Ann Inst Stat Math (2018) 70:855–887
https://doi.org/10.1007/s10463-017-0611-3

A generalized Sibuya distribution

Tomasz J. Kozubowski[1] · **Krzysztof Podgórski**[2]

Received: 11 June 2016 / Revised: 14 February 2017 / Published online: 22 June 2017

Abstract The Sibuya distribution arises 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 a trial. We study a generalization that can be viewed as the distribution of the excess random variable $N - k$ given $N > k$, where N has the Sibuya distribution and k is an integer. We summarize basic facts regarding this distribution and provide several new results and characterizations, shedding more light on its origin and possible applications. In particular, we emphasize the role Sibuya distribution plays in the extreme value theory and point out its invariance property with respect to random thinning operation.

Keywords Discrete Pareto distribution · Distribution theory · Extreme value theory · Infinite divisibility · Mixed Poisson process · Power law · Pure death process · Records · Yule distribution · Zipf's law

1 Introduction

Let X_i, $i \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$, be a sequence of independent and identically distributed (IID) continuous random variables. The first X_i, $i \in \mathbb{N} = \{1, 2, \ldots\}$, that exceeds all previous values (including the X_0) is called the first *record* value. Let I_j, $j \in \mathbb{N}$, be the associated sequence of Bernoulli random variables, indicating whether

✉ Tomasz J. Kozubowski
tkozubow@unr.edu

Krzysztof Podgórski
Krzysztof.Podgorski@stat.lu.se

[1] Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557, USA

[2] Department of Statistics, Lund University, 220 07 Lund, Sweden

or not a particular X_j is a record. It follows from the random records theory (see, e.g., Rényi 1976) that the variables I_j are mutually independent and

$$\mathbb{P}(I_j = 1) = \frac{1}{1+j}, \quad j \in \mathbb{N}.$$

Accordingly, if N denotes the waiting time for the first record to occur, then

$$\mathbb{P}(N = n) = \frac{1}{n(n+1)} = \frac{1}{1+(n-1)} - \frac{1}{1+n}, \quad n \in \mathbb{N}. \tag{1}$$

The probability distribution given by (1) is a special case with $\alpha = 1$ and $\sigma = 1$ of *discrete Pareto distribution*, which in general has the probability mass function (PMF)

$$\mathbb{P}(N = n) = \left(\frac{1}{1+\frac{n-1}{\sigma}}\right)^{\alpha} - \left(\frac{1}{1+\frac{n}{\sigma}}\right)^{\alpha}, \quad n \in \mathbb{N}, \tag{2}$$

and arises by discretization of continuous Pareto type II (Lomax) distribution with tail parameter $\alpha > 0$ and scale parameter $\sigma > 0$ (Krishna and Singh Pundir 2009; Buddana and Kozubowski 2014). The distribution given by (1) is also a special case $\alpha = 1$ of *Yule distribution* (Yule 1925), which in general case $\alpha > 0$ is given by the PMF

$$\mathbb{P}(N = n) = \frac{\alpha\Gamma(\alpha+1)\Gamma(n)}{\Gamma(\alpha+n+1)}, \quad n \in \mathbb{N}.$$

Both Yule and discrete Pareto distributions are *heavy tailed*, with *power law* behavior of their PMFs (and tails),

$$\mathbb{P}(N = n) = O\left(\frac{1}{n^{\alpha+1}}\right) \text{ as } n \to \infty. \tag{3}$$

Along with the Zipf's law, whose PMF has the same asymptotics (see, e.g., Zipf 1949 or Johnson et al. 1993), these distributions provide important modeling tools whenever empirical distributions display power-law tails. Such scaling behavior has been observed across many fields, including biology, chemistry, computer science, economics, finance, geosciences, and social science (see, e.g., Aban et al. 2006; Clauset and Newman 2009; Gabaix 2009; Newman 2005; Sornette 2006; Stumpf and Porter 2012).

In this paper, we study another generalization of (1), which is directly related to its interpretation through the record process described above. Namely, we define a discrete variable N to be the waiting time for the first success in a sequence of independent Bernoulli trials $\{I_j, j \in \mathbb{N}\}$, where the probabilities of success are given by

$$\mathbb{P}(I_j = 1) = \frac{\alpha}{\nu+j}, \quad \nu \geq 0 \quad \text{and} \quad 0 < \alpha < \nu + 1.$$

We observe that the record times correspond to $\alpha = \nu = 1$. If N is the *number of trials* until the first success, then

$$\mathbb{P}(N = n) = \left(1 - \frac{\alpha}{\nu + 1}\right) \cdots \left(1 - \frac{\alpha}{\nu + n - 1}\right) \frac{\alpha}{\nu + n}, \quad n \in \mathbb{N}. \tag{4}$$

It can be shown that, asymptotically, the probabilities (4) are also power laws of the form (3). Moreover, in the special case $\nu = 0$ and $\alpha \in (0, 1)$, we obtain the *Sibuya* distribution with the PMF

$$\mathbb{P}(N = n) = \frac{\alpha(\alpha - 1) \cdots (\alpha - n + 1)}{n!} (-1)^{n+1} = \binom{\alpha}{n} (-1)^{n+1}, \quad n \in \mathbb{N}, \tag{5}$$

which first appeared in Sibuya (1979) and was later studied in connection with discrete stable, Linnik, and Mittag-Leffler distributions (see, e.g., Christoph and Schreiber 1998, 2000; Devroye 1993; Pakes 1995; Pillai and Jayakumar 1995; Satheesh and Nair 2002). Due to this connection, we name the distribution with the PMF (4) *generalized Sibuya*.

The main goal of the paper is to account for basic properties of the generalized Sibuya distribution (4) and to show how it interrelates with its special case of Sibuya distribution (5). Additionally, we wish to emphasize the importance of the Sibuya distribution in distribution theory and provide its new characterization which goes beyond the class of generalized Sibuya variables. To this end, let us first comment on the importance of the Sibuya model in the extreme value theory. It is well known that, for any $n \in \mathbb{N}$, the quantity $[F(x)]^n$, where F is a cumulative distribution function (CDF), is also the CDF corresponding to the random variable

$$X = \max\{X_1, \dots, X_n\} = \bigvee_{j=1}^{n} X_j, \tag{6}$$

where the $\{X_j\}$ are IID with the CDF F. For a non-integer exponent $\alpha > 0$, the quantity $[F(x)]^\alpha$ is also a CDF, although in this case the relation (6) is no longer interpretable. Similarly, the quantity $[S(x)]^n$, where $S(x) = 1 - F(x)$, is the survival function (SF) corresponding to

$$Y = \min\{X_1, \dots, X_n\} = \bigwedge_{j=1}^{n} X_j,$$

although $[S(x)]^\alpha$, again a valid SF, lacks such an interpretation for fractional $\alpha > 0$. It turns out that the Sibuya distribution (5) provides a missing link, allowing an interpretation through *stochastic* maxima and minima as presented in the following result.

Proposition 1 *Let F be a CDF on $\mathbb{R}$ and S be the corresponding survival function, $S(x) = 1 - F(x)$. Further, let the distribution of X be given by the SF $[S(x)]^\alpha$ and the distribution of Y by the CDF $[F(x)]^\alpha$, where $\alpha \in (0, 1]$. Then X and Y admit the stochastic representations*

$$X \stackrel{d}{=} \bigvee_{j=1}^{N_\alpha} X_j \quad and \quad Y \stackrel{d}{=} \bigwedge_{j=1}^{N_\alpha} X_j, \tag{7}$$

where N_α has the Sibuya distribution (5) and is independent of the sequence $\{X_j\}$ of IID random variables distributed according to the CDF F.

The proof of this result can be found in Kozubowski and Podgórski (2016), where one can find additional information on random maxima and minima with Sibuya distributed number of terms as well as generalizations to random processes. The argument used there is based on somewhat more general properties of the maxima of a random number of variables. Thus, to show a more explicit argument and for the sake of completeness we provide a direct proof in "Appendix."

As we shall see in Sect. 5, one can define a pure jump random process with Sibuya marginal distributions. The laws of the jumps are related to the generalized Sibuya distribution. In particular, the size of the first jump of this process has the generalized Sibuya distribution (4) with $\nu = 1$. Such relations between the Sibuya and generalized Sibuya distributions, along with the importance of the former, provide additional motivation for studying the latter.

Let us finally provide yet another result on the Sibuya distribution, which appears to be new. It relates to the theory of birth/death Markov processes. Consider a sequence $\{X_i\}$, $i \in \mathbb{N}$, of IID random variables having continuous distribution on $\mathbb{R}_+ = (0, \infty)$. Suppose that at time $t = 0$, a population consists of a random number $N \in \mathbb{N}$ of individuals, whose future lifetimes are given by X_i, $i = 1, \ldots, N$. Then

$$N(t) = \sum_{i=1}^{N} I_{(t,\infty)}(X_i), \quad t \geq 0, \tag{8}$$

is a pure death process, describing the number of individuals alive at time t (the quantity I_A is an indicator of the set A). It turns out that if N has the Sibuya distribution (5) with some $\alpha \in (0, 1)$, then, regardless of the choice for the distribution of the $\{X_i\}$, the conditional distribution of $N(t)|N(t) > 0$ is the same as that of N. In other words, the Sibuya distribution provides a *stationary* conditional distribution of $N(t)$ for each $t \in [0, \infty)$: if it is known that the population is still alive at time $t > 0$, its size is described by *the same* Sibuya distribution. This in fact is a characterization of the Sibuya distribution, as stated in the following result, which is proven in "Appendix."

Proposition 2 *Let $\{X_i\}$, $i \in \mathbb{N}$, be a sequence of IID random variables having continuous distribution on $\mathbb{R}_+ = (0, \infty)$, and let N be a random variable on $\mathbb{N}$, independent of the $\{X_i\}$. Then N has a Sibuya distribution (5) with some $\alpha \in (0, 1)$ if and only if for each $t \in [0, \infty)$ we have the equality in distribution*

$$N \stackrel{d}{=} N(t)|N(t) > 0, \tag{9}$$

where $N(t)$ is a pure death process defined by (8).

The rest of the paper is a careful account of the properties of the generalized Sibuya model. We start with Sect. 2, where we introduce the model and derive its basic characteristics. Various stochastic representations of the model appear in Sect. 3. They are followed by account of divisibility properties in Sect. 4. In Sect. 5, we define a Sibuya random process on [0, 1] and study the structure of its sample paths. Statistical issues are briefly treated in Sect. 6. We conclude with "Appendix," containing (selected) proofs and auxiliary results.

2 Definition and basic properties

We begin with the definition of the generalized Sibuya stochastic model.

Definition 1 A random variable N with the PMF (4) is said to have a generalized Sibuya distribution with parameters $\alpha \in \mathbb{R}_+$ and $\nu \geq 0$, denoted by $GS_1(\alpha, \nu)$. The two parameters are restricted by the relation $0 < \alpha < \nu + 1$.

The subscript in the notation indicates that the distribution is supported on the set $\mathbb{N}$ of positive integers. Another version of this distribution, which is defined as the *number of failures* before the first success, shall be denoted by $GS_0(\alpha, \nu)$, i.e., for $M = N - 1$:

$$N \sim GS_1(\alpha, \nu) \quad \text{if and only if} \quad M \sim GS_0(\alpha, \nu). \tag{10}$$

The properties provided in the sequel shall be stated in terms of either one of the two distributions and can be easily re-formulated in terms of the other if needed.

