Journal of Mathematical Extension
Vol. 15, No. 2, (2021) (16)1-33
URL: https://doi.org/10.30495/JME.2021.1332
ISSN: 1735-8299
Original Research Paper

The Sine Kumaraswamy-G Family of Distributions

C. Chesneau*
University of Caen-Normandie

F. Jamal
The Islamia University of Bahawalpur

Abstract. In this paper, we introduce a new trigonometric family of continuous distributions called the sine Kumaraswamy-G family of distributions. It can be presented as a natural extension of the well-established sine-G family of distributions, with new perspectives in terms of applicability. We investigate the main mathematical properties of the sine Kumaraswamy-G family of distributions, including asymptotes, quantile function, linear representations of the cumulative distribution and probability density functions, moments, skewness, kurtosis, incomplete moments, probability weighted moments and order statistics. Then, we focus our attention on a special member of this family called the sine Kumaraswamy exponential distribution. The statistical inference for the related parametric model is explored by using the maximum likelihood method. Among others, asymptotic confidence intervals and likelihood ratio tests for the parameters are discussed. A simulation study is performed under varying sample sizes to assess the performance of the model. Finally, applications to two practical data sets are presented to illustrate its potentiality and robustness.

AMS Subject Classification: 62E15; 62H10.
Keywords and Phrases: Sine-G family of distributions; Kumaraswamy distribution; moments; practical data sets.

Received: July 2019; Accepted: June 2020
*Corresponding Author

1 Introduction

In recent years, much attention has been paid to the construction of trigonometric families of distributions. The advantages of these families are to keep a balance between a relative simplicity in their definitions, allowing a perfect comprehension of their mathematical properties, and a great applicability for modelling various kinds of practical data sets. These two points follow from an appropriate use of flexible trigonometric functions. To our knowledge, the pioneer trigonometric family of distributions is the sine-G family of distributions introduced by [12] and [20]. A brief description of this family is presented below. Let $G(x)$ be the cumulative distribution function (cdf) of an univariate continuous distribution and $g(x)$ be the corresponding probability density function (pdf). Then, the sine-G family of distributions is characterized by the cdf given by

$$F(x) = \sin\left(\frac{\pi}{2}G(x)\right), \qquad x \in \mathbb{R}. \tag{1}$$

The related pdf is given by

$$f(x) = \frac{\pi}{2}g(x)\cos\left(\frac{\pi}{2}G(x)\right), \qquad x \in \mathbb{R}.$$

Thus, simple functions are involved and it is proved in [12], [20] and [23] that the flexibility of $G(x)$ can be significantly enriched by the sine transformation. The related parametric models take advantage of these properties for a nice fitting of various kinds of data sets. By exploiting the flexible nature of various trigonometric transformations, other trigonometric families of distributions have been developed. See, for instance, the cos-G family of distributions by [20] and [24], the tan-G family of distributions by [20], [21] and [2], the sec-G family of distributions by [20] and [22], the new sine-G family of distributions by [14], the T-X-Tan-G by [1], the CS-G family of distributions by [3] and the TransSC-G family of distributions by [10].

In this paper, we propose a new trigonometric family of continuous distributions, called the sine Kumaraswamy-G family of distributions. It can be viewed as a "two power shape parameters generalization" of the former sine-G family of distributions. We describe it as follows. Let

$a > 0$, $b > 0$, $G(x)$ be the cdf of an univariate continuous distribution and $g(x)$ be the corresponding pdf. Then, the sine Kumaraswamy-G family of distributions is characterized by the cdf given by

$$F(x) = \cos\left(\frac{\pi}{2}[1 - G(x)^a]^b\right), \quad x \in \mathbb{R}. \tag{2}$$

The corresponding pdf is obtained as

$$f(x) = \frac{\pi}{2}abg(x)G(x)^{a-1}[1 - G(x)^a]^{b-1}\sin\left(\frac{\pi}{2}[1 - G(x)^a]^b\right), \quad x \in \mathbb{R}. \tag{3}$$

As indicated by its name, by using a trigonometric formula, we can show that $F(x)$ is obtained by the composition of the sine-G cdf given as (1) and the Kumaraswamy-G cdf specified by $H(x) = 1 - [1 - G(x)^a]^b$, $x \in \mathbb{R}$. Further details and applications on the Kumaraswamy-G family of distributions can be found in [4], [16], [7] and [19]. The roles of a and b are to add more flexibility to the former cdf $G(x)$, allowing the construction of models which take into account precise characteristics of various data sets. One can notice that, for $b = 1$, $F(x)$ becomes $F(x) = \sin((\pi/2)G(x)^a)$, which is the cdf of the sine exp-G family of distributions (new in the literature to the best of our knowledge, but very natural to consider) and for $a = b = 1$, we rediscover the cdf of the sine-G family of distributions. The idea of combining trigonometric and Kumaraswamy-G families of distributions finds trace in [20, Chapter 6], but for the sec-G family of distributions (not the sine-G one) and with the specific Kumaraswamy-Weibull distribution as baseline (not the general Kumaraswamy-G family of distributions, i.e., for any $G(x)$). Thus, the sine Kumaraswamy-G family of distributions remains new in the literature and deserves a complete study, which is the aim of this paper. After providing a comprehensive treatment of its mathematical properties, we focus our attention on a special member of this family, defined with the exponential distribution as baseline. It is called the sine Kumaraswamy exponential distribution. Then, we consider it as a parametric statistical model, with the estimation of the unknown parameters via the maximum likelihood method. We take advantage of the existing convergence properties of this method to present a solid model for data analysis. This is illustrated by the means of two practical sets.

In particular, we show that the proposed model is better, in some sense, to well-recognized competitive models of the literature.

The rest of the paper is organized as follows. In Section 2, the main features of the sine Kumaraswamy-G family of distributions are explored. Then, the sine Kumaraswamy exponential distribution is studied in detail in Section 3. In Section 4, it is considered as a parametric model, with a statistical inference study, including concrete applications. Conclusions are given in Section 5

2 Main features

In this section, we investigate the main features of the sine Kumaraswamy-G family of distributions. We recall that it is characterized by the cdf $F(x)$ given by (2) and the related pdf $f(x)$ specified by (3).

2.1 Main functions

We now express the main functions of interest of the sine Kumaraswamy-G family of distributions. The corresponding survival function (sf) is given by

$$S(x) = 1 - F(x) = 2\left[\sin\left(\frac{\pi}{4}[1 - G(x)^a]^b\right)\right]^2, \quad x \in \mathbb{R}.$$

We deduce the hazard rate function (hrf) sine Kumaraswamy-G family defined by

$$\begin{aligned} h(x) &= \frac{f(x)}{S(x)} \\ &= \frac{\pi}{2} abg(x)G(x)^{a-1}[1 - G(x)^a]^{b-1} \cot\left(\frac{\pi}{4}[1 - G(x)^a]^b\right), \quad x \in \mathbb{R}. \end{aligned}$$

The corresponding cumulative hazard rate function (chrf) is

$$\Omega(x) = -\log[S(x)] = -\log(2) - 2\log\left[\sin\left(\frac{\pi}{4}[1 - G(x)^a]^b\right)\right], \quad x \in \mathbb{R}.$$

Another central function of the sine Kumaraswamy-G family of distributions is the quantile function (qf) obtained as

$$Q(y) = Q_G\left(\left[1 - \left\{\frac{2}{\pi}\arccos(y)\right\}^{1/b}\right]^{1/a}\right), \quad y \in (0,1), \tag{4}$$

where $Q_G(y)$ denotes the qf corresponding to $G(x)$. Let us recall that $Q(y)$ is characterized by the non-linear equation $F(Q(y)) = Q(F(y)) = y$, $y \in (0,1)$. The median is given by

$$M = Q_G\left(\left[1 - \left\{\frac{2}{\pi}\arccos(0.5)\right\}^{1/b}\right]^{1/a}\right),$$

with $\arccos(0.5) \approx 1.04719755$. The qf is also involved in the following key result: for a random variable U having the uniform distribution on the unit interval, the random variable X defined by $X = Q(U)$ has the cdf (2). Others uses of the qf will be developed in the next.

