Iraqi Journal of Statistical Sciences

www.stats.mosuljournals.com

Using Wavelet Shrinkage to Deal with Contamination Problem in Survival Function for Weibull Distribution

Bekhal S. Sedeeq

Department of Statistics and Informatics, College of Administration and Economics, Salahaddin Uuniversity-Erbil, Erbil, Iraq

Article information

Article history:
Received January 18, 2023
Accepted April 2, 2023
Available online December 1, 2023

Keywords:
Survival function
Weibull distribution
Wavelet Shrinkage
Contaminated data

Correspondence:
Bekhal S. Sedeeq
Bikhal.sedeeq@su.edu.krd

Abstract

In this paper, the survival function of the Weibull distribution was estimated by the Classical Maximum Likelihood Estimate Method for the scale and shape parameters, and then the efficiency of the estimated parameters was calculated based on the mean square error and compared with the proposed method that deals with the contamination problem before estimating the parameters of the survival function for Weibull distribution through the use of Wavelets (Daubechies2), (Symlet3), and (Coiflit4) with several different methods of estimating the level of thresholding depending on the rule of soft. For the purpose of estimating and comparing the efficiency of the proposed method with the classical method, simulations were carried out for several different cases of the values for scale and shape parameters of Weibull distribution, contamination percentages, and different sample sizes as well as real data based on a MATLAB code designed for this purpose, the statistical program (SPSS) and the (Easy Fit) program. The study showed the efficiency of the precision parameters estimate for Weibull distribution when there was a data contamination problem when using the proposed method compared to the classical method.

Keywords: Survival function, Weibull distribution, Wavelet Shrinkage, and Contaminated data.

DOI: https://doi.org/10.33899/iqjoss.2023.0181139 ,

Introduction

Survival analysis (also called time-to-event analysis or duration analysis) is a branch of statistics aimed at analyzing the expected duration of time until one or more events happen called survival times or duration times such as death in biological organisms and failure in mechanical systems. This topic is called reliability theory or reliability analysis in engineering. Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. It is the study of time between entry into observation and a subsequent event. Now the scope of the survival analysis has become wide. Survival analysis is a set of statistical techniques used to describe and specify time to accident data. We use the term ‘failure’ in survival analysis to describe the occurrence of the task event (even though the event may actually be a ‘success’ such as recovery from therapy). The term ‘survival time’ specifies the length of time taken for failure to occur. (David, 2012. Singh, 2011).

Weibull distribution is a very beneficial distribution in survival analysis and reliability analysis. Several methods have been demonstrated to estimate the parameters of different distributions such as the method of moments method, maximum likelihood, etc. The Weibull distribution has gained much weight in the real world and is increasingly used in reliability and lifetime analysis or survival analysis. The 2-parameter Weibull distribution has the shape parameter (β) and the scale parameter (α). Most distributions such as normal, gamma, inverse gamma, and some other common distributions have two parameters which are of immense interest. (Nketiah, 2021).

The contamination of the estimation of the intercept have only a small impact on the estimation of the regression coefficients. Good leverage points are observations that are on the outskirts of the design space but are near to the regression line. They have a minor impact on the estimate of both the intercept and the regression coefficients, but they have an impact on inference. In contrast, bad leverage points are observations that are far off the regression line. Signals are typically contaminated by random noise, hence, several methods have been used to smooth noisy signals including the Fourier transform, the Svitzky Goloylocal polynomial, the mean filters, and Gussian function and so on. However, these methods usually smooth the signal to reduce the noise, but, in the process, also blure the signal. In recent years, a new method has been introduced to the method of de-noising known as wavelet shrinkage. (Taha and Saleh, 2022)

Wavelet shrinkage make de-noising a method of reducing noise in signals. Wavelet shrinkage is a signal denoising technique based on the idea of thresholding the wavelet coefficients. Donoho et al. (1995a) have introduced the method of wavelet shrinkage for general curve estimation problems. There are several good reasons why wavelet shrinkage can be used for estimation function. (Mustafa, and Taha, 2013)

Survival Analysis

Survival analysis is always treated with the analysis of data in times of accidents in life. The survival analysis and modeling the time it takes events occur, i.e. this typical event is death which is derived from the name ' survival ' analysis. Let T be a random variable that represents failure time of an event with probability of density function f (t) and cumulative distribution function

F(t) = Pr(T ≤ t) , the survival function S(t) is defined as: (David, 2012)

S(t) = Pr (T > t) = 1-F(t)

Weibull Distribution

The probability density function and the cumulative distribution function of a two parameter Weibull distribution with scale parameter, α > 0 and shape parameter, β > 0, are given by, (Nketiah, 2021)

$$f(\boldsymbol{x_i}; \alpha, \beta) = (\frac{\boldsymbol{\beta}}{\boldsymbol{\alpha}})(\frac{\boldsymbol{x_i}}{\boldsymbol{\alpha}})^{\boldsymbol{\beta-1}} \boldsymbol{exp-}\left(\frac{\boldsymbol{x_i}}{\boldsymbol{\alpha}}\right)^{\boldsymbol{\beta}} \qquad x > 0\,, \alpha > 0, \beta > 0 \tag{1}$$

The cumulative distribution function is,

$$F(\boldsymbol{x_i}; \alpha, \beta) = 1 - \boldsymbol{exp-}\left(\frac{\boldsymbol{x_i}}{\boldsymbol{\alpha}}\right)^{\boldsymbol{\beta}} \tag{2}$$

Scale and shape parameters estimated by using *Maximum Likelihood Estimation (MLE).*

