STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–23.
Published online in International Academic Press (www.IAPress.org)

Discretization of the Inverse Rayleigh-G Family of Distributions: Theoretical Properties, Machine Learning-Based Parameter Estimation via K-Nearest Neighbors algorithm, and Applications of a Novel Discrete Distribution

Adel S. Hussain[1]*, Sanaa Mohsin[2], Kawthar I. Hameed[3], Emad A. Az-Zo'bi[4], Mohammad A. Tashtoush[5,6]

[1]*IT Department, Amedi Technical Institutes, University of Duhok Polytechnic, Duhok, Iraq*
[2]*Business IT Department, College of Business Information, University of IT and Communication, Baghdad, Iraq*
[3]*College of Dentist, Al-Iraqi University, Baghdad, Iraq*
[4]*Department of Mathematics and Statistics, Faculty of Sciences, Mutah University, Al-Karak, Jordan*
[5]*Department of Basic Sciences, Al-Huson University College, Al-Balqa Applied University, Salt, Jordan*
[6]*Faculty of Education and Arts, Sohar University, Sohar, Oman*

Abstract IIn this study, the newly developed discrete distribution identified as the two-parameter discrete Inverse Rayleigh Exponential (DIRE) distribution is developed after discretization of the Inverse Rayleigh-G family of distributions was done. The probability mass function of DIRE distribution has the probability density function and it offers flexibility in its shape where shapes may be symmetric or asymmetric. It has very tractable hazard function which exhibits range of behaviors as a function of time, with the ability to be increasing, constant, uniform, monotonically increasing and even reversed J shaped. We study some features of the DIRE distribution, mean and variance, moment generating function and dispersion coefficient. The estimators discussed in this work include both the maximum likelihood estimation (MLE) approach and the K-Nearest Neighbors (K-NN) algorithm; the performance of these estimators is assessed based on large simulation studies. It is exercised on two real datasets to evaluate the efficacy of the proposed distribution practically. A comparison with other models shows that, in each case, the results of the DIRE distribution fitting are significantly higher. That K-NN algorithm has been used for parameter estimation in the case also emphasize of importance of the machine learning techniques in the modeling and application of the discrete probability distributions.

Keywords Discrete Distribution; Survival Discretization; Inverse Rayleigh-G family; Exponential Distribution; K-Nearest Neighbors algorithm.

AMS 2010 subject classifications 60J80, 60J85, 60K10

DOI: 10.19139/soic-2310-5070-2618

1. Introduction

The application of statistical distributions is essential for effectively modeling and interpreting a wide variety of data. Traditional distributions provide a robust starting point for many applications, yet they often lack the flexibility to represent the diverse patterns and intricate behaviors observed in real-world datasets. To overcome these challenges, researchers have introduced new families of both continuous and discrete distributions, broadening the scope of statistical tools available for more complex data analysis. Discrete distributions, in particular, are critical for analyzing non-continuous variables, such as event counts or durations, driving the ongoing development of

*Correspondence to: Adel S. Hussain (Email: adel.sufyan@dpu.edu.krd). IT Department, Amedi Technical Institutes, University of Duhok Polytechnic, Duhok, Iraq

ISSN 2310-5070 (online) ISSN 2311-004X (print)

novel models tailored to these specific requirements. However, existing discrete distributions often face limitations when applied to highly complex or heterogeneous data, necessitating more adaptable approaches. This has led to a growing focus on discretizing continuous models, enabling the construction of discrete analogues that extend the capabilities of traditional statistical frameworks. Recent advancements in this area have facilitated the systematic creation of discrete G families of probability distributions derived from continuous counterparts. For instance, [1] presented the discrete Gompertz-G distribution, building upon the foundational concepts outlined in [2]. Similarly, the exponential generalized-G family of continuous distributions was introduced in [3], contributing to the growing suite of flexible models. Further developments include the discrete Rayleigh-G family presented in [5], based on the methodology in [4], and the discrete analogue of the Weibull-G family introduced in [6]. These contributions highlight the significance of discretization techniques in advancing statistical methodologies, providing tools better suited to the complexities of modern data analysis.

Recently, [7] introduced an innovative methodology termed the 'transformed,' designed to systematically generate new families of probability distributions, significantly enhancing the flexibility and applicability of statistical modeling frameworks. They selected the weight function $W\left(G\left(y,\beta\right)\right)=-\log\left(1-G\left(y,\beta\right)\right)$, where $G\left(y,\beta\right)$ denotes the probability distribution of a baseline distribution corresponding to the vector β, To present the Weibull-G, beta-exponential-G, and gamma-G families of distributions.

In this work, $W\left(G\left(y,\beta\right)\right)$ is addressed in [8, 9, 10]. The Inverse Rayleigh-G distribution is characterized by its cumulative distribution function (CDF), defined as:

$$F\left(y;\lambda,\beta\right)=\int_{0}^{-\log(1-G(y,\beta))}\frac{2\lambda}{t^{3}}e^{\left(\frac{-t}{y^{2}}\right)}=1-e^{-\left(\frac{\lambda}{(-\log(1-G(y,\beta)))^{2}}\right)} \tag{1}$$

with corresponding probability density function (PDF):

$$f\left(y;\lambda,\beta\right)=\frac{2\lambda}{\left(-\log\left(1-G\left(y,\beta\right)\right)\right)^{3}}\left(\frac{g\left(y,\beta\right)}{1-G\left(y,\beta\right)}\right)e^{-\left(\frac{\lambda}{(-\log(1-G(y,\beta)))^{2}}\right)}. \tag{2}$$

The formal expressions for the hazard rate function (HRF) and the survival function (SF) are as follows:

$$S\left(y;\lambda,\beta\right)=e^{-\left(\frac{\lambda}{(-\log(1-G(y,\beta)))^{2}}\right)}, \tag{3}$$

$$h\left(y;\lambda,\beta\right)=\frac{f\left(y;\lambda,\beta\right)}{S\left(y;\lambda,\beta\right)}=\frac{2\lambda}{\left(-\log\left(1-G\left(y,\beta\right)\right)\right)^{3}}\left(\frac{g\left(y,\beta\right)}{1-G\left(y,\beta\right)}\right). \tag{4}$$

In the last few decades, many discretized forms of continuous probability distributions have been constructed to capture different discrete data plans. Among these approaches, survival discretization method is the innovative one that has been initiated and proposed by [10, 11]. This method just adopts a survival function of a continuous distribution to make the corresponding discrete distribution. As outlined in [10], the probability mass function (PMF) for a discrete distribution derived through this approach is defined as follows:

$$P\left(Y=y\right)=S\left(y\right)-S\left(y+1\right),\ x=0,1,2,\ldots \tag{5}$$

where $S(x)$ represents the survival function of the underlying continuous distribution, mathematically expressed as:

$$S\left(y\right)=P\left(Y\geq y\right)=1-F\left(y,\theta\right), \tag{6}$$

Where $F\left(y,\theta\right)$ is a CDF of the continuous distribution and θ This approach has recently been generalized to define several new discrete families of probability distribution. For example, [1] proposed the discrete Gompertz-G family of distributions: discrete Gompertz-exponential, discrete Gompertz-Weibull, discrete Gompertz-inverse Weibull.

[5] has examined the discrete Rayleigh family and has suggested the discrete Rayleigh-Weibull distribution. In [11, 12, 13] the odd Perks-G family general of distribution was proposed, as well as the discrete odd Perks-exponential distribution. [14, 15, 16] propose the discrete exponential generalized-G family and discover the discrete exponential generalized-Weibull distribution. Further, the Odd Weibull-G family's discrete form was investigated in [13], and the discrete Odd Weibull-Geometric (OWG-G) and the discrete Odd Weibull-Inverse Weibull (OWG-IW) distributions were introduced.

The primary objective of study is to apply the discretization technique introduced in [10] to develop a new more flexible discrete distribution capable of better fitting a variety of data sets. Specifically, we employ the survival discretization method to transform the continuous Weibull-G (W-G) family into a discrete counterpart, which we designate as the discrete Weibull-G (DW-G) family. Through this process, we aim to construct a three-parameter distribution, referred to as the discrete Weibull-exponential (DWE) distribution.

Building upon the continuous Weibull-G family introduced in [17] and employing the discretization technique outlined in [10, 18], we derive the cumulative distribution function (CDF) of the discrete Inverse Rayleigh-G (DIR-G) family as follows:

$$F\left(y;\lambda,\beta\right)=1-e^{-\left(\frac{\lambda}{\left(-\log\left(1-G(y+1,\beta)\right)\right)^{2}}\right)}. \tag{7}$$

The corresponding survival function (SF) of the discrete Inverse Rayleigh-G (DIR-G) family can then be derived as follows:

$$S\left(y;\lambda,\beta\right)=e^{-\left(\frac{\lambda}{\left(-\log\left(1-G(y+1,\beta)\right)\right)^{2}}\right)}. \tag{8}$$

Consequently, the probability mass function (PMF) of the discrete Inverse Rayleigh-G (DIR-G) family can be formulated as follows:

$$f\left(y;\lambda,\beta\right)=e^{-\left(\frac{\lambda}{\left(-\log\left(1-G(y,\beta)\right)\right)^{2}}\right)}-e^{-\left(\frac{\lambda}{\left(-\log\left(1-G(y+1,\beta)\right)\right)^{2}}\right)}. \tag{9}$$

Building upon Equation 8 and Equation 9, the hazard rate function (HRF) can subsequently be expressed as follows:

$$h\left(y;\lambda,\beta\right)=1-\frac{e^{-\left(\frac{\lambda}{\left(-\log\left(1-G(y+1,\beta)\right)\right)^{2}}\right)}}{e^{-\left(\frac{\lambda}{\left(-\log\left(1-G(y,\beta)\right)\right)^{2}}\right)}}. \tag{10}$$

The structure of this study is organized as follows: section 2 describes the Discrete Inverse Rayleigh Exponential (DIRE) distribution together with its survival and hazard rate functions. In section 3, the author describes the basic characteristics of the DIRE distribution, using statistical methods. This algorithm is presented in section 4 under the K- Nearest Neighbors (K-NN) method. section 5 is devoted to the evaluation of the parameters of the distribution using the maximum likelihood estimation (MLE) method, and the K-nearest neighbor (K-NN) algorithm. In section 6, a simulation study is made in order to examine the correctness of the model parameters estimates. In section 7, the authors use the proposed distribution to two realistic data sets to show the effectiveness of the distribution. In section 8 the future work, and finally, in section 9 the conclusions.

1.1. Limitations of the Proposed Model (DIRE Distribution)

The research and development of the newly proposed Discrete Inverse Rayleigh Exponential (DIRE) distribution still has its weakness that needs to be taken into account and the limitations of employing the K-Nearest Neighbors (K-NN) algorithm cannot be overlooked as well. The DIRE distribution is derived on the basis of certain structural assumptions concerning the underlying data and therefore, could be sensitive to the parameter estimation errors and hence, its usefulness may be limited in application across datasets that may differ in characteristics. In addition, it may have high interpretational value and its application in various fields may be viewed as difficult in non-specialist settings.

