Research article

A discrete Ramos-Louzada distribution for asymmetric and over-dispersed data with leptokurtic-shaped: Properties and various estimation techniques with inference

Ahmed Sedky Eldeeb[1,2,*], Muhammad Ahsan-ul-Haq[3,4] and Mohamed S. Eliwa[5]

[1] Department of Business Administration, College of Business, King Khalid University, Saudi Arabia
[2] Department of Statistics, Mathematics and Insurance, Alexandria University, Egypt
[3] College of Statistical & Actuarial Sciences, University of the Punjab, Lahore, Pakistan
[4] Quality Enhancement Cell, National College of Arts, Lahore, Pakistan
[5] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

* **Correspondence:** Email: asedky@kku.edu.sa.

Abstract: In this paper, a flexible probability mass function is proposed for modeling count data, especially, asymmetric, and over-dispersed observations. Some of its distributional properties are investigated. It is found that all its statistical properties can be expressed in explicit forms which makes the proposed model can be utilized in time series and regression analysis. Different estimation approaches including maximum likelihood, moments, least squares, Anderson's-Darling, Cramer von-Mises, and maximum product of spacing estimator, are derived to get the best estimator for the real data. The estimation performance of these estimation techniques is assessed via a comprehensive simulation study. The flexibility of the new discrete distribution is assessed using four distinctive real data sets "coronavirus-flood peaks-forest fire-Leukemia". Finally, the new probabilistic model can serve as an alternative distribution to other competitive distributions available in the literature for modeling count data.

1. Introduction

The count data sets emerge in various fields like the yearly number of destructive earthquakes, number of patients of a specific disease in a hospital ward, failure of machines, number of patients due to coronavirus, number of monthly traffic accidents, hourly bacterial growth, and so on. Various discrete probability models have been utilized to model these kinds of data sets. Poisson and negative binomial distributions are frequently for modeling count observations. On the other hand, in the advanced scientific eon, the data generated from different fields is getting complex day by day, however, existing discrete models do not provide an efficient fit.

Discretization of continuous distribution can be applied by using different approaches (survival discretization-mixed-Poisson-infinite series). The most widely used technique is the survival discretization approach by [1]. In literature, several discrete distributions are introduced and studied based on the discretization survival function. See for example, discrete Rayleigh [2], discrete a mixture of gamma and exponential [3], discrete Burr and Pareto [4], discrete Rayleigh generator [5], discrete Lindley [6], discrete Burr-XII [7], discrete Burr III [8], two-parameter discrete Lindley [9], discrete log-logistic [10], discrete Gompertz [11], discrete alpha power inverse Lomax [12], discrete Poisson and Ailamujia [13], discrete Half-Logistic [14], discrete Marshall-Olkin Weibull [15], discrete Gompertz-G family [16], discrete Burr-Hatke [17], three-parameter discrete Lindley [18], new discrete Lindley [19], exponentiated discrete Lindley [20], discrete generalized Lindley [21], discrete Gumble [22], discrete inverted Topp-Leone [23], and references cited therein.

Although various distributions are available in literate to analyze count observations, there is still a need to introduce a more flexible and suitable distribution under different conditions. The fundamental purpose of this paper is to propose discrete Ramos-Louzada distribution, which is a one-parameter lifetime distribution introduced by [24]. The proposed one-parameter distribution herein has distinctive properties which makes it among the best choice for modeling over-dispersed and positively skewed data with leptokurtic-shaped. A continuous random variable X is said to have Ramos-Louzada distribution if its probability density function (pdf) can be written as

$$g(x; \lambda) = \frac{1}{\lambda^2(\lambda-1)}(\lambda^2 - 2\lambda + x)\, e^{-\frac{x}{\lambda}}; \ x \geq 0, \lambda \geq 2, \tag{1}$$

where λ is the shape parameter. The corresponding survival function (sf) to Eq (1) can be formulated as

$$G(x; \lambda) = \frac{\lambda^2 - \lambda + x}{\lambda(\lambda-1)} e^{-\frac{x}{\lambda}}; \ x \geq 0. \tag{2}$$

In this article, the discrete version of Ramos and Louzada distribution is proposed and studied in detail. The following are some interesting features of the proposed distribution: Its statistical and reliability characteristics can be expressed as closed forms. Its failure rate is showing an increasing pattern. The suggested distribution evaluated time and count data sets more effectively than competing distributions. As a result, we feel that the proposed model is the greatest option for attracting a wider range of applications and industries.

The rest of the study is organized as follows: In Section 2, we introduce a new distribution using survival discretization methodology. Different mathematical properties are derived in Section 3. Parameter estimation and simulation study are presented in Section 4. Four data sets are utilized to show the flexibility of the proposed model in Section 5. Finally, Section 6 provides some conclusions.

2. Synthesis of discrete Ramos-Louzada distribution

Let Y be a continuous random variable with sf $G(y; \boldsymbol{\eta})$, then the pmf of the discrete random variable $X = \lfloor Y \rfloor$ can be expressed as

$$Pr(X = x; \boldsymbol{\eta}) = G(x; \boldsymbol{\eta}) - G(x + 1; \boldsymbol{\eta}); \; x \in \mathbb{Z}_+,$$

where $\lfloor \cdot \rfloor$ denotes the floor function, which returns the highest integer value smaller or equal than its argument, and $\boldsymbol{\eta}$ is a parameter vector $1 \times k$. If the random variable Y have Ramos-Louzada (RL) distribution, then the pmf of discrete RL (DRL) distribution can be written as

$$\Pr(X = x; \lambda) = p(x; \lambda) = \frac{e^{-\frac{x}{\lambda}}}{\lambda(\lambda-1)}\left[(\lambda^2 - \lambda + x)(1 - e^{-\frac{1}{\lambda}}) - e^{-\frac{1}{\lambda}}\right]; \; x = 0,1,2, ..., \tag{3}$$

where $\lambda \geq 2$ is the shape parameter. The pmf $p(x + 1; \lambda)$ can be expressed as a weight from the pmf $p(x; \lambda)$ as follows

$$p(x + 1; \lambda) = \frac{e^{-\frac{1}{\lambda}}\left[(\lambda^2 - \lambda + x + 1)\left(1 - e^{-\frac{1}{\lambda}}\right) - e^{-\frac{1}{\lambda}}\right]}{\left[(\lambda^2 - \lambda + x)\left(1 - e^{-\frac{1}{\lambda}}\right) - e^{-\frac{1}{\lambda}}\right]} p(x; \lambda).$$

