# SINE-WEIBULL DISTRIBUTION: MATHEMATICAL PROPERTIES AND APPLICATION TO REAL DATASETS

[1]Muhammad Umar Faruk, [2]Alhaji Modu Isa, [3]Aishatu Kaigama

•

[1,2,3]Department of Mathematics and Computer Science, Borno State University, Nigeria
alhajimoduisa@gmail.com

**Abstract**

*New parameters can be added to expand families of distribution for greater flexibility or to construct covariate models in several ways. In this study, a trigonometric-type distribution called Sine-Weibull distribution was developed by adopting the Weibull distribution as the baseline distribution and Sine-G Family as the generator to generate a flexible probability distribution without the need for extra parameters. The moment, moment generating function, entropy, and order statistics are some of the mathematical aspects of this distribution that were derived. The Maximum Likelihood approach was used to estimate the new distribution's parameters. Using actual datasets, the Sine-Weibull distribution's applicability was demonstrated.*

**Keywords:** Sine-G Family, Weibull Distribution, Probability Distribution, Maximum Likelihood Estimator

## I. Introduction

Distribution functions, their properties and interrelationships play a significant role in modeling naturally occurring phenomena. For this reason, a large number of distribution functions, which were found applicable to many events in real life, have been proposed and defined in literature. Various methods exist in defining statistical distributions. Many of these arose from the need to model naturally occurring events. For example, the Normal distribution addresses real-valued variables that tend to cluster at a single mean value, while the Poisson distribution models discrete rare events. Yet few other distributions are functions of one or more distributions.

To explain real world phenomena, statistical distributions are widely applied. Their theory is widely studied due to the utility of statistical distributions, and new distributions are developed. In the field of probability theory and statistics, the search for creating a more effective and scalable distribution of probability remains high [1]. Numerous standard distributions have been extensively used over the past decades for modeling data in several fields such as Engineering, Economics, Finance, Biological, Environmental and Medical Sciences etc. However, generalizing these standard distributions has produced several compound distributions that are more flexible compared to the baseline distributions. For this reason, several methods for generating new families of distributions have been studied.

Weibull distribution is a continuous probability distribution. It is one of different distributions used to describe particle size with major application in survival analysis, weather forecast and reliability engineering. The Weibull distribution is a continuous probability distribution. It was named after Swedish mathematician Waloddi Weibull, who describe it in detail in 1951, although it was first recognized by [2] and first applied by [3] to describe a unit size of distribution. Weibull distribution exist with scale and shape parameters. This distribution has become very popular in analyzing lifetime data and for many applications where a skewed distribution is required. Inducing of a new shape parameter(s) introduces a model into greater family of distributions and can give significantly skewed and heavy-tailed distributions and also provides greater flexibility in the form of new distribution.

Even when there is uncertainty about the future in real life, decisions still need to be taken. Thus, uncertainty issues must be dealt with by decision-making processes. Probability is one of the frequently employed strategies for addressing uncertainty in planning and management. In order to create a family of hybrid distributions that are more effective than their parent distributions, many researchers have focused on the idea of combining two or more probability distributions. By adding one or more parameters, these distributions become more flexible and can track a variety of random phenomena that are difficult to model using their parent distributions. The laws of generality, which state that when a particular distribution has more than four parameters, it undermines the performance of the model, can sometimes be breached by such compounding or extended distributions.

Many researchers have come up with new families of trigonometric in recent times. Some of these families include: exponentiated sine-generated family of distributions by [4], Sin-G class of distributions by [5], Sec-G Class by [7], Sine Square distribution by [8], Sine Inverse Lomax Generated Family by [9], Sine Burr XII by [10], Sine Kumaraswamy-G family of distributions by [11], Sine Topp-Leone family by [12], Sine-Exponential Distribution by [13] and Sine Power Lomax distribution by [14] (2021).

The quest for developing more efficient and flexible probability distribution remains strong in the field of probability theory and statistics. However, there is no single probability distribution that is suitable for different data sets. Therefore, there is a need to come up with their extended forms to give substitutive adaptable models or as to form a better representation of the data. Thus, this has triggered the need to extend the existing classical Weibull distributions. Therefore, this gives a gap of coming up with a distribution (Sine-Weibull Distribution) capable of handling a dataset that behaved negatively or positively skewed. Hence, this research is aimed at developing a new probability distribution function called Sine-Weibull Distribution.

