EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
2025, Vol. 18, Issue 2, Article Number 6144
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

A Novel Discrete Probability Distribution with Theoretical and Inferential Insights: Cutting-Edge Approaches to Sustainable Dispersion Data Modeling

Mohamed S. Eliwa[1,*], Weed E. Alghanem[1], Manahel A. Alenizi[1]

[1] *Department of Statistics and Operations Research, College of Science, Qassim University, Saudi Arabia*

Abstract. As real-world data becomes increasingly complex, there is a growing need for advanced probability models to support sustainable discrete data analysis. This study presents a novel and flexible extension of the discrete Gompertz distribution, developed within the framework of the half-logistic model. Named the discrete Gompertz half-logistic (DGzHLo) model, this new formulation enhances the adaptability of existing discrete distributions to better capture intricate data patterns. To understand its theoretical foundation, key mathematical and statistical properties are examined, including the probability mass function, cumulative distribution function, reliability function, and hazard rate function. Measures such as dispersion, skewness, and kurtosis offer insights into the model's behavior, while entropy and order statistics further reveal its structural characteristics. The model's parameters are estimated using the maximum likelihood estimation method, and a thorough simulation study assesses the accuracy and efficiency of the estimators across various sample sizes. To illustrate its practical relevance, the model is applied to three real-world datasets, demonstrating its superior flexibility and robustness in capturing complex data structures compared to existing models. These findings underscore the DGzHLo model's effectiveness in advancing discrete data modeling and analysis.

2020 Mathematics Subject Classifications: 62E15, 62F10

Key Words and Phrases: Statistical model, Order statistics, Reliability theory, Estimation, Simulation, Data analysis

1. Introduction

Sustainable dispersion data modeling has become increasingly essential in response to the exponential surge in real-world data, driven by technological advancements and digital innovation. The growing complexity of data structures presents significant challenges, as conventional probability models often fail to capture the intricate dependencies,

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v18i2.6144

Email address: m.eliwa@qu.edu.sa (M. S. Eliwa)

stochastic variations, and nonlinear trends observed across diverse fields such as engineering, epidemiology, finance, and actuarial sciences. This limitation underscores the need for advanced probabilistic frameworks that provide greater flexibility and adaptability for robust data analysis. Among the well-explored models in statistical and applied probability research is the Gompertz (Gz) distribution a continuous probability distribution that extends the exponential model by introducing an additional shape parameter. This modification enhances its ability to model a wide range of real-world phenomena, particularly those characterized by dynamic hazard rates. Due to its structural advantages, the Gz distribution has been extensively applied across scientific and practical domains, particularly in modeling lifetime and failure rates. In demography and actuarial science, it is frequently employed to describe adult lifespan distributions, offering critical insights into mortality patterns and life expectancy. Similarly, in biology and gerontology, researchers utilize the Gz distribution for survival analysis, especially in aging studies and biological lifespan modeling. Formally, a random variable T is said to follow the Gz model with a shape parameter $\varpi > 0$ and scale parameter $\tau > 0$ if its cumulative distribution function (CDF) is given by

$$L(t; \varpi, \tau) = 1 - e^{-\frac{\varpi}{\tau}(e^{\tau t}-1)}; \quad t > 0. \tag{1}$$

This formulation underscores the versatility of the Gz distribution in modeling diverse data patterns, particularly in contexts characterized by increasing failure rates and non-constant hazard functions. Beyond its conventional applications, the Gz distribution has been increasingly utilized in computational sciences to model software failure rates, contributing to software reliability assessment. For a comprehensive exploration of its applications, see Cordeiro et al. [5], Roozegar et al.[23], Mazucheli et al. [8], Eliwa et al. [19], Bantan et al. [22], Wang and Guo [10], Alizadeh et al. [13], Hassan and Abdelghaffar [26], among others.

Building upon the fundamental structure of the Gz distribution, Alizadeh et al. [12] introduced the Gz-G family using the methodological framework developed by Alzaatreh et al. [1]. The latter proposed the transformed-transformer family, a general approach for constructing new probability distributions with enhanced adaptability. Accordingly, a random variable Y follows the Gz-G family if its CDF is given by

$$\Theta(y; \varpi, \tau, \mathbf{\Omega}) = 1 - e^{-\frac{\varpi}{\tau}\left\{[1-G(y;\mathbf{\Omega})]^{-\tau}-1\right\}}; \quad y > 0, \tag{2}$$

where $\mathbf{\Omega}$ is a vector of parameters $(1 \times m; m = 1, 2, 3, ...)$, and $G(y; \mathbf{\Omega})$ is the baseline CDF. The reliability function (RF) of the Gz-G family can be formulated as

$$\overline{\Theta}(y; \varpi, \tau, \mathbf{\Omega}) = e^{-\frac{\varpi}{\tau}\left\{[1-G(y;\mathbf{\Omega})]^{-\tau}-1\right\}}; \quad y > 0. \tag{3}$$

The probability density function (PDF) corresponding to Equation (2) can be written as

$$\psi(y; \varpi, \tau, \mathbf{\Omega}) = \varpi g(y; \mathbf{\Omega}) \left[\overline{G}(y; \mathbf{\Omega})\right]^{-(\tau+1)} e^{-\frac{\varpi}{\tau}\left\{[1-G(y;\mathbf{\Omega})]^{-\tau}-1\right\}}; \quad y > 0, \tag{4}$$

where $g(y; \mathbf{\Omega})$ is the baseline PDF. Several researchers have built upon the framework introduced by Alzaatreh et al. [1] to develop both univariate and bivariate probability distribution families. Significant contributions in this area include studies by El-Morshedy and Eliwa [15], Alizadeh et al. [11], Abdelall et al. [29], Alsolmi [17], Atchade et al. [18], among others. These works have played a crucial role in advancing probability modeling by introducing more flexible and adaptable distributions.

In recent years, there has been an increasing focus on discretizing continuous probability distributions, particularly for applications where measuring lifetimes on a continuous scale is impractical. This issue commonly arises in areas such as reliability analysis and quality control, where product or system lifetimes are naturally recorded in discrete units like days, cycles, or events. To address this, various discrete probability models have been proposed by researchers such as Roy [25], Afify et al. [30], Abd EL-Hady et al. [3], Eliwa et al. [20], Gómez-Déniz and Calderín-Ojeda [4], Yousof et al. [6], Nakagawa and Osaki [27], Ibrahim et al. [16], Hamdi et al. [28], among others. Despite these advancements, there remains considerable scope for developing new discrete probability distributions tailored to specific data structures and real-world scenarios. This study introduces a new discrete probability model, termed the discrete Gompertz half-logistic (DGzHLo) model, derived from the discrete version of the Gz-G family originally formulated by Eliwa et al. [21]. The CDF of the discrete version of the Gz-G (DGz-G) family is given by

$$H_Z(z;\gamma,\tau,\mathbf{\Omega}) = 1-\gamma^{\frac{1}{\tau}\left\{[1-G(z+1;\mathbf{\Omega})]^{-\tau}-1\right\}};\quad z\in\mathbb{N}_0, \tag{5}$$

where $\gamma = e^{-\varpi}$, $0<\gamma<1$, $\tau>0$ and $\mathbb{N}_0=\{0,1,2,3,...\}$. Incorporating the DGz-G structure with the half-logistic (HLo) framework enhances the model's flexibility in capturing discrete lifetime data, where the CDF of the HLo model is given by

$$G(s;\theta) = \frac{1-e^{-\theta s}}{1+e^{-\theta s}};\quad s\geq 0, \tag{6}$$

where $\theta>0$. The motivation for introducing the DGzHLo model stems from the need to:

- Generating models with varying skewness characteristics (negative, positive, or symmetric) to capture diverse data distributions in sustainability studies.
- Accommodating increasing hazard rate functions, particularly in engineering and medical contexts, to improve predictions and decision-making in sustainable development.
- Modeling over-dispersed, equi-dispersed, and under-dispersed data to provide more accurate insights into environmental and economic sustainability metrics.
- Accounting for different levels of kurtosis, including leptokurtic and platykurtic distributions, to better represent the variability observed in sustainability data.
- Handling zero-inflation and non-inflation data, particularly under extreme observations, to improve the robustness of models in the context of rare but impactful sustainability events.

This paper is organized as follows: Section 2 introduces the DGzHLo distribution, outlining its formulation and key properties. Section 3 examines its statistical and reliability features, while Section 4 discusses parameter estimation using the maximum likelihood estimation method. A simulation study assessing the performance of the estimators across various sample sizes is presented in Section 5. Section 6 demonstrates the practical application of the DGzHLo model through the analysis of three real-world datasets, comparing its performance to other existing distributions. Finally, Section 7 provides a summary of the main findings and suggests potential avenues for future research.

