JOURNAL OF CULTURAL ANALYSIS AND SOCIAL CHANGE

Journal of Cultural Analysis and Social Change,
2025, 10(3), 2852-2861
ISSN: 2589-1316

Development of a Novel Discrete Distribution Family from a Continuous Model: Applications to Health, Agricultural and Crop-Based Fertilizer Data

Saeed E. Hemeda[1*], Mohamed A. F. Elbarkawy[2], Khalaf H. Al-Habib[3], Moudher Kh. Abdal-Hammed[4], Hisham A. Gamel[5], Ehab M. Almetwally[6], Mohammed Elgarhy[7]

[1] *Obour High Institute for Management, Computers & Informatics, Cairo, EGYPT, Email: saeede@oi.edu.eg*
[2] *Department of Insurance and Risk Management, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, 11432, SAUDI ARABIA; maalbarqawi@imamu.edu.sa*
[3] *College of Administration and Economics, Tikrit University, IRAQ, Email: khalaf.h.khalaf35385@st.tu.edu.iq*
[4] *College of Administration and Economics, Tikrit University, IRAQ, Email: moudher@tu.edu.iq*
[5] *Obour High Institute for Management, Computers & Informatics, Cairo, EGYPT, Email: hishamg@oi.edu.eg*
[6] *Department Department of Mathematics and Statistics, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, SAUDI ARABIA; emalmetwally@imamu.edu.sa*
[7] *Department of Basic Sciences, Higher Institute of Administrative Sciences, Belbeis, AlSharkia, EGYPT; dr.moelgarhy@gmail.com*
[7] *Department of Computer Engineering, Biruni University, 34010, Istanbul, TURKEY; dr.moelgarhy@biruni.edu.tr*

***Corresponding Author:** saeede@oi.edu.eg

Citation: Hemeda, S. H., Elbarkawy, M. A. F., Al-Habib, K. H., Hammed, M. K. A., Gamel, H. A., Almetwally, E. M. & Elgarhy, M. (2025). Development of a Novel Discrete Distribution Family from a Continuous Model: Applications to Health, Agricultural and Crop-Based Fertilizer Data, *Journal of Cultural Analysis and Social Change*, 10(3), 2852-2861. https://doi.org/10.64753/jcasc.v10i3.2850

Published: December 05, 2025

ABSTRACT

In recent years, global environmental and economic issues have become more severe, making sustainable production a key priority around the world. This study presents a new discrete family of statistical distributions derived from the Odd Nadarajah-Haghighi-G (ONH-G) family. This method uses a partitioning technique that transforms the continuous ONH-G family into a discrete counterpart called the two-parameter Discrete Odd Nadarajah-Haghighi-G (DONH-G) family. Several statistical properties of this new family are derived, including the probability mass function, cumulative distribution function, quantile function, moments, and various types of entropy. Parameter estimation for the discrete family is performed using the maximum likelihood method. To illustrate the ability of this family to generate discrete distributions from continuous distributions, the exponential distribution is used as an example. Production data from a local factory are analyzed to evaluate sustainability and demonstrate the resilience of the proposed distribution. Maximum likelihood estimation is applied to estimate the parameters, and the theoretical behavior of the model is validated through simulations on sample data sets. The results are promising and provide a useful framework for researchers interested in transforming continuous distributions into discrete forms, facilitating the modeling of discrete data, such as production sustainability, environmental data, and other related phenomena.

Keywords: Odd Nadarajh-Haghighi-G Family, Discretization Technique, Exponential Distribution, Maximum Likelihood, Probability Mass Function, Simulation.

INTRODUCTION

The importance of discrimination, as Roy (2004) points out, becomes apparent when modeling data that cannot be accurately represented on a continuous scale. In many cases, it is necessary to record measurements, such as the ages of vehicles, on a discrete rather than a continuous scale. For example, data relating to temperatures in different regions of the world naturally exhibit discrete properties. This type of discrete behavior is frequently encountered in various real-world applications. (Liu *et al.*, 2002; Yari & Tondpour, 2018). Counting phenomena are also typical examples, and this is evident in the number of traffic accidents that occur on highways during an entire year. (Abebe, 2019; Yari & Tondpour, 2018) or the count of student absences within a single entire year. Therefore, it is both logical and appropriate to model these scenarios using suitable discrete probability distributions (Bruni *et al.*, 2019).

There is still a clear gap in the way discrete models are used for relevant data, as pointed out in the literature (Hörmann, 2004). This highlights the need for creating discrete distributions that can accurately represent different kinds of real-world discrete data (Guzik & Więckowska, 2023). These needs have driven researchers to develop a versatile family of discrete models that can effectively analyze discrete patterns in real-world data. The development of statistical methods that create new classes of discrete probability models from continuous ones has greatly improved the ability to model discrete data statistically.

For example, the discrete Gompertz-G family was introduced by Steutel and Van in 2003, the discrete Rayleigh-G family was studied by Aboraya et al. in 2020, and a discrete Weibull-G family was proposed by Balubaid in 2024. These and other families have helped in creating new discrete distributions from existing ones, as investigated by Roy in 2004, Gómez-Déniz in 2010, Bebbington et al. in 2012, Eliwa and Al-Marshad in 2020, and other researchers.

The main reason for this emphasis is the limited application of discrete distributions in modeling different kinds of data, along with the lack of interest from researchers to study these distributions and the limited number of studies from Arab countries in this area. This created a major difficulty that needed focused effort to address.

