Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v18n1p183

Weibull Sine Generalized Distribution Family: Fundamental Properties, Sub-model, Simulations, with Biomedical Applications

By Obulezi et al.

15 March 2025

Electronic Journal of Applied Statistical Analysis
Vol. 18, Issue 01, march 2025, 183-212
DOI: 10.1285/i20705948v18n1p183

Weibull Sine Generalized Distribution Family: Fundamental Properties, Sub-model, Simulations, with Biomedical Applications

Okechukwu J. Obulezi*[a], Happiness O. Obiora-Ilouno[a], George A. Osuji[a], Mohamed Kayid[b], and Oluwafemi Samson Balogun[c]

[a]*Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, , P.O. Box 5025 Awka, Nigeria.*
[b]*Department of Statistics and Operations Research, College of Science, King Saud University, , P.O. Box 2455, Riyadh 11451, Saudi Arabia*
[c]*Department of Computing, University of Eastern Finland, , FI-70211, Finland*

15 March 2025

This study explores the integration of trigonometric functions into traditional statistical models, focusing on the development of the Weibull Sine Generalized (WSG-G) family of distributions. A special case was formulated name Weibull Sine generalized exponential (WSG-E) distribution. This new distribution extends the baseline exponential distribution, accommodating heavier tails and outliers, thereby effectively modeling positively skewed data. Key statistics such as mean, variance, skewness, and kurtosis indicate the distribution's capacity to handle clustered data. A simulation study demonstrates the performance of Maximum Likelihood Estimation (MLE), revealing convergence in the mean squared error and root mean squared error for the parameter α with increasing sample sizes, although convergence is less evident for other parameters. The WSG-E distribution's applicability is further illustrated through its fitting of medical datasets on bladder cancer remission times and growth hormone deficiency in children, both characterized by extreme values. Overall, the WSG-E distribution proves to be a robust model for skewed data, and future research could extend this framework to additional continuous distributions.

*Corresponding authors: oj.obulezi@unizik.edu.ng

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

keywords: Biomedical data, Lomax tangent generalized Family, Weibull Distribution, Weibull Sine Generalized Exponential distribution, Weibull Sine Generalized Family.

1 Introduction

Merging trigonometric functions with traditional statistical models has the potential to improve analytical techniques and inference methods, thereby enriching the landscape of modern statistical practice, Zaidi et al. (2024). In the study by Zaidi et al. (2024), $\tan\left(\frac{\pi x}{2}\right)$ was used as a generalizer in the T-X generator proposed by Alzaatreh et al. (2013). The function $\tan\left(\frac{\pi x}{2}\right)$ possesses several unique properties that make it notable in mathematical analysis. One key feature is its periodicity, as it repeats its values every 2 units. Additionally, the function has vertical asymptotes at odd integers, where it approaches infinity. This boundary behavior is particularly interesting, as it transitions from $-\infty$ to $+\infty$ as x approaches these odd integers. Furthermore, the range of $\tan\left(\frac{\pi x}{2}\right)$ encompasses all real numbers, indicating that it can take any value. The transformation also compresses or stretches the input to the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ for x within the interval $(0, 1)$, altering its standard behavior. These characteristics enhance its utility in various mathematical and statistical applications.

However, in this study, it is our choice to use the function $f(x) = \sin\left(\frac{\pi x}{4}\right)$ which also exhibits several notable properties, including periodicity with a period of 8, meaning it repeats every 8 units along the x-axis, (Kreyszig (2021); Hogg et al. (2013); Besicovitch and Besicovitch (1954)), and an amplitude of 1, indicating values are bounded between -1 and 1, (Parumasur and Mika (2005)). It has zeros at regular intervals, specifically $x = 0, 4, 8, 12, \ldots$, where it crosses the x-axis due to $\sin(n\pi) = 0$ for integers n, (Suslov (2003)). Additionally, it is an odd function, demonstrating symmetry with respect to the origin, which enhances its predictable behavior, (Li et al. (2013); Alzahrani et al. (2023)). The function plays a significant role in various mathematical applications, such as signal processing, where sine waves represent periodic signals like sound and radio waves, aiding in the analysis of these signals, (Eugene and Manfred (2019)). It also arises in differential equations describing oscillatory phenomena, including harmonic oscillators that model vibrating systems and the wave equation for wave propagation. The sine function plays a crucial role in Fourier series, which break down periodic functions into sums of sine and cosine terms, widely utilized in both mathematics and engineering for approximation and analysis, see Bracewell et al. (2000). In quantum mechanics, sine functions are integral to the solutions of the Schrödinger equation, where they describe particle probability distributions that align with the wave-like characteristics of quantum systems, (Griffiths and Schroeter (2018)). Additionally, in control theory, sine functions are employed to model oscillatory systems and assess their responses to periodic inputs, (Nise (2020)). In data analysis, these functions are vital for time-series analysis, aiding in the identification of periodic patterns using Fourier techniques, which are significant in fields such as economics and engineering, see Hamilton (2020). Furthermore, in geometry and trigonometry, $\sin\left(\frac{\pi x}{4}\right)$ is essential for calculating angles

and distances in right triangles, with practical applications in navigation and engineering, see Stewart (2011). In electrical engineering, it models alternating current circuits, capturing the oscillatory behavior of electrical signals, which is fundamental for design and analysis, see William et al. (2007). Finally, sine functions are used in graphics to create smooth transitions and animations, simulating waves and periodic motion in visual representations.

Several researchers have successfully integrated various trigonometric functions into probability distributions to enhance existing methods. Among them, Oramulu et al. (2024) introduced the sine generalized family of distributions, while Al-Babtain et al. (2020) proposed the Sine Topp-Leone-G family. Algarni (2020) presented the Sine power Lomax model, applying it to bladder cancer data. Chakraborty et al. (2012) developed a new skew logistic distribution, and Chesneau and Jamal (2020) contributed with the sine Kumaraswamy-G family of distributions. Additionally, Chesneau et al. (2019) devised a new class of distributions by combining sine and cosine functions, whereas Kharazmi et al. (2022) designed the Arctan-based family of distributions. Kharazmi et al. (2017) explored the hyperbolic sine-Weibull distribution, and Nadarajah and Kotz (2006) constructed the beta trigonometric distributions. Further contributions include Nagarjuna et al. (2021), who examined the effectiveness of the sine power Lomax model for various datasets, and Raab and Green (1961), who investigated cosine approximation for the normal distribution. Mahmood et al. (2019) proposed the new sine-G family of distributions by transforming two cumulative distribution functions (CDF) using trigonometric methods. Muse et al. (2021) modified the classical log-logistic distribution to create the log-logistic tangent distribution, using it to model COVID-19 mortality rates in Somalia. Finally, Shrahili et al. (2021a) developed the sine inverted exponential distribution, estimating its parameters under a censored scheme, and Shrahili et al. (2021b) introduced the sine half-logistic inverse Rayleigh distribution.

We provide this study because there is a need for more flexible statistical models that can precisely describe complicated data structures, especially for cases with heavy tails, skewness, and outliers. In fact, such data would often be tricky for classical probability distributions, resulting in poor fit with unreliable inference. It has now become apparent in recent advances in statistical modeling that the incorporation of trigonometric functions into probability distributions improves flexibility and applicability. For this purpose, we introduce the Weibull Sine Generalized (WSG-G) family of distributions, placing particular emphasis on the selective case of the Weibull Sine Generalized Exponential (WSG-E) distribution here. The WSG-E distribution rewards the integration of the sine function in the generalization with an opportunity to be more flexible toward modeling datasets from nature in, particularly medical and reliability studies. This motivation is strengthened due to having the capacity to represent oscillatory properties, periodicity, and bounded characteristics of the input, resulting in a major boost in the statistical inference and parameter estimation. The paper also aims to test the performance of Maximum Likelihood Estimate (MLE) of the new distribution under simulation studies. Practical relevance with a better fit against already existing models is demonstrated through the application of the WSG-E distribution to the medical datasets of bladder cancer remission times and growth hormone deficiency cases. This

research not only is a significant advancement in statistical methodology but will also serve as a valuable tool for practitioners dealing with complex datasets positively skewed in nature.

The remaining parts of this article are organized as follows; in section (2), the new family of distributions is constructed; section (3) discusses the general properties; in section (4), a special sub-model is studied; section (5), the properties of the sub-model are discussed; section (6) is dedicated to the estimation of the parameters of the sub-model using the maximum likelihood method; in section (7), we carry out simulation study to determine the reliability, consistency and accuracy of the estimators of the parameters; in section (8), two medical records are used to illustrate the viability of the proposed sub-model and a final remark with recommendation for future studies are contained in section (9).

2 Model Construction

Alzaatreh et al. (2013) presented a technique for constructing continuous univariate distributions and their families by substituting the beta probability density function (PDF) with the PDF $\psi(t)$ of a continuous random variable T. This method employs the function $W\{G(x)\}$, which must satisfy the following criteria:

1. $W\{G(x)\} \in [a, b]$.
2. W is differentiable and monotonically non-decreasing.
3. $Z\{G(x)\} \to a$ as $x \to -\infty$ and $W\{G(x)\} \to b$ as $x \to \infty$.

