Mathematical Biosciences and Engineering

MBE, 20(8): 14061–14080.
DOI: 10.3934/mbe.2023628
Received: 27 February 2023
Revised: 27 May 2023
Accepted: 12 June 2023
Published: 25 June 2023

http://www.aimspress.com/journal/MBE

Research article

Classical and Bayesian inference for the discrete Poisson Ramos-Louzada distribution with application to COVID-19 data

Ibrahim Alkhairy*

Department of Mathematics, Al-Qunfudah University College, Umm Al-Qura University, Mecca, Saudi Arabia

*** Correspondence:** Email: ihkhairy@uqu.edu.sa.

Abstract: The present study is based on the derivation of a new extension of the Poisson distribution using the Ramos-Louzada distribution. Several statistical properties of the new distribution are derived including, factorial moments, moment-generating function, probability moments, skewness, kurtosis, and dispersion index. Some reliability properties are also derived. The model parameter is estimated using different classical estimation techniques. A comprehensive simulation study was used to identify the best estimation method. Bayesian estimation with a gamma prior is also utilized to estimate the parameter. Three examples were used to demonstrate the utility of the proposed model. These applications revealed that the PRL-based model outperforms certain existing competing one-parameter discrete models such as the discrete Rayleigh, Poisson, discrete inverted Topp-Leone, discrete Pareto and discrete Burr-Hatke distributions.

Keywords: Ramos-Louzada distribution; poisson mixture, estimation; Markov Chain Monte Carlo; COVID-19; data analysis

1. Introduction

Data modeling has become extremely complicated in recent years as a result of the massive amount of data collected from many sectors, mainly in engineering, medicine, ecology, and renewable energy. The most popular option for analyzing count data sets is the Poisson distribution. The Poisson distribution has the drawback of being unable to represent overdispersed data sets. Overdispersion happens when the variation exceeds the mean. For count data sets, many researchers have presented mixed-Poisson distributions such as Poisson inverse Gaussian by [1], Conway–Maxwell–Poisson [2], Generalized Poisson Lindley [3], Poisson Weibull [4], Poisson Ishita [5], Poisson quasi-Lindley [6], Poisson Xgamma [7,8], Poisson XLindley [9], Poisson Moment Exponential [10], among authors. Even though there are several discrete models in the literature, there is still plenty of room to suggest a new discretized model that is acceptable under a variety of scenarios.

Let X be a random variable having Ramos and Louzada distribution [11] with the probability density function (PDF) given by

$$f(x;\lambda)=\frac{(\tau^2-2\tau+x)}{\tau^2(\tau-1)}e^{-\left(\frac{x}{\tau}\right)}, \qquad \tau\geq 2, x>0. \tag{1}$$

where τ is the scale parameter.

In this study, a new one-parameter discrete distribution for modeling count observations is introduced by compounding the Poisson distribution with Ramous-Louzada (RL) distribution. The resulting model is called the Poisson Ramous-Louzada (PRL) distribution. The major reason for the selection of the RL distribution as a compounding distribution is because of its simple form, which is needed to compute the statistical properties of the proposed distribution and estimate the unknown parameter. The proposed model may be used to model count datasets, which are frequently seen in real-world data modeling. To build a mixed Poisson model, it is assumed that the Poisson model's parameter is a random variable (RV) with a continuous distribution, and the count variable is drawn from the Poisson distribution conditional on the random parameter. As a result, the count variable's marginal distribution is a mixed Poisson distribution.

The remainder of the paper is structured as follows: The new model is described in Section 2 and gives graphical representations of PMF, and HRF. Section 3 deduces several mathematical characteristics. Section 4 estimates the PRL parameter using the following classical estimation methods, maximum likelihood estimation (MLE), Anderson Darling (AD), Cramer von Mises (CVM), ordinary least-squares (OLS) and weighted least squares (WLS), and a simulation study is also given. Section 5 additionally discusses the Bayesian model formulation for the suggested distribution. Section 6 examines three real-world data sets to demonstrate the versatility of the PRL distribution. Section 6 also includes a Bayesian study of real-world data sets using Markov chain Monte Carlo methods. Section 7 concludes with some recommendations.

2. The Structure of the new model

A random variable X is said to follow the Poisson Ramos-Louzada distribution if it possesses the following stochastic representation

$$(X|\theta)\sim Poisson\left(g(\theta)\right)$$

$$(\theta|\tau)\sim RL(\tau)$$

We call the marginal distribution of X the Poisson Ramos-Louzada distribution. The model is denoted by $PRL(\tau)$.

Theorem 1: The PMF of PRL distribution is given by

$$P(X=x,\tau)=\frac{\left(1+\frac{1}{\tau}\right)^{-x}(x-1+\tau(\tau-1))}{(\tau-1)(1+\tau)^2};\quad x=0,1,2,3,\ldots\,\&\,\tau\geq 2$$

Proof: The PMF of the new probability model can be obtained as

$$g(x|\theta)=\frac{e^{-\theta}\theta^x}{x!};\ \ x=0,1,2,3,\ldots\,\&\,\theta>0$$

when its parameter θ follows RL distribution

$$f(\theta;\tau)=\frac{(\tau^2-2\tau+\theta)}{\tau^2(\tau-1)}e^{-\left(\frac{\theta}{\tau}\right)}$$

We have

$$P(X=x,\tau)=\int_0^\infty g(x|\theta)\,f(\theta;\tau)d\theta$$

$$=\frac{1}{x!\,\tau^2(\tau-1)}\int_0^\infty e^{-\theta}\theta^x(\tau^2-2\tau+\theta)\,e^{-\left(\frac{\theta}{\tau}\right)}d\theta$$

$$=\frac{1}{x!\,\tau^2(\tau-1)}\left((\tau^2-2\tau)\int_0^\infty e^{-\theta}\theta^x\,e^{-\left(\frac{\theta}{\tau}\right)}d\theta+\int_0^\infty e^{-\theta}\theta^{x+1}\,e^{-\left(\frac{\theta}{\tau}\right)}d\theta\right)$$

$$=\frac{1}{x!\,\tau^2(\tau-1)}\left((\tau^2-2\tau)\left(1+\frac{1}{\tau}\right)^{-x-1}\Gamma(1+x)+\left(1+\frac{1}{\tau}\right)^{-2-x}\Gamma(2+x)\right)$$

$$P(X=x,\tau)=\frac{\left(1+\frac{1}{\tau}\right)^{-x}(x-1+\tau(\tau-1))}{(\tau-1)(1+\tau)^2};\quad x=0,1,2,3,\ldots\,\&\,\tau\geq 2. \tag{2}$$

The PMF behavior of the Poisson Ramos-Louzada distribution for various parameter values is shown in Figure 1.