2. Mathematical Framework and Visual Depiction

Recall, Equations (5) and (6), the CDF and RF of the DGzHLo can be formulated as

$$F_Z(z;\gamma,\tau,\theta) = 1 - \gamma^{\frac{1}{\tau}\left\{\left[1-\frac{1-e^{-\theta(z+1)}}{1+e^{-\theta(z+1)}}\right]^{-\tau}-1\right\}}; \quad z \in \mathbb{N}_0, \tag{7}$$

$$\overline{F}_Z(z;\gamma,\tau,\theta) = \gamma^{\frac{1}{\tau}\left\{\left[1-\frac{1-e^{-\theta(z+1)}}{1+e^{-\theta(z+1)}}\right]^{-\tau}-1\right\}}; \quad z \in \mathbb{N}_0, \tag{8}$$

respectively. The probability mass function (PMF) corresponding to Equation (7) can be expressed as

$$\begin{aligned} f_z(z;\gamma,\tau,\theta) &= \overline{F}(z;\gamma,\tau,\theta) - \overline{F}(z+1;\gamma,\tau,\theta) \\ &= \gamma^{-\frac{1}{\tau}}\left[\gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}} - \gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta(z+1)}}{1+e^{-\theta(z+1)}}\right]^{-\tau}}\right]; \; z \in \mathbb{N}_0. \end{aligned} \tag{9}$$

For further details on the survival discretization approach, refer to Roy [25]. The hazard rate function (HRF) can be formulated as

$$h(z;\gamma,\tau,\theta) = 1 - \gamma^{\frac{1}{\tau}\left\{\left[1-\frac{1-e^{-\theta(z+1)}}{1+e^{-\theta(z+1)}}\right]^{-\tau}-\left[1-\frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}\right\}}; \quad z \in \mathbb{N}_0, \tag{10}$$

where $h(z;\gamma,\tau,\mathbf{\Omega}) = \frac{f_z(z;\gamma,\tau,\theta)}{\overline{F}_Z(z-1;\gamma,\tau,\theta)}$. The PMF in Equation (9) is log-concave for all parameter values of the model. Specifically, the ratio $\frac{f(z+1;\gamma,\tau,\theta)}{f(z;\gamma,\tau,\theta)}$ is a decreasing function of z for all parameter values. As a result, the distribution is strongly unimodal, possesses all of its moments, and the hazard rate functions (HRFs) are increasing. Figures 1 and 2 illustrate the PMF and HRF of the DGzHLo distribution for different values of the model parameters. As shown in Figures 1 and 2, the PMF of the DGzHLo distribution proves to be a valuable statistical tool for analyzing unimodal data with varying degrees of skewness. This characteristic makes the PMF particularly beneficial for sustainable dispersion data, where it can accommodate different forms of data distributions, ensuring accurate modeling of real-world phenomena across multiple domains. Additionally, the HRF of the DGzHLo distribution exhibits an increasing trend, enhancing its applicability in diverse

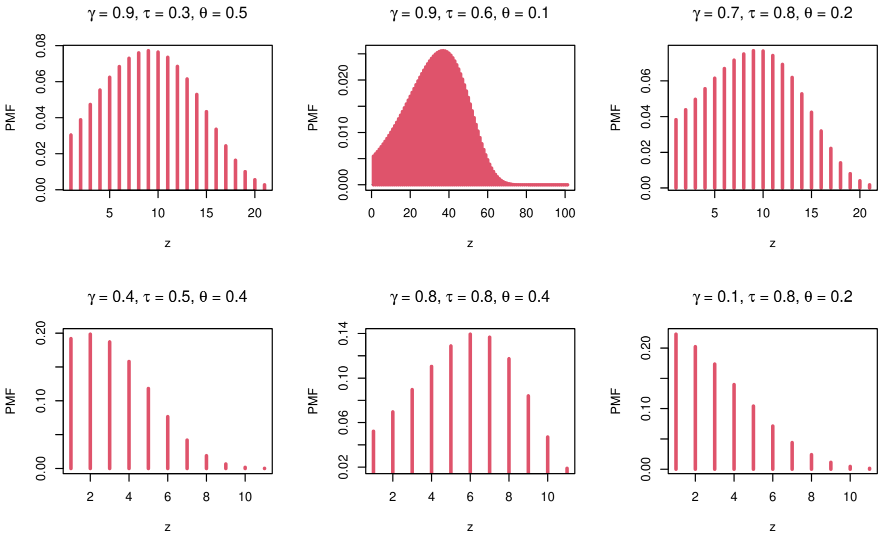

Figure 1: The PMF of the DGzHLo distribution.

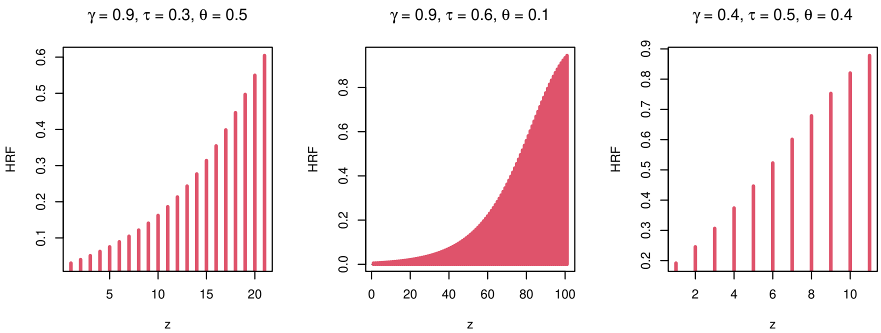

Figure 2: The HRF of the DGzHLo distribution.

fields, including engineering and medicine. The increasing HRF is especially useful for modeling lifetime data, reliability analysis, and risk assessment, where the hazard rate tends to increase over time such as in the analysis of engineering systems, medical conditions, or treatment effectiveness. Therefore, the DGzHLo distribution provides a robust and adaptable framework for analyzing and interpreting data in both sustainability and health-related fields.

3. Different Statistical Properties

3.1. Moments, cumulants, and associated measures

Moments and cumulants are essential statistical tools that provide valuable insights into the shape, variability, and relationships within a probability distribution. Moments, including the mean, variance, and index of dispersion, offer information about the central tendency and spread of the distribution. Cumulants, in contrast, capture higher-order features such as skewness and kurtosis, providing a deeper understanding of the distribution's characteristics. These measures are integral to statistical modeling, reliability analysis, and parameter estimation. In this section, we will derive the moments, cumulants, and other related measures for the DGzHLo model. These quantities are crucial for gaining a comprehensive understanding of the model's key properties and its practical applications. Let Z be a non-negative random variable following the DGzHLo distribution, denoted as $Z \sim \text{DGzHLo}(z; \gamma, \tau, \theta)$. The rth moment of Z can be expressed as

$$\begin{aligned} \mu'_r &= E(Z^r) = \sum_{z=0}^{\infty} z^r f_z(z; \gamma, \tau, \theta) \\ &= \sum_{z=1}^{\infty} \left[z^r - (z-1)^r\right] \overline{F}_Z(z-1; \gamma, \tau, \theta) \\ &= \gamma^{-\frac{1}{\tau}} \sum_{z=1}^{\infty} \left[z^r - (z-1)^r\right] \gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}}. \end{aligned} \tag{11}$$

Using Equation (11), the mean (μ'_1) and variance, say $\text{Var}(Z)$, can be respectively listed as

$$\mu'_1 = \gamma^{-\frac{1}{\tau}} \sum_{z=1}^{\infty} \gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}},$$

$$\text{Var}(Z) = \gamma^{-\frac{1}{\tau}} \sum_{z=1}^{\infty} (2z-1)\, \gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}} - (\mu'_1)^2. \tag{12}$$

The dispersion index (DsI) is the ratio of variance to the mean, serving as a tool to assess whether a model is suitable for over-dispersed or under-dispersed data. It is widely used in ecology as a standard metric to evaluate patterns of clustering (over-dispersion) or repulsion (under-dispersion). When $\text{DsI} > 1$, the distribution is considered over-dispersed, whereas $\text{DsI} < 1$ indicates under-dispersion. The DsI for the DGzHLo model is expressed as

$$\text{DsI}(Z) = \frac{\sum_{z=1}^{\infty} (2z-1)\, \gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}}}{\sum_{z=1}^{\infty} \gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}}} - \sum_{z=1}^{\infty} \gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}}. \tag{13}$$

Alternatively, the moment generating function (MGF) can be expressed as

$$\begin{aligned} M_Z(t) &= \sum_{z=0}^{\infty} e^{zt} f_z(z; \gamma, \tau, \theta) \\ &= \gamma^{-\frac{1}{\tau}} \left[\sum_{z=0}^{\infty} e^{zt} \gamma^{\Lambda(z;\tau)} - \sum_{z=0}^{\infty} e^{zt} \gamma^{\Lambda(z+1;\tau)}\right] \\ &= \gamma^{-\frac{1}{\tau}} [\Big(\gamma^{\Lambda(0;\tau)} + e^{t}\gamma^{\Lambda(1;\tau)} + e^{2t}\gamma^{\Lambda(2;\tau)} + e^{3t}\gamma^{\Lambda(3;\tau)} + \ldots\Big) \end{aligned}$$

$$- \left(\gamma^{\Lambda(1;\tau)} + e^{t}\gamma^{\Lambda(2;\tau)} + e^{2t}\gamma^{\Lambda(3;\tau)} + e^{3t}\gamma^{\Lambda(4;\tau)} + ...\right)]$$

$$= \gamma^{-\frac{1}{\tau}} \left[1 + \sum_{z=1}^{\infty} \left(e^{zt} - e^{(z-1)t}\right) \gamma^{\Lambda(z;\tau)}\right], \tag{14}$$

where

$$\Lambda(z;\tau) = \frac{1}{\tau}\left[1 - \frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}.$$

The first four derivatives of Equation (14), with respect to t at $t = 0$, yield the first four moments about the origin, i.e.

$$E(Z^r) = \frac{d^r}{dt^r} M_Z(t)|_{t=0}.$$

Moreover, utilizing Equation (11) or (14), the skewness, say $\text{Sk}(Z)$ and kurtosis, say $\text{Ku}(Z)$, can be reported as

$$\text{Sk}(Z) = (\mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3)/(\text{Var}(Z))^{3/2},$$

$$\text{Ku}(Z) = (\mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4)/(\text{Var}(Z))^2,$$

respectively. In probability theory, cumulants, denoted as k_n, are a set of quantities that offer an alternative to the moments of a probability model. In certain cases, working with cumulants can simplify the theoretical treatment of problems compared to using moments. The cumulant generating function (CGF) is defined as the logarithm of the MGF. Therefore, the cumulants k_n can be derived from the moments as follows

$$k_n = \frac{d^n}{dt^n} \log M_Z(t)|_{t=0};\ n = 1, 2, 3, ... \tag{15}$$

Additionally, the cumulants are related to the moments through the following recursive formula

$$k_n = \mu_n' - \sum_{m=1}^{n-1} \binom{n-1}{m-1} \mu_{n-m}' k_m. \tag{16}$$

The first cumulant corresponds to the mean, the second cumulant represents the variance, and the third cumulant is related to the third central moment. However, starting from the fourth cumulant, the direct relationship with central moments no longer holds, and higher-order cumulants no longer directly correspond to the central moments. For the DGzHLo distribution, the rth moment cannot be expressed in a closed form. Therefore, we utilize Maple software to investigate and compute some of its statistical properties. Table 1 presents descriptive statistics derived from the DGzHLo model for different parameter values of θ and γ with $\tau = 0.5$, highlighting that the DGzHLo model provides a strong framework for modeling sustainability data with varying kurtosis, skewness, and dispersion characteristics. This makes it an invaluable tool across a wide range of fields, from engineering to social sciences, where sustainability is a key focus.

