Stochastic Modeling and Applications
Vol. 24 No. 1 (January-June, 2020

Received: 20th March 2019 *Revised: 05th April 2020* *Accepted: 17th April 2020*

MODIFIED GEOMETRIC DISTRIBUTION: PROPERTIES AND APPLICATIONS

C. SATHEESH KUMAR AND HARISANKAR. S

Abstract. Here we obtain a modified version of the well-known "geometric distribution" through compounding generalized beta distribution with an extended form of the geometric distribution and investigate its important properties. The maximum likelihood estimation of the parameters of the distribution is discussed and the generalized likelihood ratio test procedure is considered for testing the significance of the parameters of the model and a simulation study is performed to compare the performance of the estimates in terms of bias and mean square error. In addition, the proposed mixture model is fitted to three real life data sets for showing its suitability in various fields.

1. Introduction

Count data modeling have achieved a great deal of attention of many researchers working in different areas like insurance, economics, social sciences and biometrics. However,often it has been found that count data exhibits over-dispersion (variance $>$ mean) and corresponding distribution function shows long tail behavior. In last two decades, many methods have been proposed by the researchers to develop new models for count data, one such method is mixtures of distributions. Application of mixture models spread over astronomy, biology, genetics, medicine, psychiatry, economics, engineering, marketing and other fields in the biological, physical and social sciences. For details see McLachlan and Peel [7]. In these applications, finite mixture models supports major areas of statistics including cluster and latent class analysis, discriminant analysis, image analysis and survival analysis. There is a vast literature available on finite mixture models. For example, see Everitt and Hand [2], Titterington et al. [12].

Here first we propose a distribution which we obtained through compounding an extended form of geometric and generalized beta distribution and named it as the modified geometric distribution (MGD) for creating more flexibility in modelling aspects. Some of the well-known distributions are special cases of MGD. The manuscript is organized as follows: In Section 2, we introduced the MGD and provide its important properties. In section 3 we discuss the estimation of the parameters of the MGD by utilizing the maximum likelihood procedure and illustrated the procedure in section 4. Section 5 contains a test procedure useful for testing the additional parameter of the proposed model and in section 6 a brief

---

2000 *Mathematics Subject Classification.* Primary 60E10;Secondary 60E15.
Key words and phrases. Probability generating function; Model Selection; Maximum Likelihood Estimation; Generalized Likelihood Ratio Test; Simulation.

simulation study is included for efficiency comparison of the the performance of the maximum likelihood estimators of the parameters of MGD.

Throughout this paper we adopt the following shorter notation, for j= 0,1,2,.. .

$$\Omega_j^{-1} = {}_2F_1(1+j, k+j, \rho+k+j+1; \gamma_1+\gamma_2), \tag{1.1}$$

where ${}_2F_1(.)$ is the Gaussian hyper-geometric function. For more details, see Slater (1966) or Mathai and Haubold (2008). Further, we need the following series representations in the sequel.

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\Delta_{n,m} = \sum_{m=0}^{\infty}\sum_{n=0}^{m}\Delta_{n,m-n}, \tag{1.2}$$

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\Delta_{n,m} = \sum_{m=0}^{\infty}\sum_{n=0}^{[\frac{m}{2}]}\Delta_{n,m-2n} \tag{1.3}$$

and

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\Delta_{n,m} = \sum_{m=0}^{\infty}\sum_{n=0}^{m}\Delta_{n,m}. \tag{1.4}$$

2. A Genesis of the MGD and its properties

Let X be an extended geometric random variable with probability generating function (p.g.f)

$$G(t) = \frac{{}_1F_0[1; -; \theta(\gamma_1\ t + \gamma_2\ t^2)]}{{}_1F_0[1; -; \theta(\gamma_1+\gamma_2)]}, \tag{2.1}$$

Assume that the parameter θ follows a generalized beta distribution with parameters ρ, k, γ_1 and γ_2, with probability density function (p.d.f)

$$f(\theta) = \frac{\Omega_0\ \theta^{k-1}\ (1-\theta)^{\rho}\ {}_1F_0[1; -; \theta(\gamma_1+\gamma_2)]}{B(k, \rho+1)},$$

in which $k > 0$ $\rho > -1$. Then the unconditional distribution of X is obtained as

$$\begin{aligned} Q(t) &= \frac{\Omega_0}{B(k,\rho+1)}\int_0^1 \theta^{k-1}\ (1-\theta)^{\rho}\ {}_1F_0(1; -; \theta(\gamma_1\ t+\gamma_2\ t^2)\ d\theta \\ &= \Omega_0\ {}_2F_1(1, k, \rho+k+1; \gamma_1\ t+\gamma_2\ t^2), \end{aligned} \tag{2.2}$$

using the identity (1.104) of Johnson, Kemp and Kotz (2005).

