Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n1p50

Bayesian and maximum likelihood inference approaches for the discrete generalized Sibuya distribution with censored data
By Freitas et al.

Published: 20 May 2022

Electronic Journal of Applied Statistical Analysis
Vol. 01, Issue 15, May 2022, 50-74
DOI: 10.1285/i20705948v15n1p50

Bayesian and maximum likelihood inference approaches for the discrete generalized Sibuya distribution with censored data

Bruno Caparroz Lopes de Freitas[a], Marcos Vinicius de Oliveira Peres[b], Jorge Alberto Achcar[b], and Edson Zangiacomi Martinez *[b]

[a] *State University of Maringá, Master Program in Biostatistics, Maringá, Brazil*
[b] *State University of Paraná, Department of Mathematics, Brazil*
[b] *Ribeirão Preto Medical School, University of São Paulo (USP), Ribeirão Preto, Brazil*

Published: 20 May 2022

This paper presents inferences under classical (maximum likelihood, ML) and Bayesian approaches for the parameters of the generalized Sibuya (GS) probability distribution considering complete and right censored lifetime data. Under a Bayesian approach, the joint posterior probability distributions of interest are estimated using Markov Chain Monte Carlo (MCMC) simulation methods. A comprehensive simulation study is carried out to assess the performance of the estimation procedure. The usefulness of the GS model is also assessed with applications to two real data sets. Despite its merits, one limitation of the generalized Sibuya distribution is that it does not present great flexibility of fit of the hazard function as compared to other existing lifetime models.

keywords: Survival analysis, censored data, Sibuya distribution, maximum likelihood estimation, Bayesian inference.

1 Introduction

The introduction of different parametric models for discrete time-to-event data has increased significantly in recent years, as observed in the literature, such as in educational

*Corresponding author: edson@fmrp.usp.br

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

research (Singer and Willett, 1993), sociology (Allison, 1982), clinical psychology (Herzog et al., 1997), medical studies (Scheike and Jensen, 1997), human fertility research (Maity et al., 2014; Vanegas et al., 2017), and addictive behavior research (Northrup et al., 2015), among many others. Under this formulation, a random variable denoted by T representing the time until the occurrence of an event of interest is considered as a discrete variable. In this way, the probability function and the cumulative distribution function (CDF) of T are respectively defined as $f(t) = P(T = t)$ and

$$F(t) = P(T \leq t) = \sum_{i=1}^{t} f(i), \ \forall t \in \mathbb{N}^*.$$

The survival function defined as the probability that an individual will survive at least until time t, is given by

$$S(t) = 1 - F(t) = P(T > t) = 1 - \sum_{i=1}^{t} f(i), \ \forall t \in \mathbb{N}^*,$$

(Tutz and Schmid, 2016) and the corresponding hazard function is given by

$$h(t) = P(T = t \,|T \geq t\,) = \frac{P(T = t)}{P(T \geq t)} = \frac{f(t)}{S(t-1)}. \tag{1}$$

Among many discrete probability distributions introduced in the literature to analyze lifetime data, we can mention the Yule-Simon distribution (Yule, 1924; Simon, 1955; Gallardo et al., 2017), the discrete Weibull distribution (Nakagawa and Osaki, 1975), the exponentiated discrete Weibull distribution (Nekoukhou and Bidram, 2015; Cardial et al., 2020; Freitas et al., 2021), the Weibull-Conway-Maxwell-Poisson distribution (Gupta and Huang, 2017), the discrete Ramos-Louzada distribution (Eldeeb et al., 2022), and the discrete Weibull geometric distribution (Jayakumar and Babu, 2018).

The standard Sibuya probability distribution, denoted by $Sibuya(\alpha)$, has probability function given by

$$f(t) = P(T = t) = \binom{\alpha}{t}(-1)^{t+1}, \quad t \in \mathbb{N}, \tag{2}$$

where $\alpha \in (0, 1)$ (Sibuya, 1979; Devroye, 1993; Letac, 2019). Kozubowski and Podgórski (2018) defined the standard Sibuya distribution as "*the distribution of the waiting time for the first success in Bernoulli trials, where the probabilities of success are inversely proportional to the number of trials*". Many generalizations and extensions of the Sibuya distribution have been proposed in literature (see for example, Christoph and Schreiber (2000), Bouzar (2008) and Kozubowski and Podgórski (2018)).

In the present article, we consider the use of the generalized Sibuya distribution introduced by Kozubowski and Podgórski (2018) in the analysis of time-to-event data, including the presence of right-censored data. The rest of the paper is organized as follows. The generalized Sibuya distribution along with the shapes of its survival and hazard functions are given in Section 2. The maximum likelihood and Bayesian approaches are considered to obtain the estimators. In Section 3, we conduct a simulation study to compare between the presented estimation technique and to assess the performance of the maximum likelihood estimators. Also in section 3, the potentiality of the model is studied by means of two real datasets. Concluding remarks are given in Section 4. Computational details are given in the appendices at the end of the article.

2 Methods

The generalized Sibuya (GS) distribution for a discrete random variable T, introduced by Kozubowski and Podgórski (2018) has probability function given by

$$f(t) = P(T = t) = \frac{\alpha}{\beta + t}\prod_{j=1}^{t-1}\left(1 - \frac{\alpha}{\beta + j}\right) = \frac{\alpha}{\beta + t}\frac{(\beta + 1 - \alpha)_{t-1}}{(\beta + 1)_{t-1}}, \tag{3}$$

where $\beta \geq 0$ and $0 < \alpha < \beta + 1$. In this expression, $(x)_t$ is the Pochhammer's symbol given by

$$(x)_t = \frac{\Gamma(x + t)}{\Gamma(x)} = x(x + 1)(x + 2)...(x + t - 1),$$

where t is a non-negative integer and $\Gamma(x)$ denotes the gamma function (Abramowitz and Stegun, 1965). If $t = 0$, we have $(x)_0 = 1$. A random variable that follows a generalized Sibuya distribution will be denoted by $GS(\alpha, \beta)$. The mean of a generalized Sibuya probability distribution is given by

$$E(T) = \frac{\beta}{\alpha - 1}, \quad \beta > 0, \quad 1 < \alpha < \beta + 1, \tag{4}$$

and the variance is

$$Var(T) = \frac{\alpha\beta\,(\beta + 1 - \alpha)}{(\alpha - 1)^2\,(\alpha - 2)}, \quad \beta > 1, \quad 2 < \alpha < \beta + 1.$$

The survival function for this model is given by

$$S(t) = P(T > t) = \prod_{j=1}^{t}\left(1 - \frac{\alpha}{\beta + j}\right) = \frac{(\beta + 1 - \alpha)_t}{(\beta + 1)_t}$$

and, from (1), the corresponding hazard function is

$$h(t) = \frac{\alpha}{\beta + t}.$$

If $\beta = 0$ and $\alpha \in (0, 1)$, we have the standard Sibuya distribution denoted as $Sibuya(\alpha)$ with probability function given by

$$f(t) = \frac{\alpha}{t!}(1 - \alpha)_{t-1}. \tag{5}$$

Note that the expressions (2) and (5) are similar to each other.

Kozubowski and Podgórski (2018) noted that if $\alpha = 1$, we have a probability function given by

$$f(t) = \frac{1}{\beta + t} \frac{(\beta)_{t-1}}{(\beta+1)_{t-1}}.$$

In this case, and considering that $(w)_t = w(w+1)_{t-1}$ and $(w)_t/(w)_{t-1} = w + t - 1$, we obtain

$$f(t) = \frac{1}{\beta+t}\left(\frac{\beta}{\beta+t-1}\right) = \frac{\beta}{\beta+t-1} - \frac{\beta}{\beta+t} = \frac{1}{1+\frac{t-1}{\beta}} - \frac{1}{1+\frac{t}{\beta}},$$

which corresponds to a discrete Pareto distribution with a tail parameter equal to 1 and scale parameter $\beta \geq 0$ (Buddana and Kozubowski, 1965). This defines a heavy-tailed distribution with non-finite mean.

Figure 1 contains plots of the survival and hazard functions of the generalized Sibuya distribution for some arbitrary values of α and β. We can note that the hazard function $h(t)$ is always a decreasing function in t.