Table 1. Some descriptive statistics using the DGzHLo model.

Measure	$\theta \downarrow \gamma \rightarrow$	0.1	0.3	0.5	0.7	0.9
Mean	0.5	0.7444	1.4266	2.2267	3.4741	6.5495
	1.5	0.0488	0.2107	0.4477	0.8444	1.8556
	2.0	0.0080	0.0801	0.2365	0.5210	1.2703
Var	0.5	0.8149	1.7661	3.0468	5.2333	10.719
	1.5	0.0465	0.1752	0.3374	0.6008	1.2373
	2.0	0.0079	0.0738	0.1859	0.3448	0.7183
DsI	0.5	1.0946	1.2379	1.3682	1.5063	1.6366
	1.5	0.9524	0.8316	0.7536	0.7114	0.6667
	2.0	0.9919	0.9206	0.7860	0.6617	0.5655
Sk	0.5	1.1576	0.8163	0.5891	0.3497	−0.0376
	1.5	4.1937	1.5999	0.9045	0.4929	0.0216
	2.0	11.0353	3.0963	1.3395	0.6264	0.0731
Ku	0.5	3.9593	3.1603	2.7533	2.4635	2.3412
	1.5	18.646	4.0924	2.8785	2.4486	2.3122
	2.0	122.778	10.6124	3.0893	2.4472	2.2957

The DGzHLo model proves to be highly versatile in analyzing a wide range of data types, including both negatively and positively skewed data, and symmetric data with varying levels of kurtosis. As seen in Table 1, this model can be adapted to handle data distributions with leptokurtic (high kurtosis) and platykurtic (low kurtosis) characteristics, making it an excellent tool for capturing the nuances of real-world data. One of the major advantages of the DGzHLo model lies in its ability to handle overdispersed and underdispersed data, which are common in many fields, including economics, engineering, and epidemiology. Overdispersion occurs when the variance exceeds the mean, while underdispersion is observed when the variance is smaller than the mean. The DGzHLo model's flexibility in capturing these features makes it particularly useful for data sets where traditional models, like the Poisson or negative binomial distributions, may fall short. Furthermore, the DGzHLo model can also accommodate equidispersed data, where the variance is approximately equal to the mean. This is crucial for applications in various scientific and engineering domains where data often follow such distributions, and accurate modeling of these types can lead to better decision-making and more robust predictions. In the context of sustainable development and engineering sustainability data modeling, the DGzHLo model offers a significant advantage. Sustainable data often exhibit complex patterns, with varying dispersion, skewness, and kurtosis, especially in long-term sustainability studies where factors influencing the data may change over time. The model's capacity to handle different dispersion levels and its application to a wide range of data types ensures that it can provide insights into the sustainability of processes, resources, and systems.

3.2. Entropy

Entropy is a measure of the uncertainty or unpredictability associated with a random variable, denoted as Z. It quantifies the level of disorder or information content within the distribution of Z (see Rényi, [24]). In the context of sustainable dispersion data, entropy plays a vital role in assessing the variability and uncertainty in sustainability metrics, such as resource consumption, environmental impact, or social sustainability indicators. By capturing the degree of spread or concentration of data, entropy helps to quantify the uncertainty inherent in these measures, which is essential for making informed decisions about sustainable practices. The Rényi entropy of the discrete random variable can be expressed as

$$I_\eta(Z) = \frac{1}{1-\eta} \log \sum_{z=0}^{\infty} f_z^\eta(z; \gamma, \tau, \theta),$$

where $\eta \in]0, \infty[$ and $\eta \neq 1$. The Shannon entropy can be defined by $E\left[-\log f(Z; \gamma, \tau, \theta)\right]$. It is observed that the Shannon entropy can be calculated as a special case of the Rényi entropy when $\eta \to 1$. For the DGzHLo model, the $I_\eta(Z)$ can be formulated as

$$I_\eta(Z) = \frac{1}{1-\eta} \left\{ -\frac{\eta}{\tau} \log \gamma + \log \sum_{z=0}^{\infty} \left[\gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta z}}{1+e^{-\theta z}}\right]^{-\tau}} - \gamma^{\frac{1}{\tau}\left[1-\frac{1-e^{-\theta(z+1)}}{1+e^{-\theta(z+1)}}\right]^{-\tau}} \right]^\eta \right\}. \quad (17)$$

The DGzHLo model, with its ability to handle diverse dispersion characteristics, including overdispersion, underdispersion, and equidispersion, can be effectively applied to sustainability data. It provides a robust framework for modeling the uncertainty and variation in sustainability metrics, considering various levels of kurtosis and skewness. In fields such as environmental science, economics, and social policy, where sustainability data often exhibit complex dispersion patterns, the DGzHLo model offers a flexible and powerful tool for analyzing these uncertainties. Specifically, entropy in sustainable dispersion data can be used to understand how the probability distribution of a sustainability measure is spread out or concentrated. This can provide valuable insights into the stability or variability of the data, helping stakeholders make decisions that align with long-term sustainability goals. For example, entropy could be used to assess the unpredictability of environmental changes or the effectiveness of sustainable practices, guiding policymakers and businesses toward strategies that mitigate risk and uncertainty.

3.3. Mean time to failure, mean time between failure and availability

In the context of discrete random variables, the DGzHLo model offers a valuable framework for evaluating system performance based on reliability metrics such as mean time to failure (MTTF), mean time between failures (MTBF), and availability (Av). MTTF represents the average time until the first failure of a non-repairable system, and for the DGzHLo model, it can be estimated using the expected failure time derived from the distribution's parameters. MTBF, which applies to repairable systems, indicates the average time between two successive failures, and the DGzHLo model helps estimate this by analyzing failure rates across multiple cycles.

When applied to sustainable dispersion data, the DGzHLo model's capacity to handle varying dispersion characteristics including overdispersion, underdispersion, and equidispersion becomes particularly useful. For instance, it can help analyze and model the variability in the time between system failures or the time until a certain sustainability threshold is reached. This is crucial in sustainability assessments, where understanding the lifespan and reliability of systems (such as renewable energy infrastructures or sustainable production systems) is essential. The DGzHLo model, therefore, not only provides insights into the reliability and performance of such systems but also helps quantify the uncertainty and variation inherent in sustainability metrics, offering a robust tool for long-term planning and decision-making in sustainable practices. If $T \sim \text{DGzHLo}(t; \gamma_1, \tau_1, \theta_1)$, then the MTBF is given as

$$\text{MTBF}(T) = \frac{-t}{\ln(\gamma_1^{\frac{1}{\tau_1}\left\{\left[1-\frac{1-e^{-\theta_1(t+1)}}{1+e^{-\theta_1(t+1)}}\right]^{-\tau_1}-1\right\}})}; \; t > 0, \tag{18}$$

whereas if $T \sim \text{DGzHLo}(t; \gamma_2, \tau_2, \theta_2)$, then the MTTF can be listed as

$$\text{MTTF}(T) = \gamma_2^{-\frac{1}{\tau_2}} \sum_{t=1}^{\infty} \gamma_2^{\frac{1}{\tau_2}\left[1-\frac{1-e^{-\theta_2 t}}{1+e^{-\theta_2 t}}\right]^{-\tau_2}}; \; t > 0. \tag{19}$$

In the context of sustainability dispersion data, availability measures the proportion of time a system is operational, factoring in both its operational time and the downtime caused by repairs. For systems characterized by the DGzHLo distribution, availability can be assessed by analyzing the system's uptime and the time between failures, both of which are influenced by the model's parameters. This evaluation is particularly useful in sustainability assessments, where understanding system reliability and performance is essential for optimizing resource usage and ensuring long-term sustainability. The DGzHLo model provides a framework to quantify and manage these dynamics, offering valuable insights into system availability in sustainable practices, where

$$\text{Av}(T) = \frac{-1}{t}\left[\gamma_2^{-\frac{1}{\tau_2}} \sum_{z=1}^{\infty} \gamma_2^{\frac{1}{\tau_2}\left[1-\frac{1-e^{-\theta_2 z}}{1+e^{-\theta_2 z}}\right]^{-\tau_2}}\right]\left[\ln(\gamma_1^{\frac{1}{\tau_1}\left\{\left[1-\frac{1-e^{-\theta_1(t+1)}}{1+e^{-\theta_1(t+1)}}\right]^{-\tau_1}-1\right\}})\right]; \; t > 0.$$

When utilized within the DGzHLo framework, these metrics provide a comprehensive understanding of a system's reliability and performance, particularly in scenarios where failure times are discrete and repairs are necessary. Availability represents the probability that a component is operational at a given time and can be expressed as the ratio of MTTF to MTBF. This approach allows for a detailed analysis of system behavior, especially in systems that require maintenance and have discrete failure events.