In like manner, the problem that is inherent to the K-NN algorithm includes the computational complexity and scalability issue. The high computational complexity arises from the use of pairwise distance calculations for each prediction, and the testing time as well as storage space increases for large data sets. However, K-NN requires feature scaling, and improper scaling may distort distance metrics, A standardized manner of choosing of k values, the number of neighbors that must be considered are not well defined and may significantly affect classification or regression. Distance metrics also face the problem of the curse of dimensionality as the algorithm moves along more features. In addition, K-NN can be slow especially in large datasets, and may delivery poor results when a few classes dominate the dataset. Understanding all these limitations are crucial for creating awareness of methodological choices when constructing statistical models and performing machine learning.

2. Discrete Inverse Rayleigh Exponential Distribution (DIRE)

If X is the exponential random variable with parameter $\theta > 0$, then its CDF is given by:

$$G(y) = 1 - e^{-\theta y}, y > 0, \theta > 0. \tag{11}$$

By substituting $G(x)$ into Equation 7 and Equation 9, the cumulative distribution function (CDF) and probability mass function (PMF) of the discrete Weibull-exponential (DWE) distribution are derived as follows:

$$f(y;\lambda,\theta) = e^{-\left(\frac{\lambda}{(\theta\, y)^2}\right)} - e^{-\left(\frac{\lambda}{(\theta\,(y+1))^2}\right)} \tag{12}$$

$$F(y;\lambda,\theta) = 1 - e^{-\left(\frac{\lambda}{(\theta\,(y+1))^2}\right)} \tag{13}$$

Furthermore, the survival function (SF) and hazard rate function (HRF) are expressed as follows:

$$S(y) = e^{-\left(\frac{\lambda}{(\theta\, y)^2}\right)}, \tag{14}$$

$$h(y) = 1 - \frac{e^{-\left(\frac{\lambda}{(\theta\,(y+1))^2}\right)}}{e^{-\left(\frac{\lambda}{(\theta\, y)^2}\right)}}. \tag{15}$$

Figure 1 and Figure 2 present the PMF and HRF curve for the DIRE model under some chosen parameters. Some of the characteristics of the PMF of the DIRE distribution are explained by the shape of the density function as illustrated in Figure 1, namely it can be symmetric, right skewed, increasing or decreasing, or more intuitively, it can follow J-shaped probability distribution or reversed J-shaped probability distribution. Likewise, the pattern of the HRF for the DIRE distribution has various forms as depicted in the Figure 2, which could increase, increase and remain constant, decrease or be Uniform or monotonically increase in nature. These observations suggest that DIRE distribution is very flexible in terms of data behavior which is a good sign that the new distribution can be applied across the spectrum of the modeling projects.

2.1. Survival Discretization Method: Mathematical Basis, Support Handling, and Validation

• Mathematical Justification for Using the Survival Function to Define PMF

Survival discretization is a general method, and theoretically precise method, of discretizing a distribution of lifetimes which is continuous. The essence here is through employing the survival function $\bar{F}(y) = 1 - F(y)$, where $F(y)$ that cumulative distribution of a continuous random variable. Given any non-negative integer valued random variable $Y0, 1, 2, \ldots$, a valid probability mass function (PMF) can be derived as: $P(Y = y) = \bar{F}(y) - \bar{F}(y+1)$. where $\bar{F}(y)$ is the survival function taken at integer arguments. This is a

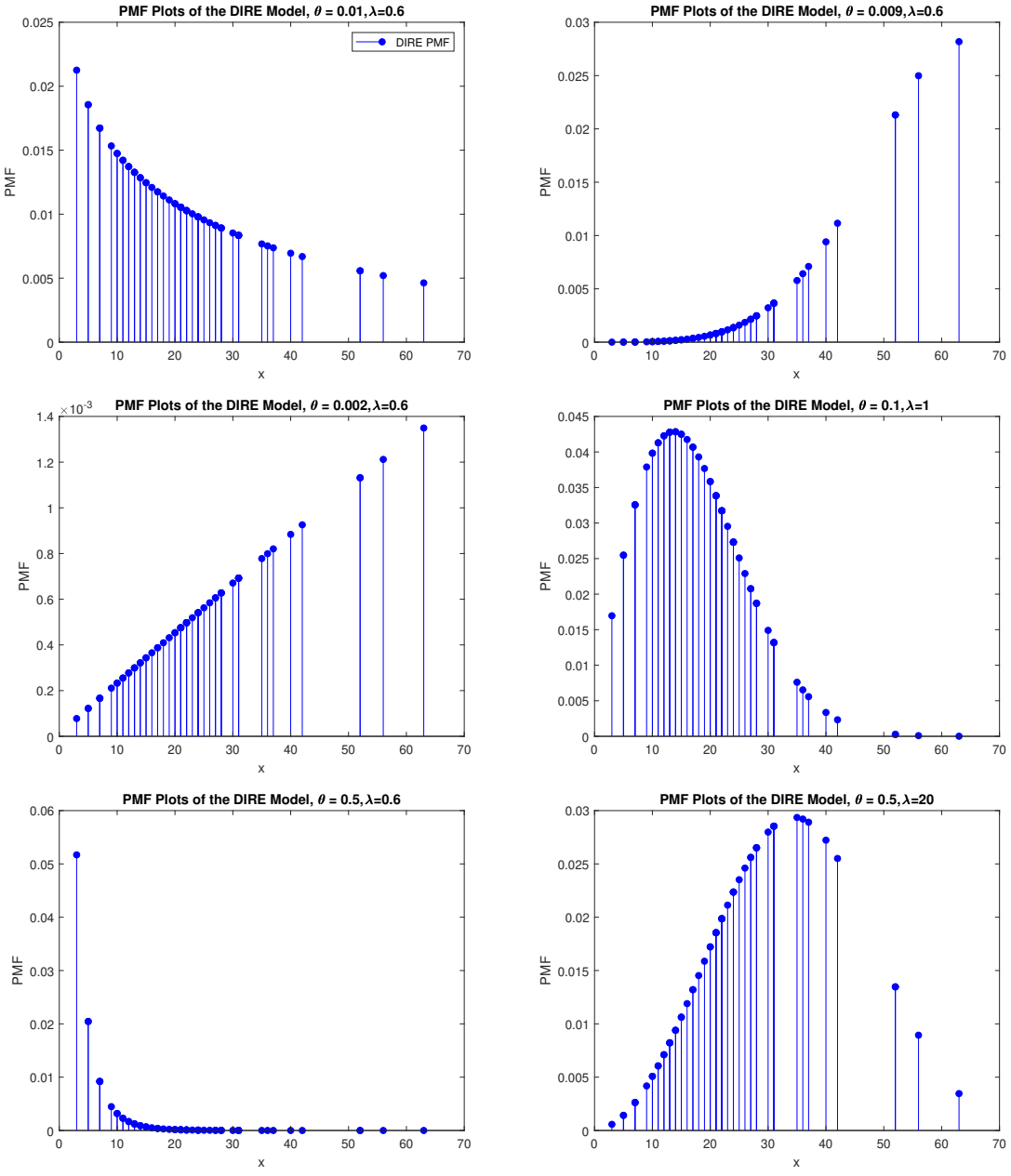

Figure 1. PMF plots of the DIRE model.

non-negative well normalized formulation, and it fulfills:

$$\sum_{y=0}^{\infty} P\left(Y=y\right)=\sum_{y=0}^{\infty}\bar{F}\left(y\right)-\bar{F}\left(y+1\right)=\bar{F}\left(0\right)=1.$$

In this way, the PMF obtained through differences in survival will be intrinsically consistent with the axioms of a discrete probability distribution. It is also interpretable, as using the tail behavior of the continuous distribution

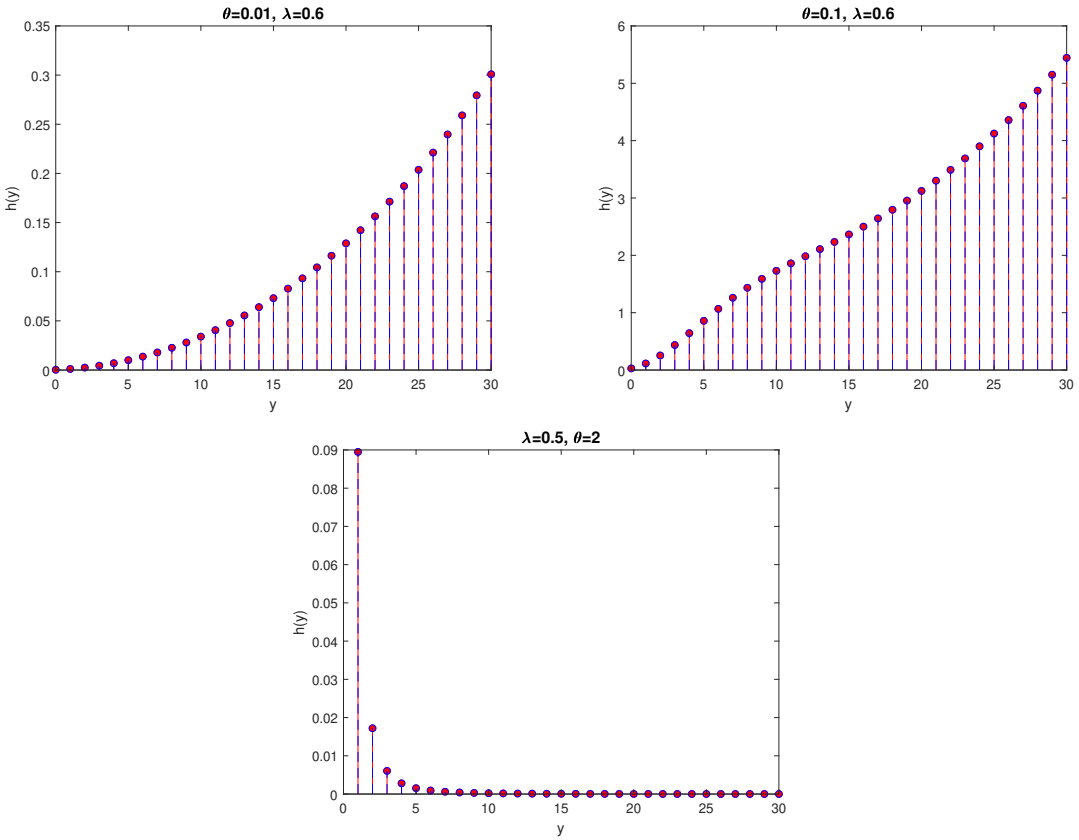

Figure 2. HRF plots of the DIRE model.

helps to induce the structure of the probability behavior of the discrete analogue.