The method of MLE is a commonly used procedure for estimating parameters. Assume $x_1, x_2, \dots, x_n$ be a random sample of size n obtained from a population with pdf, $f(x, \theta)$ where $\underline{\theta}$ is a hidden vector of parameter, $\underline{\theta} = (\beta, \alpha)$ likelihood function is given as,

$$L = \prod_{i=1}^{n} f(x_i; \theta) \tag{3}$$

The MLE of θ is the value of θ that maximizes the likelihood function or the log-likelihood function where

$$\frac{\partial logL}{\partial \theta} = 0 \tag{4}$$

By applying Eqn.(3) to the Weibull probability density function 1n Eqn.(1)

the likelihood function will be:

$$L(x_i, \alpha, \beta) = \prod_{i=1}^{n} \left(\frac{\beta}{\alpha}\right) (\frac{x_i}{\alpha})^{\beta-1} \, exp-\left(\frac{x_i}{\alpha}\right)^{\beta} \tag{5}$$

$$= \left(\frac{\beta}{\alpha}\right)^{n} \left(\frac{1}{\alpha}\right)^{n\beta-n} \prod_{i=1}^{n} (x_i)^{\beta-1} \, exp-\left(\frac{x_i}{\alpha}\right)^{\beta} \tag{6}$$

Taking the algorithm of both sides, we get:

$$lnL = \ln\{\left(\frac{\beta}{\alpha}\right)^{n}\left(\frac{1}{\alpha}\right)^{n\beta-n}\prod_{i=1}^{n}(x_i)^{\beta-1}\, exp-\sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\beta}\} \tag{7}$$

Differentiating β, α , we obtain the estimating equations as follows:

$$\frac{\partial lnL}{\partial\beta} = \frac{n}{\beta} - n\ln\alpha + \sum_{i=1}^{n}\ln x_i\left(1-\left(\frac{x_i}{\alpha}\right)^{\beta}\right) \tag{8}$$

$$\frac{\partial lnL}{\partial\alpha} = -\frac{n}{\alpha} + \frac{1}{\alpha^{\beta+1}}\sum_{i=1}^{n} x_i{}^{\beta} \tag{9}$$

Equations 8 and 9 are solved numerically to obtain the estimated parameters.

4. Contaminated Data (Contamination Outlier and Noise)

The data come from two types of distributions the first of which is called Basic Distribution $f_0(.)$ that generates good data while the second of which is called Contamination Distribution $f_1(.)$ and P is a ratio of contamination then the distribution of an arbitrary observation is (Hawkins, 1980):

$$f(.) = (1-P) * f_0(.) + P * f_1(.) \tag{10}$$

In literature on data mining and statistics, outliers are sometimes known as abnormalities, discordant, deviants or anomalies. In the majority of applications, the data is produced by one or more producing processes which may either reflect system activity or observations made about entities. Noisy data is data that have been made due to the presence of too much variation. It is presumed that the signal or observation is presented and disguised by noise. The difficulty of separating the noise from the signal or observation has long been a focus in statistics. So, the useful data need to be used to inform researchers. However, the percentage of noisy data that is relevant is frequently too small to be useful. (Taha and Saleh, 2022)

Wavelet Shrinkage

Wavelet shrinkage is a well-established technique for removing the noise present in the observation while preserving the significant features of the original data (Donoho, 1994). The wavelet shrinkage is based on thresholding of the wavelet coefficients. The wavelet shrinkage has several good properties that gained this popularity in statistics nearly minimax for a wide range of loss function and for general function classes; simple, practical and fast; adaptable to spatial and frequency in homogeneities; readily extendable to high dimensions; applicable to various problems such as density estimation and inverse problems. In statistics, applications of wavelets arise mainly in the tasks involving non-parametric regression, density estimation, assessment of scaling, functional data analysis and stochastic processes. (Donoho and Johnstone , 1995)

1. Wavelet

Wavelets are small waves that can be grouped together to form larger waves or different waves (Ali, et al, 2022). A few fundamental waves were used, i.e. they were stretched in infinitely many ways, and moved in infinitely many ways to produce a wavelet system that could make an accurate model of any wave.

Consider generating an orthogonal wavelet basis for functions $f \in L^2(R)$ (the space of square integrable real functions), starting with two parent wavelets: the scaling function $\emptyset$ (also called farther wavelet) and the mother wavelet ψ . Other wavelets are then generated by $\emptyset$ and ψ (Donald et al., 2004). The dilation and translation of the functions are defined by formulas (11) and (12).

$$\varphi_{k,q}(y) = 2^{k/2}\varphi(2^k y - q)k, q \in z \tag{11}$$

$$\psi_{k,q}(y) = 2^{k/2}\psi(2^k y - q)k, q \in z \tag{12}$$

The discrete wavelet transform (DWT) is a broadly applicable observation of processing algorithm which is benefit in several applications, for e.g. science, engineering, mathematics and computer science. DWT decomposes an observation by using scaled and shifted versions of a compact supported basis function (mother wavelet), and provides multiresolution representation of the observation. It gives a vector of observations y consisting of 2^k observations where k is an integer and the DWT of y due to formula (13). (Ali, et al, 2022)

$$W = wy \tag{13}$$

Where w is wavelet matrix with (n × n) dimension, W is a vector with (n × 1) dimension including both scaling and wavelet coefficients. The vector of wavelet coefficients can be organized into (k+1) elements. $W= [W_1, W_2, \ldots, W_k, V_{k0}]^T$ at each DWT, the approximate coefficients are divided into bands using the same wavelet as before with the result that the details are appended with the details of the latest decomposition as in the following formula (Taha and Saleh, 2022):

$$y = w^T W = \sum_{j=1}^{k} W_j^T W_j + V_{k_0}^T V_{k_0} \tag{14}$$

At each level (k) the observations can be reconstructed from the de-noise data (reducing the contamination) by the inverse DWT (Ramazan et al., 2002).