Figure 1 illustrates some pmf plots of the DRL models based on different values of the model parameter λ. It is found that the pmf can be used a positively skewed data with a uni-modal shape.
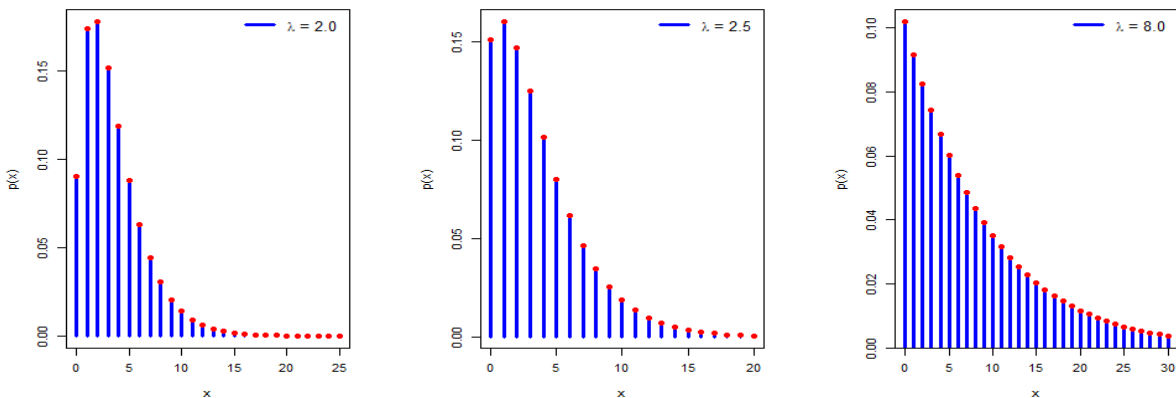

Figure 1. The pmf plots of the DRL distribution.

Based on $F(x; \lambda) = \Pr(X \leq x; \lambda) = 1 - G(x; \lambda) + \Pr(X = x; \lambda)$, the cumulative distribution function (cdf) of the DRL distribution can be formulated as

$$F(x; \lambda) = 1 - \frac{\lambda^2 - \lambda + x + 1}{\lambda(\lambda-1)}e^{-\frac{x+1}{\lambda}}; \; x = 0,1,2, ... \tag{4}$$

The sf of the DRL model can be expressed as

$$S(x; \lambda) = \frac{\lambda^2 - \lambda + x + 1}{\lambda(\lambda-1)}e^{-\frac{x+1}{\lambda}}; \; x = 0,1,2, ... \tag{5}$$

The hazard rate function (hrf) of the DRL model is given by

$$h(x; \lambda) = \frac{(\lambda^2 - \lambda + x)\left(1 - e^{-\frac{1}{\lambda}}\right) - e^{-\frac{1}{\lambda}}}{\lambda^2 - \lambda + x}; \ x = 0,1,2, \dots, \tag{6}$$

where $h(x; \lambda) = \frac{\Pr(X=x;\lambda)}{1-F(x-1;\lambda)}$. Mathematically, the shape of the hrf of the DRL model is always increasing, which makes it an effective statistical tool for modeling data, especially in the engineering and medical fields. Figure 2 shows the hrf plots of the new discrete model based on various values of the model parameter.
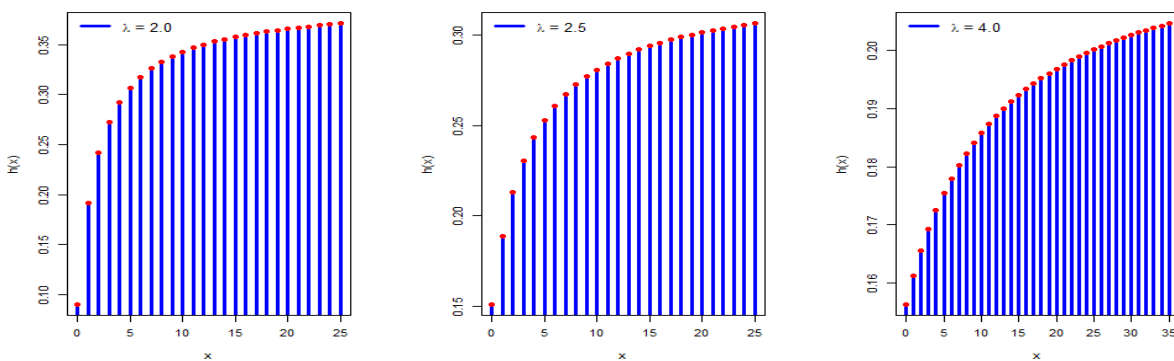

Figure 2. The hrf plots of the DRL distribution.

The reversed hazard rate function (rhrf) and the second rate of failure are given as

$$\check{r} = \frac{e^{-\frac{x}{\lambda}}\left[(\lambda^2 - \lambda + x)(1 - e^{-\frac{1}{\lambda}}) - e^{-\frac{1}{\lambda}}\right]}{\lambda(\lambda-1) - (\lambda^2 - \lambda + x + 1)e^{-\frac{x+1}{\lambda}}}; \ x = 0,1,2, \dots \tag{7}$$

and

$$r^*(x) = log\left[\frac{(\lambda^2 - \lambda + x)e^{\frac{1}{\lambda}}}{\lambda^2 - \lambda + x + 1}\right]; \ x = 0,1,2, \dots, \tag{8}$$

where $\check{r} = \frac{\Pr(X=x;\lambda)}{F(x;\lambda)}$ and $r^*(x) = \log\left[\frac{S(x;\lambda)}{S(x+1;\lambda)}\right]$. Mathematically, and after simple algebra steps, it is found that the shape of the rhrf of the DRL model is decreasing only. Figure 3 shows some rhrf plots of the proposed model based on specific parameter values.
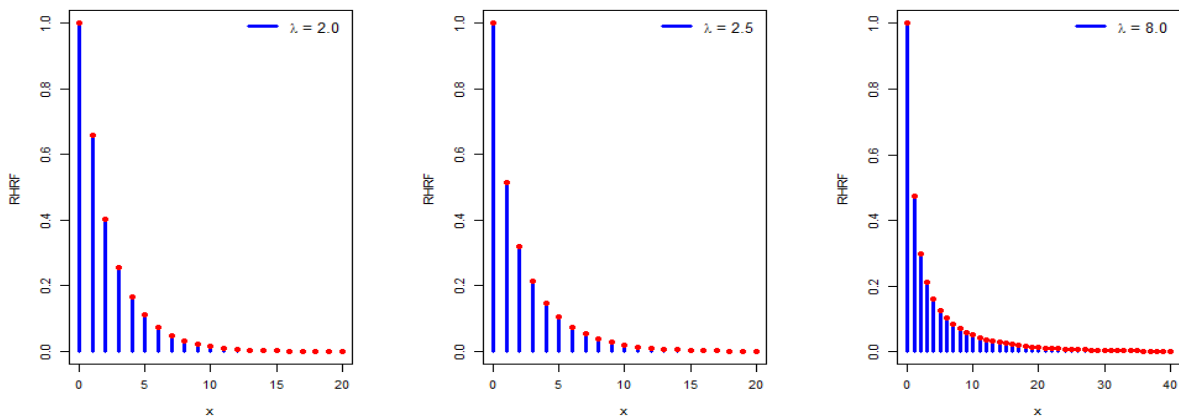
Figure 3. The rhrf plots of the DRL distribution.

