

We refer to this as beta transformation of the Poisson-Lindley distribution, denote BTPL(θ,β).

Remark 2.1. *When β takes large values, the beta transformation Y will tend to shifting one step to the right of the original non-negative integer random variable X.*

Alternatively, the probability mass function (2.1) can be written as follows

$$\begin{aligned} p_k &= P(Y=k) \\ &= \left(\frac{\theta^2+3\theta+1}{\beta(\theta+1)^3}\right)^{\delta_0(y)}\left[\frac{\theta^2(\theta+k+2)}{\beta(\theta+1)^{k+1}}\left(\frac{\beta(\theta+1)(\theta+k+1)}{\theta+k+2}-1\right)\right]^{1-\delta_0(y)}, \end{aligned} \tag{2.2}$$

for all $k \in \mathbb{N}$, where

$$\delta_0(y) = \begin{cases} 1, & y = 0, \\ 0 & \text{otherwise}. \end{cases}$$

is the indicator function at 0.

Proposition 2.1. *The corresponding cumulative distribution function (c.d.f.) of Y is given by*

$$\begin{aligned} P(Y \leq k) &= \frac{\theta^2 + 3\theta + 1}{\beta(\theta+1)^3} + \frac{\theta\left((\theta+1)^{[k]} - 1\right)}{(\theta+1)^{[k]} - 1}\left(1 - \frac{\theta+2}{\beta(\theta+1)^2}\right) \\ &+ \left(\frac{(\theta+1)^{[k]} - \theta[k] - 1}{(\theta+1)^{[k]+1}}\right)\left(1 - \frac{1}{\beta(\theta+1)}\right), \end{aligned} \tag{2.3}$$

where $[k]$ is the integer part of k.

Proof.

$$\begin{aligned} P(Y \leq k) &= \frac{\theta^2 + 3\theta + 1}{\beta(\theta+1)^3} + \sum_{k=1}^{[k]} \frac{\theta^2(\theta+k+1)}{(\theta+1)^{k+2}} - \frac{1}{\beta}\sum_{k=1}^{[k]} \frac{\theta^2(\theta+k+2)}{(\theta+1)^{k+3}} \\ &= \frac{\theta^2 + 3\theta + 1}{\beta(\theta+1)^3} + \frac{\theta^2}{\theta+1}\sum_{k=1}^{[k]} \frac{1}{(\theta+1)^{[k]}} + \frac{\theta^2}{(\theta+1)^2}\sum_{k=1}^{[k]} \frac{k}{(\theta+1)^{[k]}} \\ &\quad -\frac{1}{\beta}\left(\frac{\theta^2(\theta+2)}{(\theta+1)^3}\sum_{k=1}^{[k]} \frac{1}{(\theta+1)^{[k]}} + \frac{\theta^2}{(\theta+1)^3}\sum_{k=1}^{[k]} \frac{k}{(\theta+1)^{[k]}}\right) \\ &= \frac{\theta^2 + 3\theta + 1}{\beta(\theta+1)^3} + \frac{\theta\left((\theta+1)^{[k]} - 1\right)}{(\theta+1)^{[k]}} + \frac{(\theta+1)^{[k]} - \theta[k] - 1}{(\theta+1)^{[k]+1}} \\ &\quad -\frac{1}{\beta}\left(\frac{\theta(\theta+2)\left((\theta+1)^{[k]} - 1\right)}{(\theta+1)^{[k]+2}} + \frac{(\theta+1)^{[k]} - \theta[k] - 1}{(\theta+1)^{[k]+2}}\right) \\ &= \frac{\theta^2 + 3\theta + 1}{\beta(\theta+1)^3} + \frac{\theta\left((\theta+1)^{[k]} - 1\right)}{(\theta+1)^{[k]} - 1}\left(1 - \frac{\theta+2}{\beta(\theta+1)^2}\right) \\ &\quad + \left(\frac{(\theta+1)^{[k]} - \theta[k] - 1}{(\theta+1)^{[k]+1}}\right)\left(1 - \frac{1}{\beta(\theta+1)}\right). \end{aligned}$$

□

2.2. **Probability and moment generating functions.** A probability distribution's properties are directly related to its probability generating function (pgf) and moment generating function (mgf). From [5], the probability generating function of BTPL(θ, β) is given by:

$$G_Y(t) = \frac{1}{\beta} - \frac{(1-\beta)\theta^2}{\beta(\theta+1)}\frac{(\theta+2-t)}{(\theta+1-t)^2}. \tag{2.4}$$

The mgf can be easily calculated from the pgf by means of the relationship $G_Y(e^t) = E(e^{tY}) = M_Y(t)$.

