



SURVIVAL REGRESSION OF COVID-19 PATIENTS USING WEIBULL PROBABILITY DISTRIBUTION MODEL

Usman Mohammed and Sani Ibrahim Doguwa

Department of Statistics, Ahmadu Bello University, Zaria

Corresponding Author: umohammed67@yahoo.com

ABSTRACT

This study introduces the Sine-Odd Fréchet-Weibull (S-OFr-W) distribution, a novel statistical model designed to enhance flexibility in data modeling. In dealing with real-world data, classical or conventional probability distributions such as Poisson, Exponential, etc. fail to capture the characteristics of the data fully. The flexibility of the distributions is crucial for extracting accurate insights from complex datasets. However, classical probability distributions often fail to adequately capture multimodal behaviours, heavy-tailed data or extreme variations, highlighting the need for more robust and adaptable models. The proposed distribution extends the classical Weibull distribution by incorporating the Sine Odd Fréchet family of distribution, providing improved adaptability for different datasets. Mathematical properties, that include moments, moment generating functions, entropy, order statistics and reliability measures are derived. Parameter estimation is performed using the maximum likelihood estimation (MLE) technique, and a simulation study is conducted to evaluate the consistency and efficiency of the estimates. In addition, a Sine-OFr-Weibull Survival Regression (S-OFr-WSR) model is developed for survival data analysis. The model's performance is assessed using real-datasets, including COVID-19 fatality ratio (%) of India and US COVID-19 survival data. Comparative analysis based on the Akaike Information Criterion (AIC) demonstrates that the S-OFr-Weibull distribution outperforms the competing models of Odd Fréchet Inverse Exponential, Topp-Leone Kumaraswami Exponential and Weighted Weibull Exponential distributions. Furthermore, the log-S-OFr-Weibull survival regression model exhibits a lower AIC values compared to log-Weibull regression model, confirming its superior performance in survival data modeling.

1. INTRODUCTION

Semi-parametric model focuses on the influence of covariates on hazard, but fully parametric model can also compute the distributional form of survival time. The advantages of parametric model in survival analysis include; the distribution of survival time can be estimated; full maximum likelihood can be used to estimate parameters and residuals can represent the difference between observed and estimated values of time. (Zhang, 2016).

In parametric methods it is a requirement that the distribution of survival times should be known. The probability distributions that are often used in survival analysis research are Weibull, exponential and log-logistic. The parametric model assumes that the distribution underlying survival time follows a certain distribution. Survival analysis is a branch of statistics that analyse the expected duration of time until one or more events happen. This analysis focuses attention on a group consisting of a collection of individuals where each individual has a survival time, the time interval until the expected event occurs.

The following three things must be fulfilled to determine survival time (1) there is a start time, (2) a clear time measurement scale, (3) the definition of an event must be clear. In survival analysis, data from individuals or subjects who have not yet experienced the desired event are referred to as censored data (Cox and Oakes, 1984).

Probability distributions are fundamental tools in mathematics and statistics that help in modelling, analysing, and interpreting real-world uncertainty across different applied fields such as applied sciences, medicine, engineering, and biological. Probability distributions model the time until an event occurs. For example, in survival analysis research, survival time used Weibull or exponential distributions. In biological sciences, probability distribution help to model the traits in population genetics. However, in dealing with real-world data, there is often issues presents in which classical or conventional probability distributions (like Normal, Poisson, Exponential, etc.) fail to capture fully (Mohammed & Ugwouwa, 2020). The flexibility of these distributions is crucial for extracting accurate insights from complex datasets. However, classical probability distributions often fail to adequately capture multimodal behaviours, heavy-tailed data or extreme variations, highlighting the need for more robust and adaptable models. To address these limitations, researchers have developed generalized families of distributions, often referred to as generators. These generators enhance the flexibility of existing distributions by introducing additional parameters, such as scale or shape parameters (Isa *et al.*, 2025). Typical examples of such generalized families of distributions includes; Marshall-Olkin Weibull Generalized family (Klakattawi, *et al.*, 2023), Marshall- Olkin Exponentiated half logistic-Generalized family (Gabanakgosi, & Otladisa, (2024).) and the Secant Kumaraswamy family (Nanga *et al.*, 2024).

Trigonometric families of distributions play a greater role with its ability to improve the flexibility of existing distributions without introducing additional parameters. This allows parameters estimation easily and derivation of distribution properties. A popular example is the Sine- G family of distributions, proposed by (Kumar *et al.*, 2015), which has been used to extend various existing distributions. Some of the models that were derived from this family include Sine Exponential distribution (Isa *et al.*, 2022), Sine Lomax distribution (Mustapha *et al.*, 2023) Sine Weibull distribution (Faruq *et al.*, 2024), Sine Type II Topp- Leone Burr XII (Isa *et al.*, 2024), Cosine Marshal-Olkin distribution (Adepoju *et al.*, 2024), have outperformed their parent distributions.

Weibull distribution is a statistical probability distribution used in reliability and life time data analysis. It is a popular distribution because it can handle decreasing, increasing and constant failure rates, making it flexible in modeling various types of datasets. However, the limitations of Weibull in accommodating complex data behaviour led to the occurrence of various extensions, like Topp-Leone Weibull model (Tuoyo *et al.*, 2021,) Topp-Leone modified Weibull model (Alyani *et al.* 2022) Cosine-Weibull model (Wu *et al.*,2023). These extensions introduce additional features, enabling the Weibull distribution to effectively capture diverse hazard rate shapes and complex patterns (Isa *et al.*, 2025). This study proposes an extension of the Weibull distribution based on the Sine-Odd Fréchet family, named as Sine-Odd Fréchet-Weibull (SOFr-W) distribution. The model possesses the advantage of the properties of the Sine-Odd-Fréchet family to address the limitations of the classical Weibull distribution. The study derives the properties of the proposed distribution, including moments, quantile functions and hazard rate behaviour. The performance of the model is evaluated using real datasets, demonstrate its performance in fitting complex data.

2. Materials and Methods

2.1 Sine G Family of Distribution

Let $Q(x)$ be the CDF of a univariate continuous distribution and $q(x)$ be the PDF, then, the Sine-G family of distributions according to (Kumar *et al.*, 2015) is defined by the CDF given as:

$$F(x, \xi) = \int_0^{\frac{\pi}{2} Q(x, \xi)} \cos(t) dt = \sin \left\{ \frac{\pi}{2} Q(x, \xi) \right\} \tag{1}$$

and its corresponding PDF given as:

$$f(x, \xi) = \frac{\pi}{2} g(x, \xi) \cos \left\{ \frac{\pi}{2} Q(x, \xi) \right\} \tag{2}$$

Where $Q(x, \xi)$ and $g(x, \xi)$ are the PDF and CDF of any baseline distribution and ξ is a vector parameter of the baseline distribution.

2.2 Sine-Odd Frechet G Family of Distribution

Equation (3) gives the CDF of sine-Odd Frechet G family as:

$$F(x, \lambda, \xi) = \sin \left\{ \frac{\pi}{2} \exp \left[- \left(\frac{1 - G(x, \xi)}{G(x, \xi)} \right)^\lambda \right] \right\} \quad x, \lambda > 0 \tag{3}$$

and the corresponding PDF is given as:

$$f(x, \lambda, \xi) = \frac{\pi \lambda g(x, \xi) (1 - G(x, \xi))^{\lambda - 1} \exp \left[- \left(\frac{1 - G(x, \xi)}{G(x, \xi)} \right)^\lambda \right]}{2 G(x, \xi)^{\lambda + 1}} \times \cos \left\{ \frac{\pi}{2} \exp \left[- \left(\frac{1 - G(x, \xi)}{G(x, \xi)} \right)^\lambda \right] \right\} \quad x, \lambda > 0 \tag{4}$$

Where λ is the shape parameter which controls the skewness and kurtosis of the distribution. $G(x, \xi)$ and $g(x, \xi)$ are the CDF and PDF of the baseline distribution.

