Model Assisted Statistics and Applications 14 (2019) 255–268
DOI 10.3233/MAS-190465
IOS Press

Use of a discrete Sushila distribution in the analysis of right-censored lifetime data

Ricardo Puziol de Oliveira*, Marcos Vinicius de Oliveira Peres, Edson Zangiacomi Martinez and Jorge Alberto Achcar
Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil

Abstract. In this paper it is introduced a new two-parameter discrete distribution derived from the continuous Sushila distribution (Shanker et al., 2013). Its mathematical properties and estimation procedures for the parameters of the proposed model are presented assuming complete and right-censored data. This new model, in the same way as the continuous Sushila distribution, has the discrete Lindley distribution as a special case. An extensive simulation study is carried out to examine the bias and the roots of the mean squared errors for the maximum likelihood estimators as well the moments and Bayesian estimators of the proposed model parameters. Some examples using simulated data and real datasets are considered to show that the new proposed model performs at least as good as its particular case and some other traditional discrete models as the Poisson and geometric distributions.

Keywords: Bayesian analysis, discrete Lindley, discrete Sushila, Monte Carlo simulation, MCMC methods, right-censoring, survival data

1. Introduction

In the recent decades, the construction of new discrete distributions using discretization methods has been widely used in the literature. Basically, the main purpose of the discretization is to generate distributions that can be used for the analysis of strictly discrete data as alternatives to the traditional discrete distributions as the Poisson and geometric distributions. A field of study where the discretization process is widely needed is in the analysis of lifetime data where it is common the use of continuous distributions to model lifetime data, which could be discrete, usually in presence of censored data. Several applications where continuous distributions are used to model discrete data can be found in Klein and Moeschberger (1997), Meeker and Escobar (1998), Kalbfleisch and Prentice (2002), Lee and Wang (2003), Lawless (2003), Collett (2003), Hamada et al. (2008), among many others. A complete survey regarding all discretization methods introduced in the literature and some discretized distributions can be found in Chakraborty (2015).

One of the first proposed discretization methods presented in the literature is based on the definition of a probability mass function based on a infinite series. The first foundations of this method were presented by Good (1953), that proposed the discrete Good distribution to model population frequencies of species. Such approach was considered by other authors to define discrete analogues, for example: Haight (1957) proposed the discrete Pearson III distribution to model queues with baking; Siromoney (1964) introduced the Dirichlet's Series distribution as an alternative to model the frequency of wet days (rain-spells); Kemp (1997) formally introduced the discrete normal distribution and derived its main mathematical properties; the discrete exponential distribution was proposed by Sato

*Corresponding author: Ricardo Puziol de Oliveira, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, S.P., Brazil. E-mail: rpuziol.oliveira@gmail.com.

ISSN 1574-1699/19/$35.00

et al. (1999) to describe the defect count frequencies on wafers or chips; Bi et al. (2001) introduced the discrete log-normal distribution; Inusah and Kozubowski (2006) presented the discrete Laplace distribution also arguing that, in comparison to the discrete normal distribution, the proposed model has closed forms for the probability mass function, for the generating functions and for the central moments; the skewed version of the discrete Laplace distribution was proposed by Kozubowski and Inusah (2006); Doray and Luong (1997) presented efficient estimators for the parameters of the Good distribution family; Kemp (2008) introduced the discrete half-normal distribution also presenting its relation with other existing distributions and Nekoukhou et al. (2012, 2013) proposed the discrete generalized exponential distribution as an attempt to model rank frequencies of graphemes in the Slovene language. The method of discretization by infinite series is characterized by the following definition.

Definition 1. Let X be a continuous random variable. If X has probability density function $f_X(x;\boldsymbol{\theta})$ with support on $\mathbb{R}$, then the corresponding discrete random variable Y has probability mass function given by

$$\Pr(Y = y|\boldsymbol{\theta}) = \frac{f_X(y|\boldsymbol{\theta})}{\sum\limits_{j=-\infty}^{\infty} f_X(j|\boldsymbol{\theta})}, \quad y \in \mathbb{Z}, \tag{1}$$

being $\boldsymbol{\theta}$ the vector of parameters indexing the distribution of X. Observe that if the random variable X is defined on $\mathbb{R}_+$, the probability mass function of Y becomes

$$\Pr(Y = y|\boldsymbol{\theta}) = \frac{f_X(y|\boldsymbol{\theta})}{\sum\limits_{j=0}^{\infty} f_X(j|\boldsymbol{\theta})}, \quad y \in \mathbb{Z}_+. \tag{2}$$

One of the most recent examples of the use of this method is provided by the discrete analogue of the generalized exponential probability distribution introduced by Nekoukhou et al. (2012) having probability mass function given by,

$$\Pr(Y = y|\alpha,\lambda) = \lambda^{x-1}(1-\lambda^x)^{\alpha-1}\left[\sum_{i=1}^{\infty}\binom{\alpha-1}{j}\frac{(-1)^j\lambda^j}{1-\lambda^{1+j}}\right]^{-1}, \quad y \in \mathbb{Z}_+, \tag{3}$$

for $\alpha \in \mathbb{R}_+$ and $\lambda \in (0,1)$.

The main goal of this paper is to use the infinite series method to propose a discrete analogue for the Sushila distribution which is a two-parameter lifetime model proposed by Shanker et al. (2013). In this way, it is expected that the proposed discrete Sushila distribution could be a suitable alternative for survival data, especially in the presence of right-censored data. The Sushila distribution has the one-parameter Lindley distribution (Ghitany et al., 2008) as a special particular case and can also be written as a mixture of probability distributions in the same way as the Lindley distribution.

Le X be a continuous random variable with a Sushila probability distribution. The Sushila probability density function (pdf) is given by,

$$f_X(x|\alpha,\theta) = \frac{\theta^2}{\alpha(\theta+1)}\left(1+\frac{x}{\alpha}\right)\exp\left\{-\frac{\theta}{\alpha}x\right\}, \quad x \in \mathbb{R}_+, \tag{4}$$

where $\theta \in \mathbb{R}_+$ and $\alpha \in \mathbb{R}_+$. Shanker et al. (2013) have shown that this model is a two component mixture of an exponential distribution with scale parameter θ/α and a gamma distribution having shape parameter equal to 2 and a scale parameter θ/α with mixing proportion given by $\theta/(\theta+1)$. A comprehensive discussion about the mathematical properties of the Sushila distribution such as moments, hazard function, stochastic orderings, parameter estimation, among others is also presented by Shanker et al. (2013). The corresponding survival function is given by

$$S_X(x|\alpha,\theta) = \frac{\alpha(\theta+1)+\theta x}{\alpha(\theta+1)}\exp\left\{-\frac{\theta}{\alpha}x\right\}, \quad x \in \mathbb{R}_+. \tag{5}$$

This paper is organized as follows: In Section 2, it is introduced the discrete Sushila distribution and derived the main mathematical properties. In Section 3, the estimation of the parameters and inference procedures under Classical and Bayesian approaches are presented. In Section 4, a Monte Carlo simulation study is carried out to evaluate the performance of the presented estimators. In Section 5, applications of the proposed model to real datasets are considered to illustrate its usefulness. Some concluding remarks are presented in Section 6.

2. The discrete Sushila distribution

A first approach of the discrete Sushila distribution was introduced in the literature by Borah and Saikia (2016). These authors used the discretization idea based on the survival function proposed by Nakagawa and Osaki (1975) in the discretization of the Weibull distribution to introduce a new discrete Sushila distribution. In this paper, it is introduced another approach to construct a discrete Sushila distribution based on the discretization method by infinite series previously described.

Let X be a continuous random variable following Sushila distribution with parameters α and θ and pdf given by Eq. (4). Thus using the discretization approach given in Eq. (2), the probability mass function (pmf) of the discrete Sushila (DS) distribution with parameters α and θ is given by,

$$\Pr(X = x|\alpha,\theta) = \frac{(\alpha+x)\gamma^{\theta x}(\gamma^{-\theta}-1)^2}{\gamma^{-\theta}(\gamma^{-\theta}\alpha-\alpha+1)}, \tag{6}$$

where $x \in \mathbb{Z}_+$, $\alpha,\theta \in \mathbb{R}_+$ and $0 < \gamma = \exp\{-1/\alpha\} < 1$. Observe that, when $\alpha = 1$, the pmf given in Eq. (6) reduces to the discrete Lindley distribution introduced by Oliveira et al. (2017) which is expected since the continuous Sushila distribution has the continuous Lindley distribution as a particular case.

