EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
2025, Vol. 18, Issue 3, Article Number 6080
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

Sustainable Data Analysis: Developing a Novel Probability Mass Function With Mathematical and Inferential Foundations

Mohamed S. Eliwa[1,*], Lama A. Alqueer[1], Hussah Y. Alseilo[1]

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

Abstract. This study presents a flexible extension of the discrete Gompertz distribution, developed within the framework of the exponentiated geometric model. The resulting formulation, termed the discrete Gompertz exponentiated geometric (DGzExGc) model, enhances the ability of existing discrete distributions to model complex and diverse data structures more effectively. The key mathematical and statistical properties of the model are derived, including the probability mass function, cumulative distribution function, reliability function, and hazard rate function. Additional measures, such as the index of dispersion, skewness, and kurtosis, are explored to assess the model's behavior. Furthermore, entropy and order statistics are examined to provide deeper insights into its structural characteristics. The model accommodates both positively skewed and symmetric distributions, as well as unimodal and bimodal structures, making it highly applicable across various domains. It effectively captures a range of hazard rate functions, such as increasing, decreasing, bathtub-shaped, and increasing-constant trends, which are particularly useful for reliability assessments in medicine, engineering, and environmental studies. A critical feature of sustainable data analysis is the model's ability to handle varying dispersion levels, and the DGzExGc model performs exceptionally well in modeling equi-dispersed, over-dispersed, and under-dispersed count data. Additionally, it demonstrates the ability to represent a range of kurtosis levels, from leptokurtic to platykurtic distributions, proving its robustness in capturing diverse data structures. Parameter estimation is conducted using the maximum likelihood method, and a comprehensive simulation study evaluates the performance of the estimators across different sample sizes. To demonstrate the model's flexibility and applicability, three real-world datasets are analyzed, showcasing its superior adaptability and robustness compared to existing models.

2020 Mathematics Subject Classifications: 62E15, 62F10

Key Words and Phrases: Statistical model, Discrete Gompertz-G family, Dispersion index, Maximum likelihood method, L-moment statistics, Simulation, Data analysis.

---

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v18i3.6080

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

1. Introduction

The rapid expansion of sustainable data, fueled by technological advancements and digital transformation, has led to increasingly complex data structures that demand sophisticated analytical approaches. Traditional probability models often fall short in capturing the intricate dependencies, patterns, and variations inherent in modern datasets, particularly in fields such as engineering, epidemiology, finance, and actuarial sciences. As a result, there is a growing need for advanced probability models that offer enhanced flexibility and adaptability to ensure sustainable data analysis. One such model that has gained prominence in probability and statistics is the Gompertz (Gz) distribution, a continuous probability distribution that extends the exponential (Ex) distribution by incorporating an additional shape parameter. This extension enhances its ability to model diverse real-world phenomena, particularly in cases involving dynamic hazard rates. Due to its flexibility, the Gz model has been widely utilized in survival analysis, reliability engineering, and actuarial sciences, proving to be a valuable tool for sustainable data-driven decision-making across various disciplines. The random variable T is said to have the Gz distribution with shape parameter $\zeta > 0$ and scale parameter $\rho > 0$, if its cumulative distribution function (CDF) is given by

$$K(t;\zeta,\rho) = 1 - e^{-\frac{\zeta}{\rho}(e^{\rho t}-1)}; \quad t > 0. \tag{1}$$

The Gz distribution (see Gompertz, [21]) is often applied to describe the distribution of adult lifespans by demographers and actuaries. Related fields of science such as biology and gerontology also considered the Gz distribution for the analysis of survival. More recently, computer scientists have also started to model the failure rates of computer codes by the Gz distribution. In marketing science, it has been used as an individual-level simulation for customer lifetime value modeling. For more details see, Melnikov and Romaniuk [27], Ohishi et al. [29], Bemmaor et al. [7], Cordeiro et al. [11], Roozegar et al. [31], Eliwa et al. [17], Mazucheli et al. [26], Wang and Guo [33], along with other references cited therein.

Alizadeh et al. [4] introduced the Gz-G family based on Alzaatreh et al. [5] technique, where Alzaatreh et al. [5] introduced a general form to generate a new family named transformed-transformer family. Thus, the random variable X is said to have the Gz-G family if its CDF is given by

$$V(y;\zeta,\rho,\boldsymbol{\Upsilon}) = 1 - e^{-\frac{\zeta}{\rho}\left\{[1-G(y;\boldsymbol{\Upsilon})]^{-\rho}-1\right\}}; \quad y > 0, \tag{2}$$

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

$$\overline{V}(y;\zeta,\rho,\boldsymbol{\Upsilon}) = e^{-\frac{\zeta}{\rho}\left\{[1-G(y;\boldsymbol{\Upsilon})]^{-\rho}-1\right\}}; \quad y > 0. \tag{3}$$

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

$$v(y;\zeta,\rho,\boldsymbol{\Upsilon}) = \zeta g(y;\boldsymbol{\Upsilon}) \left[\overline{G}(y;\boldsymbol{\Upsilon})\right]^{-(\rho+1)} e^{-\frac{\zeta}{\rho}\left\{[1-G(y;\boldsymbol{\Upsilon})]^{-\rho}-1\right\}}; \quad y > 0, \tag{4}$$

where $g(y; \boldsymbol{\Upsilon})$ is the baseline PDF. Several researchers have applied the technique introduced by Alzaatreh et al. [5] to develop univariate families. Notable examples include the works of Alizadeh et al. [3], El-Morshedy et al. [18], Eghwerido and Agu [12], Zaidi et al. [34], Kajuru et al. [23], Eghwerıdo and Nzeı [13], Lekhane et al. [25], along with other references cited therein.

In recent years, discretizing continuous distributions has gained significant attention in statistical research. This process becomes relevant when measuring the lifespan of a product or device on a continuous scale is either impractical or impossible. In such cases, recording lifetimes on a discrete scale is preferred. As a result, numerous discrete distributions have been introduced in the literature, Gómez-Déniz [20], Bebbington et al. [6], Nekoukhou and Bidram [28], Chandrakant et al. [9], Eliwa and El-Morshedy [14, 15], Ahsan-ul-Haq et al. [1], El-Morshedy et al. [19], Akkanphudit [2], Chuncharoenkit et al. [10], along with other references cited therein.

Despite the existing range of discrete distributions, there remains ample room for developing new discretized models tailored to various conditions. In this study, we introduce a novel discrete probability model based on the discrete analogue of the Gz-G class, originally proposed by Eliwa et al. [16]. This new family, termed the discrete Gz-G (DGz-G) family, is characterized by a random variable Z following the DGz-G distribution, with its cumulative distribution function (CDF) defined by

$$W_Z(z; \nu, \rho, \boldsymbol{\Upsilon}) = 1 - \nu^{\frac{1}{\rho}\left\{[1-G(z+1;\boldsymbol{\Upsilon})]^{-\rho}-1\right\}}; \quad z \in \mathbb{N}_0, \tag{5}$$

where $\nu = e^{-\zeta}$, $0 < \nu < 1$, $\rho > 0$ and $\mathbb{N}_0 = \{0, 1, 2, 3, ...\}$. So, the RF of the DGz-G family can be represented as

$$\overline{W}_Z(z; \nu, \rho, \boldsymbol{\Upsilon}) = \nu^{\frac{1}{\rho}\left\{[1-G(z+1;\boldsymbol{\Upsilon})]^{-\rho}-1\right\}}; \quad z \in \mathbb{N}_0. \tag{6}$$

Assume that

$$G(z; \alpha, \beta) = [1 - (1 - \alpha)^z]^\beta; \ ; \quad z \in \mathbb{N}_0, \tag{7}$$

represents the exponentiated geometric (ExGc) distribution with parameters $0 < \alpha < 1$ and $\beta > 0$. Based on the DGz-G and exponentiated geometric (ExGe) models, a new probability model can be formulated, referred to as the discrete Gompertz exponentiated geometric (DGzExGc) model. The reasons for introducing the DGzExGc model are as follows:

- The PMF of the DGzExGc model is highly flexible, making it applicable to data sets with either a positively skewed or symmetric shape. Additionally, it can represent both unimodal and bimodal distributions, which is a feature that is not commonly found in the literature. This flexibility allows the model to accommodate a wide range of data distributions.
- The HRF of the DGzExGc model is versatile, capable of exhibiting various trends, including increasing, decreasing, bathtub-shaped, and increasing-constant behaviors. This wide range of possible hazard rate patterns makes the model suitable for different applications where the nature of the hazard rate is complex and diverse.

- The DGzExGc model is specifically designed to model data sets with different dispersion characteristics. It is appropriate for analyzing equi-dispersed, over-dispersed, and under-dispersed data, providing a flexible tool for handling various types of variability in real-world data. This adaptability is crucial for accurately modeling and analyzing different data structures.

- The model is also well-suited for analyzing data with different kurtosis levels. It can be applied to leptokurtic data (with high kurtosis, indicating heavy tails) as well as platykurtic data (with low kurtosis, indicating a flatter distribution). This makes the DGzExGc model a valuable tool for handling data with varying degrees of peakedness or flatness in the tails.

This paper is organized as follows: Section 2 presents the DGzExGc distribution, detailing its definition and formulation. In Section 3, the statistical and reliability properties of the distribution are thoroughly examined, highlighting its key characteristics. Section 4 focuses on the application of the maximum likelihood estimation method for estimating the parameters of the DGzExGc model. A simulation study is conducted in Section 5 to assess the performance of the estimators across different sample sizes, providing insight into the accuracy and efficiency of the parameter estimation process. In Section 6, the practical applicability of the DGzExGc model is demonstrated by analyzing three real-world datasets, where the model's performance is compared to several well-established distributions, showing its superior fit. Finally, Section 7 provides a comprehensive conclusion, summarizing the key findings and potential implications of the study.

