Austrian Journal of Statistics
January 2023, Volume 52, 39–50.
http://www.ajs.or.at/
doi:10.17713/ajs.v52i1.1344

A New Poisson Generalized Lindley Regression Model

Yupapin Atikankul
Rajamangala University of Technology Phra Nakhon

Abstract

In this paper, a new count distribution is introduced. It is a mixture of the Poisson and generalized Lindley distributions. Statistical properties of the proposed distribution including the factorial moments, probability generating function, moment generating function and moments are studies. Maximum likelihood estimators of unknown parameters are derived. Moreover, an alternative count regression model based on the proposed distribution is presented. Finally, the proposed model is applied for real data and compared with other well-known models.

Keywords: count regression, overdispersion, maximum likelihood estimation, mixed Poisson distribution.

1. Introduction

Count data appear in a wide range of many fields, for example, insurance, medicine and agriculture. The traditional distribution for analyzing count data is the Poisson distribution with the restriction its mean and variance relationship, equidispersion. In regression analysis if a count response variable is equidispersion, the Poisson regression model can be used to model it. Unfortunately, practical count data exhibit other features, the variance larger than mean, overdispersion and the variance smaller than mean, underdispersion. Most of count data present overdispersion; however, the Poisson model cannot be applied for them.

Various techniques are proposed to deal with overdispersed count data, such as weighted distributions and mixtures of distributions. A well-known and widely technique for allowing overdispersed in count data is the mixed Poisson distribution. The most popularity of mixed Poisson distribution is the negative binomial (NB) distribution. It can be derived from a mixture of the Poisson and gamma distributions (Greenwood and Yule 1920). Some mixed Poisson distributions are reviewed in this paper. For example, the discrete Poisson-Lindley distribution (Sankaran 1970), the generalized Poisson-Lindley distribution (Mahmoudi and Zakerzadeh 2010), the Poisson-weighed Lindley distribution (Abd El-Monsef and Sohsah 2014), the two parameter Poisson-Lindley distribution (Shanker and Mishra 2014), the new generalized Poisson-Lindley distribution (Bhati, Sastry, and Qadri 2015), the Poisson-generalised Lindley distribution (Wongrin and Bodhisuwan 2016) and the new Poisson mixed weigthed Lindley distribution (Atikankul, Thongteeraparp, and Bodhisuwan 2020) were introduced to encounter overdispersion phenomena.

Although many mixed Poisson distributions have been proposed to overcome overdispersion in count data, a few distributions have been developed for regression models. The NB regression model and the Poisson-inverse Gaussian regression model (Willmot 1987) are well-known mixed Poisson regression models. Other mixed Poisson regression models, such as the Poisson-Weibull regression model (Cheng, Geedipally, and Lord 2013), the Poisson-weighted exponential regression model (Zamani, Ismail, and Faroughi 2014), the Poisson exponential-inverse Gaussian regression model (Gómez-Déniz and Calderín-Ojeda 2016), the Poisson-transmuted exponential regression model (Bhati, Kumawat, and Gómez-Déniz 2017), the generalized Poisson-Lindley linear model (Wongrin and Bodhisuwan 2017) and the Poisson quasi-Lindley regression model (Altun 2019) were proposed for modeling count data. However, they may not be suitable for some situations and some of them are not explicit forms.

Abouammoh, Alshangiti, and Ragab (2015) introduced a new generalized Lindley (GL) distribution. It is a mixture of gamma distributions. Its probability density function (pdf) with shape parameter α and scale parameter θ is

$$g(x) = \frac{\theta^{\alpha} x^{\alpha-2}}{(\theta+1)\Gamma(\alpha)}(x+\alpha-1)e^{-\theta x},$$

for $x > 0$, $\alpha > 1$ and $\theta > 0$.

The GL distribution contains the Lindley distribution (Lindley 1958) is a special case. It has a closed-form expression for the pdf and it provides a good fit for real life data. Several properties of a mixed Poisson distribution depends on its mixing distribution. In this paper, the PGL distribution is considered as a mixing distribution.

Here, we propose a new count distribution for overdispersed data. The proposed distribution is the mixture of the Poisson and GL distributions. Its probability mass function (pmf) is in explicit form with gamma function. Furthermore, we also present a new count regression model based on the proposed distribution.

This paper is organized as follows. In Section 2, the new count distribution and its shape are proposed. Some properties of the proposed distribution are derived in Section 3. Random variate generation is given in Section 4. Next, parameter estimation is discussed by the method of maximum likelihood. In Section 6, the count regression model based on the proposed distribution is presented. In Section 7, applications of the proposed distribution and the regression model in overdispersed count data are illustrated. Finally, the paper is concluded in Section 8.

