Applications of Some Classical Distributions in Modelling Lifetimes Datasets with Varying Shapes

J. Abdullahi[1], A. S. Mohammed[1*], B. Abba[2], N. Umar[3], A. I. Ishaq[1] and Y. Zakari[1]

[1]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria
[2]Department of Mathematics, Yusuf Maitama University, Kano, Nigeria
[3]Department of Mathematics and Statistics, University of Strathclyde, Glasgow UK.
Corresponding author's email: aminusmohammed@gmail.com

Abstract

Statistical distributions are used in various fields such as reliability engineering, survival analysis, computer science, social sciences, among others, in modeling specialized data. The Exponential and Weibull distributions are more widely used in practice than the Gamma distribution. Both Gamma and the Weibull distributions have the same number of parameters and have some advantages over the conventional Exponential distribution in that they possess varying hazard rate shapes. The nature of Exponential, Weibull, and Gamma distributions were studied using different graphs of their probability density functions and cumulative distribution functions. The expressions for some of their properties, including hazard and survival functions are presented. Five data sets with varying shapes ranging from right-skewed to left-skewed were considered in the analysis. The total time on the test (TTT) plot for the data sets indicate increasing, decreasing, unimodal, and bathtub shapes. The performance of the competing models was assessed using some Goodness-of-fit measures and the results show that in some cases exponential distribution provided a better fit than the Weibull and Gamma distributions most especially when the data sets skewed to the right whereas in other cases Weibull and Gamma distributions provided a better fit than the exponential distribution.

Keywords: Exponential distribution, Weibull distribution, Gamma distribution, hazard rate.

1. Introduction

Statistical distributions are used in various fields such as reliability engineering, survival analysis, computer science, and social sciences, among others, in modeling specialized data. Continuous or discrete probability distributions are the two types of probability distributions. A continuous distribution has an infinitely wide range of values, making it uncountable. Exponential, Gamma, and Weibull distributions are some examples of continuous distributions that can be used to model lifetime data. The Exponential and Weibull distributions are more widely used in practice than the Gamma distribution. Both the Gamma and the Weibull distributions have the same number of parameters and have some advantages over the conventional Exponential distribution in that they possess varying hazard rate shapes. According to the event of interest in all areas, the times to the occurrences of events are known as Lifetimes (Singer & Willett, 1991). Statistics experts and researchers in fields including engineering, medicine, and biology have all shown a keen interest in the statistical analysis of lifetime data. Lifetime distributions are used in a variety of fields, such as engineering and biomedical sciences to study human disorder (Berezin &Achilefu, 2010).

This research examined the features of the Exponential, Weibull and Gamma distributions by creating various probability density graphs for some parameter values. A number of real-life data sets from many scientific disciplines have been taken into consideration, and an effort has been made to compare the goodness-of-fit of the exponential, Weibull and Gamma distributions to determine which is better. When estimating the amount of time until the next event, such as a success, failure, or arrival, exponential distribution is frequently utilized. For instance, you can use exponential distribution to forecast: How long it will take a consumer to make a purchase at your store (success).In studies on the life spans of produced goods to other fields, the exponential distribution was the first lifetime distribution model that was frequently utilized (Davis,1952). The studies on survival or remission times in chronic diseases can be proper handled by an Exponential distribution (Feigl & Zelen, 1965).

Shi et al., (2013) proposed a two-parameter Weibull distribution model to describe the aggregate gradation, and the applicability of the Weibull model and the fractal model was compared for different gradations. Comparison of eight methods of Weibull distribution for determining the best-fit distribution parameters with wind data measured from the met-mast (Yaniktepe et al., 2023). A study of freeze–thaw damage evolution equation and a residual strength prediction model for porous concrete based on the Weibull distribution function(Qu et al., 2023), Alrashidi (2023) study the estimation of Weibull distribution parameters for wind speed characteristics using neural network algorithm by (Alrashidi, 2023). This research focused in exploring the flexibility of Exponential, Weibull and Gamma distributions in modelling lifetime data sets with varying density shapes. The data sets were gotten from (Shanker et al., 2015).

2. Material and Methods

2.1. Exponential Distribution

The Exponential distribution (ED) is considered to be an important lifetime distribution and has been used by many researchers in diverse research areas including engineering, insurance among others. The distribution and density ED are, respectively, given as;

$$F(w;\alpha) = 1 - e^{-\alpha w} \tag{1}$$

and

$$f(w;\alpha) = \alpha e^{-\alpha w} \tag{2}$$

For, $w, \alpha > 0$

2.2. Some Properties of Exponential Distribution

The Survival and Hazard functions are respectively given as;

$$S(w) = 1 - F(w)$$

$$= 1 - \left(1 - e^{-\alpha w}\right)$$

$$S(w) = e^{-\alpha w} \tag{3}$$

and

$$h(w) = \frac{f(w)}{1 - F(w)}$$

$$h(w) = \frac{\alpha e^{-\alpha w}}{e^{-\alpha w}}$$

$$h(w) = \alpha \tag{4}$$

The Mean and Variance are, respectively, given as;