The corresponding cumulative distribution function (cdf) and survival function (sf) of the DS distribution are given, respectively, by,

$$\Pr(X \leqslant x|\alpha,\theta) = 1 - \frac{(\alpha+x)\gamma^{\theta(x-1)} - (x+\alpha-1)\gamma^{\theta x}}{\gamma^{-\theta}\alpha-\alpha+1}, \tag{7}$$

and,

$$\Pr(X > x|\alpha,\theta) = \frac{(\alpha+x)\gamma^{\theta(x-1)} - (x+\alpha-1)\gamma^{\theta x}}{\gamma^{-\theta}\alpha-\alpha+1}. \tag{8}$$

The pmf Eq. (6) does not involves complicated expressions and therefore, the probabilities can be straightforwardly computed, as for example,

$$\Pr(X = 0|\alpha,\theta) = \frac{\alpha(\gamma^{-\theta}-1)^2}{\gamma^{-\theta}(\gamma^{-\theta}\alpha-\alpha+1)},$$

for $\alpha,\theta \in \mathbb{R}_+$. Figure 1 illustrates the behavior of the pmf Eq. (6) for selected values of α and θ.

On other hand, the pmf Eq. (6) satisfies the the log-concave inequality $\Pr^2(X = x) \geqslant \Pr(X = x-1)\Pr(X = x+1)$ for $x \geqslant 1$ which implies unimodality (see Keilson & Gerber, 1971). The relationship between log-concavity, unimodality and increasing hazard rate of discrete distributions has been discussed by Grandell (1997). In this way, the mode of DS distribution is given by,

$$r(\alpha,\theta) = \begin{cases} \left\lceil \frac{1+\alpha-\alpha\gamma^{-\theta}}{\gamma^{-\theta}-1} \right\rceil, \text{ if } \frac{1+\alpha-\alpha\gamma^{-\theta}}{\gamma^{-\theta}-1} \notin \mathbb{Z}_+ \text{ and } 0<\theta<1 \\ \frac{1+\alpha-\alpha\gamma^{-\theta}}{\gamma^{-\theta}-1}, \quad \text{if } \frac{1+\alpha-\alpha\gamma^{-\theta}}{\gamma^{-\theta}-1} \in \mathbb{Z}_+ \text{ and } 0<\theta<1, \end{cases} \tag{9}$$

where $\lceil\cdot\rceil$ is the ceiling function. For instance, if $(\alpha,\theta) = (1.5, 0.7)$ then $r(\alpha,\theta) \approx 0.1816$ and hence, the mode is 1, as can be seen in the left-panel of the Fig. 1. Moreover, the following relations between pmf and mode are easily obtained from Eqs (6) and (9),

i) $\Pr(X = x+1;\theta) < \Pr(X = x;\theta)$ if $x > r(\alpha,\theta)$;
ii) $\Pr(X = x+1;\theta) = \Pr(X = x;\theta)$ if $x = r(\alpha,\theta)$;
iii) $\Pr(X = x+1;\theta) > \Pr(X = x;\theta)$ if $x < r(\alpha,\theta)$.

It follows from Eqs (6) and (8) that the hazard rate function (hf) of the DS distribution is concave and increasing in $x \in \mathbb{Z}_+$ and is given by,

$$h(x|\alpha,\theta) = \frac{\Pr(X = x|\alpha,\theta)}{\Pr(X > x|\alpha,\theta)} = \frac{(\alpha+x)\gamma^{\theta x}(\gamma^{-\theta}-1)^2}{\gamma^{-\theta}[(\alpha+x)\gamma^{\theta(x-1)-(x+\alpha-1)\gamma^{\theta x}}]}, \tag{10}$$

where $h(0|\alpha,\beta) = \Pr(X = 0|\alpha,\theta)$ and $h(\infty) = 1-\gamma^{\theta} = 1-\exp\{-\theta/\alpha\} < 1$ for $\alpha,\theta > 0$. Therefore, the hf of the DS distribution is limited to the interval (0,1). Figure 1 also illustrates the behavior of the hf Eq. (10) for selected values of α and θ.

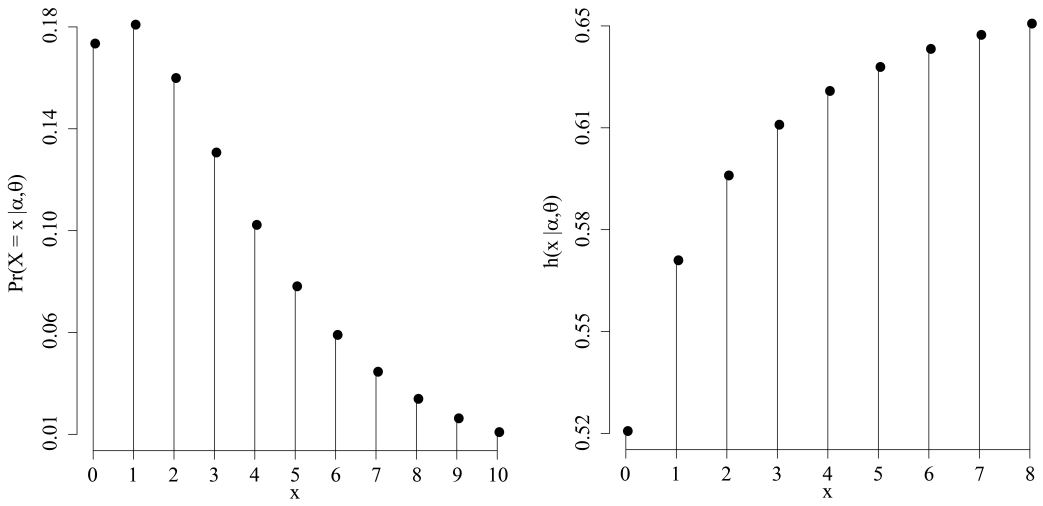

Fig. 1. Behavior of the pmf (left-panel: $(\alpha, \theta) = (1.5, 0.7)$) and hf (right-panel: $(\alpha, \theta) = (1.5, 1.7)$) of the DS distribution.

The quantile function of the DS distribution, say $Q(u)$, defined by $F(Q(u)) = u$ depends on the Lambert W function (Lambert, 1758; Jodrá, 2010) and is given by,

$$Q(u|\alpha,\theta) = \left\lfloor -\frac{\alpha}{\theta} W_{-1}\left(\frac{\theta(1-u)(\gamma^{-\theta}\alpha-\alpha+1)\exp\left\{-\frac{\theta(\gamma^{-\theta}\alpha-\alpha+1)}{\alpha(\gamma^{-\theta}-1)}\right\}}{\alpha(\gamma^{-\theta}-1)}\right) - \frac{\gamma^{-\theta}\alpha-\alpha+1}{\gamma^{-\theta}-1}\right\rfloor, \tag{11}$$

where $0 < u < 1$, $\lfloor\cdot\rfloor$ is the floor function and $W_{-1}(\cdot)$ is the Lambert W function with negative branch.

The last property that will be discussed in this paper is related to the moments of the DS distribution. Let X be a discrete random variable such that $X \sim \text{DS}(\alpha, \theta)$, then the r^{th} raw moment for the DS distribution depends on the polylogarithm $L_s(x)$ (for more details, see Miller, 2008) and it is given by,

$$\mathbb{E}[X^r] = \frac{(\gamma^{-\theta}+\gamma^{\theta}-2)[\alpha Li_{-r}(\gamma^{\theta}) + Li_{-1-r}(\gamma^{\theta})]}{\gamma^{-\theta}\alpha-\alpha+1}, \tag{12}$$

from where, taking $r = 1, 2, 3, 4$, the first four moments around the origin (raw moments) of the DS distribution are obtained, respectively, as,

$$\begin{aligned}
\mathbb{E}[X] &= \frac{\gamma^{-\theta}+\gamma^{-\theta}\alpha-\alpha+1}{(\gamma^{-\theta}\alpha-\alpha+1)(\gamma^{-\theta}-1)},\\
\mathbb{E}[X^2] &= \frac{4\gamma^{-\theta}+(\alpha+1)\gamma^{-2\theta}-\alpha+1}{(\gamma^{-\theta}\alpha-\alpha+1)(\gamma^{-\theta}-1)^2},\\
\mathbb{E}[X^3] &= \frac{(11-3\alpha)\gamma^{-\theta}+(3\alpha+11)\gamma^{-2\theta}+(\alpha+1)\gamma^{-3\theta}-\alpha+1}{(\gamma^{-\theta}\alpha-\alpha+1)(\gamma^{-\theta}-1)^3},\\
\mathbb{E}[X^4] &= \frac{(26-10\alpha)\gamma^{-\theta}+66\gamma^{-2\theta}+(10\alpha+26)\gamma^{-3\theta}+(\alpha+1)\gamma^{-4\theta}-\alpha+1}{(\gamma^{-\theta}\alpha-\alpha+1)(\gamma^{-\theta}-1)^2(\gamma^{-2\theta}-2\gamma^{-\theta}+1)}.
\end{aligned}$$

