Computational Journal of Mathematical and Statistical Sciences
4(1), 1–16
DOI:10.21608/cjmss.2024.307097.1065
https://cjmss.journals.ekb.eg/

Research article

A Novel Version of Geometric Distribution: Method and Application

Alaa R. El-Alosey,[1], **Ahmed M. Gemeay**[2*]

[1] Department of Mathematics, Faculty of Science, Tanta University, Tanta 31527, Egypt; alaa.abdelmaksoud@science.tanta.edu.eg

[2] Department of Mathematics, Faculty of Science, Tanta University, Tanta 31527, Egypt; ahmed.gemeay@science.tanta.edu.eg

* **Correspondence:** ahmed.gemeay@science.tanta.edu.eg

Abstract: This paper introduces a new family of discrete distributions, and investigates some of their statistical properties. The geometric distribution is utilized as a baseline for this new family, resulting in the derivation of a new discrete distribution, termed the generalized geometric distribution. This new distribution exhibits a wider range of shapes in its probability mass function and hazard rate function than the geometric distribution. Several mathematical properties of the proposed model are derived, and three estimation methods, namely maximum likelihood, moments, and proportion estimation, are employed to determine estimators for the new model. The performance of these estimators is evaluated using simulated data sets, demonstrating their accuracy and reliability in estimating the parameters of the generalized geometric distribution. The proposed model is applied to a real data set, and its flexibility in fitting the data is compared to other well-known discrete distributions in the literature. Our results suggest that the generalized geometric distribution provides a better fit for the data than the existing models.

Keywords: application, simulation, generalized distribution.
Mathematics Subject Classification: 62E10; 62E15
Received: 25 July 2024; **Revised:** 12 August 2024; **Accepted:** 5 September 2024; **Online:** 12 October 2024.

1. Introduction

Discrete distributions are fundamental concept in probability theory and statistical analysis. In contrast to continuous distributions, which represent probabilities over a range of values, discrete distributions describe probabilities for a finite set of values. These distributions are essential in modeling and analyzing random events that have a countable number of outcomes, such as the number of times a coin lands heads in a series of flips or the number of defective items in a production run. The study of discrete distributions is essential in many fields, including finance, engineering, and social sciences, as they provide valuable insights into the probabilities of various events occurring. One of the essential properties of discrete distributions is the probability mass function (PMF), which describes the probability of each possible outcome. The PMF is a crucial tool in understanding the behavior of discrete random variables and can be used to calculate various statistics, such as the mean, variance, and standard deviation. Discrete distributions also have various types, such as the Bernoulli distribution, binomial distribution, Poisson distribution, and geometric distribution. Each of these distributions has its own unique characteristics and applications. Understanding the different types of discrete distributions and their properties is essential in selecting the appropriate distribution for a particular problem and interpreting the results correctly. Several discrete distributions have been offered as analogues of continuous distributions in recent decades to provide a better alternative model for understanding count data sets with complicated behaviour. Traditional probability models such as the negative binomial, Poisson, Geometric, and binomial distributions are used to depict count data sets, however they may not always provide the best fit. As a result, more adaptive distributions are required.

Numerous researchers across several disciplines random variables that are discrete in character or are used frequently. In life testing trials, for example, continuously measuring a device's life span can be difficult or inconvenient can be difficult or inconvenient to continuously measure a device's life span. For instance, the lifespan of a switch has discrete values in the case of an on/off switching mechanism. When describing reliability data, it's common to talk about how many shocks, runs, or cycles a device can withstand before failing. In a survival analysis, we might count the days that lung cancer patients have lived after starting treatment, or we might count the days that pass between remission and relapse. In this context, the gamma and exponential distributions, respectively, are regarded as discrete alternatives to the negative binomial and geometric distributions. These discrete distributions contain monotonic hazard rate functions, which makes them inappropriate in some circumstances. Poisson and Geometric counted data models, on the other hand, can only handle positive integers and zero values. The first is to research a distinctive quality of a continuous distribution and create a comparable quality in discrete time. The second is to think of discrete lifespan as the integer component of continuous lifetime; for more information, see [19, 5, 21, 22]. By examining the connections between several reliability measures, Roy [20] was able to determine a bivariate geometric distribution in a special way using the bivariate extension of a univariate characterizing nature. He emphasized that the survival function can be used to express the univariate Geometric distribution as a discrete concentration of an equivalent exponential distribution.

All recent discrete distribution were derived by using the discretization method for a continuous distribution. Discretization is a technique used in data preprocessing and analysis to transform continuous data into discrete values. This is typically done by dividing the range of possible values into a set of intervals or bins and assigning each data point to the appropriate bin. Discretization is often used when dealing with data that is too complex or noisy to be easily analyzed in its original form, or when dealing with data that has too many distinct values to be effectively used in statistical models. For more information see [5] which make a survey on discrete models. Hossain [6] presented a family for obtaining basic discrete distributions existed in the literature such as Poisson, negative binomial, Bernoulli, Geometric, and binomial distributions. Aboraya et. al. [15] discussed a family of discrete distribution by using discretization of continuous families. The G-negative binomial family: general properties and applications was derived by [16]. Mardia and Sriram [17] introduced families of discrete circular distributions with some applications. An extension of Panjer's family of discrete distributions by derived the recursion formula for the probabilities of corresponding compound distributions for one such family see [18].

Our motivation in this paper is to derive a new family of discrete distributions without need of discretization method and that was introduced in Section 2, and to introduce a generalized discrete model which introduced in Section 3. We aim that the new discrete distribution will have more features which makes it is preferable than its baseline model. Different estimation methods of the new model were used to determine its unknown parameters in Section 4 along with study these methods behavior in Section 5. Also, in Section 6, we will show that the new model has a superiority for modelling real data sets than its baseline model and other well-known discrete model in the literature.

2. New family formulation

In this section, we derive a new family of discrete distributions, its cumulative distribution function (CDF) is defined as follows

$$F(x) = 1 - [1 - \zeta G(x)]^{x+1}, \; 0 < \zeta < 1, \; x = 0, 1, \ldots, \tag{2.1}$$

where $G(x)$ is the CDF of the baseline model. The PMF ofthe new family is defined as follows

$$P(X = x) = F(x) - F(x-1) = [1 - \zeta G(x-1)]^{x} - [1 - \zeta G(x)]^{x+1}. \tag{2.2}$$

Lemma 1. $P(x)$ of Equation (2.2) is a PMF.