2.3 Weibull Distribution

Given that the baseline distribution is Weibull with CDF and PDF respectively given as:

$$G(x, k, \theta) = 1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right) \quad x, k, \theta > 0 \tag{5}$$

$$g(x, k, \theta) = \frac{k}{\theta} \left(\frac{x}{\theta} \right)^{k - 1} \exp \left(- \left(\frac{x}{\theta} \right)^k \right) \quad x, k, \theta > 0 \tag{6}$$

where as k is the shape parameter and θ is the scale parameter

2.4 The Proposed Sine-Odd Frechet-Weibull Distribution

The CDF of Sine-Odd Frechet-Weibull distribution was obtained by inserting the CDF of the Weibull distribution in Equation (5) into to the CDF of the Sine-Odd Frechet- *G* family in Equation (3) as follows:

$$F(x, \lambda, k, \theta) = \sin \left\{ \frac{\pi}{2} \exp \left[- \frac{\left(1 - \left(1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right) \right) \right)^\lambda}{1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right)} \right] \right\} \quad x, \lambda, \theta, k > 0 \quad (7)$$

and the corresponding PDF is also obtained by substituting the PDF and CDF of the Weibull distribution in equation (5) and (6) respectively into the PDF of the Sine-Odd Frechet-*G* family in equation (4) as follows:

$$f(x, \lambda, k, \theta) = \frac{\pi}{2} \frac{\lambda \frac{k}{\theta} \left(\frac{x}{\theta} \right)^{k-1} \exp \left(- \left(\frac{x}{\theta} \right)^k \right) \left(1 - \left(1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right) \right) \right)^{\lambda-1}}{1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right)^{\lambda+1}} \times \exp \left[- \frac{\left(1 - \left(1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right) \right) \right)^\lambda}{1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right)} \right] \cos \left\{ \frac{\pi}{2} \exp \left[- \frac{\left(1 - \left(1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right) \right) \right)^\lambda}{1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right)} \right] \right\} \quad x, \lambda, \theta, k > 0 \quad (8)$$

where *k* is the shape parameter while θ is a scale parameter. These parameters control the flexibility and the behaviour of the distribution.

The survival function *S*(*x*), hazard function *h*(*x*) and the quantile function *Q*(*u*) are given in equation (9), (10) and (11).

$$S(x) = 1 - \sin \left\{ \frac{\pi}{2} \exp \left[- \frac{\left(1 - \left(1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right) \right) \right)^\lambda}{1 - \exp \left(- \left(\frac{x}{\theta} \right)^k \right)} \right] \right\} \quad (9)$$

$$\begin{aligned}
 h(x) &= \frac{\pi}{2} \frac{\lambda \left(\frac{x}{\theta}\right)^{k-1} \exp\left(-\left(\frac{x}{\theta}\right)^k\right) \left(1 - 1 - \exp\left(-\left(\frac{x}{\theta}\right)^k\right)\right)^{\lambda-1}}{1 - \exp\left(-\left(\frac{x}{\theta}\right)^k\right)^{\lambda+1}} \exp\left[-\frac{\left(1 - 1 - \exp\left(-\left(\frac{x}{\theta}\right)^k\right)\right)^{\lambda}}{1 - \exp\left(-\left(\frac{x}{\theta}\right)^k\right)}\right] \\
 &\times \cos\left\{\frac{\pi}{2} \exp\left[-\frac{\left(1 - 1 - \exp\left(-\left(\frac{x}{\theta}\right)^k\right)\right)^{\lambda}}{1 - \exp\left(-\left(\frac{x}{\theta}\right)^k\right)}\right]\right\} \left\{1 - \sin\left[\frac{\pi}{2} \exp\left[-\frac{\left(1 - 1 - \exp\left(-\left(\frac{x}{\theta}\right)^k\right)\right)^{\lambda}}{1 - \exp\left(-\left(\frac{x}{\theta}\right)^k\right)}\right]\right]\right\}^{-1} \quad (10)
 \end{aligned}$$

$$Q(u) = \theta \left\{ -\log \left[1 + \left[1 - \left[\log \left(\frac{\sin^{-1}(u)}{\frac{\pi}{2}} \right) \right]^{\frac{1}{\lambda}} - 1 \right]^{\frac{1}{\lambda}} \right] \right\}^{\frac{1}{k}} \quad (11)$$