2.3. **Moments.** Like the probability generating function, the mean and variance of the BTPL(θ, β) distribution with are given by

$$E(Y) = 1 + (1-\beta^{-1})\frac{\theta+2}{\theta(\theta+1)}$$

and

$$V(Y) = (1-\beta^{-1})\frac{\theta^3+4\theta^2+6\theta+2}{\theta^2(\theta+1)^2} + \frac{2\beta^{-1}(\theta+2)}{\theta(\theta+1)} + (1-\beta^{-1})\beta^{-1}\frac{(\theta^2+4\theta+6}{\theta^2(\theta+1)}.$$

3. Maximum likelihood estimators of the parameters

In this part the maximum likelihood estimators of BTPL(θ, β) are considered, where both parameters are unknown. Let $y_1, y_2, \ldots, y_n$ be a random sample of size n from the BTPL distribution (2.2), the log-likelihood function is of the form

$$\begin{aligned}
l(\theta,\beta) &= \sum_{i=1}^{n}\delta_0(y_i)\log\left(\frac{\theta^2+3\theta+1}{\beta(\theta+1)^3}\right) \\
&\quad + \sum_{i=1}^{n}(1-\delta_0(y_i))\left[\log\left(\frac{\theta^2+3\theta+1}{\beta(\theta+1)^{y_i+3}}\right) + \log\left(\frac{\beta(\theta+1)(\theta+y_i+1)}{\theta+y_i+2}\right) - 1\right] \\
&= \sum_{i=1}^{n}\delta_0(y_i)\log\left(\frac{\theta^2+3\theta+1}{\beta(\theta+1)^3}\right) \\
&\quad + \sum_{i=1}^{n}(1-\delta_0(y_i))\left[2\log\theta - \log\beta - (y_i+3)\log(\theta+1)\right] \\
&\quad + \sum_{i=1}^{n}(1-\delta_0(y_i))\log\left[\beta(\theta+1)(\theta+y_i+1) - (\theta+y_i+2)\right].
\end{aligned}$$

We obtain the score by deriving the corresponding log-likelihood function with respect to the unknown parameters as

$$\begin{aligned}\frac{\partial l}{\partial \theta} &= \frac{\theta^2-4\theta}{(\theta+1)(\theta^2+3\theta+1)}\sum_{i=1}^{n}\delta_0(y_i)+\sum_{i=1}^{n}(1+\delta_0(y_i))\left[\frac{2}{\theta}-\frac{y_i+3}{\theta+1}\right]\\ &+ \sum_{i=1}^{n}(1+\delta_0(y_i))\left[\frac{\beta(\theta+y_i+1)+\beta(\theta+1)-1}{\beta(\theta+1)(\theta+y_i+1)-(\theta+y_i+2)}\right]\end{aligned}$$

and

$$\frac{\partial l}{\partial \beta} = -\frac{1}{\beta}\sum_{i=1}^{n}\delta_0(y_i)-\sum_{i=1}^{n}(1+\delta_0(y_i))\left[\frac{1}{\beta}+\frac{(\theta+1)(\theta+y_i+1)}{\beta(\theta+1)(\theta+y_i+1)-(\theta+y_i+2)}\right].$$

The maximum likelihood estimators of θ and β are obtained by solving numerically the non-linear equations $\frac{\partial l}{\partial \theta}=0$ and $\frac{\partial l}{\partial \beta}=0$. We compute these estimators using the maxLik package for the r statistical environment (see [7] for more details).

4. Applications and goodness of fit

Table 1. Accidents to 647 women working on high explosive shells in 5 weeks. Comparaison of POISSON, NB, NGPL,NGPSD and BPLT distributions .