2. The proposed distribution

A random variable $Y|\lambda$ is said to have the Poisson distribution if its pmf

$$f(y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \tag{1}$$

for $y = 0, 1, 2, \ldots$.

Suppose λ follows the GL distribution with pdf

$$g(\lambda) = \frac{\theta^{\alpha}\lambda^{\alpha-2}}{(\theta+1)\Gamma(\alpha)}(\lambda+\alpha-1)e^{-\theta\lambda}, \tag{2}$$

for $\lambda > 0$, $\alpha > 1$ and $\theta > 0$.

Then the unconditional random variable Y has the Poisson generalized Lindley (PGL) distribution. Its pmf is given in Theorem 1.

Theorem 1. *If random variable Y follows the PGL distribution, then the pmf of the PGL distribution is*

$$f(y) = \frac{\theta^{\alpha}\Gamma\left(\alpha+y-1\right)\left(\alpha(\theta+2)-\theta+y-2\right)}{\Gamma(\alpha)\Gamma(y+1)(\theta+1)^{\alpha+y+1}},$$

for $y = 0, 1, 2, ..., \alpha > 1$ *and* $\theta > 0$.

Proof. If $Y \sim \text{PGL}(\alpha, \theta)$, then the pmf of Y is given by

$$f(y) = \int_0^\infty f(y|\lambda)g(\lambda)d\lambda. \tag{3}$$

Substituting Equation (1) and Equation (2) into Equation (3), then

$$\begin{aligned} f(y) &= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!}\frac{\theta^\alpha \lambda^{\alpha-2}}{(\theta+1)\Gamma(\alpha)}(\lambda+\alpha-1)e^{-\theta\lambda}d\lambda \\ &= \frac{\theta^\alpha}{(\theta+1)\Gamma(\alpha)\Gamma(y+1)}\int_0^\infty e^{-(\theta+1)\lambda}\lambda^{y+\alpha-2}(\lambda+\alpha-1)d\lambda \\ &= \frac{\theta^\alpha}{(\theta+1)\Gamma(\alpha)\Gamma(y+1)}\left(\frac{\alpha\Gamma(\alpha+y-1)}{(\theta+1)^{\alpha+y-1}} - \frac{\Gamma(\alpha+y-1)}{(\theta+1)^{\alpha+y-1}} + \frac{\Gamma(\alpha+y)}{(\theta+1)^{\alpha+y}}\right) \\ &= \frac{\theta^\alpha \Gamma\left(\alpha+y-1\right)\left(\alpha(\theta+2)-\theta+y-2\right)}{\Gamma(\alpha)\Gamma(y+1)(\theta+1)^{\alpha+y+1}}. \end{aligned}$$