Here, $[a, b]$ represents the range of the random variable T, where $-\infty \leq a < b \leq \infty$. This formulation leads to the noteworthy $T - X\{W\}$ generator of distribution families, with the cumulative distribution function (CDF) given by:

$$F(x) = \int_a^{W\{G(x)\}} r(t)dt = R\{W(G(x))\}, \tag{1}$$

where R denotes the CDF of the random variable T, functioning as the transformer. The PDF corresponding to equation (1) is expressed in equation (2)

$$f(x) = \left\{ \frac{\partial}{\partial x} W(F(x)) \right\} r\{W(G(x))\}. \tag{2}$$

Definition 2.1 (Weibull Sine generalized (WSG-G) family of distributions)**.** *Let T be a random variable defined for $0 < t < \infty$ with a probability density function given by $r(t) = ck\left[1 + \left(\frac{t}{s}\right)\right]^{-k-1}$. Zaidi et al. (2024) applied a trigonometric generalization of this distribution by introducing the function $W[G(x)] = \tan\left(\frac{\pi}{2}G(x)\right)$, which satisfies the conditions of the T-X generalization approach in (2). Using this, the cumulative distribution function (CDF) and probability density function (PDF) for the Lomax tangent generalized (LT-G) family are expressed in equations (3) and (4)*

$$F(x) = \int_0^{\tan\left(\frac{\pi G(x)}{2}\right)} r(t)\,dt = 1 - \left[1 + \left(\frac{\tan\left(\frac{\pi}{2}G(x)\right)}{s}\right)\right]^{-k}, \tag{3}$$

and

$$f(x) = \frac{\pi k}{2s} g(x) \sec^2\left(\frac{\pi}{2}G(x)\right) \left[1 + \left(\frac{\tan\left(\frac{\pi}{2}G(x)\right)}{s}\right)\right]^{-k-1}. \tag{4}$$

respectively. The above interesting model form the background for this study. In the present study, we utilize a different trigonometric generalizer given by,

$$W[G(x)] = \sin\left(\frac{\pi}{4}G(x)\right). \tag{5}$$

The usefulness of the function represented in equation (5) has been discussed in the introductory section of this article with wide-ranging applications in mathematics and science. From signal processing and differential equations to quantum mechanics and control theory, this function plays a crucial role in modeling and solving problems that involve periodic behavior. Its versatility makes it indispensable in both theoretical and applied mathematics. When the generalizer of the LT-G family is replaced with equation (5) as well as using the Weibull distribution as the transformer, so that $r(t) = \left(\frac{\nu}{\rho}\right)\left(\frac{t}{\rho}\right)^{\nu-1} e^{-\left(\frac{t}{\rho}\right)^{\nu}}$, a new family is produced, having its CDF as

$$F(x) = \int_0^{\sin\left(\frac{\pi}{4}G(x)\right)} r(t)\,dt = 1 - e^{-\left\{\frac{\sin\left(\frac{\pi}{4}G(x)\right)}{\rho}\right\}^{\nu}}, \quad x > 0, \quad \nu > 0 \tag{6}$$

with the associated PDF given as

$$f(x) = \frac{\nu\pi g(x)}{4\rho^{\nu}} \cos\left(\frac{\pi}{4}G(x)\right) \left\{\sin\left(\frac{\pi}{4}G(x)\right)\right\}^{\nu-1} e^{-\left\{\frac{\sin\left(\frac{\pi}{4}G(x)\right)}{\rho}\right\}^{\nu}}, \quad x > 0, \quad \nu > 0. \tag{7}$$

Equations (6) and (7) are the CDF and PDF of the Weibull Sine generalized (WSG-G) family of distributions. It is important to note that this new formulation also satisfies the conditions of the T-X generator of families of continuous distributions in (2).

2.1 Mixture Representation

To represent the PDF of the WSG-G family using a mixture of power series and binomial expansions, the power series expansion of $\cos(z)$ around $z = 0$ is expressed as $\cos(z) = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!}$. Similarly, $\sin(z) = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!}$. The binomial expansion of $(1+z)^r = \sum_{k=0}^{\infty} \binom{r}{k} z^k$. The power series expansion of $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$. Putting these together, the PDF of the WSG-G family can be denoted as in equation (8)

$$f(x) = \frac{\nu\pi g(x)}{4\rho^{\nu}} \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty} C_{n,k,m,j} G(x)^{2n+k(2m+1)+\nu j}, \tag{8}$$

where $C_{n,k,m,j} = \frac{(-1)^{n+m+j}}{(2n)!(2m+1)!j!}\binom{\nu-1}{k}\left(\frac{\pi}{4}\right)^{2n+k(2m+1)}\frac{1}{\beta^{\nu j}}$.

3 Generic Properties of the WSG-G Family

We study some commonly used mathematical properties of the proposed family in this section.

3.1 The p-th Quantile function

Let $p \in [0,1]$ represent a fixed cumulative probability, the quantile function $Q(p) = F_X^{-1}(p)$ inverts the CDF and gives the value of x such that $F_X(x) = p$. $U \sim U(0,1)$ is a uniform random variable, and we can generate samples from X using the relation $X = Q(U)$, where U is a realization from $U(0,1)$. Therefore, an expression for the quantile function of the WSG-G family is

$$x_p = G^{-1}\left\{\frac{4}{\pi}\arcsin\left[\rho\left\{-\ln(1-p)\right\}^{\frac{1}{\nu}}\right]\right\}. \tag{9}$$

Further, equation (9) can be used to derived the median and other quantiles of distribution whose parent is $G(x)$ by replacing p with $\frac{1}{2}$.

3.2 Ordinary Moment

Let $X \sim$ WSG-G(ν, ρ, Θ), where Θ is the vector of the parameters of the parent distribution with CDF $G(x)$, the rth ordinary moment can be obtained by building upon the result in equation (8), hence

$$\mu_r^{'} = \frac{\nu\pi}{4\rho^{\nu}} \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty} C_{n,k,m,j} \int_0^{\infty} x^r g(x) G(x)^{2n+k(2m+1)+\nu j}\, dx. \tag{10}$$

3.3 Shape of the WSG-G Family

we start by determining the value of x that maximizes $f(x)$. The mode is defined as the point where the probability density function reaches its maximum, which involves differentiating $f(x)$ with respect to x and setting the result equal to zero.

We differentiate $f(x)$ using the product and chain rules, focusing particularly on the functions $g(x)$ and $G(x)$. This gives us the expression

$$\frac{df(x)}{dx} = \frac{\nu\pi}{4\rho^{\nu}}\left(g'(x)\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty}C_{n,k,m,j}G(x)^{2n+k(2m+1)+\nu j}+\right.$$
$$\left.g(x)\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty}C_{n,k,m,j}\frac{d}{dx}\left(G(x)^{2n+k(2m+1)+\nu j}\right)\right).$$

Setting this derivative equal to zero, we obtain the condition for the mode:

$$g'(x)\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty}C_{n,k,m,j}G(x)^{2n+k(2m+1)+\nu j}+$$
$$g(x)\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty}C_{n,k,m,j}\frac{d}{dx}\left(G(x)^{2n+k(2m+1)+\nu j}\right)=0.$$

Solving this equation will yield the mode x_{mode}, though the specific form of the solution may depend on the functions $g(x)$ and $G(x)$. Thus, we can summarize the expression for the mode as follows:

$$x_{\text{mode}}: \quad g'(x)\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty}C_{n,k,m,j}G(x)^{2n+k(2m+1)+\nu j}+$$
$$g(x)\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty}C_{n,k,m,j}\frac{d}{dx}\left(G(x)^{2n+k(2m+1)+\nu j}\right)=0.$$

This equation can be further analyzed based on the specific characteristics of $g(x)$ and $G(x)$ to find the mode.

3.4 Moment generating function

The moment-generating function (MGF) of a random variable X is a function defined as the expected value of the exponential function of tX, where t is a real number. Mathematically, it is expressed as

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx,$$

for continuous random variables. The MGF, if it exists in a neighborhood around $t = 0$, can be used to derive all moments of the distribution, as the $r-th$ moment is given by the $r-th$ derivative of the MGF evaluated at $t = 0$, that is $\mu_n = M_X^{(n)}(0)$. Furthermore, the MGF uniquely determines the distribution of X when it exists. For the random variable X that assumes the WSG-G family, its MGF is given as

$$M(t) = \frac{\nu\pi}{4\rho^{\nu}}\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty}C_{n,k,m,j}\int_{-\infty}^{\infty}e^{tx}g(x)G(x)^{2n+k(2m+1)+\nu j}\,dx.$$

3.5 Maximum Likelihood Estimation in Generic Sense

Maximum Likelihood Estimation (MLE) is a method for estimating the parameters $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_k)$ of a probability distribution by maximizing the likelihood function. Given a sample $\{x_1, x_2, \ldots, x_n\}$, the likelihood function is defined as:

$$L(\Theta) = \prod_{i=1}^{n} f(x_i; \Theta),$$

where $f(x_i; \Theta)$ is the PDF or probability mass function (PMF) of the data, and Θ represents the parameters. The MLE for Θ is obtained by finding the parameter values that maximize the log-likelihood function:

$$\hat{\Theta} = \arg\max_{\Theta} \log L(\Theta) = \arg\max_{\Theta} \sum_{i=1}^{n} \log f(x_i; \Theta).$$

For a sample $\{x_1, x_2, \ldots, x_n\}$, the likelihood function is:

$$L(\nu, \rho) = \prod_{i=1}^{n} f(x_i; \nu, \rho, g, G),$$

where $f(x_i; \nu, \rho, g, G)$ is the probability density function involving the functions $g(x)$ and $G(x)$.

Taking the logarithm of the likelihood function:

$$\ell(\nu, \rho) = \sum_{i=1}^{n} \log \left(\frac{\nu \pi g(x_i)}{4\rho^{\nu}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \sum_{j=0}^{\infty} C_{n,k,m,j} G(x_i)^{2n+k(2m+1)+\nu j} \right).$$

To estimate ν, differentiate the log-likelihood function with respect to ν:

$$\frac{\partial \ell(\nu, \rho)}{\partial \nu} = \sum_{i=1}^{n} \left(\frac{1}{\nu} - \frac{1}{\rho} \log \left(g(x_i) \sum_{n,k,m,j} C_{n,k,m,j} G(x_i)^{2n+k(2m+1)+\nu j} \right) \right) = 0.$$

Similarly, differentiate the log-likelihood function with respect to ρ:

$$\frac{\partial \ell(\nu, \rho)}{\partial \rho} = \sum_{i=1}^{n} \left(-\frac{\nu}{\rho} + \frac{1}{\rho^2} \left(g(x_i) \sum_{n,k,m,j} C_{n,k,m,j} G(x_i)^{2n+k(2m+1)+\nu j} \right)^{\nu} \right) = 0.$$

The system of equations that provides the Maximum Likelihood Estimates (MLEs) for ν and ρ is as follows:

$$\sum_{i=1}^{n} \left(\frac{1}{\nu} - \frac{1}{\rho} \log \left(g(x_i) \sum_{n,k,m,j} C_{n,k,m,j} G(x_i)^{2n+k(2m+1)+\nu j} \right) \right) = 0,$$

$$\sum_{i=1}^{n}\left(-\frac{\nu}{\rho}+\frac{1}{\rho^2}\left(g(x_i)\sum_{n,k,m,j}C_{n,k,m,j}G(x_i)^{2n+k(2m+1)+\nu j}\right)^{\nu}\right)=0.$$

This system of equations generally needs to be solved numerically, as closed-form solutions for ν and ρ may not be available, especially due to the presence of complex sums and power series expansions.

