



A new model for over-dispersed count data: Poisson quasi-Lindley regression model

Emrah Altun¹

Received: 2 April 2019 / Accepted: 3 July 2019 / Published online: 10 July 2019
© The Author(s) 2019

Abstract

In this paper, a new regression model for count response variable is proposed via re-parametrization of Poisson quasi-Lindley distribution. The maximum likelihood and method of moment estimations are considered to estimate the unknown parameters of re-parametrized Poisson quasi-Lindley distribution. The simulation study is conducted to evaluate the efficiency of estimation methods. The real data set is analyzed to demonstrate the usefulness of proposed model against the well-known regression models for count data modeling such as Poisson and negative-binomial regression models. Empirical results show that when the response variable is over-dispersed, the proposed model provides better results than other competitive models.

Keywords Count data · Poisson regression · Negative-binomial regression · Maximum Likelihood · Method of moments · Over-dispersion

Mathematics Subject Classification 62E15 · 62J05

Introduction

The interest on count data modeling has been greatly increased in the last decade. The widely used distribution for modeling the count data sets is Poisson distribution. The well-known property of Poisson distribution is that its mean and variance are equal. Therefore, Poisson distribution does not work in the case of over-dispersion or under-dispersion. Poisson distribution is widely used in many research fields such as actuarial, environmental, actuarial and economics sciences in spite of its weakness. The reason for that comes from its simple form and easy implementation and software support. To remove the drawback of Poisson distribution, researchers have shown great interest to introduce mixed-Poisson distributions for modeling the over-dispersed or under-dispersed count data sets such as Bhati et al. [1], Imoto et al. [7], Mahmoudi and Zakerzadeh [9], Gencturk and Yigiter [5], Wongrin and Bodhisuwan [15], Déniz [3], Cheng et al. [2], Lord and Geedipally [8], Zamani et al. [16],

Sáez-Castillo and Conde-Sánchez [12], Rodríguez-Avi et al. [10], Shmueli et al. [11], Shoukri et al. [13].

As mentioned above, Poisson distribution is insufficient to model the over-dispersed count data sets. The main motivation of this study is to introduce an alternative regression model for modeling the over-dispersed count data sets. Therefore, a re-parametrization of Poisson quasi-Lindley distribution, proposed by Grine and Zeghdoudi [6], is introduced and its statistical properties are studied comprehensively such as mean, variance and estimation problem of the model parameters. The maximum likelihood (ML) and method of moments (MM) estimation methods are considered to estimate the unknown parameters of the re-parametrized PQL distribution. The efficiencies of the estimation methods are compared with extensive simulation study. Using the re-parametrized Poisson quasi-Lindley distribution, a new regression model for over-dispersed count data sets is introduced. To demonstrate the effectiveness of proposed regression model, a real data set on days of absence of the high school students are analyzed with Poisson, negative-binomial and PQL regression models.

The rest of the paper is organized as follows: In “Re-parametrization of Poisson quasi-Lindley distribution” section, the statistical properties of the re-parametrized Poisson quasi-Lindley distribution are obtained. In “Estimation” section, ML

✉ Emrah Altun
emrahaltun123@gmail.com

¹ Department of Mathematics, Bartın University,
74100 Bartın, Turkey

and (MM) estimation methods are considered to estimate the unknown model parameters. In “Simulation” section, finite sample performance of estimation methods is compared via a Monte Carlo simulation study. In “Poisson quasi-Lindley regression model” section, a new regression model is introduced. In “Empirical study” section, a real data set is analyzed to demonstrate the usefulness of proposed model against the Poisson and negative-binomial regression models. “Conclusion” section contains the concluding remarks.

