

STATISTICAL MODELLING AND APPLICATIONS OF THE ET-EXPONENTIAL DISTRIBUTION: EVIDENCE FROM BIOMEDICAL AND ENGINEERING DATA

T. Kamaleshwar^{1*}, R. Seethalakshmi², and S. Mythily³

^{1*}Associate Professor, Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai. Tamil Nadu, India.

²Assistant Professor, Department of Mathematics, Rajalakshmi Institute of Technology, Kuthambakkam, Chennai- 600124.

³Department of Statistics, Periyar University, Salem-11, Tamil Nadu, India.
kamalesh4u2@gmail.com, seethamaths79@gmail.com, mythilyshanmugam03@gmail.com

Abstract

The current study utilizes the exponential distribution as the baseline cumulative function and applies the T-x transformation on the exponential family in order to develop a new class of lifetime distribution. The main objective of this study is to develop and evaluate a flexible distribution that may capture various kinds of lifespan features present in real-world applications. A specific instance, the ET-Exponential (ET-Exp) distribution, is thoroughly examined. The structural properties of the paper, such as the moments, cumulative distribution function, and probability density function, are derived. Parameter estimation is performed via the maximum likelihood method. The suggested approach is used on real datasets, such as carbon fiber breaking stress data and hypertension survival times. The performance of the model is assessed by using statistical criteria such as Akaike Information criteria (AIC), Bayesian Information Criteria (BIC), Kolmogorov-Simron (KS), and Anderson-Darling (AD) tests, which were among the statistical tests used to estimate the goodness of fit of each distribution's essential parameters. The results demonstrate that the proposed ET-Exponential distribution is a flexible and effective tool for modeling lifetime data and shows promise for applications in both biomedical and industrial fields.

Keywords: Exponential Transformation, ET-Exponential Distribution, Lifetime Data Modeling, Reliability Analysis, Maximum likelihood estimation

I. Introduction

As real-world phenomena become increasingly complex, classical lifetime distributions often fail to capture the intricate patterns observed in empirical data. This has motivated the development of more flexible statistical models that enhance data representation and improve model fitting. Over recent decades, generalized families of lifetime distributions have broadened the scope of classical models, particularly in survival analysis, biomedical research, and reliability engineering. Among these, transformation-based families have demonstrated significant potential in increasing model adaptability. The Exponential Transformation (ET) family applies an exponential transformation to baseline cumulative distribution functions, resulting in more flexible lifetime models. A key member

of this family, the ET-Exponential distribution, introduces an additional shape parameter that allows it to model diverse hazard rate behaviors such as increasing, decreasing, and bathtub shapes. This enhanced flexibility makes the ET-Exponential distribution effective in accurately representing complex survival and reliability data, improving interpretability and predictive performance across various fields.

Over the years, various extensions and generalizations of classical lifetime distributions have been proposed to improve flexibility and fitting accuracy. Mudholkar and Srivastava [1] improved modeling flexibility by introducing the exponentiated class by increasing baseline distributions to a power. This strategy was expanded by Marshall and Olkin [2] by include shape parameters in survival functions, while Gupta and Kundu [3] developed parameter estimation methods especially for exponential-type models, making it simple to use in survival analysis, including studies of hypertension. Shaw and Buckley [4] introduced the transmuted family, which Mansour et al. [5] expanded into the transmuted-transmuted-G (TT-G) family to account for more kinds of risk forms. In the field of medicine, Tahmasbi and Rezaei [6] utilized generalized exponential and exponential models for evaluating arterial blood pressure variability, along with Almalki and Yuan [7] used modified Weibull approaches, which provide more flexibility than traditional exponential models, to analyze chronic diseases such as hypertension. A more recent development is the T-X family, which was presented by Alzaatreh et al. [8] and offers a generic framework for creating new distributions by transforming preexisting ones. Maurya et al. [9] suggested logarithmic transformations, which Aslam et al. [10] expanded upon to improve model flexibility further.