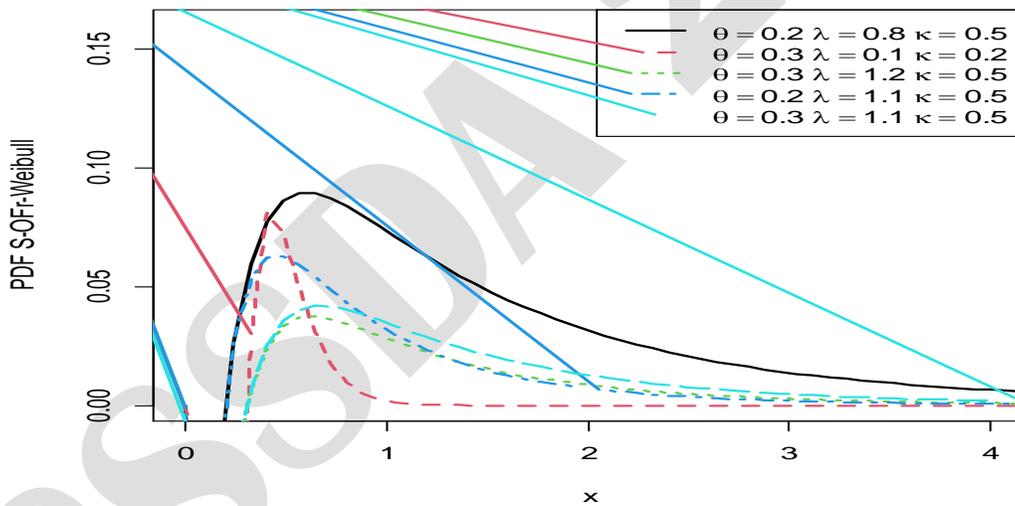


Figure 1: PDF Plot of Sine-Odd Fréchet-Weibull Distribution

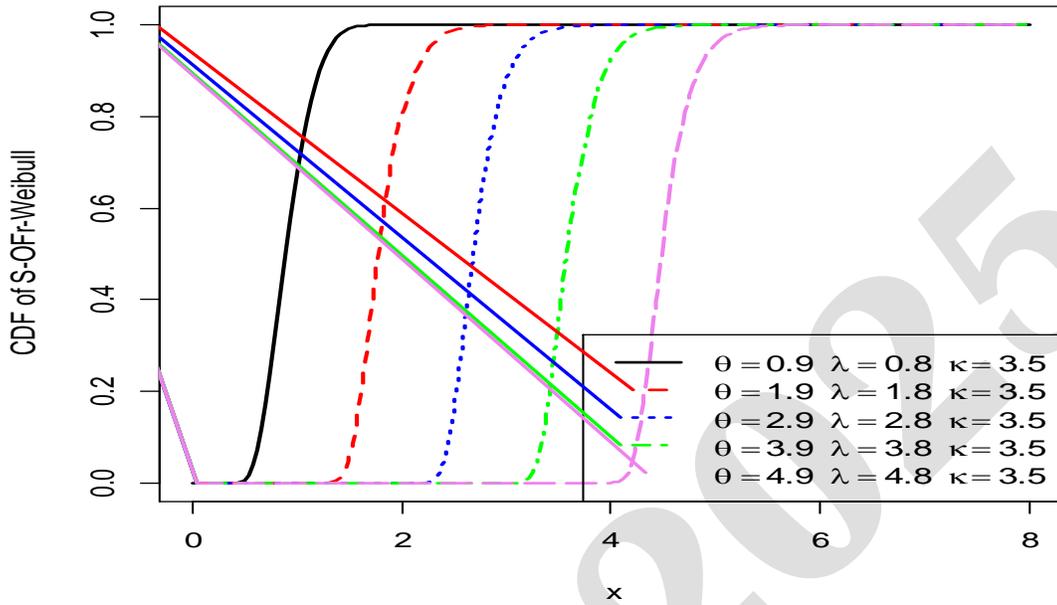


Figure 1: CDF Plot of Sine-Odd Fréchet-Weibull Distribution

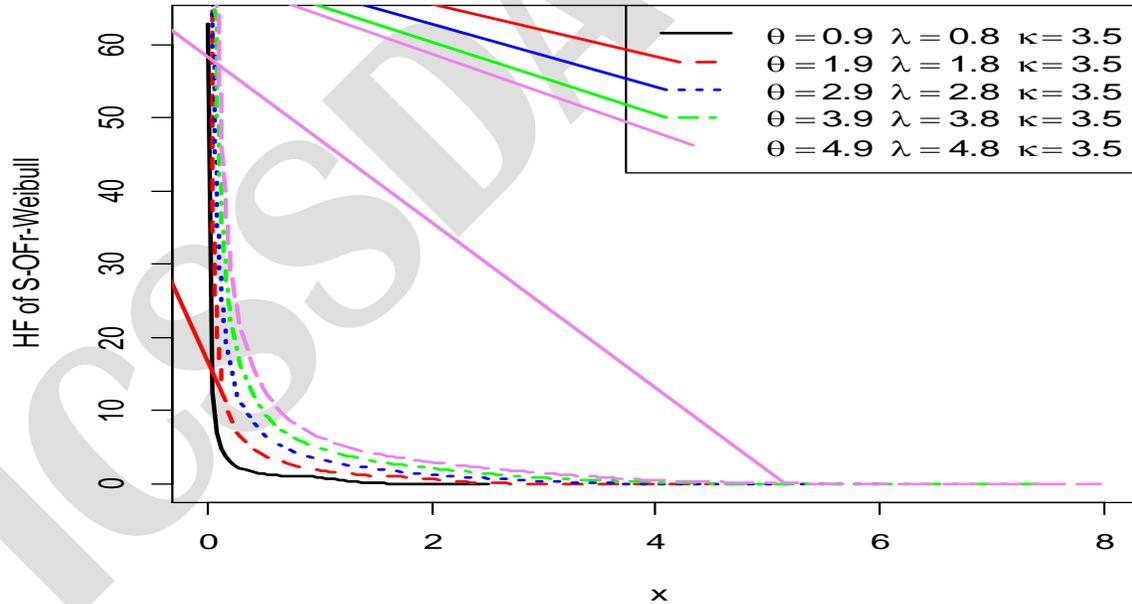


Figure 4: Hazard Function Plot of Sine-Odd Fréchet-Weibull Distribution

Figure 1 and 2 represents the PDF and CDF of the S-OFr-W distribution for different parameter values. While figure 3 displays the hazard rate function of the S-OFr-W distribution, illustrating how the risk of failure changes over time

2.5 The Probability Density Expansion

The PDF and CDF of the proposed S-OFr-W distribution can be expanded using Taylor’s Series Expansion and Binomial Expansion as follows:

Given the CDF of S-OFr-G family of distribution in Equation (3), the CDF can be expanded as follows:

$$\sin \left\{ \frac{\pi}{2} \exp \left[- \left(\frac{1 - G(x, \xi)}{G(x, \xi)} \right)^\lambda \right] \right\} = \sum_{i=0}^{\infty} (-1)^i \frac{\left(\frac{\pi}{2} \right)^{2i+1}}{(2i+1)!} \left[\exp \left(- \left(\frac{1 - G(x, \xi)}{G(x, \xi)} \right)^\lambda \right) \right]^{2i+1} \tag{12}$$

$$= \sum_{i=0}^{\infty} (-1)^i \frac{\left(\frac{\pi}{2} \right)^{2i+1}}{(2i+1)!} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left(\frac{1 - G(x, \xi)}{G(x, \xi)} \right)^{\lambda j} \tag{13}$$

Now, $\left(\frac{1 - G(x, \xi)}{G(x, \xi)} \right)^{\lambda j} = G(x, \xi)^{-\lambda j} (1 - G(x, \xi))^{\lambda j}$

Again, $(1 - G(x, \xi))^{\lambda j} = \sum_{k=0}^{\infty} (-1)^k \binom{\lambda j}{k} G(x, \xi)^{\lambda j k}$

$$F(x) = \sum_{i=0}^{\infty} (-1)^i \frac{\left(\frac{\pi}{2} \right)^{2i+1}}{(2i+1)!} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \sum_{k=0}^{\infty} (-1)^k \binom{\lambda j}{k} G(x, \xi)^k G(x, \xi)^{-\lambda j} \tag{14}$$

$$= \sum_{i=0}^{\infty} (-1)^i \frac{\left(\frac{\pi}{2} \right)^{2i+1}}{(2i+1)!} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \sum_{k=0}^{\infty} (-1)^k \binom{\lambda j}{k} G(x, \xi)^{-\lambda j+k}$$

$$= \sum_{i,j,k=0}^{\infty} \frac{(-1)^{i+j+k} \pi^{2i+1}}{(2i+1)! 2^{2i+1}} G(x, \xi)^{-\lambda j+k} \tag{15}$$

Let $\eta = \sum_{i,j,k=0}^{\infty} \frac{(-1)^{i+j+k} \pi^{2i+1}}{(2i+1)! 2^{2i+1}}$

Therefore, the reduced form of CDF of S-OFr-G is given as:

$$F(x, \lambda, \xi) = \eta G(x, \xi)^m \tag{16}$$

where $m = -\lambda j + k$. Differentiating the reduced form of CDF in Equation (16) gives the reduced form of PDF as:

$$f(x, \lambda, \xi) = \eta m g(x, \xi) G(x, \xi)^{m-1} \tag{17}$$

2.6 Mathematical Properties of S-OFr-Weibull Distribution

The mathematical properties of the S-OFr-Weibull distribution such as the moment, moment generating function, entropy and order statistics are derived as follows:

2.6.1 Moment of S-OFr-Weibull Distribution

The moment of the S-OFr-Weibull distribution is obtained as follows:

The r^{th} moment of a random variable X that follows S-OFr- G family of the distribution is given as:

$$\mu_r^{\downarrow} = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx \tag{18}$$

$$= \int_0^{\infty} x^r \eta m g(x, \xi) G(x, \xi)^{m-1} dx \tag{19}$$

Let

$$\kappa = g(x, \xi) G(x, \xi)^{m-1} dx \tag{20}$$

$$\mu_r^{\downarrow} = E(X^r) = \eta m \int_0^{\infty} x^r \kappa dx \tag{21}$$

2.6.2 Moment Generating Function of S-OFr-W Distribution

The moment generating function of a random variable can be obtained as follows

$$M_x(t) = E(e^{tx}) = \int_0^{\infty} e^{tx} f_{S-OFr-G}(x) dx \tag{22}$$

$$= \eta m \int_0^{\infty} e^{tx} g(x, \xi) G(x, \xi)^{m-1} dx \tag{23}$$

2.6.3 Entropy of S-OFr-W Distribution

The entropy of a random variable X that follows S-OFr- G family of the distribution is given as:

$$I_R(\theta) = \frac{1}{1-\theta} \log \left(\int_0^{\infty} f_{S-OFr-G}(x)^{\theta} dx \right) \tag{24}$$

$$I_R(\theta) = \frac{1}{1-\theta} \log \left(\int_0^{\infty} (\eta m g(x, \xi) G(x, \xi)^{m-1})^{\theta} dx \right) \tag{25}$$

Let

$$W = g(x, \xi) G(x, \xi)^{m-1} \tag{26}$$

$$I_R(\theta) = \frac{1}{1-\theta} \log \left(\int_0^{\infty} (\eta m W)^{\theta} dx \right) \tag{27}$$

$\theta > 0$ and $\theta \neq 1$

2.6.4 Order Statistics of S-OFr-W Distribution

Given that X_1, X_2, \dots, X_n is a random sample from S-OFr -G family of distribution and $X_{i:n}$ represent the i^{th} order statistics then,

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} f(x) [F(x)]^{i-1} [1-F(x)]^{n-i} \tag{28}$$

$$\tag{29}$$

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} \eta mg(x, \xi) G(x, \xi)^{m-1} [\eta G(x, \xi)^m] [1-\eta G(x, \xi)^m]^{n-i}$$