Re-parametrization of Poisson quasi-Lindley distribution

Let the random variable X follows a Poisson distribution. The probability mass function (pmf) is

$$P(x; \lambda) = \frac{\exp(-\lambda)\lambda^x}{x!}, \quad x = 0, 1, 2, \dots,$$

where $\lambda > 0$. The mean and variance of Poisson distribution are $E(X) = \lambda$ and $\text{Var}(X) = \lambda$, respectively. So, the dispersion index, shortly DI , for Poisson distribution is $DI = \text{Var}(X)/E(X) = \lambda/\lambda = 1$. As seen from the dispersion index of Poisson distribution, the over-dispersed or under-dispersed data sets cannot be modeled by Poisson distribution. Note that when the variance is greater than mean, the over-dispersion occurs; otherwise, it is called as under-dispersion. Grine and Zeghdoudi [6] introduced a new mixed-Poisson distribution, called Poisson quasi-Lindley (PQL), by compounding Poisson distribution with quasi-Lindley distribution, introduced by Shanker and Mishra [14]. The pmf of PQL distribution is given by

$$P(Y = y) = \frac{\theta}{\alpha + 1} \left[\frac{\alpha(\theta + 1) + \theta(y + 1)}{(\theta + 1)^{y+2}} \right], \quad y = 0, 1, 2, \dots, \tag{1}$$

where $\theta > 0$ and $\alpha > -1$. Hereafter, the random variable Y will be denoted as $PQL(\theta, \alpha)$. The corresponding cumulative distribution function (cdf) to 1 is

$$F(y) = P(Y \leq y) = 1 - \frac{\alpha + 2\theta + \alpha\theta + \theta y + 1}{(1 + \alpha)(1 + \theta)^{y+2}}. \tag{2}$$

The mean and variance of PQL distribution are given by, respectively,

$$E(Y) = \frac{2 + \alpha}{(1 + \alpha)\theta} \tag{3}$$

$$\text{Var}(Y) = \frac{2 + 4\alpha + \alpha^2 + \theta(\alpha + 2)(\alpha + 1)}{(\alpha + 1)^2\theta^2} \tag{4}$$

Here, the re-parametrization of PQL distribution is considered. The motivation of re-parametrization for PQL

distribution comes from the generalized linear model approach.

Proposition 1 *Let $\theta = (2 + \alpha)/[(1 + \alpha)\mu]$, then the pdf of PQL distribution is*

$$P(Y = y) = \frac{2 + \alpha}{(\alpha + 1)^2\mu} \left\{ \frac{\alpha + (2 + \alpha)(\mu + \alpha\mu)^{-1}(y + \alpha + 1)}{[1 + (2 + \alpha)(\mu + \alpha\mu)^{-1}]^{y+2}} \right\}, \tag{5}$$

$y = 0, 1, 2, \dots,$

where $\alpha > 0$ and $\mu > 0$. The mean and variance of 5 are given by, respectively,

$$E(Y) = \mu, \tag{6}$$

$$\text{Var}(Y) = \underbrace{\mu}_I + \underbrace{\frac{\mu^2}{(2 + \alpha)^2}(2 + 4\alpha + \alpha^2)}_{II}.$$

Note that the parameter α should be greater than zero to ensure the positive variance. The other statistical properties of PQL distribution, such as probability and moment generating functions, mode and its cdf, under the above re-parametrization can be obtained following the results in Grine and Zeghdoudi [6]. As seen from 6, since the second part of variance equation for PQL distribution is greater than zero for all values of the parameters α and μ , the variance of PQL distribution is always greater than its mean. Therefore, PQL distribution can be a good choice for modeling the over-dispersed data sets.

Figure 1 displays the dispersion index and possible shapes of PQL distribution. When the parameters α and μ increase, the dispersion of PX distribution increases. Note that the effect of the parameter μ on dispersion is higher than that of parameter α . As seen from left side of Fig. 1, PQL distribution can be a good choice for modeling extremely right-skewed data sets.

Generating random variables from Poisson-xgamma distribution

Here, a general algorithm and corresponding code written in R software are given to generate random variables from PQL distribution. The below code can be used for all discrete distributions such as Poisson, Poisson–Lindley, negative-binomial.

