Asian Journal of Statistical Sciences
Vol. 2, No-2, (2022), pp. 159-171
© ARF India, Gurgaon, India
ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

Generalized geometric distribution of order *k*

C. Satheesh Kumar[a] and Harisankar, S[b]

[a] *Department of Statistics, University of Kerala, Thiruvananthapuram - 695 581, India.*
[b] *Asian School of Business, Thiruvananthapuram - 695 316, India.*

ARTICLE HISTORY
Compiled March 30, 2023

Received 15 May 2022; Accepted 22 December 2022

ABSTRACT
Here we introduce an order k version of the generalized geometric distribution of Kumar and Harisankar (Journal of Statistical Computation and Simulation, 2019) and investigate some of its important properties by deriving an expression for its probability generating function and probability mass function. Certain recurrence relation for its probabilities, raw moments and factorial moments are also obtained, and the maximum likelihood estimation of its parameters is discussed. Certain test procedures are developed for testing the significance of the additional parameters of the model. All these procedures discussed in the paper are illustrated with the help of real life data sets. A simulation study is also considered for assessing the performance of the estimators.

KEYWORDS
Generalized geometric Distribution; Generalized Likelihood Ratio Test; Maximum Likelihood Estimation; ; Model Selection: Probability generating function; Simulation

1. Introduction

Count data modelling becomes very popular in many areas like Insurance, Ecology, Environmental Science, Health etc., because of larger variance is than mean. Because of the applications and elegant and mathematical tractable distributional form geometric distribution attracts many research for count data modeling. Thus, recently several methods have been proposed to construct new discrete distributions with higher flexibility to model such data. Some of such generalization you can find in Jain and Consul(1971), Philippou and Georghiou(1983), Tripahti et al.(1987). However researchers still continuing to propose new generalization of these standard distributions. There has been renewed interest in the study of discrete distributions of order k in the literature. Philippou (1984) introduced and studied a negative binomial distribution of order k, which is also known as the type I waiting time distribution of order k. Further, Philippou and Muwafi (1982) discussed the relationship between the geometric distribution of order k and the Fibonacci sequence of order k. Philippou, A., Georghio, C and Philippou, G (1983) derived the distribution of order k for the generalized ge-

CONTACT: C. Satheesh Kumar[a]. Email:drcsatheeshkumar@gmail.com

ometric distribution, the negative binomial distribution and the Poisson distribution was derived as a limiting form of the corresponding negative binomial distribution. Uppuluri and Patil (1983) provided another derivation of this probability using p.g.f. Panaretos and Xekalaki (1986) obtained logarithmic distribution of order k as a limiting form of the gamma-mixed Poisson distribution of order k. Recently Kumar (2009, 2010), Kumar and Shibu (2013), Kumar and Nair (2013a, 2013b) and Kumar and Riyaz (2015, 2016) studied intervened Poisson distribution of order k, hyper-Poisson distribution of order k and zero-inflated logarithmic series distribution of order k respectively. For a detailed account of order k distributions and their applications see Section 10.7 of Johnson et al. (2005).

Kumar and Harisankar (2019) proposed a generalized class of geometric distribution namely "the generalized geometric distribution (GGD)" with the following p.g.f, for $x = 0, 1, 2, \cdots$, $\rho > -1$, $\lambda > 0$ and $0 < \theta < 1$.

$$H_1(t) = {}_2F_1(1, \lambda; \rho + \lambda + 1; \theta\ (t - 1)), \tag{1}$$

where

$${}_2F_1(a_1, a_2, b; \theta) = \sum_{r=0}^{\infty} \frac{(a_1)_r (a_2)_r}{(b)_r} \frac{\theta^r}{r!},$$

is the Gaussian hypergeometric function, in which $(a)_r = a(a + 1) \cdots (a + r - 1)$ and r is any positive integer with $(a)_0 = 1$. For details regarding Gaussian hypergeometric function see Mathai and Haubold (2008) or Slater (1966). Note that the GGD with p.g.f (1) belongs to the generalized hypergeometric family of distributions and it also belongs to the Kemp family of distributions, studied by Kumar (2009).

