REVISTA INVESTIGACIÓN OPERACIONAL Vol. 45, No. 5, 766-791, 2024

DEVELOPMENT AND APPLICATIONS OF DISTRIBUTIONAL MODELS BASED ON AN ORIGINAL POLYNOMIAL-SINE SCHEME

Christophe Chesneau[1]*, Tomy Lishamol[2] and G Veena[3]
[1]Université de Caen, LMNO, Campus II, France.
[2]Department of Statistics, Deva Matha College, India.
[3]Department of Statistics, St.Thomas College, India.

ABSTRACT

Recent advances in modeling and data analysis have been greatly influenced by absolutely continuous trigonometric distributions. In this study, we develop a new polynomial-sine transformation strategy to improve the statistical capabilities of absolutely continuous baseline distributions. We thus create a modified polynomial-sine-generated family of distributions, which can be viewed as a polynomial extension of the famous sine-generated family. Analytical and practical properties of the family are examined. Subsequently, with the data-fitting aim in mind, the maximum likelihood estimation method is employed to estimate the parameters. A special three-parameter lifetime member of the family, defined with the Weibull distribution as the baseline distribution, is highlighted. It constitutes a three-parameter lifetime distribution whose main functions benefit from trigonometric functionalities for reaching flexible shapes. It has a right-skewed probability density function with peakness ranging from platykurtic to leptokurtic as a specificity. The associated hazard rate function may also take versatile forms, such as linear, increasing, decreasing, and the inverted bathtub shape. The distribution is studied in depth and tested on two real-world datasets from the medical field (COVID-19 datasets) to show its efficacy. It provides a superior fit in comparison to the other three competing Weibull distributions, including the sine-Weibull distribution as the main concurrent.

KEYWORDS: trigonometric distribution; sine function, polynomial function, Weibull distribution, lifetime data analysis
MSC: 60G22, 91G20

RESUMEN

*christophe.chesneau@gmail.com

Los avances recientes en el modelado y el análisis de datos se han visto muy influenciados por las distribuciones trigonométricas absolutamente continuas. En este estudio, desarrollamos una nueva estrategia de transformación polinomio-sinusoidal para mejorar las capacidades estadísticas de distribuciones de base absolutamente continuas. Por lo tanto, se genera una familia de distribuciones polinomio-sinusoidal, que puede verse como una extensión polinómica de la familia generada por senos. Se examinan las propiedades analíticas y prácticas de la familia. Posteriormente, teniendo en cuenta el objetivo de ajuste de los datos, se emplea el método de estimación máximo verosímil para estimar los parámetros. Se considera un miembro especial de la familia que está dado por tres parámetros de tiempo de vida, definido con la distribución de Weibull como distribución de referencia. Este elemento constituye una distribución de vida útil de tres parámetros cuyas propiedades trigonométricas les permiten alcanzar formas flexibles. Tiene una función de densidad de probabilidad sesgada a la derecha con picos que van de platicúrtica a leptocúrtica como especificidad. La función de tasa de riesgo asociada también puede adoptar formas versátiles, como lineal, creciente, decreciente y la forma de bañera invertida. La distribución se estudia en profundidad y se prueba en dos conjuntos de datos del mundo real del campo médico (conjuntos de datos de COVID-19) para mostrar su eficacia. Proporciona un ajuste superior en comparación con las otras tres distribuciones de Weibull que compiten entre sí, incluida la distribución sinusoidal-Weibull como la principal concurrente.
PALABRAS CLAVE: distribución trigonométrica, función seno, distribución Weibull, análisis de tiempo de vida

1. INTRODUCTION

1.1. Context

Real-world data may exhibit significant skewness and kurtosis that cannot be correctly captured with the standard model. In order to analyze such types of data efficiently, new strategies for creating models with few parameters but great flexibility are required. The statistical literature has recently taken a fresh turn, thanks to trigonometric distributions. These distributions enable us to describe data with adaptable probability density functions (pdfs) and hazard rate functions (hrfs) that take advantage of trigonometric functionalities. There are a number of recent trigonometric distributions that can be found in the literature, such as the sine distribution developed in [6], the Gilbert sine distribution created in [10], the trigonometric beta distribution created by [18], the Von Mises distribution studied in [12], the sine square distribution published in [3], the sin-skewed logistic distribution developed in [8], the sine-exponential distribution proposed by [15], the extended sine-Weibull examined in [20] and the special cosine-sine distribution initiated in [9]. General families of trigonometric distributions have also seen the light. On this topic, it is avoidable to mention the recent sine-generated (SinG for short) family first introduced in [22], which will be at the center of the study. An overview of this family is given below.

1.2. SinG family of distributions

The originality of the SinG family is to be centered around the simple sine transformation, without the addition of new parameters. The idea is to use the trigonometric functionalities of the sine function

to wrap up those of an absolutely continuous baseline distribution for new modeling purposes. On the mathematical side, the cumulative density function (cdf) and pdf of the SinG family are given by

$$K(x) = \sin\left(\frac{\pi}{2}G(x)\right)$$

and

$$k(x) = \frac{\pi}{2}g(x)\cos\left(\frac{\pi}{2}G(x)\right),$$

respectively, where $G(x)$ is the cdf of the baseline distribution, and $g(x)$ its associated pdf. The variable x is supposed to belong to $\mathbb{R}$, but changes can be made depending on the support of the baseline distribution. The merit of the SinG family is that it can produce simple and flexible pdfs and hrfs, beyond the shape possibilities of the pdf and hrf of the baseline distribution. New modal, skewness and kurtosis capabilities can be observed.

Many new works have been introduced as a result of the development of SinG-based distributions. Among them, we may refer to the sine-power Lomax distribution, conjointly studied in [4] and [19]. The contribution in [19] shows that the distribution can generate models for both symmetric and skewed datasets and validate them on nine datasets. In [2], the sine-Topp-Leone family was tested on two datasets and compared to 20 existing models. In [11], the sine-inverse Lomax-generated family was created, and the applicability of this family with the Lomax distribution as the baseline distribution was also studied. The sine-Burr XII distribution was examined in [13]. This new distribution was compared with the Burr XII and exponential Bur XII distributions on two datasets. One may also mention the sine-modified Lindley distribution studied in [24] and [25], which proves to be a better alternative than the former Lindley distribution for a wide panel of biomedical datasets.

1.3. Contributions

This study is devoted to the development and study of a new trigonometric family based on an original polynomial-sine transformation. It is called the modified polynomial-sine-generated (MPSG) family. According to our knowledge, it is the only family to cover both the baseline distribution and the SinG family, and this with only one additional parameter. In this sense, the study of the MPSG family is a pioneer in its field. Subsequently, the modified polynomial-sine Weibull (MPSW) distribution is introduced, with the Weibull distribution as the baseline distribution of the MPSG family. The MPSW distribution benefits from the functionalities of both the polynomial-sine transformation and the Weibull distribution, reaching an unexplored modeling aim. Based on this, its statistical inference is examined through the standard maximum likelihood (ML) estimation method for complete samples. On real-world datasets related to COVID, the new distribution is shown to have a better statistical fit than the conventional Weibull distribution and its trigonometric extensions. In this practical setting, the advantage of the MPSW model over the Weibull model and well-known trigonometric model versions of the Weibull distribution is obvious. We prove this claim using strict statistical criteria.

1.4. Paper organization

The rest of this paper is structured as follows: In Section 2., we present the definition and functions of the considered polynomial-sine family. Section 3. emphasizes useful series expansions, moment

analysis, and other useful properties. The ML estimates of the relevant parameters are examined in Section 4.. Section 5. is devoted to the MPSW distribution and its qualities in terms of the statistical model. Section 6. contains the significance of the MPSW model by taking into account two real-world datasets. Section 7. concludes the paper.

2. THE MODIFIED POLYNOMIAL-SINE FAMILY

2.1. Mathematical foundation

We begin with a theoretical result presenting a valid cdf defined with an original polynomial-sine transformation.

Proposition 2.1. *Let us consider an absolutely continuous distribution defined with a cdf denoted by $G(x)$. Then, we set the following parametric function:*

$$F(x;\theta) = G(x) + \theta \sin(\pi G(x)), \quad x \in \mathbb{R},$$

where $\theta \in [-1/\pi, 1/\pi]$. Then $F(x;\theta)$ has the properties that characterize a cdf.

