REGULAR PAPER

A new discrete XLindley distribution: theory, actuarial measures, inference, and applications

Ahmed Sedky Eldeeb[1,2] · Muhammad Ahsan-ul-Haq[3] · Ayesha Babar[4]

Abstract

This study proposes a discrete probability model for the modeling of count datasets. Some important properties are derived, including characteristic and moment-generating functions, mean, variance, index of dispersion, skewness, kurtosis, and risk measures (actuarial measures). The maximum likelihood estimation approach has been used to estimate parameter estimates for the proposed distribution. The convergence of the estimators is assessed via a simulation study. In the end, applications to four practical datasets are given to show the usefulness of the proposed distribution over famous discrete distributions. It is manifest that the proposed model provides a better fit than competitive models.

Keywords XLindley distribution · Discretization · Inference · Count data · Data analysis

1 Introduction

Count data occur in a variety of sectors, such as the number of patients, medical visits, catastrophic earthquakes each year, traffic accidents in a month, trees in the forest, number of fires, and the number of microorganisms growing in an hour. The Poisson distribution is a standard distribution for modeling count observations. Sometimes the existing models do not follow the properties of datasets and do not provide efficient results. So, a more flexible probability distribution is required to analyze count datasets with varying behavior. Several probability models are introduced using different discretization approaches. The reader can consult a comprehensive review by [9] on discrete models, datasets, and discretization techniques. Some examples are: Poisson Lindley [25], discrete Pareto [19], discrete Lindley [15],

discrete inverse Weibull [17], Poisson Ailamujia [16], discrete Ramus-Louzada [13], Poisson XLindley [3], Poisson moment exponential [2], discrete power-Ailamujia [5], discrete moment exponential [1], Poisson Mirra [21] and new discrete Ramos-Louzada [4].

Assume a random variable X having XLindley distribution [10] with the probability density function (PDF) and cumulative distribution function (CDF), respectively:

$$f(x) = \frac{\theta^2(2+\theta+x)e^{-\theta x}}{(1+\theta)^2}, \, x, \theta > 0 \tag{1}$$

and

$$F(x; \theta) = 1 - \left(1 + \frac{\theta x}{(1+\theta)^2}\right)e^{-\theta x}, \, x, \theta > 0 \tag{2}$$

The XLindley distribution attracted a lot of attention due to its adaptability. Various authors further generalized it for more complicated and different types of datasets. For example, for unit interval datasets, the Unit-XLindley [14] distribution; for count observations, the Poisson XLindley [3]; and for continuous datasets, the Power XLindley [22].

Let a random variable X follow a continuous random variable with PDF over range R, then the resulting probability mass function (PMF) for a new discrete random variable Y is obtained using relation:

✉ Muhammad Ahsan-ul-Haq
  ahsanshani36@gmail.com

1 Department of Business Administration, College of Business, King Khalid University, Abha, Saudi Arabia

2 Faculty of Business, Department of Statistics, Mathematics and Insurance, Alexandria University, Alexandria, Egypt

3 College of Statistical Sciences, University of Punjab, Lahore, Pakistan

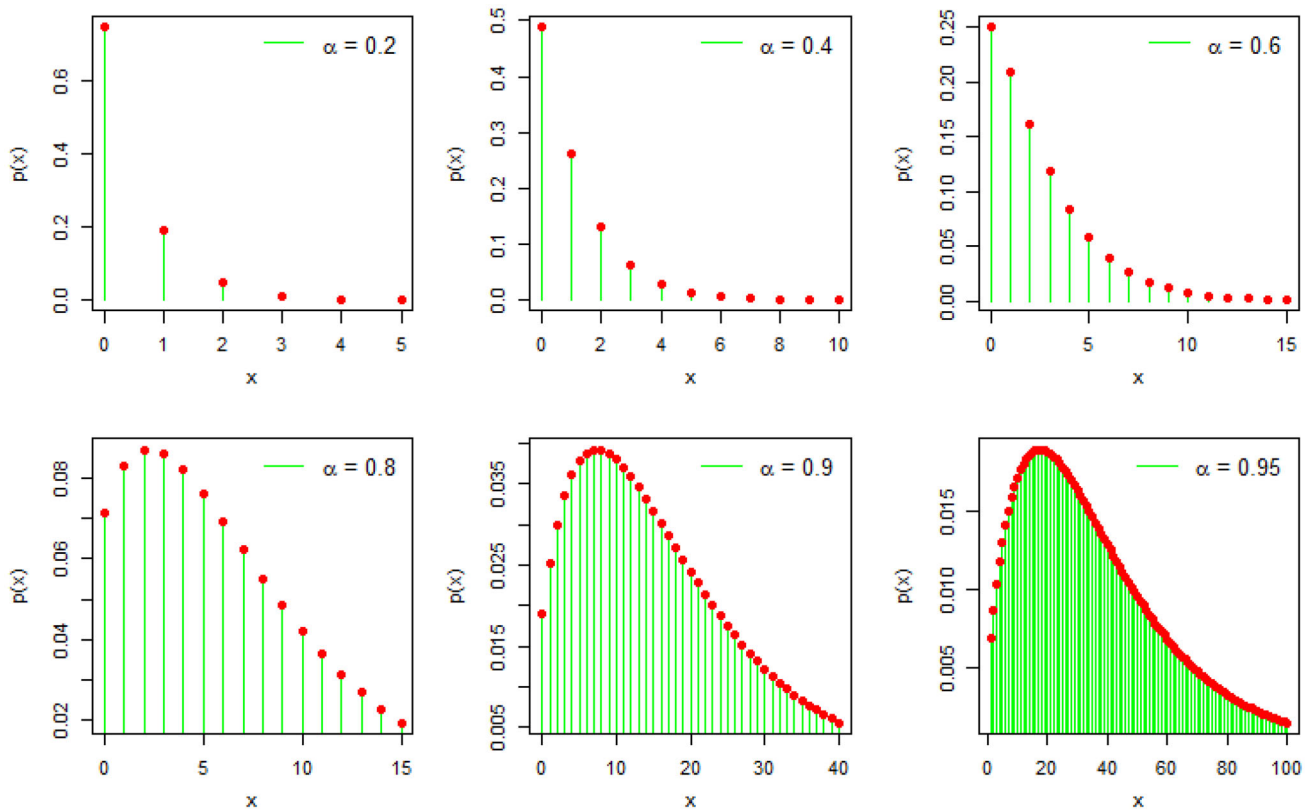4 School of Statistics, Minhaj University Lahore, Lahore, Pakistan