3. Distributional properties

In this Section, the probability generating function (pgf) as well as its rth moment are investigated. Assume the random variable X have a DRL model, then the pgf can be expressed as

$$W_X(z) = \sum_{x=0}^{\infty} z^x \Pr(X = x; \lambda) = 1 + (z - 1) \sum_{x=1}^{\infty} z^{x-1} S(x; \lambda)$$

$$= 1 + e^{-\frac{1}{\lambda}}(z - 1) \left[\frac{1}{\left(1 - ze^{-\frac{1}{\lambda}}\right)} + \frac{1}{\lambda(\lambda - 1)\left(1 - ze^{-\frac{1}{\lambda}}\right)^2} \right]. \tag{9}$$

On replacing z by e^z in Eq (9), the moment generating function (mgf) can be written as

$$M_X(z) = 1 + e^{-\frac{1}{\lambda}}(e^z - 1) \left[\frac{1}{\left(1 - e^z e^{-\frac{1}{\lambda}}\right)} + \frac{1}{\lambda(\lambda - 1)\left(1 - e^z e^{-\frac{1}{\lambda}}\right)^2} \right]. \tag{10}$$

The first four moments around the origin $(\mu'_1, \mu'_2, \mu'_3, \mu'_4)$ can be written as

$$\mu'_1 = \frac{\lambda - \lambda^2 + (\lambda^2 - \lambda + 1)e^{\frac{1}{\lambda}}}{\lambda(\lambda - 1)\left(e^{\frac{1}{\lambda}} - 1\right)^2},$$

$$\mu'_2 = \frac{(\lambda(\lambda - 1) + 1)e^{\frac{2}{\lambda}} + 3e^{\frac{1}{\lambda}} - \lambda(\lambda - 1)}{\lambda(\lambda - 1)\left(e^{\frac{1}{\lambda}} - 1\right)^3},$$

$$\mu'_3 = \frac{[\lambda(\lambda-1)+1]e^{\frac{3}{\lambda}} + [3\lambda(\lambda-1)+10]e^{\frac{2}{\lambda}} - [3\lambda(\lambda-1)-7]e^{\frac{1}{\lambda}} - \lambda(\lambda-1)}{\lambda(\lambda-1)\left(e^{\frac{1}{\lambda}}-1\right)^4}$$

and

$$\mu'_4 = \frac{[\lambda(\lambda-1)+1]e^{\frac{4}{\lambda}} + [10\lambda(\lambda-1)+25]e^{\frac{3}{\lambda}} + 67e^{\frac{2}{\lambda}} + [10\lambda(\lambda-1)-3]e^{\frac{1}{\lambda}} - \lambda(\lambda-1)}{\lambda(\lambda-1)\left(e^{\frac{1}{\lambda}}-1\right)^5}.$$

Based on the rth moments, the variance can be expressed as

$$\sigma^2 = \frac{[\lambda^4-2\lambda^3+2\lambda^2-\lambda]e^{\frac{3}{\lambda}} - [2\lambda^4-4\lambda^3+2\lambda^2-1]e^{\frac{2}{\lambda}} + [\lambda^4-2\lambda^3+\lambda]e^{\frac{1}{\lambda}}}{\lambda^2(\lambda-1)^2\left(e^{\frac{1}{\lambda}}-1\right)^4}. \tag{11}$$

The dispersion index (di) is defined by variance to mean ratio. The di indicates that the reported model is suitable for under-, equi- or over-dispersed data sets. Using the derived moments, the coefficients skewness and kurtosis can be listed in closed forms. Some numerical computations for mean, variance, di, skewness, and kurtosis based on DRL parameters are listed in Table 1.

Table 1. Mean, variance, di, skewness, and kurtosis of the DRL distribution.

Measure	λ								
	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	8.0
Mean	3.500	3.678	4.014	4.414	4.847	5.299	5.762	6.234	8.652
Variance	8.079	11.80	15.71	20.03	24.82	30.10	35.87	42.13	80.90
di	2.308	3.207	3.913	4.538	5.121	5.680	6.224	6.757	9.351
Skewness	1.395	1.519	1.629	1.708	1.764	1.806	1.837	1.862	1.928
Kurtosis	5.940	6.353	6.821	7.199	7.495	7.727	7.910	8.057	8.484

According to Table 1, it is noted that the DRL model can be used effectively to model overdispersion data as di is greater than one, which makes it a proper probability tool to discuss actuarial data. Moreover, the new discrete probabilistic model can be utilized to analyze positively skewed data with leptokurtic-shaped.

4. Estimation methods

In this section, six estimation methods are used to estimate the unknown parameter of DRL distribution. The considered estimation methods are maximum likelihood estimation (mle), method of moments (mom), least-squares estimation (lse), Anderson-Darling estimation (ade), Cramer von-Misses estimation (cvme), and maximum product of spacing estimator (mpse).

4.1. Maximum likelihood estimation

Assume a random sample $x_1, x_2, \ldots, x_n$ from the DRL model, then the log-likelihood function can be expressed as

$$L(\lambda|x) = -\frac{1}{\lambda}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \ln\left[(\lambda^2 - \lambda + x_i)\left(1 - e^{-\frac{1}{\lambda}}\right) - e^{-\frac{1}{\lambda}}\right] - n\ln\lambda - n\ln(\lambda - 1). \quad (12)$$

Differentiating the Eq (12) with respect to the parameter λ, we get the non-linear equation as follows

$$\frac{\partial L(\lambda|x)}{\partial \lambda} = \frac{1}{\lambda^2}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{\left[(2\lambda-1)\left(1-e^{-\frac{1}{\lambda}}\right) - \left(1-\frac{1}{\lambda}+\frac{1}{\lambda^2}+\frac{x_i}{\lambda^2}\right)e^{-\frac{1}{\lambda}}\right]}{\left[(\lambda^2-\lambda+x_i)\left(1-e^{-\frac{1}{\lambda}}\right) - e^{-\frac{1}{\lambda}}\right]} - \frac{n}{\lambda} - \frac{n}{\lambda-1}, \quad (13)$$

the exact solution of Eq (13) is not easy, so we will maximize it by using optimization approaches, for example, the Newton-Raphson approach using R software.