As can be seen, the PMF has a positively skewed and can be used to discuss the count data that is positively skewed. The corresponding CDF of the discrete Poisson Ramos-Louzada distribution is given as

$$F(X=x)=p_r(X\leq x)=1-\sum_{v=x+1}^{\infty}P(v)$$

$$=1-\frac{\left(1+\frac{1}{\tau}\right)^{-x}\tau\left(x+\tau^2\right)}{(\tau-1)(1+\tau)^2};\quad x=0,1,2,\ldots;\tau\geq 2. \tag{3}$$

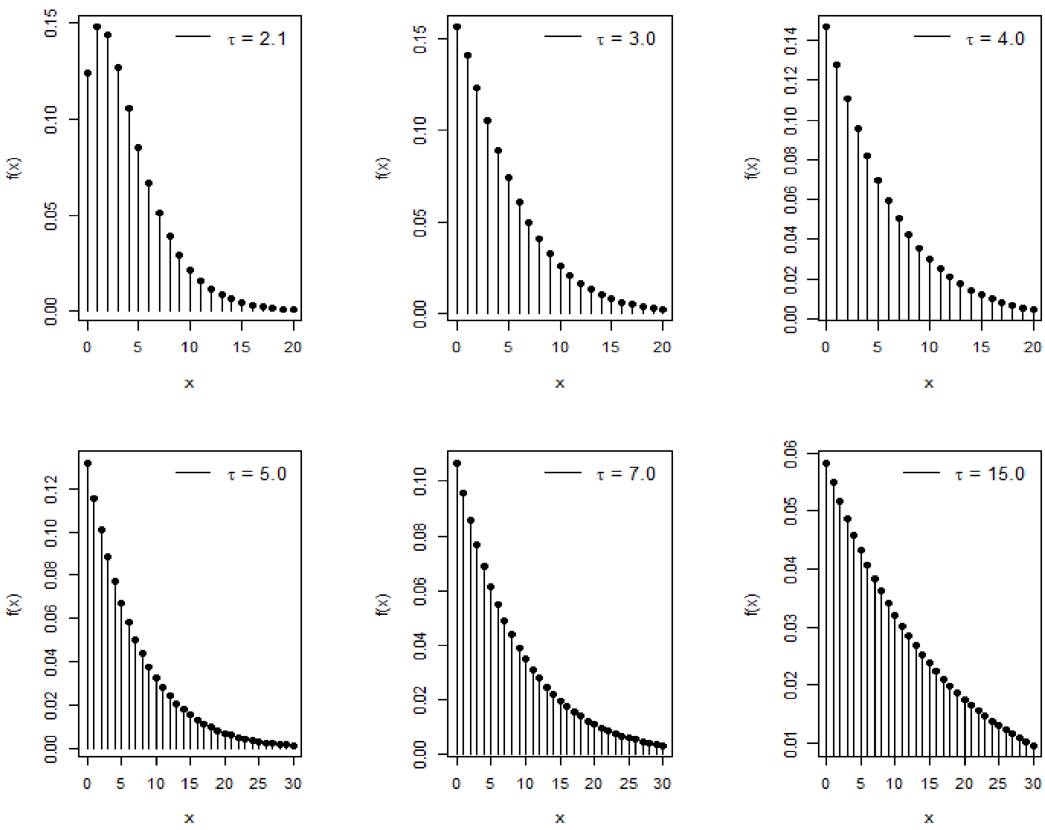

Figure 1. PMF visualization plots for the PRL distribution.

The corresponding survival function is

$$S(x;\tau)=\frac{\left(1+\frac{1}{\tau}\right)^{-x}\tau(x+\tau^2)}{(\tau-1)(1+\tau)^2}, \quad (4)$$

The hazard rate function (HRF), and reversed hazard rate function can be expressed as

$$h(x;\tau)=\frac{x+\tau(\tau-1)-1}{\tau(x+\tau^2)}, \quad (5)$$

and

$$r(x;\tau)=\frac{1-x+\tau-\tau^2}{x\tau+\tau^3-\left(1+\frac{1}{\tau}\right)^{x}(\tau-1)(1+\tau)^2}. \quad (6)$$

The graphs below depict the behavior of the HRF of the discrete PRL distribution for various parameter values.

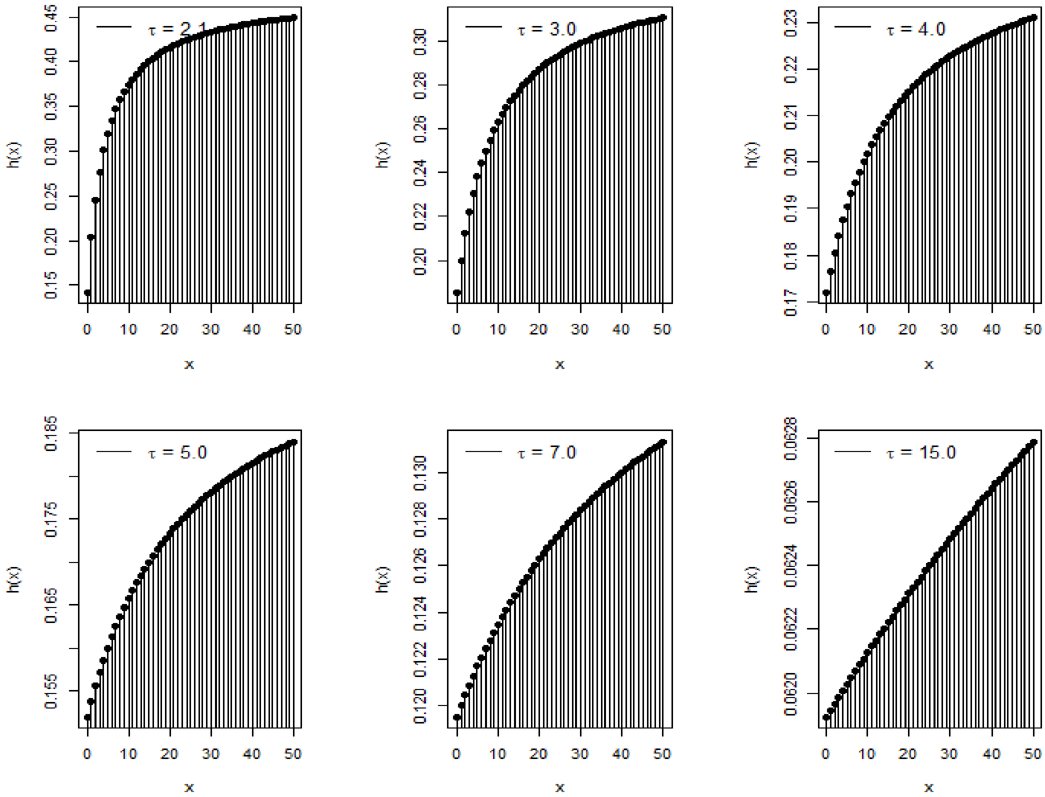

Figure 2. HRF visualization plots for the PRL distribution.

3. Statistical properties of PRL distribution

This section has examined some statistical measures of the PRL distribution. Moments, the moment generating function (MGF), and the probability generation function are among them (pgf).

3.1. Moments of PRL distribution

Assume X is a PRL random variable, the r^{th} factorial moments can be derived as

$$\mu'_{(r)} = E\left[E\left(X^{(r)}\middle|\theta\right)\right], \text{ where } X^{(r)} = X(X-1)(X-2)\ldots(X-r+1)$$

$$= \frac{1}{\tau^2(\tau-1)}\int_0^\infty \left[\sum_{x=0}^{\infty} x^{(r)}\frac{e^{-\theta}\theta^x}{x!}\right](\tau^2-2\tau+\theta)e^{-\left(\frac{\theta}{\tau}\right)}d\theta$$

$$= \frac{1}{\tau^2(\tau-1)}\int_0^\infty \left[\theta^r\sum_{x=r}^{\infty}\frac{e^{-\theta}\theta^{x-r}}{(x-r)!}\right](\tau^2-2\tau+\theta)e^{-\left(\frac{\theta}{\tau}\right)}d\theta$$

Taking $x+r$ in place of x within the bracket, we get

$$\mu'_{(r)} = \frac{1}{\tau^2(\tau-1)}\int_0^\infty \left[\theta^r \sum_{x=0}^{\infty} \frac{e^{-\theta}\theta^x}{x!}\right](\tau^2 - 2\tau + \theta)e^{-\left(\frac{\theta}{\tau}\right)}d\theta$$

$$= \frac{1}{\tau^2(\tau-1)}\int_0^\infty \theta^r\,(\tau^2 - 2\tau + \theta)e^{-\left(\frac{\theta}{\tau}\right)}d\theta$$

$$= \frac{\tau^r(-1+r+\tau)\Gamma(1+r)}{\tau-1}. \quad (7)$$

The first four factorial moments can be expressed as

$$\mu'_{(1)} = \frac{\tau^2}{\tau-1},$$

$$\mu'_{(2)} = \frac{2\tau^2(1+\tau)}{\tau-1},$$

$$\mu'_{(3)} = \frac{6\tau^3(2+\tau)}{\tau-1},$$

and

$$\mu'_{(4)} = \frac{24\tau^4(3+\tau)}{\tau-1}.$$

The first four moments about the mean of the PRL distribution are obtained.