# II. Methods

## 2.1 The Weibull Distribution

A continuous random variable $X$ is said to have followed a Weibull distribution if its cdf is expressed as;

$$H(x,k,\lambda) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}, \quad x > 0 \tag{1}$$

and the pdf is also expressed as;

$$h\,(x,k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{(k-1)} e^{-\left(\frac{k}{\lambda}\right)^k} \quad x > 0 \tag{2}$$

## 2.2 Sine G Family of Probability Distribution

Let $H(x)$ be the cumulative distribution function (cdf) of a univariate continuous distribution and $h(x)$ be the corresponding probability density function (pdf), then, the Sine-G family of probability distribution according to [5] Kumar *et al.,* (2015) is given by:

$$F(x,\xi) = \int_0^{\frac{\pi}{2}H(x,\xi)} \cos t \, dt = \sin\left\{\frac{\pi}{2} H(x,\xi)\right\} \tag{3}$$

and its corresponding pdf is given by:

$$f(x,\xi) = \frac{\pi}{2} h(x,\xi) \, \cos\left\{\frac{\pi}{2} H(x,\xi)\right\} \tag{4}$$

where $H(x,\xi)$ and $h(x,\xi)$ are the cdf and the pdf of any baseline distribution with vector parameter $\xi$.

## 2.3 The New Sine Weibull Distribution

The pdf and cdf of the new sine Weibull distribution are given in equation (5) and (6):

$$f(x,k,\lambda) = \frac{\pi}{2}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \cos\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\} \tag{5}$$

And

$$F(x,k,\lambda) = \sin\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\} \tag{6}$$

The survival function $S(x)$, hazard function $h(x)$, reverse hazard function $r(x)$ and the quantile function $Q(u)$ are given below:

$$S(x) \; = \; 1 - F(x) = 1 - \sin\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\} \tag{7}$$

$$h\,(x) = \frac{f(x)}{1 - F(x)} = \frac{\frac{\pi}{2}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \cos\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\}}{1 - \sin\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\}} \tag{8}$$

$$r(x) = \frac{f(x)}{F(x)} = \frac{\pi}{2}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \cot\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\} \tag{9}$$

$$Q(U) = F^{-1}\,\lambda\left\{-log\left(1 - \frac{2\sin^{-1} U}{\pi}\right)\right\}^{\frac{1}{k}} \tag{10}$$

## 2.4. Parameter Estimation

The parameters of the newly developed Sine-Weibull distribution will be estimated using the method of maximum likelihood (MLE). Moment and moment generating function (mgf) will be used in determine the mean, variance, skewness and kurtosis, among other properties, of the proposed distribution.

### 2.4.1 Method of Maximum Likelihood

Let $Y_1, Y_2, \dots, Y_n$ independent, identically distributed ($iid$) random sample of a random variable $Y$ with $pdf$ given by $f(y/\delta)$, then the likelihood function $L(\delta: y)$ of $Y_1, Y_2, \dots, Y_n$ is the joint density function when regarded as a function of the parameter. That is
$L(\delta: y) = \Pi_{i=1}^{n} f(y_i, \delta)$

It is more convenient to use the log likelihood.

$$l(\delta : y) = lnL(\delta, y)$$

The estimate of the parameter can be obtained by taking the derivative of the log likelihood function with respect to the parameter and equating to zero, that is

$$\frac{\partial y}{\partial \delta}\ lnL(\delta, y) = 0 \tag{11}$$

## 2.4.2 Maximum Likelihood of Sine-Weibull Distribution

Let $X_1, X_2, \dots, X_n$ be a random sample of size $n$ from a Sine-Weibull distribution with a $pdf$ given by (1.1), the likelihood function$L(\lambda : x)$ of this sample is given as

$$L(\lambda : x) = \Pi_{i=1}^{n} f(x_i, \lambda,) = \Pi_{i=1}^{n} \frac{\pi}{2}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \cos\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\}$$

$$L(\lambda : x) = \left(\frac{\pi}{2}\right)^n \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^{k-1} \cos\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x_i}{\lambda}\right)^k}\right]\right\} e^{-\sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k}$$

Taking the log of the likelihood function gives

$$l(\lambda, x) = \ln\left(\left(\frac{\pi}{2}\right)^n \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^{k-1} \cos\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x_i}{\lambda}\right)^k}\right]\right\} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k}\right)$$

$$l(\lambda, x) = \mathrm{n}\ ln\left(\frac{\pi}{2}\right) + (k-1) ln \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right) + ln \sum_{i=1}^{n} \cos\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x_i}{\lambda}\right)^k}\right]\right\} - \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^k$$

$$\cos\left\{\frac{\pi}{2}\left[1 - e^{-\left(\frac{x_i}{\lambda}\right)^k}\right]\right\} = 0, \qquad \text{because} \qquad \cos\left(\frac{\pi}{2}\right) = 0$$

