Applications of a Stationary Weibull Process method with an R-package

Debarghya Jana[1]

Roll: 221306

Supervisor: Dr. Debasis Kundu[2]

A Project Report Submitted in the Requirements of the course MTH599A for the Degree of

MASTER OF SCIENCE

in

STATISTICS

to

DEPARTMENT OF MATHEMATICS AND STATISTICS

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

NOVEMBER, 2023

[1]final year Student, M.Sc. Statistics, IIT Kanpur

[2]Professor, Department of Mathematics and Statistics, IIT Kanpur

ABSTRACT

This report will introduce a novel stochastic process characterized by discrete time intervals and a continuous state space, denoted as $\{X_n; n = 1, 2, \ldots\}$. The random variable X_n in this process follows a two-parameter Weibull distribution, and most importantly, these variables exhibit temporal dependence with a non-zero probability of consecutive values, i.e., X_n being equal to X_{n+1}. The motivation came from several instances where time-series data came from a lag-1 stationary process in which $X_n = X_{n+1}$, which can not be ignored. There are several positive valued stationary processes available in the existing literature but none of them can be applied for the case $X_n = X_{n+1}$, because in all these cases $Pr(X_n = X_{n+1}) = 0$. Our work delves into various aspects of this new Weibull process, unveiling its distinctive characteristics. Notably, we discover that the joint cumulative distribution function of X_n and X_{n+1} exhibits an elegant copula structure. This finding facilitates the exploration of diverse dependence properties and the computation of dependence measures within this framework. While explicit forms for the maximum likelihood estimators prove elusive, we introduce a straightforward profile likelihood method to effectively compute these estimators. To validate the applicability of our model, we apply it to the analysis of two synthetic datasets. Our results demonstrate a compelling fit of the proposed model to the observed data, underscoring its utility in capturing the unique characteristics of such datasets. Parametric bootstrap is also used to find the test statistic and corresponding p-values for the models which we build up using the bootstrap samples, for the Goodness of fit test by which we can detect which model provides a good fit to the particular dataset. Ultimately, we build an R-package which will take time-series data from the user and will pass on both models and give the output of test statistics from both models after finding the MLEs and will also provide the p-value to understand how good the fit is, corresponding to the test statistic. It will also provide the histograms of generated test statistics for both of the models which we obtained from parametric bootstrapping.

Keywords: Stochastic process, Weibull distribution, Copula structure, Profile likelihood method, Parametric bootstrap..

Contents

1. Introduction

In the realm of stochastic processes, a groundbreaking approach has been unveiled in the work of Kundu (2022) as they introduce an innovative Markov Weibull process. This novel process operates in discrete time with a continuous state space, characterized by its remarkable flexibility and unique attributes, rendering it applicable across various domains and disciplines. This method is designed in such a way that it ensures the marginal distributions exhibit the desirable characteristics of non-negativity and follow a two-parameter Weibull distribution. Additionally, it establishes a system where the random variables, X_n, exhibit dependence on one another. We know that the autocorrelation function (ACF) in a time series refers to the coefficient of correlation between two data points. For instance, in a time series denoted as z_t, the ACF is calculated as $\text{Corr}\,(z_t, z_{t-k})$. The parameter k represents the time gap being analyzed and is termed the lag. In specific terms, a lag-1 autocorrelation (where $k = 1$) quantifies the correlation between data points separated by a single time period. In a more general context, a lag-k autocorrelation assesses the correlation between data points that are separated by $\boldsymbol{k}$ time periods. In this context, the Weibull process is established as a lag-1 process through a minimization approach, and it exhibits the property where the values of X_n are equal to X_{n+1} with a non-zero probability. Consequently, when dealing with non-negative time series data characterized by positive skewness and the likelihood of consecutive values being equal, the proposed model emerges as a valuable tool for effectively analyzing such datasets.

Instances where $X_n = X_{n+1}$ are significant and should not be overlooked. It is worth noting that numerous positively valued stationary processes have been documented in the literature. For instance, the exponential process introduced in Tavares (1980), the Weibull and Gamma processes presented in Sim (1986), Arnold introduced the logistic process, Arnold and Hallet discussed the Pareto process Arnold (1993), Pillai proposed the semi-Pareto process Pillai (1991), and Jayakumar and Girish Babu introduced the generalized Weibull process Jayakumar and Babu (2015). References to Arnod and Robertson (1989), Jose, Ristić and Joseph (2011), and their respective citations also provide insights into these processes. However, it's crucial to emphasize that none of these processes is suitable for the current scenario because in all these cases, the probability of $P(X_n = X_{n+1})$ is equal to zero. In the main paper Kundu (2022), there is already mentioned regarding different properties of the proposed process X_n. It is to be noted that this Weibull process has one shape parameter and two scale parameters.

Various distinctive properties of the proposed process $\{X_n\}$ have been elucidated. The Weibull process $\{X_n\}$ features a single shape parameter and two scale parameters. Notably, when the shape parameter equals one, it transforms into a stationary exponential process. The generation process of $\{X_n\}$ is straightforward, facilitating the convenient execution of diverse simulation experiments.Furthermore, the joint distribution of X_n and X_{n+1} exhibits a highly advantageous copula structure, simplifying the computation of various dependence properties and dependence measures. The characterization of this process is comprehensively provided, with the marginals and the joint probability density function capable of assuming a wide array of shapes. However, it is worth noting that the autocovariance and autocorrelation of this process do not lend themselves to convenient forms. We have included an appendix section to provide the necessary proof of the theorems for a better understanding. Our aim of this paper is to make an R package documentation of this stationary process which will create a function that will return the test statistic and the associated p-value for the established models and it will detect the best fitted model for the data. Moreover, the function in this R-package will also return the histograms of the generated test statistics for both of the established models which are obtained from each bootstrap sample. We made an appendix section for providing the test run of the R-package. In section 2.1, we discussed the Weibull process. section 2.1.1 we have provided all the theorems related to the Weibull process and discussed the several properties. Insection 2.2, we have provided the mathematical details maximum likelihood estimation process for two cases and described the profile likelihood method. In section 2.3, we have provided the algorithm of the parametric bootstrap method which is used to find the test statistic and we described the testing of the hypothesis method. In section 3, we generated two synthetic data sets in different model specifications and we did simulation studies to find MLEs of the parameter used in the model in both specifications. Finally, in section 4, we described our R-package which we have built for easy implementation of these models.

2. Methodology

2.1. Description of Weibull Process:

In the paper Kundu (2022), the following notions have been adopted. A Weibull random variable, characterized by the shape parameter $\alpha > 0$ and the scale parameter $\lambda > 0$, possesses the following probability density function (PDF):

$$f_{WE}(x;\alpha,\lambda) = \begin{cases} \alpha\lambda x^{\alpha-1}e^{-\lambda x^{\alpha}} & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

This distribution will be represented as WE(α, λ). For $x > 0$, it is associated with the following cumulative distribution function and hazard function, respectively. The cumulative distribution function and hazard function for the Weibull random variable with parameters α and λ, denoted as WE(α, λ), are as follows:

$$F_{WE}(x;\alpha,\lambda) = 1 - e^{-\lambda x^{\alpha}},$$
$$h_{WE}(x;\alpha,\lambda) = \alpha\lambda x^{\alpha-1}.$$

The mean and variance of WE(α, λ) are given by:

$$\frac{1}{\lambda^{1/\alpha}}\Gamma\left(\frac{1}{\alpha}+1\right) \quad \text{and} \quad \frac{1}{\lambda^{2/\alpha}}\left[\Gamma\left(\frac{2}{\alpha}+1\right) - \left(\Gamma\left(\frac{1}{\alpha}+1\right)\right)^2\right],$$

respectively. Additionally, we denote a uniform random variable on the interval $(0, 1)$ as $U(0, 1)$. With these definitions, we are now ready to introduce the Weibull process.

Now we will define the Weibull process in details which has been exclusively published in the paper Kundu (2022).Consider a sequence of independent identically distributed (i.i.d.) random variables $U_0, U_1, \ldots$, each following a uniform distribution $U(0, 1)$. Now, for given parameters $\lambda_0 > 0$, $\lambda_1 > 0$, and $\alpha > 0$, let's introduce a new sequence of random variables $\{X_n; n = 1, 2, \ldots\}$ defined as:

$$X_n = \min\left\{\left[-\frac{1}{\lambda_0}\ln U_n\right]^{\frac{1}{\alpha}}, \left[-\frac{1}{\lambda_1}\ln U_{n-1}\right]^{\frac{1}{\alpha}}\right\}.$$

This sequence of random variables, denoted as a Weibull process, is referred to as WEP $(\alpha, \lambda_0, \lambda_1)$. In this context, α represents the shape parameter, while λ_0 and λ_1 are the scale parameters. It's worth noting that it is possible to incorporate a location parameter into this model, although such an extension has not been explored here. The term **Weibull process** is derived from the following outcomes.

2.1.1. Theorems:

There are several theorems of this Weibull process which have been discussed in the paper Kundu (2022). We will now describe those details.

Theorem 1: If $\{X_n\}$ is defined as in Equation (3),then:

(a) $\{X_n\}$ constitutes a stationary process.

(b) The random variable X_n follows a Weibull distribution WE $(\alpha, \lambda_0 + \lambda_1)$.

Proof: The verification of Part (a) is obvious from the given definition. To establish Part (b), observe that:

$$\begin{aligned}
P\left(X_n > x\right) &= P\left\{\left[-\frac{1}{\lambda_0}\ln U_n\right]^{\frac{1}{\alpha}} > x, \left[-\frac{1}{\lambda_1}\ln U_{n-1}\right]^{\frac{1}{\alpha}} > x\right\} \\
&= P\left\{U_n < e^{-\lambda_0 x^\alpha}, U_{n-1} < e^{-\lambda_1 x^\alpha}\right\} \\
&= P\left\{U_n < e^{-\lambda_0 x^\alpha}\right\} P\left\{U_{n-1} < e^{-\lambda_1 x^\alpha}\right\} \\
&= e^{-(\lambda_0+\lambda_1)x^\alpha}.
\end{aligned}$$

Theorem 2:

Suppose we have a random variable X_1 that follows a Weibull distribution WE $(\alpha, \lambda_0 + \lambda_1)$. Additionally, let U_i be a sequence of independent and identically distributed (i.i.d.) random variables, all drawn from an absolutely continuous distribution function $F(\cdot)$ defined on the interval $(0,1)$. Then, the process defined in Equation (3) assumes the characteristics of a strictly stationary Markov process if and only if U_0 follows a uniform distribution $U(0,1)$.