$$Mean = \frac{1}{\alpha} \text{ and } variance = \frac{1}{\alpha^2}$$

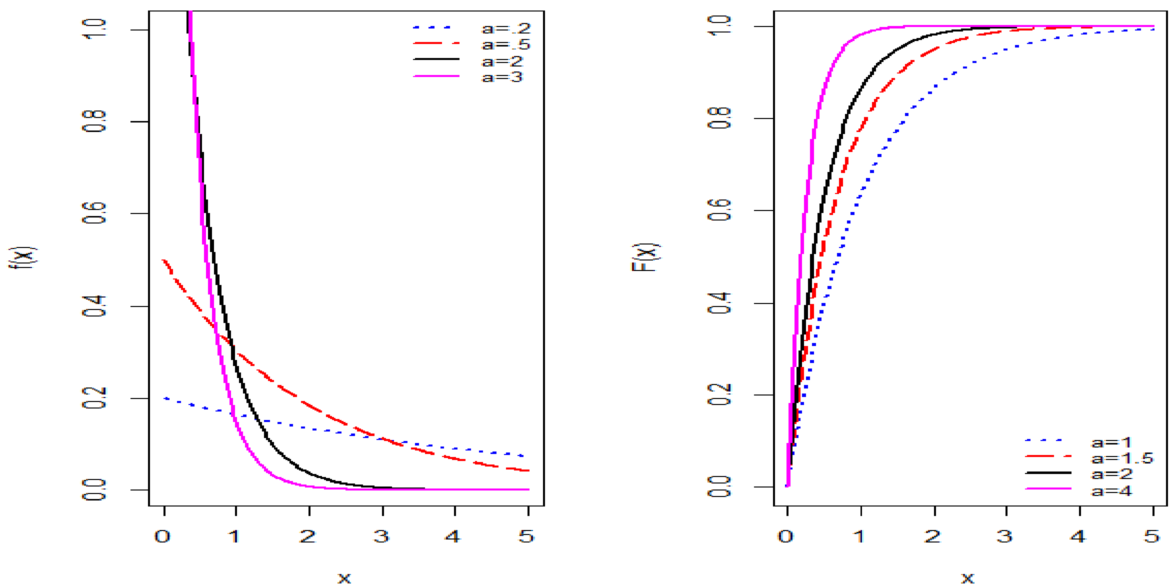

Fig.1. The PDF and CDF of Exponential Distribution

2.3. Weibull Distribution

The Weibull distribution (WD) is another lifetime distribution that is considered to be powerful when modelling a skewed data set, and its distribution and density functions are respectively given as;

$$F(w;\alpha,\beta) = 1 - e^{-\left(\frac{w}{\alpha}\right)^{\beta}}$$

and

$$f(w;\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{w}{\alpha}\right)^{\beta-1} e^{-\left(\frac{w}{\alpha}\right)^{\beta}}$$

For $w, \alpha, \beta > 0$

If beta is greater than 1, it means that the failure rate falls over time. This occurs when there is a high rate of "infant mortality," or when defective products fail prematurely, with the failure rate gradually declining over time as the defective products are eliminated from the population. When β = 1, the failure rate is said to be consistent over time. This can imply that arbitrary external circumstances are causing failure or fatality. The Weibull distribution reduces to an exponential distribution; A value of β > 1 indicates that the failure rate increases with time. This happens if there is an "aging" process, or parts that are more likely to fail as time goes on.

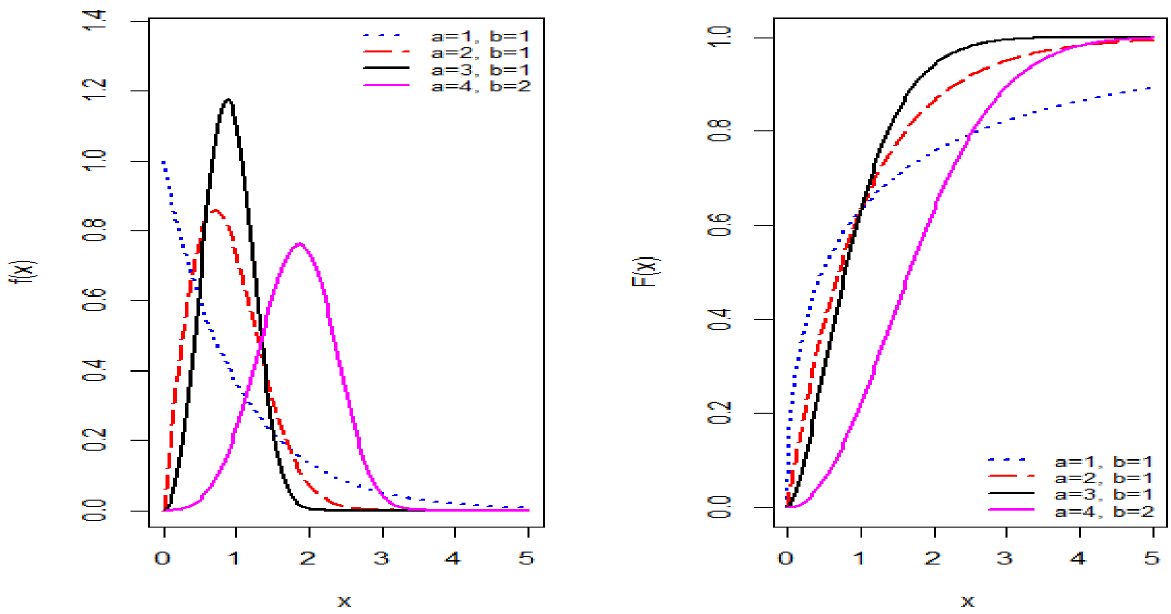

Fig.2. The PDF and CDF of WD

2.4. Some Properties of Weibull Distribution

The mean and variance are respectively as;

$$Mean = \alpha\,\Gamma\left(1 + \frac{1}{\beta}\right)$$

And

$$Variance = \alpha^{2}\left(\Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right)\right)^{2}\right)$$