$$\mu_2 = \frac{\tau^2(\tau^2+\tau-3)}{(\tau-1)^2}, \quad (8)$$

$$\mu_3 = \frac{\tau^2(2\tau^4+3\tau^3-14\tau^2+4\tau+7)}{(\tau-1)^3}, \quad (9)$$

$$\mu_4 = \frac{\tau^2(9\tau^6+18\tau^5-92\tau^4+41\tau^3+77\tau^2-41\tau-15)}{(\tau-1)^4}, \quad (10)$$

Using Eqs (8)–(10), the Index of Dispersion (ID), coefficient of skewness (CS), and coefficient of Kurtosis (CK) can be derived in closed forms,

$$ID(X) = \frac{Var(X)}{Mean(X)} = \frac{\tau^2+\tau-3}{\tau-1}, \quad (11)$$

$$CS(X) = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{\tau^2(7+4\tau-14\tau^2+3\tau^3+2\tau^4)}{(\tau-1)^3\left(\frac{\tau^2(-3+\tau+\tau^2)}{(-1+\tau)^2}\right)^{3/2}}, \quad (12)$$

and

$$CK(X) = \frac{9\tau^6+18\tau^5-92\tau^4+41\tau^3+77\tau^2-41\tau-15}{\tau^2(\tau^2+\tau-3)^2}. \quad (13)$$

The moment-generating function of RV X can be expressed as

$$M_X(s) = \sum_{x=0}^{\infty} e^{xs}\, P(X = x, \tau)$$

$$= \frac{\tau(\tau - e^s(\tau-2)-1)-1}{(\tau-1)(1+\tau-e^s\tau)^2}. \tag{14}$$

The probability-generating function of PRL distribution can be derived as

$$P_X(t) = \sum_{x=0}^{\infty} t^x\, P(X = x, \tau)$$

$$= \frac{-1-\tau+2t\tau+\tau^2-t\tau^2}{(-1+\tau)(-1-\tau+t\tau)^2}. \tag{15}$$

Table 1 displays some computational statistics of the PRL distribution for sundry parameter values.

Table 1. Some computational statistics of PRL distribution.

τ	E(X)	Var(X)	CS(X)	CK(X)	ID(X)	CV(X)
2	4.00000	12.0000	1.44338	6.08333	3.00000	0.86603
3	4.50000	20.2500	1.67901	7.05761	4.50000	1.00000
4	5.33333	30.2222	1.79405	7.66025	5.66667	1.03078
5	6.25000	42.1875	1.85607	8.02222	6.75000	1.03923
6	7.20000	56.1600	1.89348	8.25493	7.80000	1.04083
7	8.16667	72.1389	1.91786	8.41326	8.83333	1.04002
8	9.14286	90.1224	1.93468	8.52586	9.85714	1.03833
9	10.1250	110.1094	1.94678	8.60883	10.87500	1.03638
10	11.1111	132.0988	1.95579	8.67174	11.88889	1.03441
15	16.0714	272.0663	1.97871	8.83698	16.92857	1.02632
20	21.0526	462.0499	1.98750	8.90278	21.94737	1.02103

4. Parameter estimation

In this section, the parameter of PRL distribution is examined using some classical estimation approaches. The considered estimation approaches are maximum likelihood, Anderson-Darling, Cramer von Mises, least squares, and weighted least squares.

4.1. Maximum likelihood estimation

Let $X_1, X_2, X_3, \ldots X_n$ be a random sample of failure times from PRL distribution, and the likelihood function for the parameter τ can be written as

$$L(\tau|x) = \prod_{i=1}^{n} \frac{\left(1+\frac{1}{\tau}\right)^{-x_i}(x_i - 1 + \tau(\tau-1))}{(\tau-1)(1+\tau)^2}, \quad (16)$$

and log-likelihood function is specified by

$$l(\tau|x) = \sum_{i=1}^{n} \log\left(1+\frac{1}{\tau}\right)^{-x_i} + \sum_{i=1}^{n} \log(x_i - 1 + \tau(\tau-1)) - n\log(\tau-1) - n\log(1+\tau)^2. \quad (17)$$

We get the following equation by deriving Eq (17) with regard to parameter τ:

$$\frac{\partial l}{\partial \tau} = \sum_{i=1}^{n} \frac{x_i}{\left(1+\frac{1}{\tau}\right)\tau^2} + \sum_{i=1}^{n} \frac{2\tau-1}{x_i+\tau(\tau-1)-1} - \frac{n}{(\tau-1)} - \frac{2n}{(\tau+1)}. \quad (18)$$

The ML estimate is obtained by equating the above equation to zero and solving it for parameter τ. However, the ensuing expression has not a closed-form result and the required results can be obtained using iterative procedures.

4.2. Anderson darling estimation

The Anderson-Darling (AD) estimator $\hat{\tau}$ of parameter τ can be defined by minimizing the following expression

$$\mathrm{AD}(\tau) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log\left(F\left(x_{(i:n)}|\tau\right)\right) + \log\left(1 - F\left(x_{(i:n)}|\tau\right)\right)\right],$$

$$\mathrm{AD}(\tau) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log\left(1 - \frac{\left(1+\frac{1}{\tau}\right)^{-x_{(i:n)}}\tau\left(x_{(i:n)}+\tau^2\right)}{(\tau-1)(1+\tau)^2}\right) + \log\left(\frac{\left(1+\frac{1}{\tau}\right)^{-x_{(i:n)}}\tau\left(x_{(i:n)}+\tau^2\right)}{(\tau-1)(1+\tau)^2}\right)\right],$$

Alternatively, the estimator can also be obtained by solving the following nonlinear equation

$$\sum_{i=1}^{n}(2i-1)\left[\frac{\phi\left(x_{(i:n)}|\tau\right)}{F\left(x_{(i:n)}|\tau\right)} - \frac{\phi\left(x_{(n+1-i:n)}|\tau\right)}{1 - F\left(x_{(n+1-i:n)}|\tau\right)}\right] = 0$$

where $\phi(x_{i:n}|\tau) = \frac{d}{d\tau}F\left(x_{(i:n)}|\tau\right)$ and it reduces to

$$\phi(x_{i:n}|\tau) = \frac{\left(1+\frac{1}{\tau}\right)^{-x_{(i:n)}}\left(-x_{(i:n)}{}^2(\tau-1) - (\tau-3)\tau^2 - x_{(i:n)}\left(-1+\tau-3\tau^2+\tau^3\right)\right)}{(\tau-1)^2(1+\tau)^3} \quad (19)$$

4.3. Ordinary least squares estimation

The ordinary least-square (OLS) estimator of the PRL model parameter can be obtained by minimizing

$$\mathrm{LSE}(\tau) = \sum_{i=1}^{n}\left[F\left(x_{(i:n)}|\tau\right) - \frac{i}{n+1}\right]^2,$$

with respect to the parameter τ. Moreover, the LSE of τ is also obtained by solving

$$\sum_{i=1}^{m}\left[1-\frac{i}{1+n}-\frac{\left(1+\frac{1}{\tau}\right)^{-x_{(i:n)}}\tau\left(x_{(i:n)}+\tau^{2}\right)}{(\tau-1)(1+\tau)^{2}}\right]\phi(x_{i:n}|\tau)=0,$$

4.4. Weighted least-square estimation

The WLS estimate (WLSE) of τ, say $\hat{\tau}$, can be determined by minimizing

$$\mathrm{WLSE}(\tau)=\sum_{i=1}^{n}\frac{(n+1)^{2}(n+2)}{i(n-i+1)}\left[F\left(x_{(i:n)}|\tau\right)-\frac{i}{n+1}\right]^{2},$$

with respect to τ. The WLSE of τ can also be obtained by solving

$$\sum_{i=1}^{n}\frac{(1+n)^{2}\,(2+n)}{i(n-i+1)}\left[1-\frac{i}{1+n}-\frac{\left(1+\frac{1}{\tau}\right)^{-x_{(i:n)}}\tau\left(x_{(i:n)}+\tau^{2}\right)}{(\tau-1)(1+\tau)^{2}}\right]\phi(x_{i:n}|\tau)=0,$$

In which $\phi(x_{i:n}|\tau)$ is presented in (19).

