JIRSS (2021)
Vol. 20, No. 02, pp 129-152
DOI:10.52547/jirss.20.2.129

A Discrete Kumaraswamy Marshall-Olkin Exponential Distribution

Jiju Gillariose [1], **Lishamol Tomy** [2], **Farrukh Jamal** [3], **and Christophe Chesneau** [4]

[1] Department of Statistics, CHRIST (Deemed to be University), Hosur Road, Bangalore, Karnataka, 560029, India.
[2] Department of Statistics, Deva Matha College, Kuravilangad, Kerala, 686633, India.
[3] Department of Statistics, The Islamia University of Bahawalpur, Punjab 63100, Pakistan.
[4] Université de Caen, LMNO, Campus II, Science 3, 14032, Caen, France.

Received: 24/07/2020, Revision received: 22/01/2022, Published online: 01/02/2022

Abstract. Finding new families of distributions has become a popular tool in statistical research. In this article, we introduce a new flexible four-parameter discrete model based on the Marshall-Olkin approach, namely, the discrete Kumaraswamy Marshall-Olkin exponential distribution. The proposed distribution can be viewed as another generalization of the geometric distribution and enfolds some important distributions as special cases. Some properties of the new distribution are derived. The model parameters are estimated by the maximum likelihood method, with validation through a complete simulation study. The usefulness of the new model is illustrated via count-type real data sets.

Keywords. Discrete Distributions, Exponential Distribution, Generalized Family, Geo-

Jiju Gillariose(jijugillariose@yahoo.com)
Lishamol Tomy(lishatomy@gmail.com)
Farrukh Jamal(drfarrukh1982@gmail.com)
Corresponding Author: Christophe Chesneau(christophe.chesneau@gmail.com)

[DOI: 10.52547/jirss.20.2.129]

metric Distribution, Marshall-Olkin Extended Distribution, Maximum Likelihood Estimation

MSC: 60E05, 62E15, 62F10.

1 Introduction

The Marshall-Olkin family of distributions, introduced by Marshall and Olkin (1997), has appeared and reappeared in the statistical literature for two decades. As of now, a substantial part of the literature has proposed various univariate distributions belonging to the Marshall-Olkin family of distributions. This family often provides a better fit than other generalized families with the same baseline distribution, and the generated distributions have wider applications than the baseline distribution itself. For more details, one can refer to a review study by Tahir and Nadarajah (2015) which provides a list of statistical distributions derived through the Marshall-Olkin technique.

A part from that, researchers have concentrated on new families that generalize the Marshall-Olkin family of distributions. There is a strong need to introduce such useful new models, because they have a strong potential for modeling data in many areas, such as economics, biological studies, environmental sciences, physics, computer science, insurance, public health, medical, engineering, biology, industry, communications, life-testing, and many others. For this purpose, Alizadeh et al. (2015) proposed the Kumaraswamy Marshall-Olkin family. It uses the scheme of the Kumaraswamy generalized family by Cordeiro and de Castro (2011). The survival function of the Kumaraswamy Marshall-Olkin family is given by

$$\bar{G}(x;\alpha,a,b)=\left\{1-\left[\frac{F(x)}{1-\bar{\alpha}\bar{F}(x)}\right]^{a}\right\}^{b};\ \ x\in\mathbb{R};\alpha,a,b,>0, \tag{1.1}$$

which enfolds a wider family of continuous distributions. When $a=1$ and $b=1$, equation (1.1) reduces to the Marshall-Olkin family. When $\alpha=0$, equation (1.1) becomes the exponentiated Marshall-Olkin family of distributions (Jayakumar and Thomas, 2008). This family includes the Kumaraswamy family of distributions (Cordeiro and de Castro, 2011) when $\alpha=0$, the proportional reversed hazard rate distribution by Gupta and Gupta (2007) (when $a=1,\alpha=0$) and the proportional hazard rate distribution (when $b=1,\alpha=0$) by Gupta et al. (1998).

In life testing experiments, there are many situations where it is impossible or inconvenient to quantify the lifespan of a component or a device on a continuous scale. Discretized statistical models have been used to handle discrete lifetime data and also count data in a wide variety of disciplines such as biological and medical sciences, physical sciences, engineering, agriculture, and so on. The classical discrete distributions have constricted applicability as models for reliability, failure times, counts, etc. This has led to the evolution of some discrete distributions based on popular continuous models. For example, Lisman and van Zuylen (1972) pioneered the discrete version of the continuous normal distribution. A discrete analogue of the two parameter continuous Weibull distribution has been introduced by Nakagawa and Osaki (1975). Stein and Dattero (1984) devised a new discrete Weibull distribution by treating lifetime as the integer component of the continuous Weibull distribution. Sato et al. (1999) proposed a discrete exponential distribution and applied it to model defect count in a semiconductor deposition equipment as well as defect count distribution per chip. Krishna and Pundir (2009) introduced the discrete Burr distribution, which together with Chakraborty and Chakravarty (2016), invented the discrete gamma distribution. However, there is still a need to develop new discretized distributions that are desirable under diverse scenarios. In this paper, we use the Kumaraswamy Marshall-Olkin scheme to introduce a new discrete family of distributions and focus on a new model as its particular case.

The first motivation for introducing the new distribution is based on the fact that, compared to the volume of literature in the continuous case, only a few papers have been inscribed about the discrete version of the continuous family of distributions. The second motivation is related to the flexibility of the proposed distribution to model complex positive real data sets; that is, the new distribution can deliver decreasing, increasing, unimodal, and bathtub-shaped hazard rate functions (hrfs). In addition to this, the distribution proposed here can be considered as a generalization of many existing distributions in the literature. Also, the performance of the model is examined by using three examples from real data sets. As it can be seen later, the new distribution introduced in this manuscript produces a better fit to data than the well-known geometric and an extension of the geometric distributions previously considered. Finally, due to its ability to model monotone as well as non-monotone hrfs, which are quite common in reliability and biological studies, the new distribution becomes very important.

The rest of the article is organized as follows: Section 2 presents the discrete Kumaraswamy Marshall-Olkin family. Section 3 provides an elaborated study of a special case of the derived new discrete discrete distribution and its probabilistic

[DOI: 10.52547/jirss.20.2.129]

properties. Maximum likelihood estimation of the unknown parameters is presented in Section 4, completed by a simulation procedure. Utilization of the newly developed model is discussed in Section 5. The paper is concluded in Section 6.