Proof:

To establish the only if part of the proof, we first establish some key relationships. We assume that $S(x) = 1 - F(x)$ and denote the derivative of $F(x)$ as $f(x)$ for $x > 0$. Thus, for $x > 0$, we have:

$$e^{-(\lambda_0+\lambda_1)x^\alpha} = F\left(e^{-\lambda_0 x^\alpha}\right) F\left(e^{-\lambda_1 x^\alpha}\right).$$

Now, if we introduce the transformation $y = e^{-x^\alpha}$, where $0 < y < 1$, we can express this relationship as:

$$y^{\lambda_0+\lambda_1} = F\left(y^{\lambda_0}\right) F\left(y^{\lambda_1}\right).$$

This leads us to the following crucial step:

$$\frac{F\left(y^{\lambda_0}\right)}{y^{\lambda_0}} \times \frac{F\left(y^{\lambda_1}\right)}{y^{\lambda_1}} = 1 \Rightarrow \frac{F(y)}{y} = 1 \Rightarrow F(y) = y.$$

With this result in hand, we proceed to present the joint distribution of X_n and X_{n+m}, where $m \geq 1$.

Theorem 3: If $\{X_n\}$ satisfies (3), then the joint survival function of X_n and X_{n+m}, $S_{n,n+m}(x,y) = P\left(X_n > x, X_{n+m} > y\right)$ is

$$S_{n,n+m}(x,y) = \begin{cases} e^{-(\lambda_0+\lambda_1)x^\alpha} e^{-(\lambda_0+\lambda_1)y^\alpha} & \text{if} \quad m \geq 2 \\ e^{-\lambda_1 x^\alpha} e^{-\lambda_0 y^\alpha} g(x,y) & \text{if} \quad m = 1, \end{cases} \tag{2.1}$$

where, $g(x,y) = \min\left\{e^{-\lambda_0 x^\alpha}, e^{-\lambda_1 y^\alpha}\right\}$. We just skip the proof in this paper. The aforementioned theorem reveals the interdependence between two consecutive terms, X_n and X_{n+m}, within the Weibull process. Specifically, they exhibit a dependence for $m = 1$, while for $m > 1$, they become statistically independent, thus resembling a lag-1 process. To delve deeper into the dependence properties of the Weibull process, it is crucial to examine the joint distribution function of X_n and X_{n+1}. So, the joint survival function takes the form:

$$S_{n,n+1}(x,y) = \begin{cases} e^{-\lambda_1 x^\alpha} e^{-(\lambda_0+\lambda_1)y^\alpha} & \text{if} \quad \lambda_0 x^\alpha < \lambda_1 y^\alpha \\ e^{-(\lambda_0+\lambda_1)x^\alpha} e^{-\lambda_0 y^\alpha} & \text{if} \quad \lambda_0 x^\alpha > \lambda_1 y^\alpha \\ e^{-z\frac{\lambda_1^2+\lambda_0^2+\lambda_0\lambda_1}{\lambda_0\lambda_1}} & \text{if} \quad \lambda_0 x^\alpha = \lambda_1 y^\alpha = z \end{cases} \tag{2.2}$$

These expressions provide a comprehensive understanding of the dependence properties of the Weibull process and enable a detailed analysis of the joint survival probabilities between two consecutive terms, X_n and X_{n+1}.

Therefore, if $\lambda_0 = \lambda_1 = \lambda$, then

$$S_{n,n+1}(x,y) = \begin{cases} e^{-\lambda x^\alpha} e^{-2\lambda y^\alpha} & \text{if} \quad x < y \\ e^{-2\lambda x^\alpha} e^{-\lambda y^\alpha} & \text{if} \quad x > y \\ e^{-3\lambda z^\alpha} & \text{if} \quad x = y = z \end{cases} \tag{2.3}$$

It may be mentioned that eq. (2.3) is the joint survival function of the Marshall-Olkin bivariate Weibull distribution, and its properties have been well studied in the literature. In the papers Kundu and Dey (2009) and Kundu and Gupta (2006) it is described with examples. It can be easily seen that the $S_{n,n+1}(x,y)$, for $0 < \delta < 1$, has the following survival copula function is,

$$\widetilde{C}(u,v) = \begin{cases} u^\delta v & \text{if} \quad u^\delta > v^{1-\delta} \\ uv^{1-\delta} & \text{if} \quad u^\delta \leq v^{1-\delta} \end{cases} \tag{2.4}$$

Then corresponding copula density function becomes,

$$c(u,v) = \frac{\partial^2}{\partial u \partial v}\widetilde{C}(u,v) = \begin{cases} \delta u^{\delta-1} & \text{if } u^\delta > v^{1-\delta} \\ (1-\delta)v^{-\delta} & \text{if} \quad u^\delta \leq v^{1-\delta} \end{cases} \tag{2.5}$$

So, Spearman's ρ and Kendall's τ can be obtained as based on the copula density function respectively is,

$$\begin{aligned}\rho &= \frac{3\delta(1-\delta)}{\delta^2-\delta+2}\\ \tau &= \frac{\delta(1-\delta)(1-\delta(1-\delta))}{\delta^3+\delta(1-\delta)+\delta^2\left(1-\delta^2\right)+(1-\delta)^3}\end{aligned} \tag{2.6}$$

We will now introduce the following regions,

$$\begin{aligned} S_1 &= \{(x,y); x>0, y>0, \beta x<y\}\\ S_2 &= \{(x,y); x>0, y>0, \beta x>y\}\\ C &= \{(x,y); x>0, y>0, \beta x=y\}\end{aligned} \tag{2.7}$$

Here, $\beta = (\lambda_0/\lambda_1)^{1/\alpha}$, and it may be noted that the curve C has the parametric form $(t, \gamma(t))$, for $0 < t < \infty$, where $\gamma(t) = \beta t$. The following results are needed for further development.

Theorem 1. *If X_n satistfies eq. (2.3) ,then joint survival function of X_n and X_{n+1} can be written as*

$$S_{n,n+1}(x,y) = pS_a(x,y) + (1-p)S_s(x,y) \tag{2.8}$$

here, $p = \frac{\lambda_0^2+\lambda_1^2}{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}$

$$S_s(x,y) = (g(x,y))^{\frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_0\lambda_1}} \tag{2.9}$$

and $S_a(x,y)$ can be obtained by subtraction, i.e.

$$S_a(x,y) = \begin{cases} \frac{1}{p}e^{-\lambda_1 x^\alpha}e^{-(\lambda_0+\lambda_1)y^\alpha} - \frac{1-p}{p}e^{-\frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_0}y^\alpha} & \textit{if } \beta x < y\\ \frac{1}{p}e^{-(\lambda_0+\lambda_1)x^\alpha}e^{-\lambda_0 y^\alpha} - \frac{1-p}{p}e^{-\frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_1}x^\alpha} & \textit{if } \beta x > y\end{cases} \tag{2.10}$$

We now present the joint probability density function (PDF) for the random variables X_n and X_{n+1}. Due to the Markov property, it becomes valuable to compute the joint PDF for the entire sequence $X_1, X_2, \ldots, X_n$. However, it's worth noting that, in this context, the joint distribution (survival) function does not conform to the classical definition of an absolutely continuous distribution function, which would be characterized by the two-dimensional

Lebesgue measure.

In this particular scenario, we must approach the issue of the joint PDF differently, much like the approach discussed in Bemis, Bain and Higgins (1972). Here, we define our dominating measure as the two-dimensional Lebesgue measure applied to the union of sets S_1 and S_2 complemented by a one-dimensional Lebesgue measure along the curve C. This unique dominating measure accommodates the intricate nature of our problem. This approach allows us to define the joint PDF within the context of the complex measures involved in our problem, ensuring an accurate representation of the interdependencies among the Weibull process random variables X_n and X_{n+1} while accounting for the non-standard distribution characteristics.Based on this specific dominating measure, we can express the joint PDF for X_n and X_{n+1}, where $x > 0$ and $y > 0$ as follows:

Theorem 2. *If X_n satisfies eq. (2.3) then the joint pdf of X_n and X_{n+1} is*

$$f_{n,n+1}(x,y) = \begin{cases} f_1(x,y) & \textit{if } \beta x < y \\ f_2(x,y) & \textit{if } \beta x > y \\ f_0(x) & \textit{if } \beta x = y \end{cases} \tag{2.11}$$

where,

$$\begin{aligned} f_1(x,y) &= f_{WE}\left(x;\alpha,\lambda_1\right) f_{WE}\left(y;\alpha,\lambda_0+\lambda_1\right), \\ f_2(x,y) &= f_{WE}\left(x;\alpha,\lambda_0+\lambda_1\right) f_{WE}\left(y;\alpha,\lambda_0\right), \\ f_0(x) &= \frac{\alpha\lambda_0}{\beta} x^{\alpha-1} e^{-\frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_1}x^{\alpha}}. \end{aligned} \tag{2.12}$$

The following conditional PDF will be very useful for prediction purposes.So,the conditional PDF of X_{n+1} given X_n can be written as,

$$f_{X_{n+1}|X_n=x}(y) = \begin{cases} \frac{\lambda_1}{\lambda_0+\lambda_1} e^{\lambda_0 x^{\alpha}} f_{WE}\left(y;\alpha,\lambda_0+\lambda_1\right) & \text{if } \beta x < y \\ f_{WE}\left(y;\alpha,\lambda_0\right) & \text{if } \beta x > y \\ \frac{\lambda_0}{\lambda_0+\lambda_1} e^{-\frac{\lambda_0^2}{\lambda_1}x^{\alpha}} & \text{if } \beta x = y \end{cases} \tag{2.13}$$

When $\alpha = 1$, then it is called the exponential process.So, when $\lambda_0 = \lambda_1 = \lambda$,then the conditional pdf of X_n and X_{n+1} for an Weibull process will be,

$$f_{X_{n+1}|X_n=x}(y) = \begin{cases} \frac{1}{2}e^{\lambda x^\alpha} f_{WE}(y;\alpha,2\lambda) & \text{if } x < y \\ f_{WE}(y;\alpha,\lambda) & \text{if } y > x \\ \frac{1}{2}e^{-\lambda x^\alpha} & \text{if } x = y \end{cases} \tag{2.14}$$

It follows that both for the Weibull and exponential processes,

$$P\left(X_n = X_{n+1}\right) = P\left(X_n < X_{n+1}\right) = P\left(X_n > X_{n+1}\right) = \frac{1}{3}. \tag{2.15}$$

8

Now,the joint PDF of $X_1, \ldots, X_n$ can be obtained as,

$$\begin{aligned} f_{X_1,\ldots,X_n}\left(x_1,\ldots,x_n\right) &= f_{X_n|X_{n-1},\ldots,X_1}\left(x_n\right) \times \ldots \times f_{X_2|X_1}\left(x_2\right) \times f_{X_1}\left(x_1\right) \\ &= f_{X_n|X_{n-1}}\left(x_n\right) \times \ldots \times f_{X_2|X_1}\left(x_2\right) \times f_{X_1}\left(x_1\right) \\ &= \frac{f_{X_{n-1},X_n}\left(x_{n-1},x_n\right)}{f_{X_{n-1}}\left(x_{n-1}\right)} \times \ldots \times \frac{f_{X_1,X_2}\left(x_1,x_2\right)}{f_{X_1}\left(x_1\right)} \times f_{X_1}\left(x_1\right) \\ &= \frac{\prod_{i=1}^{n-1} f_{X_i,X_{i+1}}\left(x_i,x_{i+1}\right)}{\prod_{i=2}^{n-1} f_{X_i}\left(x_i\right)} \end{aligned} \tag{2.16}$$

Let's now delve into the concept of stopping time. It's important to note that the concept of stopping time has been extensively explored in the field of time series analysis. For a comprehensive overview, one can refer to works such as Christensen (2012),Novikov and Shiryaev (2007) and the associated references.