Through this paper we propose an order k version of the GGD which we termed as " stuttering generalized geometric distribution (SGGD)". The paper is organized as follows. In Section 2 we present the genesis of the SGGD and derive its important properties such as its p.g.f., probability mass function (p.m.f), expressions of mean and variance, recursion formulae for its probabilities, raw moments and factorial moments. In Section 3, we discuss the estimation of the parameters of the SGGD by the method of maximum likelihood. In Section 4, we derive the estimation of the parameters of the SGGD by the method of maximum likelihood. In section 5 we discuss the applicability of the model by fitting two real life data sets. In Section 6, we carried out a simulation study for examining the performance of the maximum likelihood estimators of the parameters of the distribution. Since the SGGD possess a stopped sum structure, they will be helpful for modeling real world phenomena arising from various fields of research such as actuarial science, biological sciences, operations research and physical sciences. For details regarding the stopped sum distributions, see chapter 9 of Johnson et.al. (2005).

Further we need the following series representation in the sequel.

$$\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} A_1(s, r) \quad = \quad \sum_{r=0}^{\infty} \sum_{s=0}^{r} A_1(s, r - s) \tag{2}$$

For $i = 0, 1, 2, \cdots$, define

$$\Omega_i^{-1} = {}_2F_1(1+i, \lambda+i, \rho+\lambda+i+1; -\sum_{j=1}^{k} \theta_j)$$

2. Derivation of the SGGD

Consider a sequence $\{Y_n, n \geq 1\}$ of independent and identically distributed discrete random variables, where Y_n has a k-point distribution with probability generating function (p.g.f.) $G(t) = \sum_{j=0}^{k} v_j t^j$, where $v_j \geq 0$ for $j = 1, 2, ..., k$ such that $v_k \neq 0$ and $\sum_{j=1}^{k} v_j = 1$. Let X be a nonnegative integer valued random variable following GGD with p.g.f (1). Define $v_j = \theta^{-1}\, \theta_j$, for j=1,2,...,k with $\theta = \sum_{j=1}^{k} \theta_j$. Suppose that $\{Y_n, n \geq 1\}$ and X are statistically independent. Define $S_0 = 0$, then the p.g.f of $S_X = \sum_{n=0}^{X} Y_n$ is

$$\begin{aligned} P(t) &= E(S^{T_X}) \qquad (3) \\ &= {}_2F_1(1, \lambda, \rho+\lambda+1; \sum_{j=1}^{k} \theta_j (t^j - 1)) \end{aligned}$$

Clearly, When k=1 the p.g.f (3) reduces to the p.g.f of the GGD as given in (1) and when k=1, ρ=0 and λ=1, the p.g.f (3) reduces to the p.g.f of the zero-inflated logarithmic series distribution (ZILSD) studied by Kumar and Riyaz (2016)

3. Properties

Let W be a random variable distributed as the SGGD with p.g.f. (3). Here, first weobtain an expression for the p.m.f. of the SGGD through the following result.

Result 1. The p.m.f $h_x = P(W = x)$ of the SGGD with p.g.f (3) is the following, for $x = 0, 1, 2 \cdots$.

$$h_x = \sum_{J_x} \frac{(\lambda)_x\, \Omega_x^{-1}}{(\rho+\lambda+1)_x} \prod_{j=1}^{k} \frac{x!\, \theta_j^{x_j}}{x_j!}, \qquad (4)$$

where $\sum_{J_x}$ denote the k-tuple sum over the test $J_x = (x_1, x_2, ..., x_k) : \sum_{j=1}^{k} j x_j = x$ and $n = \sum_{j=1}^{k} x_j$.

Proof. From (3) we have the following:

$$H(t) = {}_2F_1(1,\lambda;\rho+\lambda+1;\sum_{j=1}^{k}\theta_j(t^j-1)) \tag{5}$$

$$= \sum_{x=0}^{\infty} h_x\ t^x \tag{6}$$

On expanding the gauss hypergeometric function in (5), we get

$$H(t) = \sum_{x=0}^{\infty}\frac{(1)_x(\lambda)_x}{(\rho+\lambda+1)_x}\frac{\left(\sum_{j=1}^{k}\theta_j(t^j-1)\right)^x}{x!} \tag{7}$$

$$H(t) = \sum_{x=0}^{\infty}\frac{(1)_x(\lambda)_x}{(\rho+\lambda+1)_x}\sum_{r=0}^{x}\binom{x}{r}\frac{\left(\sum_{j=1}^{k}\theta_j t^j)^{x-r}(-\sum_{j=1}^{k}\theta_j)^r\right)}{x!} \tag{8}$$