□

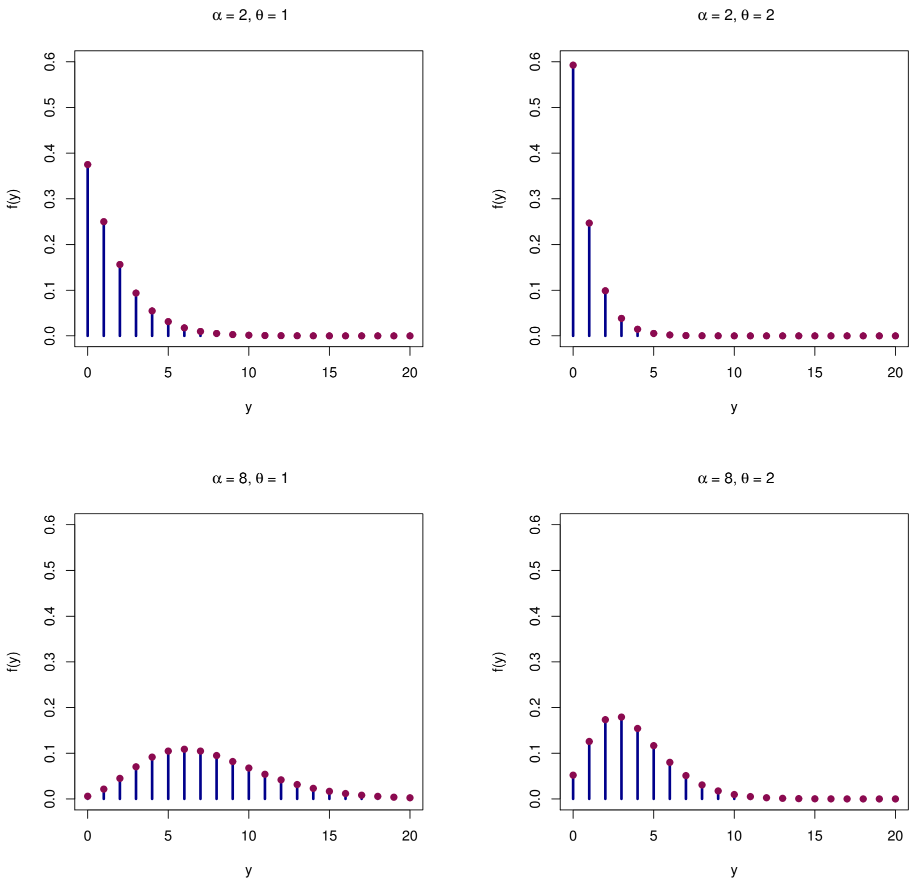

Figure 1: Pmf plots of the PGL distribution with different parameter values

Since

$$\frac{f(y+1)}{f(y)} = \frac{\alpha+y-1}{(y+1)(\theta+1)}\left(1+\frac{1}{\alpha(\theta+2)-\theta+y-2}\right)$$

and

$$\frac{f(y+2)f(y)}{f^2(y+1)} = \frac{(y+1)(\alpha+y)}{(y+2)(\alpha+y-1)}\left(\frac{1+\frac{1}{\alpha(\theta+2)-\theta+y-1}}{1+\frac{1}{\alpha(\theta+2)-\theta+y-2}}\right) < 1,$$

the distribution is log-concave; thus, the PGL distribution has an increasing failure rate (Johnson, Kemp, and Kotz 2005). The PGL distribution is unimodal because the GL distribution is unimodal (see Holgate 1970). Moreover, some pmf plots of the PGL distribution are shown in Figure 1. The figure shows that the proposed distribution is unimodal.

Remark. *If $\alpha = 2$, then the PGL distribution can be reduced to the discrete Poisson-Lindley distribution (Sankaran 1970) with pmf*

$$f(y) = \frac{\theta^2(y+\theta+2)}{(\theta+1)^{y+3}},$$

for $y = 0, 1, 2, ...,$ and $\theta > 0$.

3. Properties

This section presents some theoretical properties of the PGL distribution, including the factorial moments, probability generating function, moment generating function and moments.

Proposition 1. *Let $Y \sim PGL(\alpha, \theta)$, then the kth factorial moment of Y is given by*

$$\mu'_{[k]} = \frac{\Gamma(\alpha+k-1)}{(\theta+1)\Gamma(\alpha)\theta^k}\left[\alpha\theta+\alpha-\theta+k-1\right].$$

Proposition 2. *Let $Y \sim PGL(\alpha, \theta)$, then the probability generating function of Y is defined by*

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

where $|z| < \theta + 1$.

Proposition 3. *Let $Y \sim PGL(\alpha, \theta)$, then the moment generating function of Y is given by*

$$M(t) = \frac{\theta^\alpha\left(\theta-e^t+2\right)}{(\theta+1)(\theta-e^t+1)^\alpha},$$

where $t < log(\theta + 1)$.

Moments are important properties for any distribution and they can be used for parameter estimation. The kth moment about the origin of Y can be obtained by taking the k derivative from the moment generating function with respect to t and setting t to zero. Hence, the first four moments about the origin of Y are

$$\begin{aligned}
\mu &= \frac{(\alpha-1)\theta+\alpha}{\theta(\theta+1)}, \\
\mu'_2 &= \frac{(\alpha-1)\theta(\alpha+\theta)+\alpha(\alpha+\theta+1)}{\theta^2(\theta+1)}, \\
\mu'_3 &= \frac{(\alpha-1)\theta^3+\alpha(3\alpha-2)\theta^2+\alpha(\alpha+1)(\alpha+2)\theta+\alpha(\alpha+1)(\alpha+2)}{\theta^3(\theta+1)}, \\
\mu'_4 &= \frac{(\alpha-1)\theta^4+\alpha(7\alpha-6)\theta^3+\alpha(\alpha+1)(6\alpha+1)\theta^2+\alpha(\alpha+1)(\alpha+2)(\alpha+5)\theta}{\theta^4(\theta+1)} \\
&\quad +\frac{\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{\theta^4(\theta+1)}.
\end{aligned}$$

The second moment about the mean or the variance of Y is given by

$$\sigma^2 = \frac{(\alpha-1)\theta+\alpha}{\theta(\theta+1)} + \frac{(\alpha-1)\theta^2+2\alpha\theta+\alpha}{\theta^2(\theta+1)^2}.$$

The index of dispersion (ID) is the ratio of the variance to the mean. It can be given by

$$ID(Y) = \frac{\sigma^2}{\mu} > 1. \tag{4}$$

Figure 2 shows the mean, variance and ID plots of the proposed distribution. The mean and variance are increasing as α increases but they are decreasing as θ increases. The ID is increasing as θ and α decrease. From the Equation (4) and Figure 2, we can see that the ID is greater than one. Thus, the PGL distribution is overdispersed.

