

A NEW TRIGONOMETRIC DISTRIBUTION WITH BOUNDED SUPPORT AND AN APPLICATION

AHMED M. T. ABD EL-BAR, HASSAN S. BAKOUCH, AND SHOVAN CHOWDHURY

ABSTRACT. In this paper, we introduce a new bounded distribution by using trigonometric functions, named the cosine-sine distribution. A comprehensive study of its statistical properties is presented along with an application to a unit-interval data set, namely firms risk management cost-effectiveness data. The proposed distribution has increasing, bathtub and v hazard rate shapes. Further, we show that the distribution can be viewed as a truncated exponential sine distribution.

1. INTRODUCTION

In statistical literature, many extensions of the lifetime distributions have been studied by several researchers over the last two decades. Most of these extensions concern distributions with unbounded support, while there is a great scarcity of distributions with bounded support (see Marshall and Olkin [14]). In many practical situations, a common issue is dealing with uncertain phenomena observed in the bounded interval $(0, 1)$. For example, in real-life situations we often encounter measures such as percentages, proportions or fractions. In reliability and survival analysis, failure of a device during the warranty period may not be counted. Moreover, test conditions, cost or other constraints may lead many reliability systems to be bounded above (Lai and Jones [12], Lai and Mukherjee [13], Mukherjee and Islam [15], Jiang [9]). Also, models defined on the unit interval are recommended when the reliability is measured as the ratio of the number of successful trials to the number of total trials (Genç [5]). Further, continuous distributions having a finite support are highly recommended to model and analyze reliability data because of physical reasons, such as the finite lifetime of a component system or the bounded signals occurring in industrial systems (see Jiang [9]). In insurance, a probability distribution with domain on $(0, 1)$ can be used as a distortion function to define a premium principle (Gómez-Déniz et al. [6], Chowdhury and Nanda [3]). Further, Papke and Wooldridge [16] claim that variables whose values lie in $(0, 1)$ arise naturally in many economic settings, such as the fraction of land area allocated to

2020 *Mathematics Subject Classification.* 60E05, 62E15, 62F10.

Key words and phrases. Trigonometric distributions; moments; residual life function; entropy; goodness-of-fit statistics.

agriculture, proportion of income spent on non-durable consumption, pension plan participation rates, etc. Moreover, examples of proportions in empirical finance can be found in Cook et al. [4].

All existing bounded probability distributions are mostly specialized to the beta distribution. On the other hand, power function density, finite range density, truncated Weibull distribution, Kumaraswamy and Log-Lindley distributions are used occasionally to model unit-interval data.

Further, the statistical literature has a lack of distributions based on trigonometric functions. Among the distributions that used trigonometric functions are the von Mises distribution (see Evans et al. [8, pp. 189–191]), the circular Cauchy distribution introduced by Kent and Tyler [10], and the sine square distribution explored by Al-Faris and Khan [2].

Motivated by the preceding discussion, there is a need to introduce other unit-interval distributions besides the ones already existing in the literature; hence we introduce a new trigonometric distribution with bounded support called cosine-sine (CS) distribution. This model, when compared to beta, Kumaraswamy, and Log-Lindley distributions, offers more flexibility in modeling and evaluating actual data, as will be demonstrated later by a practical example. A comprehensive account of statistical properties of the distribution is investigated along with an application to a unit-interval data set, namely firms risk management cost-effectiveness data. There are other features that motivate the proposed distribution. First, it has a closed form expression for the cumulative distribution function (cdf) and various shapes of its hazard rate function, namely increasing, bathtub and v-shapes. Also, it can be displayed as a truncated exponential sine distribution. Moreover, the probability density of this model decays slowly and rapidly to 0 at the end of the distribution support based on the distribution parameter values, which may be a desired feature for some real-life models.

The remainder of this paper is structured as follows. Section 2 proposes the CS distribution with some related functions. In Section 3, various probabilistic properties of the CS model are obtained, including moments, conditional moments, and moment generating function. Shannon entropy and Rényi entropy are investigated in Section 4 with some numerical results. In Section 5, we discuss the residual life and reversed residual life random variables, and get some of their related statistical functions, such as survival function, mean and variance of the (reversed) residual life function for the CS distribution. In Section 6, the maximum likelihood estimator of the parameter of the CS model is obtained. The CS model and other distributions are fitted to a practical data set, and the usefulness of the model is illustrated by exhibiting a data set for which the CS model fits better than other existing parametric models, as indicated by several goodness-of-fit measures. Concluding remarks are presented in Section 7.

