ONLINE FIRST ARTICLE
doi.org/10.36253/ijas-16753

Statistica Applicata - Italian Journal of Applied Statistics

Received: 30.09.22/ **Accepted:** 20.03.25
Published online: 02.09.25

ON A WIDE CLASS OF GENERALIZED GEOMETRIC DISTRIBUTION FOR OVER AND UNDER DISPERSED DATA SETS

C. Satheesh Kumar[1]
Department of Statistics, University of Kerala, Thiruvananthapuram, India.

Harisankar, S
Asian School of Business, Technocity, Thiruvananthapuram, India.

Abstract *In this paper, a wide class of generalized geometric distribution is introduced and we name this class of distributions as "the alpha generalized geometric distribution (AGGD)". Several important distributions are obtained as a special cases of this proposed model. Important distributional properties such as generating functions, moments, recursive relations of the proposed distribution are examined. Parameter estimation using maximum likelihood is discussed. Three well-known data sets, having long tails, are analyzed and the results of fitting by various models are provided. Further, the generalized likelihood ratio test procedure is considered for testing the significance of the parameters of the GGD. Finally, performance of the different estimation methods are compared by means of a Monte Carlo simulation.*

Keywords: *Generalized Likelihood Ratio Test; Model Selection; Maximum Likelihood Estimation; Probability generating function; Simulation*

1. Introduction

Many generalizations of the geometric distribution have been attempted by researchers. Notable among them are (Jain and Consul, 1971) obtained a generalized negative binomial distribution. (Philippou et al., 1983) developed a generalized geometric distribution and some of its properties. (Tripathi et al., 1987) examined some generalizations of the geometric distribution derived by length biasing some generalized versions of the log-series distribution. (Makcutek, 2008) constructed a generalization of right truncated geometric distribution to modeling rank frequencies of graphemes. (Gomez-Deniz, 2010) proposed and studied a new generalization of the geometric distribution by using the (Marshall and Olkin, 1997) scheme.

[1] drcsatheeshkumar@gmail.com

(Kumar and Harisankar, 2020) considered a three parameter family of discrete distribution, namely “the doubly generalized Yule distribution (DGYD)" which we developed by compounding a geometric distribution with the generalized beta distribution, which combines two well known class of distributions namely GYD of (Mishra, 2009) and EYD of Martinez Rodriguez et al. (2011). We derived several important theoretical properties of DGYD which will help us to derive similar properties of both the classes of GYD and EYD. The p.g.f of the DGYD is the following.

$$G(t) = A_0 \, {}_2F_1(1, \eta; \gamma + \eta + 1; \lambda t), \tag{1}$$

where

$$A_0 = {}_2F_1(1, \eta, \gamma + \eta + 1; \lambda)^{-1}, \tag{2}$$

in which $\gamma > -(\eta + 1)$, $\eta > 0$ and $0 < \lambda < 1$.

Here,

$${}_2F_1(a, b, c; \lambda) = \sum_{r=0}^{\infty} \frac{(a)_r (b)_r}{(c)_r} \frac{\lambda^r}{r!},$$

is the Gaussian hypergeometric function (GHF); in which $(a)_r = a(a+1)\cdots(a+r-1)$ any positive integer r with $(a)_0 = 1$. For more details regarding GHF, see (Slater, 1966) or (Mathai and Haubold, 2008).
(Kumar and Harisankar, 2021) suggested another general class of distributions namely “the generalized geometric distribution (GGD)" which are more suitable for over-dispersed data sets, which obtained through compounding geometric and beta distribution and named it as “the generalised geometric distribution (GGD)". Further, we suggest an extended form of it and termed it as “the extended generalized geometric distribution (EGGD)" for creating more flexibility in modelling aspects. Here it can be seen that the standard geometric distribution is a special case of both the models. The p.g.f of the GGD is the following for $\gamma > -1$, $\eta > 0$ and $0 < \lambda < 1$.

$$H(t) = \, {}_2F_1[1, \eta; \gamma + \eta + 1; \lambda(t-1)], \tag{3}$$

Through this paper we propose a flexible count distribution as a generalization of both the DGYD and the GGD which we named as “the alpha generalized geometric distribution (AGGD)" and study some of its important properties. Since the proposed class of distribution enjoys both over-dispersed and under-dispersed

nature it can be viewed as a useful model in several practical situation. Its application in modeling emperical modeling of other count data is assessed by conducting comparative data fitting experiments. It is found that the proposed four parameter generalized geometric distribution performs well in all the examples considered here. In section 2, we present the definition of the AGGD and derive its p.m.f., expression of generating functions, mean, variance, recursion formulae for its probabilities, raw moments and factorial moments. In section 3, we discuss the estimation of the parameters of the AGGD by the method of maximum likelihood and consider generalized likelihood ratio test procedure for testing the significance of the parameters of the AGGD. So after investigating various useful properties, three data sets are considered here to justify its applicability. In section 4, we carried out a simulation study for examining the performance of the maximum likelihood estimators of the parameters of the distribution.
Further we need the following series representation in the sequel.

$$\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}A(s,r)=\sum_{r=0}^{\infty}\sum_{s=0}^{r}A(s,r-s) \tag{4}$$

2. Properties

Here, first we present the definition of the AGGD and discuss some of its important statistical properties.