4.5. Cramer Von-Misses estimator

The Cramer von Mises (CVM) is a minimum distance-based estimator. The CVM of the PRL distribution can be obtained by minimizing

$$\mathrm{CVM}(\tau)=\frac{1}{12n}+\sum_{i=1}^{n}\left[\log\left(F\left(x_{(i:n)}|\tau\right)\right)\ -\frac{2i-1}{2n}\right]^{2},$$

with respect to the parameter τ.

The CVME of τ is also obtained by solving

$$\sum_{i=1}^{n}\left[1-\frac{2i-1}{2\,n}-\frac{\left(1+\frac{1}{\tau}\right)^{-x_{(i:n)}}\tau\left(x_{(i:n)}+\tau^{2}\right)}{(\tau-1)(1+\tau)^{2}}\right]\phi(x_{i:n}|\tau)=0.$$

4.6. Simulation

In this section, we performed a simulation study to evaluate the accuracy of all considered estimators. In the simulation run, we generate 10,000 samples of size n = 10, 25, 50, 100, 200, and 300 from PRL distribution and then calculate the average estimates (AE), absolute bias (AB), mean relative error (MRE) and mean square error (MSE). For this purpose, we consider the six sets of values of parameter τ. The simulation results are presented in Tables 2–7.

Table 2. Parameter Estimates based on simulated samples for the parameter $\tau = 2.1$.

Measures	n	MLE	OLSE	WLSE	ADE	CVME
AE	10	2.4683	3.0734	2.7620	3.0486	2.7156
	25	2.3209	2.4309	2.1472	2.5131	2.3049
	50	2.2357	2.1799	2.1001	2.2501	2.1488
	100	2.1808	2.1098	2.1000	2.1307	2.1045
	200	2.1416	2.1002	2.1000	2.1013	2.1004
	300	2.1274	2.1000	2.1000	2.1002	2.1000
AB	10	0.3683	0.9734	0.6620	0.9486	0.6156
	25	0.2209	0.3309	0.0472	0.4131	0.2049
	50	0.1357	0.0799	0.0001	0.1501	0.0488
	100	0.0808	0.0098	0.0000	0.0307	0.0045
	200	0.0416	0.0002	0.0000	0.0013	0.0004
	300	0.0274	0.0000	0.0000	0.0002	0.0000
MRE	10	0.1754	0.4635	0.3152	0.4517	0.2931
	25	0.1052	0.1576	0.0225	0.1967	0.0976
	50	0.0646	0.0381	0.0001	0.0715	0.0233
	100	0.0385	0.0046	0.0000	0.0146	0.0021
	200	0.0198	0.0001	0.0000	0.0006	0.0002
	300	0.0130	0.0000	0.0000	0.0001	0.0000
MSE	10	0.7743	3.3333	2.0870	3.4064	2.3300
	25	0.2951	0.8431	0.0912	1.1611	0.5520
	50	0.1391	0.1614	0.0001	0.3712	0.0985
	100	0.0627	0.0157	0.0000	0.0651	0.0068
	200	0.0271	0.0003	0.0000	0.0025	0.0006
	300	0.0169	0.0000	0.0000	0.0006	0.0000

Table 3. Parameter Estimates based on simulated samples for the parameter $\tau = 3.0$.

Measures	n	MLE	OLSE	WLSE	ADE	CVME
AE	10	3.1396	4.3862	4.3030	4.1871	4.2023
	25	3.0268	4.0219	3.9440	3.8787	3.9351
	50	2.9950	3.9277	3.9411	3.7879	3.8614
	100	2.9948	3.9041	4.0457	3.7762	3.8804
	200	2.9964	3.9231	4.2116	3.8175	3.9187
	300	2.9970	3.9378	4.3245	3.8419	3.9348
AB	10	0.1396	1.3862	1.3030	1.1871	1.2023
	25	0.0268	1.0219	0.9440	0.8787	0.9351
	50	0.0050	0.9277	0.9411	0.7879	0.8614
	100	0.0052	0.9041	1.0457	0.7762	0.8804
	200	0.0036	0.9231	1.2116	0.8175	0.9187
	300	0.0030	0.9378	1.3245	0.8419	0.9348

Continued on next page

Measures	n	MLE	OLSE	WLSE	ADE	CVME
MRE	10	0.0465	0.5763	0.5672	0.5519	0.5599
	25	0.0089	0.4499	0.4713	0.4301	0.4473
	50	0.0017	0.3944	0.4427	0.3736	0.3907
	100	0.0017	0.3580	0.4347	0.3302	0.3570
	200	0.0012	0.3318	0.4495	0.3062	0.3321
	300	0.0010	0.3237	0.4659	0.2977	0.3217
MSE	10	1.5854	5.3423	4.9727	4.6497	4.8640
	25	0.7594	2.7425	2.7410	2.4596	2.6652
	50	0.4343	1.9387	2.1702	1.7096	1.8726
	100	0.2391	1.4819	1.9558	1.2643	1.4561
	200	0.1248	1.1972	1.9742	1.0271	1.1933
	300	0.0832	1.0952	2.0686	0.9387	1.0861

Table 4. Parameter Estimates based on simulated samples for the parameter $\tau = 4.0$.

Measures	n	MLE	OLSE	WLSE	ADE	CVME
AE	10	3.9375	5.2163	5.2168	5.1234	5.0765
	25	3.9262	4.9401	4.9726	4.8907	4.8748
	50	3.9620	4.9088	5.0659	4.8573	4.8859
	100	3.9713	4.8963	5.1787	4.8401	4.8697
	200	3.9834	4.8858	5.3031	4.8439	4.8789
	300	3.9988	4.8862	5.3711	4.8506	4.8756
AB	10	0.0625	1.2163	1.2168	1.1234	1.0765
	25	0.0738	0.9401	0.9726	0.8907	0.8748
	50	0.0380	0.9088	1.0659	0.8573	0.8859
	100	0.0287	0.8963	1.1787	0.8401	0.8697
	200	0.0166	0.8858	1.3031	0.8439	0.8789
	300	0.0012	0.8862	1.3711	0.8506	0.8756
MRE	10	0.0156	0.4834	0.4807	0.4714	0.4854
	25	0.0185	0.3538	0.3585	0.3405	0.3536
	50	0.0095	0.2878	0.3167	0.2742	0.2889
	100	0.0072	0.2466	0.3054	0.2321	0.2424
	200	0.0042	0.2264	0.3260	0.2157	0.2256
	300	0.0003	0.2229	0.3428	0.2138	0.2205
MSE	10	2.9043	6.2750	6.0663	5.8402	6.1414
	25	1.3856	3.1215	3.0986	2.8793	3.0677
	50	0.7547	1.9790	2.2677	1.8032	2.0032
	100	0.3717	1.3758	1.8782	1.2184	1.3353
	200	0.1822	1.0527	1.9039	0.9610	1.0504
	300	0.1202	0.9624	2.0149	0.8867	0.9453

Table 5. Parameter Estimates based on simulated samples for the parameter $\tau = 5.0$.