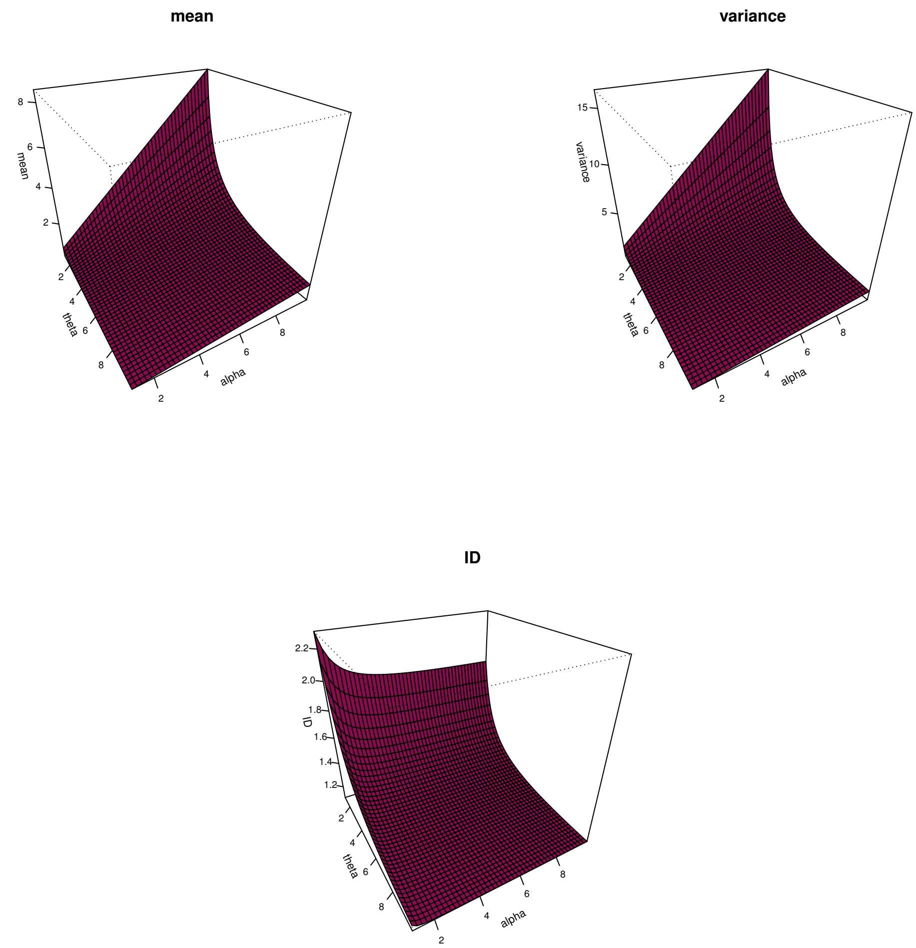

Figure 2: Mean, variance and ID plots of the PGL distribution

4. Random variate generation

In this section, a PGL random variate generation is presented. We can use the acceptance-rejection method to generate random variate of the GL(α, θ) distribution. The steps of PGL random variate generation are as following:

Step1: Generate $u_i, i = 1, \ldots, n$ from $U(0, 1)$.

Step2: If $u_i \leq \frac{1}{\theta+1}$, generate λ from Gamma(α, θ);

otherwise, generate λ from Gamma$(\alpha - 1, \theta)$.

Step3: Generate y_i from Poisson(λ).

5. Parameter estimation

The parameter estimators of the PGL distribution are obtained in this section. In this paper, the maximum likelihood estimation is applied.

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from the PGL distribution with parameters, α and θ, and $y_1, y_2, \ldots, y_n$ be the observed values. The likelihood function of the PGL distribution is given by

$$L(\alpha,\theta) = \prod_{i=1}^{n} \frac{\theta^{\alpha}\Gamma\left(\alpha+y_i-1\right)\left(\alpha(\theta+2)-\theta+y_i-2\right)}{\Gamma(\alpha)\Gamma(y_i+1)(\theta+1)^{\alpha+y_i+1}}.$$

The associated log-likelihood function can be written as

$$\begin{aligned}\log L(\alpha,\theta) &= n\alpha\log(\theta) + \sum_{i=1}^{n}\log\left(\alpha(\theta+2)-\theta+y_i-2\right) + \sum_{i=1}^{n}\log\left(\Gamma(\alpha+y_i-1)\right) \\ &\quad -n\log\left(\Gamma(\alpha)\right) - \sum_{i=1}^{n}\log\left(\Gamma(y_i+1)\right) - \sum_{i=1}^{n}\left(\alpha+y_i+1\right)\log\left(\theta+1\right).\end{aligned}$$