Definition 2.1 *A non-negative integer valued random variable W is said to follow AGGD if its p.g.f is of the following form, in which* $\gamma > -(\eta+1)$, $\eta > 0$, $0 < \lambda < 1$, $-1 < \alpha < 1$ *and* $0 < \lambda + \alpha < 1$.

$$G_W(t) \quad = \quad A_0^* \, {}_2F_1(1,\eta;\gamma+\eta+1;\alpha+\lambda t), \tag{5}$$

where

$$A_i^* = [{}_2F_1(1+i,\eta+i,\gamma+\eta+1+i;\alpha+\lambda)]^{-1}, \quad for\ i=0,1,2,... \tag{6}$$

The following table gives some of the special cases of AGGD.

Table 1: Special cases of the AGGD

No.	*Distribution*	*Parameter restriction*
1	Doubly generalized Yule distribution (DGYD) of (Kumar and Harisankar, 2020a)	$\alpha = 0$
2	Generalized geometric distribution(GGD) of (Kumar and Harisankar, 2020b)	$\alpha = -\lambda$
3	Hyper-logarithmic distribution (HLD) of (Tripathi and Gurland, 1985)	$\eta = 1$, $\alpha = 0$, $\gamma = \lambda - 1$
4	Generalized geometric distribution (GGD) of (Tripathi and Gupta, 1987)	$\eta = 2$, $\alpha = 0$, $\gamma = a - 2$
5	Extended Yule distribution(EYD) of (Martinez-Rodriguez et al., 2011)	$\eta = 1$, α=0
6	Generalized Yule distribution(GYD) of (Mishra, 2009)	α=0, $\lambda \longrightarrow 1$
7	Alpha logarithmic series distribution (ALSD) (Kumar and Riyaz, 2013a)	$\gamma = 0$, η=1
8	Alternative zero-inflated logarithmic series distribution (AZILSD) of (Kumar and Riyaz, 2013b)	$\alpha = -\lambda$, γ=0, $\eta = 1$
9	Yule distribution(YD) of (Yule, 1925)	$\eta = 1$, $\lambda \longrightarrow 1$, α=0
10	Geometric distribution	$\gamma = -1$, α=0
11	Logarithmic series distribution	$\gamma = 0$, η=1, α=0

Result 2.1 *The p.m.f* g_w *of AGGD with p.g.f* (5) *is the following, for* $w = 0, 1, 2, ...,$ $\gamma > -(\eta+1)$, $\eta > 0$, $0 < \lambda < 1$ *and* $-1 < \alpha < 1$ *with* $0 < \lambda + \alpha < 1$.

$$g_w = A_0^* \ \frac{(\eta)_{w)}\ \lambda^w\ \beta_w(\alpha)}{(\gamma+\eta+1)_w}, \tag{7}$$

where $\beta_w(\alpha) = {}_2F_1(1+w, \eta+w, \gamma+\eta+1+w; \alpha)$ *and* A_0^* *is as defined in* (6).

Proof. From (5) we have the following:

$$G_W(t) = A_0^*\ {}_2F_1(1, \eta; \gamma+\eta+1; \lambda\ t+\alpha) \tag{8}$$

$$= \sum_{w=0}^{\infty} g_w\ t^w \tag{9}$$

On expanding the Gauss hypergeometric function in (8), we get

$$G_W(t) = A_0^* \sum_{w=0}^{\infty} \frac{(\eta)_w}{(\gamma+\eta+1)_w}\ [\lambda\ t+\alpha]^w. \tag{10}$$

By applying (4) and binomial theorem in (10), we get the following.

$$
\begin{aligned}
G_W(t) &= A_0^* \sum_{w=0}^{\infty} \frac{(\eta)_w}{(\gamma+\eta+1)_w} \sum_{n=0}^{w} \binom{w}{n} (\lambda\, t)^{w-n}\, \alpha^n \\
&= A_0^* \sum_{w=0}^{\infty}\sum_{n=0}^{\infty} \frac{(\eta)_{w+n}}{(\gamma+\eta+1)_{w+n}} \binom{w+n}{n} (\lambda\, t)^{w}\, \alpha^n. \qquad (11)
\end{aligned}
$$

On equating the coefficients of t^w on the right hand side expressions of (9) and (11) we get the following.

$$
G_W(t) = A_0^* \frac{(\eta)_w \lambda^w}{(\gamma+\eta+1)_w} \sum_{n=0}^{\infty} \frac{(1+w)_n\, (\eta+w)_n\, \alpha^n}{(\gamma+\eta+1+w)_n\, n!}. \qquad (12)
$$

Hence the result. □

Result 2.2 *The moment generating function $M_W(t)$ of the AGGD is the following, for any $t \in R$*

$$
M_W(t) = A_0^*\, {}_2F_1(1,\eta;\gamma+\eta+1;\alpha+\lambda\, e^t). \qquad (13)
$$