2. Thresholding

The simplest method of non-linear wavelet de-noising is thresholding in which the wavelet coefficient is sub divided into two sets one of which represents signal while the other represents noise. To apply the thresholds of the wavelet coefficients, there are different rules and several different methods for choosing a threshold value exist such as:

I. SURE

The SURE threshold proposed by Donoho and Johonstone (1994), which is based upon the minimization of Stein's risk estimator. In SURE threshold method specifies a threshold estimate of δ_k at each level k for the wavelet coefficients and then for the soft threshold estimator, we have.

$$SURE(\delta, W) = n - 2\#\{k: |W_k| \le \delta\} - \sum_{k=1}^{d} min(|W_k|, \delta)\hat{}2 \tag{15}$$

Where $[W_k: k = 1,2, \ \ , d]$ be a wavelet coefficient in the kth level, and then select δ^S that minimizes SURE (δ, W)

$$\delta^S = arg\, min\, SURE(\delta, W) \tag{16}$$

II. Minimax

The optimal minimax threshold method is submitted by Donoho and Johonstone (1994) as an improvement to the universal threshold method. Minimax is based on an estimator $\tilde{f}$ that attains to the minimax risk as:

$$\tilde{R}(F) = \inf_{\tilde{f}} \sup_{f \in \tilde{R}(F)} R(\tilde{f}, f) \tag{17}$$

Where

$$R(\tilde{f}, f) = \frac{1}{n} \sum_{i=1}^{n} E[\tilde{f}, f]^2 \tag{18}$$

Where $f = f(x_i)$ and $\tilde{f} = \tilde{f}(x_i)$, denote the vectors of true and estimated sample values. The threshold minimax estimator is different from universal counter parts in which the minimax threshold method concentrates on reducing the overall mean square error (MSE) but the estimates are not over-smoothing.

III. Universal Threshold

Donoho and Johnstone (1994) proposed universal threshold which is given by

$$\eta^U = \tilde{\sigma}_{(MAD)} \sqrt{2 \log N} \tag{19}$$

Where N is the data length series and $\tilde{\sigma}_{(MAD)}$ is the estimator of standard deviation of details coefficients, which is estimated as:

$$\tilde{\sigma}_{(MAD)} = \frac{MAD}{0.6745} \tag{20}$$

MAD is the median absolute deviation of the wavelet coefficients at the finest scale

3. Thresholding Rules

There are two main thresholding rules:

1- Soft Thresholding

It was proposed by Donoho & Johnostone in 1995. Soft thresholding zeros all the signal values smaller than δ followed by subtracts δ from the values larger than δ which is defined as follows:

$$Wn^{(s)} = sign\{Wn\}(|Wn| - \delta)_+ \tag{21}$$

Where

$$Sign\{Wn\} = \begin{bmatrix} +1 & if \quad Wn > 0 \\ 0 & if \quad Wn = 0 \\ -1 & if \quad Wn < 0 \end{bmatrix} \tag{22}$$

and

$$(|Wn| - \delta)_+ = \begin{bmatrix} (|Wn| - \delta) & if \quad (|Wn| - \delta) \geq 0 \\ 0 & if \quad (|Wn| - \delta) < 0 \end{bmatrix} \tag{23}$$

2-Hard Thresholding

Donoho and Johnstone proposed Hard thresholding which is the simplest thresholding technique based on the premise of (keep or kill). Hard thresholding zeroes out all the signal values smaller than δ. The wavelet coefficient is set to the vector Wn^{HT} with element. "Quotation" (Donoho, and Johnstone , 1995)

$$Wn^{(HT)} = \begin{bmatrix} 0 & if \quad |Wn| \leq \delta \\ Wn & if \quad |Wn| > \delta \end{bmatrix} \tag{24}$$

Proposed Method

The proposed method included dealt with the contamination problem of Weibull distribution in survival analysis using Wavelet Shrinkage. First, compute the DWT coefficients $W(t)$ for a wavelet $w(t)$ (Daubechies, Symlets, and Coiflets wavelets). Second, the threshold level δ is estimated by one of the methods (e.g. SURE, Minimax, and Universal threshold). Third, Thresholding rules (Soft) is used to keep or kill the discrete wavelet coefficients. Thus, we get the modified DWT coefficients $MW(t)$, then it is used to compute the inverse of the modified DWT (Taha and Jwana, 2022) as in formula (25).

$$w(t)^* = Inv\big(MW(t)\big) \tag{25}$$

Finally, the data for Weibull distribution which have less contamination are used to estimate the shape and scale parameter of the Weibull distribution using the method of maximum likelihood and then analyze the survival function on this basis.

Evaluation Criterion

To measure the accuracy of the estimated parameters (scale and shape) of the Weibull distribution, the mean squared error (MSE) can be used as in the following formula:

$$MSE\big(\hat{\theta}\big) = \frac{\sum_{i=1}^{m}\big(\theta_i - \hat{\theta}_i\big)^2}{m} \tag{26}$$

m: number of samples.

Experimental and Application

To compare between the classical and the proposed method in terms of efficiency and accuracy of the estimated parameters for Weibull distribution and reliability function, an experimental aspect was done by simulating the Weibull distribution, then an applied aspect of the real data based on MSE criterion and by designing a program in MATLAB (version 2020a) dedicated to this purpose (Appendix).