3. Sine-Odd Frechet-Weibull Survival Regression (S-OFr-WSR)

The S-OFr-WSR) was obtained using the one-to-one transformation defined in equation (29) is used to change the probability density function of X directly in terms of

$$g(y) = f(x) \left| \frac{dx}{dy} \right| \tag{30}$$

Covariates or exposure variables in clinical trials such as age, sex, tumor size etc at a start of a treatment can have impact on survival time in a variety of real application. Investigating the casual relationship between survival time and these factors is crucial. It is possible to employ a survival regression model to capture the effect of these factors on the survival times of individuals or subjects. Consider a group of location scale model, where the vector $X = (X_1, X_2, \dots, X_p)^T$ of the covariates is associated with the response variable $Y = \log(X)$ using a regression framework. Let X be a random variable with S-OFr-W probability density function defined by Equation (29), where θ is the scale parameter and k is the shape parameter. By replacing the scale parameter θ with e^μ and shape parameter k with $\frac{1}{\sigma}$, and then using the log transformation yield the following:

$$g(y) = \frac{\pi}{2} \frac{\lambda}{\sigma \exp\{\mu\}} \left(\frac{\exp\{y\}}{\exp\{\mu\}} \right)^{\frac{1}{\sigma}-1} \exp \left\{ \left(\frac{\exp\{y\}}{\exp\{\mu\}} \right)^{\frac{1}{\sigma}} \right\} \left[\frac{1 - \left(1 - \exp \left(- \left(\frac{\exp\{y\}}{\exp\{\mu\}} \right)^{\frac{1}{\sigma}} \right) \right)^{\lambda-1}}{\left(1 - \exp \left(- \left(\frac{\exp\{y\}}{\exp\{\mu\}} \right)^{\frac{1}{\sigma}} \right) \right)^{\lambda+1}} \right]^{\lambda} \exp \left\{ - \left[\frac{1 - \left(1 - \exp \left(- \left(\frac{\exp\{y\}}{\exp\{\mu\}} \right)^{\frac{1}{\sigma}} \right) \right)^{\lambda}}{\left(1 - \exp \left(- \left(\frac{\exp\{y\}}{\exp\{\mu\}} \right)^{\frac{1}{\sigma}} \right) \right)^{\lambda}} \right] \right\} \tag{31}$$

$$* \cos \left\{ \frac{\pi}{2} \exp \left\{ - \left[\frac{1 - \left(1 - \exp \left(- \left(\frac{\exp\{y\}}{\exp\{\mu\}} \right)^{\frac{1}{\sigma}} \right) \right)^{\lambda}}{\left(1 - \exp \left(- \left(\frac{\exp\{y\}}{\exp\{\mu\}} \right)^{\frac{1}{\sigma}} \right) \right)^{\lambda}} \right] \right\} \right\} |\exp\{y\}|$$

$$\begin{aligned}
 g(y) &= \frac{\pi \lambda}{2 \sigma} \exp\left\{\frac{y-\mu}{\sigma}\right\} \exp\left(-\left(\exp\left\{\frac{y-\mu}{\sigma}\right\}\right)\right) \frac{\left[1-\left(1-\exp\left(-\left(\exp\left\{\frac{y-\mu}{\sigma}\right\}\right)\right)\right)\right]^{\lambda-1}}{\left(1-\exp\left(-\left(\exp\left\{\frac{y-\mu}{\sigma}\right\}\right)\right)\right)^{\lambda+1}} \\
 &\times \exp\left\{-\left[\frac{1-\left(1-\exp\left(-\left(\exp\left\{\frac{y-\mu}{\sigma}\right\}\right)\right)\right)}{\left(1-\exp\left(-\left(\exp\left\{\frac{y-\mu}{\sigma}\right\}\right)\right)\right)}\right]^{\lambda}\right\} \cos\left\{\frac{\pi}{2} \exp\left\{-\left[\frac{1-\left(1-\exp\left(-\left(\exp\left\{\frac{y-\mu}{\sigma}\right\}\right)\right)\right)}{\left(1-\exp\left(-\left(\exp\left\{\frac{y-\mu}{\sigma}\right\}\right)\right)\right)}\right]\right\}\right\}
 \end{aligned}
 \tag{32}$$

where $\lambda, \sigma, \mu > 0$ and $-\infty < \mu < \infty$. The distribution defined in Equation (32) is called the Log-Sine Odd Fréchet-Weibull (S-OFr-W) distribution i.e $Y \sim \text{Log-S-OFr-W}(\lambda, \sigma, \mu)$ and the survival function of the transformed PDF is presented in equation (33).

$$S(y) = 1 - \sin\left\{\frac{\pi}{2} \exp\left[-\left(\frac{1-\left(1-\exp\left\{\frac{y-\mu}{\sigma}\right\}\right)}{1-\exp\left\{\frac{y-\mu}{\sigma}\right\}}\right)^{\lambda}\right]\right\}
 \tag{33}$$

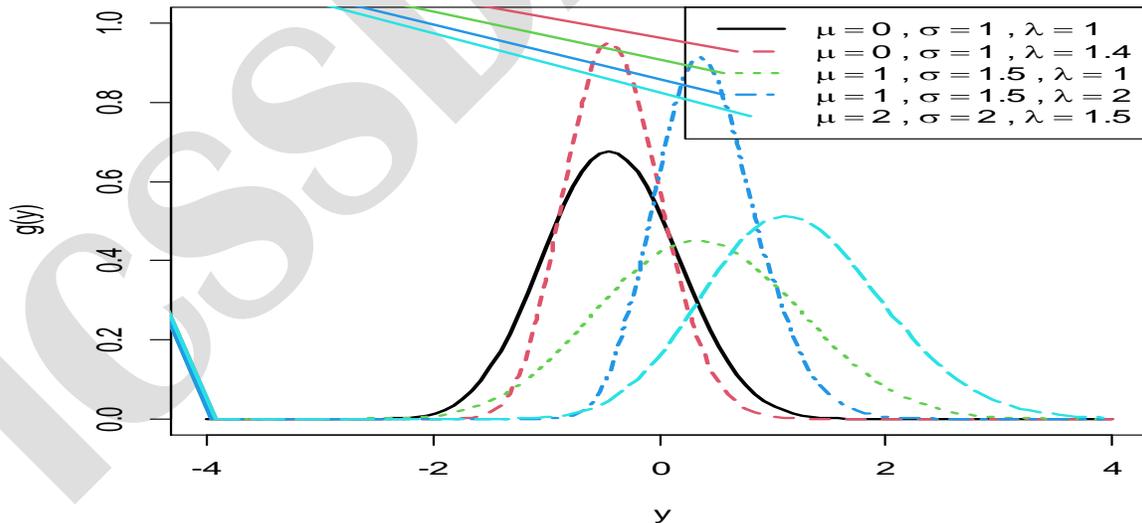


Figure 5: PDF Plot of Log S-OFr-Weibull Distribution

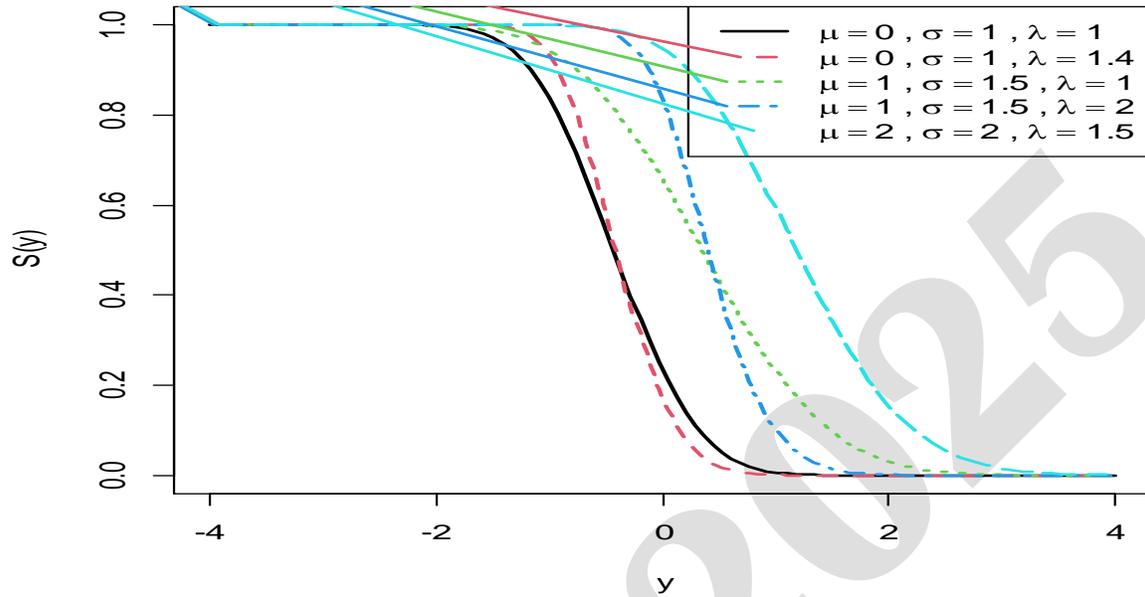


Figure 6: Survival Plot of Log S-OFr-Weibull Distribution

Equation (32) can be written as a log-linear model:

$$Y = \mu + \sigma Z \tag{34}$$

Let $Z = \left(\frac{\mu - y}{\sigma}\right)$ be the standardized random variable, then, the random variable has PDF given by:

$$g(y) = \frac{\pi \lambda}{2 \sigma} e^Z e^{-e^{\left(\frac{1 - (1 - e^{-e^Z})^{\lambda-1}}{(1 - (1 - e^{-e^Z}))^{\lambda+1}}\right)}} e^{-\left(\frac{1 - (1 - e^{-e^Z})}{(1 - (1 - e^{-e^Z}))}\right)^\lambda} \cos \left(\frac{\pi}{2} e^{-\left(\frac{1 - (1 - (1 - e^{-e^Z}))}{(1 - e^{-e^Z})}\right)} \right) \tag{35}$$

