Scientific Inquiry and Review (SIR)
Volume 6 Issue 1, Spring 2022
ISSN $_{(P)}$: 2521-2427, ISSN $_{(E)}$: 2521-2435
Homepage: https://journals.umt.edu.pk/index.php/SIR

Article QR

Title: **Application of Univariate Probability Distributions Fitting using Monte Carlo Simulation**

Author (s): Muhammad Ilyas[1], Shaheen Abbas[2], Afzal Ali[2]

Affiliation (s): [1]Government College University, Hyderabad
[2]Mathematical Sciences Research Centre Federal Urdu University of Arts, Sciences and Technology, Karachi, Pakistan

DOI: https://doi.org/10.32350/sir.61.01

History: Received: September 20, 2021, Last Revised: January 8, 2022, Accepted: March 15, 2022

Citation: Ilyas M, Abbas S, Ali A. Application of univariate probability distributions fitting using Monte Carlo simulation. *Sci Inquiry Rev.* 2022;6(1):01-22. https://doi.org/10.32350/sir.61.01

Conflict of Interest: Author(s) declared no conflict of interest

A publication of
School of Science
University of Management and Technology, Lahore, Pakistan

Application of Univariate Probability Distributions Fitting using Monte Carlo Simulation

Muhammad Ilyas[1], Shaheen Abbas*[2] and Afzal Ali[2]

[1]Department of Mathematics, Government College University, Hyderabad
[2]Laboratory of Applied Mathematics and Data Analysis, (Al-Khwarizmi DIUI Unit), Mathematical Sciences Research Centre Federal Urdu University of Arts, Sciences and Technology, Karachi, Pakistan.

*shaheen.abbas@fuuast.edu.pk

Abstract

In this study, we showcased a univariate probability distribution through the application of the three sub- and super-exponential heavier-longer and lighter-shorter tails fitting. The univariate family includes the Lognormal, Gamma, and Weibull distributions. The adequacy of distribution tails was determined using Adequate Fitting Tests (ADTs) and a descriptive criterion. For this purpose, this study investigated the logarithm population data sets of the Karachi region from 1729 to 1946 and again from 1951 to 2018. The data contained both irregular and regular lengths and peaks. We considered three Lognormal, Gamma, and Weibull distributions for tails fitting, statistical methods for validation, and normality tests to verify stochastic behavior in both data sets. Weibull and Lognormal distribution tails were found to possess heavier distribution using two validation tests (maximum likelihood estimation and probability of correct selection). Furthermore, the univariate probability distributions were applied to Monte Carlo simulation in order to generate the actual population data. The results indicated that heavy-tailed Lognormal and Weibull distributions were adequate as compared to the more commonly used lighter tailed Gamma distribution. Hence, it was determined that the Monte Carlo Simulation performs appropriate Lognormal and Weibull distributions for irregular and regular data. Our results indicated that Lognormal-Weibull distributions are suitable for the prediction of both long-term and short-term forecasting.

Keywords*:* Anderson-Darling Test (ADT), Adequate Fitting Tests (AFT), Kolmogorov-Smirnov D-test (KST), Maximum Likelihood Estimation (MLE), probability distribution, Probability of Correct Selection (PCS)

Introduction

The Univariate Probability Distributions is a mathematical function, which determines the occurrence of probabilities of different outcomes in any random phenomenon. The set of real numbers of a sample space is known as univariate probability distribution [1,2]. A univariate probability distribution is used to disperse a probability of several outcomes of a random phenomenon. When a set of possible outcomes to any random phenomenon is countable, the probability distribution can be defined in a Probability Mass Function (*p. m. f*) for discrete or continuous random variable. The univariate probability distribution is combination of Discrete-Continuous distribution. The relationship between the distributions is shown in Figure 1.1. as showcased by Leemis [1, 3-7]. Figure 1.1 displays the relationship of five univariate probability distributions with their properties.

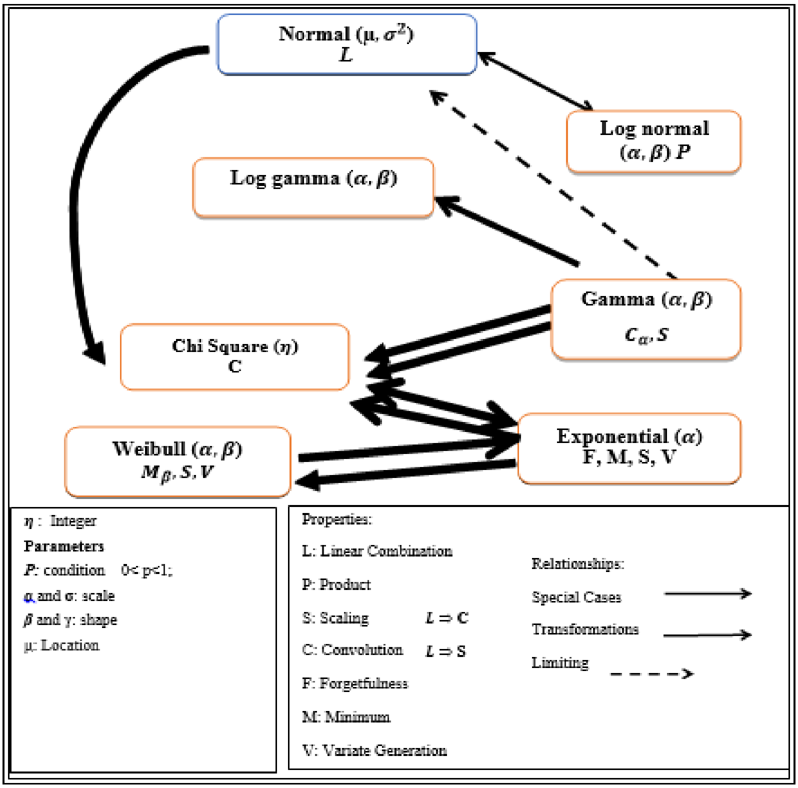

Figure 1. Connotation of Univariate Probability Distributions.

2. Univariate Distribution Properties

Linear Combination (*L*)

The property of L indicates that random independent variables for normal distribution, if $x_i \sim N(\mu_{i,\sigma^2})$ for i =1, 2, …, n.