2 Discrete Kumaraswamy Marshall-Olkin Family

In this section, we introduce a new discrete family of distributions, namely, the discrete Kumaraswamy Marshall-Olkin family of distributions. Some new discrete distributions have appeared in the literature by using the continuous Marshall-Olkin scheme. For example, we can cite the works of Jayakumar and Sankaran (2017a) and Jayakumar and Sankaran (2017b). A detailed survey of the methods and constructions of discrete analogs of continuous distributions is discussed in Chakraborty (2015). If the underlying continuous life time X has the survival function $\bar{K}(x) = P(X > x)$, the pmf of the discrete random variable associated with that continuous distribution can be written as

$$P(X = x) = \bar{K}(x) - \bar{K}(x+1); \;\; x = 0, 1, 2 \ldots. \tag{2.1}$$

Alizadeh et al. (2015) proposed a generalization of the Marshall-Olkin family of distributions by adding parameters $a > 0$ and $b > 0$ in the Marshall-Olkin scheme with the survival function given as (1.1). The new family is generated by discretizing the continuous survival function of the Kumaraswamy Marshall-Olkin family using (2.1). We obtain a new family of discrete distributions with the pmf $g(x)$ given by

$$g(x) = \left\{1 - \left[\frac{F(x)}{1 - \bar{\alpha}\bar{F}(x)}\right]^a\right\}^b - \left\{1 - \left[\frac{F(x+1)}{1 - \bar{\alpha}\bar{F}(x+1)}\right]^a\right\}^b; \;\; x = 0, 1, 2 \ldots. \tag{2.2}$$

The survival function of the discrete random variable having the pmf (2.2) is given by

$$\bar{G}(x) = \left\{1 - \left[\frac{F(x+1)}{1 - \bar{\alpha}\bar{F}(x+1)}\right]^a\right\}^b; \;\; x = 0, 1, 2 \ldots.$$

We explore one member of this family, namely, the discrete Kumaraswamy Marshall-Olkin exponential (DKMOE) distribution in the next section and present its properties in detail. The exponential distribution was chosen because it is the simplest and widely used model. In practice, other distributions can be used to model real data.

[DOI: 10.52547/jirss.20.2.129]

3 Discrete Kumaraswamy Marshall-Olkin Exponential Distribution

3.1 Definition

Let the parent distribution be exponential with parameter $\lambda > 0$ and survival function $\bar{F}(x) = e^{-\lambda x}$. We set $p = e^{-\lambda}$, satisfying $0 < p < 1$. Then, the pmf of the new model using (2.2) is given by

$$g(x) = \left\{1 - \left[\frac{1-p^x}{1-\bar{\alpha}p^x}\right]^a\right\}^b - \left\{1 - \left[\frac{1-p^{x+1}}{1-\bar{\alpha}p^{x+1}}\right]^a\right\}^b; \; x = 0, 1, 2 \ldots, \tag{3.1}$$

where $a > 0$ is a location parameter, $b > 0$ is a scale parameter, and $\alpha > 0$ is a shape parameter (for the sake of conciseness, we will note $g(x)$ instead of $g(x; a, b, \alpha, p)$, the same for all the coming functions). We call this new distribution the discrete Kumaraswamy Marshall-Olkin exponential (DKMOE) distribution with parameters a, b, α, and p. Note that, when $a = 1$ and $b = 1$, the distribution with pmf (3.1) reduces to discrete Marshall-Olkin distribution discussed in Supanekar and Shirke (2015), when $b = 1$, the distribution with pmf (3.1) reduces to discrete generalized Marshall-Olkin distribution given by Jayakumar and Sankaran (2017a) and when $a = 1, b = 1$ and $\alpha = 1$, (3.1) becomes geometric distribution. Also note that, if $x \to \infty$, then $\lim_{x\to\infty} g(x) = 0$, and if $x \to 0$, then $\lim_{x\to\infty} g(x) = 0$ when $p \to 1$ or $a \to \infty$. The corresponding survival function is given by

$$\bar{G}(x) = \left\{1 - \left[\frac{1-p^{x+1}}{1-\bar{\alpha}p^{x+1}}\right]^a\right\}^b; \; x = 0, 1, 2 \ldots,$$

and the hrf corresponding to the cdf is given by

$$h(x) = \frac{\left\{1 - \left[\frac{1-p^x}{1-\bar{\alpha}p^x}\right]^a\right\}^b - \left\{1 - \left[\frac{1-p^{x+1}}{1-\bar{\alpha}p^{x+1}}\right]^a\right\}^b}{\left\{1 - \left[\frac{1-p^{x+1}}{1-\bar{\alpha}p^{x+1}}\right]^a\right\}^b}.$$

Figure 1 shows the plots of the pmf of the DKMOE distribution for various values of a, b, α and p. The pmf can be increasing, decreasing, or upside-down bathtub shaped. Also, Figure 2 illustrates some of the possible shapes of the hrf of the DKMOE distribution for selected values of the parameters a, b, α, and p, respectively. From this figure, it is clear that the hrf can be increasing, decreasing, bathtub, or upside-down bathtub

[DOI: 10.52547/jirss.20.2.129]

shaped. Indeed, the DKMOE distribution can be suitable for modeling various data sets.

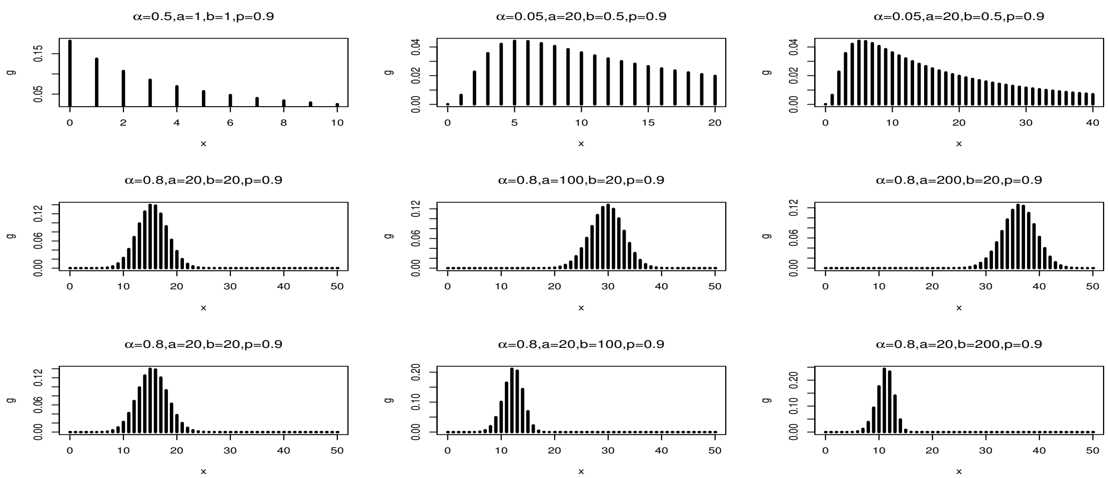

Figure 1: Pmfs of the DKMOE distribution for some parameter values

Some well-known discrete distributions arise as special cases of the DKMOE(α, a, b, p) distribution, such as the geometric distribution, Kumaraswamy-geometric distribution and other sub-families. We provide special models of the DKMOE distribution in Table 1.

