



# **Discrete Additive Weibull Geometric Distribution**

K. Jayakumar<sup>1</sup> and M. Girish Babu<sup>2,\*</sup>

<sup>1</sup>Department of Statistics, University of Calicut, Malappuram, Kerala-673 635, India <sup>2</sup>Department of Statistics, Government Arts and Science College, Meenchanda, Kozhikode, Kerala-673 018, India

#### **ARTICLE INFO**

#### ABSTRACT

Article History Received 02 August 2017 Accepted 13 February 2018

#### Keywords

Additive Weibull distribution discrete Weibull distribution geometric distribution hazard rate function order statistics Weibull distribution Discretizing a continuous distribution has received much attention among researchers recently. Discrete analogue of the wellknown continuous distributions such as Normal, Exponential, Weibull, Laplace, Rayleigh, and so on, are available in the literature. In this paper, we introduce a discrete version of the additive Weibull geometric distribution of Elbatal *et al.* [1]. Discrete Weibull, discrete modified Weibull, discrete Weibull geometric, discrete exponential geometric, discrete Rayleigh distribution, and so on, are sub models of this distribution. We study some properties of the new distribution. The hazard rate function of the new distribution is monotonically increasing or decreasing or bathtub shape based on the values of the shape parameters. The method of maximum likelihood estimation is used for estimating the model parameters. A simulation study is carried out to show the performance of the maximum likelihood estimate of parameters of the new distribution. An application of this distribution to a real data set is also presented.

© 2019 *The Authors*. Published by Atlantis Press SARL. This is an open access article distributed under the CC BY-NC 4.0 license (http://creativecommons.org/licenses/by-nc/4.0/).

## 1. INTRODUCTION

There are situations where continuous random variables may not necessarily always be measured on a continuous scale but may often be counted as discrete random variable. For example, in military service, the weapons like tanks, what is more important is the number of times it fires until failure than the life of the weapon. Similar situations frequently occur in reliability and survival analysis. By discretizing the continuous distribution, several discrete lifetime distributions are developed in the literature. Some of them are discrete Weibull distribution in Nakagawa and Osaki [2], a second type of discrete Weibull distribution in Stein and Dattero [3], a third type of discrete Weibull distribution in Padgett and Spurrier [4], discrete exponential distribution in Sato *et al.* [5], discrete normal distribution in Roy [6], discrete Rayleigh distribution in Roy [7], discrete Laplace distributions in Krishna and Pundir [10], discrete inverse Weibull distribution in Jazi *et al.* [11], discrete generalized exponential distribution in Gómez-Déniz [12], discrete additive Weibull (AW) distribution in Bebbington *et al.* [15], discrete Lindley distribution in Bakouch *et al.* [16], discrete Gumbel distribution in Chakraborty and Chakravarty [17], exponentiated geometric distribution in Chakraborty and Gupta [18], discrete distribution related to generalized gamma distribution in Chakraborty [19], transmuted geometric distribution in Chakraborty and Bhati [20], discrete Weibull geometric (DWG) distribution in Jayakumar and Babu [21], and so on.

Discretization plays a vital role in variable selection method, in addition to transforming the continuous variable to discrete variable. This method can significantly make an impact on the performance of classification algorithms applied in the analysis of high-dimensional biomedical data. While constructing the discrete version of a continuous distribution, one may preserve one or more characteristic properties of the continuous one. There are different methodologies available in the literature about the discretization of a continuous distribution (see Bracquemond and Gaudoin [22], Chakraborty [23]).

Discretization of the distribution of a continuous random variable *X*, to its discrete analogue, say *Y*, using the method of survival functions is given by

$$P(Y = y) = P(X \ge y) - P(X \ge y + 1) = S_X(y) - S_X(y + 1); y = 0, 1, 2, ...,$$
(1)

where, Y = [X] = largest integer less than or equal to X and  $S_X(.)$  is the survival function of the random variable X.

<sup>\*</sup> Corresponding author. Email: giristat@gmail.com

Xie and Lai [24] proposed the AW distribution by combining the failure rates of two Weibull distributions of which one has a decreasing failure rate and the other has an increasing failure rate. The cumulative distribution function (cdf) of *AW* distribution is given by

$$F(x;\alpha,\beta,\gamma,\delta) = 1 - e^{-\left(\alpha x^{\beta} + \gamma x^{\delta}\right)},$$
(2)

where  $\alpha > 0$ ,  $\gamma > 0$  and  $\beta > \delta > 0$  or ( $\delta > \beta > 0$ ), which gives identifiability to the model. Here  $\alpha$  and  $\gamma$  are scale parameters, and  $\beta$  and  $\delta$  are shape parameters. Lemonte *et al.* [25] examined some structural properties of *AW* distribution.

Suppose  $X_1, X_2, ..., X_N$  are *N* independent and identically distributed (iid) random variables from *AW* distribution with cdf given in Eq. (2). Let *N* be a discrete random variable following geometric distribution (truncated at zero) with probability mass function (pmf) given by

$$P(N = n) = (1 - p) p^{n-1}; n = 1, 2, ...; 0 
(3)$$

Let  $X_{(1)} = Min \{X_i\}_{i=1}^N$ . Then the cdf of  $X_{(1)}|N = n$ , is given by,

$$G_{\{X_{(1)}|N=n\}}(x) = 1 - [1 - F(x)]^n = 1 - e^{-n\left(\alpha x^{\beta} + \gamma x^{\delta}\right)}.$$
(4)

Hence, the cdf of  $X_{(1)}$  is

$$F(x;\alpha,\beta,\gamma,\delta,p) = (1-p)\sum_{n=1}^{\infty} p^{n-1} \left[ 1 - e^{-n\left(\alpha x^{\beta} + \gamma x^{\delta}\right)} \right]$$
$$= \frac{1 - e^{-\left(\alpha x^{\beta} + \gamma x^{\delta}\right)}}{1 - p e^{-\left(\alpha x^{\beta} + \gamma x^{\delta}\right)}},$$
(5)

where x > 0,  $0 , <math>\alpha > 0$ ,  $\beta > 0$ ,  $\gamma > 0$  and  $\delta > 0$ . The distribution of  $X_{(1)}$  is called AW geometric and its survival function is given by,

$$S(x;\alpha,\beta,\gamma,\delta,p) = \frac{(1-p)e^{-(\alpha x^{\beta} + \gamma x^{\delta})}}{1 - pe^{-(\alpha x^{\beta} + \gamma x^{\delta})}}$$
(6)

This distribution is studied by Elbatal et al. [1].

The contents of the paper are arranged as follows: In Section 2, the discrete AW geometric (DAWG) distribution is introduced and in Section 3, various properties of this distribution including the structure of hazard rate function are studied. In Section 4, the maximum likelihood estimation (MLE) method is used for parameter estimation. Also a simulation study is carried out to study the performance of the maximum likelihood estimates of the new distribution. Application of this distribution in real data modeling is illustrated in Section 5 and conclusions are presented in Section 6.