2.1 Special cases

Note that at the boundary of the parameter space, where $0 < \alpha = \nu+1$, the distribution collapses to a point mass at 1. This exceptional case shall be omitted from most considerations. The Sibuya distribution (5) arises as a special case of $GS_1(\alpha, \nu)$ with $\alpha \in (0, 1)$ and $\nu = 0$. This distribution is often described through its probability generating function (PGF), which, compared with the general case discussed in the sequel, is of a particularly simple form:

$$G_N(s) = \sum_{n=1}^{\infty} \binom{\alpha}{n} (-1)^{n+1} s^n = 1 - (1 - s)^{\alpha}, \quad 0 < s < 1. \tag{11}$$

In the further special case $\alpha = 1/2$, we have that $GS_0(1/2, 0)$ is a *discrete Mittag-Leffler* distribution with the PGF $G(s) = [1 + (1 - s)^{\alpha}]^{-1}$ (see, e.g., Pillai and Jayakumar 1995).

We have already noted the special case $\alpha = \nu = 1$ of the $GS_1(\alpha, \nu)$ distribution, where the PMF simplifies to (1) and we obtain a particular case of the discrete Pareto and the Yule distributions. For $\alpha = 1$ and general $\nu > 0$, the generalized Sibuya PMF (4) becomes

$$\mathbb{P}(N=n)=\left(1-\frac{1}{\nu+1}\right)\cdots\left(1-\frac{1}{\nu+n-1}\right)\frac{1}{\nu+n}=\frac{\nu}{\nu+n-1}-\frac{\nu}{\nu+n},$$

and we also obtain a case of discrete Pareto distribution (2) with $\alpha=1$ and $\sigma=\nu$.

Remark 1 Let us note that our distribution is different than that studied in Huillet (2016), who used the name "generalized Sibuya" to denote a three-parameter family given by the PGF $\phi_{\alpha,\beta,\lambda}(s)=(1-\lambda(1-s)^{\alpha})^{\beta}$. This can be see from the form of the PGF of our generalization of the Sibuya distribution that is presented in (6) in Sect. 2.4. The name "positive generalized Sibuya" was also used in Huillet (2012) to denote a random variable given by the *Laplace transform* $\phi(t)=1-\gamma\log(1+t^{\delta}/\gamma)$.

2.2 Distribution and survival functions

The CDF and the SF of a generalized Sibuya random variable $N\sim GS_1(\alpha,\nu)$ are straightforward to derive. Indeed, the SF for any $n\in\mathbb{N}$ is given by

$$\mathbb{P}(N>n)=\mathbb{P}(I_j=0,\ j=1,\ldots,n)=\left(1-\frac{\alpha}{\nu+1}\right)\cdots\left(1-\frac{\alpha}{\nu+n}\right). \tag{12}$$

It follows that the SF and the PMF of $N\sim GS_1(\alpha,\nu)$ are linked as follows:

$$\mathbb{P}(N=n)=\frac{\alpha}{n+\nu-\alpha}\mathbb{P}(N>n),\quad n\in\mathbb{N}. \tag{13}$$

We now consider the conditional distribution of $N-m$ given $N>m$, $m\in\mathbb{N}_0=\mathbb{N}\cup\{0\}$. Straightforward algebra incorporating the above results shows that

$$\begin{aligned}&\mathbb{P}(N-m=n|N>m)\\&\quad=\left(1-\frac{\alpha}{\nu+m+1}\right)\cdots\left(1-\frac{\alpha}{\nu+m+n-1}\right)\frac{\alpha}{\nu+m+n},\quad n\in\mathbb{N}.\end{aligned}$$

The above is recognized as a generalized Sibuya probability as well, with parameters α and $\nu+m$. In particular, if N has Sibuya distribution (5), i.e., $\nu=0$, then the corresponding excess $N-m$ conditionally on $N>m$ is generalized Sibuya $GS_1(\alpha,m)$. Thus, the class of generalized Sibuya distributions is closed with respect to the operation of taking the excess, as summarized in the result below.

Proposition 3 *If* $m\in\mathbb{N}_0$ *and* $N\sim GS_1(\alpha,\nu)$*, then* $N-m|N>m\sim GS_1(\alpha,\nu+m)$.

2.3 Moments and tail behavior

As shown in Christoph and Schreiber (2000), the Sibuya probabilities (5) admit the asymptotic representation

$$\mathbb{P}(N=n)\sim\frac{1}{\pi}\sin(\alpha\pi)\Gamma(1+\alpha)\frac{1}{n^{\alpha+1}}\ \text{ as } n\to\infty, \tag{14}$$

where $f(x) \sim g(x)$ means that $f(x)/g(x) \to 1$ as $x \to \infty$. Thus, if $N \sim G_1(\alpha, 0)$ (ordinary Sibuya), where necessarily $\alpha \in (0, 1)$, then we have (3). As shown below, the latter asymptotic relation holds for the generalized Sibuya distribution as well.

Proposition 4 *If $N \sim GS_1(\alpha, \nu)$, then*

$$\mathbb{P}(N = n) \sim \frac{\Gamma(\nu+1)}{\Gamma(\nu+1-\alpha)} \frac{\alpha}{n^{\alpha+1}} \quad \text{as } n \to \infty. \tag{15}$$

Remark 2 Note that if we set $\nu = 0$ in (15) and use two well-known properties of the gamma function,

$$\Gamma(\alpha)\Gamma(1-\alpha) = \frac{\pi}{\sin(\alpha\pi)}, \quad \Gamma(1+\alpha) = \alpha\Gamma(\alpha),$$

then we recover (14).

In view of the link (13) between the generalized Sibuya survival function and its probabilities, the above result immediately provides the asymptotics of the tail, stated below.

Corollary 1 *If $N \sim GS_1(\alpha, \nu)$, then*

$$\mathbb{P}(N > n) \sim \frac{\Gamma(\nu+1)}{\Gamma(\nu+1-\alpha)} \frac{1}{n^{\alpha}} \quad \text{as } n \to \infty. \tag{16}$$

Because of the power-law asymptotics (16) of its tail, the moments of order α and above of the generalized Sibuya distribution do not exist.

Corollary 2 *Let $\gamma \in \mathbb{R}_+$. If $N \sim GS_1(\alpha, \nu)$, then $\mathbb{E}N^{\gamma} < \infty$ if and only if $\gamma \in (0, \alpha)$.*

In particular, the expectation of $N \sim GS_1(\alpha, \nu)$ exists whenever $\alpha > 1$ (so that necessarily $\nu > 0$), while the variance exists if and only if $\alpha > 2$ (so that $\nu > 1$). Perhaps the most convenient way to obtain these, along with other moments, is through the stochastic representations of Sect. 3. For example, an application of Proposition 7, along with tower property for conditional expectations, leads to

$$\mathbb{E}N^{\delta} = \mathbb{E}f_{\delta}(X), \quad \delta \in \mathbb{R}_+, \tag{17}$$

where $N \sim GS_0(\alpha, \nu)$, X is a random variable defined in (27), and $f_{\delta}(t) = \mathbb{E}\{[N(t)]^{\delta}\}$, with $\{N(t),\ t \geq 0\}$ being a standard Poisson process, independent of X. This is useful mainly for integer values of δ, as fractional moments of Poisson distribution are not available in close forms. Indeed, we have the following well-known Touchard polynomial representation for the Poisson raw moments of an integer order:

$$\mathbb{E}\{[N(t)]^n\} = \sum_{i=1}^{n} t^i \left\{ \begin{matrix} n \\ i \end{matrix} \right\}, \quad n \in \mathbb{N}, t \in \mathbb{R}_+, \tag{18}$$

where

$$\left\{ \begin{matrix} n \\ i \end{matrix} \right\} = \frac{1}{i!} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} j^n \tag{19}$$

are the Stirling numbers of the second kind (representing the number of ways to partition a set of n objects into i non-empty subsets). Straightforward calculations, which take into account relations (17) and (19), lead to the following result.

Proposition 5 *Let $N \sim GS_0(\alpha, \nu)$ and $n \in \mathbb{N}$, where $n < \alpha < \nu + 1$. Then*

$$\mathbb{E}N^n = \sum_{i=1}^{n} e_i(\alpha, \nu) \left\{ \begin{matrix} n \\ i \end{matrix} \right\},$$

where

$$e_i(\alpha, \nu) = \frac{\Gamma(\alpha - i)\Gamma(i+1)\Gamma(i + \nu - \alpha + 1)}{\Gamma(\alpha)\Gamma(\nu - \alpha + 1)}, \quad i < \alpha.$$

This above result leads to explicit expressions for the classical moment characteristics of the generalized Sibuya distribution, such as the mean, the variance, the coefficient of skewness

$$\gamma_N = \mathbb{E}\left(\frac{N - \mathbb{E}N}{\sigma_N}\right)^3 = \frac{\mathbb{E}N^3 - 3(\mathbb{E}N)(\mathbb{E}N^2) + 2(\mathbb{E}N)^3}{(\mathbb{V}ar N)^{3/2}}, \tag{20}$$

and (excess) kurtosis

$$\kappa_N = \mathbb{E}\left(\frac{N - \mathbb{E}N}{\sigma_N}\right)^4 - 3 = \frac{\mathbb{E}N^4 - 4(\mathbb{E}N)(\mathbb{E}N^3) + 6(\mathbb{E}N)^2(\mathbb{E}N^2) - 3(\mathbb{E}N)^4}{(\mathbb{V}ar N)^2} - 3. \tag{21}$$

These are summarized in the following result, whose straightforward albeit tedious derivation shall be omitted.

Corollary 3 *Let $N \sim GS_0(\alpha, \nu)$. The mean and the variance of N exist whenever $\alpha > 1$ and $\alpha > 2$, respectively, in which case we have*

$$\mathbb{E}N = \frac{\nu - \alpha + 1}{\alpha - 1}, \qquad \mathbb{V}ar N = \frac{\nu - \alpha + 1}{(\alpha - 1)^2} \frac{\alpha \nu}{\alpha - 2}. \tag{22}$$

Further, the coefficient of skewness (20) exists whenever $\alpha > 3$, in which case we have

$$\gamma_N = \sqrt{\frac{\alpha - 2}{\alpha}} \frac{\alpha + 1}{\alpha - 3} \left(\sqrt{\frac{\nu}{1 - \alpha + \nu}} + \sqrt{\frac{1 - \alpha + \nu}{\nu}} \right).$$

Finally, the (excess) kurtosis (21) exists whenever $\alpha > 4$, in which case we have

$$\kappa_N = \frac{\alpha - 2}{\alpha(\alpha - 3)(\alpha - 4)} \left(6(\alpha^3 + \alpha^2 - 6\alpha - 2) + \frac{(\alpha + 5)\alpha(\alpha - 1)^2(\alpha - 2)}{(\nu - \alpha + 1)\nu} \right).$$

Remark 3 We note that the expectation of generalized Sibuya distribution is straightforward, as it does not involve any special functions or infinite series, which is not the case with discrete Pareto distribution (2), which has the same asymptotics of the tail. For example, the expectation of $N \sim GS_0(\alpha, \nu)$ with $\nu = 1$ and $\alpha = 1 + p$, where $0 < p < 1$, is equal to $\mathbb{E}N = (1 - p)/p$ and coincides with that of geometric variable with parameter p.

Remark 4 The mean of $N \sim GS_1(\alpha, \nu)$ also exists whenever $\alpha > 1$, in which case we have $\mathbb{E}N = \nu/(\alpha - 1)$. When $\alpha > 2$, its variance exists as well and coincides with the expression in (22).