Fig. 1 PMF graphs of the DXLD for various parameter settings

$$P(y; \theta) = \frac{f_X(y, \theta)}{\sum_{i=-\infty}^{\infty} f_X(i, \theta)}, \ y \epsilon \mathbb{Z} \tag{3}$$

In this study, we used the discretization approach given in Eq. (3) to propose a new discrete XLindley (DXL) distribution. The DXL model included some intriguing characteristics, including closed-form formulas for the mean, variance, and moment-generating function. It is an excellent candidate for modeling over-dispersed nature datasets. In the end, we validated the importance of the proposed distribution using four datasets from diverse fields.

The structure of the study is as follows: We proposed a new probability model in Sect. 2. In Sect. 3, statistical properties are derived. The proposed distribution's parameter estimation is covered in Sect. 4. In Sect. 5, the adaptability of the new distribution is demonstrated by the analysis of four datasets. Finally, Sect. 6 brings our research to a conclusion.

2 Derivation of new distribution

A new discrete probability distribution is derived using the discretization methodology stated in Eq. (3). The PMF of DXL distribution is given below:

$$P(y; \alpha) = \frac{(1-\alpha)^2(2-\ln(\alpha)+y)\alpha^y}{[(2-\ln(\alpha))(1-\alpha)+\alpha]}, \ y = 0, 1, 2, \ldots. \tag{4}$$

where $0 < \alpha = e^{-\theta} < 1$. The DXL PMF behavior for different values of a parameter is given in Fig. 1.

It is found that the PMF exhibits declining, increasing and then, decreasing and unimodal form. So asymmetric datasets can be modeled using the suggested approach. The DXL model's CDF can be written as follows:

$$F(y) = 1 - \frac{\alpha^{y+1}[(2-\ln(\alpha)+y)(1-\alpha)+1]}{[(2-\ln(\alpha))(1-\alpha)+\alpha]}; \ y = 0, 1, 2, \ldots \tag{5}$$

where $\alpha > 0$. The corresponding survival function is:

$$S(y) = \frac{\alpha^{y+1}[(2-\ln(\alpha)+y)(1-\alpha)+1]}{[(2-\ln(\alpha))(1-\alpha)+\alpha]}; \ y = 0, 1, 2, \ldots. \tag{6}$$
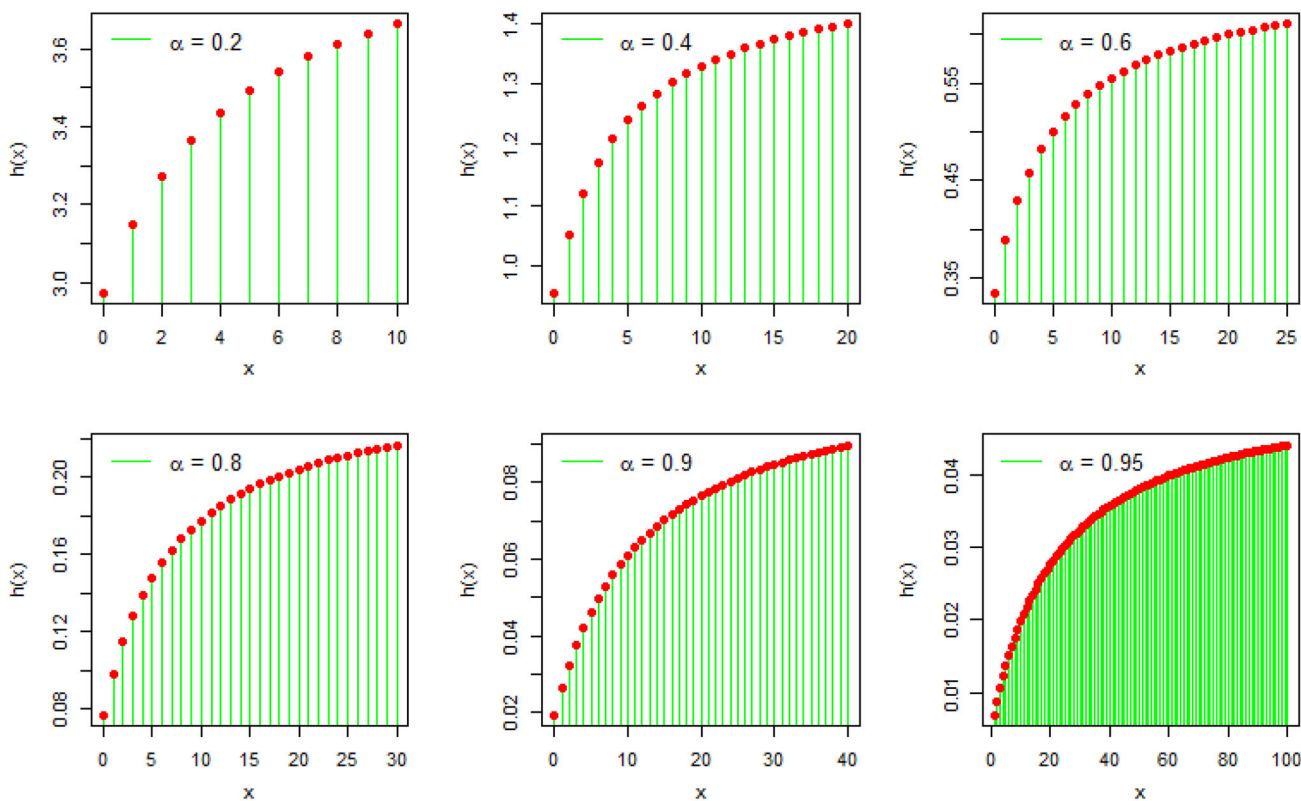
The hazard function (HF) is:

Fig. 2 Visualization of HF for various parameter values

$$h(y) = \frac{(1-\alpha)^2(2-\ln(\alpha)+y)}{\alpha[(2-\ln(\alpha)+y)(1-\alpha)+1]}; \quad y = 0, 1, 2, \ldots. \tag{7}$$

Figure 2 makes available the HF visualization of the DXL distribution for some choices of parameters. We observe that the failure rate pattern of the proposed model is increasing.