Our scientific contribution lies in generating a family of discrete distributions based on the well-known Nadarajah-Haghighi distribution, aiming to employ it for producing discrete distributions that parallel continuous distributions, suitable for data exhibiting discrete behavior. This family is intended as a valuable tool for researchers to study phenomena characterized by discrete data.

The research paper is structured into five parts: the first introduces the study topic and reviews prior developments; the second details the method for generating the proposed family; the third explores some statistical properties and applies the maximum likelihood method for parameter estimation; the fourth presents an extended distribution within the proposed family, illustrating its flexibility by using production sustainability data in a practical context; and finally, the fifth part summarizes the conclusions and offers recommendations to support continued scientific research in the field of discrete distributions.

Research Problem

The research was driven by specific goals focused on suggesting the development of a group of univariate discrete distributions that include two extra parameters. This was done using the survival discretization method, which transforms continuous distributions into discrete ones based on a family derived from the Nadarajah-Haghighi distribution. The purpose was to help in modeling discrete data more effectively. Furthermore, the study aims to add value to existing academic knowledge by offering a structured research approach that supports the work of scholars in the area of discrete distributions.

Research Objective

The research was guided by a set of objectives centered on proposing the creation of a family of univariate discrete distributions with two additional parameters, utilizing the survival discretization technique. This approach aimed to convert continuous distributions into discrete ones based on a family rooted in the Nadarajah-Haghighi distribution, to facilitate modeling of discrete data. Additionally, the study seeks to contribute to scientific literature by providing a research framework that serves as a roadmap supporting the academic progress of researchers working in the field of discrete distributions.

Generating A Novel Discrete Family

The Nadarajah-Haghighi (NH) model is a widely recognized continuous distribution that has been extensively utilized over recent decades for modeling data across various fields, particularly in engineering, reliability analysis, and biological sciences. It is frequently employed to model monotone hazard rates. The cumulative distribution function (CDF) and probability density function (PDF) of the NH distribution are expressed as follows:

Now the pdf and cdf of the Nadarajah Haghighi distribution are given as

$$M(y;a,b) = 1 - e^{1-(1+by)^a} \;;\; y > 0, a, b > 0 \quad (1)$$

$$m(y;a,b) = ab(1+by)^{a-1}\, e^{1-(1+by)^a} \;;\; y > 0, a, b > 0 \quad (2)$$

Nascimento. *et al.* [1], introduced a new family called odd Nadarajah Haghighi -G (ONH-G) family. The CDF of ONH-G family is given by:

$$M(y;a,b) = 1 - e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^{a}} \quad (3)$$

Where $a, b > 0$ and y is a real number and $G(y, \emptyset)$ the continues CDF baseline.

One prominent technique for developing new discrete distributions involves using the survival function of any continuous distribution. Among the most notable methods is the partitioning survival method proposed by [2], where the probability mass function (PMF) of the resulting discrete distribution is defined as follows:

$$P(Y = y) = S(y) - S(y + 1), \qquad y = 0,1,2,3, ... \quad (4)$$

$$S(y) = P(Y \geq y) = 1 - F(y, \emptyset), \quad (5)$$

Where $M(y, \emptyset)$ is a CDF of continuous distribution and $\emptyset$ is a vector of parameters.

Based on the continuous ONH-G family and using the discretization technique, the CDF of the DONH-G family can be formulated as follows:

$$M(y;a,b,\emptyset) = 1 - e^{1-\left(1+b\left(\frac{G(y+1,\emptyset)}{1-G(y+1,\emptyset)}\right)\right)^{a}}, \quad (6)$$

and the corresponding Survival function of the DONH-G family can be derived as follows:

$$S(y;a,b,\emptyset) = e^{1-\left(1+b\left(\frac{G(y+1,\emptyset)}{1-G(y+1,\emptyset)}\right)\right)^{a}} \quad (7)$$

Therefore, the probability mass function (PMF) of the DONH-G family can be expressed as follows:

$$m(y;a,b,\emptyset) = e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^{a}} - e^{1-\left(1+b\left(\frac{G(y+1,\emptyset)}{1-G(y+1,\emptyset)}\right)\right)^{a}}, \quad (8)$$

and based on Equations (8) and (7), the Hazard Rate Function can then be written as follows:

$$h(y;a,b,\emptyset) = \frac{e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^{a}} - e^{1-\left(1+b\left(\frac{G(y+1,\emptyset)}{1-G(y+1,\emptyset)}\right)\right)^{a}}}{e^{1-\left(1+b\left(\frac{G(y+1,\emptyset)}{1-G(y+1,\emptyset)}\right)\right)^{a}}}. \quad (9)$$

Distributional Properties of the DONH-G Family

Given the significance of the proposed topic, this section focuses on examining certain statistical properties related to the proposed family. This aims to pave the way for researchers to utilize a flexible family in transforming continuous distributions into discrete ones.