$a_1, a_2, \dots a_{n,}$ are constants and $X_1, X_2, \dots, X_{n,}$ are independent formerly.

$$\sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i \,, \sum_{i=1}^{n} {a_i}^2 \, {\sigma_i}^2). \tag{1.1}$$

Product (*P*)

The property of P indicates that the dependent random variable is considered lognormal distribution derived from the normal distribution.

$$\prod_{i=1}^{n} X_i \sim Log\ normal\ (\sum_{i=1}^{n} \mu_i \,, \sum_{i=1}^{n} {\sigma_i}^2). \tag{1.2}$$

Forgetfulness (*F)*

The property of F is normally considered as the memory less property and singular form of the residual property, which demonstrates the conditional distribution of a variable is similar distribution. This property is applicable only on exponential distribution. Exponential distribution is associated to Weibull distribution.

Minimum (*M*)

The property of M indicates that the least of independent and equally distributed random variables from distribution are derived from the similar distribution

$$min\{X_1, X_2, \dots X_{n,}\} \sim\ exponential\ \left(\frac{1}{\sum_{i=1}^{n}\left(\frac{1}{a_i}\right)}\right). \tag{1.3}$$

Variate generation (*V*)

According to the variate v, the converse CDF of any random variable can be articulated in closed form as

$$F^{-1}(u) = -alog(1-u),\ 0 < u < 1. \tag{1.4}$$

Convolution (C) and Scaling (*S*)

The property of L implies that the C and S are not recorded, which is linear combination (L) property. Also,

$$F \Rightarrow property \quad R. \tag{1.5}$$

The Weibull distribution is unbiased when shape parameter is static. The Weibull distribution M_β indicated that the convolution and scaling is effective in this limited case. With respect to overhead univariate distributions discussion, Normal distribution (ND) can be transformed into Log normal and Gamma Distribution, Gamma Distribution in to Log gamma. In turn, Log gamma can transform in to chi-square, which can transform to Exponential and Weibull distribution, respectively. It all can be converted into Normal and Log normal Distribution [**8**-**11**]**.** To determine the suitability of the univariate distribution of the data, the aberration from regularity needs to be patterned using the adequacy test. The adequate fitting test is also used to check the fitting of the univariate tail distributions.

2.1 The Adequate Fitting Test

The adequate fitting test is performed by using Anderson-Darling test (ADT) and Kolmogorov-Smirnov test (KST) ranking, KST and ADT is formulated as:

$$E_n = \frac{n(ic)}{N}, \tag{1.6}$$

$$\mathrm{D} = \max_{0 \le i \le N} \left| F(x_i) - \frac{n(i)}{N} \right|, \tag{1.7}$$

$$A^2 = N - S, \tag{1.8}$$

$$\mathrm{S} = \sum_{i=1}^{N} \frac{(2i-1)}{N} \left[lnF(x_i) + \left(1 - F(x_{N+1-i})\right) \right]. \tag{1.9}$$

H_o the null hypothesis is rejected if, KS statistic D > critical value. The value of Probability >D is the p-value. The AD test value is greater than the critical value, prob >A^2 is the p-value, while p-value < the significance alpha level.

The adequate test detected divergences in the tails. Generally, it is used to detect the discrepancy of these tails [12-14]**.** However, KS adequate test is dependent upon the Empirical cumulative distribution function (CDF) of the hypothesized distribution. These tail inconsistencies are detected to be not fit, KS is not the most appropriate test for tail analysis [10,15]**.** An -AD test is better as compared to the KS test, since it stretches the perception evidence linked to the distribution tails. Furthermore, the fitness of this test is not dependent on the quantity of intervals. By using the selective univariate distributions and its tails analysis, the adequate fitting test is confirmed. The lower value of -AD test suggests the suitability of tail fitting, as well as sub and super (heavy and lighter) tail fitting, all of which are applicable in our data intervals.

2.1.1 Descriptive Criterion for Unbiasedness

The mean, variance is supposed that total data in a duration of sample size x, x ≤ x is stated as

$$F(x,t) = \sum_{r=0}^{\infty} P_r(t) S^r(x). \quad (1.10)$$

$P_r(t)$ =increasing population probability of r, w.r.t in the observed data interval

S(x) = Data interval probability; P≤ x,

$S^r(x)$ = the distribution function r^{th} intricacy in Data interval S(x),

where mean and variance are obtained as

$m_{f=}\ m_p m_s$ Then $\sigma_F^2 = m_p \sigma_s^2 + m_s m_p$.

Mean = $E[\theta] = e^{\left\{m+\frac{\sigma^2}{2}\right\}}$; Variance = var[x] = $\left\{e^{(2m+\sigma^2)}\right\}.\left\{e^{(\sigma^2)-1}\right\}$.

The Skewness determines the notch of irregularity of the data's

$$\text{Skewness}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^3}{(n-1)S^3}. \quad (1.11)$$

The Kurtosis trials are shown peakness of data. It is determined as

$$\text{Kurtosis} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^4}{(n-1)S^4}\ , \quad (1.12)$$

where $\bar{X}$, *S*, and *n* denote the mean, standard deviation and total samples [16]**.** The standard deviation and mean squared error defined.

Let, $\theta = (\theta_1, \theta_2, \ldots, \theta_i \ldots, \theta_m)$ be the parameter vector of a Univariate distribution. If *i* θ is selected to estimate the parameter θi, where $\theta_i \in \theta$, formerly the bias, is used to denote tails.

$$\text{bias}(\hat{\theta}_i) = E(\hat{\theta}_i) - \theta_i, \quad (1.13)$$

$$\text{MSE}(\hat{\theta}_i) = E[(\hat{\theta}_i) - \theta_i)^2], \quad (1.13)\ a$$

$$\text{MSE}(\hat{\theta}_i) = E[(\hat{\theta}_i)^2] - [E(\hat{\theta}_i)]^2 + [E(\hat{\theta}_i) - \theta_i)]^2, \quad (1.13)\ b$$

$$\mathrm{MSE}(\hat{\theta}_i) = V(\hat{\theta}_i) + [bias(\hat{\theta}_i)]^2, \quad (1.13)\ c$$

$$\mathrm{MSE}(\hat{\theta}_1) < \mathrm{MSE}\ (\hat{\theta}_2)\ \text{ then }\ \hat{\theta}_1. \quad (1.13)\ d$$

We have investigated Gamma, Lognormal and Weibull distributions tail fitting. Their appropriateness is based on the selected validation methods (PDF, PCS and MLE). The performance of the Lognormal, Weibull, and Gamma distributions are obtained by the results of mean square error (MSE) Criterion.