#### 2. DAWG DISTRIBUTION

Marshall and Olkin [26] introduced a method of adding a parameter into a family of distributions. According to them if  $\overline{F}(x)$  denote the survival function of a continuous random variable *X*, then the usual device of adding a new parameter results in another survival function  $\overline{G}(x)$  is defined by

$$\overline{G}(x) = \frac{\theta \overline{F}(x)}{1 - \theta \overline{F}(x)}, -\infty < x < \infty, \ \theta > 0,$$
(7)

where  $\overline{\theta} = 1 - \theta$ . In particular when  $\theta = 1$ ,  $\overline{G}(x) = \overline{F}(x)$ .

Let *Y* be the discrete analogue of the continuous random variable *X* with survival function defined in Eq. (7). Gómez-Déniz [12] obtained the discrete analogue of Marshall–Olkin scheme by applying Eq. (7) in Eq. (1). The corresponding random variable *Y* has the pmf,

$$p_{Y}(y) = P(Y = y) = \frac{\theta\left[\overline{F}(y) - \overline{F}(y+1)\right]}{\left[1 - \overline{\theta F}(y)\right]\left[1 - \overline{\theta F}(y+1)\right]}.$$
(8)

Now, we apply the AW geometric distribution with survival function defined Eq. (6) in Eq. (8) and after re-parametrizations as  $\rho = e^{-\alpha}$  and  $\eta = e^{-\gamma}$ , then the pmf becomes,

$$p_{Y}(y) = \frac{\theta(1-p)\left[\rho^{y^{\beta}}\eta^{y^{\delta}} - \rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}}\right]}{\left[\left(1-\theta(1-p)\right)\rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}}\right]}, \quad y = 0, 1, 2, ...,$$
(9)

where  $\theta > 0, 0 \delta > 0$  (or  $\delta > \beta > 0$ ). We call this distribution as the generalized DAWG distribution. When  $\theta = 1$ , Eq. (9) becomes,

$$p_{Y}(y;p,\rho,\eta,\beta,\delta) = \frac{(1-p)\left(\rho^{y^{\beta}}\eta^{y^{\delta}} - \rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}}\right)}{\left(1 - p\rho^{y^{\beta}}\eta^{y^{\delta}}\right)\left(1 - p\rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}}\right)}, \quad y = 0, 1, 2, \dots,$$
(10)

where  $0 , <math>0 < \rho < 1$ ,  $0 < \eta < 1$ ,  $\beta > \delta > 0$  (or  $\delta > \beta > 0$ ). We call this distribution as DAWG distribution with parameters p,  $\rho$ ,  $\eta$ ,  $\beta$ , and  $\delta$  and is denoted as  $DAWG(p, \rho, \eta, \beta, \delta)$ . We have the following cases:

- 1. When  $\rho \uparrow 1$  or  $\eta \uparrow 1$ , then Eq. (10) reduces to DWG distribution introduced in Jayakumar and Babu [21].
- 2. When  $\eta = \rho$ ,  $\delta = \beta$ , then also it becomes DWG distribution with parameters  $\rho^2$  and  $\beta$ .
- 3. When  $\beta = 1$  and  $\eta = 1$ , it becomes discrete exponential geometric distribution.
- 4. When  $p \downarrow 0$  and  $\beta = 1$ , it becomes discrete modified Weibull distribution.
- 5. When  $p \downarrow 0$  and  $\eta = 1$ , then it becomes discrete Weibull distribution (Nakagawa and Osaki [2]) with parameters  $\rho$  and  $\beta$ .
- 6. When  $p \downarrow 0$ ,  $\beta = 2$ , and  $\eta = 1$ , then it becomes discrete Rayleigh distribution (Roy [7]).
- 7. When  $p \downarrow 0$ ,  $\beta = 1$ , and  $\eta = 1$ , then it becomes geometric distribution with parameter  $\rho$ .

### 3. STRUCTURAL PROPERTIES OF DAWG ( $p, \rho, \eta, \beta, \delta$ ) DISTRIBUTION

Figure 1, provides pmf plots of  $DAWG(p, \rho, \eta, \beta, \delta)$  distribution for various choices of parameter values. The probabilities can be calculated recursively using the following relation:

$$p_{Y}(y+1) = \frac{\left(1 - p\rho^{y^{\beta}}\eta^{y^{\delta}}\right) \left(\rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}} - \rho^{(y+2)^{\beta}}\eta^{(y+2)^{\delta}}\right)}{\left(1 - p\rho^{(y+2)^{\beta}}\eta^{(y+2)^{\delta}}\right) \left(\rho^{y^{\beta}}\eta^{y^{\delta}} - \rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}}\right)} p_{Y}(y).$$
(11)

From Gupta *et al.* [27], we have the distribution having pmf  $p_Y(y)$  is log-concave if and only if  $\left\{\frac{p_Y(y+1)}{p_Y(y)}\right\}_{y\geq 0}$  is decreasing and log-convex if and only if  $\left\{\frac{p_Y(y+1)}{p_Y(y)}\right\}_{y\geq 0}$  is increasing. Also, if the sequence  $\left\{\frac{p_Y(y+1)}{p_Y(y)}\right\}_{y\geq 0}$  is constant, then the hazard rate is constant and the distribution is geometric.

The cdf of *DAWG* ( $p, \rho, \eta, \beta, \delta$ ) distribution is,

$$F(y; p, \rho, \eta, \beta, \delta) = P(Y \le y) = 1 - S_X(y) + P(Y = y) = \frac{1 - \rho^{(y+1)^{\beta}} \eta^{(y+1)^{\delta}}}{1 - p \rho^{(y+1)^{\beta}} \eta^{(y+1)^{\delta}}},$$
(12)

where  $y = 0, 1, 2, ...; \beta > \delta > 0$  (or  $\delta > \beta > 0$ ),  $0 and <math>0 < \eta < 1$ . Here note that,  $F(0) = \frac{1-\rho\eta}{1-\rho\rho\eta}$  and the proportion of positive values is  $\frac{\rho\eta(1-\rho)}{1-\rho\rho\eta}$ .