Remark 5 According to the above result, for $2 < \alpha < \nu + 1$, we have

$$\frac{\mathbb{E}N}{\mathbb{V}ar N} = \frac{\alpha - 1}{\nu}\frac{\alpha - 2}{\alpha} < 1, \tag{23}$$

so that every generalized Sibuya distribution is over-dispersed (its variance is larger than the mean). Moreover, the ratio (23) is monotonically decreasing (to zero) as $\nu \to \infty$ (so the distribution becomes more and more over-dispersed) and it is monotonically increasing in α (so the distributions becomes less and less over-dispersed as α gets larger) with the limits of 0 and $(\nu - 1)/(\nu + 1)$ as $\alpha \to 2^+$ and $\alpha \to (\nu + 1)^-$, respectively.

Remark 6 The skewness coefficient γ_N takes on only positive values and is a decreasing function of ν with the limit of $2(\alpha + 1)\sqrt{\alpha - 2}/[(\alpha - 3)\sqrt{\alpha}]$ as $\nu \to \infty$. Additionally, it is straightforward to see that this limiting value is a decreasing function of α on $(3, \infty)$ as well, with the limiting value of 2 at infinity. The latter provides a lower bound for the skewness. On the other hand, γ_N is not a monotonic function of α, as can be seen by checking its derivative with respect to α. Further, it can be shown that when α is close to its lower boundary of 3, then γ_N is decreasing, while it is increasing when α is close to its upper boundary of $\nu + 1$, and the limiting values of γ_N at these boundaries are both ∞.

Remark 7 The excess kurtosis coefficient κ_N is a decreasing function of ν with the limit of $\frac{6(\alpha-2)(\alpha^3+\alpha^2-6\alpha-2)}{\alpha(\alpha-3)(\alpha-4)}$ as $\nu \to \infty$. Additionally, it is straightforward to see that this limiting value is increasing to infinity for large α $(\alpha > 9)$.

2.4 The probability generating function

The probability generating function of generalized Sibuya distribution can be obtained via the mixed Poisson representation (30), coupled with the relation (33). The following result provides relevant details.

Proposition 6 *If $N \sim GS_0(\alpha, \nu)$, then the PGF of N is given by*

$$G_N(s) = \frac{\alpha}{\nu + 1}(1 - s)^\alpha F(\nu + 1, \alpha + 1; \nu + 2; s), \quad 0 < s < 1, \tag{24}$$

where

$$F(a,b;c;t)=\sum_{n=0}^{\infty}\frac{(a)_n\,(b)_n}{(c)_n}\,\frac{t^n}{n!},\ \ |t|<1, \tag{25}$$

is the Gauss hypergeometric function and

$$(x)_n=\begin{cases} x(x+1)\cdots(x+n-1) & \text{for } n\geq 1\\ 1 & \text{for } n=0\end{cases}$$

is the (rising) Pochhammer symbol.

Remark 8 If $N\sim GS_1(\alpha,\nu)$, then, due to the relation (10), its PGF is given by (24) multiplied by s. In particular, when $\nu=0$, we obtain the PGF (11) of Sibuya distribution with the PMF (5).

Remark 9 In the case when $\nu=k\in\mathbb{N}_0$, we get the explicit form

$$G_N(t)=\frac{k!}{\Gamma(\alpha)\Gamma(1-\alpha+k)}\frac{\pi}{\sin(\pi\alpha)}\frac{1}{s^{1+k}}\left\{\sum_{j=0}^{k}\frac{s^j(-\alpha)_j}{j!}-(1-s)^{\alpha}\right\},\ \ 0<s<1, \tag{26}$$

which can be obtained similarly to the general result using the integration formula **3.228.6** on p. 321 of Gradshteyn and Ryzhik 2007. Further, let us note that for integer values of α, the quantity $\sin(\pi\alpha)$ in the denominator of the right-hand side of (26) becomes zero. In this case, the expression for the PGF is understood in the limiting sense. For example, by taking the limit as $\alpha\to 1$ of the right-hand side of (26) with $k=1$, we find that the PGF of $GS_0(1,1)$ distribution is given by

$$G_N(s)=\frac{s+(1-s)\log(1-s)}{s},\ \ 0<s<1.$$

Remark 10 When $\alpha>1$, the expectation of generalized Sibuya distribution can also be computed via the relation

$$\mathbb{E}N=\left.\frac{d}{ds}G_N(s)\right|_{s=1}.$$

However, this is not a convenient way of getting the mean. For example, for non-integer values of $\alpha>1$ and $\nu=k\in\mathbb{N}_0$, this leads the expression

$$\mathbb{E}N=\frac{k!}{\Gamma(\alpha)\Gamma(1-\alpha+k)}\frac{\pi}{\sin(\pi\alpha)}\left\{\sum_{j=0}^{k-1}\frac{(-\alpha)_{j+1}}{j!}-(k+1)\sum_{j=0}^{k}\frac{(-\alpha)_j}{j!}\right\},$$

which is not immediately seen to coincide with the formula in (22).

3 Stochastic representations

Below we provide an account of several stochastic representations of generalized Sibuya random variables, involving a randomly stopped Poisson process, mixtures of geometric distributions, and a discretization scheme.

3.1 Randomly stopped Poisson process

Consider a random variable

$$X \stackrel{d}{=} \frac{E}{T_{\alpha,\nu}}, \tag{27}$$

where E and $T_{\alpha,\nu}$ are independent, E is standard exponential, and $T_{\alpha,\nu}$ has a beta distribution of the second kind, given by the PDF

$$f(x) = \frac{1}{B(\alpha, \nu-\alpha+1)} \frac{x^{\alpha-1}}{(1+x)^{\nu+1}}, \quad x \in \mathbb{R}_+ \ \ (\nu \geq 0, \ \ 0 < \alpha < \nu+1), \tag{28}$$

where

$$B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b) \tag{29}$$

stands for the beta function.

Proposition 7 *If $N \sim GS_0(\alpha, \nu)$, then*

$$N \stackrel{d}{=} N(X), \tag{30}$$

where X is given by (27) and is independent of a standard Poisson process $\{N(t), t > 0\}$.

Remark 11 Note that, by general results concerning the ratio of two independent standard gamma variables (see, e.g., Johnson et al. 1994, p. 351), $T_{\alpha,\nu}$ admits the stochastic representation

$$T_{\alpha,\nu} \stackrel{d}{=} \frac{X_\alpha}{X_{1-\alpha+\nu}}, \tag{31}$$

where the variables on the right-hand side of (31) are independent and X_β denotes standard gamma variable with shape parameter β (and unit scale). This shows that generalized Sibuya distribution is a special case $a = 1, b = 1-\alpha+\nu, c = \alpha$ of the *generalized hyperbolic distribution* of type B3 with the PMF (64), defined via the stochastic representation (30) with

$$X \stackrel{d}{=} \frac{X_a X_b}{X_c}, \tag{32}$$

where the three variables on the right-hand side of (32) are independent and have standard gamma distributions (see, e.g., Sibuya 1979; Sibuya and Shimizu 1981, or Devroye 1993).

Remark 12 The result of Proposition 7 in case of (shifted) Sibuya distribution $GS_0(\alpha, 0)$ was noted in Devroye (1993) in connection with the problem of random variate generation from this distribution. Note that if $N \sim GS_1(\alpha, 0)$, then N does not admit the representation (30) with any X; instead, in this case we have $N \stackrel{d}{=} 1 + N(X)$.

Remark 13 It can be easily seen that if N admits the stochastic representation (30), then the PGF of N must be of the form

$$G_N(s) = \mathbb{E}s^N = \phi_X(1-s), \quad s \in (0, 1), \tag{33}$$

where $\phi_X(\cdot)$ is the Laplace transform (LT) of X (see, e.g., Steutel and Harn 2004, p. 367). This allows for a derivation of one of the functions, $G_N(\cdot)$ or $\phi_X(\cdot)$, from the other one. It can be shown (see the proof of Proposition 6 in "Appendix") that the Laplace transform of X defined by (27) is of the form

$$\phi_X(t) = \frac{\alpha}{\nu+1} t^\alpha F(\nu+1, \alpha+1; \nu+2; 1-t), \quad t \in \mathbb{R}_+, \tag{34}$$

where for $t \geq 2$ the Gauss hypergeometric function F is defined through the analytic extension of its defining series (25). For the case of integer $\nu = k \in \mathbb{N}_0$, we have

$$\phi_X(t) = \frac{k!}{\Gamma(\alpha)\Gamma(1-\alpha+k)} \frac{\pi}{\sin(\pi\alpha)} \frac{1}{(1-t)^{1+k}} \left\{ \sum_{j=0}^{k} \frac{(-\alpha)_j (1-t)^j}{j!} - t^\alpha \right\}, \quad t \in \mathbb{R}_+. \tag{35}$$

The PDF of X can be written as

$$f_X(x) = \frac{\Gamma(\nu+1)}{\Gamma(\alpha)\Gamma(1-\alpha+\nu)} \int_0^\infty e^{-\nu x} \frac{t^\alpha}{(1+t)^{\nu+1}} \mathrm{d}t, \quad x \in \mathbb{R}_+ \ (\nu \geq 0, \ 0 < \alpha < \nu+1).$$

Relations (33–35) lead to the PGF of generalized Sibuya distribution $GS_0(\alpha, \nu)$, given in Proposition 6. In case of (shifted) Sibuya distribution $GS_0(\alpha, 0)$, the function (35) reduces to

$$\phi_X(t) = \frac{1-t^\alpha}{1-t}, \quad t \in \mathbb{R}_+,$$

which can also be recovered from the PGF of $N \sim GS_0(\alpha, 0)$ via $\phi_X(t) = G_N(1-t)$. However, if $N \sim GS_1(\alpha, 0)$ with the PGF (11), then $G_N(1-t)$ does not lead to a valid Laplace transform, as noted by Satheesh and Nair (2002).

3.2 Randomly mixed geometric variable

Our second representation shows that a generalized Sibuya distribution can be thought of as a mixed geometric distribution. The result below, which follows from the theory of generalized hypergeometric distributions of type B3 (see, e.g, Sibuya 1979; Sibuya and Shimizu 1981), can be proven directly from the representation (30) and standard conditioning arguments.

Proposition 8 *Let Y have a beta distribution with parameters α and $\beta = 1 - \alpha + \nu$, where $\nu \geq 0$ and $0 < \alpha < \nu + 1$. Further, assume that, conditionally on $Y = p$, N has a geometric distribution with parameter p, i.e.,*

$$\mathbb{P}(N = n | Y = p) = p(1-p)^n, \quad n \in \mathbb{N}_0.$$

Then, unconditionally, $N \sim GS_0(\alpha, \nu)$.

Remark 14 The $GS_1(\alpha, \nu)$ version of generalized Sibuya distribution is also mixed geometric with the same stochastic probability of success, but with a shifted-by-one version of the geometric variable.

3.3 Discretization scheme

A generalized Sibuya variable arises also by a discretization scheme of the form $N = [W]$, where a discrete counterpart of a continuously distributed W is the integer part of W. A discrete counterpart of exponential distribution in this scheme is a geometric variable, while discretization of continuous Pareto II (Lomax distribution) leads to discrete Pareto distribution (see, e.g., Buddana and Kozubowski 2014).

Proposition 9 *If W is a mixed exponential variable of the form*

$$W \stackrel{d}{=} \frac{E}{V_{\alpha,\nu}},$$

where E and $V_{\alpha,\nu}$ are independent, E is standard exponential, and $V_{\alpha,\nu}$ has the PDF

$$g(x) = \frac{\Gamma(\nu+1)e^{-\nu x}}{\Gamma(\alpha)\Gamma(1-\alpha+\nu)}(e^x - 1)^{\alpha-1}, \quad x \in \mathbb{R}_+ \ (\nu \geq 0, \ 0 < \alpha < \nu + 1), \tag{36}$$

then $N = [W] \sim GS_0(\alpha, \nu)$.