Table 1: Some special distributions

a	b	α	Reduced Model
1	1	0	geometric distribution
1	-	0	geometric distribution (with probability of success p^b)
-	1	0	geometric distribution (with probability of success as $1 - p^a$)
-	-	0	Kumaraswamy-geometric (KG) distribution (Akinsete et al. , 2014)
-	1	0	exponentiated-exponential-geometric distribution (Alzaatreh et al. , 2012)
1	1	-	discrete Marshall-Olkin exponential distribution (Gómez-Déniz, 2010)
1	-	-	discrete generalized Marshall-Olkin exponential distribution (Jayakumar and Sankaran, 2017a)

[DOI: 10.52547/jirss.20.2.129]

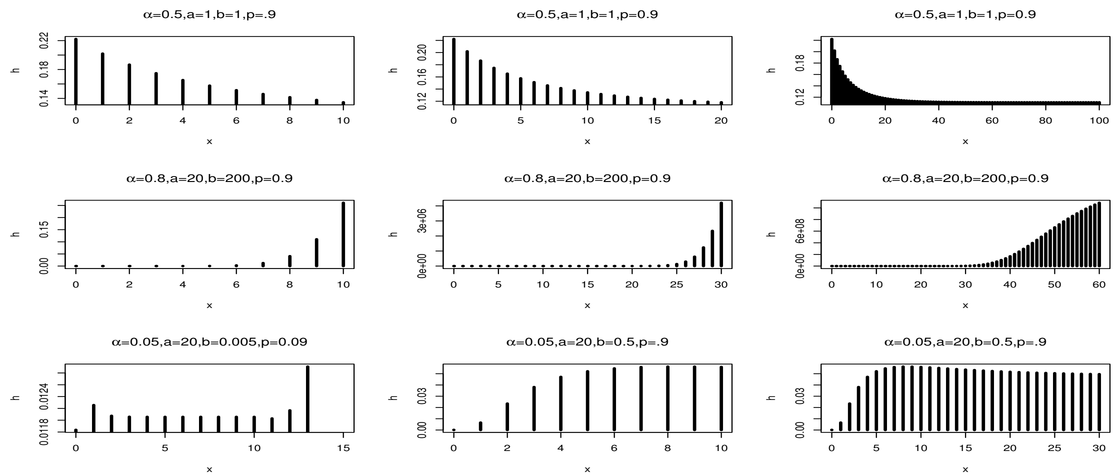

Figure 2: Hrfs of the DKMOE distribution for some parameter values

In the rest of the section, we study some general properties of the DKMOE distribution.

3.2 Quantiles, Probability Generating Function, Mean and Variance

Suppose that X follows the DKMOE distribution with cdf $G(x) = 1 - \bar{G}(x)$. Then, the corresponding quantile function $Q(u)$, $0 < u < 1$ is given as

$$Q(u) = \left\{ \frac{1}{\log p} \log \left[\frac{(1-\bar{u}^{1/b})^{1/a} - 1}{\bar{\alpha}(1-\bar{u}^{1/b})^{1/a} - 1} \right] \right\} - 1, \tag{3.2}$$

where $\bar{u} = 1 - u$. Equation (3.2) can be used to simulate values from X. First, simulate a random variable u following the uniform distribution on the unit interval and compute the value of $Q(u)$ in (3.2), which is not necessarily an integer. A DKMOE random variate x is the largest integer inferior or equal to $Q(u)$, denoted as $[x]$.

In particular, the median M is presented by

$$M = \left\{ \frac{1}{\log p} \log \left[\frac{[1-(1/2)^{1/b}]^{1/a} - 1}{\bar{\alpha}[1-(1/2)^{1/b}]^{1/a} - 1} \right] \right\} - 1.$$

[DOI: 10.52547/jirss.20.2.129]

The probability generating function (pgf) of the DKMOE distribution is given by

$$P(s) = 1 + (s-1)\sum_{x=1}^{\infty} s^{x-1}\left\{1-\left[\frac{1-p^x}{1-\bar{\alpha}p^x}\right]^a\right\}^b.$$

The mean and variance of the DKMOE distribution do not have compact forms, but can be set as

$$E(X) = \sum_{x=1}^{\infty}\left\{1-\left[\frac{1-p^x}{1-\bar{\alpha}p^x}\right]^a\right\}^b,$$

and

$$V(X) = \sum_{x=1}^{\infty}(2x-1)\left\{1-\left[\frac{1-p^x}{1-\bar{\alpha}p^x}\right]^a\right\}^b - \left[\sum_{x=1}^{\infty}\left\{1-\left[\frac{1-p^x}{1-\bar{\alpha}p^x}\right]^a\right\}^b\right]^2.$$

Using statistical software, the mean and variance of the DKMOE distribution for different values of α, a, b and p are calculated in Table 2. From this, we can say that the mean increases with p and α for different values of a and b. Moreover, depending on the values of α and p, the mean of the distribution can be smaller or greater than its variance. Therefore, the parameters of the DKMOE distribution can be used to model different data sets.

3.3 Infinite Divisibility

Infinite divisibility has a close relationship to the Central Limit Theorem and waiting time distributions. According to Steutel and van Harn (2004), pp.56, if $p(x), x \in \mathbb{N}_0$, is infinitely divisible, then $p(x) < e^{-1}$ for all $x \in \mathbb{N}$. For the DKMOE distribution with parameters $a = 10$, $b = 0.1$, $\alpha = 10$ and $p = 0.002$, we can see that $g(1) = 0.3851099 > e^{-1} = 0.367$. Hence, we can conclude that the DGMOE distribution is not infinitely divisible. In addition, since the classes of self decomposable and stable distributions, in their discrete concepts, are subclasses of infinitely divisible distributions, we can conclude that the DKMOE distribution can be neither self-decomposable nor stable, in general.

3.4 Shannon Entropy

In a probabilistic context, Shannon entropy is a measure of the variation of uncertainty, with higher entropy corresponding to less information. For further details in this

[DOI: 10.52547/jirss.20.2.129]

Table 2: The mean(variance) of the DKMOE distribution for different values of parameters