2. Mathematical Framework of the DGzExGc Model

Recall, Equations (5), (6) and (7), the CDF and RF of the DGzExGc can be listed as

$$F_Z(z;\nu,\rho,\alpha,\beta) = 1-\nu^{\frac{1}{\rho}\left\{\left[1-\left[1-(1-\alpha)^{z+1}\right]^{\beta}\right]^{-\rho}-1\right\}};\quad z\in\mathbb{N}_0, \tag{8}$$

$$\overline{F}_Z(z;\nu,\rho,\alpha,\beta) = \nu^{\frac{1}{\rho}\left\{\left[1-\left[1-(1-\alpha)^{z+1}\right]^{\beta}\right]^{-\rho}-1\right\}};\quad z\in\mathbb{N}_0, \tag{9}$$

respectively. Let $Z_1, Z_2, ..., Z_n$ be non-negative independent and identically distributed integer valued random variables and $X = \min(Z_1, Z_2, ..., Z_n)$, then $X \sim \text{DGzExGc}(z;\nu^n,\rho,\alpha,\beta)$ model provided $Z_i(i = 1, 2, ..., n) \sim \text{DGzExGc}(z;\nu,\rho,\alpha,\beta)$ model where

$$\overline{F}_X(z;\nu,\rho,\alpha,\beta) = \prod_{i=1}^{n}\Pr\left[Z_i \geq z\right] = \left(\Pr\left[Z_1 \geq z\right]\right)^n = \nu^{\frac{n}{\rho}\left\{\left[1-\left[1-(1-\alpha)^{z+1}\right]^{\beta}\right]^{-\rho}-1\right\}}. \tag{10}$$

The probability mass function (PMF) corresponding to Equation (8) can be expressed as

$$\begin{aligned} f_z(z;\nu,\rho,\alpha,\beta) &= \overline{F}(z;\nu,\rho,\alpha,\beta) - \overline{F}(z+1;\nu,\rho,\alpha,\beta) \\ &= \nu^{-\frac{1}{\rho}}\left[\nu^{\frac{1}{\rho}\left[1-\left[1-(1-\alpha)^{z}\right]^{\beta}\right]^{-\rho}} - \nu^{\frac{1}{\rho}\left[1-\left[1-(1-\alpha)^{z+1}\right]^{\beta}\right]^{-\rho}}\right];\ z\in\mathbb{N}_0. \end{aligned} \tag{11}$$

For further information on the survival discretization approach, see studies including that of Roy [32]. The hazard rate function (HRF) can be formulated as

$$h(z;\nu,\rho,\alpha,\beta) = 1 - \nu^{\frac{1}{\rho}\left\{\left[1-[1-(1-\alpha)^{z+1}]^{\beta}\right]^{-\rho}-\left[1-[1-(1-\alpha)^{z}]^{\beta}\right]^{-\rho}\right\}}; \quad z \in \mathbb{N}_0. \tag{12}$$

The PMF presented in Equation (11) exhibits diverse shapes depending on the parameter values, including decreasing, unimodal, and bimodal behaviors. This flexibility allows the distribution to effectively capture different patterns observed in real-world data. Similarly, the HRF in Equation (12) demonstrates a wide range of possible shapes, such as increasing, decreasing, bathtub-shaped, and increasing-constant trends. These variations in the HRF indicate the model's capability to represent different failure rate behaviors, making it suitable for applications in reliability analysis and risk assessment. To illustrate these characteristics, Figures 1 and 2 provide graphical representations of the PMF and HRF of the DGzExGc distribution for various parameter settings, highlighting its adaptability in modeling diverse data structures.

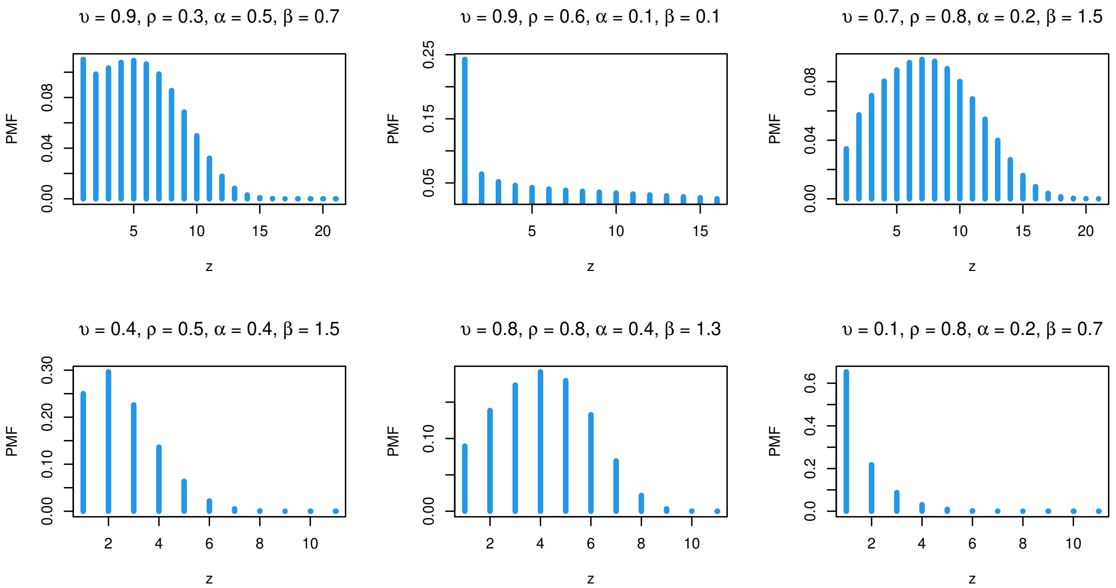

Figure 1. The PMF of the DGzExGc distribution.

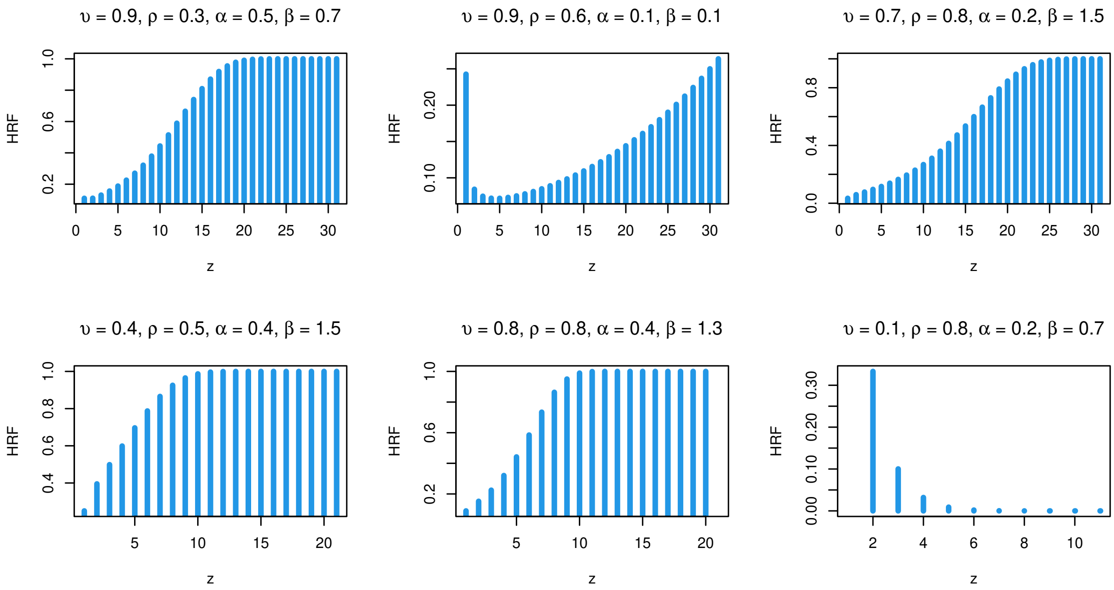

Figure 2. The HRF of the DGzExGc distribution.

3. Different Statistical Properties of the DGzExGc Model

3.1. Moments, cumulants, and associated measures

Moments and cumulants are key statistical tools that help us understand the shape, variability, and dependencies within a probability distribution. Moments like the mean, variance, and index of dispersion describe the central tendency and spread of the distribution. Cumulants, however, provide a different perspective, capturing features such as skewness and kurtosis. These measures are vital for tasks like statistical modeling, reliability analysis, and parameter estimation. In this section, we will derive the moments, cumulants, and related measures for the DGzExGc model, which are essential for grasping its core properties and practical applications. Assume non-negative random variable $Z \sim$ DGzExGc$(z;\nu,\rho,\alpha,\beta)$, then the rth moment of Z can be expressed as

$$
\begin{aligned}
\mu_r' = E(Z^r) &= \sum_{z=0}^{\infty} z^r f_z(z;\nu,\rho,\alpha,\beta) \\
&= \sum_{z=1}^{\infty} \left[z^r - (z-1)^r\right] \overline{F}_Z(z-1;\nu,\rho,\alpha,\beta) \\
&= \nu^{-\frac{1}{\rho}} \sum_{z=1}^{\infty} \left[z^r - (z-1)^r\right] \nu^{\frac{1}{\rho}\left[1-\left[1-(1-\alpha)^z\right]^\beta\right]^{-\rho}}.
\end{aligned} \tag{13}
$$

Using Equation (13), the mean (μ_1') and variance, say $\mathrm{Var}(Z)$, can be respectively listed as

$$\mu_1' = \nu^{-\frac{1}{\rho}} \sum_{z=1}^{\infty} \nu^{\frac{1}{\rho}\left[1-[1-(1-\alpha)^z]^\beta\right]^{-\rho}},$$

$$\mathrm{Var}(Z) = \nu^{-\frac{1}{\rho}} \sum_{z=1}^{\infty} (2z-1)\, \nu^{\frac{1}{\rho}\left[1-[1-(1-\alpha)^z]^\beta\right]^{-\rho}} - (\mu_1')^2. \tag{14}$$

The dispersion index (DsI) is the ratio of variance to the mean and is used to determine if a model is appropriate for over- or under-dispersed data. It is commonly applied in ecology as a standard measure to assess clustering (over-dispersion) or repulsion (under-dispersion). When DsI > 1, the distribution is considered over-dispersed, while DsI < 1 indicates under-dispersion. The DsI for the DGzExGc model is given by

$$\mathrm{DsI}(Z) = \frac{\sum_{z=1}^{\infty} (2z-1)\, \nu^{\frac{1}{\rho}\left[1-[1-(1-\alpha)^z]^\beta\right]^{-\rho}}}{\sum_{z=1}^{\infty} \nu^{\frac{1}{\rho}\left[1-[1-(1-\alpha)^z]^\beta\right]^{-\rho}}} - \sum_{z=1}^{\infty} \nu^{\frac{1}{\rho}\left[1-[1-(1-\alpha)^z]^\beta\right]^{-\rho}}. \tag{15}$$