Result 2.3 *The cumulant generating function $K_W(t)$ of the AGGD is the following, for any $t \in R$ and $i = \sqrt{-1}$.*

$$
K_W(t) = \ln[A_0^*] + \ln[{}_2F_1(1,\eta;\gamma+\eta+1;\alpha+\lambda\, e^{it})]. \qquad (14)
$$

Result 2.4 *The characteristic function $\psi_W(t)$ of the AGGD is the following, for any $t \in R$ and $i = \sqrt{-1}$.*

$$
\psi_W(t) = A_0^*\, {}_2F_1(1,\eta;\gamma+\eta+1;\alpha+\lambda\, e^{it}). \qquad (15)
$$

Result 2.5 *The factorial moment generating function $F_W(t)$ of the AGGD with p.g.f is given by*

$$
F_W(t) = A_0^*\, {}_2F_1[1,\eta;\gamma+\eta+1;\ \alpha+\lambda\, (t+1)]. \qquad (16)
$$

Result 2.6 *The mean and variance of the AGGD are the following,*

$$
Mean = \frac{A_0^*\, \eta\, \lambda}{A_1^*(\gamma+\eta+1)} \qquad (17)
$$

and

$$
Variance = \frac{A_0^*\, (\eta)\lambda}{A_1^*(\gamma+\eta+1)} \left[\frac{(\eta+1)\, \lambda}{A_2^*(\gamma+\eta+2)} + 1 - \frac{A_0^*\, \eta\, \lambda}{A_1^*(\gamma+\eta+1)} \right]. \qquad (18)
$$

Proof. It follows from the fact that,

$$Mean = G_W^1(1)$$

and

$$Variance = G_W^2(1) + G_W^1(1) - [G_W^1(1)]^2,$$

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

Result 2.7 *From (17) and (18), it can be seen that the AGGD is under-dispersed when $-(\eta+1) < \gamma < -\eta$ and over-dispersed when $\gamma > -\eta$ provided*

$$\frac{\eta\ (\eta+1)}{A_2^*(\gamma+\eta+2)} > \frac{A_0^*\ \eta}{A_1^*(\gamma+\eta+1)}$$

Next we obtain certain recurrence relations for probabilities, raw moments and factorial moments of the AGGD. For the same we need the following notations, for $i = 0, 1, 2, ...$

$$\gamma^* + i \quad = \quad (1+i, \eta+i, \gamma+\eta+1+i) \tag{19}$$

$$R_i \quad = \quad \frac{(1+i)(\eta+i)}{(\gamma+\eta+1+i)} \tag{20}$$

Result 2.8 *For $w \geq 1$, the following is a simple recurrence relation for probabilities $g_w = g_w(\gamma^*)$ of the AGGD with p.g.f (5).*

$$(w+1)\ g_{w+1}(\gamma^*) = \frac{A_0^*\ R_0\ \lambda}{A_1^*}\ g_w(\gamma^*+1), \tag{21}$$

in which R_0 and γ^* are defined in (19) and (20) respectively.

Proof. From (5), we have

$$G_W(t) = \sum_{w=0}^{\infty} g_w(\gamma^*)\ t^w = A_0^*\ {}_2F_1(1, \eta; \gamma+\eta+1; \lambda\ t+\alpha). \tag{22}$$

Differentiating the equation (22) with respect to t, we get

$$\sum_{w=0}^{\infty} (w+1)\ g_{w+1}(\gamma^*)\ t^w = A_0^*\ R_0\ \lambda\ \ {}_2F_1(2, \eta+1; \gamma+\eta+2; \alpha+\lambda\ t). \tag{23}$$

In (22) by replacing γ^* with γ^*+1 respectively, we obtain

$$\sum_{w=0}^{\infty} g_w(\gamma^*+1)\ t^w = A_1^*\ {}_2F_1(2, \eta+1; \gamma+\eta+2; \alpha+\lambda\ t). \tag{24}$$

Substitute (24) in (23) to get

$$\sum_{w=0}^{\infty}(w+1)\;g_{w+1}(\gamma^*)\;t^w = \frac{A_0^*\;R_0\;\lambda}{A_1^*}\;\sum_{w=0}^{\infty}g_w(\gamma^*+1)\;t^w. \tag{25}$$

Equating the coefficient of t^w on both sides of (25), we get (21). □

Result 2.9 *The following is a simple recurrence relation for raw moments* $\mu_r^{'} = \mu_r^{'}(\gamma^*)$ *of the AGGD, for* $r \geq 0$.

$$\mu_{r+1}^{'}(\gamma^*) = \frac{A_0^* R_0\;\lambda}{A_1^*}\sum_{s=0}^{r}\binom{r}{s}\;\mu_{r-s}^{'}(\gamma^*+1) \tag{26}$$