Where $-\infty < z < \infty$. Consider the framework of regression model based on the Log-S-OFr-Weibull distribution in equation (29). Linking the response variable Y to the vector of explanatory variables X, and let $\mu_i = \beta^T x_i$ in the Log-Linear model of equation (34) so that the model of Y given X can be represented by:

$$y_i = \beta^T x_i + z_i \quad i = 1, 2, \dots, n \tag{36}$$

Where z_i are random errors which follows the Log-S-OFr-Weibull distribution in equation (32), $\lambda, \sigma > 0$ and $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ are unknown parameters, therefore, equation (36) is referred to as the Log-SOFr-W survival regression model, with survival function given in equation (33). Following this transformation, the model can be expressed as a log-linear regression model, and making it easier to interpret the effects of the exposure variables on the survival times of the subjects.

3.1 Simulation Study of Sine-Odd Fréchet-Weibull Distribution

A Monte Carlo simulation study was conducted to evaluate the performance of the proposed S-OFr-Weibull distribution. The data was generated using the quantile function of the distribution for different sample sizes of (20, 50, 75, 100, 150, 250), with parameters set at $\theta = 1.0, \lambda = 1.0$ and $k=1.1$ Each sample size was replicated 1000 times. The mean, bias, and mean square error (MSE) of the maximum likelihood estimates were computed. The results in Table 1, indicate the accuracy and consistency of the parameter estimates for different sample sizes.

Table 1: Results of the Simulated Data from S-OFr-Weibull Distribution

<i>n</i>	Properties	$\theta=1.0$	$\lambda=1.0$	$\kappa=1.1$
20	Estimate	1.0454	1.3231	0.9946
	Bias	0.0454	0.3231	-0.1054
	MSE	0.2066	0.3885	0.043
50	Estimate	1.0379	1.2089	1.0043
	Bias	0.0379	0.2089	-0.0957
	MSE	0.1352	0.1932	0.0260
75	Estimate	1.0376	1.1706	1.0076
	Bias	0.0376	0.1706	-0.0924
	MSE	0.1026	0.1421	0.0203
100	Estimate	1.0363	1.1548	1.0096
	Bias	0.0363	0.1548	-0.0904
	MSE	0.0874	0.1206	0.0172
150	Estimate	1.0296	1.1333	1.0089
	Bias	0.0296	0.1333	-0.0911
	MSE	0.0659	0.081	0.0144
250	Estimate	1.0234	1.1239	1.0085
	Bias	0.0234	0.1239	-0.0915
	MSE	0.0487	0.0652	0.0129

Table 1 represents the results of the Monte Carlo simulation study, with bias and MSE of the parameter estimates for various sample sizes. The results revealed that, the values of biases and MSEs tend to zero as shown in Table 1, and the estimates tend to the true values of parameter as the sample size increases, indicating that the estimates are efficient and consistent.

3.2 Application to Real Dataset

The dataset of size 54 represents the COVID-19 fatality ratio (%) of India, from 1st February to 1st April of year 2022. The data was collected from official site of World Health Organization (WHO) <https://cvoterindia.com/trackers>

1.067, 1.757, 1.705, 1.849, 1.131, 1.595, 1.524, 1.014, 1.266, 1.678, 1.758, 1.898, 1.459, 3.370, 1.283, 1.753, 1.840, 2.134, 2.293, 2.217, 2.365, 1.485, 2.603, 2.952, 2.164, 3.142, 4.880, 2.885, 1.513, 2.704, 3.169, 2.485, 6.080, 2.462, 1.508, 1.078, 3.777, 3.407, 2.363, 5.893, 3.421, 7.211, 2.001, 2.087, 3.487, 3.457, 4.925, 2.469, 10.48, 2.514, 2.779, 2.514, 2.285, 3.895

Table 2: MLE Estimates, Information Criteria and Goodness of Fit Test for the Dataset

Model	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{\kappa}$	$\hat{\beta}$	$\hat{\sigma}$	LL	AIC	KS (<i>p</i> -value)
SOFrW		3.6366	6.9934	0.343			82.1509	172.302	0.048 (0.9996)
TLOFrW	2.8405	3.3985	5.6366	0.3149			83.0024	174.008	0.054 (0.9975)
OFrIE	1.2687		1.7469				84.6486	173.532	0.092 (0.7512)
EGFr	1.8105	1.7872			0.8079	82.9668	83.9668	173.934	0.058 (0.9940)
TLKE	7.4829	0.1893	2.2714		2.3251		85.0746	178.149	0.083 (0.8513)
WWE	0.0523	5.1858	1.5544				96.4071	198.814	0.172 (0.0834)

Table 2 presents the parameter estimates, information criteria and goodness-of-fit measures for competing models from the COVID-19 fatality ratio (%) of India dataset. The performance metric used to assess the fit is AIC. The distribution model with the lowest value for the metric indicates the best fit. In this case, Sine-Odd Fréchet Weibull (S-OFr-W) distribution demonstrates superior performance compared to the other models considered in the dataset. The study employed Kolmogorov–Smirnov (KS) test, along with *p*-values to compare the competing distributions. The distribution with the highest *p*-value is deemed the most optimal for the dataset. From Table 2, it is clear that the S-OFr-W model with the highest *p*-value provides a better fit than the other models considered. Therefore, the S-OFr-W model is selected as the appropriate model for analyzing the case fatality ratio (%) of India dataset.

Figure 1 displays the histogram of the fitted density plots for various probability distributions applied to the COVID-19 fatality ratio (in %) of India. All the distributions exhibit a little bit right-skewed shape. The closer the fit, the more likely that specific distribution accurately represents the real-world data. The results of analysis reveals that the S-OFr-W distribution, represented by the purple colour curve, is the best-fitting model for this dataset. Therefore, this study concludes that the proposed family of distributions is suitable for modelling COVID-19 fatality ratio (%) of India.