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

$$\begin{aligned}
M_Z(t) &= \sum_{z=0}^{\infty} e^{zt} f_z(z;\nu,\rho,\alpha,\beta) \\
&= \nu^{-\frac{1}{\rho}} \left[\sum_{z=0}^{\infty} e^{zt} \nu^{\Lambda(z;\rho)} - \sum_{z=0}^{\infty} e^{zt} \nu^{\Lambda(z+1;\rho)} \right] \\
&= \nu^{-\frac{1}{\rho}} [\left(\nu^{\Lambda(0;\rho)} + e^{t}\nu^{\Lambda(1;\rho)} + e^{2t}\nu^{\Lambda(2;\rho)} + e^{3t}\nu^{\Lambda(3;\rho)} + \ldots \right) \\
&- \left(\nu^{\Lambda(1;\rho)} + e^{t}\nu^{\Lambda(2;\rho)} + e^{2t}\nu^{\Lambda(3;\rho)} + e^{3t}\nu^{\Lambda(4;\rho)} + \ldots \right)] \\
&= \nu^{-\frac{1}{\rho}} \left[1 + \sum_{z=1}^{\infty} \left(e^{zt} - e^{(z-1)t} \right) \nu^{\Lambda(z;\rho)} \right],
\end{aligned} \tag{16}$$

where

$$\Lambda(z;\rho) = \frac{1}{\rho} \left[1 - [1-(1-\alpha)^z]^\beta \right]^{-\rho}.$$

The first four derivatives of Equation (16), 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 (13) or (16), the skewness, say $\mathrm{Sk}(Z)$ and kurtosis, say $\mathrm{Ku}(Z)$, can be reported as

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

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

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

$$k_n = \frac{d^n}{dt^n} \log M_Z(t)|_{t=0};\ n = 1, 2, 3, ... \tag{17}$$

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

$$k_n = \mu_n' - \sum_{m=1}^{n-1} \binom{n-1}{m-1} \mu_{n-m}' k_m. \tag{18}$$

The first cumulant corresponds to the mean, the second cumulant represents the variance, and the third cumulant aligns with the third central moment. However, starting from the fourth cumulant, the relationship with central moments breaks down, and higher-order cumulants no longer directly correspond to central moments. For the DGzExGc distribution, it is not possible to express the rth moment in a closed form. As a result, we use Maple software to explore and calculate some of its statistical properties. Table 1 presents some descriptive statistics derived from the DGzExGc model for different parameter values α and β with $\nu = 0.5$ and $\rho = 0.2$. The DGzExGc distribution is a highly versatile and flexible model, making it suitable for a wide range of data sets. Specifically, it can effectively model over-, equi-, and under-dispersed data, making it a valuable tool for handling various types of variability in real-world data. Additionally, this distribution is appropriate for modeling positively skewed data, providing flexibility in representing different shapes of data distributions. The DGzExGc distribution can also model leptokurtic data (with kurtosis greater than 3) when the parameter β is small, and platykurtic data (with kurtosis less than 3) when β tends to infinity, while other parameters remain constant. This adaptability allows the distribution to accommodate a range of data characteristics, making it a useful choice for statistical modeling across diverse applications.

Table 1. Some descriptive statistics using the DGzExGc model.

Measure	$\alpha\downarrow\beta\rightarrow$	0.1	0.3	0.5	1.5	2.5
Mean	0.1	1.4665	4.3010	6.5214	13.3899	17.357
	0.5	0.1221	0.4232	0.68855	1.6224	2.2134
	0.9	0.00578	0.0315	0.06160	0.2073	0.3307
Var	0.1	13.6779	41.2747	59.9455	99.6672	113.216
	0.5	0.2160	0.7752	1.2234	2.3320	2.6999
	0.9	0.00581	0.0315	0.0607	0.1855	0.2677
DsI	0.1	9.3268	9.5963	9.1921	7.4434	6.5224
	0.5	1.7684	1.8317	1.7768	1.4373	1.2198
	0.9	1.0056	1.0001	0.9860	0.8948	0.8094
Sk	0.1	4.8796	2.1688	1.6891	1.0545	0.8932
	0.5	3.8269	2.5972	1.9483	1.0626	0.8540
	0.9	13.2575	5.6254	3.9395	1.8413	1.2270
Ku	0.1	21.4932	8.5369	6.1866	4.0700	3.7193
	0.5	32.2581	10.8216	7.1720	4.0500	3.6637
	0.9	180.5851	34.5500	17.990	5.4425	3.5515

This analysis highlights the critical role of parameters α and β in shaping the distribution's behavior concerning dispersion, skewness, and kurtosis key aspects in sustainable data modeling. The ability to adjust these parameters enables the DGzExGc model to capture a wide range of data structures, making it a powerful tool for analyzing sustainable data in various applications. By accommodating different degrees of spread (dispersion), asymmetry (skewness), and tail behavior (kurtosis), the model ensures adaptability across multiple real-world scenarios. In sustainability-focused data applications, accurately modeling dispersion is essential for understanding variability in environmental studies, medical research, and engineering systems. For instance, in medical data, capturing varying dispersion levels helps improve predictive accuracy in disease progression and treatment outcomes. Similarly, in engineering applications, a flexible probability model allows for better failure time predictions, contributing to efficient resource utilization and sustainable system design. The ability to model both positively skewed and symmetric data further enhances the DGzExGc distribution's applicability, ensuring that sustainable data-driven decision-making remains reliable across different disciplines. Additionally, the influence of α and β on kurtosis ensures that datasets with either heavy tails (leptokurtic) or flat tails (platykurtic) can be accurately represented. This flexibility is particularly useful in financial and risk analysis, where extreme events must be accounted for, as well as in environmental studies, where extreme weather conditions or ecological changes impact long-term sustainability planning. By leveraging the adaptability of the DGzExGc model, sustainable data analysis can be significantly enhanced, ensuring efficient and accurate decision-making processes in fields where data complexity continues to grow.

3.2. Rényi 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. In various fields, such as econometrics, quantum information, information theory, survival analysis, and computer science, entropy plays a crucial role in understanding and managing uncertainty (Rényi, [30]). For example, in survival analysis, entropy can help assess the uncertainty in the lifetime of a product or the time to an event. Similarly, in information theory, it is used to measure the amount of information in a signal or data set. The variation in the uncertainty of a random variable Z can be expressed through its entropy, which captures how spread out or concentrated the probability distribution is. Specifically, entropy is calculated as the negative sum of the probabilities of all possible outcomes, weighted by the logarithm of those probabilities. This measure is particularly valuable in decision-making processes, where understanding the level of uncertainty can guide better choices and predictions. The measure of variation of the uncertainty of the random variable Z can be expressed as

$$
\begin{aligned}
I_\eta(Z) &= \frac{1}{1-\eta}\log\sum_{z=0}^{\infty} f_z^{\eta}(z;\nu,\rho,\alpha,\beta) \\
&= \frac{1}{1-\eta}\left\{-\frac{\eta}{\rho}\log\nu + \log\sum_{z=0}^{\infty}\left[\nu^{\frac{1}{\rho}\left[1-[1-(1-\alpha)^{z}]^{\beta}\right]^{-\rho}} - \nu^{\frac{1}{\rho}\left[1-[1-(1-\alpha)^{z+1}]^{\beta}\right]^{-\rho}}\right]^{\eta}\right\}, \quad (19)
\end{aligned}
$$

where $\eta \in]0,\infty[$ and $\eta \neq 1$. The Shannon entropy can be defined by $E\left[-\log f(Z;\nu,\rho,\alpha,\beta)\right]$. It is observed that the Shannon entropy can be calculated as a special case of the Rényi entropy when $\eta \to 1$.

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

In the context of discrete random variables, the DGzExGc model provides a framework for evaluating the performance of systems based on reliability metrics such as Mean Time to Failure (MTTF), Mean Time Between Failures (MTBF), and Availability (Av). MTTF refers to the average time until the first failure of a non-repairable system, and in the case of the DGzExGc model, it can be estimated by the expected failure time derived from the distribution's parameters. MTBF is used for repairable systems, indicating the average time between two successive failures, and the DGzExGc model helps estimate this by analyzing the failure rates over multiple cycles. If $T \sim \text{DGzExGc}(t;\nu_1,\rho_1,\alpha_1,\beta_1)$, then the MTBF is given as

$$
\text{MTBF}(T) = \frac{-t}{\ln\left(\nu_1^{\frac{1}{\rho_1}\left\{\left[1-[1-(1-\alpha_1)^{t+1}]^{\beta_1}\right]^{-\rho_1}-1\right\}}\right)};\ t>0, \quad (20)
$$

whereas if $T \sim \text{DGzExGc}(t;\nu_2,\rho_2,\alpha_2,\beta_2)$, then the MTTF can be listed as

$$
\text{MTTF}(T) = \nu_2^{-\frac{1}{\rho_2}}\sum_{z=1}^{\infty}\nu_2^{\frac{1}{\rho_2}\left[1-[1-(1-\alpha_2)^{z}]^{\beta_2}\right]^{-\rho_2}};\ t>0. \quad (21)
$$

On the other hand, availability measures the proportion of time a system is functioning, taking into account both the time it is operating and the downtime due to repairs. For systems described by the DGzExGc distribution, availability can be evaluated by looking at the system's uptime and the time between failures, which are both influenced by the model's parameters.

$$\mathrm{Av}(T) = \frac{-1}{t}\left[\nu_2^{-\frac{1}{\rho_2}}\sum_{z=1}^{\infty}\nu_2^{\frac{1}{\rho_2}\left[1-[1-(1-\alpha_2)^z]^{\beta_2}\right]^{-\rho_2}}\right]\left[\ln(\nu_1^{\frac{1}{\rho_1}\left\{\left[1-[1-(1-\alpha_1)^{t+1}]^{\beta_1}\right]^{-\rho_1}-1\right\}})\right];\ t>0.$$

When applied within the DGzExGc framework, these metrics offer a thorough understanding of a system's reliability and performance, especially in situations where failure times are discrete and repairs are involved. Availability is essentially the probability that the component is operating at a given time, and it can be expressed as the ratio of MTTF to MTBF.