3 Statistical properties

In this section, some statistical properties of DXL distribution are derived and studied.

3.1 Moment-generating function (mgf)

The mgf of DXL distribution with parameter α is given by:

$$M_y(t) = \frac{(1-\alpha)^2\left[(2-\ln(\alpha))\left(1-\alpha e^t\right)+\alpha e^t\right]}{[(2-\ln(\alpha))(1-\alpha)+\alpha](1-\alpha e^t)^2}. \tag{8}$$

The first four moments about the origin of DXL are:

$$\mu_1' = \frac{\alpha(3-\alpha-\ln(\alpha)+\alpha\ln(\alpha))}{(1-\alpha)[(2-\ln(\alpha))(1-\alpha)+\alpha]},$$

$$\mu_2' = \frac{\alpha\left(\alpha^2\ln(\alpha)-\alpha^2+4\alpha-\ln(\alpha)+3\right)}{(1-\alpha)^2[(2-\ln(\alpha))(1-\alpha)+\alpha]},$$

$$\mu_3' = \frac{\alpha\left(\alpha^3-3-17\alpha-5\alpha^2-(\alpha-1)\left(1+4\alpha+\alpha^2\right)\ln(\alpha)\right)}{(\alpha-1)^3(2-\alpha+(\alpha-1)\ln(\alpha))},$$

and

$$\mu_4' = \frac{\begin{array}{c}\alpha(\alpha^4\ln(\alpha)-\alpha^4+10\alpha^3\ln(\alpha)+6\alpha^3+66\alpha^2\\+46\alpha-10\alpha\ln(\alpha)-\ln(\alpha)+3)\end{array}}{(1-\alpha)^4[(2-\ln(\alpha))(1-\alpha)+\alpha]}.$$

The first four moments about the mean can be derived using the following relation $\mu_r = E\left(Y-\mu_1'\right)^r$.

The Dispersion Index and Coefficient of variation can be obtained by using formulas:
$DI = \frac{Variance}{Mean}$ and $CV = \frac{SD}{Mean}$

The formula to calculate the coefficient of skewness and kurtosis are:

$$CS = \frac{\mu_3'-3\mu_2'\mu+2\mu^3}{\left(\sigma^2\right)^{\frac{3}{2}}}, \quad CK = \frac{\mu_4'-4\mu_3'\mu+6\mu_2'\mu^2-3\mu^4}{\left(\sigma^2\right)^2}.$$

The descriptive measures given in Table 1 compute numerically using some parameter values to illustrate the behavior of the DXL distribution.

Table 1 Some computational statistics for different choices of parameter

α	E(X)	Var	CS	CK	DI	CV
0.2	0.33097	0.42740	2.40895	10.47247	1.29135	1.97528
0.3	0.59712	0.89685	2.09107	8.90658	1.50197	1.58599
0.4	0.97678	1.73853	1.90051	8.01228	1.77987	1.34988
0.5	1.54154	3.33136	1.76276	7.38359	2.16106	1.18401
0.6	2.43497	6.61571	1.65232	6.89955	2.71696	1.05632
0.7	3.99171	14.4250	1.56044	6.52114	3.61375	0.95148
0.8	7.21381	38.5957	1.48633	6.24025	5.35025	0.86120
0.9	17.1042	178.302	1.43447	6.06248	10.4244	0.78068
0.95	37.0511	758.151	1.41960	6.01594	20.4623	0.74315

According to the results tabulated in Table 1, the DXL model can be an appropriate choice to investigate asymmetric "positively skewed" and dispersion data having a leptokurtic shape.

3.2 Actuarial measures

One of the most difficult challenges in the field of actuarial sciences is estimating market risk. When buying and selling anything, a risk estimate is necessary. A risk estimate is required when purchasing and selling anything. We evaluated the value at risk (VaR) and tail value at risk (TVaR), two significant actuarial variables for the DXL distribution.

The VaR of DXL distribution is attained as $y_p = F(y)$, where y is gained by solving the nonlinear equation given below:

$$\frac{\alpha^{y+1}[(2 - \ln(\alpha) + y)(1 - \alpha) + 1]}{[(2 - \ln(\alpha))(1 - \alpha) + \alpha]} = 1 - p$$

TVaR stands for conditional tail expectation and is calculated as follows:

$$TVaR = \frac{1}{1 - F(y_p)} \sum_{y=y_p}^{\infty} y p(y)$$

$$TVaR = \frac{\alpha^{y_p}(2y_p(\alpha - 1) - y_p^2(\alpha - 1)^2 + (\alpha - 3)\alpha + (y_p(\alpha - 1) - \alpha)(\alpha - 1)\ln(\alpha))}{((\alpha - 1)(2 - \alpha + (\alpha - 1)\ln(\alpha)))(1 - F(y_p))}.$$

(9)

Table 2 shows the values of Value at Risk and Tail Value at Risk for some choices of parameters.

Table 2 Some VaR and TVaR values for the DXL distribution

A	Significance level	VaR_p	$TVaR_p$
0.2	0.75	0.005170	1.334776
	0.80	0.163748	1.967961
	0.85	0.367160	2.794348
	0.90	0.652068	3.975490
	0.95	1.134926	6.030100
	0.99	2.240058	10.91669
0.4	0.75	0.996260	3.910052
	0.80	1.297106	4.546909
	0.85	1.679901	5.373250
	0.90	2.211287	6.544759
	0.95	3.102234	8.557670
	0.99	5.112269	13.24370
0.6	0.75	3.089965	7.477126
	0.80	3.659892	8.328814
	0.85	4.379046	9.420512
	0.90	5.369009	10.94754
	0.95	7.013888	13.52952
	0.99	10.68628	19.41472
0.8	0.75	9.694302	17.58379
	0.80	11.03643	19.13533
	0.85	12.72116	21.10695
	0.90	15.02889	23.84066
	0.95	18.84439	28.41909
	0.99	27.31835	38.73834
0.9	0.75	23.11139	37.50574
	0.80	25.97005	40.46149
	0.85	29.55440	44.21034
	0.90	34.45917	49.39851
	0.95	42.56035	58.07059
	0.99	60.53395	77.57300