Proof. First and foremost, because $G(x)$ is continuous, $F(x;\theta)$ is also continuous by composition. Let us now denote the support of the distribution related to $G(x)$ by $[a, b]$ (the bounds $a = -\infty$ and $b = +\infty$ are not excluded). Since $\lim_{x \rightarrowtail a} G(x) = 0$ and $\lim_{x \rightarrowtail a} G(x) = 1$, we have

$$\lim_{x \rightarrowtail a} F(x;\theta) = 0 + \theta \sin(\pi \times 0) = 0, \quad \lim_{x \rightarrowtail b} F(x;\theta) = 1 + \theta \sin(\pi \times 1) = 1,$$

which are the expected limits for a valid cdf. Finally, let us examine the monotonicity of $F(x;\theta)$. Let $g(x)$ be the pdf corresponding to $G(x)$. So, because $g(x) \geq 0$, $\theta \in [-1/\pi, 1/\pi]$ and $\cos(x) \leq 1$, we have

$$F'(x;\theta) = g(x)\left[1 + \theta\pi \cos(\pi G(x))\right] \geq g(x)\left[1 - \cos(\pi G(x))\right] \geq 0.$$

We conclude that $F'(x;\theta)$ is non-decreasing, bringing the proof to a close. □

To our knowledge, the cdf presented in Proposition 2.1 is new in the literature. Thus, we define the MPSG family by the following cdf:

$$F(x;\theta) = G(x) + \theta \sin(\pi G(x)), \quad x \in \mathbb{R}, \tag{2.1}$$

by adopting the assumptions made in Proposition 2.1, i.e., $\theta \in [-1/\pi, 1/\pi]$. Along with its unique definition, the MPSG family also has the arguments below in its favor.

- The cdf of the MPSG family is analytically simple and extends the baseline distribution through a simple trigonometric transformation scheme. Indeed, taking $\theta = 0 \in [-1/\pi, 1/\pi]$ yields $F(x;\theta) = G(x)$.
- In some senses, the MPSG family realizes a generalized mixture between the baseline distribution, and the SinG family proposed by [22]. It can also be interpreted as an unification.

- The MPSG family follows the same spirit as the Burr XI generated (BXIG) family proposed by [5], which is defined with the following cdf:

$$F^*(x;r) = \left[G(x) - \frac{1}{2\pi}\sin(2\pi G(x))\right]^r, \quad x \in \mathbb{R}.$$

  The MPSG and BXIG families both combine $G(x)$ with its trigonometric counterpart. However, the BXIG family is not permitted to refind the baseline distribution or modulate the sine term. Moreover, the exponentiated operation adds some complexity from a mathematical viewpoint. In this sense, the MPSG family is preferable for the consideration of sophisticated baseline distributions.

- The distributions of the MPSG family can be used quite efficiently as models for various data analyses. This practical aspect will be motivated later with concrete graphics and numerical examples.

Remark 2.2. *An extended MPSG family can be defined by the following cdf:*

$$F(x;\theta,m) = G(x) + \frac{\theta}{m}\sin(m\pi G(x)), \quad x \in \mathbb{R},$$

where $\theta \in [-1/\pi, 1/\pi]$ and m denotes any positive integer. However, the choice of m is somewhat subjective, so we focus on the simple case where $m = 1$, which corresponds to the (simple) MPSG family.

2.2. Probabilistic functions

The probabilistic functions of the MPSG family are now described. First, the pdf is given by

$$f(x;\theta) = g(x)\left[1 + \theta\pi\cos(\pi G(x))\right], \quad x \in \mathbb{R}. \tag{2.2}$$

We recall that $g(x)$ is the pdf corresponding to $G(x)$. The survival function (sf) and hrf listed below are two additional significant MPSG family functions. The sf is

$$S(x;\theta) = 1 - G(x) - \theta\sin(\pi G(x))$$

and the hrf is

$$h(x;\theta) = \frac{g(x)\left[1 + \theta\pi\cos(\pi G(x))\right]}{1 - G(x) - \theta\sin(\pi G(x))}, \quad x \in \mathbb{R}.$$

Understanding the modeling capabilities of the relevant distribution requires knowledge of the analytical behavior of these functions. The main theory of the MPSG family is based on these functions, as described in the next section.

3. THEORY

Here, we discuss the asymptotic properties, expansion of the pdf and cdf in series form, moment analysis, and reliability parameter of the MPSG family.

3.1. Asymptotic properties and critical points of the pdf

The asymptotic properties of the pdf given in Equation (2.2) are as follows: When $x \rightarrowtail -\infty$, we have $f(x;\theta) \sim g(x)(1+\theta\pi)$, and when, $x \rightarrowtail +\infty$, we have $f(x;\theta) \sim g(x)(1-\theta\pi)$. Therefore, for these equivalences, the choice of the baseline distribution is determinant; θ just plays the role of a scale parameter.
On the other hand, the critical point(s) for $f(x;\theta)$ is (are) given by the solution(s) of the following nonlinear equation: $d\log[f(x;\theta)]/dx = 0$, with

$$\frac{\log[f(x;\theta)]}{dx} = \frac{g'(x)}{g(x)} - \frac{\theta\pi^2 \sin(\pi G(x))\, g(x)}{1+\theta\pi\cos(\pi G(x))}.$$

We can remark that θ has an important impact on the definition of the critical point(s). Regardless of the complexity of $G(x)$ and $g(x)$, such point(s) can be identified numerically.

3.2. Series expansion

In this part, we develop alternative expressions for the pdf and cdf of the MPSG family by using a series functional representation.

Proposition 3.3. *The cdf of the MPSG family can be represented as*

$$F(x;\theta) = \sum_{k=0}^{+\infty} a_k G(x)^{2k+1},$$

where $a_0 = 1+\theta\pi$ *and* $a_k = \theta(-1)^k\pi^{2k+1}/(2k+1)!$ *for* $k \geq 1$.
Furthermore, we have

$$f(x;\theta) = \sum_{k=0}^{+\infty} a_k(2k+1)g(x)G(x)^{2k}. \tag{3.1}$$

Proof. For any $x \in \mathbb{R}$, the following series expansion is established:

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

Based on the definition of $F(x;\theta)$, we immediately obtain

$$F(x;\theta) = G(x) + \theta\sum_{k=0}^{+\infty} \frac{(-1)^k\pi^{2k+1}}{(2k+1)!}G(x)^{2k+1} = \sum_{k=0}^{+\infty} a_k G(x)^{2k+1}.$$

The expansion of the pdf follows by differentiating this expression with respect to x, provided the mathematical validity of this operation, which depends on the definition of $G(x)$ mainly. This ends the proof. □

The appeal of Proposition 3.3 is that $F(x;\theta)$ and $f(x;\theta)$ can be expressed simply in terms of simple functions, defined primarily as the exponentiated version of $G(x)$. Such exponentiated baseline distributions have been the subject of numerous studies and can thus be used directly to derive properties of

the MPSG family. Moreover, the following analytical approximation can be useful for computational purposes:

$$f(x;\theta) \approx \sum_{k=0}^{\tau} a_k (2k+1) g(x) G(x)^{2k},$$

where τ denotes a chosen large integer.
An alternative series expansion is proposed below, with the use of the sf of the baseline distribution, i.e., $S(x) = 1 - G(x)$.

Proposition 3.4. *The cdf of the MPSG family can be represented as*

$$F(x;\theta) = 1 + \sum_{k=0}^{+\infty} a_k^* S(x)^{2k+1},$$

where $a_0^* = -1 + \theta\pi$ *and* $a_k^* = a_k = \theta(-1)^k \pi^{2k+1}/(2k+1)!$ *for* $k \geq 1$.
Furthermore, we have

$$f(x;\theta) = \sum_{k=0}^{+\infty} a_k^{**} (2k+1) g(x) S(x)^{2k}, \tag{3.2}$$

where $a_k^{**} = -a_k^*$.

Proof. For any $x \in \mathbb{R}$, by noticing that $G(x) = 1 - S(x)$ and, by the following standard trigonometric formula: $\sin(\pi G(x)) = \sin(\pi S(x))$, we obtain

$$F(x;\theta) = 1 - S(x) + \theta \sum_{k=0}^{+\infty} \frac{(-1)^k \pi^{2k+1}}{(2k+1)!} S(x)^{2k+1} = \sum_{k=0}^{+\infty} a_k^* S(x)^{2k+1}.$$

The expansion of the pdf follows by differentiating this expression with respect to x, ending the proof. □

The interest of Proposition 3.4 is that, for some baseline distributions, the sf is more manageable than the cdf from a mathematical viewpoint.
Hence, the discussed series expansions of the pdf, cdf, and sf can be used to make mathematical computation easier, and will be employed as it in the next.

3.3. Moment analysis

The moment analysis of the MPSG family is now examined, such as the mean, variance, measure of skewness, and kurtosis.