2.2 Asymptotic properties

Let us now investigate the asymptotic properties of the functions $F(x)$, $f(x)$ and $h(x)$. As $G(x) \to 0$, by using the equivalence $(1-y^a)^b \sim 1-by^a$ when $y \to 0$, we have

$$F(x) \sim \frac{\pi}{2}bG(x)^a, \quad f(x) \sim \frac{\pi}{2}abg(x)G(x)^{a-1}, \quad h(x) \sim \frac{\pi}{2}abg(x)G(x)^{a-1}.$$

As $G(x) \to 1$, by using $\cos(y) \sim 1 - y^2/2$ when $y \to 0$, we have

$$F(x) \sim 1 - \frac{\pi^2}{8}[1 - G(x)^a]^{2b}, \quad f(x) \sim \frac{\pi^2}{4}abg(x)[1 - G(x)^a]^{2b-1},$$

$$h(x) \sim 2abg(x)[1 - G(x)^a]^{-1}.$$

The convergence and limits of $f(x)$ and $h(x)$ can not be determined in full generality; they depend on a, b and the definition of $G(x)$ (and $g(x)$ a fortiori).

2.3 Critical points

Any critical point of $f(x)$, say x_0, satisfies the following equation: $[\log(f(x))' \mid_{x=x_0}= 0$, i.e.,

$$\frac{g'(x_0)}{g(x_0)} + (a-1)\frac{g(x_0)}{G(x_0)} - (b-1)\frac{ag(x_0)G(x_0)^{a-1}}{1-G(x_0)^a} - \frac{\pi}{2}abg(x_0)G(x_0)^{a-1}[1-G(x_0)^a]^{b-1}\cot\left(\frac{\pi}{2}[1-G(x_0)^a]^b\right) = 0. \quad (5)$$

By investigating the sign of $\tau = [\log(f(x))'' \mid_{x=x_0}$, we can determine the nature of x_0; it corresponds to a maximum point if $\tau < 0$, a minimum point if $\tau > 0$ and a point of inflection if $\tau = 0$.

Similarly, any critical point of $h(x)$, say x_*, satisfies the following equation: $[\log(h(x))' \mid_{x=x_*}= 0$, i.e.,

$$\frac{g'(x_*)}{g(x_*)} + (a-1)\frac{g(x_*)}{G(x_*)} - (b-1)\frac{ag(x_*)G(x_*)^{a-1}}{1-G(x_*)^a} + \frac{\pi}{2}abg(x_*)G(x_*)^{a-1}[1-G(x_*)^a]^{b-1}\times \left[\cot\left(\frac{\pi}{4}[1-G(x_*)^a]^b\right) - \cot\left(\frac{\pi}{2}[1-G(x_*)^a]^b\right)\right] = 0. \quad (6)$$

Also, the sign of $\theta = [\log(h(x))'' \mid_{x=x_*}$ is informative concerning the nature of x_*.

2.4 Linear representations

Here, some linear representations for $F(x)$ and $f(x)$ are determined. It follows from the series expansion of the cosine function that

$$F(x) = \cos\left(\frac{\pi}{2}[1-G(x)^a]^b\right) = \sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k)!}\frac{\pi^{2k}}{2^{2k}}[1-G(x)^a]^{2bk}.$$

Furthermore, the generalized binomial formula gives

$$[1-G(x)^a]^{2bk} = \sum_{\ell=0}^{+\infty}\binom{2bk}{\ell}(-1)^\ell G(x)^{a\ell},$$

where $\binom{2bk}{\ell} = 2bk(2bk-1)\ldots(2bk-\ell+1)/\ell!$. We immediately deduce the following linear representation for $F(x)$:

$$F(x) = \sum_{\ell=0}^{+\infty} a_\ell G(x)^{a\ell}, \quad a_\ell = (-1)^\ell \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} \frac{\pi^{2k}}{2^{2k}} \binom{2bk}{\ell}. \tag{7}$$

Upon differentiation, we obtain the following linear representation for $f(x)$:

$$f(x) = \sum_{\ell=1}^{+\infty} a_\ell [a\ell g(x) G(x)^{a\ell-1}]. \tag{8}$$

Thus, some mathematical properties of the sine Kumaraswamy-G family of distributions can be derived from these expansions and the properties of the exp-G family of distributions.

Alternatively, one can investigate linear representations for $F(x)$ and $f(x)$ in terms of the sf related to $G(x)$, i.e., $S_G(x) = 1 - G(x)$. This can be more useful if $S_G(x)$ is more tractable than $G(x)$. By using the generalized binomial formula, we have

$$G(x)^{a\ell} = \sum_{m=0}^{+\infty} \binom{a\ell}{m} (-1)^m S_G(x)^m.$$

It follows from (7) that

$$F(x) = \sum_{m=0}^{+\infty} b_m S_G(x)^m, \quad b_m = (-1)^m \sum_{\ell=0}^{+\infty} \binom{a\ell}{m} a_\ell. \tag{9}$$

Upon differentiation, we obtain the following linear representation for $f(x)$:

$$f(x) = \sum_{m=1}^{+\infty} b_m^* \left[m g(x) S_G(x)^{m-1} \right], \quad b_m^* = -b_m. \tag{10}$$

Applications of (9) and (10) will be proposed in Section 3 for a given cdf $G(x)$.

2.5 Moments

Hereafter, it is supposed that all the presented quantities exist (integral, sum...), and that the exchange of the integral and sum signs is valid.

Let r be an integer. Then, the r-th moment of the sine Kumaraswamy-G family of distributions is given by

$$\mu'_r = \int_{-\infty}^{+\infty} x^r f(x)dx$$
$$= \int_{-\infty}^{+\infty} x^r \frac{\pi}{2} abg(x)G(x)^{a-1}[1-G(x)^a]^{b-1} \sin\left(\frac{\pi}{2}[1-G(x)^a]^b\right) dx.$$

By applying the change of variable $x = Q(y)$, where $Q(y)$ denotes the qf expressed as (4), we get

$$\mu'_r = \int_0^1 Q(y)^r dy = \int_0^1 \left[Q_G\left(\left[1-\left\{\frac{2}{\pi}\arccos(y)\right\}^{1/b}\right]^{1/a}\right)\right]^r dy.$$

This integral may be not expressed simply with standard integral techniques. However, in most of the cases, for given $G(x)$, a, b and r, it can be evaluated numerically by the use of a modern mathematical software.

Alternatively, linear representations of μ'_r can be derived to (8) or (10), according to the definition of $G(x)$. Indeed, by using (8), we have

$$\mu'_r = \sum_{\ell=1}^{+\infty} a_\ell \int_{-\infty}^{+\infty} x^r \left[a\ell g(x)G(x)^{a\ell-1}\right] dx$$
$$= \sum_{\ell=1}^{+\infty} a_\ell \int_0^1 \left[a\ell y^{a\ell-1} Q_G(y)^r\right] dy.$$

Similarly, by using (10), we obtain

$$\mu'_r = \sum_{m=1}^{+\infty} b^*_m \int_{-\infty}^{+\infty} x^r \left[mg(x)S_G(x)^{m-1}\right] dx$$
$$= \sum_{m=1}^{+\infty} b^*_m \int_0^1 \left[my^{m-1} Q_G(1-y)^r\right] dy. \tag{11}$$

Especially, the mean is given by $\mu = \mu'_1$ and the variance is defined by $\sigma^2 = \mu'_2 - \mu^2$. Also, the r-th central moment is given by

$$\mu_r = \int_{-\infty}^{+\infty} (x-\mu)^r f(x)dx = \sum_{k=0}^{r} \binom{r}{k} (-1)^k (\mu'_1)^k \mu'_{r-k}$$

and the r-th descending factorial moment is given as

$$\mu'_{(r)} = \int_{-\infty}^{+\infty} x(x-1)(x-r+1)f(x)dx = \sum_{k=0}^{r} s_{sti}(r,k)\mu'_k,$$

where $s_{sti}(r,k)$ denotes the Stirling number of the first kind defined by $s_{sti}(r,k) = (1/k!)[x(x-1)\ldots(x-r+1)]^{(k)} \mid_{x=0}$. We end this subsection by mentioning that the moment generating function can be obtained by invoking arguments similar to those used for μ'_r.