2. THE CS DISTRIBUTION

The parametric class that we will study in this paper is comprised of laws indexed by a parameter $\lambda \in R^+$. Each law in the class is that of a random variable

$X = \arcsin\left(\frac{\pi}{2}Y\right)$, where Y has a truncated exponential distribution $(0, 1)$ with parameter λ , i.e., Y has density

$$g(y) = \frac{\lambda e^{-\lambda y}}{1 - e^{-\lambda}} I_{(0,1)}(y), \quad 0 < y < 1.$$

Straightforward calculations, using the change of variable formula, lead to the following expression of the CS density function:

$$f(x; \lambda) = \frac{\pi\lambda}{2(1 - e^{-\lambda})} \cos(\pi x/2) e^{-\lambda \sin(\pi x/2)} I_{(0,1)}(x), \quad 0 < x < 1, \lambda > 0, \quad (2.1)$$

and the cumulative distribution function (cdf) of the CS model is

$$F(x; \lambda) = \frac{1}{1 - e^{-\lambda}} \left(1 - e^{-\lambda \sin(\pi x/2)}\right) I_{(0,1)}(x) + I_{[1,\infty)}(x), \quad 0 < x < 1,$$

which has a closed form. The corresponding hazard rate (hr) function is

$$h(x) = \frac{\pi\lambda}{2} \cos(\pi x/2) e^{-\lambda \sin(\pi x/2)} \left(e^{-\lambda \sin(\pi x/2)} - e^{-\lambda}\right)^{-1}, \quad 0 < x < 1. \quad (2.2)$$

In Figures 1 and 2, we plot the pdf and hrf of the CS model for different values of the parameter λ . From Figure 1, we note that the pdf $f(x; \lambda)$ decays rapidly to 0

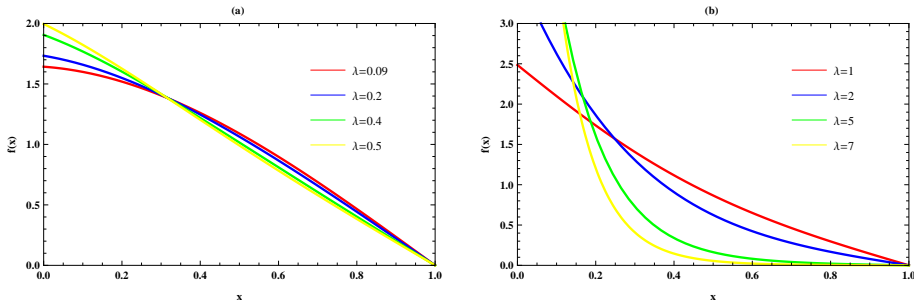


FIGURE 1. Plots of the CS pdfs.

as x goes to 1 with $\lambda < 1$ and slowly when $\lambda \geq 1$. Figure 2 shows various shapes for the hrf, including increasing, bathtub and v-shapes. This variety of shapes may be due to trigonometric functions in expression (2.2), which makes the distribution capable of analyzing many practical data.

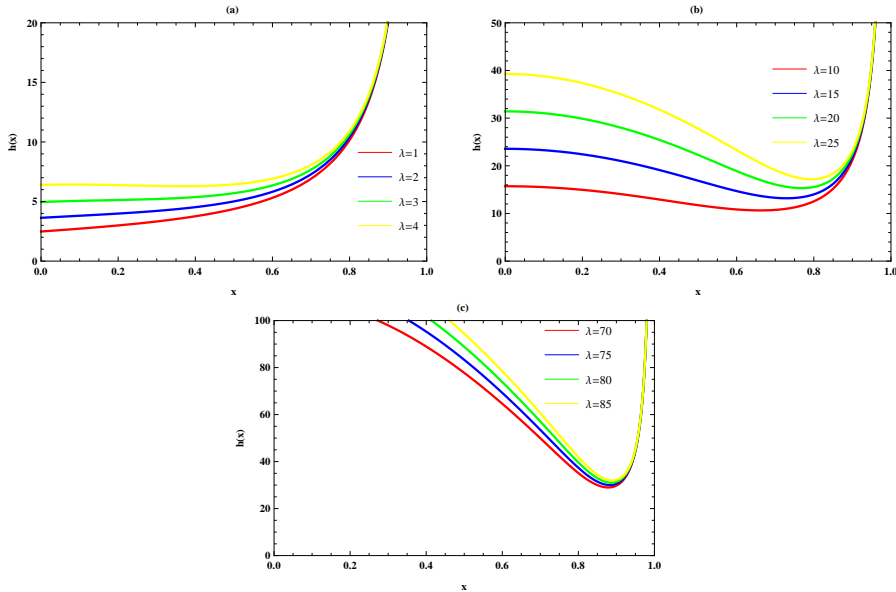


FIGURE 2. Plots of the CS hrfs

Throughout the rest of the paper we will use the following definitions and results. For $m > 0$ and $r = 1, 2, \dots$, we define

$$\begin{aligned}
 v(m; r, \lambda) &= \int_0^m x^r f(x; \lambda) dx \\
 &= \frac{2^r}{\pi^r (1 - e^{-\lambda})} \\
 &\quad \times \sum_{k=0}^{\infty} \varphi_{r,k} \frac{1}{\lambda^{r+2k}} [\Gamma(r + 2k + 1) - \Gamma(r + 2k + 1, \lambda \sin(\pi m/2))].
 \end{aligned}
 \tag{2.3}$$

For $k = 1, 2, \dots$, let $b_k = \frac{(2k)!}{2^{2k}(k!)^2(2k+1)}$; for any fixed r we define $\varphi_{r,0} = b_0^r$ and $\varphi_{r,k} = (kb_0)^{-1} \sum_{n=1}^k [n(r+1) - k] b_n \varphi_{r,k-n}$, where $\Gamma(\cdot)$ is the gamma function and $\Gamma(\cdot, \cdot)$ denotes the incomplete gamma function (see Abramowitz and Stegun [1]).