The Survival and Hazard functions are respectively given as;

$$S(w) = 1 - F(w)$$

$$=1-\left(1-e^{-\left(\frac{w}{\alpha}\right)^{\beta}}\right)$$

$$S(w)=e^{-\left(\frac{w}{\alpha}\right)^{\beta}}$$

and

$$h(w)=\frac{f(w)}{1-F(w)}$$

$$=\frac{\frac{\beta}{\alpha}\left(\frac{w}{\alpha}\right)^{\beta-1}e^{-\left(\frac{w}{\alpha}\right)^{\beta}}}{e^{-\left(\frac{w}{\alpha}\right)^{\beta}}}$$

$$h(w)=\frac{\beta w^{\beta-1}}{\alpha^{\beta}}$$

2.5. Gamma Distribution

The Gamma distribution (GD) is another lifetime distribution that is considered to be powerful when modelling a skewed data set and its distribution and density functions are, respectively, given as;

$$F(w;\alpha,\beta)=\frac{\alpha^{\beta}}{\Gamma\beta}\gamma\left(\beta,\alpha w\right)$$

and

$$f(w;\alpha,\beta)=\frac{\alpha^{\beta}}{\Gamma\beta}w^{\beta-1}e^{-\alpha w}$$

For $w,\alpha,\beta>0$.

When β = 1, then the distribution collapse to an exponential distribution.

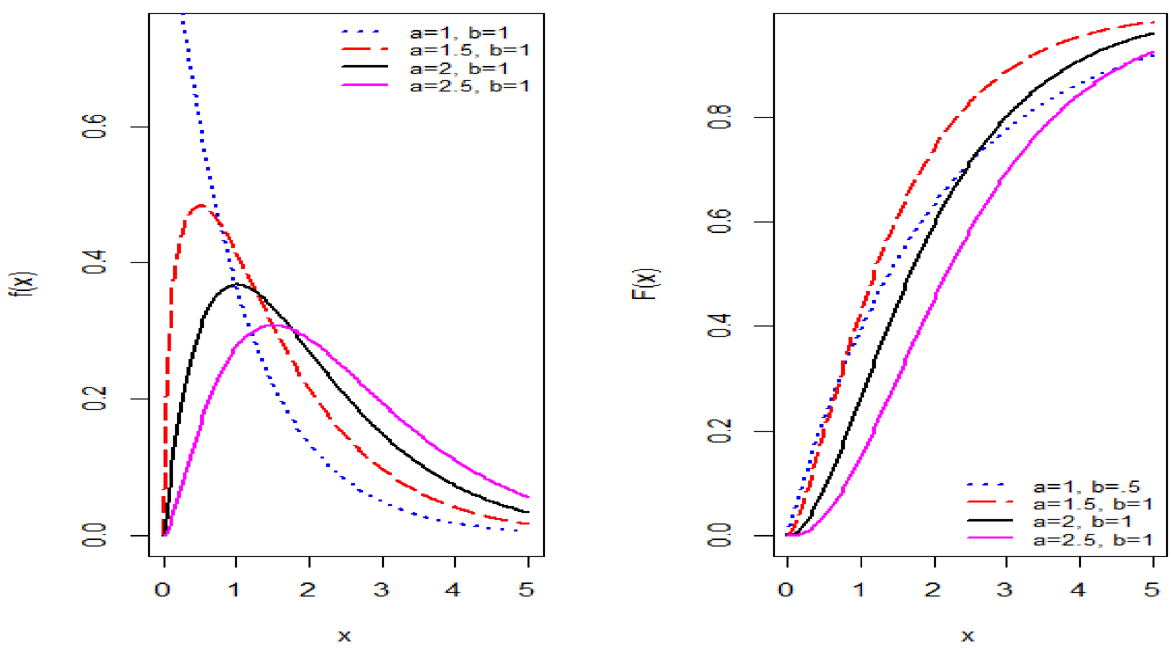

Fig.3. The pdf and cdf of Gamma distribution

The Mean and Variance of Gamma distribution are given as;

$$Mean = \frac{\beta}{\alpha}$$

and

$$Variance = \frac{\beta}{\alpha^2}$$

3. The Total Time on Test (TTT) Function

The failure rate function of a random variable T can exhibit a wide range of values. To acquire empirical behavior of the failure function, graph of total test time (TTT curve) may be utilized (Aarset, 1987).