The score function is obtained by taking the first partial derivative of the log-likelilhood function with respect to each parameter, hence

$$\begin{aligned}\frac{\partial \log L(\alpha,\theta)}{\partial\alpha} &= n\log(\theta) + \sum_{i=1}^{n}\frac{\theta+2}{\left(\theta+2\right)\alpha+y_i-\theta-2} + \sum_{i=1}^{n}\left(\Psi(\alpha+y_i-1)\right) \\ &\quad -n\left(\Psi(\alpha)\right) - \sum_{i=1}^{n}\log\left(\theta+1\right), \\ \frac{\partial \log L(\alpha,\theta)}{\partial\theta} &= \frac{n\alpha}{\theta} + \sum_{i=1}^{n}\frac{\alpha-1}{\alpha\left(\theta+2\right)-\theta+y_i-2} - \sum_{i=1}^{n}\frac{y_i+\alpha+1}{\theta+1}.\end{aligned}$$

The maximum likelihood estimates can be obtained by setting the score functions to zero and solving them. Although they are complicate, we can get the maximum likelihood estimates by the numerical methods. In this paper, the optim function for R language (R Core Team 2020) is applied to estimate the parameters of the PGL distribution.

The second partial derivative of the log-likelihood are

$$\begin{aligned}\frac{\partial^2 \log L(\alpha,\theta)}{\partial\alpha^2} &= -\sum_{i=1}^{n}\frac{(\theta+2)^2}{\left((\theta+2)\,\alpha+y_i-\theta-2\right)^2} + \sum_{i=1}^{n}\Psi_1\left(\alpha+y_i-1\right) - n\,\Psi_1\left(\alpha\right), \\ \frac{\partial^2 \log L(\alpha,\theta)}{\partial\alpha\partial\theta} &= \frac{n}{\theta} + \sum_{i=1}^{n}\frac{y_i}{\left((\alpha-1)\,\theta+y_i+2\alpha-2\right)^2} - \sum_{i=1}^{n}\frac{1}{\theta+1}, \\ \frac{\partial^2 \log L(\alpha,\theta)}{\partial\theta^2} &= -\frac{n\alpha}{\theta^2} - \sum_{i=1}^{n}\frac{(\alpha-1)^2}{\left(\alpha\left(\theta+2\right)-\theta+y_i-2\right)^2} + \sum_{i=1}^{n}\frac{y_i+\alpha+1}{\left(\theta+1\right)^2}.\end{aligned}$$

The asymptotic variances-covariances of maximum likelihood estimators are calculated by the elements of the inverse of the Fisher information matrix. Due to the complexity of the Fisher information matrix, it can be instead of the observed information matrix. The observed information matrix of the maximum likelihood estimators of the parameters is

$$J(\alpha,\theta) = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix},$$

where $J_{11} = -\frac{\partial^2 \log L(\alpha,\theta)}{\partial\alpha^2}$, $J_{12} = J_{21} = -\frac{\partial^2 \log L(\alpha,\theta)}{\partial\alpha\partial\theta}$ and $J_{22} = -\frac{\partial^2 \log L(\alpha,\theta)}{\partial\theta^2}$.
The standard error (S.E.) of the maximum likelihood estimates can be obtianed by the square roots of the diagonal elements in the variance-covariance matrix.

6. The proposed regression model

In this section, the PGL regression model for count data is presented. It is developed by the PGL distribution when y_i is count response variable.
Since $\mu = \frac{(\alpha-1)\theta+\alpha}{\theta(\theta+1)}$, we parameterize $\alpha = \frac{\theta(\theta\mu+\mu+1)}{\theta+1}$ to the pmf of the PGL distribution. Thus

$$f(y) = \frac{\theta^{\frac{\theta(\theta\mu+\mu+1)}{\theta+1}}\Gamma\left(\frac{\theta(\theta\mu+\mu+1)}{\theta+1}+y-1\right)\left(\frac{\theta(\theta\mu+\mu+1)}{\theta+1}(\theta+2)-\theta+y-2\right)}{\Gamma(\frac{\theta(\theta\mu+\mu+1)}{\theta+1})\Gamma(y+1)(\theta+1)^{\frac{\theta(\theta\mu+\mu+1)}{\theta+1}+y+1}}.$$