3.4. Order statistics and L-moment statistics

Order statistics (OS) play a significant role in various areas of statistical theory and practice. Consider a random sample $Z_1, Z_2, ..., Z_n$ drawn from the DGzHLo$(z; \tau, \gamma, \theta)$ model, and let $Z_{1:n}, Z_{2:n}, ..., Z_{n:n}$ represent the corresponding order statistics. The CDF for the ith order statistic $Z_{i:n}$, for an integer value of z, can be expressed as follows

$$\begin{aligned} F_{i:n}(z; \gamma, \tau, \theta) &= \sum_{k=i}^{n} \binom{n}{k} [F_i(z; \gamma, \tau, \theta)]^k [1 - F_i(z; \gamma, \tau, \theta)]^{n-k} \\ &= \sum_{k=i}^{n} \sum_{j=0}^{n-k} \Delta_{(n,k)}^{j} [F_i(z; \gamma, \tau, \theta)]^{k+j} \\ &= \sum_{k=i}^{n} \sum_{j=0}^{n-k} \Delta_{(n,k)}^{j} \left[1 - \gamma^{\frac{1}{\tau}\left\{\left[1 - \frac{1-e^{-\theta(z+1)}}{1+e^{-\theta(z+1)}}\right]^{-\tau} - 1\right\}} \right]^{k+j}, \end{aligned} \quad (20)$$

where $\Delta_{(n,k)}^{j} = (-1)^j \binom{n}{k} \binom{n-k}{j}$. The corresponding PMF of the ith OS, as given by Equation (9), can be represented as

$$f_{i:n}(z; \gamma, \tau, \mathbf{\Omega}) = \sum_{k=i}^{n} \sum_{j=0}^{n-k} \sum_{m=0}^{k+j} \Delta_{(n,k)}^{(m,j)} f(z; \gamma^m, \tau, \theta), \quad (21)$$

where

$$\Delta_{(n,k)}^{(m,j)} = (-1)^{j+m} \binom{n}{k} \binom{n-k}{j} \binom{k+j}{m}.$$

Thus, the uth moment of $Z_{i:n}$ can be written as

$$\Psi_{i:n}^{u} = \mathbf{E}(Z_{i:n}^{u}) = \sum_{z=0}^{\infty} \sum_{k=i}^{n} \sum_{j=0}^{n-k} \sum_{m=0}^{k+j} \Delta_{(n,k)}^{(m,j)} z^u f(z; \gamma^m, \tau, \theta). \quad (22)$$

L-moments (LMs) are calculated as linear combinations of order statistics (OS). Introduced by Hosking and Wallis [9], LMs serve to summarize both theoretical distributions and observed data samples. LM statistics are widely used to compute sample statistics for data from specific regions or to assess the homogeneity or heterogeneity of proposed groupings of sites. Let $Z(i|n)$ be ith largest observations in sample of size n, then the LMs can be take the form

$$\lambda_r^* = \frac{1}{r} \sum_{s=0}^{r-1} (-1)^s \binom{r-1}{s} \mathbf{E}\left(Z_{r-s:r}\right). \quad (23)$$

From Equation (23), we get $\lambda_1^* = \mathbf{E}(Z_{1:1})$, $\lambda_2^* = \frac{1}{2}\mathbf{E}\left(Z_{2:2} + Z_{1:2}\right)$, $\lambda_3^* = \frac{1}{3}[\mathbf{E}\left(Z_{3:3} - Z_{2:3}\right) - \mathbf{E}\left(Z_{2:3} + Z_{1:3}\right)]$ and $\lambda_4^* = \frac{1}{4}\left\{\mathbf{E}\left[(Z_{4:4} - Z_{3:4}) + (Z_{2:4} - Z_{1:4})\right] - 2\mathbf{E}\left(Z_{3:4} - Z_{2:4}\right)\right\}$. Then, we can define some statistical measures such as LM of mean, LM coefficient of variation, LM coefficient of Sk and LM coefficient of ku in the form λ_1^*, $\frac{\lambda_2^*}{\lambda_1^*}$, $\frac{\lambda_3^*}{\lambda_2^*}$ and $\frac{\lambda_4^*}{\lambda_2^*}$, respectively. The implementation of L-moments in the DGzHLo model involves using linear combinations of order statistics to capture key features of the distribution, such as location, dispersion, skewness, and kurtosis. These moments are particularly useful for modeling discrete probability distributions, as they provide a robust way to characterize the distribution's shape, especially in the presence of heavy tails or skewed data. In the context of the DGzHLo model, L-moments are calculated from the order statistics of the data, which helps in estimating the model's parameters. By comparing the sample L-moments to the theoretical L-moments of the DGzHLo model, the parameters are adjusted to best fit the data. This method enhances the model's ability to accurately represent sustainability-related or other real-world data while maintaining robustness to extreme values.

4. Maximum Likelihood Estimation

In this section, we estimate the unknown parameters of the DGzHLo model using the maximum likelihood method (MLE). Suppose $Z_1, Z_2, ..., Z_n$ be a random sample from the DGzHLo model. Then, the log-likelihood function (L) can be listed as

$$L = -\frac{1}{\tau}\ln(\gamma) +_{i=1}^{n} \ln\left(\gamma^{\frac{1}{\tau}\left\{\left[1-\frac{1-e^{-\theta z_i}}{1+e^{-\theta z_i}}\right]^{-\tau}\right\}} - \gamma^{\frac{1}{\tau}\left\{\left[1-\frac{1-e^{-\theta (z_i+1)}}{1+e^{-\theta (z_i+1)}}\right]^{-\tau}\right\}}\right). \tag{24}$$

The MLEs of the parameters γ, τ and θ can be derived by solving the nonlinear likelihood equations obtained by differentiating (24) with respect to γ, τ and θ. The components of the score vector, $\mathbf{V}(\gamma,\tau,\theta) = (\frac{\partial L}{\partial \gamma}, \frac{\partial L}{\partial \tau}, \frac{\partial L}{\partial \theta})^T$ are

$$V_\gamma = \frac{-n}{\tau\gamma} + \frac{1}{\tau\gamma}{}_{i=1}^{n} \frac{g_2(z_i) - g_2(z_i+1)}{g_1(z_i)}, \tag{25}$$

$$V_\tau = \frac{-n\ln(\gamma)}{\tau^2} - \frac{\ln(\gamma)}{\tau^2}{}_{i=1}^{n} \frac{g_2(z_i)\left[\tau\ln(1-\frac{1-e^{-\theta z_i}}{1+e^{-\theta z_i}})+1\right] - g_2(z_i+1)\left[\tau\ln(1-\frac{1-e^{-\theta(z_i+1)}}{1+e^{-\theta(z_i+1)}})+1\right]}{g_1(z_i)} \tag{26}$$

and

$$V_\theta =_{i=1}^{n} \frac{g_2(z_i)\left[1-\frac{1-e^{-\theta z_i}}{1+e^{-\theta z_i}}\right]^{-1}[G(z_i;\theta)]_\theta - g_2(z_i+1)\left[1-\frac{1-e^{-\theta(z_i+1)}}{1+e^{-\theta(z_i+1)}}\right]^{-1}[G(z_i+1;\theta)]_\theta}{g_1(z_i)}, \tag{27}$$

where

$$[G(z_i;\theta)]_\theta = \frac{\partial}{\partial\theta}\left[\frac{1-e^{-\theta z_i}}{1+e^{-\theta z_i}}\right],$$

$$g_1(z_i) = \gamma^{\frac{1}{\tau}\left\{\left[1-\frac{1-e^{-\theta z_i}}{1+e^{-\theta z_i}}\right]^{-\tau}\right\}} - \gamma^{\frac{1}{\tau}\left\{\left[1-\frac{1-e^{-\theta (z_i+1)}}{1+e^{-\theta (z_i+1)}}\right]^{-\tau}\right\}},$$

and

$$g_2(z_i) = \gamma^{\frac{1}{\tau}\left\{\left[1-\frac{1-e^{-\theta z_i}}{1+e^{-\theta z_i}}\right]^{-\tau}\right\}} \left\{\left[1-\frac{1-e^{-\theta z_i}}{1+e^{-\theta z_i}}\right]^{-\tau}\right\}.$$

Setting the Equations (25-27) to zero and solving them gives the maximum likelihood estimates (MLEs) for the DGzHLo family parameters. Since these equations cannot be solved analytically, an iterative method such as the Newton-Raphson algorithm is needed to solve them numerically. In R, this can be implemented using packages like *stats* or *optim*, which provide functions for numerical optimization and solving such equations.

5. Simulation Results: Estimator Behavior

This section presents the results from simulations conducted to evaluate the performance of the estimators for the parameters of the DGzHLo model. Using synthetic data generated from the proposed model with known parameters, we assess how accurately the maximum likelihood estimates (MLEs) perform in terms of bias, variance, and efficiency. The study considers various sample sizes and different parameter values to examine the robustness and consistency of the estimators under different conditions. The evaluation incorporates additional criteria such as mean-squared errors (MSEs), mean relative error (MRE) and other relevant performance metrics to provide a comprehensive understanding of the MLEs' accuracy and reliability. These findings offer valuable insights into the practical applicability of the estimators for real-world data. The assessment is based on a thorough simulation study, which includes a range of scenarios to explore the behavior of the DGzHLo model's estimators.

(i) Generate 1000 samples of size $n = 20, 50, 150, 300, 500$ from three Scheme, Schema I: DGzHLo$(0.1, 0.5, 0.7)$, Schema II: DGzHLo$(0.3, 0.7, 0.3)$ and Schema III: DGzHLo$(0.5, 0.2, 0.9)$.

(ii) Compute the MLEs for the 1000 samples, say $\widehat{a}_j$ and $\widehat{b}_j$ for $j = 1, 2, ..., 1000$.

(iii) Compute the biase and MSEs, and MREs where

$$\text{bias}(\varpi) = \frac{1}{1000}\sum_{j=1}^{1000}\left(\widehat{\varpi_j} - \varpi_j\right),$$

$$\text{MSE}(\varpi) = \frac{1}{1000}\sum_{j=1}^{1000}\left(\widehat{\varpi_j} - \varpi_j\right)^2,$$

$$\text{and} \quad \text{MRE}(\varpi) = \frac{1}{1000}\sum_{j=1}^{1000}\left|\frac{\widehat{\varpi_j} - \varpi_j}{\varpi_j}\right|.$$

The empirical results are shown in Figures 3-5.