	$p \longrightarrow$ α $\downarrow$	0.25	0.5	0.75
$a = 0.5$	0.25	0.0016(0.0017)	0.0131(0.0177)	0.0886(0.1859)
$b = 2$	0.50	0.0058(0.0063)	0.0406(0.0569)	0.2242(0.5180)
	0.75	0.0117(0.0129)	0.0739(0.10572)	0.3594(0.8858)
$a = 2$	0.25	0.7422 (1.6148)	1.8539(6.8015)	5.1011(39.8962)
$b = .5$	0.50	1.0173(1.9875)	2.4760(8.0756)	6.6505(46.7974)
	0.75	1.2109(2.1908)	2.8933(8.7451)	7.6686(50.5051)
$a = 2$	0.25	0.0230(0.0249)	0.1574(0.2041)	0.7969(1.5060)
$b = 2$	0.50	0.0747(0.0782)	0.4009(0.483)	1.5728(2.9167)
	0.75	0.0746(0.0782)	1.5727(2.916)	1.5727(2.9169)
$a = 0.5$	0.25	0.3775(0.9482)	0.9602(4.2549)	2.7074(26.1278)
$b = 0.5$	0.50	0.5252(1.2559)	1.3155(5.514)	3.6432(33.4067)
	0.75	0.6336(1.4624)	1.5677(6.3341)	4.2915(38.0939)
$a = 4$	0.25	0.0056(0.0056)	0.1274(0.1238)	0.9200(0.8603)
$b = 4$	0.50	0.0451 (0.0435)	0.4632(0.3533)	1.8707(1.6287)
	0.75	0.1225(0.1095)	0.7720(0.4756)	2.5916(2.2269)

regard, we refer the reader to Amigo et al. (2018). For a discrete random variable X with pmf $g(x)$, the Shannon entropy is defined as

$$S(x) = -\sum_{x=0}^{\infty} g(x) \log_2 g(x). \tag{3.3}$$

Combining (3.3) and (3.1), it gives

$$S(x) = -\sum_{x=0}^{\infty} \left(\left\{ 1 - \left[\frac{1-p^x}{1-\bar{\alpha}p^x} \right]^a \right\}^b - \left\{ 1 - \left[\frac{1-p^{x+1}}{1-\bar{\alpha}p^{x+1}} \right]^a \right\}^b \right) \times$$
$$\log_2 \left\{ 1 - \left[\frac{1-p^x}{1-\bar{\alpha}p^x} \right]^a \right\}^b - \left\{ 1 - \left[\frac{1-p^{x+1}}{1-\bar{\alpha}p^{x+1}} \right]^a \right\}^b .$$

[DOI: 10.52547/jirss.20.2.129]

Now, consider the another representation of pmf of the DKMOE distribution:

$$g(x) = \left\{1 - \left[\frac{1-p^x}{1-\bar{\alpha}p^x}\right]^a\right\}^b \left\{1 - \left[\frac{1 - \left[\frac{1-p^{x+1}}{1-\bar{\alpha}p^{x+1}}\right]^a}{1 - \left[\frac{1-p^x}{1-\bar{\alpha}p^x}\right]^a}\right]^b\right\}.$$

Note that when $b \to \infty$ and $p \to 0$ then $S(x) \to 0$. This indicates that smaller values of b increase the uncertainty in the distribution, while higher values of b increase the amount of information measured in terms of the probability when $b > 1$. The numerical values of the entropies for different values of parameters are given in Table 3. It is understood that 0.0000 represents a value < 0.0001.

Table 3: The entropy of the DKMOE distribution for different values of parameters

	$p \longrightarrow$ b $\downarrow$	0.2	0.4	0.6	0.8	0.9
$a = 2$ $\alpha = 0.5$	2	0.2089	0.5909	1.1505	2.3086	3.2807
	5	0.0045	0.0966	0.4471	1.2379	2.3794
	10	0.0000	0.0031	0.1102	0.7061	1.5826
	20	0.0000	0.0000	0.0000	0.3329	0.9743
$a = 0.2$ $\alpha = 5$	2	0.1306	0.3212	0.5877	1.063	1.4974
	5	0.0010	0.0106	0.0424	0.1411	0.2941
	10	0.0000	0.0000	0.0004	0.0046	0.0197
	20	0.0000	0.0000	0.0000	0.0000	0.0000
$a = 2$ $\alpha = 5$	2	0.8416	1.7214	2.7153	3.9254	3.9708
	5	0.5536	1.0490	2.0367	3.3834	3.9276
	10	0.3517	0.6970	1.5471	2.9416	2.9255
	20	0.0778	0.5482	1.0890	2.4872	2.0837
$a = 2$ $\alpha = 5$	0.2	3.7464	3.6515	2.8747	1.3622	0.3946
	0.4	2.3150	3.2627	3.8911	3.0142	1.1416
	0.6	1.8574	2.8369	3.6997	3.4990	1.5080
	0.9	1.4486	2.4312	3.3752	3.8541	1.9514

[DOI: 10.52547/jirss.20.2.129]

4 Estimation

4.1 Maximum Likelihood Method

In order to estimate α, a, b and p, assume that $X_1, X_2, \ldots, X_n$ is a random sample of size n from a DKMOE distribution, and consider the corresponding observations denoted by $x_1, x_2, \ldots, x_n$. We can write the log likelihood of the DKMOE model in the following form:

$$\log l = \sum_{i=1}^{n} \log \left\{1 - \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^a\right\}^b - \left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\}^b .$$

Differentiating the log-likelihood function with respect to the parameters, we get

$$\frac{\partial \log l}{\partial a} = \sum_{i=1}^{n} \frac{\mathcal{B}_i - \mathcal{A}_i}{\left\{1 - \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^a\right\}^b - \left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\}^b},$$

$$\frac{\partial \log l}{\partial b} = \sum_{i=1}^{n} \frac{\mathcal{C}_i - \mathcal{D}_i}{\left\{1 - \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^a\right\}^b - \left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\}^b},$$

$$\frac{\partial \log l}{\partial \alpha} = \sum_{i=1}^{n} \frac{\mathcal{E}_i - \mathcal{F}_i}{\left\{1 - \left[\frac{1-p_i^{x}}{1-\bar{\alpha}p^{x_i}}\right]^a\right\}^b - \left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\}^b},$$

and

$$\frac{\partial \log l}{\partial p} = \sum_{i=1}^{n} \frac{\mathcal{G}_i - \mathcal{H}_i}{\left\{1 - \left[\frac{1-p_i^{x}}{1-\bar{\alpha}p^{x_i}}\right]^a\right\}^b - \left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\}^b},$$

where

$$\mathcal{A}_i = b\left\{1 - \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^a\right\}^{b-1} \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^a \log\left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right],$$

$$\mathcal{B}_i = b\left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\}^{b-1} \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a \log\left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right],$$

[DOI: 10.52547/jirss.20.2.129]

$$\mathcal{C}_i = \left\{1 - \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^a\right\}^b \log\left\{1 - \left[\frac{1-p_i^x}{1-\bar{\alpha}p^{x_i}}\right]^a\right\},$$

$$\mathcal{D}_i = \left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\}^b \log\left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\},$$

$$\mathcal{E}_i = ab\left\{1 - \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^a\right\}^{b-1} \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^a \frac{p^{x_i}}{1-\bar{\alpha}p^{x_i}},$$

$$\mathcal{F}_i = ab\left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\}^{b-1} \left\{\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right\}^a \frac{p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}},$$

$$\mathcal{G}_i = \frac{ab\alpha x_i p^{x_i-1}}{1-\bar{\alpha}p^{x_i}}\left\{1 - \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^a\right\}^{b-1} \left[\frac{1-p^{x_i}}{1-\bar{\alpha}p^{x_i}}\right]^{a-1},$$

and

$$\mathcal{H}_i = \frac{ab\alpha(x_i+1)p^{x_i}}{1-\bar{\alpha}p^{x_i}}\left\{1 - \left[\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right]^a\right\}^{b-1} \left\{\frac{1-p^{x_i+1}}{1-\bar{\alpha}p^{x_i+1}}\right\}^{a-1}.$$

The maximum likelihood estimate (MLE) of the parameter vector $\xi = (a, b, p, \alpha)^T$, say $\hat{\xi} = (\hat{a}, \hat{b}, \hat{p}, \hat{\alpha})^T$, can be numerically obtained by solving $\partial \log l/\partial a = 0$, $\partial \log l/\partial b = 0$, $\partial \log l/\partial \alpha = 0$ and $\partial \log l/\partial p = 0$. These equations can be solved using statistical software.