2.6 Skewness and kurtosis

In the context of distributions, let us recall that the skewness corresponds to the asymmetry and the kurtosis corresponds to the tailedness. A useful skewness measure is

$$CS = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu'_3 - 3\mu'_2\mu + 2\mu^3}{\sigma^3}. \tag{12}$$

Also, a kurtosis measure is

$$CK = \frac{\mu_4}{\mu_2^2} = \frac{\mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 3\mu'_4}{\sigma^4}. \tag{13}$$

If the moments do not exist (mainly depending on the definition of $G(x)$), we can envisage measures of skewness and kurtosis depending on the qf given by (4). For instance, for a skewness measure, we can use the Bowley skewness defined by

$$B = \frac{Q(3/4) + Q(1/4) - 2Q(2/4)}{Q(3/4) - Q(1/4)}.$$

See [11]. For a kurtosis measure, we can use the Moors kurtosis defined by

$$M = \frac{Q(3/8) - Q(1/8) + Q(7/8) - Q(5/8)}{Q(6/8) - Q(2/8)}.$$

Details and applications on them can be found in [15].

2.7 Incomplete mean and consorts

Let $t \in \mathbb{R}$. The incomplete mean of the sine Kumaraswamy-G family of distributions is given as

$$\mu^*(t) = \int_{-\infty}^{t} x f(x) dx$$
$$= \int_{-\infty}^{t} x \frac{\pi}{2} abg(x) G(x)^{a-1} [1 - G(x)^a]^{b-1} \sin\left(\frac{\pi}{2}[1 - G(x)^a]^b\right) dx.$$

Equivalently, we have

$$\mu^*(t) = \int_0^{\cos\left(\frac{\pi}{2}[1-G(t)^a]^b\right)} \left[Q_G\left(\left[1 - \left\{\frac{2}{\pi}\arccos(y)\right\}^{1/b}\right]^{1/a}\right)\right]^r dy.$$

For given $G(x)$, a, b and t, this integral can be evaluated numerically.

Alternatively, we can use the linear representation given by (8) and (10). Indeed, by using (8), we have

$$\mu^*(t) = \sum_{\ell=1}^{+\infty} a_\ell \int_{-\infty}^{t} x \left[a\ell g(x) G(x)^{a\ell-1}\right] dx$$
$$= \sum_{\ell=1}^{+\infty} a_\ell \int_0^{G(t)} \left[a\ell y^{a\ell-1} Q_G(y)\right] dy.$$

Similarly, by using (10), we obtain

$$\mu^*(t) = \sum_{m=1}^{+\infty} b_m^* \int_{-\infty}^{t} x \left[m g(x) S_G(x)^{m-1}\right] dx$$
$$= \sum_{m=1}^{+\infty} b_m^* \int_{S_G(t)}^{1} \left[m y^{m-1} Q_G(1-y)\right] dy.$$

From these expressions, several probabilistic quantities involving $\mu^*(t)$ can be expressed. This is the case for the mean deviation about the mean expressed as

$$\delta_1 = \int_{-\infty}^{+\infty} |x - \mu| f(x) dx = 2\mu F(\mu) - 2\mu^*(\mu)$$
$$= 2\mu \cos\left(\frac{\pi}{2}[1 - G(\mu)^a]^b\right) - 2\mu^*(\mu).$$

One can also mention the mean deviation about the median given by $\delta_2 = \int_{-\infty}^{+\infty} |x - M| f(x)dx = \mu - 2\mu^*(M)$, the mean residual life given by $m(t) = [1 - \mu^*(t)]/S(t) - t$, the mean waiting time defined by $M(t) = t - \mu^*(t)/F(t)$, the Bonferroni curve specified by $B(y) = \mu^*(Q(y))/(y\mu)$ with $y \in (0, 1)$ and the Lorenz curve given by $L(y) = \mu^*(Q(y))/\mu$ with $y \in (0, 1)$.

2.8 Probability weighted moments

Let r and s be two integers. We now investigate the (r, s)-th probability weighted moment of the sine Kumaraswamy-G family of distributions defined by

$$\mu'_{r,s} = \int_{-\infty}^{+\infty} x^r F(x)^s f(x)dx$$
$$= \int_{-\infty}^{+\infty} x^r \left[\cos\left(\frac{\pi}{2}[1 - G(x)^a]^b\right)\right]^s \frac{\pi}{2} abg(x)G(x)^{a-1}[1 - G(x)^a]^{b-1} \times$$
$$\sin\left(\frac{\pi}{2}[1 - G(x)^a]^b\right) dx.$$

Note that $\mu'_{r,0} = \mu'_r$. Such probability weighted moments naturally appear in the determination of the moments of the order statistics, as we will see later. Another expression of $\mu'_{r,s}$ is given by

$$\mu'_{r,s} = \int_0^1 y^s \left[Q_G\left(\left[1 - \left\{\frac{2}{\pi}\arccos(y)\right\}^{1/b}\right]^{1/a}\right)\right]^r dy.$$

For given $G(x)$, a, b, r and s, this integral can be evaluated numerically.

Alternatively, one can also investigate a linear representation for $\mu'_{r,s}$ in terms of (raw) moments. Indeed, by applying a result established by

[8, Paragraph 0.314], we have

$$\begin{aligned} F(x)^{s+1} &= \left[\cos\left(\frac{\pi}{2}[1-G(x)^a]^b\right)\right]^{s+1} \\ &= \left[\sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!}\frac{\pi^{2k}}{2^{2k}}[1-G(x)^a]^{2bk}\right]^{s+1} \\ &= \sum_{k=0}^{+\infty} c_{s,k}[1-G(x)^a]^{2bk}, \end{aligned}$$

where $c_{s,0} = 1$ and, for any $k \geq 1$,

$$c_{s,k} = \frac{1}{k}\sum_{\ell=1}^{k}[\ell(s+2)-k]\frac{(-1)^\ell}{(2\ell)!}\frac{\pi^{2\ell}}{2^{2\ell}}c_{s,k-\ell}.$$

The generalized binomial formula gives

$$[1-G(x)^a]^{2bk} = \sum_{\ell=0}^{+\infty}\binom{2bk}{\ell}(-1)^\ell G(x)^{a\ell}.$$

So,

$$F(x)^{s+1} = \sum_{\ell=0}^{+\infty} d_{s,\ell}G(x)^{a\ell}, \quad d_{s,\ell} = (-1)^\ell\sum_{k=0}^{+\infty}c_{s,k}\binom{2bk}{\ell}.$$

Hence, by differentiation, we have

$$F(x)^s f(x) = \sum_{\ell=1}^{+\infty} d^*_{s,\ell}\left[a\ell g(x)G(x)^{a\ell-1}\right], \quad d^*_{s,\ell} = \frac{d_{s,\ell}}{s+1}.$$

Therefore,

$$\begin{aligned} \mu'_{r,s} &= \sum_{\ell=1}^{+\infty} d^*_{s,\ell}\int_{-\infty}^{+\infty} x^r\left[a\ell g(x)G(x)^{a\ell-1}\right]dx \\ &= \sum_{\ell=1}^{+\infty} d^*_{s,\ell}\int_0^1\left[a\ell y^{a\ell-1}Q_G(y)^r\right]dy. \end{aligned} \tag{14}$$

In terms of $S_G(x)$, by using the generalized binomial formula, we have

$$F(x)^{s+1} = \sum_{m=0}^{+\infty} e_{s,m} S_G(x)^m, \quad e_{s,m} = (-1)^m \sum_{\ell=0}^{+\infty} d_{s,\ell} \binom{\alpha \ell}{m}.$$

Hence, by differentiation, we have

$$F(x)^s f(x) = \sum_{m=1}^{+\infty} e_{s,m}^* \left[m g(x) S_G(x)^{m-1} \right], \quad e_{s,m}^* = -\frac{e_{s,\ell}}{s+1}.$$

So,

$$\begin{aligned} \mu'_{r,s} &= \sum_{m=1}^{+\infty} e_{s,m}^* \int_{-\infty}^{+\infty} x^r \left[m g(x) S_G(x)^{m-1} \right] dx \\ &= \sum_{m=1}^{+\infty} e_{s,m}^* \int_0^1 \left[m y^{m-1} Q_G(1-y)^r \right] dy. \end{aligned} \tag{15}$$