In order to guarantee that the mean of response variable is positive, the log link function is used to link the explanatory variable to the mean of the response variable. It can be expressed as

$$\log(\mu_i) = \mathbf{x}_i^T\boldsymbol{\beta},$$

where $\mathbf{x}_i^T$ is the vector of explanatory variable or covariate, $(x_{1i}, x_{2i}, \ldots, x_{pi})$, and $\boldsymbol{\beta}$ is the vector of unknown regression coefficients.
The log-likelihood function of the PGL regression model is

$$\begin{aligned}\log L(\theta,\boldsymbol{\beta}) \quad = \quad & n\frac{\theta(\theta\mu_i+\mu_i+1)}{\theta+1}\log(\theta) + \sum_{i=1}^{n}\log\left(\frac{\theta(\theta\mu_i+\mu_i+1)}{\theta+1}(\theta+2)-\theta+y_i-2\right) \\ & +\sum_{i=1}^{n}\log\left(\Gamma\left(\frac{\theta(\theta\mu_i+\mu_i+1)}{\theta+1}+y_i-1\right)\right) - n\log\left(\Gamma\left(\frac{\theta(\theta\mu_i+\mu_i+1)}{\theta+1}\right)\right) \\ & -\sum_{i=1}^{n}\log\left(\Gamma(y_i+1)\right) - \sum_{i=1}^{n}\left(\frac{\theta(\theta\mu_i+\mu_i+1)}{\theta+1}+y_i+1\right)\log\left(\theta+1\right).\end{aligned}$$

The numerical methods can be applied to obtain the maximum likelihood estimates. In this paper, the optim function for R language (R Core Team 2020) is employed.

7. Applications

In this section, applications of the PGL distribution and PGL regression model are applied for real data sets. Additionally, they are also compared wih classical models.

7.1. Distributions for count data

Two overdispersed data sets are analyzed to show the usefulness of the PGL distribution. The first data set is the the number of mistakes in copying groups of random digits (Kemp and Kemp 1965). This data set is overdispersed with the ID 1.605, The second data set is Pyrausta nublilalis (Beall 1940) with the ID 1.758. The PGL distibution are compared with the Poisson and NB distributions. The criteria for model selection are the lowest of the Akaike information criterion (AIC) (Akaike 1974), the lowest of the Bayesian information criterion (BIC), the highest of the log-likelihood and the highest of p-value based on the discrete Anderson-Darling (AD) goodness of fit test (Arnold and Emerson 2011). The result are shown in the Table 1 and Table 2.

Observed and expected frequencies of the Poisson, NB and PGL distributions for the number of mistakes in copying groups of random digits and Pyrausta nublilalis are shown in Table 1

Table 1: Observed and expected frequencies of the number of mistakes in copying groups of random digits

Count	Observed frequencies	Expected frequencies		
		Poisson	NB	PGL
0	35	27.4128	33.9511	34.3868
1	11	21.4734	14.4900	13.9063
2	8	8.4104	6.3899	6.3512
3	4	2.1961	2.8481	2.9219
4	2	0.4301	1.2762	1.3359
Estimated parameters		$\hat{\lambda} = 0.7833$	$\hat{r} = 0.9376$	$\hat{\theta} = 1.3878$
(S.E.)		(0.1143)	(0.5089)	(0.6340)
			$\hat{p} = 0.5448$	$\hat{\alpha} = 1.6705$
			(0.1432)	(0.5589)
log-likelihood		−77.5456	−73.3683	−73.2291
AIC		157.0912	150.7366	150.4582
BIC		159.1855	154.9253	154.6469
AD statistic		2.2732	0.1541	0.1139
p-value		0.0494	0.8290	0.8831

Table 2: Observed and expected frequencies of Pyrausta nublilalis

Count	Observed frequencies	Expected frequencies		
		Poisson	NB	PGL
0	33	26.4525	32.6422	32.8219
1	12	19.8394	13.0975	12.7768
2	6	7.4398	5.6731	5.7172
3	3	1.8599	2.5175	2.5864
4	1	0.3487	1.1306	1.1645
5	1	0.0523	0.5113	0.5206
Estimated parameters		$\hat{\lambda} = 0.7500$	$\hat{r} = 0.8628$	$\hat{\theta} = 1.4136$
(S.E.)		(0.1157)	(0.4629)	(0.6315)
			$\hat{p} = 0.5350$	$\hat{\alpha} = 1.6466$
			(0.1434)	(0.5282)
log-likelihood		−71.5824	−66.8893	−66.8307
AIC		145.1647	137.7787	137.6615
BIC		147.1901	141.8294	141.7122
AD statistic		1.7813	0.0179	0.0099
p-value		0.0898	0.9970	0.9992

and Table 2, respectively. We can see that the p-values based on the discrete AD test for the Poisson distribution are less than the 5% significance level for the first data set. Therefore, this data set cannot be fitted by the Poisson distribution. The PGL distribution provides the smallest AIC, the smallest BIC, the largest log-likelihood, and the largest p-value based on the discrete AD test for both data set.