To maximize equation(11), we take the derivative with respect to $\lambda$ and equate to zero

$$\frac{\partial l}{\partial \lambda} = -(k-1) \sum_{i=1}^{n} \left(\frac{1}{\lambda}\right) - \sum_{i=1}^{n} \left(\frac{-kx^k}{\lambda^{k+1}}\right) = 0$$

$$\frac{\partial l}{\partial \lambda} = -(k-1) \sum_{i=1}^{n} \left(\frac{1}{\lambda}\right) + k \sum_{i=1}^{n} \left(\frac{x^k}{\lambda^{k+1}}\right) = 0$$

$$k \sum_{i=1}^{n} \left(\frac{x^k}{\lambda^{k+1}}\right) = (k-1) \sum_{i=1}^{n} \left(\frac{1}{\lambda}\right)$$

$$k \sum_{i=1}^{n} x^k = (k-1) \sum_{i=1}^{n} \lambda^{-1}\, \lambda^{k+1}$$

$$\sum_{i=1}^{n} \lambda^k = \frac{k}{(k-1)} \sum_{i=1}^{n} x^k$$

$$\lambda^k = \frac{1}{n}\ \ \frac{k}{(k-1)} \sum_{i=1}^{n} x^k$$

$$\hat{\lambda} = \sqrt[k]{\frac{1}{n}\ \ \frac{k}{(k-1)} \sum_{i=1}^{n} x^k} \tag{12}$$

Equation (12) gives the maximum likelihood estimator of the parameter $\lambda$

## 2.5. Some Mathematical Properties

### 2.5.1 Moment

Moments plays a vital role in the field of statistical analysis, particularly when it comes to real applications. Suppose that X is a random variable and r is a non-negative integer, the r[th] moment of X is the quantity $E(X^k)$ provided it expectation exist. The r[th] is given by:

$$E(x^r) = \int_{x=0}^{\infty} xf(x)dx$$

The r[th] moment of proposed Sine-Weibull distribution is derive as follows:

$$E(x^r) = \frac{\pi}{2}\int_{x=0}^{\infty} x^r \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \cos\left\{\frac{\pi}{2}\left[1-e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\} dx$$

$$E(x^r) = \frac{\pi}{2\lambda^{k-1}}\int_{x=0}^{\infty} (x)^{k-r+1} e^{-\left(\frac{x}{\lambda}\right)^k} \cos\left\{\frac{\pi}{2}\left[1-e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\} dx$$

$$E(x^r) = \frac{\pi}{2\lambda^{k-1}}(\lambda)^{k-1+r}\int_{x=0}^{\infty} \left(\frac{x}{\lambda}\right)^{k-1+r} e^{-\left(\frac{x}{\lambda}\right)^k} \cos\left\{\frac{\pi}{2}\left[1-e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\} dx$$

$$E(x^r) = \int_{x=0}^{\infty} \left(\frac{x}{\lambda}\right)^{k-1+r} e^{-\left(\frac{x}{\lambda}\right)^k} \cos\left\{\frac{\pi}{2}\left[1-e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\} dx = 1\ (it\ is\ a\ pdf)$$

$$\Rightarrow E(x^r) = \frac{\pi}{2\lambda^{k-1}}(\lambda)^{k-1+r}$$

$$E(x^r) = \frac{\pi}{2}\lambda^r \tag{13}$$

The first and second moments (when r = 1 and r = 2) are therefore given below,

$$E(x) = \frac{\pi}{2}\lambda \tag{14}$$

$$E(x^2) = \frac{\pi}{2}\lambda^2 \tag{15}$$

The variance is given below

$$V(x) = E(x^2) - [E(x)]^2 = \frac{\pi}{2}\lambda^2 - \left(\frac{\pi}{2}\lambda\right)^2$$

$$V(x) = \frac{\pi}{2}\lambda^2\left(1-\frac{\pi}{2}\right) \tag{16}$$

$$\text{Standard Deviation (S)} = \sqrt{\frac{\pi}{2}\lambda^2\left(1-\frac{\pi}{2}\right)} \tag{17}$$

