AL-Bahir
Journal for Engineering and Pure Sciences
BJEPS

Volume 6 | Issue 1 Article 9

Poisson Lindley-Quasi XGamma Distribution For Count Data: Properties And Applications

Seth Borbye
Department of Statistics and Actuarial Science, School of Mathematical Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo

Suleman Nasiru
Department of Statistics and Actuarial Science, School of Mathematical Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo

Kingsley Kuwubasamni Ajongba
Department of Statistics and Acturial Science, School of Mathematical Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo

Sampson Wiredu
Department of Statistics and Acturial Science, School of Mathematical Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo

Poisson Lindley-Quasi XGamma Distribution For Count Data: Properties And Applications

Source of Funding

This research received no specific funding from any funding agency.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this work.

Data Availability

The (COVID-19 of Australia and Survival Times) data used to support the findings of this study are included within the article.

Author Contributions

Conceptualization: Seth Borbye, Suleman Nasiru, Kingsley Kuwubasamni Ajongba, Sampson Wiredu. Formal analysis: Seth Borbye, Suleman Nasiru, Kingsley Kuwubasamni Ajongba. Methodology: Seth Borbye, Suleman Nasiru, Kingsley Kuwubasamni Ajongba, Sampson Wiredu. Software: Seth Borbye, Suleman Nasiru, Kingsley Kuwubasamni Ajongba. Writing -original draft: Seth Borbye, Suleman Nasiru, Kingsley Kuwubasamni Ajongba. Writing - review & editing: Seth Borbye, Suleman Nasiru, Kingsley Kuwubasamni Ajongba, Sampson Wiredu.

REVIEW

Poisson Lindley-quasi XGamma Distribution for Count Data: Properties and Applications

Seth Borbye*, Suleman Nasiru, Kingsley K. Ajongba, Sampson Wiredu

Department of Statistics and Actuarial Science, School of Mathematical Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo, Ghana

Abstract

The literature contains innumerable probability distributions for modeling over-dispersed and under-dispersed count datasets from various fields of study. However, some of these proposed distributions are inadequate due to empirical or theoretical characteristics. Therefore, minimizing information loss during modeling has prompted the demand to modify the classical discrete distributions. A new two-parameter count distribution is proposed by combining Poisson and Lindley-quasi XGamma distributions via a continuous mixture technique. Some statistical properties have been derived and studied, including factorial moments, raw moments, probability generating function, moment generating function, characteristic function, mean, variance, dispersion index, skewness, and kurtosis. The shape of the PMF and dispersion index suggest that the proposed distribution is right-skewed with a heavy tail, over-dispersed, and approximately equi-dispersed. The unknown parameters of the proposed model are estimated using both maximum likelihood and Bayesian techniques. The usefulness and flexibility of the proposed distribution are measured using two distinctive datasets. The application results reveal that the developed distribution provides maximum fit to the given datasets compared to the other eight standard discrete distributions. The Poisson Lindley-quasi XGamma distribution should therefore be considered by researchers when modeling over-dispersed count data from all fields of study.

Keywords: Discrete distribution, Over-dispersion, Poisson distribution, Moments, Mixed-Poisson distributions

1. Introduction

Over the past few decades, statistical distributions have gained significant attention from researchers due to their crucial role in modeling data across various fields of study, including medicine, transportation, engineering, epidemiology, public health, and agriculture, among others. Count data from these fields, such as the number of deaths caused by COVID-19 in Ghana, the survival times (in weeks) of patients with acute myelogenous leukemia, or the survival times of lung cancer patients, the number of times judgment is passed in the court are conveniently modeled well using discrete distributions and the prominent among them is the Poisson distribution.

The Poisson distribution is a widely used conventional discrete distribution for modeling count data in the literature. However, the distribution has a unit variance-to-mean ratio property (equi-dispersed) [1], where the variance-to-mean ratio is fixed at one. Meanwhile, in real-life situations, count data often exhibit overdispersion meaning that, the variance-to-mean ratio is greater than one. This drawback of equal variance and mean assumption of the Poisson distribution has made the distribution inflexible for modeling count data with unequal variance and mean.

Moreover, this deficiency of Poisson's variance-mean equality has caught the attention of many researchers to develop more plausible mixed-Poisson distributions in the literature to address the

Received 22 October 2024; revised 15 January 2025; accepted 17 January 2025.
Available online 27 February 2025

* Corresponding author.
E-mail addresses: sborbye.stu@cktutas.edu.gh (S. Borbye), sulemanstat@gmail.com, snasiru@cktutas.edu.gh (S. Nasiru), kajongba.stu@cktutas.edu.gh (K.K. Ajongba), swiredu@cktutas.edu.gh (S. Wiredu).

https://doi.org/10.55810/2313-0083.1088

issues of over-dispersion in count observations and notable among them in the literature include; Poisson-Xgamma distribution [2], Poisson Chris-Jerry distribution [3], Poisson Akash distribution [4], Poisson Weibull distribution [5], Binomial-Discrete Poisson-Lindley distribution [6], Poisson-quasi-Xgamma distribution [7], Poisson-Mirra distribution [8], Poisson Ramos-Louzada distribution [9], Poisson quasi-Lindley distribution [10], size-biased Poisson-Pranav distribution [11], Poisson new X-Lindley distribution [12], Poisson 2S-Lindley distribution [13], and the Poisson Epanechnikov-Exponential distribution [14].

Though these mixed-Poisson distributions exist in the literature, distributional assumptions play a fundamental role in selecting an appropriate parametric model for analysis. Choosing the right parametric model for analysis depends heavily on the underlying distribution of the data [15]. New forms of data with complex characteristics emerge daily that require distributions with those complex traits and offers the best fit with minimal loss of information. This has created a gap in literature necessitating the continues development of new count distributions for modeling data. This study introduces a new mixed-Poisson distribution called the Poisson Lindley-quasi Xgamma (PLQXG) distribution. The new distribution is developed by amalgamating the Poisson and Lindley-quasi-Xgamma distributions. The motivations for this study stems from the fact that complex observations often arise daily from the non-deterministic activities of humans, animals, and organisms. The motivations are:

i. The Lindley Quasi-Xgamma distribution displays desirable properties for modeling a wide range of complex continuous data compared to the existing Lindley class of distributions such as Lindley, Quasi-Xgamma and Xgamma distributions.
ii. The probability mass function (PMF) of the PLQXG distribution displays desirable properties with different degrees of kurtosis.
iii. The dispersion index (DI) of the PLQXG model exhibits that the distribution is flexible for modeling over-dispersed count data.
iv. Evaluate the utility of the PLQXG distribution using both classical and Bayesian methods.

The rest of this study is organized as follows: Section 2 details the development of the PLQXG distribution, Section 3 presents some statistical properties of developed distribution, Section 4 presents the maximum likelihood (ML) estimation technique for estimating the parameters of the PLQXG distribution, Section 5 details the Monte Carlo simulation results of the ML estimation technique, Section 6 presents the classical and Bayesian usefulness of the developed distribution, the conclusion of the study is finally presented in Section 7.

2. Poisson Lindley-quasi XGamma distribution

Suppose the random variable X follows the Poisson distribution with probability mass function (PMF) given by

$$\mathbb{P}(x;\alpha)=\frac{e^{-\alpha}\alpha^{x}}{x!},x=0,1,2,\ldots, \tag{1}$$

where the rate parameter $\alpha > 0$ is a random variable that follows the Lindley-quasi Xgamma distribution [16] with probability density function (PDF) given by

$$f(\alpha;\lambda,\eta)=\frac{\eta e^{-\eta\alpha}}{(\lambda+\eta)^{2}}\left[\left(\lambda+\eta\right)\left(\lambda+\frac{\alpha^{2}\eta^{2}}{2}\right)+\eta(\eta-1)(1+\lambda\alpha)\right],x>0,\lambda>0,\eta>0, \tag{2}$$

where η is a scale parameter and λ is a shape parameter.

Proposition 1. The PMF of the Poisson Lindley-quasi Xgamma (PLQXG) distribution is given by

$$\mathbb{P}(X=x)=\frac{\eta}{(\lambda+\eta)^{2}(1+\eta)^{x+1}}\left[\left(\lambda+\eta\right)\left(\lambda+\frac{\eta^{2}(x+1)(x+2)}{2(1+\eta)^{2}}\right)+\eta(\eta-1)\left(1+\frac{\lambda(x+1)}{1+\eta}\right)\right], \tag{3}$$

where $\eta > 0$ is a scale parameter, $\lambda > 0$ is a shape parameter and $x = 0, 1, 2, \ldots$.