Figure 3 displays observed and the expected frequencies of the PGL distribution. The figure shows that the expected frequencies of the PGL distribution close to the observed frequencies for both data sets. From Table 1, Table 2 and Figrure 3 reveal that the PGL distribution provides good fit to overdispersed data.

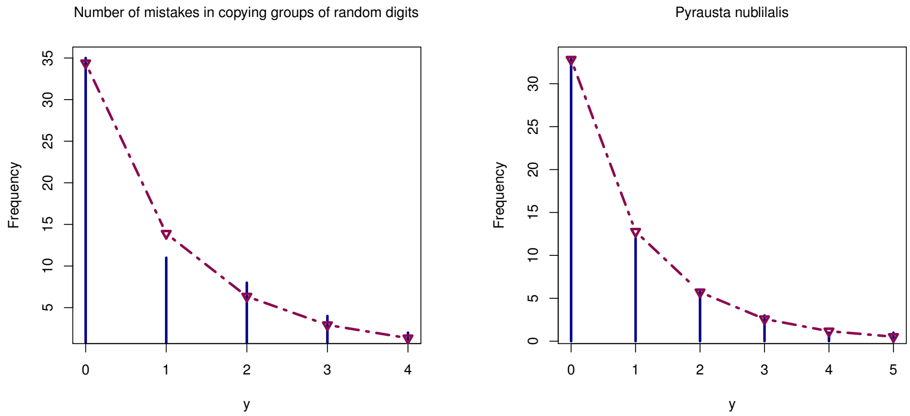

Figure 3: Plots of the observed and expected frequencies for the PGL distribution

7.2. Count regression models

The PGL regression model with the log link function is applied for the azpro data set (Hilbe 2011). The data frame consist of 3589 cardiovascular patients in 1991 in Arizona, USA. The variables used are as follows:

- *Los*, length of stay in hospital
- *Procedure*, standard cardiovascular procedures (1= CABG; 0 = PTCA)
- *Sex*, (1 Male; 0 = Female)
- *Admit*, (1 =Urgent/Emerg; 0 = Elective (type of admission)
- *Age*, (1 = Age > 75; 0 = < = 75).

The predictor variable is *Los* and four covariates are *Procedure*, *Sex*, *Admit* and *Age*. The PGL regression model is compared with the Poisson and NB regression models. Table 3 shows the parameter estimates, S.E. and the p-values based on t-test for the PGL and the other two regression models. The log-likelihood, AIC, and BIC criteria are used to select the best model.

Table 3: Parameter estimates, S.E. and the p-values for the azpro data set

Variable	Poisson		NB		PGL	
	Estimate (S.E)	p-value	Estimate (S.E.)	p-value	Estimate (S.E)	p-value
Intercept	1.4560 (0.0159)	<0.0001	1.4177 (0.0236)	<0.0001	1.4822 (0.0234)	<0.0001
Procedure	0.9603 (0.0122)	<0.0001	0.9811 (0.0183)	<0.0001	0.9423 (0.0181)	<0.0001
Sex	-0.1239 (0.0118)	<0.0001	-0.1264 (0.0191)	<0.0001	-0.1141 (0.0177)	<0.0001
Admit	0.3266 (0.0121)	<0.0001	0.3707 (0.0190)	<0.0001	0.3000 (0.0179)	<0.0001
Age	0.1222 (0.0125)	<0.0001	0.1201 (0.0202)	<0.0001	0.1154 (0.0187)	<0.0001
Dispersion	-	-	6.246 (0.254)	<0.0001	0.7663 (0.0288)	<0.0001

Table 4: The measures for model selection

Criterion	Poisson	NB	PGL
log-likelihood	−11189.9	−9973.543	−9970.605
AIC	22389.8	19959.09	19953.210
BIC	22420.7	19996.2	19990.32

Table 4 displays log-likelihood, AIC and BIC for the Poisson, NB and PGL regression models. The PGL regression model provides the highest log-likelihood, the lowest AIC, and the lowest BIC. Hence, we conclude that the PGL regression model is the best model among the competing regression models.

8. Conclusion

In this work, we have proposed the Poisson generalize Lindley distribution, called PGL distribution, for overdispersed data. Some statistical properties have been studied, such as the factorial moments, probability generating function, moment generating function and moments. The parameter estimation has been approached by the method of maximum likelihood. The PGL distribution has been compared with classical distributions, namely the Poisson and NB distributions. Some real data sets have been used to illustrate the performance of the proposed distribution. The results show that the PGL distribution provides great flexibility in modeling real data. Furthermore, The PGL regression model based on the proposed distribution has been developed. The proposed regression model is more suitable than some well-known regression models. Therefore, it can be applied for the count regression model.