4 Parameter estimation

Assume $y_1, y_2, \ldots, y_n$ to be a random sample of size n from the DXL distribution. The Log-likelihood function is given by:

$$\ell = 2n\ln(1-\alpha) - n\ln[(2-\ln\alpha)(1-\alpha)+\alpha]$$
$$+ \ln\alpha \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \ln(2-\ln\alpha+y_i), \quad (10)$$

The log-likelihood equation is obtained by differentiating the above equation with respect to the parameter α:

$$\frac{\partial \ell}{\partial \alpha} = \frac{-2n}{(1-\alpha)} + \frac{n(1-\alpha\ln\alpha)}{\alpha[(2-\ln\alpha)(1-\alpha)+\alpha]}$$
$$+ \frac{1}{\alpha} \sum_{i=1}^{n} y_i - \frac{1}{\alpha} \sum_{i=1}^{n} \frac{1}{(2-\ln\alpha+y_i)}. \quad (11)$$

It is noted that Eq. (11) cannot have an explicit solution. This goal must be solved numerically using an iterative technique like Newton–Raphson. For this purpose, we use the *fitdistrplus (version 1.1–8)* package of the R (version 4.2.2, 2022) software [23].

5 Simulation study

We run a simulation study with finite sample sizes to test the long-term correctness of the MLEs of the DXLD parameter. Using various parameter values, we created samples of n = 5, 10, 25, 50, 100, and 200 from the DXLD. We study the five parameter scenarios as follows: α = 0.2, 0.4, 0.6, 0.8, and 0.9. In this scenario, the iteration is repeated 10,000 times. As a result, we computed the average estimate (AVEs), absolute average bias (AABs), and mean square error (MSEs) given by

$$AABS_\alpha = \sum_{i=1}^{N} \frac{|\widehat{\alpha}_j - \alpha|}{N}, \text{ and } MSE_\alpha = \sum_{i=1}^{N} \frac{(\widehat{\alpha}_j - \alpha)^2}{N}$$

Table 3 summarizes the findings. As can be observed, the MSEs associated with each estimate fall as the sample size increases. This demonstrates the MLEs' consistent performance.

On the basis of simulation criteria, the MLE approach performs well in estimating the DXL distribution parameter.

6 Empirical study

The DXL distribution will be examined via four datasets originating from various domains. We evaluate our model's effectiveness by contrasting it with the Poisson distribution (PD), the Poisson XLindley distribution (PXLD) [3], the discrete inverted Topp–Leone distribution (DITLD) [12], the discrete Bilal distribution (DBD) [6], the discrete Burr–Hatke distribution (DBHD) [11], the discrete Rayleigh distribution (DRD) [24], and the discrete Pareto distribution (DPrD) [19]. The MLE approach is used to estimate the parameters. Moreover, different discrimination criteria, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC), are employed to identify the best-fit probability distribution. Furthermore, Kolmogorov–Smirnov (KS) statistics and Chi-Square are used to assess the suitability of competing models. The mathematical expressions of Chi-Square and Kolmogorov–Smirnov statistics are given as follows:

$$\chi^2 = \sum_{k=1}^{m} \frac{(o_k - e_k)^2}{e_k}, \text{ KS} = \max\left\{\frac{k}{m} - z_k, z_k - \frac{k-1}{m}\right\},$$

where e_k and o_k are expected and observed frequencies of the kth class, respectively.

Example I The first dataset is about a biological experiment, originally analyzed [7]. The data are given in Table 5. In an experiment conducted at random on 8 hills with 15 repetitions, the investigator tallies the number of borers per hill of corn. Table 4 portrays MLEs with respective standard errors (SE) of competitive distributions. We also compute the 95% confidence intervals (CI) for the estimates. Moreover, to get a closer picture, we reported the observed frequencies (OF) and empirical expected frequencies (EF) with respective goodness-of-fit (GF) measures in Table 5.

According to Table 5, our proposed distribution provides the minimum values of the mentioned discriminant criteria and the highest p values for the biological experiment dataset. Figures 3 and 4 show the fitted PMF, CDF, and PP plots, which back up the empirical data.

Example II The second application is associated with the failure of 15 electronic machines during an accelerated life examination [20]. The data are: 1, 5, 6, 11, 12, 19, 20, 22, 23, 31, 37, 46, 54, 60, and 66. Table 6 provides the MLEs with their SE for all fitted models along with their 95% CI. Furthermore, in Table 7, the observed and empirically expected frequencies with respective GF measures are reported.

From the findings listed in Table 7, it is found that the DXL distributions work quite well for discussing the second dataset. The fitted CDF and PP plots are displayed in Fig. 5, which supports the empirical results provided in Table 7.

Table 3 Simulation results for different choices of parameter

Measures	n	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 0.9$
AVEs	5	0.1699	0.3593	0.5612	0.7756	0.8875
	10	0.1817	0.3761	0.5766	0.7862	0.8931
	20	0.1896	0.3833	0.5847	0.7908	0.8960
	50	0.1935	0.3890	0.5889	0.7933	0.8972
	100	0.1957	0.3911	0.5901	0.7943	0.8977
	200	0.1961	0.3918	0.5910	0.7946	0.8981
AABs	5	0.0301	0.0407	0.0388	0.0244	0.0125
	10	0.0183	0.0239	0.0234	0.0138	0.0069
	20	0.0104	0.0167	0.0153	0.0092	0.0040
	50	0.0065	0.0110	0.0111	0.0067	0.0028
	100	0.0043	0.0089	0.0099	0.0057	0.0023
	200	0.0039	0.0082	0.0090	0.0054	0.0019
MSEs	5	0.0164	0.0205	0.0144	0.0054	0.0016
	10	0.0086	0.0096	0.0067	0.0022	0.0006
	20	0.0044	0.0047	0.0031	0.0010	0.0003
	50	0.0017	0.0018	0.0012	0.0004	0.0001
	100	0.0009	0.0009	0.0006	0.0002	0.0001
	200	0.0004	0.0005	0.0004	0.0001	0.0000