2.1 Maximum-likelihood estimation

Let T_1, T_2, ..., T_n be a random sample of size n of some distribution $f(t)$ depending on a unknown vector of parameters θ. The likelihood function for the parameter vector θ is given by

$$L\left(\theta | t_1, t_2, ..., t_n\right) = \prod_{i=1}^{n} f(t_i).$$

Assuming the generalized Sibuya (GS) distribution defined by (3) for a discrete random variable T, we have $\theta = (\alpha, \beta)$ and

$$\begin{aligned} L(\alpha, \beta | \mathbf{t}) &= \prod_{i=1}^{n} \frac{\alpha}{\beta+t_i}\left(\frac{\beta+1-\alpha}{\beta+1} \times \frac{\beta+2-\alpha}{\beta+2} \times ... \times \frac{\beta+t_i-\alpha}{\beta+t_i} \times \frac{\beta+t_i-1-\alpha}{\beta+t_i-1}\right) \\ &= \prod_{i=1}^{n} \frac{\alpha}{\beta+t_i}\left(\prod_{j=1}^{t_i-1} \frac{\beta+j-\alpha}{\beta+j}\right), \end{aligned}$$

where $\mathbf{t} = (t_1, t_2, ..., t_n)$. In this case, the log-likelihood function is given by

$$\begin{aligned} \ell(\alpha, \beta | \mathbf{t}) &= n\ln(\alpha) - \sum_{i=1}^{n} \ln\left(\beta+t_i\right) + \sum_{i=1}^{n}\sum_{j=1}^{t_i-1} \ln\left(\beta+j-\alpha\right) - \sum_{i=1}^{n}\sum_{j=1}^{t_i-1} \ln\left(\beta+j\right) \\ &= n\ln(\alpha) - \sum_{i=1}^{n}\sum_{k=1}^{t_i} \ln(\beta+k) + \sum_{i=1}^{n}\sum_{j=1}^{t_i-1} \ln(\beta+j-\alpha). \end{aligned} \quad (6)$$

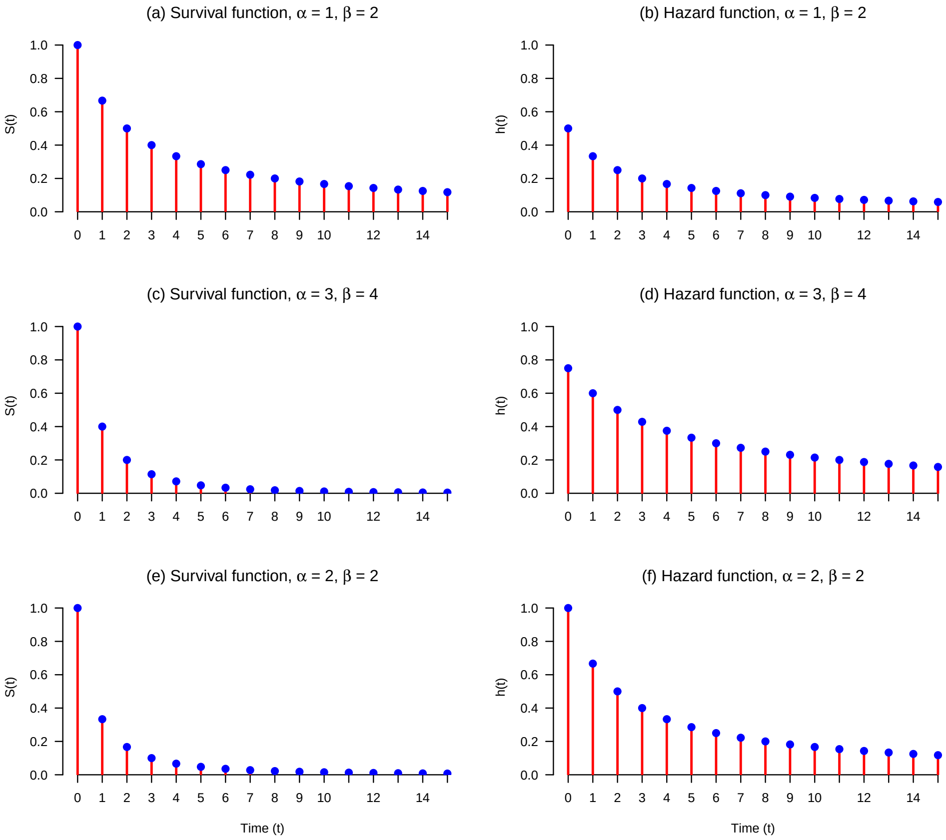

Figure 1: Plots of the survival and hazard functions of the generalized Sibuya distribution for some arbitrary values of α and β.

Therefore, the maximum likelihood estimators (MLE) for α and β which maximize (6) must satisfy the following equations:

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \sum_{j=1}^{t_i-1} \frac{1}{\beta + j - \alpha} = 0$$

and

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \sum_{j=1}^{t_i-1} \frac{1}{\beta + j - \alpha} - \sum_{i=1}^{n} \sum_{k=1}^{t_i} \frac{1}{\beta + k} = 0.$$

The maximum likelihood estimator $\widehat{\theta}_{ML} = (\widehat{\alpha}_{ML}, \widehat{\beta}_{ML})$ is thus obtained by solving the above system of nonlinear equations. Although we cannot obtain explicit expressions for the MLEs for the parameters α and β, they can be estimated numerically using existing iterative algorithms such as the Newton-Raphson method and its variants. In this article, we use the maxLik package in the statistical software R, which numerically maximizes the likelihood function (Henningsen and Toomet, 2011). The R code used in the present article is provided in an Appendix at the end of the paper.

Considering large sample sizes we can get inferences of interest for the parameters of the model, as hypothesis tests and confidence intervals, using the asymptotic bivariate normal distribution of the MLE $(\widehat{\alpha}_{ML}, \widehat{\beta}_{ML})$ given by

$$\begin{bmatrix} \widehat{\alpha}_{ML} \\ \widehat{\beta}_{ML} \end{bmatrix} \sim N\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \begin{bmatrix} \widehat{V}_{11} & \widehat{V}_{12} \\ \widehat{V}_{21} & \widehat{V}_{22} \end{bmatrix} \right),$$

where $\widehat{V}_{11}$ and $\widehat{V}_{22}$ are the estimated variances of $\widehat{\alpha}_{ML}$ and $\widehat{\beta}_{ML}$, respectively, and $\widehat{V}_{12} = \widehat{V}_{21}$ is the covariance between $\widehat{\alpha}_{ML}$ and $\widehat{\beta}_{ML}$. In this way, approximate $100(1-\gamma)\%$ Wald-type confidence intervals for α and β are, respectively, given by

$$\widehat{\alpha}_{ML} \mp z_{\gamma/2}\sqrt{\widehat{V}_{11}} \quad \text{and} \quad \widehat{\beta}_{ML} \mp z_{\gamma/2}\sqrt{\widehat{V}_{22}}, \tag{7}$$

where $z_{\gamma/2}$ denotes the upper γ-th percentile of the standard normal distribution.

The asymptotic variances of the MLEs are given by the elements of the inverse of the Fisher's information matrix. The expected Fisher's information matrix is given by

$$I(\alpha, \beta) = \begin{bmatrix} -E\left(\frac{\partial^2 \ell}{\partial \alpha^2}\right) & -E\left(\frac{\partial^2 \ell}{\partial \alpha \beta}\right) \\ -E\left(\frac{\partial^2 \ell}{\partial \alpha \beta}\right) & -E\left(\frac{\partial^2 \ell}{\partial \beta^2}\right) \end{bmatrix},$$

where the second partial derivatives of the log-likelihood function are given as follows:

$$\frac{\partial^2 \ell}{\partial \alpha^2} = -\frac{n}{\alpha^2} - \sum_{i=1}^{n}\sum_{j=1}^{t_i-1} \frac{1}{(\beta + j - \alpha)^2},$$

$$\frac{\partial^2 \ell}{\partial \beta^2} = \sum_{i=1}^{n}\sum_{k=1}^{t_i} \frac{1}{(k+\beta)^2} - \sum_{i=1}^{n}\sum_{j=1}^{t_i-1} \frac{1}{(\beta + j - \alpha)^2}$$

and

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta} = \sum_{i=1}^{n}\sum_{j=1}^{t_i-1} \frac{1}{(\beta + j - \alpha)^2}.$$

The asymptotic confidence intervals defined by the expression (7) may lead to negative lower bounds. To overcome this problem, we can use the first-order delta method (Oehlert, 1992) and the logarithmic transformation to get the asymptotic normal distributions for $\ln(\widehat{\alpha}_{ML})$ and $\ln(\widehat{\beta}_{ML})$ given by

$$\sqrt{n}\left(\ln(\widehat{\alpha}_{ML}) - \ln(\alpha)\right) \xrightarrow{D} N\left(0, \frac{\sigma_\alpha^2}{\alpha^2}\right)$$

and

$$\sqrt{n}\left(\ln(\widehat{\beta}_{ML}) - \ln(\beta)\right) \xrightarrow{D} N\left(0, \frac{\sigma_\beta^2}{\beta^2}\right),$$

respectively ($\xrightarrow{D}$ denotes convergence in distribution), where $\sigma_\alpha^2 = nVar(\widehat{\alpha}_{ML})$ and $\sigma_\beta^2 = nVar(\widehat{\beta}_{ML})$. Thus, asymptotic $100(1-\gamma)\%$ confidence intervals for $\ln(\alpha)$ and $\ln(\beta)$ are given respectively by

$$\ln(\widehat{\alpha}_{ML}) \mp z_{\gamma/2}\frac{\sqrt{Var(\widehat{\alpha}_{ML})}}{\widehat{\alpha}_{ML}} \equiv (L_\alpha, U_\alpha)$$

and

$$\ln(\widehat{\beta}_{ML}) \mp z_{\gamma/2}\frac{\sqrt{Var(\widehat{\beta}_{ML})}}{\widehat{\beta}_{ML}} \equiv (L_\beta, U_\beta).$$

Using the inverse logarithmic transformation, the approximate confidence intervals for α and β are respectively obtained as $(e^{L_\alpha}, e^{U_\alpha})$ and $(e^{L_\beta}, e^{U_\beta})$.