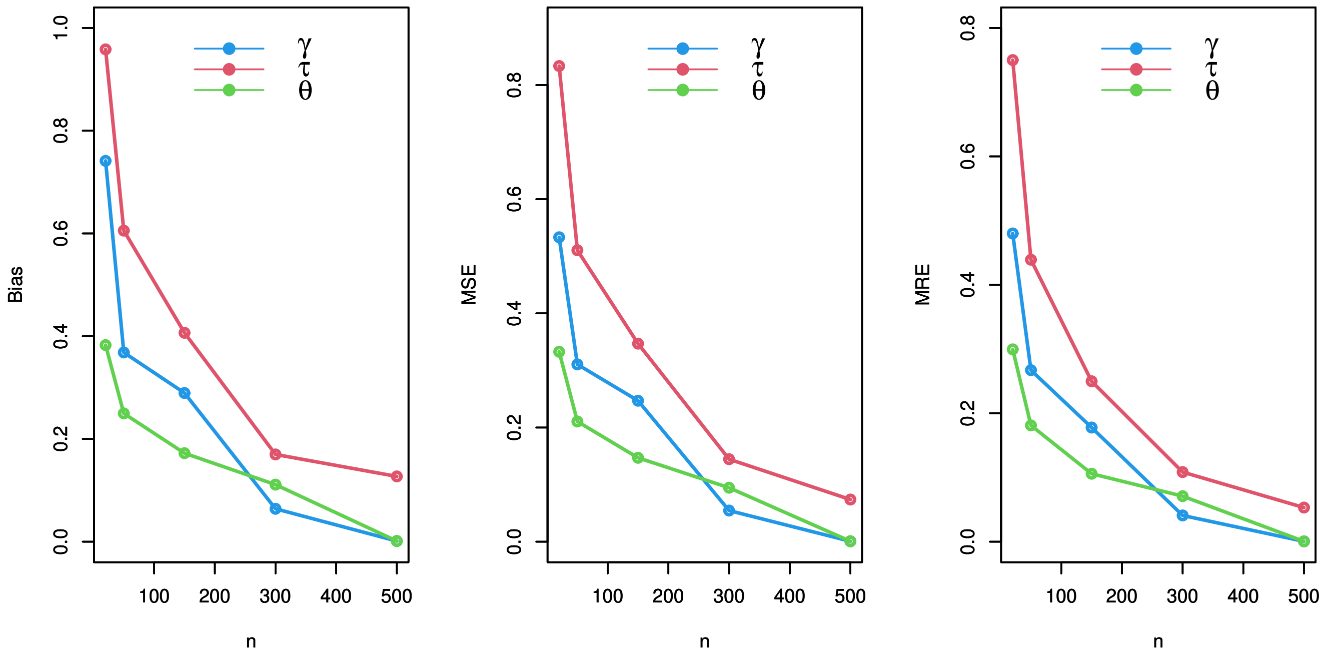

Figure 3: The parameter biases, MSE, and MRE for Schema I.

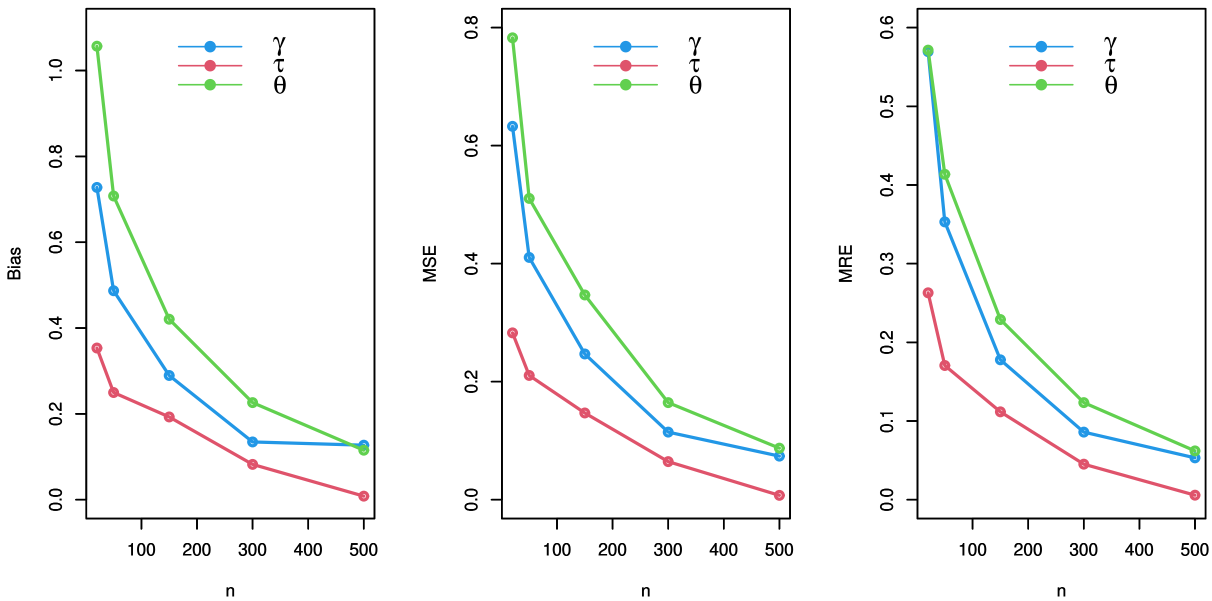

Figure 4: The parameter biases, MSE, and MRE for Schema II.

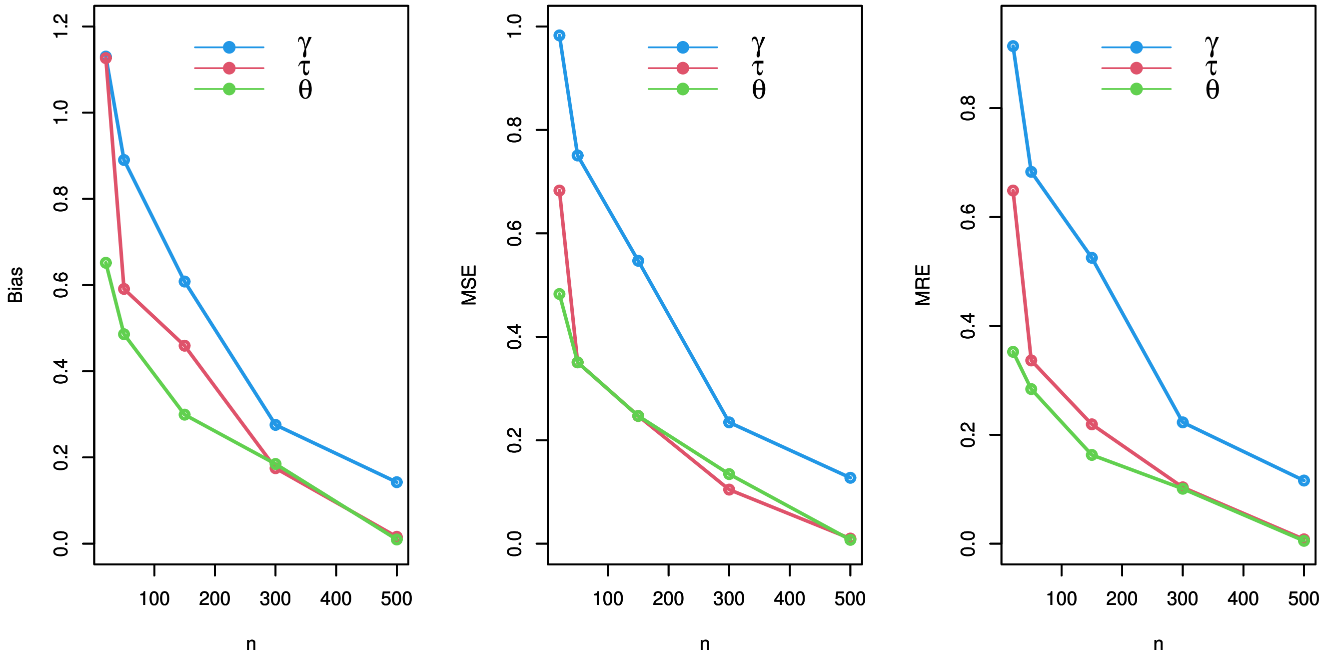

Figure 5: The parameter biases, MSE, and MRE for Schema III.

The simulation results highlight the excellent performance of the estimators in terms of bias, MSE, and MRE for larger sample sizes. In Schemas I, II, and III, the bias of the estimators decreases as the sample size increases, with bias values approaching zero, indicating that the estimators become nearly unbiased as the sample size grows. In Schema I, the estimators exhibit minimal bias across all sample sizes, and both the MSE and MRE steadily decrease as n increases, suggesting improved accuracy and precision with larger datasets. Similarly, in Schemas II and III, the bias remains consistently low and continues to decrease as sample sizes increase, demonstrating the robustness of the model's estimators. The MSE and MRE also show a clear reduction as n grows, indicating that the estimators become more efficient with larger samples. Overall, across all schemas, the estimators perform exceptionally well, showing low bias, MSE, and MRE, especially as sample sizes increase. These findings underscore the effectiveness and reliability of the MLE method for the DGzHLo model in practical applications, highlighting the robustness of the estimation process and its suitability for modeling real-world data.

6. Sustainable Data Analysis

Sustainable dispersion data modeling using discrete probability distributions helps manage variability in count data within engineering applications, especially in reliability analysis, system maintenance, and quality control. By applying discrete distributions, it effectively captures fluctuations in failure rates and component lifespans. These models account for overdispersion, where the variance surpasses the mean, a typical feature of engineering systems with high variability. In this section, we demonstrate the practical significance of the DGzHLo distribution through two real data applications. The performance of the fitted models is assessed using multiple criteria, including $-L$, Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (CAIC), Chi-square (χ^2) along with its P-value, and the Kolmogorov-Smirnov (K-S) test with its corresponding P-value. The DGzHLo distribution will be compared against several competing models presented in Table 2.

Table 2. The competitive models of the DGzHLo distribution.