$$H(t) = \sum_{x=0}^{\infty}\frac{(1)_{(x+r)}(\lambda)_{(x+r)}}{(\rho+\lambda+1)_{(x+r)}}\sum_{r=0}^{\infty}\binom{x+r}{r}\frac{\left(\sum_{j=1}^{k}\theta_j t^j)^{x}(-\sum_{j=1}^{k}\theta_j)^r\right)}{(x+r)!} \tag{9}$$

$$H(t) = \sum_{x=0}^{\infty}\frac{(1)_x(\lambda)_x}{(\rho+\lambda+1)_x}\sum_{r=0}^{\infty}\frac{(1+x)_r(\lambda+x)_r(-\sum_{j=1}^{k}\theta_j)^r}{(\rho+\lambda+x+1)_r\ r!}\frac{\left(\sum_{j=1}^{k}\theta_j t^j)^x\right)}{x!} \tag{10}$$

Now by applying multinomial theorem in (7) to get the following.

$$H(t) = \sum_{x=0}^{\infty}\sum_{I_x}\frac{(\lambda)_x\ \Omega_x^{-1}}{(\rho+\lambda+1)_x}\prod_{j=1}^{k}\frac{x!\ \theta_j^{x_j}}{x_j!}t^{\delta} \tag{11}$$

where $\delta = \sum_{j=1}^{k} jx_j$ and $\sum_{I_x}$ denote the k-tuple sum over the test $I_x = \{(x_1,x_2,...,x_k) : \sum_{j=1}^{k} x_j = x\}$ On equating the coefficients of t^x on the right hand side expressions of (6) and (11) we get (4). □

Next we derive expressions for the mean and variance of the SGGD through the following result.

Result 2. The mean and variance of the SGGD are the following, in which

$$Mean = b_0\ \sum_{j=1}^{k} j\theta_j \tag{12}$$

and

$$Variance = b_0 \ (\sum_{j=1}^{k} j\theta_j)[b_1 \ (\sum_{j=1}^{k} j\theta_j) + 1 - b_0 \ (\sum_{j=1}^{k} j\theta_j)] \tag{13}$$

for $i = 0, 1, 2..., \ b_i = \frac{(1+i)(\lambda+i)}{(\rho+\lambda+1+i)}$

Proof. It follows from the fact that,

$$Mean = H^1(1)$$

and

$$Variance = H^2(1) + H^{'}(1) - [H^{'}(1)]^2,$$

where $H^r(t) = \frac{d^r \ H(t)}{dt^r}/t = 1.$ □

Result 3. For $x \geq 1$, the following is a simple recursion formula for probabilities $h_x = h_x(1, \lambda; \rho + \lambda + 1)$ of the SGGD with p.g.f (3).

$$(x+1) \ h_{x+1}(1, \lambda; \rho + \lambda + 1) = \frac{\lambda \ (\sum_{j=1}^{k} j\theta_j)}{\rho + \lambda + 1} \ h_{x-j+1}(2, \lambda + 1; \rho + \lambda + 2) \tag{14}$$

Proof. From (3), we have

$$H(t) = \sum_{x=0}^{\infty} h_x(1, \lambda; \rho + \lambda + 1)t^x = \ {}_2F_1(1, \lambda; \rho + \lambda + 1; \sum_{j=1}^{k} \theta_j(t^j - 1)) \tag{15}$$

Differentiating the equation (15) with respect to t, we get

$$\sum_{x=0}^{\infty} \ x \ h_x(1, \lambda; \rho+\lambda+1) \ t^{x-1} = \frac{\lambda \ \sum_{j=1}^{k} j\theta_j t^{j-1}}{\rho + \lambda + 1} \ {}_2F_1(2, \lambda+1; \rho+\lambda+2; \sum_{j=1}^{k} \theta_j(t^j - 1)) \tag{16}$$

In (15) by replacing 1, λ and $\rho + \lambda + 1$ with 2, $\lambda + 1$ and $\rho + \lambda + 2$ respectively, we obtain

$${}_2F_1(2, \lambda + 1; \rho + \lambda + 2; \sum_{j=1}^{k} \theta_j(t^j - 1)) = \sum_{x=0}^{\infty} h_x(2, \lambda + 1; \rho + \lambda + 2) \ t^x \tag{17}$$

Substitute (17) in (16) to get

$$\sum_{x=0}^{\infty} (x+1) \ h_{x+1}(1, \lambda; \rho+\lambda+1) \ t^x = \frac{\lambda}{\rho + \lambda + 1} \ \sum_{j=1}^{k} j\theta_j \ \sum_{x=0}^{\infty} h_{x-j+1}(2, \lambda+1; \rho+\lambda+2) \ t^x. \tag{18}$$

□

On equating the coefficients of t^x on both sides of (18) we get (14).