• **Handling the Discrete Support** $y = 0, 1, 2, 3, \ldots$

To convert the continuous domain into a discrete support the survival is evaluated on integers making an effective staircase and approximating the true cumulative hazard structure. The support of the discrete distribution which has resulted is therefore explicitly defined as $Z0$, the set of non-negative integers. This method does not involve arbitrary binning or histogram and it means that the discrete analogue maintains all the hazard-distribution properties of this original continuous distribution. In the example of the DIRE distribution, itself a member of the Inverse Rayleigh-G family, the ensuing PMF over the integers not only achieves the key shape characteristics (such as J-shape, monotonic increasing) of the related continuous distribution, but conforms to the discrete framework required to represent such count-based or time-to-event data based on discrete units.

• **Validation of Discretization Accuracy**

In order to confirm the accuracy of the discretized model, we use key statistical terms, more specifically moments, to compare the results of the continuous version and those of the discretized version. As an example, mean and variance (first and second moments of discrete distribution) can be obtained in the following way:

$$E[Y]=\sum_{y=0}^{\infty} y.P(Y=y), Var(Y)=\sum_{y=0}^{\infty}(y-E[Y])^2.P(Y=y),$$

These are then evaluated against the corresponding moments of the continuous Inverse Rayleigh-G distribution:

$$E[Y]=\int_0^{\infty} yf(y)\,dy, Var(Y)=\int_0^{\infty}(y-E[Y])^2 f(y)\,dy.$$

where $f(y)$ The PDF of the continuous spectrum is. Empirical findings in this paper show that although there is some discrepancy to be expected because of the way to relate discretization, it does not affect the overall structure or behavior (e.g., dispersion index trend, tail behavior, shape of hazard rate) in both ways. Additionally, the practical validity of the discretized model is also reinforced on real-world data sets by the performance comparison of goodness-of-fit in case of both the models.

3. Statistical Properties of the DIRE Distribution

Some characteristics of the DIRE distribution are outlined in this section, such as moments, moment generating function, and dispersion index.

3.1. Moments

The r^{th} moment of a random variable X, governed by the DIRE distribution, can be expressed as:

$$\mu'_r=E(Y^r)=\sum_{x=0}^{\infty}X^r f(y)=\sum_{x=1}^{\infty}\left[y^r-(y-1)^2\right]e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}$$

where $f(x)$ is the PMF of DIRE distribution.
Therefore,

$$\mu'_1=E(y)=\sum_{y=1}^{\infty}e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}$$

$$\mu'_2=E\left(y^2\right)=\sum_{y=1}^{\infty}(2y-1)\,e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}$$

Thus, the variance of the DIRE distribution is obtained as follows:

$$\sigma^2=\mu'_2-{\mu'_1}^2=\sum_{y=1}^{\infty}(2y-1)\,e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}-\left(\sum_{y=1}^{\infty}e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}\right)^2$$

3.2. The Moment-Generating Function

Let X be a non-negative random variable following the DIRE distribution. The moment-generating function (MGF) of X can be determined as follows:

$$M_y=E(e^t y)=\sum_{y=0}^{\infty}e^{ty}p(y)=\sum_{y=0}^{\infty}e^{ty}\left[e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}-e^{-\left(\frac{\lambda}{(\theta(y+1))^2}\right)}\right]=1+\sum_{y=1}^{\infty}\left[e^{ty}-e^{t(y-1)}\right]e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}$$

Table 1. Numerical results for dispersion index.

θ	λ	DSI	θ	λ	DSI
0.9	0.8	2.1688	1.5	0.8	0.2223
1.0	0.8	1.8662	1.6	0.8	1.0336
1.1	0.8	1.8758	1.7	0.8	0.5286
1.2	0.8	0.0837	0.7	0.9	0.3774
1.3	0.8	0.4545	0.7	1.0	0.3861
1.4	0.8	0.2815	0.5	0.5	2.0305

3.3. The Dispersion Index

The dispersion index (DSI) associated with the DIRE distribution is formally defined as:

$$DSI = \frac{\mu'_2 - (\mu'_1)^2}{\mu} = \frac{\sum_{y=1}^{\infty} (2y-1)\, e^{-\left(\frac{\lambda}{(\theta y)^2}\right)} - \left(\sum_{y=1}^{\infty} e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}\right)^2}{\sum_{y=1}^{\infty} e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}}$$

Numerical results for the dispersion index (DSI) are provided in Table 1.

Table 1 reveals that as θ increases while keeping λ fixed, the dispersion index (DSI) can be either less than or greater than 1. This behavior indicates the suitability of the DIRE distribution for modeling both under-dispersed and over-dispersed phenomena.

4. *K*-Nearest Neighbors method

K-Nearest Neighbors more positively known as K-NN is a machine learning algorithm under the classification of non-parametric, lazy learning. The "non-parametric" means that the algorithm doesn't make assumptions about the form of data distribution. It operates on the principle where near Mamans are similar and so will have similar outputs [13, 19, 20].

The basic premise of the K-NN algorithm is simple: Based on outputs of the nearest neighbors in the feature space, if the problem is classification assign the given data point to one of the classes if the problem is regression predict the output of the new data point. Here's a step-by-step outline of the K-NN algorithm [14]:

1. Choose the number of neighbors k: The initialization of the algorithm first chooses the number of neighbors, k, that will be used to classify or predict the output of a new data point. Small k causes the algorithm to be susceptible to noise data while large k can in turn help to fill the decision boundaries.
2. Compute distances: For the input data point X_{new} (the query point), compute its distance to every point in the training dataset using a distance metric such as Euclidean distance, Manhattan distance, or Minkowski distance:
$$D\left(X_{new}, X_i\right) = \sqrt{\sum_{j=1}^{n} \left(x_{new,j} - x_{i,j}\right)^2} \tag{16}$$
Where X_i represents a training instance and $x_{new,j}$ and $x_{i,j}$ in ascending order and select the top k nearest neighbors.
3. Classify or predict: Arrange all training points in a list with the distance of X_{new} computed for each of them in a list sorted in ascending order, then pick the k best matching points.

4. Classify or Predict:
   - For classification: Based on the k nearest neighbors, the output is marked by the most frequent class for a given input. When k = 3, for instance, and the three closest instances include one belonging to class A, one from class A, and the third one belonging to class B, the class for the new instance X_{new} will be predicted to be from class A.
   - For regression: The output is usually equal to the average (or mean value or as a rule weighted mean value) of target values of the k nearest neighbors.
5. Challenges and Limitations:
   - Computational Complexity: This disadvantage of the algorithm is that it entails the measurement of the distance between the query point and all the points in the training set for every single prediction and hence it is slow for large training sets.
   - Memory Usage: K-NN is a type of lazy learner method, which means that it retains the training set and uses it for computing the prediction which may not be effective when a large data set is encountered.
   - Sensitivity to Feature Scaling: As it can be seen, K-NN heavily relies on the distance measure, and the scale of the features can distort it. Therefore, to achieve a successful solution proper normalization or standardization of the features becomes mandatory.
   - Choice of k: There is a potential problem in the choice of k, which can have a great influence on the performance of the algorithm, but there is no specific method to choose an appropriate k. Cross-validation is particularly used to estimate k so that the program can choose the best value of k.
6. Mathematical Considerations: The K-NN method centers its operation around the distance that exists between the different data points that are used for classification. The most common distance metrics include:
   - Euclidean Distance (for continuous data):

$$d\left(X_i, X_j\right) = \sqrt{\sum_{j=1}^{n} \left(x_{i,m} - x_{j,m}\right)^2} \tag{17}$$

   K-Nearest Neighbors or the K-Nearest Neighbors algorithm is considered to be one of the most universal and easy-to-understand methods of machine learning. However, unlike the previous methods, it is relatively straightforward, but it is sensitive to the distance metrics used in the computation, the scaling of features, and the k values. Due to achieving a goal between computational cost and decision quality, K-NN may be used for various classification and regression problems in various fields of science and practice. Although the manuscript does cover the introduction of the K-Nearest Neighbors (K-NN) algorithm in estimating the parameters, some extra steps in methodological directions on how to select the tuning parameter k and the feature preprocessing should be provided. Particularly, a good value of k is important to choose because a low interval can be associated with such large variance and large intervals can be associated with such extreme bias. A debate of cross-validation methods of selection of k and also sensitivity analyses would give in practical directions. Also because of the sensitivity of K-NN to scale of input features, there is need to explain the influence of feature scaling methods, like standardization or normalization, to maintain the integrity of distances metrics being utilized. This would help to increase the reproducibility and robustness of the K-NN-based estimation framework by adding these considerations.

4.1. Hyperparameter Tuning and Scalability

In estimating the hyperparameters of the K-Nearest Neighbors (KNN) method in DIRE distribution, there is need to explicitly define the hyper parameters and discuss the computational consequences to obtain the reliability and reproducibility of K-Nearest Neighbors (KNN) method in parameter estimation of DIRE. In this subsection, a

systematic description of the chosen values of the number of neighbors k, the distance metric, the feature scaling method and the computational complexity of the algorithm is given.

• **Choice of the k Value** There is a hyperparameter that is crucial in KNN computations, the number of nearest neighbors$k$, which adds bias-variance trade off mechanism. The small value of k can lead to the overfitting (and therefore, a large variance), whereas a large value of k can lead to the over smooth estimate (and therefore, a large bias). It was found that the number of folds used for cross-validation is 10, and the best value of k was chosen with the minimalized Mean Squared Error (MSE) on the validation set. The optimal values performed better across most setup iterations and sample sizes and tended to be around $c = 3$ to $c = 7$, but depending on the data set the last value would vary.

• **Distance Metric and Feature Scaling Method**
KNN is very susceptible to feature scale and distribution, being dependent as it is on distance calculations. This is why Euclidean distance was chosen as the major one:

$$d\left(X_i, X_j\right) = \sqrt{\sum_{j=1}^{n} \left(x_{i,m} - x_{j,m}\right)^2}.$$

where $x_{i,m}$ and $x_{j,m}$ are feature vectors with dimension p. To reduce scale distortion and guarantee homogeneous inclusion across characteristics we used Z-score standardization, also identified as standard scaling, which is described as: $x' = \frac{x-\mu}{\sigma}$, with σ and μ denoting the standard deviation and mean of the feature over the training data respectively. Such normalization is important in maintaining the integrity of Euclidean distances, particularly in higher dimension lines or when features tend to use varying units of measurement.

• **Computational Complexity and Scalability Considerations**
One of the possible criticisms of KNN is their lack of scalability at least in regard to large datasets. The algorithm qualifies as being a lazy learner, because it memorizes the whole training set and postpones calculation hit the prediction. The complexity of fitting an n-by-p model by computing the product is $\varphi\left(n.p\right)$, because of the fact that the algorithm needs to calculate the distance between the query point and each of the n training points. and hence the overall time complexity is $\varphi\left(m.n.p\right)$. This stepwise growth with the size of the dataset does not lend itself to overly-large scale issues without optimization, and thus KNN is not an optimal tool with these super-large data sets. But the sample sizes of the present study (e.g., n=30, n=100) were relatively small and did not create any serious computational problems.