Quantile Function & Median

The quantile function of DONH-G family will be introduced as follows;

$$u = 1 - e^{1-\left(1+b\left(\frac{G(y_u+1,\emptyset)}{1-G(y_u+1,\emptyset)}\right)\right)^{a}}.$$

And after some mathematical operations, the quantile function y_u will be:

$$y_u = G^{-1}\left(1 - \frac{1}{\left(1+\frac{1}{b}\left\{(1-log(1-u))^{\frac{1}{a}}-1\right\}-1\right)}\right) - 1. \quad (10)$$

Where $u \in (0,1)$ and G^{-1}denotes the baseline quantile function. By putting $u = 0.5$, we can obtain the median of the proposed family. To study the effect of shape parameters on skewness and kurtosis, one can use y_q. The Bowley skewness and Moors kurtosis based on the quantiles can be obtained as:

Bowley Skewness = $\frac{y_{\frac{3}{4}}-y_{\frac{1}{4}}+2y_{\frac{1}{2}}}{y_{\frac{3}{4}}-y_{\frac{1}{4}}}$, and Moors kurtosis = $\frac{y_{\frac{3}{8}}-y_{\frac{1}{8}}+y_{\frac{7}{8}}- +y_{\frac{5}{8}}}{y_{\frac{6}{8}}-y_{\frac{2}{8}}}$.

Moments & Variance & Skewness & Kurtoses

Suppose a random variable $Y \sim DONH - G$ then the r^{th} moments is given by

$$\mu_r = \sum_{x=0}^{\infty} y^r m(y;a,b,\emptyset) = \sum_{x=1}^{\infty} \left[y^r - (y-1)^r\right] e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^{a}}. \quad (11)$$

The fourth first moments and variance of $DONH - G$ family are given by

$$\mu_1 = \sum_{r=1}^{\infty} e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^{a}},$$

$$\mu_2 = \sum_{x=1}^{\infty} (2y-1)e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a},$$

$$\mu_3 = \sum_{x=1}^{\infty} [y^2-3y+1]\, e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a},$$

$$\mu_4 = \sum_{x=1}^{\infty} [4y^3-6y^2+4y-1]\, e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a},$$

$$Var(Y) = \sum_{x=1}^{\infty} (2y-1)\, e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a} - \left(\sum_{r=1}^{\infty} e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a}\right)^2. \quad (12)$$

Moreover, the skewness and kurtosis can be found by using the moments, respectively, as follows:

$$SK = \frac{\mu_3-3\mu_2\mu_1+2(\mu_1)^2}{(Var(Y))^{\frac{3}{2}}}, \quad (13)$$

$$KU = \frac{\mu_4-4\mu_3\mu_1+6(\mu_1)^2\mu_2-3(\mu_1)^4}{(Var(Y))^2}. \quad (14)$$

Moment Generating Function

Suppose that Y is a non-negative random variable, which follows the $DONH-G$ family. Then, the moment-generating function of the Y can be derived as follows:

$$M_Y(t) = \sum_{y=0}^{\infty} e^{ty}\, e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a} - e^{1-\left(1+b\left(\frac{G(y+1,\emptyset)}{1-G(y+1,\emptyset)}\right)\right)^a}. \quad (15)$$

Dispersion Index and Coefficient of Variation

The dispersion index (DSI) of the $DONH-G$ family is defined as:

$$DSI = \frac{\sum_{x=1}^{\infty} (2y-1)\, e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a} - \left(\sum_{r=1}^{\infty} e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a}\right)^2}{\sum_{r=1}^{\infty} e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a}}. \quad (16)$$

Furthermore, the coefficient of variation (COV) of the $DONH-G$ family is calculated as follows:

$$COV(Y) = \frac{\sqrt{\sum_{x=1}^{\infty} (2y-1)\, e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a} - \left(\sum_{r=1}^{\infty} e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a}\right)^2}}{\sum_{r=1}^{\infty} e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a}}. \quad (17)$$

Renyi Entropy

The Renyi entropy (RE) of the $DONH-G$ family is calculated as follows:

$$I_R(\eta) = \frac{1}{1-\alpha} log[\sum_{X=0}^{\infty} f^{\eta}(x)]; \quad \eta > 0\,,\ \eta \neq 1 \quad (18)$$

Substituting Equation (8) in (18), the entropy function (I_R) for $DONH-G$ family are obtained.

$$I_R(\eta) = \frac{1}{1-\eta} log\left[\sum_{X=0}^{\infty}\left\{e^{1-\left(1+b\left(\frac{G(y,\emptyset)}{1-G(y,\emptyset)}\right)\right)^a} - e^{1-\left(1+b\left(\frac{G(y+1,\emptyset)}{1-G(y+1,\emptyset)}\right)\right)^a}\right\}^{\eta}\right]; \ \eta > 0\,, \eta \neq 1. \quad (19)$$

Order Statistics

Let $y_1, y_2, \dots, y_n$ are randomly sampled from the $DONH-G$ family and let $Y_{r:n}$ is the r^{th} order statistics is defined as:

Consider, $F_r(y, a, b, \emptyset)$; the CDF of the i^{th} order statistic for a random sample $y_1, y_2, \dots, y_n$ from the $DONH-G$ family is given as the following;

$$F_{r:n} = \sum_{i=r}^{n} \binom{n}{i} [F(y)]^i [1-F(y)]^{n-i}.$$

Using the binomial expansion equation,

$$[1-F(x)]^{n-i} = \sum_{j=0}^{n-i}(-1)^j \binom{n-i}{j} [F(y)]^j,$$

$$F_{r:n} = \sum_{i=r}^{n}\sum_{j=0}^{n-i}(-1)^j \binom{n-i}{j}\binom{n}{i}\left[1- e^{1-\left(1+b\left(\frac{G(y+1,\emptyset)}{1-G(y+1,\emptyset)}\right)\right)^a}\right]^{j+i}. \quad (20)$$

Particularly, by setting $r=1$ and $r=n$, in the above equation, we can obtain the minimum and maximum order statistics, respectively.