Distribution	Abbreviation
Discrete Weibull	DW
Exponentiated discrete Weibull	EDW
Discrete inverse Weibull	DIW
Discrete exponential	Geo
Discrete generalized exponential type II	DGEx-II
Discrete Rayleigh	DR
Discrete inverse Rayleigh	DIR
Discrete Lindley	DLi
Exponentiated discrete Lindley	EDLi
Discrete Lindley type II	DLi-II
Discrete extension Xgamma	DExg
Discrete log-logistic	DLLc
Discrete Lomax	DLo
Two- parameter discrete Burr type XII	DB-XII
Discrete Pareto	DPa
Discrete inverted Nadarajah–Haghighi	DINH
Poisson	Poi

6.1. Dataset I: Computer breakdowns

Industrial engineering focuses on optimizing complex systems and processes to enhance efficiency and productivity across various sectors. It involves analyzing workflows, reducing waste, and ensuring the effective utilization of resources. A key aspect of industrial engineering is managing machine performance, where engineers monitor and address issues such as computer breakdowns that occur over time. These breakdowns can disrupt production schedules and cause downtime, significantly impacting overall productivity. To mitigate this, industrial engineers employ predictive maintenance strategies and data analysis to anticipate equipment failures, thus minimizing downtime and improving the reliability of manufacturing systems. This section presents a dataset that tracks the number of computer breakdowns over 128 consecutive weeks of operation (see Hand et al., [2]). Figure 6 displays the distribution of dataset I. The results show a moderate rightward skew, likely due to some extreme observations, as well as a relatively high level of variability. The kurtosis suggests that the distribution has slightly heavier tails than a normal distribution, indicating some overdispersion. Table 3 presents the goodness-of-fit

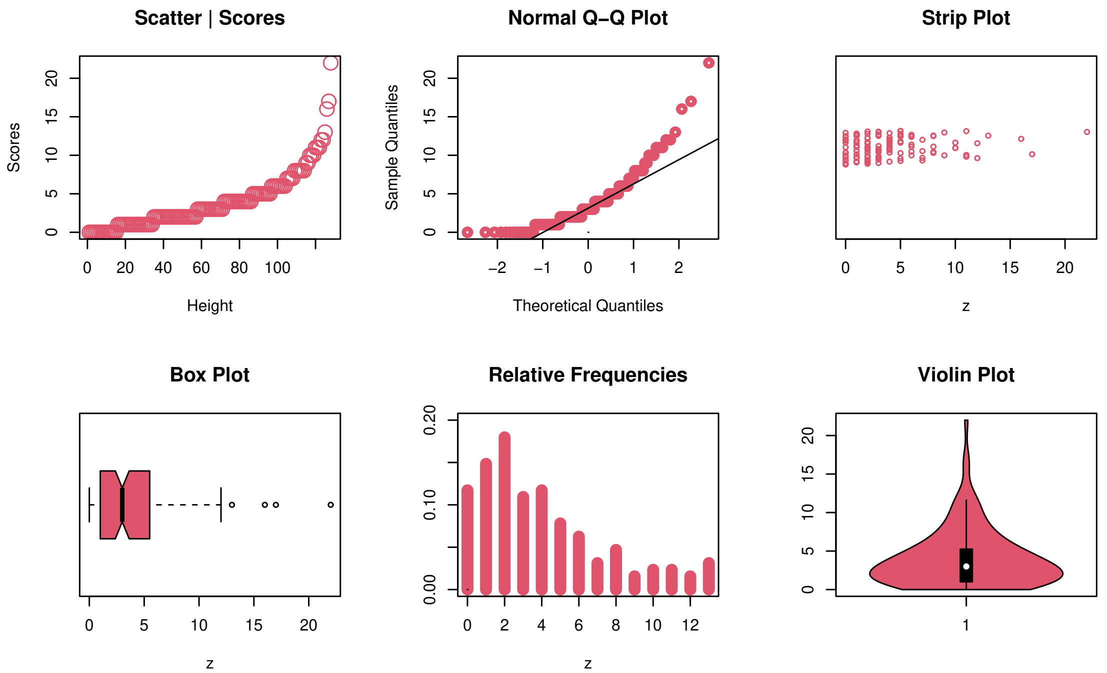

Figure 6: Non-parametric plots for dataset I.

test results for dataset I at a significance level of 0.05, evaluating the performance of several statistical models applied to the data. Figure 7 visually complements these findings by illustrating how each model fits the observed data. It compares the predicted values of each model with the actual data, highlighting any discrepancies and providing a clear overview of the overall fit. Although the DGzHLo, DExg, and Geo models all perform adequately in modeling the data, the DGzHLo model proves to be the most effective. It provides the best fit and demonstrates superior accuracy in capturing the underlying distribution of the data, offering a more precise representation than the other models.

Table 3. The goodness-of-fit test for dataset I.

X	OF	DGzHLo	DExg	DINH	DIW	DB-XII	Geo	Poi	DPa	DR
0	15	16.32	19.04	12.40	9.69	19.89	25.53	2.31	47.81	3.64
1	19	19.47	16.47	30.16	33.09	36.59	20.43	9.27	19.19	10.30
2	23	18.35	16.06	19.94	23.10	19.05	16.36	18.61	10.76	15.31
3	14	15.86	15.23	12.78	14.48	10.69	13.09	24.91	7.02	18.04
4	15	13.23	13.61	8.72	9.54	6.83	10.49	25.01	5.00	18.44
5	10	10.81	11.52	6.29	6.64	4.76	8.39	20.09	3.77	16.92
6	8	8.67	9.33	4.73	4.84	3.52	6.72	13.44	2.97	14.17
7	4	6.82	7.28	3.69	3.65	2.72	5.38	7.71	2.40	10.94
8	6	5.26	5.52	2.95	2.85	2.17	4.31	3.87	1.99	7.84
9	2	3.97	4.09	2.41	2.27	1.77	3.46	1.74	1.69	5.23
10	3	2.92	2.97	2.01	1.85	1.48	2.75	0.69	1.45	3.26
11	3	2.10	2.12	1.69	1.53	1.25	2.21	0.25	1.26	1.89
12	2	1.49	1.49	1.46	1.28	1.08	1.77	0.08	1.11	1.04
+13	4	2.73	3.27	18.76	13.183	16.19	7.11	0.01	21.59	0.99
Total	128	128	128	128	128	128	128	128	128	128
$-l$		310.657	318.953	331.931	330.446	342.581	320.703	384.974	369.766	347.148
MLE_γ		0.869	0.497	1.244	0.076	0.785	0.801	4.016	0.509	0.972
MLE_τ		0.045	0.513	1.634	1.235	3.309	–	–	–	–
MLE_θ		1.432	–	–	–	–	–	–	–	–
χ^2		4.788	7.093	20.199	18.709	40.183	11.651	88.995	94.971	49.897
P.value		0.686	0.527	0.003	0.005	0	0.234	0	0	0

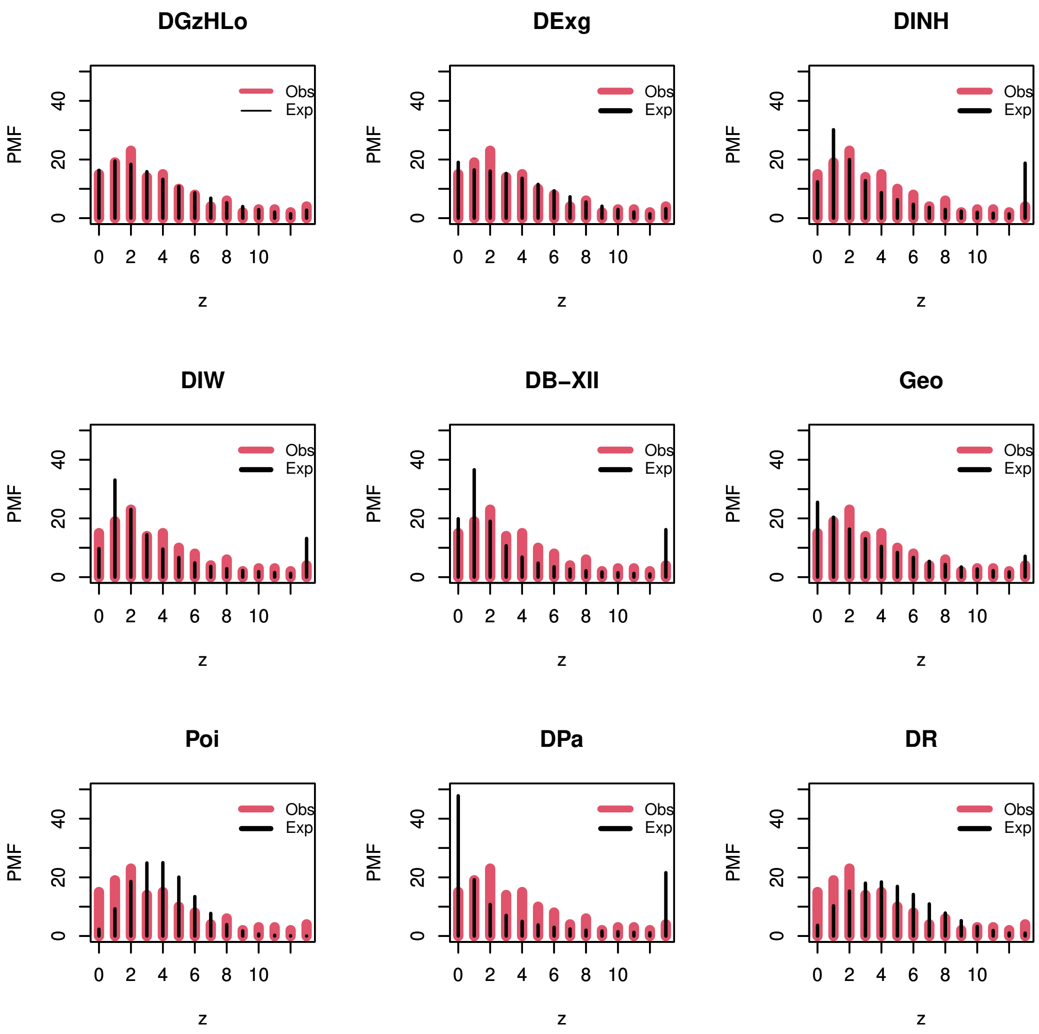

Figure 7: The fitted PMFs for dataset I.

