International Journal of Data Science and Analytics
https://doi.org/10.1007/s41060-023-00382-z

REGULAR PAPER

A new one-parameter discrete probability distribution with its neutrosophic extension: mathematical properties and applications

Muhammad Ahsan-ul-Haq[1] · **Javeria Zafar**[1]

Received: 24 November 2022 / Accepted: 14 January 2023

Abstract

Count data modeling's significance and its applicability to real-world occurrences have been emphasized in a number of research studies. The purpose of this work is to introduce a new one-parameter discrete distribution for the modeling of count datasets. Some mathematical properties, such as reliability measures, characteristic function, moment-generating function, and associated measurements, such as mean, variance, skewness, kurtosis, and index of dispersion, have been derived and studied. The nature of the probability mass function and failure rate function has been studied graphically. The model parameter is estimated using renowned maximum likelihood estimation methods. A neutrosophic extension of the new model is also introduced for the modeling of interval datasets. In addition, the proposed distribution's applicability was compared to that of other discrete distributions. The study's findings show that the novel discrete distribution is a very appealing alternative to some other discrete competitive distributions.

Keywords Infinite discretization · Ramos–Louzada distribution · Neutrosophic statistics · Count data · Analysis

1 Introduction

Count data modeling is used to examine non-negative integer outcomes in various disciplines of study such as insurance, medicine, psychology, and engineering. Various datasets may possess different characteristics and hence must have different count data models. Many count data mostly follow binomial, Poisson, geometric, truncated Poisson, and negative binomial distributions.

In the last few decades, the discretization of continuous probability distribution gets great attention. Moreover, many authors introduced new discrete models, some are given as discrete Weibull [28], discrete Rayleigh [28], discrete Lindley [18], discrete Xgamma [26], discrete Quasi-Xgamma [27], discrete Burr–Hatke [15], Poisson Ailamujia [22], discrete natural Lindley [6], discrete Nadharajah and Haghighi [14], discrete Ramos–Louzada [17], discrete Inverted Topp–Leone [16], Poisson XLindley [5], Poisson moment exponential [4], discrete moment exponential [1], discrete Power Ailamujia [8], and references therein.

A one-parameter continuous probability distribution for the analysis of instantaneous failures [30]. It is later known as Ramos–Louzada (RL) distribution. The random variable (r.v.) X follows RL distribution if its probability density function (PDF) and cumulative distribution function (CDF) are given by

$$f(y;\,\upsilon)=\frac{\upsilon^2-2\upsilon+y}{\upsilon^2(\upsilon-1)}e^{-\frac{y}{\upsilon}},\quad y>0,\ \ \upsilon\geq 2,\tag{1}$$

and

$$F(y;\,\upsilon)=1-\frac{\upsilon^2-2\upsilon+y}{\upsilon(\upsilon-1)}e^{-\frac{y}{\upsilon}}.\tag{2}$$

Al-Mofleh et al. [7] extend it by incorporating a new parameter to extend its flexibility for the modeling of datasets.

The infinite series discretization approach is when the continuous random variable (r.v.) of interest is defined on $R+$. Thus, if the r.v. variable Y is defined on $R+$, the PMF of X becomes

$$P(X=x,\,\xi)=\frac{f_Y(x;\,\xi)}{\sum_{i=0}^{\infty}f_Y(i;\,\xi)},\quad x\in\mathbb{Z}.\tag{3}$$

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

1 College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan

1.1 Neutrosophic Statistics

The concept of neutrosophic probability as a function NP $:\to [0, 1]^3$ was originally presented by [32], where U is a neutrosophic sample space and defined the probability mapping to take the form $NP(S) = (ch(S), ch(neutS), ch(antiS)) = (\eta, \beta, \tau)$ where $0 \leq \eta, \beta, \tau \leq 1$ and $0 \leq \eta + \beta + \tau \leq 3$. Furthermore, many scholars have studied various neutrosophic probability models such as Poisson, binomial, exponential, uniform, normal, Weibull, Kumaraswamy, generalized Pareto, Maxwell, Lognormal, and Gamma, see [2, 9, 11, 23–25, 29, 31]. In many cases, researchers investigate goodness-of-fit tests, neutrosophic time series prediction, and modeling, such as neutrosophic logarithmic models, neutrosophic moving averages, and neutrosophic linear models, as shown in [3, 10, 13].

Recently, many authors have begun to investigate the concept of the neutrosophic random variable (see Definition 1.3). Zeina and Hatip [34] introduced the first concept of neutrosophic random variables, in addition to fundamental ideas. Far ahead, Granados [19] demonstrated new ideas about neutrosophic random variables, and Granados [20] investigated the independence of neutrosophic random variables.

1.2 Groundworks

In this subsection, we will obtain some well-known concepts that will be useful in the development of this paper. The term Ψ represents the set of sample space, R represents the set of real numbers, and Υ denotes a sample space event, X_N and Y_N denote neutrosophic r.v. Furthermore, we demonstrate certain renowned definitions and characteristics of neutrosophic probability and logic that will be important in creating this neutrosophic probability model.

Definition 1.3.1 Consider the real-valued crisp r.v. X, which has the following definition:

$$X : \Psi \to \mathbb{R}$$

where Ψ is the event space and X_N neutrosophic r.v. as follows:

$$X_N : \Psi \to \mathbb{R}(I)$$

and

$$X_N = X + I$$

The term I represents indeterminacy.

Theorem 1.3.1 (See Granados [21]) Let the neutrosophic r.v. $X_N = X + I$ and the CDF of X is $F_X(x) = P(X \leq x)$. The following assertions are correct:

$$F_{X_N}(x) = F_X(x - I),$$
$$f_{X_N}(x) = f_X(x - I),$$

where F_{X_N} and f_{X_N} are the CDF and PDF of a neutrosophic r.v. X_N, respectively.

Theorem 1.3.2 (See Granados [21]) Let $X_N = X + I$, is the neutrosophic r.v., then the expected value and variance can be derived as follows: $E(X_N) = E(X) + I$ and $V(X_N) = V(X)$.

The main motivation behind this work is to introduce a new flexible discrete probability model using infinite series for the analysis of count observations. The model is named "New Discrete Ramos–Louzada Distribution-NDRL." The new distribution contained compact expressions of its probability mass function (PMF), CDF, moments, and some associated measures. The MLE approach is used to estimate the NDRL distribution parameter. Three datasets from different fields were analyzed using the NDRL distribution. In the end, to study count datasets with indeterminacy, a neutrosophic extension of this model is also presented.