Proof. Since $P(x)$ is between zero and one, to justify its PMF status, we will prove that $\sum_{x=0}^{\infty} p(x) = 1$ as follow

$$\sum_{x=0}^{\infty} P(x) = [1 - \zeta G(-1)]^{0} - [1 - \zeta G(0)] + [1 - \zeta G(0)] - [1 - \zeta G(1)]^{2} + \cdots + [1 - \zeta G(\infty)]^{\infty} = 1,$$

this completes the proof. □

The survival function and the hazard rate functions of the new family are, respectively, defined as follows

$$S(x) = [1 - \zeta G(x)]^{x+1}, \; h(x) = \frac{[1 - \zeta G(x-1)]^{x}}{[1 - \zeta G(x)]^{(x+1)}} - 1.$$

The family probability generating function is

$$\begin{aligned} PG(z) &= \sum_{x=0}^{\infty} z^{x} P(x) = [1 - \zeta G(-1)]^{0} - [1 - \zeta G(0)] + z[1 - \zeta G(0)] - z[1 - \zeta G(1)]^{2} + \cdots \\ &= 1 + (z-1) \sum_{x=0}^{\infty} z^{x} [1 - \zeta G(x)]^{x+1} = 1 + (z-1) \sum_{x=0}^{\infty} z^{x} S(x). \end{aligned}$$

Its moment generating function is

$$M(z) = 1 + (e^{z} - 1) \sum_{x=0}^{\infty} e^{zx} S(x). \tag{2.3}$$

3. Generalized Geometric distribution

The CDF and PMF of the Geometric distribution (GoD) [5] are, respectively, defined as follows

$$G(x) = 1 - \gamma^{x+1}, \; x = 0, 1, 2, \ldots, \; 0 < \gamma < 1, \tag{3.1}$$

$$P(X = x) = (1 - \gamma)\,\gamma^{x}, \tag{3.2}$$

where $(1 - \gamma)$ is the probability of success.

In this section, we derived a special model of our proposed family by replacing our baseline CDF $G(x)$ in Equation (2.1) by the CDF of the GoD in Equation (3.1). The new model will be known as generalized Geometric distribution (GGoD), its CDF and PMF are, respectively, defined as follows

$$F(x) = 1 - \left[1 - \zeta\left(1 - \gamma^{x+1}\right)\right]^{x+1}, \; 0 < \zeta < 1, \; 0 < \gamma < 1, \; x = 0, 1, \ldots, \tag{3.3}$$

$$P(X = x) = \left[1 - \zeta\left(1 - \gamma^{x}\right)\right]^{x} - \left[1 - \zeta\left(1 - \gamma^{x+1}\right)\right]^{x+1}. \tag{3.4}$$

The survival and hazard rate (HR) functions of the GGoD are, respectively, defined as follows

$$S(x) = \left[1 - \zeta\left(1 - \gamma^{x+1}\right)\right]^{x+1}, \tag{3.5}$$

$$h(x) = \frac{\left[\zeta\left(\gamma^{x} - 1\right) + 1\right]^{x}}{\left[\zeta\left(\gamma^{x+1} - 1\right) + 1\right]^{x+1}} - 1. \tag{3.6}$$

Different shapes of the PMF (3.4) are presented in Figure 1. It can easy see that the PMF (3.4) can be a unimodal or a decreasing function which is an advantage for the GGoD compared with the GoD which has only decreasing shape. Possible shapes of the HRF (3.6) are presented in Figure 2 and it has more shapes than the GGoD which has only constant HRF.

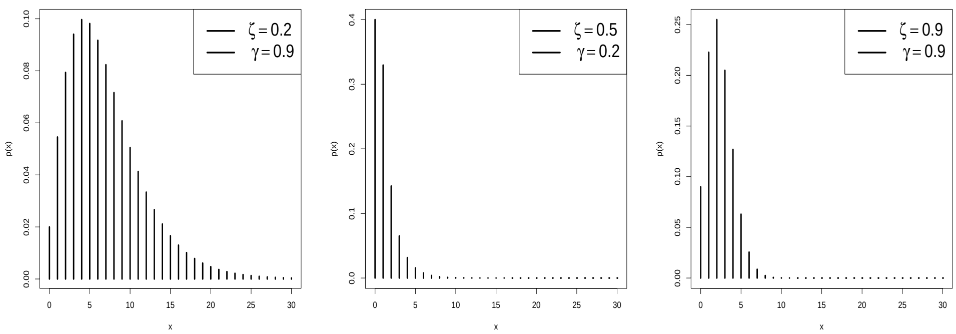

Figure 1. Plots of the PMF of the GGoD.

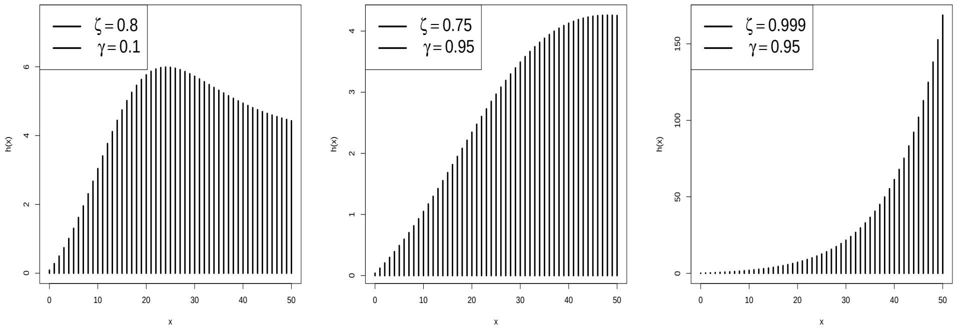

Figure 2. Plots of the HRF of the GGoD.

3.1. Moments

Let X be a random variable follows the GGoD, then its moment generating function is determined by using Equations (2.3) and (3.5) as follows

$$M(z) = 1 + (e^z - 1) \sum_{x=0}^{\infty} e^{zx} \Delta(x), \tag{3.7}$$

where $\Delta(x) = \left[1 - \zeta\left(1 - \gamma^{x+1}\right)\right]^{x+1}$.

By using the relationship between moment generating function and moments, we have the moments of the GGoD as follows

$$\mu_r = \frac{d^r}{dz^r} M(z)|_{z=0}. \tag{3.8}$$

Now, we will determine the first four moments of the GGoD, respectively, as follows

$$\frac{d}{dz} M(z) = \sum_{x=0}^{\infty} \Delta(x) e^{xz} \left[x(e^z - 1) + e^z\right], \mu_1 = \sum_{x=0}^{\infty} \Delta(x),$$

$$\frac{d^2}{dz^2} M(z) = \sum_{x=0}^{\infty} \Delta(x) e^{xz} \left[e^z + 2xe^z + x^2(e^z - 1)\right], \mu_2 = \sum_{x=0}^{\infty} \Delta(x)(2x + 1),$$

$$\frac{d^3}{dz^3} M(z) = \sum_{x=0}^{\infty} \Delta(x) e^{xz} \left[e^z + 3xe^z + 3x^2 e^z + x^3(e^z - 1)\right], \mu_3 = \sum_{x=0}^{\infty} \Delta(x) \left[1 + 3x + 3x^2\right],$$

$$\frac{d^4}{dz^4} M(z) = \sum_{x=0}^{\infty} \Delta(x) e^{xz} \left[e^z + 4xe^z + 6x^2 e^z + 4x^3 e^z + x^4(e^z - 1)\right], \mu_4 = \sum_{x=0}^{\infty} \Delta(x) \left[1 + 4x + 6x^2 + 4x^3\right].$$

By using the first four moments of the GGoD, we can obtain its variance (σ^2), coefficients of skewness (β_1) and kurtosis (β_2), respectively, by the following relations

$$\sigma^2 = \mu_2 - (\mu_1)^2, \beta_1 = \frac{(\mu_3 - 3\mu_2\mu_1 + 2(\mu_1)^3)^2}{(\mu_2 - (\mu_1)^2)^3}, \beta_2 = \frac{\mu_4 - 4\mu_3\mu_1 + 6\mu_2(\mu_1)^2 - 3(\mu_1)^4}{(\mu_2 - (\mu_1)^2)^2}.$$