3.4. Order statistics: The PMF, moments, 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 DGzExGc$(z;\rho,\nu,\alpha,\beta)$ 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

$$\begin{aligned} F_{i:n}(z;\nu,\rho,\alpha,\beta) &= \sum_{k=i}^{n}\binom{n}{k}\left[F_i(z;\nu,\rho,\alpha,\beta)\right]^k\left[1-F_i(z;\nu,\rho,\alpha,\beta)\right]^{n-k} \\ &= \sum_{k=i}^{n}\sum_{j=0}^{n-k}(-1)^j\binom{n}{k}\binom{n-k}{j}\left[F_i(z;\nu,\rho,\alpha,\beta)\right]^{k+j} \\ &= \sum_{k=i}^{n}\sum_{j=0}^{n-k}\sum_{m=0}^{k+j}\Delta_{(n,k)}^{(m,j)}F(z;\rho,\nu^m,\alpha,\beta), \end{aligned} \tag{22}$$

where $\Delta_{(n,k)}^{(m,j)} = (-1)^{j+m}\binom{n}{k}\binom{n-k}{j}\binom{k+j}{m}$. The corresponding PMF of the ith OS can be expressed as

$$f_{i:n}(z;\nu,\rho,\mathbf{\Upsilon}) = \sum_{k=i}^{n}\sum_{j=0}^{n-k}\sum_{m=0}^{k+j}\Delta_{(n,k)}^{(m,j)}f(z;\nu^m,\rho,\alpha,\beta). \tag{23}$$

The uth moment of $Z_{i:n}$ can be written as

$$\Psi_{i:n}^{u} = \mathbf{E}(Z_{i:n}^{u}) = \sum_{z=0}^{\infty}\sum_{k=i}^{n}\sum_{j=0}^{n-k}\sum_{m=0}^{k+j}\Delta_{(n,k)}^{(m,j)}z^u f(z;\nu^m,\rho,\alpha,\beta). \tag{24}$$

L-moments (LMs) are derived as linear combinations of order statistics (OS). Hosking and Wallis [22] introduced LMs to summarize both theoretical distributions and observed samples. LM statistics are commonly used to compute sample statistics for data from specific regions or to test 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

$$\lambda_r^* = \frac{1}{r}\sum_{s=0}^{r-1}(-1)^s \begin{pmatrix} r-1 \\ s \end{pmatrix} \mathbf{E}\left(Z_{r-s:r}\right). \tag{25}$$

From Equation (25), 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 $\lambda_4^* = \frac{1}{4}\left\{\mathbf{E}\left[\left(Z_{4:4}-Z_{3:4}\right)+\left(Z_{2:4}-Z_{1:4}\right)\right]-2\mathbf{E}\left(Z_{3:4}-Z_{2:4}\right)\right\}$. 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.

4. Maximum Likelihood Estimation (MLE)

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

$$L = -\frac{1}{\rho}\ln(\nu) + \sum_{i=1}^{n}\ln\left(\nu^{\frac{1}{\rho}\left\{\left[1-[1-(1-\alpha)^{z_i}]^{\beta}\right]^{-\rho}\right\}} - \nu^{\frac{1}{\rho}\left\{\left[1-[1-(1-\alpha)^{z_i+1}]^{\beta}\right]^{-\rho}\right\}}\right). \tag{26}$$

The MLEs of the parameters ν, ρ, α and β can be derived by solving the nonlinear likelihood equations obtained by differentiating (26) with respect to ν, ρ, α and β. The components of the score vector, $\mathbf{V}(\nu,\rho,\boldsymbol{\Upsilon}) = (\frac{\partial L}{\partial \nu}, \frac{\partial L}{\partial \rho}, \frac{\partial L}{\partial \boldsymbol{\Upsilon}})^T$, where $\boldsymbol{\Upsilon} = (\alpha,\beta)^T$ are

$$V_\nu = \frac{-n}{\rho\nu} + \frac{1}{\rho\nu}\sum_{i=1}^{n}\frac{g_2(z_i)-g_2(z_i+1)}{g_1(z_i)}, \tag{27}$$

$$V_\rho = \frac{-n\ln(\nu)}{\rho^2} - \frac{\ln(\nu)}{\rho^2}\sum_{i=1}^{n}\frac{g_2(z_i)\left[\rho\ln(1-[1-(1-\alpha)^{z_i}]^{\beta})+1\right]-g_2(z_i+1)\left[\rho\ln(1-[1-(1-\alpha)^{z_i+1}]^{\beta})+1\right]}{g_1(z_i)} \tag{28}$$

$$V_{\boldsymbol{\Upsilon}_j} = \sum_{i=1}^{n}\frac{g_2(z_i)\left[1-[1-(1-\alpha)^{z_i}]^{\beta}\right]^{-1}[G(z_i;\boldsymbol{\Upsilon})]_{\boldsymbol{\Upsilon}_j} - g_2(z_i+1)\left[1-[1-(1-\alpha)^{z_i+1}]^{\beta}\right]^{-1}[G(z_i+1;\boldsymbol{\Upsilon})]_{\boldsymbol{\Upsilon}_j}}{g_1(z_i)} \tag{29}$$

where

$$[G(z_i;\boldsymbol{\Upsilon})]_{\boldsymbol{\Upsilon}_j} = \frac{\partial}{\partial \boldsymbol{\Upsilon}_j}[1-(1-\alpha)^{z_i}]^{\beta};\ \boldsymbol{\Upsilon}_j = \alpha,\beta,$$

$$g_1(z_i) = \nu^{\frac{1}{\rho}\left\{\left[1-[1-(1-\alpha)^{z_i}]^{\beta}\right]^{-\rho}\right\}} - \nu^{\frac{1}{\rho}\left\{\left[1-[1-(1-\alpha)^{z_i+1}]^{\beta}\right]^{-\rho}\right\}},$$

and

$$g_2(z_i) = \nu^{\frac{1}{\rho}\left\{\left[1-[1-(1-\alpha)^{z_i}]^{\beta}\right]^{-\rho}\right\}}\left\{\left[1-[1-(1-\alpha)^{z_i}]^{\beta}\right]^{-\rho}\right\}.$$

The MLE for the DGzExGc model parameters is obtained by setting the first-order partial derivatives of the log-likelihood function to zero and solving the resulting system of nonlinear equations. These equations, denoted as(27-29), define the conditions under which the likelihood function reaches its peak, ensuring that the parameter estimates maximize the probability of observing the given data. In real-world applications, solving these nonlinear equations analytically is often impractical, making numerical optimization techniques essential. The R programming environment provides powerful tools for this purpose, particularly through the stats and optim packages. These packages offer iterative algorithms such as the Newton-Raphson method, which updates parameter estimates using the Hessian matrix and the gradient of the likelihood function.

Implementing these techniques in R involves defining the log-likelihood function, applying numerical optimization with functions like **optim** or **nlm**, and assessing convergence to ensure reliable parameter estimates. This process is crucial for accurately modeling complex datasets across various disciplines. Sustainable data analysis greatly benefits from efficient parameter estimation in advanced probability models like DGzExGc. In medicine, precise MLEs enhance survival analysis and hazard rate modeling, supporting better patient prognosis and treatment planning. In engineering, accurate failure time modeling enables predictive maintenance, optimizing resource allocation and improving infrastructure sustainability. By integrating advanced statistical methodologies with computational tools, sustainable data analysis ensures more precise and interpretable models, reinforcing the DGzExGc distribution's value in tackling complex real-world challenges.

5. Simulation Results: Estimator Behavior

This section presents the results from simulations to assess the behavior of the estimators for the parameters of the DGzExGc model. By generating synthetic data based on known parameters, we examine how well the maximum likelihood estimates (MLEs) perform in terms of bias, variance, and efficiency. Various sample sizes and different values of the model parameters are considered to evaluate the robustness and consistency of the estimators. The results help to understand the accuracy and reliability of the MLEs in practical applications. The assessment is based on a simulation study:

(i) Generate 1000 samples of size $n = 25, 50, 150, 300, 500$ from two scheme, Schema I: DGzExGc$(0.1, 0.5, 0.5, 0.8)$ and Schema II: DGzExGc$(0.3, 0.7, 0.8, 0.9)$.

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

(iii) Compute the biases, mean-squared errors (MSEs), and mean relative error (MRE) where

$$\text{bias}(\zeta) = \frac{1}{1000}\sum_{j=1}^{1000}\left(\widehat{\zeta}_j - \zeta\right), \ \text{MSE}(\zeta) = \frac{1}{1000}\sum_{j=1}^{1000}\left(\widehat{\zeta}_j - \zeta\right)^2, \ \text{MRE}(\zeta) = \frac{1}{1000}\sum_{j=1}^{1000}\left|\frac{\widehat{\zeta}_j - \zeta}{\zeta}\right|.$$

$$\text{MRE}(\zeta) = \frac{1}{1000}\sum_{j=1}^{1000}\left|\frac{\widehat{\zeta}_j - \zeta}{\zeta}\right|.$$

The empirical results are shown in Figures 3-6.

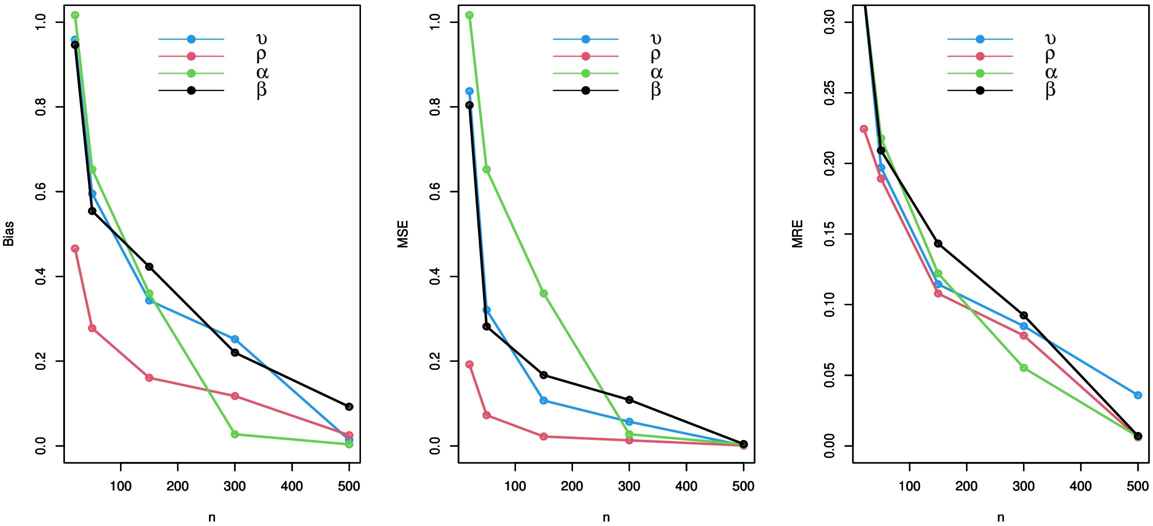

Figure 3. Simulation results for Schema I.