4 Divisibility properties

4.1 Infinite divisibility

Recall that a random variable X (and its distribution) is *infinitely divisible* (ID) if for each $n \in \mathbb{N}$ it can be decomposed into the sum

$$X \stackrel{d}{=} X_{n,1} + \cdots + X_{n,n} \tag{37}$$

of IID random variables $\{X_{n,j}\}$ $(1 \leq j \leq n)$. Further, an integer-valued random variable X supported on $\mathbb{N}_0$ is *discrete infinitely divisible* if it is ID and the variables $\{X_{n,j}\}$ in (37) are integer-valued and supported on $\mathbb{N}_0$ as well. It is well known that (shifted) Sibuya distribution $GS_0(\alpha, 0)$ is discrete ID (see, e.g., Christoph and Schreiber 2000), implying that Sibuya distribution $GS_1(\alpha, 0)$ is ID (but not discrete ID). Similar properties hold for generalized Sibuya distribution and follow from their representations as mixtures of geometric distributions, as the latter are ID (see, e.g., Steutel and Harn 2004, Theorem 7.8, p. 381). The following result summarizes these facts.

Proposition 10 *If $N \sim GS_0(\alpha, \nu)$, then the distribution of N is discrete ID (and thus ID). Further, the distribution of $N + 1 \sim GS_1(\alpha, \nu)$ is ID (but not discrete ID).*

This property allows us to build a continuous-time discrete-value stochastic processes based on the generalized Sibuya distribution. In particular, we can define a *Lévy motion* $\{N(t),\quad t > 0\}$, a process with stationary, independent increments, where $N(1)$ is $GS_0(\alpha, \nu)$ with PGF G given by (24), and, for each $t > 0$, the PGF of $N(t)$ is G^t.

4.2 Self-decomposability

A discrete-valued random variable N supported on $\mathbb{N}_0$ is discrete self-decomposable (DSD) if for each $c \in (0, 1)$ it can be decomposed as

$$N \stackrel{d}{=} c \odot N + N_c, \tag{38}$$

where the variable N_c is also discrete-valued and supported on $\mathbb{N}_0$, and is independent of $c \odot N$ (see, e.g., Steutel and Harn 1979). The dot product $c \odot N$ is the *discrete multiplication* (also known as *thinning*), defined as

$$c \odot N \stackrel{d}{=} \sum_{j=1}^{N} I_j, \quad c \in (0, 1), \tag{39}$$

where the $\{I_j\}$ are IID Bernoulli variables with parameter c, independent of N. In terms of the PGFs, the condition (38) can be stated as

$$G_N(s) = G_N(1 - c + cs)G_c(s), \quad s \in (0, 1), \tag{40}$$

where G_N is the PGF of N, $G_N(1 - c + cs)$ is the PGF of the dot product (39), and G_c is the PGF of N_c. It was shown by Christoph and Schreiber (2000) that the (shifted) Sibuya distribution $GS_0(\alpha, 0)$ is DSD for each $\alpha \in (0, 1)$. The following result provides an extension to the generalized Sibuya case.

Proposition 11 *If $N \sim GS_0(\alpha, \nu)$, then the distribution of N is discrete self-decomposable.*

Remark 15 Let us note that if $N \sim GS_1(\alpha, \nu)$ then N is not DSD, since $\mathbb{P}(N = 0) = 0$. In particular, Sibuya distribution (5) is not DSD. However, for $c \in (0, 1)$, the *scaled* Sibuya variable

$$N^{(c)} \stackrel{d}{=} c \odot N, \tag{41}$$

where $N \sim GS_1(\alpha, 0)$, may be DSD, depending on the value of c. Indeed, as shown in Christoph and Schreiber (2000), the variable (41) is DSD if and only if

$$0 < c \leq \left(\frac{1-\alpha}{1+\alpha}\right)^{1/\alpha}.$$

Moreover in the same work, it is also shown that $N^{(c)}$ is (discrete) infinitely divisible if and only if $0 < c \leq (1-\alpha)^{1/\alpha}$.

4.3 Invariance properties

In this section, we present an important new characterization of the Sibuya distribution, which is connected with the thinning operation (39) and (partially) explains the characterization of this distribution stated in Proposition 2. Let N have Sibuya distribution $GS_1(\alpha, 0)$, given by the PMF (5). As observed by several authors (see, e.g., Christoph and Schreiber 2000), the probability distribution corresponding to the scaled Sibuya variable $N^{(c)}$, defined by (41), is a mixture of a point mass at zero (with probability $1 - c^\alpha$) and the original distribution of N (with probability c^α). In other words, we can write

$$c \odot N \stackrel{d}{=} I^{(c)} \cdot N, \quad c \in (0, 1), \tag{42}$$

where $I^{(c)}$ is a Bernoulli random variable with parameter $p_c = c^\alpha$, independent of N. A natural question is whether the property (42) is unique to Sibuya distribution, that is whether there is any other variable N supported on $\mathbb{N}$ for which we have (42) with *some* $p_c \in (0, 1)$. As shown below, there is no such distribution other than the Sibuya distribution.

Proposition 12 *If a random variable N supported on $\mathbb{N}$ satisfies the relation* (42)*, where $I^{(c)}$ is a Bernoulli random variable with some parameter $p_c \in (0, 1)$, independent of N, then N must have Sibuya distribution $GS_1(\alpha, 0)$ and $p_c = c^\alpha$.*

Observe that whenever we have (42), then for $n \in \mathbb{N}$

$$\mathbb{P}(c \odot N = n | c \odot N > 0) = \frac{\mathbb{P}(I^{(c)} \cdot N = n)}{\mathbb{P}(I^{(c)} \cdot N > 0)} = \frac{\mathbb{P}(I^{(c)} = 1)\mathbb{P}(N = n)}{1 - \mathbb{P}(I^{(c)} = 0)} = \mathbb{P}(N = n), \tag{43}$$

so that

$$c \odot N | c \odot N > 0 \stackrel{d}{=} N, \quad c \in (0, 1). \tag{44}$$

In other words, the distribution of the thinned random variable $c \odot N$, conditioned on being positive, is the same as that of N, regardless of the thinning parameter $c \in (0, 1)$. Note that, for $c \in (0, 1)$ and any integer-valued variable N supported on $\mathbb{N}_0$, we have $\mathbb{P}(c \odot N = 0) = G_N(1 - c) = 1 - \mathbb{P}(c \odot N > 0)$. Thus, if an integer-valued variable N supported on $\mathbb{N}$ satisfies (44), then it also satisfies (42) with

$$p_c = \mathbb{P}(I^{(c)} = 1) = 1 - G_N(1 - c), \quad c \in (0, 1).$$

Thus, in view of Proposition 12, the only distributions that are *stable with respect to the operation of thinning*, in the sense of (44), are Sibuya distributions.

Corollary 4 *Within the class of all probability distributions supported on $\mathbb{N}$, the stability property* (44) *is unique to Sibuya distributions $GS_1(\alpha, 0)$, defined by the PMF* (5).

Let us relate these properties to the characterization of the Sibuya distribution given in Proposition 2. Consider again the pure death process (8), connected with the population of N individuals, whose lifetimes $\{X_j\}$ are IID with the common CDF F. In terms of the operation of thinning, we have

$$N(t) \stackrel{d}{=} c(t) \odot N, \tag{45}$$

where $c(t) = \mathbb{P}(X_j > t) = 1 - F(t)$ is a function on $\mathbb{R}_+$ with the range coinciding with the interval $(0, 1)$. In view of this, the condition (9) is essentially a restatement of (44), which, according to Corollary 4, is known to characterize the Sibuya distribution.

5 A Sibuya random process on [0, 1]

The Sibuya distribution with parameter α less than one arises as the marginal distribution of the *Sibuya random process* that we define as follows. Consider a sequence of IID uniform random variables U_n, $n \in \mathbb{N}$, and set

$$N(t) = \min\{n \in \mathbb{N} : nU_n \leq t\}, \quad t \in [0, 1], \tag{46}$$

with the convention that the minimum over an empty set is infinity, so that $N(0) = \infty$. Since for each $n \in \mathbb{N}$ we have $\mathbb{P}(nU_n \leq t) = t/n$, the variable $N(t)$ has the Sibuya distribution $GS_1(\alpha, 0)$ given by the PMF (5) with $\alpha = t$.

This process can be conveniently described through the classical concept of records. Let $\mathbf{x} = \{x_n\}$, $n \in \mathbb{N}$, be a sequence of positive numbers and consider the pairs

$$(k_i, r_i) = (k_i(\mathbf{x}), r_i(\mathbf{x})), \quad i \in \mathbb{N},$$

where the k_i is the time (index) at which the ith record occurs among the $\{x_i\}$, while $r_i = x_{k_i}$ is the size of that record. Here, a value that is *smaller* than all the previous values sets a new record, and x_1 is also considered to be a record, so that $k_1 = 1$ and $r_1 = x_1$. Further, assume that $x_1 \leq 1$, so that all the $\{r_i\}$ are less than one (while the $\{x_i\}$ are not required to be such). Moreover, let $\delta_i = r_{i-1} - r_i$ (with $r_0 = 1$) represent the differences between successive record values and let $\tau_i = k_i - k_{i-1}$ (with $k_0 = 0$) be the inter-arrival times between successive records. Under this notation, define

$$N_{\mathbf{x}}(t) = 1 + \sum_{i=1}^{\infty} \tau_{i+1} I_{(t,1]}(r_i), \quad t \in [0, 1],$$

where, as before, I_A is an indicator function of the set A.

Clearly, the $N(t)$ defined by (46) is the same as the $N_{\mathbf{x}}(t)$ above if we take $\mathbf{x} = \{nU_n\}$. We see that, looking from right to left, the random process $N(t)$ initially "starts" with the value of one at $t = 1$ and then jumps up at every record value r_i, with the size of the jump being τ_{i+1}. Further, by the definition of the process, the values of $N(t)$ are constant on the intervals $[r_n, r_{n-1})$, and $N(r_n) = k_n$. The following result provides basic properties of the Sibuya random process $\{N(t),\ t \in [0, 1]\}$ discussed above.

Proposition 13 *For each $t \in [0, 1]$, the marginal distribution $N(t)$ is Sibuya given by (5) with $\alpha = t$. Further, $N(t)$ is a right-continuous, pure jump, and non-increasing random process. Moreover, for any $\delta \in (0, 1)$, the number of jumps is finite on the interval $[\delta, 1]$ and infinite on the interval $[0, \delta]$.*

Remark 16 One application of the Sibuya process is a construction of an extremal process on [0, 1] (and beyond) via Proposition 1, as discussed in Kozubowski and Podgórski (2016). For example, if $\{X_n\}$ is a sequence of IID random variables with the common CDF F and we let

$$Y(t) = \bigwedge_{n=1}^{N(t)} X_n, \quad t \in [0, 1],$$

where $N(t)$ is the Sibuya process defined above, independent of the $\{X_n\}$, then the CDF of $Y(t)$ is given by F^t for each $t \in [0, 1]$. This extends the notion of an extreme value of n IID random variables to fractional values of n.

We now look at the sample path structure of the Sibuya process. For convenience, we will look at a time-reversed process $S(t) = N(1-t)$, as it is more natural to follow the evolution of the sample paths from left to right. In Figure 1, we schematically present a part of a sample path of $S(t)$.