A Wald-type confidence interval for the mean of a random variable that follows a generalized Sibuya distribution can be obtained by applying the delta method and using the observed information matrix for $\theta = (\alpha, \beta)$ (Oehlert, 1992). Let us define the mean (4) as a function of the parameters (α, β), or say,

$$w(\alpha, \beta) = \frac{\beta}{\alpha - 1}, \tag{8}$$

where $\beta > 0$ and $1 < \alpha < \beta + 1$. Using a first-order Taylor series, an approximation for the variance of $w(\alpha, \beta)$ is given by

$$Var\left[w(\alpha,\beta)\right] \approx \begin{bmatrix} \frac{\partial w(\alpha,\beta)}{\partial \alpha} & \frac{\partial w(\alpha,\beta)}{\partial \beta} \end{bmatrix} \boldsymbol{\Sigma}(\alpha,\beta) \begin{bmatrix} \frac{\partial w(\alpha,\beta)}{\partial \alpha} \\ \frac{\partial w(\alpha,\beta)}{\partial \beta} \end{bmatrix},$$

where

$$\frac{\partial w(\alpha,\beta)}{\partial \alpha} = -\frac{\beta}{(\alpha-1)^2}, \qquad \frac{\partial w(\alpha,\beta)}{\partial \beta} = \frac{1}{\alpha - 1},$$

and $\boldsymbol{\Sigma}(\alpha, \beta)$ is the maximum-likelihood estimated variance-covariance matrix. Thus, an approximate $100(1-\gamma)\%$ Wald-type confidence interval for the mean is given by

$$w(\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}) \mp z_{\gamma/2}\sqrt{Var\left[w(\widehat{\alpha}_{ML}, \widehat{\beta}_{ML})\right]},$$

where $z_{\gamma/2}$ denotes the upper γ-th percentile of the standard normal distribution.

In this article, the Akaike (AIC) and Bayesian (BIC) information criteria were used to compare the proposed models based on standard Sibuya, generalized Sibuya and discrete Pareto distributions (Anderson and Burnham, 2002). The AIC value is given by $AIC = -2\ln(L) + 2k$, where L is the maximum value of the likelihood function for each candidate model, and k is the number of estimated parameters of a candidate distribution function. The BIC is defined as $BIC = -2\ln(L) + k\ln(n)$, where n is the the sample size. AIC and BIC allow for a direct comparison between models, with lower values indicating a better model fit (Akaike, 1974).

2.2 Maximum-likelihood estimation including censored data

Censored data are common in many areas of application when the primary endpoint of interest is time to a certain event, such as time to death. Right-censored data refers to individuals who have not yet experienced the event of interest at the end of the follow-up period. In this case, the contribution of the ith subject for the likelihood function under a right-censoring mechanism is given by

$$L_i = [f(t_i)]^{d_i}\,[S(t_i)]^{1-d_i},$$

where $d_i = 1$ denotes a complete observation and $d_i = 0$ denotes a censored observation (Klein and Moeschberger, 2003). Considering the generalized Sibuya distribution (3), the likelihood function is given by

$$\begin{aligned} L(\alpha,\beta|\mathbf{t},\mathbf{d}) &= \prod_{i=1}^{n}\left(\frac{\alpha}{\beta+t_i}\right)^{d_i}\left[\frac{(\beta+1-\alpha)_{t_i-1}}{(\beta+1)_{t_i-1}}\right]^{d_i}\left[\frac{(\beta+1-\alpha)_{t_i}}{(\beta+1)_{t_i}}\right]^{1-d_i} \\ &= \prod_{i=1}^{n}\left(\frac{\alpha}{\beta+t_i}\right)^{d_i}\left[\frac{\beta+t_i-1-\alpha}{\beta+t_i-1}\right]^{d_i}\left[\prod_{j=1}^{t_i-1}\frac{\beta+j-\alpha}{\beta+j}\right], \end{aligned}$$

where $\mathbf{d} = (d_1, d_2, ..., d_n)$. The log-likelihood function becomes

$$\begin{aligned} \ell(\alpha,\beta|\mathbf{t},\mathbf{d}) &= \ln\alpha\sum_{i=1}^{n} d_i - \sum_{i=1}^{n} d_i\ln(\beta+t_i) + \sum_{i=1}^{n} d_i\ln(\beta+t_i-1-\alpha) \\ &\quad -\sum_{i=1}^{n} d_i\ln(\beta+t_i-1) + \sum_{i=1}^{n}\sum_{j=1}^{t_i-1}\left[\ln(\beta+j-\alpha)-\ln(\beta+j)\right]. \end{aligned}$$

Deriving the log-likelihood function $\ell(\alpha,\beta|\mathbf{t},\mathbf{d})$ with respect to α and β, we have the following expressions:

$$\frac{\partial\ell}{\partial\alpha} = \frac{1}{\alpha}\sum_{i=1}^{n} d_i - \sum_{i=1}^{n}\frac{d_i}{\beta+t_i-1-\alpha} - \sum_{i=1}^{n}\sum_{j=1}^{t_i-1}\frac{1}{\beta+j-\alpha}$$

and

$$\begin{aligned} \frac{\partial\ell}{\partial\beta} &= -\sum_{i=1}^{n}\frac{d_i}{\beta+t_i} + \sum_{i=1}^{n}\frac{d_i}{\beta+t_i-1-\alpha} \\ &\quad -\sum_{i=1}^{n}\frac{d_i}{\beta+t_i-1} + \sum_{i=1}^{n}\sum_{j=1}^{t_i-1}\left[\frac{1}{\beta+j-\alpha}-\frac{1}{\beta+j}\right]. \end{aligned}$$

Setting these expressions equal to zero, we get the corresponding score equations whose numerical solution leads to the maximum likelihood estimators (MLE). The approximate standard errors for the maximum likelihood estimates can be obtained by computing the inverse of the observed Fisher information matrix at the estimated parameters, where the second partial derivatives of the log-likelihood function with respect to the parameters are given by

$$\frac{\partial^2 \ell}{\partial \alpha^2} = -\frac{1}{\alpha^2}\sum_{i=1}^{n} d_i - \sum_{i=1}^{n} \frac{d_i}{(\beta + t_i - 1 - \alpha)^2} - \sum_{i=1}^{n}\sum_{j=1}^{t_i - 1} \frac{1}{(\beta + j - \alpha)^2},$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \beta^2} &= \sum_{i=1}^{n} \frac{d_i}{(\beta + t_i)^2} - \sum_{i=1}^{n} \frac{d_i}{(\beta + t_i - 1 - \alpha)^2} \\ &\quad + \sum_{i=1}^{n} \frac{d_i}{(\beta + t_i - 1)^2} + \sum_{i=1}^{n}\sum_{j=1}^{t_i - 1}\left[\frac{1}{(\beta + j)^2} - \frac{1}{(\beta + j - \alpha)^2}\right]\end{aligned}$$

and

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta} = \sum_{i=1}^{n} \frac{d_i}{(\alpha - \beta - t_i + 1)^2} + \sum_{i=1}^{n}\sum_{j=1}^{t_i - 1} \frac{1}{(\beta + j - \alpha)^2}.$$

Wald-type 95% confidence intervals for the parameters can be obtained from the respective estimators of the standard errors.

2.3 Bayesian analysis

Since the Bayesian framework considers the parameters of a model as random variables, to estimate the unknown parameters α and β of the GS distribution it is necessary to specify their prior distributions (Gelman et al., 2013). According to the Bayes theorem, we can write the joint posterior density by combining the joint prior distribution with the likelihood function for these parameters of interest. Considering that α and β are positive parameters, we can assume a gamma prior distributions for these parameters. We then set $\alpha \sim Uniform(0, \beta + 1)$ and $\beta \sim Gamma(a, b)$, where a and b are known hyperparameters. From the Bayes theorem the posterior distribution function is proportional to the product of the likelihood function given in the previous sections and the joint prior distribution for these parameters.