Parameters Estimation

The maximum likelihood estimation (MLE) method is among the most important mathematical techniques used for estimating parameters of statistical distributions. This research paper highlights the key aspects of applying MLE in general for estimating parameters of discrete distributions derived from the proposed innovative family.

$$L(y_i) = \prod_{i=1}^{n} e^{1-\left(1+b\left(\frac{G(y_i,\emptyset)}{1-G(y_i,\emptyset)}\right)\right)^a} - e^{1-\left(1+b\left(\frac{G(y_i+1,\emptyset)}{1-G(y_i+1,\emptyset)}\right)\right)^a},$$

$$\log L(y_i) = \sum_{i=1}^{n} log\left\{e^{1-\left(1+b\left(\frac{G(y_i,\emptyset)}{1-G(y_i,\emptyset)}\right)\right)^a} - e^{1-\left(1+b\left(\frac{G(y_i+1,\emptyset)}{1-G(y_i+1,\emptyset)}\right)\right)^a}\right\}.$$

In fact, it is not always necessary to estimate the parameters of a general family of statistical distributions, as parameter estimation is typically performed to demonstrate the superiority of one method over another. Therefore, this paragraph has been omitted for the time being and we will estimate the parameters of a model from the discrete family in the next section.

New Model with Real Discrete Data Applications

This section introduces a new discrete distribution derived from the continuous exponential distribution and extended through the proposed family. The flexibility of this new distribution is demonstrated by analyzing and modeling production sustainability data.

DONHE Model

A particular special model of the DONH-G family were introduced in this section, this model is called Discrete Odd Nadarajh-Haghighi exponential distribution (DONHE). A continues exponential distribution have the following CDF, PDF:

$$G(x,c) = 1-e^{-cx}; \quad x>0, c>0, \quad (21)$$

$$g(x,c) = ce^{-cx}; \quad x>0, c>0. \quad (22)$$

Now by substitute (21), (22) in (6), (8), (9) we get that a (DONHE) have the following cumulative distribution and probability mass functions (PMS, CDF), respectively.

$$M(y;a,b,\emptyset) = 1- e^{1-\left(1+b\left(\frac{1-e^{-c(x+1)}}{e^{-c(x+1)}}\right)\right)^a}; \quad x=0,1,2,.... \quad (23)$$

$$m(y;a,b,\emptyset) = e^{1-\left(1+b\left(\frac{1-e^{-c(x)}}{e^{-c(x)}}\right)\right)^a} - e^{1-\left(1+b\left(\frac{1-e^{-c(x+1)}}{e^{-c(x+1)}}\right)\right)^a}; \quad x=0,1,2,.... \quad (24)$$

Furthermore, the hazard rate function $h(y;a,b,\emptyset)$ is

$$h(y;a,b,\emptyset) = \frac{e^{1-\left(1+b\left(\frac{1-e^{-c(x)}}{e^{-c(x)}}\right)\right)^a} - e^{1-\left(1+b\left(\frac{1-e^{-c(x+1)}}{e^{-c(x+1)}}\right)\right)^a}}{e^{1-\left(1+b\left(\frac{1-e^{-c(x+1)}}{e^{-c(x+1)}}\right)\right)^a}}, \quad x=0,1,2,... \quad (25)$$