4 A Special Case: The Weibull Sine Generalized-Exponential (WSG-E) Distribution

The exponential distribution is one of the standard distributions in the literature that is widely used especially in survival analysis and reliability engineering. The works of Crathorne (1925); Bru and Bru (2022) contain detail discussion and applications of the exponential distribution. The CDF and PDF are respectively denoted as $G(x)=1-e^{-\lambda x}$ and $g(x)=\lambda e^{-\lambda x}$. We utilize the distribution as the parent model to realize a more robust and versatile distribution. We substitute the CDF and PDF of the exponential distribution into equations (6) and (7) to obtain the CDF and PDF of the WSG-E distribution given respectively given in equations (11) and (12)

$$F(x)=1-e^{-\left\{\frac{\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x}\right)\right)}{\rho}\right\}^{\nu}};\quad x>0, \tag{11}$$

and

$$f(x)=\frac{\nu\pi\lambda e^{-\lambda x}}{4\rho^{\nu}}\cos\left(\frac{\pi}{4}\left(1-e^{-\lambda x}\right)\right)\left\{\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x}\right)\right)\right\}^{\nu-1}e^{-\left\{\frac{\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x}\right)\right)}{\rho}\right\}^{\nu}}, \tag{12}$$

where $\lambda>0$, $\nu>0$ and $\rho>0$. The new distribution has two scale parameters λ and ρ while its shape parameter is ν.

The hazard function is given in equation (13)

$$h(x)=\frac{\nu\pi\lambda e^{-\lambda x}}{4\rho^{\nu}}\cos\left(\frac{\pi}{4}\left(1-e^{-\lambda x}\right)\right)\left\{\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x}\right)\right)\right\}^{\nu-1}. \tag{13}$$

Figures (1) is the plots of the density function showing varying patterns namely reversed-bath-tub, positively skewed, leptokutic, platikurtic and mesokurtic shapes.

Figures (2), (3), (4), (5), (6) and (7) are different shapes of the hazard function which are strictly increasing, right-skewed, left-skewed and reversed L-shape, reversed bath-tub and strictly decreasing shapes respectively. The different shapes of the plots are virtual evidence that the proposed WSG-E distribution can be useful in modelling data of varying behaviours. Further justification will be provided with application to real-life datasets.

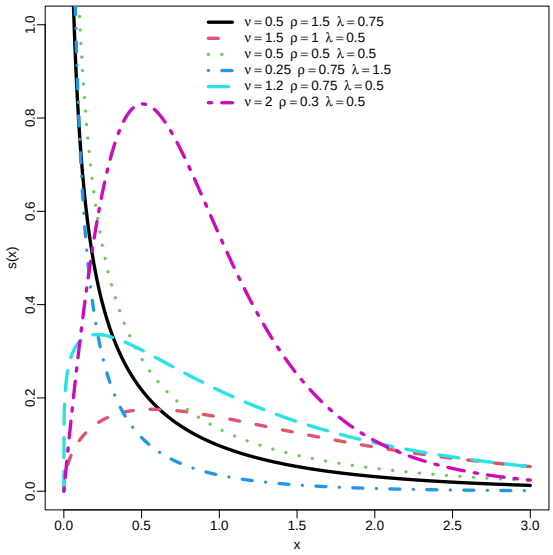

Figure 1: Plots of the PDF of WSG-E Distribution

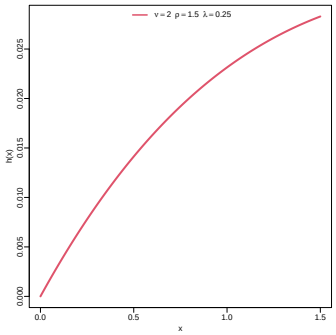

Figure 2: Hazard function

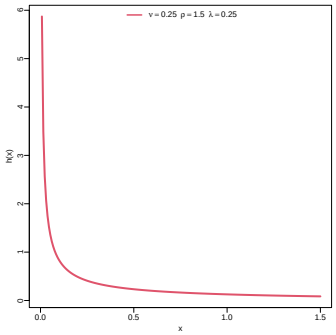

Figure 3: Hazard function

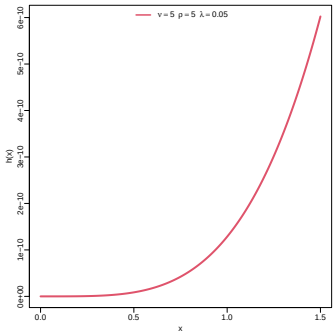

Figure 4: Hazard function

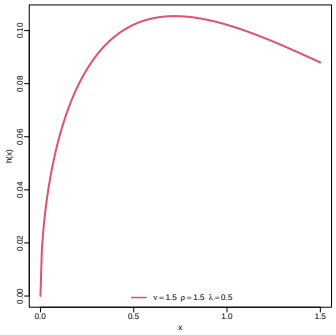

Figure 5: Hazard function

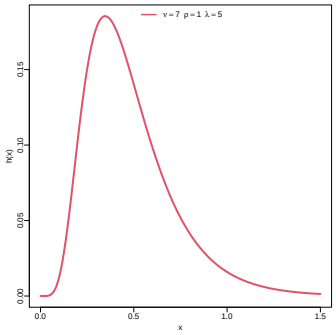

Figure 6: Hazard function

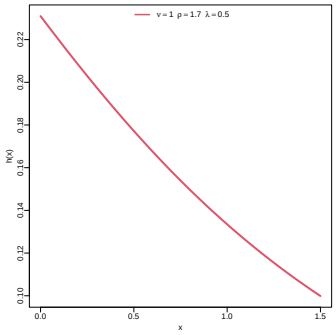

Figure 7: Hazard function

4.1 Behavior of WSG-E Distribution under Extreme Conditions

Here, we analyze the asymptotic behavior of the WSG-E distribution under extreme conditions as $x \to 0$ and $x \to \infty$.

As $x \to 0$, the term $e^{-\lambda x} \to 1$, leading to simplifications in the PDF. Using the Taylor series expansion of sine and cosine functions around zero, we have:

$$\cos\left(\frac{\pi}{4}(1-e^{-\lambda x})\right) \approx 1, \quad \sin\left(\frac{\pi}{4}(1-e^{-\lambda x})\right) \approx \frac{\pi}{4}(1-e^{-\lambda x}).$$

For small x, $1-e^{-\lambda x} \approx \lambda x$. Substituting these approximations into the PDF, we find:

$$f(x) \approx \frac{\nu\pi\lambda}{4\rho^{\nu}}\left(\frac{\pi}{4}\lambda x\right)^{\nu-1}.$$

Thus, the PDF behaves asymptotically as:

$$f(x) \propto x^{\nu-1}.$$

This leads to different behaviors depending on the value of ν:

- If $\nu > 1$, $f(x) \to 0$ as $x \to 0$.
- If $\nu = 1$, $f(x)$ approaches a constant as $x \to 0$.
- If $\nu < 1$, $f(x)$ diverges as $x \to 0$.

As $x \to \infty$, $e^{-\lambda x} \to 0$, and the sine and cosine terms approach constants:

$$\cos\left(\frac{\pi}{4}(1-e^{-\lambda x})\right) \to \frac{1}{\sqrt{2}}, \quad \sin\left(\frac{\pi}{4}(1-e^{-\lambda x})\right) \to \frac{1}{\sqrt{2}}.$$

For large x, the PDF is dominated by $e^{-\lambda x}$:

$$f(x) \propto e^{-\lambda x}.$$

As $x \to \infty$, the PDF decays exponentially with rate λ.

The WSG-E distribution exhibits power-law behavior as $x \to 0$, and exponential decay as $x \to \infty$. Specifically, near $x = 0$:

$$f(x) \propto x^{\nu-1},$$

and as $x \to \infty$:

$$f(x) \propto e^{-\lambda x}.$$

These asymptotic behaviors highlight the WSG-E distribution's capacity to model extreme values through both power-law and exponential tails.

4.2 Series Expansion of WSG-E Distribution

To express the PDF of the WSG-E distribution using power series and binomial expansions, we utilize well-known expansions for trigonometric, exponential, and binomial functions. This approach allows us to decompose the original function into summations, simplifying its analysis.

The Taylor series for the cosine function around $z = 0$ is:

$$\cos(z) = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!}.$$

Similarly, the sine function is expanded as:

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

For terms like $\sin(z)^{\nu-1}$, we apply the binomial expansion:

$$(1+z)^r = \sum_{k=0}^{\infty} \binom{r}{k} z^k,$$

valid for $|z| < 1$. The exponential function is expanded as:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Substituting these expansions into the original function allows us to express each term as an infinite sum over powers of z. This decomposes the function into summations involving powers of the cosine, sine, binomial, and exponential functions, simplifying its analysis. Based on this, the PDF of the WSG-E model in equation (12) can be compactly represented as:

$$f(x) = \frac{\nu\pi\lambda}{4\rho^{\nu}} \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty} C_{n,k,m,j} x^n, \tag{14}$$

where the coefficient $C_{n,k,m,j}$ is given by:

$$C_{n,k,m,j} = \frac{(-1)^{n+m+j}}{(2n)!(2m+1)!j!} \binom{\nu-1}{k} \left(\frac{\pi}{4}\right)^{2n+k(2m+1)} \frac{1}{\rho^{\nu j}}.$$

5 Properties of the WSG-E Distribution

In the instance of the sub-model WSG-E, some properties are studied in this subsection.

5.1 The rth Crude Moment

Let X be WSG-E(ν, ρ, λ) distributed, then the expression for the crude moment $E[X^r]$ is given as

$$E[X^r] = \frac{\nu\pi\lambda\Gamma(r+1)}{4\rho^{\nu}\lambda^{r+1}} \sum_{n=0}^{\infty}\sum_{k=0}^{\nu-1} \frac{(-1)^{n+k}}{(2n)!} \left(\frac{\pi}{4}\right)^{2n} \binom{\nu-1}{k} \frac{1}{(1+m)^{r+1}}; \quad r = 1, 2, \cdots \quad (15)$$

Substituting $r = 1, 2, 3$ and 4 provides the first, second, third and fourth moments necessary for computing summary statistics such as the mean, variance, standard deviation, skewness, kurtosis and coefficient of variation. The proof of equation (15) is trivial as sufficient information are provided in the generic moment contained in subsection 3.2.

5.2 The pth Quantile function for the WSG-E model

The expression for the quantile function $Q(p)$ is

$$Q(p) = -\frac{1}{\lambda}\log\left(1 - \frac{4}{\pi}\arcsin\left(\rho(-\log(1-p))^{1/\nu}\right)\right), \quad (16)$$

where $p \in (0, 1)$. Equation (16) is utilized in generating random samples that assume the WSG-E distribution in section 7.