### 2.5.2 Skewness and Kurtosis of the Sine-Weibull Distributions

The skewness and kurtosis of the sine-Weibull distribution are obtained using the third and fourth moment respectively with the power of the standard deviation of the distribution These approaches is the measure of kurtosis ($\alpha_3$) and skewness($\alpha_4$) based on moments

$$(\alpha_3) = \frac{E(x^3)}{S^3}$$

$$(\alpha_3) = \frac{\frac{\pi}{2}\lambda^3}{\left(\sqrt{\frac{\pi}{2}\lambda^2(1-\frac{\pi}{2})}\right)^3} \tag{18}$$

$$(\alpha_4)=\frac{E(x^4)}{S^4}=\frac{\frac{\pi}{2}\lambda^4}{\left(\sqrt{\frac{\pi}{2}\lambda^2\left(1-\frac{\pi}{2}\right)}\right)^4}=\frac{\frac{\pi}{2}\lambda^4}{\left(\frac{\pi}{2}\lambda^2\left(1-\frac{\pi}{2}\right)\right)^2}$$

$$(\alpha_4)=\frac{1}{\frac{\pi}{2}(1-\frac{\pi}{2})^2} \tag{19}$$

### 2.5.3 Entropy

The entropy of is a measure of variation of the uncertainty. There are many entropy measures studied and discussed in literature but the Renyi entropy is perhaps one of the most popular. Renyi entropy of with proposed density function is given by

$$i_{R(\rho)}=\frac{1}{1-\rho}log\left(\int_0^\infty f(x)^\rho dx\right) \tag{20}$$

where $\rho > 0$ *and* $\rho \neq 0$. Inserting equation (4) into (20)

$$i_{R(\rho)}=\frac{1}{1-\rho}log\left(\int_0^\infty\left(\frac{\pi}{2}\left(\frac{x}{\lambda}\right)^{k-1}e^{-\left(\frac{x}{\lambda}\right)^k}\cos\left\{\frac{\pi}{2}\left[1-e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\}\right)^\rho dx\right) \tag{21}$$

### 2.5.4 Order Statistics

Suppose that $x_1, x_2, \dots, x_n$ are random samples of size n from probability distribution with pdf f(x) and cdf $F(x)$ as defined in (3) and (4) respectively, the p^th^ order statistic can be expressed

$$f_n(x)=\frac{n!\,f(x)}{(p-1)!\,(n-1)!}F(x)^{p-1}[1-F(x)]^{n-p} \tag{22}$$

The order statistics of the proposed Sine-Weibull distribution is given by:

$$f_n(x)=\frac{n!\left(\frac{\pi}{2}\left(\frac{x}{\lambda}\right)^{k-1}e^{-\left(\frac{x}{\lambda}\right)^k}\cos\left\{\frac{\pi}{2}\left[1-e^{-\left(\frac{x}{\lambda}\right)^k}\right]\right\}\right)^\rho}{(p-1)!\,(n-1)!}\left\{\sin\left[\frac{\pi}{2}\left(1-e^{-\left(\frac{x}{\lambda}\right)^k}\right)\right]\right\}^{p-1}$$
$$\times\left\{1-\sin\left[\frac{\pi}{2}\left(1-e^{-\left(\frac{x}{\lambda}\right)^k}\right)\right]\right\}^{n-p} \tag{23}$$

## III. Results

### 3.1 Application

Specifically, AIC is aimed to obtain the best approximating model to the unknown true data generating process. Superficially, BIC differs from AIC only in the first term which depends on sample size n. Models that minimize the BIC are selected. From a Bayesian perspective, BIC is designed to find the most probable model given the data.

### 3.1.1 Dataset

One dataset was considered for illustrative purposes and comparison with the baseline distribution and other competitors. The comparison was done with Weibull distribution and Lomax distribution. We estimated the unknown parameters of the distribution by the maximum-likelihood method. We obtain the values of the Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC) for the newly developed distribution as well as the competitors. The dataset consists of thirty successive values of March precipitation (in inches) in Minneapolis/St [16]. The data are as follows:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.77 | 1.74 | 0.81 | 1.2 | 1.95 | 1.2 | 0.47 | 1.43 | 3.37 | 2.2 | 3.0 | 3.09 |
| 1.51 | 2.1 | 0.52 | 1.62 | 1.31 | 0.32 | 0.59 | 0.81 | 2.81 | 1.87 | 1.18 | 1.35 |
| 4.75 | 2.48 | 0.96 | 1.89 | 0.9 | 2.05 | | | | | | |

**Table 1:** *Summary Statistics of the dataset*

| Data | Minimum | $Q_1$ | Media | Mean | $Q_3$ | Maximu |
|---|---|---|---|---|---|---|
| Dataset | 0.92 | 1.302 | 1.544 | 1.658 | 1.814 | 5.306 |

Table 1 gives the summary statistics of the data sets such as the mean, the median, the first and third quartile, the minimum and the maximum values.