2 Derivation of new model and its properties

The new probability model is derived using the approach given in Eq. (3), and the PMF is

$$P(x; \upsilon) = \frac{(1-\upsilon)^2\left(1 + 2\log(\upsilon) + x(\log(\upsilon))^2\right)\upsilon^x}{\left[(1 + 2\log(\upsilon))(1-\upsilon) + \upsilon(\log(\upsilon))^2\right]}, \quad x = 0, 1, 2, 3, \ldots. \tag{4}$$

where $0.61 \leq \upsilon \leq 1$.

Remark 1 The first derivative of PMF is.

$$\frac{\mathrm{d}p(x)}{\mathrm{d}x} = \frac{(\alpha-1)^2\alpha^x\log(\upsilon)\left(1 + 3\log(\upsilon) + x(\log(\upsilon))^2\right)}{1 - \alpha - 2(\alpha-1)\log(\upsilon) + \alpha(\log(\upsilon))^2},$$

which provides the critical point

$$\widehat{x} = \frac{-1 - 3\log(\upsilon)}{(\log(\upsilon))^2},$$

For $\upsilon < \left(e^{-\frac{1}{3}} = 0.716531\right)$, the critical point is $\left\{\frac{-1-3\log(\upsilon)}{(\log(\upsilon))^2}\right\}$ which is the maximum point of $p(\widehat{x}, \upsilon)$, and for $\upsilon \geq e^{-\frac{1}{3}}$, the probability mass function is a declining function of x. Further, the 2$^{\text{nd}}$ derivative is given by

$$\frac{\mathrm{d}^2p(x)}{\mathrm{d}x^2} = \frac{(\upsilon-1)^2\upsilon^x(\log(\upsilon))^2\left(1 + 4\log(\upsilon) + x(\log(\upsilon))^2\right)}{1 - \upsilon - 2(\upsilon-1)\log(\upsilon) + \upsilon(\log(\upsilon))^2},$$

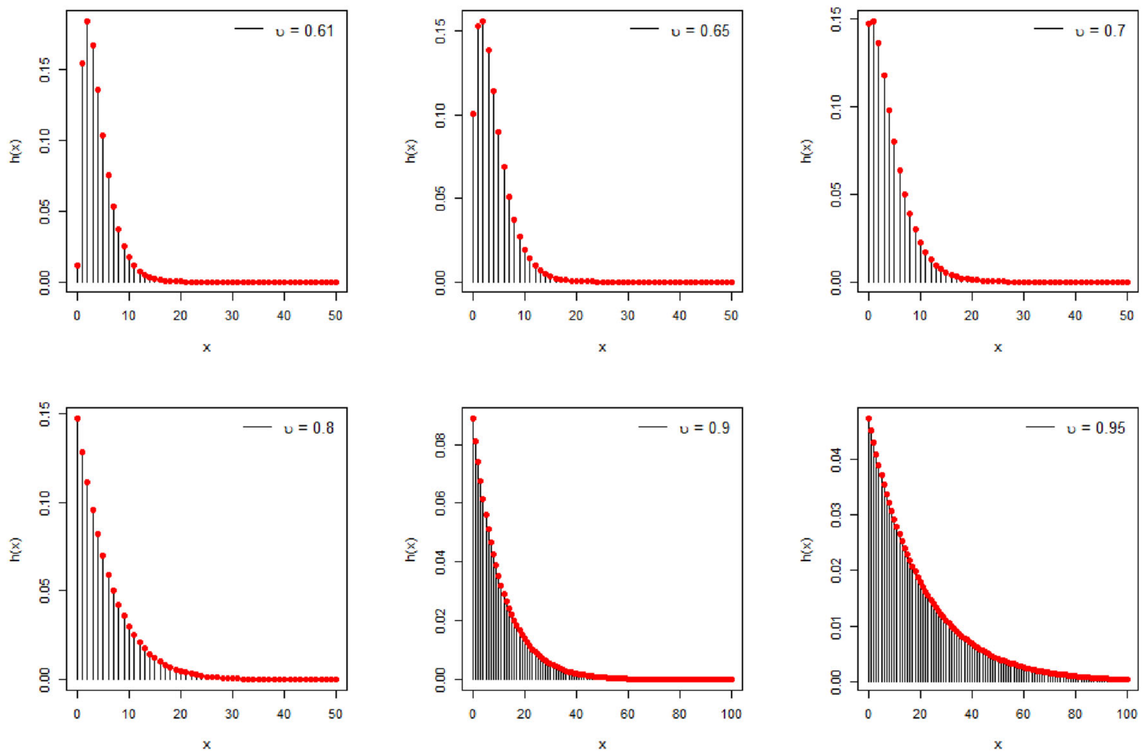

Fig. 1 PMF plots for selected parameter values

Table 1 Mode values

υ	0.61	0.62	0.63	0.64	0.65	0.67	0.68	0.71
Mode	2	2	2	2	2	1	1	0

Therefore, the mode of NDRL distribution is

$$\text{Mode}(X) = \begin{cases} \frac{-1-3\log(\upsilon)}{(\log(\upsilon))^2}, & \text{for} \quad \upsilon < e^{-\frac{1}{3}} \\ 0 & \text{otherwise} \end{cases}$$

The mode values are presented in Table 1. The PMF plots for NDRL distribution are obtainable in Fig. 1.

The cumulative distribution function corresponding to Eq. (4) is

$$F(x;\,\upsilon) = 1 - \frac{\upsilon^{x+1}\left[\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)(1-\upsilon)+(\log(\upsilon))^2\right]}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]}. \tag{5}$$

The survival function of NDRLD is

$$S(x;\,\upsilon) = \frac{\upsilon^{x+1}\left[\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)(1-\upsilon)+(\log(\upsilon))^2\right]}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]}. \tag{6}$$

The hazard rate function (hrf) is

$$h(x;\,\upsilon) = \frac{(1-\upsilon)^2\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)}{\upsilon\left[\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)(1-\upsilon)+(\log(\upsilon))^2\right]}. \tag{7}$$

The behavior of the hrf may be examined using the Glaser technique and the PMF of the NDRL distribution.

$$\rho(x) = -\frac{p'(x)}{p(x)} = -\frac{\log(\upsilon)\left(1+3\log(\upsilon)+x(\log(\upsilon))^2\right)}{1+2\log(\upsilon)+x(\log(\upsilon))^2},$$

and it follows that

$$\rho'(x) = \frac{(\log(\upsilon))^4}{\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)^2} > 0$$

As $\rho'(x) > 0$, the hrf of the NDRL distribution is an increasing function of x. The hrf plots for NDRL distribution are obtainable in Fig. 2.