Definition 2.1. A non-negative integer valued random variable Y is said to follow the modified geometric distribution (MGD) if its p.g.f is of the following form, in which $\rho > -1$, $k > 0$, $\gamma_1 > 0$ and $\gamma_2 \geq 0$.

$$Q(t) = \Omega_0\ {}_2F_1[1, k; \rho+k+1; \gamma_1 t + \gamma_2 t^2] \tag{2.3}$$

Clearly certain well-known models can be obtained as special cases of MGD, as given below:
1).When $\gamma_2 = 0$ the p.g.f (2.3) reduces to the p.g.f of the doubly generalized Yule distribution (DGYD) introduced by Kumar and Harisankar (2019).
2). When ρ=0 and k=1, the p.g.f (2.3) reduces to the p.g.f of the modified zero-inflated logarithmic series distribution (MZILSD) studied by Kumar and Riyaz (2013), which reduces to zero-inflated logarithmic series distribution when γ_2=0.
3). When $\gamma_2 = 0$, k=1 the p.g.f (2.3) reduces to the p.g.f of the extended version of yule distribution (EYD) introduced by Martinez-Rodriguez (2011).
4). When $\gamma_2 = 0$ and γ_1 approaches to 1 the p.g.f (2.3) reduces to the p.g.f of the Waring distribution (WD), which reduces to Yule distribution when k=1.
5). When ρ approaches to -1, the p.g.f (2.3) reduces to the p.g.f of the extended geometric distribution (EGGD) having p.g.f (2.1) with $\theta = 1$, which further reduces to standard geometric distribution when γ_2=0.
Now we obtain the p.m.f of the MGD through the following result.

Result 2.2. The p.m.f q_y of the MGD with p.g.f (2.3) is the following, for $y = 0, 1, 2\cdots$, $\rho > -1$, $k > 0$, $\gamma_1 > 0$ and $\gamma_2 \geq 0$ such that $\gamma_1 + \gamma_2 < 1$.

$$q_y = \Omega_0 \sum_{j=0}^{[\frac{y}{2}]} \frac{(1)_{y-j}\ (k)_{y-j}}{(\rho+k+1)_{y-j}} \frac{\gamma_1^{y-2j}\ \gamma_2^j}{(y-2j)!\ j!} \tag{2.4}$$

where Ω_0 is as defined in (1.1).

Proof. From (2.3) we have the following:

$$Q(t) = \Omega_0\ {}_2F_1[1, k; \rho+k+1; \gamma_1\ t + \gamma_2\ t^2] \tag{2.5}$$

$$= \sum_{y=0}^{\infty} q_y\ t^y \tag{2.6}$$

On expanding the gauss hyper-geometric function in (2.5) to get

$$Q(t) = \Omega_0 \sum_{y=0}^{\infty} \frac{(k)_y}{(\rho+k+1)_y} [\gamma_1\ t + \gamma_2\ t^2]^y. \tag{2.7}$$

By applying binomial theorem in (2.7), we obtain the following.

$$Q(t) = \Omega_0 \sum_{y=0}^{\infty} \frac{(k)_y}{(\rho+k+1)_y} \sum_{j=0}^{y} \binom{y}{j} (\gamma_1\ t)^{y-j}\ (\gamma_2\ t^2)^j$$

$$= \Omega_0 \sum_{y=0}^{\infty}\sum_{j=0}^{\infty} \frac{(k)_{y+j}}{(\rho+k+1)_{y+j}} \binom{y+j}{j} (\gamma_1\ t)^y\ (\gamma_2\ t^2)^j. \tag{2.8}$$

Apply (1.3) in (2.8) to obtain the following.

$$Q(t) = \Omega_0 \sum_{x=0}^{\infty} \sum_{j=0}^{[\frac{y}{2}]} \frac{(1)_{y-j}\ (k)_{y-j}}{(\rho+k+1)_{y-j}\ (y-2j)!\ j!} \gamma_1^{y-2j}\ \gamma_2^j\ t^y \tag{2.9}$$

On equating the coefficients of t^y on the right hand side expressions of (2.6) and (2.9) we get (2.4).