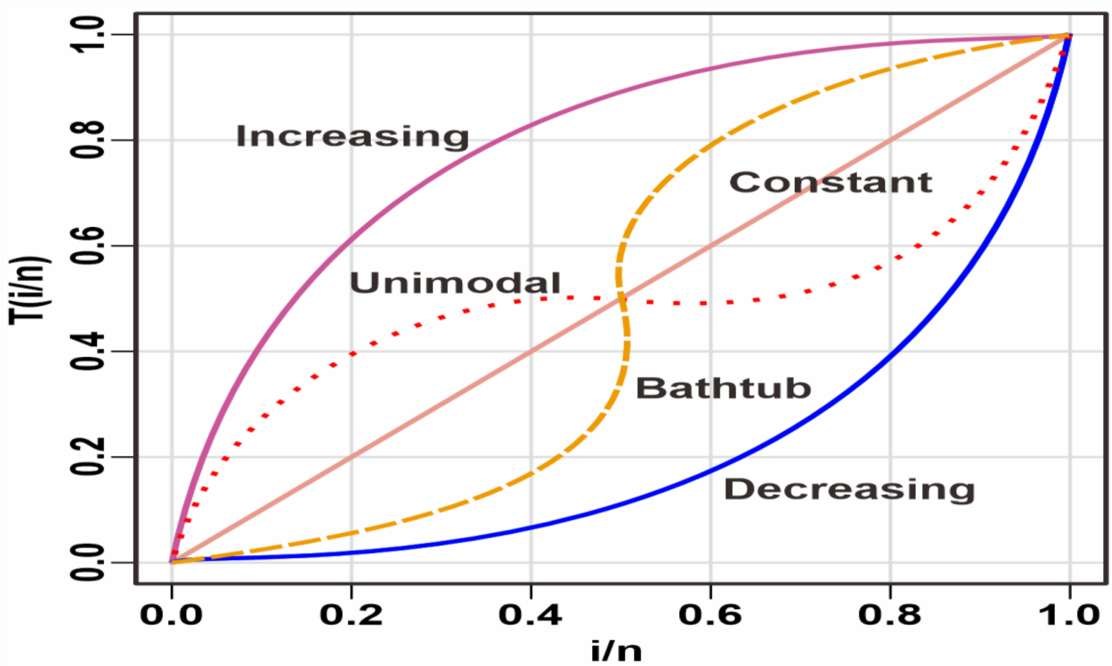

Fig.4. The plot of TTT function

According to Aarset (1987), a constant failure rate is sufficient if the curve approaches a straight diagonal function; if the curve is convex or concave, a monotonically increasing or decreasing failure rate function is sufficient; if the failure rate function is both convex and concave, a format U failure rate function is sufficient; if not, a unimodal failure rate function is more appropriate.

3.1. The Nature of the Data Sets

The first data is the ordered remission times (in months) of a random sample of 128 bladder cancer patients. 0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06,7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47,14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31,0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90 ,2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12,46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05,1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34,5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85 ,8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07 ,3.36, 6.93, 8.65, 12.63, 22.69.

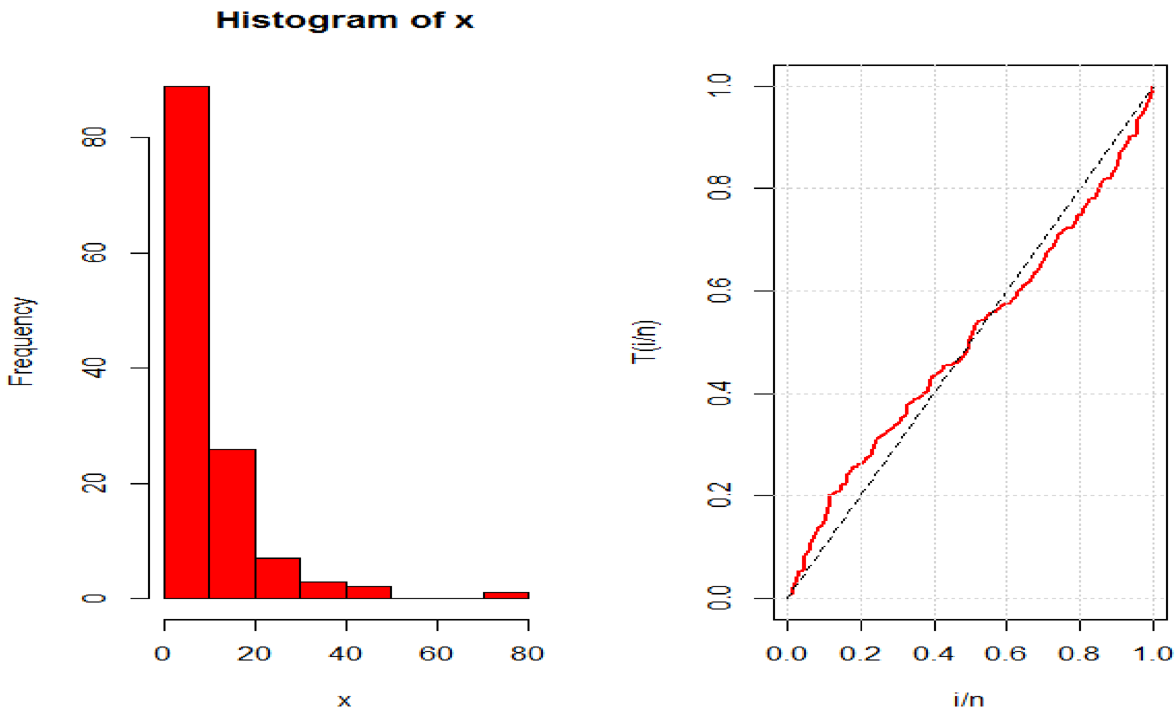

Fig.5. The histogram and TTT plot for Data 1.

The second data represents the strength of 1.5cm glass fibers measured at the National Physical Laboratory, England.
0.55,0.93,1.25,1.36,1.49,1.52,1.58,1.61,1.64,1.68,1.73,1.81,2.00,0.74,1.04,1.27,1.39,1.49,1.53,1.59,1.61,1.66,1.68,1.76,1.82,2.01,0.77,1.11,1.28,1.42,1.50,1.54,1.60,1.62,1.66,1.69,1.76,1.84,2.24,0.81,1.13,1.29,1.48,1.50,1.55,1.61,1.62,1.66,1.70,1.77,1.84,0.84,1.24,1.30,1.48,1.51,1.55,1.61,1.63,1.67,1.70,1.78,1.89.