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

$$\psi_W(t) \;=\; \sum_{r=0}^{\infty}\mu_r^{'}(\gamma^*)\;\frac{(it)^r}{r!} \tag{27}$$

$$=\; A_0^*\;{}_2F_1(1,\eta;\gamma+\eta+1;\alpha+\lambda\;e^{it}). \tag{28}$$

By using (15) with γ^* replaced by γ^*+1 respectively, we obtain

$$\sum_{r=0}^{\infty}\mu_r^{'}(\gamma^*+1)\frac{(it)^r}{r!} = A_1^*\;{}_2F_1(2,\eta+1;\gamma+\eta+2;\alpha+\lambda\;e^{it}). \tag{29}$$

Differentiate (27) and (28) with respect to t, to get

$$\sum_{r=0}^{\infty} i\;\mu_{r+1}^{'}(\gamma^*)\frac{(it)^r}{r!} = A_0^*\;i\;R_0\;\lambda\;e^{it}\;{}_2F_1(2,\eta+1;\gamma+\eta+2;\alpha+\lambda\;e^{it}),$$

which on simplification gives

$$\sum_{r=0}^{\infty}\mu_{r+1}^{'}(\gamma^*)\frac{(it)^r}{r!} = \frac{A_0^*\;\;R_0\;\lambda\;e^{it}}{A_1^*}\sum_{r=0}^{\infty}\mu_r^{'}(\gamma^*+1)\frac{(it)^r}{r!}. \tag{30}$$

On expanding the exponential functions in (30) and applying (4) to obtain

$$\sum_{r=0}^{\infty}\mu_{r+1}^{'}(\gamma^*)\frac{(it)^r}{r!} = \frac{A_0^*\;\;R_0\;\lambda}{A_1^*}\left[\sum_{r=0}^{\infty}\sum_{s=0}^{r}\mu_{r-s}^{'}(\gamma^*+1)\frac{(it)^r}{(r-s)!s!}\right]. \tag{31}$$

Equating the coefficients of $(it)^r\;(r!)^{-1}$ on both sides of (31), we get (26). □

Result 2.10 *The following is a simple recurrence relation for factorial moments* $\mu_{[r]} = \mu_{[r]}(\gamma^*)$ *of the AGGD, for* $r \geq 0$.

$$\mu_{[r+1]}(\gamma^*) = \frac{A_0^* \, R_0 \, \lambda}{A_1^*} \, \mu_{[r]}(\gamma^* + 1). \tag{32}$$

Proof. The factorial moment generating function $F_{W_1}(t)$ of the AGYD is given by

$$F_W(t) = \sum_{r=0}^{\infty} \mu_{[r]} \frac{t^r}{r!} \tag{33}$$

$$= A_0^* \, {}_2F_1[1, \eta; \gamma + \eta + 1; \alpha + \lambda \ (t+1)] \tag{34}$$

From (16) with γ^* changed by $\gamma^* + 1$ respectively, we have

$$\sum_{r=0}^{\infty} \mu_{[r]}(\gamma^* + 1) \frac{t^r}{r!} = A_1^* \, {}_2F_1[2, \eta + 1; \gamma + \eta + 2; \alpha + \lambda \ (t+1)]. \tag{35}$$

On differentiating (33) and (34) with respect to t, we get

$$\sum_{r=0}^{\infty} \mu_{[r+1]}(\gamma^*) \frac{t^r}{r!} = A_0^* \, R_0 \, \lambda \, {}_2F_1[2, \eta + 1; \gamma + \eta + 2; \alpha + \lambda \ (t+1)]. \tag{36}$$

Applying (35) in (36) to obtain

$$\sum_{r=0}^{\infty} \mu_{[r+1]}(\gamma^*) \frac{t^r}{r!} = \frac{A_0^* \, R_0 \, \lambda}{A_1^*} \left[\sum_{r=0}^{\infty} \mu_{[r]}(\gamma^* + 1) \frac{t^r}{r!} \right]. \tag{37}$$

By equating the coefficients of $t^r \ (r!)^{-1}$ on both sides of (37), we get (32). □

2.1. Estimation and Testing of Hypothesis

In this section we discuss the estimation of the parameters γ, η, λ and α of the AGGD by MLE and thereafter the GLRT procedure is utilized for testing the significance of the parameters γ, η, λ and α of the AGGD.

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

$$L = \prod_{w=0}^{y} [g_w]^{a(w)}, \tag{38}$$

which implies

$$\ln L = \sum_{w=0}^{y} a(w) \ln g_w. \tag{39}$$

Let $\hat{\gamma}$, $\hat{\eta}$, $\hat{\lambda}$ and $\hat{\alpha}$ be the MLEs of γ, η, λ and α respectively. Now the MLEs of the parameters are obtained by solving the following likelihood equations, obtained from (39) on differentiation with respect to γ, η, λ and α respectively and equating to zero. Then