Result 4. The following is a simple recursion formula for raw moments $\mu_r = \mu_r(1, \lambda; \rho + \lambda + 1)$ of the SGGD, for $r \geq 0$.

$$\mu_{r+1}(1, \lambda; \rho + \lambda + 1) = \frac{\lambda}{\rho + \lambda + 1} \sum_{s=0}^{r} \sum_{j=1}^{k} \jmath^{s+1} \, \theta_j \; \mu_{r-s}(2, \lambda; \rho + \lambda + 2) \tag{19}$$

Proof. By definition, the characteristic function of the SGGD is given by

$$\begin{aligned} \psi(t) &= \sum_{r=0}^{\infty} \mu_r(1, \lambda; \rho + \lambda + 1) \, \frac{(it)^r}{r!} \\ &= {}_2F_1(1, \lambda; \rho + \lambda + 1; \sum_{j=1}^{k} \theta_j (e^{itj} - 1)) \end{aligned} \tag{20}$$

By using (20) with 1, λ and $\rho+\lambda+1$ replaced by 2, $\lambda+1$ and $\rho+\lambda+2$ respectively, we obtain

$${}_2F_1(2, \lambda + 1; \rho + \lambda + 2; \sum_{j=1}^{k} \theta_j (e^{itj} - 1)) = \sum_{r=0}^{\infty} \mu_r(2, \lambda + 1; \rho + \lambda + 2) \frac{(it)^r}{r!}. \tag{21}$$

Differentiate (20) with respect to t, to get

$$\begin{aligned} \sum_{r=0}^{\infty} i \; \mu_{r+1}(1, \lambda; \rho + \lambda + 1) \frac{(it)^r}{r!} &= \frac{i \;\; \lambda \; \sum_{j=1}^{k} \theta_j e^{itj}}{\rho + \lambda + 1} \\ &\quad \times {}_2F_1(2, \lambda + 1; \rho + \lambda + 2; \sum_{j=1}^{k} \theta_j (e^{itj} - 1)) \end{aligned} \tag{22}$$

which on simplification gives

$$\begin{aligned} \frac{\rho + \lambda + 1}{\lambda} \sum_{r=0}^{\infty} \mu_{r+1}(1, \lambda; \rho + \lambda + 1) \frac{(it)^r}{r!} &= \sum_{j=1}^{k} \theta_j e^{itj} \\ &\quad \times \sum_{r=0}^{\infty} \mu_r(2, \lambda + 1; \rho + \lambda + 2) \frac{(it)^r}{r!}. \end{aligned} \tag{23}$$

On expanding the exponential functions in (23) and applying (2) to obtain

$$\begin{aligned} &\frac{\rho + \lambda + 1}{\lambda} \sum_{r=0}^{\infty} \mu_{r+1}(1, \lambda; \rho + \lambda + 1) \frac{(it)^r}{r!} = \\ &\sum_{j=1}^{k} \sum_{s=0}^{r} \binom{r}{s} j^{s+1} \theta_j \; \mu_{r-s}(2, \lambda + 1; \rho + \lambda + 2) \end{aligned} \tag{24}$$

Equating the coefficients of $(it)^r \; (r!)^{-1}$ on both sides of (24), we get (19). □

Result 5. The following is a simple recursion formula for factorial moments $\mu_{[r]} = \mu_{[r]}(1, \lambda; \rho + \lambda + 1)$ of the SGGD, for $r \geq 0$.

$$\left(\frac{\rho+\lambda+1}{\lambda}\right)\mu_{[r+1]}(1,\lambda;\rho+\lambda+1) = \sum_{j=1}^{k}\sum_{m=0}^{j-1}\binom{j-1}{m}\; j\;\theta_j\; r^{(m)}\;\mu_{r-m}(2,\lambda;\rho+\lambda+2) \tag{25}$$