A notable recent development is the Exponential Transformation (ET) family proposed by Al-Nasser and Hanandeh [11], which demonstrated superior performance in real-data applications compared to traditional lifetime models. Jin and Qian [12] introduced the Mixture Exponential Family Distribution Posterior Network (MEFDPN), leveraging mixtures of exponential families to improve uncertainty estimation and machine learning accuracy. Chaturvedi et al. [13] developed the Generalized Positive Exponential Family, which includes Weibull and gamma distributions, showing superior fitting on wind speed data. Finally, Mohsin and Kalt [14] proposed the Alpha Logarithm Family, a two-parameter extension of the exponential distribution, enhancing flexibility through logarithmic transformation and outperforming classical models in empirical analysis. Based on empirical data analysis alone, the model demonstrated better adaptability and accuracy than classical distributions.

The main objective of this paper is to propose a new ET-Exponential distribution based on the T-X transformation framework. The study aims to derive and analyze its mathematical properties and parameter estimation methods. It also compares the ET-Exponential model's performance with existing lifetime distributions. The model is applied to real-life datasets, including hypertension survival times and carbon fibre breaking stress. Overall, the paper demonstrates the flexibility and practical applicability of the ET-Exponential distribution in modeling complex lifetime data.

The remainder of the paper is organized as follows: Section 2 introduces the ET-Exp model and its theoretical formulation. Section 3 discusses its mathematical properties. Section 4 presents parameter estimation via maximum likelihood. Section 5 applies the model to real datasets. Finally, Section 6 concludes the study with key findings and future research directions.

II. ET-Exponential Distribution

A nonnegative continuous random variable Y with a baseline cumulative distribution function (CDF) $F(y, \lambda)$ and a probability distribution function (PDF) $f(y, \lambda)$ where $\lambda \in (0, \infty)$ is a real valued distribution parameter can be used to construct a more flexible family of lifetime distributions. The proposed CDF's stochastic presentation to generate a new class of distributions can be defined as,

$$G(y, \lambda, \rho) = F(y, \lambda)e^{-\rho F(y, \lambda)} \quad y \geq 0; \rho \geq 0 \quad (1)$$

Where $\bar{F}(y, \lambda) = 1 - F(y, \lambda)$. Noting that when $\rho = 0$, then the proposed distribution is exactly the same as the baseline distribution.

This family will be called as exponential transformation (ET), i.e., $ET(y, \lambda, \rho)$. Now, the PDF of ET family can be obtained by finding the first derivative of equation (1):

$$g(y, \lambda, \rho) = f(y, \lambda)e^{-\rho F(y, \lambda)}(1 + \rho F(y, \lambda)); \quad y \geq 0; \rho \geq 0 \quad (2)$$

According to Alzaatreh et al. (2013), this family can be a part of the T-X family in the following ways: For a continuous probability distribution represented by $F(y, \lambda)$ with a general baseline CDF, a new CDF with the form,

$$G(y, \lambda, \rho) = \int_{\lambda}^{F(y, \rho)} r(t) dt$$

where the related CDF is $R(t)$ and $r(t)$ is a PDF defined over $(0, 1)$. Thus, the PDF $g(y, \lambda, \rho)$ can be computed as

$$g(y, \lambda, \rho) = f(y, \rho)r(F(y, \lambda))$$

If we used $r(t)$ in the following functional form:

$$r(t) = (1 + \rho t)e^{-\rho(1-t)}$$

Then, the proposed family could be considered as member of the T-X family. As an illustration of the proposed family, the exponential distribution will be considered as baseline distribution.

I. Asymptotic Properties of the CDF and PDF

Assuming that Y is a continuous random variable belongs to the ET family as described in (1), it is known that this distribution family corresponds to the distribution functions and the Kolmogorov axioms. For example, the limit property of $G(y, \lambda, \rho)$ clearly satisfies the CDF property.

$$\lim_{y \rightarrow \infty} G(y, \lambda, \rho) = \lim_{y \rightarrow \infty} F(y, \lambda)e^{-\rho F(y, \lambda)} = 1 \text{ and}$$

$$\lim_{y \rightarrow 0} G(y, \lambda, \rho) = \lim_{y \rightarrow 0} F(y, \lambda)e^{-\rho F(y, \lambda)} = 0.$$

Hence, the total probability is equal to one. Also, it is monotone right increasing function of y and $0 \leq G(y, \lambda, \rho) \leq 1; \forall y$. Therefore, $G(y, \lambda, \rho)$ is an absolute continuous distribution function.