To perform asymptotic inference for the parameter vector ξ, the normal approximation of the MLE of ξ can be used. Indeed, under some regular conditions stated in Cox and Hinkley (1974), we have a $\hat{\xi}$ that is approximately normally distributed with a mean of ξ and an asymptotic variance-covariance matrix Σ_ξ. The asymptotic behavior remains valid if Σ_ξ is approximated by I^{-1}, the inverse of the matrix I, which is given as

$$I = -\left.\begin{bmatrix} \frac{\partial^2 \log l}{\partial a^2} & \frac{\partial^2 \log l}{\partial a \partial b} & \frac{\partial^2 \log l}{\partial a \partial p} & \frac{\partial^2 \log l}{\partial a \partial \alpha} \\ \frac{\partial^2 \log l}{\partial b \partial a} & \frac{\partial^2 \log l}{\partial b^2} & \frac{\partial^2 \log l}{\partial b \partial p} & \frac{\partial^2 \log l}{\partial b \partial \alpha} \\ \frac{\partial^2 \log l}{\partial p \partial a} & \frac{\partial^2 \log l}{\partial p \partial b} & \frac{\partial^2 \log l}{\partial p^2} & \frac{\partial^2 \log l}{\partial p \partial \alpha} \\ \frac{\partial^2 \log l}{\partial \alpha \partial a} & \frac{\partial^2 \log l}{\partial \alpha \partial b} & \frac{\partial^2 \log l}{\partial \alpha \partial p} & \frac{\partial^2 \log l}{\partial \alpha^2} \end{bmatrix}\right|_{\xi=\hat{\xi}}.$$

From I^{-1}, we get the asymptotic variances and covariances of the MLEs $\hat{a}$, $\hat{b}$, $\hat{\alpha}$ and $\hat{p}$. Also, the multivariate normal $N_4(0_4, I^{-1})$ distribution can be used to construct

[DOI: 10.52547/jirss.20.2.129]

approximate confidence intervals for the parameters a, b, p and α whose lower and upper bounds are given, respectively, by $\hat{a} \pm z_{\eta/2} \times [\hat{Var}(\hat{a})]^{\frac{1}{2}}$, $\hat{b} \pm z_{\eta/2} \times [\hat{Var}(\hat{b})]^{\frac{1}{2}}$, $\hat{p} \pm z_{\eta/2} \times [\hat{Var}(\hat{p})]^{\frac{1}{2}}$ and $\hat{\alpha} \pm z_{\eta/2} \times [\hat{Var}(\hat{\alpha})]^{\frac{1}{2}}$, where $\hat{Var}(\hat{a})$ is the diagonal element of I^{-1} corresponding to the parameter a, the same for the other parameters, and $\eta/2$ is the quantile $100(1-\eta/2)\%$ of the standard normal distribution.

4.2 Simulation

We now examine the performance of the MLEs using a Monte Carlo simulation study. The following sets of target parameters are considered: Set I ($a = 4.5$, $b = 10$, $\alpha = 9.5$, $p = 0.8$), Set II ($a = 5$, $b = 3$, $\alpha = 10$, $p = 0.6$) and Set III ($a = 8$, $b = 10$, $\alpha = 5$, $p = 0.9$). Also, we take into account the following different sample sizes $n = 50, 100, 150, 200, 250$ and 300, and the process of value generation is repeated $N = 1000$ times for each value of n.

Then, we consider the following measures. The average estimate (AEs) defined as

$$\text{AE}_n(\xi) = \frac{1}{N}\sum_{k=1}^{N} \hat{\xi}_k,$$

where $\hat{\xi}_k$ denotes the MLE of ξ with $\xi = a, b, \alpha$ or p, obtained at the k-th repetition with the sample size of n, the bias is defined as

$$\text{Bias}_n(\xi) = \left(\frac{1}{N}\sum_{k=1}^{N} \hat{\xi}_k\right) - \xi,$$

and the mean squared error (MSE) is specified by

$$\text{MSE}_n(\xi) = \frac{1}{N}\sum_{k=1}^{N} (\hat{\xi}_k - \xi)^2.$$

The calculations are performed via the R software. The obtained numerical results are listed in Tables 4, 5 and 6 for Sets I, II and III, respectively. From these tables, we note that, as the sample size increases, the biases and the MSEs decrease. That is, the considered estimation method performs well for estimating the parameters of the DKMOE model. Figures 3 and 4 illustrate this claim graphically by plotting the biases and MSEs of the model parameters with respect to n for the three sets of parameters.

[DOI: 10.52547/jirss.20.2.129]

Table 4: Average estimates, biases and MSEs for the parameters of the DKMOE model at Set I ($a = 4.5$, $b = 10$, $\alpha = 9.5$, $p = 0.8$)

	Set I ($a = 4.5$, $b = 10$, $\alpha = 9.5$, $p = 0.8$)		
n	AEs	Biases	MSEs
50	7.0915	1.0915	4.1617
	19.4726	1.4726	6.5296
	12.0229	1.0229	3.8967
	0.7851	-0.0149	0.0086
100	5.0972	0.3972	2.0602
	15.0237	1.0237	5.1462
	11.2768	1.0768	3.6440
	0.7948	-0.0052	0.0047
150	5.3061	0.3061	1.9459
	14.8374	0.8374	2.8372
	8.9247	0.9247	1.0586
	0.7936	-0.0064	0.0050
200	4.6982	0.3018	1.2246
	11.1761	0.1761	1.8667
	10.1263	0.1263	0.8996
	0.7857	-0.0043	0.0054
250	4.5052	0.1948	0.8355
	10.4367	0.2367	0.6592
	9.7691	0.1691	0.7498
	0.7847	-0.0053	0.0045
300	4.5009	-0.1591	0.7371
	10.0095	0.0595	0.3293
	9.0510	0.0510	0.4189
	0.7997	-0.0025	0.0023

[DOI: 10.52547/jirss.20.2.129]

Table 5: Average estimates, biases and MSEs for the parameters of the DKMOE model at Set II ($a = 5$, $b = 3$, $\alpha = 10$, $p = 0.6$)