5.3 Moment generating function of WSG-E Distribution

The MGF can also be used to identify the distribution of a random variable, as different distributions generally have distinct MGFs. It plays an important role in limit theorems and approximations in probability, such as in proving the Central Limit Theorem. Let $X \sim$ WSG-E (ν, ρ, λ), then the Moment generating function is given as

$$E[e^{tX}] = \frac{\nu\pi\lambda}{4\rho^{\nu}(\lambda(m+1)-t)} \sum_{n=0}^{\infty}\sum_{k=0}^{\nu-1} \frac{(-1)^{n+k}}{(2n)!} \left(\frac{\pi}{4}\right)^{2n} \binom{\nu-1}{k}. \quad (17)$$

The proof of equation (17) is straight-forward as provided in subsection 3.4.

Figures 8, 9, 10 and 11 represent 3-dimensional plots of the Mean, Variance, Skewness and Kurtosis of the WSG-E distribution. We present some useful statistics which includes the mean, variance, standard deviation, skewness, kurtosis and coefficient of variation across various sample sizes and choices of the parameter values in Table 1. Notice that the skewness for parameter value combinations are all positive and that the kurtosis greater than 3, showing leptokurtic or highly peaked distribution. This justifies the PDF plot with purple color in Figure 1. Again, the values of the coefficient of variation (CV) are each less than 1, and this implies that this distribution can be applied to datasets whose points are tightly clustered around the mean.

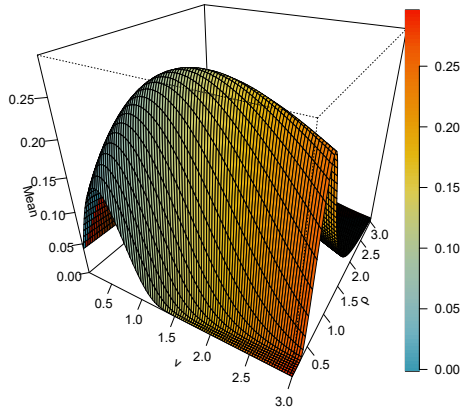

Figure 8: Mean of WSG-E(ν, ρ, λ)

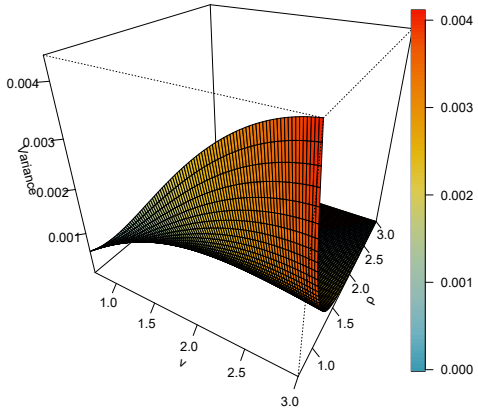

Figure 9: Variance of WSG-E(ν, ρ, λ)

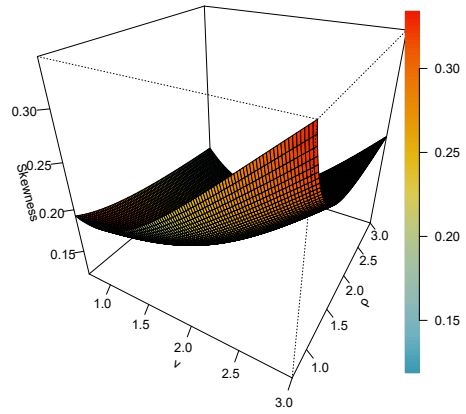

Figure 10: Skewness of WSG-E(ν, ρ, λ)

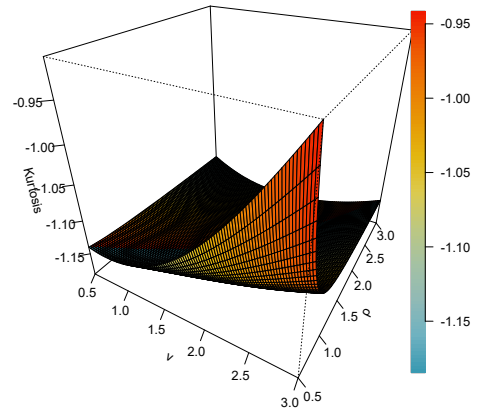

Figure 11: Kurtosis of WSG-E(ν, ρ, λ)

Table 1: Some useful Statistics from WSG-E Model using different parameter values

n	Mean	σ^2	σ	Skewness	Kurtosis	CV
Case I: $\nu = 1.5$, $\rho = 0.5$ and $\lambda = 0.75$						
25	1.835147	2.738789	1.654929	1.288238	4.464861	0.901796
50	1.449331	2.048560	1.431279	1.599601	5.706253	0.987545
100	1.300215	1.588066	1.260185	1.761409	6.842002	0.969213
150	1.342163	1.693192	1.301227	1.607486	5.685367	0.969500
Case II: $\nu = 1.5$, $\rho = 1.0$ and $\lambda = 1.75$						
25	0.786492	0.503043	0.709255	1.288238	4.464861	0.901796
50	0.621142	0.376266	0.613405	1.599601	5.706253	0.987545
100	0.557235	0.291686	0.540079	1.761409	6.842002	0.969213
150	0.575213	0.310995	0.557669	1.607486	5.685367	0.969500
Case III: $\nu = 2.0$, $\rho = 1.5$ and $\lambda = 0.5$						
25	2.752720	6.162275	2.482393	1.288238	4.464861	0.901796
50	2.173996	4.609260	2.146919	1.599601	5.706253	0.987545
100	1.950322	3.573148	1.890277	1.761409	6.842002	0.969213
150	2.013244	3.809683	1.951841	1.607486	5.685367	0.969500
Case IV: $\nu = 2.5$, $\rho = 2.0$ and $\lambda = 1.0$						
25	1.376360	1.540569	1.241197	1.288238	4.464861	0.901796
50	1.086998	1.152315	1.073459	1.599601	5.706253	0.987545
100	0.975161	0.893287	0.945139	1.761409	6.842002	0.969213
150	1.006622	0.952421	0.975920	1.607486	5.685367	0.969500
Case V: $\nu = 0.5$, $\rho = 2.0$ and $\lambda = 0.25$						
25	5.505441	24.649102	4.964786	1.288238	4.464861	0.901796
50	4.347992	18.437040	4.293837	1.599601	5.706253	0.987545
100	3.900644	14.292591	3.780554	1.761409	6.842002	0.969213
150	4.026488	15.238731	3.903682	1.607486	5.685367	0.969500

5.4 Distribution of the $\tau - th$ Order Statistics

In statistical analysis, the concept of order statistics plays a crucial role, particularly in understanding the distribution of sample data. The τ-th order statistic, denoted as $X_{(\tau)}$, represents the τ-th smallest value in a sample of size n drawn from a continuous distribution characterized by its cumulative distribution function (CDF) $F(x)$ and probability density function (PDF) $f(x)$.

The CDF of the τ-th order statistic $X_{(\tau)}$ provides the probability that the τ-th smallest value in the sample is less than or equal to a specified value x. This can be expressed mathematically as:

$$F_{X_{(k)}}(x) = P(X_{(\tau)} \leq x) = \sum_{j=\tau}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j} \tag{18}$$

The PDF of the τ-th order statistic, derived from its CDF, gives the likelihood of observing the τ-th smallest value at x. It is defined as:

$$f_{X_{(\tau)}}(x) = \binom{n}{\tau} [F(x)]^{\tau-1} [1 - F(x)]^{n-\tau} f(x). \tag{19}$$

So that

$$f_{X_{(\tau)}}(x) = \frac{\nu\pi\lambda}{4\rho^{\nu}} \binom{n}{\tau} \sum_{i=0}^{\infty}\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty}\sum_{f=0}^{\infty}\sum_{h=0}^{\infty}\sum_{l=0}^{\infty} \frac{(-1)^{i+f+h}}{f!h!} \binom{\tau-1}{i} \binom{(f+h)\nu}{l}$$
$$\times \left(\sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{\pi}{4}\left(1 - e^{-\lambda x}\right)\right)^{2n+1}}{(2n+1)!} \right)^{l} \frac{i^f (n-\tau)^h}{\rho^{(f+h)\nu}} C_{n,k,m,j} x^n,$$

where the coefficient $C_{n,k,m,j}$ is as defined earlier.

6 Inference on the Parameters

The likelihood function $L(\Theta)$ for a sample $x_1, x_2, \ldots, x_n$ from a distribution with PDF $f(x; \Theta)$ is expressed as:

$$L(\Theta) = \prod_{i=1}^{n} f(x_i; \Theta).$$

Here, Θ represents the parameters of the distribution. The maximum likelihood estimates are found by maximizing the likelihood function, which is equivalent to maximizing the log-likelihood function:

$$\ell(\Theta) = \log L(\Theta) = \sum_{i=1}^{n} \log f(x_i; \Theta)$$

The log-likelihood function when equation (8) is plugged in is

$$\ell(\nu,\lambda,\rho)=\sum_{i=1}^{n}\left(\log\left(\frac{\nu\pi\lambda}{4\rho^{\nu}}\right)-\lambda x_i+\log\left(\cos\left(\frac{\pi}{4}\left(1-e^{-\lambda x_i}\right)\right)\right)+\right.$$
$$\left.(\nu-1)\log\left(\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x_i}\right)\right)\right)-\left\{\frac{\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x_i}\right)\right)}{\rho}\right\}^{\nu}\right)$$

Differentiating ℓ with respect to ν:

$$\frac{\partial\ell}{\partial\nu}=\sum_{i=1}^{n}\left(\frac{1}{\nu}-\frac{\log(\rho)}{4}+\log\left(\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x_i}\right)\right)\right)-\right.$$
$$\left.\left\{\frac{\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x_i}\right)\right)}{\rho}\right\}^{\nu}\log\left(\frac{\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x_i}\right)\right)}{\rho}\right)\right)$$

Differentiating ℓ with respect to λ:

$$\frac{\partial\ell}{\partial\lambda}=-\sum_{i=1}^{n}\left(x_i-\frac{\pi}{4}x_i e^{-\lambda x_i}\left(\tan\left(\frac{\pi}{4}(1-e^{-\lambda x_i})\right)-(\nu-1)\cot\left(\frac{\pi}{4}(1-e^{-\lambda x_i})\right)\right)\right)$$

Differentiating ℓ with respect to ρ:

$$\frac{\partial\ell}{\partial\rho}=-\sum_{i=1}^{n}\left(-\frac{\nu}{\rho}+\left\{\frac{\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x_i}\right)\right)}{\rho^2}\right\}^{\nu}\log\left\{\frac{\sin\left(\frac{\pi}{4}\left(1-e^{-\lambda x_i}\right)\right)}{\rho}\right\}\right).$$

Setting the derivatives to Zero

$$\frac{\partial\ell}{\partial\nu}=0,\quad\frac{\partial\ell}{\partial\lambda}=0,\quad\frac{\partial\ell}{\partial\rho}=0.$$

By solving this system of equations, the MLEs for $\hat{\nu}$, $\hat{\lambda}$, and $\hat{\rho}$ are obtained.