Measures	n	MLE	OLSE	WLSE	ADE	CVME
AE	10	4.8793	6.2159	6.2417	6.1934	6.0518
	25	4.9283	5.9406	6.0539	5.9254	5.8979
	50	4.9480	5.9061	6.0826	5.8767	5.8738
	100	4.9839	5.8739	6.1814	5.8623	5.8668
	200	4.9940	5.8508	6.2729	5.8430	5.8573
	300	4.9858	5.8588	6.3355	5.8436	5.8443
AB	10	0.1207	1.2159	1.2417	1.1934	1.0518
	25	0.0717	0.9406	1.0539	0.9254	0.8979
	50	0.0520	0.9061	1.0826	0.8767	0.8738
	100	0.0161	0.8739	1.1814	0.8623	0.8668
	200	0.0060	0.8508	1.2729	0.8430	0.8573
	300	0.0142	0.8588	1.3355	0.8436	0.8443
MRE	10	0.0241	0.4372	0.4343	0.4289	0.4335
	25	0.0143	0.2971	0.3010	0.2896	0.2975
	50	0.0104	0.2376	0.2514	0.2249	0.2334
	100	0.0032	0.1979	0.2415	0.1921	0.1968
	200	0.0012	0.1769	0.2549	0.1739	0.1775
	300	0.0028	0.1737	0.2671	0.1705	0.1714
MSE	10	4.5278	8.0561	7.8298	7.6115	7.7488
	25	1.9727	3.5841	3.6415	3.3645	3.5746
	50	0.9619	2.1944	2.3526	1.9809	2.1332
	100	0.4770	1.4498	1.9140	1.3619	1.4272
	200	0.2355	1.0642	1.8717	1.0091	1.0642
	300	0.1600	0.9522	1.9495	0.9070	0.9279

Table 6. Parameter Estimates based on simulated samples for the parameter $\tau = 7.0$.

Measures	n	MLE	OLSE	WLSE	ADE	CVME
AE	10	6.8852	8.3751	8.3080	8.2719	8.0785
	25	6.9508	7.9654	8.0939	7.9661	7.9015
	50	6.9772	7.9105	8.0665	7.8996	7.8877
	100	6.9808	7.8671	8.1001	7.8491	7.8347
	200	6.9902	7.8377	8.2073	7.8376	7.8279
	300	6.9994	7.8330	8.2567	7.8323	7.8247
AB	10	0.1148	1.3751	1.3080	1.2719	1.0785
	25	0.0492	0.9654	1.0939	0.9661	0.9015
	50	0.0228	0.9105	1.0665	0.8996	0.8877
	100	0.0192	0.8671	1.1001	0.8491	0.8347
	200	0.0098	0.8377	1.2073	0.8376	0.8279
	300	0.0006	0.8330	1.2567	0.8323	0.8247

Continued on next page

Measures	n	MLE	OLSE	WLSE	ADE	CVME
MRE	10	0.0164	0.3971	0.3834	0.3728	0.3801
	25	0.0070	0.2536	0.2487	0.2396	0.2494
	50	0.0033	0.1922	0.1930	0.1835	0.1908
	100	0.0027	0.1539	0.1695	0.1468	0.1502
	200	0.0014	0.1307	0.1741	0.1286	0.1298
	300	0.0001	0.1240	0.1797	0.1228	0.1229
MSE	10	8.0853	13.221	12.275	11.590	11.818
	25	3.1995	5.2182	4.9697	4.6070	5.0214
	50	1.5333	2.9054	2.8651	2.6487	2.8488
	100	0.7477	1.7950	2.0349	1.6308	1.7225
	200	0.3819	1.2139	1.8654	1.1544	1.1984
	300	0.2538	1.0264	1.8495	0.9997	1.0176

Table 7. Parameter Estimates based on simulated samples for the parameter $\tau = 15.0$.

Measures	n	MLE	OLSE	WLSE	ADE	CVME
AE	10	14.870	16.725	16.755	16.639	16.487
	25	14.975	16.192	16.194	16.166	16.071
	50	15.003	15.979	16.045	16.015	15.948
	100	15.014	15.881	16.023	15.876	15.818
	200	14.992	15.818	16.066	15.853	15.801
	300	15.014	15.810	16.081	15.826	15.813
AB	10	0.1302	1.7254	1.7551	1.6394	1.4871
	25	0.0254	1.1917	1.1941	1.1658	1.0705
	50	0.0026	0.9791	1.0445	1.0145	0.9477
	100	0.0142	0.8806	1.0226	0.8755	0.8180
	200	0.0084	0.8184	1.0661	0.8532	0.8008
	300	0.0135	0.8103	1.0806	0.8263	0.8125
MRE	10	0.0087	0.3328	0.3258	0.3120	0.3262
	25	0.0017	0.2138	0.2007	0.1975	0.2060
	50	0.0002	0.1496	0.1444	0.1445	0.1491
	100	0.0009	0.1114	0.1076	0.1043	0.1085
	200	0.0006	0.0832	0.0879	0.0814	0.0824
	300	0.0009	0.0733	0.0817	0.0710	0.0729
MSE	10	27.483	42.816	40.605	37.786	40.567
	25	11.128	16.789	14.977	14.517	15.753
	50	5.4177	8.2091	7.5370	7.6602	8.2052
	100	2.7917	4.5139	4.1585	3.9290	4.2486
	200	1.3835	2.4799	2.6592	2.3612	2.4262
	300	0.8964	1.8851	2.2170	1.7610	1.8652

4.7. Bayesian analysis

The Bayesian parameter estimation technique is an alternate to classical maximum likelihood estimation. In Bayesian estimation, a prior distribution must be defined for each unknown parameter. Consider a set of data $x = x_1, x_2, \dots, x_n$ taken from discrete PRL distribution and the likelihood function is provided by

$$L(\tau|x) = \prod_{i=1}^{n} \frac{\left(1+\frac{1}{\tau}\right)^{-x_i}(x_i-1+\tau(\tau-1))}{(\tau-1)(1+\tau)^2}. \tag{20}$$

The Bayesian model is constructed by stating the prior distribution for the model parameter and then multiplying it with the likelihood function for the provided data using the Bayes theorem to generate the posterior distribution function. The prior distribution of parameter τ is denoted as $p(\tau)$.

$$p(\tau|x) \propto L(\tau|x)p(\tau).$$

For the proposed distribution, the gamma distribution is considered a prior distribution with known hyperparameters such as $\tau \sim Gamma(\alpha, \beta)$. The posterior expression, up to proportionality, may be found by multiplying the likelihood by the prior, and this can be represented as

$$p(\tau|x) \propto \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} \exp(-\tau\beta) \prod_{i=1}^{n} \frac{\left(1+\frac{1}{\tau}\right)^{-x_i}(x_i-1+\tau(\tau-1))}{(\tau-1)(1+\tau)^2}$$

The posterior density is not mathematically tractable; for inference purposes, we will utilize the Markov Chain Monte Carlo (MCMC) approach to mimic posterior samples, allowing for easy sample-based conclusions.