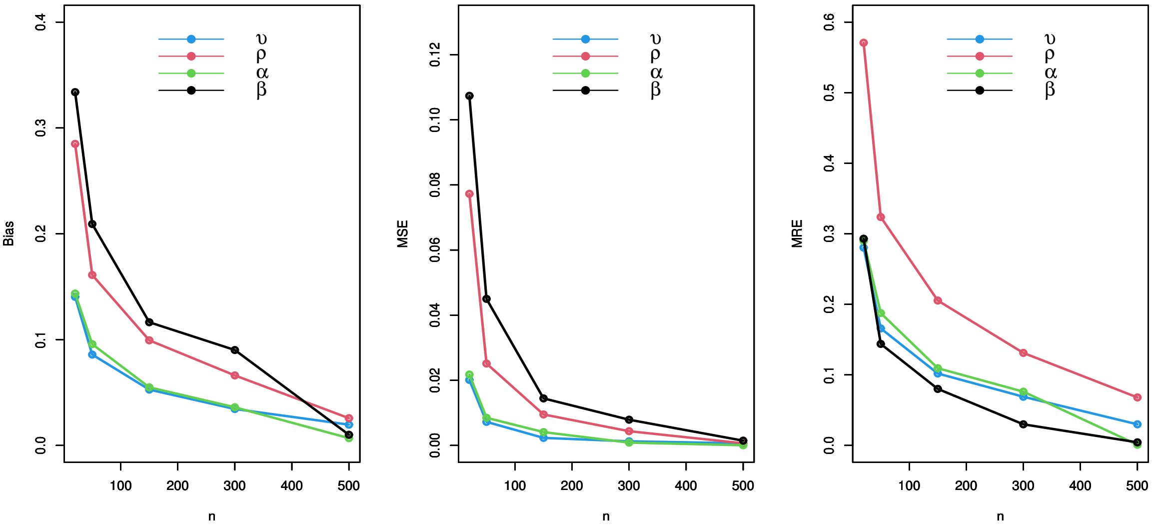

Figure 4. Simulation results for Schema II.

The simulation results demonstrate the excellent performance of the estimators in terms of bias, MSE, and MRE for large sample sizes. In both Schema I and Schema II, the bias of the estimators decreases as the sample size increases, with the bias values approaching zero, which indicates that the estimators become unbiased as the sample size grows. In Schema I, the estimators show minimal bias across all sample sizes. Both MSE and MRE steadily decrease as the sample size n increases, reflecting a marked improvement in the accuracy and precision of the estimators for larger datasets. This trend suggests that as the sample size grows, the estimators become more reliable and provide more accurate estimates. Similarly, in Schema II, the bias remains small and decreases further with increasing sample sizes, underscoring the robustness of the model's estimators. MSE and MRE also demonstrate a clear reduction as n increases, indicating that the estimators become more efficient and exhibit better performance with larger samples.

Overall, in both schemas, the estimators perform exceptionally well with low bias, MSE, and MRE, particularly as the sample sizes increase. These findings highlight the effectiveness and reliability of the MLE method when applied to the DGzExGc model, confirming its suitability for practical applications.

6. Sustainable Data Analysis

Sustainable data practices are essential in both medicine and engineering, significantly enhancing decision-making, resource optimization, and long-term advancements. In medicine, sustainable data management improves patient care and healthcare efficiency through predictive analytics, AI-driven diagnostics, and optimized clinical trials, while also strengthening public health surveillance and epidemic response. In engineering, sustainable data enables the development of smart infrastructure, predictive maintenance strategies, renewable energy management, and industrial process optimization, leading to increased efficiency and reduced environmental impact. By ensuring the ethical, efficient, and long-term use of data, both fields can foster innovation, minimize waste, and support sustainable development, ultimately benefiting both society and the environment.

In this section, the empirical relevance of the DGzExGc distribution is demonstrated through three real-world applications in sustainable data analysis within the fields of medicine and engineering. These applications highlight the model's flexibility and effectiveness in addressing complex data structures and improving decision-making in diverse scientific and technological domains. The fitted models are compared using some criteria, namely, L, Akaike information criterion (AIC), correct Akaike information criterion (CAIC), Chi-square (χ^2) with degree of freedom (d.f) and its P-value, Kolmogorov-Smirnov (K-S) and its P-value. We shall compare the DGzExGc distribution with some competitive models provided in Table 2.

Table 2. The competitive models of the DGzExGc distribution.

Distribution	Abbreviation
Discrete Weibull	DW
Exponentiated discrete Weibull	EDW
Discrete inverse Weibull	DIW
Discrete exponential	DEx
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 log-logistic	DLLc
Discrete Lomax	DLo
Two- parameter discrete Burr type XII	DB-XII
Discrete Pareto	DPa
Discrete Gompertz Weibull	DGzW
Poisson	Poi

6.1. Data set I

In the context of sustainability, sustainable data analysis is fundamental in reliability engineering and failure time studies, as it enables the accurate modeling of failure behavior, which directly influences the longevity and efficiency of systems. By understanding how devices or components fail over time, engineers can optimize designs and maintenance schedules to extend product life and reduce waste. Data set I, which represents the failure times (in weeks) of 50 devices subjected to a life test (see Bebbington et al., [6]), offers crucial insights into the durability and performance of these devices under controlled conditions. In light of sustainability, this data is pivotal for making informed decisions that enhance product reliability and efficiency. To evaluate the best fit for this failure time data, the DGzExGc distribution will be compared with several competitive models. Non-parametric plots, presented in Figure 5, illustrate the behavior of the data, providing a visual understanding of the failure patterns and helping identify the most appropriate model for accurate analysis and sustainable decision-making.

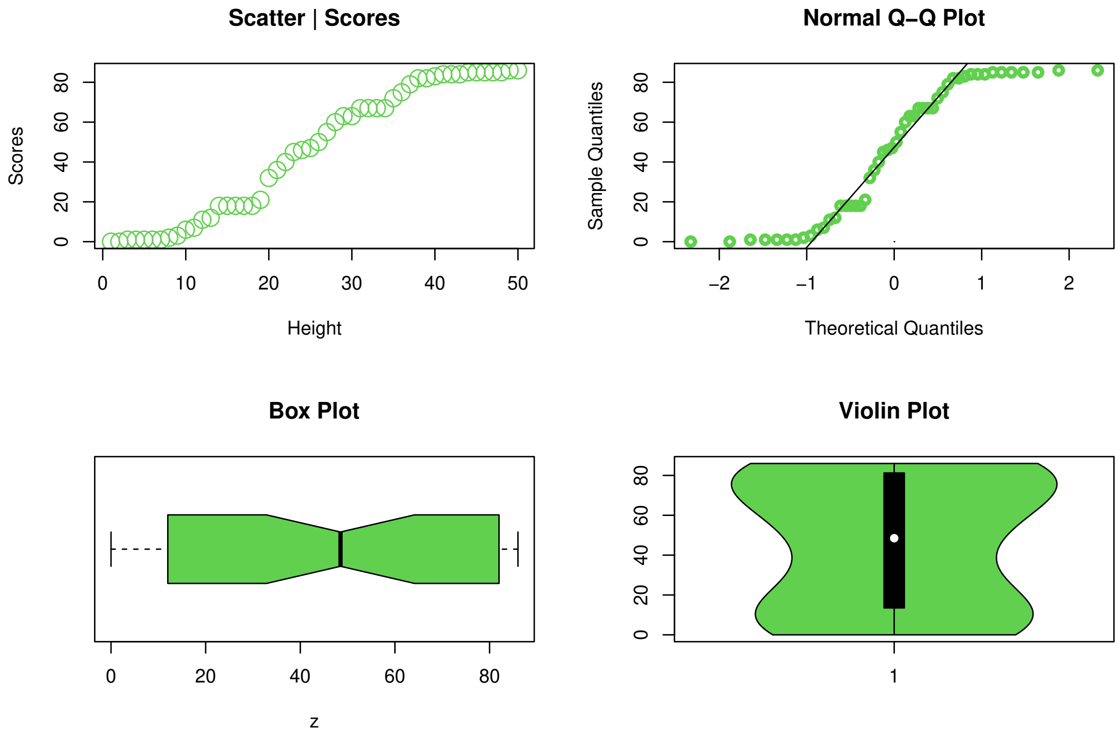

Figure 5. Non-parmetric plots for dataset I.

The MLEs with their corresponding standard errors (Std-er), and the goodness of fit statistics are reported in Tables 3 and 4, respectively.

Table 3. The MLEs with their corresponding Std-er for data set I.

Model ↓ Parameter →	ν		ρ		α		β	
	MLE	Std-er	MLE	Std-er	MLE	Std-er	MLE	Std-er
DGzExGc	0.962	0.037	0.5087	0.386	0.053	0.039	0.125	0.171
DGzW	0.938	0.444	0.499	3.709	0.364	2.683	0.620	0.163
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 4. The goodness-of-fit statistics for data set I.

Statistic ↓ Model →	DGzExGc	DGzW	EDW	DW	DIW	DLi-II	EDLi	DLLc	DPa
$-L$	229.6	233.1	240.2	241.6	261.9	240.6	240.3	294.9	275.9
AIC	467.3	474.1	486.7	487.2	527.8	485.2	484.6	593.8	553.7
CAIC	468.1	474.9	487.2	487.5	528.1	485.4	484.8	594.0	553.8
K-S	0.136	0.161	0.195	0.187	0.258	0.186	0.195	0.535	0.335
P-value	0.311	0.149	0.045	0.061	0.0026	0.064	0.045	< 0.001	< 0.001

In reference to Table 4, the analysis of sustainability-related failure time data reveals that the DGzExGc, DGzW, EDW, DW, DLi-II, and EDLi models effectively capture the underlying failure patterns. Among these, the DGzExGc model stands out, demonstrating the best fit based on statistical criteria such as the $-L$, AIC, CAIC, K-S statistics, and P-values. The superiority of the DGzExGc model, indicated by its lowest $-L$, AIC, CAIC, and K-S statistics along with the highest P-value, suggests its robustness in evaluating failure time data, which is crucial for sustainable system performance and risk management. The importance of selecting the optimal model in failure time analysis extends to various sustainability-driven applications, including predictive maintenance, resource optimization, and lifespan estimation of materials and infrastructure. By providing a reliable statistical framework, the DGzExGc model enhances decision-making in industries such as renewable energy, healthcare, and transportation, where efficient failure prediction and system sustainability are paramount. Furthermore, Figure 6, along with the HRF plot, further corroborates the findings presented in Table 4, reinforcing the DGzExGc model's effectiveness in analyzing failure time data for sustainable system development.