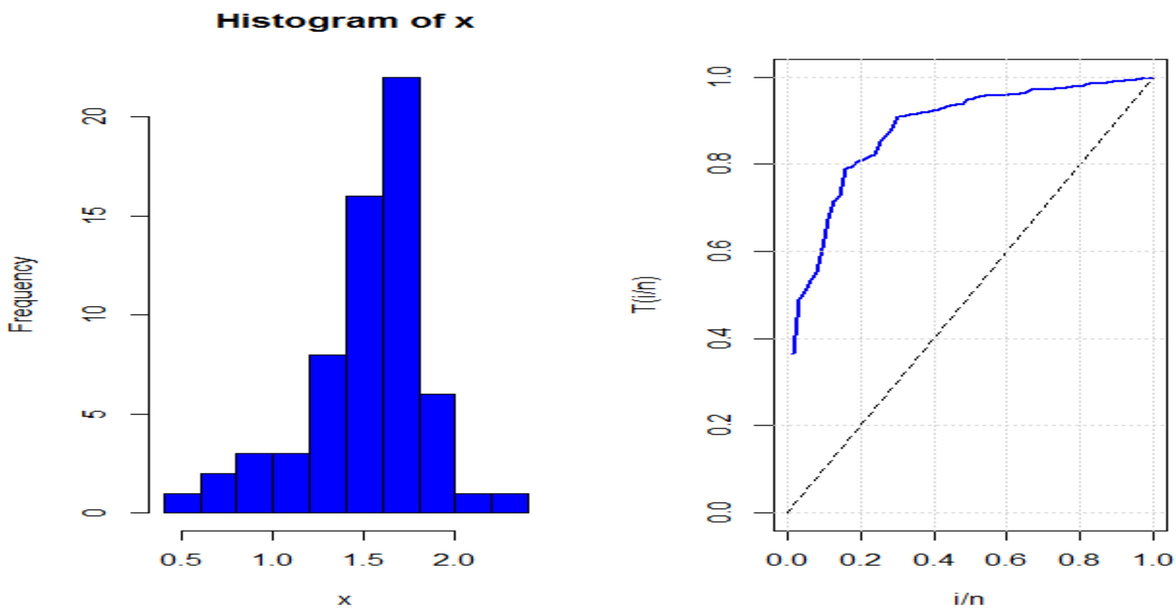

Fig.6. The histogram and TTT plot for Data 2.

The third data is for the times between successive failures of air conditioning equipment in a Boeing 720 airplane.74,57,48,29,502,12,70,21,29,386,59,27,153,26,326.

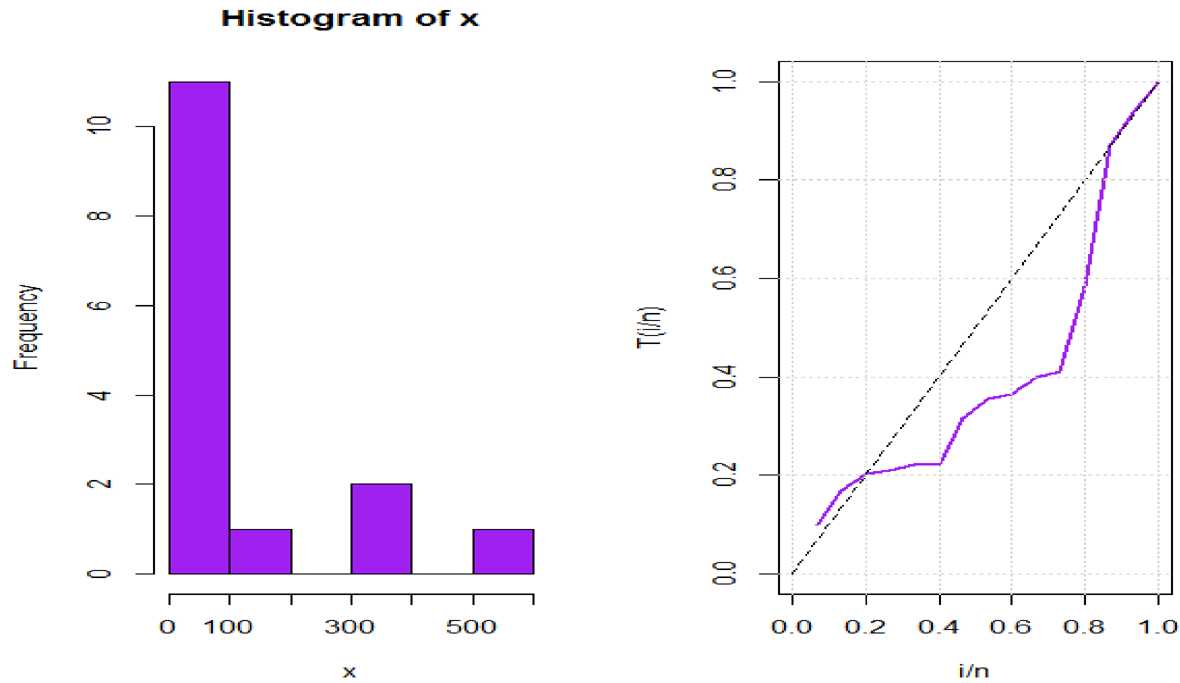

Fig.7. The histogram and TTT plot for data 3.