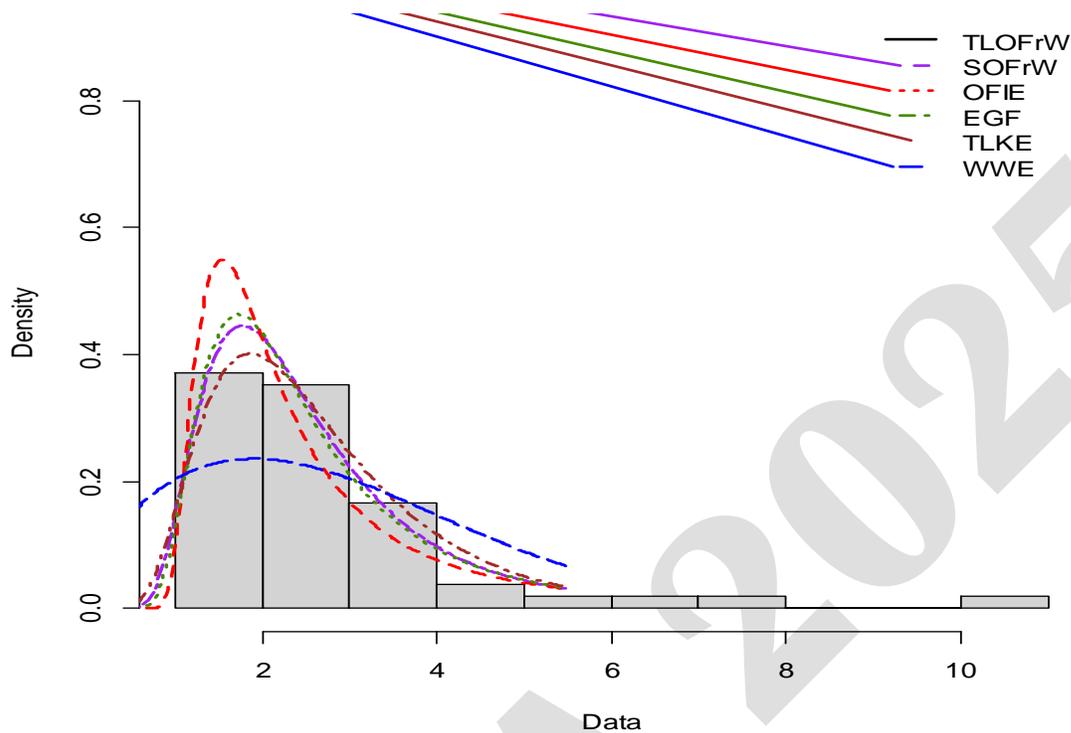


Figure 1: The curves of Density of the Fitted Models of COVID-19 fatality ratio (%) of India

3.3 Fitting Sine-Odd Fréchet-Weibull Survival Regression with US COVID-19 Dataset

The US COVID-19 dataset was obtained from World Health Organisation (WHO, 2023). The dataset contains a sample size of 425 of COVID-19 patients, admitted to US health facilities with COVID-19 from January to June 2022 was considered. 61.6% were censored while 38.4% were uncensored (those who experienced the event COVID-19). The response variable y is the natural logarithm of the observed survival time in days; d = censoring (0 = patient is alive at study end, lost to follow-up or withdraw from the study. 1 = death due to COVID-19).

Given the following exposure variables: x_1 : the patient's age is less than or equal 30 years; x_2 : the patient's age group is 31-60 years; x_3 : the patient that is over 60 years; x_4 : the patient is male; x_5 : the patient is a female; x_6 : respiratory distress syndrome; x_7 : arrhythmia; x_8 : coughing; x_9 : fever; x_{10} : difficulty in breathing; x_{11} : travel a lot; x_{12} : chronic disease; x_{13} : cardiovascular disease; x_{14} : diabetes; x_{15} : hypertension.

Table 3: MLE of the Parameters for the Two Fitted Survival Regression Models to the COVID-19 Dataset

Parameter	Log-SOFWSR		Log-WSR	
	Estimate	P-value	Estimate	P-value
λ	29.8908	0.65822	-	-
σ	37.2738	0.66325	1.09289	2.2e-16
β_0	-5.7899	0.91202	1.74684	1.1e-15
β_1	7.70875	0.71486	2.71119	0.00091
β_2	7.34642	0.72767	-0.0070	0.98323
β_3	7.18619	7.3e-16	-2.0074	2.8e-09
β_4	12.8549	5.9e-12	2.25519	2.2e-16
β_5	12.9969	3.6e-12	2.34166	2.2e-16
β_6	-0.0168	0.91218	-0.8254	0.00047
β_7	-0.1786	0.52649	-1.0515	0.01877
β_8	-0.4830	0.00945	-1.0515	0.72643
β_9	0.0322	0.00412	0.07711	0.85932
β_{10}	1.1724	0.00234	-0.2003	0.75997
β_{11}	-0.8925	0.00929	2.00175	0.75997
β_{12}	0.52007	0.48605	1.23378	0.29406
β_{13}	-0.0689	0.92834	-1.1283	0.34929
β_{14}	-0.0124	0.98697	-1.8765	0.11404
β_{15}	-0.3607	0.63152	-1.4300	0.22230
LL	182.431		742.675	
AIC	218.4305		776.6747	
Rank	1		2	

Table 3 presents the results of the fitted log-S-OFrWSR and log-WSR models for the COVID-19 dataset. The survival regression models for the COVID-19 dataset based on the AIC statistics were compared. The model with the lowest AIC value is preferred. From Table 3, Log-SOFrWSR has a lower value of AIC followed by the comparator Log-WSR, indicating that, even though the Log-WSR model is more appropriate, but Log-SOFrWSR model fit the dataset better than the log-WSR model. The results for the best-fitted survival regression model i.e. Log-SOFrWSR show that there are statistically significant effects of the exposure variables except for (x1, x2, x6, x7, x12, x13, x14 and x15), that a COVID-19 patient will survive beyond time t , at 5% significance level.

3.4 Cox-Snell Residuals Analysis for COVID-19

An important step after constructing a survival regression model is analyzing residuals to assess the adequacy of the model. The computation of the residuals enables the assessment of how well the model

fits the survival data by comparing the observed and predicted survival probabilities or survival curves. The study focuses on the Cox-Snell residual method, introduced by (Cox and Snell, 1968), which is computed as the negative of the natural logarithm of the estimated survival probability, comparing observed and estimated survival times. It is defined as:

$$e_i = -\log[S(y_i / x_i)], \quad i = 1, 2, \dots, n \tag{37}$$

where $S(t_i)$ is the survival function obtained from the fitted survival regression models. The Cox-Snell residuals of the proposed log- S-OFr-Weibull survival regression models is given as:

$$e_i = -\log \left[1 - \sin \left\{ \frac{\pi}{2} \exp \left[- \left(\frac{1 - \left(1 - \exp \left\{ \frac{y - X_i^T \pi}{\sigma} \right\} \right)^\lambda}{1 - \exp \left\{ \frac{y - X_i^T \pi}{\sigma} \right\}} \right) \right] \right\} \right] \quad i = 1, 2, \dots, n \tag{38}$$

If the fitted models are adequate, the residuals will follow standard exponential distribution

Table 4: MLE of the Parameters for the Two Fitted Survival Regression Models to the Cox-Snell Residuals for the COVID-19 Dataset

Parameter	Log-SOFrWSR		Log-WSR	
	Estimate	P-value	Estimate	P-value
λ	51.3679633	2.1e-16	-	-
σ	51.7794156	2.1e-16	0.32139269	2.2e-16
β_0	8.6883743	0.2949	-3.0921779	0.96361
β_1	6.0121795	2.0e-16	13.0396916	0.74678
β_2	6.6094807	2.0e-16	1.7469209	0.91106
β_3	6.5827022	2.0e-16	0.43857449	0.97763
β_4	6.5003533	0.0035	3.43053729	0.95503
β_5	6.5326976	0.0047	3.38040361	0.95569
β_6	-0.0700489	0.0083	-0.6972067	2.2e-16
β_7	-0.3531873	-0.0496	-0.9060465	4.5e-11
β_8	-0.0078264	-0.9723	-0.5626327	0.00029
β_9	0.1935397	0.3461	0.25762963	0.01176
β_{10}	0.2232416	0.4762	0.51086219	0.00772
β_{11}	-0.3467378	0.0325	3.79610623	0.67187
β_{12}	-0.0531207	0.9286	0.15041894	0.38554
β_{13}	0.2230333	0.0132	0.0406741	8.3e-07
β_{14}	0.2012324	0.0081	-0.73167	7.2e-05
β_{15}	0.0399	0.9467	-0.2371095	0.17014
LL	127.3741		271.4474	
AIC	163.3741		305.4474	
Rank	1		2	

The Cox-Snell residuals analysis evaluates the fit of the regression models by comparing the observed survival data with the expected survival data under the assumed model. The proposed model and the competing model were assessed and the results revealed that, the proposed SOFr-WSR yielded a lower AIC value (163.3741) compared to the log-WSR (305.4474) model, indicating that the log-SOFr-Weibull model provides a better fit to the US COVID-19 survival data. Significant parameter estimates, particularly for variables, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_{11}, \beta_{13}, \beta_{14}$ in the log-SOFr-WSR model, suggest these variables strongly influence survival time of the individuals.