Table 4 MLEs, SE, and 95% CI for the first dataset

Parameter		Model							
		DXLD	PXLD	DBD	DBHD	DRD	DPrD	DITLD	PD
α	MLE	0.4915	0.8661	2.3767	0.8655	1.8743	1.1112	1.9840	1.4833
	SE	0.0242	0.0822	0.1598	0.0385	0.0874	0.1027	0.1832	0.1112
95% CI	LCI	0.4441	0.7050	2.0635	0.7900	1.7030	0.9099	1.6249	1.2653
	UCI	0.5389	1.0272	2.6899	0.9410	2.0456	1.3125	2.3431	1.7013

Table 5 Observed, expected, and GF for the first dataset

Expected frequencies

| X	OF	DXLD	PXLD	DBD	DBHD	DRD	DPrD	DITLD	PD
0	43	44.979	47.124	32.741	68.070	15.920	64.451	52.189	27.227
1	35	30.264	29.231	39.589	21.966	36.171	20.149	30.424	40.385
2	17	18.884	17.796	24.275	10.514	34.577	9.6859	14.112	29.952
3	11	11.252	10.679	12.505	5.9829	21.025	5.6470	7.4663	14.809
4	5	6.4988	6.3345	5.9678	3.7540	8.8891	3.6801	4.3900	5.4916
5	4	3.6701	3.7225	2.7359	2.5074	2.7044	2.5797	2.7906	1.6291
6	1	2.0378	2.1706	1.2256	1.7488	0.6021	1.9040	1.8811	0.4027
7	2	1.1166	1.2574	0.5414	1.2588	0.0990	1.4603	1.3271	0.0853
8	2	1.2978	1.6861	0.4193	4.1983	0.0133	10.443	5.4195	0.0189
Total	120	120	120	120	120	120	120	120	120
$-l$		200.42	200.63	204.68	214.05	235.23	220.62	205.15	219.19
AIC		402.84	403.26	411.35	430.10	472.45	443.24	412.30	440.38
BIC		405.63	406.04	414.14	432.89	475.24	446.02	415.09	443.16
Chi-Square		1.4628	1.8289	9.6431	25.196	60.049	36.243	6.9771	21.760
p value		0.8332	0.7672	0.0469	0.0000	0.0000	0.0000	0.0000	0.0000
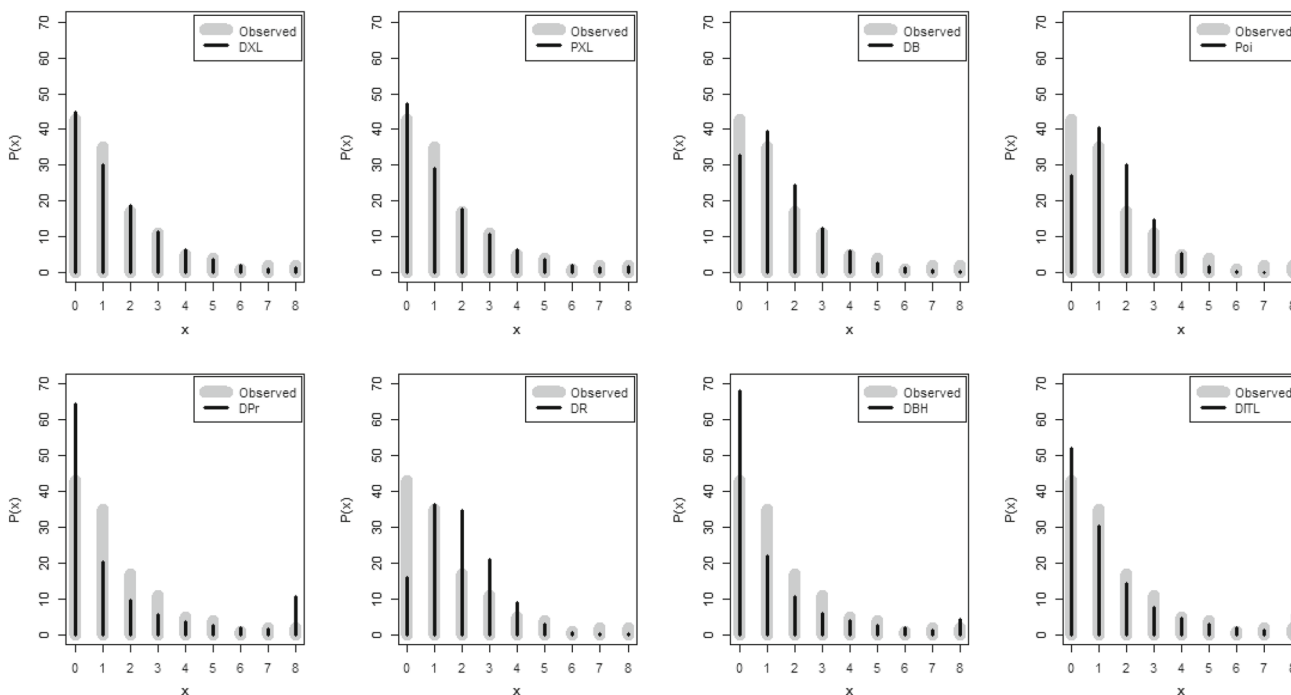
Fig. 3 Fitted PMF for the first dataset

Fig. 4 Fitted CDF (left panel) and PP (right panel) plots for the first dataset
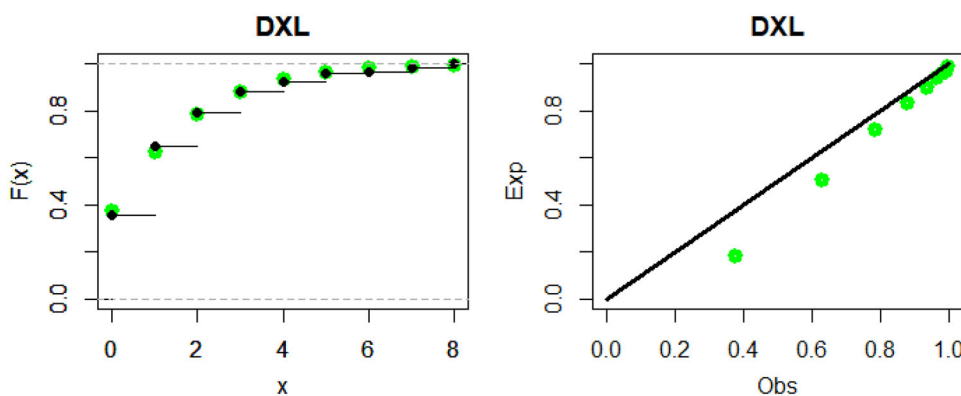
Example III The next discrete dataset is about the number of epileptic seizure tallies [8]. Similarly, the MLEs, SE, and 95% confidence interval for this dataset are presented in Table 8. The observed and expected frequencies and GF are given in Table 9.