The index of dispersion (ID) and the coefficient of variation (CV) of the GGoD are

$$ID = \frac{\sigma^2}{\mu_1}, \; CV = \frac{\sigma}{\mu_1}.$$

From Table 1, we conclude that

- The mean is a measure of central tendency that indicates the average value of the data. As ζ increases, the mean tends to decrease for a given γ. This behavior is consistent with the trend of decreasing values of mean as the shape parameter ζ becomes larger, resulting in a more skewed distribution.
- Variance quantifies the spread or dispersion of the data points around the mean. Similar to the mean, as ζ increases, the variance tends to decrease for a given γ. Lower values of ζ lead to larger variances, indicating greater variability in the data.
- Skewness measures the asymmetry of the data distribution. As ζ increases, the coefficient of skewness generally decreases. A larger ζ tends to make the distribution more symmetric, reducing the skewness.
- Kurtosis measures the "tailedness" of the distribution. As ζ increases, the coefficient of kurtosis tends to decrease. Higher values of ζ lead to distributions with lighter tails, indicating fewer extreme values compared to a normal distribution.
- The index of dispersion measures the relationship between the mean and variance of the distribution. As ζ increases, the index of dispersion tends to decrease. This indicates that the variance decreases at a faster rate than the mean as ζ increases, implying a reduction in variability relative to the mean.
  The ability of the GGoD to cover situations of over-dispersion, equi-dispersion, and under-dispersion, as demonstrated by the varied index of dispersion values in Table 1, underscores its versatility in modeling different types of data distributions. This property makes the GGoD a valuable tool for capturing a wide range of dispersion patterns, which is of significance in various statistical analyses and applications.
- The coefficient of variation is the ratio of the standard deviation to the mean and is a measure of relative variability. As ζ increases, the coefficient of variation tends to decrease. Larger values of ζ lead to distributions with lower relative variability compared to the mean.

In summary, the statistical measures presented in Table 1 demonstrate consistent trends as the shape parameter ζ varies. As ζ increases, the distribution becomes less skewed, has lighter tails, reduced dispersion relative to the mean, and lower relative variability. These trends reflect the impact of the ζ parameter on shaping the characteristics of the data generated by the GGoD. Additionally, for fixed values of ζ, as the γ parameter varies, these measures also change, reflecting the sensitivity of the GGoD model to changes in the shape parameter γ.

Table 1. A numerical representation for some statistical measures of the GGoD.

γ	measure	ζ								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	μ	9.01208	4.02363	2.36803	1.54529	1.05543	0.731807	0.503018	0.333363	0.20302
	σ^2	89.7993	19.8422	7.65982	3.66897	1.95322	1.09608	0.626641	0.354162	0.190367
	β_1	4.03454	4.11962	4.22919	4.33638	4.41112	4.42095	4.34195	4.20671	4.32829
	β_2	9.03771	9.14181	9.29407	9.46564	9.60812	9.63802	9.41184	8.70612	7.39346
	ID	9.96433	4.93142	3.23468	2.3743	1.85064	1.49777	1.24576	1.06239	0.937679
	CV	1.05151	1.10707	1.16875	1.23955	1.32418	1.43062	1.57372	1.78519	2.14911
0.2	μ	9.02976	4.05676	2.41462	1.60358	1.12388	0.809048	0.587827	0.42466	0.299845
	σ^2	89.5193	19.6428	7.5263	3.58905	1.91654	1.09396	0.651773	0.400409	0.252586
	β_1	4.06589	4.19942	4.31774	4.34284	4.19963	3.83241	3.24506	2.56201	2.08618
	β_2	9.07472	9.25574	9.46852	9.61552	9.56466	9.14551	8.17559	6.57126	4.64492
	ID	9.91381	4.84199	3.11698	2.23815	1.70529	1.35216	1.10878	0.942892	0.842387
	CV	1.04781	1.0925	1.13617	1.1814	1.2318	1.29279	1.3734	1.49008	1.67613
0.3	μ	9.0564	4.10428	2.47843	1.68	1.20994	0.902364	0.68648	0.527101	0.404826
	σ^2	89.1216	19.3914	7.38018	3.5179	1.89785	1.11103	0.692258	0.455207	0.315042
	β_1	4.10818	4.28628	4.37175	4.2457	3.84078	3.17491	2.37523	1.64797	1.18574
	β_2	9.12703	9.39421	9.63108	9.65784	9.29597	8.41367	7.0074	5.28848	3.68696
	ID	9.84073	4.72466	2.97777	2.09399	1.56855	1.23125	1.00842	0.863605	0.778215
	CV	1.0424	1.07292	1.09612	1.11643	1.13859	1.1681	1.21201	1.28	1.38649
0.4	μ	9.09806	4.17456	2.56811	1.78257	1.32076	1.0181	0.804778	0.646276	0.523697
	σ^2	88.5442	19.0778	7.23098	3.46962	1.90974	1.15696	0.755558	0.525397	0.385682
	β_1	4.16548	4.36966	4.35049	3.99518	3.33143	2.50941	1.72771	1.13475	0.776875
	β_2	9.20222	9.55401	9.73497	9.50218	8.73451	7.48525	5.98082	4.52992	3.376
	ID	9.73221	4.57001	2.81568	1.94641	1.44595	1.13639	0.93884	0.812959	0.73646
	CV	1.03426	1.04629	1.04709	1.04495	1.04632	1.05649	1.08009	1.12157	1.18586
0.5	μ	9.16664	4.28268	2.69846	1.92462	1.46809	1.1668	0.95248	0.791594	0.665854
	σ^2	87.6854	18.6992	7.1035	3.47078	1.97348	1.24762	0.854163	0.622186	0.476002
	β_1	4.24246	4.4207	4.18574	3.54703	2.70927	1.9041	1.27142	0.841344	0.579593
	β_2	9.31183	9.71343	9.68374	9.02975	7.86008	6.46676	5.14813	4.08193	3.30176
	ID	9.56572	4.36624	2.63243	1.80336	1.34426	1.06927	0.896778	0.785991	0.714875
	CV	1.02154	1.00971	0.987691	0.967986	0.956897	0.957293	0.970319	0.996455	1.03616
0.6	μ	9.28777	4.45832	2.89681	2.1301	1.67302	1.36749	1.14728	0.979911	0.847645
	σ^2	86.3771	18.2848	7.05902	3.57423	2.12855	1.41266	1.0119	0.766349	0.605199
	β_1	4.3405	4.36699	3.78081	2.89821	2.05763	1.41002	0.962337	0.668519	0.480006
	β_2	9.47113	9.79534	9.29983	8.13517	6.75192	5.50163	4.52351	3.81016	3.29781
	ID	9.30009	4.10128	2.43683	1.67796	1.27228	1.03303	0.882	0.782061	0.713976
	CV	1.00066	0.959123	0.917175	0.887545	0.872049	0.869148	0.876799	0.893361	0.917772
0.7	μ	9.52525	4.76705	3.22025	2.44819	1.97898	1.65962	1.42582	1.2458	1.10201
	σ^2	84.3714	17.9828	7.25497	3.89745	2.46088	1.71845	1.28336	1.00487	0.814743
	β_1	4.43594	4.04347	3.0531	2.13281	1.47664	1.04277	0.756456	0.563248	0.429563
	β_2	9.68354	9.5682	8.33118	6.83894	5.60232	4.70692	4.08013	3.63477	3.30648
	ID	8.85766	3.77231	2.25292	1.59197	1.24351	1.03544	0.900086	0.806602	0.739322
	CV	0.96432	0.889567	0.836428	0.80639	0.79269	0.789875	0.794529	0.804646	0.819074
0.8	μ	10.0758	5.38472	3.81484	3.00419	2.49743	2.14492	1.88254	1.67792	1.51281
	σ^2	81.6135	18.4261	8.16139	4.76841	3.21505	2.35994	1.83115	1.47725	1.22652
	β_1	4.36202	3.16964	2.07061	1.41569	1.02928	0.78431	0.617437	0.497258	0.407176
	β_2	9.80679	8.46628	6.66275	5.41131	4.62683	4.11806	3.76655	3.50859	3.30949
	ID	8.09998	3.42193	2.13938	1.58725	1.28735	1.10024	0.972703	0.880409	0.810759
	CV	0.896609	0.797176	0.748869	0.726874	0.717962	0.716207	0.718816	0.724364	0.732073
0.9	μ	11.8521	7.01373	5.25204	4.29139	3.66955	3.22665	2.89133	2.62642	2.41052
	σ^2	82.4411	23.0769	11.9525	7.72144	5.57227	4.29602	3.46044	2.8756	2.44591
	β_1	3.25064	1.72196	1.14837	0.873294	0.709807	0.598967	0.517326	0.453797	0.402441
	β_2	8.5546	5.93723	4.79434	4.23803	3.90993	3.68915	3.52709	3.40083	3.29819
	ID	6.95581	3.29025	2.27578	1.79929	1.51851	1.33142	1.19683	1.09487	1.01468
	CV	0.766084	0.68492	0.658265	0.647517	0.643284	0.642363	0.643381	0.645653	0.648799