For our discussion, consider a fixed real number denoted as L where $L > 0$. We introduce a new random variable, which we'll denote as N. This random variable is defined as follows, for any integer k greater than or equal to 1:

$$\{N = k\} \Leftrightarrow \{X_1 > L, \ldots, X_{k-1} > L, X_k \leq L\}. \tag{2.17}$$

In other words, the random variable N takes on the value k when we have observed a sequence of k consecutive terms in the Weibull process $(X_1, X_2, ...)$ where each term is greater than L, but the kth term is less than or equal to L. This provides a formal and mathematical way to describe the stopping time N in the context of the Weibull process. Then clearly, N is a stopping time. Now for $\lambda = \max\{\lambda_0, \lambda_1\}$, first we obtain,

$$\begin{aligned} P\left(X_1 > L, \ldots, X_k > L\right) &= P\left(U_0 < e^{-\lambda_1 L^\alpha}, U_1 < e^{-\lambda_0 L^\alpha}, \ldots U_{k-1} < e^{-\lambda_1 L^\alpha}, U_k < e^{-\lambda_0 L^\alpha}\right) \\ &= e^{-(\lambda_0+\lambda_1)L^\alpha} - e^{-(k-1)\lambda L^\alpha} \end{aligned} \tag{2.18}$$

$$P(N=1) = P\left(X_1 \le L\right) = 1 - e^{-(\lambda_0+\lambda_1)L^\alpha} \tag{2.19}$$

$$\begin{aligned} P(N=2) &= P\left(X_1 > L, X_2 \le L\right) = P\left(X_1 > L\right) - P\left(X_1 > L, X_2 > L\right) \\ &= e^{-(\lambda_0+\lambda_1)L^\alpha} - P\left(U_1 < e^{-\lambda_0 L^\alpha}, U_0 < e^{-\lambda_1 L^\alpha}, U_2 < e^{-\lambda_0 L^\alpha}, U_1 < e^{-\lambda_1 L^\alpha}\right) \\ &= e^{-(\lambda_0+\lambda_1)L^\alpha}\left(1 - e^{-\lambda L^\alpha}\right) \end{aligned} \tag{2.20}$$

$$\begin{aligned} P(N=3) &= P\left(X_1 > L, X_2 > L, X_3 \le L\right) = P\left(X_1 > L, X_2 > L\right) - P\left(X_1 > L, X_2 > L, X_3 > L\right) \\ &= e^{-(\lambda_0+\lambda_1)L^\alpha} e^{-\lambda L^\alpha}\left(1 - e^{-\lambda L^\alpha}\right) \end{aligned} \tag{2.21}$$

$$P(N=k) = e^{-(\lambda_0+\lambda_1)L^\alpha} e^{-(k-2)\lambda L^\alpha}\left(1 - e^{-\lambda L^\alpha}\right). \tag{2.22}$$

The probability generating function of N can be obtained as,

$$\begin{aligned} G(z) = E\left(z^N\right) &= \sum_{k=1}^{\infty} P(N=k) z^k \\ &= \frac{z\left(1 - e^{-(\lambda_0+\lambda_1)L^\alpha}\right) + z^2\left(e^{-(\lambda_0+\lambda_1)L^\alpha} - e^{-\lambda L^\alpha}\right)}{\left(1 - z e^{-\lambda L^\alpha}\right)} \end{aligned} \tag{2.23}$$

Utilizing the probability generating function, one can readily derive various moments and other essential properties.

2.2. Maximum Likelihood Estimators:

Let us consider, the maximum likelihood estimators of the unknown parameters of a Weibull process based on a sample of size n.So, $x_1, x_1, ..., x_n$ is a sample values.

CASE - I: $\lambda_0 = \lambda_1$

In this case it is assumed that $\lambda_0 = \lambda_1 = \lambda$. Our problem is to estimate α and λ based on $\mathcal{D} = \{x_1, \ldots, x_n\}$. We use the following notations: $I = \{1, \ldots, n-1\}$ $I_1 = \{i : i \in I, x_i < x_{i+1}\}, I_2 = \{i : i \in I, x_i > x_{i+1}\}, I_0 = \{i : i \in I, x_i = x_{i+1}\}$ The number of elements in I_0, I_1 and I_2 are denoted by n_0, n_1 and n_2, respectively. Based on the joint PDF (16), the log-likelihood function can be written as,

$$
\begin{aligned}
l(\alpha, \lambda \mid \mathcal{D}) &= \sum_{i \in I_0 \cup I_1 \cup I_2} \ln f_{X_i, X_{i+1}}(x_i, x_{i+1}) - \sum_{i=2}^{n} \ln f_{X_i}(x_i) \\
&= (n_1 + n_2 + 1) \ln \alpha + (n_1 + n_2 + 1) \ln \lambda + (\alpha - 1) \left(\ln x_1 + \sum_{i \in I_1 \cup I_2} \ln x_{i+1} \right) - \lambda g_1(\alpha \mid \mathcal{D})
\end{aligned} \tag{2.24}
$$

where,

$$
g_1(\alpha \mid \mathcal{D}) = \sum_{i \in I_1} \left(x_i^\alpha + 2x_{i+1}^\alpha \right) + \sum_{i \in I_2} \left(2x_i^\alpha + x_{i+1}^\alpha \right) + 3 \sum_{i \in I_0} x_i^\alpha - 2 \sum_{i=2}^{n-1} x_i^\alpha \tag{2.25}
$$

It is immediate that for any $\alpha, g_1(\alpha \mid \mathcal{D}) > 0$. Hence, for a given α, the MLE of λ, say $\widehat{\lambda}(\alpha)$ can be obtained as,

$$
\widehat{\lambda}(\alpha) = \frac{n_1 + n_2 + 1}{g_1(\alpha \mid \mathcal{D})} \tag{2.26}
$$

and the MLE of α, say $\widehat{\alpha}$ can be obtained by maximizing

$$\begin{aligned} h(\alpha) = &(n_1 + n_2 + 1) \ln \alpha + (n_1 + n_2 + 1) \left(\ln (n_1 + n_2 + 1) - \ln g(\alpha \mid \mathcal{D}) \right) \\ &+ \alpha \left(\ln x_1 + \sum_{i \in I_1 \cup I_2} \ln x_{i+1} \right). \end{aligned} \tag{2.27}$$

After obtaining the estimate $\widehat{\alpha}$, the maximum likelihood estimate (MLE) of λ, denoted as $\widehat{\lambda}$, can be determined as a function of $\widehat{\alpha}$. Despite the intricate nature of the function $h(\alpha)$, establishing its unimodal character is a challenging task. However, in our data analysis, we have observed that $h(\alpha)$ exhibits unimodal behavior, as we will explain further. In the context of an exponential process, the MLE of λ can be obtained as,

$$\widehat{\lambda} = \frac{n_1 + n_2 + 1}{g_1(1 \mid \mathcal{D})} \tag{2.28}$$

CASE 2: $\lambda_0 \neq \lambda_1$

In this section we consider the case when λ_0 and λ_1 are arbitrary. We use the following notations

$$\begin{aligned} I_1(\beta) &= \{i : i \in I, \beta x_i < x_{i+1}\} \\ I_2(\beta) &= \{i : i \in I, \beta x_i > x_{i+1}\} \\ I_0(\beta) &= \{i : i \in I, \beta x_i = x_{i+1}\} \end{aligned} \tag{2.29}$$

and $n_0(\beta) = |I_0(\beta)|, n_1(\beta) = |I_1(\beta)|$ and $n_2(\beta) = |I_2(\beta)|$. Here, β is same as defined before.

Based on the data vector $\mathcal{D}$, the log-likelihood function of λ_0, λ_1 and α becomes

$$
\begin{aligned}
l\left(\lambda_0, \lambda_1, \alpha \mid \mathcal{D}\right) = & \left(n_1(\beta)+n_2(\beta)+1\right) \ln \alpha + n_1(\beta) \ln \lambda_1 + \left(n_2(\beta)+n_0(\beta)\right) \ln \lambda_0 + \\
& \left(n_1(\beta)+n_2(\beta)+2-n\right) \ln \left(\lambda_0+\lambda_1\right) + (\alpha-1)\left\{\ln x_1 + \sum_{i \in I_1(\beta) \cup I_2(\beta)} \ln x_{i+1}\right\} - \\
& \lambda_1 \sum_{i \in I_1(\beta)} x_i^\alpha - \left(\lambda_0+\lambda_1\right) \sum_{i \in I_2(\beta)} x_i^\alpha \frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_1} \sum_{i \in I_0(\beta)} x_i^\alpha - \\
& \left(\lambda_0+\lambda_1\right) \sum_{i \in I_1(\beta)} x_{i+1}^\alpha - \lambda_0 \sum_{i \in I_2(\beta)} x_{i+1}^\alpha - n_0(\beta) \ln \beta + \left(\lambda_0+\lambda_1\right) \sum_{i=2}^{n-1} x_i^\alpha .
\end{aligned} \tag{2.30}
$$

It is not very easy to maximize the eq. (2.3) in a usual method. So, we need to do re-parameterization.We use the following transformed parameters, $(\gamma, \lambda_1, \alpha)$, where $\gamma = \frac{\lambda_0}{\lambda_1}$ and based on the transformed parameters, the log-likelihood function can be written as,

$$
\begin{aligned}
l\left(\gamma, \lambda_1, \alpha\right) = & \left(n_1(\beta)+n_2(\beta)+1\right)\left(\ln \lambda_1 + \ln \alpha\right) - \lambda_1 g_2(\alpha, \gamma \mid \mathcal{D}) - \left(n_0(\beta)-1\right) \ln(1+\gamma) + \\
& \left(\left(n_0(\beta) - \frac{1}{\alpha}\right) + n_2(\beta)\right) \ln(\gamma) + (\alpha-1) h_2(\mathcal{D})
\end{aligned} \tag{2.31}
$$

where,

$$
\begin{aligned}
g_2(\alpha, \gamma \mid \mathcal{D}) = & \sum_{i \in I_1(\beta)} x_i^\alpha + (1+\gamma)\left(\sum_{i \in I_1(\beta)} x_{i+1}^\alpha + \sum_{i \in I_2(\beta)} x_i^\alpha\right) + \gamma \sum_{i \in I_2(\beta)} x_{i+1}^\alpha \\
& + \left(\gamma^2+\gamma+1\right) \sum_{i \in I_0(\beta)} x_i^\alpha - (1+\gamma) \sum_{i=2}^{n-1} x_i^\alpha \qquad (2.32) \\
h_2(\mathcal{D}) = & \ln x_1 + \sum_{i \in I_1(\beta) \cup I_2(\beta)} \ln x_{i+1} .
\end{aligned}
$$

So, now we will use the profile likelihood method to maximize (22). For fixed γ and α (β is also fixed in that case), first, we maximize (22) with respect to λ_1, say $\widehat{\lambda}_1(\gamma, \alpha)$, and it can be obtained in explicit form as,

$$
\widehat{\lambda}_1(\gamma, \alpha) = \frac{n_1(\beta)+n_2(\beta)+1}{g_2(\alpha, \gamma \mid \mathcal{D})} \tag{2.33}
$$

The MLEs of γ and α, say $\widehat{\gamma}$ and $\widehat{\alpha}$, respectively, can be obtained by maximizing $l\left(\gamma, \widehat{\lambda}_1(\gamma, \alpha), \alpha\right)$. Finally, the MLE of λ_1 can be obtained as $\widehat{\lambda}_1(\widehat{\gamma}, \widehat{\alpha})$. We will denote this as $\widehat{\lambda}_1$. Due to the complicated nature of the function $g_2(\alpha, \gamma \mid \mathcal{D})$, it is difficult to prove that it has a unique maximum. But in our data analysis it is observed from the contour plot that $g_2(\alpha, \gamma \mid \mathcal{D})$ is an unimodal function. In both cases we have suggested parametric bootstrap method to construct confidence intervals of the unknown parameters. They can be very easily implemented in practice. In case of exponential process, the MLE of λ_1 for a given γ can be obtained as,

$$\widehat{\lambda}_1(\gamma) = \frac{n_1(\beta) + n_2(\beta) + 1}{g_2(1, \gamma \mid \mathcal{D})}, \tag{2.34}$$

and the MLE of γ can be obtained by maximizing$l\left(\gamma, \widehat{\lambda}_1(\gamma), 1\right)$. It is a one-dimensional optimization problem.