2.2 Sub and Super Exponential Distribution Through Validation Methods

The most fitted tail fitting probability is generally sectioned into two categories [7, 17, 18]**.** The exponetial distribution inclining to zero is found the super exponetial categrory instad to zero is sub categrory [17, 19, 20]**.** By appling these four methods for validation , the sub-super exponential distributions are articulated using tail analyses. The four methods include probability density function (PD), Exceedance Probability Function (EPF), Probability of correct selection (PCS) and Maximum Likelihood Estimation (MLE).

The distribution function F of sub exponential is stated as

$$\lim_{x\to\infty} \frac{1-F(x)}{\exp(x/\beta)} = \infty\,, \quad \forall \quad \beta > 0\,, \quad (1.14)$$

The investigation of heavy tail is more applicable for real data modeling; if the length of data interval is large size to heavy tail is more fitted as compared to light tail. Let the population and nonzero population sample be p and $p|p > 0$, conforming tail distribution for nonzero data sample, its symbolization, is distinct as per

$$P(X > p/p > 0 = \overline{F}p\backslash p > 0(x) = 1 - Fp|p > 0(p), \quad (1.15)$$

wher $Fp|p > 0(p)$. Subsequently, we only consider in the constant fragment of the distribution. Similarly, on the right tail, for representation ease, we neglect the subscript in $\overline{F}p/p > 0(p)$, symbolizing the conditional exceedance Probability Function as $\overline{F}p$ [19]. The distribution functions F and G with sustenance boundless to the right are known as tail-equivalent [21]. If $\lim x$, then 0

$$\lim_{x\to\infty} \frac{\bar{F}(x)}{\bar{G}(x)} = C\ with\ 0 < c < \infty, \quad (1.16)$$

$$\text{erf}(x) = 2\pi^{-1/2}\int_{x}^{\infty} e^{-t^2 dt}, \quad (1.16)\text{ a}$$

Probability Density Function (PDF) and Exceedance Probability Function (EPF) for the Lognormal distribution are as follows:

$$f_{LN}(X) = \frac{1}{\sqrt{\pi\gamma x}} e^{\left(ln^2(\frac{x}{\beta})^{1/\gamma}\right)}, \quad (1.17)$$

$$\bar{F}_{LN}(x) = \frac{1}{2}\,\text{erfc}\left(ln\left(\frac{x}{\beta}\right)\right)^{1/\gamma}, \quad (1.17)\text{ a}$$

The Lognormal distribution is considered a heavy-tail distribution. It includes the two scale and shape parameters $\beta > 0$ and $\gamma > 0$ that pedal the performance of the tail. Lognormal distribution belongs to the sub exponential family.

The Weibull distribution may be formed as per the interpretation of the exponential distribution. Its Probability Density Function and Exceedance Probability Function are as follows:

$$f_W(x) = \left(\frac{\gamma}{\beta}\left(\frac{x}{\beta}\right)\right)^{\frac{1}{\gamma}-1} e^{\left(-(\frac{x}{\beta})^{\gamma}\right)}, \quad (1.18)$$

$$f_W(x) = e^{\left(-(\frac{x}{\beta})^{\gamma}\right)}. \quad (1.18)\text{ a}$$

The parameters scale and shape are $\beta > 0$ and $\gamma > 0$ tail's comportment. The $\gamma < 1$ is super tail distributions [21]. For example, the generalized exponential, the Gamma and the Normal. The Weibull distribution fit in to the exponential intimate. The PDF and EPF are stated, in

$$f_G(x) = \left(\frac{1}{\beta\Gamma(\gamma)}\left(\frac{x}{\beta}\right)\right)^{\frac{1}{\gamma}} e^{\left(-(\frac{x}{\beta})\right)}, \quad (1.19)$$

$$\bar{F}_G(x) = \frac{\Gamma\left(\gamma,\frac{x}{\beta}\right)}{\Gamma(\gamma)}, \quad (1.19)\text{ a}$$

$$\Gamma(s,x) = \int_{x}^{\omega} t^{s-1}e^{-t}\,dt \text{ and } \Gamma(s) = \int_{x}^{\omega} t^{s-1}e^{-t}\,dt\;, \quad (1.20)$$

Usually, we assumed that the gamma distribution tail acts similar to the exponential tail. However, it is not tail-equivalent by the exponential.

$$\lim_{n\to\infty}\frac{\bar{F}(x)}{\bar{G}(x)} = f(x) = \begin{cases} 0 & 0<\gamma<1 \\ 1 & \gamma<1 \\ \infty & \gamma<1 \end{cases} \quad (1.21)$$

In Case, $0 < \gamma < 1$, the Gamma distribution takes the lighter tail; whereas for $\gamma > 1$, it takes the heavier tail.

3. Maximum Likelihood Estimation and Propability of Correct Selection

The Lognormal and Weibull distributions belong to sub exponential heavy-tailed and, sub- super exponential, which are reliant on the shape parameter value, Gamma distribution has a light exponential tail distribution[2]**.** Statistics suggest that heavier-tailed distributions Lognormal and Weibull are better than lighter-tailed complements. Furthermore, these distributions are used to determine the probability of correct selection (PCS), which is used to pick the accurate model.

Now, The PDF of a Lognormal by scale and shape parameter $\eta > 0$ and $\sigma > 0$ for Lognormal and Weibull is

$$f_{LN}(x;\sigma,\eta) = \frac{1}{\sqrt{2\pi x\sigma}}e^{-\frac{1}{2}\left(\frac{lnx-ln\eta}{\sigma}\right)^2}; x > 0, \quad (1.20)$$

$$f_{WE}(x;\beta,\theta) = \beta\theta^{\beta}x^{(\beta-1)}e^{-(\theta x)^{\beta}}; x > 0. \quad (1.21)$$

It is denoted in LN (σ, η) and WE (β, θ) correspondingly. The PDF of a Weibull distribution, by shape $\beta > 0$ and scale $\theta > 0$ parameters is [22-24]. We take ($\widehat{\beta}\,\widehat{\theta}$) and($\widehat{\sigma}\,\hat{\eta}$) as Maximum Likelihood Functions of the WE (β, θ) and LN (σ, η) parameters [25].