Proof. The PMF of the PLQXG distribution is obtained using the continuous mixtures technique given by

$$\mathbb{P}(X=x)=\int_{0}^{\infty}\mathbb{P}(X=x;\alpha)f(\alpha;\lambda,\eta)d\alpha, \tag{4}$$

where $\mathbb{P}(X=x;\alpha)$ and $f(\alpha;\lambda,\eta)$ are the PMF of the Poisson distribution and the PDF of the Lindley-quasi Xgamma distribution respectively presented in Eqs. (1) and (2). Therefore

$$\mathbb{P}(X=x)=\int_0^\infty \frac{\eta\alpha^x e^{-\alpha(1+\eta)}}{x!(\lambda+\eta)^2}\left[(\lambda+\eta)\left(\lambda+\frac{\alpha^2\eta^2}{2}\right)+\eta(\eta-1)(1+\lambda\alpha)\right]d\alpha.$$

After some algebraic manipulation, we obtain

$$\mathbb{P}(X=x)=\frac{\eta}{(\lambda+\eta)^2(1+\eta)^{x+1}}\left[\lambda(\lambda+\eta)+\eta(\eta-1)+\frac{(\lambda+\eta)\eta^2}{2(1+\eta)^2}(x+1)(x+2)+\frac{\lambda\eta(\eta-1)}{1+\eta}(x-1)\right].$$

Further simplification yields the PMF of the PLQXG distribution as

$$\mathbb{P}(X=x)=\frac{\eta}{(\lambda+\eta)^2(1+\eta)^{x+1}}\left[(\lambda+\eta)\left(\lambda+\frac{\eta^2(x+1)(x+2)}{2(1+\eta)^2}\right)+\eta(\eta-1)\left(1+\frac{\lambda(x+1)}{1+\eta}\right)\right].$$

Fig. 1 displays the PMF plot of the PLQXG distribution for some selected parameter values. It is observed from Fig. 1 that the PMF of the developed distribution is right-skewed.

3. Statistical properties

This section outlines some statistical properties of the PLQXG distribution such as factorial moments, ordinary moments, probability generating function, moment generating function and characteristic function.

3.1. Moments

3.1.1. Factorial moments

Proposition 2. The r^{th} factorial moment about the origin of the PLQXG distribution is given by

$$\mu'_{[r]}=\frac{\Gamma(r+1)}{(\lambda+\eta)^2\eta^r}\left[(\lambda+\eta)\left(\lambda+\frac{(r+1)(r+2)}{2}\right)+(\eta-1)(\eta+\lambda(r+1))\right], r=1,2,3,\ldots. \tag{5}$$

Proof. *By definition, the factorial moments of the PLQXG distribution is obtain by*

$$\mu'_{[r]}=E[E(X^r|\lambda,\eta)]=\int_0^\infty\left[\sum_{x=0}^\infty\frac{x^r\alpha^x e^{-\alpha}}{x!}\right]f(\alpha;\lambda,\eta)d\alpha,$$

where $\sum_{x=0}^\infty\frac{x^r\alpha^x e^{-\alpha}}{x!}=\alpha^r$ is the factorial moment of the Poisson distribution. Thus

$$\mu'_{[r]}=\int_0^\infty \alpha^r f(\alpha;\lambda,\eta)d\alpha.$$

Therefore

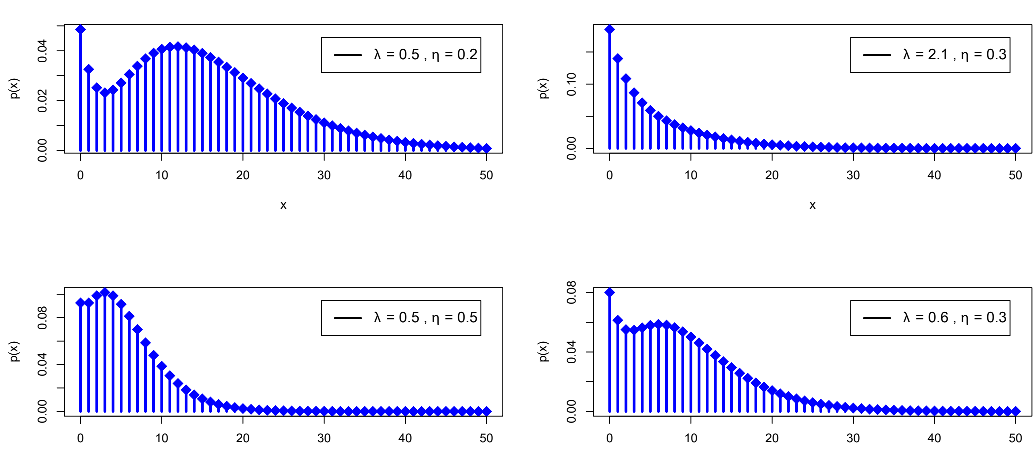

Fig. 1. PMF plots of the PLQXG distribution.

$$\mu'_{[r]} = \int_0^\infty \alpha^r \frac{\eta e^{-\eta\alpha}}{(\lambda+\eta)^2}\left[\left(\lambda+\eta\right)\left(\lambda+\frac{\alpha^2\eta^2}{2}\right) + \eta(\eta-1)(1+\lambda\alpha)\right]d\alpha.$$

Further

$$\mu'_{[r]} = \frac{1}{(\lambda+\eta)^2\eta^r}\left[(\lambda(\lambda+\eta)+\eta(\eta-1))\Gamma(r+1) + \frac{(\lambda+\eta)}{2}\Gamma(r+3)+\lambda(\eta-1)(r+1)\right].$$

Hence,

$$\mu'_{[r]} = \frac{\Gamma(r+1)}{(\lambda+\eta)^2\eta^r}\left[(\lambda+\eta)\left(\lambda+\frac{(r+1)(r+2)}{2}\right) + (\eta-1)(\eta+\lambda(r+1))\right].$$

The first four factorial moments of the PLQXG distribution are given as;

$$\mu'_{[1]} = \frac{1}{(\lambda+\eta)^2\eta}[(\lambda+\eta)(\lambda+3)+(\eta-1)(\eta+2\lambda)],$$

$$\mu'_{[2]} = \frac{2}{(\lambda+\eta)^2\eta^2}[(\lambda+\eta)(\lambda+6)+(\eta-1)(\eta+3\lambda)],$$

$$\mu'_{[3]} = \frac{6}{(\lambda+\eta)^2\eta^3}[(\lambda+\eta)(\lambda+10)+(\eta-1)(\eta+4\lambda)]$$

and

$$\mu'_{[4]} = \frac{24}{(\lambda+\eta)^2\eta^4}[(\lambda+\eta)(\lambda+15)+(\eta-1)(\eta+5\lambda)].$$

3.1.2. Ordinary moments

Using the general relationship between factorial moments about the origin and ordinary moments, the first two ordinary moments of the PLQXG distribution are given as;

$$\mu'_1 = \frac{1}{(\lambda+\eta)^2\eta}[(\lambda+\eta)(\lambda+3)+(\eta-1)(\eta+2\lambda)]$$

and

$$\mu'_2 = \frac{1}{(\lambda+\eta)^2\eta}\left[\frac{\lambda+\eta}{\eta}(\eta(\lambda+3)+2(\lambda+6)) + \frac{\eta-1}{\eta}(\eta(\eta+2\lambda)+2(\eta+3\lambda))\right].$$

The dispersion index (DI) of the PLQXG distribution is defined as

$$DI = \frac{Var(X)}{\mu'_1}. \tag{6}$$

Hence, the DI of the PLQXG distribution is given by

$$DI = \frac{A}{(\lambda+\eta)(\lambda+3)+(\eta-1)(\eta+2\lambda)}, \tag{7}$$

where

$$A = \frac{\lambda+\eta}{\eta}(\eta(\lambda+3)+2(\lambda+6)) + \frac{\eta-1}{\eta}(\eta(\eta+2\lambda)+2(\eta+3\lambda)) - \frac{1}{\eta(\lambda+\eta)^2}[(\lambda+\eta)(\lambda+3)+(\eta-1)(\eta+2\lambda)]^2.$$

Table 1 displays the first four ordinary moments, coefficient of variation (CV) and DI of the PLQXG distribution for the following set of parameter values I:$\lambda = 0.5$ and $\eta = 0.6$, II: $\lambda = 2$ and $\eta = 0.5$, III: $\lambda = 0.9$ and $\eta = 5$, IV: $\lambda = 100$ and $\eta = 200$. It is observed from Table 1 that the PLQXG distribution is flexible for modeling over-dispersed and approximately equi-dispersed count data.

Fig. 2 presents the plot of the DI of the PLQXG distribution for $\lambda \in [0.5, 20]$ and $\eta \in [0.01, 40]$. It is

Table 1. First four ordinary moments, CV and DI of the PLQXG distribution.