7 Simulation Study

In this study, we conducted simulations to evaluate the performance of the parameters of the distribution defined by the CDF in equation (11). We generated simulated data using the quantile function associated with the CDF for various sample sizes. The specific sample sizes considered were $n=25,50,100,200,500,1000$, and for each sample size, we performed $10,000$ replicates. The algorithm for generating random samples is as follows;

1. Define the quantile function $Q(p)$ corresponding to the CDF $F(x)$ as in equation (16).

2. Generate uniform random variables $U_i \sim U(0,1)$ for $i = 1, 2, \ldots, n$.

3. Use the quantile function to obtain the random samples:

$$X_i = Q(U_i), \quad \text{for } i = 1, 2, \ldots, n.$$

In each replicate, we derived the parameters $\hat{\nu}$, $\hat{\rho}$, and $\hat{\lambda}$ from the generated samples. To assess the quality of these estimates, we calculated the Mean Squared Error (MSE):

$$\text{MSE}(\hat{\Theta}) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\Theta}_r - \Theta)^2,$$

where $\hat{\Theta}_r$ indicates the estimate for the r-th replicate, Θ is the actual parameter value, and R signifies the number of replicates. We also evaluated the Root Mean Squared Error (RMSE):

$$\text{RMSE}(\hat{\Theta}) = \sqrt{\text{MSE}(\hat{\Theta})}.$$

The findings include average parameter estimates for each sample size, as well as the MSE and RMSE for each parameter across $10,000$ replicates. This extensive simulation study enables a thorough examination of the robustness and precision of the parameter estimates across different sample sizes.

The simulation results presented in Table 2 indicate that as the sample size increases, both the mean squared error (MSE) and root mean squared error (RMSE) for the parameter ν decrease, demonstrating convergence. This trend suggests that the estimates for ν become more accurate and consistent with larger samples, which is a desirable outcome in statistical inference. The reduction in these error metrics confirms that the estimator for ν is reliable and consistent, making it suitable for inference as more data is collected. In contrast, the other two parameters, ρ and λ, do not exhibit the same pattern of convergence. Despite increasing sample sizes, the MSE and RMSE for these parameters remain higher, indicating difficulties in accurately estimating them. This discrepancy may be attributed to various factors, such as the sensitivity of the estimators, the model structure, parameter correlation, or numerical challenges in estimation. As a result, while inference on ν appears robust, caution should be exercised when interpreting the estimates for ρ and λ, as further investigation or advanced methods may be needed to improve their estimation.

Wherever the sample sizes are identical from the associated parameters, the RMSE values are also identical. This connotes that the errors which contribute to the computation of RMSE across all parameters are fairly uniform. Such behavior could be the outcome of an estimation process which follows a consistent trend with respect to bias and variance, hence squaring errors which do not show much variability between parameters. This is especially true in Maximum Likelihood Estimation (MLE) in which the accuracy of its parameter estimates tends to be somewhat close under the same conditions.

Table 2: Statistics from the Simulation Study for $\nu = \rho = 1$ and $\lambda = 1.5$

n	Parameter	Avg. Est	ν_{MSE}	ρ_{MSE}	λ_{MSE}	ν_{RMSE}	ρ_{RMSE}	λ_{RMSE}
	ν	1.442750	0.250123	0.903014	0.792012	0.500123	0.950270	0.889952
25	ρ	0.022943	0.260894	0.925013	0.804501	0.510718	0.961264	0.897942
	λ	0.103123	0.280421	0.970511	0.830456	0.529489	0.985151	0.911309
	ν	1.411493	0.190245	0.920146	0.815678	0.435960	0.959030	0.903146
50	ρ	0.020217	0.200138	0.940157	0.824561	0.447367	0.969683	0.908056
	λ	0.087200	0.218742	0.980294	0.842419	0.467665	0.990126	0.917796
	ν	1.394907	0.160130	0.930217	0.825314	0.400163	0.964470	0.908467
100	ρ	0.018746	0.170542	0.945236	0.834812	0.412877	0.972280	0.913626
	λ	0.080140	0.187692	0.985342	0.852108	0.433228	0.992648	0.922029
	ν	1.381132	0.140216	0.940315	0.835109	0.374405	0.969577	0.914904
200	ρ	0.018576	0.150528	0.950284	0.844310	0.388012	0.974812	0.918890
	λ	0.079520	0.166812	0.988342	0.862502	0.408434	0.994157	0.928640
	ν	1.377678	0.130121	0.950467	0.845003	0.360717	0.974796	0.919288
500	ρ	0.018362	0.140210	0.960402	0.854120	0.374447	0.980000	0.923124
	λ	0.078439	0.155671	0.990412	0.872500	0.394547	0.995199	0.933573
	ν	1.377311	0.128110	0.955602	0.847901	0.357984	0.977526	0.921844
1000	ρ	0.018344	0.138123	0.965710	0.856309	0.371553	0.982674	0.925477
	λ	0.078452	0.152600	0.992345	0.876123	0.390614	0.996166	0.936020

The RMSE values were derived from many replications, which means that the RMSE values for all parameters were estimated from the same underlying error distributions and hence converging towards the same values. This effect becomes more pronounced with a large number of replications used because the influence of random fluctuations fades away and continues to stabilize at certain RMSE estimates. When RMSE values remain indistinguishable in this sense, it is characteristic about the structure of the model.

8 Application to Medical Data

The first dataset comprises remission times (in months) for a random sample of 128 bladder cancer patients, previously utilized by Lee (2003) and investigated by Ekemezie et al. (2024). The data is presented in Table 3. The second dataset pertains to children diagnosed with growth hormone deficiency, sourced from the Programa Hormonal de Crescimento da Secretaria da Saúde de Minas Gerais. This dataset includes the estimated duration (in years) from the initiation of growth hormone treatment until the children reached their target height, as reported by Bakouch et al. (2021) and further examined by Aljohani (2024). The relevant information is detailed in Table 7. In both scenarios, we draw inferences by comparing the proposed Weibull Sine Generalized Ex-

ponential (WSG-E) distribution with the Type-I Heavy-Tailed Exponential (TIHTE) distribution studied by Nwankwo et al. (2024), the Weibull distribution by Weibull (1939), the Gamma distribution, the Log-Normal (Lnorm) distribution, the Exponentiated Weibull (EW) distribution by Pal et al. (2006), and the Exponentiated Inverted Exponential (EIE) distribution by Fatima and Ahmad (2017). We first diagnosed the two medical datasets to understand their peculiarities; further details can be found in Tables 4 and 8. Summary statistics reveal that Data-I is positively skewed with a high peak of 15.2. Additional diagnostics indicate that both datasets contain outliers, with Data-I being particularly vulnerable, as illustrated in Figures 20 and 21. This prompted the inclusion of Cramér-von Mises statistics (W) and Anderson-Darling statistics (A), given that Akaike Information Criterion (AIC), Consistent AIC (CAIC), Bayesian Information Criterion (BIC), and the Hannan-Quinn Information Criterion (HQIC) are sensitive to outliers. Model performance and fitness metrics for Data-I and Data-II are presented in Tables 5 and 9, respectively. For Data-I, the proposed distribution demonstrates exceptional fit, achieving a Kolmogorov-Smirnov (KS) P-value of 0.9827. In Data-II, the proposed model ranks as the first runner-up in adequacy, competing favorably with the other models. The Maximum Likelihood Estimates (MLEs) for the fitted distributions are displayed in Tables 6 and 10, with standard errors of the estimates provided in parentheses. It is evident that the WSG-E distribution outperforms the others, exhibiting the lowest standard errors for all parameters in both datasets. Graphical representations of the fit between the proposed distribution and the two medical datasets are shown in Figures 12 and 13. These figures depict the empirical density plots superimposed on histograms, the empirical cumulative distribution function (CDF) versus the CDF of the WSG-E model, the survival function, and the Time-To-Event (TTT) plots for both data instances. Figures 14, 15, and 16 illustrate the profile log-likelihood of ν_{MLE}, ρ_{MLE}, and λ_{MLE} for Data-I, while Figures 17, 18, and 19 correspond to Data-II. These sequences of figures correspond to the MLE values obtained in Tables 6 and 10. Based on the characteristics of the two datasets utilized to demonstrate the utility of the WSG-E distribution, we can preliminarily classify the WSG-E distribution as a heavy-tailed distribution.

(6) gives the maximum likelihood estimations (MLE) for different models fitted to Bladder Cancer Data. Values are given in parentheses for standard errors. Here, the parameters estimates from the WSG-E model are $\hat{\nu} = 1.3205$, $\hat{\rho} = 0.1783$, and $\hat{\lambda} = 0.0293$. The standard errors for $\hat{\nu}$ and $\hat{\rho}$ are rather small, supporting the stability of this estimate; however, the standard error of $\hat{\lambda}$ is large, perhaps suggesting some uncertainness in the estimate.

For the TIHTE, the fitted parameter $\hat{\nu} = 0.0914$ is truly small, and very large standard error $\hat{\rho} = 1.4533$ is representing it at 1.0092. This implies that the model is probably not fitting the underlying data very well since large standard error suggests an instability in estimating parameters. Gamma provided the estimates $\hat{\nu} = 1.1360$ and $\hat{\rho} = 7.7597$, the latter with a very large standard error (2.3454) indicating variability in the fit. Similarly, for the Lognormal model, the estimates $\hat{\nu} = 1.7777$ and $\hat{\rho} = 1.0664$, where again both estimates have high standard errors, pointing at problems in estimating these parameters.