6.2. Dataset II: Failure times

This dataset represents the failure times (in weeks) of 50 devices subjected to a life test, as discussed in Bebbington et al. [14], focusing on sustainable dispersion data. In this analysis, we will compare the fit of the DGzHLo distribution with several competing models to assess its performance. Non-parametric plots, shown in Figure 8, provide a visual representation of the data's distribution. It is important to note that the data exhibits asymmetry, with some outlier observations that may influence the model fitting and interpretation. These characteristics will be considered when evaluating the effectiveness of the DGzHLo distribution compared to other models.

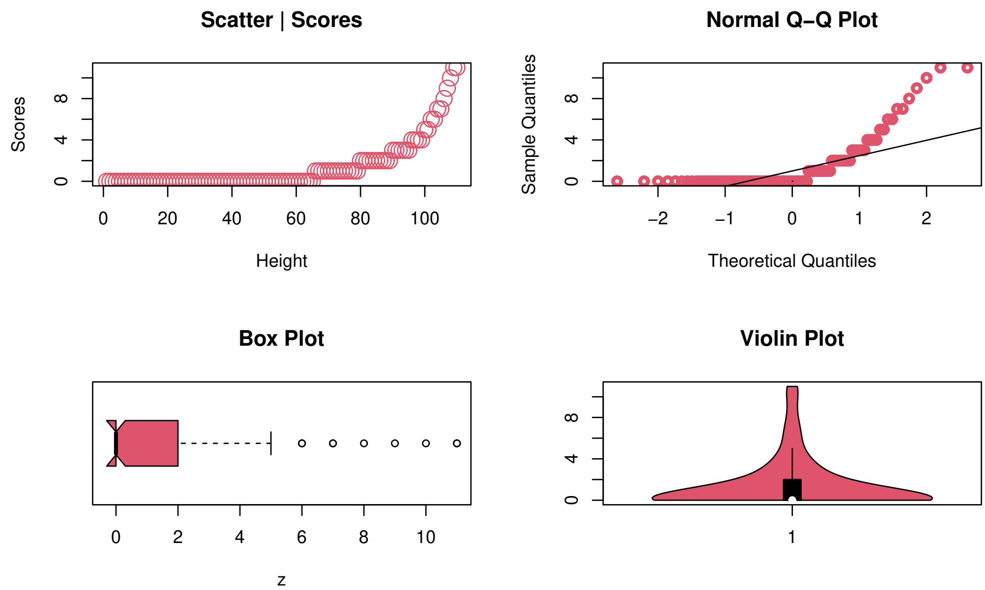

Figure 8: Non-parametric plots for dataset II.

The MLEs, along with their corresponding standard errors (Std-er), and the goodness-of-fit statistics are presented in Tables 4 and 5 at a significance level of 0.05, respectively.

Table 4. The MLEs with their corresponding Std-er for data set II.

Model ↓ Parameter →	γ		τ		θ	
	MLE	Std-er	MLE	Std-er	MLE	Std-er
DGzHLo	0.153	0.028	3.211	0.117	0.009	0.001
EDW	0.989	0.164	1.139	3.227	0.784	3.053
DW	0.981	0.011	1.023	0.131	–	–
DIW	0.018	0.013	0.582	0.061	–	–
DLi-II	0.969	0.005	0.058	0.027	–	–
EDLi	0.972	0.005	0.480	0.087	–	–
DLLc	1.0	0.321	0.439	0.062	–	–
DPa	0.739	0.032	–	–	–	–

Table 5. The goodness-of-fit statistics for data set II.

Statistic ↓ Model →	DGzHLo	EDW	DW	DIW	DLi-II	EDLi	DLLc	DPa
$-L$	235.0	240.2	241.6	261.9	240.6	240.3	294.9	275.9
AIC	476.0	486.7	487.2	527.8	485.2	484.6	593.8	553.7
CAIC	476.6	487.2	487.5	528.1	485.4	484.8	594.0	553.8
K-S	0.171	0.195	0.187	0.258	0.186	0.195	0.535	0.335
P-value	0.106	0.045	0.061	0.0026	0.064	0.045	< 0.001	< 0.001

Regarding Table 5, it is evident that the DGzHLo, EDW, DW, DLi-II, and EDLi models all perform reasonably well in analyzing this data, with the DGzHLo model showing a P-value greater than 0.05. However, our goal is always to identify the best model for optimal data evaluation. Based on the $-L$, AIC, CAIC, K-S statistics, and their corresponding P-values, we conclude that the DGzHLo model provides the best fit among all the tested models. This is evident as it has the smallest values for $-L$, AIC, CAIC, and K-S statistics, along with the highest P-value. Figure 9 further supports the findings presented in Table 5, where "PP" refers to the probability-probability plot.

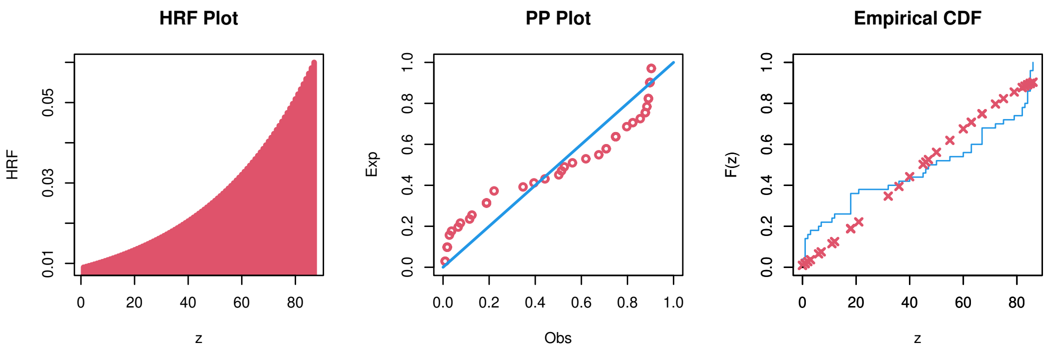

Figure 9: The HRF, PP plot, and estimated CDFs for dataset II.

It is evident that the dataset, dataset II, is best modeled by the DGzHLo distribution.

6.3. Dataset III: Electronic components

This data, reported in Lawless [7], presents the failure times for a sample of 15 electronic components in an accelerated life test, focusing on sustainable dispersion data. Non-parametric plots are shown in Figure 10. The data is asymmetric, without any extreme observations.

For this dataset, we will compare the fit of the DGzHLo distribution with several competitive models. The MLEs along with their corresponding Std-er, as well as the goodness-of-fit statistics, are presented in Tables 6 and 7, respectively.

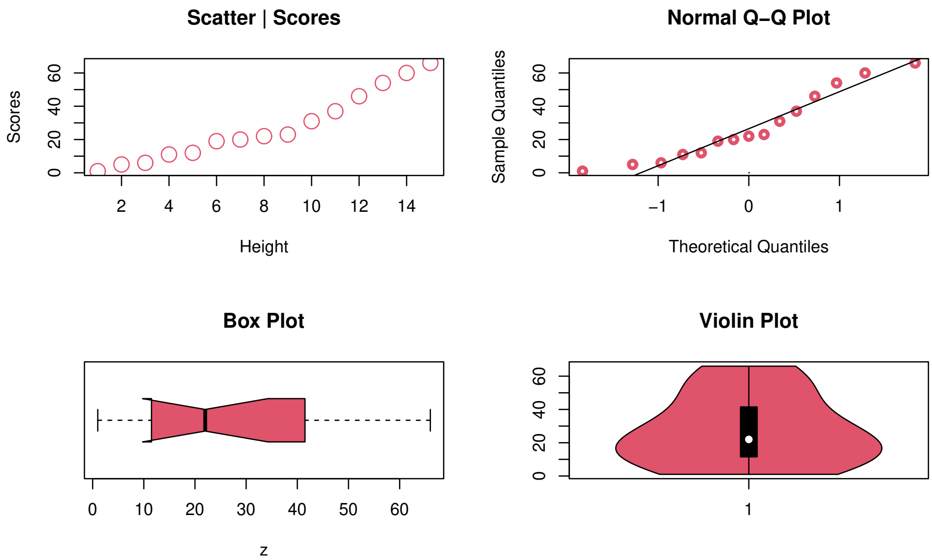

Figure 10: Non-parametric plots for dataset III.

Table 6. The MLEs with their corresponding Std-er for data set III.

Model ↓ Parameter →	γ		τ		θ	
	MLE	Std-er	MLE	Std-er	MLE	Std-er
DGzHLo	0.959	0.019	0.037	0.026	0.550	0.004
DGzEx	0.587	0.023	0.588	0.041	0.039	0.002
DEx	0.965	0.009	–	–	–	–
DGEx-II	0.956	0.013	1.491	0.535	–	–
DR	0.999	2.58×10^{-4}	–	–	–	–
DIR	1.8×10^{-7}	0.055	–	–	–	–
DIW	2.2×10^{-4}	7.75×10^{-4}	0.875	0.164	–	–
DLo	0.012	0.039	104.506	84.409	–	–
DB-XII	0.975	0.051	13.367	27.785	–	–
DPa	0.720	0.061	–	–	–	–

Table 7. The goodness-of-fit statistics for data set III.

Statistic	Model									
	DGzHLo	DGzEx	Geo	DGEx-II	DR	DIR	DIW	DLo	DB-XII	DPa
$-L$	62.934	63.804	65.000	64.420	66.394	89.096	68.703	65.864	75.724	77.402
AIC	131.868	133.608	134.000	134.839	134.788	180.192	141.406	135.728	155.448	156.805
CAIC	134.049	135.789	136.308	135.839	136.096	180.499	142.406	136.728	156.448	157.112
K-S	0.116	0.120	0.177	0.129	0.216	0.698	0.209	0.205	0.388	0.405
P-value	0.974	0.963	0.673	0.937	0.433	9.1×10^{-7}	0.482	0.491	0.015	0.009

Regarding Table 7, the DGzHLo distribution emerges as the best model among all the tested models. Figure 11 further supports the results presented in Table 7. It is clear that the dataset likely originates from the DGzHLo model.