Additionally, the fact that $g(y, \lambda, \rho)$ is a non-negative real valued PDF for all y can additionally be immediately confirmed. In the case of an explosion of growth, for example:

$$\lim_{y \rightarrow \infty} g(y, \lambda, \rho) = 0 \text{ and}$$

$$\lim_{y \rightarrow 0} g(y, \lambda, \rho) = \lambda e^{-\rho}.$$

Since both parameters are positive this indicates that $g(y, \lambda, \rho)$ is a unimodal distribution. Now,

the functional form given in (2) satisfied the PDF property:

$$\int_0^{\infty} g(y, \lambda, \rho) dx = \int_0^{\infty} f(y, \theta) e^{-\rho \overline{F}(y, \lambda)} (1 + \rho F(y, \lambda)) dx$$

Assuming that the exponential distribution with mean $1/\theta$ is the baseline distribution, we obtain the ET-Exp distribution to demonstrate the use of the novel stochastic representation provided in (1) and the related PDF provided in (2).

III. Mathematical Properties

I. Moments

Some Moments can be used to study some of a statistical distribution's most crucial properties, including skewness, kurtosis, dispersion, and tendency. Assuming that computing the expected value of $k(y)$ yields the moments of $ET(y, \lambda, \rho)$; where

$$k(y) = \begin{cases} y^r, & \text{for moment of order } r, \\ e^{ty}, & \text{for moment of generating function,} \\ e^{ity}, & \text{for characteristic function.} \end{cases}$$

Hence,

$$\begin{aligned} E(k(y)) &= \int_0^{\infty} k(y) f(y, \lambda) e^{-\rho \overline{F}(y, \lambda)} (1 + \rho F(y, \lambda)) dx \\ &= \int_0^{\infty} k(y) f(y, \lambda) (1 + \rho F(y, \lambda)) \left(\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-1)^{j+k} \frac{\rho^j}{j!} \binom{j}{k} F(y, \lambda)^k \right) dx, \end{aligned}$$

which is equivalent to the expected value based on the baseline distribution

$$E_F(k(y) e^{-\rho \overline{F}(y, \lambda)} (1 + \rho F(y, \overline{F}(y, \lambda))))$$

Then the expected value can be obtained using expansion technique or by using integral estimation.

IV. Parameter Estimation

In this section, estimation of the unknown parameters of the $ET(y, \lambda, \rho)$ family of distributions based on complete samples are determined using method of moment (MOM) and maximum likelihood estimation (MLE) method. Let y_1, y_2, \dots, y_n be the observed values from $ET(y, \lambda, \rho)$ family.

I. Method of Moment

The MOM estimator can be obtained by solving the following equations:

$$\begin{aligned} E_p(y e^{-\rho \Gamma(y, v)} (1 + \rho F(y, \lambda))) &= \frac{\sum_n y_i}{n}, \\ E_p(y^2 e^{-\rho \Gamma(y, v)} (1 + \rho F(y, \lambda))) &= \frac{\sum_n y^2}{n}. \end{aligned}$$

The first moment of ET-Exp distribution can be expressed as:

$$E(Y) = \frac{1}{\omega \vartheta} \left(1 - e^{-\rho + \rho \left(\int_0^\infty \frac{\sinh(t)}{t} dt - \int_0^\infty \frac{\cosh(t)-1}{y} dt \right)} \right)$$

$$= \frac{1}{\omega \vartheta} \left(1 - e^{-\rho + \rho (\log(\rho) - \text{Chi}(\rho) + \text{Shs}(\rho) + \gamma)} \right)$$

while the second moment can be replaced by the following formula:

$$E(Y^2) = \frac{2(\rho^2, F, (1,1,1; 2,2,2; -\rho) + \log(\rho) + \Gamma(\Omega, \rho) + \gamma)}{\vartheta \omega^2}$$

where γ is Euler's constant, with numerical value ≈ 0.577216 , the incomplete gamma function satisfies

$$\Gamma(0, \infty) = \int_\infty^\infty \frac{e^{-t}}{t} dt$$

and ${}_3F_3(1,1,1; 2,2,2; -\infty)$ is the generwhoed hypergeometric function.