2.1 Moment-generating function (mgf)

The mgf of r.v. X is thus represented by the notation $M_X(t)$ and is derived as follows:

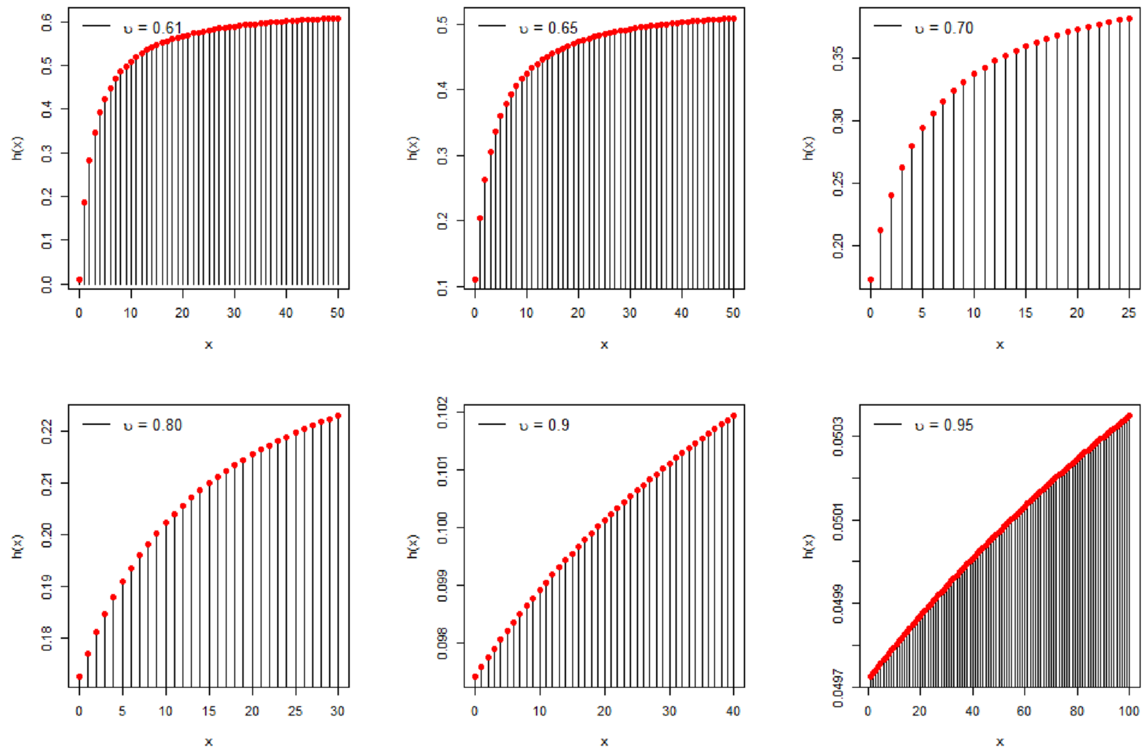

Fig. 2 The hrf visualization plots for NDRL distribution

$$
\begin{aligned}
M_X(t) &= E\left(e^{xt}\right) = \sum_{x=1}^{\infty} e^{xt} P(x) \\
M_X(t) &= \sum_{x=0}^{\infty} e^{xt} \frac{(1-\upsilon)^2\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)\upsilon^x}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]} \\
M_X(t) &= \frac{(1-\upsilon)^2}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]} \\
&\quad\times \sum_{x=0}^{\infty}\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)\left(\upsilon e^t\right)^x \\
M_X(t) &= \frac{(1-\upsilon)^2}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]} \\
&\quad\times \left[(1+2\log(\upsilon))\sum_{x=0}^{\infty}\left(\upsilon e^t\right)^x + (\log(\upsilon))^2 \sum_{x=0}^{\infty} x\left(\upsilon e^t\right)^x\right] \\
M_X(t) &= \frac{(1-\upsilon)^2}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]} \\
&\quad\times \left[\frac{(1+2\log(\upsilon))\left(1-\upsilon e^t\right)+\upsilon e^t(\log(\upsilon))^2}{\left(1-\upsilon e^t\right)^2}\right] \\
M_X(t) &= \frac{(1-\upsilon)^2\left[(1+2\log(\upsilon))\left(1-\upsilon e^t\right)+\upsilon e^t(\log(\upsilon))^2\right]}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]\left(1-\upsilon e^t\right)^2}.
\end{aligned}
\tag{8}
$$

The first four moments about the origin are given below

$$\mu_1' = \frac{\upsilon\left[(1+2\log(\upsilon))(1-\upsilon)+(1+\upsilon)(\log(\upsilon))^2\right]}{(1-\upsilon)\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]}$$

$$\mu_2' = \frac{\upsilon\left[(1+2\log(\upsilon))\left(1-\upsilon^2\right)+(1+2\upsilon)^2(\log(\upsilon))^2\right]}{(1-\upsilon)^2\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]}$$

$$\mu_3' = \frac{\upsilon\left[(1+2\log(\upsilon))\left(1+3\upsilon-3\upsilon^2-\upsilon^3\right)+(\log(\upsilon))^2\left(1+11\upsilon+20\upsilon^2+4\upsilon^3\right)\right]}{(1-\upsilon)^3\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]}$$

$$\mu_4' = \frac{\upsilon\left[(1+2\log(\upsilon))\left(1+10\upsilon-10\upsilon^3-\upsilon^4\right)+(\ln\upsilon)^2\left(1+27\upsilon+115\upsilon^2+116\upsilon^3+20\upsilon^4\right)\right]}{(1-\upsilon)^4\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]}$$

The variance is obtained as

$$
\begin{aligned}
\sigma^2 &= \frac{\upsilon\left[(1+2\log(\upsilon))\left(1-\upsilon^2\right)+(1+2\upsilon)^2(\log(\upsilon))^2\right]}{(1-\upsilon)^2\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]} \\
&\quad - \left(\frac{\upsilon\left[(1+2\log(\upsilon))(1-\upsilon)+(1+\upsilon)(\log(\upsilon))^2\right]}{(1-\upsilon)\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]}\right)^2.
\end{aligned}
$$

The following formula could be employed to compute the dispersion index (DI):

$$\mathrm{DI} = \frac{\mathrm{Var}(X)}{\mathrm{Mean}(X)}$$

Table 2 lists a few descriptive statistics, including mean, variance (Var), dispersion index (DI), coefficients of skewness (CS), kurtosis (CK), and variance (CV).