It is important to point out that the mean of the DS distribution is always greater than the mode, that is, the DS distribution is positively skewed. The mean and the variance for the DS distribution are given, respectively, by,

$$\begin{aligned}
\mu = \mathbb{E}[X] &= \frac{\gamma^{-\theta}+\gamma^{-\theta}\alpha-\alpha+1}{(\gamma^{-\theta}\alpha-\alpha+1)(\gamma^{-\theta}-1)},\\
\sigma^2 = \mathbb{E}[X^2] - \mathbb{E}[X] &= \frac{(1+\alpha)(\gamma^{-\theta}\alpha-\alpha+1)\gamma^{-2\theta}-(\gamma^{-\theta}\alpha+(1-\alpha)\alpha+2)\gamma^{-\theta}}{(\gamma^{-\theta}\alpha-\alpha+1)^2(\gamma^{-\theta}-1)^2},
\end{aligned} \tag{13}$$

where $\sigma^2 > \mu$ if $\theta < \epsilon$, $\sigma^2 < \mu$ if $\theta > \epsilon$ and $\sigma^2 = \mu$ if $\theta = \epsilon$ where ϵ is given by,

$$\epsilon = \ln\left[\frac{(\alpha+1+\sqrt{2})(\alpha-1)}{\alpha^2+2\alpha-1}\right]\alpha. \tag{14}$$

Now, taking the ratio between the variance and the expected value, one can define the dispersion index and the coefficient of variation of the DS distribution, respectively, as,

$$\text{DI} = \frac{\sigma^2}{\mu} = \frac{(1+\alpha)(\gamma^{-\theta}\alpha-\alpha+1)\gamma^{-2\theta} - (\gamma^{-\theta}\alpha+(1-\alpha)\alpha+2)\gamma^{-\theta}}{(\gamma^{-\theta}\alpha-\alpha+1)(\gamma^{-\theta}-1)(\gamma^{-\theta}\alpha+\gamma^{-\theta}-\alpha+1)}, \tag{15}$$

$$\text{CV} = \frac{\sigma}{\mu} = \frac{\sqrt{\frac{(1+\alpha)(\gamma^{-\theta}\alpha-\alpha+1)\gamma^{-2\theta}-(\gamma^{-\theta}\alpha+(1-\alpha)\alpha+2)\gamma^{-\theta}}{(\gamma^{-\theta}\alpha-\alpha+1)^2(\gamma^{-\theta}-1)^2}}(\gamma^{-\theta}\alpha-\alpha+1)(\gamma^{-\theta}-1)}{\gamma^{-\theta}+\gamma^{-\theta}\alpha-\alpha+1}. \tag{16}$$

The asymmetry degree and the flatness of a distribution are usually measured by their coefficients of skewness and kurtosis, respectively. The first one can be computed by the third central moment normalized by the variance raised to the power $3/2$ and the latter is given by the fourth central moment divided by the square of the variance. These coefficients are quite important to characterize the shape of a distribution but, for the DS model, extensive and very complicated expressions are obtained for such measures. For this reason, the expressions of these coefficients will be omitted of this study.

3. Estimation methods

In this section, it is estimated the unknown parameters of the DS distribution using the methods of moments, maximum likelihood method and the Bayesian method. For maximum likelihood and Bayesian methods, it is also presented the estimation procedure in presence of right-censored data.

3.1. Method of moments

The method of moments is the simplest technique commonly used in parameter estimation when the raw moments has a closed analytical form. For the DS distribution, first, let $m_1 = 1/n\sum_{i=1}^n x_i$ and $m_2 = 1/n\sum_{i=1}^n x_i^2$ be the first two sample moments, respectively. The method of moments estimates (MOM) $\hat{\alpha}$ and $\hat{\theta}$ for α and θ are obtained by solving the equations,

$$\mathbb{E}[X|\hat{\alpha},\hat{\theta}] = m_1 \quad \text{and} \quad \mathbb{E}[X^2|\hat{\alpha},\hat{\theta}] = m_2, \tag{17}$$

where $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ are the first two raw moments of DS distribution given in Eq. (13). Solving both equations simultaneously, the MOM of θ has closed form and is given by,

$$\hat{\theta}_{MOM} = \ln\left(\frac{2\bar{x}\hat{\alpha}-\bar{x}+\hat{\alpha}+1+\sqrt{\bar{x}^2+6\bar{x}\hat{\alpha}+\hat{\alpha}^2-2\bar{x}+2\hat{\alpha}+1}}{2\bar{x}\hat{\alpha}}\right)\hat{\alpha}. \tag{18}$$

For the MOM estimator of α, there is no closed analytical form for the moment estimator. Thus, in this paper, the MOM estimator of α was computed using the *nleqslv* package of the R software considering 'Newton' as optimization method.

3.2. Maximum likelihood method

3.2.1. Complete data

Considering $x_1,\ldots,x_n$ a random sample of the DS distribution with parameters α and θ, and pmf given by Eq. (6), the likelihood function could be written as,

$$L(\alpha,\theta|\mathbf{x}) = \prod_{i=1}^n \frac{(\alpha+x_i)\gamma^{\theta x_i}(\gamma^{-\theta}-1)^2}{\gamma^{-\theta}(\gamma^{-\theta}\alpha-\alpha+1)}. \tag{19}$$

From Eq. (19), the log-likelihood can be written as,

$$\ell(\alpha,\beta|\mathbf{x}) = \sum_{i=1}^{n} \ln(\alpha + x_i) - \frac{\theta n \bar{x}}{\alpha} + 2n\ln(\gamma^{-\theta} - 1) - \frac{n\theta}{\alpha} - n\ln(\gamma^{-\theta}\alpha - \alpha + 1), \tag{20}$$

which is maximized solving numerically, in α and θ, the non-linear system of the equations,

$$U_n = \begin{cases} \frac{\partial \ell}{\partial \alpha} = \sum\limits_{i=1}^{n} \frac{1}{\alpha + x_i} + \frac{\theta n \bar{x}}{\alpha^2} - \frac{2n\theta\gamma^{-\theta}}{\alpha^2(\gamma^{-\theta}-1)} + \frac{n\theta}{\alpha^2} - \frac{n\left(-\frac{\theta\gamma^{-\theta}}{\alpha} + \gamma^{-\theta} - 1\right)}{\gamma^{-\theta}\alpha - \alpha + 1} = 0 \\ \frac{\partial \ell}{\partial \theta} = -\frac{n\bar{x}}{\alpha} + \frac{2n\gamma^{-\theta}}{\alpha(\gamma^{-\theta}-1)} - \frac{n}{\alpha} - \frac{n\gamma^{-\theta}}{\gamma^{-\theta}\alpha - \alpha + 1} = 0 \end{cases}. \tag{21}$$

From Eq. (21), it is obtained that $\bar{x} = \mu$, where μ is the mean of the DS distribution, which implies that $\hat{\theta}_{MLE} = \hat{\theta}_{MOM}$, that is,

$$\hat{\theta}_{MLE} = \ln\left(\frac{2\bar{x}\hat{\alpha} - \bar{x} + \hat{\alpha} + 1 + \sqrt{\bar{x}^2 + 6\bar{x}\hat{\alpha} + \hat{\alpha}^2 - 2\bar{x} + 2\hat{\alpha} + 1}}{2\bar{x}\hat{\alpha}}\right) \hat{\alpha} = \hat{\theta}_{MOM}. \tag{22}$$

On other hand, there is no closed analytical form for the maximum likelihood estimator (MLE) for α. Thus, it is needed in the estimation procedure the use of standard numerical optimization algorithms such the Newton-Raphson, BFGS or Nelder-Mead methods. In this paper, the estimation was done by the *fitdistplus* package of the R software and the BFGS optimization method.