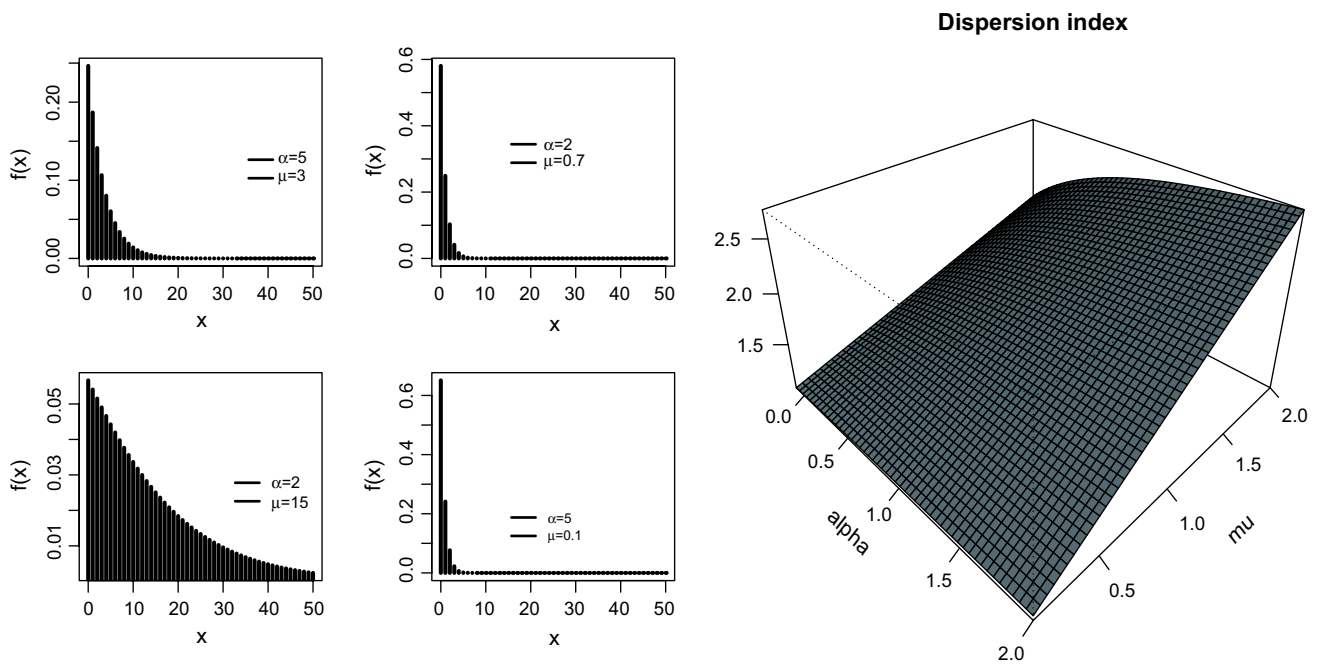


Fig. 1 The dispersion index (right) and the pmf shapes of PQL distribution (left) for some values of α and μ

```

ppois=function(x, alpha ,mu){
theta=(2+alpha)/((1+alpha)*mu)
f=1-(alpha+2*theta+alpha*theta+theta*x+1)/((1+alpha)*(1+theta)^(x+2))
return(f)
}
r.pois <- function(n, alpha,mu){
U <- runif(n)
X <- rep(0,n)
for(i in 1:n)
{if(U[i] < ppois(0,alpha ,mu))
{X[i] <- 0
} else{
B = FALSE
I = 0
while(B == FALSE)
{int <- c( ppois(I, alpha,mu), ppois(I+1,alpha,mu) )
if( (U[i] > int[1]) & (U[i] < int[2]) ){
X[i] <- I+1
B = TRUE
} else{
I=I+1}}}}
return(X)}

```

Estimation

In this section, ML and MM estimation methods are considered to estimate the unknown parameters of PQL distribution.

Maximum likelihood estimation

Let X_1, X_2, \dots, X_n be independent and identically distributed PQL random variables. The log-likelihood function is

$$\begin{aligned} \ell(\alpha, \mu) &= n \ln \left(\frac{2 + \alpha}{(\alpha + 1)^2 \mu} \right) \\ &+ \sum_{i=1}^n \ln [\alpha + (2 + \alpha)(\mu + \alpha \mu)^{-1} (y_i + \alpha + 1)] \\ &- \ln ([1 + (2 + \alpha)(\mu + \alpha \mu)^{-1}]) \sum_{i=1}^n (y_i + 2). \end{aligned} \tag{7}$$