The entropy formulae that we will derive later will make use of the following functions.

(i) The modified Bessel function of the first kind is given by

$$I_\nu(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k+\nu}}{k! \Gamma(\nu + k + 1)},$$

where $\nu \in \mathbb{R}$, $z \in \mathbb{C}$ and

$$I_\nu^{(1,0)}(z) = I_\nu(z) \log(z/2) - \sum_{k=0}^{\infty} \frac{\varpi(k + \nu + 1)}{k! \Gamma(\nu + k + 1)},$$

where $\varpi(\cdot)$ is the polygamma function defined as

$$\varpi^{(n)}(z) = \frac{d^n}{dz^n} \varpi(z) = \frac{d^{n+1}}{dz^{n+1}} \ln \Gamma(z), \quad n = 1, 2, 3, \dots$$

(see Abramowitz and Stegun [1]).

(ii) The modified Struve function is defined by

$$L_\nu(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k+\nu+1}}{\Gamma(k + 3/2) \Gamma(k + \nu + 3/2)},$$

and

$$L_\nu^{(1,0)}(z) = L_\nu \log(z/2) - \sum_{k=0}^{\infty} \frac{\varpi(k + \nu + 3/2)}{\Gamma(k + 3/2) \Gamma(k + \nu + 3/2)}.$$

3. MOMENTS

In this section, we derive an expression for the marginal and conditional moments of the random variable X as a function of λ .

Theorem 3.1. *Let X be a random variable with CS density function (2.1). Then the r -th moment about the origin of X is given, for $r = 1, 2, \dots$, by*

$$E(X^r) = \frac{2^r}{\pi^r (1 - e^{-\lambda})} \sum_{k=0}^{\infty} \varphi_{r,k} \frac{1}{\lambda^{r+2k}} [\Gamma(r + 2k + 1) - \Gamma(r + 2k + 1, \lambda)]. \quad (3.1)$$

Proof. Using the definition of the r -th moment of X around the origin, we have

$$\begin{aligned} E(X^r) &= \frac{\pi \lambda}{2(1 - e^{-\lambda})} \int_0^1 x^r \cos(\pi x/2) e^{-\lambda \sin(\pi x/2)} dx \\ &= \frac{2^r \lambda}{\pi^r (1 - e^{-\lambda})} \int_0^1 (\sin^{-1} u)^r e^{-\lambda u} du. \end{aligned}$$

Using the expressions

$$\sin^{-1} x = \sum_{k=0}^{\infty} \frac{(2k)!}{2^{2k} (k!)^2 (2k + 1)} x^{2k+1}$$

and

$$\left(\sum_{k=0}^{\infty} b_k u^{2k+1} \right)^r = \sum_{k=0}^{\infty} \varphi_{r,k} u^{2k+r}$$

(see Gradshteyn and Ryzhik [7]) in the successive substitutions, Eq. (3.1) is attained. In particular, using (2.3), the mean of the CS distribution follows as

$$E(X) = \frac{2}{\pi (1 - e^{-\lambda})} \sum_{k=0}^{\infty} \varphi_{1,k} \frac{1}{\lambda^{2k+1}} [\Gamma(2k + 2) - \Gamma(2k + 2, \lambda)]. \quad (3.2) \quad \square$$

In connection with lifetime distribution, it is important to determine the conditional moments $E(X^r | X > t)$, $r = 1, 2, \dots$, which are of interest in predictive inference.

Proposition 3.2. *The conditional moments of the CS distribution are given by*

$$E(X^r | X > t) = \frac{1 - e^{-\lambda}}{e^{-\lambda \sin(\pi t/2)} - e^{-\lambda}} [E(X^r) - v(t; r, \lambda)], \quad t \in (0, 1),$$

where $E(X^r)$ is given in (3.1) and $v(t; r, \lambda)$ is given by (2.3).

Proof. The proof follows by applying the identity

$$E(X^r | X > t) = \frac{1}{S(t)} \left[E(X^r) - \int_0^t x^r f(x; \lambda) dx \right]. \quad \square$$

4. STOCHASTIC ORDERING AND ENTROPIES

In this section we show that the CS model can be ordered in terms of the likelihood ratio order, which is a powerful tool in parametric models. We also provide expressions for the Shannon and Rényi entropies for this model.

4.1. Stochastic ordering. The next theorem shows that the CS model can be ordered in terms of the likelihood ratio ordering.