2.9 Order statistics

Here, we focus on the order statistics related to the sine Kumaraswamy-G family of distributions. Let $X_1, \ldots, X_n$ be random sample having the sine Kumaraswamy-G cdf given as (2) and $X_{i:n}$ be the i-th order statistic, i.e., the i-th random variable such that, by arranging $X_1, \ldots, X_n$ in increasing order, we have $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$. The complete theory about order statistics can be found in [6]. In particular, in our mathematical context, the cdf of $X_{i:n}$ is obtained as

$$\begin{aligned} F_{i:n}(x) &= \frac{n!}{(i-1)!(n-i)!} \sum_{k=0}^{n-i} \frac{(-1)^k}{k+i} \binom{n-i}{k} F(x)^{k+i} \\ &= \frac{n!}{(i-1)!(n-i)!} \sum_{k=0}^{n-i} \frac{(-1)^k}{k+i} \binom{n-i}{k} \left[\cos\left(\frac{\pi}{2} [1 - G(x)^a]^b \right) \right]^{k+i}, \\ & x \in \mathbb{R}. \end{aligned}$$

The corresponding pdf is specified by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!}\sum_{k=0}^{n-i}(-1)^k \binom{n-i}{k} f(x)F(x)^{k+i-1}$$
$$= \frac{n!}{(i-1)!(n-i)!} f(x)F(x)^{i-1}S(x)^{n-i}$$
$$= \frac{n!}{(i-1)!(n-i)!}2^{n-i-1}\pi abg(x)G(x)^{a-1}[1-G(x)^a]^{b-1}\times$$
$$\sin\left(\frac{\pi}{2}[1-G(x)^a]^b\right)\left[\cos\left(\frac{\pi}{2}[1-G(x)^a]^b\right)\right]^{i-1}\times$$
$$\left[\sin\left(\frac{\pi}{4}[1-G(x)^a]^b\right)\right]^{2(n-i)}.$$

Several kinds of moments can be obtained from $f_{i:n}(x)$. In particular, the r-th moment of $X_{i:n}$ is given by

$$\mu_r^o = \mathbb{E}(X_{i:n}^r) = \int_{-\infty}^{+\infty} x^r f_{i:n}(x)dx.$$

It can be calculated at least numerically for given $G(x)$, a, b and r. Alternatively, it can be expressed via the probability weighted moments given by (14). Indeed, we have

$$\mu_r^o = \int_{-\infty}^{+\infty} x^r f_{i:n}(x)dx$$
$$= \frac{n!}{(i-1)!(n-i)!}\sum_{k=0}^{n-i}(-1)^k \binom{n-i}{k} \int_{-\infty}^{\infty} x^r f(x)F(x)^{k+i-1}dx$$
$$= \frac{n!}{(i-1)!(n-i)!}\sum_{k=0}^{n-i}(-1)^k \binom{n-i}{k} \mu'_{r,k+i-1}. \quad (16)$$

Again, this integral can be evaluated numerically.

3 The sine Kumaraswamy exponential distribution

This section is devoted to a special member of the sine Kumaraswamy-G family of distributions called the sine Kumaraswamy exponential (SKE) distribution.

3.1 Definition and main functions

As indicated by its name, the SKE distribution is the member of the sine Kumaraswamy-G family of distributions defined with the exponential distribution with parameter $\lambda > 0$ as baseline. Hence, it is characterized by the cdf in (2) with the baseline cdf $G(x) = 1 - e^{-\lambda x}$, $x > 0$, i.e.,

$$F(x) = \cos\left(\frac{\pi}{2}[1-(1-e^{-\lambda x})^a]^b\right), \quad x > 0. \tag{17}$$

One can remark that, for $a = b = 1$, we have $F(x) = \cos\left((\pi/2)e^{-\lambda x}\right)$, the cdf of the SE distribution introduced by [12].

The pdf corresponding to (17) is given by

$$f(x) = \frac{\pi}{2} ab\lambda e^{-\lambda x}(1-e^{-\lambda x})^{a-1}[1-(1-e^{-\lambda x})^a]^{b-1} \times \\ \sin\left(\frac{\pi}{2}[1-(1-e^{-\lambda x})^a]^b\right), \quad x > 0. \tag{18}$$

The related sf is obtained as

$$S(x) = 2\left[\sin\left(\frac{\pi}{4}[1-(1-e^{-\lambda x})^a]^b\right)\right]^2, \quad x > 0.$$

Also, the corresponding hrf is

$$h(x) = \frac{\pi}{2} ab\lambda e^{-\lambda x}(1-e^{-\lambda x})^{a-1}[1-(1-e^{-\lambda x})^a]^{b-1} \times \\ \cot\left(\frac{\pi}{4}[1-(1-e^{-\lambda x})^a]^b\right), \quad x > 0$$

and the corresponding chrf is defined by

$$\Omega(x) = -\log(2) - 2\log\left[\sin\left(\frac{\pi}{4}[1-(1-e^{-\lambda x})^a]^b\right)\right], \quad x > 0.$$

The related qf is given by

$$Q(y) = -\frac{1}{\lambda}\log\left(1-\left[1-\left\{\frac{2}{\pi}\arccos(y)\right\}^{1/b}\right]^{1/a}\right), \quad y \in (0,1). \tag{19}$$

Median, quartiles and octiles can be derived, as well as other results.

3.2 Some properties

All the properties exhibited in Section 2 for the general sine Kumaraswamy-G family of distributions can be applied for the SKE distribution with the functions $G(x) = 1 - e^{-\lambda x}$, $x > 0$, $g(x) = \lambda e^{-\lambda x}$ and $Q_G(y) = -(1/\lambda)\log(1-y)$, $y \in (0,1)$. The most significant of them, with numerical illustrations, are presented below.

As $x \to 0$, we have

$$F(x) \sim \frac{\pi}{2} b\lambda^a x^a, \quad f(x) \sim \frac{\pi}{2} ab\lambda^a x^{a-1}, \quad h(x) \sim \frac{\pi}{2} ab\lambda^a x^{a-1}.$$

We can remark that, if $a < 1$, we have $f(x) \to +\infty$, if $a = 1$, we have $f(x) \to (\pi/2)b\lambda$, and if $a > 1$, we have $f(x) \to 0$. The same limits hold for $h(x)$.

As $x \to +\infty$, we have

$$F(x) \sim 1 - \frac{\pi^2}{8} a^{2b} e^{-2b\lambda x}, \quad f(x) \sim \frac{\pi^2}{4} b\lambda a^{2b} e^{-2b\lambda x}, \quad h(x) \sim 2ab\lambda.$$

Therefore, for all the values of the parameters, we have $f(x) \to 0$ and $h(x) \to 2ab\lambda$.

By using (5) and (6), any critical point of $f(x)$, say x_0, satisfies the following equation:

$$\begin{aligned}&-\lambda^2 + (a-1)\frac{\lambda e^{-\lambda x_0}}{1-e^{-\lambda x_0}} - (b-1)\frac{a\lambda e^{-\lambda x_0}(1-e^{-\lambda x_0})^{a-1}}{1-(1-e^{-\lambda x_0})^a}\\ &-\frac{\pi}{2} ab\lambda e^{-\lambda x_0}(1-e^{-\lambda x_0})^{a-1}[1-(1-e^{-\lambda x_0})^a]^{b-1}\times\\ &\cot\left(\frac{\pi}{2}[1-(1-e^{-\lambda x_0})^a]^b\right) = 0\end{aligned}$$

and any critical point of $h(x)$, say x_*, satisfies the following equation:

$$\begin{aligned}&-\lambda^2 + (a-1)\frac{\lambda e^{-\lambda x_*}}{1-e^{-\lambda x_*}} - (b-1)\frac{a\lambda e^{-\lambda x_*}(1-e^{-\lambda x_*})^{a-1}}{1-(1-e^{-\lambda x_*})^a}\\ &+\frac{\pi}{2} ab\lambda e^{-\lambda x_*}(1-e^{-\lambda x_*})^{a-1}[1-(1-e^{-\lambda x_*})^a]^{b-1}\times\\ &\left[\cot\left(\frac{\pi}{4}[1-(1-e^{-\lambda x_*})^a]^b\right) - \cot\left(\frac{\pi}{2}[1-(1-e^{-\lambda x_*})^a]^b\right)\right] = 0.\end{aligned}$$

They can be evaluated numerically. We illustrate the shapes of $f(x)$ and $h(x)$ in Figure 1 for selected values of a, b and λ.