2.3. Goodness of fit:

In this section, we provide a goodness of fit test, so that whether a given data set comes from a Weibull process or not can be tested. Suppose $\{x_1, \ldots, x_n\}$ is a sample from a stationary sequence $\{X_1, \ldots, X_n\}$. We want to test the following null hypothesis

$$H_0 : \{X_1, \ldots, X_n\} \sim \text{WEP}\left(\alpha, \lambda_0, \lambda_1\right) \tag{2.35}$$

Let us use the following notations. We denote $X_{1:n} < \ldots < X_{n:n}$ as the ordered $\{X_1, \ldots, X_n\}$, similarly, $x_{1:n} < \ldots < x_{n:n}$ as the ordered $\{x_1, \ldots, x_n\}$, and $a_1 = E_{H_0}\left(X_{1:n}\right), \ldots, a_n = E_{H_0}\left(X_{n:n}\right)$ as their ordered expected values under H_0. Here, $a_1, \ldots, a_n$ depend on $\alpha, \lambda_0, \lambda_1$, but we do not make it explicit. We use the following statistic for goodness of fit test.

$$W_n = \max_{1 \le i \le n} \left| X_{i:n} - a_i \right|. \tag{2.36}$$

It is expected that if H_0 is true, then W_n should be small. Hence, we use the following test criterion for a given level of significance $0 < \beta < 1$ Reject H_0 if $W_n > c_n(\beta)$,where $c(\beta)$ is such that

$$P_{H_0}\left(W_n > c_n(\beta) = \beta.\right. \tag{2.37}$$

Note that $c_n(\beta)$ also depends on $\alpha, \lambda_0, \lambda_1$, but we do not make it explicit for brevity. It is difficult to obtain $c_n(\beta)$ theoretically even for large n. Hence, we will use the parametric bootstrap technique to approximate $c_n(\beta)$ from a given observed sample $\{x_1, \ldots, x_n\}$.

Algorithm 1 Algorithm for Parametric Bootstrap

1: Obtain $\widehat{\alpha}, \widehat{\lambda}_0, \widehat{\lambda}_1$, the MLEs of $\alpha, \lambda_0, \lambda_1$, respectively, based on $\{x_1, \ldots, x_n\}$

2: Generate a sample of size n from a WEP$\left(\widehat{\alpha}, \widehat{\lambda}_0, \widehat{\lambda}_1\right)$, order them. Let us denote them as $(x^1_{1:n}, \ldots, x^1_{n:n})$. Repeat this procedure B times, and obtain $\left\{\left(x^b_{1:n}, \ldots, x^b_{n:n}\right); b = 1, \ldots, B\right\}$.

3: Obtain estimates of $a_1, \ldots, a_n$ as

$$\widehat{a}_i = \frac{1}{B}\sum_{b=1}^{B} x^b_{i:n}; \quad i = 1, \ldots, n$$

4: Compute

$$w^b = \max_{1 \le i \le n}\left\{\left|x^b_{i:n} - \widehat{a}_i\right|\right\}; \quad b = 1, \ldots, B$$

5: Order $\left\{w^1, \ldots, w^B\right\}$ as $w^{(1)} < \ldots < w^{(B)}$, then $\widehat{c}_n(\beta) = w^{[(100(1-\beta))]}$ is an estimate of $c_n(\beta)$.

Now, if $w_n = \max_{1 \le i \le n}|x_{i:n} - \widehat{a}_i| > \widehat{c}_n(\beta)$, then we reject the null hypothesis with β% level of significance, otherwise we accept the null hypothesis.

3. Data and exploratory analysis:

3.1. Synthetic Data 1:

This synthetic dataset has been generated using the following model specification: $\alpha = 2.0, \lambda_0 = \lambda_1 = 1.0, n = 75$. The data set is presented in Figure 1. Based on the profile maximization we obtain the MLEs of α and λ as $\widehat{\alpha} = 1.889$ and $\widehat{\lambda} = 1.025$. The associate 95% confidence intervals are $(1.743, 2.056)$ and $(0.863, 1.264)$, respectively.

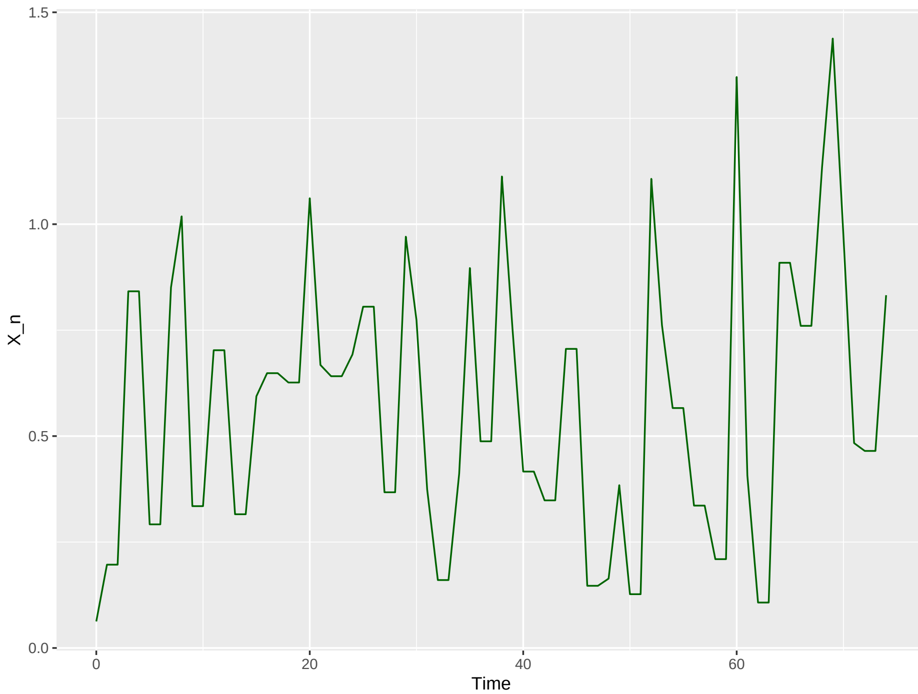

FIG 1. *Synthetic data set with* $\alpha = 2.0, \lambda_0 = \lambda_1 = 1$

3.1.1. Simulation Study:

Now we have done a simulation study and find the MLEs of α and λ for that model specification where $\lambda_0 = \lambda_1 = \lambda$ for different sample size i.e for $n = 75, 100, 125, 150, 175, 200$ in such a manner that we first draw a sample from the model specification where $\alpha = \widehat{\alpha}$ and $\lambda = \widehat{\lambda}$.Then we find MLEs of α and λ from here and again use those MLEs of α and λ in the model specification to generate the dataset. Then, We repeat this procedure 1000 times where we find MLEs of α and λ at each time and then use those MLEs in the model specification to generate another dataset and use that new dataset to find the MLEs of α and λ again.We have repeated this procedure for different sample sizes i.e for $n = 75, 100, 125, 150, 175, 200$. Then we plotted the histogram of MLEs of α and λ for different sample sizes in Figure 2 and Figure 3.

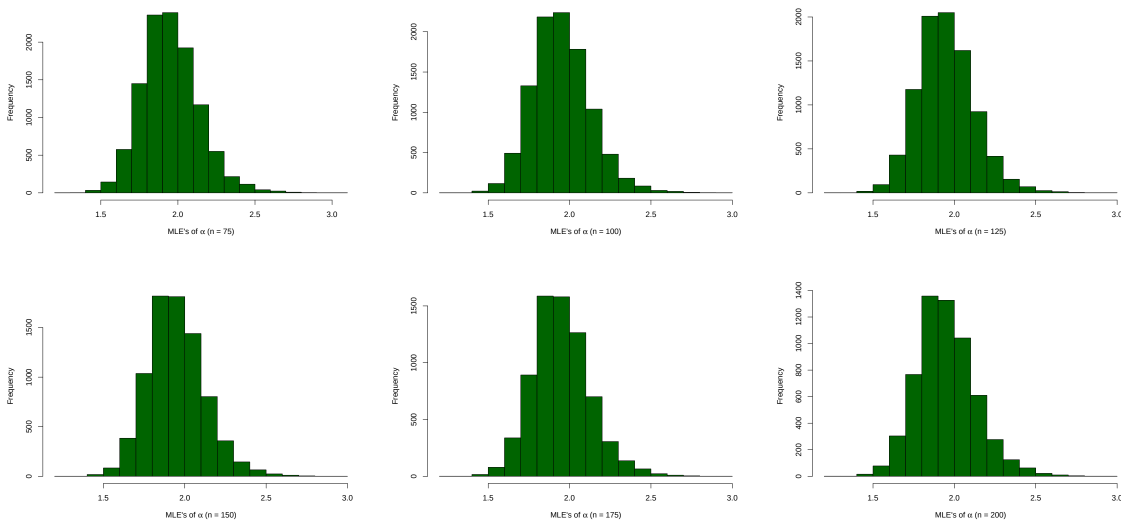

FIG 2. *Histograms of MLE of α*

3.2. Synthetic Data 2:

It has been generated using the following model specification: $\alpha = 3.0, \lambda_0 = 0.15, \lambda_1 = 0.04$ and $n = 75$. The data set is presented in Figure 4. Based on the profile maximization we obtain the MLEs of α, λ_0 and λ_1 become $\widehat{\alpha} = 3.295, \widehat{\lambda_0} = 0.1429$ and $\widehat{\lambda_1} = 0.024$. The associated 95% bootstrap confidence intervals become $(2.876, 3.954), (0.124, 0.152)$ and $(0.014, 0.035)$, respectively.