Proposition 3.5. *Let* X *be a random variable following a fixed distribution from the MPSG family. Then, if existence, the expectation of* $\Phi(X)$ *can be expressed as*

$$\Theta_\Phi = E\left[\Phi(X)\right] = \int_0^1 \Phi(G^{-1}(u)) du + \theta\pi \int_0^1 \Phi(G^{-1}(u)) \cos(\pi u) du.$$

Proof. The proof is based on the definition of $f(x;\theta)$ and an appropriate change of variables, that is

$$\begin{aligned}\Theta_\Phi &= \int_{-\infty}^{+\infty} \Phi(x) f(x;\theta) dx = \int_{-\infty}^{+\infty} \Phi(x) g(x) \left[1 + \theta\pi \cos(\pi G(x))\right] dx \\ &= \int_{-\infty}^{+\infty} \Phi(x) g(x) dx + \theta\pi \int_{-\infty}^{+\infty} \Phi(x) g(x) \cos(\pi G(x)) dx \\ &= \int_0^1 \Phi(G^{-1}(u)) du + \theta\pi \int_0^1 \Phi(G^{-1}(u)) \cos(\pi u) du.\end{aligned}$$

This ends the proof. □

Depending on the complexity of the function $G^{-1}(u)$, the two integral terms in Proposition 3.5 can be computed analytically.

In all circumstances, for a given baseline distribution, $\Theta_\Phi(x)$ can be determined using numerical methods provided in any mathematical software.

Additionally, Equations (3.1) and (3.2) implement the following different analytical approaches

$$\Theta_\Phi = \sum_{k=0}^{+\infty} a_k (2k+1) \int_{-\infty}^{+\infty} \Phi(x) g(x) G(x)^{2k} dx$$

and

$$\Theta_\Phi = \sum_{k=0}^{+\infty} a_k^{**} (2k+1) \int_{-\infty}^{+\infty} \Phi(x) g(x) S(x)^{2k} dx,$$

respectively. In order to provide an acceptable approximation of $\Theta_\Phi(x)$, the sum can be truncated to a large enough integer.

Moments can be used to compute basic statistical measures. For any positive integer m, the mth ordinary moment is indicated as $\mu_m^\iota = E(X^m)$. The basic measures are $Mean = E(X)$ and $Variance = \sigma^2 = \mu_2$. In addition, the definition of the mth central moment and the cumulant are given by

$$\mu_m = \sum_{k=0}^{m} \binom{m}{k} (-1)^k \, Mean^k \mu_{m-k}^\iota$$

and

$$K_m = \mu_m^\iota - \sum_{k=1}^{m-1} \binom{m-1}{k-1} K_k \mu_{m-k}^\iota,$$

where $K_1 = Mean$, respectively. The coefficient of skewness and coefficient of kurtosis are expressed as $C_{sk} = \mu_3 / Variance^{3/2}$ and $C_{ku} = \mu_4 / Variance^2$, respectively.

For any positive integer m, the mth incomplete moment of X is defined as $\mu_m(y) = E\left(X^m I\left(X \leq y\right)\right) = \int_{-\infty}^{y} x^m f(x;\theta) dx$. For empirical purposes, the shape of many distributions can be usefully described by the incomplete moments. The importance of the incomplete moments can be seen in calculating inequality measures and mean deviations. As a last remark, we can express $\mu_m(y)$ as a series expansion via Equations (3.1) and (3.2).

3.4. Reliability parameter

The lifespan of a component is determined by a certain stress-strength model, which is employed in the field of reliability. It makes use of a random strength (represented by a random variable, say X_1) and a random stress (represented by a random variable, say X_2). The component fails when the amount of stress placed on it exceeds its capacity, and it will function effectively whenever $X_1 > X_2$. The probability $R = P(X_2 < X_1)$ is thus a component reliability metric. The readers are directed to [16] for a more in-depth understanding.
The next finding relates to how R for the MPSG family is calculated in a particular environment.

Proposition 3.6. *Let $\theta_1 \in [-1/\pi, 1/\pi]$ and $\theta_2 \in [-1/\pi, 1/\pi]$ (the case $\theta_1 = \theta_2$ is allowed). We assume that X_1 has the pdf given as $f(x;\theta_1)$, and X_2 has the cdf given as $F(x;\theta_2)$, and X_1 and X_2 are independent. Then, we have*

$$R = \frac{1}{2} + \frac{2}{\pi}(\theta_1 - \theta_2).$$

Proof. With the independence of X_1 and X_2, the definitions of $f(x;\theta_1)$ and $F(x;\theta_2)$ and an appropriate change of variables, we get

$$\begin{aligned} R = P(X_2 < X_1) &= \int_{-\infty}^{+\infty} F(x;\theta_2) f(x;\theta_1) dx \\ &= \int_{-\infty}^{+\infty} \left[G(x) + \theta_1 \sin\left(\pi G(x)\right)\right] g(x) \left[1 + \pi\theta_2 \cos\left(\pi G(x)\right)\right] dx \\ &= \int_0^1 \left[u + \theta_1 \sin\left(\pi u\right) + \pi\theta_2 u \cos(\pi u) + \pi\theta_2 \cos\left(\pi u\right) \theta_1 \sin\left(\pi u\right)\right] du \\ &= \frac{1}{2} + \frac{2}{\pi}(\theta_1 - \theta_2). \end{aligned}$$

This ends the proof. □

It is worth noting that the expression for R is quite simple; it is dependent only on the parameters θ_1 and θ_2 and not on the baseline distribution. When X_1 and X_2 in Proposition 3.6 have identical distributions, i.e., $\theta_1 = \theta_2$, we obtain the expected value of 1/2.
In a statistical setting, estimating R from data based on the estimates of θ_1 and θ_2 is also useful when using Proposition 3.6 and the substitution technique. Indeed, if $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimates of θ_1 and θ_2, respectively, then the substitution method offers the following accurate estimate:

$$\hat{R} = \frac{1}{2} + \frac{2}{\pi}(\hat{\theta}_1 - \hat{\theta}_2).$$

However, further study is required to see how this algorithm applies to real-world data.

4. ESTIMATION THEORY

Because of its general simplicity and theoretical guarantees of good convergence properties on the resulting estimates, the ML method is frequently used in parametric estimation.

4.1. ML estimation

Let us now discuss the ML estimation method for the parameter(s) for the MPSG family, with the pdf $f(x;\theta)$ given in Equation (2.2). A specific vector v represents the parameters of the baseline distribution. Given n observations of a random variable X following a fixed distribution from the MPSG family, say $x_1, x_2, \ldots, x_n$, the likelihood function for the parameter vector $\delta = (\theta, v)$ is

$$L(\underline{x};\delta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} g(x_i)\left[1+\theta\pi\cos\left(\pi G(x_i)\right)\right].$$

where $\underline{x} = \{x_1, x_2, \ldots, x_n\}$. The vector δ that maximizes $L(\underline{x};\delta)$ yields the ML vector estimate of δ. If it is well-defined, it can be obtained by maximizing the log-likelihood function given by

$$\log\left(L(\underline{x};\delta)\right) = \sum_{i=1}^{n}\log\left(g(x_i)\right) + \sum_{i=1}^{n}\log\left[1+\theta\pi\cos\left(\pi G(x_i)\right)\right].$$

To this aim, statistical programs can be employed, such as R, Mathcad, SAS, Mathematica and Ox programs. On the mathematical aspect, the ML vector estimate can be obtained by solving the following nonlinear equations: $(U_\theta, U_v) = (0, 0_*)$, where 0_* is a vector with the same dimension to v, and

$$U_\theta = \frac{d\log\left(L(x;\delta)\right)}{d\theta} = \sum_{i=1}^{n}\frac{\pi\cos\left(\pi G(x_i)\right)}{1+\theta\pi\cos\left(\pi G(x_i)\right)}.$$

and, by denoting $g_v(x) = dg(x)/dv$ and $G_v(x) = dG(x)/dv$ (provided that they exist),

$$U_v = \frac{d\log\left(L(x;\delta)\right)}{dv} = \sum_{i=1}^{n}\frac{g_v(x_i)}{g(x_i)} - \sum_{i=1}^{n}\frac{\theta\pi^2 G_v(x_i)\sin\left(\pi G(x_i)\right)}{1+\theta\pi\cos\left(\pi G(x_i)\right)}.$$

In general, these equations cannot be solved analytically, but they can be solved numerically by software. The ML vector estimate of δ is denoted as $\hat{\delta} = (\hat{\theta}, \hat{v})$, and $\hat{\theta}$ and $\hat{v}$ are called the ML estimates of θ and v, respectively.

5. AN EXAMPLE: THE MPSW DISTRIBUTION

We now focus on an original member of the MPSG family defined with the Weibull distribution as the baseline distribution.