We selected the Weibull or Lognormal as the suitable fitted tail distributions if

$$T_n = L_{WE}(\widehat{\beta},\hat{\theta}) - L_{LN}(\hat{\sigma}\hat{\eta}) \quad > 0 \text{ or } < 0. \quad (1.22)$$

The parameters($\widehat{\beta}\,\widehat{\theta}$) and ($\widehat{\sigma}\,\hat{\eta}$) symbolize the log-likelihood functions of the Lognormal, Weibull may be formed as paths, if

$$L_{WE}(\beta,\theta) = \sum_{i=1}^{r} lnf_{WE}(X_i;\beta,\theta) + (n-r)\ln\left(1 - F_{WE}\left(X_{(r)};\beta,\theta\right)\right). \quad (1.23)$$

$$L_N(\sigma,\eta) = \sum_{i=1}^{r} lnf_{LN}(X_i;\sigma,\eta) + (n-r)\ln\left(1 - F_{LN}\left(X_{(r)};\sigma,\eta\right)\right). \quad (1.24)$$

We assume, LN (σ, η) and WE (β, θ) are fitted for data sets, in equation (3.27), appropriate tail Tn by mean $\mu_{LN}(T_n)$ and $\mu_{WE}(T_n)$ and variance $\sigma_{LN}(T_n)$ and $\sigma_{WE}(T_n)$. If the data shown is best fitting for LN (σ, η) then, T_n is designate $\mu_{LN}(T_n)$ and $\sigma_{LN}(T_n)$. Hence, it is proven that the $PCS_{LN} = P(T_n \leq 0)$ and $PCS_{WE} = P(T_n > 0)$. Weibull is selected if Tn > c and Lognormal if otherwise The PCS and Mean square error estimations method of validation confirmed that heavier-tailed Lognormal and Weibull distributions contract in the actual /real population data sets and are called lighter tailed distribution.

3.1 Normality Test

The Cumulative Hazard Function (CHF) plot is comprised the data inconsistencies and the physical spectacles is intricate [14, 26]. The application of P-P plots is dependent upon the assessment of the Cumulative Distribution Function (CDF) in the distribution of a fitted curve. The CHF plot is indicated the rapid probability, data inconsistencies and the physical spectacles is intricate. The CDF and CHF plots are

The CDF and CHF are indicated $FX(x) = P(X \leq x)$ and

$$H(t)X \overset{def}{=} \int_0^t h(u)\, du \; t > 0.$$

It tried to find shapes and stimulate novel structures [14, 26]. The descriptive Criterion for unbiasedness tests is compared for modelling and data simulation using the Monte Carlo probability distribution simulation technique. Then, formerly seeing the Monte Carlo simulation, some concepts associated to the estimators are distinct and how one can choose a good estimator.

4. The univariate Monte Carlo simulation erudiotion

Monte Carlo probability simulation computational technique can be used to recognize on ecological systems, such as population forecasting, financial, project management, forecasting models. These are valuable simulation models and are used when forecasting an indefinite future. In a Monte Carlo simulation is depend on the array of evaluations. This simulation is computes to the parameters of the model millions of times. Each time, it randomly selects different values [27, 28]. In this research data, are simulated using the above selected univariate distributions. Additionally, Lognormal and Weibull distributions are applied to evaluate the parameter values. In this section, normality tests are compared with modelling and data simulation using the Monte Carlo probability distribution simulation technique. These simulation methods can be used to generate samples of various data sizes from Weibull and Lognormal models using the initial parameters. The performances of the defined estimators were compared with maximum likelihood estimators using simulation. For this purpose, we used several criteria for simulation and estimations, that is, bias, mean squared error, mean absolute deviation, mean relative total error and running time of the algorithm. The proposed techniques have been illustrated using real data sets of irregular and regular logarithm population data sets of Karachi region for the last three centuries (from 1729 to 1946 and from 1951 to 2018).

5. Results and Discussion

Univariate Distributions Fitting to the Irregular and Regular Datasets

Table 1. The Descriptive statistics of Lognormal appropriate distribution tails in to the 1729-2018 samples of annual population data

Data intervals / sample	Adequacy Fitting test ADT, KST		Mean	SD	Variance	Coefficient of variance	Skewnes	Kurtosis	SD
1729-1946	0.28722	0.14414	1.0123	2.7787	5.72124	7.4491	2.07639	0.0248	2.7787
1951-2018	2.437	0.16153	1.3663	3.3223	2.8327	0.8360	3.09229	0.7963	3.3223
1729-1798	2.2291	0.31678	5.9292	4.9176	2.4183	82.938	5.70771	2.2409	4.9176
1810-1897	1.1789	0.2681	2.9658	16.684	2.7836	5.6253	1.94891	1.2084	16.684
1901-1946	0.2872	0.14414	2.6835	14.896	2.2191	0.5551	1.83640	1.5393	14.896
1951-2000	2.152	0.1412	5.1496	4.0317	1.6255	0.7829	2.82869	1.6666	4.0317
2001-2018	0.18526	0.13096	1.0610	21.9315	4.8099	0.0206	0.0646	0.00745	21.9315

These results were determined using two real data sets including logarithm population irregular centuries and regular decades wise intervals (from 1729 to 1946 and from 1951to 2018). Selective univariate distributions and its right-skewed (right heavier-lighter tails) analysis was used to evaluate the fitted lognormal, gamma and Weibull distribution to forecast and irregularity. For this purpose, adequate fitting Tests such as Anderson-

Darling (ADT) test and Kolmogorov-Smirnov *D*-test (KST) were used. The statistical results for these selective univariate distributions are depicted in Table 1 to 3.