4.2. Method of moment estimation

Based on the mom definition, we must equate the sample mean to the corresponding population mean, and then solve the non-linear equation for the parameter

$$\frac{\lambda - \lambda^2 + (\lambda^2 - \lambda + 1)e^{\frac{1}{\lambda}}}{\lambda(\lambda-1)\left(e^{\frac{1}{\lambda}} - 1\right)^2} = \frac{1}{n}\sum_{i=1}^{n} x_i. \quad (14)$$

To solve Eq (14), the **uniroot** function should be utilized.

4.3. Least squares estimation

To estimate the parameter minimizing the sum of squares of residuals, a standard approach like the lse should be used. For the estimation of the parameter of DRL distribution, the lse can be obtained by minimizing

$$\text{lse}(\lambda) = \sum_{i=1}^{n}\left[1 - \frac{(\lambda^2-\lambda+x_{i:n})}{\lambda(\lambda-1)}e^{-\frac{x_{i:n}+1}{\lambda}} - \frac{i}{n+2}\right]^2, \quad (15)$$

with respect to the parameter λ.

4.4. Anderson-Darling estimation

The ade of the parameter can be derived by minimizing the following equation

$$\text{ade}(\lambda) = -n - \frac{1}{n}\sum_{i=1}^{n}(2_i - 1)\left[\log\left(1 - \frac{(\lambda^2-\lambda+x_{i:n})}{\lambda(\lambda-1)}e^{-\frac{x_{i:n}+1}{\lambda}}\right) + \log\left(\frac{(\lambda^2-\lambda+x_{i:n})}{\lambda(\lambda-1)}e^{-\frac{x_{i:n}+1}{\lambda}}\right)\right]^2, \quad (16)$$

with respect to the parameter λ.

4.5. Cramer von-Misses estimation

The cvme is an estimation method. This method is derived as the difference between the empirical cdf and fitted cdf where

$$\text{cvme}(\lambda) = \frac{1}{12n} + \sum_{i=1}^{n} \left[1 - \frac{(\lambda^2 - \lambda + x_{i:n})}{\lambda(\lambda-1)} e^{-\frac{x_{i:n+1}}{\lambda}} - \frac{2i-1}{2n} \right]^2. \tag{17}$$

4.6. Maximum product of spacing estimator

For $u = 1,2,3,\dots, h+1$, assume $D_u(\lambda) = F\big(x_{(u)}|\lambda\big) - F\big(x_{(u-1)}|\lambda\big)$, be the uniform spacings of a random sample from the DRL model, where $F\big(x_{(0)}|\lambda\big) = 0$, $F\big(x_{(h+1)}|\lambda\big) = 1$ and $\sum_{r=1}^{h+1} D_u(\lambda) = 1$. The mpse of the parameter λ, say $\hat{\lambda}$, can be estimated by maximizing the geometric mean of the spacings

$$\text{mpse}(\lambda) = [\textstyle\prod_{u=1}^{h+1} D_u(\lambda)]^{\frac{1}{h+1}}, \tag{18}$$

with respect to the parameter λ.

5. Simulation

In this section, we discussed the results of the simulation study to compare the estimation performance of the proposed estimators based on the DRL model. The performance of considered estimators is evaluated via absolute biases and mean square errors. We simulate 10,000 random samples from the DRL distribution using the following sample sizes $n = 10, 20, 50, 100, 200$, and 500. The results calculated for parameter values $\lambda = (2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 8.0, 10.0)$ using R Software. The absolute bias and mse for the parameter λ are presented in Tables 2 and 3.

Table 2. Simulation results of the DRL distribution for different parameter values.

Para.	n	Bias						mse					
λ		mle	mom	ade	cvme	lse	mpse	mle	mom	ade	cvme	lse	mpse
	10	0.099	0.072	0.878	0.606	0.956	0.801	0.762	1.496	2.897	2.059	3.034	1.797
	20	0.034	0.035	0.486	0.259	0.444	0.435	0.307	1.103	1.408	0.749	1.212	0.626
2.0	50	0.017	0.160	0.122	0.037	0.065	0.172	0.055	0.811	0.304	0.086	0.148	0.111
	100	0.013	0.226	0.015	0.003	0.005	0.086	0.017	0.697	0.033	0.006	0.009	0.026
	200	0.009	0.277	0.000	0.000	0.000	0.050	0.007	0.620	0.000	0.000	0.000	0.008
	500	0.002	0.338	0.000	0.000	0.000	0.028	0.003	0.569	0.000	0.000	0.000	0.002
	10	0.081	0.131	1.109	1.140	1.397	0.958	1.105	2.032	3.694	3.919	4.392	2.609
	20	0.065	0.239	0.837	0.869	1.060	0.622	0.485	1.487	2.368	2.560	2.765	1.173
2.5	50	0.058	0.293	0.595	0.695	0.798	0.320	0.186	1.063	1.436	1.723	1.785	0.391
	100	0.027	0.298	0.475	0.602	0.672	0.195	0.091	0.816	1.082	1.373	1.428	0.165
	200	0.013	0.255	0.374	0.525	0.589	0.098	0.045	0.602	0.864	1.139	1.185	0.065
	500	0.007	0.146	0.273	0.484	0.532	0.047	0.018	0.316	0.674	0.979	1.014	0.022

Continued on next page

Para.	n	Bias						mse					
λ		mle	mom	ade	cvme	lse	mpse	mle	mom	ade	cvme	lse	mpse
3.0	10	0.019	0.224	1.085	1.103	1.304	0.996	1.574	2.613	3.976	4.160	4.504	3.186
	20	0.005	0.268	0.877	0.924	1.044	0.639	0.751	1.793	2.525	2.701	2.865	1.512
	50	0.003	0.193	0.738	0.844	0.893	0.357	0.313	1.003	1.540	1.728	1.760	0.537
	100	0.004	0.130	0.754	0.868	0.872	0.222	0.163	0.567	1.160	1.352	1.341	0.242
	200	0.001	0.053	0.775	0.887	0.896	0.141	0.084	0.238	0.920	1.112	1.106	0.113
	500	0.001	0.006	0.811	0.929	0.940	0.070	0.035	0.066	0.769	0.965	0.980	0.042
4.0	10	0.049	0.209	1.105	1.023	1.181	1.100	2.450	3.694	5.254	5.458	5.552	4.850
	20	0.043	0.169	0.896	0.891	0.985	0.748	1.339	2.065	3.000	3.234	3.315	2.269
	50	0.031	0.049	0.835	0.847	0.899	0.400	0.585	0.763	1.608	1.739	1.789	0.815
	100	0.008	0.014	0.849	0.878	0.891	0.255	0.291	0.328	1.161	1.240	1.285	0.377
	200	0.007	0.016	0.838	0.877	0.882	0.156	0.147	0.159	0.911	1.005	1.010	0.174
	500	0.004	0.002	0.831	0.877	0.878	0.066	0.058	0.064	0.773	0.861	0.862	0.062

Table 3. Simulation results of the DRL distribution for different parameter values.