μ'_r	I	II	III	IV
μ'_1	4.4215	3.2800	0.2885	0.0061
μ'_2	33.3930	24.7200	0.4528	0.0062
μ'_3	332.6125	265.3600	0.9176	0.0064
μ'_4	4074.5317	3628.5600	2.4010	0.0067
CV	13.8435	13.9616	0.3696	0.0062
DI	3.1310	4.2566	1.2812	1.0057

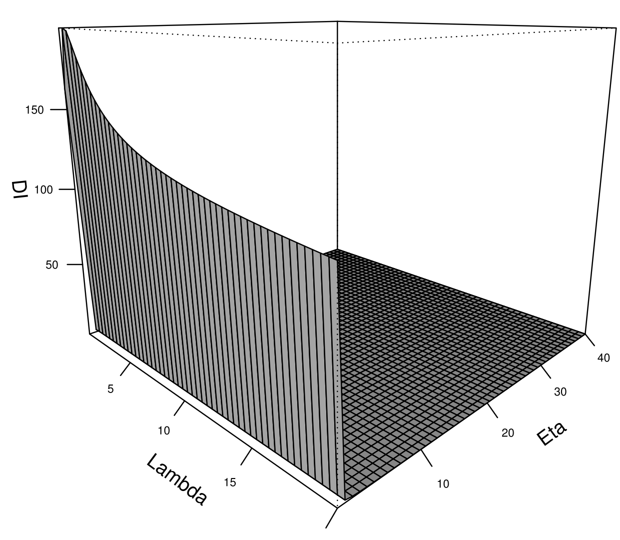

Fig. 2. DI plot of the PLQXG distribution.

observed from Fig. 2 that the PLQXG distribution can be over-dispersed.

Fig. 3 presents the plots of the coefficient of skewness (CS) and coefficient of kurtosis (CK) of the PLQXG distribution for $\lambda \in [0.5, 20]$ and $\eta \in [0.01, 40]$. It is observed from Fig. 3 that the PLQXG distribution can be right-skewed and leptokurtic.

3.2. Probability generating function

Propositions 3. The probability generating function G(s) of the PLQXG distribution is given by

$$G(s) = \frac{(\lambda(\lambda+\eta)+\eta(\eta-1))(\eta-s-1)^2+(\lambda+\eta)\eta^2\lambda\eta(\eta-1)(\eta-s+1)}{(\eta-s+1)^3}. \tag{8}$$

Proof. By definition

$$G(s) = E\left[E(s^X|\lambda,\eta)\right] = \int_0^\infty \left[\sum_{x=0}^{\infty}\frac{s^x\alpha^x e^{-\alpha}}{x!}\right] f(\alpha;\lambda,\eta)d\alpha,$$

where $\sum_{x=0}^{\infty}\frac{s^x\alpha^x e^{-\alpha}}{x!} = e^{\alpha(s-1)}$ is the probability generating function of the Poisson distribution. Thus

$$G(s) = \int_0^\infty e^{\alpha(s-1)} f(\alpha;\lambda,\eta)d\alpha.$$

Therefore

$$G(s) = \int_0^\infty \frac{\eta e^{-\alpha(\eta-s+1)}}{(\lambda+\eta)^2}\left[\left(\lambda+\eta\right)\left(\lambda+\frac{\alpha^2\eta^2}{2}\right) + \eta(\eta-1)(1+\lambda\alpha)\right]d\alpha.$$

After some algebraic manipulation, we obtain

$$G(s) = \frac{\lambda(\lambda+\eta)+\eta(\eta-1)}{\eta-s+1} + \frac{(\lambda+\eta)\eta^2}{2(\eta-s+1)^3}\Gamma(3) + \frac{\lambda\eta(\eta-1)}{(\eta-s+1)^2}\Gamma(2).$$

Hence,

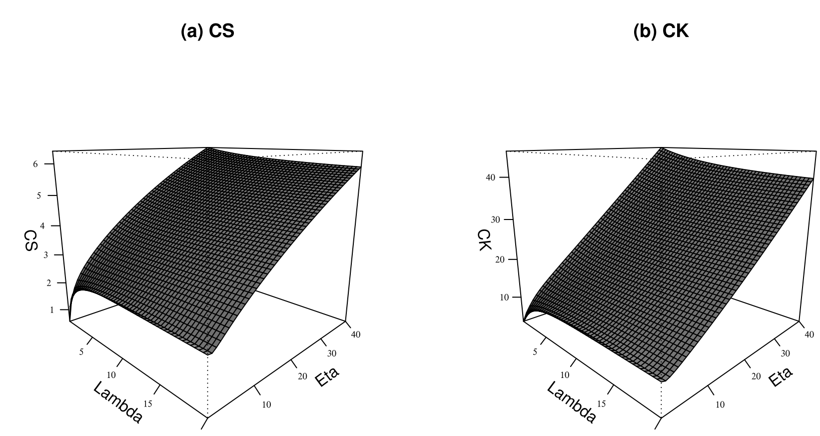

Fig. 3. CS and CK plots of the PLQXG distribution.

$$G(s)=\frac{(\lambda(\lambda+\eta)+\eta(\eta-1))(\eta-s-1)^2+(\lambda+\eta)\eta^2+\lambda\eta(\eta-1)(\eta-s+1)}{(\eta-s+1)^3}.$$

3.3. Moment generating function

Proposition 4. The moment generating function (mgf) of the PLQXG distribution if it may exist is given as

$$M_X(t)=\frac{(\lambda(\lambda+\eta)+\eta(\eta-1))(\eta-e^t-1)^2+(\lambda+\eta)\eta^2+\lambda\eta(\eta-1)(\eta-e^t+1)}{(\eta-e^t+1)^3}. \quad (9)$$

Proof. By definition

$$M_X(t)=E\big[E(e^{tX}|\lambda,\eta)\big]=\int_0^\infty\left[\sum_{x=0}^{\infty}\frac{e^{tx}\alpha^x e^{-\alpha}}{x!}\right]f(\alpha;\lambda,\eta)d\alpha,$$

where $\sum_{x=0}^{\infty}\frac{e^{tx}\alpha^x e^{-\alpha}}{x!}=e^{\alpha(e^t-1)}$ is the moment generating function (mgf) of the Poisson distribution. Thus

$$M_X(t)=\int_0^\infty e^{\alpha(e^t-1)}f(\alpha;\lambda,\eta)d\alpha.$$

Therefore

$$M_X(t)=\int_0^\infty\frac{\eta e^{-\alpha(\eta-e^t+1)}}{(\lambda+\eta)^2}\left[\left(\lambda+\eta\right)\left(\lambda+\frac{\alpha^2\eta^2}{2}\right)+\eta(\eta-1)(1+\lambda\alpha)\right]d\alpha.$$

After some algebraic manipulation

$$M_X(t)=\frac{(\lambda(\lambda+\eta)+\eta(\eta-1))(\eta-e^t-1)^2+(\lambda+\eta)\eta^2+\lambda\eta(\eta-1)(\eta-e^t+1)}{(\eta-e^t+1)^3}.$$

3.4. Characteristic function

Proposition 5. The characteristic function *Q(t)* of the PLQXG distribution is given by

$$Q(t)=\frac{(\lambda(\lambda+\eta)+\eta(\eta-1))(\eta-e^{it}-1)^2+(\lambda+\eta)\eta^2+\lambda\eta(\eta-1)(\eta-e^{it}+1)}{(\eta-e^{it}+1)^3}, i=\sqrt{-1}. \quad (10)$$

Proof. By definition

$$Q(t)=E\big[E(e^{itX}|\lambda,\eta)\big]=\int_0^\infty\left[\sum_{x=0}^{\infty}\frac{e^{itx}\alpha^x e^{-\alpha}}{x!}\right]f(\alpha;\lambda,\eta)d\alpha,$$

where $\sum_{x=0}^{\infty}\frac{e^{itx}\alpha^x e^{-\alpha}}{x!}=e^{\alpha(e^{it}-1)}$ is the characteristic function of the Poisson distribution. Thus

$$Q(t)=\int_0^\infty e^{\alpha(e^{it}-1)}f(\alpha;\lambda,\eta)d\alpha.$$

Therefore

$$M_X(t)=\int_0^\infty\frac{\eta e^{-\alpha(\eta-e^{it}+1)}}{(\lambda+\eta)^2}\left[\left(\lambda+\eta\right)\left(\lambda+\frac{\alpha^2\eta^2}{2}\right)+\eta(\eta-1)(1+\lambda\alpha)\right]d\alpha.$$

After some algebraic manipulation

$$Q(t)=\frac{(\lambda(\lambda+\eta)+\eta(\eta-1))(\eta-e^{it}-1)^2+(\lambda+\eta)\eta^2+\lambda\eta(\eta-1)(\eta-e^{it}+1)}{(\eta-e^{it}+1)^3}.$$

4. Parameter estimation

In this section, the ML estimation technique is employed for estimating the parameters of the PLQXG distribution. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n obtained from the PLQXG distribution and $x_1, x_2, x_3, \ldots, x_n$ be the observed values of $X_1, X_2, X_3, \ldots, X_n$. Then, the log-likelihood function of the PLQXG distribution is given by

$$\ell(\lambda,\eta;x) = n\log(\eta) - 2n\log(\lambda+\eta) - \sum_{i=1}^{n}(x_i+1)\log(1+\eta) + \sum_{i=1}^{n}\log\left[(\lambda+\eta)\left(\lambda+\frac{\eta^2(x_i+1)(x_i+2)}{2(1+\eta)^2}\right) + \eta(\eta-1)\left(1+\frac{\lambda(x_i+1)}{1+\eta}\right)\right]. \tag{11}$$

Taking the partial derivatives of the log-likelihood function presented in equation (11) with respect to λ and η produce the score functions of the PLQXG distribution denoted as

$$S_\lambda = \frac{\partial \ell(\lambda,\eta;x)}{\partial \lambda} \tag{12}$$

and

$$S_\eta = \frac{\partial \ell(\lambda,\eta;x)}{\partial \eta}. \tag{13}$$

The ML estimates of the PLQXG distribution are obtained by setting $S_\lambda = S_\eta = 0$ and solving simultaneously. Since Eqs. (12) and (13) are not in closed form, the ML estimates of the PLQXG distribution are obtain using numerical methods in R-software [17].