Table 2. The Descriptive statistics of Gamma appropriate distribution tails into the 1729-2018 samples of population data

Data intervals	Adequacy Fitting test ADT, KST		Mean	SD	Variance	Coefficient of variance	Skewness	Kurtosis
1729-1946	6.4362	0.42195	9.2204	2.9161	8.5041	3.1627	6.3255	6.018
1951-2018	3.3567	0.17004	6.1108	3.5323	1.2477	0.5780	1.1560	2.0048
1729-1798	3.3948	0.47423	1.2075	3.0694	9.4214	2.5418	5.0836	3.8764
1810-1897	4.148	0.56957	1.1167	3.8469	1.4799	3.4448	6.8895	7.1199
1901-1946	0.3199	0.14218	2.6939	1.5529	2.4117	0.5764	1.1529	1.9938
1951-2000	3.1424	0.16102	5.0310	3.0619	9.3756	0.6086	1.2172	2.2224
2001-2018	0.16334	0.11918	1.0610	2.1873	4.7845	0.02061	0.0412	0.0025

The descriptive statistical tests (Mean, Median, Standard Deviations, Variance, Skewness and Kurtosis) and ADT are computed for well fitted distribution. All of the data intervals illustrate prolonged Right-Tail.

Table 3. The Descriptive statistics of Weibull appropriate distribution tails into the 1729-2018 Samples of annual population data

Data interval Sample size	Adequacy Fitting test ADT, KST		Mean	SD	Variance	Coefficient of variance	Skewness	Kurtosis
1729-1946	1.3585	0.14263	2.7835	7.5014	5.6271	2.6949	0.8713	1.6413
1951-2018	1.7646	0.13694	6.1661	4.0287	1.6231	0.6533	0.9998	0.1587
1729-1798	2.5985	0.33042	2.7658	1.6520	2.7292	5.9730	3.3668	3.4848
1801-1898	2.4208	0.20246	5.9682	6.613	4.3733	1.1080	2.3295	8.4068
1901-1946	0.52703	0.17481	2.4542	1.3361	1.7872	0.5447	0.6934	0.3680
1951-2000	1.5815	0.130111	4.9726	3.1106	9.6762	0.6255	0.9215	0.9269
2001-2018	0.31661	0.14476	1.0568	2.6218	6.8740	0.0248	-1.0268	1.8868

Subsequently, Lognormal and Weibull distributions are found heavy right long tail distribution for the total intervals while mean parameter value is greater than other distributions that displayed a high value of Skewness or kurtosis. The Anderson-Darling (AD) test value also confirmed this finding. The AD test confirmed Mean-Tail analysis for all selected data intervals. In our case, the data values showed heavy right tails.

The Sub and Super Exponential Right Tail-Fitting for Entire Datasets

Table 4. The Sub and Super Exponential Right Tail-Fitting

Interval	Model	Shape/Location Parameter	Scale Parameter	Log likelihood	SSE	Fitted Model
1729-1946	Lognormal	2.5741	3.607	-99.43	2.151	Log-normal
	Weibull	1.0682	0.44024	-99.23	2.910	
1729-1798	Lognormal	2.9726	1.9669	-87.20	0.002	Log-normal
	Weibull	24.11	0.28673	-86.50	0.208	
1810-1897	Lognormal	1.867	3.9495	-70.34	0.032	Log-normal
	Weibull	56.856	0.90391	-70.70	0.298	
1901-1946	Lognormal	0.51828	5.458	-55.00	0.015	Log-normal
	Weibull	2.7663	1.9106	-56.28	0.230	
1951-2000	Lognormal	0.691432	15.2154	-43.14	3.754	Log-normal
	Weibull	5.55827	1.64022	-43.64	5.91	
2001-2018	Lognormal	0.01979	16.177	-33.14	4.393	Log-normal
	Weibull	9.7358	24.542	-34.64	3.790	
1951-2018	Lognormal	0.728	1.5402	-98.43	4.33	Weibull
	Weibull	6.8623	1.564	-97.03	5.91	

The Gamma, Weibull and lognormal distribution are obtained by the help of validation tests, so these three distributions are considered sub-super exponential heavy light tails analysis are depicted in Table 4 and Figures 2(a-c) and 3(a-c). The lognormal and Weibull distribution is found Right-Heavy tails and Gamma is Right-Light tail. We have noted that heavy tails incline, particularly once the data is minor. Therefore, even when the Gamma tail performed well, heavy-tailed distributions were preferred for long span of prediction.

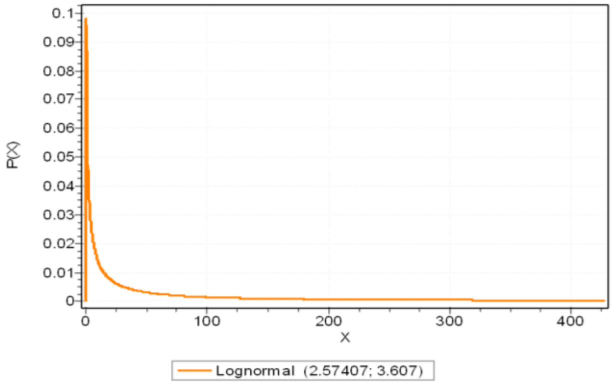

Figure 2(a). The Irregular interval shape parameters of Distribution tails.