1. Experimental Aspect

Four cases were selected for scale parameter (0.5 and 1) and shape parameter (5 and 10), the sample size (50 and 100) and the addition of contamination percentages (10% and 20%) has a Cauchy distribution (α = 0 and β = 0.5). For the first experimental with n = 100, figure (1) is shown.

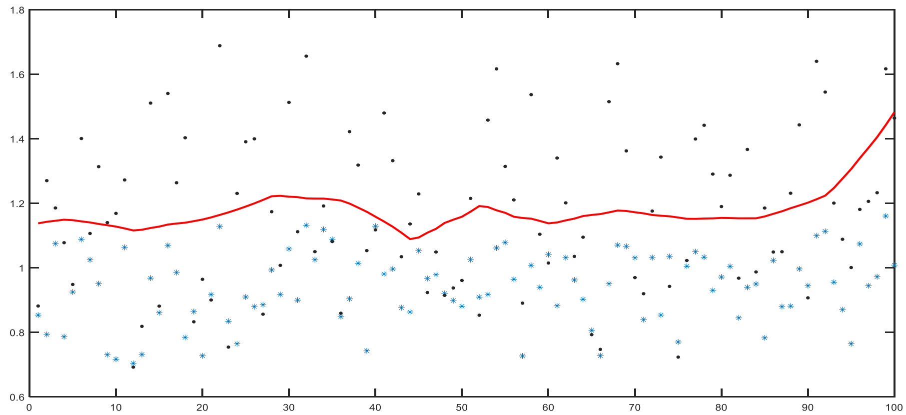

Figure (1): The Original data (*), Contamination data (.), and De-noise data (-)

Figure (1) shows the scatter plot of the data generated from the Weibull distribution (*) at scale parameter (0.5) and shape parameter (5), and the values of the scatter of the contamination data (.) at 10% contamination, and then the data processed from the contamination (-) using the (Sym3) wavelet with universal threshold and soft rule. The Survival function of the Weibull distribution for the contaminated and treated data is shown in Fig. 2 and 3 respectively.

For the purpose of the comparison between the proposed and classical method in estimating the parameters of the Weibull distribution, the experiment was repeated to (1000) times and the average criterion for MSE was calculated. Three wavelets (Db2), (Sym3), and (Coif4) were used with different methods in estimating the threshold level (SURE, Minimax, and Universal), with threshold rule (Soft), and for different samples (50, and 100) and percentage of contamination (10% and 20%). The results are summarized in tables (1-4) for the average of (MSE) criterion when at

$$t \sim Wei(0.5,5), \quad t \sim Wei(1,5), \quad t \sim Wei(0.5,10), and\ t \sim Wei(1,10)$$

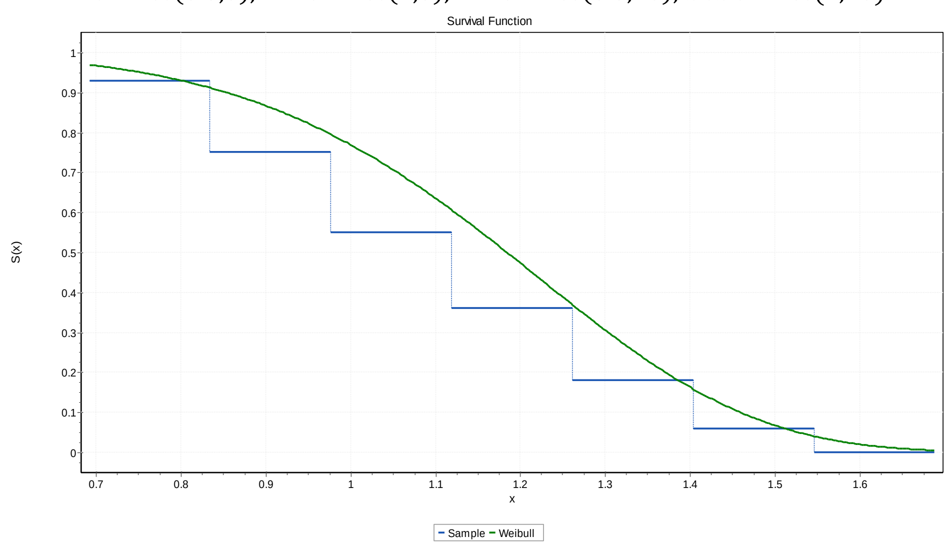

Figure (2): Survival Function of the Weibull Distribution for the Contaminated Data

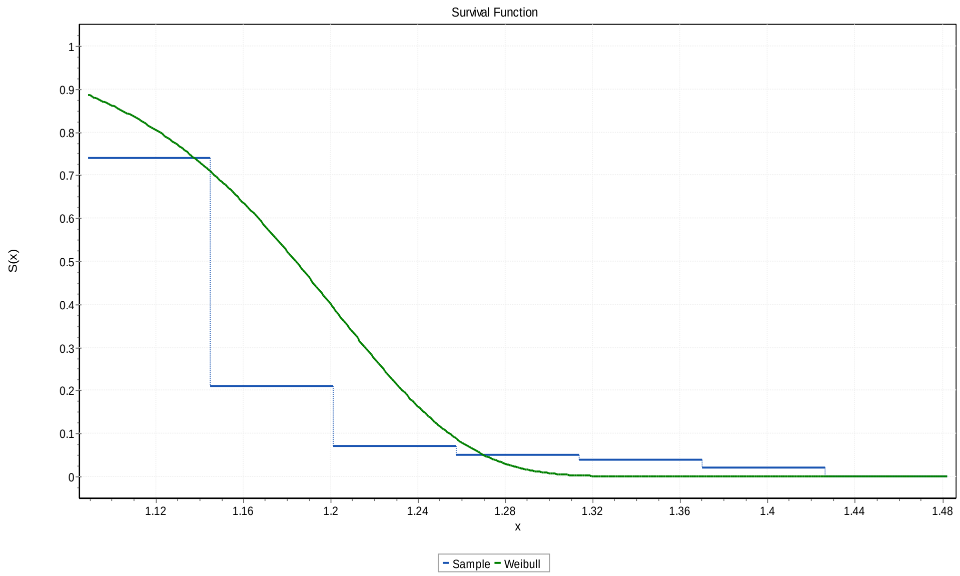

Figure (3): The Survival Function of the Weibull Distribution for the Treated Data