Given that the joint posterior distribution is analytically intractable, it is not easy to generate samples from it. Therefore, in order to get information about the posterior distributions of the parameters of interest, we used MCMC methods, especially the Gibbs sampling and Metropolis-Hastings algorithms (Gelfand and Smith, 1990; Chib and Greenberg, 2000) available in the MCMCpack, an R package that contains functions to perform Bayesian analysis (Martin and Quinn, 2006; Martin et al., 2011). The model was ran for 1,000,000 iterations with a burn-in phase of 10,000 simulated samples and a thinning interval of size 200. The R code used to specify the model is given in an Appendix at the end of this paper.

Highest (posterior) density intervals (usually abbreviated as HPD intervals or HDI) with 95% coverage were obtained for all parameters of interest. A HPD interval (or HDI) is the shortest interval among all of the Bayesian credible intervals (Kruschke, 2014). Convergence was assessed visually from traceplots of each MCMC chain and quantitatively via Geweke criterion, which is based on a test for equality of the means of the first and last part of a Markov chain (Geweke, 1992). The MCMCpack package considers by default the first 10% and the last 50%. Assuming asymptotic independence between these two parts, the Geweke's Z-score is given by the difference between the two sample means divided by its estimated standard error. If the MCMC samples are drawn from a stationary distribution, this statistics has an asymptotically standard normal distribution.

To compare Bayesian models based on different distributions, the logarithm of the pseudo marginal likelihood (LPML) was used as a measure of goodness-of-fit statistic. This Bayesian model comparison criterion is based on the conditional predictive ordinate (CPO) statistics (Geisser and Eddy, 1979; Maity et al., 2021). Let $\mathbf{y} = \{y_1, \cdots, y_n\}$ be an observed sample from $f(\cdot|\theta)$. A Monte Carlo approximation of the CPO_i by using a MCMC sample $\{\theta_1, \cdots, \theta_M\}$ from the posterior distribution $\pi(\theta|\mathbf{y})$ is given by

$$\widehat{CPO}_i = \left\{ \frac{1}{M} \sum_{i=1}^{M} \frac{1}{f(y_i|\theta_m)} \right\}^{-1}$$

(Dey et al., 1997). Since the CPO_i is defined for each observation, the LPML value is given by

$$LPML = \sum_{i=1}^{n} \log \widehat{CPO}_i.$$

Models with larger LPML values are preferred over models with lower LPML values.

3 Results

3.1 Applications to simulated data

In order to exemplify the application of the proposed model and to compare the results from the frequentist and Bayesian approaches, samples from the GS distribution were simulated for different sizes and values for the parameters. We simulated samples of size n = 30, 60, 100, 250, 500, and 100 from the GS distribution for (a) $(\alpha, \beta) = (1, 2)$, (b) $(\alpha, \beta) = (3, 4)$ and (c) $(\alpha, \beta) = (2, 2)$. All the samples were simulated using the R code provided in the Appendix, including the presence of censored data. The seed of the random number generator of the R program was arbitrarily set to a value of 8, to allow the creation of simulations that can be reproduced.

Table 1 shows the ML and Bayesian estimates of α and β based on the simulated samples. Confidence intervals based on the ML approach were obtained using the delta method and the logarithmic transformation, as described in the previous section. Under the Bayesian estimation framework, prior distributions of α and β were given by $\alpha \sim Uniform(0, \beta+1)$ and $\beta \sim Gamma(0.1, 0.1)$, respectively, that is, approximately non-informative priors. The Bayesian estimates are approximately equal to the ML estimates, except when the sample size is relatively small (n less than 100). Both Bayesian and ML parameter estimates of α and β tend to approach the nominal level as n becomes larger, as it is expected. Also it is observed that the 95% confidence intervals for the parameters of the model are very large when compared to the 95% HPD's assuming a Bayesian approach for small sample sizes. It is important to point out that the classical confidence intervals are obtained using asymptotical distributions that depend on large sample sizes. This is a important point in favor of the Bayesian approach, since in many applications, especially in medical studies, we have small sample sizes.

3.2 Simulation study

In this subsection, we present a simulation study to illustrate the performance of the ML approach presented in the previous sections. In this way, the following criteria are adopted: the average bias, the mean squared error (MSE) and the coverage probabilities of the confidence intervals for the parameters α and β based on the delta method. The bias in the estimation of a parameter θ is estimated by

$$\widehat{Bias}(\widehat{\theta}_{ML}) = \frac{1}{B}\sum_{b=1}^{B}\left(\widehat{\theta}_{ML}^{(b)} - \theta_N\right)$$

and the corresponding MSE is estimated by

$$\widehat{MSE}(\widehat{\theta}_{ML}) = \frac{1}{B}\sum_{b=1}^{B}\left(\widehat{\theta}_{ML}^{(b)} - \theta_N\right)^2,$$

where $\widehat{\theta}_{ML} \in (\widehat{\alpha}_{ML}, \widehat{\beta}_{ML})$ is the ML estimate for a given parameter, $\widehat{\theta}_{ML}^{(b)}$ is the ML estimate obtained for θ considering the b-th simulated sample, θ_N is the corresponding nominal value for θ, $\theta \in (\alpha, \beta)$, and B is the number of simulated samples. The bias informs to what extent the average of an estimated value is far from the true parameter, and the MSE is a measure of the average of the squared errors. Values of bias and MSE that are closer to zero are desirable.

We generated $B = 1,000$ random samples each of size $n = 25, 30, 35, 40, ..., 400$, using the R functions provided in the Appendix at the end of this article, including censored data. The results from the simulation study are presented in Figure 2, considering a nominal confidence coefficient of 95% and three sets of arbitrary values for the parameters of the GS distribution, given by for (a) $(\alpha, \beta) = (1, 2)$, (b) $(\alpha, \beta) = (3, 4)$ and (c) $(\alpha, \beta) = (2, 2)$.

Table 1: Maximum likelihood and Bayesian estimates for the parameters of the model based on GS distribution.

Nominal Values	n	Parameter	Maximum likelihood estimates: Estimate	Std. error	95%CI	Bayesian estimates: Estimate	Std. dev.	95%HDI
$GS(1,2)$	30	α	2.2574	0.8045	(1.0556, 4.8304)	1.1820	0.4011	(0.5061, 1.9934)
		β	5.2225	2.4389	(1.9199, 14.2149)	1.9490	1.1084	(0.2472, 4.2501)
	60	α	0.7706	0.2400	(0.4707, 1.2614)	0.7511	0.1789	(0.4497, 1.1115)
		β	1.0216	0.8346	(0.2914, 3.5808)	0.9920	0.6096	(0.0024, 2.1566)
	100	α	0.9301	0.2146	(0.5893, 1.4683)	0.9166	0.2042	(0.5617, 1.3351)
		β	1.7807	0.8633	(0.6878, 4.6107)	1.7755	0.8333	(0.4265, 3.5248)
	250	α	0.9671	0.1088	(0.7427, 1.2596)	0.9630	0.1327	(0.7268, 1.2364)
		β	1.9121	0.4233	(1.1141, 3.2826)	1.9170	0.5453	(0.9413, 3.0135)
	500	α	1.2589	0.1164	(1.0366, 1.5293)	1.2580	0.1381	(0.9897, 1.5232)
		β	2.8919	0.4560	(2.0567, 4.0690)	2.9010	0.5597	(1.9043, 4.0442)
	1000	α	1.0772	0.0756	(0.9522, 1.2186)	1.0759	0.0735	(0.9301, 1.2167)
		β	2.3281	0.3123	(1.8426, 2.9423)	2.3292	0.3056	(1.7468, 2.9291)
$GS(3,4)$	30	α	2.8092	2.7296	(0.1770, 44.5887)	1.4193	0.5374	(0.6434, 2.5069)
		β	3.3978	4.5817	(0.0727, 158.6691)	1.1909	0.8718	(0.0003, 2.8640)
	60	α	2.3900	1.2588	(0.9439, 6.0493)	1.5716	0.5236	(0.7110, 2.6106)
		β	2.7227	2.0666	(0.7147, 10.3675)	1.4364	0.8448	(0.0006, 3.0411)
	100	α	3.5399	1.6230	(1.1068, 11.2486)	3.2627	2.1510	(0.6462, 7.4184)
		β	4.3054	2.5274	(0.9702, 18.9424)	3.9512	3.4260	(0.0001, 10.5278)
	250	α	2.2289	0.5489	(1.3435, 3.6960)	2.2606	0.5919	(1.2636, 3.4180)
		β	2.5317	0.9393	(1.1798, 5.4285)	2.6095	1.0304	(0.9387, 4.6491)
	500	α	2.9686	0.5036	(1.8872, 4.6889)	2.9996	0.6422	(1.9129, 4.3007)
		β	4.0217	0.9122	(2.1916, 7.4208)	4.0945	1.1675	(2.1189, 6.4703)
	1000	α	2.7683	0.4075	(1.8982, 4.0377)	2.7766	0.3688	(2.1023, 3.5153)
		β	3.6252	0.7357	(2.1523, 6.1072)	3.6475	0.6688	(2.5161, 5.0590)
$GS(2,2)$	30	α	1.5980	1.1190	(0.3738, 6.8326)	1.2600	0.3525	(0.6594, 1.9484)
		β	1.3883	1.7682	(0.0987, 19.5321)	0.9264	0.5476	(0.0523, 1.9809)
	60	α	0.9234	0.3825	(0.4083, 2.0882)	1.1188	0.2643	(0.7154, 1.6751)
		β	0.3364	0.5882	(0.0107, 10.5448)	0.6845	0.4186	(0.0487, 1.5316)
	100	α	2.5278	1.2319	(0.9231, 6.9249)	2.3529	1.5287	(0.6456, 5.2016)
		β	2.3752	1.6842	(0.5487, 10.2871)	2.1885	2.1600	(0.0001, 6.0814)
	250	α	1.8726	0.4403	(1.1722, 2.9915)	1.8645	0.4655	(1.0429, 2.7914)
		β	1.6267	0.6485	(0.7352, 3.5993)	1.6314	0.6923	(0.4839, 3.0658)
	500	α	1.9657	0.3429	(1.3964, 2.7670)	1.9704	0.3486	(1.3515, 2.6725)
		β	1.9629	0.5505	(1.1329, 3.4010)	1.9795	0.5638	(0.9932, 3.1171)
	1000	α	1.7861	0.1863	(1.4317, 2.2281)	1.7851	0.1989	(1.4116, 2.1764)
		β	1.6274	0.2916	(1.1129, 2.3799)	1.6295	0.3117	(1.0586, 2.2638)