Theorem 4.1. *Let X and Y be CS random variables with pdf $f(x; \lambda_1)$ and $f(x; \lambda_2)$, respectively. If $\lambda_2 < \lambda_1$, then X is said to be smaller than Y in the likelihood ratio order (denoted by $X \leq_{lr} Y$).*

Proof. The density ratio is given by

$$\frac{f_X(x)}{f_Y(x)} = \frac{\lambda_1 (1 - e^{-\lambda_2})}{\lambda_2 (1 - e^{-\lambda_1})} e^{(\lambda_2 - \lambda_1) \sin(\frac{\pi x}{2})}.$$

Hence, it follows that

$$\frac{d}{dx} \left[\frac{f_X(x)}{f_Y(x)} \right] = (\lambda_2 - \lambda_1) \frac{\pi}{2} \cos\left(\frac{\pi x}{2}\right) \frac{f_X(x)}{f_Y(x)}.$$

Since $\lambda_2 < \lambda_1$, we have $\frac{d}{dx} \left[\frac{f_X(x)}{f_Y(x)} \right] < 0$. Hence $\frac{f_X(x)}{f_Y(x)}$ is decreasing in x . That is, $X \leq_{lr} Y$, which completes the proof. □

4.2. Entropies. Entropy is a measure of uncertainty variation of a system modeled by a probability distribution $f(x)$. Two popular entropy measures of $f(x)$ are the Shannon entropy (Shannon [18]) and Rényi entropy (Rényi [17]), defined as

$$\Lambda = E(-\log[f(x)]) \tag{4.1}$$

and

$$\Omega = \frac{1}{1 - \delta} \log \left[\int_R f^\delta(x) dx \right], \quad \delta > 0, \delta \neq 1, \tag{4.2}$$

respectively. The next two theorems explore explicit expressions of such entropies for the CS model.

Theorem 4.2. *The Shannon entropy for the CS model is given by*

$$\begin{aligned} \Lambda &= \log [2 (1 - e^{-\lambda})] - \log [\pi\lambda] + \frac{1 - (1 + \lambda) e^{-\lambda}}{\lambda (1 - e^{-\lambda})} \\ &\quad - \frac{\sqrt{2\pi^3\lambda^3}}{4(1 - e^{-\lambda})} \left(I^{(1,0)} \left[\frac{1}{2}, \lambda \right] - L^{(1,0)} \left[\frac{1}{2}, \lambda \right] \right) \\ &\quad - \frac{\pi\lambda}{4(1 - e^{-\lambda})} [(2\gamma - \log [4] + 2 \log [\lambda]) (\cosh (\lambda) - \sinh (\lambda) - 1)], \end{aligned}$$

where $\gamma \approx 0.577216$ is the Euler–Mascheroni constant.

Proof. Using the definition of Shannon entropy (4.1), we have

$$\Lambda = \log [2 (1 - e^{-\lambda})] - \log [\pi\lambda] + \lambda E ((\sin \pi X/2)) - E (\log [\cos (\pi X/2)]).$$

The expression $E (\sin (\pi X/2))$ is immediate from the fact that $X = \frac{2}{\pi} \arcsin (Y)$, where Y follows a truncated exponential distribution. Hence,

$$E (\sin (\pi X/2)) = \frac{1 - (1 + \lambda) e^{-\lambda}}{\lambda (1 - e^{-\lambda})}.$$

Also, the quantity $E (\log [\cos (\pi X/2)])$ can be computed as

$$\begin{aligned} E (\log [\cos (\pi X/2)]) &= \frac{\pi\lambda}{2(1 - e^{-\lambda})} \int_0^1 \log [\cos (\pi x/2)] \cos (\pi x/2) e^{-\lambda \sin (\pi x/2)} dx \\ &= \frac{\pi\lambda}{4(1 - e^{-\lambda})} [(2\gamma - \log [4] + 2 \log [\lambda]) (\cosh (\lambda) - \sinh (\lambda) - 1)] \\ &\quad + \frac{\pi\lambda}{4(1 - e^{-\lambda})} \left[\sqrt{2\pi\lambda} \left(I^{(1,0)} \left[\frac{1}{2}, \lambda \right] - L^{(1,0)} \left[\frac{1}{2}, \lambda \right] \right) \right]. \end{aligned}$$

Applying all the above results in the expression for Λ , we have the desired proof. \square

Theorem 4.3. *The Rényi entropy for the CS distribution is*

$$\Omega = \frac{1}{1 - \delta} \log [\xi (\delta, \lambda)] + \frac{1}{1 - \delta} \log [\Gamma[(1 + \delta) / 2] (I [\delta/2, \delta\lambda] - L [\delta/2, \delta\lambda])],$$

where $\xi (\delta, \lambda) = \frac{(\pi\lambda)^\delta}{2^\delta (1 - e^{-\lambda})^\delta \sqrt{\pi}} \left(\frac{2}{\delta\lambda} \right)^{\delta/2}$.