The survival function of  $DAWG(p, \rho, \eta, \beta, \delta)$  distribution is given by,

$$S(y) = P(Y > y) = 1 - P(Y \le y) = \frac{(1-p)\rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}}}{1-p\rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}}}.$$
(13)

The hazard rate function of  $DAWG(p, \rho, \eta, \beta, \delta)$  distribution is

$$h(y) = P(Y = y/Y \ge y) = \frac{P(Y = y)}{P(Y \ge y)} = \frac{1 - \rho^{(y+1)^{\beta} - y^{\beta}} \eta^{(y+1)^{\delta} - y^{\delta}}}{1 - p\rho^{(y+1)^{\delta}} \eta^{(y+1)^{\delta}}},$$
(14)



**Figure 1** Plots of the probability mass function (pmf) of discrete additive Weibull geometric (*DAWG*)(p,  $\rho$ ,  $\eta$ ,  $\beta$ ,  $\delta$ ) distribution.

provided,  $P(Y \ge y) > 0$ . In Figure 2, we present the plot of hazard rate function of  $DAWG(p, \rho, \eta, \beta, \delta)$  distribution for various parameter values. When  $y \to 0$ , we have from Eq. (14)

$$h(y) \rightarrow \frac{1-\rho\eta}{1-\rho\rho\eta} = p_Y(0)$$

Now to study the limit of h(y) as  $y \to \infty$ , we consider the following five cases based on the values of the shape parameters  $\beta$  and  $\delta$ : **Case (i).** When  $\beta > 1$  and  $\delta > 1$  (provided  $\beta > \delta$  or  $\beta < \delta$ ).

Here note that  $\lim_{y\to\infty} h(y) = 1$ . In this case  $h(0) = \frac{1-\rho\eta}{1-\rho\rho\eta}$ ,  $h(1) = \frac{1-\rho^{2^{\beta}-1}\eta^{2^{\delta}-1}}{1-\rho\rho^{2^{\beta}}\eta^{2^{\delta}}}$ ,

 $h(2) = \frac{1-\rho^{3^{\delta}-2^{\delta}}\eta^{3^{\delta}-2^{\delta}}}{1-\rho\rho^{3^{\delta}}\eta^{3^{\delta}}}, \dots$  That is,  $h(0) < h(1) < h(2) < \dots < 1$ . Therefore, h(y) is an increasing function increases from  $\frac{1-\rho\eta}{1-\rho\rho\eta}$  to 1.

**Case (ii).** When  $\beta > 1$  and  $\delta = 1$ .

Here note that  $\lim_{y\to\infty} h(y) = 1$ . Also it can be seen that h(0) < h(1) < h(2) < ... < 1. Therefore, h(y) is an increasing function increases from  $\frac{1-\rho\eta}{1-\rho\rho\eta}$  to 1.



**Figure 2** Plots of the hazard rate function of discrete additive Weibull geometric (*DAWG*)(p,  $\rho$ ,  $\eta$ ,  $\beta$ ,  $\delta$ ) distribution.

**Case (iii).** When  $0 < \beta < 1$  and  $\delta > 1$ .

Here also  $\lim_{y\to\infty} h(y) = 1$ . But here h(y) is initially decreases from h(0) to the minimum point h(m) and then increases to 1. The minimum point *m* can be numerically identified by solving the conditions,  $h(m) - h(m-1) \le 0$  and  $h(m+1) - h(m) \ge 0$ .

#### **Case (iv).** When $0 < \beta < 1$ and $\delta = 1$ .

In this case  $\lim_{y\to\infty} h(y) = 1 - \eta$ . Also  $h(0) > h(1) > h(2) > ... > 1 - \eta$ . That is, h(y) is a decreasing function.

**Case (v).** When 
$$0 < \beta < 1$$
 and  $0 < \delta < 1$  (provided  $\beta > \delta$  or  $\beta < \delta$ ).

Here  $\lim_{y\to\infty} h(y) = 0$ . It can be shown that h(0) > h(1) > h(2) > ... > 0. That is, in this case also, h(y) is decreasing.

Figure 3, shows a comparison of all the five cases explained above.

The reverse hazard rate function is

$$h^{*}(y) = P(Y = y/Y \le y) = \frac{(1-p)\left(\rho^{y^{\beta}}\eta^{\delta} - \rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}}\right)}{\left(1 - p\rho^{y^{\beta}}\eta^{\delta}\right)\left(1 - \rho^{(y+1)^{\beta}}\eta^{(y+1)^{\delta}}\right)}.$$
(15)

The second rate of failure is

$$h^{**}(y) = \log\left\{\frac{S(y)}{S(y+1)}\right\} = \log\left\{\frac{\left(\frac{1}{\rho}\right)^{(y+2)^{\beta}} \left(\frac{1}{\eta}\right)^{(y+2)^{\delta}} - p}{\left(\frac{1}{\rho}\right)^{(y+1)^{\beta}} \left(\frac{1}{\eta}\right)^{(y+1)^{\delta}} - p}\right\}.$$
(16)



Figure 3 Plots of the hazard rate functions for the five cases.

The accumulated hazard function, H(y) is given by,

$$H(y) = \sum_{t=0}^{y} h(t) = \sum_{t=0}^{y} \frac{1 - \rho^{(t+1)^{\beta} - t^{\beta}} \eta^{(t+1)^{\delta} - t^{\delta}}}{1 - p \rho^{(t+1)^{\beta}} \eta^{(t+1)^{\delta}}}.$$
(17)

The mean residual life function (MRLF) is given by,

$$L(y) = E[(Y-y)|Y \ge y] = \frac{\sum_{j>y} S(j)}{S(y)} = \sum_{j\ge y} \prod_{t=y}^{j} (1-h(t))$$
$$= \sum_{j\ge y} \prod_{i=y}^{j} \frac{\rho^{(i+1)^{\beta}} \eta^{(i+1)^{\delta}} (1-p\rho^{i^{\beta}} \eta^{i^{\delta}})}{\rho^{i^{\beta}} \eta^{i^{\delta}} (1-p\rho^{(i+1)^{\beta}} \eta^{(i+1)^{\delta}})}; \quad y = 0, 1, 2, \dots.$$
(18)

#### 3.1. Quantile Function

Since the cdf of *DAWG* distribution is not invertible, we use the method discussed in Lemonte *et al.* [25] to obtain the quantile function. We take

$$F(y) = \frac{1 - \rho^{(y+1)^{\beta}} \eta^{(y+1)^{\delta}}}{1 - p \rho^{(y+1)^{\beta}} \eta^{(y+1)^{\delta}}} = u,$$

where  $u \in (0, 1)$ . This implies,

$$\left(y+1\right)^{\beta}\ln\left(\rho\right)+\left(y+1\right)^{\delta}\ln\left(\eta\right)=\ln\left(\frac{1-u}{1-up}\right).$$
(19)