By Proposition 13, $S(t)$ is a pure jump process whose sample paths (which start at $S(0) = 1$ almost surely) are continuous from the left and non-decreasing. We already know from the above construction that the jumps of this process and the waiting times between them are closely related to record inter-arrival times and record sizes connected with a random sequence $\{nU_n\}$. Here, the locations of the jumps occur at an increasing sequence $\{T_i\}$, where $T_i = 1 - R_i$ and the R_i is the location of the ith jump of the process $N(t)$ (counted from right to left). Our first result describes the joint distribution of the locations of the jumps.

Proposition 14 *Let $S(t) = N(1-t)$, $t \in [0, 1]$, where $\{N(u),\ u \in [0, 1]\}$ is a Sibuya process defined by (46), and let $\{\Gamma_n\}$ be the successive arrival times of a standard Poisson process. Then for each $n \in \mathbb{N}$ we have*

$$(T_1, \ldots, T_n) \stackrel{d}{=} (H(\Gamma_1), \ldots, H(\Gamma_n)), \tag{47}$$

where $0 < T_1 < \cdots < T_n < 1$ are the (random) locations of the first n jumps of $S(t)$ and H is the CDF of standard exponential distribution.

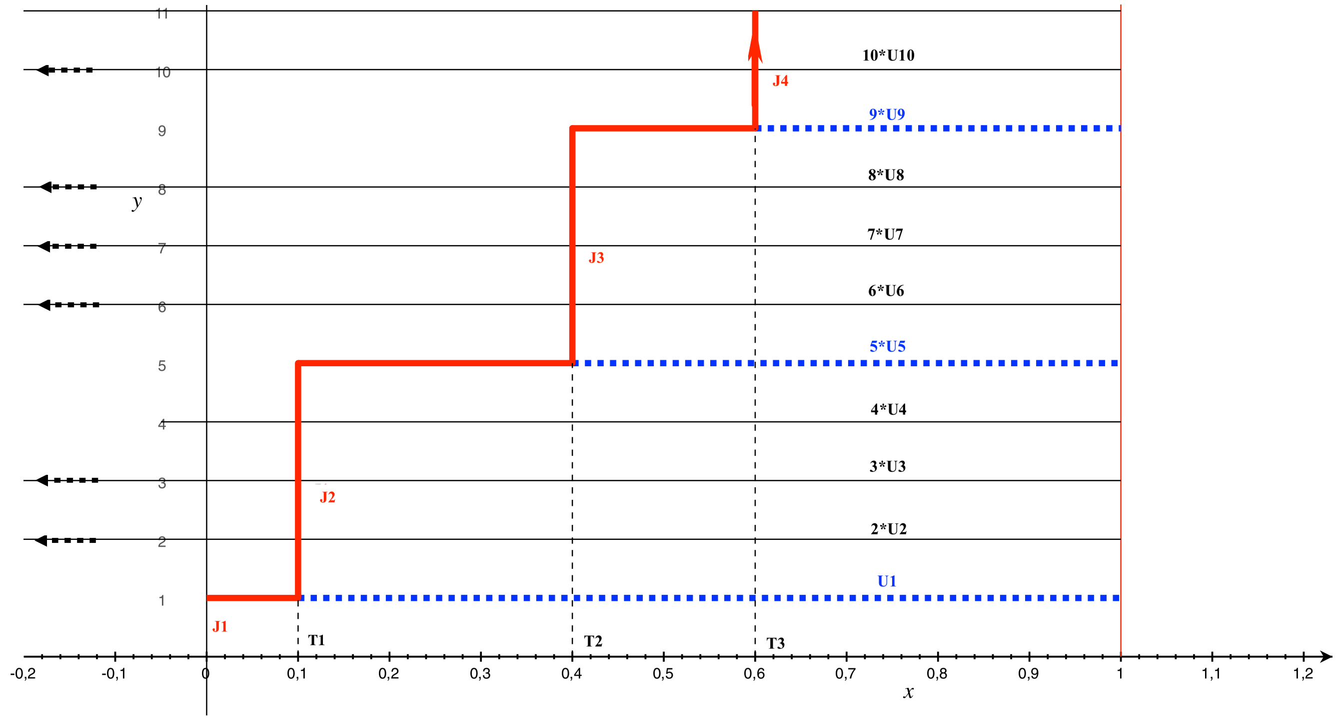

Fig. 1 A schematic sketch of an individual sample of the Sibuya process $S(t)$. The *horizontal lines* represent the values of nU_n, with the *doted lines* marking records at which the process jumps. Three locations (T_1, T_2, T_3) and corresponding jumps (J_1, J_2, J_3) are shown and the trajectory is marked by a *thick line*. The *arrows indicate* that the lengths of lines extend beyond the figure due to large values of nU_n and the jumps size J_4

Remark 17 It follows that the location of the first jump of the process $S(t)$ has a standard uniform distribution, while for $n \geq 2$, the joint distribution of the locations of its first n jumps is given by the PDF (see the proof of Proposition 14 in "Appendix")

$$g(t_1, \ldots, t_n) = \frac{1}{(1-t_1)(1-t_2)\cdots(1-t_{n-1})}, \quad 0 < t_1 < \cdots < t_n < 1. \tag{48}$$

An equivalent description of this is through the conditional distributions: for each $n \in \mathbb{N}$, the conditional distribution of T_n given the $n-1$ values $0 < t_1 < \cdots < t_{n-1} < 1$ of the previous jump locations has a uniform distribution on the interval $(t_{n-1}, 1)$. This distribution is known as *random division of the unit interval*.

Remark 18 It follows that if the time has not been reversed, the jumps of the Sibuya process (46), when viewed from right to left, occur at the points $\exp(-\Gamma_n)$, $n \in \mathbb{N}$. Moreover, if the time line is stretched to $(0, \infty)$ via logarithmic transformation $t \to -\log(1-t)$, the locations of the jumps of $S(t)$ will coincide with those of the standard Poisson process.

Our final result, concerning the joint conditional distribution of the jump sizes and their locations, shades light on the probabilistic structure of the time-reversed Sibuya process. Using the above notation connected with the record process, we shall look at the evolution of the sequence of random points (T_i, K_i), $i \in \mathbb{N}$, where K_i is the time of the ith record connected with the sequence $\{nU_n\}$. As illustrated in Figure 1, $S(t)$ is a pure jump process started at $S(0) = 1$, with the first jump occurring at the random location T_1. Regarding the first random point (T_1, K_1), we have $K_1 = 1$ and, by Proposition 14, the variable T_1 is standard uniform. We now consider the second pair (T_2, K_2), conditioned on the event $\mathcal{B}_1 = \{T_1 = t_1, K_1 = 1\}$, and consider the joint distribution of (T_2, J_2), where $J_2 = K_2 - K_1$ is the size of the jump of $S(t)$ at $t = t_1$. By the construction of the process $S(t)$, for $t_1 < t < 1$ we have

$$\mathbb{P}(T_2 > t, J_2 = n|\mathcal{B}_1) = \mathbb{P}(2U_2 > 1-t_1, \ldots, nU_n > 1-t_1, (n+1)U_{n+1} < 1-t)),$$

so that

$$\mathbb{P}(T_2 > t, J_2 = n|\mathcal{B}_1) = p(r_1, 1, n)\frac{1-t}{1-t_1}, \tag{49}$$

where $r_1 = 1 - t_1$ and

$$p(r, k, n) = \left(1 - \frac{r}{\nu+1}\right)\cdots\left(1 - \frac{r}{\nu+n-1}\right)\frac{r}{\nu+n}, \quad n \in \mathbb{N},$$

represents the probability $\mathbb{P}(S = n)$ with $S \sim GS_1(r, \nu)$. In view of (49) and the fact that the fraction on the right-hand side in (49) is the probability $\mathbb{P}(T_2 > t|T_1 = t_1)$, we conclude that, conditioned on $\mathcal{B}_1$, the variables T_2 and J_2 are independent, with the latter having the generalized Sibuya distribution $GS_1(1 - t_1, 1)$. These calculations extend in a straightforward way beyond the second pair (T_2, J_2), leading to the following result.

Proposition 15 *In the above setting, conditioned on $\mathcal{B}_n = \{T_1 = t_1, \dots, T_n = t_n, K_1 = k_1, \dots K_n = k_n\}$, the variables T_{n+1} and J_{n+1} are independent, with T_{n+1} having uniform distribution on $(t_n, 1)$ and with $J_{n+1} \sim GS_1(1 - t_n, k_n)$.*

According to the above result, the conditional distributions of the jumps of the time-reversed Sibuya process $S(t)$ have generalized Sibuya distributions.

6 Estimation

Although a comprehensive discussion of parameter estimation for the generalized Sibuya distributions is beyond the scope of this work, we still provide several basic results connected with the two common estimation methods: the method of moments and the maximum likelihood method. In doing so, we shall use an alternative parameterization of the generalized Sibuya distribution, where instead of $\alpha > 0$ and $\nu \geq 0$ we have

$$\beta = \frac{1}{\nu + 1} \in [0, 1] \text{ and } \theta = \frac{\alpha}{\nu + 1} \in (0, 1].$$

Here, we have the following special boundary cases: $\theta = 1$ corresponds to the point mass at 1, $\beta = 0$ corresponds to geometric distribution with parameter θ, and $\beta = 1$ yields the original Sibuya distribution. Note that we exclude the case $\theta = 0$, which does not correspond to a valid probability distribution. In the following two sections, the parameter estimation is based on a random sample $X_1, \dots, X_n$ from a generalized Sibuya distribution supported on the set of positive integers $\mathbb{N}$, which in the above parameterization is denoted by $GS_1(\beta, \theta)$. The results easily extend to the case of the shifted distribution $GS_0(\beta, \theta)$ via the relation (10).

6.1 Method of moments

We assume that $\alpha > 2$ or, equivalently $\theta > 2\beta$, so that the first two moments are well defined, and we let $M_1 = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $M_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ be the first two sample moments. To derive the method of moment estimators (MMEs) of the two parameters we set the first two moments of the $GS_1(\beta, \theta)$ distribution (computation of which is aided by Proposition 3) equal to the sample moments $\{M_i\}$, which results in the following system of two equations in two unknowns:

$$\begin{aligned} M_1 &= \frac{\beta - 1}{\beta - \theta}, \\ M_2 &= \frac{(\beta - 1)\ (2\,\beta + \theta - 2)}{2\,\beta^2 - 3\,\beta\,\theta + \theta^2}. \end{aligned} \tag{50}$$

Solving the first equation for β produces

$$\beta = \frac{M_1\theta - 1}{M_1 - 1}, \tag{51}$$

and subsequent substitution of the above into the second equation in (50) leads to the following equation for the parameter θ:

$$\theta[M_2 - 2M_1^2 + M_1 + M_1(M_2 - M_1)] = 2(M_2 - M_1^2). \tag{52}$$

It is not hard to see the equation (52) admits a unique solution in the open interval (0, 1) if the following key condition holds:

$$M_2 - 2M_1^2 + M_1 > 0. \tag{53}$$

Further, the solution is an explicit one, and, along with (51), leads to the following MMEs of the two parameters:

$$\hat{\beta}_n = \frac{M_2 - 2M_1^2 + M_1}{M_2 - 2M_1^2 + M_1 + M_1(M_2 - M_1)}, \quad \hat{\theta}_n = \frac{2(M_2 - M_1^2)}{M_2 - 2M_1^2 + M_1 + M_1(M_2 - M_1)}. \tag{54}$$

The following result summarizes this discussion.