Para.	n	Bias						mse					
λ		mle	mom	ade	cvme	lse	mpse	mle	mom	ade	cvme	lse	mpse
5.0	10	0.106	0.077	1.058	1.013	1.242	1.164	3.881	4.545	6.333	6.748	7.579	6.421
	20	0.061	0.036	0.955	0.915	1.007	0.824	2.026	2.335	3.527	3.991	4.014	3.056
	50	0.046	0.015	0.864	0.869	0.904	0.447	0.821	0.839	1.793	1.906	1.970	1.070
	100	0.013	0.002	0.857	0.866	0.857	0.270	0.399	0.411	1.244	1.325	1.303	0.513
	200	0.013	0.006	0.833	0.843	0.849	0.161	0.198	0.214	0.946	1.006	1.009	0.238
	500	0.001	0.003	0.832	0.846	0.843	0.082	0.079	0.078	0.792	0.829	0.822	0.086
6.0	10	0.167	0.115	1.192	1.073	1.169	1.260	5.339	5.663	8.785	8.977	9.398	8.321
	20	0.049	0.027	0.946	0.952	0.975	0.890	2.794	2.926	4.236	4.861	4.686	3.860
	50	0.008	0.019	0.873	0.889	0.881	0.499	1.081	1.079	2.081	2.225	2.292	1.417
	100	0.007	0.016	0.859	0.843	0.870	0.294	0.523	0.517	1.392	1.421	1.522	0.631
	200	0.007	0.006	0.845	0.839	0.846	0.181	0.261	0.269	1.030	1.061	1.065	0.308
	500	0.001	0.001	0.832	0.832	0.835	0.085	0.107	0.108	0.823	0.841	0.847	0.112
8.0	10	0.141	0.089	1.294	1.077	1.320	1.622	8.569	8.656	12.77	12.61	13.63	13.41
	20	0.074	0.009	1.050	0.991	1.111	1.110	4.265	4.178	6.399	6.736	7.355	6.250
	50	0.029	0.003	0.883	0.860	0.918	0.577	1.685	1.630	2.726	3.019	3.118	2.151
	100	0.015	0.014	0.867	0.861	0.859	0.372	0.834	0.814	1.787	1.862	1.859	1.029
	200	0.008	0.019	0.837	0.808	0.842	0.219	0.416	0.422	1.202	1.204	1.285	0.493
	500	0.006	0.001	0.826	0.814	0.820	0.097	0.169	0.175	0.877	0.879	0.902	0.176
10.0	10	0.106	0.086	1.365	1.333	1.446	1.868	12.89	12.67	17.66	19.87	19.51	20.26
	20	0.001	0.063	1.143	1.079	1.067	1.308	6.232	6.279	8.767	9.553	9.777	9.056
	50	0.025	0.036	0.899	0.892	0.933	0.686	2.483	2.373	3.774	4.076	4.248	3.246
	100	0.018	0.016	0.868	0.886	0.872	0.457	1.247	1.246	2.255	2.421	2.464	1.489
	200	0.001	0.009	0.852	0.827	0.829	0.267	0.636	0.599	1.467	1.501	1.501	0.712
	500	0.002	0.007	0.823	0.801	0.818	0.119	0.246	0.240	0.967	0.971	1.005	0.255

Based on the simulation criteria, it is observed that all estimation approaches work quite well in estimating the parameter λ of the DRL distribution.

6. Data analysis

In this Section, the importance of the proposed distribution is discussed by using data sets from different areas. We shall compare the fits of the DRL distribution with different competitive distributions such as Poisson (Poi), discrete Pareto (DPr), discrete Rayleigh (DR), discrete inverse Rayleigh (DIR), discrete Burr-Hatke (DBH), discrete Bilal (DBi), discrete Lindley (DL), new discrete Lindley (NDL), and discrete Burr-XII (DBXII) distributions. The fitted probability distributions are compared using some criteria, namely, the negative log-likelihood ($-L$), Akaike information criterion (aic), and Kolmogorov-Smirnov (ks) test with its p-value.

6.1. Data set I: Covid-19 data in Pakistan

The first data set represents the number of deaths due to coronavirus in Pakistan during the period March 18, 2020, to April 30, 2020, which were obtained from the public reports of the National Institute of Health (NIH), Islamabad, Pakistan (https://covid.gov.pk/stats/pakistan). The mean, variance, and di of data set I are 9.4773, 102.39, and 10.804, respectively. The mle(s) along with standard error(s) "se(s)" and goodness-of-fit measures for this data are presented in Table 4.

Table 4. The mle(s), se(s), and goodness-of-fit measures for data set I.

Model	λ		δ		Goodness-of-fit measures			
	mle	se	mle	se	$-L$	aic	ks	p-value
DRL	8.8686	1.5033	-	-	145.22	292.43	0.156	0.2300
Poi	9.4773	0.4641	-	-	283.94	569.89	0.391	<0.0001
DPr	0.5021	0.0757	-	-	162.19	326.38	0.401	<0.0001
DR	9.9883	0.7535	-	-	168.85	339.70	0.339	<0.0001
DIR	7.4291	1.2625	-	-	166.31	334.61	0.382	<0.0001
DBH	0.9950	0.0115	-	-	175.37	352.74	0.647	<0.0001
DBi	11.838	1.2932	-	-	151.29	304.59	0.213	0.0370
DL	0.8313	0.0165	-	-	149.17	300.33	0.184	0.1000
NDL	0.1640	0.0161	-	-	148.44	298.89	0.237	0.0140
DBXII	0.9536	0.0434	11.907	11.305	150.70	305.40	0.302	0.0007