Fourth data was collected from Aarset, M. V. (1987). 0.1, 0.2, 1.0, 1.0, 1.0, 1.0, 1.0, 2.0, 3.0, 6.0, 7.0, 11.0, 12.0, 18.0, 18.0, 18.0,18.0, 18.0, 21.0, 32.0, 36.0, 40.0, 45.0, 46.0, 47.0, 50.0, 55.0, 60.0, 63.0, 63.0, 67.0, 67.0, 67.0, 67.0, 72.0, 75.0, 79.0, 82.0, 82.0, 83.0, 84.0, 84.0, 84.0, 85.0, 85.0, 85.0, 85.0, 85.0, 86.0, 86.0.

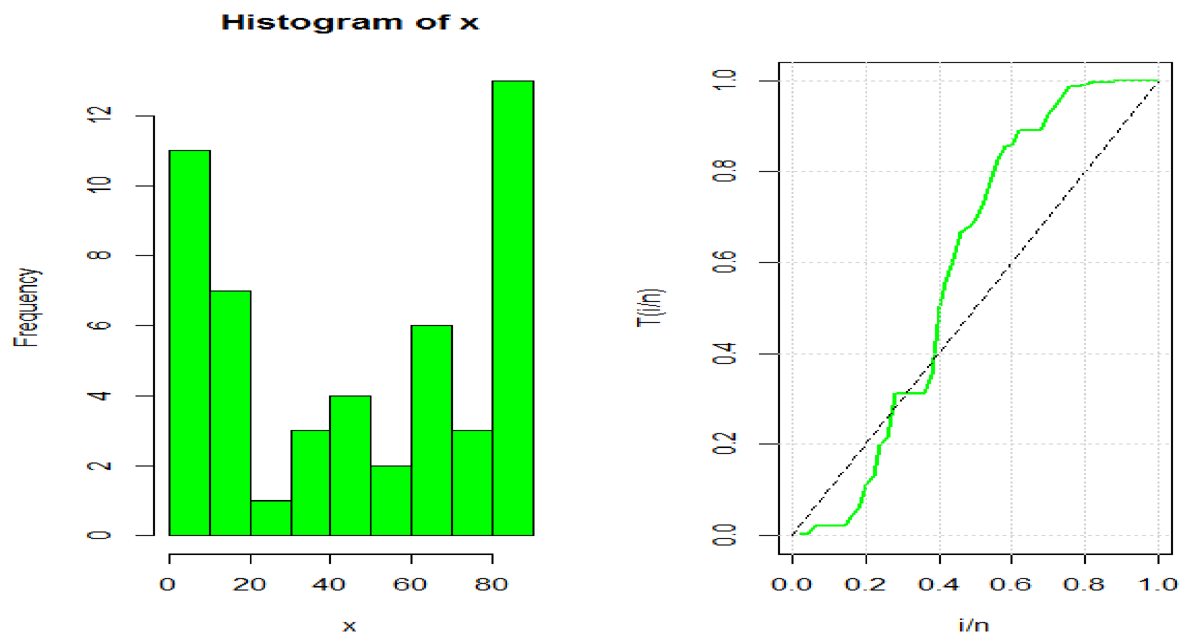

Fig.8. The histogram and TTT plot for data 4.

Fifth data set represents the lifetime's data relating to relief times (in minutes) of 20 patients receiving an analgesic and is given as;
1.1,1.4,1.3,1.7,1.9,1.8,1.6,2.2,1.7,2.7,4.1,1.8,1.5,1.2,1.4,3,1.7,2.3,1.6,2

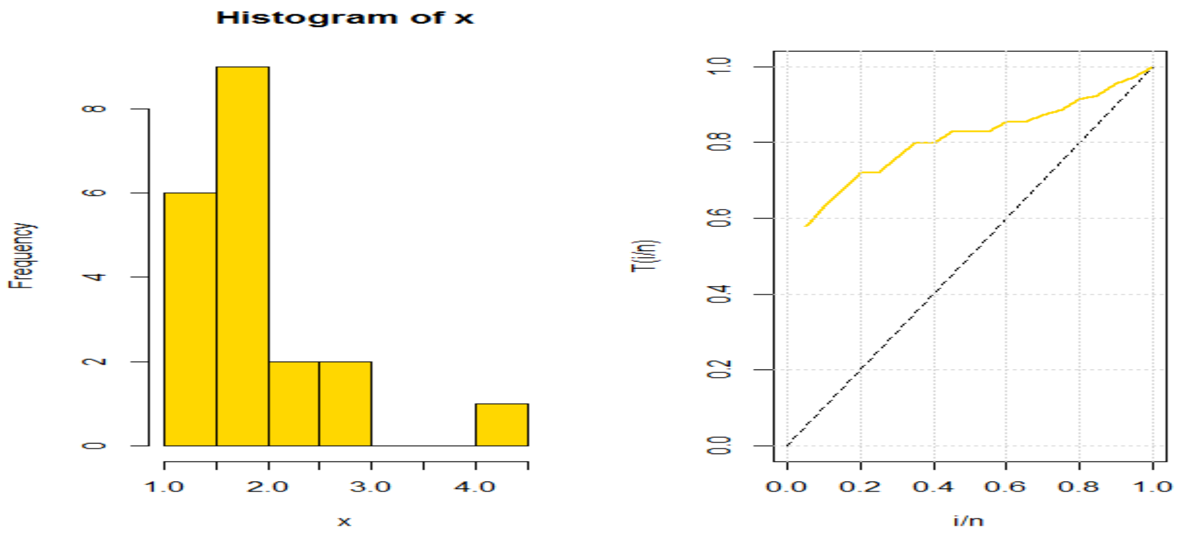

Fig. 9. The histogram and TTT plot for data 5.