Now, under suitable regularity conditions (see, Lehmann & Casella, 1998, pp. 461–463), the asymptotic distribution of the maximum likelihood estimator $(\hat{\alpha}, \hat{\theta})$ is a multivariate Normal distribution with mean (α, θ) and covariance matrix $\mathbf{\Sigma}(\hat{\mathbf{\Omega}}), \hat{\mathbf{\Omega}} = (\hat{\alpha}, \hat{\theta})$, which can be consistently estimated by the inverse of the observed Fisher's information matrix I_0 given by,

$$I_0(\hat{\alpha}, \hat{\theta}) = \begin{pmatrix} -\frac{\partial^2 \ell}{\partial^2 \alpha^2} & -\frac{\partial^2 \ell}{\partial^2 \alpha \theta} \\ -\frac{\partial^2 \ell}{\partial^2 \theta \alpha} & -\frac{\partial^2 \ell}{\partial^2 \theta^2} \end{pmatrix}_{\hat{\alpha},\hat{\theta}}, \tag{23}$$

where,

$$\frac{\partial^2 \ell}{\partial^2 \theta^2} = \frac{2n\gamma^{-\theta}}{\alpha^2(\gamma^{-\theta}-1)}\left(1 - \frac{\gamma^{-\theta}}{\gamma^{-\theta}-1}\right) - \frac{n\gamma^{-\theta}}{\gamma^{-\theta}\alpha - \alpha + 1}\left(\frac{1}{\alpha} - \frac{\gamma^{-\theta}}{\gamma^{-\theta}\alpha - \alpha + 1}\right),$$

$$\frac{\partial^2 \ell}{\partial^2 \alpha^2} = -\sum_{i=1}^{n} \frac{1}{(\alpha + x_i)^2} - \frac{2\theta n\bar{x}}{\alpha^3} + \frac{4n\theta\gamma^{-\theta}}{\alpha^3(\gamma^{-\theta}-1)} + \frac{2n\theta\gamma^{-\theta}}{\alpha^4(\gamma^{-\theta}-1)}\left(1 - \frac{\gamma^{-\theta}}{\gamma^{-\theta}-1}\right)$$
$$- \frac{2n\theta}{\alpha^3} - \frac{n\theta^2\gamma^{-\theta}}{\alpha^3(\gamma^{-\theta}\alpha - \alpha + 1)} + \frac{n\left(-\frac{\theta\gamma^{-\theta}}{\alpha} + \gamma^{-\theta} - 1\right)^2}{(\gamma^{-\theta}\alpha - \alpha + 1)^2},$$

$$\frac{\partial^2 \ell}{\partial^2 \alpha\theta} = \frac{\partial^2 \ell}{\partial^2 \theta\alpha} = \frac{n\bar{x}}{\alpha^2} + \frac{2n\gamma^{-\theta}}{\alpha^2(\gamma^{-\theta}-1)}\left(\frac{\theta\gamma^{-\theta}}{\alpha(\gamma^{-\theta}-1)} - \frac{\theta}{\alpha} - 1\right) + \frac{n}{\alpha^2} + \frac{n\theta\gamma^{-\theta}}{\alpha^2(\gamma^{-\theta}\alpha - \alpha + 1)}$$
$$+ \frac{n\left(-\frac{\theta\gamma^{-\theta}}{\alpha} + \gamma^{-\theta} - 1\right)\gamma^{-\theta}}{(\gamma^{-\theta}\alpha - \alpha + 1)^2}.$$

On other hand, for the expected Fisher's information matrix I_E, observe that the I_E differs from I_0 only for the term $\sum_{i=1}^{n} 1/(\alpha + x_i)^2$ since $\mathbb{E}[\bar{x}] = n\mu$ where μ is the mean of the DS distribution. In this case, observe that,

$$\mathbb{E}\left[\frac{1}{(\alpha + X)^2}\right] = \frac{(\gamma^{-\theta}-1)^2\Phi(1/\gamma^{-\theta}, 1, \alpha)}{(\gamma^{-\theta}\alpha - \alpha + 1)\gamma^{-\theta}}, \tag{24}$$

where $\Phi(\cdot)$ is the Lerch transcendent function (Hassani, 2007; Ferreira et al., 2017). Thus, the expected Fisher's information matrix I_E is the same as the observed Fisher's information matrix I_E replacing the terms $\sum_{i=1}^{n} 1/(\alpha + x_i)^2$ by $\mathbb{E}[1/(\alpha + X)^2]$ given above and $\mathbb{E}[\bar{x}] = n\mu$. For the interval estimates, it could be used large sample approximations for the $100 \times (1-\eta)\%$ two sided confidence interval (CI), that is, $\hat{\alpha} \pm z_{\frac{\eta}{2}}\,\widehat{\text{se}}(\hat{\alpha})$ and $\hat{\theta} \pm z_{\frac{\eta}{2}}\,\widehat{\text{se}}(\hat{\theta})$, where z_η is the upper η^{th} percentile of the standard Normal distribution and the standard error (SE) is estimated as the squared root of the variance of $\hat{\alpha}$ and $\hat{\theta}$ obtained from the expected Fisher's information matrix $I_E(\hat{\alpha}, \hat{\theta})$.

3.2.2. Right-censored data

Let us consider the situation when the lifetime, X_i, is not completely observed and may be subject to right censoring. Let C_i be the censoring time for the ith individual. From a sample of size n, it is observed $X_i = \min\{X_i, C_i\}$ and $\delta_i = I(X_i < C_i)$, where $\delta_i = 1$ if X_i is a complete observed lifetime and $\delta_i = 0$ if it is a right censored lifetime. In this case, the log-likelihood function considering the DS distribution with pmf defined in Eq. (6), can be written as,

$$\begin{aligned}\ell(\alpha,\beta|\mathbf{x}) = &\sum_{i=1}^{n}\delta_i \ln(\alpha+x_i) - \frac{\theta r n\bar{x}}{\alpha} + 2rn\ln(\gamma^{-\theta}-1) - \frac{rn\theta}{\alpha} - nr\ln(\gamma^{-\theta}\alpha-\alpha+1) + \\ &\sum_{i=1}^{n}(1-\delta_i)\ln\left[\frac{(\alpha+x_i)\gamma^{\theta(x_i-1)} - (x_i+\alpha-1)\gamma^{\theta x_i}}{\gamma^{-\theta}\alpha-\alpha+1}\right],\end{aligned} \tag{25}$$

where $r = \sum_{i=1}^{n}\delta_i$ is the number of uncensored observations and $n\bar{x} = \sum_{i=1}^{n}x_i$ is the sample mean. In this case, the MLEs $\hat{\alpha}$ and $\hat{\theta}$ for the unknown parameters α and θ obtained by maximizing the log-likelihood function defined in Eq. (25), have no closed analytical form which differs from the complete data case where the MLE for θ was easily obtained. In addition, under suitable regularity conditions, the observed Fisher's information matrix I_0 is given by,

$$I_0(\hat{\alpha},\hat{\theta}) = \begin{pmatrix} -\frac{\partial^2\ell}{\partial^2\alpha^2} & -\frac{\partial^2\ell}{\partial^2\alpha\theta} \\ -\frac{\partial^2\ell}{\partial^2\theta\alpha} & -\frac{\partial^2\ell}{\partial^2\theta^2}\end{pmatrix}_{\hat{\alpha},\hat{\theta}}, \tag{26}$$