5. Simulation study

In this section, the performance of the ML estimator of the parameters of the PLQXG distribution is examined via Monte Carlo simulation experiments. The experiments were carried out using different values of the parameters. That is (λ,η)=(0.6,0.2), (λ,η)=(0.7,0.5) and (λ,η)=(0.8,0.8). The simulations were replicated 5000 times for each sample size n = 30, 100, 200, 300, 400 and 500. The sample sizes were chosen in a manner that enables us to examine the performance of the estimator of the parameters in small, moderate, and large samples. The random samples were generated using the inversion method and the performance of the estimator is examined using the average estimate (AE), absolute bias (AB), relative absolute bias (RAB), root mean square error (RMSE), and coverage probability (CP). The AB, RAB, RMSE and CP are respectively given as

$$AB = \frac{1}{N}\sum_{i=1}^{N}|\hat{\theta}_i - \theta|,$$

$$RAB = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{\theta}_i - \theta}{\theta}\right|,$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2}{N}}$$

and

$$CP = P(L(X;\theta) < \theta < U(X;\theta)).$$

From Table 2, the AEs of the parameters approach the true value as the sample size increases. This is an indication that the estimator is asymptotically unbiased. The ABs, RABs and RMSEs of the parameters decrease as the sample size increases. Hence, the maximum likelihood estimator of the parameter is consistent. The 95 % confidence interval CPs are quite high and approaches the nominal value of 0.95 as the sample size increases. It can therefore be concluded that the maximum likelihood estimator performs well in estimating the parameters of the PLQXG distribution.

6. Applications

In the section, we demonstrate the classical and Bayesian Applications of the PLQXG distribution using two distinctive datasets.

6.1. Classical applications

The classical usefulness of the PLQXG distribution is demonstrated and compared with the Poisson (P) distribution, Piosson-Mirra (PMi) distribution [8], Poisson-quasi-Xgamma (PQXG) distribution [7],

Table 2. Monte Carlo simulation results for PLQXG distribution.

	I:(λ, η)=(0.6,0.2)					
Parameter	n	AE	AB	RAB	RMSE	CP
	30	0.6993	0.1910	0.3183	0.8489	0.90167
	100	0.6047	0.0605	0.1009	0.0792	0.93100
λ	200	0.6023	0.0423	0.0705	0.0535	0.93600
	300	0.6009	0.0337	0.0562	0.0425	0.94433
	400	0.6017	0.0301	0.0501	0.0380	0.94300
	500	0.6015	0.0262	0.0437	0.0332	0.95067
	30	0.2021	0.0233	0.1166	0.0312	0.94033
	100	0.2011	0.0121	0.0603	0.0153	0.94700
η	200	0.2003	0.0085	0.0424	0.0107	0.94467
	300	0.2003	0.0067	0.0337	0.0085	0.95600
	400	0.2003	0.0059	0.0296	0.0074	0.95100
	500	0.2001	0.0052	0.0260	0.0066	0.94733
	II:(λ, η)=(0.7,0.5)					
Parameter	n	AE	AB	RAB	RMSE	CP
	30	0.9372	0.4725	0.6750	1.3844	0.9063
	100	0.7428	0.1755	0.2508	0.4927	0.9273
λ	200	0.7069	0.1057	0.1509	0.1378	0.9377
	300	0.7030	0.0843	0.1204	0.1082	0.9437
	400	0.7039	0.0711	0.1015	0.0901	0.9563
	500	0.7004	0.0649	0.0927	0.0820	0.9363
	30	0.5133	0.0842	0.1683	0.1103	0.9457
	100	0.5027	0.0444	0.0889	0.0569	0.9480
η	200	0.5026	0.0308	0.0615	0.0387	0.9503
	300	0.5022	0.0248	0.0496	0.0312	0.9523
	400	0.5011	0.0210	0.0420	0.0265	0.9543
	500	0.5015	0.0189	0.0377	0.0239	0.9523
	III: (λ, η)=(0.7,0.5)					
Parameter	n	AE	AB	RAB	RMSE	CP
	30	1.2720	1.0104	1.2630	2.0769	0.9553
	100	1.0271	0.5406	0.6758	1.3719	0.9410
λ	200	0.8706	0.2936	0.3670	0.7489	0.9470
	300	0.8344	0.2169	0.2711	0.4885	0.9530
	400	0.8159	0.1798	0.2248	0.3281	0.9470
	500	0.8136	0.1607	0.2009	0.2078	0.9493
	30	0.8349	0.1778	0.2222	0.2234	0.9570
	100	0.8058	0.1030	0.1288	0.1328	0.9517
η	200	0.8046	0.0717	0.0896	0.0933	0.9490
	300	0.8023	0.0569	0.0712	0.0725	0.9563
	400	0.8041	0.0490	0.0613	0.0621	0.9603
	500	0.8012	0.0446	0.0558	0.0562	0.9480

Poisson Ram Awadh (PRA) distribution [18], Poisson Prakaamy (PP) distribution [19], discrete inverted Topp-Leone (DITL) distribution [20], Discrete inverse Rayleigh (DIR) distribution [21] and Poisson XRani (PXR) distribution [22] based on their log-likelihood (ℓ), Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC).

6.1.1. *Dataset I: COVID-19 of Australia*

The first dataset is a 32 days daily new cases of COVID-19 from Australia recorded from 3rd September to October 4, 2020, studied by Gillariose et al. [23], Almetwally and Ibrahim [24]. The data are: 6, 15, 59, 11, 5, 9, 8, 11, 7, 9, 6, 7, 6, 0, 8, 8, 5, 7, 5, 2, 35, 2, 8, 1, 2, 3, 7, 4, 2, 2, 3. Table 3 displays the mean, standard deviation (SD), DI, CS and excess kurtosis of the first dataset, it is observed that the dataset is over-dispersed, right-skewed and leptokurtic.

Table 4 presents the ML estimates with their accompanying standard errors in brackets and model selection criteria for the fitted models. It is observed that the PLQXG distribution produce the best fit with the highest log-likelihood value and the least AIC, AICc and BIC values.

Fig. 4 presents the PMF plots of empirical and fitted models for dataset I. It is observed from Fig. 4 that the PLQXG distribution performs better than the other eight computing models since it closely mimics the empirical PMF.

6.1.2. *Dataset II: survival times*

The second dataset is the survival times (in weeks) for 33 patients suffering from acute myelogenous leukemia. The data is studied by Afify et al. [25] and the data are: 3, 3, 30, 3, 8, 4, 2, 4, 4, 65, 100, 108, 121, 4, 134, 16, 39, 26, 22, 1, 143, 56, 1, 5, 65, 17, 7, 16, 56, 65, 22, 43, 156. Table 5 displays the mean, SD, DI, CS and excess kurtosis of the second dataset, it is observed that the dataset is over-dispersed, right-skewed and platykurtic.

Table 6 presents the ML estimates with their accompanying standard errors in brackets and model selection criteria for the fitted models. It is observed that the PLQXG distribution produce the best fit with the highest log-likelihood value and the least AIC, AICc and BIC values.

Fig. 5 displays the PMF plots of the empirical and fitted models for dataset II. It is seen from Fig. 5 that the PLQXG distribution provides the maximum fit since the PMF of the PLQXG distribution closely mimics the empirical PMF.

6.2. *Bayesian applications*

This section presents the Bayesian application of the PLQXG distribution. In Bayesian estimation, the parameters of the distribution are assumed to be random variables that follow a given distribution

Table 3. Descriptive statistics of dataset I.

Minimum	Maximum	Median	Mean	SD	DI	CS	Kurtosis
0.0000	59.0000	6.0000	7.3750	9.9895	13.5309	4.2743	19.3338

Table 4. ML estimates and model selection criteria for dataset I.