4. Estimation of the GGoD parameters

We'll explore the utilization of various estimation methods namely, maximum likelihood estimation method (MLEM), moment estimation method (MEM), and proportion estimation method (PEM) as approaches for determining estimators of GGoD.

The log-likelihood function of PMF (3.4) is

$$\log(L) = \sum_{i=1}^{n} \log\left\{[1-\zeta(1-\gamma^{x_i})]^{x_i} - \left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i+1}\right\}, \tag{4.1}$$

now, we will equating the first derivative of Equation (4.1) with respect to GGoD's parameters (ζ, γ) to zero as follows

$$\begin{aligned}\frac{\partial \log(L)}{\partial \zeta} &= \sum_{i=1}^{n}\left\{[1-\zeta(1-\gamma^{x_i})]^{x_i} - \left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i+1}\right\}^{-1} \\ &\left(-\gamma^{x_i+1}\left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i} + \left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i}\right. \\ &- x_i\gamma^{x_i+1}\left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i} + x_i\left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i} \\ &\left. + x_i\gamma^{x_i}[1-\zeta(1-\gamma^{x_i})]^{x_i-1} - x_i[1-\zeta(1-\gamma^{x_i})]^{x_i-1}\right) = 0,\end{aligned} \tag{4.2}$$

$$\begin{aligned}\frac{\partial \log(L)}{\partial \gamma} &= \sum_{i=1}^{n}\left\{[1-\zeta(1-\gamma^{x_i})]^{x_i} - \left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i+1}\right\}^{-1} \\ &\left(\zeta x_i^2\gamma^{x_i-1}[1-\zeta(1-\gamma^{x_i})]^{x_i-1} - \zeta\gamma^{x_i}\left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i}\right. \\ &\left. - \zeta x_i^2\gamma^{x_i}\left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i} - 2\zeta x_i\gamma^{x_i}\left[1-\zeta\left(1-\gamma^{x_i+1}\right)\right]^{x_i}\right) = 0.\end{aligned} \tag{4.3}$$

For estimating the GGoD unknown parameters $\hat{\alpha}$ and $\hat{\gamma}$ by MEM, we should solve the following two equations with respect to our parameters.

$$\frac{1}{n}\sum_{i=1}^{n} x_i - \sum_{x=1}^{\infty}\left[1-\zeta\left(1-\gamma^{x+1}\right)\right]^{x+1} = 0, \tag{4.4}$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \sum_{x=1}^{\infty}\left[1-\zeta\left(1-\gamma^{x+1}\right)\right]^{x+1} + 2\sum_{x=1}^{\infty}(x-1)\left[1-\zeta\left(1-\gamma^{x+1}\right)\right]^{x+1} = 0. \tag{4.5}$$

Consider a random sample from the GGoD which has two parameters (ζ, γ), then, we will two indicators as follows

$$I_0(x_i) = \begin{cases} 1, & x_i = 0 \\ 0, & otherwise, \end{cases} \tag{4.6}$$

$B_0 = \sum_{i=1}^{n} I_0(x_i)$.

$$I_1(x_i) = \begin{cases} 1, & x_i = 1 \\ 0, & otherwise, \end{cases} \tag{4.7}$$

$B_1 = \sum_{i=1}^{n} I_1(x_i)$.

From Equations (3.4), (4.6), and (4.7), we have the following two equations

$$P(X=0) = \zeta(1-\gamma) = \frac{B_0}{n}, \tag{4.8}$$

$$P(X=1) = 1 - \left[\left(\gamma^2 - 1\right)\zeta + 1\right]^2 + (\gamma - 1)\zeta = \frac{B_1}{n}. \tag{4.9}$$

By solving Equations (4.8) and (4.9) with respect to parameters ζ and γ, then, we have the estimated parameters $\hat{\zeta}$ and $\hat{\gamma}$ by PEM.

5. Numerical simulation for the GGoD

This section tests the behavior for the estimators of the GGoD by using simulated data sets. We produce distinct random samples (with a total of $N = 2000$ samples) from the GGoD, employing various sample sizes $(50, 150, 300, 500)$. This is accomplished by initiating the generation process with different parameter values ($\zeta = 0.5, 0.25, 0.75, 0.9, 0.3, 0.5$ and $\gamma = 0.75, 0.4, 0.3, 0.2, 0.95, 0.5$). After each generation we determine $\overline{AVEs}$ along with $\overline{|BIAS|}$, $\overline{MSE}$, and $\overline{MRE}$, which are calculated as follows

$$\overline{AVEs} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{\Phi}},\ \overline{|BIAS|} = \frac{1}{N}\sum_{i=1}^{N}|\widehat{\mathbf{\Phi}} - \mathbf{\Phi}|,$$

$$\overline{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})^2,\ \overline{MRE} = \frac{1}{N}\sum_{i=1}^{N}|\widehat{\mathbf{\Phi}} - \mathbf{\Phi}|/\mathbf{\Phi},\ \Phi = (\zeta,\ \gamma).$$

Finally, these numerical results are presented in Tables 2 and 3, we conclude that

- The estimators $\hat{\zeta}$ and $\hat{\gamma}$ has the consistency property.
- $\overline{|BIAS|}$, $\overline{MSE}$, and $\overline{MRE}$ exhibit a declining trend as the sample size progressively grows.
- All estimation methods behaves well for all parameters values.