$$\frac{\partial \ln L}{\partial \gamma} = 0$$

or equivalently

$$\begin{aligned}
&\sum_{w=0}^{y} a(w) \left[-A_0^* \sum_{r=0}^{\infty} \frac{(\eta)_r \, (\alpha+\lambda)^r}{(\gamma+\eta+1)_r} [\psi(\gamma+\eta+1) - \psi(\gamma+\eta+r+1)] \right. \\
&-\frac{1}{(\gamma+\eta+1)_w} \left(\psi(\gamma+\eta+1) - \psi(\gamma+\eta+w+1)\right) \\
&+\frac{1}{\beta_w} \sum_{r=0}^{\infty} \frac{(1+w)_r \, (\eta+w)_r \, \alpha^r}{(\gamma+\eta+w+1)_r \, r!} \, [\psi(\gamma+\eta+w+1) \\
&\left. -\psi(\gamma+\eta+w+r+1)] \right] = 0
\end{aligned} \tag{40}$$

$$\frac{\partial \ln L}{\partial \eta} = 0$$

or equivalently

$$\begin{aligned}
&\sum_{w=0}^{y} a(w) \left[-A_0^* \sum_{r=0}^{\infty} \frac{(\eta)_r \, (\alpha+\lambda)^r}{(\gamma+\eta+1)_r} [\psi(\gamma+\eta+1) - \psi(\gamma+\eta+r+1) \right. \\
&+\psi(\eta+r) - \psi(\eta)] + \frac{1}{(\eta)_w} [\psi(\eta+w) - \psi(\eta)] - \frac{1}{(\gamma+\eta+1)_w} [\psi(\gamma+\eta+w+1) \\
&-\psi(\gamma+\eta+w+r+1)] + \frac{1}{\beta_w} \sum_{r=0}^{\infty} \frac{(1+w)_r \, (\eta+w)_r \, (\alpha)^r}{(\gamma+\eta+w+1)_r \, r!} \\
&\left. \times [\psi(\gamma+\eta+w+1) - \psi(\gamma+\eta+w+r+1) + \psi(\eta+w+r) - \psi(\eta+w)] \right] = 0
\end{aligned} \tag{41}$$

$$\frac{\partial \ln L}{\partial \lambda} = 0$$

or equivalently

$$\sum_{w=0}^{y} a(w) \left[\frac{w}{\lambda} - \frac{A_0^* \, \eta}{(\gamma+\eta+1) \, A_1^*} \right] = 0 \tag{42}$$

and

$$\frac{\partial \ln L}{\partial \alpha} = 0$$

or equivalently

$$\sum_{w=0}^{y} a(w)\,[-\frac{A_0^*\,\eta}{A_1^*\,(\gamma+\eta+1)} - \frac{\beta_{w+1}\,(1+w)\,(\eta+w)}{\beta_w\,(\gamma+\eta+w+1)}] = 0, \tag{43}$$

where

$$\psi(\gamma) \quad = \quad [\Gamma(\gamma)]^{-1}\,\frac{d\,\Gamma(\gamma)}{d\gamma}.$$

It is important to note that the likelihood equations do not always guarantee a unique solution, and in some cases, the maximum of the likelihood function may occur at the boundary of the parameter domain. To address this, we computed the second-order partial derivatives of lnL with respect to the parameters α, λ, γ and η. Using MATHEMATICA, we verified that these second-order partial derivatives are consistently negative for the parameter ranges $\gamma > -(\eta+1)$, $\eta > 0$, $0 < \lambda < 1$ and $-1 < \alpha < 1$ with $0 < \lambda + \alpha < 1$. This confirms that the AGGD is log-concave within these restrictions, ensuring that the MLEs of α, λ, γ and η are unique and correspond to the global maximum of the likelihood function.

To further ensure that the MLEs do not occur at the boundary of the parameter space and that the regularity conditions for the validity of the asymptotic Chi-squared distribution are satisfied, we computed the Hessian matrix of lnL and verified that it is negative definite, further supporting the log-concavity of the likelihood function. Thus, the uniqueness and validity of the MLEs are established under the specified parameter restrictions. We have used the FindMaximum function of MATHEMATICA 11.2 then we get the maximum likelihood estimators of the parameters α, λ, γ and η of the AGGD.

Here we present the GLRT procedure for testing the significance of the parameters of the AGGD. Here we consider the following tests:

Test 1: ${H_0}^{(1)} : \alpha = 0$ against ${H_1}^{(1)} : \alpha \neq 0$

Test 2: ${H_0}^{(2)} : \alpha = 0, \eta = 1$ against ${H_1}^{(2)} : \alpha \neq 0, \eta \neq 1$

Test 3: ${H_0}^{(3)} : \alpha = 0, \lambda = 1$ against ${H_1}^{(3)} : \alpha \neq 0, \lambda \neq 1$

The test statistic is

$$-2\ln\Lambda = 2\left(\ln L(\underline{\hat{\Lambda}}; w) - \ln L(\underline{\hat{\Lambda}}^*; w)\right), \tag{44}$$

in which $\hat{\Lambda}$ is the MLE of $\underline{\Lambda} = (\alpha, \lambda, \gamma, \eta)$ with no restriction and $\hat{\underline{\Lambda}}^*$ is the MLE of $\underline{\Lambda}$ under H_0 . The test statistic (44) is asymptotically distributed as a χ^2 with one degree of freedom in case of Test 1 and two degrees of freedom in case of Test 2 and 3 respectively. We have computed the values of $\ln L(\hat{\underline{\Lambda}}; w)$, $\ln L(\hat{\underline{\Lambda}}^*; w)$ and the test statistic in case of all the three data sets and inserted in Table 5.