In the light of equation (2.3), we obtain the following results with regard to the MGD. □

Result 2.3. The characteristic function $\phi(t)$ of the MGD is the following for any $t \in R$ and $i = \sqrt{-1}$.

$$\phi(t) = \Omega_0 \ {}_2F_1[1, k; \rho + k + 1; \gamma_1 \ e^{it} + \gamma_2 \ e^{2it}] \tag{2.10}$$

Result 2.4. The factorial moment generating function $\Lambda(t)$ of the MGD with p.g.f is given by

$$\Lambda(t) = \Omega_0 \ {}_2F_1[1, k; \rho + k + 1; \ (\gamma_1 + 2 \ \gamma_2)t + \gamma_2 \ t^2 + (\gamma_1 + \gamma_2)]. \tag{2.11}$$

Result 2.5. The mean and variance of the MGD are

$$Mean = \Omega_0 \ \frac{k \ (\gamma_1 + 2\gamma_2)}{(\rho + k + 1)} \ \Omega_1 \tag{2.12}$$

and

$$Variance = \frac{\Omega_0 k(\gamma_1 + 2\gamma_2)}{\rho + k + 1} \left[\frac{2(k+1)(\gamma_1 + 2\gamma_2)\Omega_2}{\rho + k + 2} + \Omega_1\Omega_0(1 - \frac{\Omega_1 k(\gamma_1 + 2\gamma_2)}{\rho + k + 1}) \right]. \tag{2.13}$$

Remark 2.6. From (2.12) and (2.13), it can be seen that the MGD is over-dispersed for all values of ρ, k,γ_1 and γ_2.

Result 2.7. For $y \geq 0$, the following is a simple recursion formula for probabilities $q_y = q_y(1, k; \rho + k + 1)$ of the MGD with p.g.f (2.3).

$$(y + 1) \ q_{y+1}(1, k; \rho + k + 1) = \Omega_0 \ \frac{k \ (\gamma_1 + 2 \ \gamma_2)}{\rho + k + 1} \ q_y(2, k + 1; \rho + k + 2) \tag{2.14}$$

Proof. From (2.3), we have

$$Q(t) = \sum_{y=0}^{\infty} q_y(1, k; \rho + k + 1) \ t^y = \ {}_2F_1[1, k; \rho + k + 1; \gamma_1 \ t + \gamma_2 \ t^2] \tag{2.15}$$

Differentiating the equation (2.15) with respect to t, we get

$$\sum_{y=0}^{\infty} (y+1) \ q_{y+1}(1, k; \rho+k+1) \ t^y = \frac{k \ (\gamma_1 + 2\gamma_2)}{\rho + k + 1} \ {}_2F_1[2, k+1; \rho+k+2; \gamma_1 \ t + \gamma_2 \ t^2]. \tag{2.16}$$

In (2.15) by replacing $1, \ k$ and $\rho + k + 1$ with $2, \ k + 1$ and $\rho + k + 2$ respectively, we obtain

$${}_2F[2, k + 1; \rho + k + 2; \gamma_1 \ t + \gamma_2 \ t^2] = \sum_{y=0}^{\infty} q_y(2, k + 1; \rho + k + 2) \ t^y. \tag{2.17}$$

Substitute (2.17) in (2.16) to get

$$\sum_{y=0}^{\infty}(y+1) \ h_{y+1}(1, k; \rho+k+1) \ t^y = \frac{k \ (\gamma_1 + 2\gamma_2)}{(\rho + k + 1)} \ \sum_{y=0}^{\infty} q_y(2, k+1; \rho+k+2) \ t^y. \tag{2.18}$$

On equating the coefficients of t^y on both sides of (2.18), we get (2.14). □

Result 2.8. The following is a simple recursion formula for raw moments $\mu_r = \mu_r(1, k; \rho + k + 1)$ of the MGD, for $r \geq 0$.

$$\mu_{r+1}(1, k; \rho{+}k{+}1) = \frac{\Omega_0\ k}{\rho + k + 1} \sum_{s=0}^{r} \binom{r}{s} (\gamma_1{+}2^{s+1}\ \gamma_2)\ \mu_{r-s}(2, k; \rho{+}k{+}2) \quad (2.19)$$

Proof. For $t \in R = (-\infty, \infty)$ and $i = \sqrt{-1}$, the characteristic function of the MGD is given by

$$\begin{aligned} \phi(t) &= \sum_{r=0}^{\infty} \mu_r(1, k; \rho + k + 1)\ \frac{(it)^r}{r!} \\ &= \Omega_0\ {}_2F_1[1, k; \rho + k + 1; \gamma_1\ e^{it} + \gamma_2\ e^{2it}] \end{aligned} \quad (2.20)$$