5.1. Motivation

Swedish physicist Waloddi Weibull, in [26], is credited with devising the Weibull distribution. He used it to model material yield strength. Since this first study, the Weibull distribution has been widely used in many domains, including engineering, reliability, failure analysis, lifetime analysis, material science, quality control, physics, medicine, meteorology, hydrology, and others. The conventional Weibull distribution has, however, some limitations, such as a relative rigidity in the left-tail and overall kurtosis, and a hrf that has only monotonic shapes (increasing, constant, and decreasing

shapes). For this reason, some generalized Weibull distributions have seen the light in recent years, and were applied in a variety of situations.
On the mathematical plan, the cdf of the Weibull distribution is given by

$$G_W(x;\alpha,\beta) = 1 - e^{-\alpha x^\beta}, \quad x > 0, \tag{5.1}$$

where $\alpha > 0$ and $\beta > 0$, and the associated pdf is

$$g_W(x;\alpha,\beta) = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}, \quad x > 0. \tag{5.2}$$

As all the lifetime distributions, we have $G_W(x;\alpha,\beta) = g_W(x;\alpha,\beta) = 0$ for $x \leq 0$.
In the setting of the trigonometric families, the Weibull distribution has often been considered as a natural baseline distribution, yielding successful statistical models. In particular, for the SinG and cosine-generated (CosG) families, the sine-Weibull (SW) and cosine-Weibull (CW) distributions have seen the light in [22] and [23], respectively. One may also mention the the sine-exponential distribution in [15], which constitutes a special case of the SW distribution. The SW and CW distributions are recalled below.

- The SW distribution is defined with the cdf and pdf given as

  $$F_{SW}(x;\alpha,\beta) = \cos\left(\frac{\pi}{2} e^{-\alpha x^\beta}\right), \quad x > 0$$

  and

  $$f_{SW}(x;\alpha,\beta) = \frac{\pi}{2}\alpha\beta x^{\beta-1} e^{-\alpha x^\beta} \sin\left(\frac{\pi}{2} e^{-\alpha x^\beta}\right), \quad x > 0,$$

  respectively. It is understood that $F_{SW}(x;\alpha,\beta) = f_{SW}(x;\alpha,\beta) = 0$ for $x \leq 0$.

- The CW distribution is defined with the cdf and pdf given as

  $$F_{CW}(x;\alpha,\beta) = 1 - \sin\left(\frac{\pi}{2} e^{-\alpha x^\beta}\right), \quad x > 0$$

  and

  $$f_{CW}(x;\alpha,\beta) = \frac{\pi}{2}\alpha\beta x^{\beta-1} e^{-\alpha x^\beta} \cos\left(\frac{\pi}{2} e^{-\alpha x^\beta}\right), \quad x > 0,$$

  respectively. It is understood that $F_{CW}(x;\alpha,\beta) = f_{CW}(x;\alpha,\beta) = 0$ for $x \leq 0$.

It is demonstrated that these extended trigonometric Weibull distributions have improved some of the skewness and kurtosis capabilities of the former Weibull distribution, showing better fits for important datasets. In light of this, we naturally decide to introduce the MPSW distribution by selecting the Weibull distribution as the baseline distribution for the MPSG family.

5.2. Definition of the MPSW distribution

Based on Equations (2.1) and (5.1), and the following classical trigonometric formula: $\sin(\pi x) = \sin(\pi(1-x))$, the MPSW distribution is defined by the following cdf:

$$F_{MPSW}(x;\alpha,\beta,\theta) = 1 - e^{-\alpha x^\beta} + \theta \sin(\pi e^{-\alpha x^\beta}), \quad x > 0,$$

and $F_{MPSW}(x;\alpha,\beta,\theta) = 0$ when $x \leq 0$. We recall that $\beta > 0$ is a shape parameter, $\alpha > 0$ is a scale parameter, and $\theta \in [-1/\pi, 1/\pi]$. Obviously, when $\theta = 0$, we get $F_{MPSW}(x;\alpha,\beta,\theta) = F_W(x;\alpha,\beta)$. Based on Equations (2.2), (5.1) and (5.2), the pdf of the MPSW distribution is given by

$$f_{MPSW}(x;\alpha,\beta,\theta) = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} \left[1 - \pi\theta \cos\left(\pi e^{-\alpha x^\beta}\right)\right], \quad x > 0 \tag{5.3}$$

and $f_{MPSW}(x;\alpha,\beta,\theta) = 0$ when $x \leq 0$.
The corresponding sf is given by

$$S_{MPSW}(x;\alpha,\beta,\theta) = e^{-\alpha x^\beta} - \theta \sin(\pi e^{-\alpha x^\beta}), \quad x > 0,$$

and $S_{MPSW}(x;\alpha,\beta,\theta) = 1$ when $x \leq 0$, and the corresponding hrf is obtained as

$$h_{MPSW}(x;\alpha,\beta,\theta) = \frac{\alpha\beta x^{\beta-1} e^{-\alpha x^\beta} \left[1 - \pi\theta \cos\left(\pi e^{-\alpha x^\beta}\right)\right]}{e^{-\alpha x^\beta} - \theta \sin\left(\pi e^{-\alpha x^\beta}\right)}, \quad x > 0,$$

and $h_{MPSW}(x;\alpha,\beta,\theta) = 0$ when $x \leq 0$.
In the next section, we examine the asymptotic and shape properties of the pdf, and other functions.

5.3. Shape behavior of the MPSW distribution

The MPSW distribution has the following asymptotic property: when $x \rightarrowtail 0$, $f_{MPSW}(x;\alpha,\beta,\theta) \sim \alpha\beta(1+\pi\theta)x^{\beta-1}$. We look into the three following cases:

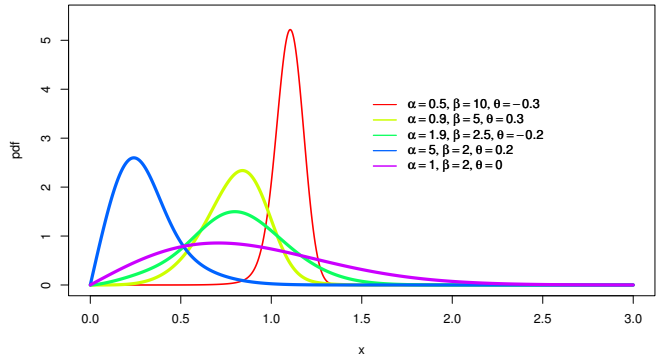
- when $\beta < 1$, $f_{MPSW}(x;\alpha,\beta,\theta) \rightarrowtail +\infty$,
- when $\beta = 1$, $f_{MPSW}(x;\alpha,\beta,\theta) \sim \alpha(1+\pi\theta)$,
- when $\beta > 1$, $f_{MPSW}(x;\alpha,\beta,\theta) \rightarrowtail 0$.

Hence, the parameter β is discriminative in this regard.
On the other hand, when $x \rightarrowtail +\infty$, $f_{MPSW}(x;\alpha,\beta,\theta) \rightarrowtail 0$ for all the values of the parameters.
The solutions to the following equation give the critical points of $f_{MPSW}(x;\alpha,\beta,\theta)$:
$d\log\left(f_{MPSW}(x;\alpha,\beta,\theta)\right)/dx = 0$, which is equivalent to

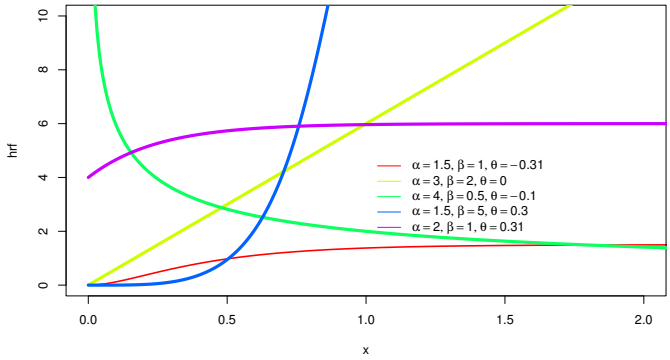
$$\frac{\beta-1}{x} - \alpha\beta x^{\beta-1} - \frac{\pi^2\alpha\beta\theta x^{\beta-1} e^{-\alpha x^\beta} \sin(\pi e^{-\alpha x^\beta})}{1 - \pi\theta\cos(\pi e^{-\alpha x^\beta})} = 0. \tag{5.4}$$

Numerical investigations show that the MPSW distribution can have one mode at maximum, and its value can be determined using any statistical software.

We complete this section with a graphical analysis to provide a thorough examination of the shape properties of the pdf and hrf of the MPSW distribution in Figure 1.