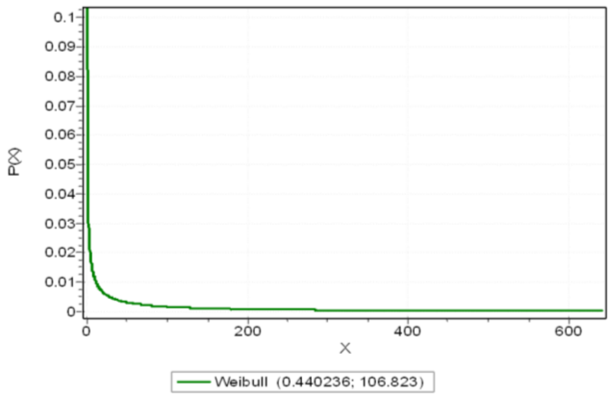

Figure 2(b). The Irregular interval shape parameters of Distribution tails

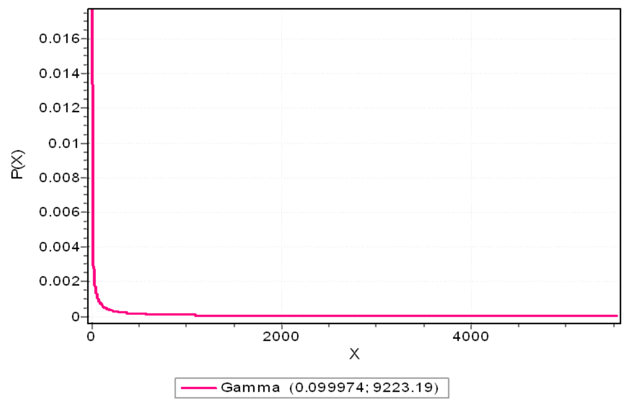

Figure 2(c). The Irregular interval shape parameters of Distribution tails

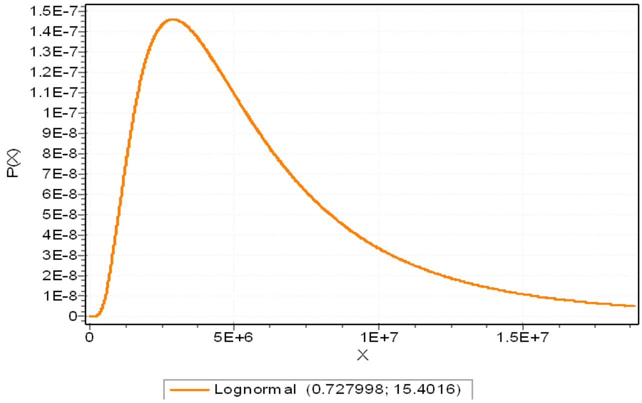

Figure 3(a). The regular interval shape parameters of Distribution tails

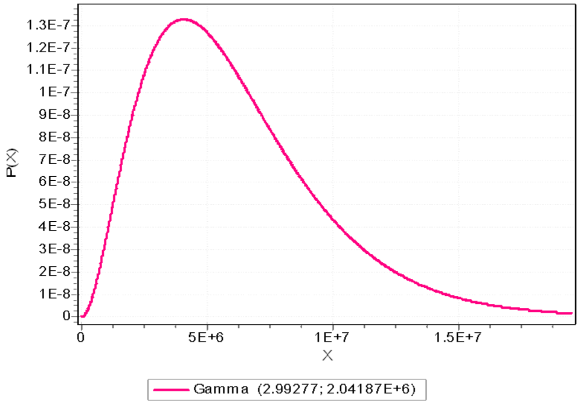

Figure 3(b). The regular interval shape parameters of Distribution tails

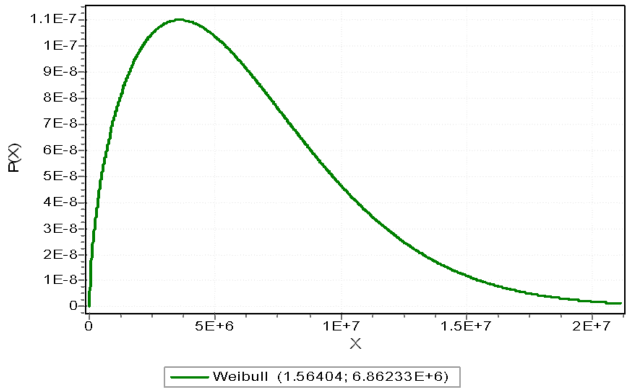

Figure 3(c): The regular interval shape parameters of Distribution tails

Afterwards, we applied methods of validations to asses reliability and application of Sub-Super heavy and light tails analysis. Table 4 results indicate that all intervals performed Lognormal and Weibull heavy tails. The Lognormal ($\hat{\sigma}$, $\hat{\eta}$)= (2.5741, 3.607); Weibull (β, θ) = (1.0682, 0.44024) for regular data set, for total irregular data interval is Lognormal ($\hat{\sigma}, \eta$)= (0.728, 1.5402) and Weibull (β, θ) = (6.8623, 1.564). The log–likelihood values of lognormal (-99.43) for irregular, Weibull (-99.23) distributions, total regular Lognormal (-98.43) as well as Weibull (-98.03) $WE_{1729-1946}$ **(1.0682, 0.44024),** $LN_{1729-1946}$ **(2.5741, 3.607),** $LN_{1951-2018}$ **(0.728, 0.5402) and** $WE_{1951-2018}$ **(6.8623, 1.564)** are depicts in Table 4 and Figure 4(a-b), the familiarity of the distribution functions.

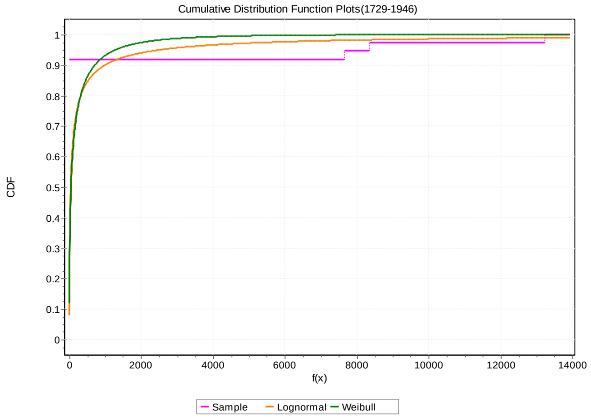

Figure 4(a). The CDF for irregular interval and regular intervals.

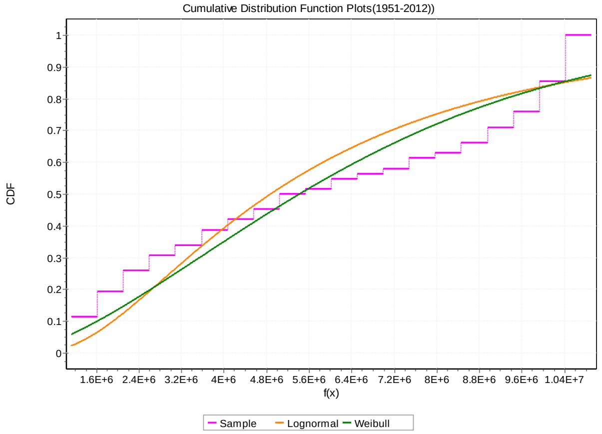

Figure 4(b). The CDF for irregular interval and regular intervals.