3.2.1. Simulation Study:

Now we have done a simulation study and find the MLEs of α, λ_0 λ_1 for that model specification where λ_0 and λ_1 are different for different sample sizes i.e for $n = 50, 75, 100, 125, 150, 175,$ 200 in such a manner that we first draw a sample from the model specification where $\alpha = \widehat{\alpha}$,$\lambda_0 = \widehat{\lambda_0}$ and $\lambda_1 = \widehat{\lambda_1}$.Then we find MLEs of α,λ_0,λ_1 from here and again use those MLEs of α,λ_0 and λ_1 in the model specification to generate the dataset. Then, We repeat this procedure 1000 times where we find MLEs of α and λ at each time and then use those MLEs in the model specification to generate an another dataset and use that new dataset

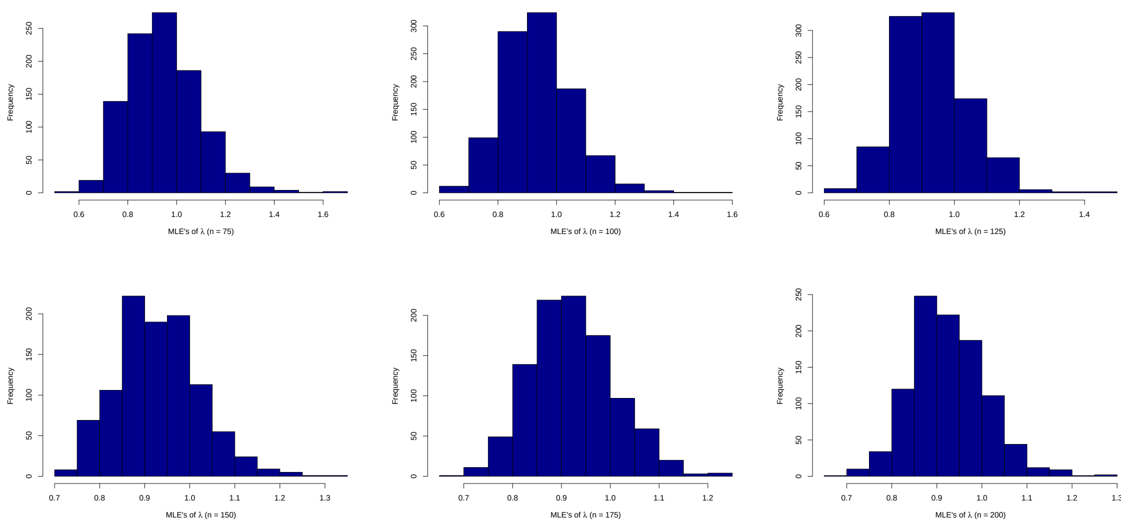

FIG 3. *Histograms of MLE of λ*

to find the MLEs of α,$\lambda_0$andλ_1 again.We have repeated this procedure for different sample sizes i.e. for $n = 50, 75, 100, 125, 150, 175, 200$. Now, we will plot the histogram of MLEs of α, λ_0 and λ_1 for different sample sizes in Figure 5, Figure 6 and Figure 7.

4. Creating R Package:

Our main goal was to make an R-Package where the user will provide time-series data and our function in the package will implement the dataset in the two above-mentioned models and will find MLEs of α and λ for the model with specific settings that $\lambda_0 = \lambda_1 = \lambda$ and also will find the MLE's of α, λ_0 and λ_1 for the model with specific setting where λ_0 is not equal to λ_1. Then it will generate test statistic using the parametric bootstrap method and will give histograms for both models. Ultimately, it will provide the test statistics which is used to test whether the data is coming from model 1 or it is coming from model 2. Then it will provide the corresponding p-values to measure how good the fit is. We made a GitHub repository of our R-package.The name of our package is **stnweib**.

Github link for our R-package: https://github.com/debarghya2000/stnweib

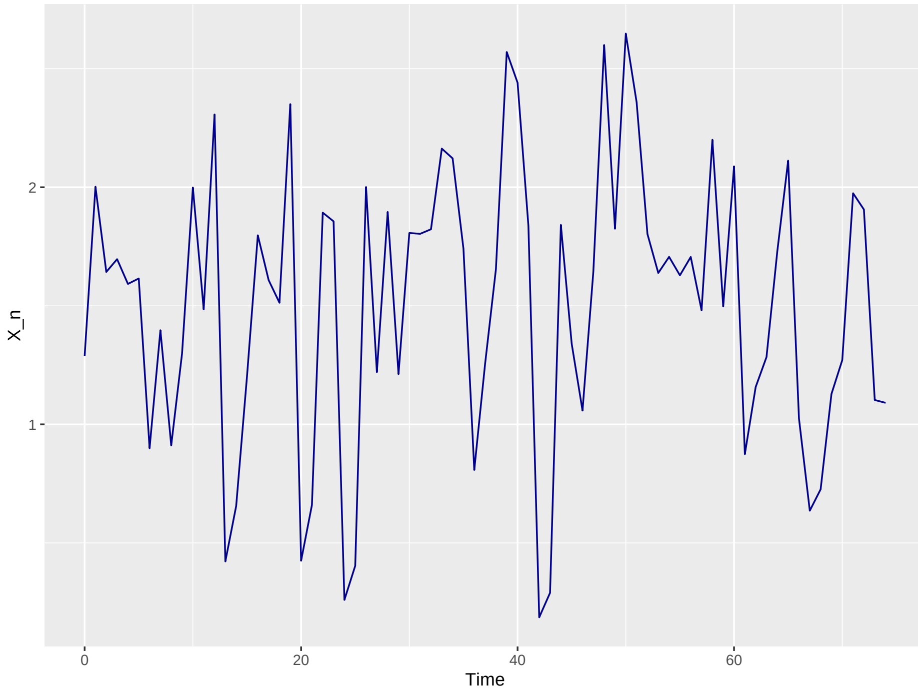

FIG 4. *Figure 4: Synthetic data set with* $\alpha = 3.0, \lambda_0 = 0.15$ *and* $\lambda_1 = 0.04$.

5. Conclusion:

In this paper, we introduce a novel discrete-time, continuous-state-space stochastic process based on Weibull distributions. A distinctive feature of this proposed process is the incorporation of a positive probability that X_n equals X_{n+1} for some n, making it particularly well-suited for scenarios involving ties between two consecutive time points. We have conducted a comprehensive study of various properties of this novel process and have provided inference procedures for estimating the unknown parameters.

It is worth noting that while the proposed stochastic process is a lag-1 process, it can be readily extended to a lag-q process. This extension can be achieved as follows: Assuming that $U_0, U_1, U_2, \ldots$ are independently and identically distributed (i.i.d.) uniform $U(0,1)$ random variables, and given $\alpha > 0, \lambda_0 > 0, \lambda_1, \ldots, \lambda_q > 0$, we define:

$$X_n = \min\left\{\left[-\frac{1}{\lambda_0}\ln U_n\right]^{\frac{1}{\alpha}}, \ldots, \left[-\frac{1}{\lambda_q}\ln U_{n-q}\right]^{\frac{1}{\alpha}}\right\} \tag{5.1}$$

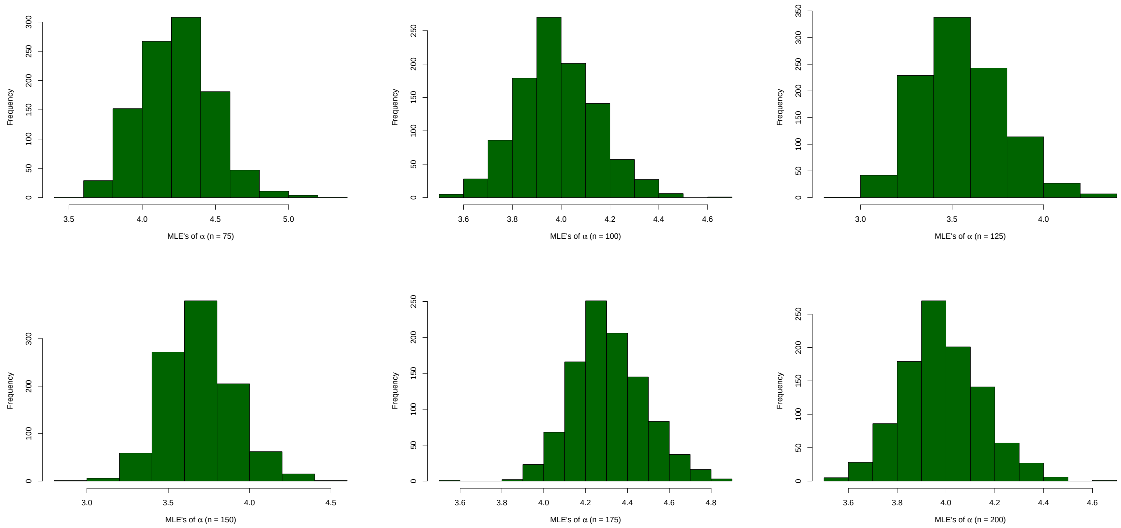

FIG 5. *Histograms of MLE of α*

This formulation results in a lag-q stationary Weibull process. Importantly, it can be easily verified that there is a positive probability that $X_n = X_{n+1} = \ldots = X_{n+k}$ for some n and for values of k ranging from 1 to q. Additionally, this extended process exhibits a convenient copula structure. Future work in this direction may involve the exploration of further properties and the development of classical methodologies related to this lag-q stationary Weibull process. So, more work is genuinely needed in this direction.