Table 3: The remission times (in months) of a sample of 128 bladder cancer patients (Data-I)

3.88	5.32	7.39	10.34	14.83	34.26	0.90	2.69	4.18	5.34	7.59	10.66
15.96	36.66	1.05	2.69	4.23	5.41	7.62	10.75	16.62	43.01	1.19	2.75
4.26	5.41	7.63	17.12	46.12	1.26	2.83	4.33	5.49	7.66	11.25	17.14
79.05	1.35	2.87	5.62	7.87	11.64	17.36	1.40	3.02	4.34	5.71	7.93
0.08	2.09	3.48	4.87	6.94	8.66	13.11	23.63	0.20	2.23	3.5	4.98
6.97	9.02	13.29	0.40	2.26	3.57	5.06	7.09	9.22	13.80	25.74	0.50
2.46	3.64	5.09	7.26	9.47	14.24	25.82	0.51	2.54	3.70	5.17	7.28
9.74	14.76	26.31	0.81	2.62	3.82	5.32	7.32	10.06	14.77	32.15	2.64
11.79	18.10	1.46	4.40	5.85	8.26	11.98	19.13	1.76	3.25	4.50	6.25
8.37	12.02	2.02	3.31	4.51	6.54	8.53	12.03	20.28	2.02	3.36	6.76
12.07	21.73	2.00	3.36	6.93	8.65	12.63	22.69				

Table 4: Summary Statistics from Data-I

n	$\bar{X}$	σ	Median	Trimmed $\bar{X}$	MAD	Min	Max	Range	Sk	Ku	S_e
128	9.37	10.51	6.39	7.42	5.46	0.08	79.05	78.97	3.25	15.2	0.93

Table 5: Model Comparison and Fitness to Bladder Cancer Patients Data

Distr.	NLL	AIC	CAIC	BIC	HQIC	W	A	KS	p-value	Rank
WSG-E	410.40	826.792	826.986	835.349	830.269	0.032	0.199	0.041	0.9827	1
TIHTE	415.40	834.884	834.980	840.588	837.202	0.039	0.249	0.069	0.5784	5
Gamma	413.37	831.360	831.456	837.064	833.678	0.123	0.735	0.068	0.5895	4
Lnorm	415.09	834.265	834.361	839.969	836.583	0.122	0.826	0.053	0.8682	3
EW	410.68	827.360	827.554	835.916	830.837	0.044	0.289	0.045	0.9574	2
EIE	413.65	831.300	831.396	837.004	833.617	0.118	0.725	0.078	0.4247	6
Weibull	414.09	832.366	832.462	838.070	834.683	0.129	0.773	0.084	0.3265	7

The Exponentiated Weibull (EW) model provided the estimates of $\hat{\nu} = 2.7939$, $\hat{\rho} = 0.6546$, and $\hat{\lambda} = 3.3486$. In this model, $\hat{\nu}$ and $\hat{\rho}$ have moderate standard errors, while $\hat{\lambda}$ shows large standard error 1.8873, indicating some uncertainty in the estimate. The Exponentiated Inverse Exponential (EIE) model has rather extreme low estimates $\hat{\nu} = 0.0228$ and $\hat{\rho} = 0.2110$, with correspondingly small standard errors indicating stable estimates, but possibly the model may be fitting extreme observations in the data. The Weibull model estimates showed $\hat{\nu} = 1.0468$ and $\hat{\rho} = 9.9190$, with moderate standard errors, implying a reasonable fit. However, the large $\hat{\rho}$ value suggests a very spread-out distribution.

Table 6: MLEs for the models fitted on Bladder Cancer Data

Distr.	$\hat{\nu}_{\text{MLE}}$	$\hat{\rho}_{\text{MLE}}$	$\hat{\lambda}_{\text{MLE}}$
WSG-E	1.3205 (0.1383)	0.1783 (0.0558)	0.0293 (0.1111)
TIHTE	0.0914 (0.0708)	1.4533 (1.0092)	
Gamma	1.1360 (0.7003)	7.7597 (2.3454)	
Lnorm	1.7777 (0.8711)	1.0664 (0.6782)	
EW	2.7939 (1.2600)	0.6546 (0.1345)	3.3486 (1.8873)
EIE	0.0228 (0.0175)	0.2110 (0.1595)	
Weibull	1.0468 (0.5602)	9.9190 (0.7262)	

The comparison among the different models shows that the WSG-E model has the advantage of stable parameter estimates, with small standard errors compared to many other models. The standard errors from the Gamma and Lognormal models were much larger, showing possible difficulties in accurately estimating the parameters. Large standard error values for $\hat{\rho}$ suggest that the TIHTE might not really fit the data well. Well, EW gives some flexibility with their higher variances on some parameters which raises doubts about their robustness. Fairly, the Weibull and EIE would still be more stable, and the much larger uncertainty may compromise the selection of Weibull. In all, from the parameter stability and standard errors' perspective, the WSG-E is the more suitable one in the modeling of Bladder Cancer Data but more verification measures through goodness-of-fit would be necessary to claim its strengths.

Table 7: Data on children diagnosed with growth hormone deficiency (Data-II)

2.15	2.20	2.55	2.56	2.63	2.74	2.81	2.90	3.05	3.41	3.43	3.43
3.84	4.16	4.18	4.36	4.42	4.51	4.60	4.61	4.75	5.03	5.10	5.44
5.90	5.96	6.77	7.82	8.00	8.16	8.21	8.72	10.40	13.20	13.70	

Table 8: Summary Statistics for Data-II

n	$\bar{X}$	σ	Median	Trimmed $\bar{X}$	MAD	Min	Max	Range	Sk	Ku	S_e
35	5.31	2.91	4.51	4.88	2.16	2.15	13.7	11.55	1.31	1.15	0.49

(10) summarizes the Maximum Likelihood Estimates (MLEs) for various models fitted to growth hormone deficiency data, including standard errors in parentheses. For the WSG-E model, the fitted parameters are $\hat{\nu} = 4.5587$, $\hat{\rho} = 0.5042$, and $\hat{\lambda} = 0.2048$. Parameters $\hat{\rho}$ and $\hat{\lambda}$ appear to be estimated with quite small standard errors, whereas $\hat{\nu}$ has a rather large estimated standard error of 1.6782, therefore carrying some uncer-

Table 9: Model Comparison and Fitness for the data on growth hormone deficiency

Distr.	NLL	AIC	CAIC	BIC	HQIC	W	A	KS	p-value	Rank
WSG-E	79.71	160.412	160.186	162.078	160.022	0.012	0.138	0.093	0.9233	1
TIHTE	85.39	174.777	175.152	177.888	175.851	0.071	0.482	0.200	0.1225	7
Gamma	80.11	164.236	164.611	167.347	165.310	0.101	0.655	0.132	0.5732	5
Lnorm	78.49	161.004	161.379	164.114	162.077	0.055	0.382	0.101	0.8700	4
EW	77.73	161.201	161.976	165.867	162.812	0.034	0.249	0.088	0.9473	2
EIE	78.39	160.778	161.153	163.889	161.852	0.057	0.393	0.099	0.8826	3
Weibull	82.49	169.066	169.441	172.177	170.140	0.168	1.052	0.160	0.3282	6

Table 10: MLEs for the models fitted to the data on growth hormone deficiency

Distr.	$\hat{\nu}_{\text{MLE}}$	$\hat{\rho}_{\text{MLE}}$	$\hat{\lambda}_{\text{MLE}}$
WSG-E	4.5587 (1.6782)	0.5042 (0.0983)	0.2048 (0.0874)
TIHTE	0.1622 (0.0690)	1.3740 (0.5062)	
Gamma	4.2438 (1.0403)	1.2625 (1.1054)	
Lnorm	1.5590 (0.9111)	0.4854 (0.0695)	
EW	975.034 (765.2402)	0.3394 (0.0317)	0.0130 (0.0091)
EIE	4.2593 (1.6912)	9.3175 (1.8039)	
Weibull	2.0636 (1.1701)	6.1198 (2.3255)	

tainty. Be that as it may, the WSG-E model appears to balance well across the various parameter estimates.

In the TIHTE model, $\hat{\nu} = 0.1622$ and $\hat{\rho} = 1.3740$, whose estimation uncertainty is fairly moderate with a standard error of 0.5062. The low value of $\hat{\nu}$ indicates that this model might not be able to capture the variability in the data as well as others. For the Gamma model, $\hat{\nu} = 4.2438$ and the estimate of $\hat{\rho} = 1.2625$ has a high standard error of 1.1054, implying estimation instability. The Lognormal (Lnorm) model estimates $\hat{\nu} = 1.5590$ and $\hat{\rho} = 0.4854$, with reasonable stability indicated by the standard errors.

Exceedingly, the Exponentiated Weibull (EW) model obtains a very large value for $\hat{\nu} = 975.034$ and a massive standard error of 765.2402, suggesting its estimates are questionable due to either overfitted or numerically unstable conditions. On the other hand, $\hat{\rho} = 0.3394$ and $\hat{\lambda} = 0.0130$ show reasonably small standard errors, indicating that these parameters are well estimated, but practically, the huge estimated value of $\hat{\nu}$ will cast shadows on its applicability.

In the Exponentiated Inverse Exponential (EIE) model, $\hat{\nu} = 4.2593$ and $\hat{\rho} = 9.3175$, high uncertainty attended $\hat{\rho}$ with a standard error of 1.8039. So, while the shape parameter is probably relatively stable, much uncertainty attaches to the scale parameter. In contrast, the Weibull model yields $\hat{\nu} = 2.0636$ and $\hat{\rho} = 6.1198$, with $\hat{\rho}$ suffering

from a large standard error of 2.3255, raising concerns regarding robustness of estimate realization.

For the overall summary, WSG-E maintains a very good balance of stability with reasonable standard errors, hence a solid contender for modelling the growth hormone deficiency data. The parameter estimates for the Gamma and Lognormal models are thus also in some way deemed reasonably stable, though uncertainties in the estimates arise due to the greater standard error of $\hat{\rho}$ in the Gamma model. On account of the extreme value of $\hat{\nu}$, we deem EW quite unreliable. The EIE and Weibull models pose some questions regarding parameter stability. Because of these observations, it is important that different tests of model fit and tests of predictive performance be carried out to discover which model can be considered a good fit for the data at hand.