According to Table 8, it is found that the DXL distribution appears to be the best among all the competitive models considered. The fitted PMF is graphically displayed in Figs. 6 and 7 which supports the results provided in Table 9.

Example IV The fourth dataset is related to the number of forest fires in Greece between July 1 and August 31, 1998. This dataset is reported in [18]. Table 10 shows the MLEs for the competing models, standard errors, and 95% CI for

the estimations. Table 11 shows the GF measures for the distributions that were examined.

From Table 11, it can be observed that the DXL distribution appears to be the best choice for analyzing the fourth dataset. In Fig. 8, the fitted CDF and PP plots are plotted, which also supports the findings listed in Table 11.

7 Conclusion

The discrete XLindley distribution is a novel discrete distribution derived in this article. The proposed model may be the

Table 6 MLEs, SE, and 95% CI for the second dataset

Parameter		Model						
		DXLD	DITLD	DBHD	PD	DPrD	DBD	DRD
α	MLE	0.9343	0.4178	0.9992	25.533	0.3283	0.9707	24.382
	SE	0.0116	0.1079	0.0076	1.3548	0.0847	0.0054	3.1481
95% CI	LCI	0.9116	0.2063	0.9843	24.878	0.1622	0.9602	18.212
	UCI	0.9571	0.6293	1.0000	30.188	0.4946	0.9812	30.552

Table 7 GF measures for the second dataset

Statistic	Model						
	DXLD	DITLD	DBHD	DP	DPrD	DBD	DRD
$-l$	64.429	74.491	91.368	151.21	77.402	64.784	66.394
AIC	130.86	150.98	184.74	304.41	156.80	131.57	134.79
BIC	131.57	151.69	185.44	305.12	157.51	132.28	135.50
KS	0.1070	0.3590	0.7910	0.3810	0.4061	0.1140	0.2160
p value	0.9900	0.0310	0.0000	0.0180	0.0098	0.9710	0.4300

Fig. 5 Fitted CDF (left panel) and PP (right panel) plots for the second dataset
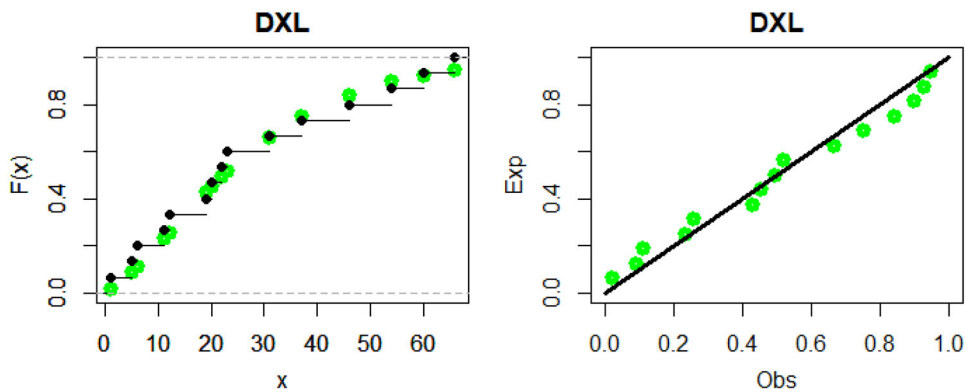

Table 8 MLE, SE, and 95% CI for the third dataset

Parameter		Model							
		DXLD	PXLD	DBD	DBHD	DRD	DPrD	DITLD	PD
α	MLE	0.5007	0.8361	2.4564	0.8702	1.8678	1.0787	1.9045	1.5441
	SE	0.0140	0.0459	0.0964	0.0220	0.0510	0.0582	0.1027	0.0663
CI	LCI	0.4732	0.7463	2.2675	0.8271	1.7678	0.9646	1.7032	1.4142
	UCI	0.5281	0.9259	2.6453	0.9133	1.9678	1.1928	2.1058	1.6740

best solution for modeling asymmetric data with overdispersion phenomena. Several properties of the new model have been derived. It was discovered that all of its attributes can be stated in closed forms, which makes the new model more appealing because it can be used in many studies, particularly time series and regression. To estimate the model parameter, the maximum likelihood estimation approach is applied. Actuarial indicators such as the value at risk and tail value at risk of the proposed distribution are calculated to quantify market risk in a portfolio of instruments. To illustrate the flexibility of the proposed discrete model, four distinctive