Taking partial derivatives of (7) with respect to α and μ , we have

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= n(\alpha + 2)^{-1}(\alpha + 1)^2 \mu \left(\frac{1}{(\alpha + 1)^2 \mu} - \frac{2(\alpha + 2)}{(\alpha + 1)^3 \mu} \right) \\ &- \left(\frac{\alpha + 2}{\alpha \mu + \mu} + 1 \right)^{-1} \left(\frac{1}{\alpha \mu + \mu} - \frac{(\alpha + 2)\mu}{(\alpha \mu + \mu)^2} \right) \\ &\sum_{i=1}^n (y_i + 2) \\ &+ \sum_{i=1}^n \left[\left(\frac{\alpha + 2}{\alpha \mu + \mu} + \frac{(\alpha + 2)\mu(y_i + \alpha + 1)}{(\alpha \mu + \mu)^2} \right. \right. \\ &\left. \left. + \frac{(y_i + \alpha + 1)}{\alpha \mu + \mu} + 1 \right) \right. \\ &\left. \left(\alpha + \frac{(\alpha + 2)(y_i + \alpha + 1)}{\alpha \mu + \mu} \right) \right] \end{aligned} \tag{8}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \mu} &= \sum_{i=1}^n \frac{(\alpha + 1)(\alpha + 2)(y_i + 2)}{(\alpha \mu + \mu)^2 \left(\frac{\alpha + 2}{\alpha \mu + \mu} + 1 \right)} \\ &- \sum_{i=1}^n \frac{(\alpha + 1)(\alpha + 2)(y_i + \alpha + 1)}{(\alpha \mu + \mu)^2} \\ &\left(\frac{\alpha + 2(y_i + \alpha + 1)}{\alpha \mu + \mu} + \alpha \right) - \frac{n}{\mu} \end{aligned} \tag{9}$$

The ML estimates of (α, λ) can be obtained by means of simultaneous solutions of 8 and 9. It is not possible to obtain explicit forms of ML estimates of PQL distribution since the likelihood equations contain nonlinear functions. For this reason, nonlinear minimization tools are needed to solve these equations. The nonlinear minimization (**nlm**) function of R software is used for this purpose. The corresponding interval estimations of the parameters are obtained by means of observed information matrix which is given by

$$I_F(\boldsymbol{\tau}) = - \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\mu} \\ I_{\mu\alpha} & I_{\mu\mu} \end{pmatrix}.$$

The elements of observed information matrix are upon request from the authors. It is well known that under the

regularity conditions that are fulfilled for the parameters, the asymptotic joint distribution of $(\hat{\alpha}, \hat{\mu})$, as $n \rightarrow \infty$ is a bi-variate normal distribution with mean (α, μ) and variance–covariance $I_F^{-1}(\boldsymbol{\tau})$. Using the asymptotic normality, the asymptotic $100(1 - p)\%$ confidence intervals for the parameters α and μ , respectively, are given by

$$\hat{\alpha} \pm z_{p/2} \sqrt{\text{Var}(\hat{\alpha})}, \quad \hat{\mu} \pm z_{p/2} \sqrt{\text{Var}(\hat{\mu})},$$

where $z_{p/2}$ is the upper $p / 2$ quantile of the standard normal distribution.

Method of moments

The MM estimators of the parameters α and μ can be obtained by equating the mean and variance of PQL distribution to sample mean and variance, given as follows

$$\bar{y} = \mu, \tag{10}$$

$$s^2 = \mu + \frac{\mu^2}{(2 + \alpha)^2} (2 + 4\alpha + \alpha^2), \tag{11}$$

where \bar{y} and s^2 are the sample mean and variance, respectively. For simultaneous solution of (10) and (11), we have

$$\hat{\mu}_{MM} = \bar{y}, \tag{12}$$

$$\hat{\alpha}_{MM} = \sqrt{\frac{2\bar{y}^2}{\bar{y}^2 + \bar{y} - s^2}} - 2 \tag{13}$$

Theorem 1 For fixed values of μ , MM estimator $\hat{\alpha}_{MM}$ of α is consistent and asymptotically normal distributed:

$$\sqrt{n}(\hat{X} - \mu) \rightarrow N(0, v^2(\theta)), \tag{14}$$

where

$$v^2(\theta) = \frac{2 + 4\alpha + \alpha^2 + \theta(\alpha + 2)(\alpha + 1)}{(\alpha + 1)^2 \theta^2}. \tag{15}$$

The detailed information about asymptotic properties of MM estimators can be found in Farbod and Arzideh [4].