4. Applications

Here, the Exponential, Weibull and Gamma distributions were applied to these data sets and their performances are computed in terms of some goodness-of-fits measures. The results are tabulated as follows;

Table 1. Shows the goodness-of-fits measures for the competing distributions

Data	Model	AIC	CAIC	BIC	Rank
1	Exponential	830.6838	830.7155	833.5358	1
	Weibull	832.1738	832.2698	837.8779	3
	Gamma	830.7356	830.8316	836.4396	2
2	Exponential	179.6606	179.7262	181.8038	3
	Weibull	34.4137	34.6137	38.7	1
	Gamma	51.9031	52.1031	56.1894	2
3	Exponential	175.9398	176.2475	176.6478	1
	Weibull	177.5315	178.5315	178.9476	2
	Gamma	177.8521	178.8521	179.2682	3
4	Exponential	484.1792	484.2625	486.0912	1
	Weibull	486.0036	486.259	489.8277	2
	Gamma	484.3805	484.6359	179.2682	3
5	Exponential	67.6742	67.8964	68.6699	3
	Weibull	45.1728	45.8787	47.1643	2
	Gamma	39.6372	40.3431	41.6287	1

From the results, it was observed that an Exponential distribution outperformed other competing models and raked first on data sets 1, 3 and 4 while Weibull distribution performed well on data set 2 thereby has a better fit as compared to the others and finally Gamma distribution outperformed the other models on data set 5.