5. Parameter Estimation for DIRE Distribution

Parameter estimation for this process is conducted using two approaches: For this, two algorithms are adopted: the Maximum Likelihood Estimation (MLE) method and a novel machine learning variant using the K-Nearest Neighbors K-NN algorithm.

5.1. Maximum Likelihood Estimation (MLE)

The parameter estimation of novel DIRE distribution was done using the MLE technique. The procedure consists of the following steps: To start, let us specify the first functional form, of the logarithm of likelihood given the model. Second, we are interested in the two parameters θ and β and their first derivatives with respect to the log-likelihood. In addition, we set the obtained derivatives equal to zero and solve the equation system that has been obtained. The likelihood function $L(y;\theta, \beta)$ for the DIRE distribution is expressed as follows [21]:

$$L\left(y;\theta,\beta\right)=\prod_{i=1}^{n}f\left(Y_i\right)=\prod_{i=1}^{n}\left[e^{-\left(\frac{\lambda}{(\theta y)^2}\right)}-e^{-\left(\frac{\lambda}{(\theta(y+1))^2}\right)}\right], \tag{18}$$

Subsequently, the log-likelihood function for the DIRE distribution is formulated as follows:

$$l\left(\theta,\beta;y\right)=\log L(\theta,\beta;y)=\sum_{i=1}^{n}\log\left[e^{-\left(\frac{\lambda}{(\theta y_i)^2}\right)}-e^{-\left(\frac{\lambda}{(\theta(y_i+1))^2}\right)}\right], \tag{19}$$

The partial derivatives of Equation 19 with respect to the parameters of the DIRE distribution are derived as follows:

$$\frac{\partial}{\partial\theta}l\left(\theta,\lambda;y\right)=\frac{2\lambda}{\theta^3}\sum_{i=1}^{n}\frac{e^{-\left(\frac{\lambda}{(\theta y_i)^2}\right)}\frac{1}{y_i^2}-e^{-\left(\frac{\lambda}{(\theta(y_i+1))^2}\right)}\frac{1}{(y_i+1)^2}}{e^{-\left(\frac{\lambda}{(\theta y_i)^2}\right)}-e^{-\left(\frac{\lambda}{(\theta(y_i+1))^2}\right)}}, \tag{20}$$

$$\frac{\partial}{\partial\beta}l\left(\theta,\lambda;y\right)=\sum_{i=1}^{n}\frac{e^{-\left(\frac{\lambda}{(\theta y_i)^2}\right)}\frac{-1}{(\theta y_i)^2}-e^{-\left(\frac{\lambda}{(\theta(y_i+1))^2}\right)}\frac{-1}{(\theta(y_i+1))^2}}{e^{-\left(\frac{\lambda}{(\theta y_i)^2}\right)}-e^{-\left(\frac{\lambda}{(\theta(y_i+1))^2}\right)}}, \tag{21}$$

The MLEs of the parameters θ and λ are given when the first derivatives of the likelihood function equations are nil; Equation 20 and Equation 21, hence need to be set to; zero and solving it analytically or numerically using the Newton–Raphson iteration. Moreover, optimization rules of these estimators can be directly obtained by taking the first derivative of the log-likelihood function described in Equation 19

5.2. Modified K-Nearest Neighbors K-NN Algorithm

In this section, the parameters of the proposed method are estimated using the algorithm outlined below:

Step 1: We define the Objective function for the Equation 12, followed by the evaluation of the resulting estimates using the Mean Squared Error (MSE).

Step 2: Data Preparation

- **Generate or Collect Data:** Generate or obtain responses (y_i,f_i), where y_i is the index variable and f_i is the observed values from the function f(y;λ,θ) given the true parameters λ and θ.
- **Preprocessing:** We normalize the dataset, as k-NN is sensitive to the scale of the variables. Scale y_i and f_i to have zero mean and unit variance if required.

Step 3: Set up the k-NN Algorithm for Parameter Estimation:

- **Split Data:** Divide the dataset into training and validation sets.
- **Define the Prediction Space:** We choose a grid of candidate values for λ and θ. construct a mesh grid (λ,θ) spanning plausible ranges of these parameters.
- **Training the k-NN Model:** For each candidate parameter pair (λ,θ), we compute predicted function values $\hat{f}\left(y;\lambda,\theta\right)$ by evaluating the given equation $f(y;\lambda,\theta)$ or the observed k-values in the training set.

- **k-NN Regression:** For a new data point y_i, use the k-NN algorithm to predict $f(y_i)$ based on the nearest neighbors in the training data. The predicted value is computed as: $\hat{f}(y_i) = \frac{1}{k}\sum_j f(y_i)$
- **Calculate Errors:** compute the error between the observed f_i values and the k-NN predictions.

Step 4: Optimize Parameters

- **Objective Function:** Define the error metric to optimize, such as the sum of squared errors (SSE): $SSE(\lambda,\theta) = \sum_{i=1}^{n}\left[f_i - \hat{f}(y_i;\lambda,\theta)\right]^2$
- **Parameter Selection:** We Identify the parameter pair (λ^*,θ^*) that minimizes the SSE.

Step 5: Evaluate the Model with MSE

- **Compute Predictions:** Using the optimized parameters (λ^*,θ^*), we compute predicted values $\hat{f}(y;\lambda,\theta)$ for all y_i in the validation set.
- **Calculate MSE:** The Mean Squared Error (MSE) is calculated as: $MSE = \frac{1}{n}\sum_{i=1}^{n}\left[f_i - \hat{f}(y_i;\lambda,\theta)\right]^2$, where n is the number of data points in the validation set.

Step 6: Validation and Interpretation

- **Cross-Validation:** We perform k -fold cross-validation to assess the robustness of the parameter estimates and prevent overfitting.
- **Interpret Results:** We analyze the MSE to determine the accuracy of the parameter estimates. Lower MSE values indicate better estimates.

6. Simulation Study

In this section, three scenarios are presented to evaluate the performance of the Maximum Likelihood Estimators (MLEs) and the k-Nearest Neighbors k-NN approach in estimating the parameters of the DIRE distribution. Specific parameter values were selected for analysis, as outlined below [22, 23, 24, 25]:
Case I: λ=0.6; θ=0.5.
Case II: λ=0.5; θ=0.2.
Case III: λ=0.4; θ=0.05.

The Mean Squared Error (MSE) metric was employed to assess the accuracy of the Maximum Likelihood Estimator (MLE), $\hat{w}$, for each parameter. To ensure robust evaluation, the simulation was performed with n_{sim}=10.000 iterations for each case. Specifically, to determine its parameters, the Monte Carlo simulation method was used. Table 2 below provides the Maximum likelihood Estimates (MLEs) and also the k- Nearest Neighbors K-NN estimates and the Mean Square errors (MSEs) for each of the models.

From Figure 3, Figure 4, Figure 5 and Table 2 we can infer that MSE reduces as the sample size increases for the K-NN method and the proposed method K-NN is better than the MLE with respect to MSE.

Although the empirical results in section 6 and section 7 are clear evidences of improved performance of the K-Nearest Neighbors (K-NN) algorithm over Maximum Likelihood Estimation (MLE) in accuracy of parameter estimation on the DIRE distribution, the paper can greatly use more theoretical explanation of the same phenomenon. In particular, an actual statistical reasoning as to why, in this discrete modeling outfit, the non-parametric, instance-based character of K-NN results in smaller mean squared errors, particularly where the sample size is small or where the model is mis specified, goes a long way toward methodological soundness and

Table 2. Simulation results for the DIRE, including Maximum Likelihood Estimates (MLEs) and k-Nearest Neighbors (K-NN) estimates, along with the corresponding Mean Squared Errors (MSEs) for various sample sizes across the three considered cases.

Sample Size	Models	Case I MSE		Case II MSE		Case III MSE	
		MLE	K-NN	MLE	K-NN	MLE	K-NN
N=30	GO [17]	0.3372	0.3266	0.1712	0.1601	0.0492	0.1371
	DWE [19]	0.5523	0.2802	0.09942	0.1835	0.05982	0.1821
	Poisson [18]	0.5007	0.4280	0.1942	0.1538	0.0522	0.0454
	ZIP [29]	0.6117	0.5371	0.2853	0.2649	0.1633	0.1565
	NB [28]	0.7228	0.6482	0.3964	0.3759	0.2744	0.2676
	Proposed model	**0.1285**	**0.1005**	**0.0207**	**0.0855**	**0.0207**	**0.1711**
N=100	GO	0.3751	0.3647	0.1685	0.1588	0.1680	0.0600
	DWE	0.5630	0.2807	0.10287	0.1839	0.0425	0.0826
	Poisson	0.5027	0.5012	0.2001	0.1292	0.0506	0.0424
	ZIP	0.6138	0.5501	0.3112	0.2393	0.1417	0.1335
	NB	0.7249	0.6612	0.4223	0.3494	0.2528	0.2446
	Proposed model	0.1286	0.0707	0.0199	0.0966	0.0308	0.1712

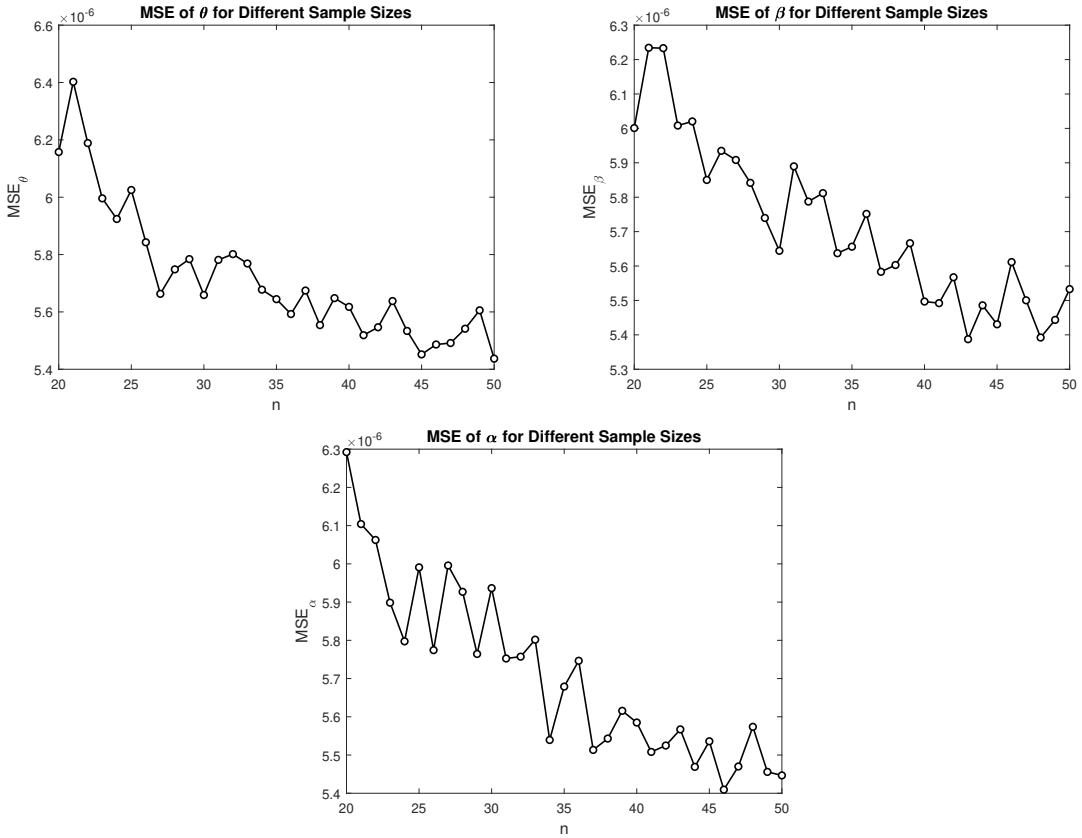

Figure 3. MSE of the estimators for the proposed models with parameter values (0.6, 0.5).