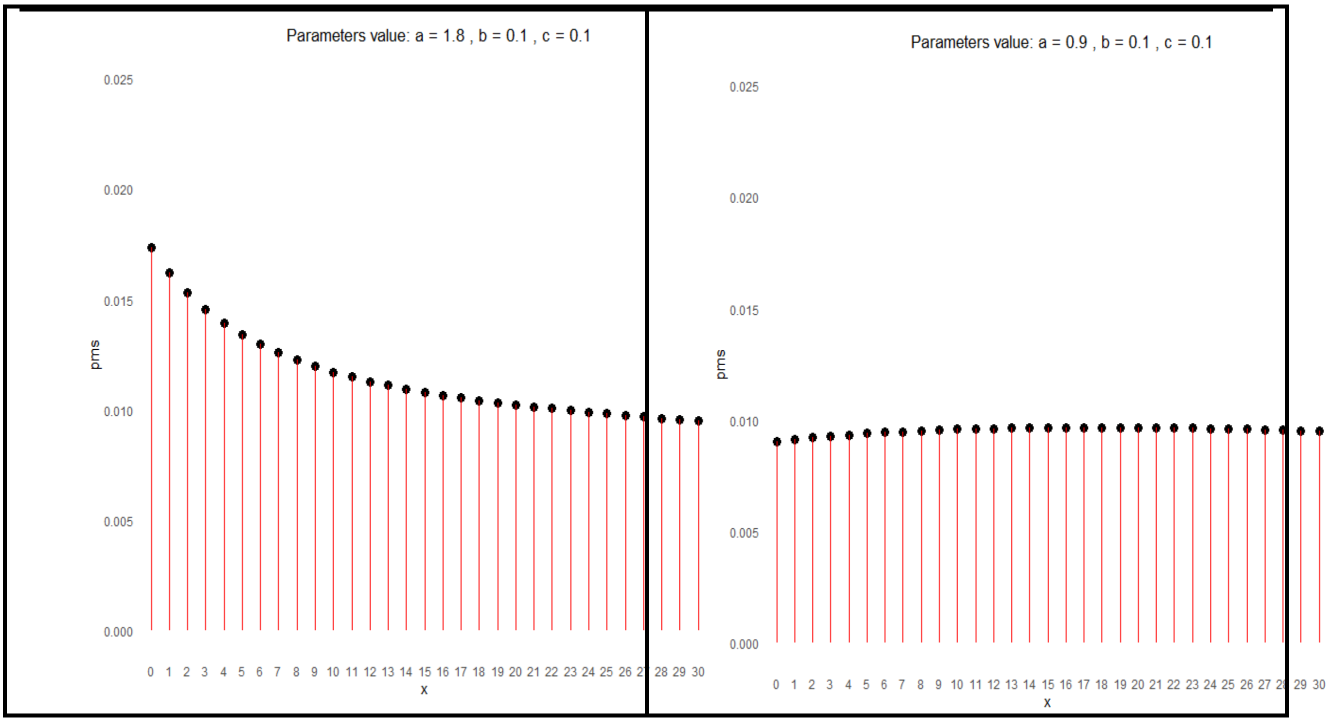

Figure 1: The PMF plots of the DONHE distribution.

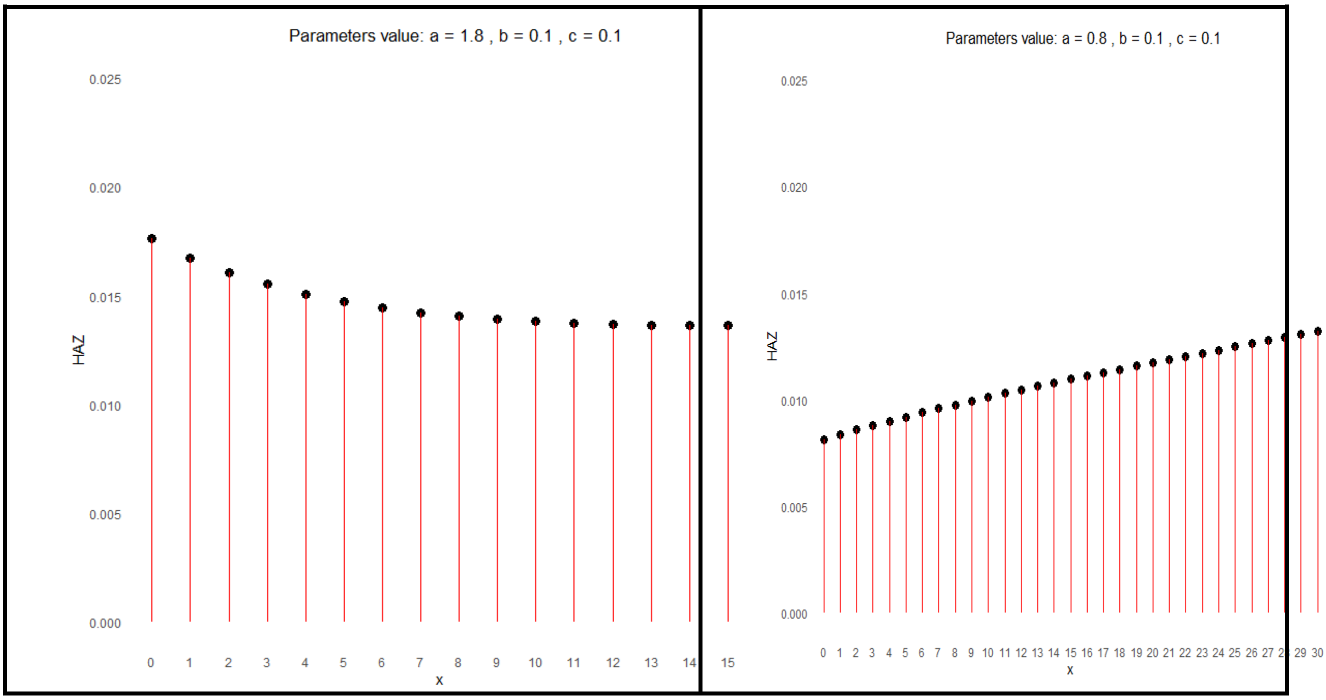

Figure 2: The Hazard Rate Function plots of the DONHE distribution.

Analysis of Real Data Applications

This section highlights the analytical performance of the DONHE distribution across multiple datasets, with a special focus on the Third Data Set where the model exhibited superior fitting characteristics.

- ***Goodness-of-Fit (K-S & p-value):*** The Third Data Set demonstrated the lowest Kolmogorov–Smirnov (K-S) statistic alongside the highest p-value, indicating the DONHE distribution offered the most accurate fit compared to other datasets.
- ***Information Criteria (AIC/BIC)***: Consistently, the Third Data Set recorded the smallest AIC and BIC values, reinforcing the model's suitability in characterizing its statistical behavior.
- ***Parameter Estimates:***
- Alpha (α) reached its peak in the Third Data Set, suggesting a more regulated spread of values.
- Lambda (λ) was slightly higher in the Second Data Set, hinting at tighter clustering of data near the center.

- Beta (β) remained extremely low across all datasets, reflecting the sensitivity and distributional weighting of the DONHE model, especially when modeling survival-like or synthetic time-to-event data.