where,

$$\begin{aligned}\frac{\partial^2\ell}{\partial^2\theta^2} = &\frac{2rn\gamma^{-\theta}}{\alpha^2(\gamma^{-\theta}-1)}\left(1-\frac{\gamma^{-\theta}}{\gamma^{-\theta}-1}\right) - \frac{rn\gamma^{-\theta}}{\gamma^{-\theta}\alpha-\alpha+1}\left(\frac{1}{\alpha}-\frac{\gamma^{-\theta}}{\gamma^{-\theta}\alpha-\alpha+1}\right) + \\ &\sum_{i=1}^{n}\frac{(1-\delta_i)\gamma^{-\theta}}{\gamma^{-\theta}\alpha-\alpha+1}\left[\frac{\gamma^{-\theta}}{\gamma^{-\theta}\alpha-\alpha+1}-\frac{1}{\alpha}\right] + \sum_{i=1}^{n}\frac{(1-\delta_i)}{\alpha^2[(\alpha+x_i)\gamma^{\theta(x_i-1)}-(x_i+\alpha-1)\gamma^{\theta x_i}]} \\ &\left\{\left[(\alpha+x_i)(x_i-1)^2\gamma^{\theta(x_i-1)}-(x_i+\alpha-1)x_i^2\gamma^{\theta x_i}\right] - \frac{[(\alpha+x_i)(x_i-1)\gamma^{\theta(x_i-1)}-(x_i+\alpha-1)x_i\gamma^{\theta x_i}]^2}{\alpha^2[(\alpha+x_i)\gamma^{\theta(x_i-1)}-(x_i+\alpha-1)\gamma^{\theta x_i}]}\right\}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\ell}{\partial^2\alpha^2} = &-\sum_{i=1}^{n}\frac{\delta_i}{(\alpha+x_i)^2} - \frac{2\theta rn\bar{x}}{\alpha^3} + \frac{4rn\theta\gamma^{-\theta}}{\alpha^3(\gamma^{-\theta}-1)} + \frac{2rn\theta\gamma^{-\theta}}{\alpha^4(\gamma^{-\theta}-1)}\left(1-\frac{\gamma^{-\theta}}{\gamma^{-\theta}-1}\right) - \frac{2n\theta}{\alpha^3} - \frac{n\theta^2\gamma^{-\theta}}{\alpha^3(\gamma^{-\theta}\alpha-\alpha+1)} + \\ &\frac{n\left(-\frac{\theta\gamma^{-\theta}}{\alpha}+\gamma^{-\theta}-1\right)^2}{(\gamma^{-\theta}\alpha-\alpha+1)^2} + \sum_{i=1}^{n}\frac{(1-\delta_i)}{\gamma^{-\theta}\alpha-\alpha+1}\left[\frac{\left(\gamma^{-\theta}-\frac{\theta}{\alpha}\gamma^{-\theta}-1\right)^2}{\gamma^{-\theta}\alpha-\alpha+1} - \frac{\theta^2\gamma^{-\theta}}{\alpha^3}\right] + \\ &\sum_{i=1}^{n}\frac{(1-\delta_i)\theta[(x_i^2\theta+((\theta-2)\alpha-\theta)x_i-2\alpha)x_i\gamma^{\theta x_i}-(x_i^2\theta+((\theta-2)\alpha-\theta)x_i-\alpha\theta)(x_i-1)\gamma^{\theta(x_i-1)}]}{\alpha^4[(\alpha+x_i)\gamma^{\theta(x_i-1)}-(x_i+\alpha-1)\gamma^{\theta x_i}]} - \\ &\sum_{i=1}^{n}\frac{(1-\delta_i)\{\gamma^{\theta(x_i-1)}+\alpha^{-2}[\theta(\alpha+x_i)(x_i-1)\gamma^{\theta(x_i-1)}]-\gamma^{\theta x_i}-\alpha^{-2}[\theta(\alpha+x_i-1)x_i\gamma^{\theta x_i}]\}^2}{[(\alpha+x_i)\gamma^{\theta(x_i-1)}-(x_i+\alpha-1)\gamma^{\theta x_i}]^2},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\ell}{\partial^2\alpha\theta} = \frac{\partial^2\ell}{\partial^2\theta\alpha} = &\frac{rn\bar{x}}{\alpha^2} + \frac{2rn\gamma^{-\theta}}{\alpha^2(\gamma^{-\theta}-1)}\left(\frac{\theta\gamma^{-\theta}}{\alpha(\gamma^{-\theta}-1)}-\frac{\theta}{\alpha}-1\right) + \frac{rn}{\alpha^2} + \frac{rn\theta\gamma^{-\theta}}{\alpha^2\left(\gamma^{-\theta}\alpha-\alpha+1\right)} + \\ &\frac{rn\left(-\frac{\theta\gamma^{-\theta}}{\alpha}+\gamma^{-\theta}-1\right)\gamma^{-\theta}}{\left(\gamma^{-\theta}\alpha-\alpha+1\right)^2} + \sum_{i=1}^{n}\frac{(1-\delta_i)\gamma^{-\theta}}{\gamma^{-\theta}\alpha-\alpha+1}\left[\frac{\left(\gamma^{-\theta}-\frac{\theta}{\alpha}\gamma^{-\theta}-1\right)}{\gamma^{-\theta}\alpha-\alpha+1} + \frac{\theta}{\alpha^2}\right] + \\ &\sum_{i=1}^{n}\frac{(1-\delta_i)\theta\left[(x_i^2\theta+((\theta-1)\alpha-\theta)x_i+\alpha)x_i\gamma^{\theta x_i}-(x_i^2\theta+((\theta-1)\alpha-\theta)x_i-\alpha\theta)(x_i-1)\gamma^{\theta(x_i-1)}\right]}{\alpha^3\left[(\alpha+x_i)\gamma^{\theta(x_i-1)}-(x_i+\alpha-1)\gamma^{\theta x_i}\right]} -\end{aligned}$$

$$\sum_{i=1}^{n} \frac{(1-\delta_i)\left\{\left[(x_i^2\theta+(\alpha-1)\theta x_i+\alpha^2\right]\gamma^{\theta x_i}-\left[x_i^2\theta+(\alpha-1)\theta x_i+\alpha(\alpha-\theta)\right]\gamma^{\theta(x_i-1)}\right\}}{\alpha^3\left[(\alpha+x_i)\gamma^{\theta(x_i-1)}-(x_i+\alpha-1)\gamma^{\theta x_i}\right]^2\left\{(x_i+\alpha-1)x_i\gamma^{\theta x_i}-(x_i-1)(\alpha+x_i)\gamma^{\theta(x_i-1)}\right\}^{-1}}.$$

Different from complete data, for censored data, the expected Fisher's information matrix I_E cannot be obtained in an analytical way. Then, in this case, the asymptotic distribution of the maximum likelihood estimator $(\hat{\alpha}, \hat{\theta})$ is a multivariate Normal distribution with mean (α, θ) and covariance matrix $\boldsymbol{\Sigma}(\hat{\boldsymbol{\Omega}}), \hat{\boldsymbol{\Omega}} = (\hat{\alpha}, \hat{\theta})$, which can be consistently estimated by the inverse of the observed Fisher's information matrix I_0.

3.3. Bayesian method

The Bayesian paradigm is based on specifying a probability distribution for the observed data D given a vector of unknown parameters $\boldsymbol{\eta}$ and it provides a method for updating the new information using the Bayes' rule given the prior distribution specifying the uncertainty about the parameter (see Ibrahim et al., 2005).

In order to determine the Bayes estimators for the unknown parameters of the DS model based on the squared error loss function, $L(\eta, a) = (\eta - a)^2$, suppose that the parameters α and θ have non-informative independent Uniform(0, k) prior distributions where k is a positive integer such that $\pi(\alpha), \pi(\theta) \propto 1$. The Bayes estimate of any function of (α, θ), say $\omega(\alpha, \theta)$, assuming the squared error loss function is given by,

$$\hat{\mu}_B = \frac{\int_0^k \int_0^k \omega(\alpha,\theta) L(\alpha,\theta)\pi(\alpha)\pi(\theta) d\alpha d\theta}{\int_0^k \int_0^k L(\alpha,\theta)\pi(\alpha)\pi(\theta) d\alpha d\theta} \tag{27}$$

Since the kernel of the likelihood function is proportional to a gamma distribution, the Bayes estimate of $\omega(\alpha, \theta)$ could also be obtained assuming independent gamma prior distributions given respectively, by,

$$\begin{aligned} \pi_1(\alpha) &= \frac{\beta^\lambda}{\Gamma(\lambda)}\alpha^{(\lambda-1)}\exp\{-\beta\alpha\}, \\ \pi_2(\theta) &= \frac{\beta^\lambda}{\Gamma(\lambda)}\theta^{(\lambda-1)}\exp\{-\beta\theta\}. \end{aligned} \tag{28}$$

Observe that since it is not possible to compute Eq. (27) analytically, it is used MCMC methods to get the posterior summaries of interest. In this way, without loss of generality, it is used the Gibbs sampling algorithm to generate samples from the posterior distribution of interest from which it is computed the Monte Carlo Bayes estimators under the squared error loss function. The Gibbs sampling algorithm steps are given by,

- Step 1: Choose initial values, $\alpha^{(0)}$ and $\theta^{(0)}$ for α and θ. Denote the values of α and θ at the i^{th} step by $\alpha^{(i)}, \theta^{(i)}$.
- Step 2: Generate $\alpha^{(i)}$, $\theta^{(i+1)}$ from the conditional posterior distributions needed for the Gibbs sampling algorithm obtained directly from the joint posterior distribution.
- Step 3: Repeat step 2, N times.
- Step 4: Calculate the Monte Carlo Bayes estimate of $\omega(\alpha, \theta)$ using the expression given by $(1/(N - B))$ $\sum_{i=B+1}^{N} \omega(\alpha^{(i)}, \theta^{(i)})$ where $B =$ 5,000 is the burn-in period.

The posterior summaries of interest are computed using the package **R2jags** (Su & Yajima, 2015) from R software (R Core Team, 2018) considering a "burn-in sample" of size 5,000 to eliminate the effect of the initial values and a final Gibbs sample of size 2,000 taking every 100th sample from 200,000 simulated Gibbs samples. Furthermore, the convergence of the Gibbs Sampling algorithm was monitored using standard graphical methods, as the trace plots of the simulated samples.