By using (2.10) with 1, k and $\rho + k + 1$ replaced by 2, $k + 1$ and $\rho + k + 2$ respectively, we obtain

$$\Omega_0\ {}_2F_1[2, k + 1; \rho + k + 2; \gamma_1\ e^{it} + \gamma_2\ e^{2it}] = \sum_{r=0}^{\infty} \mu_r(2, k + 1; \rho + k + 2) \frac{(it)^r}{r!}. \quad (2.21)$$

Differentiate (2.21) with respect to t to get

$$\begin{aligned} \sum_{r=0}^{\infty} i\mu_{r+1}(1, k; \rho + k + 1) \frac{(it)^r}{r!} &= \Omega_0 \frac{i(\gamma_1 e^{it} + 2\gamma_2 e^{2it})\ k}{\rho + k + 1} \\ &\quad \times {}_2F_1[2, k + 1; \rho + k + 2; \gamma_1 e^{it} + \gamma_2 e^{2it}], \end{aligned} \quad (2.22)$$

which on simplification gives

$$\begin{aligned} \frac{\rho + k + 1}{k} \sum_{r=0}^{\infty} \mu_{r+1}(1, k; \rho + k + 1) \frac{(it)^r}{r!} &= \Omega_0 \Big[(\gamma_1 e^{it} + 2\gamma_2 e^{2it}) \\ &\quad \sum_{r=0}^{\infty} \mu_r(2, k + 1; \rho + k + 2) \frac{(it)^r}{r!} \Big] \end{aligned} \quad (2.23)$$

On expanding the exponential function in (2.23) and applying (1.2) to obtain

$$\begin{aligned} = \Omega_0 \Bigg[& \gamma_1 \sum_{r=0}^{\infty} \sum_{s=0}^{r} \mu_{r-s}(2, k + 1; \rho + k + 2) \frac{(it)^r}{(r - s)!s!} \\ & + \gamma_2 \sum_{r=0}^{\infty} \sum_{s=0}^{r} 2^{s+1} \mu_{r-s}(2, k + 1; \rho + k + 2) \frac{(it)^r}{(r - s)!s!} \Bigg] \end{aligned} \quad (2.24)$$

Equating the coefficients of $(it)^r\ (r!)^{-1}$ on both sides of (2.24), we get (2.19). □

Result 2.9. The following is a simple recursion formula for factorial moments $\psi_{[m]} = \psi_{[m]}(1, k; \rho + k + 1)$ of the MGD, for $m \geq 0$

$$\left(\frac{\rho + k + 1}{k}\right) \psi_{[m+1]}(1, k; \rho + k + 1) = \Omega_0 \, [(\gamma_1 + 2 \, \gamma_2) \, \psi_{[m]}(2, k + 1; \rho + k + 2) + 2 \, m \, \gamma_2 \, \psi_{[m-1]}(2, k + 1; \rho + k + 2)] \tag{2.25}$$

Proof. The factorial moment generating function $\Lambda(t)$ of the MGD (2.11) is given by

$$\Lambda(t) = \sum_{m=0}^{\infty} \psi_{[m]} \frac{t^m}{m!} \tag{2.26}$$

$$= \Omega_0 \, {}_2F_1[1, k; \rho + k + 1; \gamma_1 \, (t + 1) + \gamma_2[(t + 1)^2]]. \tag{2.27}$$

From (2.11) with 1, k and $\rho + k + 1$ changed by 2, $k + 1$ and $\rho + k + 2$ respectively, we have

$$\Omega_0 \, {}_2F_1[2, k+1; \rho+k+2; \gamma_1 \, (t+1)+\gamma_2(t+1)^2] = \sum_{m=0}^{\infty} \psi_m(2, k+1; \rho+k+2) \frac{t^m}{m!}. \tag{2.28}$$

By differentiating (2.26) with respect to t, we get

$$\sum_{m=0}^{\infty} \psi_{[m+1]}(1, k; \rho + k + 1) \frac{t^m}{m!} = \Omega_0 \, [\frac{k \, (\gamma_1 + 2\gamma_2 \, (t + 1))}{\rho + k + 1} \times {}_2F_1[2, k + 1; \rho + k + 2; \gamma_1 \, t + \gamma_2(t + 1)^2] \tag{2.29}$$