Table 2 Mean, Var, DI, CS, CK, and CV for some selected values of the parameter

υ	Mean	Var	DI	CS	CK	CV
0.61	4.0539	8.0898	1.4365	6.0724	1.9956	1.42528
0.65	3.8955	10.761	1.4200	6.0096	2.7623	1.18753
0.69	3.9776	13.661	1.5234	6.3722	3.4346	1.07615
0.73	4.2573	17.555	1.6337	6.8426	4.1234	1.01611
0.75	4.4759	20.153	4.5025	1.6848	7.0856	0.99703
0.77	4.7563	23.411	1.7323	7.3258	4.9221	0.98301
0.81	5.5607	33.115	1.8156	7.7827	5.9552	0.96631
0.85	6.8767	51.237	1.8835	8.1915	7.4508	0.96070
0.89	9.2301	91.830	1.9361	8.5356	9.9489	0.96320
0.93	14.367	218.43	1.9735	8.7984	15.204	0.97207
0.97	33.365	1144.5	1.9950	8.9601	34.301	0.98626
0.99	100.01	10,100	100.99	1.9994	8.9953	0.99514

2.2 Characteristic function

The characteristic function can be derived as

$$\phi_X(t) = E\left(e^{ixt}\right) = \sum_{x=0}^{\infty} e^{ixt} P(x)$$

$$\phi_X(t) = \sum_{x=0}^{\infty} e^{ixt} \frac{(1-\upsilon)^2\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)\upsilon^x}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]}$$

$$\phi_X(t) = \frac{(1-\upsilon)^2}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]} \times \sum_{x=0}^{\infty} e^{ixt}\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)\upsilon^x$$

$$\phi_X(t) = \frac{(1-\upsilon)^2}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]} \times \sum_{x=0}^{\infty}\left(1+2\log(\upsilon)+x(\log(\upsilon))^2\right)\left(\upsilon e^{it}\right)^x$$

$$\phi_X(t) = \frac{(1-\upsilon)^2\left[\left[(1+2\log(\upsilon))\left(1-\upsilon e^{it}\right)+\upsilon e^{it}(\log(\upsilon))^2\right]\right]}{\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]\left(1-\upsilon e^{it}\right)^2}. \tag{9}$$

3 Parameter estimation and simulation

In this section, we will look at the MLE approach for estimating the parameter of the NDRL model. Let $x_1, x_2, x_3, \ldots, x_n$ is a random sample from the NDRL model, then its log-likelihood function (l) is

$$l(x;\upsilon) = 2n\log(1-\upsilon) - n\log\left((1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right) + \log(\upsilon)\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\left(1+2\log(\upsilon)+x_i(\log(\upsilon))^2\right), \tag{10}$$

The following non-linear equation results from differentiating Eq. (10) with respect to the parameter.

$$\frac{\partial l(x;\upsilon)}{\partial \upsilon} = \frac{-2n}{(1-\upsilon)} + \frac{n\left(2-3\upsilon+\upsilon(\log(\upsilon))^2\right)}{\upsilon\left[(1+2\log(\upsilon))(1-\upsilon)+\upsilon(\log(\upsilon))^2\right]} + \frac{1}{\upsilon}\sum_{i=1}^{n} x_i + \frac{2}{\upsilon}\sum_{i=1}^{n}\frac{1+x_i(\log(\upsilon))}{\left(1+2\log(\upsilon)+x_i(\log(\upsilon))^2\right)}. \tag{11}$$

Since there is no exact solution to the aforementioned equation, estimates can be obtained by employing an iterative process (Table 3).

The performance of ML estimations of the NDRL distribution was evaluated using a simulation study, and the exercise was carried out with various parameter values, that is $\upsilon = \{0.61, 0.65, 0.70, 0.75\}$. The inverse transformation strategy was used to generate random observations from the NDRL distribution with sample sizes of $n = 10, 20, 50, 100$ and 200. The experiment was reproduced for $N = 10{,}000$ times for each combination of parameters and sample size. The mean square error (MSE), absolute bias (AB), and average estimations (AE) were computed using below measures;

$$\mathrm{AE}(\upsilon) = \frac{1}{10000}\sum_{i=1}^{10000}\widehat{\upsilon_i},$$

$$\mathrm{AB}(\upsilon) = \frac{1}{10000}\sum_{i=1}^{10000}\left(\widehat{\upsilon_i}-\upsilon\right),$$

$$\mathrm{MSE}(\upsilon) = \frac{1}{10000}\sum_{i=1}^{10000}\left(\widehat{\upsilon_i}-\upsilon\right)^2.$$

4 Application

In this part, we demonstrate the modeling flexibility of the NDRL model using datasets from different fields. We will

Table 3 Results of simulation for some parameter values

n	Est.	$\upsilon = 0.61$	$\upsilon = 0.65$	$\upsilon = 0.68$	$\upsilon = 0.70$	$\upsilon = 0.75$
10	AVE	0.623730	0.649356	0.677385	0.696783	0.760320
	AB	0.013730	0.000644	0.002615	0.003217	0.010320
	MSE	0.001069	0.001485	0.001458	0.001568	0.000950
20	AVE	0.618544	0.650478	0.677600	0.698703	0.757954
	AB	0.008544	0.000478	0.002400	0.001297	0.007954
	MSE	0.000582	0.000939	0.000783	0.000663	0.000433
50	AVE	0.614244	0.648942	0.678957	0.698277	0.756162
	AB	0.004244	0.001058	0.001043	0.001723	0.006162
	MSE	0.000273	0.000411	0.000299	0.000258	0.000211
100	AVE	0.612101	0.649055	0.679783	0.698991	0.754200
	AB	0.002101	0.000945	0.000217	0.001009	0.004200
	MSE	0.000162	0.000205	0.000152	0.000123	0.000119
200	AVE	0.611259	0.649806	0.679609	0.700209	0.751064
	AB	0.001259	0.000194	0.000391	0.000209	0.001064
	MSE	0.000095	0.000099	0.000081	0.000056	0.000037

compare the fits of the NDRL distribution to other competing models such as Poisson (Poi), discrete inverse Rayleigh (DIR), natural discrete Lindley (NDL), discrete Pareto (DPr), discrete Burr–Hatke (DBH), discrete inverted Topp–Leone (DITL), and Poisson Ailamujia (PA) distribution. The PMFs of these distributions are presented below;