We obtain the nonlinear equation,  $at^{\beta} + ct^{\delta} = x$ , where  $a = \ln(\rho), c = \ln(\eta), x = \ln\left(\frac{1-u}{1-up}\right)$  and t = y + 1. We can expand  $t^{\beta}$  in Taylor series as  $t^{\beta} = \sum_{k=0}^{\infty} (\beta)_k (t-1)^k / k! = \sum_{k=0}^{\infty} f_j t^j$ , where  $f_j = \sum_{k=j}^{\infty} (-1)^{k-j} \binom{k}{j} (\beta)^{[k]} / k!$ ,  $(\beta)_k = \beta (\beta - 1) \dots (\beta - k + 1)$  is the

falling factorial and  $(\beta)^{[k]} = \beta (\beta + 1) \dots (\beta + k - 1)$  is the ascending factorial. Analogously, we can expand  $t^{\delta}$  as  $t^{\delta} = \sum_{j=0}^{\infty} g_j t^j$ , where

$$g_j = \sum_{k=j}^{\infty} (-1)^{k-j} \binom{k}{j} \left(\delta\right)^{[k]} / k! . \text{Now,}$$

$$x = H(t) = \sum_{j=0}^{\infty} \left( af_j + cg_j \right) t^j = \sum_{j=0}^{\infty} h_j t^j,$$
(20)

where  $h_j = af_j + cg_j$ . To obtain an expansion for the quantile function of *DAWG* distribution we use the Lagrange's theorem. Now suppose that if the power series expansion holds

$$x = H(t) = h_0 + \sum_{j=1}^{\infty} h_j t^j, \quad h_1 = H'(t)|_{t=0} \neq 0,$$

where H(t) is analytic at a zero point, then the inverse power series  $t = H^{-1}(x)$  exists and is single-valued in the neighbourhood of the point x = 0, and is given by

$$t = H^{-1}(x) = \sum_{j=1}^{\infty} v_j x^j,$$

where the coefficients  $v_i$  are given by

$$v_{j} = \frac{1}{j!} \left( \frac{d^{j-1}}{dt^{j-1}} \left[ \phi(t) \right]^{j} \right) |_{t=0}, \quad \phi(t) = \frac{t}{H(t) - h_{0}}$$

Hence, the quantile function can be expressed as

$$Q(u) = \sum_{j=1}^{\infty} v_j \left( \ln \left( \frac{1-u}{1-up} \right)^j \right) - 1.$$
(21)

#### 3.2. Moments

The *r*<sup>th</sup> raw moment about origin is given by,

$$\mu'_{r} = E(Y^{r}) = \sum_{y=0}^{\infty} y^{r} \frac{\left(1-p\right) \left(\rho^{y^{\beta}} \eta^{y^{\delta}} - \rho^{\left(y+1\right)^{\beta}} \eta^{\left(y+1\right)^{\delta}}\right)}{\left(1-p \rho^{y^{\beta}} \eta^{y^{\delta}}\right) \left(1-p \rho^{\left(y+1\right)^{\beta}} \eta^{\left(y+1\right)^{\delta}}\right)}.$$

Since this expansion is not in a tractable form, for given values of p,  $\rho$ ,  $\eta$ ,  $\beta$  and  $\delta$ , the moments can be numerically computed using R programming. Table 1 shows the moments, skewness and kurtosis for *DAWG* distribution for given values of parameters.

**Table 1** Moments, skewness, and kurtosis for p = 0.9,  $\rho = 0.8$ ,  $\eta = 0.9$ , and various choices of  $\beta$  and  $\delta$ .

Parameter	Raw moments	Central moments	Skewness	Kurtosis
$\beta = 1.5$ $\delta = 2$	$\mu_1' = 0.27  \mu_2' = 0.45  \mu_3' = 0.98  \mu_4' = 2.70$	$\mu_2 = 0.38$ $\mu_3 = 0.65$ $\mu_4 = 1.82$	2.79	12.59
$\beta = 1.5$ $\delta = 1$	$\begin{array}{l} \mu_1' = 0.32 \\ \mu_2' = 0.73 \\ \mu_3' = 2.37 \\ \mu_4' = 10.24 \end{array}$	$\mu_2 = 0.63$ $\mu_3 = 1.73$ $\mu_4 = 7.59$	3.50	19.44
$\beta = 0.5$ $\delta = 1.5$	$\begin{array}{l} \mu_1' = 0.46 \\ \mu_2' = 1.76 \\ \mu_3' = 10.13 \\ \mu_4' = 76.97 \end{array}$	$\mu_2 = 1.55$ $\mu_3 = 7.88$ $\mu_4 = 60.28$	4.09	25.18
$\beta = 0.2$ $\delta = 0.9$	$\mu'_1 = 1.38\mu'_2 = 27.89\mu'_3 = 1092.96\mu'_4 = 63894.82$	$\mu_2 = 26.00 \\ \mu_3 = 983.02 \\ \mu_4 = 58185.39$	7.41	86.06

#### 3.3. Order Statistics

Let  $Y_1, Y_2, ..., Y_n$  be a random sample from  $DAWG(p, \rho, \eta, \beta, \delta)$  distribution. Also, let  $Y_{(1)}, Y_{(2)}, ..., Y_{(n)}$ , denotes the corresponding order statistics. Then the pmf and the cdf of  $k^{th}$  order statistic, say,  $Z = Y_{(k)}$ , are

$$f_{Z}(z) = \frac{n!}{(k-1)!(n-k)!} F^{k-1}(z) [1 - F(z)]^{n-k} f(z)$$

$$= \frac{n!}{(k-1)!(n-k)!} \frac{(1-p)^{(n-k+1)} \rho^{(n-k)(z+1)^{\beta}} \eta^{(n-k)(z+1)^{\delta}}}{(1-p\rho^{(z+1)^{\beta}} \eta^{(z+1)^{\delta}})^{n}}$$

$$= \frac{(\rho^{z^{\beta}} \eta^{z^{\delta}} - \rho^{(z+1)^{\beta}} \eta^{(z+1)^{\delta}}) (1-\rho^{(z+1)^{\beta}} \eta^{(z+1)^{\delta}})^{k-1}}{(1-p\rho^{z^{\beta}} \eta^{z^{\delta}})}, \qquad (22)$$

and

$$F_{Z}(z) = \sum_{i=k}^{n} \binom{n}{j} F^{j}(z) \left[1 - F(z)\right]^{n-j}$$
  
=  $\sum_{j=k}^{n} \binom{n}{j} \frac{\left(1 - p\right)^{n-j} \rho^{(n-j)(z+1)^{\beta}} \eta^{(n-j)(z+1)^{\delta}} \left(1 - \rho^{(z+1)^{\beta}} \eta^{(z+1)^{\delta}}\right)^{j}}{\left(1 - p\rho^{(z+1)^{\beta}} \eta^{(z+1)^{\delta}}\right)^{n}},$  (23)

respectively.