Table (1): Average of MSE Criterion when $t \sim Wei(0.5,5)$

Method	Sample size	Percentage of Contamination	Wavelet	Threshold Method	MSE(α)	MSE(β)
Proposed	50	10%	Db2	SURE	29.9383	1.8361
				Minimax	29.8261	1.4718
				Universal	29.6699	1.3710
			Sym3	SURE	29.9890	1.9961
				Minimax	29.8394	1.4203
				Universal	29.6371	1.3449
			Coif4	SURE	29.9009	1.6753
				Minimax	29.7546	1.3025
				Universal	**29.6123**	**1.2506**
Classical					31.1941	9.3698
Proposed	50	20%	Db2	SURE	120.1897	2.1030
				Minimax	119.7780	1.7198
				Universal	119.1890	1.5535
			Sym3	SURE	120.3456	2.2526
				Minimax	119.8419	1.6561
				Universal	119.0830	1.4870
			Coif4	SURE	120.0310	1.8744
				Minimax	119.5262	1.4863
				Universal	**118.9887**	**1.3718**
Classical					122.8904	10.2371
Proposed	100	10%	Db2	SURE	28.8237	2.6367
				Minimax	28.6013	2.6987
				Universal	28.4428	3.5660
			Sym3	SURE	28.9553	2.3568
				Minimax	28.7226	2.2413
				Universal	28.5333	3.0492
			Coif4	SURE	28.7310	3.5742
				Minimax	28.4770	3.7580

				Universal	**28.3180**	**2.2137**
Classical					31.2688	9.4165
Proposed	100	20%	Db2	SURE	115.9454	2.2129
				Minimax	115.1003	2.1284
				Universal	114.4698	2.8339
			Sym3	SURE	116.4356	2.0469
				Minimax	115.5527	1.8101
				Universal	114.8186	2.4264
			Coif4	SURE	115.5530	2.8968
				Minimax	114.6063	2.9944
				Universal	**113.9868**	**1.7810**
Classical					123.1614	10.2919

Table (2): Average of MSE Criterion when $t \sim Wei(1,5)$

Method	Sample size	Percentage of Contamination	Wavelet	Threshold Method	$MSE(\alpha)$	$MSE(\beta)$
Proposed	50	10%	Db2	SURE	29.1860	1.5416
				Minimax	29.0536	1.2177
				Universal	28.8879	1.2386
			Sym3	SURE	29.2510	1.7314
				Minimax	29.0638	1.1804
				Universal	28.8460	1.2830
			Coif4	SURE	29.1386	1.4764
				Minimax	28.9685	1.1649
				Universal	**28.8163**	**1.1589**
Classical					31.1308	8.0982
Proposed	50	20%	Db2	SURE	117.6636	1.8820
				Minimax	117.2345	1.5141
				Universal	116.6199	1.4014
			Sym3	SURE	117.8677	2.0424
				Minimax	117.2856	1.4626
				Universal	116.4954	1.3669
			Coif4	SURE	117.5177	1.7053
				Minimax	116.9569	1.3329
				Universal	**116.3993**	**1.2675**
Classical					122.2504	9.5304
Proposed	100	10%	Db2	SURE	27.9865	3.7141
				Minimax	27.7598	3.9826
				Universal	27.5976	5.2004
			Sym3	SURE	28.1462	3.1542
				Minimax	27.8891	3.2646
				Universal	27.6915	4.4804
			Coif4	SURE	27.8984	4.9873
				Minimax	27.6301	5.4365
				Universal	**27.4674**	**3.1415**
Classical					31.2109	8.1348
Proposed	100	20%	Db2	SURE	113.2828	2.5646
				Minimax	112.4170	2.5755
				Universal	111.7902	3.4094
			Sym3	SURE	113.7934	2.2832
				Minimax	112.8932	2.1179

				Universal	112.1464	2.9146
			Coif4	SURE	112.9306	3.3937
				Minimax	111.9261	3.5960
				Universal	**111.2990**	**2.1087**
Classical					122.5386	9.5784