Proof. The factorial moment generating function F(t) of the SGGD is given by

$$\begin{aligned} F(t) &= \sum_{r=0}^{\infty}\mu_{[r]}\frac{t^r}{r!} \qquad (26)\\ &= {}_2F_1[1,\lambda;\rho+\lambda+1;\sum_{j=1}^{k}\theta_j((t+1)^j-1)]. \end{aligned}$$

From (26) with 1, λ and $\rho+\lambda+1$ changed by 2, $\lambda+1$ and $\rho+\lambda+2$ respectively, we have

$${}_2F_1(2,\lambda+1;\rho+\lambda+2;\sum_{j=1}^{k}\theta_j((t+1)^j-1)]) = \sum_{r=0}^{\infty}\mu_r(2,\lambda+1;\rho+\lambda+2)\frac{t^r}{r!}. \tag{27}$$

On differentiating (26) with respect to t, we get

$$\begin{aligned} \sum_{r=0}^{\infty}\mu_{[r+1]}(1,\lambda;\rho+\lambda+1)\frac{t^r}{r!} &= \frac{\lambda}{\rho+\lambda+1}\sum_{j=1}^{k}j\theta_j(t+1)^{j-1}\\ &\quad\times{}_2F_1[2,\lambda+1;\rho+\lambda+2;\sum_{j=1}^{k}\theta_j((t+1)^j-1)] \end{aligned} \tag{28}$$

Equations (27)and (28) together implies

$$\begin{aligned} \sum_{r=0}^{\infty}\mu_{[r+1]}(1,\lambda;\rho+\lambda+1)\frac{t^r}{r!} &= \frac{\lambda}{\rho+\lambda+1}\sum_{r=0}^{\infty}\sum_{j=1}^{k}\sum_{m=0}^{j-1}\binom{j-1}{m}j\theta_j\\ &\quad\times\mu_r(2,\lambda+1;\rho+\lambda+2)\frac{t^{r+m}}{r!} \end{aligned} \tag{29}$$

Applying the series representation (2) in (28) to obtain

$$\begin{aligned}\frac{\rho+\lambda+1}{\lambda}\sum_{r=0}^{\infty}\mu_{[r+1]}(1,\lambda;\rho+\lambda+1)\frac{t^r}{r!} &= \sum_{r=0}^{\infty}\sum_{j=1}^{k} j\theta_j \sum_{m=0}^{j-1}\binom{j-1}{m} r^{(m)} \\ &\quad \times\mu_{r-m}(2,\lambda+1;\rho+\lambda+2)\frac{t^r}{r!}\end{aligned} \tag{30}$$

By equating the coefficients of $t^r\ (r!)^{-1}$ on both sides of (30), we get (25). □

4. Estimation

In this section we discuss the estimation of the parameters ρ, λ, θ_1, θ_2,...,θ_k (for a fixed value of k) of the SGGD by the method of maximum likelihood. Let $a(x)$ be the observed frequency of x events based on the observations from a sample with independent components and let y be the highest value of the x observed. The likelihood function of the sample is

$$L=\prod_{x=0}^{y}[h_x]^{a(x)}, \tag{31}$$

Taking logarithm on both sides of (31) we get

$$\ln L=\sum_{x=0}^{y}a(x)\ln h_x \tag{32}$$

Let $\hat{\rho}$, $\hat{\lambda}$, $\hat{\theta_1}$, $\hat{\theta_2}$, $\hat{\theta_3}$,...,$\hat{\theta_k}$ be the MLEs of ρ, λ, θ_1, θ_2, θ_3,...,θ_k respectively. Now the MLEs of the parameters are obtained by solving the following likelihood equations, obtained from (32) on differentiation with respect to ρ, λ, θ_i's respectively and equating to zero. Then

$$\frac{\partial\log L}{\partial\rho}=0 \tag{33}$$

or equivalently

$$\begin{aligned}&\sum_{x=0}^{y}a(x)[\upsilon(\rho+\lambda+x+1)-\upsilon(\rho+\lambda+1)\\ &+\Omega_x^{-1}\sum_{r=0}^{\infty}\frac{(1+x)_r(\lambda+x)_r(\sum_{j=1}^{k}-\theta_j)^r}{(\rho+\lambda+x+1)_r r!}[\upsilon(\rho+\lambda+x+1)\\ &-\upsilon(\rho+\lambda+x+r+1)]]=0,\end{aligned}$$