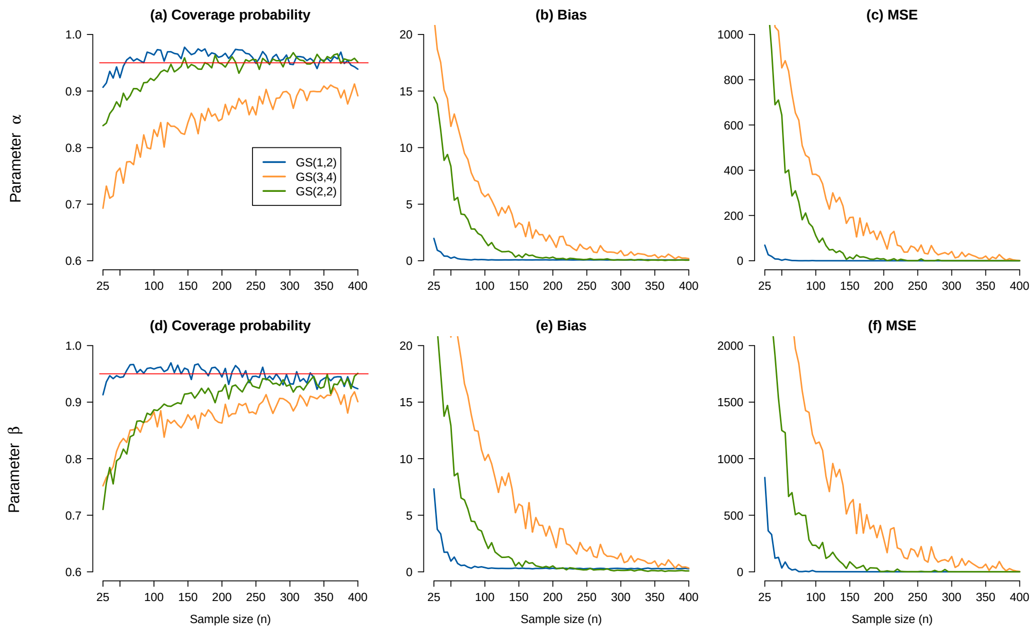

Figure 2: Results from the simulation study. The graphs show the estimates of the coverage probabilities of the 95% confidence intervals for the parameters α and β based on the delta method, the average bias, and the mean squared error (MSE).

We can note in Figure 2 that in all simulations, the biases and the MSE always approached zero as the sample size increased, as it is expected. However, the simulations suggest that a better performance of the ML approach is observed for $(\alpha, \beta) = (1, 2)$, where we can notice a better approximation of the coverage probabilities to the nominal value of 95% even when the sample size is relatively small. When considering $(\alpha, \beta) = (3, 4)$, the coverage probability is below the nominal level even for large sample sizes, and the bias and MSE need relatively large sample sizes to reach values close to zero. In summary, this simulation study suggests that the ML approach has a worse performance for higher parameter values.

3.3 Applications to real data sets

3.3.1 Special educators data

Now let us consider a real dataset introduced by Singer and Willett (1993), where they evaluated the number of consecutive years in teaching for 3,941 special educators hired in Michigan public schools between 1972 and 1978. The event time is right-censored if an educator was still teaching when the data collection was finished. Approximately 44% of the observations are censored. Table 2 shows the ML estimates, standard errors and Wald-type confidence intervals obtained from the analyses based on the standard

Sibuya, generalized Sibuya and discrete Pareto distributions. Considering the model based on the GS distribution, it was not possible to estimate the mean of the number of consecutive years in teaching because the correspondent ML estimate for α is lower than 1.

Table 2: Maximum likelihood estimates, standard errors, 95% confidence intervals (95%CI), AIC and BIC values for the models based on the standard Sibuya, generalized Sibuya and discrete Pareto distributions, applied to the special educators data.

Distribution	Parameter	Estimate	Std. error	95%CI	AIC	BIC
Sibuya ($\beta = 0$)	α	0.2310	0.0045	$(0.2223, 0.2400)$	15335	15341
Generalized Sibuya	α	0.7869	0.0282	$(0.7335, 0.8442)$	14667	14679
	β	5.0520	0.2536	$(4.5787, 5.5743)$		
Discrete Pareto ($\alpha = 1$)	β	7.0950	0.1748	$(6.7345, 7.4731)$	14674	14680

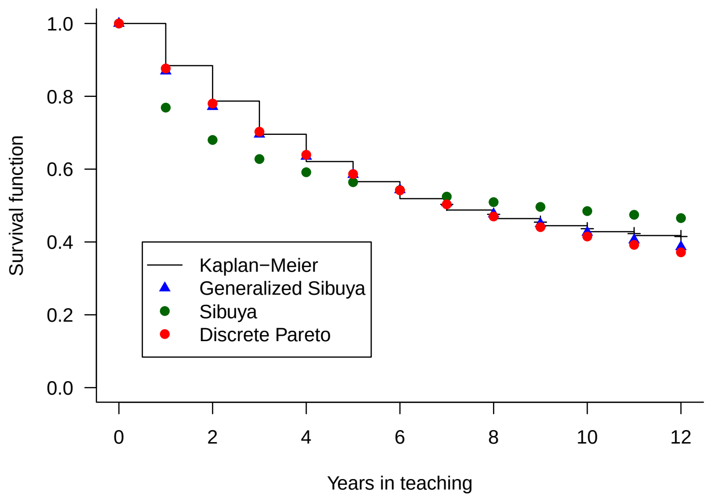

Figure 3: Survival function for the special educators data estimated by the Kaplan-Meier method and by using the models based on the generalized Sibuya, standard Sibuya and discrete Pareto distributions.

Table 3: Bayesian estimates, 95% HDI and LPML values for the models based on the standard Sibuya, generalized Sibuya and discrete Pareto distributions, applied to the special educators data.

Distribution	Parameter	Posterior mean	Std. dev.	95%HDI interval	Geweke's Z-score	LPML
Sibuya ($\beta = 0$)	α	0.2311	0.0045	$(0.2225, 0.2399)$	-0.2700	-7667.17
Generalized Sibuya	α	0.7699	0.0540	$(0.6718, 0.8789)$	-0.4318	-7333.35
	β	4.8360	0.5311	$(3.8177, 5.8678)$	-0.5891	
Discrete Pareto ($\alpha = 1$)	β	7.0622	0.2001	$(6.6868, 7.4627)$	-0.5172	-7336.74