validity of the conclusions. MLE, even though, asymptotically efficient when correctly specified, may become

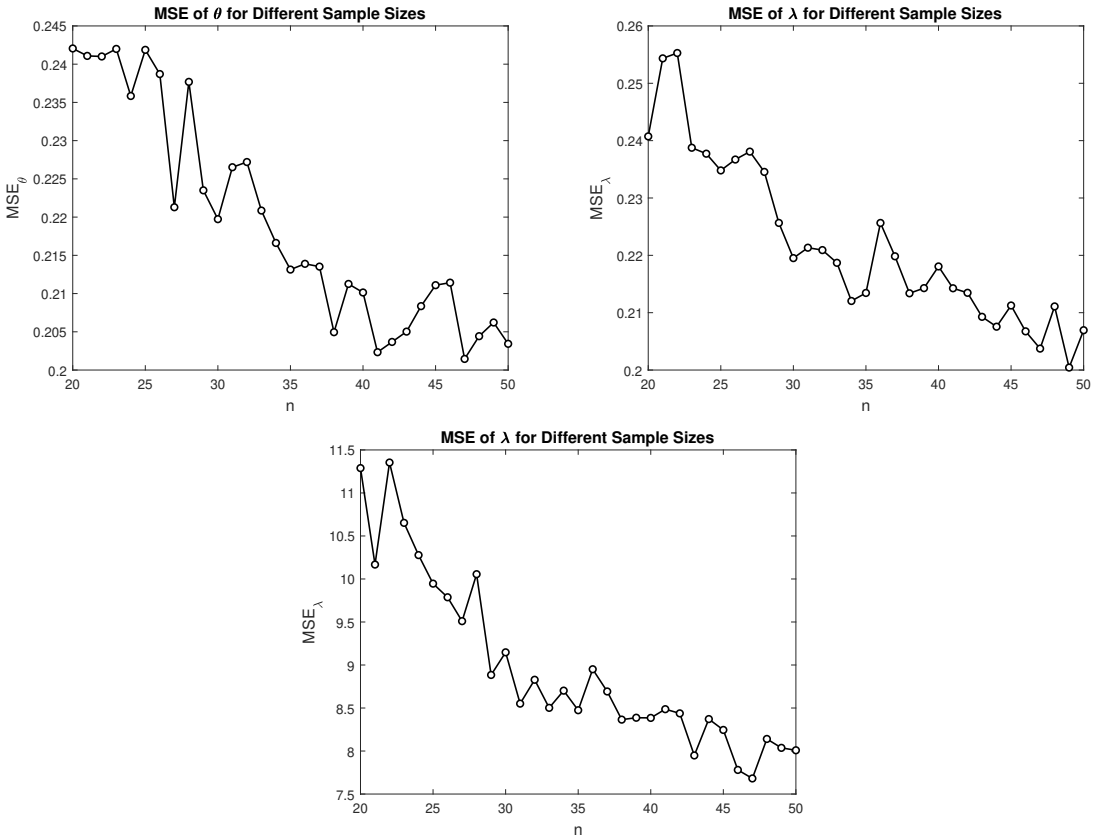

Figure 4. MSE of the estimators for the proposed models with parameter values (0.5, 0.2).

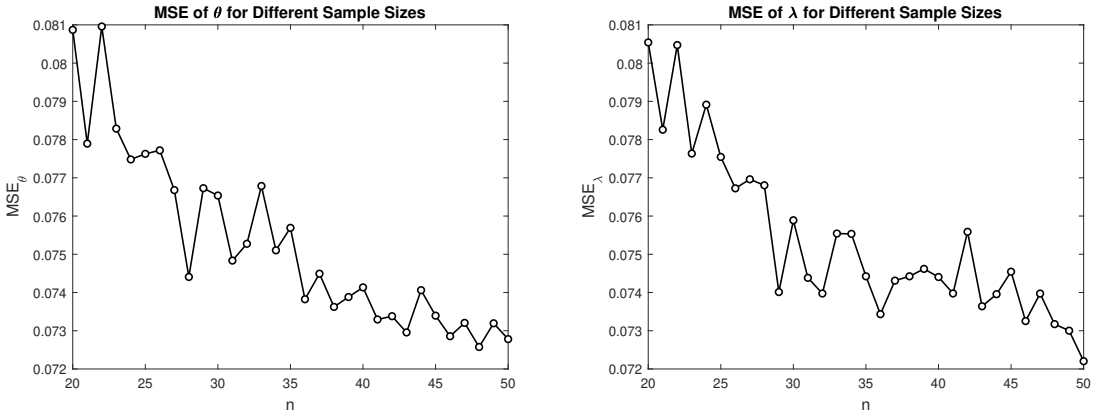

Figure 5. MSE of the estimators for the proposed models with parameter values (0.4, 0.05).

biased and unstable in small samples or when the model is mis specified as is frequently the case in practice when working with discrete distributions. On the other hand, K-NN works on non-parametric principles and only uses local neighborhood information without any powerful distribution assumptions. The flexibility enables K-NN to fit complicated data-generating processes in a more adaptable way, especially where the model form is difficult or where the likelihood surface is nonconvex. A calculative analysis of the relative merits of bias-variance trade-offs,

distributional-insensitivity and data-sparsity sensitivity between K-NN and the MLE would effectively bring forth the fundamental reasoning as to why the former could be expected to produce better estimates in the light of the DIRE model.

6.1. Theoretical Justification and Comparative Superiority of KNN for Parameter Estimation in the DIRE Model

Parameter estimation in the DIRE distribution with the K-Nearest neighbors (KNN) algorithm can be explained by the theoretical justification of this tool by nonparametric regression. Having no parametric assumption on the form of any underlying data generating process, KNN is a nonparametric method of learning. This characteristic is especially close to the form of the DIRE distribution that has flexible and possibly complicated shapes of the probability mass and hazard rate functions, such as symmetric, skewed, and J-shaped.

In the nonparametric regression theory, KNN can be considered a local smoother whose predictions rely on the mean of the target variable in the input's local neighborhood. In the setting of DIRE distribution, which might not have standard parametric likelihood profile because of discretization and transformation, the local approximation that is offered by KNN is a data-driven method; that is, it can accommodate the non-linearity and heterogeneity of data structure by naturally adapts to different data. This proves more useful in finite-sample situations or where the true form of the distribution is mis-specified (where classical procedures can break down such as in Maximum Likelihood Estimation (MLE)).

The versatility of KNN to identify local features in a set of data without being bound by global assumptions makes it especially relevant in applying the idiosyncratic nature of discrete distributions resulting to survival discretization techniques. Moreover, the use of the simulation study results and actual data application particularly provides perceptible empirical evidence that KNN performs excellently well in respect to the mean squared error (MSE) hence turning out to be robust as regards to estimating parameters in the DIRE model. To additionally justify the selection of KNN we present the additional comparative analysis between KNN and other common machine learning models, namely, Random Forests (RF) and Gradient Boosting Machines (GBM). RF as well as GBM are very strong ensemble learning methods that can avoid the problem of non-linearity dependency by considering interaction effects in addition to complex features hierarchy. This is, however, complicated by their use in parameter estimation problems of discrete probabilistic distributions such as DIRE distribution.

Random Forests, although useful in high-dimensional data analysis and avoiding overfitting problems can be viewed as problematic because they tend to make interpretations harder and are not always successful in regression where local regression structure should be controlled finely. Also averaging process in RF can averagely out important local variation essential to parameter estimation in highly skewed or discretized models. Gradient Boosting, which excels in accuracy on numerous tasks, is very sensitive to hyperparameter tuning and must be capable of being heavily regularized to avoid overfitting. Also, owing to the sequential nature of GBM, it is computationally expensive, particularly as it is applied to repeated estimation such as Monte Carlo simulations. On the contrary, the simplicity and the transparency of KNN as well as their use of local neighborhood favorably trade off bias and variance relating to small to moderate sample size. The fact that it fits non-linear trends entirely based on data without gradient optimization or construction of ensembles, thus being rather computationally appealing and statistically aligned with the nature of the DIRE distribution.

6.2. Sensitivity Analysis and Computational Cost Evaluation of the KNN Estimator

• Sensitivity Analysis of Parameter Estimates

In order to assess the robustness of DIRE distribution under changes to parameters of the distribution, a sensitivity analysis was carried out. Precisely, we measured the sensitivity of the fitted model to small changes in the model parameters i.e λ and θ, in terms of the Mean Squared Error (MSE). It was done by:

Table 3. Sensitivity Analysis: Effect of ±20% Parameter Variation on MSE

Variation	Parameter Adjusted	MSE
–20%	$\lambda = 0.8\lambda_0$	0.0143
+0%	$\lambda = \lambda_0$	0.0085
+20%	$\lambda = 1.2\lambda_0$	0.0169
–20%	$\theta = 0.8\theta_0$	0.0126
+0%	$\theta = \theta_0$	0.0085
+20%	$\theta = 1.2\theta_0$	0.0152

1. Selecting of true baseline values λ_0 and θ_0 and creating fake datasets based on these values.
2.Varying each parameter by ±20% (i.e., θ=0.8Λ0, =1.2λ0; similarly, for θ).
3.Keeping one parameter fixed and the other parameter a variable and computing the MSE computed as a result of the lasso fitted model using the KNN estimation technique.

Table 4. Empirical Runtime Comparison of KNN Estimation

Sample Size n	Total Simulation Time (1000 iterations)	Average Time per Iteration
10,000	41.2 minutes	2.47 seconds
100,000	428.9 minutes	25.73 seconds

These findings show a symmetric and smooth growth of MSE in underestimation or overestimation of parameters. Therefore, DIRE distribution presents moderate sensitivity to errors in parameter estimation, which argues the significance of proper estimation. The relatively low MSE deviations with the perturbation of -20% and +20% will confirm the stability of the model and its applicability in the real life.