(a)



(b)

Figure 1: Shapes of the (a) pdf and (b) hrf of the MPSW distribution for different values of α, β and θ

The pdf shape can be spiked, nearly symmetrical bell-shaped, or skew to the left or right with varying degrees of kurtosis. In comparison to the pdf of the Weibull distribution, the pdf of the MPSW distribution seems to benefit from more kurtosis nuances, especially for the mesokurtic and leptokurtic cases, and is able to capture a small left-tail.
In contrast, the hrf can take on increasing, decreasing, (almost) linear, or J shapes.

5.4. Moments for MPSW distribution

In the next result, we develop a manageable series formula for the moments of a random variable X following the MPSW distribution.

Proposition 5.7. *Let X be a random variable following the MPSW distribution, and m be a positive*

integer. Then, the mth moment of X can be expressed as

$$\mu_m^{\iota} = E\left(X^m\right) = \alpha^{-m/\beta} \sum_{k=0}^{+\infty} a_k^{**} \frac{\Gamma\left(m/\beta + 1\right)}{(2k+1)^{m/\beta}},$$

where $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$, $x > 0$, *refers to the standard gamma function, and* a_k^{**} *is defined as in Equation* (3.2).

Proof. By the definition of moments and Equation (3.2), using the inversion of the sum and integral symbols which is possible by the Lebesgue theorem, we have

$$\begin{aligned}
\mu_m = E\left(X^m\right) &= \int_{-\infty}^{+\infty} x^m f_{MPSW}(x;\alpha,\beta,\theta)dx \\
&= \sum_{k=0}^{+\infty} a_k^{**}(2k+1) \int_0^{+\infty} x^m g_W(x;\alpha,\beta)[1 - G_W(x;\alpha,\beta)]^{2k} dx \\
&= \sum_{k=0}^{+\infty} a_k^{**}(2k+1) \int_0^{+\infty} x^m \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} e^{-2\alpha k x^{\beta}} dx \\
&= \alpha\beta \sum_{k=0}^{+\infty} a_k^{**}(2k+1) \int_0^{+\infty} x^{m+\beta-1} e^{-\alpha(2k+1)x^{\beta}} dx \\
&= \alpha^{-m/\beta} \sum_{k=0}^{+\infty} a_k^{**} \frac{\Gamma\left(m/\beta + 1\right)}{(2k+1)^{m/\beta}}.
\end{aligned}$$

This ends the proof. □

Owing to Proposition 5.7, the following finite sum approximation is valid:

$$\mu_m^{\iota} \approx \alpha^{-m/\beta} \sum_{k=0}^{\tau} a_k^{**} \frac{\Gamma\left(m/\beta + 1\right)}{(2k+1)^{m/\beta}},$$

where τ denotes a chosen large integer. All in this sum are simply calculable; the gamma function being implemented in most of the mathematical softwares.
From the moments, as for the general MPSG family, the mean, variance and measures of skewness and kurtosis of X can be calculated.
Table 1 indicates these measures for different values of α, β and θ.

Table 1: Moment measures, with skewness and kurtosis of the MPSW distribution for different values of α, β and θ.

		$\alpha = 0.25$	$\alpha = 2.5$	$\alpha = 5$	$\alpha = 10$
$\beta = 1$, $\theta = 0.1$	μ	3.259225	0.3259225	0.16296126	0.081480630
	σ^2	12.871419	0.1287142	0.03217855	0.008044637
	C_{sk}	2.371628	2.3716282	2.37162817	2.371628199
	C_{ku}	11.405947	11.4059473	11.40594736	11.405947379
		$\beta = 2$	$\beta = 3.5$	$\beta = 10$	$\beta = 25$
$\alpha = 1$, $\theta = -0.2$	μ	1.0807118	1.02082761	0.998877091	0.998353233
	σ^2	0.2024495	0.06661115	0.009316317	0.001609406
	C_{sk}	0.3776614	-0.23512304	-0.950718762	-1.046773542
	C_{ku}	3.2150591	3.19186361	4.857201893	6.562156397
		$\theta = -1/\pi$	$\theta = -0.1$	$\theta = 0$	$\theta = 0.2$
$\alpha = 0.3$, $\beta = 0.5$	μ	41.024680	28.129189	22.222223	10.40829
	σ^2	4186.354572	3084.789150	2469.135753	1028.47276
	C_{sk}	5.127459	5.922299	6.618781	10.16347
	C_{ku}	54.543873	71.261059	87.719809	202.19471

Table 1 illustrates numerically the observations made on Figure 1: We see that C_{sk} can be negative or positive (but is mainly positive). This indicates that the different levels of skewness that can be reached by the MPSW distribution. In addition, C_{ku} can be close to 3 or superior, for the considered values, corresponding to the mesokurtic and leptokurtic case. We conclude that the MPSW distribution is ideal to use as model for the analysis of data having analogeous empirical features.

5.5. ML estimation

We now investigate the estimation of the parameters of the MPSW distribution by the ML method, as described in a general setting in [7]. We thus aim to calculate the ML estimates for the parameters α, β, and θ, via the method described in Subsection 4.1.. Given n observations of a random variable X following the MPSW distribution, say $x_1, x_2, \ldots, x_n$, the log-likelihood function for the parameter vector $\delta = (\alpha, \beta, \theta)$ is

$$\log\left(L(\underline{x};\delta)\right) = n\log\alpha + n\log\beta + (\beta-1)\sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n}x_i^{\beta} + \sum_{i=1}^{n}\log\left[1-\theta\pi\cos\left(\pi e^{-\alpha x_i^{\beta}}\right)\right], \tag{5.5}$$

where $\underline{x} = \{x_1, x_2, \ldots, x_n\}$. The ML vector estimate of δ is obtained by maximizing this equation with respect to the vector parameter δ. It is denoted by $\hat{\delta} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})$, and $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\theta}$ are the ML estimates of α, β and θ, respectively. In practice, the log-likelihood function can be maximized simply by statistical programs, namely, R, Mathcad, SAS, Mathematica, and Ox programs. The ML estimates can also be obtained by solving the following nonlinear equations: $(U_\alpha, U_\beta, U_\theta) = (0, 0, 0)$, where

$$U_\alpha = \frac{d\log\left(L(\underline{x};\delta)\right)}{d\alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n}x_i^{\beta} - \pi^2\theta\sum_{i=1}^{n}\frac{x_i^{\beta}e^{-\alpha x_i^{\beta}}\sin\left(\pi e^{-\alpha x_i^{\beta}}\right)}{1-\theta\pi\cos\left(\pi e^{-\alpha x_i^{\beta}}\right)},$$

$$U_\beta = \frac{d\log\left(L(\underline{x};\delta)\right)}{d\beta} = \frac{n}{\beta} + \sum_{i=1}^{n}\log(x_i) - \alpha\sum_{i=1}^{n}x_i^{\beta}\log(x_i) - \sum_{i=1}^{n}\frac{\pi^2\alpha\theta x_i^{\beta}\log(x_i)e^{-\alpha x_i^{\beta}}\sin\left(\pi e^{-\alpha x_i^{\beta}}\right)}{1-\theta\pi\cos\left(\pi e^{-\alpha x_i^{\beta}}\right)}$$

and

$$U_\theta = \frac{d\log\left(L(\underline{x};\delta)\right)}{d\theta} = -\sum_{i=1}^{n}\frac{\pi\cos\left(\pi e^{-\alpha x_i^{\beta}}\right)}{1-\theta\pi\cos\left(\pi e^{-\alpha x_i^{\beta}}\right)}.$$

These equations cannot be solved analytically, but they can be solved numerically by software such as R, Mathematica, and Python. It is understood that $\hat{\delta}$ satisfies the properties of consistance and asymptotically normality under certain well-identified regularity conditions (see [7]).

6. PRACTICE

The MPSW distribution is now employed to fit several datasets of importance.

6.1. Methodology

In this section, we use two real-world datasets to decipher the applicability of the MPSW distribution over other the Weibull and trigonometric extended Weibull distributions.
Table 2 displays the distributions along with their cdfs competing against the MPSW distribution.