Distribution	ML Estimates (SE)	ℓ	AIC	AICc	BIC
PLQXG	$\lambda = 25.9750$ (59.5988)	−97.8900	199.7738	200.1876	202.7053
	$\eta = 0.1412$ (0.0286)				
Poisson	$\lambda = 7.3750$ (0.4801)	−164.9200	331.8332	331.9665	333.2989
PMi	$\alpha = 1.5017$ (3.6764)	−98.9200	201.8398	202.2536	204.7713
	$\theta = 0.0663$ (0.0663)				
PQXG	$\theta = 0.1424$ (0.0314)	−97.9300	199.8602	200.2740	202.7916
	$\alpha = 38.7782$ (109.8474)				
PRA	$\theta = 0.8146$ (0.0789)	−107.0300	216.0687	216.2021	217.5345
PP	$\theta = 0.8137$ (0.0787)	−107.0100	216.0211	216.1544	217.4868
DITL	$\theta = 0.7562$(0.1338)	−106.3100	214.6111	214.7445	216.0769
DIR	$\alpha = 10.9096$(2.0231)	−102.6100	207.2138	207.3471	208.6795
PXR	$\lambda = 0.6761$(0.0674)	−104.2400	210.4793	210.6127	211.9451

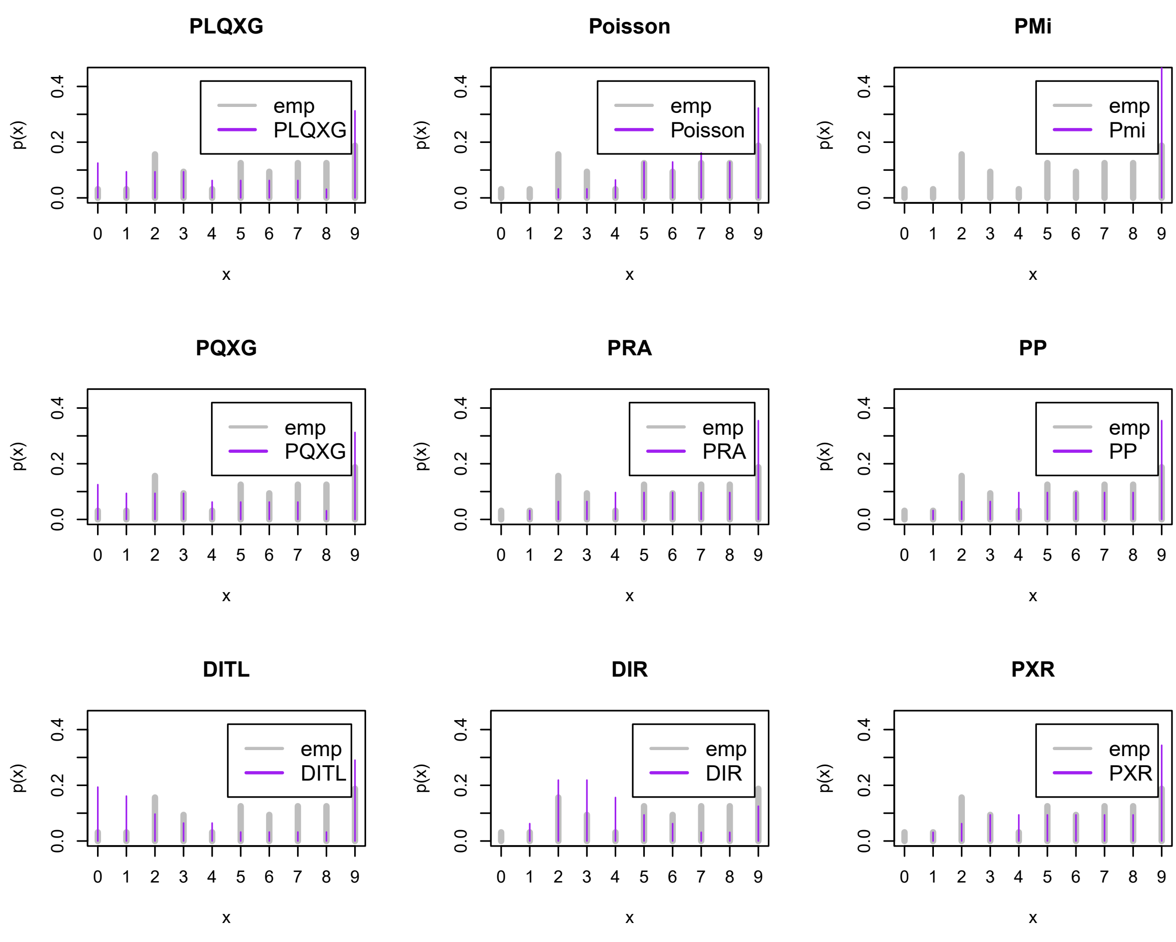

Fig. 4. PMF plots of the empirical and fitted models for dataset I.

Table 5. Descriptive statistics of dataset II.

Minimum	Maximum	Median	Mean	SD	DI	CS	Kurtosis
1.0000	156.0000	22.0000	40.8800	46.7030	53.3571	1.1120	−0.0642

Table 6. ML estimates and model selection criteria for dataset II.

Distribution	ML Estimates (SE)	ℓ	AIC	AICc	BIC
PLQXG	$\lambda = 1.2521$ (0.4950)	−152.9400	309.8700	310.2700	312.8630
	$\eta = 0.0424$ (0.0077)				
Poisson	$\lambda = 40.8790$ (1.1130)	−872.9600	1747.9200	1748.0500	1749.4200
PMi	$\alpha = 0.0010$ (0.0014)	−155.3500	314.7080	315.1080	317.7010
	$\theta = 0.0421$ (0.0113)				
PQXG	$\theta = 0.0362$ (0.0084)	−154.9800	313.9690	314.3690	316.9620
	$\alpha = 3.1656$ (2.6364)				
PRA	$\theta = 0.1468$ (0.0112)	−231.7000	465.4011	465.5302	466.8977
PP	$\theta = 0.1468$ (0.0112)	−231.6900	465.3879	465.5169	466.8844
DITL	$\theta = 0.4204$ (0.0732)	−163.27	328.5455	328.6745	330.042
DIR	$\alpha = 26.3444$ (4.8211)	−193.51	389.0106	389.1397	390.5071
PXR	$\lambda = 0.1223$ (0.0101)	−216.700	435.3995	435.5286	436.8961

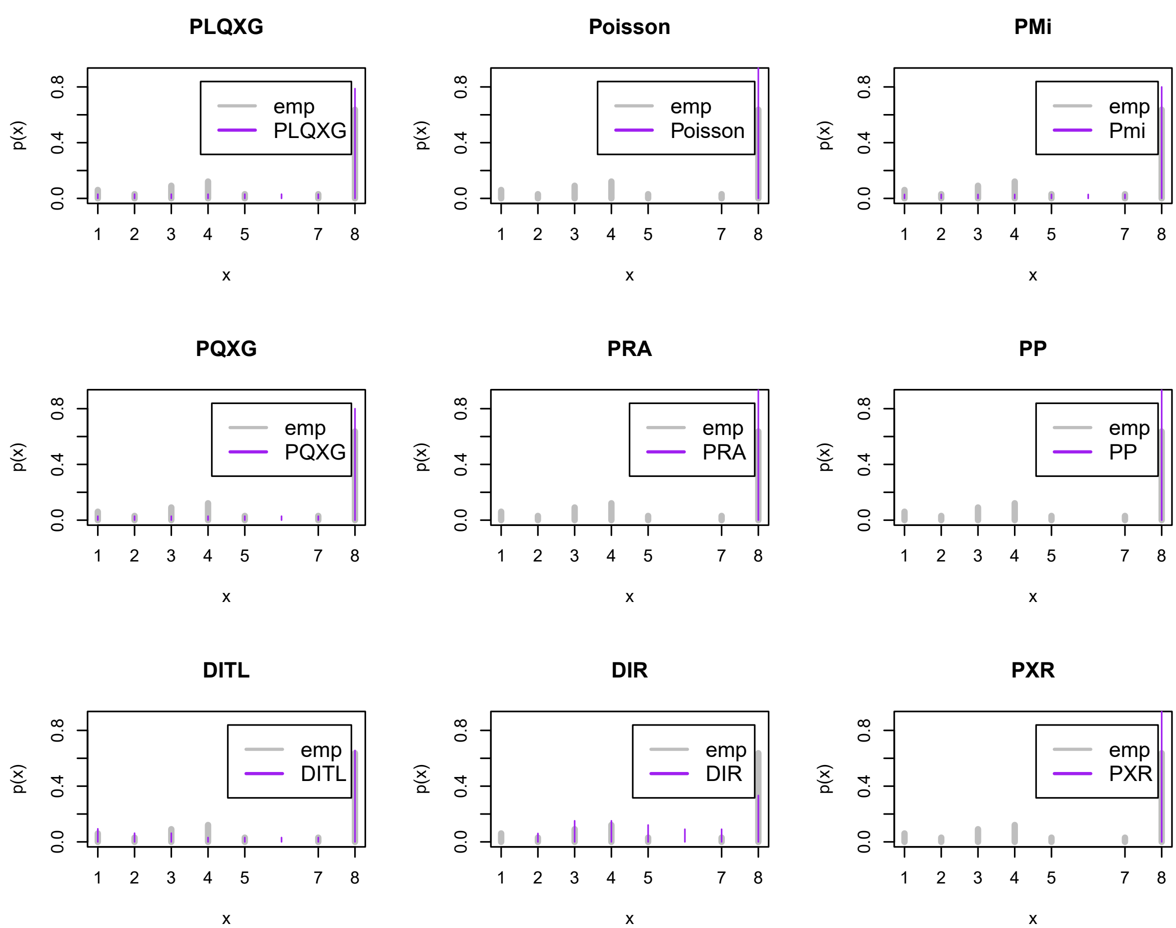

Fig. 5. PMF plots of the empirical and fitted models for dataset II.