(a) (b)

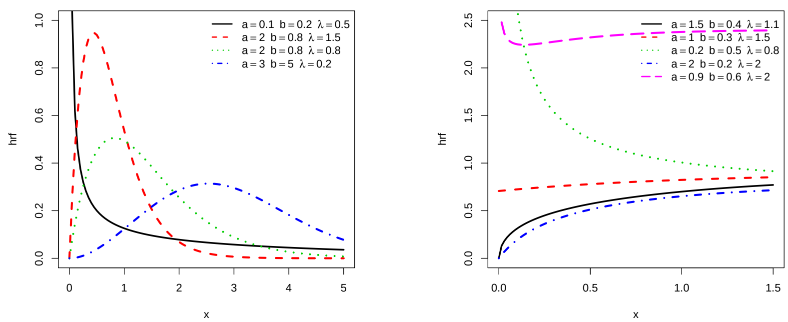

Figure 1: Plots of some (a) SKE pdfs and (b) SKE hrfs for selected values of a, b and λ.

Also, by (10), we can express $f(x)$ as an infinite linear combinations of exponential pdfs, i.e.,

$$f(x) = \sum_{m=1}^{+\infty} b_m^* [m\lambda e^{-\lambda m x}], \quad x > 0.$$

Let r be an integer. Then, the r-th moment of the SKE distribution exists. It can be expressed by an integral as in (11) or as the following linear representation:

$$\mu_r' = \sum_{m=1}^{+\infty} b_m^* \int_0^{+\infty} x^r [m\lambda e^{-\lambda m x}] dx = \lambda^{-r} \Gamma(r+1) \sum_{m=1}^{+\infty} b_m^* m^{-r},$$

where $\Gamma(x) = \int_0^{+\infty} u^{x-1} e^{-u} du$ (the gamma function). Also, one can remark that $\Gamma(r+1) = r!$. Table 1 presents the numerical values of the moments of order 1, 2, 3 and 4, the variance σ^2, the coefficient of skewness CS and the coefficient of kurtosis CK defined by (12) and (13), respectively, for selected values of a, b and λ.

Table 1: Some moments, variance, skewness and kurtosis of X for the SKE distribution for the following selected parameters values in order (a, b, λ); (i): $(1, 2, 5)$, (ii): $(3, 2, 5)$, (iii): $(1.5, 1, 5)$ (iv): $(1.5, 0.5, 0.5)$ (v): $(5, 6, 0.5)$ and (vi): $(30, 6, 0.5)$.

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
μ'_1	0.1984	0.0974	0.0742	0.2809	10.7531	1.3342
μ'_2	0.05637	0.0128	0.2613	2575.5120	126.3263	1.8944
μ'_3	0.02077	0.0020	0.4247	172252.4	1599.5430	2.8323
μ'_4	0.0094	0.0004	0.0005	0.9924	21625.63	4.4266
σ^2	0.0169	0.0033	0.0049	0.1823	10.6944	0.1141
CS	1.2845	1.00650	1.8457	3.1944	0.3182	0.0014
CK	5.5791	4.4545	8.1281	18.6453	3.1172	2.8808

From Table 1, we observe that the considered measures can take wide ranges of values, illustrating the flexibility of the SKE distribution on these aspects.

Other kinds of moments can be expressed. For instance, for $t \geq 0$, the r-th incomplete moment of the SKE distribution is given as

$$\mu_r^*(t) = \sum_{m=1}^{+\infty} b_m^* \int_0^t x^r [m\lambda e^{-\lambda m x}] dx = \lambda^{-r} \sum_{m=1}^{+\infty} b_m^* m^{-r} \gamma(r+1, \lambda m t),$$

where $\gamma(x, t) = \int_0^t u^{x-1} e^{-u} du$ (the lower incomplete gamma function).

Similarly, by using (15), the r-th probability weighted moment of the SKE distribution is

$$\mu'_{r,s} = \sum_{m=1}^{+\infty} e_{s,m}^* \int_0^{+\infty} x^r [m\lambda e^{-\lambda m x}] dx = \lambda^{-r} \Gamma(r+1) \sum_{m=1}^{+\infty} e_{s,m}^* m^{-r}.$$

Finally, we mention that all the results on order statistics presented in Subsection 2.9 can be applied, with the use of the probability weighted moments to express the (raw) moments of the i-th order statistic, as described in (16).

4 Estimation, simulation and applications

In this section, we investigate the SKE model governed by the cdf given by (17) (and the pdf given by (18)).

4.1 Estimation

We now examine the estimation of the parameters a, b and λ of the SKE model by using the maximum likelihood method, ensuring nice convergence properties of the obtained estimates called the maximum likelihood estimates (MLEs). Among others, they can be used to construct approximate confidence intervals for a, b and λ and test statistics. The essential of the method adapted to the SKE distribution is presented below. Let $x_1, \ldots, x_n$ be n independent observations from the SKE distribution with parameters a, b and λ. Then, the likelihood function for the vector of parameters $\Theta = (a, b, \lambda)^\top$ is defined by

$$\begin{aligned} L(\Theta) &= \prod_{i=1}^{n} f(x_i) \\ &= \left(\frac{\pi}{2} ab\lambda\right)^n \prod_{i=1}^{n} e^{-\lambda x_i}(1-e^{-\lambda x_i})^{a-1}[1-(1-e^{-\lambda x_i})^a]^{b-1} \times \\ &\sin\left(\frac{\pi}{2}[1-(1-e^{-\lambda x_i})^a]^b\right). \end{aligned}$$

Applying the logarithmic transformation, the corresponding log-likelihood function is given by

$$\begin{aligned} \ell(\Theta) = \log\left[L(\Theta)\right] &= n\log\left(\frac{\pi}{2}\right) + n\log(a) + n\log(b) + n\log(\lambda) - \lambda\sum_{i=1}^{n} x_i \\ &+ (a-1)\sum_{i=1}^{n}\log\left(1-e^{-\lambda x_i}\right) + (b-1)\sum_{i=1}^{n}\log\left[1-(1-e^{-\lambda x_i})^a\right] \\ &+ \sum_{i=1}^{n}\log\left[\sin\left(\frac{\pi}{2}[1-(1-e^{-\lambda x_i})^a]^b\right)\right]. \end{aligned}$$

Then, the related score vector is obtained as $U(\Theta) = (U_a(\Theta), U_b(\Theta), U_\lambda(\Theta))^\top$ with

$$U_a(\Theta) = \frac{\partial}{\partial a}\ell(\Theta) = \frac{n}{a} + \sum_{i=1}^{n} \log\left(1 - e^{-\lambda x_i}\right)$$
$$- (b-1)\sum_{i=1}^{n} \frac{(1-e^{-\lambda x_i})^a \log(1-e^{-\lambda x_i})}{1-(1-e^{-\lambda x_i})^a}$$
$$- \frac{\pi}{2} b \sum_{i=1}^{n} (1-e^{-\lambda x_i})^a \log(1-e^{-\lambda x_i})[1-(1-e^{-\lambda x_i})^a]^{b-1} \times$$
$$\cot\left(\frac{\pi}{2}[1-(1-e^{-\lambda x_i})^a]^b\right),$$

$$U_b(\Theta) = \frac{\partial}{\partial b}\ell(\Theta) = \frac{n}{b} + \sum_{i=1}^{n} \log\left[1-(1-e^{-\lambda x_i})^a\right]$$
$$+ \frac{\pi}{2}\sum_{i=1}^{n}[1-(1-e^{-\lambda x_i})^a]^b \log\left[1-(1-e^{-\lambda x_i})^a\right] \times$$
$$\cot\left(\frac{\pi}{2}[1-(1-e^{-\lambda x_i})^a]^b\right)$$

and

$$U_\lambda(\Theta) = \frac{\partial}{\partial \lambda}\ell(\Theta) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i + (a-1)\sum_{i=1}^{n} \frac{x_i e^{-\lambda x_i}}{1-e^{-\lambda x_i}}$$
$$- a(b-1)\sum_{i=1}^{n} \frac{x_i e^{-\lambda x_i}(1-e^{-\lambda x_i})^{a-1}}{1-(1-e^{-\lambda x_i})^a}$$
$$- \frac{\pi}{2} ab \sum_{i=1}^{n} x_i e^{-\lambda x_i}(1-e^{-\lambda x_i})^{a-1}[1-(1-e^{-\lambda x_i})^a]^{b-1} \times$$
$$\cot\left(\frac{\pi}{2}[1-(1-e^{-\lambda x_i})^a]^b\right).$$