Table (3): Average of MSE Criterion when $t \sim Wei(0.5,10)$

Method	**Sample size**	**Percentage of Contamination**	**Wavelet**	Threshold Method	$\text{MSE}(\alpha)$	$\text{MSE}(\beta)$
Proposed	50	10%	Db2	SURE	30.1136	35.2721
				Minimax	29.9956	33.6752
				Universal	29.8382	31.8721
			Sym3	SURE	30.1549	35.3359
				Minimax	30.0078	33.2520
				Universal	29.8042	30.9007
			Coif4	SURE	30.0588	33.5041
				Minimax	29.9209	31.6553
				Universal	**29.7790**	**30.0291**
Classical					31.4130	64.6975
Proposed	50	20%	Db2	SURE	120.5057	37.4422
				Minimax	120.0781	35.8988
				Universal	119.4895	34.0720
			Sym3	SURE	120.6475	37.4647
				Minimax	120.1447	35.4239
				Universal	119.3857	33.0576
			Coif4	SURE	120.3542	35.7758
				Minimax	119.8240	33.8354
				Universal	**119.2866**	**32.2075**
Classical					123.3063	67.0477
Proposed	100	10%	Db2	SURE	28.9857	20.5670
				Minimax	28.7648	17.6639
				Universal	28.6033	15.7304
			Sym3	SURE	29.1198	22.6352
				Minimax	28.8843	19.5389
				Universal	28.6943	17.1038
			Coif4	SURE	28.8839	18.6996
				Minimax	28.6369	15.4699
				Universal	**28.4779**	**13.6675**
Classical					31.4843	64.8822
Proposed	100	20%	Db2	SURE	116.2646	22.6668
				Minimax	115.3885	19.6353
				Universal	114.7571	17.6281
			Sym3	SURE	116.7477	24.6787
				Minimax	115.8434	21.5736
				Universal	115.1075	19.0601
			Coif4	SURE	115.8330	20.6174
				Minimax	114.8926	17.3623
				Universal	**114.2733**	**15.4556**
Classical					123.5680	67.2418

Table (4): Average of MSE Criterion when $t \sim Wei(1,10)$

Method	Sample size	Percentage of Contamination	Wavelet	Threshold Method	$MSE(\alpha)$	$MSE(\beta)$
Proposed	50	10%	Db2	SURE	29.5131	31.5014
				Minimax	29.3832	29.9700
				Universal	29.2157	28.1807
			Sym3	SURE	29.5710	31.7905
				Minimax	29.3903	29.5946
				Universal	29.1729	27.2522
			Coif4	SURE	29.4627	30.0089
				Minimax	29.2922	28.0110
				Universal	**29.1405**	**26.3864**
Classical					31.5393	61.0901
Proposed	50	20%	Db2	SURE	118.2867	35.6962
				Minimax	117.8332	34.1064
				Universal	117.2145	32.3038
			Sym3	SURE	118.4436	35.7407
				Minimax	117.8802	33.6804
				Universal	117.0887	31.3192
			Coif4	SURE	118.0879	33.9611
				Minimax	117.5441	32.0812
				Universal	**116.9879**	**30.4530**
Classical					123.0293	65.1388
Proposed	100	10%	Db2	SURE	28.3065	17.3308
				Minimax	28.0739	14.4667
				Universal	27.9102	12.7196
			Sym3	SURE	28.4562	19.3011
				Minimax	28.2026	16.2475
				Universal	28.0054	13.9791
			Coif4	SURE	28.2115	15.7111
				Minimax	27.9432	12.4664
				Universal	**27.7793**	**10.8975**
Classical					31.6152	61.2684
Proposed	100	20%	Db2	SURE	113.8650	20.9711
				Minimax	112.9944	18.0499
				Universal	112.3565	16.0971
			Sym3	SURE	114.3819	23.0055
				Minimax	113.4643	19.9312
				Universal	112.7144	17.4826
			Coif4	SURE	113.4534	19.0606
				Minimax	112.4899	15.8356
				Universal	**111.8630**	**14.0107**
Classical					123.3046	65.3248

Tables (1-4) shows that all the proposed methods have better efficiency than the classical method in estimating scale and shape parameters for Weibull distribution depending on the average of criterion (MSE) for all cases. Also, (Coif4) wavelet with Universal threshold method is the best efficient compared with all other proposed methods and with the classical method because it has the lowest average of the criterion ($MSE(\alpha)$ and $MSE(\beta)$). The efficiency of the estimated parameters decreases with increasing contamination percentage for all simulations. Also, the efficiency of the estimated scale parameter α is not affected by an increase in its real value, and the efficiency of the estimated shape parameter β decreases as its real value increases for all simulations.

2. Application Part

Real data represent the time of kidney failure. The distribution of the data was tested using Kolmogorov-Smirnov, and the test statistic (0.12074) is less than critical value (0.23059), that supports the hypothesis of the Weibull distribution for data (p-value 0.45096 > 0.01). The statistic test (the goodness of fit, χ^2) for the classical and proposed methods is summarized in table (5).

Table (5): The Goodness of Fit for Real Data

Method	Wavelet	Threshold Method	The Statistic
Proposed	Db2	SURE	3.8305
		Minimax	3.4331
		Universal	3.6593
	Sym3	SURE	3.8296
		Minimax	1.9165
		Universal	**1.8655**
	Coif4	SURE	3.8253
		Minimax	2.1507
		Universal	2.1214
Classical			3.8245

The proposed method (Db2 with Universal Threshold Method) is the best because it has the statistic (1.8655) with scale parameter (8.3444) and shape parameter (18.922). On this basis, the probability density function, cumulative, and survival of kidney failure data are shown in the figures (4-6):

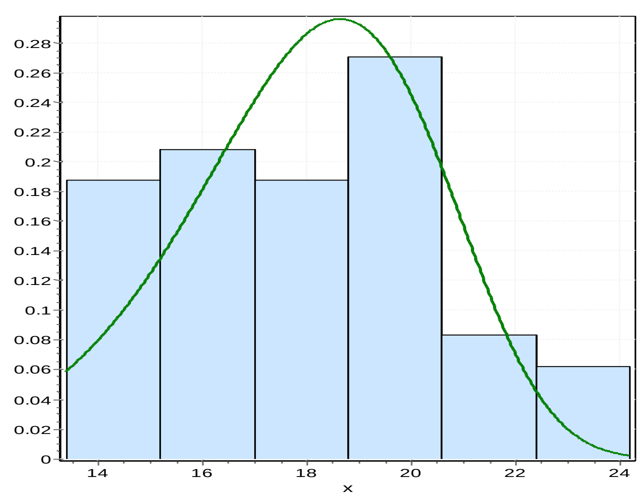

Figure (4): The Probability Density Function of the Weibull Distribution for Real Data

Figure (4) shows the Probability Function of the Weibull Distribution of Kidney Failure Data for the Time Period (12-24).