Proposition 16 *Let $X_1, \ldots, X_n$ be a random sample from a $GS_1(\beta, \theta)$ distribution with $\theta > 2\beta$, and let M_1 and M_2 be the first and the second sample moments based on the $\{X_i\}$, respectively. Then, if the inequality (53) is fulfilled, the moment equations (50) admit a unique solution, leading to the MMEs (54).*

Remark 19 When the sample size increases to infinity, then, by law of large numbers, the left-hand side in (53) converges to

$$\mathbb{E}N^2 - 2(\mathbb{E}N)^2 + \mathbb{E}N = \frac{2\beta(1-\beta)(1-\theta)}{(\theta-\beta)^2(\theta-2\beta)},$$

which is larger than 0 when the true parameters are in the interior of the parameter space. Thus, in this case, for large samples the condition (53) is expected to hold. Additionally, let us observe that the MMEs (54) always satisfy the relation $0 < 2\hat{\beta}_n < \hat{\theta}_n < 1$, which is consistent with the assumed existence of the second moment.

Remark 20 In the original parametrization, the MMEs of α and ν also exists and are unique when the condition (53) holds, in which case we have

$$\hat{\alpha}_n = \frac{2(M_2 - M_1^2)}{M_2 - 2M_1^2 + M_1}, \quad \hat{\nu}_n = \frac{M_1(M_2 - M_1)}{M_2 - 2M_1^2 + M_1}.$$

In order to derive the asymptotic behavior of the estimates, they need to be defined regardless of whether or not condition (53) is fulfilled. First, observe that if the quantity on the left-hand side of (53) is zero, then $\hat{\beta}_n$ in (54) becomes zero as well, in which case, according to (51) (as well as (54)), the other estimate becomes $1/M_1$, so that the MMEs are the pair

$$\hat{\beta}_n = 0, \quad \hat{\theta}_n = \frac{1}{M_1}. \tag{55}$$

While in the case of negative value of the expression on the left-hand side of (53) there are no admissible solutions of method of moments equations (50), it is natural to define the MMEs as in (55) in this case as well. In this setup, the asymptotic properties of the MMEs follow from standard (multivariate) delta method, leading to the result below.

Proposition 17 *Let $X_1, \ldots, X_n$ be a random sample from a $GS_1(\beta, \theta)$ distribution with $\theta > 2\beta$, and let M_1 and M_2 be the first and the second sample moments based on the $\{X_i\}$, respectively. Then, the MMEs, defined by (54) if the inequality (53) is fulfilled and by (55) when it is not, are consistent and, whenever the true parameters satisfy the condition $0 < 4\beta < \theta < 1$, asymptotically normal, in which case $\sqrt{n}[(\hat{\beta}_n, \hat{\theta}_n) - (\beta, \theta)]$ converges in distribution to a bivariate normal distribution with the (vector) mean zero and the covariance matrix*

$$\Sigma_{MME} = \frac{\theta\,(\beta-\theta)^2\,(1-2\,\beta)}{24\,\beta^3 - 26\,\beta^2\,\theta + 9\,\beta\,\theta^2 - \theta^3}\begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix},$$

where

$$\begin{aligned} w_{11} &= \frac{\left(2\,\beta^2-3\,\beta+1\right)\left(6\,\beta^2\,\theta-4\,\beta^2-4\,\beta\,\theta^2+3\,\beta\,\theta+\theta^3-\theta^2\right)}{(1-2\,\beta)\,(1-\theta)}, \\ w_{12} &= 6\,\beta^2\,\theta - 8\,\beta^2 - 4\,\beta\,\theta^2 + 4\,\beta\,\theta + \theta^3 - \theta^2, \\ w_{22} &= \frac{\left(\theta^2-3\,\theta+2\right)\left(12\,\beta^3-8\,\beta^2\,\theta-12\,\beta^2+2\,\beta\,\theta^2+6\,\beta\,\theta-\theta^2\right)}{(1-2\,\beta)\,(1-\beta)}. \end{aligned}$$

6.2 Maximum likelihood

Assume that there are $r \geq 1$ distinct values

$$1 \leq j_1 < j_2 < \cdots < j_r < \infty$$

among the $\{X_j\}$, where j_i appears with the frequency $n_i \geq 1$ and $n_1+n_2+\cdots+n_r = n$. Then, the likelihood function is of the form

$$L = L(\beta, \theta) = \prod_{i=1}^{r} p_{j_i}^{n_j}(\beta, \theta),$$

where

$$\begin{aligned} p_n(\beta, \theta) &= \mathbb{P}(X = n) \\ &= (1-\theta)\left(1 - \frac{\theta}{1+\beta}\right)\cdots\left(1 - \frac{\theta}{1+(n-2)\beta}\right)\frac{\theta}{1+(n-1)\beta}, \quad n \in \mathbb{N}. \end{aligned} \tag{56}$$

Observe that L is a rational function in its arguments (a polynomial in θ and a rational function in β), which extends continuously to $[0, 1]^2$ with the value of zero at the boundary $\theta = 0$ and is nonzero otherwise. Consequently, it has a global maximum value over its domain, and, due to its differentiability, both partial derivatives (in θ and in β) at the point where the maximum is reached must be zero as long as the point is in the open set $(0, 1)^2$. Since the behavior on the boundary of the domain is also explicit, we conclude that, while the maximum likelihood estimators (MLEs) are generally not available in closed forms, they can be obtained by fairly standard numerical search, utilizing the explicit form of the first and the second derivatives.

We conclude this section by commenting on the asymptotic optimality properties of the MLEs. For this, we shall assume that the true parameter (β_0, θ_0) is in the interior of $[0, 1]^2$. Due to the infinite differentiability of the likelihood $(\beta, \theta) \mapsto p_n(\beta, \theta)$ in the interior of its domain, one can easily check that the standard conditions for the asymptotic efficiency of the MLEs satisfied (see, e.g., Lehman 1983, Theorem 4.1, p. 429), leading to the following result.

Theorem 1 *Let* $\left(\widehat{\beta}_{MLE}, \widehat{\theta}_{MLE}\right)$ *be the MLEs of* (β_0, θ_0) *based on a random sample from the generalized Sibuya distribution* $GS_1(\beta_0, \theta_0)$*, where* $(\beta_0, \theta_0) \in (0, 1)^2$*. Then,*

$$\sqrt{n}\left[\left(\widehat{\beta}_{MLE}, \widehat{\theta}_{MLE}\right) - (\beta_0, \theta_0)\right]$$

converges in distribution to a bivariate normal distribution with the (vector) mean zero and the covariance matrix $[\mathbf{I}(\beta_0, \theta_0)]^{-1}$*, where*

$$\mathbf{I}(\beta_0, \theta_0) = \begin{bmatrix} \theta_0^{-2} + \sum_{n=1}^{\infty} a_n(\beta_0, \theta_0) p_n(\beta_0, \theta_0) & \sum_{n=1}^{\infty} b_n(\beta_0, \theta_0) p_n(\beta_0, \theta_0) \\ \sum_{n=1}^{\infty} b_n(\beta_0, \theta_0) p_n(\beta_0, \theta_0) & \sum_{n=1}^{\infty} c_n(\beta_0, \theta_0) p_n(\beta_0, \theta_0) \end{bmatrix}$$

is the Fisher information matrix, with $p_n(\beta, \theta)$ *as in (56) and*

$$a_n(\beta, \theta) = \sum_{k=0}^{n-2} \frac{1}{(1 + k\beta - \theta)^2},$$

$$b_n(\beta, \theta) = -\sum_{k=0}^{n-2} \frac{k}{(1 + k\beta - \theta)^2},$$

$$c_n(\beta, \theta) = \sum_{k=0}^{n-2} k^2 \left(\frac{1}{(1 + k\beta - \theta)^2} - \frac{1}{(1 + k\beta)^2}\right).$$

Acknowledgements The authors thank two anonymous referees for their helpful comments. Research of the first author was partially funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under Grant Agreement No 318984 - RARE. The second author was partially supported by Riksbankens Jubileumsfond Grant Dnr: P13-1024.

7 Appendix

Proof of Proposition 1. For the variables X and Y in (7), we have

$$S_X(x) = \mathbb{P}(X > x) = 1 - \mathbb{E}(\mathbb{P}(X \le x|N_\alpha)) = 1 - \mathbb{E}\left(F^{N_\alpha}(x)\right) = 1 - G_{N_\alpha}(F(x)),$$
$$F_Y(y) = \mathbb{P}(Y \le y) = 1 - \mathbb{E}(\mathbb{P}(Y > y|N_\alpha)) = 1 - \mathbb{E}\left(S^{N_\alpha}(y)\right) = 1 - G_{N_\alpha}(S(y)),$$

where G_{N_α} is the PGF given in (11). Thus,

$$S_X(x) = (1 - F(x))^\alpha = S^\alpha(x),$$
$$F_Y(y) = (1 - S(x))^\alpha = F^\alpha(x),$$

as required. □

Proof of Proposition 2. Suppose that, for some $\alpha \in (0, 1)$, N has Sibuya distribution $GS_1(\alpha, 0)$, given by the PMF (5). Then, for each $t \in \mathbb{R}_+$, the value of the process $N(t)$ defined by (8) admits the stochastic representation (45), where $c(t) = \mathbb{P}(X_j > t)$. Since for Sibuya distributed N we have (42) with $c = c(t)$, which, in turn, implies (43), $N(t)$ satisfies (9), as desired.

Next, assume that $N(t)$ satisfies equation (9). Thus, for each $t \in \mathbb{R}_+$, we have

$$\mathbb{P}(N(t) = n) = \mathbb{P}(N = n)\mathbb{P}(N(t) > 0), \quad n \in \mathbb{N}. \tag{57}$$

Using standard conditioning argument, write

$$\mathbb{P}(N(t) = n) = \sum_{k=1}^{\infty} \mathbb{P}(N(t) = n|N = k)\mathbb{P}(N = k), \quad n \in \mathbb{N}_0.$$

Noting that for $k < n$ we have $\mathbb{P}(N(t) = n|N = k) = 0$ while for $k \ge n$ the variable $N(t) = n|N = k$ is binomial with parameters k and $p = 1 - F(t)$, where F is the common CDF of the X_j's, we conclude that

$$\mathbb{P}(N(t) = n) = \sum_{k=n}^{\infty} \binom{k}{n} [1 - F(t)]^n [F(t)]^{k-n} \mathbb{P}(N = k), \quad n \in \mathbb{N}, \quad t \in \mathbb{R}_+. \tag{58}$$

For $n = 0$, we have

$$\mathbb{P}(N(t) = 0) = \sum_{k=1}^{\infty} [F(t)]^k \mathbb{P}(N = k), \quad t \in \mathbb{R}_+. \tag{59}$$

We now write $s = F(t) \in (0, 1)$ and $p_n = \mathbb{P}(N = n)$ and substitute (58) and (59) into (57), which results in the following equation

$$(1-s)^n \sum_{k=n}^{\infty} \binom{k}{n} s^{k-n} p_k = p_n \left(1 - \sum_{k=1}^{\infty} s^k p_k\right), \quad n \in \mathbb{N}, \quad s \in (0, 1). \tag{60}$$

Further, by expanding the term $(1-s)^n$ into a power series in s and changing the index of the summation on the left-hand side of (60) to $j = k - n$, we conclude that

$$\left\{\sum_{k=0}^{n} \binom{n}{k} (-1)^k s^k\right\} \cdot \left\{\sum_{j=0}^{\infty} \binom{j+n}{n} p_{j+n} s^j\right\} = p_n - \sum_{j=1}^{\infty} p_j p_n s^j, \quad n \in \mathbb{N}, \quad s \in (0, 1). \tag{61}$$