4. Monte Carlo simulation study

This section reports the results of a simulation study carried out to assess the performance of the MLEs, MOM and Bayesian estimators for the DS distribution assuming complete and censored data. The simulation study was performed in R software using the packages *fitdistrplus* for MLEs, *nleqslv* for MOM estimators and *R2jags* for Bayesian estimators. For the Bayesian estimators, independent approximately non-informative gamma prior distributions, Gamma(0.001, 0.001), were assumed for both parameters. The *BFGS* optimization method was considered as the *optim.method*. To simulate observations from DS model, the inverse transformation method for discrete case was used following the steps (Devroye, 1987):

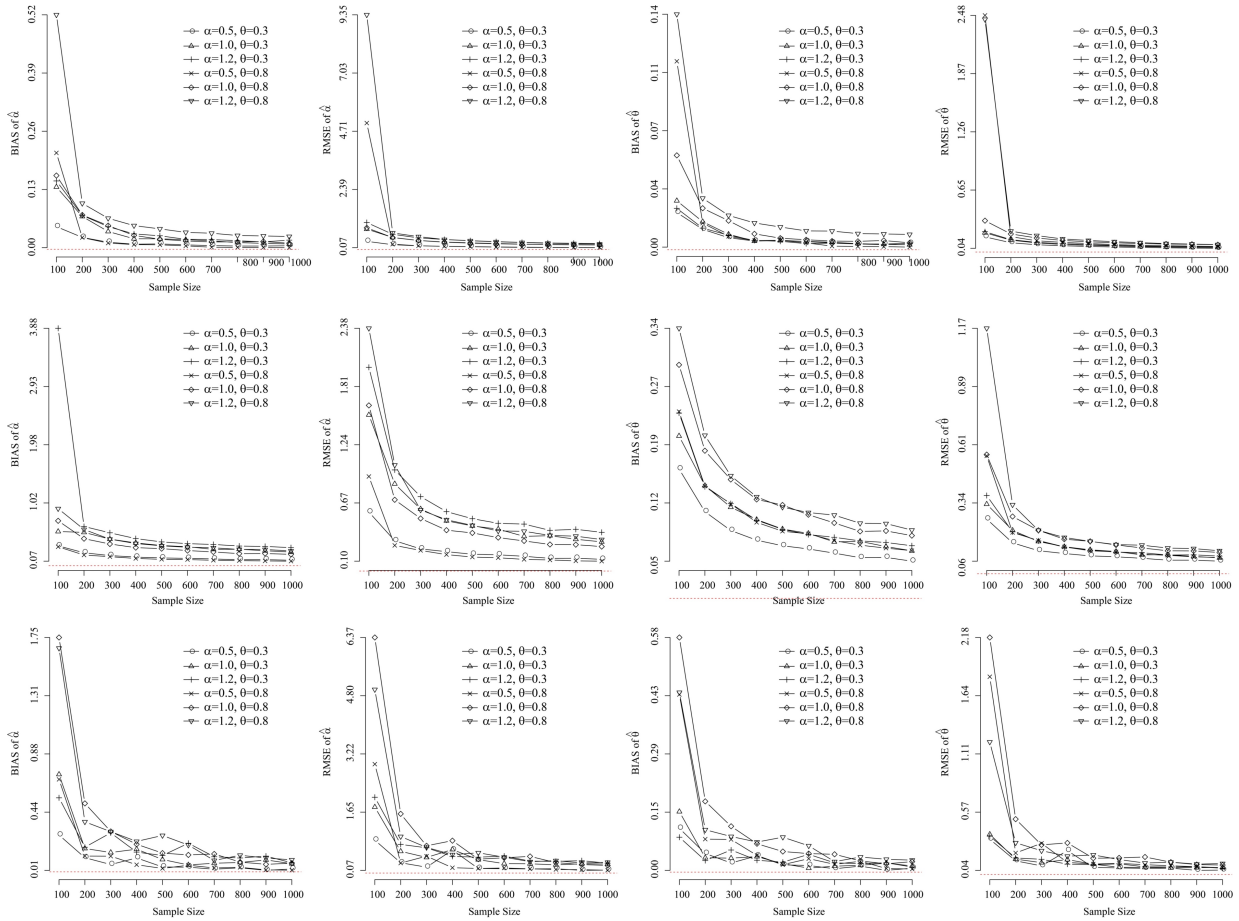

Fig. 2. The biases and RMSEs for both parameters assuming the DS distribution for each considered scenario considering the MLE (upper-panels), MOM (middle-panels) and Bayesian (lower-panels) estimators assuming complete data.

- Step 1: Generate $U \sim \textit{Uniform}(0, 1)$.
- Step 2: Define X by $F(X-1) = \sum_{i<X} p_i < U \leqslant \sum_{i \leqslant X} p_i = F(X)$ where $P(X = i) = p_i$. Set $X = 0$ and $S = p_0$.
- Step 3: While $U > S$, do $X = X + 1$ and $S = S + p_X$.
- Step 4: Return X.

The simulation study was performed under six scenarios considering the assumed parameter values as the combination of ($\alpha = 0.5, 1.0, 1.2$, $\theta = 0.3, 0.8$) for better computational stability. It was also considered the sample sizes $n = 100, \ldots, 1000$, each one involving 10,000 Monte Carlo replications. For each scenario, the biases and the RMSE for the estimated parameter component of the vector of parameters (α, θ) were computed using the expressions:

$$\textit{BIAS}(\hat{\Psi}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\Psi}_i - \Psi), \quad \textit{RMSE}(\hat{\Psi}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\Psi}_i - \Psi)^2}$$

where $N = 10{,}000$ is the number of simulations and Ψ denotes each parameter α or θ.

4.1. Simulation results

4.2. Complete data

The obtained simulation results for each scenario assuming complete data are illustrated in Fig. 2 for MLE, MOM and Bayesian, respectively. From the simulation results illustrated in Fig. 2, it is possible to conclude that,

i.) Assuming the MLE estimators, both parameters are asymptotically non-biased since $E(\theta) \approx \theta$ and $E(\alpha) \approx \alpha$ when $n \to \infty$. Moreover, the RMSE values also tend to zero when $n \to \infty$. In addition, the biases values do not exceed the value 0.55 for both parameters and the RMSEs values do not exceed the value 0.90 for both parameters, except for the fourth and the sixth scenarios assuming sample size equals to $n = 100$.

ii.) Assuming the MOM estimators, both parameters are asymptotically non-biased since $E(\theta) \approx \theta$ and $E(\alpha) \approx \alpha$ when $n \to \infty$. Moreover, the RMSE values also tend to zero when $n \to \infty$. In addition, the biases values do not exceed the value 0.65 for the parameter α and the value 0.25 for the parameter θ; and the RMSEs values do not exceed the value 1.10 for α and the value 0.35 for θ, for the third and the sixth scenarios assuming sample size equals to $n = 100$. In general, MOM estimators are quite similar to the MLE estimators, however, the MLE estimators are more computational stable than the MOM estimators.

iii.) Assuming the Bayesian estimators, both parameters are asymptotically non-biased since $E(\theta) \approx \theta$ and $E(\alpha) \approx \alpha$ when $n \to \infty$. Moreover, the RMSE values also tend to zero when $n \to \infty$. In addition, the biases values do not exceed the value 0.50 for the parameter α and the value 0.20 for the parameter θ; and the RMSEs values do not exceed the value 1.60 for α and the value 0.50 for θ, except for the sample size equals to $n = 100$. The results could be improved using another prior distributions (as Uniform, for example), but, in general, the Bayesian estimators are more suitable than the MLE and MOM in medical applications, for example.

iv.) The convergence of each estimation method was quite good even thought, for all considered samples, the parameters had been overestimated (that is, the parameters are positively biased). This result is expected since the Lindley distribution has the same property. In addition, based on these simulation results, it could be concluded that the DS distribution could be used as an alternative to other existing discrete univariate distributions to describe univariate lifetimes with good accuracy and computational aspects.

4.3. Right-censored data

In this section, it is presented the simulation results assuming censored data. To perform the simulation study, it was considered the same data generated assuming complete data case and considering that the censored observations are determined by assuming a cut point equals to 10 to obtain simple right-censored observations. The simulation study was performed under four scenarios considering the assumed parameter values as the combination of ($\alpha =$ 0.2, 0.7, $\theta =$ 0.2, 0.5) for better computational stability. Only the Bayesian estimators were considered in this case for the simulation study since they are more suitable than the MLE and MOM as stated previously. Independent approximately non-informative uniform prior distributions, Uniform(0, 1), were assumed for both parameters.

The obtained simulation results for each scenario assuming simple right-censored data are illustrated in Fig. 3. From the results presented in Fig. 3, it is possible to conclude that both parameters are asymptotically non-biased since $E(\theta) \approx \theta$ and $E(\alpha) \approx \alpha$ when $n \to \infty$. Moreover, the RMSE values also tend to zero when $n \to \infty$. In addition, the biases values do not exceed the value 0.15 for both parameters α and θ; and the RMSEs values do not exceed the value 0.25 for both parameters α and θ.

4.4. Numerical experiments

4.4.1. Complete simulated dataset

For illustration purposes, let us consider $n = 100$ univariate lifetimes generated from the DS distribution (dataset in Table 1) with arbitrary parameters values $\alpha = 0.7$ and $\theta = 0.3$ not considering the presence of right-censoring. For the statistical analysis, it is considered only MLE and Bayesian estimators. For the Bayesian approach, it is assumed approximately non-informative uniform prior distributions with hyperparameter values equal to (0, 1). The inference summaries of interest are presented in Table 2.

From the results of Table 2 and comparing to the true parameter values adopted in this simulated dataset, it is possible to conclude that the Monte Carlo Bayesian estimates are more accurate when compared to the MLE estimates by observing the standard errors and deviations as well the length of the confidence and credibility intervals. Despite the differences between both estimators, in general, the Bayesian estimators are more suitable in applications for the DS distribution.