9 Final Remarks

In this study, we developed a trigonometric class of distribution known as the Weibull Sine Generalized family of distributions, and we examined its generic properties. A specific instance, termed the Weibull Sine Generalized Exponential distribution, was identified and analyzed. This new distribution extends the baseline exponential distribution, featuring a heavier tail and the capability to model data with outliers. We derived preliminary results for key statistics, including the mean, variance, standard deviation, skewness, kurtosis, and coefficient of variation, which indicate that the distribution can effectively model positively skewed data characterized by high peakness and clustering around the mean. A simulation study was conducted to evaluate the performance of Maximum Likelihood Estimation (MLE). The results showed that as the sample size increased, both the mean squared error and root mean squared error for the parameter ν decreased, indicating convergence—an encouraging finding for inference. However, this trend was not observed for the other two parameters. To illustrate the applicability of the proposed model, we utilized two medical datasets: the first comprises remission times of 128 randomly selected bladder cancer patients, and the second pertains to children diagnosed with growth hormone deficiency. Both datasets exhibit a significant presence of extreme values (outliers), particularly in the first dataset, as well as positive skewness. To maintain data integrity, we did not alter the datasets but instead fitted the new distribution to the original data. The metrics employed demonstrate that the Weibull Sine Generalized Exponential distribution adequately fits the datasets, making it a recommended choice for modeling skewed data. Future research could extend any continuous distribution as a parent model using the Weibull Sine Generalized family of distributions proposed in this study.

The WSG-family of distributions offers substantial research potential beyond extending parent distributions. Its flexibility makes it suitable for survival analysis, reliability studies, and failure time modeling. The improved goodness-of-fit suggests applications in actuarial science, financial modeling, and risk assessment. In biomedical research, it can be used for analyzing survival rates and disease progression. Further studies could explore its integration into Bayesian inference, development of regression models, and

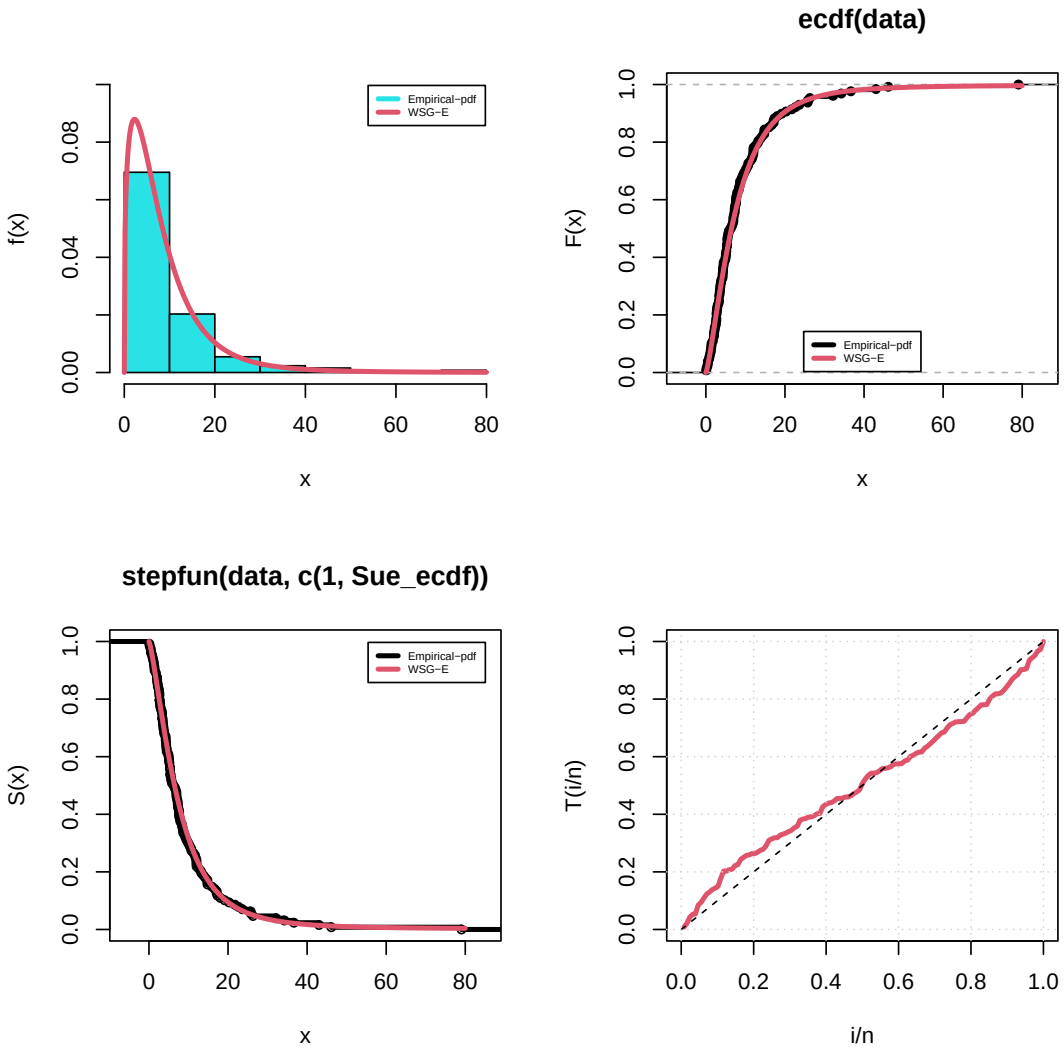

Figure 12: Density plot superimposed on histogram, CDF, Survival and TTT plots of the bladder cancer data

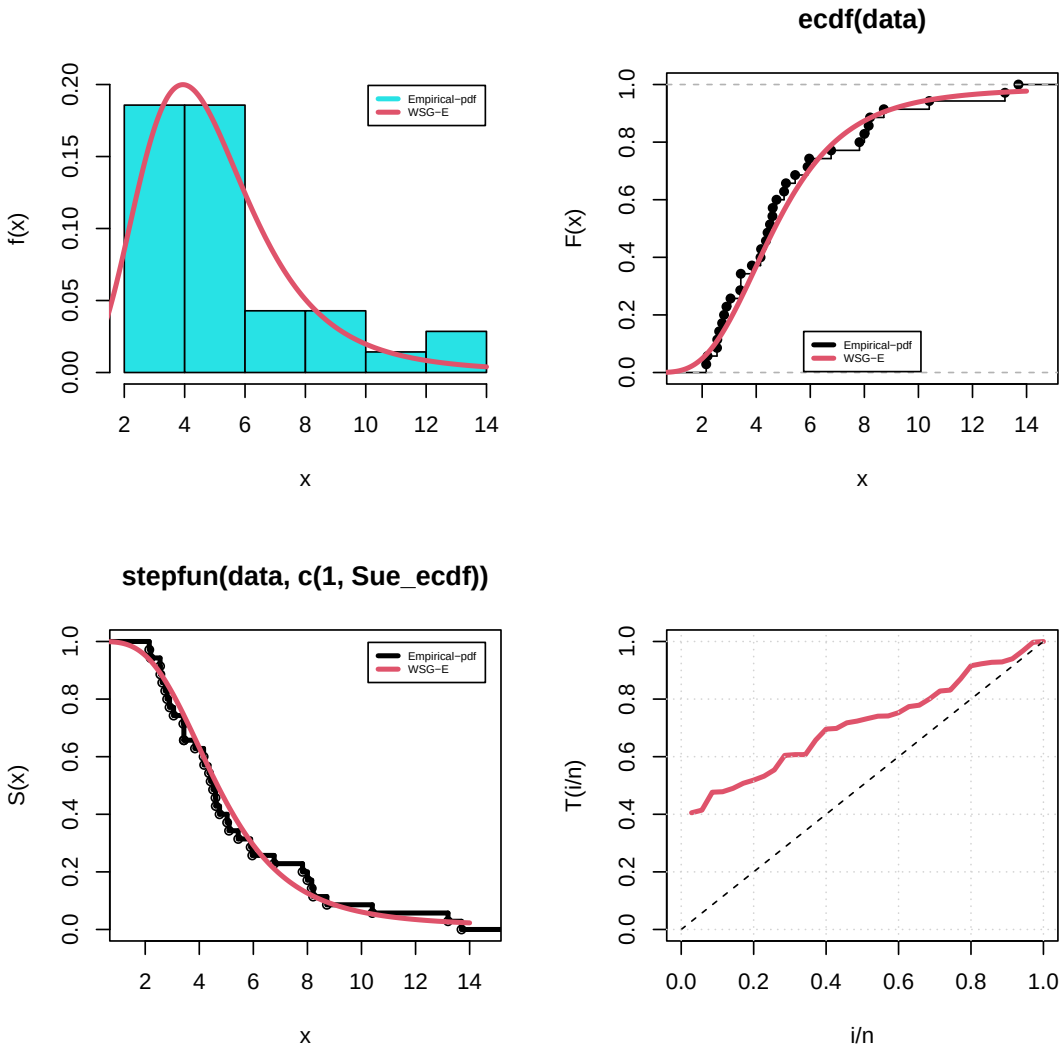

Figure 13: Density plot superimposed on histogram, CDF, Survival and TTT plots of growth hormone deficiency data