Using standard result for power series, stating that the coefficients c_k of the product

$$\sum_{k=0}^{\infty} c_k s^k = \left\{\sum_{i=0}^{\infty} a_i s^i\right\} \cdot \left\{\sum_{j=0}^{\infty} b_j s^j\right\} \tag{62}$$

are given by

$$c_k = \sum_{i=0}^{k} a_i b_{k-i},$$

following some algebra, we conclude the left-hand side of (61) is of the form (62) with

$$c_k = \sum_{j=0}^{k} \binom{n}{j} \binom{k-j+n}{n} p_{k-j+n} (-1)^j, \quad 0 \le k \le n, \quad n \in \mathbb{N}.$$

Thus, in view of the above, coupled with (61), and by the uniqueness of the power series, we conclude that

$$\sum_{j=0}^{k} \binom{n}{j} \binom{k-j+n}{n} p_{k-j+n} (-1)^j = -p_k p_n, \quad 1 \le k \le n, \quad n \in \mathbb{N}. \tag{63}$$

In particular, for $k = 1$, relation (63) reduces to

$$(n+1) p_{n+1} - n p_n = -p_1 p_n, \quad n \in \mathbb{N},$$

leading to

$$p_{n+1} = \frac{(n - p_1) p_n}{n+1}, \quad n \in \mathbb{N}.$$

It now follows by induction that the $\{p_n\}$ coincide with Sibuya probabilities (5), where $\alpha = p_1 = \mathbb{P}(N = 1)$. This concludes the proof. $\square$

Proof of Proposition 4. Since, in view of (13), the results of Proposition 4 and Corollary 1 are equivalent, it is enough to establish (16). First, by incorporating the well-known property of the gamma function,

$$\Gamma(\eta + k) = \Gamma(\eta)\eta(\eta + 1)\cdots(\eta + k - 1), \quad \eta \in \mathbb{R}_+, k \in \mathbb{N},$$

the generalized Sibuya SF (12) can be written as

$$\mathbb{P}(N > n) = \frac{1}{n^\alpha}\frac{\Gamma(\nu + 1 - \alpha + n)}{\Gamma(n)n^{\nu+1-\alpha}}\frac{\Gamma(n)n^{\nu+1}}{\Gamma(n + \nu + 1)}\frac{\Gamma(\nu + 1)}{\Gamma(\nu + 1 - \alpha)}.$$

Next, since for any $\gamma > 0$, we have the asymptotic representation of the Gamma function (see, e.g., Gradshteyn and Ryzhik 2007, formula **8.328**.2, p. 895)

$$\frac{\Gamma(\gamma + n)}{\Gamma(n)n^\gamma} \to 1 \ \text{ as } n \to \infty,$$

the right-hand side of (7) divided by the right-hand side of (16) converges to 1 with $n \to \infty$, as desired. □

Proof of Proposition 5. By Proposition 7, we have the equality in distribution $N \stackrel{d}{=} N(X)$, where $\{N(t),\ t > 0\}$ is a standard Poisson process independent of $X \stackrel{d}{=} X_1 X_{\nu-\alpha+1}/X_\alpha$, where all the three variables on the right-hand side are independent and gamma distributed with scale one and the shape parameters indicated by the subindex. The result now follows from (17), the representation (18) for the integer-order moments of $N(t)$, and and the well-known moment formulas for gamma distribution, which produce

$$\mathbb{E}(X^\delta) = \frac{\Gamma(\alpha - \delta)\Gamma(\delta + 1)\Gamma(\delta + \nu - \alpha + 1)}{\Gamma(\alpha)\Gamma(\nu - \alpha + 1)}, \quad \delta < \alpha.$$

□

Proof of Proposition 6. By Proposition 7, the PGF of N is given by (33), where $\phi_X(\cdot)$ is the LT of the variable X defined in (27). To prove the result, it is enough to show that the LT of X is given by (34). To establish the latter, we condition on $T_{\alpha,\nu}$ when computing the LT of X, leading to

$$\phi_X(t) = \int_0^\infty \mathbb{E}e^{-tE/x} f(x)\mathrm{d}x,$$

where $f(x)$ is given in (28) and E is standard exponential with the LT

$$\mathbb{E}e^{-tE} = \frac{1}{1 + t}, \quad t \in \mathbb{R}_+.$$

Thus,

$$\phi_X(t) = \frac{1}{B(\alpha, \nu - \alpha + 1)} \int_0^\infty \frac{x^\alpha}{(t + x)(1 + x)^{\nu+1}} \mathrm{d}x,$$

where $B(a, b)$ is the beta function (29). The result now follows by the integration formula **3.227**.1 p. 320 of Gradshteyn and Ryzhik (2007). □

Proof of Proposition 7. It is known (see, e.g., Devroye 1993) that the generalized hypergeometric distribution of type B3, given in (30) with X as in (32), is of the form

$$\mathbb{P}(N = n) = \frac{\Gamma(a + c)\Gamma(b + c)\Gamma(a + n)\Gamma(b + n)}{\Gamma(a)\Gamma(b)\Gamma(c)\Gamma(a + b + c + n)n!}, \quad n \in \mathbb{N}_0. \tag{64}$$

Setting $a = 1$, $b = 1 - \alpha + \nu$, and $c = \alpha$ in (64) produces the $GS_0(\alpha, \nu)$ distribution. □

Proof of Proposition 9. We proceed by showing that the PMF of the variable $[W]$ coincides with that of the $GS_0(\alpha, \nu)$ distribution. First, using standard conditioning argument, write

$$\mathbb{P}([W] = n) = \int_0^\infty \mathbb{P}([E/x] = n) g(x) \mathrm{d}x, \quad n \in \mathbb{N}_0, \tag{65}$$

where E has the standard exponential distribution and g is the PDF of $V_{\alpha,\nu}$, given by (36). Since

$$\mathbb{P}([E/x] = n) = \mathbb{P}(nx \leq E < (n + 1)x) = e^{-nx} - e^{-(n+1)x},$$

the probability (65) takes on the form

$$\mathbb{P}([W] = n) = \frac{\Gamma(\nu + 1)}{\Gamma(\alpha)\Gamma(1 - \alpha + \nu)} \left\{I_{\nu+n}(\alpha) - I_{\nu+n+1}(\alpha)\right\},$$

where

$$I_\nu(\alpha) = \int_0^\infty e^{-\nu x}(e^x - 1)^{\alpha-1} \mathrm{d}x, \quad \nu \geq 0, \quad 0 < \alpha < \nu + 1.$$

Noting that the function $g(\cdot)$ in (36) is a genuine PDF for each $\nu \geq 0$ and $0 < \alpha < \nu+1$, we conclude that

$$I_\nu(\alpha) = \frac{\Gamma(\alpha)\Gamma(1 - \alpha + \nu)}{\Gamma(\nu + 1)}, \quad \nu \geq 0, \quad 0 < \alpha < \nu + 1. \tag{66}$$

A substitution of (66) into (7), followed by some algebra, produces the $GS_0(\alpha, \nu)$ distribution. This concludes the proof. □

Proof of Proposition 11. To prove the result, we shall use the following sufficient condition for this property to hold (Bondesson 1992, p. 28): A strictly decreasing PMF $\{p_n\}$, $n \in \mathbb{N}_0$, is DSD if

$$\max_{0 \le n \le j} \frac{p_{n+1}}{p_n} \le \frac{j+2}{j+1} \frac{p_{j+1} - p_{j+2}}{p_j - p_{j+1}}, \quad j \in \mathbb{N}_0. \tag{67}$$

First, we shall show that generalized Sibuya PMF is strictly decreasing in n. To see this, note that the ratio

$$\frac{p_{n+1}}{p_n} = \frac{\mathbb{P}(N = n+1)}{\mathbb{P}(N = n)} = \frac{\nu + n + 1 - \alpha}{\nu + n + 2}, \quad n \in \mathbb{N}_0, \tag{68}$$

is strictly increasing in $n \in \mathbb{N}_0$. Indeed, the derivative of the function

$$g(x) = \frac{\nu + 1 - \alpha + x}{\nu + 2 + x}, \quad x \in \mathbb{R}_+,$$

is positive for all $x \in \mathbb{R}_+$, which can be checked by straightforward algebra. Since the ratio (68) converges to 1 as $n \to \infty$, we conclude that $p_{n+1}/p_n < 1$ for all $n \in \mathbb{N}_0$, showing the monotonicity of the sequence $\{p_n\}$, $n \in \mathbb{N}_0$. This also shows that the maximum on the left-hand side of (67) is attained for $n = j$, so that the condition (67) becomes

$$\frac{p_{j+1}}{p_j} \le \frac{j+2}{j+1} \frac{p_{j+1} - p_{j+2}}{p_j - p_{j+1}}, \quad j \in \mathbb{N}_0. \tag{69}$$

After some algebra, condition (69) can be restated as

$$(j+1)\left(1 - \frac{p_{j+1}}{p_j}\right) \le (j+2)\left(1 - \frac{p_{j+2}}{p_{j+1}}\right), \quad j \in \mathbb{N}_0. \tag{70}$$

Since

$$(j+1)\left(1 - \frac{p_{j+1}}{p_j}\right) = \frac{(j+1)(1+\alpha)}{\nu + 1 + (j+1)}, \quad j \in \mathbb{N}_0,$$

and the function

$$h(x) = \frac{x(1+\alpha)}{\nu + 1 + x} = \frac{1+\alpha}{1 + \frac{\nu+1}{x}}$$

is non-decreasing in $x \in \mathbb{R}_+$, we obtain (70). This concludes the proof. □

Proof of Proposition 12. According to the remarks following the statement of Proposition 12, condition (42) implies (44), which, in view of (45), is equivalent to (9). The result now follows from Proposition 2. □

Proof of Proposition 14. For $n = 1$, the statement is trivial. To prove the result for general $n \in \mathbb{N}$, it is enough to show the following fact:

(A) For each $n \geq 2$, the conditional distribution of T_n given the $n-1$ values $0 < t_1 < \cdots < t_{n-1} < 1$ of the previous jump locations has a uniform distribution on the interval $(t_{n-1}, 1)$.

Indeed, if (A) is true, the PDF of the joint distribution of $(T_1, \ldots, T_n)$ is easily seen to be given by (48). This, in turn, is the joint PDF of the random vector on the right-hand side of (47). To see this, consider a random vector $(\Gamma_1, \ldots, \Gamma_n)$ of successive arrivals of standard Poisson process, so that $\Gamma_i = W_1 + \cdots + W_i$, $i = 1 \ldots n$, where the $\{W_i\}$ are IID standard exponential variables. Routine calculations show that the PDF of $(\Gamma_1, \ldots, \Gamma_n)$ is of the form

$$r(x_1, \ldots, x_n) = e^{-x_n}, \quad 0 < x_1 < x_2 < \cdots < x_n.$$

Consider a one-to-one transformation $T_i = H(\Gamma_i)$, $i = 1, \ldots, n$, where $H(x) = 1 - e^{-x}$ is the common CDF of the W_i's, with the inverse of $H^{-1}(t) = -\log(1-t)$. Since the Jacobian of the inverse transformation is the product

$$J = \frac{1}{(1-t_1) \cdot \cdots \cdot (1-t_n)},$$

the PDF of $(T_1, \ldots, T_n)$ becomes

$$g(t_1, \ldots, t_n) = r(H^{-1}(t_1), \ldots, H^{-1}(t_n)) \cdot |J| = e^{\log(1-t_n)} \frac{1}{(1-t_1) \cdot \cdots \cdot (1-t_n)},$$

which produces (48).