Simulation

In this section, Monte Carlo simulation study is conducted to evaluate the finite sample performance of ML and MM estimates of PQL distribution. The following simulation procedure is used.

1. Set the sample size n and the vector of parameters $\theta = (\alpha, \mu)^T$;
2. Generate random observations from the PQL(α, μ) distribution, using the algorithm given in “Generating random variables from Poisson-xgamma distribution” section, with size n ;
3. Use the generated random observations in Step 2, and estimate θ by means of ML and MM estimation methods;
4. Repeat N times the steps 2 and 3;
5. Use $\hat{\theta}$ and θ and calculate the biases, mean relative estimates (MREs) and mean square errors (MSEs) from the following equations:

$$\text{Bias} = \sum_{j=1}^N \frac{\hat{\theta}_{ij} - \theta_i}{N}, \quad \text{MRE} = \sum_{j=1}^N \frac{\hat{\theta}_{ij} / \theta_i}{N},$$

$$\text{MSE} = \sum_{j=1}^N \frac{(\hat{\theta}_{ij} - \theta_i)^2}{N}, \quad i = 1, 2.$$

Figure 2 displays results of simulation study performed under the above procedure. The following parameters are considered: $\theta = (\alpha = 0.5, \mu = 1.5)^T$, $N = 10,000$ and

$n = 40, 45, 50, \dots, 500$. When n is sufficiently large, MREs should be closer to one and MSEs and biases should be closer to zero. As seen from Fig. 2, when the sample size, n , increases, the MSEs and biases are closer to zero and MREs approach to one for both estimation methods. The MM and ML estimation methods yield similar results for the parameter μ in view of estimated MSE, bias and MRE. However, ML estimation method provides more satisfactory results for the parameter α , especially for small sample sizes. Therefore, we suggest to use ML estimation method when the sample size is small.

Poisson quasi-Lindley regression model

The Poisson and negative-binomial are the two commonly used regression models for count data modeling. When the response variable is not equi-dispersed, the negative-binomial regression model is preferable. Here, an alternative regression model is introduced for over-dispersed response variable.

Let random variable Y follow a PQL distribution, given in (5). The mean of Y is $E(Y|\alpha, \mu) = \mu$. Therefore, the covariates can be linked to the mean of response variable, y , by means of the log-link function, given by

