Analele Universităţii Oradea
Fasc. Matematica, Tom XXXII (2025), Issue No. 2, 137–151

STATISTICAL PROPERTIES OF THE THREE - PARAMETER LOGISTIC - TYPE FUNCTION AND THEIR APPLICATION TO A REAL - LIFE DATASET

ABDULLAH OZAN ASLAN[1], HANDE UYAN[2], SEDA KARATEKE[3] AND İBRAHİM BÜYÜKYAZICI[4]

Abstract. In this paper, we study the Three-Parameter Logistic-Type (TPLT, for short) Function and its first derivative as cumulative distribution function (cdf) and probability density function (pdf), respectively. We analyze important statistical properties such as hazard rate, survival function, quantile function, moment, incomplete moments residual-life function, Renyi entropy, and q-entropy for these distributions. Additionally, we present plots illustrating the variation of the pdf and hazard rate of the TPLT with respect to the parameters. The maximum likelihood (MLE) technique for parameter estimation for this TPLT distribution is performed for both simulation data and for the Breast Cancer data set, and the results are compared with the TIHLE distribution and the exponential distribution, and it is concluded that the TPLT distribution fits the data used better than the other distributions.

1. Introduction

Traditional distributions frequently fall short in characterizing and predicting a wide range of interesting data sets. To address these limitations, a novel approach involving the generation of families of continuous distributions has been developed. This innovative method extends classical distributions by introducing additional shape parameters, thereby enhancing their flexibility and applicability across various domains. The generated families of distributions have been extensively studied and applied, offering more versatility in fitting diverse data sets. Notable examples of such generated families include the beta-G by Eugene et al. [1], gamma-G (type 1) by Zografos and Balakrishanan [2], Kumaraswamy-G by Cordeiro and de Castro [3], gamma-G (type 2) by Ristic and Balakrishanan [4], transformed-transformer (T-X) by Cordeiro et al. [5], type I half-logistic-G by Cordeiro et al. [6], Garhy-G by Elgarhy et al. [7], exponentiated Weibull-G by Hassan and Elgarhy [8], Kumaraswamy Weibull-G by Hassan and Elgarhy [9], type II half-logistic-G by Hassan et al. [10], exponentiated extended-G by Elgarhy et al. [11] and Odd Frechet-G by Haq and Elgarhy [12]. In this paper we base on [14] and use the parameter diversity of our TPLT distribution, which allows us to produce a better distribution compared to the

---

2020 *Mathematics Subject Classification.* 62N02, 60E05, 62E10, 62N05.

Key words and phrases. Three - Parameter Logistic - Type Function, Maximum likelihood estimates, Newton - Raphson Method, Moments, Quantile.

TIHLE and E distribution. We can give the cdf and pdf of the TIHLE distribution respectively as follows [14]:

$$F(x,\varphi)=\frac{1-e^{-\alpha\omega x}}{1+e^{-\alpha\omega x}}\quad \omega,\alpha>0,\ x>0,$$

where $\varphi=(\omega,\alpha)$ is the set of parameters,

$$f(x,\varphi)=\frac{2\omega\alpha e^{-\alpha\omega x}}{[1+e^{-\alpha\omega x}]^2}.$$

and also give the cdf and pdf of the E distribution respectively [14],

$$G(x)=1-e^{-\alpha x},$$

$$g(x)=\alpha e^{-\alpha x}\quad \alpha>0,\ x>0$$

where, α is the scale parameter.

This newly created TPLT distribution has shown significant improvements over conventional distributions, providing better fits and more accurate predictions. By adding one or more additional shape parameters to the underlying distribution, these models offer improved flexibility and robustness in statistical modelling. As a result, they have become invaluable tools in various fields, enabling researchers and practitioners to obtain more precise and reliable analyses.