Proof. Using Eq. (2.1), we get

$$\begin{aligned} \int_0^1 f^\delta (x; \lambda) dx &= \frac{(\pi\lambda)^\delta}{2^\delta (1 - e^{-\lambda})^\delta} \int_0^1 (\cos (\pi x/2))^\delta e^{-\lambda\delta \sin (\pi x/2)} dx \\ &= \frac{(\pi\lambda)^\delta}{2^\delta \sqrt{\pi} (1 - e^{-\lambda})^\delta} \left(\frac{2}{\delta\lambda} \right)^{\delta/2} \Gamma \left[\frac{1 + \delta}{2} \right] (I [\delta/2, \delta\lambda] - L [\delta/2, \delta\lambda]). \end{aligned}$$

Using (4.2), it follows that

$$\Omega = \frac{1}{1 - \delta} \log \left[\frac{(\pi\lambda)^\delta}{2^\delta (1 - e^{-\lambda})^\delta \sqrt{\pi}} \left(\frac{2}{\delta\lambda} \right)^{\delta/2} \right] + \frac{1}{1 - \delta} \log \left[\Gamma \left[\frac{1 + \delta}{2} \right] (I[\delta/2, \delta\lambda] - L[\delta/2, \delta\lambda]) \right],$$

which completes the proof. □

Table 1 gives some numerical values for the Shannon entropy and the Rényi entropy. It can be observed that the Shannon entropy decreases with increasing λ and takes negative values, which may mean loss of system information. Also, it can be noted that the Rényi entropy takes negative values and decreases with increasing λ .

TABLE 1. Shannon and Rényi entropies for different values of λ .

| $\lambda \downarrow$ | Shannon entropy | $\lambda \downarrow$ | Rényi entropy with $\delta = 2$ |
|----------------------|-----------------|----------------------|---------------------------------|
| 1 | -0.274 | 2 | -0.667 |
| 2 | -0.453 | 3 | -0.928 |
| 3 | -0.658 | 5 | -1.370 |
| 4 | -0.864 | 7 | -1.700 |
| 5 | -1.056 | 9 | -1.953 |

5. SOME RELIABILITY MEASURES

Residual life and reversed residual life random variables are used extensively in reliability analysis and risk theory. In this section, we get some of their related statistical functions, such as survival function (SF), mean and variance of residual life functions for the CS model.

5.1. Residual life functions. The survival function of the residual lifetime $\psi_{(t)}$ for the CS distribution is

$$S_{\psi_{(t)}}(x) = \frac{S(x+t)}{S(t)} = \frac{e^{-\lambda \sin(\pi(x+t)/2)} - e^{-\lambda}}{e^{-\lambda \sin(\pi t/2)} - e^{-\lambda}}.$$

Using this equation, the pdf and hrf of $\psi_{(t)}$ are defined, respectively, as

$$f_{\psi_{(t)}}(x) = \frac{(\pi\lambda/2) \cos(\pi(x+t)/2)}{e^{-\lambda \sin(\pi t/2)} - e^{-\lambda}} e^{-\lambda \sin(\pi(x+t)/2)}, \quad 0 \leq x+t < 1,$$

and

$$h_{\psi_{(t)}}(x) = \frac{(\pi\lambda/2) \cos(\pi(x+t)/2)}{e^{-\lambda \sin(\pi(x+t)/2)} - e^{-\lambda}} e^{-\lambda \sin(\pi(x+t)/2)}, \quad 0 \leq x+t < 1.$$

Additionally, the mean of residual lifetime (MRL) of $\psi_{(t)}$ for the CS distribution is

$$\begin{aligned} \vartheta(t) &= E(\psi_{(t)}) = \frac{1}{S(t)} \int_t^1 x f(x; \lambda) dx - t \\ &= \frac{1}{S(t)} [E(x) - v(t; 1, \lambda)] - t, \quad 0 \leq t < 1, \end{aligned}$$

where $E(x)$ is obtained by (3.2) and $v(t; 1, \lambda)$ is given by (2.3).

Finally, the variance of residual lifetime of $\psi_{(t)}$ for the CS model is given by

$$\begin{aligned} \rho(t) &= V(\psi_{(t)}) = \frac{2}{S(t)} \int_t^1 x S(x) dx - 2t\vartheta(t) - [\vartheta(t)]^2 \\ &= \frac{1}{S(t)} [\mu'_2 - v(t; 2, \lambda)] - t^2 - 2t\vartheta(t) - [\vartheta(t)]^2, \quad 0 \leq t < 1, \end{aligned}$$

where $S(\cdot)$ is the SF of the CS distribution, μ'_2 can be obtained using (3.1), and $v(t; 2, \lambda)$ is given by (2.3). Some numerical values of the MRL are displayed in Table 2 for different values of λ at the time points $t = 0.1, 0.3, 0.5, 0.7, 0.9$. From this table, the MRL decreases with increasing the time points t and decreases with increasing λ .