$$\frac{\partial\log L}{\partial\lambda}=0 \tag{34}$$

or equivalently

$$\sum_{x=0}^{y} a(x)\ [\upsilon(\lambda+x)-\upsilon(\lambda)\upsilon(\rho+\lambda+x+1)-\upsilon(\rho+\lambda+1)$$
$$-\Omega_x^{-1}\sum_{r=0}^{\infty}\frac{(1+x)_r(\lambda+x)_r(\sum_{j=1}^{k}\theta_j)^r}{(\rho+\lambda+x+1)_r\ r!}$$
$$\upsilon(\rho+\lambda+x+1)-\upsilon(\rho+\lambda+x+r+1)+\upsilon(\lambda+r+x)-\upsilon(\lambda+x)]=0,$$

$$\frac{\partial \log L}{\partial \theta_i}=0 \tag{35}$$

or equivalently

$$\sum_{x=0}^{y} a(x)\ [\Omega_x^{-1}\sum_{r=0}^{\infty}\frac{(\lambda)_r(\sum_{j=1}^{k}\theta_j)^{r-1}}{(\rho+\lambda+1)_r\ (r-1)!}+\frac{1}{\phi(x;\rho^*)}\sum_{I_x}\left[\frac{x!\ \theta_1^{x_1}\theta_2^{x_2}..\theta_i^{x_{i-1}}..\theta_k^{x_k}}{x_1!x_2!..(x_i-1)!x_i!..x_k!}\right]=0,$$

where

$$\upsilon(\rho)=[\Gamma(\rho)]^{-1}\ \frac{d\ \Gamma(\rho)}{d\rho},$$

and

$$\phi(x;\rho^*)=\sum_{I_x}\ (1)_x\ \prod_{j=1}^{k}\frac{\theta_j^{x_j}}{x_j!}.$$

On solving the log-likelihood equations by using some mathematical software say MATHEMATICA one can obtain the maximum likelihood estimators of the parameters ρ, λ, $\theta_i's$ of the SGGD.

5. Applications

For numerical illustration, we have considered two real life data application. The first data set is a sample consisting of counts of the number of eggs of an intestinal trematode, Schistosoma Mansoni, on single slides studied by Muench (1938) from 926 inhabitants in an Egyptian village. While the second data set is on the number of European red mite on each leave based on an experiment with 150 leaves from apple trees taken from Bliss et al (1953) . We have fitted the Generalized Poisson distribution(GP), the Generalized geometric distribution(GGOD) by Gomez-Deniz(2010), the GGD by Kumar and Harisankar(2019) and the SGGD to these data sets and the results obtained along with the corresponding values of the expected frequencies, Chi-square statistic, degrees of freedom (d.f), Akaike information criterion (AIC) and Bayesian information criterion (BIC), Corrected Akaike information criterion (AICc) in respect of each of the models are presented in Tables 1 and 2 respectively. Based on the computed values of the Chi-square statistic, AIC, BIC and AICc values it can be observed that the SGGD(k=3) gives better fit to both the data sets considered here compared to the existing models the GP, the GGOD, the GGD.

Table 1.: Observed frequencies and computed values of expected frequencies of the the GD, the GGD and the SGGD by the method of maximum likelihood for the first data set.

x	Observed	Expected frequency by MLE				
		GP	GGOD	GGD	SGGD	
					k=2	k=3
0	603	574.23	564.34	535.03	618.33	604.84
1	112	202.16	178.12	210.15	130.48	115.81
2	93	82.60	109.47	101.09	87.74	92.17
3	53	31.50	38.25	40.09	39.04	52.29
4	19	14.57	16.44	17.61	21.36	21.10
5	21	9.54	10.26	11.11	8.94	15.89
6	7	5.72	5.22	5.15	5.03	10.03
7	6	3.15	2.85	3.43	2.96	6.84
8	7	1.44	0.91	1.52	1.77	4.40
9	5	0.69	0.14	0.72	0.35	2.63
Total	926	926	926	926	926	926
d.f		5	4	4	3	2
Estimates of parameters		λ_1=0.52 λ_2=0.21	θ=1.15 α=0.16	ρ=-0.04 θ=0.55 λ=2.89	ρ=-0.86 θ_1=0.40 θ_2=0.24 λ=0.09	ρ=-0.91 θ_1=0.58 θ_2=0.22 θ_3=0.10 λ=0.06
χ^2-*value*		103.41	58.46	95.59	27.14	4.14
AIC		2470.70	2482.24	2572.72	2507.88	2431
BIC		2471	2483.14	2573.62	2509.08	2432.50
AICc		2470.95	2484.24	2574.72	2511.88	2441.50