Table 2. Simulation results for $\overline{AVEs}$, $\overline{|BIAS|}$, $\overline{MSE}$, $\overline{MRE}$ of GGoD.

			$\zeta = 0.5, \gamma = 0.75$			$\zeta = 0.25, \gamma = 0.4$			$\zeta = 0.75, \gamma = 0.3$		
n	Est.	$\widehat{Par.}$	MLEM	MEM	PEM	MLEM	MEM	PEM	MLEM	MEM	PEM
50	$\overline{AVEs}$	$\hat{\gamma}$	0.625432	0.370770	0.542049	0.261201	0.320162	0.274105	0.801418	0.765323	0.779546
		$\hat{\zeta}$	0.767309	0.495364	0.734678	0.396419	0.438911	0.500064	0.327368	0.298654	0.315420
	$\overline{\lvert BIAS\rvert}$	$\hat{\gamma}$	0.180257	0.214254	0.109258	0.029306	0.104378	0.035729	0.096462	0.081944	0.075222
		$\hat{\zeta}$	0.090465	0.279405	0.078466	0.150836	0.178695	0.175121	0.115097	0.095909	0.097453
	$\overline{MSE}$	$\hat{\gamma}$	0.063433	0.057076	0.021598	0.001627	0.032137	0.002365	0.015428	0.010213	0.009202
		$\hat{\zeta}$	0.012094	0.118148	0.009800	0.035113	0.046890	0.040358	0.019680	0.013178	0.014367
	$\overline{MRE}$	$\hat{\gamma}$	0.360514	0.428507	0.218515	0.117223	0.417510	0.142914	0.128616	0.109259	0.100296
		$\hat{\zeta}$	0.120620	0.372539	0.104621	0.377091	0.446738	0.437803	0.383656	0.319695	0.324844
150	$\overline{AVEs}$	$\hat{\gamma}$	0.544379	0.413882	0.522582	0.252335	0.294804	0.256827	0.766470	0.765093	0.768750
		$\hat{\zeta}$	0.758197	0.598545	0.747166	0.396272	0.428029	0.420466	0.310522	0.304479	0.317990
	$\overline{\lvert BIAS\rvert}$	$\hat{\gamma}$	0.086239	0.179149	0.067904	0.015704	0.072813	0.017496	0.048946	0.062456	0.047575
		$\hat{\zeta}$	0.054137	0.182284	0.044745	0.092233	0.154848	0.126944	0.067260	0.074719	0.064426
	$\overline{MSE}$	$\hat{\gamma}$	0.017377	0.042265	0.010570	0.000404	0.016501	0.000505	0.004430	0.006320	0.003895
		$\hat{\zeta}$	0.004663	0.054317	0.003226	0.013866	0.036104	0.023469	0.006994	0.008442	0.006686
	$\overline{MRE}$	$\hat{\gamma}$	0.172478	0.358298	0.135808	0.062815	0.291252	0.069983	0.065262	0.083275	0.063433
		$\hat{\zeta}$	0.072183	0.243046	0.059661	0.230582	0.387119	0.317360	0.224200	0.249065	0.214752
300	$\overline{AVEs}$	$\hat{\gamma}$	0.522491	0.454447	0.528288	0.251397	0.271329	0.254649	0.759488	0.761492	0.756967
		$\hat{\zeta}$	0.755137	0.662457	0.758488	0.398781	0.408517	0.408005	0.309292	0.305507	0.302542
	$\overline{\lvert BIAS\rvert}$	$\hat{\gamma}$	0.055298	0.154017	0.059310	0.010762	0.046194	0.012519	0.033435	0.046382	0.034525
		$\hat{\zeta}$	0.038378	0.125614	0.040979	0.063840	0.123124	0.103943	0.045906	0.058387	0.047076
	$\overline{MSE}$	$\hat{\gamma}$	0.006285	0.032653	0.008063	0.000182	0.006528	0.000261	0.001949	0.003691	0.001999
		$\hat{\zeta}$	0.002269	0.026329	0.002639	0.006618	0.023281	0.017027	0.003448	0.005310	0.003412
	$\overline{MRE}$	$\hat{\gamma}$	0.110596	0.308033	0.118621	0.043047	0.184778	0.050077	0.044581	0.061843	0.046033
		$\hat{\zeta}$	0.051171	0.167486	0.054639	0.159600	0.307810	0.259857	0.153020	0.194624	0.156919
500	$\overline{AVEs}$	$\hat{\gamma}$	0.513766	0.478819	0.512175	0.250877	0.263852	0.252087	0.754546	0.757352	0.755241
		$\hat{\zeta}$	0.753327	0.687542	0.749025	0.401708	0.407434	0.392937	0.301087	0.303976	0.302667
	$\overline{\lvert BIAS\rvert}$	$\hat{\gamma}$	0.041809	0.152725	0.040269	0.008589	0.034503	0.009681	0.025210	0.034535	0.027236
		$\hat{\zeta}$	0.030152	0.109429	0.030646	0.049201	0.101886	0.086395	0.034790	0.044003	0.034982
	$\overline{MSE}$	$\hat{\gamma}$	0.003211	0.033575	0.003353	0.000113	0.002932	0.000150	0.001064	0.002022	0.001129
		$\hat{\zeta}$	0.001452	0.018635	0.001462	0.003788	0.016294	0.012472	0.001904	0.003033	0.001897
	$\overline{MRE}$	$\hat{\gamma}$	0.083618	0.305450	0.080538	0.034356	0.138014	0.038724	0.033613	0.046046	0.036315
		$\hat{\zeta}$	0.040203	0.145905	0.040861	0.123001	0.254716	0.215987	0.115968	0.146676	0.116606

Table 3. Continued of Table 2.