2.2. Applications

For numerical illustration, we have considered three real-life data applications, of which first data sets deals with the distribution of the plant Lespedeza Capitaita as reported by (Thomson, 1952) using contagious distributions. later it was studied by (Katti and Gurland, 1961). To illustrate the utility of the AGGD distribution we have considered a second data about the number of public secondary schools in the 771 municipalities of Andalusia (Spain). The data is taken from (Rodriguez et al., 2017). They have obtained the data from the database of the System of Multi territorial Information of Andalusia. Whereas, the third data represents the counts of cysts of kidneys using steroids from (Chan et al., 2010). We have fitted the GGD (Kumar and Harisankar, 2020), the EYD (Martínez-Rodríguez et al, 2011), the AGGD and the Waring Distribution (WD) to these data sets and the results obtained along with the corresponding values of the expected frequencies, χ^2 statistic, d.f, AIC, BIC and AICc in respect of each of the models are presented in Tables (2), (3) and (4) respectively. Based on the computed values of the χ^2 statistic, AIC, BIC and AICc it can be observed that the AGGD gives better fit to all the three data sets considered here compared to the existing models the GGD, the EYD and the WD if the data is heavy-tailed.

Table 2: Observed frequencies and computed values of expected frequencies of the GGD, the EYD, the WD and the AGGD by MLE for the first data set.

w	Observed frequency	Expected frequency by MLE			
		GGD	EYD	WD	AGGD
0	7178	7156.29	6933.50	7128.28	7171.01
1	286	310	652.49	263.43	287.42
2	93	96.40	100.61	87.74	94.96
3	40	39.68	27.27	47.31	41.55
4	24	18.28	15.07	30.04	21.49
5	7	8.98	5.78	20.95	9.81

Continued ...

w	Observed frequency	Expected frequency by MLE			
		GGD	EYD	WD	AGGD
6	5	5.20	2.77	15.54	5.95
7	1	2.43	1.35	12.04	3.37
8	2	1.30	0.60	9.64	1.95
9	1	0.71	0.36	7.91	1.15
10	2	0.39	0.15	6.61	0.68
11	1	0.22	0.04	5.64	0.41
12	0	0.12	0.01	4.86	0.25
Total	7640	7640	7640	7640	7640
d.f		4	4	9	3
Estimates of parameters		γ=0.06 $\lambda = 0.62$ $\eta = 0.08$	γ=8.07 λ=0.99	γ=0.76 $\eta = 0.14$	γ=0.10 λ=0.65 $\eta = 0.14$ α=0.0005
χ^2 *value*		4.94	149.94	55.40	1.44
AIC		4602.80	4884.34	4608.70	4598.20
BIC		4599.98	4882.46	4606.82	4594.64
AICc		4604.80	4883.90	4609.80	4603.20

Table 3: Observed frequencies and computed values of expected frequencies of the GGD, the EYD, the WD and the AGGD by MLE for the second data set.

w	Observed frequency	Expected frequency by MLE			
		GGD	EYD	WD	AGGD
0	528	526.42	547.36	537.36	527.37
1	152	148.28	123.43	132.39	144.65
2	50	55.73	45.34	48.46	50.81
3	15	21.52	21.07	21.94	17.30
4	8	8.37	11.29	11.38	10.17
5	4	5.37	6.66	6.49	6.96
6	1	2.44	4.21	3.97	4.26
7	3	1.28	2.81	2.56	2.75
8	0	0.85	1.95	1.73	1.85

Continued ...

w	Observed frequency	Expected frequency by MLE			
		GGD	EYD	WD	AGGD
9	1	0.42	1.40	1.21	1.28
10	2	0.21	1.03	0.87	0.92
11	0	0.07	0.78	0.64	0.67
12	1	0.03	0.60	0.48	0.50
13	1	0.01	0.47	0.37	0.38
14	0	0.004	0.38	0.29	0.29
15	1	0.001	0.30	0.23	0.23
16	0	0.0007	0.25	0.18	0.18
17	0	0.0003	0.20	0.15	0.14
18	0	0.0001	0.17	0.12	0.12
19	1	0.00004	0.14	0.10	0.09
20	3	0.0000	0.12	0.08	0.08
Total	771	771	771	771	771
d.f		3	5	5	3
Estimates of parameters		γ=-0.91 $\lambda = 0.40$ $\eta = 0.22$	γ=2.39 λ=0.99	γ=3.0 $\eta = 1.31$	γ=2.82 λ=0.96 $\eta = 1.38$ α=0.0065
χ^2 *value*		17.25	11.59	9.64	2.11
AIC		1718.64	1619.08	1622.80	1616.92
BIC		1716.40	1617.72	1621.44	1614.20
AICc		1719.84	1619.48	1623.40	1619.42

Table 4: Observed frequencies and computed values of expected frequencies of the GGD, the EYD, the WD and the AGGD by MLE for the third data set.