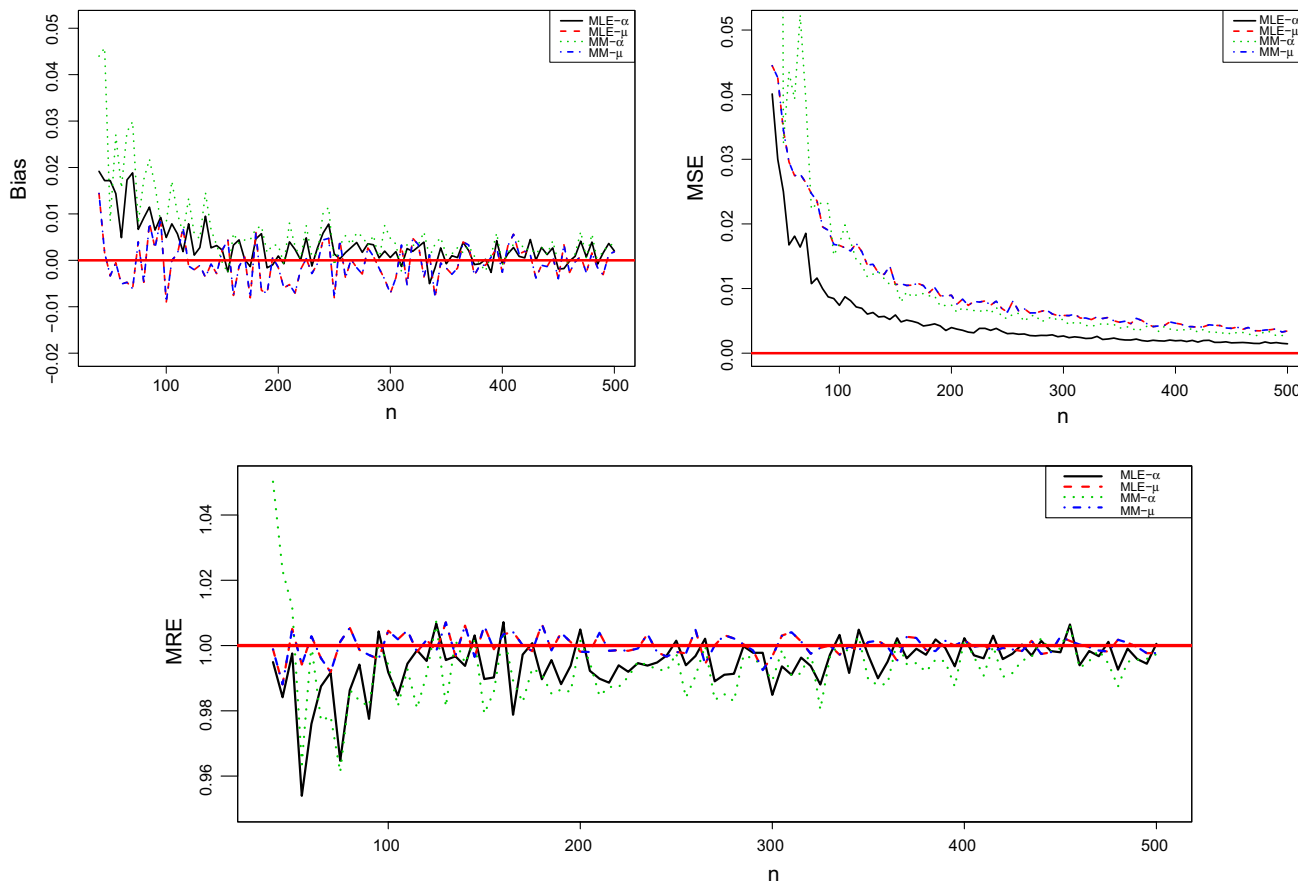


Fig. 2 Estimated biases, MSEs and MREs of the parameters of PQL distribution based on the ML and MM estimation methods

$$\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}), i = 1, \dots, n, \tag{16}$$

where $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{ik})$ is the vector of covariates and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^T$ is the unknown vector of regression coefficients. Inserting (16) in (5), the log-likelihood function can be obtained as follows

$$\begin{aligned} \ell(\boldsymbol{\tau}) = & n \ln(2 + \alpha) - \sum_{i=1}^n \ln [(\alpha + 1)^2 \exp(\mathbf{x}_i^T \boldsymbol{\beta})] \\ & + \sum_{i=1}^n \ln [\alpha + (2 + \alpha) [\exp(\mathbf{x}_i^T \boldsymbol{\beta}) \\ & + \alpha \exp(\mathbf{x}_i^T \boldsymbol{\beta})]^{-1} (y_i + \alpha + 1)] \\ & - \sum_{i=1}^n (y_i + 2) \ln [1 + (2 + \alpha) [\exp(\mathbf{x}_i^T \boldsymbol{\beta}) \\ & + \alpha \exp(\mathbf{x}_i^T \boldsymbol{\beta})]^{-1}], \end{aligned} \tag{17}$$

where $\boldsymbol{\tau} = (\alpha, \boldsymbol{\beta})^T$. The unknown parameters, α and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^T$, are obtained by maximizing (16) with the `nlm` function of R software. Under standard regularity conditions, the asymptotic distribution of $(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})$ is multivariate normal $N_{k+2}(0, J(\boldsymbol{\tau})^{-1})$, where $J(\boldsymbol{\tau})$ is the expected information matrix. The asymptotic covariance matrix $J(\boldsymbol{\tau})^{-1}$ of $\hat{\boldsymbol{\tau}}$ can be approximated by the inverse of the $(k + 2) \times (k + 2)$ observed information matrix $I(\boldsymbol{\tau})$, whose elements are evaluated numerically via most statistical packages. The approximate multivariate normal distribution $N_{k+2}(0, I(\boldsymbol{\tau})^{-1})$ for $\hat{\boldsymbol{\tau}}$ can be used to construct asymptotic confidence intervals for the vector of parameters $\boldsymbol{\tau}$.