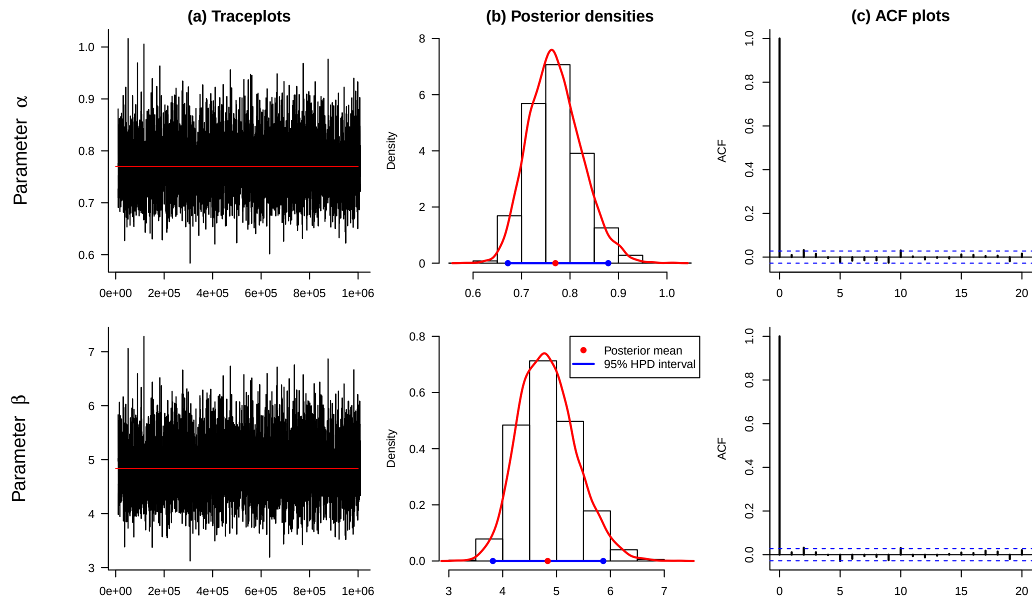

Figure 4: Posterior samples for the model parameters of the Generalized Sibuya distribution applied to the special educators data. (a) Traceplots of posterior samples, (b) histograms and posterior densities with the correspondent 95% highest density interval (HDI) intervals, and (c) auto-correlation function (ACF) plots for the posterior samples of the model parameters.

Figure 3 compares the Kaplan-Meier estimates and the ML estimates for the survival function obtained from the three assumed parametric models. We can observe that the estimates for the survival function obtained from the models based on the GS and discrete Pareto distributions are closer to those obtained by the Kaplan-Meier method, than the estimates obtained from the model based on the standard Sibuya distribution. In addition, the AIC and BIC values shown in Table 2 suggest that the GS and discrete Pareto models provided a better fit for the special educators data compared to the standard Sibuya model. In fact, as suggested by the shape of the Kaplan-Meier curve, heavy tail distributions seem to be more appropriate to be fitted by the data.

For a Bayesian analysis of the data, it was assumed $\alpha \sim Uniform(0, \beta + 1)$ and the same approximately non-informative gamma prior distribution for the parameter β as considered in the simulation study presented in Subsection 3.1. The Bayesian estimates were obtained using MCMC methods. Table 3 shows the Bayesian estimates of the parameters of the models based on the standard Sibuya, GS and discrete Pareto distributions, when applied to the special educators data, with their respective 95% HDI. As seen from Table 3, the Bayesian estimates are close to those based on the ML approach (Table 2). Table 3 also shows the Geweke's Z-scores for each parameter. Since all absolute Z-scores are less than 1.96, the Geweke's diagnostic evidenced the convergence of the correspondent chains. Similar to the frequentist analyses, the Bayesian models based on the GS and discrete Pareto distributions have the highest LPML values, suggesting that these models provide the best fit to the data.

Posterior samples for the model parameters of the GS distribution applied to the special educators data are described in Figure 4. Traceplots again indicate that the generated samples reached good convergence, and the plots of the auto-correlation function (ACF) show that the posterior samples are approximately uncorrelated. Histograms suggest that the posterior samples have slightly skewed distributions. Smooth kernel density estimates are superimposed to the histograms.

3.3.2 Breastfeeding duration data

A prospective cohort study of $1,803$ live-born children and their mothers was conducted in a tertiary obstetric hospital in Australia (Oddy et al., 2006). Let us consider a subsample of this study involving 324 overweight and obese women. The variable of interest is the age of the children in months at the moment when breastfeeding stopped. This study does not include censored observations. As the data were available in figures and not in numerical form, we used the open-source software WebPlotDigitizer, which is a web based tool to extract numerical data from images (Drevon et al., 2017; Rohatgi, 2020). Table 4 shows the ML estimates, standard errors and Wald-type confidence intervals obtained from the analyses based on the standard Sibuya, GS and discrete Pareto distributions. Considering the GS model, the mean of breastfeeding duration was obtained by the equation (8) and the respective standard error was obtained by the delta method as described in the Methods section. The model based on GS distribution showed a lower AIC and BIC values compared with the others, indicating a better fit of this model to the observed data.

Figure 5 compares the survival function estimated by the Kaplan-Meier method and fitted by parametric models using ML estimation method. The Kaplan-Meier estimates do not suggest that the time-to-event variable has a heavy-tail distribution, and in this case the GS model can produce estimates for the survival function closest to the empirical values when compared with the discrete Pareto model.

Table 4: Maximum likelihood estimates, standard errors, 95% confidence intervals (95%CI), AIC and BIC values for the models based on the standard Sibuya, generalized Sibuya and discrete Pareto distributions, applied to the breastfeeding duration data.

Distribution	Parameter	Estimate	Std. error	95%CI	AIC	BIC
Sibuya ($\beta = 0$)	α	0.4178	0.0189	$(0.3818, 0.4576)$	2004.13	2009.63
Generalized Sibuya	α	4.6308	0.3268	$(4.0326, 5.3176)$	1837.14	1848.13
	β	25.0712	0.6932	$(23.7487, 26.4672)$		
	Mean	6.9052	0.5113	$(5.9031, 7.9074)$		
Discrete Pareto ($\alpha = 1$)	β	2.9987	0.2908	$(2.4796, 3.6265)$	1871.88	1877.37

Table 5: Bayesian estimates, 95% HDI and LPML values for the models based on the standard Sibuya, generalized Sibuya and discrete Pareto distributions, applied to the breastfeeding duration data.

Distribution	Parameter	Posterior mean	Std. dev.	95%HDI interval	Geweke's Z-score	LPML
Sibuya ($\beta = 0$)	α	0.4178	0.0189	$(0.3816, 0.4554)$	-0.2284	-1001.71
Generalized Sibuya	α	3.6233	1.0763	$(1.8343, 5.7367)$	1.901	-918.54
	β	18.5591	6.9327	$(7.3648, 32.1564)$	2.090	
	Mean	7.2068	0.6164	$(6.1259, 8.4636)$	-0.897	
Discrete Pareto ($\alpha = 1$)	β	3.0082	0.2881	$(2.4599, 3.5810)$	-0.1597	-935.89

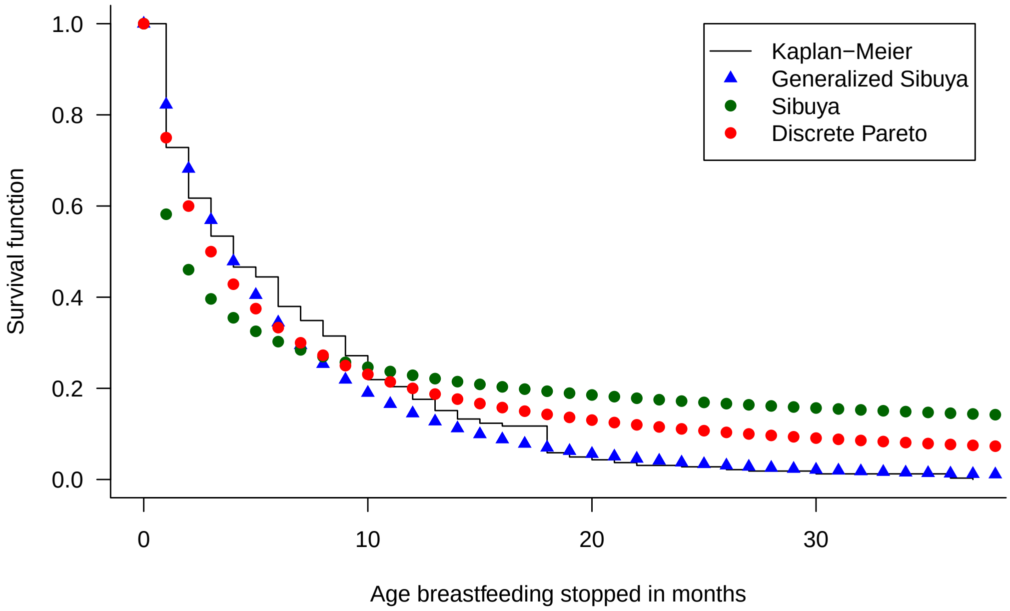

Figure 5: Survival function for the breastfeeding duration data estimated by the Kaplan-Meier method and by using the models based on the generalized Sibuya, standard Sibuya and discrete Pareto distributions.

Table 5 shows the Bayesian posterior estimates of the parametric models also assuming $\alpha \sim Uniform(0, \beta + 1)$ and approximately non-informative gamma prior distributions for the parameter β of the proposed models. These estimates are relatively close to those obtained from the ML method, showed in 4. The results of the GS model provide the highest LPML value, suggesting that this model provides the best fit to the data among these three models. The results from Figure 6 indicate satisfactory convergence for both simulated chains, despite we note in Table 5 a Geweke's Z-score a little larger than 1.96 for the β parameter.