w	Observed frequency	Expected frequency by MLE			
		GGD	EYD	WD	AGGD
0	65	57.49	61.40	63.61	65.19
1	14	19.27	20.17	20.95	14.27
2	10	12.62	10.05	9.68	9.55
3	6	8.84	5.76	5.77	6.44

Continued ...

w	Observed frequency	Expected frequency by MLE			
		GGD	EYD	WD	AGGD
4	4	5.12	4.23	3.21	4.23
5	2	3.08	3.11	2.10	3.11
6	2	1.76	1.67	1.45	2.49
7	2	0.94	1.20	1.04	1.64
8	1	0.48	0.88	0.77	1.28
9	1	0.25	0.65	0.59	0.85
10	1	0.11	0.50	0.46	0.51
11	2	0.04	0.38	0.37	0.44
Total	110	110	110	110	110
d.f		2	3	2	1
Estimates of parameters		γ=-0.54 $\lambda = 0.63$ $\eta = 2.62$	γ=0.60 λ=0.88	γ=1.91 $\eta = 1.50$	γ=-0.86 λ=0.73 $\eta = 0.15$ α=0.001
χ^2 *value*		4.85	2.55	4.86	0.52
AIC		358.88	346.84	348	342.06
BIC		356.18	345.04	346.20	338.46
AICc		360.38	348.14	349.30	347.86

Table 5: Test statistic value and table value for GLRT in case of all the three data sets

Data sets		$\ln L(\hat{\underline{\Lambda}}; w)$	$\ln L(\hat{\underline{\Lambda}}^*; w)$	Test Statistic	d.f	Tabled value
Data set 1	Test 1	-2292.10	-2295.40	6.60	1	3.84
	Test 2	-2295.10	-2440.17	290.14	2	5.99
	Test 3	-2295.10	-2302.35	14.50	2	5.99
Data set 2	Test 1	-805.46	-856.22	101.52	1	3.84
	Test 2	-805.46	-809.54	8.16	2	5.99
	Test 3	-805.46	-809.40	9.80	2	5.99
Data set 3	Test 1	-167.03	-176.44	18.82	1	3.84
	Test 2	-167.03	-171.42	8.78	2	5.99
	Test 3	-167.03	-172	9.94	2	5.99

From Table 5, it is observed that, for all three tests, the calculated value of the test statistic exceeds the critical value corresponding to the 0.95 quantile of the Chi-squared distribution. Therefore, it can be concluded that the parameters of the fitted AGGD model are significant across all three datasets at the 5% level of significance.

2.3. 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 AGGD, EYD and WD and then estimated the parameters by the maximum-likelihood method. The true parameter values used in the data generating processes for AGGD are i) γ= -0.45, $\eta = 0.55$, α=0.25, λ=0.05 (under-dispersed) ii) γ= 0.30, $\eta = 1.20$, α=0.45, λ=0.02 (over-dispersed). The true parameter values utilized in the data generation processes for EYD are i) γ= -0.25, λ=0.08 (under-dispersed) ii) γ= 0.35, λ=0.12 (over-dispersed) and The true parameter values applied in the data generation processes for WD are i) γ= 0.45, $\eta = 0.55$ (over-dispersed) ii) γ= 0.30, $\eta = 1.20$ (over-dispersed). By using simulated observations, we estimated the parameters of the AGGD, the EYD and the WD and thereby computed the values of the absolute bias and mean squared errors of each of the estimators. The results obtained are presented in Tables 6, 8 and 7 from which it can be observed that both the absolute values of bias and MSE of the estimators of the parameters are in decreasing order as the sample size increases. The simulation results demonstrate that the AGGD outperforms the other two distributions in terms of bias and mean square error (MSE), particularly as sample size increases. The lower bias indicates that the AGGD provides more accurate estimates closer to the true parameter value, while the reduced MSE reflects improved precision and accuracy, combining lower variability with minimal systematic error. This suggests that AGGD exhibits strong asymptotic properties, making it highly effective for larger datasets. Its superior performance highlights its potential as a robust and reliable method for parameter estimation, contributing to statistical methodology and practical applications where accuracy and precision are critical.

Table 6: Absolute bias and MSE (in the parenthesis) of the estimators of the parameters γ, η, α and λ of the AGGD for the simulated data sets.

		MLE			
Parameter set	*Sample size*	$\hat{\gamma}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\lambda}$
(i)	$n = 100$	0.31 (0.262)	0.40 (0.346)	0.12 (0.081)	0.06 (0.065)
	$n = 200$	0.18 (0.185)	0.24 (0.142)	0.06 (0.041)	0.01 (0.024)
	$n = 500$	0.09 (0.064)	0.16 (0.041)	0.004 (0.0012)	0.005 (0.0011)
(ii)	$n = 100$	0.17 (0.111)	0.61 (0.282)	0.15 (0.051)	0.036 (0.0013)
	$n = 200$	0.08 (0.082)	0.31 (0.104)	0.04 (0.0021)	0.0051 (0.0002)
	$n = 500$	0.03 (0.0001)	0.11 (0.024)	0.008 (0.0006)	0.0002 (0.00008)

Table 7: Absolute bias and MSE (in the parenthesis) of the estimators of the parameters γ and λ of the EYD for the simulated data sets.