Table 2: Competitive distributions considered against the MPSW distribution

Distributions	References	cdfs
Weibull (W)	[14]	$1-e^{-\alpha x^{\beta}}$
Sin-Weibull (SW)	[22]	$\cos\left(\frac{\pi}{2}e^{-\alpha x^{\beta}}\right)$
Cos-Weibull (CW)	[23]	$1-\sin\left(\frac{\pi}{2}e^{-\alpha x^{\beta}}\right)$

Initially, we start by looking at the fundamental statistical metrics of the data, such as skewness, kurtosis, and central tendency. In this study, with the use of a boxplot, a visual analysis of the datasets is carried out. We also consider a graphical method based on the total test on time (TTT) plot in order to detect the form of the hrf of the datasets. In light of this, we consider the empirical TTT plot using the following equation:

$$T\left(\frac{r}{n}\right)=\frac{\sum_{i=1}^{r}x_{(i)}+(n-r)\,x_{(r)}}{\sum_{i=1}^{n}x_{(i)}},\quad r=1,2\ldots,n,$$

where $x_{(i)}$ and $x_{(r)}$ are the ith and rth (increasing) order statistics of the data. The shapes of the associated hrf are decreasing, growing, upside-down bathtub-shaped, and convex if the empirical TTT transform is convex, concave, convex then concave, and concave then convex, respectively. See [1].
In order to analyze the statistical fitting of the considered models, (i) the parameters of the models are estimated by the ML method with the help of the optim function in R software, and (ii) we use standard goodness-of-fit measures. These measures include the $-2\log L$ ($-2\log L$), the Akaike information criterion (AIC), the Cramér-von Mises (CVM) statistic, the Anderson-Darling (AD) statistic, and the Kolmogorov-Smirnov (KS)statistic and its p-value.
Finally, we plot the fitted empirical pdfs (epdfs) and the empirical cdfs (ecdfs) of the considered models based on the datasets.

6.2. COVID-19 mortality rates

We use two real-world datasets, from the medical field. The data are related to the recent spread of the COVID-19 pandemic in 2020.

The first dataset is a 102-day COVID-19 dataset for the USA, spanning the period from March 28, 2019, to July 7, 2020. These data were gathered by dividing the daily death toll by the daily new case toll. For the instant access to the data, they are: 0.0149, 0.0235, 0.0230, 0.0159, 0.0200, 0.0413, 0.0360, 0.0378, 0.0363, 0.0399, 0.0453, 0.0436, 0.0598, 0.0624, 0.0546, 0.0607, 0.0609, 0.0521, 0.0615, 0.0928, 0.2232, 0.0620, 0.0812, 0.0629, 0.0651, 0.0840, 0.1072, 0.0821, 0.0567, 0.0559, 0.0606, 0.0380, 0.0586, 0.0980, 0.0925, 0.0631, 0.1869, 0.0049, 0.0176, 0.0495, 0.1112, 0.0890, 0.0940, 0.0600, 0.0652,

0.0413, 0.0588, 0.0665, 0.0816, 0.0753, 0.0579, 0.0436, 0.0527, 0.0382, 0.0568, 0.0613, 0.0531, 0.0767, 0.0400, 0.0406, 0.0237, 0.0471, 0.0722, 0.0595, 0.0597, 0.0389, 0.0265, 0.0518, 0.0419, 0.0566, 0.0516, 0.0390, 0.0245, 0.0266, 0.0314, 0.0701, 0.0410, 0.0436, 0.0320, 0.0255, 0.0171, 0.0268, 0.0259, 0.0333, 0.0318, 0.0188, 0.0172, 0.0112, 0.0155, 0.0229, 0.0184, 0.0621, 0.0146, 0.0114, 0.0216, 0.0103, 0.0129, 0.0134, 0.0117, 0.0143, 0.0032 and 0.0054.

The second dataset is a 127-day COVID-19 dataset for Italy, spanning the period from 1 March to 6 July 2020. The data are calculated by dividing the daily new deaths by the daily new cases. For the instant access to data, they are: 0.0107, 0.0490, 0.0601, 0.0460, 0.0533, 0.0630, 0.0297, 0.0885, 0.0540, 0.1720, 0.0847, 0.0713, 0.0989, 0.0495, 0.1025, 0.1079, 0.0984, 0.1124, 0.0806, 0.1044, 0.1212, 0.1167, 0.1255, 0.1416, 0.1315, 0.1073, 0.1629, 0.1485, 0.1453, 0.2000, 0.2070, 0.1520, 0.1628, 0.1666, 0.1417, 0.1221, 0.1767, 0.1987, 0.1408, 0.1456, 0.1443, 0.1319, 0.1053, 0.1789, 0.2032, 0.2167, 0.1387, 0.1646, 0.1375, 0.1421, 0.2012, 0.1957, 0.1297, 0.1754, 0.1390, 0.1761, 0.1119, 0.1915, 0.1827, 0.1548, 0.1522, 0.1369, 0.2495, 0.1253, 0.1597, 0.2195, 0.2555, 0.1956, 0.1831, 0.1791, 0.2057, 0.2406, 0.1227, 0.2196, 0.2641, 0.3067, 0.1749, 0.2148, 0.2195, 0.1993, 0.2421, 0.2430, 0.1994, 0.1779, 0.0942, 0.3067, 0.1965, 0.2003, 0.1180, 0.1686, 0.2668, 0.2113, 0.3371, 0.1730, 0.2212, 0.4972, 0.1641, 0.2667, 0.2690, 0.2321, 0.2792, 0.3515, 0.1398, 0.3436, 0.2254, 0.1302, 0.0864, 0.1619, 0.1311, 0.1994, 0.3176, 0.1856, 0.1071, 0.1041, 0.1593, 0.0537, 0.1149, 0.1176, 0.0457, 0.1264, 0.0476, 0.1620, 0.1154, 0.1493, 0.0673, 0.0894 and 0.0365.

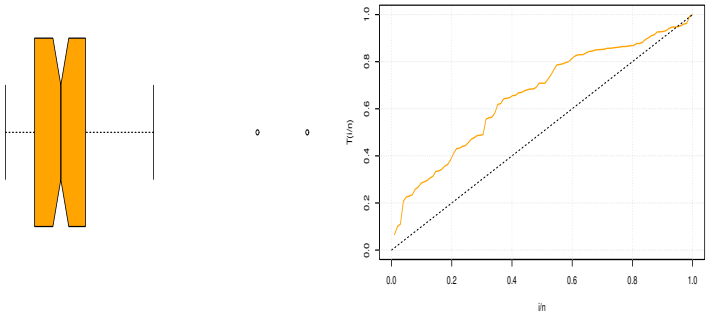
The descriptive measures of both the datasets are mentioned in Table 3.

Table 3: Descriptive statistics of the COVID-19 dataset by countries

Country	Mean	Standard deviation	Skewness	Kurtosis
USA	0.04881	0.03325986	2.152116	11.28798
Italy	0.1609	0.07547866	0.9565583	5.360532

From Table 3, we can observe that the COVID-19 dataset in the USA is right-skewed and leptokurtic in nature, whereas the dataset of Italy, is slightly right-skewed, but also leptokurtic in nature.

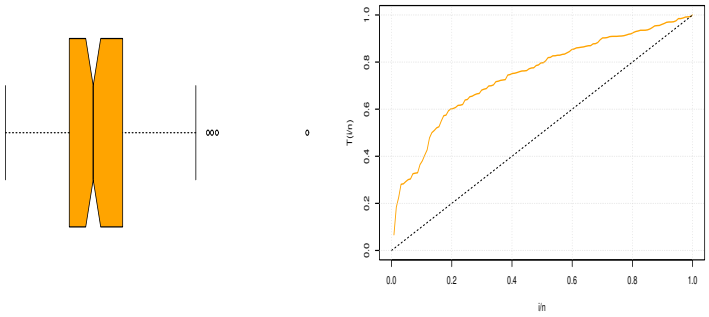
Figures 2 and 3 illustrate the boxplots and TTT plots of the COVID-19 dataset in the USA and Italy.



(a) (b)

Figure 2: Illustration of the (a) boxplot and (b) TTT plot of the COVID-19 dataset in the USA

From Figure 2, it is clear that the boxplot contains outliers, having a long right tail compared to the left tail. As the colored TTT line is concave, the associated empirical hrf for the USA dataset is increasing.



(a) (b)

Figure 3: Illustration of the (a) boxplot and (b) TTT plot of the COVID-19 dataset in the Italy

The boxplot in Figure 3 also contains outliers and has equal tail lengths, implying that the dataset is slightly right-skewed. As the colored TTT line is concave, the associated empirical hrf for the Italy dataset is increasing.

Table 4 provides the ML estimates for the datasets in the USA and Italy on the considered models, as well as their standard errors (SEs).