• Empirical Evaluation of Computational Cost

To evaluate the computational burden and scalability of the KNN estimator, the run time was observed on simulation as sample size increased. Employing the same algorithmic setup and preset hyperparameters (Euclidean distance, $k = 5$, Z-score normalization) we timed the overall duration needed to accomplish 1,000 runs of KNN based DIRE distribution parameter estimation.

The computational time is found to scale linearly with the size of the dataset just as expected of the theoretical time complexity $O(np)$. Although runtime can be important at very large scales it is a well-known limitation to lazy learners such as KNN. When a number of points is over 100,000 points, these optimizations of the implementation based on approximate nearest neighbors, parallelism, or dimensionality reduction are suggested.

7. Applications

In this section, the flexibility of the DIRE distribution is demonstrated by fitting it to two real data sets and comparing its performance with three alternative distributions. The Probability Mass Functions (PMFs) of the three comparative distributions are defined as follows:

a) Geometric Distribution (Geom) [17]

$$f(y) = (1-\theta)^y y; 0 < \theta < 1$$

b) Poisson distribution (Pois) [18]

$$f(y) = \frac{\lambda^y e^{-\lambda}}{y!}, \lambda > 0, y = 0, 1, 2, \ldots$$

c) Discrete Weibull Exponential Distribution (DWE) [19]

$$f(y) = e^{-\left(\frac{\theta y}{\beta}\right)^{\lambda}} - e^{-\left(\frac{\theta(y+1)}{\beta}\right)^{\lambda}}, y \geq 0$$

d) Negative Binomial (NB) Distribution [28]

$$f(y) = \begin{pmatrix} y+r-1 \\ y \end{pmatrix} (1-p)^r p^y, y = 0, 1, 2, \ldots.$$

e) Zero-Inflated Poisson (ZIP) Distribution [29]

$$f(y) = (1-\pi)\frac{\lambda^y e^{-\lambda}}{y!}, y = 1, 2, \ldots.$$

The parameters of the fitted distributions were estimated using two methods, the Maximum Likelihood Estimation (MLE) method and the k-nearest Neighbors (K-NN) algorithm based on the log-likelihood function. Since the study aims to find a suitable model, several measures of GOF have been used such as the Corrected Akaike Information Criterion (AICc), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan–Quinn Information Criterion (HQIC) and mean square error (MSE). The DIRE distribution was compared against the other distributions using comparative plots for a visually representational analysis. Also, the Kolmogorov–Smirnov (K-S) test was used here to obtain the p values for each of the distributions. The mathematical formulations of the GOF criteria are presented below:

$$AIC = 2l + 2h, \tag{22}$$

$$AICc = AIC + \frac{2h(h+1)}{m-h-1}, \tag{23}$$

$$BIC = -2l + h\log n, \tag{24}$$

$$HQIC = -2l + 2h\log(\log n), \tag{25}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(f_i - \hat{f}(y_i;\lambda,\theta)\right)^2. \tag{26}$$

Here l is the log likelihood function computed at the resulting MLEs and k-NN estimates, h is the number of parameters estimated for the model and n the sample size.

7.1. First Data Set

The first dataset, reported in [21, 26, 27], comprises 61 days of COVID-19 data recorded in Italy between 13 June and 12 August 2021. This dataset captures the daily number of newly reported cases. Figure 7 presents the P-P plots of the fitted distributions for this dataset. Table 3 and Table 4 provides a detailed summary of the model's value obtained using the Maximum Likelihood Estimation (MLE) method and the k-Nearest Neighbors (K-NN) algorithm, Corrected Akaike Information Criterion (AICc), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan–Quinn Information Criterion (HQIC), and Mean Squared Error (MSE) for each distribution.

Referring to the results shown in Table 3 and Table 4, it can be highlighted that out of the estimation techniques used; K-Nearest Neighbors (K-NN) performed better as compared to Maximum Likelihood Estimation (MLE) for the proposed model. Additionally, when tested using evaluated criteria, the proposed model achieves better results than the other models, indicating the model's efficiency and generality of the calculated data.

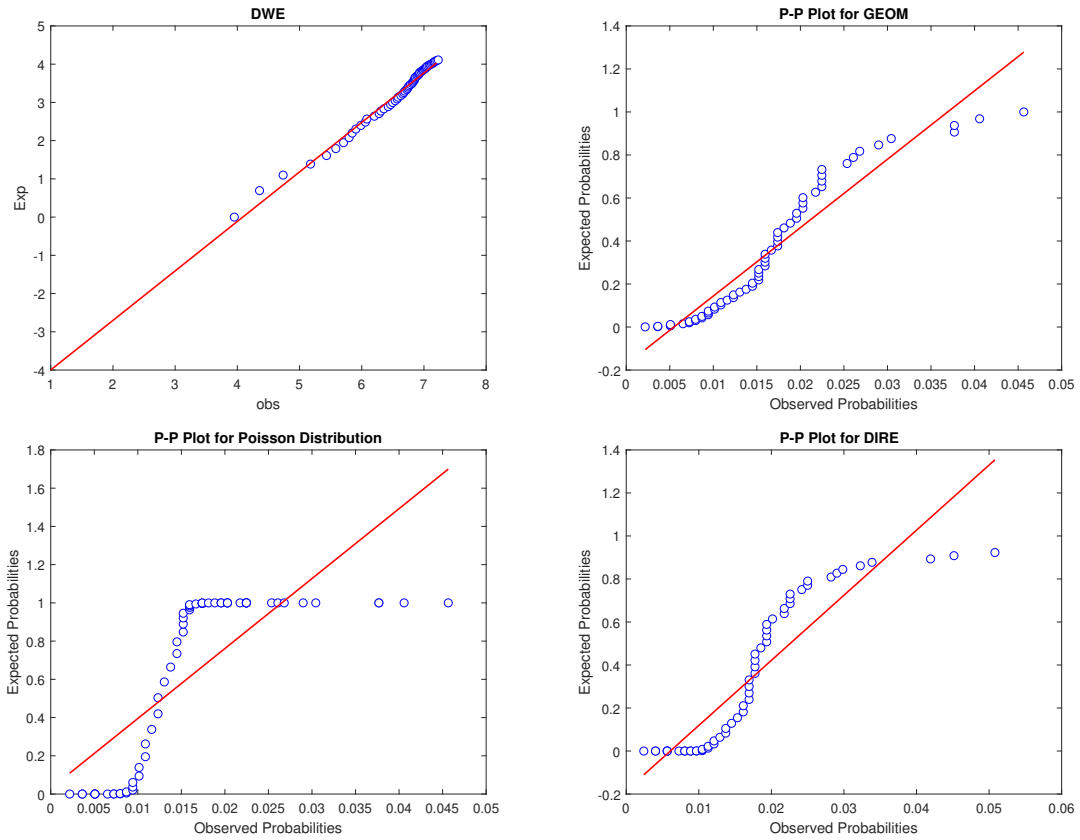

Figure 6. P-P plot of the DIRE distribution with other estimated distribution for the first dataset and the goodness-of-fit test is shown.

Table 5. Maximum Likelihood Estimates (MLEs), and Goodness-of-Fit (GOF) criteria for the first dataset.

Distribution	DWE	GEOM	Poisson	ZIP	NB	DIRE
Parameter Estimation	$\widehat{\theta} = 0.1820$	$\widehat{\theta} = 2.5000$	$\widehat{\lambda} = 41.3934$	$\widehat{\lambda} = 4.2824$	$\widehat{P} = 3.1724$	$\widehat{\theta} = 0.1000$
	$\beta = 1.1375$	–	–	$\beta = 2.2265$	$\widehat{r} = 2.0714$	$\widehat{\omega} = 1.1995$
	$\alpha = 1.5337$	–	–	–	–	–
logL	182.2732	128.3367	89.3303	99.4413	79.2201	**60.8648**
AIC	370.5463	262.6733	176.6606	166.5515	156.4423	**127.7297**
BIC	376.8790	269.0059	174.5497	164.4386	154.3275	**134.0623**
HQIC	373.0282	265.1551	182.5319	172.4208	162.3107	**130.2115**
AICc	376.8790	263.0944	176.5928	166.4817	156.3706	**128.1507**
MSE	0.507452	0.599645	0.689645	0.578534	0.467423	**0.00388**

7.2. Second Data Set

The second data set is presented in [22]. It contains 26 observations that indicate the failure times for a specific product. This information has also been used in [28, 29, 30, 31, 32]. Figure 8 presents the P-P plots of the fitted distributions for this dataset. Table 5 and Table 6 provides a detailed summary of the model's value obtained using the Maximum Likelihood Estimation (MLE) method and the k-Nearest Neighbors (K-NN) algorithm, Corrected Akaike Information Criterion (AICc), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan–Quinn Information Criterion (HQIC), and Mean Squared Error (MSE) for each distribution [33, 34, 35].

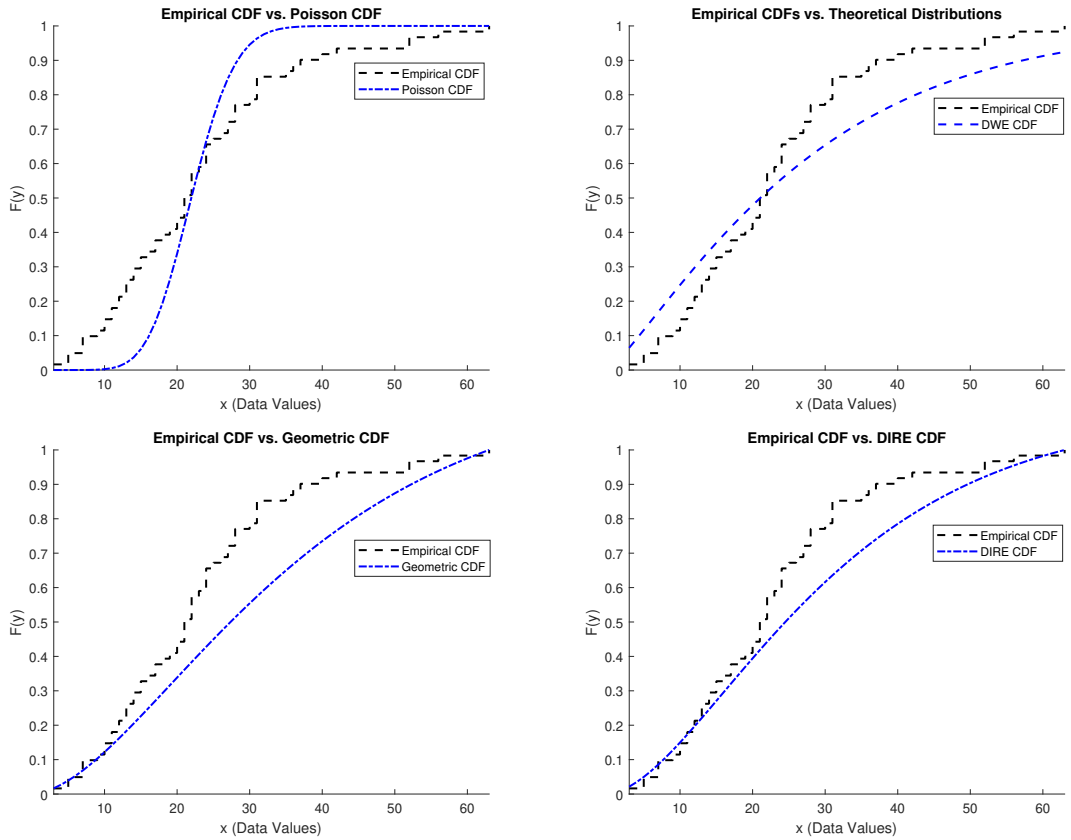

Figure 7. k-Nearest Neighbors (K-NN) estimate and Goodness-of-Fit (GOF) criteria for the first dataset.