The pmf of the minimum is,

$$f_{Y_{(1)}}(z) = \frac{n\left(1-p\right)^{n}\rho^{(n-1)(z+1)^{\beta}}\eta^{(n-1)(z+1)^{\delta}}\left(\rho^{z^{\beta}}\eta^{z^{\delta}} - \rho^{(z+1)^{\beta}}\eta^{(z+1)^{\delta}}\right)}{\left(1-p\rho^{z^{\beta}}\eta^{z^{\delta}}\right)\left(1-p\rho^{(z+1)^{\beta}}\eta^{(z+1)^{\delta}}\right)^{n}},$$
(24)

and the pmf of the maximum is,

$$f_{Y_{(n)}}(z) = \frac{n\left(1-p\right)\left(1-\rho^{(z+1)^{\beta}}\eta^{(z+1)^{\delta}}\right)^{n-1}\left(\rho^{z^{\beta}}\eta^{z^{\delta}}-\rho^{(z+1)^{\beta}}\eta^{(z+1)^{\delta}}\right)}{\left(1-p\rho^{z^{\beta}}\eta^{z^{\delta}}\right)\left(1-p\rho^{(z+1)^{\beta}}\eta^{(z+1)^{\delta}}\right)^{n}}.$$
(25)

#### 3.4. Stress–Strength Parameter

The stress-strength parameter, R = P(Y > Z) is a measure of component reliability. Suppose that, the random variable *Y* is the strength of a component which is subjected to a random stress *Z*, the estimation of *R* when *Y* and *Z* are i.i.d has been considered in the literature. One may see Kotz *et al.* [28], for a review of stress-strength model. In the discrete case, the stress-strength model is defined as,

$$R = P(Y > Z) = \sum_{y=0}^{\infty} p_Y(y) F_Z(y)$$

where,  $p_Y(y)$  and  $F_Z(y)$  denotes the pmf and cdf of the independent discrete random variables *Y* and *Z*, respectively. The stress-strength models are applied in various fields such as Engineering, Psychology and Medicine.

Let,  $Y \sim DAWG(\theta_1)$  and  $Z \sim DAWG(\theta_2)$ , where,  $\theta_1 = (p_1, \rho_1, \eta_1, \beta_1, \delta_1)^T$  and  $\theta_2 = (p_2, \rho_2, \eta_2, \beta_2, \delta_2)^T$ . Then, from Eq. (10) and Eq. (12), we have,

$$R = \sum_{y=0}^{\infty} \frac{\left(1-p_{1}\right) \left(\rho_{1}^{y^{\beta_{1}}} \eta_{1}^{y^{\delta_{1}}} - \rho_{1}^{(y+1)^{\beta_{1}}} \eta_{1}^{(y+1)^{\delta_{1}}}\right) \left(1-\rho_{2}^{(y+1)^{\beta_{2}}} \eta_{2}^{(y+1)^{\delta_{2}}}\right)}{\left(1-p_{1} \rho_{1}^{y^{\beta_{1}}} \eta_{1}^{y^{\delta_{1}}}\right) \left(1-p_{1} \rho_{1}^{(y+1)^{\delta_{1}}} \eta_{1}^{(y+1)^{\delta_{1}}}\right) \left(1-p_{2} \rho_{2}^{(y+1)^{\beta_{2}}} \eta_{2}^{(y+1)^{\delta_{2}}}\right)}.$$
(26)

Assume that,  $(y_1, y_2, ..., y_n)$  and  $(z_1, z_2, ..., z_m)$  are independent observations drawn from  $DAWG(\theta_1)$  and  $DAWG(\theta_2)$ , respectively. The total likelihood function is given by,  $L_R(\theta^*) = L_n(\theta_1) L_m(\theta_2)$ , where,  $\theta^* = (\theta_1, \theta_2)$ . The score vector is given by,

$$U_R(\partial^*) = \left(\frac{\partial L_R}{\partial p_1}, \frac{\partial L_R}{\partial \rho_1}, \frac{\partial L_R}{\partial \eta_1}, \frac{\partial L_R}{\partial \beta_1}, \frac{\partial L_R}{\partial \delta_1}, \frac{\partial L_R}{\partial p_2}, \frac{\partial L_R}{\partial \rho_2}, \frac{\partial L_R}{\partial \eta_2}, \frac{\partial L_R}{\partial \beta_2}, \frac{\partial L_R}{\partial \delta_2}\right).$$

		$p_1 = 0.8, p_2 = 0.8$		
		$\rho_1 = 0.5, \rho_2 = 0.5$	$\eta_1 = 0.5, \eta_2 = 0.5$	
$\overline{(\beta_1,\delta_1)\to(\beta_2,\delta_2)\downarrow}$	(0.5,1)	(1,1.5)	(1.5,2)	(2,2.5)
(0.5, 1)	0.9404	0.9402	0.9402	0.9401
(1, 1.5)	0.9411	0.9410	0.9409	0.9409
(1.5, 2)	0.9413	0.9413	0.9412	0.9412
(2,2.5)	0.9413	0.9413	0.9413	0.9413
		$\rho_1 = 0.2, \rho_2 = 0.6$	$\eta_1 = 0.2, \eta_2 = 0.6$	
$(\beta_1, \delta_1) \rightarrow (\beta_2, \delta_2) \downarrow$	(0.5,1)	(1,1.5)	(1.5,2)	(2,2.5)
(0.5, 1)	0.8994	0.8993	0.8993	0.8993
(1, 1.5)	0.8996	0.8996	0.8995	0.8995
(1.5, 2)	0.8997	0.8997	0.8997	0.8996
(2,2.5)	0.8977	0.8997	0.8997	0.8997
		$p_1 = 0.5, p_2 = 0.8$		
		$\rho_1 = 0.5, \rho_2 = 0.5$	$\eta_1 = 0.5, \eta_2 = 0.5$	
$(\beta_1, \delta_1) \rightarrow (\beta_2, \delta_2) \downarrow$	(0.5,1)	(1,1.5)	(1.5,2)	(2,2.5)
(0.5, 1)	0.9443	0.9438	0.9436	0.9435
(1, 1.5)	0.9457	0.9455	0.9454	0.9453
(1.5, 2)	0.9463	0.9462	0.9462	0.9461
(2,2.5)	0.9464	0.9464	0.9464	0.9463
		$\rho_1 = 0.2, \rho_2 = 0.6$	$\eta_1 = 0.2, \eta_2 = 0.6$	
$(\beta_1, \delta_1) \rightarrow (\beta_2, \delta_2) \downarrow$	(0.5,1)	(1,1.5)	(1.5,2)	(2,2.5)
(0.5, 1)	0.9002	0.9001	0.9001	0.9001
(1, 1.5)	0.9006	0.9006	0.9005	0.9005
(1.5, 2)	0.9008	0.9008	0.9008	0.9008
(2,2.5)	0.9009	0.9009	0.9009	0.9009

 Table 2
 Value of R for various choices of parameter values.

The MLE,  $\hat{\theta}^*$  may be obtained from the solution of the nonlinear equation,  $U_R(\hat{\theta}^*) = 0$ . Applying  $\hat{\theta}^*$ , in Eq. (26), the stress-strength parameter *R* can be obtained. The stress strength reliability function for different values of  $p_1$ ,  $\rho_1$ ,  $\eta_1$ ,  $\beta_1$ ,  $\delta_1$  and  $p_2$ ,  $\rho_2$ ,  $\eta_2$ ,  $\beta_2$ ,  $\delta_2$  are computed in Table 2. We see that the value of R is decreasing when  $\beta_1$  and  $\delta_1$  increases, and increasing when  $\beta_2$  and  $\delta_2$  increases.