4 Concluding remarks

In this article, we explored the potential of the GS distribution introduced by Kozubowski and Podgórski (2018) to get better analysis of discrete time-to-event data, including censored observations. Maximum likelihood (ML) and Bayesian approaches were used to estimate the parameters α and β of the model. Applications to simulated data showed that both the ML and Bayesian approaches are computationally feasible to estimate the parameters of GS distribution. However, we have noticed that ML and Bayesian are relatively distant from each other when the sample size is small, perhaps due to some sense of information provided by the choice of the prior distributions. Therefore, future studies should investigate the possibility of obtaining less informative prior distributions. In

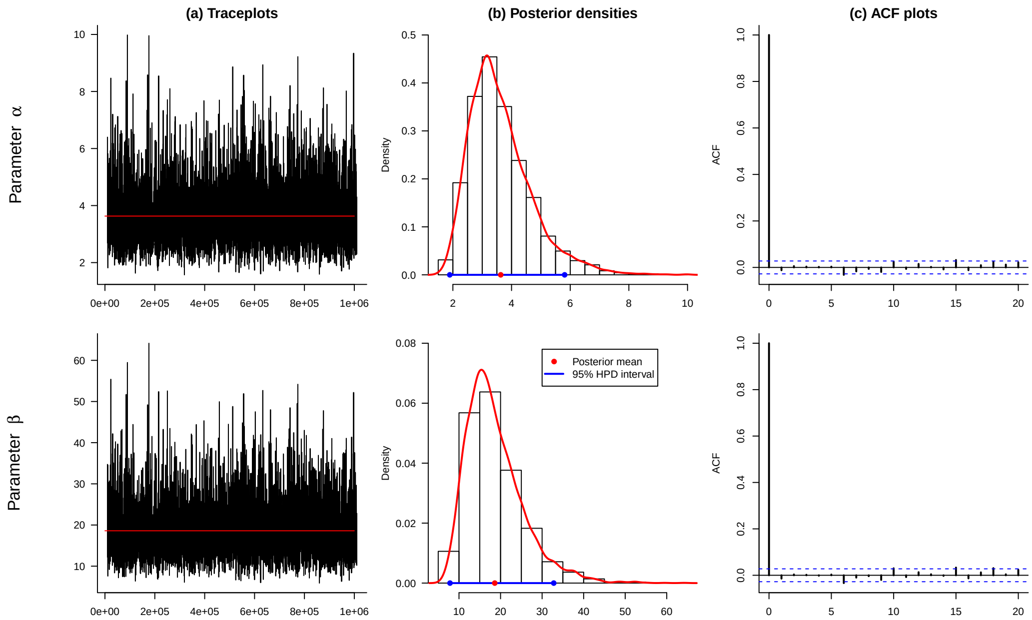

Figure 6: Posterior samples for the model parameters of the Generalized Sibuya distribution applied to the breastfeeding duration data. (a) Traceplots of posterior samples, (b) histograms and posterior densities with the correspondent 95% highest density interval (HDI) intervals, and (c) auto-correlation function (ACF) plots for the posterior samples of the model parameters.

this study, we assumed approximately non-informative gamma prior distributions for the parameters of the models in the simulation and real data applications. In practical work, we could use prior opinion of experts to get more informative priors, especially when we have small sample sizes, a situation where the elicitation of priors is an important Bayesian issue to get more accurate inference results.

The estimation performance of the ML method was investigated in a simulation study. We observed that the ML method perform well when the nominal values of the α and β parameters are relatively small, and in this case the estimated bias and MSE values are satisfactorily low even for small samples. We also conducted a simulation study to investigate the coverage probability of the Wald-type confidence intervals for α and β parameters based on expression (7) (results not shown), but we have noticed that the intervals based on the delta method and logarithmic transformation generally yields coverage probabilities closer to nominal.

Applications with real data have shown that the GS probability distribution can fits the data very well, under both the ML and Bayesian frameworks. The example that considers the special educators data shows that both GS and discrete Pareto distributions are suitable for the data analysis, considering a heavy tail distribution. However, the analysis of the breastfeeding duration data evidenced that GS model fits the data better, considering a distribution without heavy tail. In a situation like this, the GS model is found to be advantageous over the model based on the discrete Pareto distribution.

The GS model can be easily implemented in existing free computational programs as R, as showed in the Appendix. An important limitation of the generalized Sibuya distribution is to have a lack of flexibility on the hazard function. Its hazard function $h(t)$ is always a decreasing function in t, a fact that could be become a serious obstacle in the application of the model to real data.

References

Abramowitz, M. and Stegun I. A. (eds.) (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Applied Mathematics Series 55.* Dover Publications, New York.

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control*, 19:716-723.

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, 13:61-98.

Anderson, D. and Burnham, K. (2002). *Model selection and multi-model inference.* 2nd edition. Springer-Verlag, New York.

Bouzar, N. (2008). The semi-Sibuya distribution. *Annals of the Institute of Statistical Mathematics*, 60(2):459-464.

Buddana, A. and Kozubowski, T. J. (2014). Discrete Pareto distributions. *Stochastics and Quality Control*, 29(2):143-156.

Cardial, M.R.P., Fachini-Gomes, J.B., and Nakano, E. Y. (2020). Exponentiated discrete Weibull distribution for censored data. *Brazilian Journal of Biometrics*, 38(1), 35–56.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4), 327-335.

Christoph, G. and Schreiber, K. (2000). Scaled Sibuya distribution and discrete self-decomposability. *Statistics & Probability Letters*, 48(2), 181-187.

Devroye, L. (1993). A triptych of discrete distributions related to the stable law. *Statistics & Probability Letters*, 18(5), 349-351.

Dey, D. K., Chen, M. H., and Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics*, 53(4), 1239-1252.

Drevon, D., Fursa, S. R., and Malcolm, A. L. (2017). Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior Modification*, 41(2), 323-339.

Eldeeb, A. S., Ahsan-ul-Haq, M., Eliwa, M. S., and Cell, Q. E. (2022). A discrete Ramos-Louzada distribution for asymmetric and over-dispersed data with leptokurtic-shaped: Properties and various estimation techniques with inference. *AIMS Mathematics*, 7(2), 1726-1741.

Freitas, B. C. L., Oliveira-Peres, M. V., Achcar, J. A., and Martinez, E. Z. (2021). Classical and Bayesian inference approaches for the exponentiated discrete Weibull model with censored data and a cure fraction. *Pakistan Journal of Statistics and Operation Research*, 17(2), 467–481.

Gallardo, D. I., Gómez, H. W., and Bolfarine, H. (2017). A new cure rate model based on the Yule-Simon distribution with application to a melanoma data set. *Journal of Applied Statistics*, 44(7):1153-1164.

Geisser, S., and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365), 153-160.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410), 398-409.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*, 3th Edition. Chapman and Hall/CRC.

Geweke J (1992). Bayesian Statistics, volume 4, chapter Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, 4:641-649.

Graf, M. and Nedyalkova, D. (2015). GB2: Generalized Beta Distribution of the Second Kind: Properties, Likelihood, Estimation. *R Package Version 2.1*

Gupta, R. C. and Huang, J. (2017). The Weibull-Conway-Maxwell-Poisson distribution to analyze survival data. *Journal of Computational and Applied Mathematics*, 311:171-182.

Henningsen, A. and Toomet, O. (2011). maxLik: A package for maximum likelihood estimation in R. *Computational Statistics*, 26(3):443-458.

Herzog, W., Schellberg, D., and Deter, H. C. (1997). First recovery in anorexia nervosa patients in the long-term course: a discrete-time survival analysis. *Journal of Consulting and Clinical Psychology*, 65(1):169-177.

Jayakumar, K. and Babu, M. G. (2018). Discrete Weibull geometric distribution and its properties. *Communications in Statistics - Theory and Methods*, 47(7):1767-1783.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*. New York: Springer.

Kozubowski, T. J. and Podgórski, K. (2018). A generalized Sibuya distribution. *Annals of the Institute of Statistical Mathematics*, 70(4):855-887.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* Academic Press.

Letac, G. (2019). Is the Sibuya distribution a progeny?. *Journal of Applied Probability*, 56(1):52-56.

Maity, A., Williams, P. L., Ryan, L., Missmer, S. A., Coull, B. A., and Hauser, R. (2014). Analysis of in vitro fertilization data with multiple outcomes using discrete time-to-event analysis. *Statistics in Medicine*, 33(10):1738–1749.

Maity, A. K., Basu, S., and Ghosh, S. (2021). Bayesian criterion-based variable selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70(4):835–857.