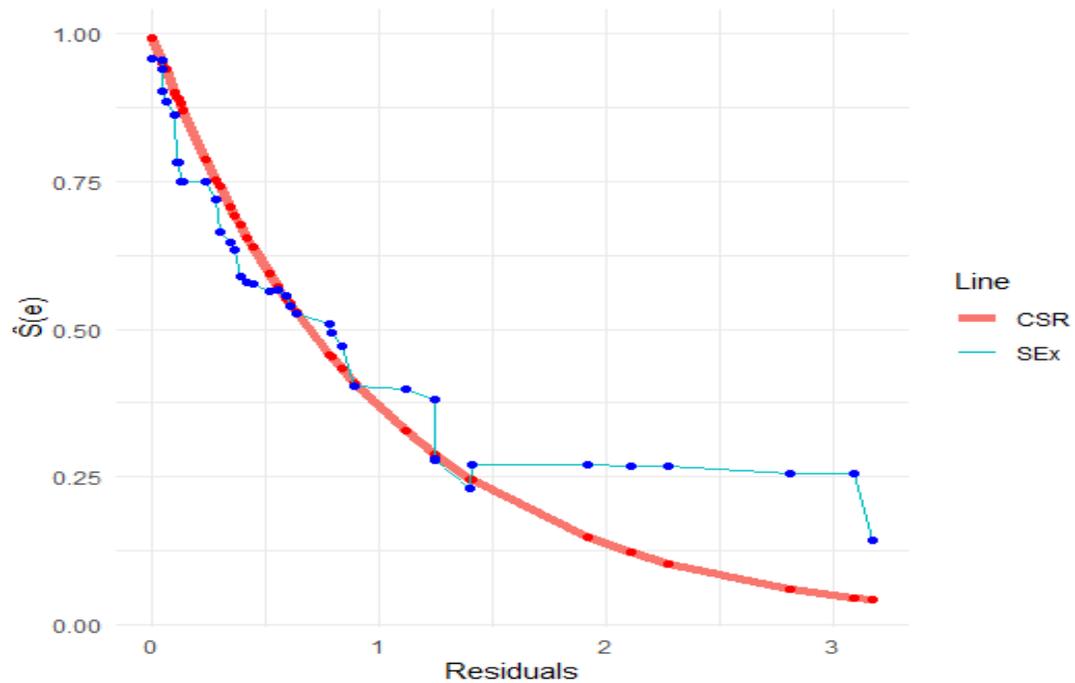


Figure 8: Cox-Snell Residuals Plots for US COVID-19 using Log-S-OFr-WSR

Using the Kolmogorov-Smirnov one-sample test, the D_n statistic yields values of 0.0292 which is less than the Table value (0.0659) at 5% significance level. This suggests that the Cox-Snell residuals for Log-SOFr-WSR model follow the standard exponential distribution. From Figure 8, it can be seen that the estimated standard exponential curve aligns with the theoretical survival curve for the model. Hence, the model fits US COVID-19 data and can be used for inference.

3.4 US COVID-19 Kaplan-Meier Survival Probability

The results of the Kaplan-Meier survival probability analysis for the US COVID-19 dataset are presented in Table 5, while the 95% confidence interval for the survival probability is shown in Figure 9. The proposed and the competing models are also plotted with the non-parametric Kaplan-Meier estimator in in figure 11 and 12.

Table 5: Kaplan-Meier Survival Probability for US COVID-19 Dataset

time	No. at risk	No. of event	Survival prob.	SE	lower 95% CI	upper 95% CI
1	424	70	0.8350	0.0180	0.8000	0.8710
2	349	22	0.7820	0.0201	0.7440	0.8230
3	327	12	0.7540	0.0210	0.7140	0.7960
4	308	18	0.7100	0.0222	0.6670	0.7540
5	288	13	0.6780	0.0229	0.6340	0.7240
6	267	16	0.6370	0.0237	0.5920	0.6850
7	230	2	0.6310	0.0238	0.5860	0.6800
8	221	3	0.6230	0.0240	0.5780	0.6720
9	202	1	0.6200	0.0240	0.5740	0.6690
10	186	4	0.6060	0.0244	0.5600	0.6560
13	118	1	0.6010	0.0248	0.5550	0.6520
18	51	1	0.5890	0.0269	0.5390	0.6450

Table 9 presents Kaplan-Meier survival probabilities for US COVID-19 patients over time, showing how survival rates decrease as time progresses. It can be observed that, as more events (deaths) occur over time, the survival probability continues to decline. The Table shows how survival trend and uncertainty were captured by SE and CIs at different time.

Kaplan-Meier Survival Curve for COVID-19

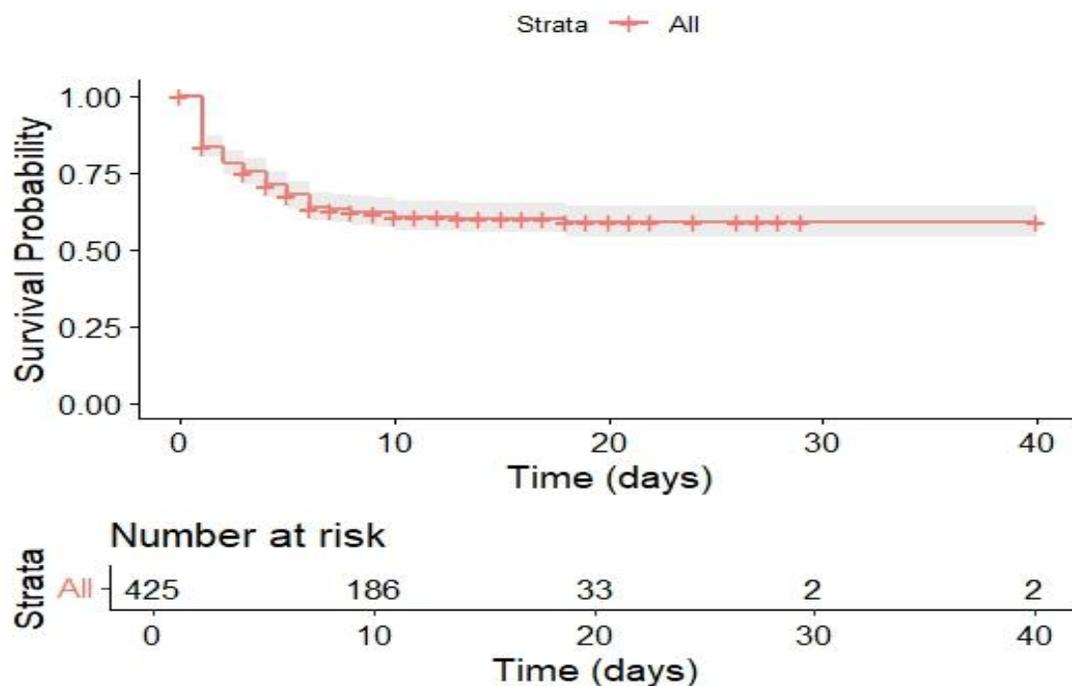


Figure 9: Kaplan-Meier Survival Curve for US COVID-19 Dataset

The 95% confidence interval for the survival probability is displayed in Figure 9. The exposure variable had improved the survival probabilities for these patients who experience the event (dead) before around

six days, and the survival probability continued as shown in Figure 9. The second graph in Figure 9, shows how many patients are still being followed (i.e., are still alive and at risk of death) at different time points. At time 0 weeks, there were 425 patients. At 10 weeks, 186 patients were still at risk. At 20 weeks only 33 patients were still alive, meaning that 392 had either died or been censored by that time. At 40 weeks, only 2 patients still at risk. This means that as time goes on, fewer patients are available for analysis either because they have had the event (death) or were censored (lost to follow-up) or they are still alive.