Model	PMF
Poisson distribution	$P(x;\upsilon) = \frac{\upsilon e^{-\upsilon x}}{x!}$
DIR distribution	$P(x;\upsilon) = e^{\left(-\frac{\upsilon}{(1+x)^2}\right)} - e^{\left(-\frac{\upsilon}{x^2}\right)}$
NDL distribution	$P(x;\upsilon) = \frac{\upsilon^2(2+x)(1-\upsilon)^x}{(1+\upsilon)}$
DPr distribution	$P(x;\upsilon) = e^{-\upsilon\log(1+x)} - e^{-\upsilon\log(2+x)}$
DBH distribution	$P(x;\upsilon) = \left(\frac{1}{x+1} - \frac{\upsilon}{x+2}\right)\upsilon^x$
DITL distribution	$P(x;\upsilon) = \frac{(1+2x)^\upsilon}{(1+x)^{2\upsilon}} - \frac{(3+2x)^\upsilon}{(2+x)^{2\upsilon}}$
PA distribution	$P(x;\upsilon) = \frac{4\upsilon^2(1+x)}{(1+2\upsilon)^{(x+2)}}$

Different criteria are used to select the best-fitted probabilistic models, such as the maximum log-likelihood (l) value, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the Kolmogorov–Smirnov (KS) test with its related p value.

The first dataset is on the number of fires in Greece's forest districts from July 1st to August 31st, 1998 [12], and it was also studied by [17]. Table 4 shows the data observations.

The MLE technique is used to estimate the model parameters for each distribution under consideration. Table 5 includes the MLEs and goodness-of-fit (GOF) measures. For each of the investigated distributions, PP plots are also shown in Fig. 3.

The second dataset is a count of COVID-19 daily fatalities in China from January 23 to March 28 [1]. The data observations are: 3, 3, 4, 5, 5, 6, 6, 7, 7, 7, 8, 8, 9, 10, 11, 11, 13, 3, 14, 15, 16, 17, 22, 22, 24, 26, 26, 27, 28, 29, 30, 31, 31, 35, 38, 38, 42, 43, 44, 45, 46, 47, 52, 57, 64, 65, 71, 73, 73, 86, 89, 97, 97, 97, 98, 105, 108, 109, 114, 118, 121, 136, 142, 143, 146 and 150. Table 6 shows the MLEs as well as the goodness-of-fit measurements. Figure 4 shows PP plots for all distributions evaluated in the second dataset.

The third dataset reflects the number of coronavirus-related fatalities in Pakistan [16]. The 44 fatalities that occurred between March 18, 2020, and April 30, 2020, are a sample. The data observations are 2, 1, 0, 2, 1, 1, 1, 1, 2, 1, 2, 7, 5, 1, 7, 6, 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 and 32. Table 7 includes the MLEs and goodness-of-fit metrics. For each of the investigated distributions, PP charts are shown in Fig. 5.

According to Tables 6 and 7, the suggested distribution is the best model for evaluating all types of datasets since it has the minimum AIC and BIC and the highest log-likelihood and KS values.

Table 4 Forest fire in Greece

Forest fires in Greece	0	1	2	3	4	5	6	7	8	9	10	11	12	15	16	20	43
No. of fires	16	13	14	9	11	9	4	3	7	6	2	4	6	3	1	1	1

Table 5 MLEs, SE, and GOF measures for the first dataset

Dist	MLEs (S.E.)	l	AIC	BIC	KS (P value)
NDRL	0.79339 (0.02390)	– 301.10	604.19	606.90	0.049 (0.315)
Poisson	5.20000 (0.21742)	– 434.16	870.32	873.02	0.282 (0.000)
DIR	3.51980 (0.37480)	– 360.90	723.80	726.50	0.413 (0.000)
NDL	0.25669 (0.01524)	– 302.73	607.47	610.17	0.169 (0.004)
DPr	0.62502 (0.05970)	– 339.05	680.10	682.80	0.352 (0.000)
DBH	0.98332 (0.01364)	– 352.42	706.85	709.55	0.532 (0.000)
DITL	0.93388 (0.08917)	– 320.45	642.90	645.60	0.277 (0.000)
PA	0.19231 (0.01526)	– 305.93	613.85	616.55	0.095 (0.280)