#### 4. MAXIMUM LIKELIHOOD ESTIMATION (MLE) OF PARAMETERS

Consider a random sample  $(y_1, y_2, ..., y_n)$  of size *n* from the *DAWG*  $(p, \rho, \eta, \beta, \delta)$  distribution. Then, the likelihood function is given by,

$$L = \frac{\left(1-p\right)^{n} \prod_{i=1}^{n} \left(\rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}} - \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}}\right)}{\prod_{i=1}^{n} \left(1-p \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}}\right) \prod_{i=1}^{n} \left(1-p \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}}\right)}.$$
(27)

The log-likelihood function is,

$$\log (L) = n \log (1-p) + \sum_{i=1}^{n} \log \left( \rho^{y_i^{\beta}} \eta^{y_i^{\delta}} - \rho^{(y_i+1)^{\beta}} \eta^{(y_i+1)^{\delta}} \right) - \sum_{i=1}^{n} \log (1-p \rho^{y_i^{\beta}} \eta^{y_i^{\delta}}) - \sum_{i=1}^{n} \log \left( 1-p \rho^{(y_i+1)^{\beta}} \eta^{(y_i+1)^{\delta}} \right).$$
(28)

The likelihood equations are the following:

$$\frac{\partial \log (L)}{\partial p} = \frac{-n}{1-p} + \sum_{i=1}^{n} \frac{\rho^{y_i^{\beta}} \eta^{y_i^{\delta}}}{1-p \rho^{y_i^{\beta}} \eta^{y_i^{\delta}}} + \sum_{i=1}^{n} \frac{\rho^{(y_i+1)^{\beta}} \eta^{(y_i+1)^{\delta}}}{1-p \rho^{(y_i+1)^{\beta}} \eta^{(y_i+1)^{\delta}}} = 0,$$
(29)

$$\frac{\partial \log (L)}{\partial \rho} = \sum_{i=1}^{n} \frac{y_{i}^{\beta} \rho^{y_{i}^{\beta}-1} \eta^{y_{i}^{\delta}} - (y_{i}+1)^{\beta} \rho^{(y_{i}+1)^{\beta}-1} \eta^{(y_{i}+1)^{\delta}}}{\rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}} - \rho^{(y_{i}+1)^{\beta}} t a^{(y_{i}+1)^{\delta}}} + p \sum_{i=1}^{n} \frac{y_{i}^{\beta} \rho^{y_{i}^{\beta}-1} \eta^{y_{i}^{\delta}}}{1 - p \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}}} + p \sum_{i=1}^{n} \frac{(y_{i}+1)^{\beta} \rho^{(y_{i}+1)^{\beta}-1} \eta^{(y_{i}+1)^{\delta}}}{1 - p \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}}} = 0,$$
(30)

$$\frac{\partial \log (L)}{\partial \eta} = \sum_{i=1}^{n} \frac{y_{i}^{\delta} \rho^{y_{i}^{\delta}} \eta^{y_{i}^{\delta}-1} - (y_{i}+1)^{\delta} \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}-1}}{\rho^{y_{i}^{\delta}} \eta^{y_{i}^{\delta}} - \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}}} + p \sum_{i=1}^{n} \frac{y_{i}^{\delta} \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}-1}}{1 - p \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}}} + p \sum_{i=1}^{n} \frac{(y_{i}+1)^{\delta} \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}-1}}{1 - p \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}}} = 0,$$
(31)

$$\frac{\partial \log (L)}{\partial \beta} = \log (\rho) \sum_{i=1}^{n} \frac{y_{i}^{\beta} \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}} \log(y_{i}) - (y_{i}+1)^{\beta} \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}} \log(y_{i}+1)}{\rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}} - \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}}} + p \log (\rho) \sum_{i=1}^{n} \frac{y_{i}^{\beta} \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}} \log(y_{i})}{1 - p \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}}} + p \log (\rho) \sum_{i=1}^{n} \frac{(y_{i}+1)^{\beta} \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}} \log(y_{i}+1)}{1 - p \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}}} = 0,$$
(32)

and

$$\begin{split} \frac{\partial \log{(L)}}{\partial \delta} &= \log{(\eta)} \sum_{i=1}^{n} \frac{y_{i}^{\delta} \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}} \log(y_{i}) - (y_{i}+1)^{\delta} \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}} \log(y_{i}+1)}{\rho^{y_{i}^{\beta}} \eta^{y_{i}^{j}} - \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}}} \\ &+ p \log{(\eta)} \sum_{i=1}^{n} \frac{y_{i}^{\beta} \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}} \log(y_{i})}{1 - p \rho^{y_{i}^{\beta}} \eta^{y_{i}^{\delta}}} \\ &+ p \log{(\eta)} \sum_{i=1}^{n} \frac{(y_{i}+1)^{\beta} \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}} \log(y_{i}+1)}{1 - p \rho^{(y_{i}+1)^{\beta}} \eta^{(y_{i}+1)^{\delta}}} = 0. \end{split}$$

(33)

These equations do not have explicit solutions and they have to be obtained numerically by using the statistical softwares like *nlm* package in R programming.

We compute the maximized unrestricted and restricted log-likelihood ratio (LR) test statistic for testing on some *DAWG* sub models. We can use the LR test statistic to check whether *DAWG* distribution for a given data set is statistically superior to the sub models. Here,  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta \neq \theta_0$  can be performed using LR test. The LR test statistic is  $\omega = 2(l(\hat{\theta}, y) - l(\hat{\theta}_0, y))$ , where  $\hat{\theta}$  and  $\hat{\theta}_0$  are the MLEs under  $H_1$  and  $H_0$ , respectively. The test statistic  $\omega$  is asymptotically (as  $n \to \infty$ ) distributed as  $\chi^2_{(k)}$ , where k is the length of the parameter vector  $\theta$  of interest. The LR test rejects  $H_0$  if  $\omega > \chi^2_{(k,\alpha)}$ , where  $\chi^2_{(k,\alpha)}$  denotes the upper 100  $(1 - \alpha)$ % quantile of the  $\chi^2_{(k)}$  distribution.

#### 4.1. Simulation Study

Here we study the performance of the MLEs of the model parameters of *DAWG* distribution using Monte Carlo simulation for various sample sizes and for selected parameter values. The algorithm for the simulation study are given below:

step 1: Input the number of replications (N);

**step 2:** Specify the sample size *n* and the values of the parameters *p*,  $\rho$ ,  $\eta$ ,  $\beta$  and  $\delta$ ;

**step 3:** Generate  $u_i \sim Uniform(0, 1), i = 1, 2, ..., n$ .;

step 4: Obtain random observations from DAWG distribution by solving for real roots of the Eq. (19) and take the floor value;

step 5: Compute the MLEs of the five parameters;

- step 6: Repeat steps 3 to 5, N times;
- step 7: Compute the average bias, mean square error (MSE) and coverage probability (CP) for each parameter.