Acknowledgement

The author is grateful to the editor and the reviewers for their review and valuable suggestions. Furthermore, the author would like to thank Rajamangala University of Technology Phra Nakhon.

References

Abd El-Monsef MME, Sohsah NM (2014). "Poisson-Weighted Lindley Distribution." *Jökull Journal*, **64**(5), 192–202.

Abouammoh AM, Alshangiti AM, Ragab IE (2015). "A New Generalized Lindley Distribution." *Journal of Statistical Computation and Simulation*, **85**(18), 3662–3678.

Akaike H (1974). "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

Altun E (2019). "A New Model for Over-Dispersed Count Data: Poisson Quasi-Lindley Regression Model." *Mathematical Sciences*, **13**(3), 241–247.

Arnold TB, Emerson JW (2011). "Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions." *The R Journal*, **3**(2), 34–39.

Atikankul Y, Thongteeraparp A, Bodhisuwan W (2020). "The New Poisson Mixed Weighted Lindley Distribution with Applications to Insurance Claims Data." *Songklanakarin Journal of Science and Technology*, **42**(1), 152–162.

Beall G (1940). "The Fit and Significance of Contagious Distributions when Applied to Observations on Larval Insects." *Ecology*, **21**(4), 460–474.

Bhati D, Kumawat P, Gómez-Déniz E (2017). "A New Count Model Generated from Mixed Poisson Transmuted Exponential Family with an Application to Health Care Data." *Communications in Statistics-Theory and Methods*, **46**(22), 11060–11076.

Bhati D, Sastry DVS, Qadri PZM (2015). "A New Generalized Poisson-Lindley Distribution: Applications and Properties." *Austrian Journal of Statistics*, **44**(4), 35–51.

Cheng L, Geedipally SR, Lord D (2013). "The Poisson–Weibull Generalized Linear Model for Analyzing Motor Vehicle Crash Data." *Safety Science*, **54**, 38–42.

Gómez-Déniz E, Calderín-Ojeda E (2016). "The Mixture Poisson Exponential–Inverse Gaussian Regression Model: An application in Health Services." *Metodoloski Zvezki*, **13**(2), 71–85.

Greenwood M, Yule GU (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.

Hilbe JM (2011). *Negative Binomial Regression*. Cambridge University Press. ISBN 978-0511973420.

Holgate P (1970). "The Modality of Some Compound Poisson Distributions." *Biometrika*, **57**(3), 666–667.

Johnson NL, Kemp AW, Kotz S (2005). *Univariate Discrete Distributions*. John Wiley & Sons. ISBN 978-0471272465.

Kemp CD, Kemp AW (1965). "Some Properties of the Hermite Distribution." *Biometrika*, **52**(3-4), 381–394.

Lindley DV (1958). "Fiducial Distributions and Bayes' Theorem." *Journal of the Royal Statistical Society: Series B (Methodological)*, **20**(1), 102–107.

Mahmoudi E, Zakerzadeh H (2010). "Generalized Poisson-Lindley Distribution." *Communications in Statistics—Theory and Methods*, **39**(10), 1785–1798.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sankaran M (1970). "The Discrete Poisson-Lindley Distribution." *Biometrics*, **26**(1), 145–149.

Shanker R, Mishra A (2014). "A Two-Parameter Poisson-Lindley Distribution." *International Journal of Statistics and Systems*, **9**(1), 79–85.

Willmot GE (1987). "The Poisson-Inverse Gaussian Distribution as an Alternative to the Negative Binomial." *Scandinavian Actuarial Journal*, **1987**(3-4), 113–127.

Wongrin W, Bodhisuwan W (2016). "The Poisson-Generalised Lindley Distribution and its Applications." *Songklanakarin Journal of Science and Technology*, **38**(6), 645–656.

Wongrin W, Bodhisuwan W (2017). "Generalized Poisson–Lindley Linear Model for Count Data." *Journal of Applied Statistics*, **44**(15), 2659–2671.

Zamani H, Ismail N, Faroughi P (2014). "Poisson-Weighted Exponential Univariate Version and Regression Model with Applications." *Journal of Mathematics and Statistics*, **10**(2), 148–154.

Affiliation:

Yupapin Atikankul
Department of Mathematics and Statistics
Rajamangala University of Technology Phra Nakhon
Bang Sue, Bangkok, Thailand
E-mail: yupapin.a@rmutp.ac.th

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 52 *Submitted:* 2021-04-19
January 2023 *Accepted:* 2021-08-02