Martin, A. D. and Quinn, K. M. (2006). Applied Bayesian inference in R using MCMCpack. *R News*, 6(1):2-7.

Martin, A. D., Quinn, K. M., and Park, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, 42(9):1-21

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

Nekoukhou, V. and Bidram, H. (2015). The exponentiated discrete Weibull distribution. *SORT Statistics and Operations Research Transactions*, 39(1): 127–146.

Northrup, T. F., Stotts, A. L., Green, C., Potter, J. S., Marino, E. N., Walker, R., Weiss, R. D., and Trivedi, M. (2015). Opioid withdrawal, craving, and use during and after outpatient buprenorphine stabilization and taper: a discrete survival and growth mixture model. *Addictive Behaviors*, 41:20-28.

Oddy, W. H., Li, J., Landsborough, L., Kendall, G. E., Henderson, S., and Downie, J. (2006). The association of maternal overweight and obesity with breastfeeding duration. *The Journal of Pediatrics*, 149(2):185-191.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, 46(1):27-29.

Ramos, P. L., Guzman, D. C., Mota, A. L., Rodrigues, F. A., and Louzada, F. (2020). Sampling with censored data: a practical guide. *arXiv preprint*, arXiv:2011.08417.

Rohatgi, A. (2020). WebPlotDigitizer (Version 4.4) [Computer software]. Retrieved from https://automeris.io/WebPlotDigitizer/

Scheike, T. H. and Jensen, T. K. (1997). A discrete survival model with random effects: an application to time to pregnancy. *Biometrics*, 53(1):318-329.

Sibuya, M. (1979). Generalized hypergeometric, digamma and trigamma distributions. *Annals of the Institute of Statistical Mathematics*, 31(3):373-390.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3-4):425-440.

Singer, J. D. and Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18(2):155-195.

Tutz, G. and Schmid, M. (2016). *Modeling discrete time-to-event data.* New York: Springer.

Vanegas, J. C., Chavarro, J. E., Williams, P. L., Ford, J. B., Toth, T. L., Hauser, R., and Gaskins, A. J. (2017). Discrete survival model analysis of a couple's smoking pattern and outcomes of assisted reproduction. *Fertility Research and Practice*, 3(1):5.

Yule, G. U. (1924). A Mathematical Theory of Evolution, based on the Conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society B*, 213(402-410):21-87.

Appendix 1: Simulating samples from a generalized Sibuya distribution

The Proposition 7 of the article by Kozubowski and Podgórski (2018) establish that if $T \sim GS(\alpha, \beta)$ we have $T \stackrel{d}{=} N(X)$, where X is given by

$$X \stackrel{d}{=} \frac{T_1}{T_2(\alpha, \beta)}$$

and is independent of a standard Poisson process $\{N(t), t > 0\}$. T_1 follows standard exponential distribution and $T_2(\alpha, \beta)$ follows a continuous beta distribution of the second kind, with probability density function given by

$$f(x) = \frac{1}{B(\alpha, \beta - \alpha + 1)} \frac{x^{\alpha-1}}{(1+x)^{\beta+1}},$$

where $B(a, b)$ denotes the beta function, $x \geq 0$, $\beta \geq 0$ and $0 < \alpha < \beta + 1$. We also consider independence between T_1 and $T_2(\alpha, \beta)$.

For generating random samples from a beta distribution of the second kind, the R package GB2 can be used (Graf and Nedyalkova, 2015). The function `rgb2(n,shape1, scale, shape2, shape3)` included in the GB2 package generates random samples from a generalized beta distribution of the second kind, which is a four-parameter distribution. In order to generate n samples from a beta distribution of the second kind, we can use `rgb2(n,1,1,shape2,shape3)`.

The following R function `rGsib` is used to generate random samples of size n from a generalized Sibuya distribution with parameters α and β.

```
# Loading the GB2 package
library(GB2)
# Simulating samples from a generalized Sibuya distribution
rGsib <- function(n,alpha,beta) {
 if (alpha >= (beta+1)) { stop("Alpha must be less than (beta - 1)") }
 if (beta<0) { stop("Beta must be greater or equal to zero") }
 if (alpha<=0) { stop("Alpha must be greater than zero") }
 T1 <- rexp(n,1)
 T2 <- rgb2(n,1, 1, alpha, beta - alpha + 1)
 X <- T1/T2
 rGsib <- rep(NA,n)
 for (i in 1:n) rsib[i] <- 1 + rpois(1,X[i])
 return(rGsib) }
```

In order to simulate a sample of size n from the generalized Sibuya distribution with randomly right-censored data, we can use the Algorithm 6 of the article by Ramos et al. (2020). To do this, we can use the following R function.

```
# Generating a discrete uniform random variable
rdu <- function(n,k) sample(1:k,n,replace=T)
# Generating a GS random variable
rGsibC <- function(n,alpha,beta) {
     x <- rGsib(n,alpha,beta)
     c <- rdu(n,max(x))
     t <- pmin(x,c)
     d <- as.numeric(x<=c)
 sampl <- data.frame(t,d)
 return(sampl) }
```

Appendix 2: Maximum likelihood estimation

Under the frequentist approach, the following R code can be used for the model using the function `maxLik` of the maxLik package (Henningsen and Toomet, 2011) for the maximization of the likelihood function.

```
# Loading the maxLik package
 library(maxLik)
# Defining the Pochhammer function
 pochhammer <- function(a,b) pochhammer <- gamma(a+b)/gamma(a)
# The likelihood function
 log.f <- function(parms) {
alpha <- parms[1]
beta  <- parms[2]
if (parms[1]<0) return(-Inf)
if (parms[2]<0) return(-Inf)
ft <- alpha/(beta+t)*pochhammer(beta+1-alpha,t-1)/pochhammer(beta+1,t-1)
St <- pochhammer(beta+1-alpha,t)/pochhammer(beta+1,t)
like <- ft^d*St^(1-d)
L <- sum(log(like))
if (is.na(L)==TRUE) {return(-Inf)} else {return(L)} }
# Obtaining the ML estimates
 mle  <- maxLik(logLik=log.f,start=c(alpha.0,beta.0))
 summary(mle)
# The AIC value
 print(paste("AIC = ", AIC(mle)))
# ML estimates
 alphaML <- mle$estimate[1]
 betaML  <- mle$estimate[2]
# Obtaining the variance-covariance matrix
 s <- vcov(mle)
# Obtaining the 95% CI for alpha e beta
 Lalpha <- round(exp(log(alphaML) - qnorm(0.975)*sqrt(s[1,1])/alphaML),4)
```

```
Ualpha <- round(exp(log(alphaML) + qnorm(0.975)*sqrt(s[1,1])/alphaML),4)
Lbeta  <- round(exp(log(betaML)  - qnorm(0.975)*sqrt(s[2,2])/betaML),4)
Ubeta  <- round(exp(log(betaML)  + qnorm(0.975)*sqrt(s[2,2])/betaML),4)
cat("Alpha = ",alphaML,"95%CI: (",Lalpha,",",Ualpha,") \n")
cat("Beta  = ",betaML, "95%CI: (",Lbeta, ",",Ubeta, ") \n")
```

In this code `ft` is the probability function, `St` is the survival function, `like` is the likelihood function, `t` is the time-to-event variable, `d` is the censoring indicator, and `alpha.0` and `beta.0` are respectively the initial values for α and β.

Appendix 3: Bayesian estimation

The following R code can be used to obtain the Bayesian estimates. The MCMCpack package also allows users to obtain the HPD intervals and the Geweke's Z-scores.

```
# Loading the MCMCpack package
library(MCMCpack)
# Defining the Pochhammer function
pochhammer <- function(a,b) pochhammer <- gamma(a+b)/gamma(a)
# The log posterior function
log.post <- function(t,d,parms) {
alpha <- parms[1]
beta  <- parms[2]
if (parms[1]<0) return(-Inf)
if (parms[2]<0) return(-Inf)
ft <- alpha/(beta+t)*pochhammer(beta+1-alpha,t-1)/pochhammer(beta+1,t-1)
St <- pochhammer(beta+1-alpha,t)/pochhammer(beta+1,t)
like     <- ft^d*St^(1-d)
log.like <- sum(log(like))
prior    <- dunif(alpha,0,beta+1)*dgamma(beta,.1,.1)
log.prior <- log(prior)
L <- log.like + log.prior
if (is.na(L)==TRUE) {return(-Inf)} else {return(L)} }
# Obtaining the MCMC estimates
posterior <- MCMCmetrop1R(log.post,
  theta.init=c(alpha=0.5,beta=1.7), burnin=10000,
  mcmc=1000000, thin=200, logfun=T, t=t, d=d, verbose=100000, tune = 1)
varnames(posterior)<-c("alpha","beta")
summary(posterior)
# Obtaining the HPD intervals
HPDinterval(posterior, prob = 0.95)
# Geweke Z scores
geweke.diag(posterior)
```