Here the expected value of the estimator is  $E(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \hat{\theta}_i$ , average bias  $= \frac{1}{N} \sum_{i=1}^{N} (\hat{\theta}_i - \theta)$ ,

$$MSE\left(\hat{\theta}\right) = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\theta}_{i} - \theta\right)^{2} \text{ and the CP} = \text{probability of } \theta_{i} \in \left(\hat{\theta}_{i} \pm 1.96\sqrt{-\frac{\partial^{2}\log(L)}{\partial\theta_{i}^{2}}}\right).$$

43

We have taken the parameter values as p = 0.8,  $\rho = 0.5$ ,  $\eta = 0.5$ ,  $\beta = 0.5$  and  $\delta = 1.5$  and generated random samples of size n = 20, 60 and 100 respectively. The MLEs of p,  $\rho$ ,  $\eta$ ,  $\beta$  and  $\delta$  are determined by maximizing the log-likelihood function in Eq. (28) using the *nlm* package of R software based on each generated samples. This simulation is repeated 500 times and the average estimates of bias, MSE and CP are computed and presented in Table 3. It can be seen that, as the sample size increases, the bias and MSE decreases. Also note that the CP values are quite closer to the 95% nominal level.

Sample size	Actual value	Estimates	Average bias	MSE	СР
	p = 0.8	0.921	0.115	0.074	0.873
	$\rho = 0.5$	0.346	-0.164	0.086	0.926
20	$\eta = 0.5$	0.723	0.213	0.017	0.932
	$\dot{\beta} = 0.5$	0.661	0.165	0.038	0.896
	$\delta = 1.5$	1.833	0.301	0.099	0.882
60	p = 0.8	0.866	0.071	0.016	0.926
	$\rho = 0.5$	0.486	-0.013	0.018	0.936
	$\eta = 0.5$	0.610	0.102	0.008	0.943
	$\dot{\beta} = 0.5$	0.612	0.110	0.012	0.912
	$\delta = 1.5$	1.598	0.096	0.073	0.917
100	p = 0.8	0.833	0.028	0.009	0.938
	$\rho = 0.5$	0.491	-0.003	0.007	0.942
	$\eta = 0.5$	0.552	0.057	0.005	0.949
	$\dot{\beta} = 0.5$	0.587	0.083	0.006	0.929
	$\delta = 1.5$	1.554	0.052	0.011	0.934

 Table 3
 The average bias, MSE, and CP for given values of parameters.

MSE, mean square error; CP, coverage probability.

## 5. APPLICATION

In this section, to show how the  $DAWG(p, \rho, \eta, \beta, \delta)$  distribution works in practice, we use the data set representing remission times (in months) of 128 bladder cancer patients taken from Lee and Wang [29]. The data are: 0.080 0.200 0.400 0.500 0.510 0.810 0.900 1.050 1.190 1.260 1.350 1.400 1.460 1.760 2.020 2.020 2.070 2.090 2.230 2.260 2.460 2.540 2.620 2.640 2.690 2.750 2.830 2.870 3.020 3.250 3.310 3.360 3.360 3.480 3.520 3.570 3.640 3.700 3.820 3.880 4.180 4.230 4.260 4.330 4.340 4.400 4.500 4.510 4.870 4.980 5.060 5.090 5.170 5.320 5.340 5.410 5.410 5.490 5.620 5.710 5.850 6.250 6.540 6.760 6.930 6.940 6.970 7.090 7.260 7.280 7.320 7.390 7.590 7.620 7.630 7.660 7.870 7.930 8.260 8.370 8.530 8.650 8.660 9.020 9.220 9.470 9.740 10.06 10.34 10.66 10.75 11.25 11.64 11.79 11.98 12.02 12.03 12.07 12.63 13.11 13.29 13.80 14.24 14.76 14.77 14.83 15.96 16.62 17.12 17.14 17.36 18.10 19.13 20.28 21.73 22.69 23.63 25.74 25.82 26.31 32.15 34.26 36.66 43.01 46.12 79.05.

Since the data set is continuous, here first we discretize the data by considering the floor value (y). The parameters are estimated by using the method of MLE. We compare the fit of the *DAWG* distribution with the discrete life time distributions:

(a) Geometric (G) distribution having pmf,

$$P(Y = y) = (1 - p) p^{y}; \quad 0$$

(b) Discrete Weibull (DW) distribution having pmf,

$$P(Y = y) = q^{y^{\beta}} - q^{(y+1)^{\beta}}; \quad 0 < q < 1, \beta > 0, y = 0, 1, 2, ....$$

(c) Discrete Logistic (DLOG) distribution (see Chakraborty and Chakravarty [30]) having pmf,

$$P(Y = y) = \frac{(1-p)p^{y-\mu}}{(1+p^{y-\mu})(1+p^{(y-\mu+1)})}; \quad 0$$

(d) Exponentiated discrete Weibull (EDW) distribution (see Nekoukhou and Bidram [31]) having pmf,

$$P(Y = y) = \left(1 - p^{(y+1)^{\alpha}}\right)^{\gamma} - \left(1 - p^{y^{\alpha}}\right)^{\gamma}; \quad 0 0, \gamma > 0, y = 0, 1, 2, \dots$$

(e) DWG distribution (see Jayakumar and Babu [21]) having pmf,

$$P(Y = y) = \frac{(1-p)\left(\rho^{y^{\alpha}} - \rho^{(y+1)^{\alpha}}\right)}{(1-p\rho^{y^{\alpha}})\left(1-p\rho^{(y+1)^{\alpha}}\right)},$$

Model	ML estimates	-log L	AIC	AICC	BIC	K-S	<i>p</i> value
G	$\hat{p} = 0.8991$	414.836	831.672	831.704	831.779	0.1000	0.1549
DW	$\hat{q} = 0.9114$ $\hat{\beta} = 1.0511$	414.556	833.112	837.304	833.326	0.1131	0.0758
DLOG	$\hat{p} = 0.8000$ $\hat{\mu} = 7.6149$	456.825	917.650	917.746	917.864	0.1860	0.0003
EDW	$\hat{p} = 0.4689$ $\hat{\alpha} = 0.5397$ $\hat{\gamma} = 4.9697$	409.766	825.532	825.726	825.854	0.1237	0.0399
DWG	$\hat{p} = 0.9529$ $\hat{\rho} = 0.9982$ $\hat{\sigma} = 1.7025$	409.277	824.554	824.748	824.876	0.0905	0.2458
DAMAG	$\hat{p} = 0.9589$ $\hat{p} = 0.9989$ $\hat{p} = 0.9005$	105 220	020.460	020.052	220.000	0.0002	0.2525
DAWG	$\hat{\beta} = 0.9995$ $\hat{\beta} = 1.7018$ $\hat{\delta} = 1.7016$	405.230	820.460	820.952	820.996	0.0882	0.2/2/

 Table 4
 Parameter estimates and goodness of fit for various models fitted for the data set.