The MLEs of a, b and λ, denoted by $\hat{a}$, $\hat{b}$ and $\hat{\lambda}$, respectively, satisfy the system of equations: $U(\hat{\Theta}) = (0,0,0)^\top$, with $\hat{\Theta} = (\hat{a}, \hat{b}, \hat{\lambda})^\top$. There are no closed forms for these estimates. However, they can be obtained numerically with efficient iterative algorithms (see [17]). Under regularity conditions, the underlying distribution of $\hat{\Theta}$ can be approximated by a 3 dimensional normal distribution with mean Θ and covariance matrix given as $J(\Theta)^{-1} \mid_{\Theta=\hat{\Theta}}$, where $J(\Theta) = -\partial^2 \ell(\Theta)/\partial\Theta\partial\Theta^T$. Then, for $h \in \{a, b, \lambda\}$, an approximate confidence interval for h at the level $100(1-\omega)\%$ is given by

$$CI_h = [\hat{h} - z_\omega s_{\hat{h}}, \hat{h} + z_\omega s_{\hat{h}}], \tag{20}$$

where $s_{\hat{h}}$ is the square-root of the diagonal element of $J(\hat{\Theta})^{-1}$ at the same position as h corresponding to the standard error (SE) of h and $z_\omega = Q_Z(1-\omega/2)$, where $Q_Z(x)$ is the qf of a standard normal random variable Z. Note that, for $\omega = 0.05$, we have $z_\omega = 1.959964$ and for $\omega = 0.01$, we have $z_\omega = 2.575829$.

The likelihood ratio (LR) statistic for testing goodness-of-fit of the SKE model with its sub-models can also be described. Thus, we can consider hypotheses of the form: $H_0 : \Theta = \Theta_0$ versus $H_1 : \Theta \neq \Theta_0$, where Θ_0 denotes a vector of 3 fixed values. In this case, the LR statistic is given by

$$LR = 2[\ell(\hat{\Theta}) - \ell(\Theta_0)], \tag{21}$$

where $\hat{\Theta}_0$ contains the MLEs of a, b and λ under H_0. Then, if H_0 is assumed to be true, the subjacent distribution of LR converges in distribution to a random variable K following the chi square distribution with r degrees of freedom, where r is equal to the difference between the number of parameters estimated in the general case and the number of parameters estimated under H_0. The corresponding p-value is defined by

$$p = \mathbb{P}(K > LR). \tag{22}$$

In our study, it is useful to check if the SKE model is superior in fitting to the SE model defined with the cdf $F(x) = \cos\left((\pi/2)e^{-\lambda x}\right)$, $x > 0$, for a given data set.

4.2 Simulation

The following result in distribution holds. For a random variable U following the uniform distribution on the unit interval, by using the qf given by (19), the random variable X defined by

$$X = Q(U) = -\frac{1}{\lambda}\log\left(1 - \left[1 - \left\{\frac{2}{\pi}\arccos(U)\right\}^{1/b}\right]^{1/a}\right)$$

follows the SKE distribution with parameters a, b and λ. Based on this result, we can simulate data distributed following the SKE distribution. Here, we use this result to evaluate the performance of the MLEs of the SKE parameters via a graphical Monte Carlo simulation study. All the computations are done by using the software R. We generate $N = 3000$ samples samples of size $n = 20, 40, ..., 500$ from the SKE distribution with true parameters values I: $a = 2.5$, $b = 5$, $\lambda = 1.5$, II: $a = 2.5$, $b = 3$, $\lambda = 1.5$ and III: $a = 2.5$, $b = 5.5$, $\lambda = 3$. We also calculate the mean square error (MSE) of the MLEs empirically. For $h \in \{a, b, \lambda\}$, we consider the empirical MSE corresponding to h defined by

$$MSE_h = \frac{1}{N}\sum_{i=1}^{N}(\hat{h}_i - h)^2,$$

where $\hat{h}_i$ denotes the MLE of h determined at the i-th repetition of the simulation. The obtained results are given in Figures 2, 3 and 4.

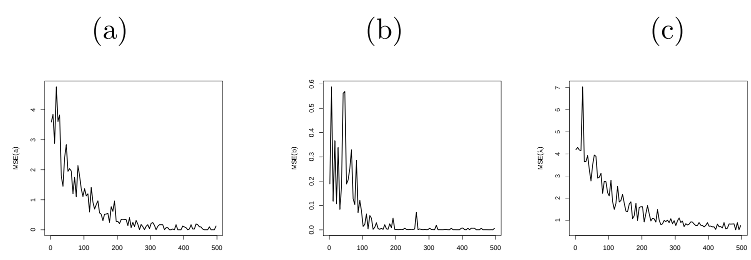

Figure 2: The MSE plots for the selected parameter values I for the SKE distribution, i.e., $a = 2.5$, $b = 5$, $\lambda = 1.5$.

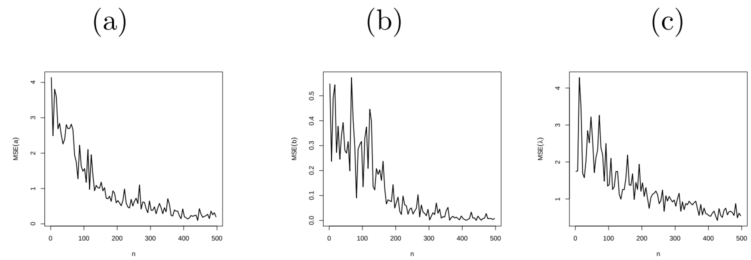

Figure 3: The MSE plots for the selected parameter values II for the SKE distribution, i.e., $a = 2.5$, $b = 3$, $\lambda = 1.5$.

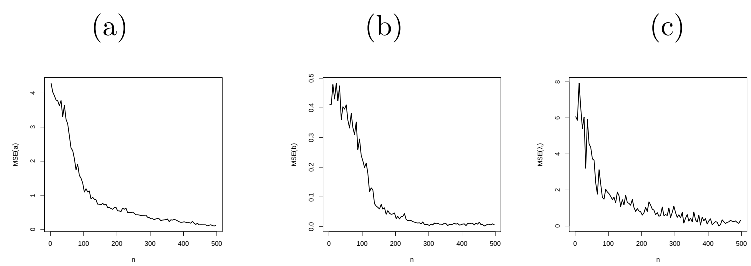

Figure 4: The MSE plots for the selected parameter values III for the SKE distribution, i.e., $a = 2.5$, $b = 5.5$, $\lambda = 3$.

In each figure, we observe that, when the sample size increases, the empirical MSEs tend to zero in all cases. This is consistent with the subjacent theory of the MLEs.

4.3 Applications

In this subsection, the flexibility of the SKE model is shown by means of two real data sets. Also, the SKE model is compared with the four competitive models listed in Table 2. The following standard statistics are used: $-\widehat{\ell}$ where $\widehat{\ell}$ denotes the maximized log-likelihood, AIC

(Akaike information criterion), BIC (Bayesian information criterion), CVM (Cramér-von Mises), AD (Anderson-Darling) and KS (Kolmogorov-Smirnov), consistent Akaike information criterion (CAIC), and Hannan-Quinn information criterion (HQIC). All the computations are done by using the software R.

Table 2: The considered competitive models of the SKE model.

Model	Reference
Kumaraswamy Weibull (KW)	[5]
Beta Weibull (BW)	[13]
CS transformation of exponential ($CS1_E$)	[3]
Exponential (E)	Standard

The first application uses a real data set given by [9]. It consists of thirty successive values of March precipitation (in inches) in Minneapolis/St Paul. The data are: 0.77, 1.74, 0.81, 1.20, 1.95, 1.20, 0.47, 1.43, 3.37, 2.20, 3.00, 3.09, 1.51, 2.10, 0.52, 1.62, 1.31, 0.32, 0.59, 0.81, 2.81, 1.87, 1.18, 1.35, 4.75, 2.48, 0.96, 1.89, 0.90, 2.05.