Table 1
Simulated lifetimes from a DS distribution with true arbitrary parameters values $\alpha = 0.7$ and $\theta = 0.3$

X	1	9	3	4	1	4	0	1	5	1	2	2	2	2	3	3	2	0	3	2
	4	4	7	5	6	8	8	4	3	1	3	1	1	3	7	2	15	7	8	4
	0	2	11	8	4	5	5	3	7	0	0	3	2	3	11	4	2	4	9	9
	2	4	1	1	2	2	0	8	12	5	5	7	9	1	2	3	2	0	5	3
	1	3	2	3	4	3	2	3	5	0	0	1	10	8	1	8	1	2	0	3

Table 2
Inference summaries for DS distribution with true arbitrary parameters values $\alpha = 0.7$ and $\theta = 0.3$

Maximum likelihood				Bayesian Approach			
Param.	Estimate	Std. error	95% Conf. Int.	Param.	Post. Mean	Std. Dev.	95% Cred. Int.
α	0.6497	0.3265	(0.0098, 1.2896)	α	0.7091	0.1844	(0.3343, 0.9872)
θ	0.2927	0.1308	(0.0363, 0.5491)	θ	0.3158	0.0765	(0.1601, 0.4439)

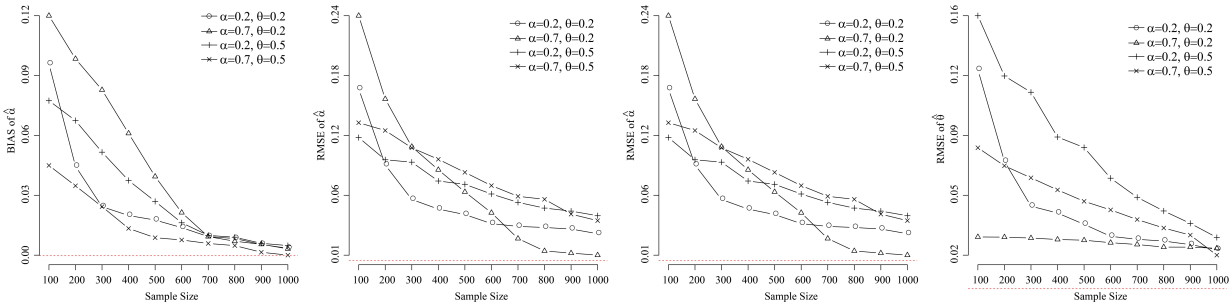

Fig. 3. The biases and RMSEs for both parameters assuming the DS distribution for each considered scenario considering the Bayesian estimators assuming simple right-censored data with cut point equals to 10.

4.4.2. Right-censored simulated data

As a second illustration, it was considered the same data generated for the numerical experiment assuming complete data from a DS distribution where the censored observations are determined by considering a cut point equals to 10 to obtain simple right-censored observations. Since the Bayesian estimators are more suitable for the DS distribution, in this illustration, they are only considered as the estimation approach assuming approximately non-informative uniform prior distributions for the parameters of the model with hyperparameter values equal to (0, 1). In Table 3, it is presented the posterior summaries of interest for the parameters of the DS distribution.

From the results of Table 3, it is observed that the lengths of the 95% credible intervals are relatively narrow and the standard deviations are estimated by small values, an indication that the DS distribution has a good performance for this right-censored simulated dataset under a Bayesian approach and could be used as an alternative to lifetime models.

5. Real data applications

5.1. Lung cancer data

To illustrate the usefulness of the proposed distribution, it is presented in this subsection, the analysis of a real medical dataset related to lung cancer assuming the DS distribution. The dataset is given by Ding et al. (2017) and corresponds to the lifetimes of Chinese patients pathologically confirmed lung cancer who received EGFR, KRAS, and BARF mutation tests at the Thoracic Cancer Institute, Tongji University from January 2012 to April 2016. The dataset consists of $n = 28$ patients with 0 censored observations for the overall survival times and 4 censored observations for the progression-free survival times.

For the statistical analysis, it is assumed two lifetimes: *the overall survival times (calculated from the date of lung cancer diagnosis to death from any reason or censored at the last follow-up date)*, and *the progression-free survival times (the times from the treatment start time until the date of systemic progression or death)*. It is important to point

Table 3
Posterior summaries for DS distribution with true arbitrary parameters values $\alpha = 0.7$ and $\theta = 0.3$ in presence of right-censoring

Param.	Post. Mean	Std. Dev.	95% Cred. Int.
α	0.7406	0.1742	(0.3657, 0.9892)
θ	0.3203	0.0706	(0.1706, 0.4355)

Table 4
Posterior summaries for the parameters and the mean (μ) of DS, Geo, DL_1 and DL_2 distributions for both lifetimes considered for the lung cancer data

	Overall survival				Progression-free survival			
Model	Par.	Mean (S.D.)	95% Cred. Int.	DIC	Par.	Mean (S.D.)	95% Cred. Int.	DIC
DS	α	0.6164 (0.2462)	(0.1257, 0.9838)	176.9	α	0.4588 (0.2494)	(0.0669, 0.9476)	146.3
	θ	0.0696 (0.0290)	(0.0141, 0.1228)		θ	0.1475 (0.0766)	(0.0228, 0.3006)	
	μ	17.3620 (2.5596)	(13.0547, 23.0751)		μ	5.7026 (0.8380)	(4.2966, 7.5695)	
Geo	α	0.9466 (0.0102)	(0.9248, 0.9646)	191.1	α	0.8475 (0.2060)	(0.7933, 0.8927)	160.2
	μ	19.7387 (3.9175)	(13.6960, 28.5863)		μ	6.7250 (1.1573)	(4.8965, 9.4240)	
DL_1	θ	0.1073 (0.0149)	(0.0814, 0.1384)	178.8	θ	0.3053 (0.0417)	(0.2256, 0.3912)	148.4
	μ	17.5071 (2.7369)	(13.1816, 23.8878)		μ	5.6960 (0.8999)	(4.1704, 7.6417)	
DL_2	θ	0.1108 (0.0146)	(0.0839, 0.1396)	177.6	θ	0.2933 (0.0395)	(0.2230, 0.3810)	150.1
	μ	17.6525 (2.6285)	(13.1717, 23.0719)		μ	5.6895 (0.9303)	(4.1579, 7.7762)	

out that both times are measured in complete months (discrete data) and the possible dependence structure between both times is not considered here.

The parameters of the DS distribution were estimated under a Bayesian approach assuming approximately non-informative independent Uniform(0, 1) prior distributions for both parameters. The corresponding results are presented in Table 4 and the fit of DS distribution was compared to the fit of the Geometric (Geo) distribution and two discrete Lindley (DL_1 introduced by Bakouch et al., 2014; and DL_2 introduced by Oliveira et al., 2017) distributions (for those distributions, uniform prior distributions for the parameters were also considered). Using the Deviance Information Criteria (DIC) (Spiegelhalter et al., 2014) to discriminate the considered distribution, it could be concluded that the DS is the best model fitted for the dataset.

A measure that plays an important role in survival analysis is the comparison of the estimated lifetime mean assuming a parametric distribution with a nonparametric estimator for the mean. In this case, the non-parametric estimators for the means obtained from the Kaplan and Meier (1958) non-parametric estimators for the survival functions are given, respectively, by 16.57 months for the overall survival time and 5.71 months for the progression-free survival time. The estimated mean and the 95% credible intervals for each assumed model are presented in Table 4 from which it could be concluded that the estimated mean for DS distribution is very close to the empirical Kaplan-Meier estimator which is a great indication of a good fit and adequacy of the DS distribution to describe both survival times associated to the lung cancer data.

5.2. Alberta fires data

Now, considering count data, in this subsection, it is presented the analysis of a real dataset related to number of fires in Alberta assuming the DS distribution. The dataset is given by Tremblay et al. (2018) and corresponds of the number of fires recorded over a seven year period (1996–2002) within a 67,000 km^2 study region of boreal forest in northeastern of Alberta, Canada. Fire records were selected from the Alberta government's Historical Wildfire Database available in the website http://wildfire.alberta.ca/resources/historical-data/historical-wildfire-database.aspx.

Forest fires are important events in many terrestrial ecosystems, including the boreal forests of North America. In many areas where they occur, fire management agencies attempt to control the growth and limit the size of these fires, to protect human lives, infrastructure, and natural resources (Tremblay et al., 2018). Thus, knowing the probability distribution of the number of fire is important for forecasting the number of fires that affects one area to draw a way to extinguish the fires.