II. Maximum Likelihood Estimation of the Parameters

Using the MLE, the point estimator of the unknown parameter can be obtained by solving the following likelihood functions

$$L = \prod_{i=1}^n f(y_i - \lambda) e^{-\rho T(y, \lambda)} (1 + \rho F(y_i, \lambda))$$

Taking the Log of the likelihood function will simplify the estimation problem:

$$\log L = \left\{ \sum_{i=1}^n \log(f(y, \lambda)) - \sum_{i=1}^n \rho F(y_i, \lambda) + \sum_{i=1}^n \log(1 + \rho F(y_i, \lambda)) \right\}$$

Now, we have to find the first order condition:

$$\frac{d \log L}{d \lambda} = \sum_{i=1}^n \frac{df(y_i, \lambda)/d \lambda}{f(y_i, \lambda)} + \sum_{i=1}^n \rho f(y_i, \lambda) + \sum_{i=1}^n \frac{\rho f(y_i, \lambda)}{1 + \rho f(y_i, \lambda)}$$

$$\frac{d \log L}{d y} = - \sum_{i=1}^n F(y_i, \lambda) + \sum_{i=1}^n \frac{F(y_i, \lambda)}{1 + \rho F(y, \lambda)}$$

Since the equations are non-linear, they are solved numerically using R software (R Core Team, 2021).

V. Results and Discussion

I. Applications

This section shows off the ET-Exp distribution's capabilities by fitting the model to two datasets: the Lindley, Generalized Exponential (GE), and Exponential distributions. The parameters associated with each distribution have been calculated using the maximum likelihood technique for these two datasets. Using the information obtained the values of Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) and -LogL are obtained from the datasets. The Anderson-Darling (AD) and Kolmogorov-Smirnov (K-S) statistics, together with their related P-value (p-Val), are also found. The mathematical formulae of these measures are given by:

The AD test statistics computed as

$$AD = -m - \frac{1}{m} \sum_{j=1}^m (2j - 1) \times [\log G(y_j; \Delta) + \log (1 - G(y_{t-m+1}; \Delta))]$$

The KS test statistic derived as

$$KS = \text{Supermium } x[G_m(y; \Delta) - G(y; \Delta)]$$

The AIC test statistics obtained as

$$AIC=2p-2L$$

The BIC test statistics derived as

$$BIC=p\log(n)-2L$$

where L is the maximized likelihood function evaluated at MLEs, m is the sample size, and p is the number of parameters in the model. For every dataset, the five-number summary, mean, variance, skewness, and kurtosis are computed as basic descriptive statistics. The best distribution has been proposed to be the one with the lowest AIC and BIC, which finds an optimal balance between model complexity and goodness-of-fit. The -Log L value help to evaluate the model's overall fit.

II. Dataset I: Hypertension Data

This dataset shows the survival periods in years before the development of hypertension for 119 patients randomly selected from the Bolgatanga Regional Hospital in Ghana's Upper East region.

The following components are contained in the dataset: 71, 5, 39, 62, 52,71, 38, 56, 35, 69,34,71,66,70, 52, 37, 35, 71, 73, 19, 74, 74, 75, 51, 76,49, 19, 76, 78, 76, 76, 49, 47, 48, 48, 46, 46, 46, 41, 40, 43, 45, 47, 47, 44, 45, 46, 42, 43, 42, 20, 28, 26, 60, 27, 24, 29, 60, 25, 60, 69, 36, 69, 69, 68, 68, 67, 67, 67, 52, 35, 66, 55, 66, 61, 61, 64, 64, 65, 65, 63, 63, 62, 39, 62, 62, 62, 59, 59, 59, 58, 58, 18, 57, 57, 56, 56, 37, 53, 53, 53, 53, 54, 54, 66, 17, 50, 75, 51, 38, 52, 66, 4, 52, 55, 19, 58, and 73.