-logL, log-likelihood function; K–S, Kolmogorov–Smirnov; AIC, Akaike Information Criterion; AICC, Akaike Information Criterion with correction; BIC, Bayesian Information Criterion; DLOG, Discrete Logistic; DAWG, discrete additive Weibull geometric; EDW, exponentiated discrete Weibull; DWG, discrete Weibull geometric; DW, discrete Weibull.



**Figure 4** Fitted cumulative distribution function's (cdf) of the data with empirical distribution.

where  $y = 0, 1, 2, ...; \alpha > 0, 0 and <math>0 < \rho < 1$ .

The values of the log-likelihood function  $(-\log L)$ , the statistics Kolmogorov–Smirnov (K - S), Akaike Information Criterion (*AIC*), Akaike Information Criterion with correction (*AICC*), and Bayesian Information Criterion (*BIC*) are calculated for the six distributions in order to verify which distribution fits better to these data. The better distribution corresponds to smaller  $-\log L$ , *AIC*, *AICC*, *BIC*, and K - S values and larger p value.

Here,  $AIC = -2 \log L + 2k$ ,  $AICC = -2 \log L + \left(\frac{2kn}{n-k-1}\right)$  and  $BIC = -2 \log L + k \log n$ , where *L* is the likelihood function evaluated at

the maximum likelihood estimates, *k* is the number of parameters, and *n* is the sample size. The K - S distance,  $D_n = \sup_{y} |F(y) - F_n(y)|$ , where,  $F_n(y)$  is the empirical distribution function.

The values in Table 4, indicates that *DAWG* distribution leads to a better fit compared to the other five models. Figure 4 shows the structure of the cdf's of the six models with the empirical distribution of the given data. Here the dotted line indicates the empirical cdf of the data. The LR test statistic is used to test the hypothesis  $H_0$ :  $\eta = 1$  versus  $H_1$ :  $\eta \neq 1$  is  $\omega = 8.094 > 5.991$  with *p* value 0.0175. So we reject the null hypothesis.

## 6. CONCLUSION

In the present study, we have introduced the generalized DAWG distribution. A particular member of this distribution, namely *DAWG* distribution is studied in detail. This discrete distribution contains the DWG, discrete exponential geometric, discrete modified Weibull, discrete Weibull, discrete Rayleigh, and geometric distribution as special cases. We have studied some basic properties of the new model and illustrated that the hazard rate function of the new model is monotonically increasing, decreasing, or bathtub shape depending on the values of the shape parameters. By fitting the *DAWG* model to a real data set, the flexibility and capacity of the new distribution in data modeling is established.

## ACKNOWLEDGMENTS

The authors would like acknowledge the comments and suggestions of the Editor and the anonymous referee on earlier version of the manuscript which resulted in substantial improvements in the original version and presentation of the article.

## REFERENCES

- 1. I. Elbatal, M.M. Mansour, M. Ahsanullah, J. Stat. Theory Appl. 15 (2016), 125–141.
- 2. T. Nakagawa, S. Osaki, IEEE Trans. Rel. 24 (1975), 300-301.
- 3. W.E. Stein, R. Dattero, IEEE Trans. Rel. 33 (1984), 196–197.
- 4. W.J. Padgett, J.D. Spurrier, IEEE Trans. Rel. 34 (1985), 253–256.
- 5. H. Sato, M. Ikota, A. Sugimoto, H. Masuda, IEEE Trans. Semicond. Manuf. 12 (1999), 409-418.
- 6. D. Roy, Commun. Stat. Theory Methods. 32 (2003), 1871–1883.
- 7. D. Roy, IEEE Trans. Rel. 53 (2004), 255-260.
- 8. S. Inusah, T.J. Kozubowski, J. Stat. Plan. Inference. 136 (2006), 1090-1102.
- 9. T.J. Kozubowski, S. Inusah, Ann. Ins. Stat. Math. 58 (2006), 555-571.
- 10. H. Krishna, P.S. Pundir, Stat. Methodol. 6 (2009), 177-188.
- 11. M.A. Jazi, C.D. Lai, M.H. Alamatsaz, Stat. Methodol. 7 (2010), 121-132.
- 12. E. Gómez-Déniz, Test. 19 (2010), 399-415.
- 13. V. Nekoukhou, M.H. Alamatsaz, H. Bidram, Commun. Stat. Theory Methods. 41 (2012), 2000–2013.
- 14. S. Chakraborty, D. Chakravarty, Commun. Stat. Theory Methods. 41 (2012), 3301-3324.
- 15. M. Bebbington, C.D. Lai, M. Wellington, R. Zitikis, Rel. Eng. Syst. Safety. 106 (2012), 37-44.
- 16. H.S. Bakouch, M.A. Jazi, S. Nadarajah, Statistics. 48 (2014), 200-240.
- 17. S. Chakraborty, D. Chakravarty, arXiv: 1410.7568[math.ST], 2014.
- 18. S. Chakraborty, R.D. Gupta, Commun. Stat. Theory Methods. 44 (2015), 1143-1157.
- 19. S. Chakraborty, Commun. Stat. Theory Methods. 44 (2015), 1691–1705.
- 20. S. Chakraborty, D. Bhati, SORT. 40 (2016), 153-176.
- 21. K. Jayakumar, M.G. Babu, Commun. Stat. Theory Methods. 47 (2018), 1767–1783.
- 22. C. Bracquemond, O. Gaudoin, Int. J. Rel. Qual. Safety Eng. 10 (2003), 69-98.
- 23. S. Chakraborty, J. Stat. Distributions Appl. 2 (2015), 1-30.
- 24. M. Xie, C.D. Lai, Rel. Eng. Syst. Safety. 52 (1995), 87–93.
- 25. A.J. Lemonte, G.M. Cordeiro, E.M.M. Ortega, Commun. Stat. Theory Methods. 43 (2014), 2066–2080.
- 26. A.W. Marshall, I. Olkin, Biometrika. 84 (1997), 641-652.
- 27. P.L. Gupta, R.C. Gupta, R.C. Tripathi, J. Stat. Plan. Inference. 65 (1997), 255-268.
- S. Kotz, Y. Lumelskii, M. Pensky, The Stress-Strength Model and its Generalizations: Theory and Applications, World Scientific Co., Singapore, 2003.
- 29. E.T. Lee, J. Wang, Statistical Methods for Survival Data Analysis, John Wiley and Sons, New York, 2003.
- 30. S. Chakraborty, D. Chakravarty, Commun. Stat. Theory Methods. 45 (2016), 492–505.
- 31. V. Nekoukhou, H. Bidram, SORT. 39 (2015), 127-146.