Empirical study

In this section, modeling ability of PQL regression model is compared with Poisson and NB regression models via an application on real data set. The data contain number of absence (daily), gender and type of instructional program of the 314 high school students from two urban high schools. The data set can be obtained from https://stats.idre.ucla.edu/stat/stata/dae/nb_data.dta. The response variable, number of absence y_i , is modeled with gender (female = 1, male = 0) (x_1) and type of instructional program (general = 1, academic = 2, vocational = 3). The vocational program is used as a baseline category for type of instructional program variable. The general and academic instructional programs are coded as x_2 and x_3 , respectively. To decide the best model, the estimated negative log-likelihood value, Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values are used. The

lowest values of these statistics show the best-fitted model for the used data set. The following regression model is fitted.

$$\mu_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}) \tag{18}$$

Figure 3 displays the distribution of days of absence. The mean of response variable is 5.955, and variance is 49.518 which is an evidence for over-dispersion.

Table 1 lists the estimated parameters of the models and corresponding SEs, estimated negative $-\ell$, AIC and BIC values. Since PQL regression model has the lowest values of these statistics, we conclude that PQL regression model provides better fits than Poisson and NB regression models, especially for over-dispersed data set.

The obtained observed information matrix of PQL regression model, $I(\boldsymbol{\tau})$, is

$$\begin{bmatrix} 269.303 & 398.549 & 37.894 & 151.312 & 0.058 \\ 398.549 & 656.873 & 53.206 & 224.836 & 0.117 \\ 37.894 & 53.206 & 37.895 & 0.001 & -0.084 \\ 151.312 & 224.836 & 0.001 & 151.305 & 0.043 \\ 0.058 & 0.117 & -0.084 & 0.043 & 0.382 \end{bmatrix}$$

The diagonal elements of the inverse of $I(\boldsymbol{\tau})$ give the variances of estimated parameters. The inverse of $I(\boldsymbol{\tau})$ is

$$\begin{bmatrix} 0.046 & -0.022 & -0.015 & -0.013 & -0.021 \\ -0.022 & 0.015 & 0.001 & 0.001 & -0.008 \\ -0.015 & 0.001 & 0.039 & 0.012 & 0.091 \\ -0.013 & 0.001 & 0.013 & 0.019 & 0.025 \\ -0.021 & -0.008 & 0.091 & 0.025 & 26.407 \end{bmatrix}$$

The asymptotic confidence intervals of regression parameters are $0.049 < \beta_1 < -0.430$, $1.349 < \beta_2 < 0.961$ and $1.215 < \beta_3 < 0.673$, respectively. As seen from estimated regression coefficients of PQL regression model, we

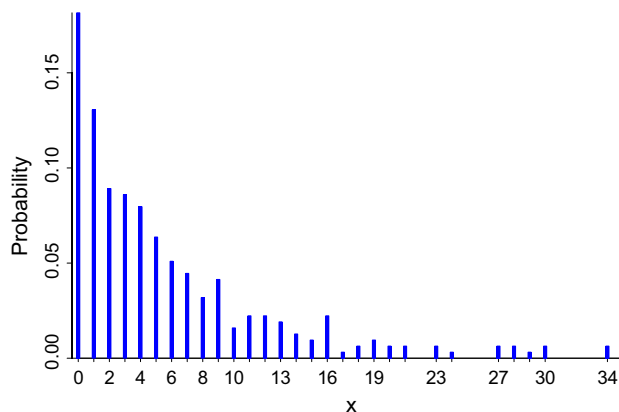


Fig. 3 The distribution of days of absence of students

Table 1 The estimated parameters of models and goodness-of-fit statistics

Covariates	Poisson			NB			PQL		
	Estimate (SE)	SE	<i>p</i> value	Estimate	SE	<i>p</i> value	Estimate	SE	<i>p</i> value
Intercept	1.323	0.089	< 0.001	1.271	0.214	< 0.001	1.273	0.215	< 0.001
Gender	− 0.234	0.047	< 0.001	− 0.193	0.123	0.118	− 0.191	0.122	0.119
General	1.374	0.076	< 0.001	1.362	0.199	< 0.001	1.348	0.198	< 0.001
Academic	0.957	0.066	< 0.001	0.949	0.140	< 0.001	0.945	0.138	< 0.001
Dispersion	–	–	–	1.017	0.104	–	4.999	5.139	–
− <i>ℓ</i>	1343.250			869.423			867.436		
AIC	2694.500			1748.846			1744.872		
BIC	2709.498			1767.593			1763.619		