6. Appendix I: Proofs

Proof of Theorem 4: Note that p and $S_a(x, y)$ can be obtained from $S_{n,n+1}(x, y)$ as follows:

$$p = \int_0^\infty \int_0^\infty \frac{\partial^2 S_{n,n+1}(x, y)}{\partial x \partial y} dx dy$$

and

$$p S_a(x, y) = \int_y^\infty \int_x^\infty \frac{\partial^2 S_{n,n+1}(u, v)}{\partial u \partial v} du dv$$

Now, from

$$\frac{\partial^2 S_{n,n+1}(x, y)}{\partial x \partial y} = \begin{cases} f_1(x, y) & \text{if} \quad (x, y) \in S_1 \\ f_2(x, y) & \text{if} \quad (x, y) \in S_2 \end{cases}$$

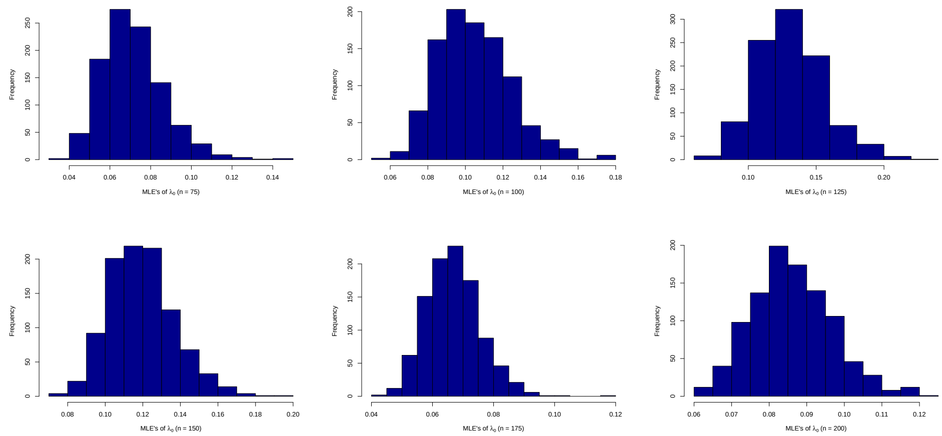

FIG 6. *Histograms of MLE of λ_0*

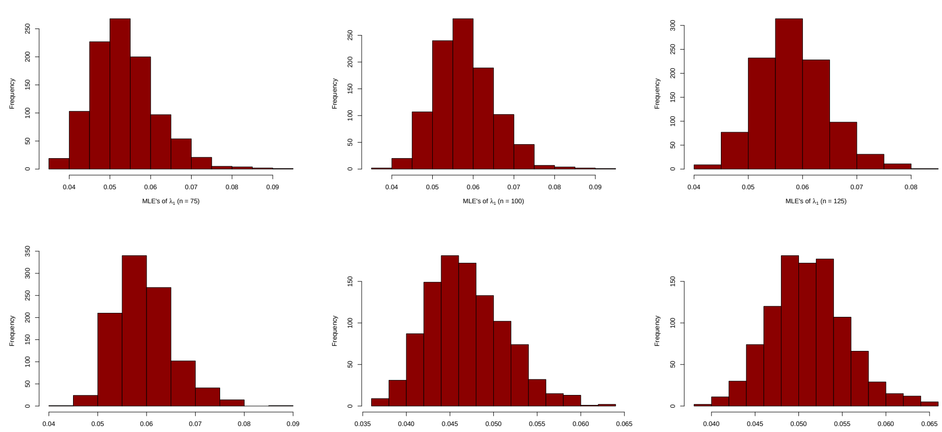

FIG 7. *Histograms of MLE of λ_1*

where

$$f_1(x,y) = f_{WE}\left(x;\alpha,\lambda_1\right) f_{WE}\left(y;\alpha,\lambda_0+\lambda_1\right)$$

$$f_2(x,y) = f_{WE}\left(x;\alpha,\lambda_0+\lambda_1\right) f_{WE}\left(y;\alpha,\lambda_0\right).$$

Since

$$\int_0^\infty \int_{\beta x}^\infty f_1(x,y)dydx = \frac{\lambda_1^2}{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1} \quad \text{and} \int_0^\infty \int_{y/\beta}^\infty f_2(x,y)dxdy = \frac{\lambda_0^2}{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1},$$

$$p = \frac{\lambda_0^2+\lambda_1^2}{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}.$$

Using this $p, S_a(x,y)$ can be obtained by simple integration, and after that $S_s(x,y)$ can be obtained by subtraction.

Alternatively, a simple probabilistic arguments also can be given as follows. Suppose A is the following event

$$A = \left\{\left[-\frac{1}{\lambda_0}\ln U_n\right] < \left[-\frac{1}{\lambda_1}\ln U_{n-1}\right]\right\} \cap \left\{\left[-\frac{1}{\lambda_1}\ln U_n\right] < \left[-\frac{1}{\lambda_0}\ln U_{n+1}\right]\right\}$$

then

$$P(A) = P\left(U_n > U_{n-1}^{\frac{\lambda_0}{\lambda_1}}, U_n > U_{n+1}^{\frac{\lambda_1}{\lambda_0}}\right) = \int_0^1 u^{\frac{\lambda_0}{\lambda_1}+\frac{\lambda_1}{\lambda_0}} du = \frac{\lambda_0\lambda_1}{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1} = 1-p.$$

Moreover,

$$P\left(X_n > x, X_{n+1} > y\right) = P\left(X_n > x, X_{n+1} > y \mid A\right) P(A) + P\left(X_n > x, X_{n+1} > y \mid A'\right) P\left(A'\right),$$

and

$$P\left(X_n > x, X_{n+1} > y \mid A\right) = P\left(U_n < e^{-\lambda_0 x^\alpha}, U_n < e^{-\lambda_1 y^\alpha}\right) = g(x,y).$$

The rest can be obtained by subtraction. Proof of Theorem 5: We need to show that that for all $0 < x, y < \infty$,

$$S_{n,n+1}(x,y) = \iint_{B_1} f_1(u,v)dudv + \iint_{B_2} f_2(u,v)dudv + \int_{h(x,y)}^\infty f_0(t)\left|\gamma'(t)\right| dt$$

here for $R(x,y) = \{(u,v); x \leq u < \infty, y \leq v < \infty\}, B_1 = R(x,y) \cap S_1, B_2 = R(x,y) \cap S_2$, and $h(x,y) = \max\left\{x, \frac{y}{\beta}\right\}$. It has already been shown in Theorem 4 that

$$\iint_{B_1} f_1(u,v)dudv + \iint_{B_2} f_2(u,v)dudv = pS_a(x,y)$$

hence, the result is proved if we can show

$$\int_{h(x,y)}^{\infty} f_0(t)\left|\gamma'(t)\right| dt = (1-p)S_s(x,y)$$

Since, $|\gamma'(t)| = \beta$ and $(1-p) = \frac{\lambda_0\lambda_1}{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}$,

$$\int_{h(x,y)}^{\infty} f_0(t)\left|\gamma'(t)\right| dt = \int_{h(x,y)} \alpha\lambda_0 t^{\alpha-1} e^{-\frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_1}t^{\alpha}} dt = (1-p)\int_{v(x,y)} e^{-u} du$$

where $v(x,y) = \max\left\{\frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_1}x^{\alpha}, \frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_0}y^{\alpha}\right\}$. Let us remember,

$$S_s(x,y) = \begin{cases} e^{-\frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_0}y^{\alpha}} & \text{if } y > \beta x \\ e^{-\frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_1}x^{\alpha}} & \text{if } y < \beta x. \end{cases}$$

Hence, the result follows.

7. Appendix II: Autocovariance and Autocorrelation Functions

In this section we provide all the expressions of the autocorrelation function of the GE process mainly for completeness purposes. First we will calculate $E\left(X_{n+1}X_n\right)$. If I_A denote the indicator function on the set A, then

$$
\begin{aligned}
E\left(X_{n+1}X_n\right) =& E\left(X_{n+1}X_n \cdot I_{\{\beta X_n<X_{n+1}\}}\right) + E\left(X_{n+1}X_n \cdot I_{\{\beta X_n>X_{n+1}\}}\right) + \\
& E\left(X_{n+1}X_n \cdot I_{\{\beta X_n=X_{n+1}\}}\right) \\
=& E_{X_n}\left(X_n E_{\{X_{n+1}|X_n\}}\left(X_{n+1} \cdot I_{\{\beta X_n<X_{n+1}\}} \mid X_n\right) + \right. \\
& E_{X_n}\left(X_n E_{\{X_{n+1}|X_n\}}\left(X_{n+1} \cdot I_{\{\beta X_n>X_{n+1}\}} \mid X_n\right) + \right. \\
& E_{X_n}\left(X_n E_{\{X_{n+1}|X_n\}}\left(X_{n+1} \cdot I_{\{\beta X_n=X_{n+1}\}} \mid X_n\right) . \right.
\end{aligned} \tag{7.1}
$$

Now,

$$
\begin{aligned}
E_{X_n}\left(X_n E_{\{X_{n+1}|X_n\}}\left(X_{n+1} \cdot I_{\{\beta X_n=X_{n+1}\}} \mid X_n\right)\right. &= \alpha\beta\lambda_0 \int_0^\infty x^{\alpha+1} e^{-\frac{\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1}{\lambda_1}x^\alpha} dx \\
&= \Gamma\left(\frac{2}{\alpha}+1\right) \frac{\left(\lambda_0\lambda_1\right)^{1+\frac{1}{\alpha}}}{\left(\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1\right)^{\frac{2}{\alpha}+1}}
\end{aligned} \tag{7.2}
$$

If we denote $\Gamma(x,a) = \int_x^\infty t^{a-1}e^{-t}dt$ and $\gamma(x,a) = \int_0^x t^{a-1}e^{-t}dt$ as incomplete gamma functions, then

$$
\begin{aligned}
E_{X_n}\left(X_n E_{\{X_{n+1}|X_n\}}\left(X_{n+1} \cdot I_{\{\beta X_n<X_{n+1}\}} \mid X_n\right)\right. &= \alpha^2\lambda_1 \int_0^\infty x^\alpha e^{-\lambda_1 x^\alpha} \left\{\int_{\beta x}^\infty y^\alpha e^{-(\lambda_0+\lambda_1)y^\alpha} dy\right\} dx \\
&= \frac{\alpha\lambda_1}{\left(\lambda_0+\lambda_1\right)^{1/\alpha}} \int_0^\infty x^\alpha e^{-\lambda_1 x^\alpha} \Gamma\left(\frac{\lambda_0\left(\lambda_0+\lambda_1\right)x^\alpha}{\lambda_1}, \frac{1}{\alpha}+1\right) dx \\
E_{X_n}\left(X_n E_{\{X_{n+1}|X_n\}}\left(X_{n+1} \cdot I_{\{\beta X_n>X_{n+1}\}} \mid X_n\right)\right. &= \alpha^2\lambda_0\left(\lambda_0+\lambda_1\right) \int_0^\infty x^\alpha e^{-(\lambda_0+\lambda_1)x^\alpha} \left\{\int_0^{\beta x} y^\alpha e^{-\lambda_0 y^\alpha} dy\right\} dx \\
&= \frac{\alpha\left(\lambda_0+\lambda_1\right)}{\lambda_0^{1/\alpha}} \int_0^\infty x^\alpha e^{-(\lambda_0+\lambda_1)x^\alpha} \gamma\left(\frac{\lambda_0^2 x^\alpha}{\lambda_1}, \frac{1}{\alpha}+1\right) dx.
\end{aligned} \tag{7.3}
$$