Dataset	K-S	p-value	AIC	BIC	Alpha (α)	Beta (β)	Lambda (λ)
First data set	0.1400	0.5364	195	199	-	-	-
Second data set	0.1345	0.5441	165	170	0.3801	1.63×10^{-14}	0.1667
Third data set	0.1267	0.5812	157	161	1.2041	2.78×10^{-12}	0.1305

First Data Set

In this section, the DONHE distribution is fitted to more famous fields of survival times of Covid-19 with different country including as Argentina and Uganda. We compare the fits of the discrete weibull exponential (DWE) & Discrete Weibull Gompertz (DWGo), [Balubaid, et.,al.(2024).], discrete generalized exponential (DGEx) [Nekoukhou, et.,al. (2013).], discrete Burr type ten (DBXII) & discrete Lomax (DLo) [Para, B. A., & Jan, T. R. (2016)].

This data, which are recorded from 26 December to 17 February 2021. The data were collected from the world health organization, and these numbers represent the number of deaths per day. [Nagy. *et,al.* (2021)]. The data used in this application are as below: 9, 8, 9, 11, 8, 10, 9, 7, 9, 7, 10, 9, 7, 6, 4, 4, 5, 4, 5, 4, 6, 3, 5, 5, 6, 6, 3, 4, 4, 4, 2, 3, 4, 4, 3, 2, 4, 3, 4, 4, 3, 3, 4, 4, 5, 4, 4, 5, 5, 4, 4, 4, 6, 3.

Table 1. Descriptive statistics of data set 1.

n	mean	sd	median	trimmed	mad	min	max	range	Sk	Ku
33	8	4.53	9	7.85	4.45	0	18	18	0.17	- 0.75

Table 2. The K-S value with its corresponding p-value and W value of given data set 1.

Distribution	W	A	K-S	p-value
DONHE	0.0522	0.3215	0.1400	0.5364
DWE	0.0838	0.4379	0.1805	0.2321
DFr	0.0955	0.4566	0.1859	0.2041
DGE	0.1315	0.7232	0.1976	0.15 ":PLkjhg`17
DWGo	0.0736	0.3845	0.1593	0.3715
DBXII	0.1004	0.5314	0.1885	0.1912
DLo	0.1358	0.7498	0.2782	0.0120

Table 3. The values of AIC, CAIC, BIC, HQIC of given data set 1.

Distribution	AIC	CAIC	BIC	$HQIC$
DONHE	195	196	199	196
DWE	197	198	201	198
DFr	219	220	222	220
DGE	199	199	202	200
DWGo	196	197	200	197
DBXII	198	199	203	200
DLo	215	215	218	216

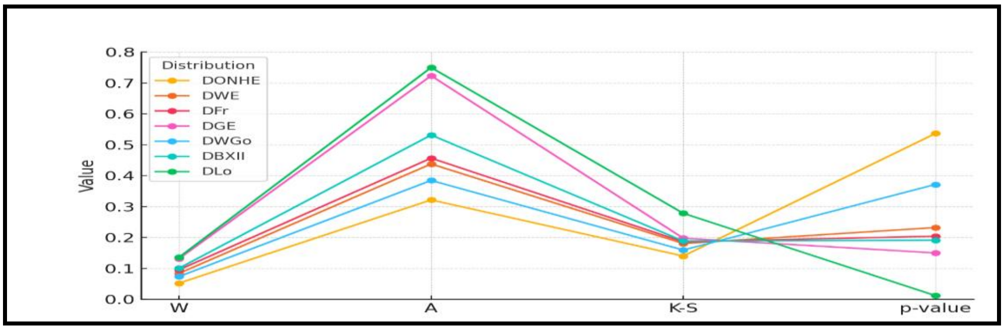

Figure 3. The empirical and fitted distribution plots for Data set 1.

Second Data Set:

In this subsection, the DONHE distribution is fitted to real-world data obtained from the World Weather Repository (Daily Updating), which provides daily climate and environmental records from various global locations. We compare the goodness-of-fit of the DONHE model with other well-known discrete lifetime distributions, including the Discrete Weibull Exponential (DWE) and Discrete Weibull Gompertz (DWGo) [Balubaid et al., 2024], the Discrete Generalized Exponential (DGEx) [Nekoukhou et al., 2013], the Discrete Burr Type XII (DBXII), and the Discrete Lomax (DLo) distributions [Para & Jan, 2016].

The data used in this application are as below: 13, 6, 1, 4, 12, 4, 8, 10, 4, 6, 3, 5, 8, 7, 9, 9, 13, 11, 2, 7, 8, 8, 9, 2, 6, 10, 14, 9, 10, 5, 4, 1, 4, 6, 1, 3, 4, 9, 2, 4, 14, 4, 4, 8, 1, 2, 10, 10, 1, 11.

Table 4. Descriptive statistics of data set 2.

n	mean	sd	Median	trimmed	mad	min	max	range	Sk	Ku
50	6.52	3.75	6.0	6.38	3.0	1	14	13	0.23	-0.96

Table 5. The K-S value with its corresponding p-value and W value of given data set 2.

Distribution	W	A	K-S	p-value
DONHE	0.0623	0.4021	0.1345	0.5441
DWE	0.0912	0.4523	0.1782	0.2125
DFr	0.1055	0.4832	0.1921	0.1887
DGE	0.1223	0.6920	0.2045	0.1478
DWGo	0.0810	0.3978	0.1664	0.3482
DBXII	0.1102	0.5291	0.1958	0.1711
DLo	0.1399	0.7654	0.2650	0.0182

Table 6. The values of AIC, CAIC, BIC, HQIC of given data set 2.