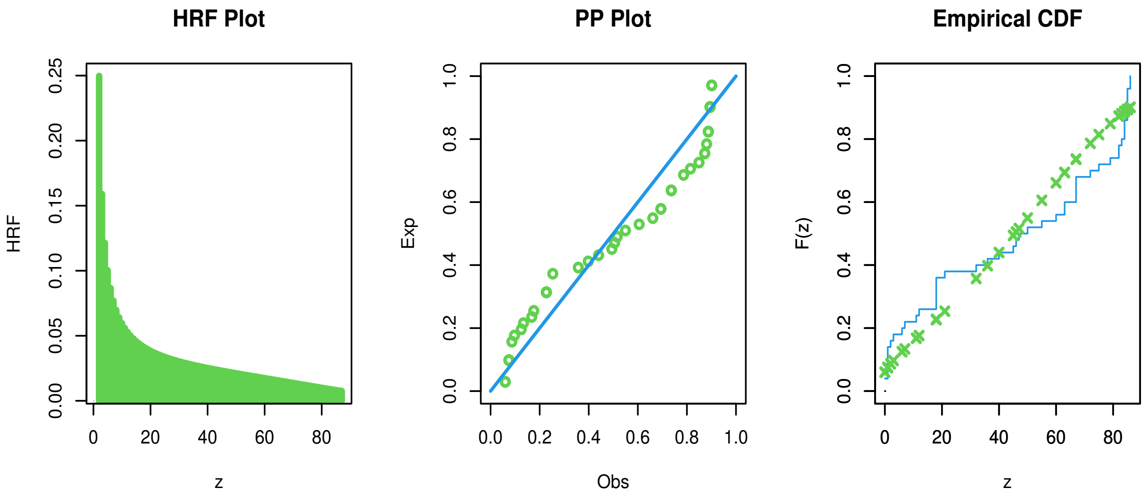

Figure 6. The HRF, PP plot, estimated CDFs for dataset I.

Based on the probability-probability (PP) plot and the empirical CDF, it is evident that the dataset likely originates from the DGzExGc model. The PP plot shows a close alignment between the observed and expected values under the DGzExGc model, indicating a good fit. Additionally, the empirical CDF closely follows the theoretical CDF of the DGzExGc model, further supporting the plausibility that the data comes from this model.

6.2. Data set II

In the context of sustainability, failure time data, such as the one reported in Lawless [24], plays a significant role in assessing the reliability and longevity of electronic components. The data, which pertains to the failure times of 15 electronic components in an acceleration life test, provides valuable insights into the performance and durability of these components under accelerated conditions. This is crucial for improving the sustainability of electronic products by enabling manufacturers to predict component lifespans, optimize maintenance schedules, and enhance resource efficiency. Non-parametric plots, presented in Figure 7, help visualize the failure patterns and variations in the data, which are essential for understanding the sustainability of the components in real-world applications. These plots allow for a more flexible analysis, without assuming a specific parametric form, which is particularly useful when the data exhibits complex or unknown underlying distributions. By effectively analyzing such failure time data, sustainability in engineering can be promoted by improving product design, reducing waste, and increasing the lifecycle of electronic components.

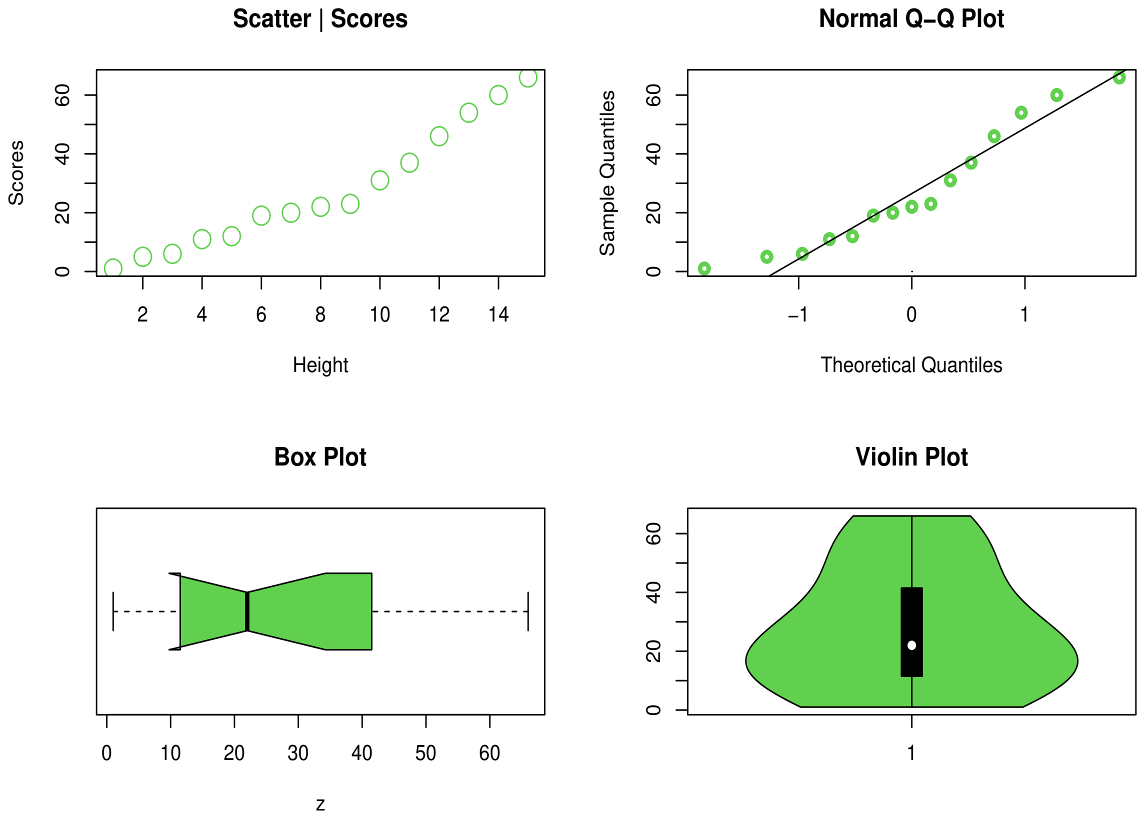

Figure 7. Non-parametric plots for dataset II.

For this data set, we shall compare the fits of the DGzExGc distribution with some competitive models. The MLEs with their corresponding Std-er, and the goodness of fit statistics are reported in Tables 5 and 6, respectively.

graphicx

Table 5. The MLEs with their corresponding Std-er for data set II.

Model ↓ Parameter →	ν		ρ		α		β	
	MLE	Std-er	MLE	Std-er	MLE	Std-er	MLE	Std-er
DGzExGc	0.554	0.083	0.616	0.883	0.035	0.437	1.028	0.565
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 6. The goodness-of-fit statistics for data set II.

	Model								
Statistic	DGzExGc	DEx	DGEx-II	DR	DIR	DIW	DLo	DB-XII	DPa
$-L$	63.802	65.000	64.420	66.394	89.096	68.703	65.864	75.724	77.402
AIC	135.606	134.000	134.839	134.788	180.192	141.406	135.728	155.448	156.805
CAIC	139.605	136.308	135.839	136.096	180.499	142.406	136.728	156.448	157.112
K-S	0.119	0.177	0.129	0.216	0.698	0.209	0.205	0.388	0.405
P-value	0.964	0.673	0.937	0.433	9.1×10^{-7}	0.482	0.491	0.015	0.009

Regarding Table 6, it is clear that the DGzExGc, DEx, DGEx-II, DR, DIW, and DLo models perform quite well for analyzing the data, with the exception of the DGzExGc model, which exhibits some differences. However, the DGzExGc distribution emerges as the best model among all the tested models. This is evident from its superior fit to the data, as indicated by various model evaluation criteria. Figure 8 further supports the findings presented in Table 6, providing a visual representation of the model's performance and reinforcing the conclusion that the DGzExGc distribution is the most suitable model for this dataset.

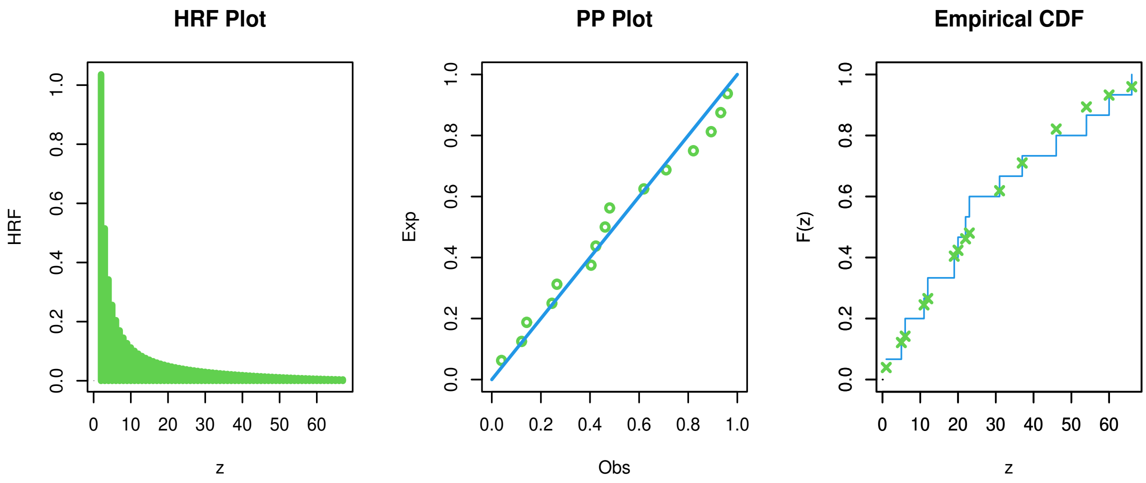

Figure 8. The HRF, PP plot, estimated CDFs for dataset II.

The analysis of both the PP plot and the empirical CDF strongly suggests that the dataset is likely derived from the DGzExGc model. The PP plot demonstrates a near-perfect alignment between the observed and expected values under this model, which indicates a good fit. Similarly, the empirical CDF closely mirrors the theoretical CDF of the DGzExGc model, providing additional evidence that the data is consistent with this distribution. These findings collectively reinforce the conclusion that the DGzExGc model is a plausible source for the dataset.