The CHF of over lapping Lognormal and Weibull distribution functions for both intervals is depicted in Figure 5(a-b). The figure CDF is shown smoothly, result is confirmed lognormal tail is more appropriate for irregular data interval.

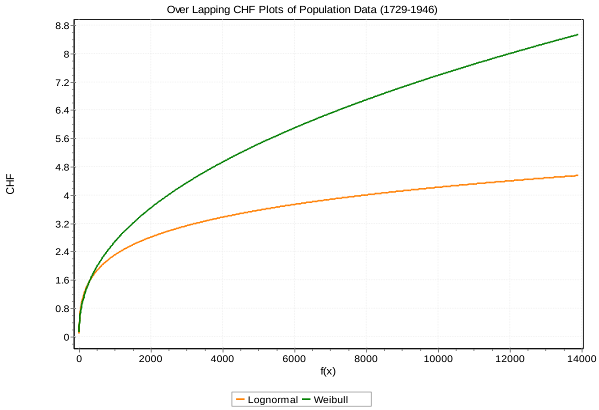

Figure 5(a). The CHF over lapping distributions WE (1.0682, 0.44024), LN (2.5741, 3.607) LN (0.45, 1.50); (1.67, 0.48) and LN (0.45, 1.50).

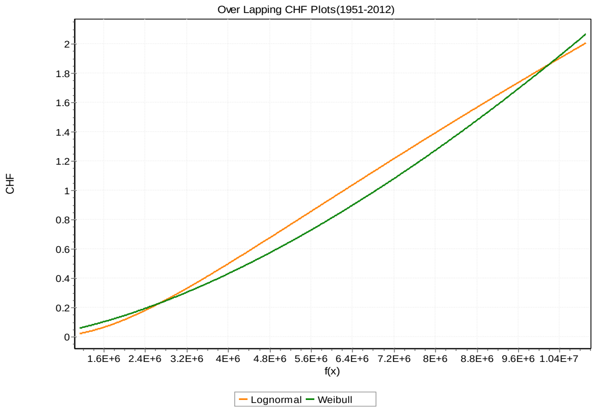

Figure 5(b). The CHF over lapping distributions WE (1.0682, 0.44024), LN (2.5741, 3.607) LN (0.45, 1.50); (1.67, 0.48) and LN (0.45, 1.50).

The Monte Carlo simulation is a beneficial procedure for modeling and evaluating real-world circumstances. Monte Carlo simulation are also used to hind cast and forecast data intervals.

6. Simulation

Monte Carlo simulation are used to simulate and forecasts the data. According to the results, the Lognormal and Weibull distributions are found to generate the 288 and 69 data values. In this section, the use of the Monte Carlo simulation method is described data samples of different sizes, according to which the Lognormal distribution and Weibull distribution seem to be the most apt distributions for the all-data intervals. For the first simulation, data from 1729 to 1929 was used and for the second simulation, data from 1951 to 2000 was used.

Table 5. Monte Carlo simulation of probability distributions and statistical analysis for population data (1729-2020)

Population Data Sets	Distribution	Parameters (α, β, γ) (σ, μ)	ADT (α=0.05)	median	mode	variance	Skewness	Kurtosis
1729-1929	LND	2.5778, 3.3179	0.82776	14	$2.18e^3$	$9.5303e^6$	3.3708	10.972
1930-1950	LND	2.529, 2.875	0.11864	0.82776	-	$1.2309e^6$	4.5172	20.567
1729-2020	LND	2.9053, 3.2819	0.20899	21.435	-	$9.3351e^6$	6.835	49.656
1951-2000	GD	474.56, 0.03206	1.0237	$4.2903e^6$	$6.77e^5$	$9.3756e^{12}$	0.38083	-1.3319
2001-2020	GD	855.88, 0.01791	0.20367	$4.4114e^6$	$2.10e^5$	$6.9994e^{12}$	0.82939	0.06987
1951-2020	WD	1.2958, $4.0119e^6$, $1.3915e^6$	0.32631	$4.3349e^6$	-	$8.0111e^{12}$	1.1929	1.5804
2001-2018	GD	3.5557, $1.5755e^6$	0.15759	$4.9150e^6$	1.62215	$8.8263e^{12}$	0.69557	-0.26682

The remaining data was used. Finally, Monte Carlo stochastic simulation forecasts data values were generated from 1930 to 1950 and 1729 to 2020 for the first data set. The lognormal distribution is fitted for the data interval from 1930 to 1946 respectively, simulation is verified. Finally, the duration of the data set taken from 1729 to 2020 was determined by using the

Lognormal from 1951 to 2020 and 2001 to 2020 for the second data set. Using Gamma and Weibull parameters data interval from 2001 to 2020, total interval 1951 to 2020 are also simulated. These Monte Carlo simulated outcomes are showed in Table. 1.5 - Figure 1.6. The mean and standard deviation of the generated data for the intervals 1729 to 1946 and 1951 to 2000, 2001 to 2020, 1951 to 2020, and 2001 to 2018 was (503.11 ± 30.620), (641.79 ± 4.154), (631.24 ± 3.932), (689.09 ± 24.737) and (677.09 ± 4.737), respectively. Mean values are greater than the modes values, indicating the presence of heavy gradually lengthier tails. Consequently, the shape of the distribution is skewed.

Our results indicated that the Lognormal distribution is most appropriate for total irregular (from 1729 to 1946) interval, while the Weibull distribution is adequate for the total regular (from 1951 to 2018) interval. Additionally, the Log Normal and Weibull distributions are found to be the most suitable for both intervals.