No of ac	Obs freq	Poisson	NB	GPL	NGPL	NGPSD	BTPL
0	447	406.3140	445.859	446.4497	441.5905	442.2160	446.9389
1	132	189.0254	134.9321	133.6980	140.1961	139.3588	131.4969
2	42	43.9692	43.9949	44.4401	44.5095	44.4633	46.1255
3	21	6.8185	14.6880	14.9290	14.1309	14.2553	15.3206
4	3	0.7930	4.9610	4.9992	4.4863	4.5694	4.9103
≥ 5	2	0.0799	2.5649	2.4841	2.0868	2.1372	2.2079
Total	647	647	647	647	647	647	647
MLE		$\widehat{\lambda}=0.465$	$\widehat{r}=0.8659$ $\widehat{p}=0.6505$	$\widehat{\theta}=2.2447$ $\widehat{\beta}=0.7361$	$\widehat{\theta}=2.1499$ $\widehat{\beta}=0.0001$	$\widehat{\theta}=2.6592$ $\widehat{\beta}=0.3019$	$\widehat{\theta}=2.7044$ $\widehat{\beta}=0.4678$
LogLik		-617.1843	-592.2671	-592.1282	-592.4798	-592.3929	-591.9387
χ^2		103.14	3.7692	3.5189	4.5221	4.3158	3.2391
df		5	5	5	5	5	5
Pvalue		0	0.5831	0.6205	0.4769	0.5049	0.6632
AIC		1236.369	1188.534	1188.256	1188.96	1188.786	1187,877

In Table 1, we consider a real data set representing the number of accidents suffered by 647 women working on explosive shells over a 5 week period. [11]

used this data set to fit the Poisson-Lindley distribution (1.1). La Table 1 shows the comparison of observed and expected frequencies for the POISSON, NB, NGPL, NGPSD and BPLT distributions. The log likelihood, χ^2, p-value and AIC values are also shown in this table. It appears that this two-parameter distribution BTPL provides a good fit.

5. Conclusion

In this paper, we have proposed a new counting probability distribution called the beta transformation of the Poisson-Lindley distribution. The beta transformation is another way to add a parameter to a distribution. The incorporation of additional parameters using the beta transformation can greatly improve the frequency approximation of the model. An attempt was made to study the goodness of fit of the BTPL distribution against count data on the number of accidents for machine operators and it was observed that BTPL provides a better fit than the other models namely Poisson, NB, GPL, NGPL and NGPSD.

References

[1] S.A. Aderoju: *A New Generalized Poisson Mixed Distribution and Its Application*, Applied Mathematical Sciences, **14**(5) (2020), 229-234.

[2] D. Bhati, D.V.S. Sastry, P.Z, Qadri Maha: *A New Generalized Poisson-Lindley Distribution: Applications and Properties*, Austrian Journal of Statistics, **44** (2015), 35-51.

[3] N.L. Johnson, A.W. Kemp, S. Kotz: *A Univariate Discrete Distributions*, third ed., Wiley, Hoboken, New Jersey, 2005.

[4] D.J. Daley, J.H. Maindonald:*A unified view of models describing the avoidance of superparasitism*, IMA Journal of Mathematics Applied in Biology and Medicine, **6** (1989), 161-178.

[5] B. Dimitrov, N. Kolev: *Beta transformation. beta type self-decomposition and related characterizations*, Brezilian Journal of Probability and Statistics, **14** (2000), 123-140.

[6] C.G. Louzayadio, R.O. Malouata, M.D. Koukouatikissa: *A weighted poisson distribution for underdispersed count data*, International Journal of Statistics and probability, **10**(4) (2021), 157-157.

[7] A. Henningsen, O. Toomet: *A package for maximum likelihood estimation in r*, Comput Stat, **26**(3) (2011), 443-458.

[8] E. Mamhoudi, H. Zakerzadeh: *Generalized Poisson-Lindley Distribution*, Communications in Statistics-Theory and Methods, **39**(10) (2010), 1785-1798.

[9] B.J.T. Morgan: *Modelling polyspermy*, Biometrics, **38** (1982), 85-98.

[10] D. MIZERE: *Contributions to the Modelling and Statistical Analysis of count data*, (Ph.D. thesis) University of Pau, France, 2006.

[11] M. SANKARAN: *The discrete Poisson-Lindley distribution*, Biometrics, **26** (1970), 145-149.

FACULTY OF ECONOMICS SCIENCES
MARIEN NGOUABI UNIVERSITY
AVENUE OF THE FIRST AFRICAN GAMES, BRAZZAVILLE,
CONGO.
Email address: gelinlouz@gmail.com

HIGHER TEACHERS TRAINING COLLEGE
MARIEN NGOUABI UNIVERSITY
AVENUE OF THE FIRST AFRICAN GAMES, BRAZZAVILLE,
CONGO.
Email address: michel.koukouatikissa@umng.cg

HIGHER TEACHERS TRAINING COLLEGE
MARIEN NGOUABI UNIVERSITY
AVENUE OF THE FIRST AFRICAN GAMES, BRAZZAVILLE,
CONGO.
Email address: onesimero@gmail.com