In this paper, we study the three-parameter logistic-type function and its first derivative, interpreting them as the cumulative distribution function (cdf) and probability density function (pdf), respectively. We analyze important statistical properties such as hazard rate, survival function, quantile function, moment, incomplete moments residual-life function, Renyi entropy, q-entropy for these distributions. Additionally, we present plots illustrating the variation of the pdf and hazard rate of the three-parameter logistic-type function with respect to the parameters. Parameter estimation for the distribution of this three-parameter logistic-type function (TPLT distribution) is carried out using maximum likelihood (MLE) technique. Furthermore, we perform parameter estimation using the maximum likelihood method for two real data sets, comparing the results with other distributions. We demonstrate the effectiveness of the TPLT distribution for modeling these selected data sets. Finally, the results of this study confirm the successful performance of the TPLT distribution on real data sets by exploiting the parameter diversity, as can be seen in the figures and tables. These results support the usability of this function as a powerful modeling tool in practical applications. Therefore it is important to note that the TPLT distribution is an important alternative in the field of statistical analysis and data modeling. In addition to the aforementioned, we will apply the analysis to the Breast Cancer data set. This data set provides valuable insights into the genetic expression patterns associated with breast cancer. By leveraging this dataset, we aim to further validate the efficacy of the TPLT distribution in modeling real-world data, specifically in the context of breast cancer research. This application underscores the versatility and relevance of the proposed statistical methodology in the domain of biomedical data analysis, contributing to the ongoing efforts in understanding and combating breast cancer.

2. TPLT DISTRIBUTION

In this section we obtain the cumulative distribution function (cdf) of the TPLT distribution given in [13] as follows:

$$F(x,\gamma)=\theta_{\lambda,a,H}(x)=\frac{H^{a(x-x_0)}}{\lambda+H^{a(x-x_0)}}, \tag{2.1}$$

where $\lambda, a>0$; $x_0\in(-\infty,\infty):H>1$ and $\gamma=(H,a,\lambda)$ is the set of parameters.

If we take the first derivative of this cdf as the probability density function (pdf) corresponding to (2.1), this pdf of the TPLT distribution as follows:

$$f(x,\gamma)=\theta'_{\lambda,a,H}(x)=\frac{\lambda a\ln(H)H^{a(x-x_0)}}{(\lambda+H^{a(x-x_0)})^2} \tag{2.2}$$

The survival function $\overline{F}(x,\gamma)$, hazard rate $h(x,\gamma)$, reversed hazard rate $\tau(x,\gamma)$ and cumulative hazard rate $H(x,\gamma)$ functions are, respectively, given by

$$\begin{aligned}\overline{F}(x,\gamma) &= 1-F(x,\gamma)\\ &= \frac{1}{1+H^{a(x-x_0)}},\\ h(x,\gamma) &= \frac{f(x,\gamma)}{\overline{F}(x,\gamma)}\\ &= \frac{a\ln(H)H^{a(x-x_0)}}{\lambda+H^{a(x-x_0)}},\\ \tau(x,\gamma) &= \frac{f(x,\gamma)}{F(x,\gamma)}\\ &= \frac{\lambda a\ln(H)}{(\lambda+H^{a(x-x_0)})^2},\end{aligned}$$

and

$$\begin{aligned}H(x,\gamma) &= -\ln(1-F(x,\gamma))\\ &= -\ln\left(\frac{\lambda}{\lambda+H^{a(x-x_0)}}\right).\end{aligned}$$

In the next section, some statistical properties of the TPLT distribution will be examined. These useful statistical features give us information about the TPLT distribution.

3. Some Statistical Properties of TPLT distribution

In this section, the statistical properties of the three-parameter logistic-type distribution (TPLT) are examined. The quantile function and median of the TPLT distribution were calculated, and expansions were made for the distribution and density functions. Important statistical measures such as moments and Rènyi and q-entropies are also discussed. Examining the statistical properties of the TPLT distribution allows more accurate estimates of this distribution on various data sets. These calculations make the TPLT distribution more flexible and applicable, increasing its use in statistical analysis and modelling.