Table 9 GF measures for the third dataset

| X	OF	DXLD	PXLD	DBD	DBHD	DRD	DPrD	DITLD	PD
				Expected frequencies					
0	126	127.69	134.01	91.510	198.28	46.867	184.81	148.06	74.940
1	80	87.683	84.746	113.87	64.122	106.28	58.874	88.349	115.71
2	59	55.795	52.559	72.068	30.775	101.22	28.629	41.886	89.338
3	42	33.891	32.113	38.301	17.569	61.198	16.832	22.551	45.982
4	24	19.951	19.390	18.847	11.063	25.678	11.042	13.452	17.750
5	8	11.482	11.595	8.9033	7.4179	7.7386	7.7831	8.6560	5.4816
6	5	6.4965	6.8783	4.1075	5.1946	1.7034	5.7714	5.8972	1.4107
7	4	3.6270	4.0530	1.8678	3.7549	0.2764	4.4446	4.1996	0.3112
8	3	4.3888	5.6536	1.5216	12.824	0.0365	32.807	17.947	0.0722
Total	351	351	351	351	351	351	351	351	351
$-l$		594.98	596.77	606.76	644.90	672.30	664.41	620.06	636.05
AIC		1192.0	1195.5	1215.5	1291.8	1346.6	1330.8	1242.1	1274.1
BIC		1195.8	1199.4	1219.4	1295.7	1350.5	1334.7	1246.0	1278.0
Chi-Square		5.3145	8.0562	30.006	110.86	174.61	136.03	46.659	80.913
p value		0.5042	0.2340	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
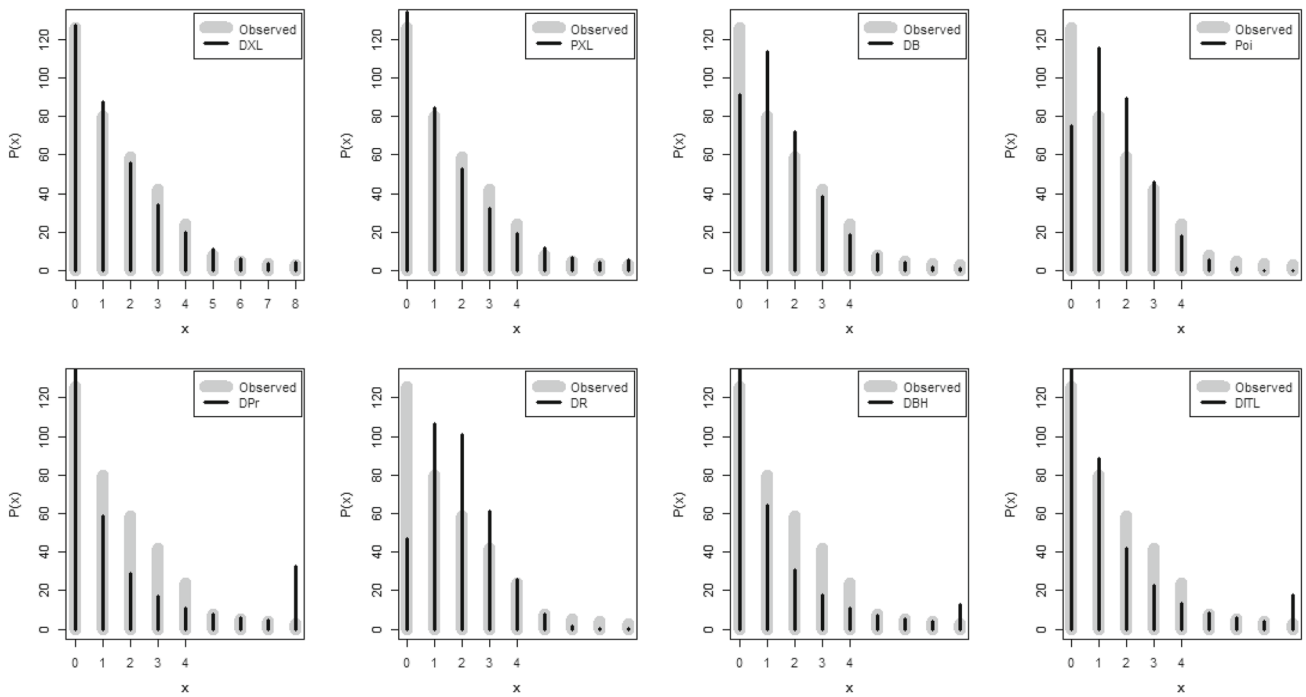
Fig. 6 Fitted PMF for the third dataset

Fig. 7 Fitted CDF (left panel) and PP (right panel) plots for the third dataset
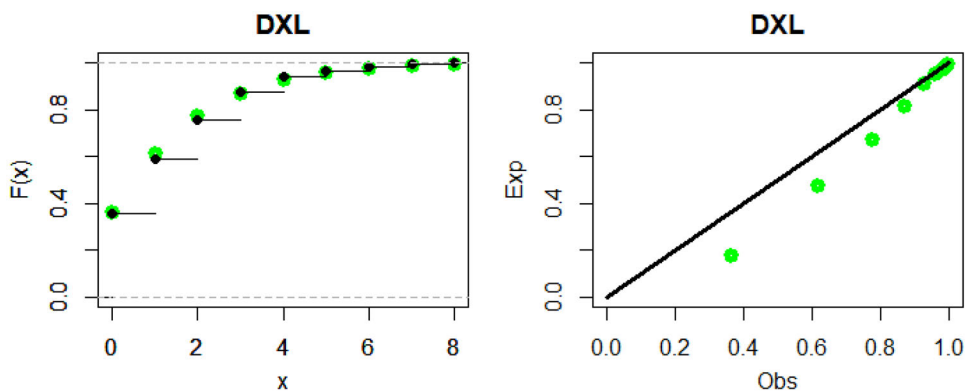

Table 10 MLEs, SE, and 95% CI for the fourth dataset

Parameter		Model						
		DXLD	DITLD	DBHD	DRD	DPrD	DBD	PD
α	MLE	0.7540	0.8976	0.9837	5.6794	0.6047	7.0506	5.3984
	SE	0.0137	0.0810	0.0127	0.2567	0.0546	0.4598	0.2095
CI	LCI	0.7272	0.7388	0.9588	5.1763	0.4977	6.1494	4.9878
	UCI	0.7809	1.0564	1.0086	6.1825	0.7117	7.9518	5.8090

Table 11 GF measures for dataset IV

Statistic	Model						
	DXLD	DITLD	DBHD	DRD	DPrD	DBD	PD
$-l$	339.80	367.52	407.16	385.25	389.64	346.90	467.83
AIC	681.60	737.05	816.31	772.49	781.27	695.80	937.65
BIC	684.42	739.86	819.12	775.31	784.08	698.62	940.47
KS	0.1060	0.2880	0.5470	0.2190	0.3550	0.0965	0.2550
p value	0.1300	0.0000	0.0000	0.0000	0.0000	0.2000	0.0000