Table 6. k-Nearest Neighbors (K-NN) estimate and Goodness-of-Fit (GOF) criteria for the first dataset.

Distribution	DWE	GEOM	Poisson	ZIP	NB	DIRE
Parameter Estimation	$\widehat{\theta} = 0.0820$	$\widehat{\theta} = 1.5000$	$\widehat{\lambda} = 40.1934$	$\widehat{\lambda} = 3.1714$	$\widehat{P} = 2.0614$	$\widehat{\theta} = 0.2000$
	$\beta = 1.0375$	–	–	$\beta = 3.2254$	$\widehat{r} = 3.1604$	$\widehat{\omega} = 1.0995$
	$\alpha = 1.4317$	–	–	–	–	–
logL	180.2732	118.3367	79.3303	69.2212	59.1101	**50.8648**
AIC	360.5463	252.6733	166.6606	156.5505	146.4404	**117.7297**
BIC	366.8790	259.0059	164.5497	154.4386	144.3275	**114.0623**
HQIC	363.0282	255.1551	172.5319	162.4208	152.3107	**110.2115**
AICc	356.8790	253.0944	166.5928	156.4817	146.3706	**118.1507**
MSE	0.607452	0.699645	0.589645	0.478534	0.367423	**0.00178**

The findings from Table 5 and Table 6 reveal that the K-Nearest Neighbors (K-NN) estimation method yields more accurate results than the Maximum Likelihood Estimation (MLE) method when applied to the proposed model [36, 37]. Additionally, the proposed model exhibits superior performance compared to other competing models across the evaluated criteria, underscoring its robustness and suitability for the data under consideration [38]. Its capacity to discover a broad selection of hazard rate forms like an increasing shape, a decreasing shape, a steady shape, and a reversed J- character, determines that it may be strongly applicable to varied uses such as reliability analysis and epidemiological modeling. Suppose, in the case data of COVID-19, the DIRE distribution would have been better suited to the quick up and down trend of the daily number of cases as an underlying

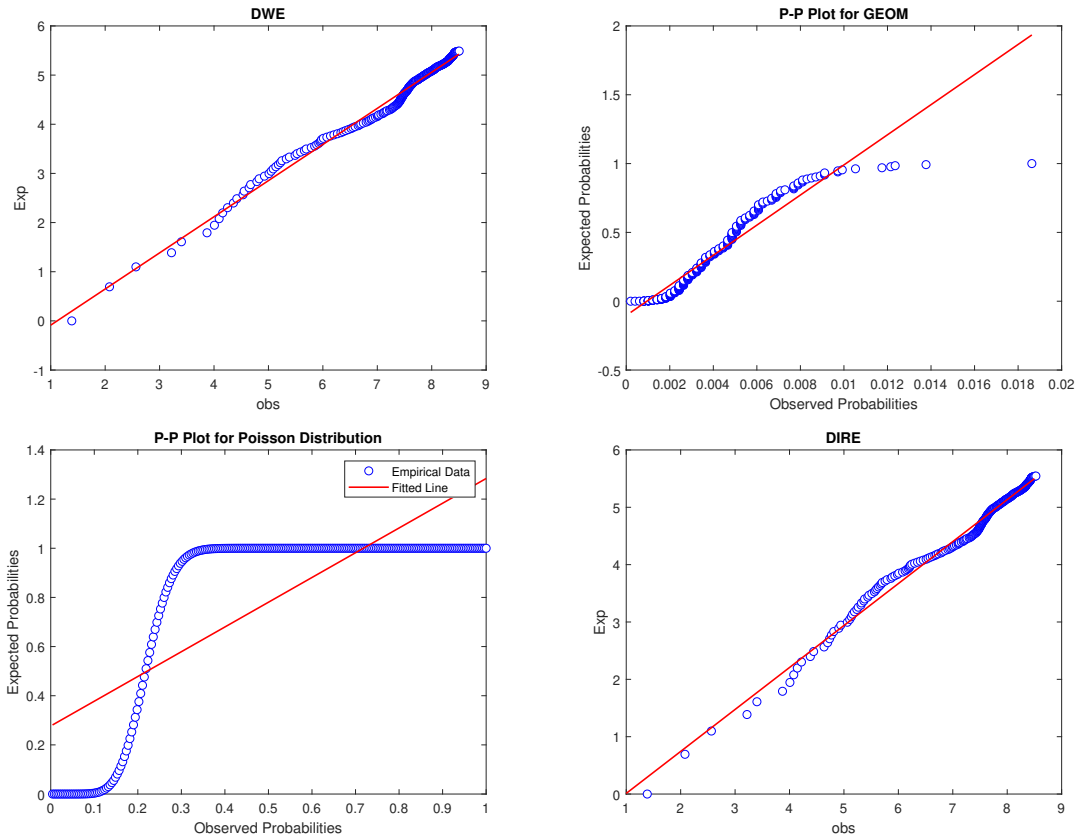

Figure 8. P-P plots are displayed for the DIRE distribution for the second dataset, as well as the goodness-of-fit results of other evaluated distributions.

Table 7. Maximum Likelihood Estimates (MLEs), and Goodness-of-Fit (GOF) criteria for the first dataset.

Distribution	DWE	GEOM	Poisson	ZIP	NB	DIRE
Parameter Estimation	$\widehat{\theta} = 0.0505$	$\widehat{\theta} = 0.0233$	$\widehat{\lambda} = 44.3934$	$\widehat{\lambda} = 33.0714$	$\widehat{P} = 20.0524$	$\widehat{\theta} = 0.1010$
	$\beta = 2.1375$	–	–	$\beta = 13.1254$	$\widehat{r} = 2.2604$	$\widehat{\omega} = 2.1995$
	$\alpha = 4.5337$	–	–	–	–	–
logL	100.2732	123.4167	114.3303	104.1212	100.1101	**98.8648**
AIC	206.5463	248.6733	230.6606	146.4515	136.3204	**127.7297**
BIC	209.8790	250.0059	232.5497	144.2376	134.4175	**134.0623**
HQIC	207.0282	249.1551	231.5319	152.2207	142.3116	**130.2115**
AICc	207.8790	248.0944	231.5928	146.4707	136.2605	**128.1507**
MSE	0.406452	0.489645	0.589645	0.177434	0.166313	**0.00277**

transmission pattern compared to discrete traditional models. In the same measure, its applicability to the failure time data shows its potential in the modeling of systems that have variable risk structure, and thus offers better opportunity in the engineering and industrial environment when making better decisions in such settings. As an example, the data on coronavirus disease 2019 (COVID-19) daily cases (the first data set) demonstrates how the DIRE distribution can provide an accurate model of discrete data with a changing hazard structure, which chronicles both the period of escalation of the outbreak and the period of its containment much better than existing

Table 8. k-Nearest Neighbors (K-NN), and Goodness-of-Fit (GOF) criteria for the first dataset.

Distribution	DWE	GEOM	Poisson	ZIP	NB	DIRE
Parameter Estimation	$\widehat{\theta} = 0.0405$	$\widehat{\theta} = 0.1233$	$\widehat{\lambda} = 43.3934$	$\widehat{\lambda} = 23.0604$	$\widehat{P} = 10.0413$	$\widehat{\theta} = 1.1010$
	$\beta = 1.1375$	–	–	$\beta = 43.0154$	$\widehat{r} = 4.1503$	$\widehat{\omega} = 3.1995$
	$\alpha = 3.5337$	–	–	–	–	–
logL	110.2732	113.4167	124.3303	114.0202	90.0101	87.8648
AIC	216.5463	238.6733	240.6606	140.3514	130.2103	117.7297
BIC	219.8790	240.0059	242.5497	124.1276	124.3075	124.0623
HQIC	217.0282	239.1551	241.5319	152.1206	122.2016	120.2115
AICc	217.8790	238.0944	241.5928	126.4505	126.1604	118.1507
MSE	0.506452	0.589645	0.689645	0.007434	0.006313	0.00167

classical models [39].

Correspondingly, when failure-time data are involved, the fact that DIRE distribution may be used to model non-monotonic hazard rates would permit it to be used to model more faithfully at least some aging components or wear-out failure processes. With this implementation of real-world application, the paper will be able to draw the line between the hypothetical growth and the practical usefulness of the model-making the latter all the more interpretable and substantial [40].

8. Future Work

The critical research directions that may be considered for future research on the Discrete Inverse Rayleigh Exponential (DIRE) distribution to take into consideration in supporting the theoretical and application-based expansion of the knowledge domain are: One such direction includes the development of a multivariate version of the test which will allow to study joint distribution properties and dependencies between several variables at once. Similar application-based works may explore the efficacy of the DIRE distribution in other contexts like reliability modelling in telecommunication systems, medical survival rates, and failure time data in engineering. Other related discrete distributions with which the DIRE distribution could be compared include the negative binomial distribution and the zero-inflated Poisson distribution and are: More comparisons between the DIRE distribution and other related discrete distributions would provide helpful information on goodness of fit, flexibility and stability of the parameters of this distribution.

In addition, the range and stability tests will be crucial for evaluating possible distribution behavior under the condition of varying data and potential model mis- specification. It is also added that fine-tuning of the parameter estimation is possible by enhancing the utilization of posterior predictive computation with advanced machine learning efficiency such as neural networks and ensemble. Possible improvements in the design of the TMB software packages would include the availability and enhanced usability of parameter estimation supported by tools for depiction of the simulation results and diagnostic evaluation for more extensive incorporation in academic analysis and applier statistics.

Lastly, further research on the study of moments, the large N limit, and relationships between the DIRE distribution and other discrete frameworks will enhance the statistical support of the model at large. The studies collectively would help improve the literature review and understanding of DIRE distribution and extend it across sciences and engineering disciplines.