We have already indicated the mean and variance of a Weibull random variable in (2). Now based on the above expressions, the autocovariance and autocorrelation functions can be obtained. In case of exponential process i.e. when $\alpha = 1$, the above expressions can be obtained in explicit forms. For example

$$
\begin{aligned}
&E_{X_n}\left(X_n E_{\{X_{n+1}|X_n\}}\left(X_{n+1}\cdot I_{\{\beta X_n = X_{n+1}\}} \mid X_n\right) = \frac{2\left(\lambda_0\lambda_1\right)^2}{\left(\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1\right)^3}\right.\\
&E_{X_n}\left(X_n E_{\{X_{n+1}|X_n\}}\left(X_{n+1}\cdot I_{\{\beta X_n < X_{n+1}\}} \mid X_n\right) =\right.\\
&\frac{\lambda_1}{\lambda_0+\lambda_1}\int_0^\infty x e^{-\lambda_1 x}\left(e^{-\frac{\lambda_0(\lambda_0+\lambda_1)}{\lambda_1}x} + \frac{\lambda_0\left(\lambda_0+\lambda_1\right)}{\lambda_1} x e^{-\frac{\lambda_0(\lambda_0+\lambda_1)}{\lambda_1}x}\right) dx =\\
&\frac{\lambda_1^3}{\left(\lambda_0+\lambda_1\right)\left(\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1\right)^2} + \frac{2\lambda_0\lambda_1^3}{\left(\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1\right)^3}\\
&E_{X_n}\left(X_n E_{\{X_{n+1}|X_n\}}\left(X_{n+1}\cdot I_{\{\beta X_n > X_{n+1}\}} \mid X_n\right) =\right.\\
&\quad\frac{\lambda_0+\lambda_1}{\lambda_0}\int_0^\infty x e^{-(\lambda_0+\lambda_1)x}\left(1 - e^{-\frac{\lambda_0^2}{\lambda_1}x} - \frac{\lambda_0^2}{\lambda_1} x e^{-\frac{\lambda_0^2}{\lambda_1}x}\right) dx =\\
&\frac{1}{\lambda_0\left(\lambda_0+\lambda_1\right)} - \frac{\lambda_1^2\left(\lambda_0+\lambda_1\right)}{\lambda_0\left(\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1\right)^2} - \frac{2\lambda_0\lambda_1^2\left(\lambda_0+\lambda_1\right)}{\left(\lambda_0^2+\lambda_1^2+\lambda_0\lambda_1\right)^3}.
\end{aligned}
\tag{7.4}
$$

Since,

$$
E\left(X_n\right) = E\left(X_{n+1}\right) = \frac{1}{\lambda_0+\lambda} \quad \text{and} \quad V\left(X_n\right) = V\left(X_{n+1}\right) = \frac{1}{\left(\lambda_0+\lambda\right)^2}, \tag{7.5}
$$

The autocovariance and autocorrelation can be obtained in explicit forms.

8. Appendix III: R-Package test run with several data sets.

When $\alpha = 3$, $\lambda_0 = 0.15$, $\lambda_1 = 0.04$ (Figure 8) then,

R-package test for WEP(n,α, λ_0, λ_1)						
	$n = 75$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$
Test Statistic	1.27073e-01	1.50618e-01	1.430506e-01	9.76388e-02	9.91640e-02	8.52095e-02
P-value	0.953	0.863	0.832	0.968	0.957	0.982

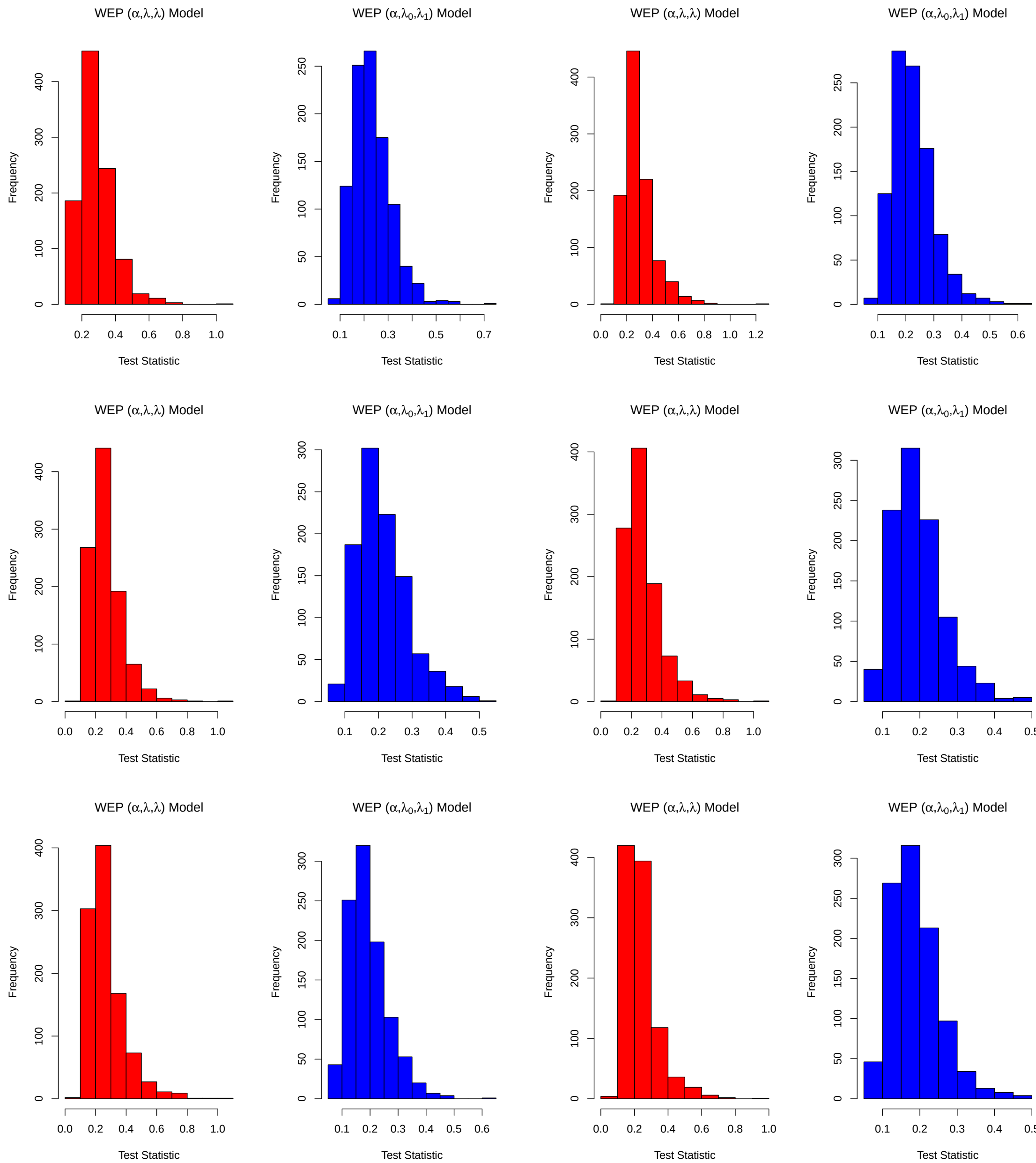

FIG 8. *Histograms of Test Statistic*

R-package test for WEP(n,α, λ, λ)						
	$n = 75$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$
Test Statistic	1.28159e-01	1.58759e-01	1.02728e-01	1.8353e-01	1.4329e-01	1.8492e-01
P-value	0.024	0.037	0.026	0.021	0.031	0.042

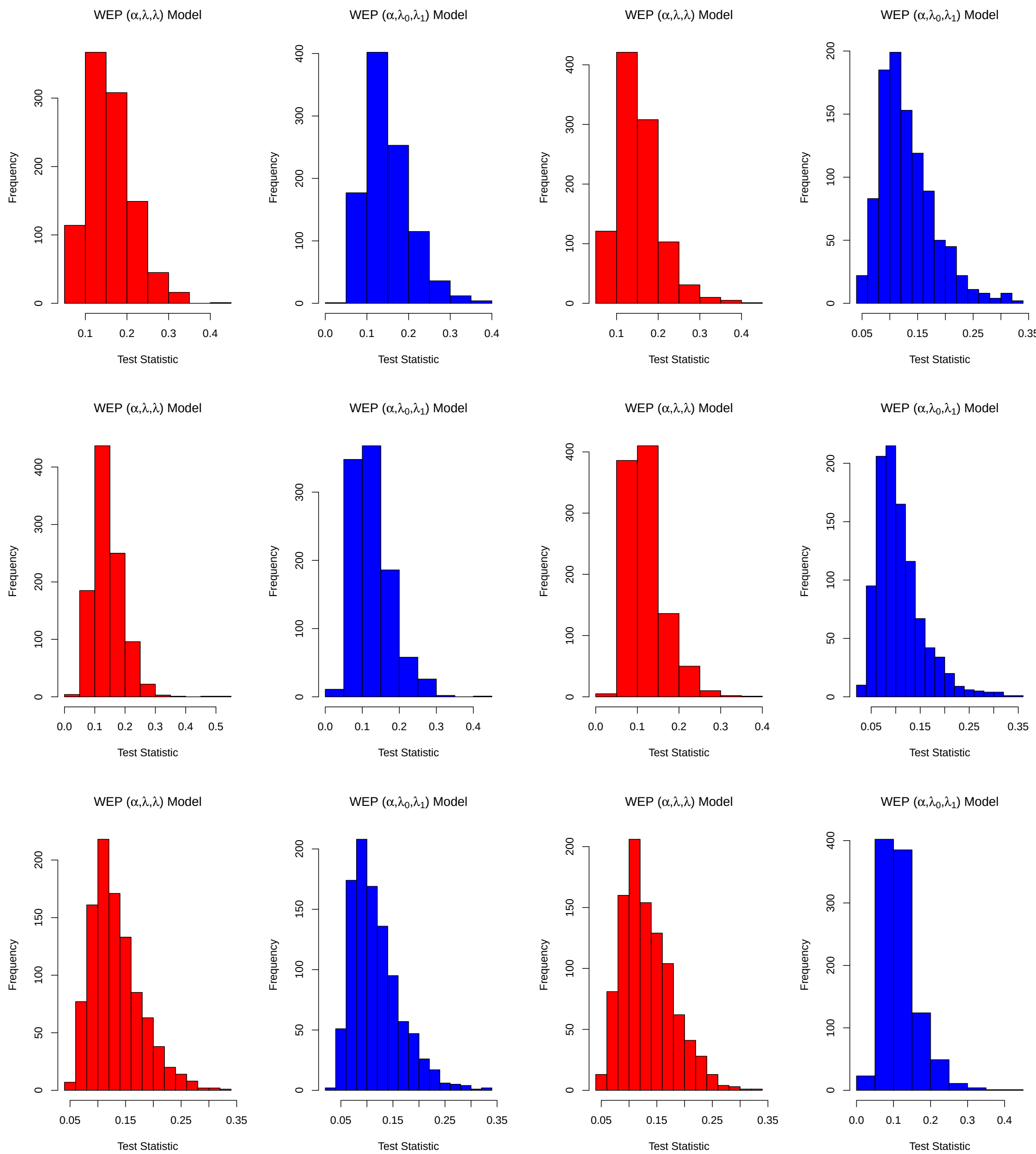

FIG 9. *Histograms of Test Statistics*

When $\alpha = 5$, $\lambda_0 = 0.25$, $\lambda_1 = 0.07$ (Figure 9) then,

R-package test for WEP(n,α, λ_0, λ_1)						
	$n = 75$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$
Test Statistic	9.48934e-02	6.34619e-02	7.11130e-02	3.70319e-02	5.06981e-02	5.62673e-02
P-value	0.861	0.965	0.9	0.997	0.975	0.93

R-package test for WEP(n,α, λ, λ)						
	$n = 75$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$
Test Statistic	1.3957e-01	1.2138e-01	1.4287e-01	1.2985e-01	1.3628e-01	1.1832e-01
P-value	0.34	0.034	0.026	0.032	0.021	0.25