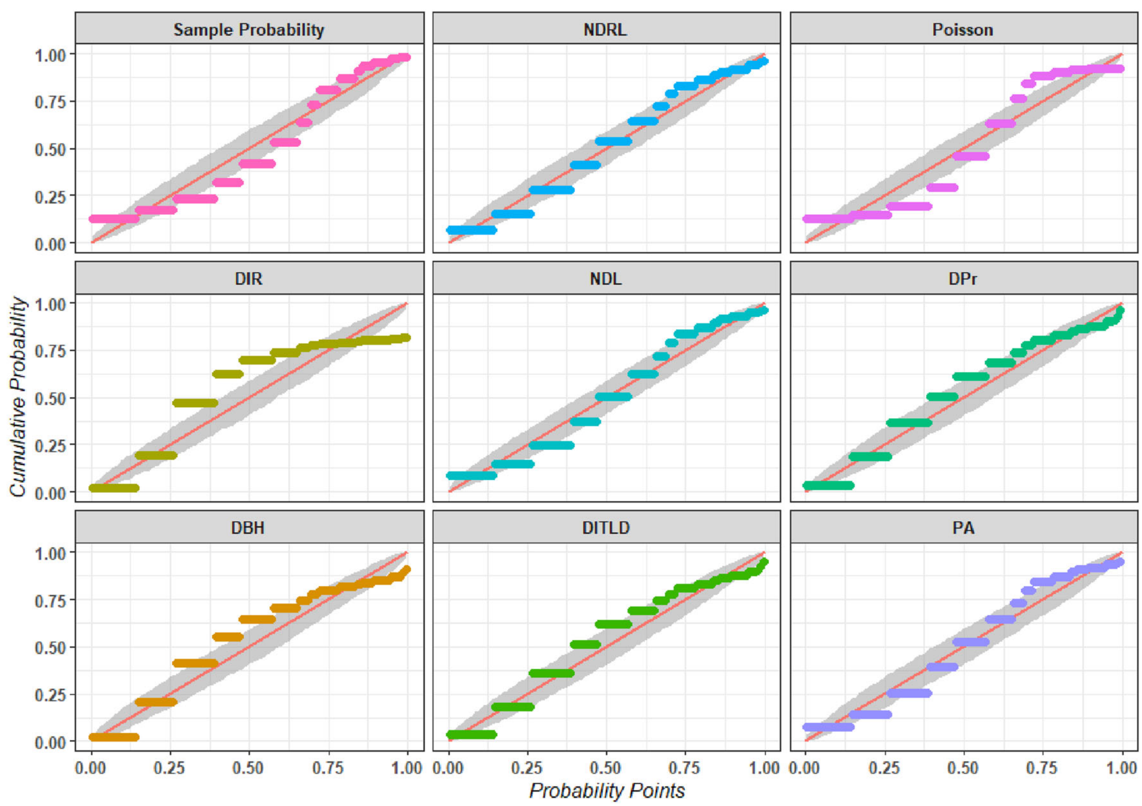

Fig. 3 PP plots for the first dataset

Table 6 MLEs, SE, and GOF measures for the second dataset

Dist	MLEs (S.E.)	l	AIC	BIC	KS (P value)
NDRL	0.97982 (0.00250)	− 324.31	650.62	652.81	0.086 (0.710)
Poisson	49.5910 (0.86682)	− 1428.1	2858.2	2860.3	0.497 (0.000)
DIR	124.080 (15.7350)	− 381.74	765.48	767.67	0.465 (0.000)
NDL	0.03808 (0.00325)	− 329.84	661.67	663.86	0.173 (0.038)
DPr	0.28780 (0.03543)	− 377.56	757.12	759.31	0.371 (0.000)
DBH	0.99974 (0.00186)	− 458.67	919.34	921.53	0.800 (0.000)
DITL	0.35601 (0.04382)	− 365.44	732.89	735.08	0.316 (0.000)
PA	0.02018 (0.00179)	− 330.84	663.69	665.88	0.169 (0.046)

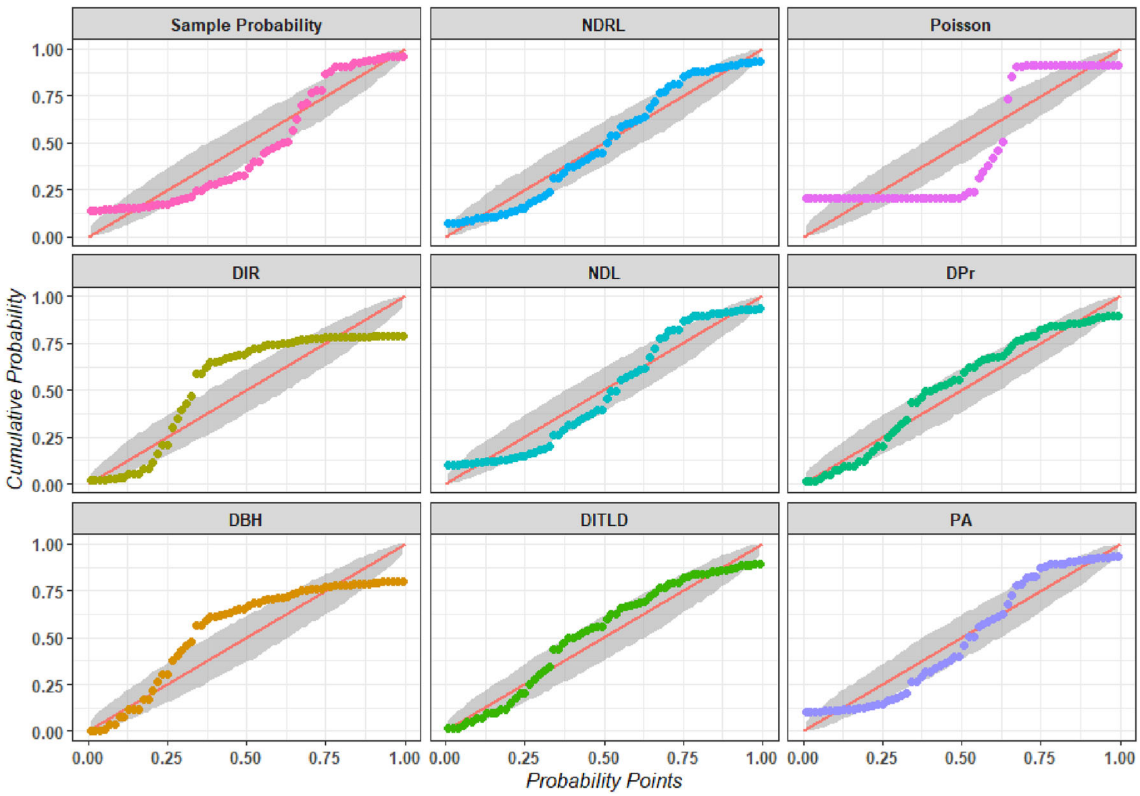

Fig. 4 PP plots for the second dataset

Table 7 MLEs, SE, and GOF measures for the third dataset

Dist	MLEs (S.E.)	l	AIC	BIC	KS (P value)
NDRL	0.89328 (0.01713)	− 145.22	292.43	294.22	0.156 (0.230)
Poisson	9.47730 (0.46410)	− 283.94	569.89	571.67	0.391 (0.000)
DIR	7.42900 (1.26240)	− 166.31	334.61	336.40	0.382 (0.000)
NDL	0.16400 (0.01615)	− 148.44	298.89	300.67	0.237 (0.014)
DPr	0.50214 (0.07575)	− 162.19	326.38	328.17	0.401 (0.000)
DBH	0.99485 (0.01148)	− 175.37	352.74	354.52	0.647 (0.000)
DITL	0.70969 (0.10706)	− 153.04	308.09	309.87	0.318 (0.001)
PA	0.10551 (0.01237)	− 150.15	302.30	304.08	0.198 (0.063)