6.3. Data set III

This dataset represents the count of kidney cysts in patients undergoing steroid treatment, originally sourced from the study by Chan et al. [8]. In the context of sustainability, analyzing such medical data is crucial for optimizing treatment strategies, minimizing adverse effects, and improving long-term patient outcomes. Sustainable data-driven approaches in healthcare ensure that medical resources are efficiently utilized while enhancing diagnostic and therapeutic decisions. Non-parametric plots, presented in Figure 9, provide insights into the distribution and structure of the data, facilitating a deeper understanding of its variability and aiding in the development of sustainable healthcare interventions.

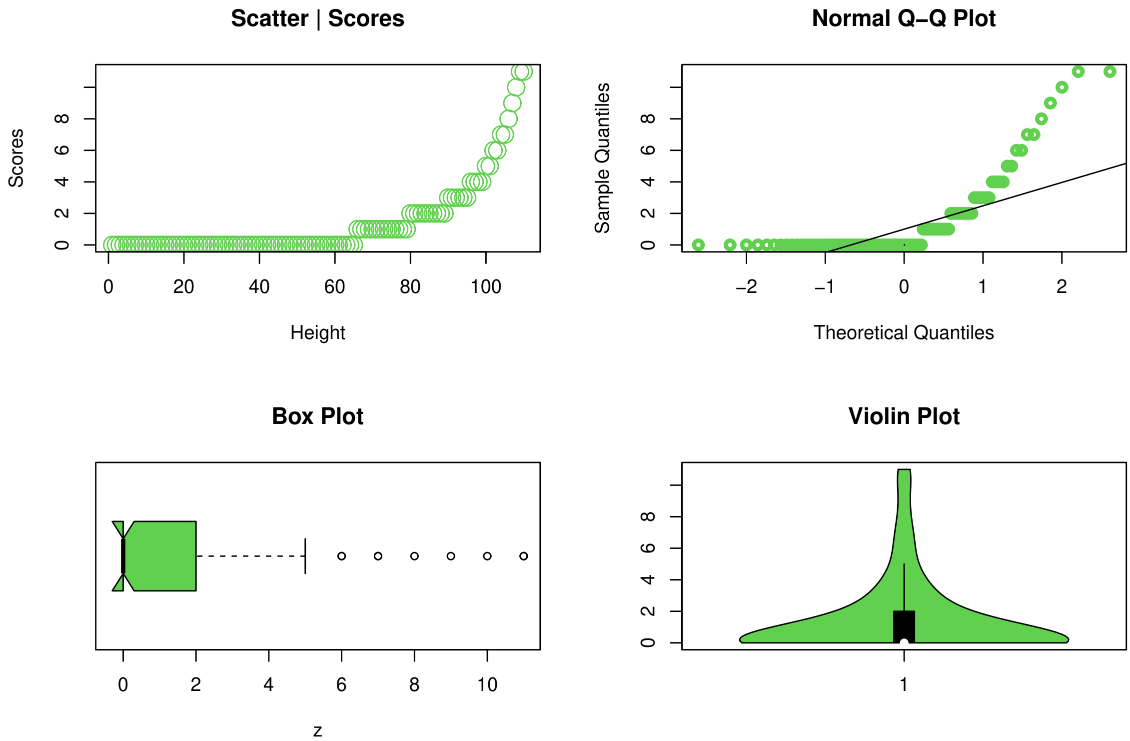

Figure 9. Non-parmetric plots for dataset III.

For this data set, we shall compare the fits of the DGzExGc distribution with some competitive models. The MLEs with their corresponding Std-er, and the goodness of fit statistics are reported in Tables 7 and 8, respectively.

Table 7. The MLEs with their corresponding Std-er for data set III.

Model ↓ Parameter →	ν		ρ		α		β	
	MLE	Std-er	MLE	Std-er	MLE	Std-er	MLE	Std-er
DGzExGc	0.449	0.577	0.474	2.151	0.122	0.501	0.248	0.238
DW	–	–	–	–	0.750	0.084	0.431	0.340
DIW	–	–	–	–	0.581	0.048	1.049	0.146
DR	–	–	–	–	0.901	0.009	–	–
DEx	–	–	–	–	0.581	0.030	–	–
DLi	–	–	–	–	0.436	0.026	–	–
DLi-II	–	–	–	–	0.581	0.045	0.001	0.058
DLo	–	–	–	–	0.150	0.098	1.830	0.951
Poi	–	–	–	–	1.390	0.112	–	–

Table 8. The goodness-of-fit statistics for data set III.

Z	Observed Frequency	Expected Frequency								
		DGzExGc	DW	DIW	DR	DEx	DLi	DLi-II	DLo	Poi
0	65	65.08	59.01	63.91	11.00	46.09	40.25	46.03	61.89	27.42
1	14	13.91	19.84	20.70	26.83	26.78	29.83	26.77	21.01	38.08
2	10	8.76	10.78	8.05	29.55	15.56	18.36	15.57	9.65	26.47
3	6	6.13	6.26	4.23	22.23	9.04	10.35	9.05	5.24	12.26
4	4	4.45	4.19	2.60	12.49	5.25	5.53	5.27	3.17	4.26
5	2	3.28	2.01	1.75	5.42	3.05	2.86	3.06	2.06	1.18
6	2	2.42	1.99	1.26	1.85	1.77	1.44	1.78	1.42	0.27
7	2	1.78	1.32	0.95	0.52	1.03	0.71	1.04	1.02	0.05
8	1	1.29	0.99	0.74	0.11	0.60	0.35	0.60	0.76	0.01
9	1	0.93	0.86	0.59	0.02	0.35	0.17	0.35	0.58	0.00
10	1	0.66	0.76	0.48	0.00	0.20	0.08	0.20	0.46	0.00
11	2	1.31	1.99	4.74	0.00	0.28	0.07	0.28	2.74	0.00
Total	**110**	**110**	**110**	**110**	**110**	**110**	**110**	**110**	**110**	**110**
$-L$		166.73	170.14	172.93	277.78	178.77	189.1	178.8	170.48	246.21
AIC		341.47	344.28	349.87	557.56	359.53	380.2	361.5	344.96	494.42
CAIC		341.85	344.39	349.98	557.59	359.57	380.3	361.6	345.07	494.46
χ^2		0.610	3.125	6.463	321.07	22.88	43.48	22.89	3.316	294.10
d.f		1	3	3	4	4	4	3	3	4
P.value		0.435	0.373	0.091	< 0.0001	0.0001	< 0.0001	< 0.0001	0.345	< 0.0001

Table 8 highlights the importance of selecting an appropriate model for sustainable data analysis, particularly in applications where accurate predictions and resource optimization are essential. The DGzExGc, DW, DIW, and DLo models exhibit strong performance in analyzing this dataset; however, the DGzExGc model emerges as the most effective choice. Based on key model evaluation criteria, including lower values for $-L$, AIC, CAIC, and χ^2 statistics, as well as higher P-values, the DGzExGc model provides the best fit. This outcome underscores the role of sustainability in statistical modeling, as selecting the most efficient model ensures better data-driven decision-making, minimizes computational waste, and enhances predictive accuracy. Figures 10 and 11 further support the findings in Table 8, offering visual confirmation of the model's superior fit. By adopting robust statistical approaches like the DGzExGc model, sustainable data analysis contributes to improved reliability, efficiency, and long-term applicability in various fields, including engineering, medicine, and environmental studies.

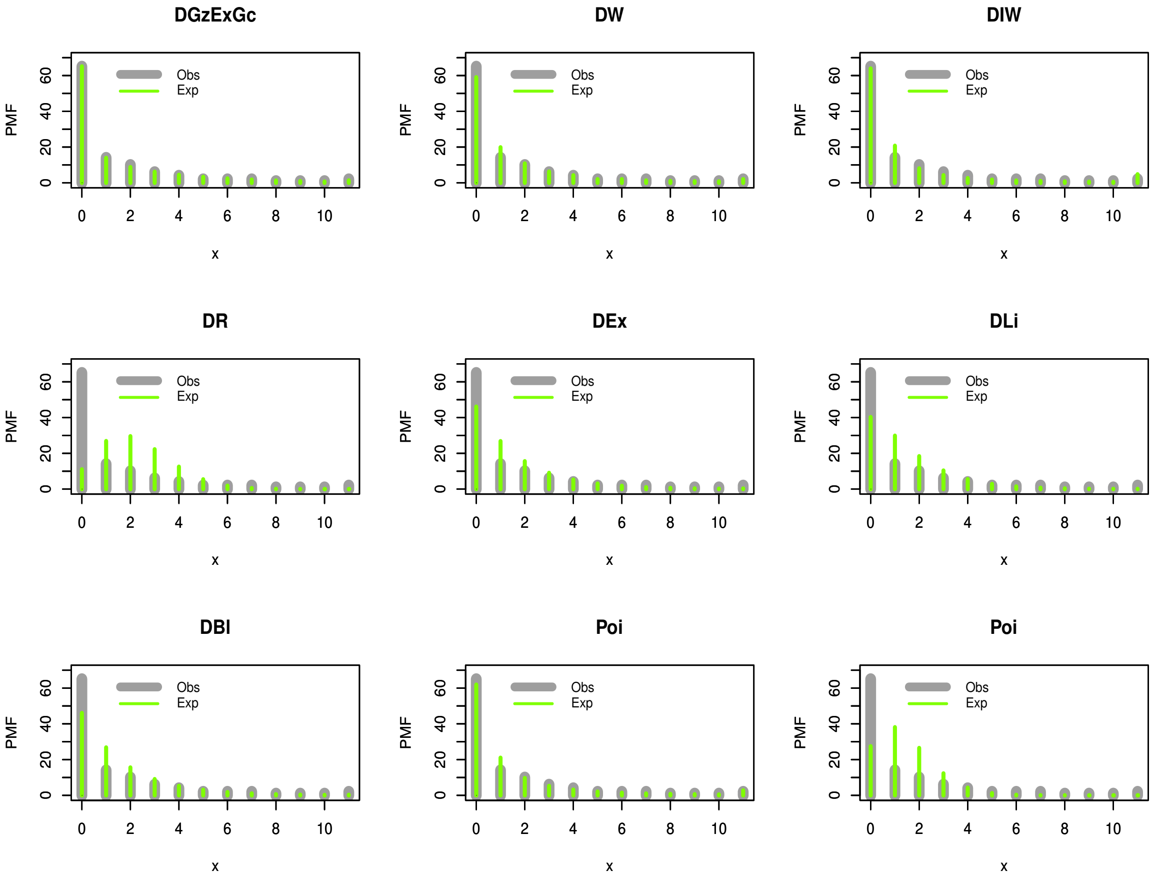

Figure 10. The fitted PMFs for dataset III.