When $\alpha = 4$, $\lambda_0 = 0.2$, $\lambda_1 = 0.2$ (Figure 10) then,

R-package test for WEP(n,α, λ, λ)						
	$n = 75$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$
Test Statistic	1.01350e-01	9.69882e-02	8.13595e-02	7.82217e-02	7.5234e-02	8.24579e-02
P-value	0.954	0.903	0.97	0.981	0.979	0.935

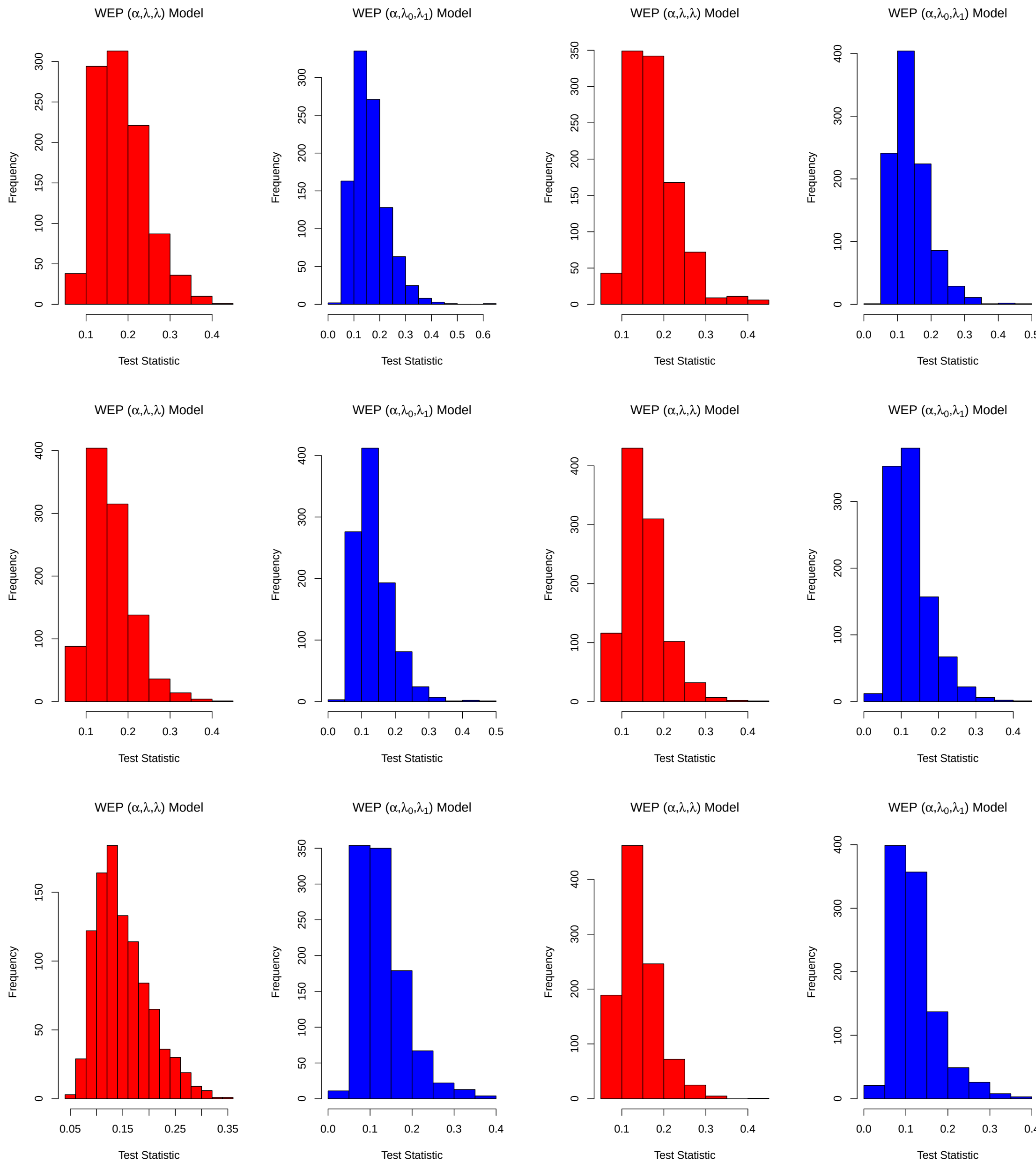

Fig 10. *Histograms of Test Statistics*

R-package test for WEP(n,α, λ_0, λ_1)						
	$n = 75$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$
Test Statistic	1.3983e-01	1.6529e-01	1.2139e-01	1.2658e-01	1.3429e-01	1.7592e-01
P-value	0.262	0.118	0.038	0.329	0.023	0.262

When $\alpha = 8$, $\lambda_0 = 0.7$, $\lambda_1 = 0.7$ (Figure 11) then,

R-package test for WEP(n,α, λ, λ)						
	$n = 75$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$
Test Statistic	1.01350e-01	9.69882e-02	8.1628e-02	7.3298e-02	7.8734e-02	8.9238e-02
P-value	0.977	0.987	0.932	0.9	0.956	0.949

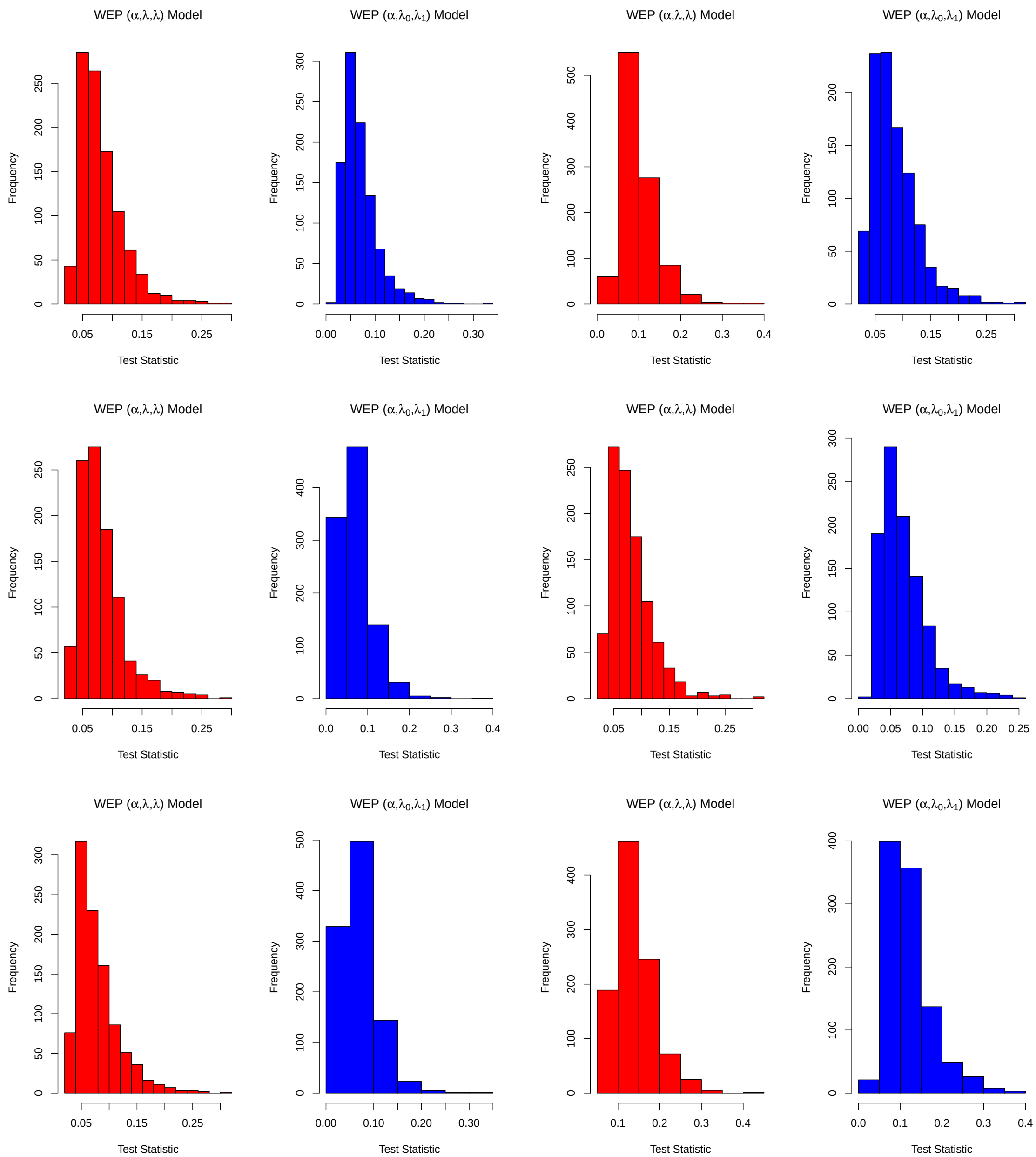

Fig 11. *Histograms of Test Statistics*

R-package test for WEP(n,α, λ_0, λ_1)						
	$n = 75$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$
Test Statistic	1.28159e-01	1.58759e-01	1.58759e-01	1.03536e-01	1.43619e-01	1.545802e-01
P-value	0.262	0.118	0.038	0.329	0.883	0.262

References

Arnod, B. C. and Robertson, C. (1989). Autoregressive logistical regression. *Journal of Applied Probabilty* **26** 524–531.

Arnold, B. C. (1993). Logistic processes involving Markovian minimization. *Communications in Statistics-Theory and Methods* **22** 1699–1707.

Bemis, B. M., Bain, L. J. and Higgins, J. J. (1972). Estimation and hypothesis testing for the parameters of a bivariate exponential distribution. *Journal of the American Statistical Association* **67** 927–929.

Christensen, S. (2012). Phase-type distributions and optimal stopping for autoregressive processes. *Journal of Applied Probability* **49** 22–39.

Jayakumar, K. and Babu, M. G. (2015). Some Generalizations of Weibull Distribution and Related Processes. *J. Stat. Theory Appl.* **14** 425–434.

Jose, K., Ristić, M. M. and Joseph, A. (2011). Marshall–Olkin bivariate Weibull distributions and processes. *Statistical Papers* **52** 789–798.

Kundu, D. (2022). A stationary Weibull process and its applications. *Journal of Applied Statistics* 1–20.

Kundu, D. and Dey, A. K. (2009). Estimating the parameters of the Marshall–Olkin bivariate Weibull distribution by EM algorithm. *Computational Statistics & Data Analysis* **53** 956–965.

Kundu, D. and Gupta, R. D. (2006). Estimation of P [Y¡ X] for Weibull distributions. *IEEE Trans. Reliab.* **55** 270–280.

Novikov, A. and Shiryaev, A. (2007). On a solution of the optimal stopping problem for processes with independent increments. *Stochastics An International Journal of Probability* *and Stochastic Processes* **79** 393–406.

Pillai, R. (1991). Semi-pareto processes. *Journal of applied probability* **28** 461–465.

Sim, C.-H. (1986). Simulation of Weibull and gamma autoregressive stationary process. *Communications in Statistics-Simulation and Computation* **15** 1141–1146.

Tavares, L. V. (1980). An exponential Markovian stationary process. *Journal of Applied Probability* **17** 1117–1120.