Fig. 8 Empirical CDF (left panel) and PP (right panel) plots for the fourth dataset
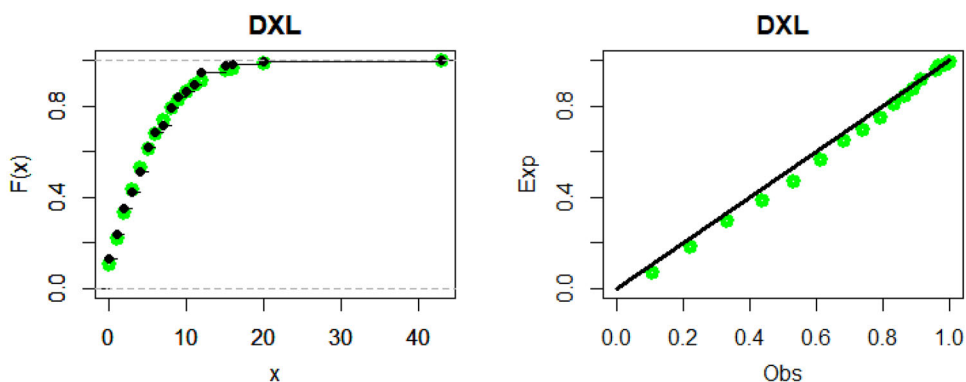
real datasets are utilized in various fields. Finally, we hope that the DXL distribution attracts a wider set of applications in various fields.

Author contributions All authors contribute equally.

Data availability The data are given in the manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Afify, A.Z., Ahsan-ul-Haq, M., Aljohani, H.M., Alghamdi, A.S., Babar, A., Gómez, H.W.: A new one-parameter discrete exponential distribution: properties, inference, and applications to COVID-19 data. King Saud Univ. J. Sci. **1**, 102199 (2022)

2. Ahsan-ul-Haq, M.: On poisson moment exponential distribution with applications. Ann. of Data Sci. **2022**, 1–16 (2022)

3. Ahsan-ul-Haq, M., Al-Bossly, A., El-Morshedy, M., Eliwa, M.S.: Poisson XLindley distribution for count data: statistical and reliability properties with estimation techniques and inference. Comput. Intell. Neurosci. **2022**, 1–16 (2022)

4. Ahsan-ul-Haq, M., Zafar, J.: A new one-parameter discrete probability distribution with its neutrosophic extension: mathematical properties and applications. J Data Sci Anal Int (2023). https://doi.org/10.1007/s41060-023-00382-z

5. Alghamdi, A.S., Ahsan-ul-Haq, M., Babar, A., Aljohani, H.M., Afify, A.Z.: The discrete power-Ailamujia distribution: properties, inference, and applications. AIMS Math. **7**(5), 8344–8360 (2022)

6. Altun, E., El-Morshedy, M., Eliwa, M.S.: A study on discrete Bilal distribution with properties and applications on integer-valued autoregressive process. Revstat. Stat. J. **18**, 70–99 (2020)

7. Beall, G.: The fit and significance of contagious distributions when applied to observations on larval insects. Ecology **21**, 460–474 (1940)

8. Chakraborty, S.: On some distributional properties of the family of weighted generalized Poisson distribution. Commun. Stat. - Theory Methods **39**, 2767–2788 (2010)

9. Chakraborty, S.: Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. J. Stat. Distrib. Appl. **2**, 1–30 (2015)

10. Chouia, S., Zeghdoudi, H.: The XLindley distribution: properties and application. J. Stat. Theory Appl. **20**, 318 (2021)

11. El-Morshedy, M., Eliwa, M.S., Altun, E.: Discrete burr-hatke distribution with properties, estimation methods and regression model. IEEE Access **8**, 74359–74370 (2020)

12. Eldeeb, A.S., Ahsan-ul-Haq, M., Babar, A.: A discrete analog of inverted topp-leone distribution: properties, estimation and applications. Int. J. Anal. Appl. **19**, 695–708 (2021)

13. Eldeeb, A.S., Ahsan-ul-Haq, M., Eliwa, M.S.: A discrete Ramos-Louzada distribution for asymmetric and over-dispersed data with leptokurtic-shaped: properties and various estimation techniques with inference. AIMS Math. **7**, 1726–1741 (2021)

14. Eliwa, M.S., Ahsan-ul-Haq, M., Al-Bossly, A., El-Morshedy, M.: A unit probabilistic model for proportion and asymmetric data: properties and estimation techniques with application to model data from SC16 and P3 algorithms. Math. Probl. Eng. **2022**, 1–13 (2022)

15. Gómez-Déniz, E., Calderín-Ojeda, E.: The discrete Lindley distribution: properties and applications. J. Stat. Comput. Simul. **81**, 1405–1416 (2011)

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

17. Jazi, M.A., Lai, C.D., Alamatsaz, M.H.: A discrete inverse Weibull distribution and estimation of its parameters. Stat. Methodol. **7**, 121–132 (2010)

18. Karlis, D., Xekalaki, E., Lipitakis, E.A.: On some discrete valued time series models based on mixtures and thinning. Proceedings of the fifth hellenic-european conference on computer mathematics and its applications, 872–877 (2001)

19. Krishna, H., Pundir, P.S.: Discrete Burr and discrete Pareto distributions. Stat. Methodol. **6**, 177–188 (2009)

20. Lawless, J.F.: Statistical models and methods for lifetime data, vol. 362. Wiley, Hoboken (2011)

21. Maya, R., Irshad, M.R., Chesneau, C., Nitin, S.L., Shibu, D.S.: On discrete poisson-mirra distribution: regression, INAR (1). Process Appl Axioms **11**, 1–27 (2022)

22. Meriem, B., Gemeay, A.M., Almetwally, E.M., Halim, Z., Alshawarbeh, E., Abdulrahman, A.T., Hussam, E.: The power XLindley distribution: statistical inference, fuzzy reliability, and COVID-19 application. J. Funct. Spaces **2022**, 1–21 (2022)

23. R Core Team. (2022). R: a language and environment for statistical computing. R foundation for statistical computing: Vienna, Austria, 2021. Available online: https://www.R-project.org/

24. Roy, D.: Discrete rayleigh distribution. IEEE Trans. Reliab. **53**, 255–260 (2004)

25. Sankaran, M.: The discrete poisson-lindley distribution. Biometrics **26**, 145–149 (1970)