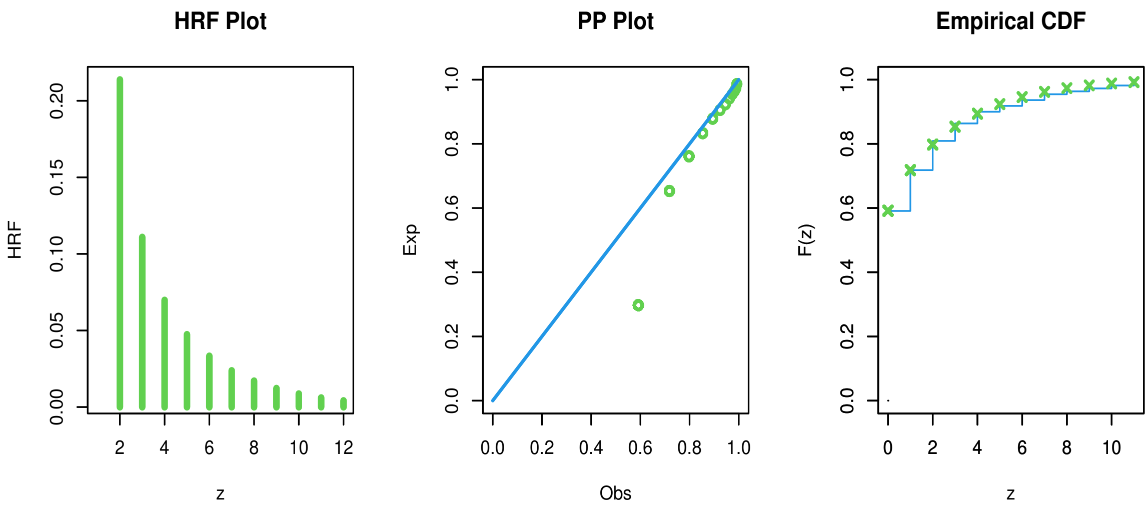

Figure 11. The HRF, PP plot, estimated CDFs for dataset III.

The analysis of both the PP plot and the empirical CDF provides compelling evidence that the dataset is likely derived from the DGzExGc model. The PP plot reveals an almost perfect alignment between the observed and expected values under this model, which strongly indicates that the model fits the data well. This close alignment suggests that the DGzExGc model accurately captures the underlying distribution of the dataset. Additionally, the empirical CDF, which represents the cumulative distribution of the observed data, closely follows the theoretical CDF of the DGzExGc model. This further corroborates the model's suitability, as it shows a high degree of consistency between the theoretical and observed data distributions. Together, these analyses strongly support the conclusion that the DGzExGc model is a plausible source for the dataset, providing a reliable and accurate representation of the data.

7. Conclusion and Perspectives

The DGzExGc model presents a highly flexible and adaptable framework for sustainable count data analysis, making it a valuable tool for diverse applications. Its ability to accommodate both positively skewed and symmetric distributions, as well as unimodal and bimodal structures, ensures comprehensive applicability across various domains. This model is particularly suited for sustainable data-driven decision-making, as it can effectively capture different hazard rate functions, including increasing, decreasing, bathtub-shaped, and increasing-constant trends, making it ideal for reliability assessments in medicine, engineering, and environmental studies. A critical aspect of sustainability in data analysis is the ability to handle varying dispersion levels, and the DGzExGc model excels in modeling equi-dispersed, over-dispersed, and under-dispersed count data. Additionally, its capability to represent different kurtosis levels ranging from leptokurtic (heavy-tailed) to platykurtic (flat-tailed) distributions demonstrates its robustness in capturing diverse data structures. These features make the model suitable for long-term, efficient data management and analysis, supporting sustainability in both predictive analytics and resource optimization. Moreover, the DGzExGc model proves effective in modeling both count and lifetime data, contributing to advancements in medical diagnostics, failure time analysis in engineering, and other data-intensive fields. The maximum likelihood estimation method has been successfully applied for parameter estimation, and simulation studies confirm the model's reliability in practical scenarios. The versatility of the DGzExGc model has been validated through its application to three distinct datasets, emphasizing its relevance for sustainable data analysis and its potential for future research in data-driven sustainability initiatives. As future work, various estimation techniques will be explored, along with simulation studies for censored datasets.

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

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-2-2025-52604) during the academic year 1446 AH / 2024 AD.

References

[1] M. Ahsan-ul-Haq, A. Al-Bossly, M. El-Morshedy, and M. S. Eliwa. Poisson XLindley distribution for count data: statistical and reliability properties with estimation techniques and inference. *Computational Intelligence and Neuroscience*, 2022(1):6503670, 2022.

[2] T. Akkanphudit. The discrete Gompertz–Weibull–Fréchet distribution: Properties and applications. *Lobachevskii Journal of Mathematics*, 44(9):3663–3672, 2023.

[3] 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(1):281–308, 2020.

[4] M. Alizadeh, G. M. Cordeiro, G. B. Luis, and I. Ghosh. The Gompertz-G family of distributions. *Journal of Statistical Theory and Practice*, 11(1):179–207, 2017.

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

[6] 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.

[7] A. C. Bemmaor and N. Glady. Modeling purchasing behavior with sudden 'death': A flexible customer lifetime model. *Management Science*, 58(5):1012–1021, 2012.

[8] S. Chan, P. R. Riley, K. L. Price, F. McElduff, and P. J. Winyard. Corticosteroid-induced kidney dysmorphogenesis is associated with deregulated expression of known cystogenic molecules, as well as Indian hedgehog. *American Journal of Physiology–Renal Physiology*, 298(2):F346–F356, 2009.

[9] K. Chandrakant, M. T. Yogesh, and M. K. Rathi. On a discrete analogue of linear failure rate distribution. *American Journal of Mathematical and Management Sciences*, 36(3):229–246, 2017.

[10] W. Chuncharoenkit, W. Bodhisuwan, and S. Aryuyuen. Discrete Gompertz-Lomax distribution and its applications. *Thailand Statistician*, 22(4):832–855, 2024.

[11] 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(2):29–50, 2016.

[12] J. T. Eghwerido and F. I. Agu. The shifted Gompertz-G family of distributions: properties and applications. *Mathematica Slovaca*, 71(5):1291–1308, 2021.

[13] J. T. Eghwerıdo and L. Nzeı. A weighted Gompertz-G family of distributions for reliability and lifetime data analysis. *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics*, 73(1):235–258, 2024.

[14] M. S. Eliwa and M. El-Morshedy. A one-parameter discrete distribution for over-dispersed data: Statistical and reliability properties with applications. *Journal of Applied Statistics*, 49(10):2467–2487, 2022.

[15] M. S. Eliwa and M. El-Morshedy. A discrete extension of the exponential type II distribution: statistical characterizations, reliability analysis, and Bayesian vs. non-Bayesian inferences for random right-censored and complete count data. *Japanese Journal of Statistics and Data Science*, 1–39, 2024. https://doi.org/10.1007/s42081-024-00277-8.

[16] 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.

[17] 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(2):321–339, 2019.

[18] M. El-Morshedy, M. S. Eliwa, and A. Z. Afify. The odd Chen generator of distributions: Properties and estimation methods with applications in medicine and engineering. *Journal of the National Science Foundation of Sri Lanka*, 48(2):113–130, 2020.

[19] M. El-Morshedy, M. S. Eliwa, and A. Tyagi. A discrete analogue of odd Weibull-G family of distributions: properties, classical and Bayesian estimation with applications to count data. *Journal of Applied Statistics*, 49(11):2928–2952, 2022.

[20] J. Gómez-Déniz. Another generalization of the geometric distribution. *Test*, 19(2):399–415, 2010.

[21] B. Gompertz. On the nature of the function expressive of the law of human mortal-

ity, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115:513–583, 1825.

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

[23] J. Y. Kajuru, H. G. Dikko, A. S. Mohammed, and A. I. Fulatan. Generalized odd Gompertz-G family of distributions: Statistical properties and applications. *Communication in Physical Sciences*, 10(2):104–116, 2023.

[24] J. F. Lawless. Statistical models and methods for lifetime data. *Wiley, New York*, 2003.

[25] O. Lekhane, B. Oluyede, L. Gabaitiri, and O. V. Mabikwa. The exponentiated-Gompertz-Marshall-Olkin-G family of distributions: Properties and applications. *Statistics, Optimization & Information Computing*, 13(5):1752–1788, 2025.

[26] J. Mazucheli, A. F. Menezes, and S. Dey. Unit-Gompertz distribution with applications. *Statistica*, 79(1):25–43, 2019.

[27] A. Melnikov and Y. Romaniuk. Evaluating the performance of Gompertz, Makeham and Lee-Carter mortality models for risk management with unit-linked contracts. *Insurance: Mathematics and Economics*, 39(3):310–329, 2006.

[28] V. Nekoukhou and H. Bidram. The exponentiated discrete Weibull distribution. *Statistics and Operations Research Transactions*, 39(1):127–146, 2015.

[29] K. Ohishi, H. Okamura, and T. Dohi. Gompertz software reliability model: estimation algorithm and empirical validation. *Journal of Systems and Software*, 82(3):535–543, 2009.

[30] A. Rényi. On measures of entropy and information. *Mathematical Statistics and Probability*, 1:547–561, 1961.

[31] R. Roozegar, S. Tahmasebi, and A. A. Jafari. The McDonald Gompertz distribution: properties and applications. *Communications in Statistics: Simulation and Computation*, 46(5):3341–3355, 2017.

[32] D. Roy. Discrete Rayleigh distribution. IEEE Transactions on Reliability, 53(2):255–260, 2004.

[33] J. Wang and X. Guo. The Gompertz model and its applications in microbial growth and bioproduction kinetics: Past, present and future. *Biotechnology Advances*, 108335, 2024.

[34] S. M. Zaidi, M. M. A. Sobhi, M. El-Morshedy, and A. Z. Afify. A new generalized family of distributions: Properties and applications. *AIMS Mathematics*, 6(1):456–476, 2021.