The results in Table 4 show that the DRL distribution provides a better fit over other competing discrete models since it has the minimum aic, and ks values with the highest p-value. Figure 4 shows the probability-probability (pp) plots for all tested models which prove the empirical results listed in Table 4.
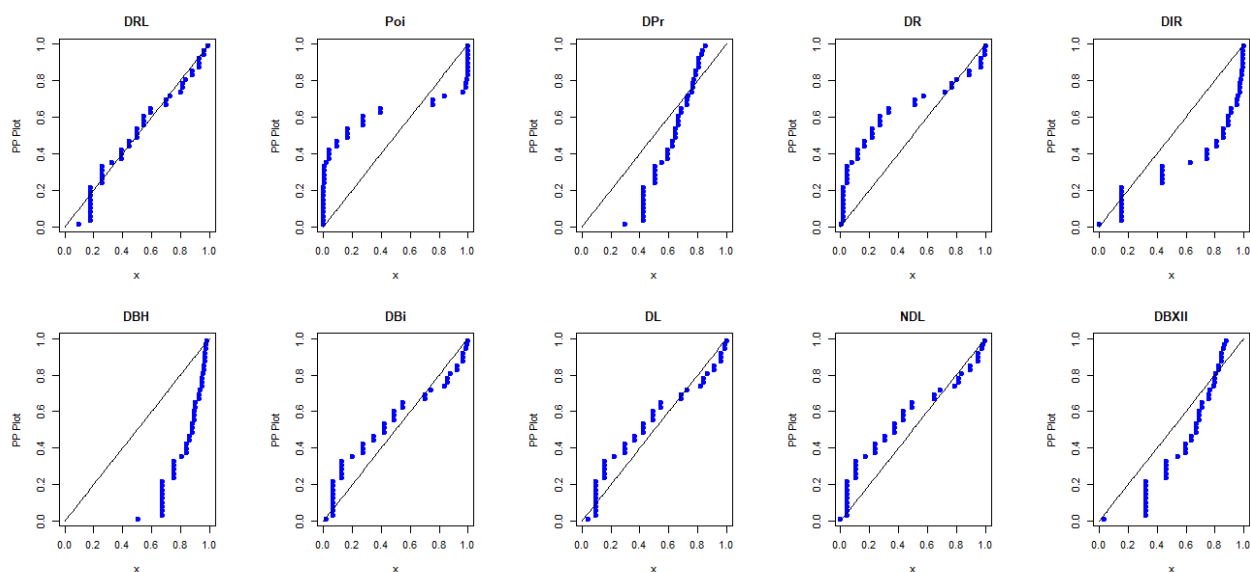
Figure 4. The pp plots for all fitted distributions for data set I.

6.2. Data set II: Hydrology data

The second data set was reported in [25], which represents the exceedance of flood peaks in m^3/s of the Wheaton River near Carcross in Yukon Territory, Canada based on the discretization concept. The mean, variance, and di of this data are 11.806, 152.38, and 12.908, respectively. The mle(s), se(s), and goodness-of-fit measures for data set II are reported in Table 5.

Table 5. The mle(s), se(s), and goodness-of-fit measures for data set II.

Model	λ		δ		Goodness-of-fit measures			
	mle	se	mle	se	$-L$	aic	ks	p-value
DRL	11.214	1.4497	-	-	252.71	507.43	0.133	0.1600
Poi	11.805	0.4049	-	-	564.38	1130.8	0.408	<0.0001
DPr	0.4770	0.0563	-	-	276.82	555.64	0.311	<0.0001
DR	12.280	0.7239	-	-	300.65	603.29	0.323	<0.0001
DIR	4.7947	0.6303	-	-	331.46	664.92	0.497	<0.0001
DBH	0.9966	0.0072	-	-	302.29	606.57	0.572	<0.0001
DBi	14.621	1.2479	-	-	272.50	546.99	0.257	0.0002
DL	0.8592	0.0109	-	-	264.30	530.59	0.232	0.0009
NDL	0.1373	0.0107	-	-	262.09	526.17	0.271	<0.0001
DBXII	0.8205	0.0591	2.6112	0.9287	270.50	544.99	0.228	0.0011

It is observed that the DRL model is the best among all competitive distributions. Figure 5 illustrates the pp plots for all tested distributions which prove the empirical results reported in Table 5.
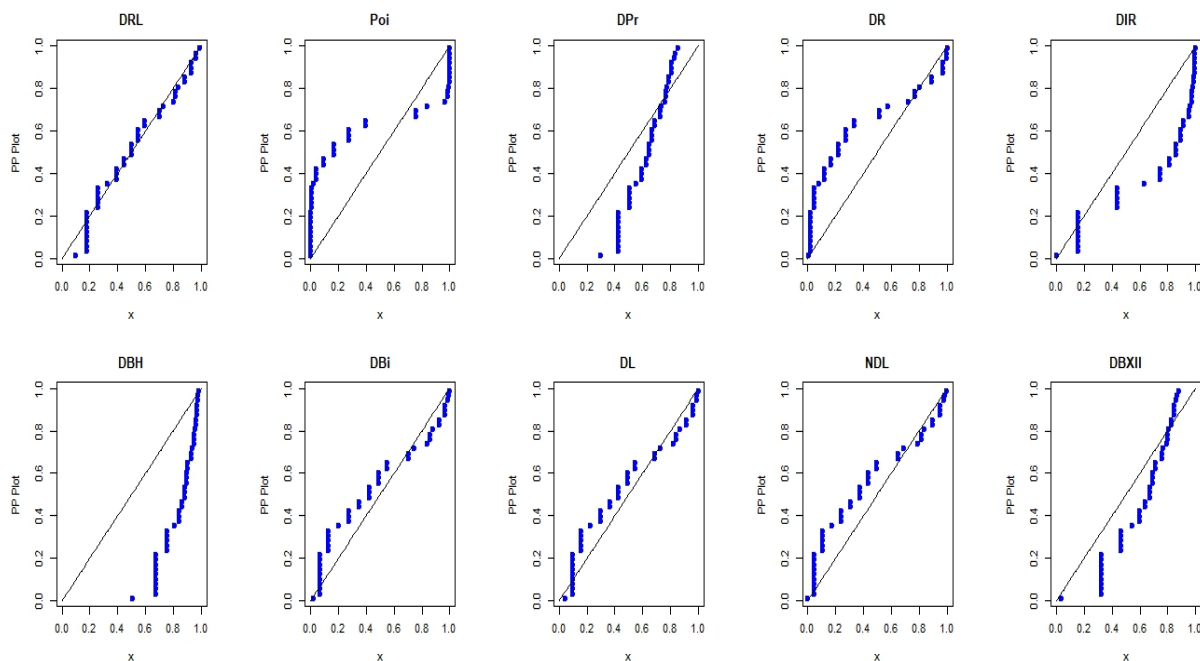
Figure 5. The pp plots for all fitted distributions for data set II.

6.3. Data set III: Forest fire in Greece

The third data set was listed in [26] and represents the number of fires in Greece forest districts for the period from 1st July 1998 to 31 August 1998. The mean, variance, and di measures are 5.2, 32.382, and 6.2272, respectively. The mle(s), se(s), and goodness-of-fit measures for data set II are listed in Table 6.

Table 6. The mle(s), se(s), and goodness-of-fit measures for data set III.