In the present study, we explore the application of MCMC algorithms implemented in the package MCMCpack of the R program to simulate samples from the joint posterior distribution. For this purpose, we generated 1006000 samples of the joint posterior distribution of interest. The effects of the initial values in the iterative process are eliminated after a burn-in phase of 6000 simulated samples. To achieve approximately independent samples, a thinning interval of size 300 was utilized. The parameter Bayes estimates were gained by taking the expected value of generated samples. Traceplots and the Geweke diagnostic were used to monitor the convergence of the simulated sequences. The asymptotic standard error of the difference divided by the difference between the two means of non-overlapping parts of a simulated Markov chain is the basis of the Geweke convergence diagnostic. We may say that a chain has reached convergence if its corresponding absolute z score is smaller than 1.96 since this z score asymptotically follows a typical normal distribution. The construction of interesting posterior summaries was done using the R software package MCMCpack.

5. Application

This section is ardent to prove the usefulness of the discrete Poisson Ramos-Louzada distribution in the modeling of three datasets. We compare the fits of the proposed distribution with some renowned one-parameter discrete distributions, discrete Raleigh [12], Poisson, discrete Pareto [13] and discrete Burr-Hatke [14], discrete Inverted Topp-Leone [15]. The Kolmogorov-Smirnov (KS) test, Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are used to compare the fitted models. We also illustrate the estimation procedures based on censored samples proposed in the previous section with three examples from the literature.

5.1. Data I (Failure times of electronic components)

A sample of the failure time of 15 electronic components in an acceleration life test [16]. The observations are 1, 5, 6, 11, 12, 19, 20, 22, 23, 31, 37, 46, 54, 60, and 66. The mean and variance of the first dataset are 27.533 and 431.94 respectively. The dispersion index value is 15.689 which indicates that the dataset is overdispersed. We determine the MLEs, standard errors (SE), and model selection measures (AIC, BIC, and KS) for the first dataset using the R software's maxLik package. These results are shown in Table 8 along with the model selection measures.

Table 8. ML Estimates and goodness-of-fit for the first dataset.

Model	MLEs (S.E.)	-LogLik.	AIC	BIC	K-S	P-value
PRL	26.455 (7.2429)	64.995	131.99	132.70	0.1770	0.6700
DR	24.382 (3.1481)	66.394	134.79	135.50	0.2160	0.4300
Poisson	27.533 (1.3548)	151.21	304.41	305.12	0.3810	0.0180
DITL	0.4178 (0.1079)	74.491	150.98	151.69	0.3590	0.0310
DPr	0.3284 (0.0848)	77.402	156.80	157.51	0.4060	0.0097
DBH	0.9992 (0.0076)	91.368	184.74	185.44	0.7910	0.0000

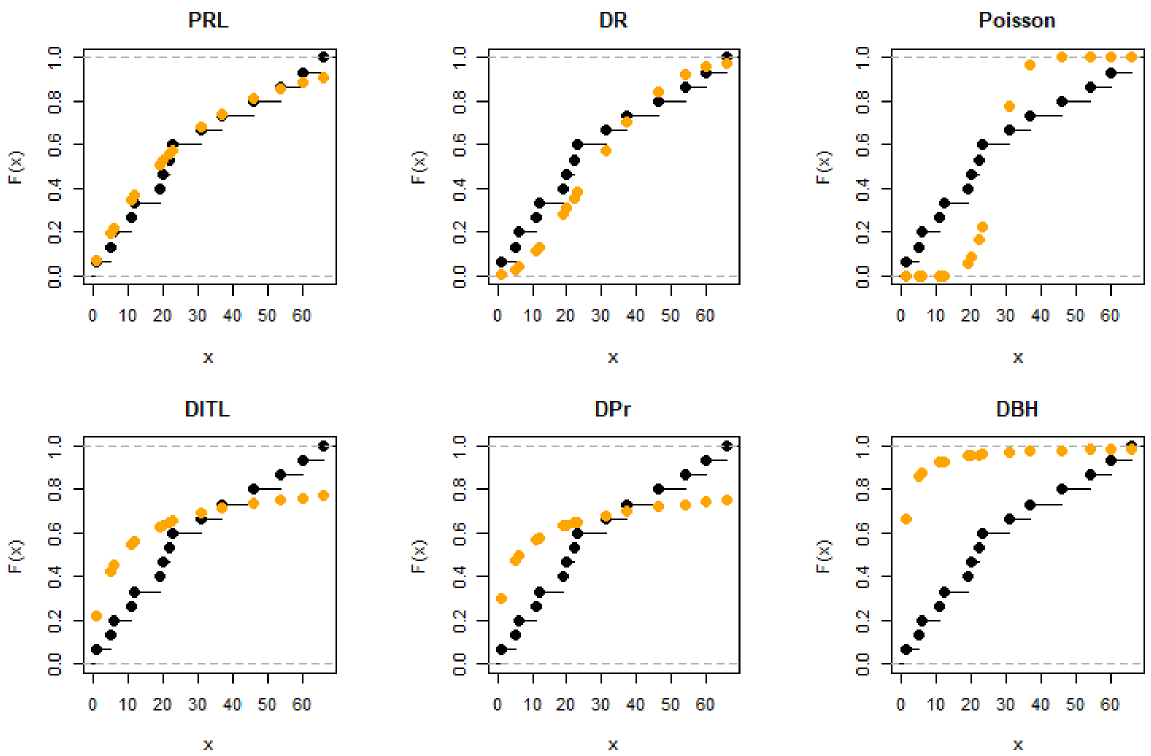

Figure 3. Plots of fitted CDFs versus empirical CDFs for the first dataset.

For Bayesian data analysis, the parameter τ of the PRL distribution was assumed to have an approximate gamma as the prior distribution, that is, $\tau \sim Gamma(0.001, 0.1)$. Figure 4 depicts posterior samples for the parameter τ. The evaluation of the MCMC draws across iterations is assessed using traceplot, posterior density, and ACF plot. From the traceplot, it is interesting to note that the samples produced attained acceptable convergence. The ACF plot indicates that the posterior samples are uncorrelated. Furthermore, the z-score of the Geweke test is –0.2498, indicating that the samples have sufficiently converged to a stable distribution. The posterior mean for τ is $\tau_{Bayes} = 13.00418$ with a standard deviation of 2.18641, and the corresponding 95% highest density interval is (9.008356, 17.3976). We observe that the ML and Bayesian estimates are quite similar.

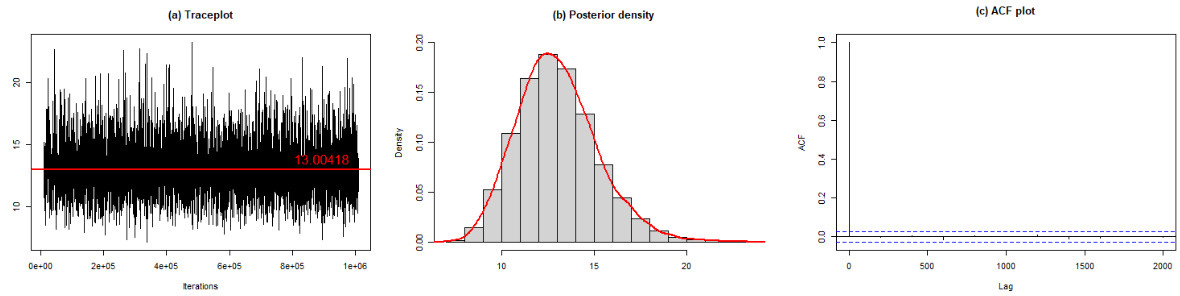

Figure 4. Traceplot, Posterior density, and ACF plot based on the first dataset.