The descriptive statistics for this data set are shown in Table 1, and the MLE and goodness of fit test results for this data set using each distribution are displayed in Table 2.

Table 1: The summary statistics for the Hypertension Data

Parameter	Min	Q1	Median	Q3	Mean	Max	Skewness	Kurtosis	Variance
Dataset1	4	43	55	65.5	52.428	78	-0.731	0.043	270.28

Table 2: MLEs and goodness of fit statistics for the Hypertension Data

Distribution	α	θ	-LogL	K-S	p-val	A-D	p-val	AIC	BIC
Exponential	-	0.019	-590.2	0.361	0.000	25.452	0.0005	1182.35	1185.12
GE	7.0179	0.047	-530.9	0.1606	0.004	6.045	0.0009	1065.96	1071.52
ET-EXP	9.4877	0.05	-522.6	0.1507	0.008	4.62	0.007	1049.95	1054.95
Lindley	-	0.037	-555.0	0.2638	0.000	15.409	0.000	1112.13	1114.91

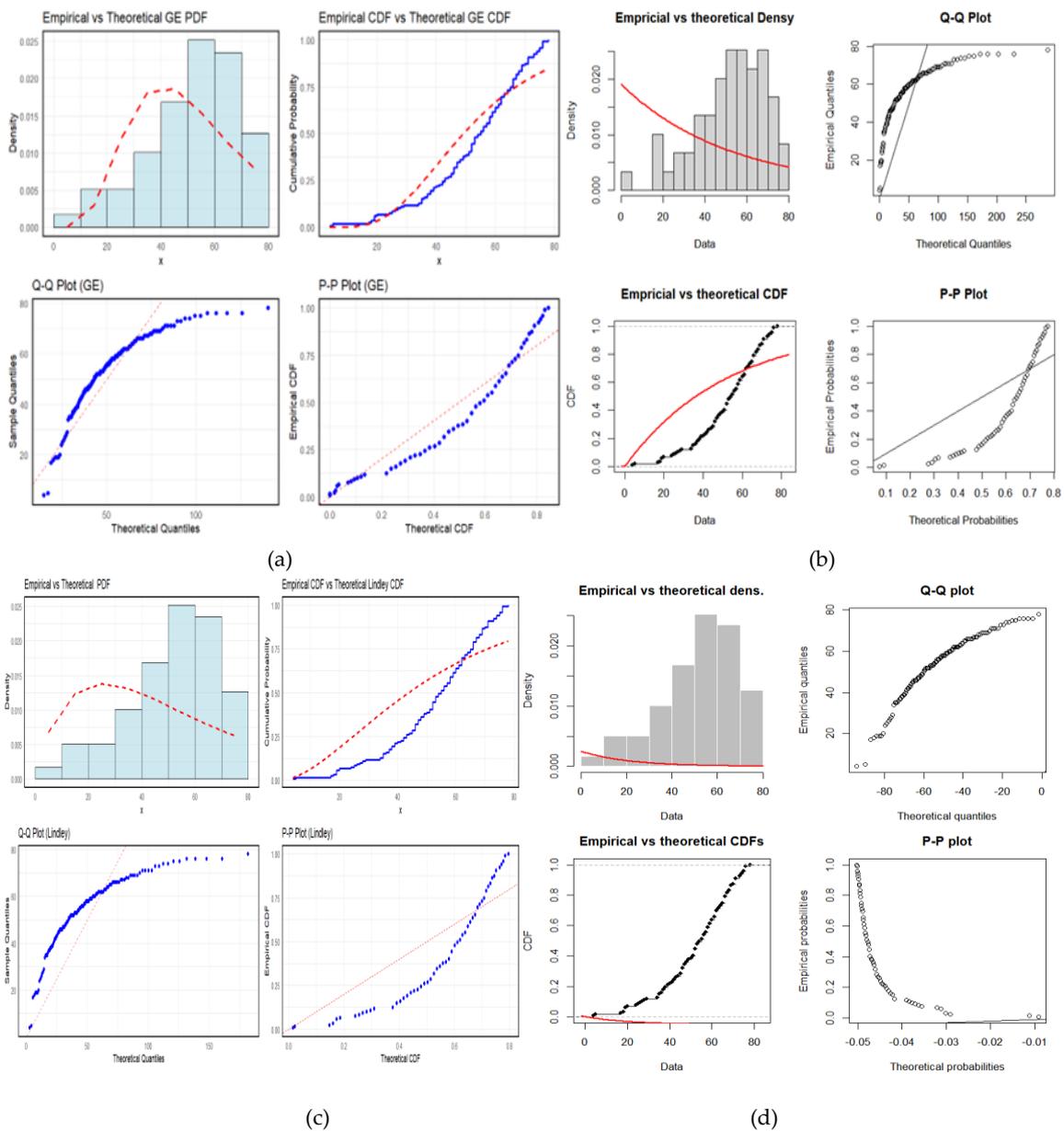


Figure 1: The empirical and theoretical PDFs, empirical and theoretical CDFs, Q-Q plots and p-p plot for (a) Exponential, (b) Generalized-Exponential, (c) Lindley and (d) ET-Exponential for the Hypertension data.

Table 2 and Figure 1 shows that the goodness-of-fit statistics for the evaluated distributions show that the Exponentiated Exponential (ET-EXP) and Generalized Exponential (GE) distributions best suit the data. ET-EXP and GE have the lowest negative log-likelihood values, as well as lower Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) statistics than the Exponential and Lindley distribution, respectively. The p-values for the K-S and A-D tests for ET-EXP and GE, although tiny, are significantly greater compared to those for the Exponential and Lindley models, showing a superior fit. Furthermore, ET-EXP and GE have the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, implying that they stabilize goodness-of-fit with model complexity more efficiently. On the other hand, the Exponential and Lindley distributions have inferior fit performance, having higher statistical significance and lower p-values, indicating a less suitable representation of data. Based on these criteria, the ET-EXP distribution exceeds the GE distribution somewhat, making it the best fit for the data, followed by the GE. Both models outperform the Lindley and Exponential distributions significantly.