Substituting (2.29) in (2.28) , we have

$$\frac{\rho + k + 1}{k} \sum_{m=0}^{\infty} \psi_{[m+1]}(1, k; \rho + k + 1) \frac{t^m}{m!} = \Omega_0[(\gamma_1 + 2\gamma_2) + 2\gamma_2 t] \times \sum_{m=0}^{\infty} \psi_{[m]}(2, k + 1; \rho + k + 2) \frac{t^m}{m!} \tag{2.30}$$

$$= \Omega_0[(\gamma_1 + 2\gamma_2) \sum_{m=0}^{\infty} \psi_{[m]}(2, k + 1; \rho + k + 2) \frac{t^m}{m!} + 2\gamma_2 \sum_{m=0}^{\infty} m\psi_{[m]}(2, k + 1; \rho + k + 2) \frac{t^{m+1}}{m!}] \tag{2.31}$$

By equating the coefficients of $t^m \, (m!)^{-1}$ on both sides of (2.31), we get (2.25). □

3. Estimation

Here we estimating the un-known parameters ρ, k, γ_1 and γ_2 of the MGD by the method of maximum likelihood and generalized ratio test procedure applied for testing the significance of the parameters of the model.

Let $a(y)$ be the observed frequency of y events based on the observations from a sample with independent components and let z be the highest value of the y observed. The likelihood function of the sample is

$$L = \sum_{y=0}^{z} [q_y]^{a(y)}, \tag{3.1}$$

which implies

$$\ln L = \sum_{y=0}^{z} a(y) \ln q_y. \tag{3.2}$$

Let $\hat{\rho}$, $\hat{k}$, $\hat{\gamma_1}$ and $\hat{\gamma_2}$ be the MLEs of ρ, k, γ_1 and γ_2 respectively. Now the MLEs of the parameters are obtained by solving the following likelihood equations, obtained from (3.2) on differentiation with respect to ρ, k, γ_1 and γ_2 respectively and equating to zero. Then

$$\frac{\partial \log L}{\partial \rho} = 0 \tag{3.3}$$

or equivalently

$$\sum_{y=0}^{z} a(y) \sum_{j=0}^{[\frac{y}{2}]} [\frac{1}{\epsilon(y;\gamma_1,\gamma_2)} \frac{\binom{y-j}{j} (k)_{y-j} \gamma_1^{y-2j} \gamma_2^{j}}{(\rho+k+1)_{y-j}} [\upsilon(\rho+k+1+y-j) - \upsilon(\rho+k+1)]$$

$$+\Omega_0 \sum_{r=0}^{\infty} \frac{(k)_r (\gamma_1+\gamma_2)^r}{(\rho+k+1)_r} [\upsilon(\rho+k+1) - \upsilon(\rho+k+r+1)]] = 0,$$

$$\frac{\partial \log L}{\partial k} = 0 \tag{3.4}$$

or equivalently

$$\sum_{y=0}^{z} a(y) [\Omega_0 \sum_{r=0}^{\infty} \frac{(k)_r (\gamma_1+\gamma_2)^r}{(\rho+k+1)_r} [-\upsilon(\rho+k+1) - \upsilon(\rho+k+r+1) + \upsilon(k+r) - \upsilon(k)]$$

$$+ \sum_{j=0}^{[\frac{y}{2}]} \frac{1}{\epsilon(y;\gamma_1,\gamma_2)} \frac{\binom{y-j}{j}(k)_{y-j}\gamma_1^{y-2j}\gamma_2^{j}}{(\rho+k+1)_{y-j}} [-\upsilon(\rho+k+1+y-j)$$

$$+\upsilon(\rho+k+1) + \upsilon(k+y-j) - \upsilon(k)] = 0,$$

$$\frac{\partial \log L}{\partial \gamma_1} = 0 \tag{3.5}$$

or equivalently

$$\sum_{y=0}^{z} a(y) \; -\Omega_0 \sum_{r=0}^{\infty} \frac{(k)_r \; r \; (\gamma_1+\gamma_2)^{r-1}}{(\rho+k+1)_r} + \sum_{j=0}^{[\frac{y}{2}]} [\frac{1}{\epsilon(y;\gamma_1,\gamma_2)} \frac{\binom{y-j}{j} (k)_{y-j} (y-2j) \gamma_1^{y-2j-1} \gamma_2^{j}}{(\rho+k+1)_{y-j}} = 0,$$

$$\frac{\partial \log L}{\partial \gamma_2} = 0 \tag{3.6}$$

or equivalently

$$\sum_{y=0}^{z} a(y) \; - \Omega_0 \sum_{r=0}^{\infty} \frac{(k)_r \; r \; (\gamma_1+\gamma_2)^{r-1}}{(\rho+k+1)_r} + \sum_{j=0}^{[\frac{y}{2}]} [\frac{1}{\epsilon(y;\gamma_1,\gamma_2)} \; \frac{\binom{y-j}{j} \; (k)_{y-j} \; \gamma_1^{y-2j} \; j \; \gamma_2^{j-1}}{(\rho+k+1)_{y-j}} = 0,$$

where

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

$$\upsilon(\rho, y-j) = \varphi(\rho) - \varphi(\rho+y-j)$$

and

$$\xi(\rho^*) = \gamma_1 \; \gamma_2 \; \frac{(1)_{y-j} \; (k)_{y-j}}{(\rho+k+1)_{y-j}} \; \Omega_{y-j}(\gamma_1+\gamma_2).$$