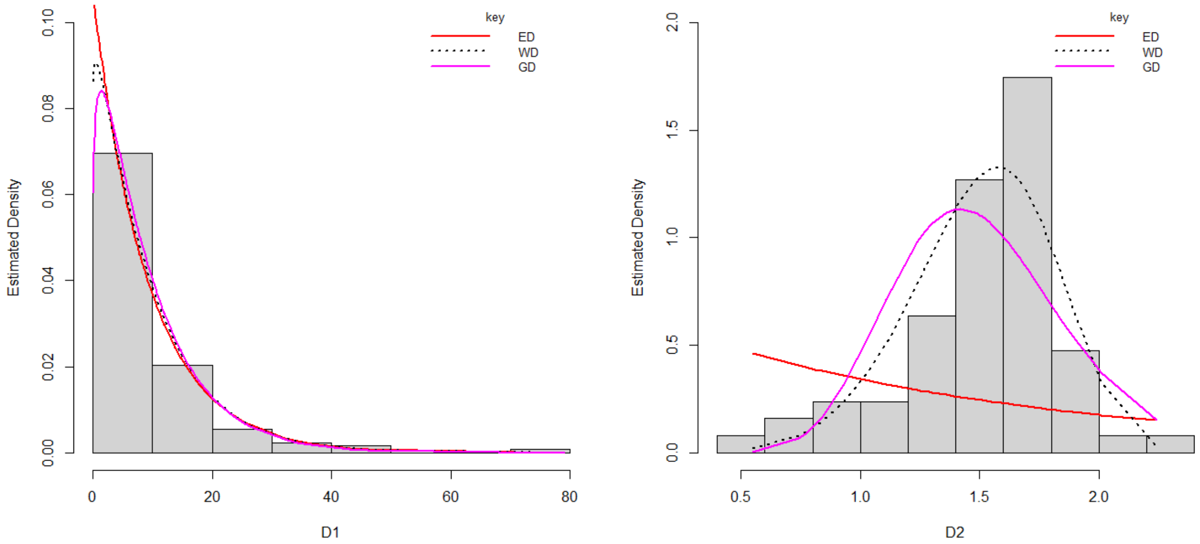

Fig.10. Estimated Densities for Data 1 and 2

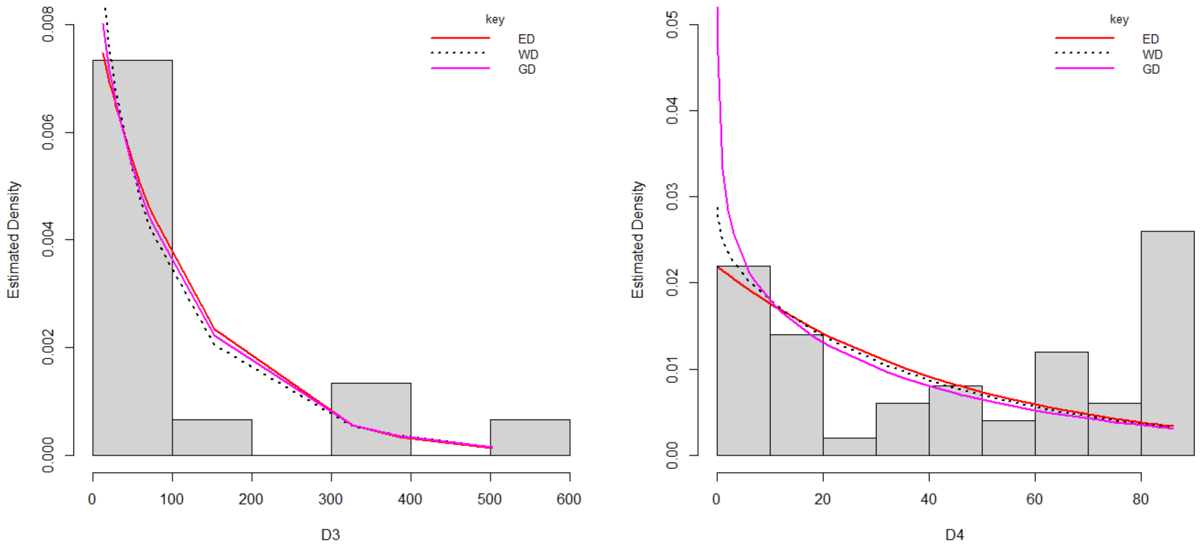

Fig.11. Estimated Densities for Data 3 and 4

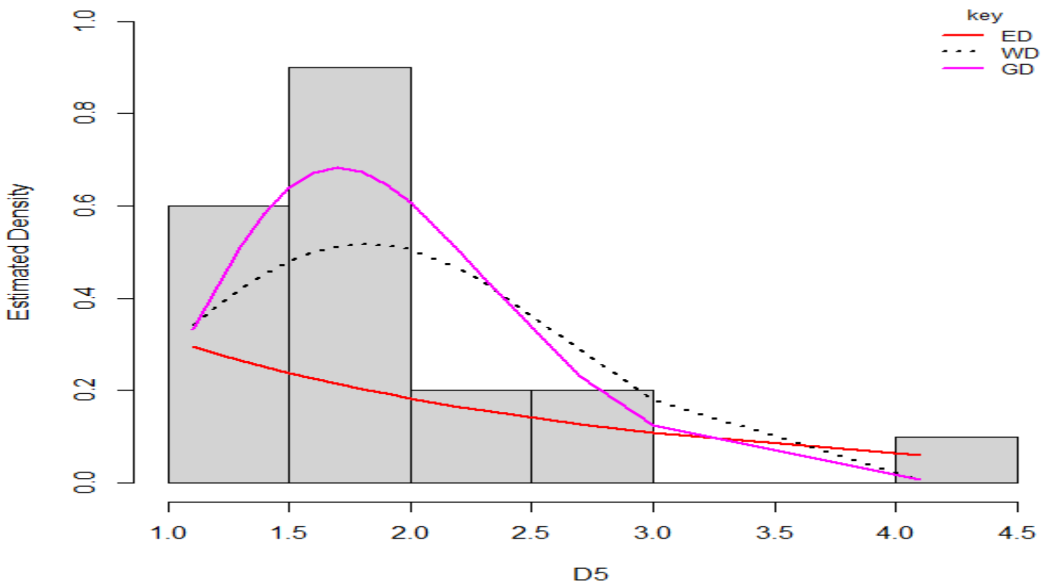

Fig.12. Estimated Density for Data 5

5. Conclusion

This research discussed the potentiality of Exponential, Weibull and Gamma distributions. Some properties of the said distributions were presented in an explicit form. Five data sets with varying shapes ranging from right-skewed to left-skewed were considered in the analysis. The total time on test (TTT) plot has shown that some data sets possessed and increasing, decreasing, unimodal and bathtub failure rates. The performance of the competing models was assessed using some Goodness-of-fit measures and the model with the lowest values of these measures is considered to be the best for that data set. When the data sets skewed to the right with a predefined shape, ED outperformed both Gamma and Weibull distributions.

References

Aarset, M. V. (1987). How to identify a bathtub hazard rate. IEEE transactions on reliability, 36(1), 106-108.

Berezin, M. Y., & Achilefu, S. (2010). Fluorescence lifetime measurements and biological imaging. Chemical reviews, 110(5), 2641-2684.

Alrashidi, M. (2023). Estimation of Weibull Distribution Parameters for Wind Speed Characteristics Using Neural Network Algorithm. *CMC Comput. Mater. Con*, *75*, 1073-1088.

Davis, D. J. F. (1952). An analysis of some failure data. *Journal of the American Statistical Association*, *47*(258), 113-150.

Feigl, P., & Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, 826-838.

Qu, G., Zheng, M., Wang, X., Zhu, R., Su, Y., & Chang, G. (2023). A freeze–thaw damage evolution equation and a residual strength prediction model for porous concrete based on the Weibull distribution function. *Journal of Materials in Civil Engineering*, *35*(5), 04023074.

Singer, J. D., & Willett, J. B. (1991). Modeling the days of our lives: using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. Psychological Bulletin, 110(2), 268.

Shanker, R., Hagos, F., & Sujatha, S. (2015). On modeling of Lifetimes data using exponential and Lindley distributions. *Biometrics & Biostatistics International Journal*, *2*(5), 1-9.

Shi, C., Yu, H., Qian, G., Li, X., Zhu, X., Yao, D., & Zhang, C. (2023). Research on the characteristics of asphalt mixture gradation curve based on Weibull distribution. *Construction and Building Materials*, *366*, 130151.

Yaniktepe, B., Kara, O., Aladag, I., & Ozturk, C. (2023). Comparison of eight methods of Weibull distribution for determining the best-fit distribution parameters with wind data measured from the met-mast. *Environmental Science and Pollution Research*, *30*(4), 9576-9590.