Table 4: Parameter estimates and SEs (under positioned) of the parameters of the MPSW model with other competitive models

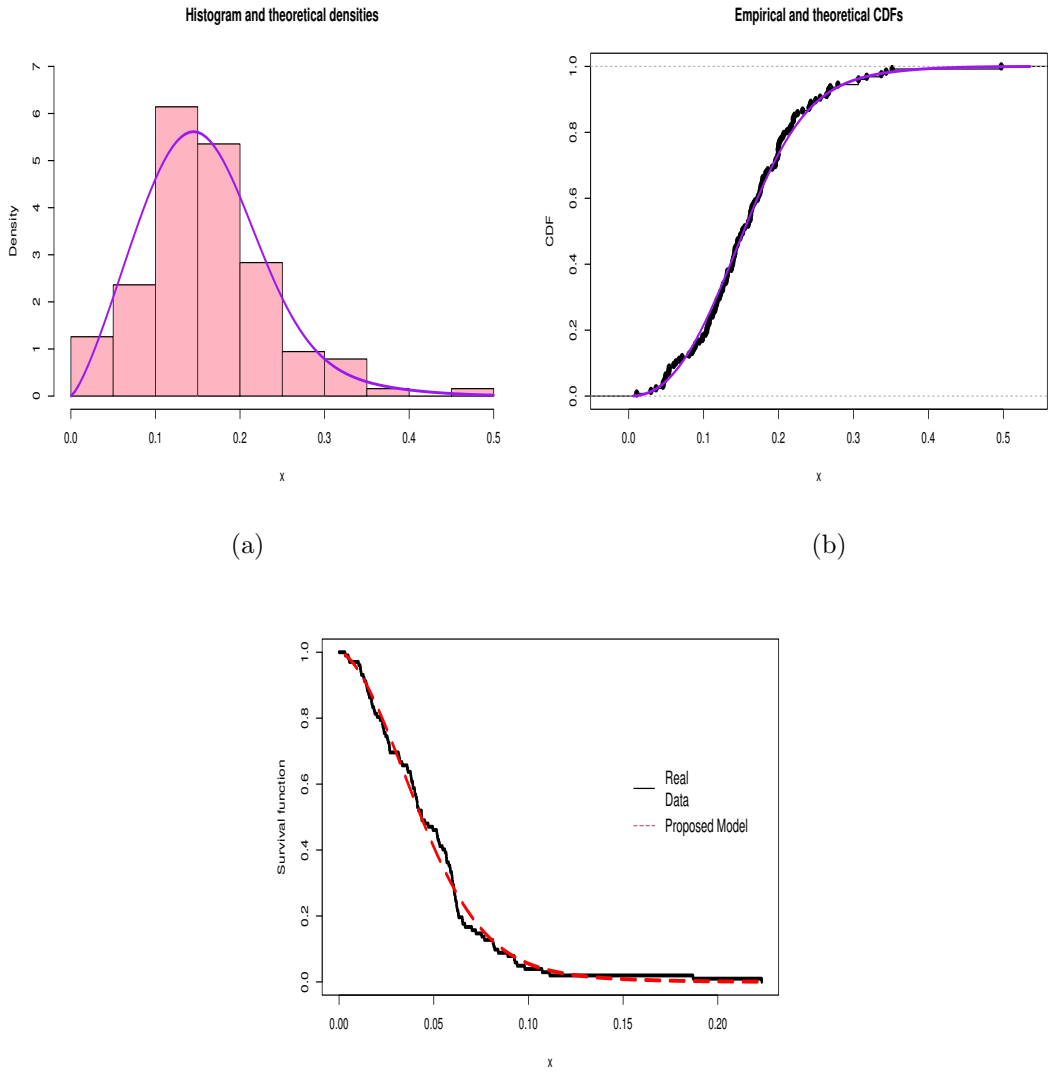
Country	ML estimates SE	MPSW	CW	W	SW
USA	$\hat{\alpha}$	60.94589	43.89717	98.47946	45.10554
		23.04937	11.10586	30.90252	13.69119
	$\hat{\beta}$	1.663611	1.161887	1.578016	1.501706
		0.1297254	0.08460758	0.113909	0.1104469
	$\hat{\theta}$	0.2843082	-	-	-
		0.02596361	-	-	-
Italy	$\hat{\alpha}$	26.77666	23.95018	45.40631	21.43757
		12.14315	4.615214	10.85249	4.9295
	$\hat{\beta}$	2.32632	1.625295	2.235264	2.123184
		0.2061552	0.1078028	0.14735064	0.142248
	$\hat{\theta}$	0.266669	-	-	-
		0.05219571	-	-	-

Table 5 provides the goodness-of-fit measures of the MPSW distribution and other models for the COVID-19 dataset by countries.

Table 5: Goodness-of-fit measures of the MPSW model and other models on COVID-19 dataset by countries

Country		MPSW	CW	W	SW
USA	KS	0.07229826	0.08368365	0.08933719	0.08769048
	p-value	0.6605885	0.4726967	0.3896192	0.4128757
	$-2\log L$	-450.0121	-445.4612	-442.4261	-442.6405
	AIC	-444.0121	-441.4612	-438.4261	-438.6405
	CVM	0.09561021	0.1556808	0.1309334	0.123618
	AD	0.5519462	0.8942431	0.8916152	0.8549048
Italy	KS	0.04632975	0.05512264	0.06246413	0.0621645
	p-value	0.9480196	0.8350289	0.704675	0.7103054
	$-2\log L$	-311.7072	-307.2932	-306.5651	-306.8347
	AIC	-305.7072	-303.2932	-302.5651	-302.8347
	CVM	0.04060696	0.09139985	0.1064359	0.5788786
	AD	0.2834435	0.6291132	0.6595867	0.634789

The MPSW model has the highest p-value for the datasets of both countries, the USA and Italy, according to Table 5. Furthermore, it has the lowest values for the $-2\log L$, AIC, CVM, and AD when compared to the other competitive models. In order to support the numerical analysis of goodness-of-fit, we present a visual representation of the datasets by countries through the MPSW model. Figures 4 and 5 illustrate the epdf, ecdf, and the Kaplan-Meier survival plot of the datasets of the MPSW model on the COVID-19 datasets in the USA and Italy.

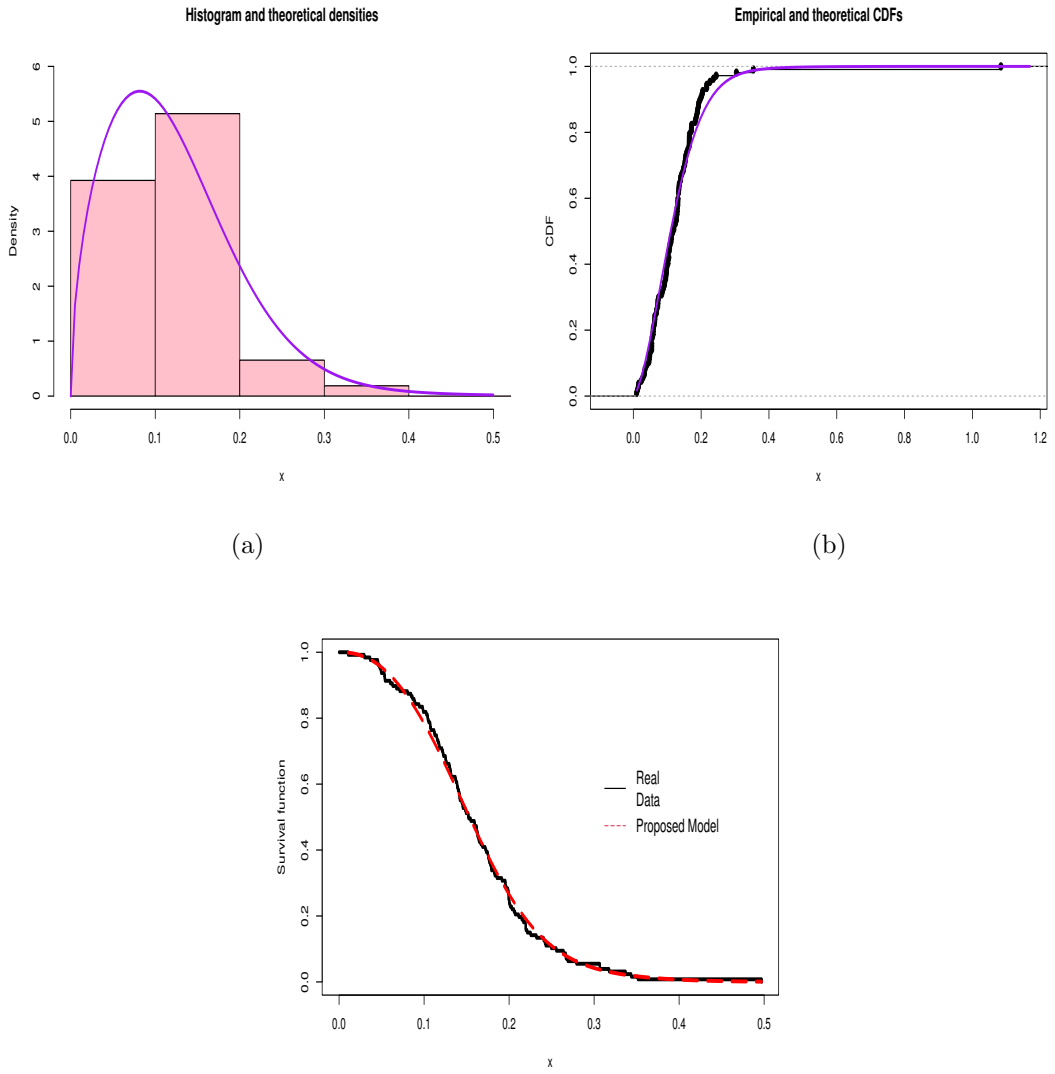


(a) (b)

(c)

Figure 4: Illustration of the (a) fitted epdf, (b) ecdf and (c) Kaplan-Meier survival plots of the MPSW model on the COVID-19 dataset in the USA

It can be observed from Figure 4 that the corresponding estimated functions of the MPSW model clearly outline the histogram, cdf and the Kaplan-Meier survival plots of the COVID-19 dataset in the USA.