		MLE	
Parameter set	*Sample size*	$\hat{\gamma}$	$\hat{\lambda}$
(i)	$n = 100$	0.45 (0.462)	0.48 (0.46)
	$n = 200$	0.38 (0.285)	0.44 (0.242)
	$n = 500$	0.21 (0.094)	0.28 (0.082)
(ii)	$n = 100$	0.55 (0.25)	0.92 (0.42)
	$n = 200$	0.44 (0.182)	0.31 (0.164)
	$n = 500$	0.24 (0.08)	0.20 (0.12)

Table 8: Absolute bias and MSE (in the parenthesis) of the estimators of the parameters γ and η of the WD for the simulated data sets.

		MLE	
Parameter set	*Sample size*	$\hat{\gamma}$	$\hat{\eta}$
(i)	$n = 100$	0.65 (0.562)	0.98 (0.66)
	$n = 200$	0.52 (0.451)	0.66 (0.352)
	$n = 500$	0.30 (0.221)	0.42 (0.149)
(ii)	$n = 100$	0.45 (0.28)	0.72 (0.38)
	$n = 200$	0.38 (0.176)	0.56 (0.155)
	$n = 500$	0.23 (0.10)	0.44 (0.13)

2.4. Summary and Conclusion

In this paper, we introduced the alpha generalized geometric distribution (AGGD), a univariate distribution that demonstrates remarkable flexibility in modeling both over-dispersed and under-dispersed datasets. The AGGD encompasses several important generalized geometric models and well-known generalized distributions as special cases, highlighting its versatility.

We derived key statistical properties of the distribution and estimated its parameters using the method of maximum likelihood estimation (MLE). A test procedure was proposed to evaluate the significance of the parameters, and the results were compared with those obtained from existing models, showcasing the advantages of the AGGD.

To validate the performance of the proposed estimators, a simulation study was conducted, confirming their effectiveness under different scenarios. The findings suggest that the AGGD provides a compelling alternative to existing models for the types of datasets considered in this study.

While this paper presents several characteristic properties and inferential aspects of the model, further exploration of its properties and applications remains an avenue for future research, which we aim to publish in due course.

References

[1] Chan, S. K., Riley, P. R., Price, K. L., McElduff, F., Winyard, P. J., Welham, S. J., Woolf, A. S., and Long, D. A. (2010). Corticosteroid-induced kidney dysmorphogenesis is associated with deregulated expression of known cystogenic molecules, as well as Indian hedgehog. American Journal of Physiology-Renal Physiology, 298(2), 346-356.

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

[3] Gupta, R. C., Gupta, P. L., Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. Commun. Statist. Theor Meth, 39(15), 2767-2788.

[4] Gupta, R. D., Kundu, D. (2007). Generalized exponential distribution: Existing results and some recent developments. J. Statist. Plann. Infer, 137, 3537-3547.

[5] Jain GC, Consul PC (1971) A generalized negative binomial distribution. SIAM J Appl Math, 21, 501-513.

[6] Katti, S.K. and Gurland, J. (1961). The Poisson Pascal Distribution. Biometrics, 17(4), 527-538.

[7] 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.

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

[9] Kumar, C.S. and Riyaz, A. (2013b). An alternative to the logarithmic-with-zeros distribution and its properties and estimation. South African Statistical Journal, 48(2), 253-266.

[10] Kumar,C.S. and Harisankar. S. (2020a). On Some Aspects of a General Class of Yule distribution and its Applications, Commun. Statist. Theor Meth, 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] Kumar, C. S. and Harisankar, S. (2020d). Modeling over-dispersed datasets using generalized geometric distributions. Journal of Statistical Computation and Simulation, 90(4), 606-623.

[13] Makcutek J (2008). A generalization of the geometric distribution and its application in quantitative linguistics. Romanian Rep Phys, 60(3), 501-509.

[14] Marshall, A.W, Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika, 84(3), 641-652.

[15] 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.

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

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

[18] Mudholkar, G. S, Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure data. IEEE Trans. Reliab, 42, 299-302.

[19] Philippou, A.N, Georghiou, C, Philippou, G.N. (1983). A generalized geometric distribution and some of its properties. Stat Probab Lett, 1, 171-175.

[20] RodrÃguez-Avi, J. and Olmo-JimÃ©nez, M. J. (2017). A regression model for overdispersed data without too many zeros. Statistical Papers, 58(3), 749-773.

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

[22] Thomson, G. W. (1952). Measures of plant aggregation based on contagious distribution. Contrib. Lab. Vert. Biol. U. Mich, 53, pp. 16.

[23] Tripathi, R. C. and Gupta, R. C. (1985). A generalization of the log-series distribution, Communications in Statistics-Theory and Methods, 14, 1779-1799.

[24] Tripathi, R. C., Gupta, R. C., and White, T. J. (1987). Some generalizations of the geometric distribution, Sankhya, Series B, 49, 218-223