On solving these log-likelihood equations by using MATHEMATICA one can obtain the MLE's of the parameters ρ, k, γ_1 and γ_2 of the MGD.

4. Applications

For numerical illustration, we have considered three real life data applications, of which the first data set taken from Greenwood and Yule (1920) is on the number of accidents experienced by each 414 industrial workers in a munitions factory over a three months period. The second data set is on the number of food stores in the Ljubljuna taken from Douglas (1980). While the third data set is on frequencies of the observed number of days that expereienced thunderstorm events at Cape Kennedy, Florida for the month of June, 1967 from Falls et al (1971). All the three data sets have been fitted by the proposed model MGD and compared with the existing models the DGYD, the EGGD, the EYD, the EZILSD and the Waring Distribution(WD). The numerical results obtained are included in the Table 1, 2 and 3. For model comparsion we have considered the information measures like AIC, BIC ... . Based on the computed values of chi-square statistic it can be seen that the MGD gives best fit to all the data sets and the values of AIC, BIC and P-value support its suitability compared to models the DGYD, the EGGD, the EYD, the EZILSD and the WD.

Table 1: The number of accidents experienced by each 414 industrial workers in a munitions factory.

x	Observed	Expected frequency by MLE					
		DGYD	EGGD	EYD	EZILSD	WD	MGD
0	296	305.19	281.65	276.87	305.90	306.15	285.58
1	74	71.44	74.60	87.28	70.09	59.06	77.07
2	26	22.85	35.25	31.11	23.14	21.72	30.12
3	8	8.59	13.44	11.59	9.69	11.35	11.10
4	4	3.15	5.50	4.42	3.24	5.88	5.03
5	4	1.56	2.19	1.71	1.10	3.98	3.13
6	1	0.72	0.88	0.66	0.53	2.66	1.41

Continued ⋯

x	Observed	Expected frequency by MLE					
		DGYD	EGGD	EYD	EZILSD	WD	MGD
7	0	0.34	0.35	0.26	0.22	1.94	0.39
8	1	0.16	0.14	0.10	0.09	1.36	0.17
Total	414	414	414	414	414	414	414
d.f		1	1	2	2	3	1
Estimates of parameters		ρ=1.02 $\gamma = 0.61$ k=1.25	γ_1=0.24 γ_2=0.047	ρ=-0.70 γ=0.41	γ_1=0.46 γ_2=0.01	ρ=-0.60 $k = 0.77$	ρ=1.77 k=0.64 γ_1=0.42 γ_2=0.02
χ^2-*value*		3.63	5.46	6.42	5.67	7.62	2.30
AIC		769.58	771.56	774.78	768.06	789.48	765.36
BIC		770.18	771.96	775.18	768.46	789.88	766.16
P Value		0.056	0.019	0.043	0.058	0.054	0.137

Table 2: Number of food stores in the Ljubljuna.

x	Observed	Expected frequency by MLE					
		DGYD	EGGD	EYD	EZILSD	WD	MGD
0	83	101.66	77.76	72.15	87.83	114.33	84.37
1	18	23.12	22.20	34.41	19.96	16.50	18.85
2	13	8.68	20.80	17.81	15.74	5.48	15.36
3	9	5.12	9.45	9.46	8.04	3.65	9.28
4	7	2.68	6.61	5.09	5.35	1.77	6.05
5	7	1.45	3.61	2.76	3.35	0.96	4.51
6	2	0.87	2.26	1.50	2.24	0.65	3.23
7	5	0.42	1.31	0.82	1.49	0.46	2.35
Total	144	144	144	144	144	144	144
d.f		1	3	3	3	1	1
Estimates of parameters		ρ=1.71 $\gamma = 0.82$ k=0.91	γ_1=0.26 γ_2=0.20	ρ=-0.84 γ=0.56	γ_1=0.46 γ_2=0.23	ρ=-0.82 k=0.19	$\rho = 1.55$ k=0.40 γ_1=0.43 γ_2=0.19
χ^2-*value*		54.49	10.59	27.15	8.16	83.74	2.09
AIC		446.10	434.54	451.34	434.16	528.64	430.90
BIC		446.31	434.75	451.55	434.37	528.85	432.11
P Value		0.00001	0.0141	0.00001	0.0428	0.00001	0.1526