Model	λ		δ		Goodness-of-fit measures			
	mle	se	mle	se	$-L$	aic	ks	p-value
DRL	4.3673	0.5510	-	-	301.11	604.21	0.1510	0.0140
Poi	5.2000	0.2174	-	-	434.16	870.32	0.282	<0.0001
DPr	0.6250	0.0597	-	-	339.05	680.10	0.352	<0.0001
DR	5.6788	0.2714	-	-	352.72	707.45	0.261	<0.0001
DIR	3.5198	0.3748	-	-	360.90	723.80	0.413	<0.0001
DBH	0.9833	0.0136	-	-	352.42	706.85	0.532	<0.0001
DBi	6.7993	0.4693	-	-	310.75	623.49	0.107	<0.0001
DL	0.7337	0.0156	-	-	303.88	609.75	0.193	0.0100
NDL	0.2567	0.0152	-	-	302.73	607.47	0.169	0.0037
DBXII	0.7486	0.0459	2.4582	0.4938	325.00	654.01	0.287	<0.0001

It is found that the new discrete model is the best among all tested distributions. Figure 6 shows the pp plots for all competitive distributions which prove the empirical results listed in Table 6.
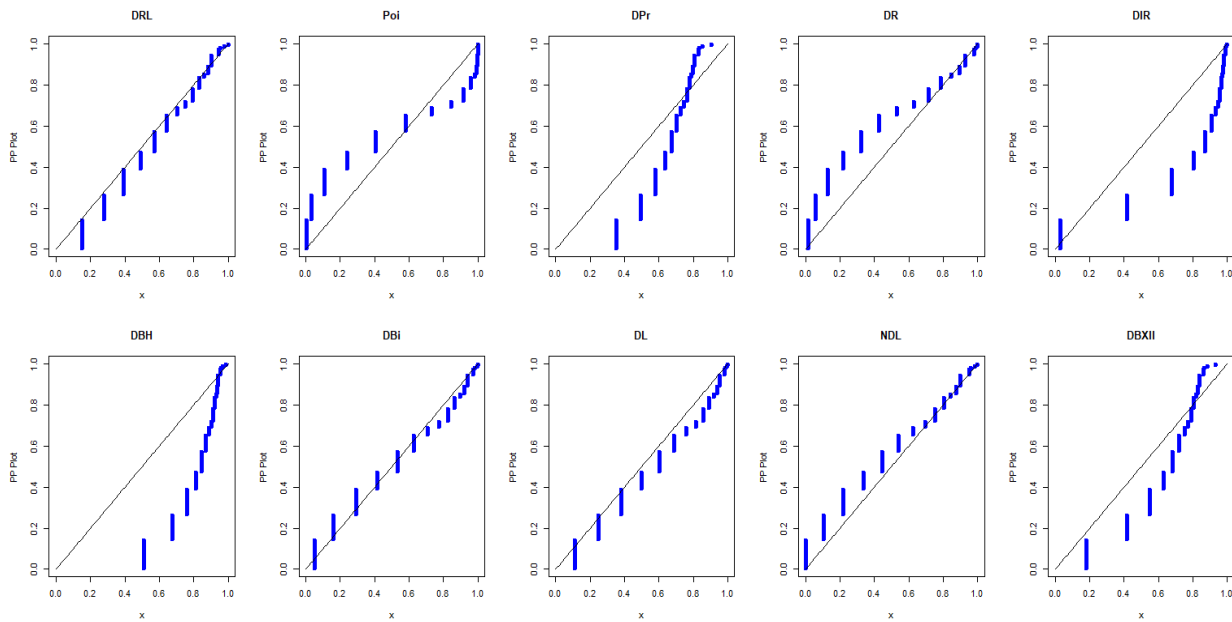
Figure 6. The pp plots for all fitted distributions for data set III.

6.4. Data set IV: AG-positive leukemia

The fourth data set represents the time to death (in weeks) of AG-positive leukemia patients [27]. The mean, variance, and di values are 62.471, 2954.3, and 47.29, respectively. The estimates and goodness-of-fit measures for all competitive distributions are listed in Table 7.

Table 7. The mle(s), se(s), and goodness-of-fit measures for data set IV.

Model	λ		δ		Goodness-of-fit measures			
	mle	se	mle	se	$-L$	aic	ks	p-value
DRL	61.943	15.273	-	-	87.425	176.85	0.152	0.8300
Poi	62.470	1.9169	-	-	475.26	952.52	0.470	0.0011
DPr	0.2838	0.0688	-	-	98.335	198.67	0.324	0.0560
DR	58.076	7.0429	-	-	96.794	195.59	0.309	0.0770
DIR	25.310	6.543	-	-	128.59	259.18	0.681	<0.0001
DBH	0.9999	0.0029	-	-	119.81	241.62	0.716	<0.0001
DBi	75.109	13.164	-	-	92.886	187.77	0.219	0.3900
DL	0.9692	0.0052	-	-	91.858	185.72	0.215	0.4100
NDL	0.0306	0.0052	-	-	91.458	184.92	0.218	0.3900
DBXII	0.9975	0.0008	117.30	45.982	96.151	196.30	0.327	0.0530

It is noted that the DRL is the best for this data. Figure 7 shows the pp plots for all tested distributions which prove the empirical results mentioned in Table 7.
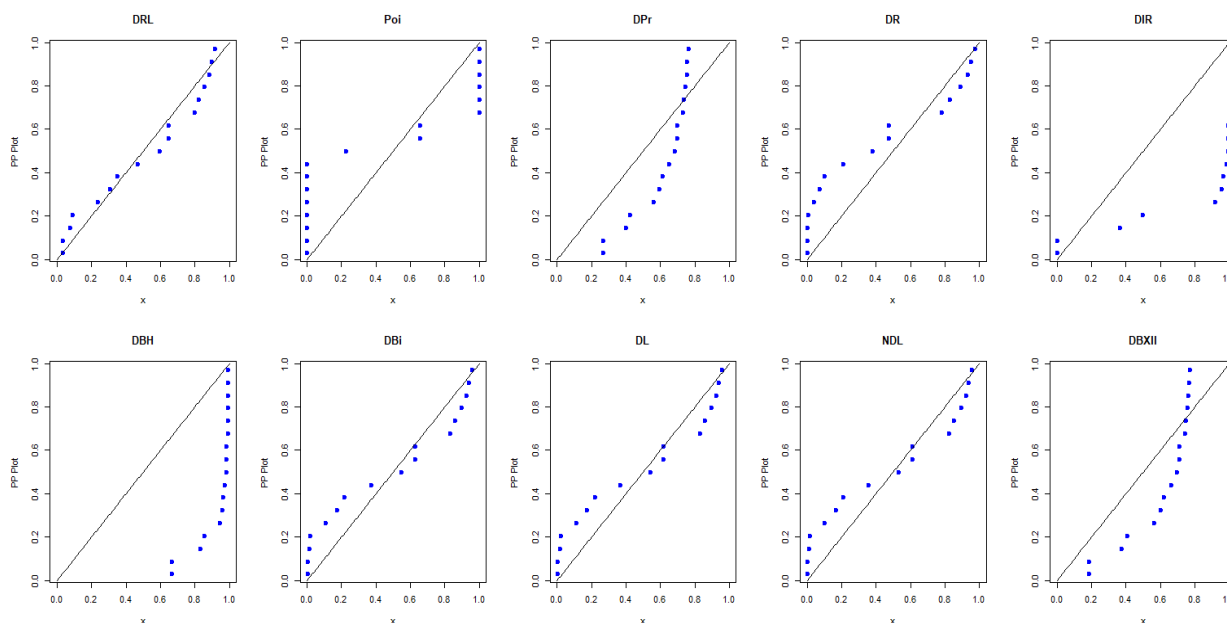
Figure 7. The pp plots for all fitted distributions for data set IV.