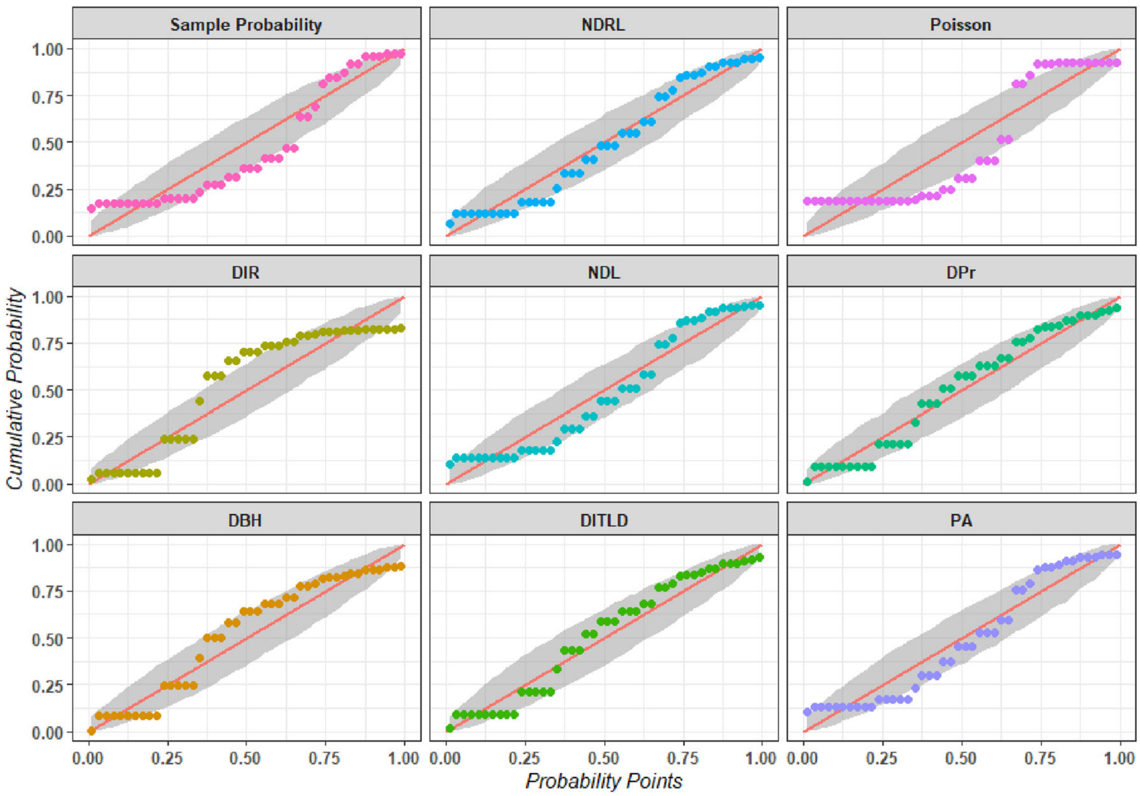

Fig. 5 PP plots for the third dataset

5 Neutrosophic extension of NDRL distribution

Assume Y_N be a neutrosophic random variable. Then say that Y_N has a neutrosophic new discrete Ramos–Louzada distribution denoted by $Y_N \sim \mathrm{NDRL}(\upsilon_N)$ where υ_N is the set with one or more elements (maybe υ_N be an interval). The neutrosophic probability function is given by

$$f_Y(y-I) = \begin{cases} \frac{(1-\upsilon_N)^2\left(1+2\log(\upsilon_N)+(y-I)(\log(\upsilon_N))^2\right)\upsilon_N^{y-I}}{\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+\upsilon_N(\log(\upsilon_N))^2\right]}, & \text{iff} \quad y = I,\ 1+I,\ 2+I, \\ 0, & \text{otherwise} \end{cases}$$

Proof

$$\sum_{y=I}^{\infty} P(y) = \sum_{y=I}^{\infty} \frac{(1-\upsilon_N)^2\left(1+2\log(\upsilon_N)+(y-I)(\log(\upsilon_N))^2\right)\upsilon_N^{y-I}}{\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+\upsilon_N(\log(\upsilon_N))^2\right]},$$

$$\sum_{y=I}^{\infty} P(y) = \frac{(1-\upsilon_N)^2}{\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+\upsilon_N(\log(\upsilon_N))^2\right]} \times \sum_{y=I}^{\infty}\left(1+2\log(\upsilon_N)+(y-I)(\log(\upsilon_N))^2\right)\upsilon_N^{y-I},$$

$$\sum_{y=I}^{\infty} P(y) = \frac{(1-\upsilon_N)^2\upsilon_N}{\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+\upsilon_N(\log(\upsilon_N))^2\right]} \times \left[(1+2\log(\upsilon_N))\frac{1}{1-\upsilon_N}+\frac{\upsilon(\log(\upsilon_N))^2}{(1-\upsilon)^2}\right]$$

$$\sum_{y=I}^{\infty} P(y) = \frac{(1-\upsilon_N)^2\upsilon_N}{\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+\upsilon_N(\log(\upsilon_N))^2\right]} \times \left[\frac{\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+\upsilon_N(\log(\upsilon_N))^2\right]}{(1-\upsilon_N)^2\upsilon_N}\right],$$

$$\sum_{y=I}^{\infty} P(y) = 1$$

By Theorem 1.3, then mean and variance can be written as

$$E(X_N) = \left\{\frac{\upsilon_N\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+(1+\upsilon_N)(\log(\upsilon_N))^2\right]}{(1-\upsilon_N)\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+\upsilon_N(\log(\upsilon_N))^2\right]}\right\}+I,$$

$$\mathrm{Var}(X_N) = \frac{\upsilon_N\left[(1+2\log(\upsilon_N))\left(1-\upsilon_N^2\right)+(1+2\upsilon_N)^2(\log(\upsilon_N))^2\right]}{(1-\upsilon_N)^2\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+\upsilon_N(\log(\upsilon_N))^2\right]} - \left(\frac{\upsilon_N\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+(1+\upsilon_N)(\log(\upsilon_N))^2\right]}{(1-\upsilon_N)\left[(1+2\log(\upsilon_N))(1-\upsilon_N)+\upsilon_N(\log(\upsilon_N))^2\right]}\right)^2.$$