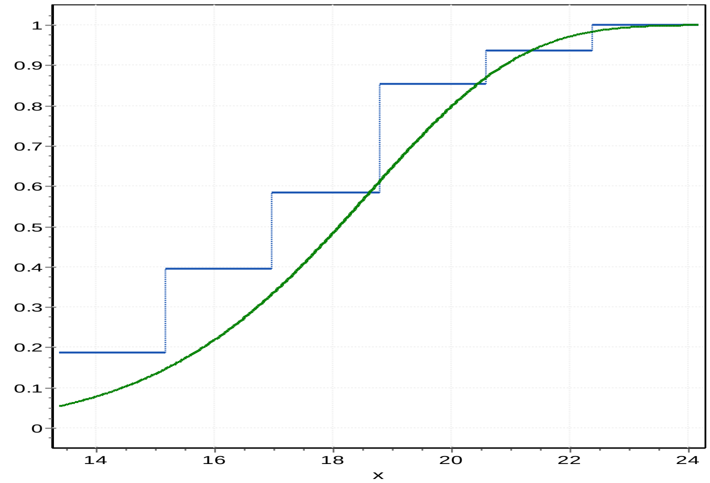

Figure (5): The Cumulative Function of the Weibull Distribution for the Real Data

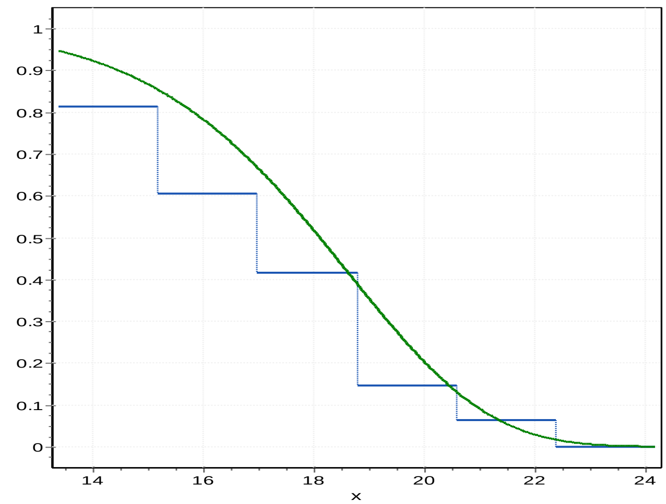

Figure (6): The Survival Function of the Weibull Distribution for Real Data

Conclusions

1. All the proposed methods have better efficiency than the classical method in estimating scale and shape parameters for Weibull Distribution depending on the average of criterion (MSE) for all cases.
2. Coif4 wavelet with Universal Threshold Method is the best efficient compared with all other proposed methods and with the classical method because it has the lowest average of the Criterion (MSE(α) and MSE(β)).
3. For real data the proposed method (Db2 with Universal threshold method) is the best

Recommendations

1.The proposed method for estimating the parameters of Weibull distribution is recommended.

2. The use of other types of orthogonal wavelets (bior, rbio, and dmey), methods for estimating the threshold level, and the thresholding rules in estimating the parameters of Weibull distribution and Survival function is also recommended.
3. Using a Bayesian approach with Wavelet Shrinkage in estimation the parameters of Weibull, Gamma, and Gomppertz distribution is recommended as well.