Table 2.: Observed frequencies and computed values of expected frequencies of the the GD, the GGD and the SGGD by the method of maximum likelihood for the second data set.

x	Observed	Expected frequency by MLE				
		GP	GGOD	GGD		SGGD
					k=2	k=3
0	1333	1273.36	1314.05	1340.16	1311.23	1334.21
1	404	446.35	423.25	411.34	414.65	401.34
2	133	156.22	159.66	135.09	140.33	130.76
3	43	54.67	38.11	40.92	50.05	46.53
4	25	19.13	12.32	14.84	22.22	20.14

Continued ···

x	Observed	Expected frequency by MLE				
		GP	GGOD	GGD	SGGD	
					k=2	k=3
5	10	6.69	7.65	8.07	12.58	11.33
6	4	3.34	4.29	5.75	5.40	7.16
7	4	1.82	1.95	3.21	3.28	4.25
8	1	0.28	0.54	1.62	1.32	2.44
9	2	0.10	0.16	0.61	0.66	1.32
10	2	0.03	0.02	0.24	0.25	0.88
11	0	0.01	0.008	0.11	0.07	0.51
12	1	0.001	0.0001	0.06	0.01	0.23
Total	1962	1962	1962	1962	1962	1962
d.f		4	4	4	3	2
Estimates of parameters		λ_1=0.72	θ=0.96	ρ=-0.53	ρ=-0.47	ρ=-0.07
		λ_2=0.24	α=0.10	θ=0.58	θ_1=0.62	θ_1=0.85
				λ=1.08	θ_2=0.10	θ_2=0.03
					λ=1.55	θ_3=0.001
						λ=2.42
χ^2-*value*		28.89	11.73	13.24	6.35	2.86
AIC		2470.70	2572.72	2464.52	2507.88	2431
BIC		2471	2573.62	2465.42	2509.08	2432.50
AICc		2470.95	2574.72	2466.52	2511.88	2441.50

6. Simulation

To examine the performance of the MLEs, a simulation procedure was conducted for different sample sizes (n= 100, 200, 500). We simulated 1000 samples from the SGGD and then estimated the parameters by the maximum-likelihood method. By using simulated observations, we estimated the parameters ρ, λ, θ_1, θ_2, θ_3 and θ_4 of the SGGD and thereby computed the values of the absolute bias and mean squared errors of each of the estimators. The results obtained are presented in presented in Table 3 and 4, from which it can be observed that both the absolute values of bias and standard errors of the estimators of the parameters are in decreasing order as the sample size increases.

Table 3.: Absolute bias and standard errors in the parenthesis of the estimators of the parameters ρ, λ, θ_1, θ_2, θ_3 and θ_4 of the SGGD for k=4 corresponding to the parameter set ρ= -0.75, λ=0.55, θ_1=0.35, θ_2=0.24, θ_3=0.09, θ_4=0.005 .

Parameter set	Sample size	MLE					
		$\hat{\rho}$	$\hat{\lambda}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
(i)	$n = 100$	0.541 (0.416)	0.28 (0.125)	0.14 (0.081)	0.12 (0.045)	0.05 (0.009)	0.08 (0.041)

Continued $\cdots$

Sample Size	(t_1, t_2)	MLE					
		$\hat{\rho}$	$\hat{k}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
	$n = 200$	0.224 (0.281)	0.10 (0.082)	0.08 (0.041)	0.03 (0.005)	0.01 (0.002)	0.004 (0.0008)
	$n = 500$	0.12 (0.064)	0.06 (0.021)	0.04 (0.006)	0.02 (0.001)	0.048 (0.0006)	0.012 (0.0004)

Table 4.: Absolute bias and standard errors in the parenthesis of the estimators of the parameters ρ, λ, θ_1, θ_2 and θ_3 of the SGGD for k=3 corresponding to the parameter set ρ= -0.92, λ=0.95, θ_1=0.24, θ_2=0.10, θ_3=0.02.