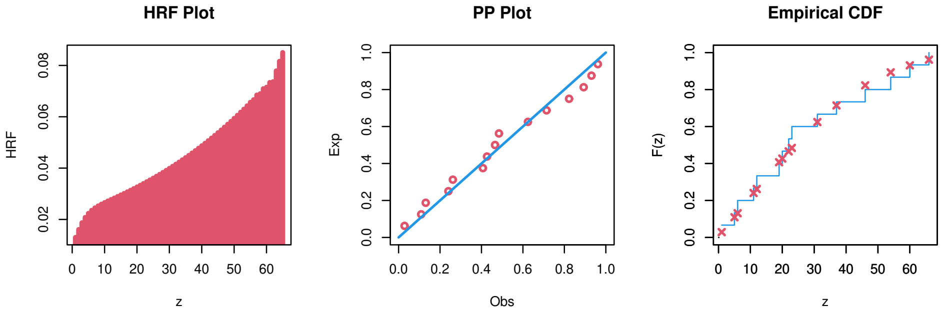

Figure 11: The HRF, P-P plot, estimated CDFs for dataset III.

7. Concluding Remarks and Future Work

This article proposes a new discrete model, the DGzHLo, and examines its statistical properties. The DGzHLo model is found to offer flexibility in modeling data with various skewness types, including negative, positive, and symmetric shapes. This versatility makes it particularly useful for modeling sustainability dispersion data, where variability often follows diverse patterns. The DGzHLo model accommodates increasing hazard rate functions, making it highly applicable for both engineering and medical applications, where such rate functions are crucial for reliability analysis and predicting system failures. Additionally, the model can handle over-dispersion, equi-dispersion, and under-dispersion, accurately capturing a wide range of data variability often observed in real-world applications. Another key feature of the DGzHLo model is its ability to account for different levels of kurtosis, including leptokurtic and platykurtic distributions. This ability makes it useful in handling data with varying tail behaviors, often encountered in sustainability studies. Furthermore, the model effectively handles zero-inflation and non-inflation data, particularly in the presence of extreme observations, a common challenge in many fields, including environmental studies and economics. The parameters of the DGzHLo model were estimated using the MLE method. A simulation study was carried out to evaluate the model's performance, and the results demonstrated that the MLE method effectively estimated the model parameters with high accuracy. The flexibility of the DGzHLo model was further showcased through the analysis of three distinct datasets, where it outperformed other models in terms of both fit and predictive accuracy. This confirmed its potential as a reliable tool for analyzing complex sustainability dispersion data. Looking ahead, future research will focus on extending the DGzHLo model to include time-varying parameters and exploring its application in additional fields such as environmental monitoring, climate change studies, and economics. Further refinements to estimation techniques, particularly through the use of Bayesian methods, will aim to improve the model's robustness and computational efficiency. Additionally, integrating the DGzHLo model with machine learning approaches will be explored to enable real-time predictive analytics across various industries, further enhancing its applicability and utility in complex data analysis.

Acknowledgements

The authors gratefully acknowledge Qassim University, represented by the Deanship of Graduate Studies and Scientific Research, on the financial support for this research under the number (QU-J-UG-2-2025-52613) during the academic year 1446 AH / 2024 AD.

- **Data Availability Statement**: The data sets are available in the paper.
- **Conflicts of Interest**: The authors declare no conflict of interests.

References

[1] A. Alzaatreh, C. Lee, and F. Famoye. A new method for generating families of continuous distributions. *Metron*, 71:63–79, 2013.

[2] D. J. Hand, F. Daly, K. M. Conway, D. Lunn, and E. Ostrowski. A handbook of small data sets, Chapman & Hall. DOI: 10.1201/9780429246579, 1993.

[3] E. Abd EL-Hady, M. A. Hegazy, and A. A. EL-Helbawy. Discrete exponentiated generalized family of distributions. *Computational Journal of Mathematical and Statistical Sciences*, 2:303–327, 2023.

[4] E. Gómez-Déniz, and E. Calderín-Ojeda. The discrete Lindley distribution: Properties and applications. *Journal of Statistical Computation and Simulation*, 81:1405–1416, 2011.

[5] G. M. Cordeiro, M. Alizadeh, D. C. Abraao, and R. Mahdi. The exponentiated Gompertz generated family of distributions: Properties and applications. *Chilean Journal of Statistics*, 7:29–50, 2016.

[6] H. M. Yousof, A. H. Al-Nefaie, N. S. Butt, G. G. Hamedani, H. Alrweili, A. Aljadani, and M. Ibrahim. A new discrete generator with mathematical characterization, properties, count statistical modeling and inference with applications to reliability, medicine, agriculture, and biology data. *Pakistan Journal of Statistics and Operation Research*, 745–770, 2024.

[7] J. F. Lawless. Statistical Models and Methods for Lifetime Data. Wiley, New York, USA, 2003.

[8] J. Mazucheli, A. F. Menezes, and S. Dey. Unit-Gompertz distribution with applications. *Statistica*, 79:25–43, 2019.

[9] J. R. Hosking and J. R. Wallis. Regional frequency analysis: An approach based on L-moments. Cambridge University Press, 1997.

[10] J. Wang, and X. Guo. The Gompertz model and its applications in microbial growth and bioproduction kinetics: Past, present, and future. *Biotechnology Advances*, 27:17–29, 2024.

[11] M. Alizadeh, A. Z. Afify, M. S. Eliwa, and S. Ali. The odd log-logistic Lindley-G family of distributions: Properties, Bayesian and non-Bayesian estimation with applications. *Computational Statistics*, 35:281–308, 2020.

[12] M. Alizadeh, G. M. Cordeiro, G. B. Luis, and I. Ghosh. The Gompertz-G family of distributions. *Journal of Statistical Theory and Practice*, 11:179–207, 2017.

[13] M. Alizadeh, M. Afshari, J. E. Contreras-Reyes, D. Mazarei, and H. M. Yousof. The extended Gompertz model: Applications, mean of order P assessment, and statistical threshold risk analysis based on extreme stresses data. *IEEE Transactions on Reliability*, DOI: 10.1109/TR.2024.3425278, 1-13, 2024.

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

[15] M. El-Morshedy, and M. S. Eliwa. The odd flexible Weibull-H family of distributions: Properties and estimation with applications to complete and upper record data. *Filomat*, 33:2635–2652, 2019.

[16] M. Ibrahim, N. S. Butt, A. H. Al-Nefaie, G. G. Hamedani, H. M. Yousof, and A. S. Mahmoud. An extended discrete model for actuarial data and value at risk analysis: Properties, applications, and risk analysis under financial automobile claims data. *Statistics, Optimization & Information Computing,* 13:27–46, 2025.

[17] M. M. Alsolmi. A new logarithmic tangent-U family of distributions with reliability analysis in engineering data. *Computational Journal of Mathematical and Statistical Sciences*, 4:258–282, 2025.

[18] M. N. Atchadé, A. A. Agbahide, T. Otodji, M. J. Bogninou, and A. Moussa Djibril. A new shifted Lomax-X family of distributions: Properties and applications to actuarial and financial data. *Computational Journal of Mathematical and Statistical Sciences*, 4:41–71, 2025.

[19] M. S. Eliwa, M. El-Morshedy, and M. Ibrahim. Inverse Gompertz distribution: Properties and different estimation methods with application to complete and censored data. *Annals of Data Science*, 6:321–339, 2019.

[20] M. S. Eliwa, M. H. Tahir, M. A. Hussain, B. Almohaimeed, A. Al-Bossly, and M. El-Morshedy. Univariate probability-G classes for scattered samples under different forms of hazard: Continuous and discrete versions with their inference tests. *Mathematics*, 11:2929, 2023.

[21] M. S. Eliwa, Z. A. Alhussain, and M. El-Morshedy. Discrete Gompertz-G family of distributions for over-and under-dispersed data with properties, estimation, and applications. *Mathematics*, 8(3), 358, 2020.

[22] R. A. Bantan, F. Jamal, C. Chesneau, and M. Elgarhy. Theory and applications of the unit gamma/Gompertz distribution. *Mathematics*, 9:1850, 2021.

[23] R. Roozegar, S. Tahmasebi, and A. A. Jafari. The McDonald Gompertz distribution: Properties and applications. *Communications in Statistics-Simulation and Computation*, 46:3341–3355, 2017.

[24] Rényi. On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. *Contributions to the Theory*

of Statistics, 547–562. University of California Press, 1961.

[25] Roy. Discrete Rayleigh distribution. *IEEE Transactions on Reliability*, 53:255–260, 2004.

[26] S. Hassan, and A. M. Abdelghaffar. Bayesian and E-Bayesian estimation of Gompertz distribution in stress-strength reliability model under partially accelerated life testing. *Computational Journal of Mathematical and Statistical Sciences*, 4:348–378, 2025.

[27] T. Nakagawa, and S. Osaki. The discrete Weibull distribution. *IEEE Transactions on Reliability*, 24:300–301, 1975.

[28] W. A. Hamdi, Y. Akdoğan, T. Erbayram, M. Alqawba, and A. Z. Afify. Statistical inference of the generalized process capability index for the discrete Lindley distribution. *Scientific Reports*, 15:6776, 2025.

[29] Y. Y. Abdelall, A. S. Hassan, and E. M. Almetwally. A new extension of the odd inverse Weibull-G family of distributions: Bayesian and non-Bayesian estimation with engineering applications. *Computational Journal of Mathematical and Statistical Sciences*, 3:359–388, 2024.

[30] Z. Afify, M. Ahsan-ul-Haq, H. M. Aljohani, A. S. Alghamdi, A. Babar, and H. W. Gómez. A new one-parameter discrete exponential distribution: Properties, inference, and applications to COVID-19 data. *Journal of King Saud University-Science*, 34:102199, 2022.