TABLE 2. MRL for different values of λ and the time points t .

| $\lambda \downarrow$ $t \rightarrow$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|--------------------------------------|-------|-------|-------|-------|-------|
| 1 | 0.271 | 0.217 | 0.160 | 0.098 | 0.033 |
| 1.5 | 0.248 | 0.205 | 0.155 | 0.097 | 0.033 |
| 1.9 | 0.229 | 0.195 | 0.151 | 0.096 | 0.033 |
| 2.2 | 0.217 | 0.188 | 0.148 | 0.095 | 0.033 |
| 2.6 | 0.201 | 0.179 | 0.145 | 0.095 | 0.033 |

5.2. Reversed residual life functions. The survival function of the reversed residual lifetime $\bar{\psi}_{(t)}$ for the CS model is

$$S_{\bar{\psi}_{(t)}}(x) = \frac{F(t-x; \lambda)}{F(t; \lambda)} = \frac{1 - e^{-\lambda \sin(\pi(t-x)/2)}}{1 - e^{-\lambda \sin(\pi t/2)}}, \quad 0 \leq x < t.$$

Therefore, using this equation, the pdf and hrf of $\bar{\psi}_{(t)}$ are, respectively,

$$f_{\bar{\psi}_{(t)}}(x) = \frac{(\pi\lambda/2) \cos(\pi(t-x)/2)}{1 - e^{-\lambda \sin(\pi t/2)}} e^{-\lambda \sin(\pi(t-x)/2)},$$

and

$$h_{\bar{\psi}_{(t)}}(x) = \frac{(\pi\lambda/2) \cos(\pi(t-x)/2)}{1 - e^{-\lambda \sin(\pi(t-x)/2)}} e^{-\lambda \sin(\pi(t-x)/2)}.$$

Moreover, the mean of reversed residual lifetime (MRRL) of $\bar{\psi}_{(t)}$ for the CS model is

$$\Lambda(t) = E(\bar{\psi}_{(t)}) = t - \frac{1}{F(t; \lambda)} \int_0^t x f(x; \lambda) dx = t - \frac{v(t; 1, \lambda)}{F(t; \lambda)},$$

where $F(\cdot)$ is the cdf of the CS model and $v(t; 1, \lambda)$ is given by (2.3).

Finally, the variance of reversed residual lifetime of $\bar{\psi}_{(t)}$ for the CS model is given by

$$\begin{aligned} \Omega(t) &= \text{Var}(\bar{\psi}_{(t)}) = 2t\Lambda(t) - [\Lambda(t)]^2 - \frac{2}{F(t; \lambda)} \int_0^t xF(x; \lambda) dx \\ &= 2t\Lambda(t) - [\Lambda(t)]^2 - t^2 + \frac{v(t; 2, \lambda)}{F(t; \lambda)}, \end{aligned}$$

where $v(t; 2, \lambda)$ is given by (2.3) for $r = 2$. In Table 3 we give some numerical values for the MRRL with different values of λ at the time points $t = 0.1, 0.3, 0.5, 0.7, 0.9$. It can be seen that the MRRL increases with increasing the time points t and increases with increasing λ .

TABLE 3. MRRL for different values of λ and the time points t .

| $\lambda \downarrow$ | $t \longrightarrow$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|----------------------|---------------------|--------|--------|--------|--------|--------|
| 1 | | 0.0514 | 0.1641 | 0.2923 | 0.4387 | 0.6073 |
| 1.5 | | 0.0521 | 0.1696 | 0.3062 | 0.4615 | 0.6366 |
| 1.9 | | 0.0526 | 0.1740 | 0.3168 | 0.4787 | 0.6583 |
| 2.2 | | 0.0529 | 0.1773 | 0.3246 | 0.4909 | 0.6736 |
| 2.6 | | 0.0535 | 0.1815 | 0.3345 | 0.5064 | 0.6925 |

6. ESTIMATION WITH DATA ANALYSIS

In this section, we obtain the maximum likelihood estimator (MLE) of the unknown parameter of the CS model. Using this estimator we check the capability of this distribution to model a practical data set.

Let X_1, X_2, \dots, X_n be a random sample of size n from the CS distribution with observed values x_1, x_2, \dots, x_n . Then, the corresponding log-likelihood function is

$$\begin{aligned} \ell &= n \log(\pi/2) + n \log(\lambda) - n \log(1 - e^{-\lambda}) \\ &\quad + \sum_{i=1}^n \log(\cos(\pi x_i/2)) - \lambda \sum_{i=1}^n \sin(\pi x_i/2). \end{aligned}$$

Differentiating with respect to λ , we have

$$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \frac{ne^{-\lambda}}{1 - e^{-\lambda}} - \sum_{i=1}^n \sin(\pi x_i/2).$$

By setting $\frac{d\ell}{d\lambda} = 0$, the MLE $\hat{\lambda}$ of the parameter λ can be obtained by solving the non-linear equation above numerically for λ using some statistical software package, such as Mathematica.