	Set II ($a = 5$, $b = 3$, $\alpha = 10$, $p = 0.6$)		
n	AEs	Biases	MSEs
50	6.2772	1.2772	4.0050
	8.6686	1.1686	5.7415
	16.9710	1.4701	4.6330
	0.4839	-0.1161	0.0684
100	4.1563	0.8437	2.2870
	7.1396	0.6396	2.0304
	14.0247	0.9247	4.0079
	0.5130	-0.0870	0.0490
150	5.0920	0.6920	1.1088
	4.7607	0.3607	1.5062
	10.5322	0.7322	2.5191
	0.6423	-0.0423	0.0322
200	5.9597	0.4597	0.9356
	3.4524	0.1476	1.4437
	9.8940	0.6940	1.9899
	0.6969	-0.0231	0.0315
250	5.9683	0.3683	0.9157
	2.7971	-0.1029	0.9959
	10.3272	0.2272	0.8023
	0.5900	-0.0100	0.0277
300	5.0339	0.1339	0.2467
	3.0592	-0.0408	0.7152
	10.0823	0.1823	0.6203
	0.6053	-0.0047	0.0236

[DOI: 10.52547/jirss.20.2.129]

Table 6: Average estimates, biases and MSEs for the parameters of the DKMOE model at Set III ($a = 8$, $b = 10$, $\alpha = 5$, $p = 0.9$).

	Set III ($a = 8$, $b = 10$, $\alpha = 5$, $p = 0.9$)		
n	AEs	Biases	MSEs
50	11.3061	1.9061	3.8637
	14.5514	1.5514	2.5878
	7.7942	0.2942	1.0430
	0.8754	-0.0246	0.0049
100	9.8854	1.8854	3.9637
	12.9366	0.9366	1.7521
	5.7678	0.2678	0.9740
	0.8927	-0.0073	0.0017
150	8.7045	0.7045	3.7028
	13.6911	0.6911	1.6143
	5.4014	0.1914	0.5535
	0.8907	-0.0093	0.0019
200	8.9095	0.7095	2.6514
	10.9972	0.3972	1.2700
	6.1110	0.1610	0.3607
	0.8843	-0.0157	0.0022
250	8.2654	0.5654	2.1727
	10.1669	0.1669	0.6530
	5.4111	0.1011	0.1546
	0.8904	-0.0096	0.0014
300	8.1883	-0.2117	1.1373
	10.0709	0.1009	0.1759
	5.0850	0.0850	0.1272
	0.8977	-0.0023	0.0012

[DOI: 10.52547/jirss.20.2.129]

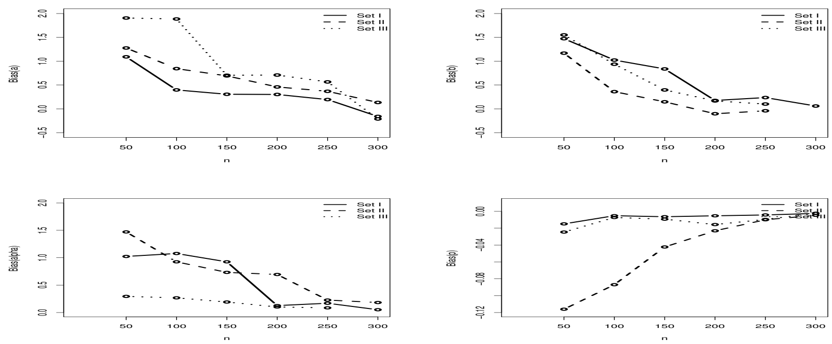

Figure 3: Plots for the biases of the parameters of the DKMOE model for Sets I, II and III.

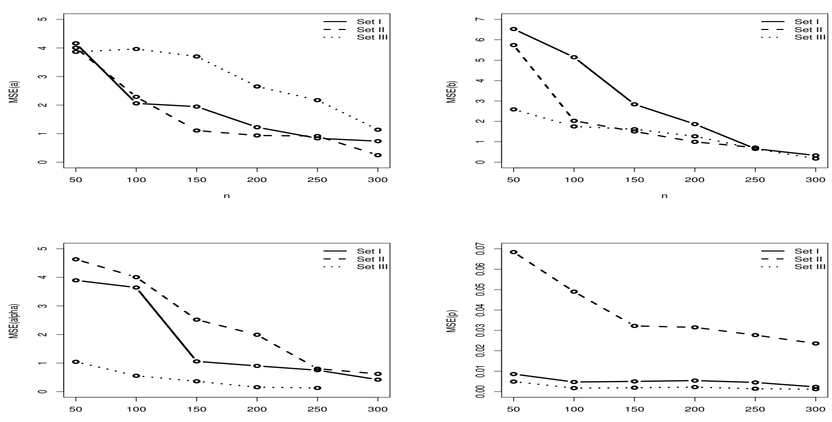

Figure 4: Plots for the MSEs of the parameters of the DKMOE model for Sets I, II and III.

5 Application

In this section, we present three data sets to assess the performance of a newly proposed DKMOE model. For comparing the superiority of the DKMOE model

[DOI: 10.52547/jirss.20.2.129]

over the KG and geometric models, we estimated the values of unknown parameters by the maximum likelihood method, standard error (SE), −log likelihood (−logL), Akaike information criterion (AIC), Bayesian information criterion (BIC), Kolmogorov-Smirnov (K-S) statistic, and K-S p-value. The three considered data sets are presented below.

Data set 1 : The first data set represents the number of shocks received prior to the failure being reported in Murthy et al. (2004, p.245). The data set is:

1 3 3 4 4 4 4 5 5 6 6 7 10 11 12 14

Data set 2: The second data set consists of the number of inspections between the discovery of defects in an industrial process and is taken from Xie and Goh (1993). The data set is:

1 1 1 1 1 1 2 2 2 2 3 3 3 4 4 4 5 5 7 9 11 13 14 14 17 18 26 29

Data set 3: The third data set contains the integer part of the lifetime of fifty devices in weeks which is given by Aarset (1987). The data set is:

0 0 1 1 1 1 1 2 3 6 7 11 12 18 18 18 18 18 21 32 36 40 45 46 47 50 55
60 63 63 67 67 67 67 72 75 79 82 82 83 84 84 84 85 85 85 85 85 86 86

Tables 7, 8 and 9 list the values of the descriptive study for the fitted DKMOE, KG and geometric models for Data set 1, 2 and 3 respectively. The smallest −logL, AIC, BIC, K-S statistic and the highest p-values are achieved for the DKMOE model. Thus, it is the best model compared to the others. Therefore, it should be preferred for fitting the current data sets. In Figures 6a, 6b and 6c, we give the estimated pmfs for Data set 1, 2 and 3, respectively. It clearly shows that the DKMOE model captures the general pattern of the histograms. On the other hand, Figures 5a, 5b and 5c show the comparison of the cdfs for each model with the empirical distribution function. These plots also indeed affirm that the DKMOE model is clearly a competitive model for the considered data sets. Thus, the new model may be an interesting alternative to the geometric and KG models for modeling positive real data.