9. Conclusion

In this research work, we attempt to discretize a continuous set of distributions that gives the discrete Inverse Rayleigh-G (DIR-G) family of distributions. This family is enriched with the new two-parameter discrete Inverse Rayleigh Exponential (DIRE) distribution. Mathematically, the PMF and HRF of DIRE distribution conform well and display versatile characteristics that unveil the severity nature inherent in real-world data. Secondly, we investigate the statistical properties of the DIRE distribution. Estimation of the parameters of the DIRE distribution was done using the maximum likelihood estimation method, MLE and the K-Nearest neighbors, K-NN. The comparison of MLE and K-NN has demonstrated that the K-NN recognition method has a higher accuracy of the parameters of the sample. To complete this comparison, we ran simulations with three different sample sizes and with a range of parameter values for the DIRE distribution. In addition, two actual datasets were used to compare the efficacy of the DIRE distribution with other frequently used distributions. The results, shown in the figures, prove the versatility and relevance of the DIRE distribution since the parameters of the distribution can be easily adjusted to a wide range of numeric values. In totality, the results indicate that the DIRE distribution might be extremely flexible and has the potential to capture data characteristics in the real world effectively.

Acknowledgment

The authors are very grateful to Dr. Mushtaq Karam, an Assistant professor at the University of Karbala, for providing access that allowed for more accurate data collection and improved the quality of this work.

Data Availability

The datasets supporting the conclusions of this article are included in the article.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

REFERENCES

1. Eliwa, M., Alhussain, Z., El-Morshedy, M. (2020). Discrete Gompertz-G family of distributions for over-and under-dispersed data with properties, estimation, and applications. *Mathematics*, **8**, 358.
2. Steutel, F.W., Van Harn, K. (2003). *Infinite Divisibility of Probability Distributions on the Real Line*. CRC Press: Boca Raton, FL, USA.
3. Yousof, H.M., Majumder, M., Jahanshahi, S., Masoom Ali, M., Hamedani, G. (2018). A new Weibull class of distributions: Theory, characterizations and applications. *J. Stat. Res. Iran JSRI*, **15**, 45–82.
4. Hussain, A. S., Saadoon, N. Q., Abdulghafour, A. S., Abduirazzaq, N. T. (2025). Parameters estimation of a suggested model non-homogeneous poisson process: A comparison of conventional and artificially intelligent approaches. *AIP Conference Proceedings*, **3282**(1), 020023.
5. Aboraya, M., M. Yousof, H., Hamedani, G., Ibrahim, M. (2020). A new family of discrete distributions with mathematical properties, characterizations, Bayesian and non-Bayesian estimation methods. *Mathematics*, **8**, 1648.
6. Ibrahim, M., Ali, M.M., Yousof, H.M. (2021). The discrete analogue of the Weibull G family: Properties, different applications, Bayesian and non-Bayesian estimation methods. *Ann. Data Sci.*, **10**, 1069–1106.

7. Hussain, A. S., Mahmood, K. B., Ibrahim, I. M., Jameel, A. F., Nawaz, S., Tashtoush, M. A. (2025). Parameters Estimation of the Gompertz-Makeham Process in Non-Homogeneous Poisson Processes: Using Modified Maximum Likelihood Estimation and Artificial Intelligence Methods.
8. Alzaatreh, A., Ghosh, I. (2015). On the Weibull-X family of distributions. *J. Stat. Theory Appl.*, **14**, 169–183.
9. Mohsin, S., Salim, B. W., Saleem, A. N. (2024). DarkNet Traffic Recognition Using Meta-Learning. *2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 95–99.
10. Hussein, H. R., Hussein, S. R., Hussain, A. S., Tashtoush, M. A. (2025). Estimating Parameters of Software Reliability Growth Models Using Artificial Neural Networks Optimized by the Artificial Bee Colony Algorithm Based on a Novel NHPP. *Mathematical Modelling of Engineering Problems*, **12**(1).
11. Elbatal, I., Alotaibi, N., Almetwally, E.M., Alyami, S.A., Elgarhy, M. (2022). On Odd Perks-G Class of Distributions: Properties, Regression Model, Discretization, Bayesian and Non-Bayesian Estimation, and Applications. *Symmetry*, **14**, 883.
12. Eliwa, M.S., El-Morshedy, M., Yousof, H.M. (2022). A Discrete Exponential Generalized-G Family of Distributions: Properties with Bayesian and Non-Bayesian Estimators to Model Medical, Engineering and Agriculture Data. *Mathematics*, **10**, 3348.
13. El-Morshedy, M., Eliwa, M., Tyagi, A. (2022). A discrete analogue of odd Weibull-G family of distributions: Properties, classical and Bayesian estimation with applications to count data. *J. Appl. Stat.*, **49**, 2928–2952.
14. Lubis, A. R., Lubis, M. (2020). Optimization of distance formula in K-Nearest Neighbor method. *Bulletin of Electrical Engineering and Informatics*, **9**(1), 326–338.
15. Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, **11**(1), 113.
16. Zhang, T. (2024). Exact MLE for Generalized Linear Mixed Models. *arXiv preprint arXiv:2410.08492*.
17. Mageed, I. A., Zhang, Q. (2023). Formalism of the Rényian Maximum Entropy (RMF) of the Stable M/G/1 queue with Geometric Mean (GeoM) and Shifted Geometric Mean (SGeoM) Constraints with Potential GeoM Applications to Wireless Sensor Networks (WSNs). *electronic Journal of Computer Science and Information Technology*, **9**(1), 31–40.
18. Hussain, A. S., Oraibi, Y. A., Sulaiman, M. S., Abdulghafour, A. S. (2024). Parameters Estimation of a Proposed Non-Homogeneous Poisson Process and Estimation of the Reliability Function Using the Gompertz Process: A Comparative Analysis of Artificially Intelligent and Traditional Methods: nonhomogenous poisson process with intelligent. *Iraqi Journal for Computer Science and Mathematics*, **5**(2), 36–47.
19. 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.
20. Siepe, B. S., Bartoš, F., Morris, T. P., Boulesteix, A. L., Heck, D. W., Pawel, S. (2024). Simulation studies for methodological research in psychology: A standardized template for planning, preregistration, and reporting. *Psychological methods*.
21. Almetwally, E.M., Abdo, D.A., Hafez, E., Jawa, T.M., Sayed-Ahmed, N., Almongy, H.M. (2022). The new discrete distribution with application to COVID-19 Data. *Results Phys.*, **32**, 104987.
22. SH, A., Fatah, K. S., Sulaiman, M. S. (2023). Estimating the Rate of Occurrence of Exponential Process Using Intelligence and Classical Methods with Application. *Palestine Journal of Mathematics*, **12**.
23. Nassar, M., Kumar, D., Dey, S., Cordeiro, G.M., Afify, A.Z. (2019). The Marshall–Olkin alpha power family of distributions with applications. *J. Comput. Appl. Math.*, **351**, 41–53.
24. Gacula, M., Jr., Kubala, J. (1975). Statistical models for shelf life failures. *J. Food Sci.*, **40**, 404–409.
25. Nassar, M., Kumar, D., Dey, S., Cordeiro, G.M., Afify, A.Z. (2019). The Marshall–Olkin alpha power family of distributions with applications. *J. Comput. Appl. Math.*, **351**, 41–53.
26. Almetwally, E.M., Abdo, D.A., Hafez, E., Jawa, T.M., Sayed-Ahmed, N., Almongy, H.M. (2022). The new discrete distribution with application to COVID-19 Data. *Results Phys.*, **32**, 104987.
27. Birnbaum, Z.W., Saunders, S.C. (1969). Estimation for a family of life distributions with applications to fatigue. *J. Appl. Probab.*, **6**, 328–347.
28. Weisstein, E. W. (2002). Binomial distribution. *MathWorld*. Available: https://mathworld.wolfram.com/.
29. Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**(1), 1–4.
30. Chupradit, S., Tashtoush, M., Ali, M., AL-Muttar, M., Sutarto, D., Chaudhary, P., Mahmudiono, T., Dwijendra, N., Alkhayyat, A. (2022). A Multi-Objective Mathematical Model for the Population-Base Transportation Network Planning. *Industrial Engineering & Management Systems*, **21**(2), 322–331.
31. Shirawia, N., Kherd, A., Bamsaoud, S., Tashtoush, M., Jassar, A., Az-Zo'bi, E. (2024). Dejdumrong Collocation Approach and Operational Matrix for a Class of Second-Order Delay IVPs: Error Analysis and Applications. *WSEAS Transactions on Mathematics*, **23**, 467–479.
32. Ibrahim, I., Taha, W., Dawi, M., Jameel, A., Tashtoush, M., Az-Zo'bi, E. (2024). Various Closed-Form Solitonic Wave Solutions of Conformable Higher-Dimensional Fokas Model in Fluids and Plasma Physics. *Iraqi Journal For Computer Science and Mathematics*, **5**(3), 401–417.
33. Az-Zo'bi, E. (2014). An approximate analytic solution for isentropic flow by an inviscid gas model. *Archives of Mechanics*, **66**(3), 203–212.
34. Az-Zo'bi, E. (2013). Construction of solutions for mixed hyperbolic elliptic Riemann initial value system of conservation laws. *Applied Mathematical Modelling*, **37**(8), 6018–6024. https://doi.org/10.1016/j.apm.2012.12.006
35. Az-Zo'bi, E. (2014). On the reduced differential transform method and its application to the generalized Burgers-Huxley equation. *Applied Mathematical Sciences*, **8**, 8823–8831. https://doi.org/10.12988/ams.2014.410835
36. Az-Zo'bi, E. (2018). A reliable analytic study for higher-dimensional telegraph equation. *The Journal of Mathematics and Computer Science*, **18**(4), 423–429. https://doi.org/10.22436/jmcs.018.04.04
37. Az-Zo'bi, E. (2019). Exact analytic solutions for nonlinear diffusion equations via generalized residual power series method. *Int. J. Math. Comput. Sci*, **14**(1), 69–78.
38. Az-Zo'bi, E., Kallekh, A., Rahman, R., Akinyemi, L. Bekir, A., Ahmad, H., Tashtoush, M., Mahariq, I. (2024). Novel topological, non-topological, and more solitons of the generalized cubic p-system describing isothermal flux. *Optical and Quantum Electronics*,

56(1), 1–16.
39. Hussain, A., Sulaiman, M., Hussein, S., Az-Zo'bi, E. Tashtoush, M. (2025). Advanced Parameter Estimation for the Gompertz-Makeham Process: A Comparative Study of MMLE, PSO, CS, and Bayesian Methods. *Statistics, Optimization & Information Computing*, **13**(6), 2316–2338.
40. Hussain, A., Pati, K., Atiyah, A., Tashtoush, M. (2025). Rate of Occurrence Estimation in Geometric Processes with Maxwell Distribution: A Comparative Study between Artificial Intelligence and Classical Methods. *International Journal of Advances in Soft Computing and Its Applications*, **17**(1), 1–15.