III. Dataset II: Carbon Fiber Breaking Stress Data

The second dataset is comprised of 100 observations. The dataset refers to breaking stress of carbon fibers, which are measured in Gba [25]. The descriptive statistics for this data set are shown in Table 3, and the MLE and goodness of fit test results for this data set using each distribution are displayed in Table 4.

Table 3: Summary Statistics of Carbon Fiber Breaking Stress Data

Parameter	Min	Q1	Median	Q3	Mean	Max	Skewness	Kurtosis	Variance
Dataset2	0.39	1.84	2.7	3.22	2.621	5.56	0.3629	0.043	1.028

Table 4: MLEs and goodness of fit statistics for the Carbon Fiber Breaking Stress Data

Distribution	α	θ	-LogL	K-S	p-val	A-D	p-val	AIC	BIC
Exponential	-	0.381	-196.36	0.26	0.000	12.315	0.000	365.50	368.106
GE	7.788	1.013	-146.17	0.97	0.000	4.218	0.007	292.70	297.918
ET-EXP	9.182	1.092	-144.35	0.10	0.196	1.222	0.259	296.35	301.568
Lindley	-	0.617	-181.75	0.32	0.000	17.301	0.000	394.73	397.339

Table 4 and Figure 2 shows that the ET-EXP distribution offers the best overall fit to the data, based on a comparison of four potential distributions: Lindley, GE, ET-EXP, and Exponential. Its suitability is demonstrated by low test statistics and high p-values in both the Anderson-Darling and Kolmogorov-Smirnov tests, indicating a statistically significant and consistent fit. The GE distribution's goodness-of-fit p-values are below 0.05, indicating a poor fit, although having a marginally lower AIC than ET-EXP. The exponential distribution does not meet the goodness-of-fit criteria, even if it is simple and has a low -LogL. The Lindley distribution performs the worst overall, with the highest AIC and BIC values and the lowest fit statistics. The ET-EXP distribution is therefore determined to be the most appropriate model for the data based on both goodness-of-fit and model selection criteria.