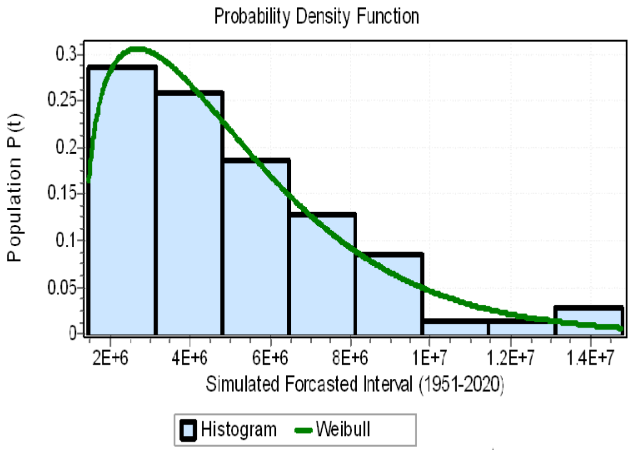

Figure 6. Weibull heavy right tail distribution fitting by Monte Carlo Simulate population data (1951 -2020).

7. Conclusion

This study is computed appropriateness of Lognormal, Gamma and Weibull distribution along with sub and super tail fitting. Our results illustrated that Lognormal and Weibull are best fitted for most intervals, while Gamma is the better fitted for irregular (small) size of data intervals. We also performed Monte Carlo probability simulation, whose outcome indicated that regular data set results are verified.

Acknowledgement

We are thankful to the **Pakistan Bureau of Statistics** for providing population data and Higher Education Commission (HEC) Pakistan for financial support (vide NRPU grant #20-4039/R&D/HEC/14/697).

References

[1] Maio C, Schexnayder C. Simulation model probability distribution functions: Relationships between input data and goodness-of-fit tests. In*IAARC/IFAC/IEEE. International symposium 1999* (pp. 103-108).

[2] Silveira FV, Gomes-Silva F, Brito CC, et al. Modelling wind speed with a univariate probability distribution depending on two baseline functions. *arXiv preprint arXiv:2101.03622*. 2021. https://arxiv.org/abs/2101.03622

[3] Leemis LM, McQueston JT. Univariate Distribution Relationships. *Am Stat*. 2008;62(1):45-53. https://doi.org/10.1198/000313008X270448

[4] Johnson NL, Kotz S, Kemp AW. *Univariate discrete distributions*. John Wiley & Sons; 2005.

[5] Johnson NL, Kotz SI, Balakrishnan N. Beta distributions. *Continuous univariate distributions*. 2nd ed. New York, NY: John Wiley and Sons. 1994:221-35.

[6] Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. John Wiley & Sons; 2011.

[7] Kale MD, Butar FB. *Fractal analysis of time series and distribution properties of Hurst exponent* (Doctoral dissertation, Sam Houston State University).

[8] Hassan D, Abbas S, Ansari MR, Jan B. The study of sunspots and K-index data in the perspective of probability distributions. *Int J Phys Social Sci*. 2014;4(1):23-41.

[9] Vose, David. *Risk analysis: a quantitative guide*. John Wiley & Sons;2008.

[10] Leemis and M. Lawrence. Relationships among common univariate distributions. *Am Stat*. 1986;40:143-46.

[11] Protassov, Rostislav S. EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ. *Stat Comput*. 2004;14:67-77. https://doi.org/10.1023/B:STCO. 0000009419.12588.da

[12] Johnson, Norman L., Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous univariate distributions,* John wiley & sons; 1995.

[13] Kim, Jin Seon, and Bong-Jin Yum. Selection between Weibull and lognormal distributions: A comparative simulation study. *Comput Stat Data Analy*. 2008;53:477-485. https://doi.org/10.1016/j.csda. 2008.08.012

[14] Metzler, Ralf, and Joseph Klafter. The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *J Phys A: Math Gen.* 2004;37:R161.

[15] Papalexiou SM, Koutsoyiannis D, Makropoulos C. How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrol Earth System Sci*. 2013;17(2):851-62.

[16] Gibbons JD, Chakraborti S. Nonparametric statistical inference. CRC press; 2014.

[17] Michael JR. The stabilized probability plot. *Biometrika*. 1983;70(1):11-7.

[18] Werner T, Upper C. Time variation in the tail behavior of Bund future returns. *J Futures Markets: Futures, Options, and Other Derivative Products*. 2004;24(4):387-98.

[19] Kundu D, Manglick A. Discriminating between the Weibull and log-normal distributions. *Naval Res Logist (NRL).* 2004;51(6):893-905. https://doi.org/10.1002/nav.20029

[20] Prabhu SS, Montgomery DC, Runger GC. A combined adaptive sample size and sampling interval X control scheme. *J Quality Technol.* 1994; 26(3):164-76. https://doi.org/10.1080/00224065.1994.11979524

[21] Corrado CJ, Su T. Implied volatility skews and stock index skewness and kurtosis implied by S&P 500 index option prices. *J Derivat*. 1997; 4(4):8-19.

[22] Shpilberg, David C. The probability distribution of fire loss amount. *J Risk Insur*. 1977;103-115. https://doi.org/10.2307/251860

[23] Améndola C, Kohn K, Reichenbach P, Seigal A. Invariant theory and scaling algorithms for maximum likelihood estimation. *SIAM J Appl Algebr Geom*. 2021;5(2):304-37.

[24] Klüppelberg, Claudia. Subexponential distributions and integrated tails. *J Appl Prob*. 1988;25:132-141. https://doi.org/10.2307/3214240

[25] Law, Averill MW. David Kelton, and W. David Kelton. *Simulation modeling and analysis*. New York: McGraw-Hill; 2000.

[26] AbouRizk, Simaan M, Daniel W, Halpin, James R, Wilson. Fitting beta distributions based on sample data. *J Construct Eng Manag*. 1994;120:288-305. https://doi.org/10.1061/(ASCE)0733-9364(1994)120:2(288)

[27] AbouRizk, Simaan M., and Daniel W. Halpin. Probabilistic simulation studies for repetitive construction processes. *J Construct Eng Manag*. 1990;116:575-594.https://doi.org/10.1061/(ASCE)0733-9364(1990)116:4(575)

[28] Bakhtiari, Hamed, Jin Zhong, and Manuel Alvarez. Predicting the stochastic behavior of uncertainty sources in planning a stand-alone renewable energy-based microgrid using Metropolis–coupled Markov chain Monte Carlo simulation. *Appl Energy*. 2021;290:116719. https://doi.org/10.1016/j.apenergy.2021.116719