References

1. Aako, O.L. and Are, S.O., (2020), comparison of Cox, Weibull, Gompertz and Log-Logistic regression models in survival analysis of hypertensive patients.
2. Donald B. Percival, Muyin Wang and James E. Overland, (2004), An Introduction to Wavelet Analysis with Application to Vegetation Time Series, University of Washington
3. Donoho D. L. and Johnstone I. M. , (1994) , Ideal Spatial Adaptation by Wavelet Shrinkage , Biometrika, 81 , 425-550
4. Donoho D. L. and Johnstone I. M. , (1995) , Adapting to unknown smoothness via Wavelet Shrinkage , J.Am. Statist.Ass.90, 1200=1224
5. David G. Kleinbaum, Klein, M.(2012). Survival Analysis. Springer, LLC, Third Edition, USA.
6. Eddy Armand Kwessi Nyandjou, (2011), Efficient Rank Regression with Wavelets Estimated Scores, dissertation submitted to the Graduate Faculty of Auburn University.
7. Hawkins, D. M.(1980), Identification of outliers. Springer (Vol. 11).
8. Jeena Joy1, Salice Peter 2, Neetha John3, (2013), de-noising using soft thresholding, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 3.
9. Nketiah, E., Chenlong, L., Yingchuan, J., and Dwumah, B.(2021). Parameter estimation of the Weibull Distribution; Comparison of the Least- Squares Method and the Maximum Likelihood estimation. Department of Statistics, Taiyuan University of Technology, China. International Journal of Advanced Engineering Research and Science (IJAERS).
10. Rogério Porto, Pedro Morettin, Donald Percival, and Elisete Aubin, (2016), Wavelet shrinkage for regression models with random design and correlated errors, Braz. J. Probab. Stat. 30(4): 614-652, DOI: 10.1214/15-BJPS296.
11. Ali, Taha Hussein, and Saleh, Dlshad Mahmood. (2022), "Proposed Hybrid Method for Wavelet Shrinkage with Robust Multiple Linear Regression Model: With Simulation Study." QALAAI ZANIST JOURNAL 7.1: 920-937.
12. Singh, R. and Mukhopadhyay, K., (2011), Survival analysis in clinical trials: Basics and must know areas. Perspectives in clinical research, 2(4), p.145.
13. Xing-cai Zhou, Ying-zhi Xu, and Jin-guan Lin, (2017), Wavelet estimation in varying coefficient models for censored dependent data, journal homepage: www.elsevier.com/locate/stapro.
14. Yogendra P. Chaubey, Hassan Doosti, Esmaeel Shirazi, B. L. S., and Prakasa Rao, (2010), Linear Wavelet-Based Estimation for Derivative of a Density under Random Censorship, Concordia University, Montr´eal, Canada, JIRSS, Vol. 9, No. 1, pp 41-51. .
15. Taha ali; Jwana Rostam Qadir. (2022), "Using Wavelet Shrinkage in the Cox Proportional Hazards Regression model (simulation study)". IRAQI JOURNAL OF STATISTICAL SCIENCES, 19, 1, 17-29. doi: 10.33899/iqjoss.2022.174328
16. Ali, Taha Hussein, and Dlshad Mahmood Saleh, (2021), "COMPARISON BETWEEN WAVELET BAYESIAN AND BAYESIAN ESTIMATORS TO REMEDY CONTAMINATION IN LINEAR REGRESSION MODEL." PalArch's Journal of Archaeology of Egypt/Egyptology 18.10: 3388-3409.
17. Ali, Taha Hussein, Saman Hussein Mahmood, and Awat Sirdar Wahdi (2022), "Using Proposed Hybrid method for neural networks and wavelet to estimate time series model." Tikrit Journal of Administration and Economics Sciences 18.57 part 3
18. Mustafa, Qais, and Taha. H. A. (2013), "Comparing the Box Jenkins models before and after the wavelet filtering in terms of reducing the orders with application." Journal of Concrete and Applicable Mathematics 11: 190-198.

Appendix

The real data

13	14	15	15	15	15	15	15	15	15
16	16	16	16	16	16	17	17	17	17
17	17	18	18	18	18	18	19	19	19
19	19	19	20	20	20	20	20	20	20
21	21	21	21	22	22	23	24		

```
clc
clear all
%rng('default');        % For reproducibility
theta=[.5 5];n=100
for i=1:10
X = wblrnd(.5,5,n,1);     % Simulated strengths
fit(i,:) = wblfit(X);E=(fit-theta).^2;
r=rand(n,1)*100;XN=.10*r+.90*X;
fit1(i,:) = wblfit(XN);E1=(fit1-theta).^2;
Xw=wdenoise(XN,'Wavelet','coif4','DenoisingMethod','SURE','ThresholdRule','soft');fit2(i,:)=
wblfit(abs(Xw));
E2=(fit2-theta).^2;
Xw=wdenoise(XN,'Wavelet','coif4','DenoisingMethod','minimax','ThresholdRule','soft');
fit3(i,:)= wblfit(abs(Xw));E3=(fit3-theta).^2;
Xw=wdenoise(XN,'Wavelet','coif4','DenoisingMethod','universal','ThresholdRule','soft');
fit4(i,:)= wblfit(abs(Xw));E4=(fit4-theta).^2;
end
MSEO=mean(E),MSE=mean(E1),MSEw1=mean(E2),MSEw2=mean(E3)
MSEw3=mean(E4)
```

إستخدام التقليص المويجي لمعالجة مشكلة التلوث في دالة البقاء لتوزيع ويبل

بيخال صمد صديق

قسم الاحصاء والمعلوماتية، كلية الإدارة والاقتصاد، جامعة صلاح الدين–اربيل، اربيل، العراق/
Bikhal.sedeeq@su.edu.krd

الخلاصة

تم في هذا البحث تقدير دالة البقاء لتوزيع ويبل بطريقة تقدير بالإمكان الأعظم التقليدية لمعلمة القياس والشكل ومن ثم حساب كفاءة المعلمات المقدرة إعتماداً على متوسط الخطأ التربيعي ومقارنتها مع الطريقة المقترحة التي تعالج مشكلة التلوث قبل تقدير معلمات دالة البقاء لتوزيع ويبل وذلك من خلال إستخدام المويجات (Daubechies2)، (Symlet3)، و (Coiflet4) مع عدة طرائق مختلفة في تقدير مستوى قطع العتبة بالإعتماد على قاعدة قطع العتبة الناعمة (Soft). لغرض التقدير والمقارنة بين كفاءة الطريقة المقترحة والتقليدية تم إجراء المحاكاة لعدة حالات مختلفة من قيم معلمات القياس والشكل لتوزيع ويبل ولنسب تلوث وأحجام عينات مختلة فضلاً عن بيانات حقيقية إعتماداً على برنامج مصمم لهذا الغرض بلغة ماتلاب والبرنامج الإحصائي الجاهز (SPSS) وبرنامج (EasyFit)، وتوصلت الدراسة إلى كفاءة تقدير معلمات الدقة لتوزيع ويبل عند وجود مشكلة تلوث البيانات بإستخدام الطريقة المقترحة مقارنةً مع الطريقة التقليدية.

الكلمات المفتاحية: دالة البقاء، توزيع وايبل ،المويجات، البيانات الملوثة.