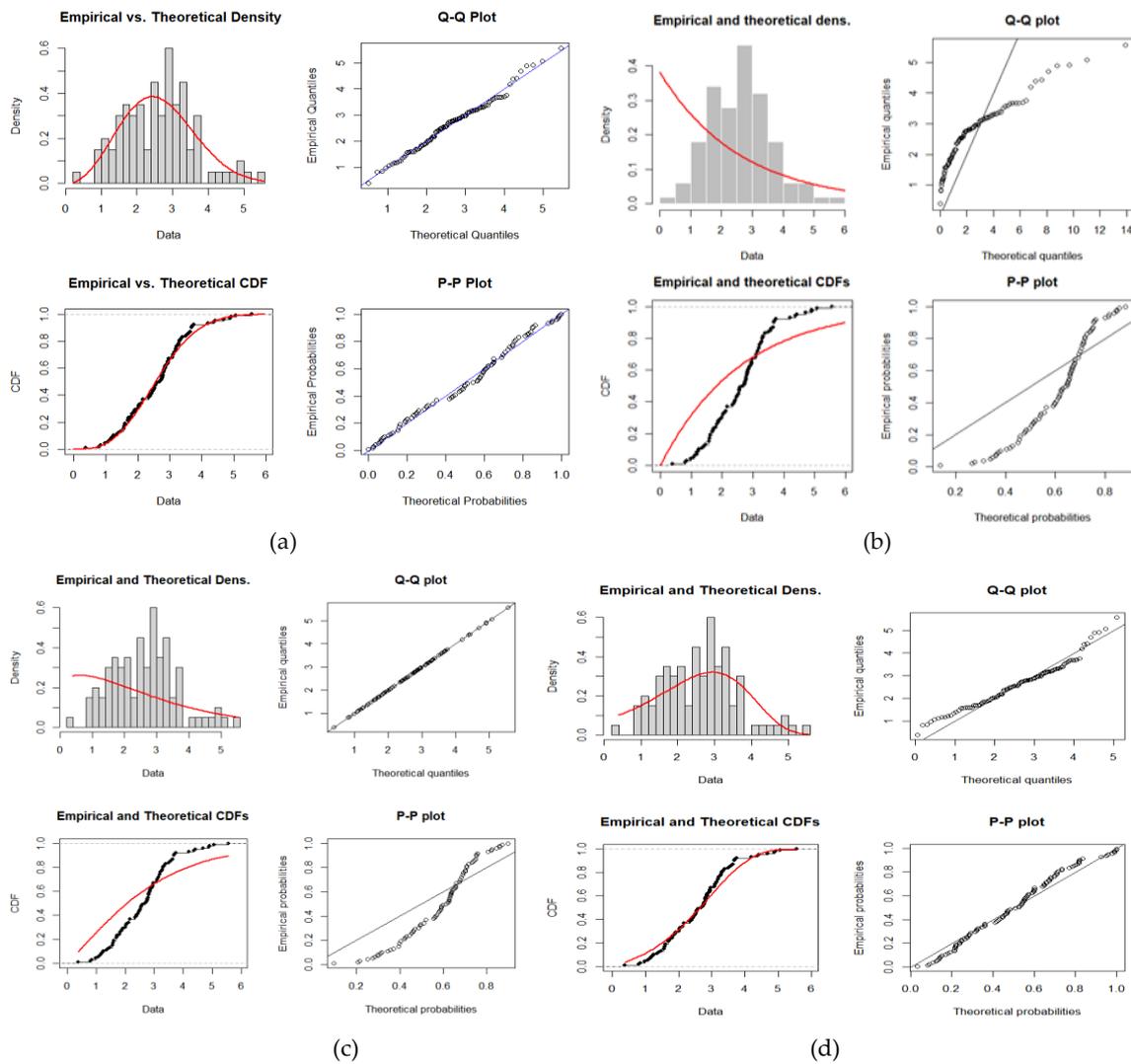


Figure 2: The empirical and theoretical PDFs, empirical and theoretical CDFs, Q-Q plots and p-p plot for (a) Exponential, (b) Generalized-Exponential, (c) Lindley and (d) ET-Exponential for the Carbon Fiber Breaking Stress Data.

VI. Conclusion

This study introduces the ET-Exp distribution as a novel and versatile model for lifetime data analysis. Across both cases, the ET-Exp model demonstrates clear superiority, achieving lower values of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), along with improved goodness-of-fit statistics, including the Kolmogorov–Smirnov (K–S) and Anderson–Darling (A–D) tests. Compared to the traditional Exponential and Lindley distributions, the ET-Exp model consistently provides a significantly better fit. Furthermore, it performs comparably to or better than the Generalized Exponential (GE) distribution, particularly in balancing fit quality with model complexity. In addition to its empirical strengths, the ET-Exp distribution offers desirable theoretical properties, including closed-form expressions for moments and entropy, which further underscore its practical utility. Overall, the ET-Exp distribution emerges as a robust, flexible, and effective tool for modeling lifetime data, with strong potential applications in both biomedical research and industrial reliability analysis.

References

- [1] Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE transactions on reliability*, 42(2): 299-302.
- [2] Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, 84(3): 641-652.
- [3] Gupta, R. D. and Kundu, D. (2001). Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 43(1):117-130.
- [4] Shaw, W. T. and Buckley, I. R. (2007). The alchemy of probability distributions: Beyond gram-charlier and cornish-fisher expansions, and skew-normal or kurtotic-normal distributions. *Submitted, Feb, 7, 64*.