5.2. Data II (COVID-19 Deaths in China)

A sample of 66 patients died due to COVID-19 in China from January 23, 2022, to March 28, 2020. The data are: 8, 16, 15, 24, 26, 26, 38, 43, 46, 45, 57, 64, 65, 73, 73, 86, 89, 97, 108, 97, 146, 121, 143, 142, 105, 98, 136, 114, 118, 109, 97, 150, 71, 52, 29, 44, 47, 35, 42, 31, 38, 31, 30, 28, 27, 22, 17, 22, 11, 7, 13, 10, 14, 13, 11, 8, 3, 7, 6, 9, 7, 4, 6, 5, 3 and 5. Some descriptive measures (mean, variance, and dispersion index) for this dataset are 47.742, 1924.8, and 38.696. We acquire the ML estimates for the parameter, and model selection metrics (AIC, BIC, and KS) for the second dataset. These results are shown in Table 9.

Table 9. ML Estimates and goodness-of-fit for the second dataset.

Model	MLEs (S.E.)	-LogLik.	AIC	BIC	K-S	P-value
PRL	48.711 (6.1847)	324.51	651.02	653.21	0.0851	0.7300
DR	47.010 (2.8934)	347.23	696.45	698.64	0.2930	0.0000
Poisson	49.743 (0.8682)	1409.8	2821.6	2823.8	0.4970	0.0000
DITL	0.3539 (0.0436)	366.91	735.81	738.00	0.3290	0.0000
DPr	0.2863 (0.0352)	379.07	760.14	762.33	0.3820	0.0000
DBH	0.9997 (0.0019)	461.02	924.04	926.23	0.8120	0.0000

For Bayesian data analysis, the parameter tau of the PRL distribution was assumed to have a gamma prior distribution. The associated Geweke z-score is –0.08203, which likewise indicates that the samples have sufficiently converged to a stable distribution. The posterior mean for τ is $\tau_{Bayes} = 32.0684$ with a standard deviation of 2.89397, and a 95% HDI of (26.20931, 37.44432). The ML and Bayesian estimates are discernibly similar to one another.

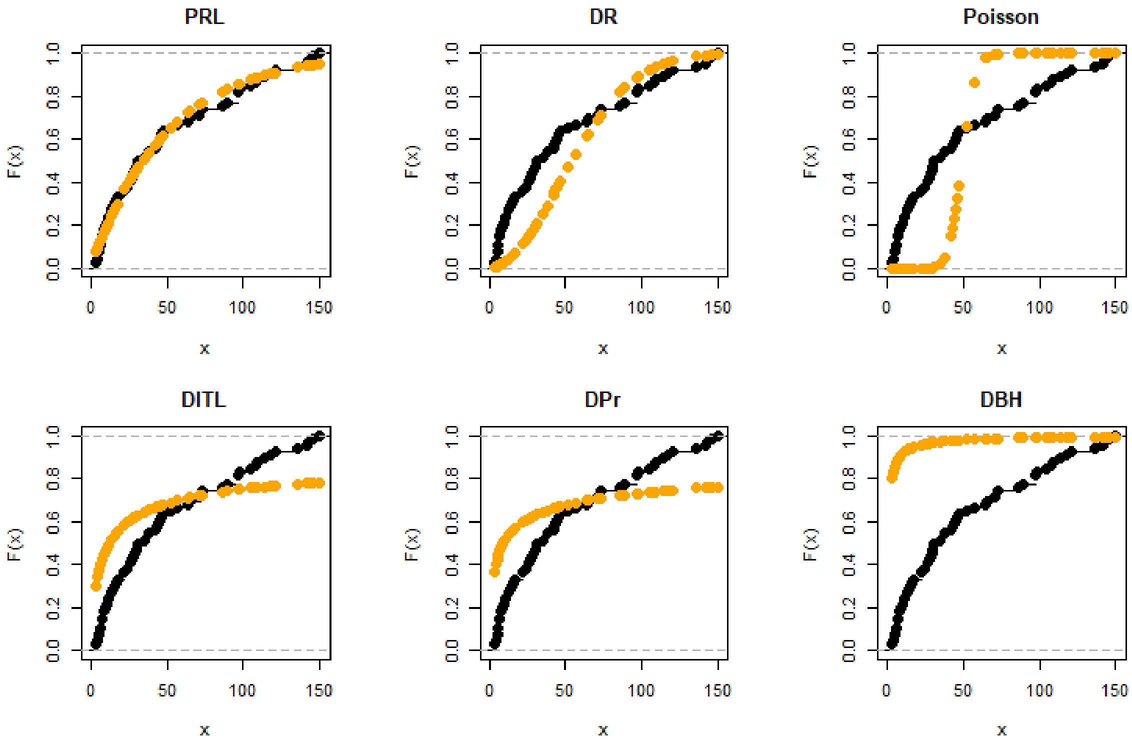

Figure 5. Plots of fitted CDFs versus empirical CDFs for the second dataset.

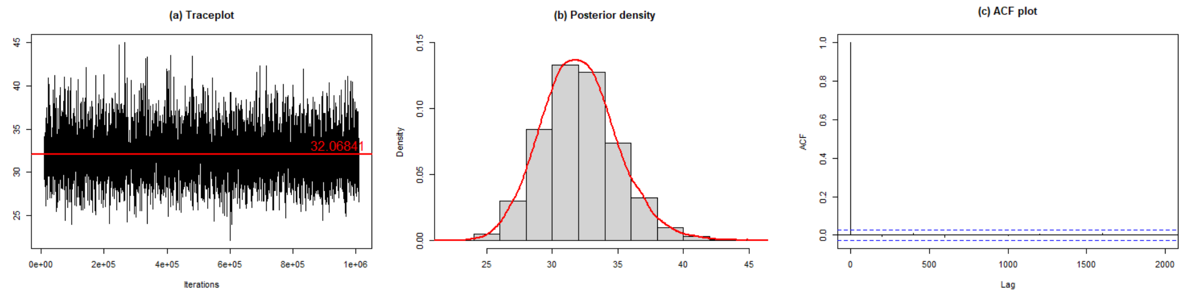

Figure 6. Traceplot, Posterior density, and ACF plot based on the second dataset.

5.3. Data set III (Deaths due to COVID-19 in Pakistan)

The third dataset is also about deaths due to COVID-19 in Pakistan from 18 March 2020 to 30 June 2020. The data are: 1, 6, 6, 4 , 4, 4, 1, 20, 5, 2, 3, 15, 17, 7, 8, 25, 8, 25, 11, 25, 16, 16, 12, 11, 20, 31, 42, 32, 23, 17, 19, 38, 50, 21, 14, 37, 23, 47, 31, 24, 9, 64, 39, 30, 36, 46, 32, 50, 34, 32, 34, 30, 28, 35, 57, 78, 88, 60, 78, 67, 82, 68, 97, 67, 65, 105, 83 , 101, 107, 88, 178, 110, 136, 118, 136, 153, 119, 89, 105, 60, 148, 59, 73, 83, 49, 137 and 91. Some computational measures, mean, variance and index of dispersion for the third dataset are; 50.057, 1758.8, and 35.135. The MLEs and goodness-of-fit measures for this dataset are given in Table 10.

Table 10. ML Estimates and goodness-of-fit for the third dataset.