To establish the claim (A) above, we start with $n = 2$, and consider the conditional probability $\mathbb{P}(T_2 > t | T_1 = t_1)$ for $t_1 < t < 1$. Using the law of total probability, we obtain

$$\mathbb{P}(T_2 > t | T_1 = t_1) = \sum_{k=2}^{\infty} \mathbb{P}(R_2 < 1-t, K_2 = k | R_1 = r_1),$$

where (K_i, R_i) are the random pairs of record times and their sizes (with $R_i = 1 - T_i$), connected with the sequence $\{nU_n\}$ (as described in Sect. 5). Note that the probability under the above sum can be written in terms of the $\{U_n\}$ as

$$\mathbb{P}(R_2 < 1-t, K_2 = k | R_1 = r_1) = \mathbb{P}(2U_2 > r_1, \ldots, (k-1)U_{k-1} > r_1, kU_k < 1-t),$$

or, equivalently, as

$$\mathbb{P}(R_2 < 1-t, K_2 = k | R_1 = r_1) = p(r_1, k) \frac{1-t}{r_1},$$

where

$$p(r_1, k) = \left(1 - \frac{r_1}{2}\right) \cdots \left(1 - \frac{r_1}{k-1}\right) \frac{r_1}{k}, \quad k \geq 2.$$

When compared with (4), $p(r_1, k)$ is recognized as the PMF of $S \sim GS_0(r_1, 1)$ and consequently,

$$\mathbb{P}(T_2 > t | T_1 = t_1) = \frac{1-t}{r_1} \sum_{k=2}^{\infty} p(r_1, k) = \frac{1-t}{1-t_1}.$$

Since the quantity on the right-hand side above is the survival function of the uniform distribution on the interval $(t_1, 1)$, the result holds for $n = 2$. The proof in the case $n > 2$ is similar.

Under the same notation and using again the law of total probability, we have

$$\begin{aligned}
&\mathbb{P}(T_n > t | A_{n-1}) \\
&= \sum_{k=n-1}^{\infty} \sum_{m=1}^{\infty} \mathbb{P}(T_n > t, K_n = k + m | K_{n-1} = k)\mathbb{P}(K_{n-1} = k), \quad t_{n-1} < t < 1,
\end{aligned}$$

where A_{n-1} denotes the condition $T_1 = t_1, \ldots, T_{n-1} = t_{n-1}$. Similarly as before, the conditional probabilities under the double sum above can be expressed as

$$\mathbb{P}(T_n > t, K_n = k + m | K_{n-1} = k) = p(r_{n-1}, k, m)\frac{1-t}{r_{n-1}},$$

where

$$p(r_{n-1}, k, m) = \left(1 - \frac{r_{n-1}}{k+1}\right) \cdots \left(1 - \frac{r_{n-1}}{k+m-1}\right) \frac{r_{n-1}}{k+m}, \quad m \in \mathbb{N},$$

is recognized as the probability $\mathbb{P}(S = m)$ with $S \sim GS_1(r_1, k)$. Since these probabilities sum up to one across the values of $m \in \mathbb{N}_0$, and so do the probabilities $\mathbb{P}(K_{n-1} = k)$ across the values of $k \geq n - 1$, we obtain

$$\begin{aligned}
\mathbb{P}(T_n > t | A_{n-1}) &= \frac{1-t}{r_{n-1}} \sum_{k=n-1}^{\infty} \mathbb{P}(K_{n-1} = k) \sum_{m=1}^{\infty} p(r_{n-1}, k, m) \\
&= \frac{1-t}{1-t_{n-1}}, \quad t_{n-1} < t < 1.
\end{aligned}$$

Since the quantity on the right-hand side above is the survival function of the uniform distribution on the interval $(t_{n-1}, 1)$, the result follows. □

Proof of Proposition 17. Write the estimators as

$$(\hat{\beta}_n, \hat{\theta}_n) = H(M_1, M_2) = (H_1(M_1, M_2), H_2(M_1, M_2)), \tag{71}$$

where

$$H_1(y_1, y_2) = \frac{y_2 - 2y_1^2 + y_1}{y_2 - 2y_1^2 + y_1 + y_1(y_2 - y_1)},$$

$$H_2(y_1, y_2) = \frac{2(y_2 - y_1^2)}{y_2 - 2y_1^2 + y_1 + y_1(y_2 - y_1)}$$

whenever $y_2 - 2y_1^2 + y_1 > 0$, while otherwise, $H_1(y_1, y_2) = 0$, $H_2(y_1, y_2) = 1/y_1$ (with $y_1, y_2 \geq 1$). To prove consistency, apply law of large numbers to the sequence $Z_i = (X_i, X_i^2)'$ and conclude that the sample mean $\overline{Z}_n = (M_1, M_2)'$ converges in distribution to the population mean $m_Z = \mathbb{E}(Z_i) = (\mu_1, \mu_2)'$, where

$$\mu_1 = \frac{1-\beta}{\theta-\beta}, \quad \mu_2 = \frac{1-\beta}{\theta-\beta}\frac{2(1-\beta)-\theta}{\theta-2\beta}$$

are the first two moments of $GS_1(\beta, \theta)$ distribution (and are well defined when $\theta > 2\beta$). Since the function H is continuous at m_Z, by continuous mapping theorem, the sequence (71) converges in distribution to $H(m_Z) = (\beta, \theta)$. The last equality follows straightforward, albeit tedious, algebra. This proves the estimators are consistent.

Next, we establish their asymptotic normality. Assuming the fourth moment of the $\{X_i\}$ is finite ($\theta > 4\beta$), by the classical multivariate central limit theorem, we have the convergence in distribution $\sqrt{n}(\overline{Z}_n - m_Z) \stackrel{d}{\to} \mathrm{N}(0, \Sigma)$, where the right-hand side denotes the bivariate normal distribution with mean vector zero and covariance matrix

$$\Sigma = \begin{bmatrix} \mathbb{V}ar(X_i) & \mathbb{C}ov(X_i, X_i^2) \\ \mathbb{C}ov(X_i, X_i^2) & \mathbb{V}ar(X_i^2) \end{bmatrix}.$$

A straightforward calculation, facilitated by Propositions 5 and 3, along with basic properties of expectation, shows that

$$\begin{aligned} \mathbb{V}ar(X_i) &= \frac{\theta\ (1-\beta)\ (1-\theta)}{(\beta-\theta)^2\ (\theta-2\,\beta)}, \\ \mathbb{C}ov(X_i, X_i^2) &= \frac{\theta\ (1-\beta)\ (1-\theta)\ (4-5\,\beta-\theta)}{(\beta-\theta)^2\ (2\,\beta-\theta)\ (3\,\beta-\theta)}, \\ \mathbb{V}ar(X_i^2) &= \frac{\theta\ (1-\beta)\ (1-\theta)\ (2-2\,\beta-\theta)\ \left(32\,\beta^2-13\,\beta\,\theta-22\,\beta-\theta^2+10\,\theta\right)}{(\beta-\theta)^2\ (2\,\beta-\theta)^2\ (3\,\beta-\theta)\ (4\,\beta-\theta)}. \end{aligned}$$

Thus, since the function H is differentiable at m_Z, standard multivariate delta method leads to the conclusion that, as $n \to \infty$, the variables

$$\sqrt{n}(H(\overline{Z}_n) - H(m_Z)) = \sqrt{n}[(\hat{\beta}_n, \hat{\theta}_n)' - (\theta, \beta)']$$

converge in distribution to a bivariate normal vector with mean vector zero and covariance matrix $\Sigma_{MME} = D\Sigma D'$, where

$$D = \left[\frac{\partial H_i}{\partial y_j}\bigg|_{(y_1, y_2) = m_Z} \right]_{i,j=1}^2 = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$$

is the matrix of partial derivatives of the vector-valued function H evaluated at m_Z. A routine, rather lengthy calculation yields

$$d_{11} = \frac{(\beta-\theta)\left(8\beta^2-4\beta\theta-6\beta-\theta^2+4\theta\right)}{2\,(1-\theta)},$$

$$d_{22} = -\frac{(\beta-\theta)\,(2\beta-\theta)^2}{2\,(1-\beta)},$$

$$d_{21} = -\frac{(\beta-\theta)\left(-8\beta^2+4\beta\theta+8\beta+\theta^2-6\theta\right)}{2\,(1-\beta)},$$

$$d_{12} = -\frac{(\beta-\theta)\,(2\beta-\theta)^2}{2\,(1-\theta)}.$$

Finally, straightforward matrix multiplication produces the asymptotic covariance matrix Σ_{MME}. □

References

Aban, I. B., Meerschaert, M. M., Panorska, A. K. (2006). Parameter estimation for the truncated Pareto distribution. *Journal of the American Statistical Association*, *101*, 270–277.

Bondesson, L. (1992). *Generalized gamma convolutions and related classes of distributions and densities*. Lecture notes in statistics (Vol. 76). Berlin: Springer.

Buddana, A., Kozubowski, T. J. (2014). Discrete Pareto distributions. *Economic Quality Control*, *29*(2), 143–156.

Christoph, G., Schreiber, K. (1998). Discrete stable random variables. *Statistics and Probability Letters*, *37*, 243–247.

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

Clauset, A., Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *2*, 661–703.

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

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, *1*, 255–293.

Gradshteyn, I. S., Ryzhik, I. M. (2007). *Tables of integrals, series, and products* (7th ed.). Amsterdam: Academic Press.

Huillet, T. E. (2012). On Linnik's continuous-time random walks (new version of Huillet (2000). On Linnik's continuous-time random walks. *Journal of Physics A*: *Mathematical and General*, *33*(14), 2631–2652. Available at http://www.reserachgate.net/publication/231129053.

Huillet, T. E. (2016). On Mittag-Leffler distributions and related stochastic processes. *Journal of Computational and Applied Mathematics*, *296*, 181–211.

Johnson, N. L., Kotz, S., Kemp, A. W. (1993). *Univariate discrete distributions*. New York: John Wiley & Sons.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous univariate distributions* (Vol. 1). New York: John Wiley & Sons.

Kozubowski, T.J., Podgórski, K. (2016). Certain bivariate distributions and random processes connected with maxima and minima. *Working papers in statistics 2016*:*9*, Department of Statistics, School of Economics and Management, Lund University.

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

Lehman, E. L. (1983). *Theory of point estimation*. New York: John Wiley & Sons.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*, 323–351.

Pakes, A. G. (1995). Characterization of discrete laws via mixed sums and Markov branching processes. *Stochastic Processes and their Applications*, *55*, 285–300.

Pillai, R. N., Jayakumar, K. (1995). Discrete Mittag-Leffler distributions. *Statistics and Probability Letters*, *23*, 271–274.

Rényi, A. (1976). On outstanding values of a sequence of observations. In: Rényi A (ed). *Selected papers of A. Rényi* (Vol. 3, pp. 50–65). Budapest: Akadémiai Kiadó.

Satheesh, S., Nair, N. U. (2002). Some classes of distributions on the non-negative lattice. *Journal of the Indian Statistical Association*, *40*(1), 41–58.

Sibuya, M., Shimizu, R. (1981). The generalized hypergeometric family of distributions. *Annals of the Institute of Statistical Mathematics*, *33*, 177–190.

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

Sornette, D. (2006). *Critical phenomena in natural sciences*: *Chaos, fractals, selforganization and disorder*: *Concepts and tools* (2nd ed.). Berlin: Springer.

Steutel, F. W., van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability*, *7*, 893–899.

Steutel, F. W., van Harn, K. (2004). *Infinitely divisibility of probability distributions on the real line*. New York: Marcel Dekker.

Stumpf, M. P. H., Porter, M. A. (2012). Critical truths about power laws. *Science*, *335*, 665–666.

Yule, G. U. (1925). A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London Series B.*, *213*, 21–87.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.