**Table 2:** *MLE, AIC, CAIC, BIC, and HQIC of the data set*

| Data Set | $MLE$ | $AIC$ | $CAIC$ | $BIC$ | $HQIC$ |
|---|---|---|---|---|---|
| Sine-Weibull | 55.61173 | 115.2235 | 115.3472 | 120.4388 | 117.3322 |
| Weibull | 150.5514 | 305.1029 | 305.2266 | 310.3132 | 307.2716 |
| Lomax | 150.5514 | 303.1029 | 303.1437 | 310.3132 | 304.1572 |

Table 2 presents the results of the analysis of the dataset. The result of the analysis of the Sine-Weibull Distribution was compared with Weibull Distribution and Lomax Distribution to test the efficiency of the model. The proposed Sine-Weibull distribution has proven to be the better model because it has the least AIC, CAIC, BIC and HQIC.

## IV. Discussion

There has been a growing interest among statisticians and applied researchers in developing flexible lifetime models for the betterment of modelling survival data. In this paper, we introduced a two-parameter Sine-Weibull distribution which is obtained by considering a Weibull distribution as the baseline. We study some of its statistical and mathematical properties. Maximum Likelihood Estimation was used in parameter estimation. The usefulness of the new distribution was illustrated via the analysis of real data sets. We hope that the proposed extended model will attract wider applications.

## References

[1] Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, *71*(1), 63-79.

[2] Frechet, M. (1927). Sur la loi de probabilit´e de l´ecart maximum. Ann. de la Soc. polonaisede Math, 6, 93–116.

[3] Rosin, P. and Rammler, E. (1933) The Laws Governing the Fineness of powdered coal. *Journal of the Institute of Fuel, 7,* 29-36

[4] Muhammad, M., Alshanbari, H.M. Alanzi, A.R.A., Liu, L., Sami, W., Chesneau, C., Jamal, F. A. (2021). New Generator of Probability Models: The Exponentiated Sine-G Family for Lifetime Studies. *Entropy, 23*, 1394.

[5] Kumar, D., Singh, U., & Singh, S. K. (2015). A new distribution using sine function-its application to bladder cancer patients' data. *Journal of Statistics Applications & Probability*, *4*(3), 417.

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

[7] Souza, L., de Oliveira, W. R., de Brito, C. C. R., Chesneau, C., Fernandes, R., & Ferreira, T. A. (2022). Sec-G class of distributions: Properties and applications. *Symmetry*, *14*(2), 299.

[8] Al-Faris, R. Q., & Khan, S. (2008). Sine square distribution: a new statistical model based on the sine function. *Journal of Applied Probability and Statistics*, *3*(1), 163-173.

[9] Fayomi, A., Algarni, A., & Almarashi, A. M. (2021). Sine Inverse Lomax Generated Family of Distributions with Applications. *Mathematical Problems in Engineering*, 1-11.

[10] Isa, A. M., Ali, B. A., & Zannah, U. (2022). Sine Burr XII Distribution: Properties and Application to Real Data Sets. *Arid Zone Journal of Basic and Applied Sciences, 1*(3), 48-58.

[11] Chesneau, C., & Jamal, F. (2020). The sine Kumaraswamy-G family of distributions. *Journal of Mathematical Extension*, *15*.

[12] Al-Babtain, A. A., Elbatal, I., Chesneau, C., & Elgarhy, M. (2020). Sine Topp-Leone-G family of distributions: Theory and applications. *Open Physics*, *18*(1), 574-593.

[13] Isa, A. M., Bashiru, S. O., Ali, B. A., Adepoju, A. A., & Itopa, I. I. (2022). Sine-Exponential Distribution: Its Mathematical Properties and Application to Real Dataset. *UMYU Scientifica*, *1*(1), 127-131.

[14] Nagarjuna, V. B. V., Vardhan, R.V. and Chesneau, C. (2021). On the Accuracy of the Sine Power Lomax Model for Data Fitting. *Modelling, 2*, 78–104.

[15] Hinkley, D. (1977). On quick choice of power transformation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *26*(1), 67-69.