Table 7. Posterior summaries of the PLQXG distribution for dataset I.

Parameter	Estimates	SD	SE	2.50 %	50 %	97.50 %	$\widehat{R}$	Neff
η	0.1753	0.0806	0.0006	0.0989	0.1527	0.4378	1.001	36000
λ	9.2355	7.6204	0.0419	0.4084	7.3351	29.3479	1.001	36000

Table 8. Stationarity and halfwidth test for dataset I.

	Parameter	Stationarity Test	p-value	Halfwidth Test	Halfwidth
Chain 1	η	Pass	0.251	Pass	0.0019
	λ	Pass	0.772	Pass	0.1448
Chain 2	η	Pass	0.3800	Pass	0.0019
	λ	Pass	0.2100	Pass	0.1416
Chain 3	η	Pass	0.9640	Pass	0.0020
	λ	Pass	0.5810	Pass	0.1406

called the prior distribution. This study uses the gamma distribution as the prior distribution for the parameters of the PLQXG distribution given as

$$\pi(\lambda) = \frac{b_1^{a_1}}{\Gamma(a_1)}\lambda^{a_1-1}e^{-b_1\lambda}, a_1>0, b_1>0, \lambda>0$$

and

$$\pi(\eta) = \frac{b_2^{a_2}}{\Gamma(a_2)}\lambda^{a_2-1}e^{-b_2\eta}, a_2>0, b_2>0, \eta>0$$

with hyper-parameter values $a_1 = a_2 = b_1 = b_2 = 0.1$. The study uses the R2jags package [26] in R-software to perform the analysis using three parallel chains with 900,000 iterations, 300,000 burn-in, and 50 thinning intervals each.

6.2.1. Dataset I: COVID-19 of Australia

The Bayesian estimation of parameters of the PLQXG distribution is demonstrated for dataset I. Table 7 presents the Bayesian estimate and other descriptives of the posterior parameters of the PLQXG distribution for dataset I. It is observed from Table 7 that potential reduction scale factor (R) is close to 1 and the effective sample size (Neff) is high indicating the convergence of the MCMC algorithm. Table 8 displays the stationarity and halfwidth tests for the posterior parameters of the PLQXG distribution for the dataset I. The results from Table 8 demonstrate that the process is from a stationary distribution.

The convergence of the chains is investigated using time series plots, running mean plots, and autocorrelation plots. The time series plot in Fig. 6 suggest a stationary pattern, hence convergence of the chains.

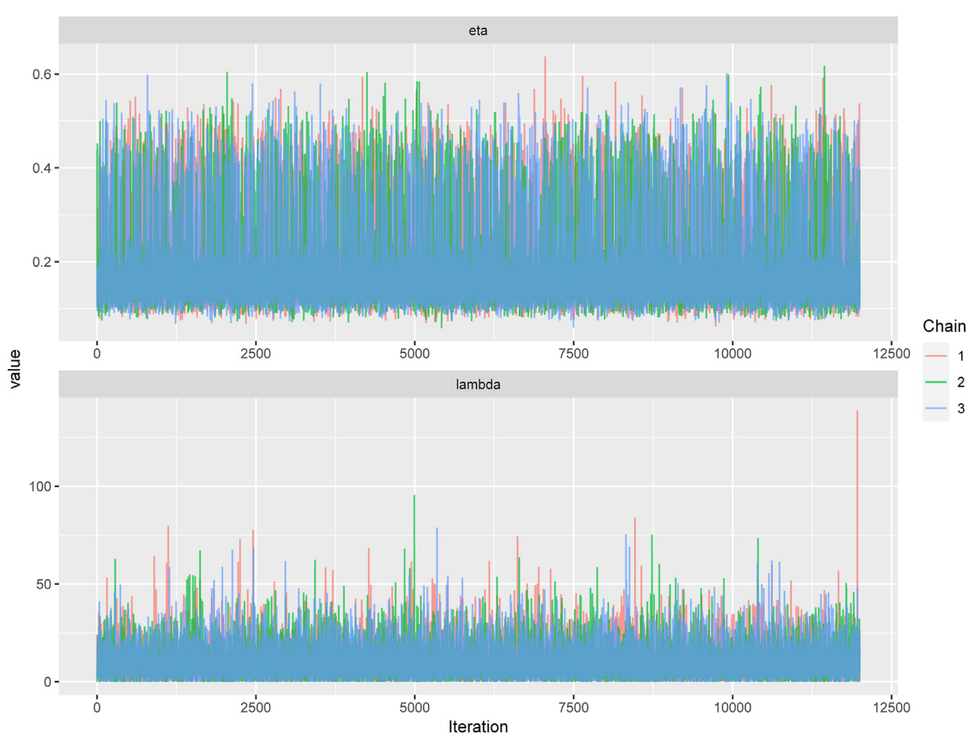

Fig. 6. Time series plots of the posterior parameters for dataset I.

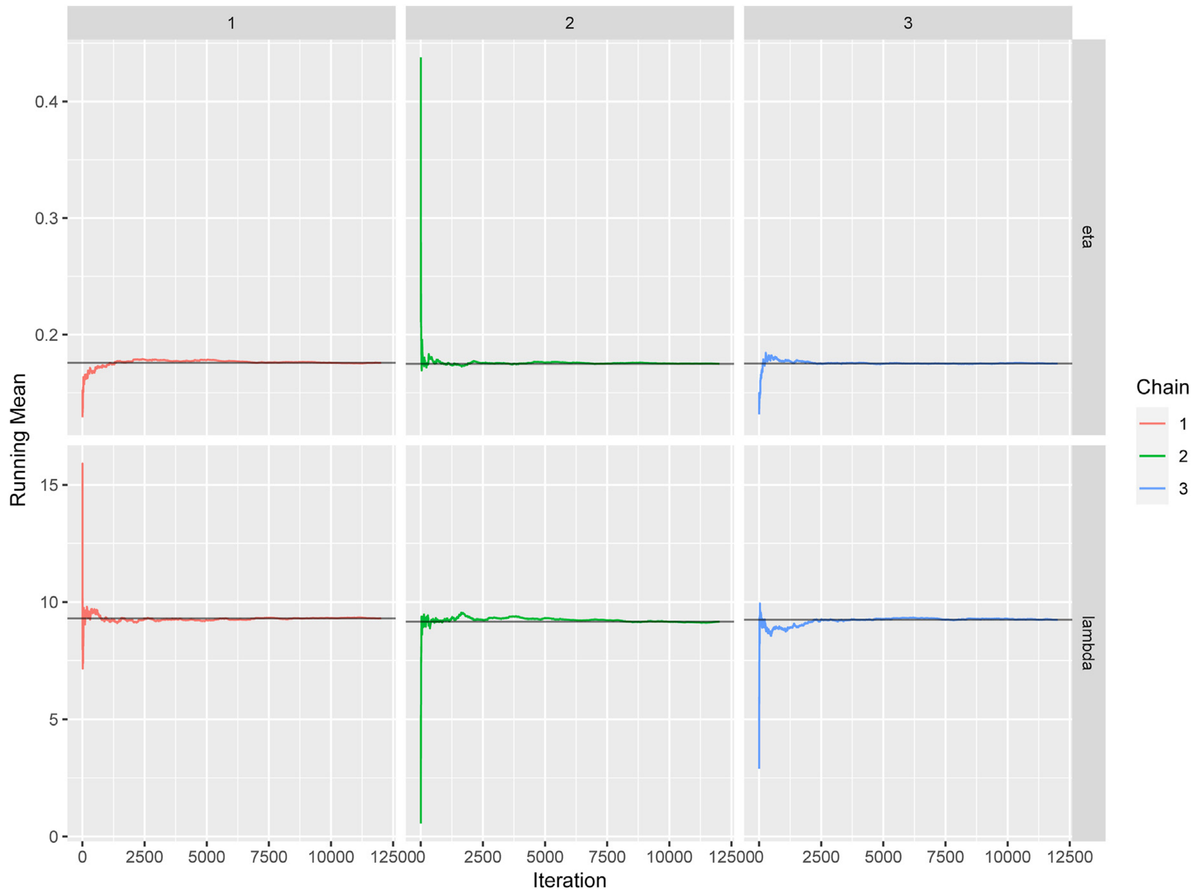

Fig. 7. Running mean plots of the posterior parameters for dataset I.

The running mean plot in Fig. 7 indicates that the chains have converged after 2500 iterations.
The swift decay of the autocorrelation plots in Fig. 8 indicates a good mixing of the chains and the convergence of the MCMC algorithm.