Model	MLEs (S.E.)	-LogLik.	AIC	BIC	K-S	P-value
PRL	49.020 (5.4201)	428.30	858.61	861.07	0.0676	0.8210
DR	46.339 (2.4841)	452.55	907.10	909.56	0.2473	0.0000
Poisson	50.058 (0.9742)	1713.0	3428.1	3430.5	0.4954	0.0000
DITL	0.3493 (0.0375)	488.14	978.28	980.75	0.3263	0.0000
DPr	0.2835 (0.0304)	503.61	1009.2	1011.7	0.3558	0.0000
DBH	0.9997 (0.0016)	613.80	1229.6	1232.1	0.7876	0.0000

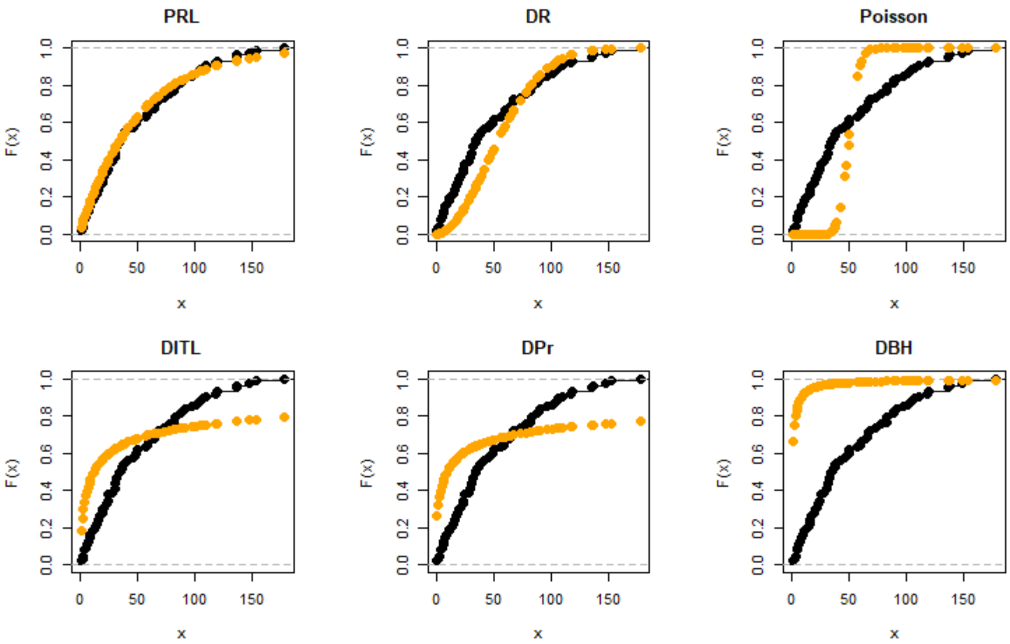

Figure 7. Plots of fitted CDFs versus empirical CDFs for the third dataset.

For the third dataset, the gamma distribution is again considered as the prior distribution, and the posterior samples for the parameter are described in Figure 8. Furthermore, the Geweke z-score is used as a diagnostic measure and its value is –0.03794, suggesting convergence of the samples to a stable distribution. The posterior mean for the third dataset is $\tau_{Bayes} = 46.96159$ with a standard deviation of 4.92385. The corresponding 95% HDI (37.94273, 57.07319). The ML and Bayes estimate is quite similar to each other.

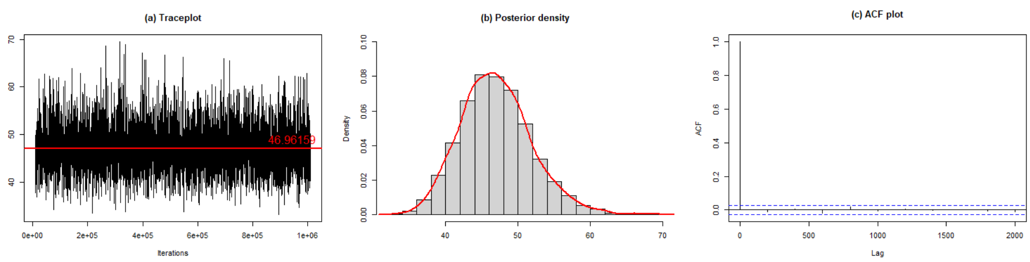

Figure 8. Traceplot, Posterior density, and ACF plot based on the third dataset.

6. Conclusions

In this paper, we introduce a one-parameter discrete distribution by compounding Poisson with the Ramos-Louzada distribution. The proposed distribution is showing unimodal and positively skewed behavior. The failure rate of new distribution is increasing pattern. Some statistical properties derived include the moment-generating function, probability-generating function, factorial moments, dispersion index, skewness and kurtosis. The model parameter is estimated using the maximum likelihood estimation approach and the behavior of the derived estimator is assessed via a simulation study. The usefulness of the proposed distribution is carried out using three real-life datasets. The proposed distribution provides more efficient results than all considered competitive distributions. The Bayesian analysis is also performed by taking the MCMC approximation approach.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. M. Shoukri, M. H. Asyali, R. VanDorp, D. Kelton, The Poisson inverse Gaussian regression model in the analysis of clustered counts data, *J. Data Sci.*, **2** (2004), 17–32. https://doi.org/10.6339/JDS.2004.02(1).135
2. G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle, P. Boatwright, A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution, *J. R. Stat. Soc. Ser. C.*, **54** (2005), 127–142. https://doi.org/10.1111/j.1467-9876.2005.00474.x
3. E. Mahmoudi, H. Zakerzadeh, Generalized poisson–lindley distribution, *Commun. Stat. Methods*, **39** (2010), 1785–1798. https://doi.org/10.1080/03610920902898514
4. L. Cheng, S. R. Geedipally, D. Lord, The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data, *Saf. Sci.*, **54** (2013), 38–42. https://doi.org/10.1016/j.ssci.2012.11.002
5. H. Hassan, S. A. Dar, P. B. Ahmad, Poisson Ishita distribution: A new compounding probability model, *IOSR J. Eng.*, **9** (2019), 38–46.
6. E. Altun, A new model for over-dispersed count data: Poisson quasi-Lindley regression model, *Math. Sci.*, **13** (2019), 241–247. https://doi.org/10.1007/s40096-019-0293-5
7. B. A. Para, T. R. Jan, H. S. Bakouch, Poisson Xgamma distribution: A discrete model for count data analysis, *Model Assist. Stat. Appl.*, **15** (2020), 139–151. https://doi.org/10.3233/MAS-200484
8. E. Altun, G. M. Cordeiro, M. M. Ristić, An one-parameter compounding discrete distribution, *J. Appl. Stat*., **49** (2022), 1935–1956. https://doi.org/10.1080/02664763.2021.1884846

9. M. Ahsan-ul-Haq, A. Al-bossly, M, El-morshedy, M. S. Eliwa, Poisson XLindley distribution for count data : Statistical and reliability properties with estimation techniques and inference, *Comput. Intell. Neurosci.*, **2022** (2022). https://doi.org/10.1155/2022/6503670
10. M. Ahsan-ul-Haq, On poisson moment exponential distribution with applications, *Ann. Data Sci.*, 2022. https://doi.org/10.1007/s40745-022-00400-0
11. P. L. Ramos, F. Louzada, A Distribution for instantaneous failures, *Stats*, **2** (2019), 247–258. https://doi.org/10.3390/stats2020019
12. D. Roy, Discrete rayleigh distribution, *IEEE Trans. Reliab.*, **53** (2004), 255–260. https://doi.org/10.1109/TR.2004.829161
13. H. Krishna, P. S. Pundir, Discrete Burr and discrete Pareto distributions, *Stat. Methodol.*, **6** (2009), 177–188. https://doi.org/10.1016/j.stamet.2008.07.001
14. M. El-Morshedy, M. S. Eliwa, E. Altun, Discrete Burr-Hatke distribution with properties, estimation methods and regression model, *IEEE Access*, **8** (2020), 74359–74370. https://doi.org/10.1109/ACCESS.2020.2988431
15. A. S. Eldeeb, M. Ahsan-ul-Haq, A. Babar, A discrete analog of inverted Topp-Leone distribution: Properties, estimation and applications. *Int. J. Anal. Appl.*, **19** (2021), 695–708. https://doi.org/10.28924/2291-8639-19-2021-695
16. J. F. Lawless, *Statistical models and methods for lifetime data*, John Wiley & Sons, 2011.