			$\zeta = 0.9, \gamma = 0.2$			$\zeta = 0.3, \gamma = 0.95$			$\zeta = 0.5, \gamma = 0.5$		
n	Est.	$\widehat{Par.}$	MLEM	MEM	PEM	MLEM	MEM	PEM	MLEM	MEM	PEM
50	$\overline{AVEs}$	$\hat{\gamma}$	0.929168	0.867006	0.882815	0.542208	0.529774	0.547173	0.544953	0.209024	0.414037
		$\hat{\zeta}$	0.211921	0.175634	0.192458	0.496077	0.452777	0.533958	0.953769	0.786267	0.945094
	$\overline{\lvert BIAS\rvert}$	$\hat{\gamma}$	0.070155	0.048127	0.039301	0.083273	0.125856	0.081214	0.297424	0.390976	0.186709
		$\hat{\zeta}$	0.076661	0.072762	0.066760	0.117378	0.160913	0.119776	0.029375	0.163733	0.027434
	$\overline{MSE}$	$\hat{\gamma}$	0.006452	0.003859	0.002617	0.015279	0.027614	0.013557	0.170858	0.108901	0.076624
		$\hat{\zeta}$	0.008825	0.007735	0.006730	0.023669	0.035705	0.022447	0.001380	0.034259	0.001138
	$\overline{MRE}$	$\hat{\gamma}$	0.077949	0.053475	0.043668	0.166547	0.251713	0.162429	0.991412	0.703254	0.622362
		$\hat{\zeta}$	0.383307	0.363808	0.333800	0.234755	0.321826	0.239552	0.030921	0.272350	0.028878
150	$\overline{AVEs}$	$\hat{\gamma}$	0.915698	0.897467	0.905049	0.506765	0.531395	0.524799	0.416747	0.163336	0.395484
		$\hat{\zeta}$	0.210689	0.192773	0.202904	0.500074	0.484630	0.530136	0.952351	0.854592	0.951681
	$\overline{\lvert BIAS\rvert}$	$\hat{\gamma}$	0.044738	0.040856	0.036513	0.034520	0.106873	0.049747	0.164507	0.145084	0.144007
		$\hat{\zeta}$	0.053438	0.050366	0.041800	0.067446	0.126822	0.071667	0.020219	0.095408	0.020045
	$\overline{MSE}$	$\hat{\gamma}$	0.003085	0.002457	0.001956	0.002810	0.021061	0.004859	0.072093	0.021880	0.050620
		$\hat{\zeta}$	0.004304	0.003883	0.002963	0.007092	0.023298	0.007904	0.000623	0.011089	0.000552
	$\overline{MRE}$	$\hat{\gamma}$	0.049709	0.045396	0.040570	0.069040	0.213745	0.099494	0.548356	0.483613	0.480022
		$\hat{\zeta}$	0.267189	0.251832	0.208999	0.134892	0.253644	0.143335	0.021283	0.100430	0.021100
300	$\overline{AVEs}$	$\hat{\gamma}$	0.908291	0.903292	0.902204	0.517412	0.520263	0.507069	0.350841	0.228354	0.339526
		$\hat{\zeta}$	0.205713	0.200269	0.196545	0.515339	0.493022	0.513134	0.950403	0.902987	0.951575
	$\overline{\lvert BIAS\rvert}$	$\hat{\gamma}$	0.032780	0.032946	0.031560	0.032020	0.077366	0.028736	0.095697	0.102494	0.077334
		$\hat{\zeta}$	0.041770	0.039259	0.036970	0.051740	0.097748	0.049620	0.014358	0.049536	0.013746
	$\overline{MSE}$	$\hat{\gamma}$	0.001784	0.001654	0.001556	0.001728	0.011878	0.001420	0.029471	0.016437	0.013639
		$\hat{\zeta}$	0.002755	0.002349	0.002143	0.004286	0.014496	0.003755	0.000319	0.004003	0.000315
	$\overline{MRE}$	$\hat{\gamma}$	0.036422	0.036607	0.035067	0.064041	0.154731	0.057471	0.318991	0.274980	0.257780
		$\hat{\zeta}$	0.208849	0.196297	0.184850	0.103480	0.195497	0.099241	0.015113	0.052144	0.014469
500	$\overline{AVEs}$	$\hat{\gamma}$	0.902149	0.904900	0.903983	0.502857	0.523449	0.510343	0.325023	0.276448	0.329630
		$\hat{\zeta}$	0.198923	0.204795	0.201269	0.498437	0.502365	0.507957	0.950230	0.915338	0.951288
	$\overline{\lvert BIAS\rvert}$	$\hat{\gamma}$	0.028246	0.028267	0.023867	0.018899	0.070037	0.021795	0.060336	0.092175	0.064312
		$\hat{\zeta}$	0.031441	0.033208	0.029830	0.035873	0.082099	0.039575	0.010976	0.043714	0.010831
	$\overline{MSE}$	$\hat{\gamma}$	0.001256	0.001218	0.000888	0.000631	0.010593	0.000759	0.009661	0.009499	0.010041
		$\hat{\zeta}$	0.001568	0.001673	0.001356	0.001900	0.010581	0.002313	0.000188	0.003314	0.000175
	$\overline{MRE}$	$\hat{\gamma}$	0.031385	0.031408	0.026519	0.037798	0.140073	0.043589	0.201121	0.107251	0.214372
		$\hat{\zeta}$	0.157203	0.166038	0.149149	0.071745	0.164198	0.079150	0.011554	0.046015	0.011401

6. Real data analysis

The objective of this section is to investigate how effectively the GGoD can be employed to fit real-world datasets. The analyzed data set consists of $n = 116$ observations which represents daily ozone concentrations that were collected in New York during May–September, 1973. It was presented by Ferreira et al. [1], it is studied by Jayakumar et al. [2]. A non-parametric plot of the analyzed real data set are displayed in Figure 3.

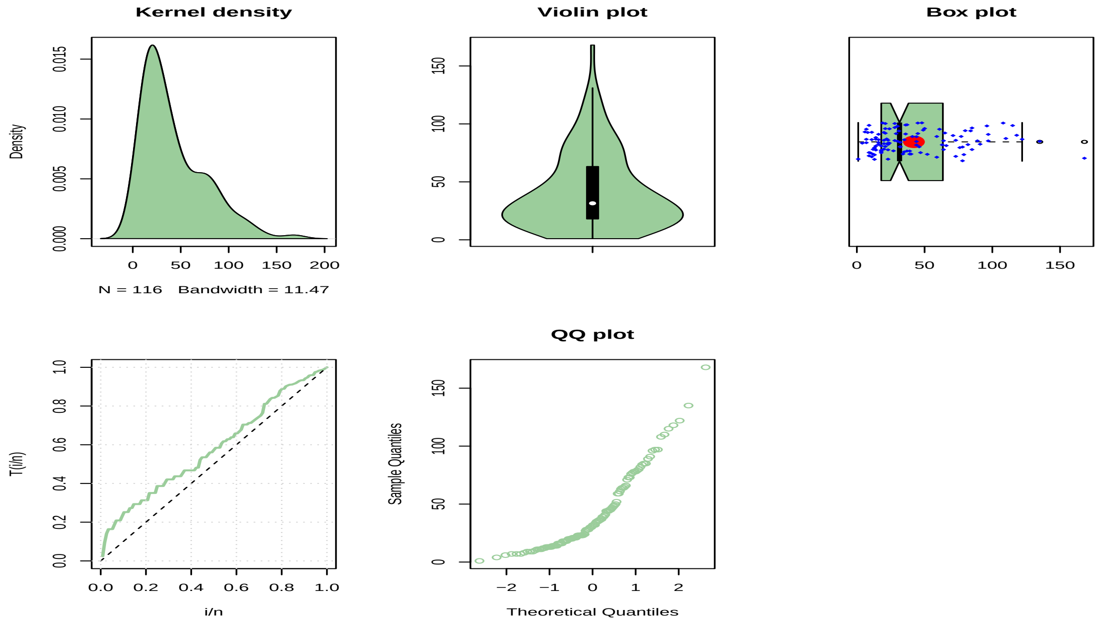

Figure 3. Non-parametric plot of ozone real data set.

Employing real-world datasets showcases the adaptability of the GGoD when compared to various established distributions, such as GoD, discrete half-logistic distribution [7] (DHLD), discrete Lindley distribution [8] (DLD), discrete gamma distribution [9] (DGD), discrete complementary Weibull-geometric distribution [2] (DCWGD), discrete Weibull distribution [10] (DWD), discrete three-parameter Burr type XII distribution [14] (D3PBD), discrete complementary exponential geometric distribution [2] (DCEGD), discrete Lomax distribution [14] (DLoD), exponentiated discrete Weibull distribution [4] (EDWD), discrete Ramos-Louzada distribution [11] (DRLD), discrete Weibull-geometric distribution [3] (DWGD), discrete mixture of gamma and exponential distribution [13] (DMGED), and discrete rayleigh distribution [12] (DRD). We used MLEM to determine the estimated parameters of all distributions. For best model selection, we used well-known statistics such as A_{IC}, CA_{IC}, B_{IC}, HQ_{IC}, KSS, with its p-value P_{KSS}.