6.2.2. Dataset II: survival times
The Bayesian estimation of parameters of the PLQXG distribution is demonstrated for dataset II. Table 9 presents the Bayesian estimates and other descriptives of the posterior parameters of the PLQXG distribution for dataset II. It is observed from Table 9 that potential reduction scale factor $(\hat{R})$ is approximately 1 and the effective sample size (Neff) is high indicating the convergence of the MCMC algorithm.
Table 10 displays the stationarity and halfwidth test for the posterior parameters of the PLQXG distribution for dataset II. The results from Table 10 demonstrate that the process is from a stationary distribution.
The time series plot in Fig. 9 suggest a stationary pattern, hence convergence of the chains for dataset II.

The running mean plot in Fig. 10 indicates that the chains have converged after 10,000 iterations.
The sharp decay of the autocorrelation plots in Fig. 11 indicates a good mixing of the chains and the convergence of the MCMC algorithm.

7. Conclusions

A new two-parameter mixed-Poisson distribution is developed in this study by compounding the Poisson distribution and the Lindley Quasi-Xgamma distribution. Some statistical properties of the developed distribution were derived. The PMF and dispersion index of the developed distribution revealed that the PLQXG distribution is flexible for modeling right-skewed, over-dispersed, and approximately equi-dispersed count datasets. Both maximum likelihood and Bayesian estimation techniques were employed to estimate the parameters of the PLQXG distribution and a simulation study was carried out to measure the performance of the maximum likelihood estimation technique. The study demonstrated the usefulness of the developed distribution using two real datasets. The

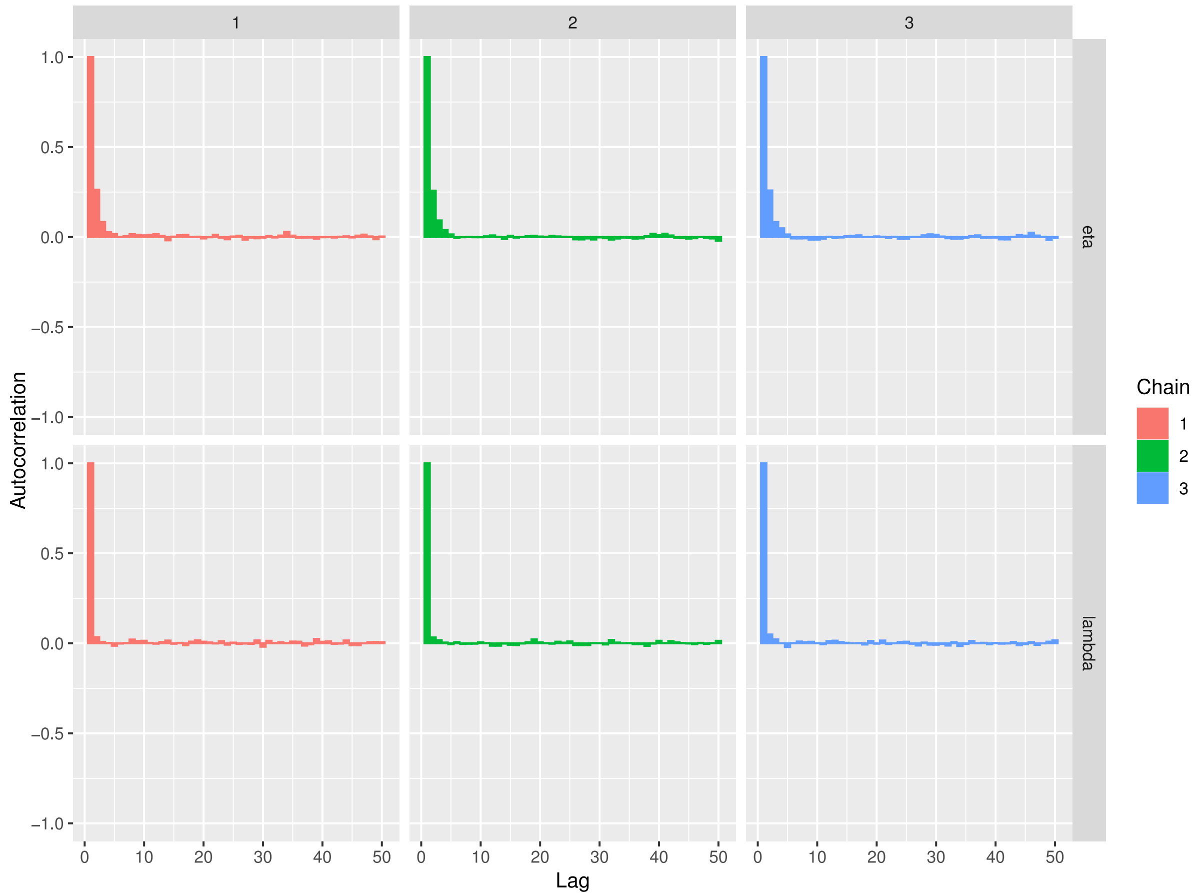

Fig. 8. Autocorrelation plots of the posterior parameters for dataset I.

Table 9. Posterior summaries of the PLQXG distribution for dataset II.

Parameter	Estimates	SD	SE	2.50 %	50 %	97.50 %	$\widehat{R}$	Neff
η	0.03950	0.0082	4.387×10^{-5}	0.0240	0.0394	0.0561	1.001	36000
λ	2.3154	2.6741	1.796×10^{-2}	0.7771	1.5614	9.0111	1.001	36000

Table 10. Stationarity and halfwidth test for dataset II.

	Parameter	Stationarity Test	p-value	Halfwidth Test	Halfwidth
Chain 1	η	Pass	0.8450	Pass	0.0002
	λ	Pass	0.9360	Pass	0.0703
Chain 2	η	Pass	0.4030	Pass	0.0002
	λ	Pass	0.2170	Pass	0.0603
Chain 3	η	Pass	0.1420	Pass	0.0001
	λ	Pass	0.4610	Pass	0.0509

PLQXG distribution was compared with eight competing models and it was revealed that the developed distribution provides a better fit than the compared models with the highest log-likelihood values and the least AIC, AICc, and BIC values. The PLQXG distribution should be considered as an alternative model when modeling over-dispersed count datasets. The proposed model has a limitation

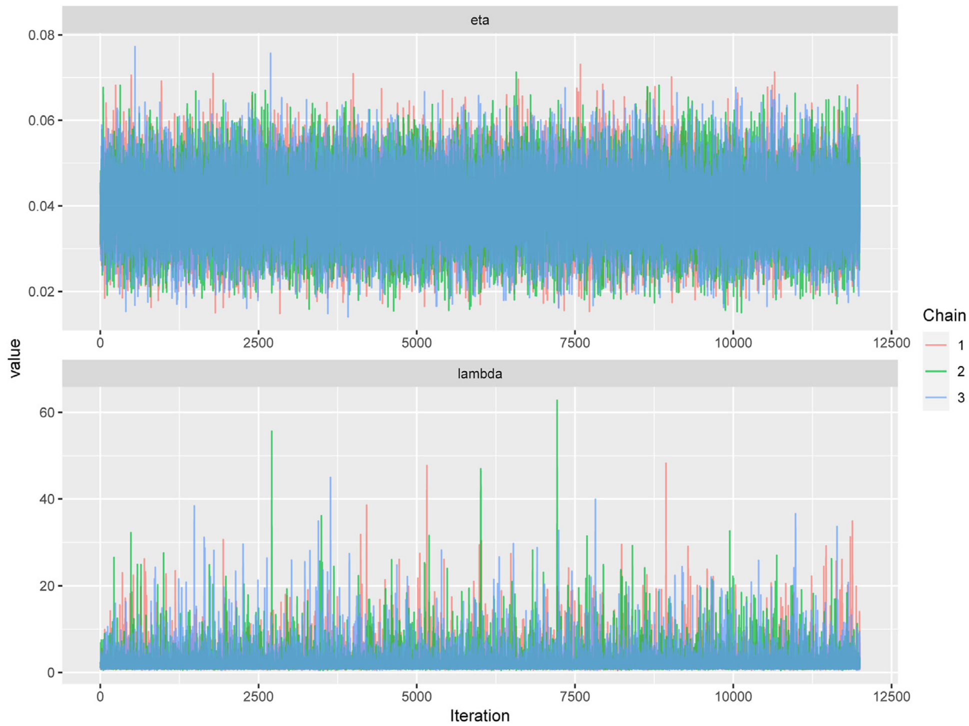

Fig. 9. Time series plots of the posterior parameters for dataset II.