For the statistical analysis considered here, it is assumed the response variable *number of fires* and the DS distribution do describe the response distribution. A Bayesian approach was assumed considering approximately non-informative independent Uniform(0, 10) prior distributions for both parameters of the DS distribution. The corresponding results are presented in Table 5 and the fit of the DS distribution was compared to the fit of the Geometric

Table 5
Posterior summaries for the parameters and the mean (μ) of DS, Geo, DL_1, DL_2, and P distributions for number of fires in northeastern of Alberta, Canada

Model	Param.	Post. Mean	Std. Dev.	95% Cred. Int.	DIC
DS	α	6.0791	1.2213	(3.9385, 8.3157)	6298.7
	θ	0.1416	0.0082	(0.1183, 0.1526)	
	μ	12.4620	0.3770	(11.7278, 13.2178)	
Geo	α	0.9255	0.0024	(0.9211, 0.9303)	6332.3
	μ	13.4457	0.4198	(12.6292, 14.2928)	
DL_1	θ	0.1449	0.0034	(0.1382, 0.1516)	6319.8
	μ	12.4392	0.3246	(11.7917, 13.0720)	
DL_2	θ	0.1489	0.0036	(0.1419, 0.1559)	6332.4
	μ	12.4666	0.3208	(11.8529, 13.1133)	
P	θ	12.4411	0.1213	(12.2142, 12.6881)	11222.7
	μ	12.4411	0.1213	(12.2142, 12.6881)	

(Geo) distribution, the two discrete Lindley (DL_1 and DL_2, see the cited authors in Application 1 for more details) and Poisson (P) (for those models, uniform prior distributions for the parameters were also considered). Using DIC criteria to discriminate the considered distribution, it could be concluded that the DS is the best model fitted by the dataset. In addition, the empirical mean of the data is 12.44 fires and the estimated mean as well the 95% credible intervals for each assumed model are presented in Table 5 from which it could be concluded that the estimated mean for DS distribution is very close to the empirical mean which is a great indication of a good fit and adequacy of the DS distribution to describe the distribution of number of fires in northeastern of Alberta, Canada.

6. Concluding remarks

In this study, it was introduced a new univariate discrete distribution, named discrete Sushila (DS) distribution, obtained using the Good (1953) discretization method to generate a discrete analogue from the Sushila distribution proposed by Shanker et al. (2013) as an alternative to existing univariate discrete distributions like the popular Poisson, the geometric and the discrete Lindley distribution to analyze univariate discrete lifetime data in presence of right-censored data. The main mathematical properties of this new distribution were also discussed in this study from where it could be stated that the DS distribution can be used to model data with over/under/equi-dispersion. Moreover, an extensive simulation study was performed to verify the effectiveness of the maximum likelihood method, moments estimator and Bayesian method assuming different fixed values for the parameters of the model and different sample sizes. The results obtained from Monte Carlo studies showed that the parameters of the DS distribution are asymptotically non-biased and the biases as well the RMSEs tends to zero when the sample size increases for complete and simple right-censored data.

In the application with real data presented in this study, it was observed that, with the use of the DS distribution, it is possible to obtain in a simple way the inferences of interest for the dataset in presence or not of right-censored data with small computational costs and a good accuracy even using non-informative priors for the parameters of the DS model, under a Bayesian approach. As pointed out in the applications, the estimated mean using DS distribution is basically equal to the empirical mean of the data considered which is a great indication that the DS distribution is adequate and describes well the distribution of the data. In general, most of the existing models could be fitted to the dataset, however, only few of these models could describe the mean in a adequate way which, in some cases, is the main interest of the researcher. These results could be of great interest for the search of appropriate univariate lifetime distributions for the analysis of right-censored especially in medical and engineering studies.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

Bakouch, H.S., Jazi, M.A., & Nadarajah, S. (2014). A new discrete distribution. *Statistics*, *48*(1), 200-240.

Bi, Z., Faloutsos, C., & Korn, F. (2001). The DGX distribution for mining massive, skewed data. in: *Proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 17-26. ACM.

Borah, M., & Saikia, K.R. (2016). Certain properties of discrete Sushila. *Statistics*, *5*(6).

Chakraborty, S. (2015). Generating discrete analogues of continuous probability distributions: A survey of methods and constructions. *Journal of Statistical Distributions and Applications*, *2*(1), 1-30.

Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman and Hall, New York, 2nd edition.

Devroye, L. (1987). *Non-Uniform Random Variate Generation*. Springer-Verlag.

Ding, X., Zhang, Z., Jiang, T., Li, X., Zhao, C., Su, B., & Zhou, C. (2017). Clinicopathologic characteristics and outcomes of Chinese patients with non-small-cell lung cancer and BRAF mutation. *Cancer Medicine*, *6*(3), 555-562.

Doray, L.G., & Luong, A. (1997). Efficient estimators for the Good family. *Communications in Statistics – Simulation and Computation*, *26*(3), 1075-1088.

Ferreira, E., Kohara, A., & Sesma, J. (2017). New properties of the Lerch's transcendent. *Journal of Number Theory*, *172*, 21-31.

Ghitany, M.E., Atieh, B., & Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and Computers in Simulation*, *78*(4), 493-506.

Good, I.J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, *40*(3-4), 237-264.

Grandell, J. (1997). *Mixed Poisson Processes*, *77*. CRC Press.

Haight, F.A. (1957). Queueing with balking. *Biometrika*, *44*(3/4), 360-369.

Hamada, M.S., Wilson, A.G., Reese, C.S., & Martz, H.F. (2008). *Bayesian reliability*. Springer Series in Statistics. Springer, New York.

Hassani, M. (2007). Approximation of the dilogarithm function. *J Inequalities in Pure and Applied Mathematics*, *8*, 1-7.

Ibrahim, J.G., Chen, M.-H., & Sinha, D. (2005). *Bayesian survival analysis*. Springer Science and Business Media.

Inusah, S., & Kozubowski, T.J. (2006). A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference*, *136*(3), 1090-1102.

Jodrá, P. (2010). Computer generation of random variables with Lindley or Poisson-Lindley distribution via the Lambert W function. *Mathematics and Computers in Simulation*, *81*(4), 851-859.

Kalbfleisch, J.D., & Prentice, R.L. (2002). *The statistical analysis of failure time data*. Wiley, New York, 2nd edition.

Kaplan, E.L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457-481.

Keilson, J., & Gerber, H. (1971). Some results for discrete unimodality. *Journal of the American Statistical Association*, *66*(334), 386-389.

Kemp, A.W. (1997). Characterizations of a discrete Normal distribution. *Journal of Statistical Planning and Inference*, *63*(2), 223-229.

Kemp, A.W. (2008). The discrete Half-Normal distribution. in: *Advances in Mathematical and Statistical Modeling*. 353-360. Birkhäuser Boston, Boston.

Klein, J.P., & Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag, New York.

Kozubowski, T.J., & Inusah, S. (2006). A skew Laplace distribution on integers. *Annals of the Institute of Statistical Mathematics*, *58*(3), 555-571.

Lambert, J.H. (1758). Observationes variae in mathesin puram. *Acta Helvetica*, *3*(1), 128-168.

Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, Hoboken, NJ, 2nd edition.

Lee, E.T., & Wang, J.W. (2003). *Statistical Methods for Survival Data Analysis*. John Wiley and Sons, Hoboken, NJ, 3rd edition.

Lehmann, E., & Casella, G. (1998). *Theory of Point Estimation*. New York.

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

Meeker, W.Q., & Escobar, L.A. (1998). *Statistical Methods for Reliability Data*. John Wiley and Sons, New York.

Miller, S.J. (2008). An identity for sums of polylogarithm functions. *Integers: Electronic Journal Of Combinatorial Number Theory*, *8*, A15.

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

Nekoukhou, V., Alamatsaz, M.H., & Bidram, H. (2012). A discrete analog of the Generalized Exponential distribution. *Communication in Statistics – Theory and Methods*, *41*(11), 2000-2013.

Nekoukhou, V., Alamatsaz, M.H., & Bidram, H. (2013). Discrete generalized Exponential distribution of a second type. *Statistics – A Journal of Theoretical and Applied Statistics*, *47*(4), 876-887.

Oliveira, R.P., Mazucheli, J., & Achcar, J.A. (2017). A comparative study between two discrete Lindley distributions. *Ciência e Natura*, *39*(3), 539-552.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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

Shanker, R., Sharma, S., Shanker, U., & Shanker, R. (2013). Sushila distribution and its application to waiting times data. *International Journal of Business Management*, *3*(2), 1-11.

Siromoney, G. (1964). The general Dirichlet's Series distribution. *Journal of the Indian Statistical Association*, *2-3*(2), 1-7.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(3), 485-493.

Su, Y.-S., & Yajima, M. (2015). *R2jags: Using R to Run 'JAGS'. R package version 0.* 5-7.

Tremblay, P.-O., Duchesne, T., & Cumming, S.G. (2018). Survival analysis and classification methods for forest fire size. *PloS One*, *13*(1), e0189860.