Table 4 presents parameters estimates (standard error) and other measures for the ozone real data set that was investigated. In contrast to all other competing distributions, the GGoD closely matches the modeled dataset. This is evident from the small values across all measures related to the GGoD, as depicted in Table 4, except for the P_{KSS} value, which stands out with the highest value.

The graphical representation in Figure 4 illustrates the behavior of the log-likelihood function concerning the estimated parameters $\hat{\zeta}$ and $\hat{\gamma}$ for the GGoD. Both figures display a unimodal shape, demonstrating maximum values at the estimated parameter values of $\hat{\zeta}$ and $\hat{\gamma}$. This indicates the presence of a global maximum for our estimated parameters. Additionally, utilizing the Nmaximize function within Wolfram Mathematica software also corroborates this, as depicted in Figure 4, showcasing the pursuit of a global maximum solution. Moreover, Figure 5 presents the probability-probability (PP) plots, comparing the proposed model with other models under consideration. These plots further validate and support the outcomes detailed in Table 4.

Table 4. Numerical analysis for the ozone real data set.

Model	A_{IC}	CA_{IC}	B_{IC}	HQ_{IC}	KSS	P_{KSS}	$\widehat{Par.}$ (SEs)
GGoD	1086.6	1086.71	1092.11	1088.84	0.0630321	0.746052	$\hat{\zeta}$ = 0.0258219 (0.00277097)
							$\hat{\gamma}$ = 0.947396 (0.0172907)
DHLD	1095.17	1095.21	1097.93	1096.29	0.105444	0.151564	$\hat{\lambda}$ = 0.0330171 (0.00255731)
GoD	1104.58	1104.62	1107.34	1105.7	0.150529	0.0104234	$\hat{\lambda}$ = 0.0231861 (0.00212767)
DLD	1231.7	1231.73	1234.45	1232.81	0.0993041	0.202762	$\hat{\alpha}$ = 0.04746 (0.00311561)
DGD	1087.62	1087.73	1093.13	1089.86	0.0823368	0.411219	$\hat{\alpha}$ = 1.76556 (0.213781)
							$\hat{\beta}$ = 0.0414169 (0.00578978)
DWD	1090.26	1090.37	1095.77	1092.5	0.0845464	0.378272	$\hat{\alpha}$ = 0.994731 (0.0021811)
							$\hat{\beta}$ = 1.36338 (0.0968541)
DRLD	1104.56	1104.59	1107.31	1105.67	0.150522	0.0104282	$\hat{\lambda}$ = 41.582 (3.95895)
DRD	1104.58	1104.62	1107.34	1105.7	0.150529	0.0104234	$\hat{\lambda}$ = 0.976814 (0.00212767)
DMGED	1096.52	1096.56	1099.28	1097.64	0.116122	0.0875726	$\hat{\alpha}$ = 0.982179 (0.00144476)
D3PBD	1091.41	1091.62	1099.67	1094.76	0.0785799	0.470909	$\hat{\alpha}$ = 106.422 (117.002)
							$\hat{\beta}$ = 0.0138368 (0.0755321)
							$\hat{c}$ = 1.54381 (0.246673)
DLoD	1110.49	1110.6	1116.	1112.73	0.162173	0.00447856	$\hat{\alpha}$ = 510.979 (174.008)
							$\hat{\beta}$ = 2.76847 × 10^{-6} (0.0000113515)
DWGD	1091.55	1091.76	1099.81	1094.9	0.0712661	0.597703	$\hat{\alpha}$ = 1.37059 (0.21253)
							$\hat{\beta}$ = 0.278118 (0.511739)
							$\hat{c}$ = 0.995683 (0.00531994)
EDWD	1088.48	1088.69	1096.74	1091.83	0.0655601	0.701051	$\hat{\alpha}$ = 0.774117 (0.235471)
							$\hat{\beta}$ = 3.2203 (2.23417)
							$\hat{c}$ = 0.896784 (0.120087)
DCWGD	1087.83	1088.05	1096.09	1091.19	0.0652087	0.707371	$\hat{\alpha}$ = 1.79708 (0.0372976)
							$\hat{\beta}$ = 4.57645 (0.350449)
							$\hat{c}$ = 0.999633 (0.0000322492)
DCWED	1096.01	1096.12	1101.52	1098.25	0.0868787	0.345345	$\hat{\beta}$ = 0.361554 (0.0436103)
							$\hat{c}$ = 0.963053 (0.00913524)

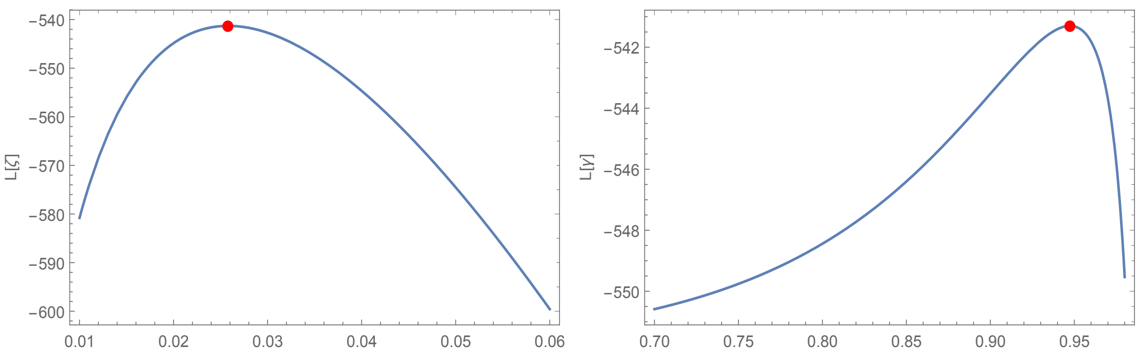

Figure 4. Plot of the log-likelihood function with the estimated parameters of the GGoD for the ozone real data set.