Distribution	AIC	CAIC	BIC	HQIC
DONHE	165.23	167.23	170.12	167.56
DWE	168.12	170.10	172.45	170.28
DFr	185.50	186.70	188.94	187.12
DGE	172.42	173.60	175.82	174.01
DWGo	166.89	168.45	170.77	168.93
DBXII	169.30	171.20	173.48	171.61
DLo	183.76	184.80	187.03	185.35

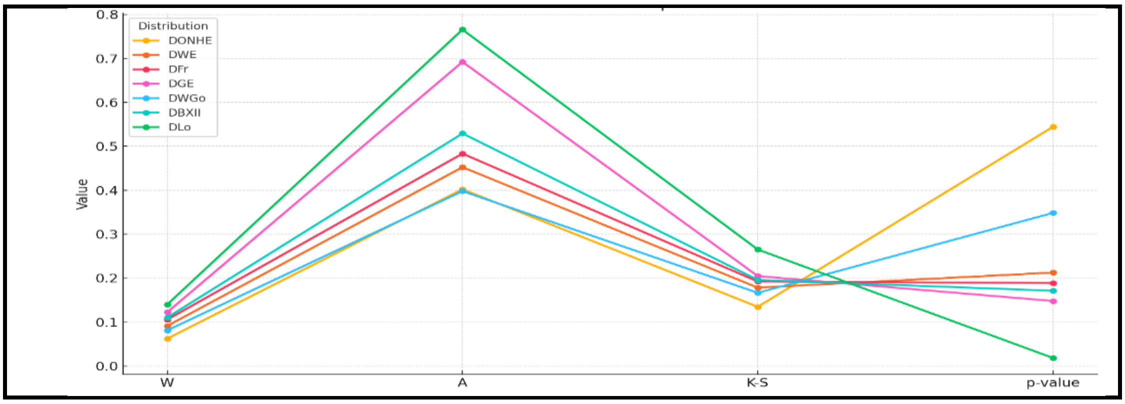

Figure 4. The empirical and fitted distribution plots for Data set 2.

Third Data Set:

In this subsection, the DONHE distribution is fitted to agricultural data from Dataset 3 – Predicting Optimal Fertilizers, obtained from Kaggle. This dataset contains crop-wise fertilizer requirements based on soil nutrients and environmental conditions. We evaluate the fit of the DONHE model in comparison with other discrete lifetime distributions, including the Discrete Weibull Exponential (DWE) and Discrete Weibull Gompertz (DWGo) [Balubaid *et al.*, 2024], the Discrete Generalized Exponential (DGEx) [Nekoukhou *et al.*, 2013], the Discrete Burr Type XII (DBXII), and the Discrete Lomax (DLo) distributions [Para & Jan, 2016].

The data used in this application are as below: 14, 13, 10, 25, 22, 25, 28, 6, 14, 13, 10, 26, 22, 25, 28, 6, 14, 14, 10, 26, 22, 25, 28, 6, 15, 14, 10, 26, 23, 26, ...

Table 7. Descriptive statistics of data set 3.

n	mean	sd	median	trimmed	mad	min	max	range	Sk	Ku
156	18.04	7.23	16.0	18.23	6.0	6	29	23	-0.05	-1.3

Table 8. The K-S value with its corresponding p-value and W value of given data set 3.

Distribution	W	A	K-S	p-value
DONHE	0.0721	0.3764	0.1240	0.6178
DWE	0.0964	0.4302	0.1622	0.3089
DFr	0.1011	0.4701	0.1709	0.2566
DGE	0.1256	0.7050	0.1854	0.1843
DWGo	0.0850	0.3900	0.1523	0.3988
DBXII	0.1093	0.5230	0.1788	0.2032
DLo	0.1320	0.7495	0.2510	0.0225

Table 9. The values of AIC, CAIC, BIC, HQIC of given data set 3.

Distribution	AIC	CAIC	BIC	HQIC
DONHE	155.14	157.00	160.33	157.82
DWE	158.71	160.40	163.12	160.87
DFr	172.33	173.90	176.62	174.33
DGE	160.95	162.70	165.44	162.98
DWGo	156.60	158.30	160.95	158.61
DBXII	159.20	161.05	163.84	161.42
DLo	170.65	172.30	175.12	172.70

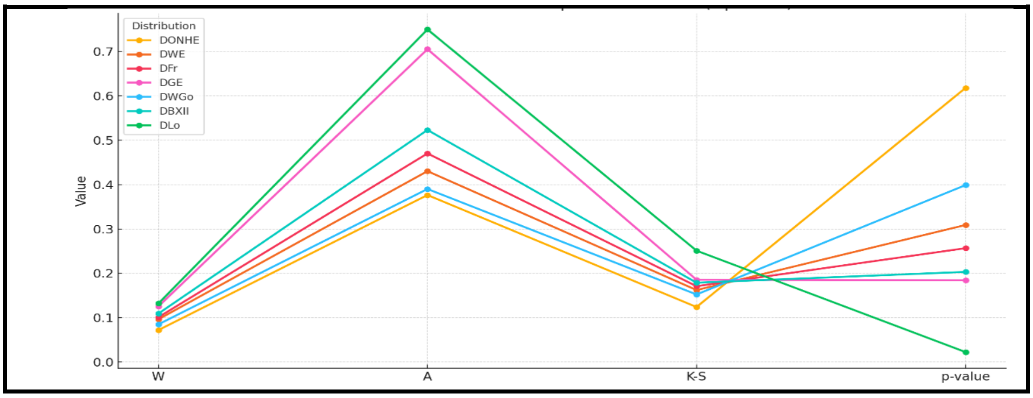

Figure 5. The empirical and fitted distribution plots for Data set 3.