conclude that the gender has no statistically significant effect on the days of absence for students. However, the days of absence for general and academic instructional program students are 1.348 and 0.945 times higher than the vocational instructional program students.

Conclusion

A re-parametrization of the Poisson quasi-Lindley distribution is introduced and studied comprehensively. The parameter estimation problem of the Poisson quasi-Lindley distribution is discussed via extensive simulation study. A new regression model for count data is proposed and compared with Poisson and negative-binomial regression models based on the real data set. We conclude that Poisson quasi-Lindley regression model exhibits better fitting performance than Poisson and negative-binomial regression models when the response variable is over-dispersed. We hope that the results given in this study will be very helpful for researchers studying in this field.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bhati, D., Kumawat, P., Gómez-Déniz, E.: A new count model generated from mixed Poisson transmuted exponential family with an application to health care data. *Commun. Stat. Theory Methods* **46**(22), 11060–11076 (2017)
- Cheng, L., Geedipally, S.R., Lord, D.: The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data. *Saf. Sci.* **54**, 38–42 (2013)
- Déniz, E.G.: A new discrete distribution: properties and applications in medical care. *J. Appl. Stat.* **40**(12), 2760–2770 (2013)

- Farbod, D., Arzideh, K.: Asymptotic properties of moment estimators for distributions generated by Levy’s law. *Int. J. Appl. Math. Stat* **20**(11), 55–59 (2010)
- Gençturk, Y., Yigiter, A.: Modelling claim number using a new mixture model: negative binomial gamma distribution. *J. Stat. Comput. Simul.* **86**(10), 1829–1839 (2016)
- Grine, R., Zeghdoudi, H.: On Poisson quasi-Lindley distribution and its applications. *J. Mod. Appl. Stat. Methods* **16**(2), 21 (2017)
- Imoto, T., Ng, C.M., Ong, S.H., Chakraborty, S.: A modified Conway–Maxwell–Poisson type binomial distribution and its applications. *Commun. Stat. Theory Methods* **46**(24), 12210–12225 (2017)
- Lord, D., Geedipally, S.R.: The negative binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. *Accid. Anal. Prev.* **43**(5), 1738–1742 (2011)
- Mahmoudi, E., Zakerzadeh, H.: Generalized Poisson–Lindley distribution. *Commun. Stat. Theory Methods* **39**(10), 1785–1798 (2010)
- Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A.J., Olmo-Jiménez, M.J., Martínez-Rodríguez, A.M.: A generalized Waring regression model for count data. *Comput. Stat. Data Anal* **53**(10), 3717–3725 (2009)
- Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., Boatwright, P.: A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *J. R. Stat. Soc. Ser. C Appl. Stat.* **54**(1), 127–142 (2005)
- Sáez-Castillo, A.J., Conde-Sánchez, A.: A hyper-Poisson regression model for overdispersed and underdispersed count data. *Comput. Stat. Data Anal.* **61**, 148–157 (2013)
- Shoukri, M.M., Asyali, M.H., VanDorp, R., Kelton, D.: The Poisson inverse Gaussian regression model in the analysis of clustered counts data. *J. Data Sci.* **2**(1), 17–32 (2004)
- Shanker, R., Mishra, A.: A quasi Lindley distribution. *Afr. J. Math. Comput. Sci. Res.* **6**(4), 64–71 (2013)
- Wongrin, W., Bodhisuwan, W.: Generalized Poisson–Lindley linear model for count data. *J. Appl. Stat.* **44**(15), 2659–2671 (2017)
- Zamani, H., Ismail, N., Faroughi, P.: Poisson-weighted exponential univariate version and regression model with applications. *J. Math. Stat.* **10**(2), 148–154 (2014)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.