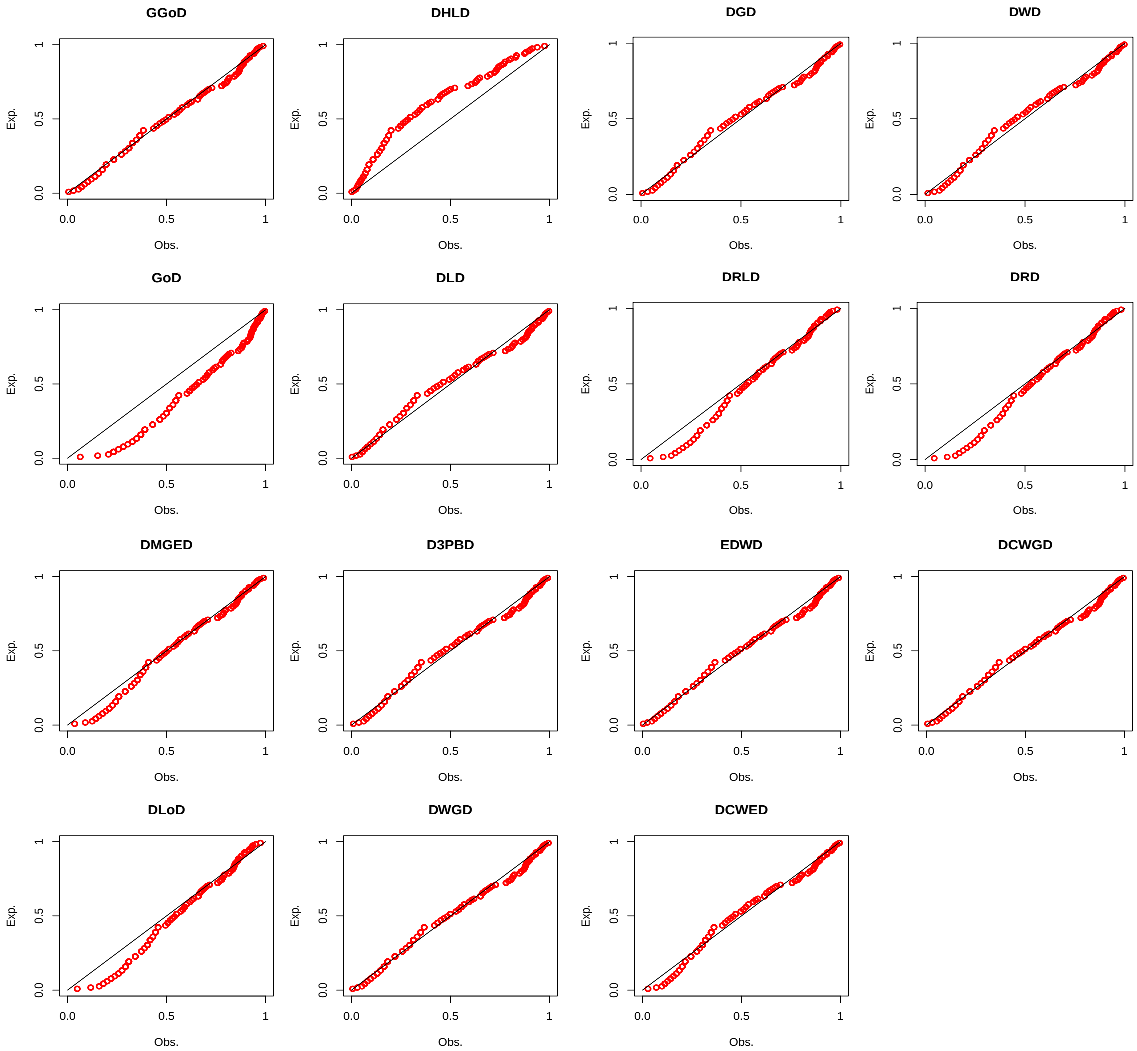

Figure 5. P-P plot of the GGoD and other compared models for the ozone real data set.

7. Conclusion

In conclusion, our study introduces a new family of discrete distributions and investigates its statistical properties. Specifically, we focus on the generalized Geometric distribution as a special model within this family and explore its key statistical characteristics. Our findings highlight the performance of the proposed estimators for the parameters of the generalized Geometric distribution. These estimators exhibit reliable and accurate performance across various scenarios. Through extensive simulations and real-world data analysis, we demonstrate the effectiveness and competitiveness of the new family in comparison to existing models.

There are various ways to expand the research presented in this article. One approach is to apply the generalized Geometric distribution to analyze and model data in other fields, such as reliability engineering, medicine, economics, survival analyses, and life testing. Another direction could involve examining Bayesian estimation of the distribution's parameters using complete and various types of censored samples under different loss functions. Additionally, a potential area of study is exploring a bivariate extension of the generalized Geometric distribution.

References

1 Ferreira, M., Gomes, M. I., & Leiva, V. (2012). On an extreme value version of the Birnbaum–Saunders distribution. REVSTAT-Statistical Journal, 10(2), 181-210.

2 Jayakumar, K., Babu, M. G., & Bakouch, H. S. (2021). General classes of complementary distributions via random maxima and their discrete version. Japanese Journal of Statistics and Data Science, 4, 797-820.

3 Jayakumar, K., & Babu, M. G. (2018). Discrete Weibull geometric distribution and its properties. Communications in Statistics-Theory and Methods, 47(7), 1767-1783.

4 Nekoukhou, V., & Bidram, H. (2015). The exponentiated discrete Weibull distribution. Sort, 39, 127-146.

5 Chakraborty, S. (2015). Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. Journal of Statistical Distributions and Applications, 2, 1-30.

6 Hossain, M. S. (2016). On a family of discrete probability distributions (fdpd). European Journal of Statistics and Probability, 4(5), 1-10.

7 Barbiero, A., & Hitaj, A. (2020, November). A discrete analogue of the half-logistic distribution. In 2020 International conference on decision aid sciences and application (DASA) (pp. 64-67). IEEE.

8 Bakouch, H. S., Jazi, M. A., & Nadarajah, S. (2014). A new discrete distribution. Statistics, 48(1), 200-240.

9 Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. Journal of Molecular evolution, 39, 306-314.

10 Nakagawa, T., & Osaki, S. (1975). The discrete Weibull distribution. IEEE transactions on reliability, 24(5), 300-301.

11 Eldeeb, A. S., Ahsan-ul-Haq, M., & Eliwa, M. S. (2022). A discrete Ramos-Louzada distribution for asymmetric and over-dispersed data with leptokurtic-shaped: Properties and various estimation techniques with inference. AIMS Mathematics, 7(2), 1726-1741.

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

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

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

15 Aboraya, M., M. Yousof, H., Hamedani, G. G., & Ibrahim, M. (2020). A new family of discrete distributions with mathematical properties, characterizations, Bayesian and non-Bayesian estimation methods. Mathematics, 8(10), 1648.

16 Percontini, A., Cordeiro, G. M., & Bourguignon, M. (2013). The G-negative binomial family: General properties and applications. Adv. Applic. Statist, 35, 127-160.

17 Mardia, K. V., & Sriram, K. (2023). Families of discrete circular distributions with some novel applications. Sankhya A, 85(1), 1-42.

18 Beknazaryan, A., & Adamic, P. (2022). On a stochastic order induced by an extension of Panjer's family of discrete distributions. Metrika, 85(1), 67-91.

19 Bracquemond, C., & Gaudoin, O. (2003). A survey on discrete lifetime distributions. International Journal of Reliability, Quality and Safety Engineering, 10(01), 69-98.

20 Roy, D. (1993). Reliability measures in the discrete bivariate set-up and related characterization results for a bivariate geometric distribution. Journal of Multivariate Analysis, 46(2), 362-373.

21 Kuş, C., Akdoğan, Y., Asgharzadeh, A., Kınacı, İ., & Karakaya, K. (2018). Binomial-discrete Lindley distribution. Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, 68(1), 401-411.

22 Sóyínká, A. T., & Olósundé, A. A. (2022). On discretization of continuous random variables for contingency tables: Discrete Johnson systems of distribution as a case study with applications. Journal of Statistical Theory and Practice, 16(4), 63.

SASAR
Open Access Journal