[DOI: 10.52547/jirss.20.2.129]

Table 7: Estimated values, −logL, AIC, BIC, K-S statistics and p-value for Data set 1

Distribution	Estimates(SE)	−logL	AIC	BIC	K-S	p-value
DKMOE	$\hat{a}$ = 5.7053(9.9333)	41.2162	90.4323	93.5227	0.1890	0.6173
	$\hat{b}$ = 2.2710(9.4137)					
	$\hat{p}$ = 0.8594(0.5445)					
	$\hat{\alpha}$ = 0.3859(1.4609)					
KG	$\hat{a}$ = 0.9732(0.0349)	43.7760	93.5520	95.8697	0.2747	0.1786
	$\hat{b}$ = 0.0599(0.0155)					
	$\hat{p}$ = 0.0808(0.0013)					
geometric	$\hat{p}$ = 0.13910.0322	46.3889	94.7778	95.5504	0.3883	0.0160

Table 8: Estimated values, −logL, AIC, BIC, K-S statistics and p-value for Data set 2

Distribution	Estimates(SE)	−logL	AIC	BIC	K-S	p-value
DKMOE	$\hat{a}$ = 1.9348(0.0192)	81.7253	171.4507	171.7795	0.1513	0.5415
	$\hat{b}$ = 0.1163(0.0231)					
	$\hat{p}$ = 0.27880(0.0039)					
	$\hat{\alpha}$ = 2.5008(1.1022)					
KG	$\hat{a}$ = 0.9375(0.0095)	83.17498	172.35	176.3466	0.17853	0.334
	$\hat{b}$ = 0.0964(0.0186)					
	$\hat{p}$ = 0.2776(0.0062)					
geometric	$\hat{p}$ = 0.1212(0.0215)	85.31595	172.6319	173.9641	0.22773	0.1096

[DOI: 10.52547/jirss.20.2.129]

Table 9: Estimated values, −logL, AIC, BIC, K-S statistics and p-value for Data set 3

Distribution	Estimates(SE)	−logL	AIC	BIC	K-S	p-value
DKMOE	$\hat{a} = 0.6137(0.1498)$	233.3173	474.6346	482.2827	0.1716	0.1054
	$\hat{b} = 1.4944(0.9124)$					
	$\hat{p} = 0.9679(0.0102)$					
	$\hat{\alpha} = 9.8793(5.3413)$					
KG	$\hat{a} = 0.4987(0.3082)$	240.1928	486.3855	492.1216	0.18549	0.06409
	$\hat{b} = 0.1129(0.1299)$					
	$\hat{p} = 0.8367(0.1758)$					
geometric	$\hat{p} = 0.02142(0.0030)$	241.6264	485.2527	487.1647	0.19310	0.048

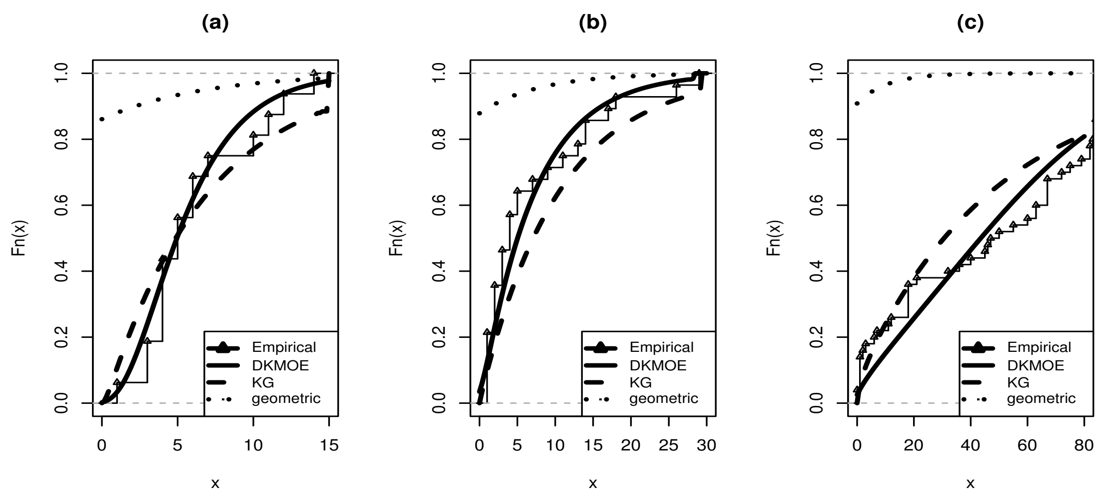

Figure 5: Plots of estimated cdfs of models for (a) Data set 1 (b) Data set 2 and (c) Data set 3

[DOI: 10.52547/jirss.20.2.129]