(a) (b)

(c)

Figure 5: Illustration of the (a) fitted epdf, (b) ecdf and (c) Kaplan-Meier survival plots of MPSW model for the COVID-19 data in Italy

Figure 5 shows that the histogram, cdf, and Kaplan-Meier survival plots of the COVID-19 dataset in Italy are clearly outlined by the corresponding estimated functions of the MPSW model.

7. CONCLUSION

The trigonometric families of distributions have proven themselves as generators of powerful statistical models, being able to handle data with a high level of complexity on the skewness and kurtosis aspects. For the very first time, this work offered a new modified polynomial-sine-generated (MPSG) family that

aims to unify the baseline distribution and the famous SinG family. The baseline-SinG switch is made possible thanks to an additional scale parameter that is quite simple to modulate. The fundamental mathematical characteristics of this family were derived. The Weibull distribution was chosen as the baseline distribution for a motivated example, resulting in the MPSW distribution. The MPSW model was empirically tested for applicability and adaptability using two real datasets (COVID-19 datasets), which is of utmost importance in the current scenario, proving that it can yield a better fit than rival models like the Sine-Weibull, Cosine-Weibull, and Weibull models. This was highlighted by using goodness-of-fit measures, such as the Akaike information criterion, Kolmogorov-Smirnov statistic with its p-value, Anderson-Darling statistic, and Cramér-von Mises statistic. Graphical examples of how the new distribution fit the two datasets were provided to clarify the meaning of the numerical measures.

Beyond the findings of this study the perspectives opened by our new modeling strategy include: (i) the investigation of the extended MPSG family as presented in Remark 2.2, (ii) the adaptation of our statistical methodology to the incomplete data, which often appear in practice; (iii) the development of bivariate trigonometric models, such as trigonometric copula models, regression models, survival models, etc.; (iv) the application of our trigonometric models in challenging problems arising in climate changes as global warming (see [21] and [17]); and (v) the development of new discrete trigonometric models.

RECEIVED: JULY, 2023.
REVISED: NOVEMBER, 2023.

REFERENCES

[1] AARSET, M. V. (1987): How to identify a Bathtub Hazard Rate. **IEEE Transactions on Reliability**, 36, 106-108. doi: 10.1109/TR.1987.5222310

[2] AL-BABTAIN, A. A., ELBATAL, I., CHESNEAU, C. and ELGARHY, M. (2020): Sine Topp-Leone-G family of Distributions: Theory and Applications. **Open Physics**, 18, 574–593. doi: 10.1515/phys-2020-0180

[3] AL-FARIS, R. Q. and KHAN, S. (2008): Sine square distribution: A new statistical model based on the sine function. **Journal of Applied Probability and Statistics**, 3, 163–173. url: http://japs.isoss.net/ isoss16a.pdf

[4] ALGARNI, A. (2020): Sine power lomax model with application to bladder cancer data. **Nanoscience and NanoTechnology Letters**, 12, 677–684. doi: 10.1166/nnl.2020.3137

[5] AMPADU, C. B. (2020): The Burr XI generated family of distribution with illustration to cancer patients data. **Global Journal of Cancer Case Reports**, 1, 1–4. doi: 10.47733/GJCCR.2020.1102

[6] BURROWS, P. M. (1986): Extreme statistics from the sine distribution. **The American Statistician**, 40, 216–217. doi: 10.1080/00031305.1986.10475396

[7] CASELLA, G. and BERGER, R. L. (1990): **Statistical Inference**, Brooks/Cole Publishing Company: Bel Air, CA, USA.

[8] CHAKRABORTY, S., HAZARIKA, P. and ALI, M. (2012): A new skew logistic distribution and its properties. **Pakistan Journal of Statistics**, 28, 513–524.

[9] CHESNEAU, C., BAKOUCH, H. S. and HUSSAIN, T. (2019): A new class of probability distributions via cosine and sine functions with applications. **Communications in Statistics-Simulation and Computation**, 48, 2287–2300. doi: 10.1080/03610918.2018.1440303

[10] EDWARDS, A. W. F. (2000): Gilberts Sine Distribution. **Teaching Statistics**, 22(3), 70–71. doi: 10.1111/1467-9639.00026

[11] FAYOMI, A., ALGARNI, A. and ALMARASHI, A. M. (2021): Sine inverse lomax generated family of distributions with applications. **Mathematical Problems in Engineering**, 1–11. doi: 10.1155/2021/1267420

[12] GATTO, R. and JAMMALAMADAKA, S. R. (2007): The generalized von mises distribution. **Statistical Methodology**, 4, 341–353. doi: 10.1016/j.stamet.2006.11.003

[13] ISA, A. M., ALI, B. A. and ZANNAH, U. (2022): Sine burr xii distribution: properties and application to real data sets. **AJBAR**, 1, 48–58. doi: 10.55639/607lkji

[14] JOHNSON, N., KOTZ, S. and BALAKRISHNAN, N. (1995): **Continuous Univariate Distributions**, Wiley, New York, USA.

[15] KUMAR, D., SINGH, U. and SINGH, S. K. (2015): A new distribution using sine function: its application to bladder cancer patients data. **Journal of Statistics Applications and Probability**, 4, 417–427. doi: 10.12785/jsap/040309

[16] KOTZ, S., LUMELSKII, Y. and PENSKEY, M. (2003): **The stress-strength model and its generalizations and applications**, World Scientific, Singapore.

[17] OLATAYO, K., MATIVENGA, P. and MARNEWICK, A. (2021): Comprehensive evaluation of plastic flows and stocks in South Africa. **Resources, Conservation and Recycling**, 170, 105567.

[18] NADARAJAH, S. and KOTZ, S. (2006): Beta trigonometric distributions. **Portuguese Economic Journal**, 5, 207–224. doi: 10.1007/s10258-006-0013-6

[19] NAGARJUNA, V. B., VARDHAN, R. V. and CHESNEAU, C. (2021): On the accuracy of the sine power lomax model for data fitting. **Modelling**, 2, 78–104. doi: 10.3390/modelling2010005

[20] PENG, X. (2018): An extended Weibull model with variable periodicity. **Journal of Systems Science and Complexity**, 31, 841–858.

[21] SOLDATENKO, S., BOGOMOLOV, A. and RONZHIN, A. (2021): Mathematical modelling of climate change and variability in the context of outdoor ergonomics. **Mathematics**, 9, 2920.

[22] SOUZA, L., JÚNIOR, W. R. O., BRITO, C. C. R., CHESNEAU, C., FERREIRA, T. A. E. and SOARES, L. G. M. (2019a): On the sin-G class of distributions: theory, model and application. **Journal of Mathematical Modeling**, 7, 3, 357–379. doi: 10.22124/jmm.2019.13502.1278

[23] SOUZA, L., JUNIOR, W. R. D. O., DE BRITO, C. C. R., FERREIRA, T. A. E. and SOARES, L. G. M. (2019b): General properties for the C=cos-G class of distributions with applications. **Eurasian Bulletin of Mathematics** (ISSN: 2687-5632), 63–79.

[24] TOMY, L., G, V. and CHESNEAU, C. (2021): The sine modified lindley distribution. **Mathematical and Computational Applications**, 26, 81.

[25] TOMY, L., G, V. and CHESNEAU, C. (2022): Applications of the sine modified Lindley distribution to biomedical data. **Mathematical and Computational Applications**, 27, 43, 1–16.

[26] WEIBULL, W. (1938): Investigate into strength properties of brittle material. **Generalstabens Litografiska Anst.**