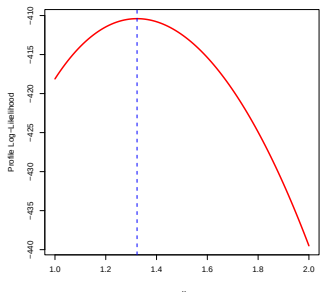

Figure 14: Profile log-like of ν for Data-I

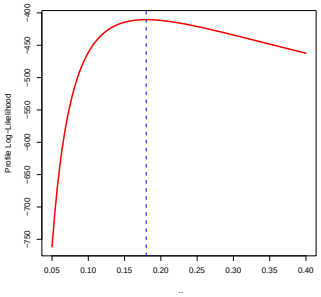

Figure 15: Profile log-like of ρ for Data-I

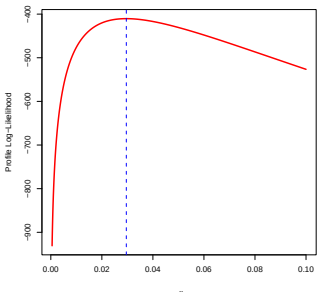

Figure 16: Profile log-like of λ for Data-I

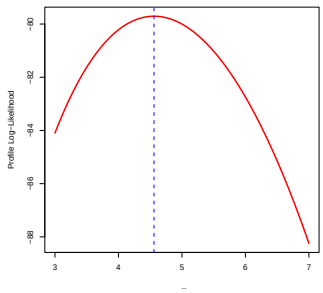

Figure 17: Profile log-like of ν for Data-II

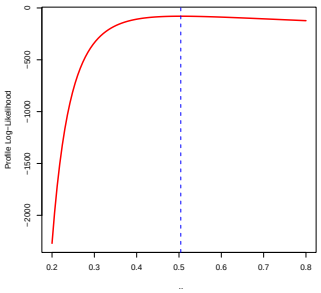

Figure 18: Profile log-like of ρ for Data-II

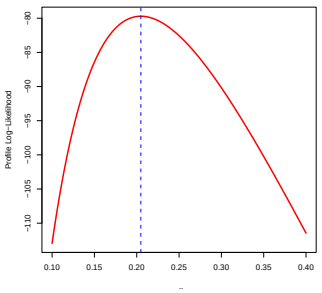

Figure 19: Profile log-like of λ for Data-II

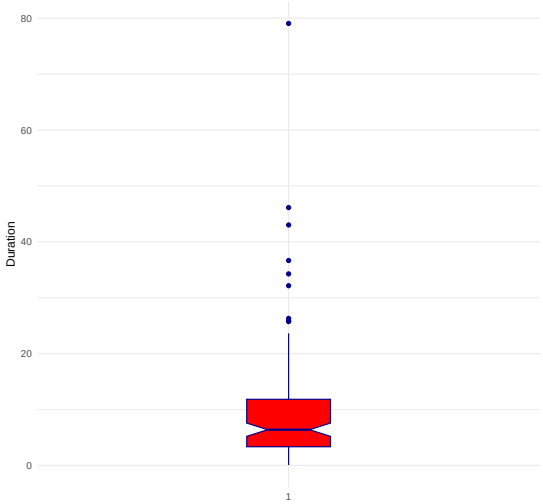

Figure 20: Boxplot for Data-I

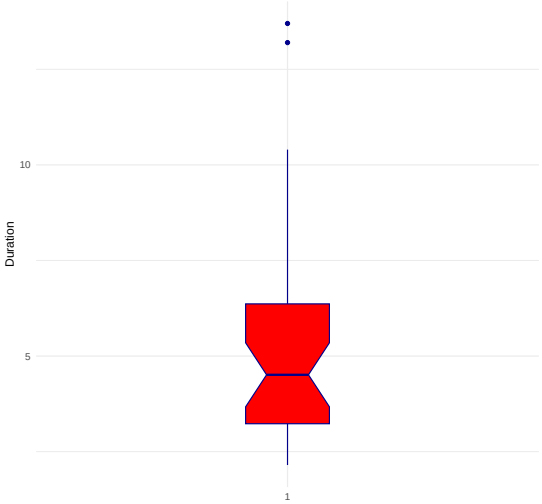

Figure 21: Boxplot for Data-II

computational efficiency in parameter estimation. The theoretical properties, including moments and entropy, also warrant deeper investigation, making the WSG-family a valuable tool for both theoretical and applied research.

Data and Code Availability Statement

Data included in article/supplementary material is referenced in the article.

Competing Interest

The authors declare that there is no competing interest.

Authors Contribution

Conceptualization: O.J.O.; **Methodology:** O.J.O; **Software:** O.J.O.; **Validation:** O.J.O., H.O.O.-I., G.A.O., M.K., O.S.B.; **Formal Analysis:** O.J.O.; **Investigation:** O.J.O., H.O.O.-I., G.A.O., M.K., O.S.B.; **Resources:** O.J.O., H.O.O.-I., G.A.O., M.K., O.S.B.; **Data Curation:** O.J.O., H.O.O.-I., G.A.O., M.K., O.S.B.; **Writing – Original Draft Preparation:** O.J.O.; **Writing – Review & Editing:** O.J.O., H.O.O.-I., G.A.O., M.K., O.S.B. ; **Visualization:** O.J.O.; **Supervision:** H.O.O.-I., G.A.O., M.K., O.S.B.; **Project Administration:** H.O.O.-I., G.A.O., M.K., O.S.B. **Funding:** M.K., O.S.B.

References

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(1):574–593.

Algarni, A. (2020). Sine power lomax model with application to bladder cancer data. *Nanoscience and NanoTechnology Letters*, 12(5):677–684.

Aljohani, H. M. (2024). Advances in medical data modeling: A new logarithmic beta generated family of distributions with theory and inference. *Alexandria Engineering Journal*, 102:339–360.

Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71(1):63–79.

Alzahrani, A. B., Abdoon, M. A., Elbadri, M., Berir, M., and Elgezouli, D. E. (2023). A comparative numerical study of the symmetry chaotic jerk system with a hyperbolic sine function via two different methods. *Symmetry*, 15(11):1991.

Bakouch, H., Chesneau, C., and Enany, M. (2021). A weighted general family of distributions: Theory and practice. *Computational and Mathematical Methods*, 3(6):e1135.

Besicovitch, A. S. and Besicovitch, A. S. (1954). *Almost periodic functions*, volume 4. Dover New York.

Bracewell, R. N., Cherry, C., Gibbons, J. F., Harman, W. W., Heffner, H., Herold, E. W., Linvill, J. G., Ramo, S., and Rohrer, R. A. (2000). Mcgraw-hill seires in electrical and computer engineering.

Bru, M.-F. and Bru, B. (2022). Émile borel's denumerable martingales, 1909–1949. In *The Splendors and Miseries of Martingales: Their History from the Casino to Mathematics*, pages 51–65. Springer.

Chakraborty, S., Hazarika, P. J., and Ali, M. M. (2012). Pak. j. statist. 2012 vol. 28 (4), 513-524 a new skew logistic distribution and its properties. *Pak. J. Statist*, 28(4):513–524.

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(8):2287–2300.

Chesneau, C. and Jamal, F. (2020). The sine kumaraswamy-g family of distributions. *Journal of Mathematical Extension*, 15.

Crathorne, A. (1925). E. borel, traité du calcul des probabilités et de ses applications.

Ekemezie, D.-F. N., Alghamdi, F. M., Aljohani, H. M., Riad, F. H., Abd El-Raouf, M., and Obulezi, O. J. (2024). A more flexible lomax distribution: Characterization, estimation, group acceptance sampling plan and applications. *Alexandria Engineering Journal*, 109:520–531.

Eugene, T. and Manfred, S.-S. (2019). Introduction to signal processing: sine wave and complex signals. *International Journal of Open Information Technologies*, 7(3):17–24.

Fatima, K. and Ahmad, S. (2017). The exponentiated inverted exponential distribution. *Journal of Applied Information Science*, 5(1).

Griffiths, D. J. and Schroeter, D. F. (2018). *Introduction to quantum mechanics*. Cambridge university press.

Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

Hogg, R. V., McKean, J. W., Craig, A. T., et al. (2013). *Introduction to mathematical statistics*. Pearson Education India.

Kharazmi, O., Alizadeh, M., Contreras-Reyes, J. E., and Haghbin, H. (2022). Arctan-based family of distributions: Properties, survival regression, bayesian analysis and applications. *Axioms*, 11(8):399.

Kharazmi, O., Saadatinik, A., and Tamandi, M. (2017). Hyperbolic sine-weibull distribution and its applications. *Int J Math Comput*, 28(3).

Kreyszig, E. (2021). Advanced engineering mathematics.

Lee, E. (2003). *Statistical methods for survival data analysis*. John Wiley & Sons.

Li, Y., Sawada, T., Shi, Y., Steinman, R. M., and Pizlo, Z. (2013). Symmetry is the sine qua non of shape. *Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective*, pages 21–40.

Mahmood, Z., Chesneau, C., and Tahir, M. H. (2019). A new sine-g family of distributions: properties and applications. *Bull. Comput. Appl. Math.*, 7(1):53–81.

Muse, A. H., Tolba, A. H., Fayad, E., Abu Ali, O. A., Nagy, M., and Yusuf, M. (2021).

Modelling the covid-19 mortality rate with a new versatile modification of the log-logistic distribution. *Computational Intelligence and Neuroscience*, 2021(1):8640794.

Nadarajah, S. and Kotz, S. (2006). Beta trigonometric distributions. *Portuguese Economic Journal*, 5:207–224.

Nagarjuna, V. B., Vardhan, R. V., and Chesneau, C. (2021). On the accuracy of the sine power lomax model for data fitting. *Modelling*, 2(1):78–104.

Nise, N. S. (2020). *Control systems engineering*. John Wiley & Sons.

Nwankwo, B. C., Obiora-Ilouno, H. O., Almulhim, F. A., SidAhmed Mustafa, M., and Obulezi, O. J. (2024). Group acceptance sampling plans for type-i heavy-tailed exponential distribution based on truncated life tests. *AIP Advances*, 14(3).

Oramulu, D. O., Alsadat, N., Kumar, A., Bahloul, M. M., and Obulezi, O. J. (2024). Sine generalized family of distributions: Properties, estimation, simulations and applications. *Alexandria Engineering Journal*, 109:532–552.

Pal, M., Ali, M. M., and Woo, J. (2006). Exponentiated weibull distribution. *Statistica*, 66(2):139–147.

Parumasur, N. and Mika, J. R. (2005). Amplitude-shape method for solving partial differential equations of chemical kinetics. *Applied numerical mathematics*, 55(4):473–479.

Raab, D. H. and Green, E. H. (1961). A cosine approximation to the normal distribution. *Psychometrika*, 26(4):447–450.

Shrahili, M., Elbatal, I., Almutiry, W., and Elgarhy, M. (2021a). Estimation of sine inverse exponential model under censored schemes. *Journal of Mathematics*, 2021(1):7330385.

Shrahili, M., Elbatal, I., and Elgarhy, M. (2021b). Sine half-logistic inverse rayleigh distribution: Properties, estimation, and applications in biomedical data. *Journal of Mathematics*, 2021(1):4220479.

Stewart, J. (2011). *Calculus, Early Transcendentals, AP ed.* Cengage Learning.

Suslov, S. K. (2003). Asymptotics of zeros of basic sine and cosine functions. *Journal of Approximation Theory*, 121(2):292–335.

Weibull, W. (1939). A statistical theory of strength of materials. *IVB-Handl.*

William, H., Kemmerly, J. E., and Durbin, S. M. (2007). *Engineering circuit analysis*. McGraw-Hill Higher Education.

Zaidi, S. M., Mahmood, Z., Atchadé, M. N., Tashkandy, Y. A., Bakr, M., Almetwally, E. M., Hussam, E., Gemeay, A. M., and Kumar, A. (2024). Lomax tangent generalized family of distributions: characteristics, simulations, and applications on hydrological-strength data. *Heliyon*.