Table 3: Frequencies of the observed number of days that expereienced thunderstorm events at Cape Kennedy, Florida for the month of June, 1967 .

x	Observed	Expected frequency by MLE					
		DGYD	EGGD	EYD	EZILSD	WD	MGD
0	187	210.31	205.89	166.44	207.72	197.44	182.68
1	77	73.98	61.20	84.29	66.88	71.21	81.67
2	40	27.97	34.84	41.43	28.22	30.74	36.64
3	17	11.26	15.62	20.83	14.54	15.06	16.62
4	6	4.20	7.34	9.78	6.93	7.58	7.37
5	2	1.64	3.41	4.84	3.69	4.85	3.46
6	1	0.64	1.60	2.39	2.02	3.12	1.56
Total	330	330	330	330	330	330	330
d.f		1	3	3	3	3	1
Estimates of parameters		ρ=-0.90 $\gamma = 0.40$	γ_1=0.29 γ_2=0.08	ρ=-1.05 γ=0.48	γ_1=0.64 $\gamma_2=2\times10^{-6}$	ρ=-0.98 k=1.70	ρ=4.85 k=3.15 γ_1=0.463 γ_2=0.0012
χ^2*-value*		11.78	7.81	7.85	10.345	7.49	1.75
AIC		795.60	803.14	818.24	808.16	806.64	786.40
BIC		795.45	803.04	816.24	806.06	806.54	788.20
P Value		0.0005	0.0501	0.0492	0.0158	0.0578	0.1958

5. Testing of hypothesis

In this section, we present the generalized likelihood ratio test (GLRT) procedure for testing the significance of the parameters of the MGD. We consider the following tests:

(1) Test 1: ${H_0}^{(1)} : \gamma_2 = 0$ against ${H_1}^{(1)} : \gamma_2 \neq 0$
(2) Test 2: ${H_0}^{(2)} : \rho = 0, k = 1$ against ${H_1}^{(2)} : \rho \neq 0, k \neq 0$
(3) Test 3: ${H_0}^{(3)} : \gamma_2 = 0, k = 1$ against ${H_1}^{(3)} : \gamma_2 \neq 0, k \neq 1$

. The test statistic is

$$-2 \ln \Lambda = 2 \left(\ln L(\hat{\underline{\Lambda}}; x) - \ln L(\hat{\underline{\Lambda}}^*; x) \right), \tag{5.1}$$

in which $\hat{\Lambda}$ is the MLE of $\underline{\Lambda} = (\gamma_1, \gamma_2, \rho, k)$ with no restriction and $\hat{\underline{\Lambda}}^*$ is the MLE of $\underline{\Lambda}$ under H_0 . The test statistic $-2 \log \Lambda$ is asymptotically distributed as a chi-square with one degree of freedom. For details of the GLRT, see Rao (1947). We have computed the values of $\ln L(\hat{\underline{\Lambda}}; x)$, $\ln L(\hat{\underline{\Lambda}}^*; x)$ and the test statistic in case of all the three data sets and inserted in Table 4.

Table 4: Test statistic value and Chi square value for GLRT in case of all the three data sets

Data sets		$\log L(\underline{\hat{\Lambda}}; x)$	$\log L(\underline{\hat{\Lambda}}^*; x)$	Test Statistic	d f	Chi square value
Data set 1	Test 1	-378.68	-381.79	6.22	1	3.84
	Test 2	-378.68	-385.39	13.22	2	5.99
	Test 3	-378.68	-382.03	6.70	2	5.99
Data set 2	Test 1	-211.95	-220.05	16.20	1	3.84
	Test 2	-211.95	-223.67	23.44	2	5.99
	Test 3	-211.95	-215.08	6.26	2	5.99
Data set 3	Test 1	-390.20	-394.80	9.20	1	3.84
	Test 2	-390.20	-406.12	31.84	2	5.99
	Test 3	-390.20	-401.08	21.76	2	5.99

Since the calculated value of the test statistic is greater than the critical value at 5% level of significance in case of all the three data sets and hence we can conclude that the parameters of the fitted model MGD is significant.