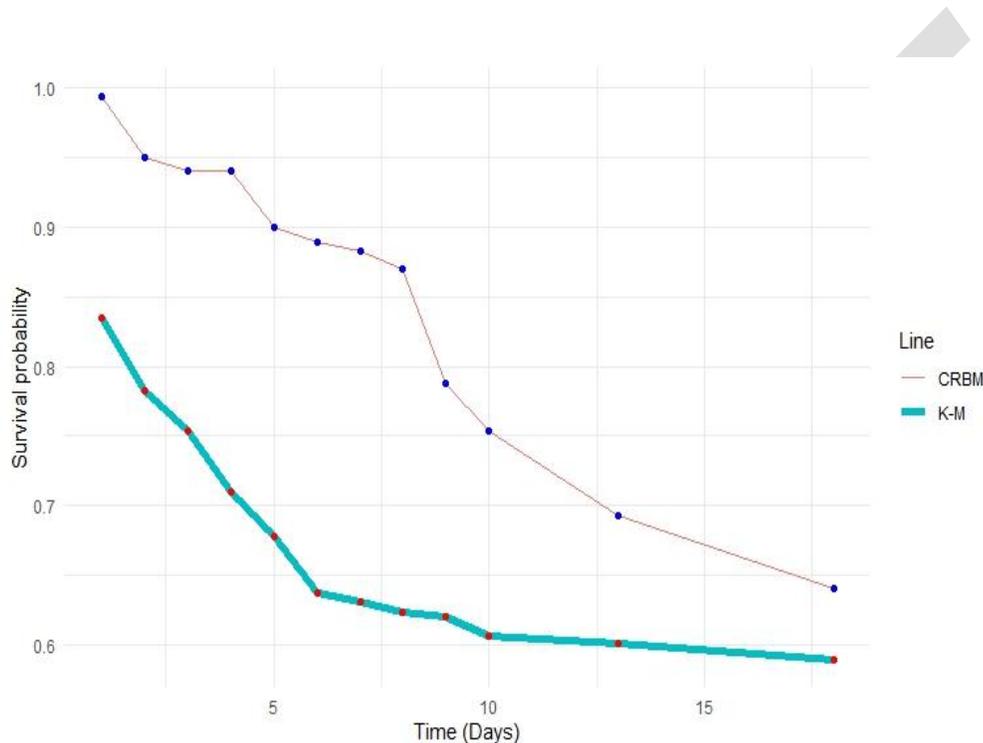


Figure 10: K-M and Log-SOFr-WSR Survival Curve for US COVID-19 dataset

The plot in Figure 10 shows the best-fitted survival regression model together with the non-parametric Kaplan-Meier survival probability. The effects of the covariates play a significant role in describing the distribution of the event times and the existence of the association between the explanatory variables and the event-time distribution. In Figure 10, the red line represents the Cox-Snell residuals best model (CRBM) line, while the blue line represents the Kaplan-Meier (K-M). The observed deviation between CRBM model and Kaplan-Meier survival estimates may be attributed to the inherent complexity of the data. This complexity stems from combination of some variables that are not captured in the data which may influence survival differences in the risk profiles of the patients. As a results, the model tends to slightly overestimates the survival probability compared to K-M estimates. Nevertheless, the modelling process applied in this study represents a systematic first efforts to understand the data structure. For future analysis, more flexible modelling strategies should be considered such as alternative parametric forms or semi-parametric approaches to better capture the observed survival pattern.

4. Conclusion

In this study, the Sine-odd Fréchet-Weibull (SOFr-W) distribution was proposed and fitted to the data on COVID-19 fatality ratio (%) of India, from 1st February to 1st April of year 2022. Maximum likelihood estimation (MLE) was applied to estimate the parameters of the proposed model, and its goodness-of-fit

was evaluated using log-likelihood (LL) and Akaike Information Criteria (AIC). The SOFr-W distribution achieved the lowest AIC value (173.302) compared to other competing models of Topp-Leone Odd Frechet Weibull (TLOFrW), Odd Frechet Inverse Exponential (OFrIE), Exponential Generalized Frechet (EGFr), Topp-Leone Kumaraswami Frechet (TLKFr) and Weighted Weibull Exponential (WWE) models, indicating that the SOFr-Weibull model provides the best fit for the COVID-19 fatality ratio data.

Furthermore, the SOFr-Weibull was transformed into log-SOFr-WSR and was also compared with the log-Weibull regression model based on their AIC values. The log-SOFr-WSR yielded a lower AIC (218.4305) compared to the log-WSR (776,6747), confirming that it is a suitable model for analyzing US COVID-19 dataset. Also, significant parameter estimates were observed for some exposure variables in the log-SOFr-WSR, indicating strong relationships between the predictor and the outcome variables. Cox-Snell residuals analysis was conducted to further assess the fit of the regression models. The analysis revealed that the log-SOFr-WSR model outperformed the log-Weibull model, with a lower AIC value (128.1833) as compared to (133.3930) respectively. The significant parameters in the proposed model ($\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_{11}, \beta_{13}, \beta_{14}$) demonstrated that these variables have significant influence on US COVID-19 patients survival time. Therefore, the results demonstrate that the SOFr-Weibull distribution and its associated survival regression model provide a robust and reliable framework for modelling COVID-19 fatality ratio dataset, providing improved accuracy compared to classical models. This study contributes in developing a novel flexible SOFr-Weibull distribution for real-data modelling and also, the study suggests further applications of the model in similar datasets.

References

- Adepoju, A. A., Isa, A. M., & Bello, O. A. (2024). Cosine Marshal-Olkin-G Family of Distribution: Properties and Applications. *Reliability: Theory & Applications*, **19**(3), 408-422
- Alyami, S. A., Elbatal, I., Alotaibi, N., Almetwally, E. M., Okasha, H. M., & Elgarhy, M. (2022). Topp-Leone modified Weibull model: Theory and applications to medical and engineering data. *Applied Sciences*, **12**(20), 10431.
- Charles, Z. (2023). <https://real-statistics.com/survival-analysis/cox-regression/real-statistics-capabilities-for-cox-regression/survival-analysis-study-data/>
- Cox D. R. and Oakes D. (1984) Analysis of Data Survival Data (London: Chapman and Hall)
- Cox, D. R., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, **30**(2), 248-265.
- Faruk, M. U., Isa, A. M., & Kaigama, A. (2024). Sine-Weibull Distribution: Mathematical Properties and Application to Real Datasets. *Reliability: Theory & Applications*, **19**(1), 65-72.
- Gabanakgosi, M., & Otladisa, M. (2024). A new family of generalized-G distributions with properties and applications. *Scientific African*, **26**, e02344

- 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*, *11*(1), 127-131.
- Isa, A. M., Doguwa, S. I., Alhaji, B. B., & Dikko, H. G. (2024). Sine Type II Topp-Leone Burr XII Distribution: Mathematical Properties and Application to Biomedical and Environmental Data. *Iraqi Statisticians journal*, 30-40.
- Isa, A. M., Doguwa, S. I., Alhaji, B. B., & Dikko, G. (2025). Extension of the Weibull Survival Regression Based on Sine Type II Topp-Leone-G Family of Distribution: Properties and Application. *Iraqi Statisticians journal*, 74-90.
- Klakattawi, H., Alsulami, D., Elaal, M. A., Dey, S., & Baharith, L. (2023). Correction: A new generalized family of distributions based on combining Marshal-Olkin transformation with T-X family. *PLoS ONE*, *18*(10), e0293100
- 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), 417-425.
- Mohammed, A. S., & Ugwuowo, F. I. (2020). A new four-parameter Weibull Distribution with application to failure time data. *FUDMA Journal of Sciences*, *4*(3), 563-575.
- Mustapha B. A., Isa A. M., Sule O. B., & Itopa I. I. (2023). Sine-Lomax distribution: properties and applications to real data sets. *FUDMA Journal of Sciences*, *7*(4), 60 - 66.
- Nanga, S., Sayibu, S. B., Angbing, I. D., Alhassan, M., Benson, A.-M., Abubakari, A. G., Nasiru, S., & Suleman, S. (2024). Secant Kumaraswamy family of distributions: Properties, regression model, and applications. *Computational and Mathematical Methods*, 8925329
- Tuoyo, D. O., Opone, F. C., & Ekhosuehi, N. (2021). The Topp-Leone Weibull distribution: Its properties and application. *Earthline Journal of Mathematical Sciences*, *7*(2), 381-401
- Wu, X., Ahmad, Z., Hussam, E., Alhelali, M. H., Aldallal, R., Almuqrin, M. A., & Riad, F. H. (2023). A new Cosine-Weibull model: Distributional properties with applications to basketball and medical sectors. *Alexandria Engineering Journal*, *66*, 751-767
- Zhang, Z. (2016). Parametric regression model for survival data: Weibull regression model as an example. *Annals of translational medicine*, *4*(24), 484. <https://cvoterindia.com/trackers> WHO (2023). Who US covid-19.