The second data set represents the tensile strength data measured in GPa for single carbon fibers. It is from [18]. The data are: 0.312, 0.314, 0.479, 0.552, 0.700, 0.803, 0.861, 0.865, 0.944, 0.958, 0.966, 0.997, 1.006, 1.021, 1.027, 1.055, 1.063, 1.098, 1.140, 1.179, 1.224, 1.240, 1.253, 1.270, 1.272, 1.274, 1.301, 1.301, 1.359, 1.382, 1.382, 1.426, 1.434, 1.435, 1.478, 1.490, 1.511, 1.514, 1.535, 1.554, 1.566, 1.570, 1.586, 1.629, 1.633, 1.642, 1.648, 1.684, 1.697, 1.726, 1.770, 1.773, 1.800, 1.809, 1.818, 1.821, 1.848, 1.880, 1.954, 2.012, 2.067, 2.084, 2.090, 2.096, 2.128, 2.233, 2.433, 2.585, 2.585.

Analysis of data set 1. For data set 1, descriptive statistics are given in Table 3. In particular, we see that the subjacent distribution of this data set is left-skewed (skewness estimated to 1.0866) with a non-negligible tail (kurtosis estimated to 1.2068). Table 4 provides the values of goodness-of-fit measures for the SKE model and other fitted models. We see that the SKE model has the lowest statistics, indicating that it provides a better fit to the considered competitors. The MLEs and their corresponding SEs (in parentheses) are listed in Table 5. The probability-probability (P-P), quantile-quantile (Q-Q), empirical probability density function (epdf) and empirical cumulative density function (ecdf) plots of the SKE are shown in Figure 5. In each case, a nice fit is observed, indicating that the SKE model is appropriate for the analysis of data set 1. To complete this analysis, we provide in Table 6 the approximation confidence intervals of the parameters of the SKE model (see (20)). The levels 95% and 99% are considered. Finally, a LR test with the hypotheses: $H_0 : a = b = 1$ versus $H_1 : a \neq 1$ or $b \neq 1$, is performed in Table 7 (the formulas (21) and (22) are used). The p-value, which is based on the chi-square distribution with 2 degree of freedom, satisfies p-value < 0.0001. This shows the importance of the parameters a and b in terms of fit for data set 1 in comparison to the former SE model.

Table 3: Some descriptive statistics for data set 1.

Statistics	N	Mean	Median	Variance	skewness	kurtosis
Data set 1	30	1.6750	1.4700	1.0012	1.0866	1.2068

Table 4: Goodness-of-fit measures for data set 1.

Model	$-\widehat{\ell}$	AIC	BIC	CAIC	HQIC	KS	CVM	AD
SKE	36.8774	81.1549	85.3585	82.0780	82.4997	0.0635	0.0112	0.1041
KW	37.9766	83.9533	89.5581	85.5533	85.7463	0.0681	0.0148	0.1065
BW	38.0700	84.1400	89.7448	85.7400	85.9330	0.0631	0.0144	0.1045
$CS1_E$	41.8522	89.7045	93.9081	90.6276	91.0493	0.0964	0.0702	0.4887
E	45.4743	92.9480	94.3499	93.0916	93.3970	0.2351	0.0195	0.1086

Table 5: MLEs and SEs (in parentheses) for data set 1.

Model	Estimates			
SKE	3.7201	0.3802	1.5250	
(a, b, λ)	(0.6010)	(0.1086)	(0.2959)	
KW	2.8788	0.1685	2.9571	1.4502
(a, b, α, β)	(1.4350)	(0.0467)	(0.1595)	(0.1688)
BW	0.3536	0.8078	4.4861	5.5074
(a, b, α, β)	(2.7762)	(0.9862)	(9.9203)	(2.1934)
$CS1_E$	0.8412	9.7350	0.5383	
$(\alpha, \theta, \lambda)$	(1.3128)	(1.5192)	(0.0865)	
E	0.5969			
(λ)	(0.1089)			

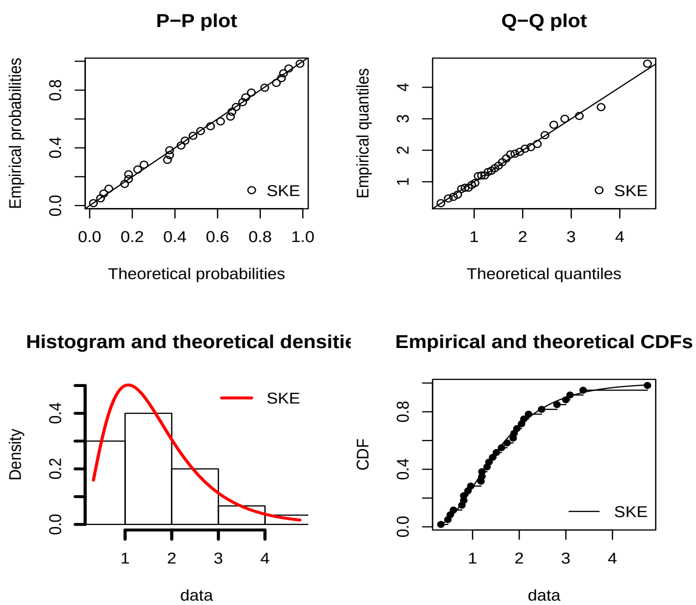

Figure 5: P-P, Q-Q, epdf and ecdf plots of the SKE distribution for data set 1.

Table 6: Confidence intervals for the parameters of the SKE model for data set 1.

CI	a	b	λ
95%	[2.5421 4.4989]	[0.1673 0.5930]	[0.9450 2.1049]
99%	[2.1695 5.2706]	[0.1000 0.6603]	[0.7615 2.2884]

Table 7: LR test for data set 1.

Idea	H_0	LR	p-value
SKE versus SE [12]	$a = b = 1$	17.1938	< 0.001 (***)

Analysis of data set 2. For data set 2, we adopt the same methodology to the one used for the analysis of data set 1. Thus, some descriptive statistics are presented in Table 8. Since the estimated skewness is close to zero, the subjacent distribution is near symmetric around its mean. The values of the goodness-of-fit measures for the SKE model and other fitted models are collected in Table 9, whereas the MLEs and their corresponding SEs are listed in Table 10. Again, we see that the SKE model has the lowest statistics, indicating that it is statistically superior to the competitors. The P-P, Q-Q, epdf and ecdf plots of the SKE are presented in Figure 6. We see nice fits, indicating that the SKE model is a good choice for the analysis of data set 2. Then, we provide the approximation confidence intervals of the parameters of the SKE model in Table 11, for the levels 95% and 99%. Finally, a LR test with the hypotheses: $H_0 : a = b = 1$ versus $H_1 : a \neq 1$ or $b \neq 1$, is performed in Table 12. The p-value satisfies p-value < 0.0001, indicating that the SKE model is again preferable to the SE model.

Table 8: Some descriptive statistics for data set 2.

Statistics	N	Mean	Median	Variance	skewness	kurtosis
Data set 2	69	1.4513	1.4780	0.2451	-0.02821	-0.05927

Table 9: Goodness-of-fit measures for data set 2.