6 Conclusion

A novel discrete probability distribution with one parameter is proposed. The mathematical properties of the new distribution are derived, including the moment-generating function, characteristic function, survival function, and hazard function. To estimate the model parameters, the maximum likelihood technique is utilized. Finally, three datasets are utilized to illustrate the flexibility of the proposed distribution over several competing distributions. The suggested distribution is shown to have better fits than the other distributions tested. We also presented a neutrosophic extension entitled "discrete neutrosophic Ramos–Louzada distribution" to evaluate datasets with uncertainty and derived some of its properties.

Thus, in future, we will further extend this distribution for interval statistics (IS), which is the generalized form of neutrosophic statistics (NS). For more detail about IS and NS, readers can consult the following reference [33].

Author contribution Both authors equally contributed to this work.

Funding No funding received.

Data availability The data are given in the manuscript.

Declarations

Conflict of interest The authors 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.: A new Cramèr–von Mises goodness-of-fit test under uncertainty a new Cramèr–von Mises goodness-of-fit test under uncertainty. Neutrosophic Sets Syst. **49**, 262–268 (2022)
3. Ahsan-ul-Haq, M.: Neutrosophic Kumaraswamy distribution with engineering application. Neutrosophic Sets Syst. **49**, 269–276 (2022)
4. Ahsan-ul-Haq, M.: On Poisson moment exponential distribution with applications. Ann. Data Sci. **2022**, 1–16 (2022)
5. 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)
6. Al-Babtain, A.A., Ahmed, A.H.N., Afify, A.Z.: A new discrete analog of the continuous Lindley distribution, with reliability applications. Entropy **22**, 1–18 (2020)
7. Al-Mofleh, H., Afify, A.Z., Ibrahim, N.A.: A new extended two-parameter distribution: properties, estimation methods, and applications in medicine and geology. Mathematics **8**, 1–20 (2020)
8. 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**, 8344–8360 (2022)
9. Alhabib, R., Ranna, M.M., Farah, H., Salama, A.A.: Some neutrosophic probability distributions. Neutrosophic Sets Syst. **22**, 30–38 (2018)
10. Alhabib, R., Salama, A.: Using moving averages to pave the neutrosophic time series. Int. J. Neutrosophic Sci. **3**, 14–20 (2020)
11. Alhasan, K.F.H., Smarandache, F.: Neutrosophic Weibull distribution and neutrosophic family Weibull distribution. Infin. Study **28**, 191 (2019)
12. Bakouch, H.S., Jazi, M.A., Nadarajah, S.: A new discrete distribution. Statistics **48**, 200–240 (2014)
13. Cruzaty, L.E.V., Tomalá, M.R., Gallo, C.M.C.: A neutrosophic statistic method to predict tax time series in ecuador. Neutrosophic Sets Syst. **34**, 33–39 (2020)
14. Dey, S.: Univariate discrete Nadarajah and Haghighi distribution: properties and different methods of estimation. Statistica **80**, 301–330 (2020)
15. 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)
16. 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)
17. 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)
18. Gómez-Déniz, E., Calderín-Ojeda, E.: The discrete lindley distribution: properties and applications. J. Stat. Comput. Simul. **81**, 1405–1416 (2011)
19. Granados, C.: New notions on neutrosophic random variables. Neutrosophic Sets Syst. **47**, 286–297 (2021)
20. Granados, C.: On independence neutrosophic random variables. Neutrosophic Sets Syst. **47**, 541–557 (2021)
21. Granados, C.: Some discrete neutrosophic distributions with neutrosophic parameters based on neutrosophic random variables. Hacet. J. Math. Stat. **51**, 1442–1457 (2022)
22. 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)
23. Khan, Z., Almazah, M., Hamood Odhah, O., Alshanbari, H.M.: Generalized pareto model: properties and applications in neutrosophic data modeling. Bioinorg. Chem. Appl. **2022**, 1–11 (2022)
24. Khan, Z., Amin, A., Khan, S.A., Gulistan, M.: Statistical development of the neutrosophic Lognormal model with application to environmental data. Neutrosophic Sets Syst. **47**, 1 (2021)
25. Khan, Z., Al-Bossly, A., Almazah, M., Alduais, F.S.: On Statistical development of neutrosophic Gamma distribution with applications to complex data analysis. Complexity **2021**, 1–8 (2021)
26. Maiti, S.S., Dey, M., Sarkar, S.: Discrete Xgamma distributions: properties, estimation and an application to the collective risk model. J. Reliab. Stat. Stud. **11**, 117–132 (2018)
27. Mazucheli, J., Bertoli, W., Oliveira, R.P., Menezes, A.F.B.: On the discrete quasi Xgamma distribution. Methodol. Comput. Appl. Probab. **22**, 747–775 (2020)
28. Nakagawa, T., Osaki, S.: The discrete Weibull distribution. IEEE Trans. Reliab. **24**, 300–301 (1975)
29. Patro, S.K., Smarandache, F.: The neutrosophic statistical distribution, more problems, more solutions. Infin. Study **12**, 1 (2016)
30. Ramos, P.L., Louzada, F.: A distribution for instantaneous failures. Stats **2**, 247–258 (2019)
31. Shah, F., Aslam, M., Khan, Z., Almazah, M., Alduais, F.S.: On neutrosophic extension of the maxwell model: properties and applications. J. Funct. Spaces **2022**, 1–9 (2022)
32. Smarandache, F.: Introduction to Neutrosophic Measure, Neutrosophic Integral, and Neutrosophic Probability. Sitech & Education Publishing, Craiova (2013)

33. Smarandache, F.: Neutrosophic statistics is an extension of interval statistics, while Plithogenic statistics is the most general form of statistics (second version). Int. J. Neutrosophic Sci. **19**, 148–165 (2022)
34. Zeina, M.B., Hatip, A.: Neutrosophic random variables. Neutrosophic Sets Syst. **39**, 4 (2021)