7. Conclusions

In this article, a new one-parameter discrete model has been proposed entitled a discrete Ramos-Louzada (DRL) distribution. The new model can be used in modeling asymmetric data with overdispersion phenomena. Some of its statistical properties have been derived. It was found that all its properties can be expressed in closed forms, which makes the new model can be utilized in different analysis, especially, in time series and regression. Various estimation techniques including maximum likelihood, moments, least squares, Anderson's-Darling, Cramer von-Mises, and maximum product of spacing estimator, have been investigated to get the best estimator for the real data. The estimation performance of these estimation techniques has been assessed via a comprehensive simulation study. The flexibility of the proposed discrete model has been tested utilizing four distinctive real data sets in various fields. Finally, we hope that the DRL distribution attracts wider sets of applications in different fields.

Conflict of interest

The authors declare that they have no conflict of interest to report regarding the present study.

References

1. D. Roy, The discrete normal distribution, *Commun. Stat. Methods*, **32** (2003), 1871–1883. doi: 10.1081/STA-120023256.
2. D. Roy, Discrete Rayleigh distribution, *IEEE Trans. Reliab.*, **53** (2004), 255–260. doi: 10.1109/TR.2004.829161.

3. M. S. Eliwa, M. El-Morshedy, A one-parameter discrete distribution for over-dispersed data: Statistical and reliability properties with applications, *J. Appl. Stat.*, 2021, 1–21. doi: 10.1080/02664763.2021.1905787.

4. H. Krishna, P. S. Pundir, Discrete Burr and discrete Pareto distributions, *Stat. Methodol*, **6** (2009), 177–188. doi: 10.1016/j.stamet.2008.07.001.

5. M. Aboraya, M. Y. Haitham, G. G. Hamedani, M. Ibrahim, A new family of discrete distributions with mathematical properties, characterizations, Bayesian and non-Bayesian estimation methods, *Mathematics*, **8** (2020), 1648. doi: 10.3390/math8101648.

6. E. Gómez-Déniz, E. Calderín-Ojeda, The discrete lindley distribution: Properties and applications, *J. Stat. Comput. Simul.*, **81** (2011), 1405–1416. doi: 10.1080/00949655.2010.487825.

7. B. A. Para, Discrete generalized Burr-Type XII distribution, *J. Mod. Appl. Stat. Methods*, **13** (2014), 244–258. doi: 10.22237/jmasm/1414815120.

8. A. A. AL-Huniti, G. R. AL-Dayian, Discrete Burr type III distribution, *Am. J. Math. Stat.*, **2 (**2012**)**, 145–152. doi: 10.5923/j.ajms.20120205.07.

9. T. Hussain, M. Aslam, and M. Ahmad, A two parameter discrete Lindley distribution, *Rev. Colomb. Estadística*, **39** (2016), 45–61.

10. B. Para, Discrete version of log-logistic distribution and its applications in genetics, *Int. J. Mod. Math. Sci.*, 2016.

11. G. R. Al-dayian, Bayesian estimation and prediction of discrete Gompertz distribution, **36** (2019), 1–21. doi: 10.9734/jamcs/2021/v36i230335.

12. E. M. Almetwally, M. I. Gamal, Discrete alpha power inverse Lomax distribution with application of COVID-19 data, *Int. J. Appl. Math.*, **9** (2020), 11–22.

13. A. Hassan, G. A. Shalbaf, S. Bilal, A. Rashid, A new flexible discrete distribution with applications to count data, *J. Stat. Theory Appl.*, 19 (2020), 102–108.

14. A. Barbiero, A. Hitaj, A discrete analogue of the half-logistic distribution, *2020 International Conference on Decision Aid Sciences and Application (DASA)*, 2020,64–67. doi: 10.1109/DASA51403.2020.9317237.

15. F. C. Opone, E. A. Izekor, I. U. Akata, F. E. U. Osagiede, A discrete analogue of the continuous Marshall-Olkin Weibull distribution with application to count data, *Earthline J. Math. Sci.*, **5 (**2020**)**, 415–428. doi: 10.34198/ejms.5221.415428.

16. M. S. Eliwa, A. A. Ziyad, M. El-Morshedy, Discrete Gompertz-G family of distributions for over- and under-dispersed data with properties, estimation, and applications, *Mathematics*, **8** (2020), 358. doi: 10.3390/math8030358.

17. 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. doi: 10.1109/ACCESS.2020.2988431.

18. M. S. Eliwa, E. Altun, M. El-Dawoody, M. El-Morshedy, A new three-parameter discrete distribution with associated INAR(1) process and applications, *IEEE Access*, **8** (2020), 91150–91162. doi: 10.1109/ACCESS.2020.2993593.

19. A. A. Al-Babtain, A. H. N. Ahmed, A. Z. Afify, A new discrete analog of the continuous Lindley distribution, with reliability applications, *Entropy*, **22** (2020), 1–18. doi: 10.3390/e22060603.

20. M. El-Morshedy, M. S. Eliwa, H. Nagy, A new two-parameter exponentiated discrete Lindley distribution: properties, estimation and applications, *J. Appl. Stat.*, **47** (2020), 354–375. doi: 10.1080/02664763.2019.1638893.

21. M. El-Morshedy, E. Altun, M. S. Eliwa, A new statistical approach to model the counts of novel coronavirus cases, *Math. Sci.*, **18** (2021), 1–14. doi: 10.1007/s40096-021-00390-9.

22. S. Chakraborty, D. Chakravarty, J. Mazucheli, W. Bertoli, A discrete analog of Gumbel distribution: properties, parameter estimation and applications, *J. Appl. Stat.*, **48** (2021), 712–737. doi: 10.1080/02664763.2020.1744538.

23. 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.

24. P. L. Ramos, F. Louzada, A distribution for instantaneous failures, *Stats*, **2** (2019), 247–258. doi: 10.3390/stats2020019.

25. V. Choulakian, M. A. Stephens, Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics*, **43** (2001), 478–484. doi: 10.1198/00401700152672573.

26. H. S. Bakouch, M. A. Jazi, S. Nadarajah, A new discrete distribution, *Statistics*, **48** (2014), 200–240. doi: 10.1080/02331888.2012.716677.

27. S. Chakraborty, D. Chakravarty, Discrete gamma distributions: Properties and parameter estimations, *Commun. Stat. Methods*, 41 (2012), 3301–3324. doi: 10.1080/03610926.2011.563014.