Model	$-\widehat{\ell}$	AIC	BIC	CAIC	HQIC	KS	CVM	AD
SKE	48.1311	104.2624	110.9647	104.6316	106.9214	0.0455	0.0211	0.1977
KW	48.7684	105.5368	114.4733	106.1618	109.0822	0.0475	0.0226	0.1984
BW	48.8954	105.7908	114.7272	106.4158	109.3362	0.0480	0.0256	0.2217
$CS1_E$	49.5405	105.0810	111.7833	105.4502	107.7400	0.0487	0.0279	0.1989
E	94.7013	191.4026	193.6367	191.4623	192.2890	0.3622	0.1238	0.8712

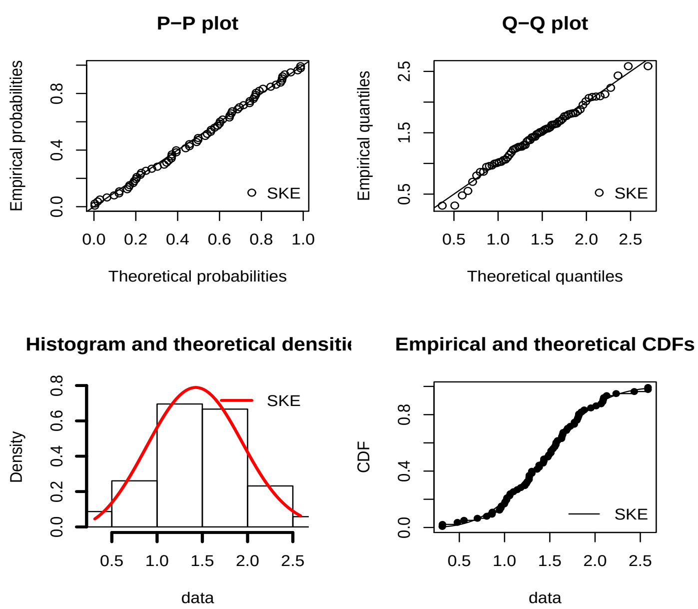

Figure 6: P-P, Q-Q, epdf and ecdf plots of the SKE distribution for data set 2.

Table 10: MLEs and SEs (in parentheses) for data set 2.

Model	Estimates			
SKE	3.5848	50.6984	0.2100	
(a, b, λ)	(0.5853)	(4.0734)	(0.1626)	
KW	0.7268	0.1621	1.0308	3.5369
(a, b, α, β)	(0.0052)	(0.0186)	(0.0218)	(0.0086)
BW	0.3585	3.7827	0.7813	5.7953
(a, b, α, β)	(2.0367)	(1.2916)	(0.4105)	(2.5127)
$CS1_E$	0.0916	10.7578	0.2785	
$(\alpha, \theta, \lambda)$	(1.0176)	(11.6449)	(0.0276)	
E	0.5969			
(λ)	(0.1089)			

Table 11: Confidence intervals for the parameters of the SKE model for data set 2.

CI	a	b	λ
95%	[2.4376 4.3433]	[42.7146 58.6822]	[0 0.5286]
99%	[2.0747 5.0948]	[40.1890 61.2077]	[0 0.6295]

5 Conclusions

In the last decade, the trigonometric families of distributions have received a lot of attention, mainly thanks to their flexible properties in terms of fitting a wide variety of real data sets. In this study, we explore a natural extension of the sine-G family of distributions, called the sine Kumaraswamy-G family of distributions. We investigate its main mathematical properties, including asymptotes, quantile function, linear representations of the cumulative distribution and probability density functions, moments, skewness and kurtosis, incomplete moments, probability

Table 12: LR test for data set 2.

Idea	H_0	LR	p-value
SKE versus SE [12]	$a = b = 1$	93.1404	< 0.001 (***)

weighted moments and order statistics. Then, a special focus is done on the sine Kumaraswamy exponential distribution, a notable member of this family. After presenting its mathematical features, we study the ability of the related model in the fitting of data sets. The maximum likelihood method is used to estimate the unknown parameters and a simulation study gives numerical guarantees of their performance. Applications to two practical data sets are presented in detail, showing that the proposed model outperforms some strong well-established competitors in the literature. We hope that the sine Kumaraswamy-G family of distributions and the related perspective of models may attract wider applications in statistics in various areas.

Acknowledgments

We thank the referees for their constructive comments which have helped to improve the paper.

References

[1] H. Al-Mofleh, On generating a new family of distributions using the tangent function, *Pakistan J. Stat. Oper. Res.*, 14 (2018), 471-499.

[2] C. B. Ampadu, The Tan-G family of distributions with illustration to data in the health sciences, *Phys. Sci. and Biophysics J.*, 3 (2019), 000125.

[3] C. Chesneau, H. S. Bakouch and T. Hussain, A new class of probability distributions via cosine and sine functions with applications, *Comm. Statist. Simulation Comput.*, 48 (2019), 2287-2300.

[4] G. M. Cordeiro and M. de Castro, A new family of generalized distributions, *J. Stat. Comput. Simul.*, 81 (2011), 883-893.

[5] G. M. Cordeiro, E. M. M. Ortega and S. Nadarajah, The Kumaraswamy Weibull distribution with application to failure data, *J. Franklin Inst.*, 347 (2010), 1399-1429.

[6] H. A. David and H. N. Nagaraja, *Order Statistics*, John Wiley and Sons, New Jersey, 2003.

[7] M. A. R. de Pascoa, E. M. M. Ortega and G. M. Cordeiro, The Kumaraswamy Weibull distribution with application to failure data, *J. Franklin Inst.*, 347 (2011), 1399-1429.

[8] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, New York, 2000.

[9] D. Hinkley, On quick choice of power transformations, *J. Roy. Statist. Soc. Ser. C*, 26 (1977), 67-69.

[10] F. Jamal and C. Chesneau, A new family of polyno-expo-trigonometric distributions with applications, *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 22 (2020), 1-15.

[11] J. F. Kenney and E. S. Keeping, *Mathematics of Statistics*, 3 edn, Chapman and Hall Ltd, New Jersey, 1962.

[12] D. Kumar, U. Singh and S. K. Singh, A new distribution using sine function- its application to bladder cancer patients data, *J. Stat. Appl. Prob.*, 4 (2015), 417-427.

[13] C. Lee, F. Famoye and O. Olumolade, Beta-Weibull distribution: some properties and applications to censored data, *J. Mod. Appl. Stat. Methods*, 6 (2007), 173-186.

[14] Z. Mahmood, C. Chesneau and M. H. Tahir, A new sine-G family of distributions: properties and applications, *Bull. Comput. Appl. Math.*, 7 (2019), 53-81.

[15] J. J. Moors, A quantile alternative for kurtosis, *J. Roy. Statist. Soc. Ser. D*, 37 (1988), 25-32.

[16] M. Pararai, G. Warahena-Liyanage and B. O. Oluyede, A new class of generalized inverse Weibull distribution with applications, *J. Appl. Math. Bioinform.*, 4 (2014), 17-35.

[17] W. H. Press, A. A. Teukolsky, W. T. Vetterling and Y. B. P. Flanner, *Numerical Recipes in C: The Art of Scientific Computing*, 3 edn, Cambridge University Press, New York, 2007.

[18] M. Raqab, T. Madi and K. Debasis, Estimation of $P(Y < X)$ for the 3-parameter generalized exponential distribution, *Comm. Statist. Theory Methods*, 37 (2008), 2854-2864.

[19] J. A. Rodrigues and A. P C. Silva, The exponentiated Kumaraswamy-exponential distribution, *Br. J. Appl. Sci. Technol.*, 10 (2015), 1-12.

[20] L. Souza, *New Trigonometric Classes of Probabilistic Distributions*, Thesis, Universidade Federal Rural de Pernambuco, 2015.

[21] L. Souza, W. R. O. Junior, C. C. R. de Brito, C. Chesneau, R. L. Fernandes and T. A. E. Ferreira, Tan-G class of trigonometric distributions and its applications, *Cubo* (to appear) (2021).

[22] L. Souza, W. R. O. Junior, C. C. R. de Brito, C. Chesneau and T. A. E. Ferreira, Sec-G class of distributions: Properties and applications, *preprint* (2019).

[23] L. Souza, W. R. O. Junior, C. C. R. de Brito, C. Chesneau, T. A. E. Ferreira and L. Soares, On the Sin-G class of distributions: theory, model and application, *J. Math. Model.*, 7 (2019), 357-379.

[24] L. Souza, W. R. O. Junior, C. C. R. de Brito, C. Chesneau, T. A. E. Ferreira and L. Soares, General properties for the Cos-G class of distributions with applications, *Eurasian Bull. J.*, 2 (2019), 63-79.

Christophe Chesneau
Department of Mathematics
Assistant Professor
LMNO, University of Caen-Normandie
Caen, France

E-mail: christophe.chesneau@gmail.com

Farrukh Jamal
Department of Statistics
Assistant Professor
The Islamia University of Bahawalpur
Punjab, Pakistan
E-mail: drfarrukh1982@gmail.com