Parameter set	Sample size	MLE				
		$\hat{\rho}$	$\hat{\lambda}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
(ii)	$n = 100$	0.26 (0.141)	0.44 (0.214)	0.19 (0.101)	0.06 (0.0013)	0.05 (0.01)
	$n = 200$	0.18 (0.082)	0.31 (0.141)	0.04 (0.054)	0.005 (0.0012)	0.032 (0.0024)
	$n = 500$	0.06 (0.035)	0.11 (0.084)	0.008 (0.004)	0.0002 (0.00068)	0.002 (0.0012)

Acknowledgement(s)

We are grateful to both the referees for their careful review and constructive comments that substantially improved the article.

References

[1] Dorogovtsev, S. N., Mendes, J. F. F. and Samukhin, A. N. (2000). Structure of growing networks with preferential linking. Physical Review Letters, 85(21), 4633 -4636.

[2] Falls, W.L., Williford, O.W. and Carter, C.M. (1971). Probability distributions for thunderstorm activity at Cape Kennedy, Florida, Journal of Applied Materorology, 10, 97-104.

[3] Gomez-Deniz, E. (2010). Another generalization of the geometric distribution. Test, 19, 399-415.

[4] Greenwood, M. and Yule, G.U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents . Journal of Royal Statistical Society, 83,255-304.

[5] Handcock, M. S. and Jones, J. H. (2004). Likelihood-based inference for stochastic models of sexual network formation. Theoretical Population Biology, 65,413-422.

[6] Jones, J. H. and Handcock, M. S. (2003). An assessment of preferential attachment as a mechanism for human sexual network formation. Proceedings of the Royal Society of London B: Biological Sciences, 270(1520), 1123-1128.

[7] Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). Univariate Discrete Distributions (III). John Wiley & Sons, New York.

[8] Kumar, C.S. and Riyaz, A. (2013a). An Extended Zero-Inflated logarithmic series distribution and its application, Journal of Applied Statistical Science, 21(1), 31-42.

[9] Kumar, C.S. and Riyaz, A. (2013b). On Zero-Inflated logarithmic series distribution and its application, Statistica , 73(4), 477-492.

[10] Kumar,C.S. and Harisankar, S. (2020a). On Some Aspects of a General Class of Yule distribution and its Applications, Communication in Statistics - Theory and Methods, 49(12), 2887-2897.

[11] Kumar,C.S. and Harisankar, S. (2020b). Modified Geometric Distribution: Properties and Applications , Stochastic Modelling and Applications, 24(1), 43-54.

[12] Levene, M., Fenner, T., Loizou, G. and Wheeldon, R. (2002). A Stochastic model for the evolution of the web.Computer Networks, 39(3), 277-287.

[13] Martínez-Rodríguez A.M., Sáez-Castillo, A. J. and Conde-Sánchez, A. (2011). Modelling using an extended Yule distribution. Computational Statistics and Data Analysis, 55(1), 863-873

[14] Mathai, A. M. and Haubold, H. J. (2008). Special Functions for Applied Statistics. Springer, New York .

[15] Mishra, A. (2009). On a generalized Yule distribution. Assam Statistical Review, 23, 140-150.

[16] Philippou, A.N. (1984). The Negative Binomial Distribution of Order k and Some of Its Properties. Biometrical Journal, 26(7), 789 - 794.

[17] Philippou, A.N., Georghiou, C. and Philippou, G.N. (1983). A generalized geometric distribution and some of its properties. Statistics & Probability Letters, 1(4), 171-175.

[18] Philippou, A.N. and Muwafi, A.A. (1982). Waiting for the k-th consecutive success and the Fibonacci sequence of order k. Fibonacci Quarterly, 20(1), 28-32.

[19] Simon, H. A. (1955). On a class of skew distribution. Biometrika, 42, 425-440.

[20] Slater, L. J. (1966). Generalized Hyper-geometric Functions: Cambridge University Press, Cambridge.

[21] Wagner, G. G., Burkhauser, R.V. and Behringer, F. (1993). The English language public use file of the German socio-economic panel. Journal of Human Resource, 28, 429-433

[22] Xekalaki, E. (1983). A property of Yule distribution and its applications. Communication in Statistics: Theory and Methods 12, 140-150.

[23] Yule, G. U. (1920). A mathematical theory of evolution based on the conclusion of Dr. J. C. Wills. Philosophical Society of Royal Society of London. Series B, 213, 21-87.