6.1. Data analysis. By making use of a practical data set, we illustrate the applicability of the CS model among a set of classical and recent models —containing beta, Kumaraswamy and Log-Lindley— which also have support on $(0, 1)$. We estimate the model parameters by using the maximum likelihood method. We compare goodness-of-fit of the models with the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Also, we apply the following goodness-of-fit tests: Cramér–von Mises (W^*), Anderson–Darling (A^*), the P-values based on Cramér–von Mises statistic and the Anderson–Darling statistic. In general, the smaller the values of these statistics and the largest P-values, the better the model's fit.

The competitive models are:

- (1) Beta (Be) distribution;
- (2) Kumaraswamy (Kw) distribution — see Kumaraswamy [11];
- (3) Log-Lindley (LL) distribution — see Gómez-Déniz et al. [6].

Description of the data is as follows.

The data set is given in Table 4. This data set is extracted from firms risk management cost-effectiveness, available at the personal web page of Professor E. Frees (Wisconsin School of Business). The data is defined as the total property and casualty premiums and uninsured losses as a percentage of the total assets leading to a bounded set on $(0, 1)$. This data set was used by Gómez-Déniz et al. [6] to illustrate the applicability of the Log-Lindley distribution.

Table 5 gives some descriptive statistics for the data set and using it we note that the data are under-dispersed (variance $<$ mean) and positively skewed. Further, the data set possesses a negative kurtosis. Description of such data adapts features of the CS distribution and this proves the suitability of this distribution for analyzing such data. From Tables 5 and 8, we find that the descriptive and the corresponding theoretical estimated measures (under the CS model) are closed approximately for the data set. Also, Table 8 shows negative values for the considered entropies that adapt the theoretical results given in Table 1.

On the other hand, comparing the CS model with other bounded distributions is investigated as follows. For the data set, we estimate the unknown parameters of each distribution by the maximum-likelihood method, and using those estimates, we obtain the statistics AIC, BIC, W^* , A^* , and the P-values based on Cramér–von Mises statistic and the Anderson–Darling statistic. The results obtained are reported in Tables 6 and 7. From these tables, the smallest values of the AIC, BIC, W^* , A^* , and the largest P-values based on Cramér–von Mises statistic and the Anderson–Darling statistic are obtained for the CS model. Hence, we conclude that the CS model provides the best fit among the compared distributions.

TABLE 4. Firms risk management cost-effectiveness data.

| | | | | |
|---------------|---------------|---------------|---------------|---------------|
| 0.0328999962 | 0.1937999916 | 0.00930000007 | 0.1753000069 | 0.65 |
| 0.0931000042 | 0.0375 | 0.10 | 0.14 | 0.08510000229 |
| 0.04070000172 | 0.1332999992 | 0.05820000172 | 0.02059999943 | 0.04309999943 |
| 0.06940000057 | 0.04579999924 | 0.09130000114 | 0.0125 | 0.00360000014 |
| 0.05349999905 | 0.1396000004 | 0.09 | 0.1192000008 | 0.07829999924 |
| 0.2886000061 | 0.00280000001 | 0.1260999966 | 0.00649999976 | 0.05090000153 |
| 0.02789999962 | 0.06079999924 | 0.02150000095 | 0.03150000095 | 0.00400000006 |
| 0.15 | 0.00939999998 | 0.2221999931 | 0.02549999952 | 0.03700000048 |
| 0.03890000105 | 0.02970000029 | 0.1271000004 | 0.08180000305 | 0.15 |
| 0.04070000172 | 0.04110000134 | 0.1597000027 | 0.1128999996 | 0.1356999969 |
| 0.04340000153 | 0.01220000029 | 0.04320000172 | 0.08850000381 | 0.1244999981 |
| 0.04329999924 | 0.06289999962 | 0.08489999771 | 0.0075999999 | 0.00200000003 |
| 0.05289999962 | 0.06119999886 | 0.0525 | 0.05710000038 | 0.18 |
| 0.07900000095 | 0.0350999999 | 0.1832999992 | 0.2912000084 | |
| 0.9755000305 | 0.02160000086 | 0.2171999931 | 0.7930000305 | |

TABLE 5. Descriptive statistics of the data set.

| | |
|-----------|---------|
| Mean | 0.1312 |
| Median | 0.1129 |
| SD | 0.0887 |
| MD-mean | 0.0765 |
| MD-median | 0.0761 |
| Skewness | 0.5776 |
| Kurtosis | -0.6515 |

7. CONCLUDING REMARKS

A new trigonometric distribution with bounded support, named cosine-sine distribution, is introduced. The probability density function of this model decays slowly and rapidly to 0 at the end of the distribution support based on the distribution parameter values, which may be a desired feature for some real-life models. Also, this model can be displayed as a truncated exponential sine distribution.