6. Simulation

As the MLE's of the parameters of the MGD are not in explicit form, we have carried out a simulation study for assessing the efficiency of the MLE's of the parameters of the distribution as follows. We have simulated data sets of size(n) 200, 300 and 500 from MGD for the two sets of parameters i)ρ=0.20, k=1.5,γ_1=0.50, γ_2=0.10 ii)ρ=-0.62, k=0.85,γ_1=0.32,γ_2=0.04. We have generated 100 independent samples of size n=200, 300 and 500 from MGD and computed the MLE's for each of the 100 samples using the statistical software MATHEMATICA. By using simulated observations, we estimated the parameters ρ, k, γ_1 and γ_2 of the MGD and thereby computed the values of the absolute bias and standard errors of each of the estimators. The estimated biases and the estimated standard errors are presented in Table 5. From Table 5, we can see 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 5: Absolute bias and standard errors in the parenthesis of the estimators of the parameters ρ, k, γ_1 and γ_2 of the MGD for the simulated data sets.

Parameter set	Sample size	MLE			
		$\hat{\rho}$	$\hat{k}$	$\hat{\gamma_1}$	$\hat{\gamma_2}$
(i)	$n = 200$	0.27 (0.344)	0.82 (1.228)	0.09 (0.034)	0.07 (0.0085)
	$n = 300$	0.11 (0.308)	0.18 (0.572)	0.05 (0.015)	0.025 (0.0033)
	$n = 500$	0.098 (0.134)	0.158 (0.244)	0.0308 (0.0086)	0.0056 (0.00043)

Continued $\cdots$

	Sample size	MLE			
		$\hat{\rho}$	$\hat{k}$	$\hat{\gamma_1}$	$\hat{\gamma_2}$
(ii)	$n = 200$	0.86 (1.44)	0.49 (1.12)	0.07 (0.024)	0.0267 (0.0024)
	$n = 300$	0.336 (0.678)	0.34 (0.88)	0.065 (0.0073)	0.0184 (0.0016)
	$n = 500$	0.19 (0.0572)	0.322 (0.213)	0.029 (0.0028)	0.0032 (0.0006)

References

1. Douglas, J.B.: Analysis with standard contagious distribution: Statistical Distributions in Scientific Work (ed.G.P.Patil), International cooperative publishing house, Fairland, (1980), 241-262.
2. Everitt, B.S and Hande, D.J.: Finite mixture distributions, Chapman and Hall, London, 1981.
3. Falls, W.L, Williford, O.W and Carter, C.M.: Probability distributions for thunderstorm activity at Cape Kennedy, Florida, Journal of Applied Materorology, 10, (1971), 97-104.
4. Greenwood, M and Yule, G.U.: 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, (1920), 255-304.
5. Johnson, N. L., Kemp, A. W and Kotz, S.: Univariate Discrete Distributions (III). John Wiley & Sons, New York, 2005.
6. Kumar, C.S and Riyaz, A.: On Zero-Inflated logarithmic series distribution and its modification, Statistica, 73(4), (2013), 477-492.
7. Kumar,C.S and Harisankar. S.: (2019) On Some Aspects of a General Class of Yule distribution and its Applications, Communication in Statistics - Theory and Methods, (2019), DOI: 10.1080/03610926.2019.1584308.
8. Martínez-Rodríguez A.M, Sáez-Castillo, A. J and Conde-Sánchez, A.: Modelling using an extended Yule distribution. Computational Statistics and Data Analysis, 55(1), (2011), 863-873.
9. Mathai, A. M. and Haubold, H. J.: Special Functions for Applied Statistics. Springer, New York, 2008.
10. McLachlan. G and Peel. D.: Finite mixture models. Wiley, New York, 2000.
11. Rao, C. R.: Minimum variance and the estimation of several parameters. Mathematical Proceedings of the Cambridge Philosophical Society, 43, (1947), 280-283.
12. Slater, L. J.: Generalized Hyper-geometric Functions: Cambridge University Press, Cambridge, 1966.
13. Titterington, D.M, Smith, A.F.M and Markov, U.E.: Statistical analysis of finite mixture distributions, Wiley, New York, 1985.
14. Yule, G. U.: A mathematical theory of evolution based on the conclusion of Dr. J. C. Wills. Philosophical Society of Royal Society of London. Series B, 213, (1920), 21-87.

C. Satheesh Kumar: Department of Statistics, University of Kerala, Thiruvananthapuram:695 581, India.
E-mail address: drcsatheesh@gmail.com

Harisankar. S: Department of Statistics, University of Kerala, Thiruvananthapuram:695 581, India.
E-mail address: harisankar83ster@gmail.com