- [5] Tahmasbi, R. and Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics and Data Analysis*, 52(8): 3889-3901.
- [6] Almalki, S. J. and Yuan, J. (2013). A new modified Weibull distribution. *Reliability Engineering and System Safety*, 111: 164-170.
- [7] Alzaatreh, A. Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71(1): 63-79.
- [8] Maurya, S. K. Kaushik, A. Singh, R. K. Singh, S. K. and Singh, U. (2016). A new method of proposing distribution and its application to real data. *Imperial Journal of Interdisciplinary Research*, 2(6):1331-1338.
- [9] Aslam, M. Ley, C. Hussain, Z. Shah, S. F. and Asghar, Z. (2020). A new generator for proposing flexible lifetime distributions and its properties. *PloS one*, 15(4): e0231908.
- [10] Ahmad, Z. Mahmoudi, E. Alizadeh, M. Roozegar, R. and Afify, A. Z. (2021). The Exponential T-X Family of Distributions: Properties and an Application to Insurance Data. *Journal of Mathematics*, 2021(1): 3058170.
- [11] Nasiru, S. and Abubakari, A. G. (2022). Marshall-Olkin Zubair-G family of distributions. *Pakistan Journal of Statistics and Operation Research*, 195-210.
- [12] Al-Nasser, A. and Hanandeh, A. (2025). Stochastic Generator of a New Family of Lifetime Distributions with Illustration. *REVSTAT-Statistical Journal*, 23(1): 139-160.
- [13] Jin, X. and Qian, Q. (2025). MEFDPN: Mixture exponential family distribution posterior networks for evaluating data uncertainty. *Expert Systems with Applications*, 262, 125593.
- [14] Chaturvedi, A. Bhatti, M. I. Bapat, S. R. and Joshi, N. (2025). Modeling wind speed data using the generalized positive exponential family of distributions. *Modeling Earth Systems and Environment*, 11(2): 1-11.
- [15] Mohsin, L. A. J. and Kalt, H. G. (2025). An application to exponential distribution of a novel family of distributions. *AIP Conference Proceedings*, 3282: 040036.