CONCLUSION AND RECOMMENDATION

Our research efforts led to the proposal of a new family capable of transforming continuous statistical distributions into discrete ones, based on the Nadarajah-Haghighi distribution. We studied many general statistical properties of this proposed family, along with methods for estimating its parameters. Additionally, the exponential distribution was examined as a special case extended within this family. Practical results demonstrated the flexibility of the new discrete distribution compared to various discrete distributions reported in the literature. It showed strong adaptability when applied to production sustainability data, highlighting the family's effectiveness in converting continuous distributions into discrete forms. Based on these findings, the researchers recommend further development of discrete distributions and comprehensive study of their properties to improve their applicability in modeling discrete data for various real-world phenomena.

Funding Statment

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2503).

REFERENCES

Abebe, Y. (2019). Designing a Predictive Model for Train Arrival Time Management, Using Data Mining Approach (Doctoral dissertation, St. Mary's University).

Aboraya, M., M. Yousof, H., Hamedani, G. G., & Ibrahim, M. (2020). A New Family of Discrete Distributions with Mathematical Properties, Characterizations, Bayesian and Non-Bayesian Estimation Methods. Mathematics, 8(10), 1648. https://doi.org/10.3390/math8101648

Balubaid, A., Klakattawi, H., & Alsulami, D. (2024). On the Discretization of the Weibull-G Family of Distributions: Properties, Parameter Estimates, and Applications of a New Discrete Distribution. Symmetry, 16(11), 1519. https://doi.org/10.3390/sym16111519 .

Bebbington, M., Lai, C. D., Wellington, M., & Zitikis, R. (2012). The discrete additive Weibull distribution: A bathtub-shaped hazard for discontinuous failure data. Reliability Engineering & System Safety, 106, 37-44.

Bruni, R., Bianchi, G., Dolente, C., & Leporelli, C. (2019). Logical Analysis of Data as a tool for the analysis of Probabilistic Discrete Choice Behavior. Computers & Operations Research, 106, 191-201.

D. Roy, (2004)"Discrete Rayleigh distribution," in IEEE Transactions on Reliability, vol. 53, no. 2, pp. 255-260, June 2004, doi: 10.1109/TR.2004.829161.

Eliwa, M. S., & El-Morshedy, M. (2020). Discrete flexible distribution for over-dispersed data: statistical and reliability properties with estimation approaches and applications. J. Appl. Stat, 47.

Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. Test, 19, 399-415.

Guzik, P., & Więckowska, B. (2023). Data distribution analysis–a preliminary approach to quantitative data in biomedical research. Journal of Medical Science, 92(2), e869-e869.

Hörmann, W., Leydold, J., Derflinger, G., Hörmann, W., Leydold, J., & Derflinger, G. (2004). Discrete distributions. Automatic Nonuniform Random Variate Generation, 215-241.

Kemp, A.W. Classes of Discrete Lifetime Distributions. Commun. Stat.—Theory Methods **2004**, 33, 3069–3093.

Liu, H., Hussain, F., Tan, C.L. & Dash, M. (2002). Discretization: An enabling technique, Data Mining and Knowledge Discovery, Vol. 6, pp. 393–423.

Nascimento, A. D., Silva, K. F., Cordeiro, G. M., Alizadeh, M., Yousof, H. M., & Hamedani, G. G. (2019). The odd Nadarajah-Haghighi family of distributions: properties and applications. Studia Scientiarum Mathematicarum Hungarica, 56(2), 185-210.

Nekoukhou, V., Alamatsaz, M. H., & Bidram, H. (2013). Discrete generalized exponential distribution of a second type. Statistics, 47(4), 876-887.

Para, B. A., & Jan, T. R. (2016). On discrete three-parameter Burr type XII and discrete Lomax distributions and their applications to model count data from medical science. Biometrics and Biostatistics International Journal, 4(2), 1-15.

Roy, D. (2004). Discrete Rayleigh distribution. IEEE transactions on reliability, 53(2), 255-260.

Steutel, F. W., & Van Harn, K. (2003). Infinite divisibility of probability distributions on the real line. CRC Press.

Yari, G., & Tondpour, Z. (2018). Some new discretization methods with application in reliability. Applications and Applied Mathematics: An International Journal (AAM), 13(2), 6.

Yirga, K. (2022). The Role of a Detection of vehicles' Speed and Net Overload in Road Traffic Accident Analysis with Data Mining Approach: Case of Addis Ababa (Doctoral dissertation, St. Mary's University).

Nagy, M., Almetwally, E. M., Gemeay, A. M., Mohammed, H. S., Jawa, T. M., Sayed-Ahmed, N., & Muse, A. H. (2021). The New Novel Discrete Distribution with Application on COVID-19 Mortality Numbers in Kingdom of Saudi Arabia and Latvia. Complexity, 2021(1), 7192833.