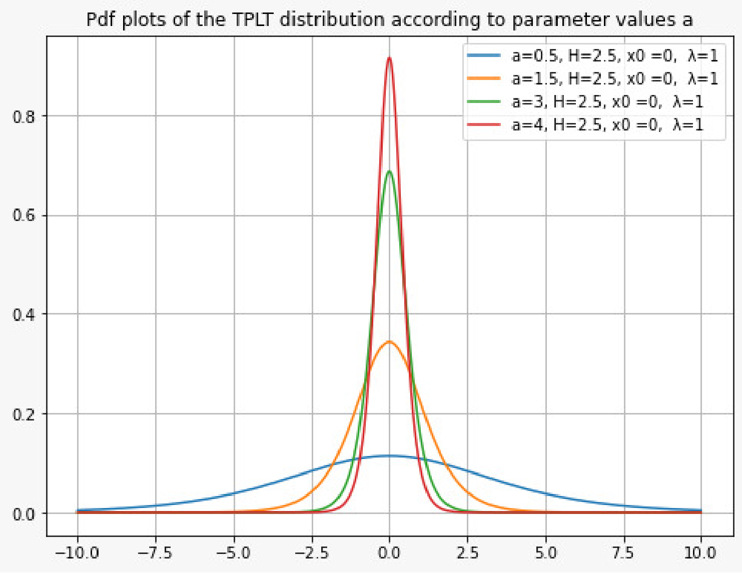


FIGURE 1. Plots of the pdf of TPLT distribution for some parameter values.

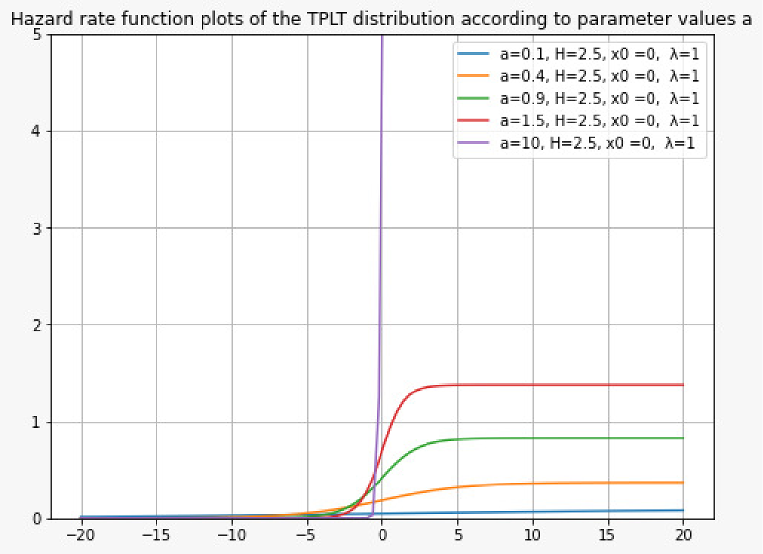


FIGURE 2. Plots of the hazard rate function of the TPLT distribution for some parameter values.

3.1. **Quantile and median.**

Quantile functions play a crucial role in both the theoretical framework of probability theory and practical applications in statistics and simulations. They are particularly valuable in simulation methods, as they enable the generation of simulated random variables for both traditional and novel continuous distributions.