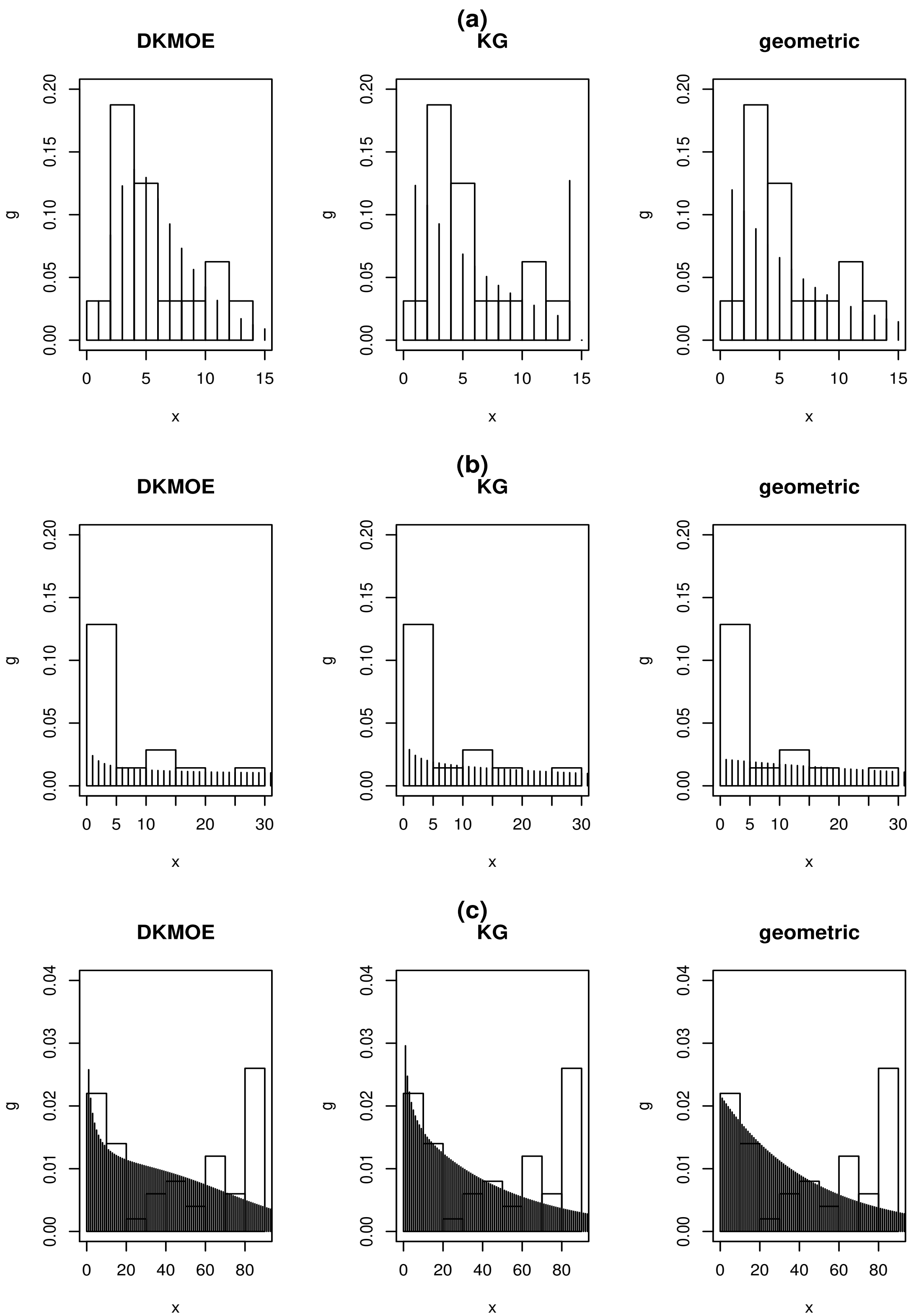

Figure 6: Plots of estimated pmfs of models for (a) Data set 1 (b) Data set 2 and (c) Data set 3

[DOI: 10.52547/jirss.20.2.129]

6 Conclusions

In this paper, we introduced a new discrete family of distributions and proposed a new discrete model as its particular case. This new distribution is really another generalization of the geometric distribution and also unfolds a number of existing distributions as sub-models. We have discussed explicit mathematical expressions for some of its basic statistical properties, such as the pmf, cdf, hrf, mean, variance, quantile function, order statistics and the entropy measure. The method of maximum likelihood estimation is used in estimating the parameters of the new DMOE model. Real data applications show the tractability of the proposed distribution in data modeling.

Acknowledgement

The authors are thankful to the associate editor and anonymous reviewers for the constructive comments.

References

Aarset, M. V. (1987), How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, **36**(1), 106-108.

Akinsete, A., Famoye, F., and Lee, C. (2014), The Kumaraswamy-geometric distribution. *Journal of Statistical Distributions and Applications*, 1, 1-21.

Alizadeh, M. Tahir, M. H., Cordeiro, G. M., Zubair, M., and Hamedani, G. G. (2015), The Kumaraswamy Marshall-Olkin family of distributions. *Journal of the Egyptian Mathematical Society*, **23**(3), 546-557.

Alzaatreh, A., Lee, C., and Famoye, F. (2012), On the discrete analogues of continuous distributions. *Statistical Methods*, **9**(6), 589-603.

Amigo, J. M., Balogh, S. G., and Hernandez, S. (2018), A brief review of generalized entropies. *Entropy*, **20**, 813.

Chakraborty, S. (2015), A new discrete distribution related to generalized gamma distribution and its properties. *Communication in Statistics-Theory and Methods*, **44**(8), 1691-1705.

[DOI: 10.52547/jirss.20.2.129]

Chakraborty, S., and Chakravarty, D. (2016), A new discrete probability distribution with integer support on $(-\infty, \infty)$. *Communication in Statistics-Theory and Method*, **45**(2), 492-505.

Cordeiro, G. M., and de Castro, M. (2011), A new family of generalized distributions. *Journal of Statistics Computation and Simulation*, **81**(7), 883-893.

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London: Chapman & Hall.

Gómez-Déniz, E. (2010), Another generalization of the geometric distribution. *Test*, **19**(2), 399-415.

Gupta, R. C., and Gupta, R. D. (2007), Proportional reversed hazard rate model and its applications. *Journal of Statistics and Planning Inference*, **137**(11), 3525-3536.

Gupta, R. C., Gupta, P. L., and Gupta, R. D. (1998), Modeling failure time data by Lehman alternatives. *Communications in Statistics-Theory Methods*, **27**(4), 887-904.

Jayakumar, K., and Thomas, M. (2008), On a generalization of Marshall-Olkin scheme and its application to Burr type XII distribution. *Statistical Papers*, **49**(3), 421-439.

Jayakumar, K., and Sankaran, K. K. (2017a), A discrete generalization of Marshall-Olkin scheme and its application to geometric distribution. *Journal of the Kerala Statistical Association*, **28**, 1-21.

Jayakumar, K., and Sankaran, K. K. (2017b), A generalization of discrete Weibull distribution. *Communications in Statistics-Simulation and Computation*, **47**(24), 6064-6078.

Krishna, H., and Pundir, P. S. (2009), Discrete Burr and discrete Pareto distributions. *Statistical Methodology*, **6**(2), 177-188.

Lisman, J. H. C., and van Zuylen, M. C. A. (1972), Note on the generation of the most probable frequency distribution. *Statistica Neerlandica*, **26**(1), 19-23.

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

Murthy, D. N. P., Xie, M., and Jiang, R. (2004), *Weibull models*. New Jersey: John Wiley and Sons.

[DOI: 10.52547/jirss.20.2.129]

Nakagawa, T., and Osaki, S. (1975), The discrete Weibull distribution. *IEEE Transactions on Reliability*, **24**(5), 300-301.

Sato, H., Ikota, M., Aritoshi, S., and Masuda, H. (1999), A new defect distribution meteorology with a consistent discrete exponential formula and its applications. *IEEE Transactions on Semiconductor Manufacturing*, **12**(4), 409-418.

Stein, W. E., and Dattero, R. (1984), A new discrete Weibull distribution. *IEEE Transactions on Reliability*, **33**(2), 196-197.

Steutel, F. W., and van Harn, K. (2004), *Infinite Divisibility of Probability Distributions on the Real Line*. New York: Marcel Dekker.

Supanekar, S. R., and Shirke, D. T. (2015), A new discrete family of distributions. *ProbStat Forum*, **8**, 83-94.

Tahir, M. H., and Nadarajah, S. (2015), Parameter induction in continuous univariate distributions: Well established G families. *Annals of the Brazilian Academy of Sciences*, **87**(2), 539-568.

Xie, M., and Goh, T. N. (1993), Improvement detection by control charts for high yield processes. *International Journal of Quality & Reliability Management*, **10**(7), 24-31.

[DOI: 10.52547/jirss.20.2.129]