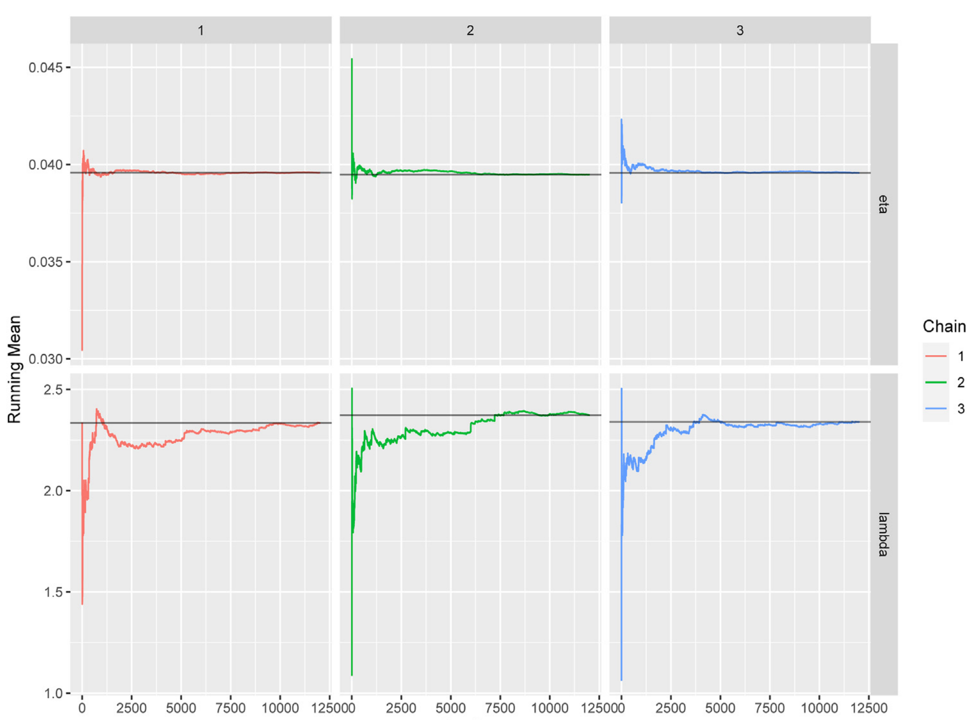

Fig. 10. Running mean plots of the posterior parameters for dataset II.

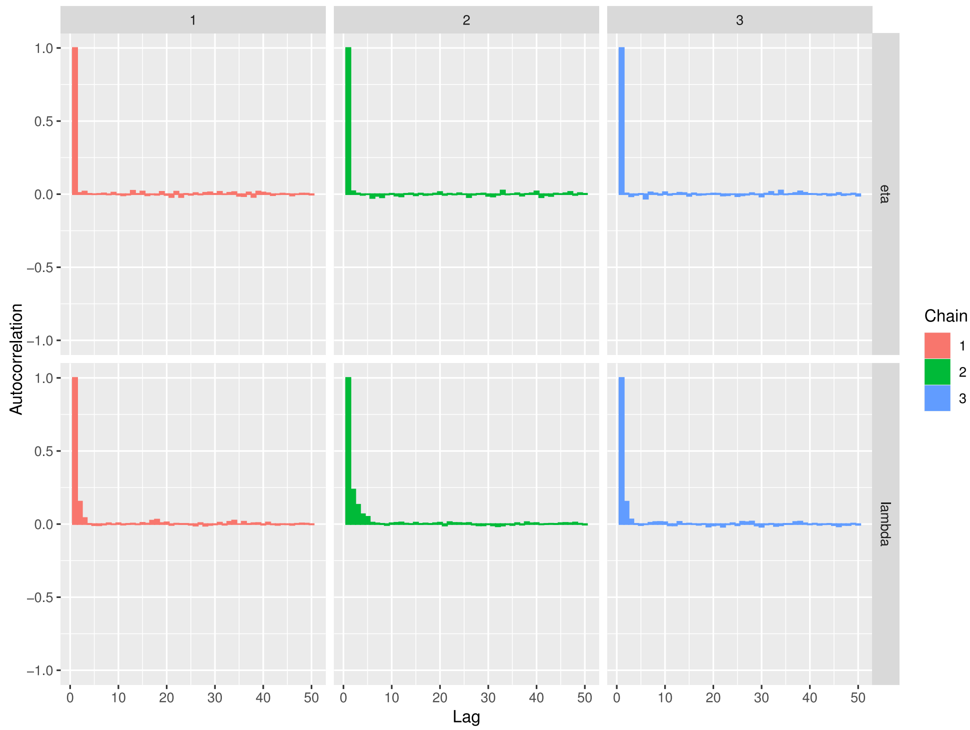

Fig. 11. Autocorrelation plots of the posterior parameters for dataset II.

of not being flexible for handling left-skewed, symmetric and under-dispersed count data. Thus, PLQXG distribution needs considerable extension and studies to render it flexible for modeling left-skewed, symmetric, bimodal and under-dispersed count data which we shall consider in our future research.

Data availability

The (COVID-19 of Australia and Survival Times) data used to support the findings of this study are included within the article.

Ethics information

This research complies with the ethical information for conducting scientific studies.

Funding

This research was conducted independently by the authors without any external funding or financial support. No grant IDs or funding sources are applicable for this study.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this work.

References

[1] Wongrin W, Bodhisuwan W. The Poisson-generalised Lindley distribution and its applications. Songklanakarin J Sci Technol 2016;38(6).
[2] Para BA, Jan TR, Bakouch HS. Poisson Xgamma distribution: a discrete model for count data analysis. Model Assisted Statistics Appl 2020;15(2):139–51.
[3] Ahmad PB, Wani MK. A new compound distribution and its applications in over-dispersed count data. Annals of Data Science 2023:1–22.
[4] Shanker R, Shukla KK, Leonida TA. A two-parameter Poisson-Akash distribution with properties and applications. Int J Probab Stat 2018;7(4):114–23.
[5] Bereta EM, Louzanda F, Franco MA. The Poisson-Weibull distribution. Adv Appl Stat 2011;22(2):107–18.
[6] Chesneau C, Bakouch H, Akdogan Y, Karakaya K. The binomial-discrete Poisson-Lindley model: modeling and applications to count regression. Commun Math Res 2022; 38(1):28–51.
[7] Altun E, Bhati D, Khan NM. A new approach to model the counts of earthquakes: INARPQX (1) process. SN Appl Sci 2021;3:1–17.
[8] Maya R, Irshad MR, Chesneau C, Nitin SL, Shibu DS. On discrete Poisson–Mirra distribution: regression, INAR (1) process and applications. Axioms 2022;11(5):193.
[9] Alkhairy I. Classical and Bayesian inference for the discrete Poisson Ramos-Louzada distribution with application to COVID-19 data. Math Biosci Eng 2023;20(8):14061–80.
[10] Grine R, Zeghdoudi H. On Poisson quasi-Lindley distribution and its applications. J Mod Appl Stat Methods 2017; 16(2):21.
[11] Shukla KK, Shanker R, Tiwari MK, Ababneh F. Size-biased Poisson-pranav distribution and its applications. Intern J Agri Stat Sci 2023;19(2).
[12] Irshad MR, Aswathy S, Maya R, Nadarajah S. New one-parameter over-dispersed discrete distribution and its

application to the nonnegative integer-valued autoregressive model of order one. Mathematics 2023;12(1):81.
[13] Chesneau C, D'cruz V, Maya R, Irshad M. A novel discrete distribution based on the mixture of Poisson and sum of two Lindley random variables. REVSTAT-Stat Journal 2023.
[14] Karakaya K. A new discrete distribution with applications to radiation, smoking, and health data. J Rad Res Appl Sci 2023; 16(4):100735.
[15] Alghamdi SM, Albalawi O, Almarzouki SM, Nagarjuna VB, Nasiru S, Elgarhy M. Different estimation methods of the modified Kies Topp-Leone model with applications and quantile regression. PLoS One 2024;19(9):e0307391.
[16] Hassan A, Wani SA, Shafi S. Lindley-quasI xgamma distribution: properties and applications. Pak J Statist 2020;36(1): 73–89.
[17] Team RC. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2020.
[18] Shukla KK, Shanker R, Tiwari MK. A new one parameter discrete distribution and its applications. J Stat Manag Syst 2022;25(1):269–83.
[19] Shukla KK, Shanker R. The Poisson-Prakaamy distribution and its applications. Aligarh J Stat 2020;40:137–50.
[20] Eldeeb AS, Ahsan-Ul-Haq M, Babar A. A discrete analog of inverted Topp-Leone distribution: properties, estimation and applications. Intern J Analysis Applications 2021;19(5): 695–708.
[21] Hussain T, Ahmad M. Discrete inverse Rayleigh distribution. P J Stat 2014;30(2).
[22] Borbye S, Nasiru S, Ajongba KK. Poisson XRani distribution: an alternative discrete distribution for overdispersed count data. Int J Math Math Sci 2024;2024(1):5554949.
[23] Gillariose J, Balogun OS, Almetwally EM, Sherwani RAK, Jamal F, Joseph J. On the discrete Weibull Marshall–Olkin family of distributions: properties, characterizations, and applications. Axioms 2021;10(4):287.
[24] Almetwally EM, Ibrahim GM. Discrete alpha power inverse Lomax distribution with application of COVID-19 data. Int J Appl Math 2020;9(6):11–22.
[25] Afify AZ, Elmorshedy M, Eliwa M. A new skewed discrete model: properties, inference, and applications. Pak J Statistics Oper Res 2021:799–816.
[26] Su YS, Yajima M. R2jags: a package for running jags from R. R package version 0.03-08. 2012.