TABLE 6. Firms risk management cost-effectiveness data: ML estimates, AIC and BIC values.

| Model | Estimates | | AIC | BIC |
|-----------------------|--------------------|-------------------|----------|----------|
| $Be(a, b)$ | $a = 0.1625$ | $b = 3.7979$ | -148.235 | -143.654 |
| $Kw(a, b)$ | $a = 0.6648$ | $b = 3.4407$ | -153.308 | -148.727 |
| $LL(\lambda, \sigma)$ | $\lambda = 0.0343$ | $\sigma = 0.6907$ | -149.208 | -144.627 |
| $CS(\lambda)$ | $\lambda = 6.2743$ | | -176.068 | -173.777 |

TABLE 7. Firms risk management cost-effectiveness data: Goodness-of-fit tests.

| Model | W^* | A^* | A^* (P-value) | W^* (P-value) |
|-----------------------|--------|--------|-----------------|-----------------|
| $Be(a, b)$ | 0.6924 | 3.9193 | 0.0096 | 0.0130 |
| $Kw(a, b)$ | 0.4984 | 3.0643 | 0.0255 | 0.0399 |
| $LL(\lambda, \sigma)$ | 1.0602 | 5.5756 | 0.0016 | 0.0017 |
| $CS(\lambda)$ | 0.0927 | 0.7884 | 0.4891 | 0.6332 |

TABLE 8. Some measures of the CS model for the data set.

| | |
|-----------------|---------|
| Mean | 0.1031 |
| Median | 0.0703 |
| SD | 0.1061 |
| MD-mean | 0.0030 |
| MD-median | 0.0720 |
| Shannon entropy | -1.2725 |
| Rényi entropy | -1.7387 |

Additionally, the new model has a closed form expression for the cumulative distribution function and various shapes of its hazard rate function. Finally, the cosine-sine model and other distributions are fitted to a practical data set and the usefulness of the model is illustrated.

ACKNOWLEDGMENTS

We would like to thank the associate editor and the referee for carefully reading our manuscript and giving constructive comments, which have helped us to substantially improve the present article.

REFERENCES

- [1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series, 55. Tenth printing, 1972.
- [2] R. Q. Al-Faris and S. Khan, Sine square distribution: a new statistical model based on the sine function, *J. Appl. Probab. Stat.* **3** (2008), no. 1, 163–173. MR 2530366.
- [3] S. Chowdhury and A. K. Nanda, A new lifetime distribution with applications in inventory and insurance, *Int. J. Qual. Reliab. Manag.* **35** (2018), 527–544.
- [4] D. O. Cook, R. Kieschnick and B. D. McCullough, Regression analysis of proportions in finance with self selection, *J. Empirical Finance* **15** (2008), 860–867.
- [5] A. Genç, Estimation of $P(X > Y)$ with Topp-Leone distribution, *J. Stat. Comput. Simul.* **83** (2013), no. 2, 326–339. MR 3015226.
- [6] E. Gómez-Déniz, M. A. Sordo and E. Calderín-Ojeda, The Log-Lindley distribution as an alternative to the beta regression model with applications in insurance, *Insurance Math. Econom.* **54** (2014), 49–57. MR 3145850.
- [7] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, sixth edition, translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, Academic Press, San Diego, CA, 2000. MR 1773820.
- [8] N. A. J. Hastings and J. B. Peacock, *Statistical Distributions*, Halsted Press, New York, 1975. MR 0359137.
- [9] R. A. Jiang, A new bathtub curve model with finite support *Reliab. Eng. System Safety* **119** (2013), 44–51.
- [10] J. T. Kent and D. E. Tyler, Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Statist.* **15** (1988) no. 2, 247–254.
- [11] P. Kumaraswamy, Generalized probability density function for double-bounded random processes, *J. Hydrology* **46** (1980), no. 1-2, 79–88.
- [12] C. D. Lai and G. Jones, Beta hazard rate distribution and applications, *IEEE Trans. Reliab.* **64** (2015), no. 1, 44–50.
- [13] C. D. Lai and S. P. Mukherjee, A note on “A finite range distribution of failure times”, *Microelectron. Reliab.* **26** (1986), no. 1, 183–189.
- [14] A. W. Marshall and I. Olkin, *Life Distributions*, Springer Series in Statistics, Springer, New York, 2007. MR 2344835.
- [15] S. P. Mukherjee and A. Islam, A finite-range distribution of failure times, *Naval Res. Logist. Quart.* **30** (1983), no. 3, 487–491. MR 0717739.
- [16] L. E. Papke and J. M. Wooldridge, Econometric methods for fractional response variables with an application to 401(K) plan participation rates, *J. Appl. Econometrics* **11** (1996), no. 6, 619–632.
- [17] A. Rényi, On measures of entropy and information, in *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, Vol. I, 547–561, Univ. California Press, Berkeley, CA, 1961. MR 0132570.
- [18] C. E. Shannon, Prediction and entropy of printed English, *Bell Syst. Tech. J.* **30** (1951), no. 1, 50–64.

A. M. T. Abd El-Bar[✉]

Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt
ahmed.abdelbar@science.tanta.edu.eg

H. S. Bakouch

Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt
hassan.bakouch@science.tanta.edu.eg

S. Choudhury

Quantitative Methods and Operations Management Area, Indian Institute of Management
Kozhikode, Kerala, India
shovanc@iimk.ac.in

Received: December 22, 2019

Accepted: December 15, 2020