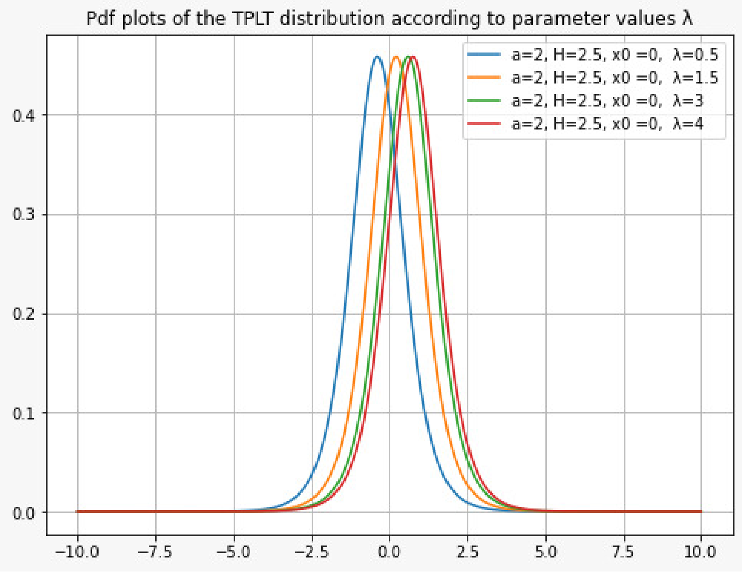


FIGURE 3. Plots of the pdf of TPLT distribution for some λ parameter values.

The quantile function, denoted by, $Q(\delta) = F^{-1}(\delta)$ of X is derived as follows

$$\delta = \frac{H^{a(Q(\delta)-x_0)}}{\lambda + H^{a(Q(\delta)-x_0)}}.$$

Then, the quantile function takes the following form

$$F^{-1}(\delta) = Q(\delta) = x_0 + \frac{1}{a\ln(H)} \ln\left(\frac{\lambda\delta}{1-\delta}\right) \tag{3.1}$$

where, δ is a uniform distribution on the interval $(0,1)$ and $F^{-1}(.)$ is the inverse function of $F(.)$. Specifically, the median can be obtained from (3.1) by setting $a = 1$ and $\delta = 0.5$. Thus, the median is given by

$$\text{median} = x_0 + \frac{\ln(\lambda)}{\ln(H)}.$$

3.2. **Expansions for distribution and density functions.** In this subsection, a useful expansion of the pdf for TPLT distribution is presented. According to the generalized binomial theorem, for $s > 0$ and $|\zeta| < 1$,

$$(\lambda + \zeta)^{-s} = \sum_{t=0}^{\infty} (-1)^t \binom{s+t-1}{t} \lambda^{-t-s} \zeta^t. \tag{3.2}$$

By applying the binomial theorem (3.2) in (2.2), the three-parameter logistic-type distribution function (TPLT distribution) can be expressed as follows:

$$f(x) = \sum_{t=0}^{\infty} \sigma_t \lambda^{-1-t} H^{a(t+1)x}, \tag{3.3}$$

where $x_0 = 0$ and $\sigma_t = a\ln(H)(-1)^t\binom{s+t-1}{t}$.

3.3. **Moments.** If X has the pdf (3.3), r^{th} moment is obtained as follows

$$\mu(r) = E(X^r) = \int\limits_0^\infty x^r f(x)dx. \tag{3.4}$$

then substituting (3.3) into (3.4) we get:

$$\mu(r) = \sum_{t=0}^{\infty} \sigma_t \lambda^{-1-t} \int\limits_0^\infty x^r H^{(a(t+1)x)} dx.$$

Then, for $H = e$,

$$\mu(r) = \sum_{t=0}^{\infty} \sigma_t \lambda^{-1-t} \int\limits_0^\infty x^r e^{-(-(a(t+1)x)} dx$$

whence

$$\mu(r) = \sum_{t=0}^{\infty} - \frac{\sigma_t \lambda^{1-t} \Gamma(r+1)}{(a(t+1))^{r+1}}.$$

For a random variable X, it is known that, the moment generating function of three-parameter logistic - type distribution (TPLT distribution) is defined as

$$M_X(k) = \sum_{r=0}^{\infty} \frac{k^r}{r!} E(X^r) = \sum_{t,r=0}^{\infty} - \frac{k^r}{r!} \frac{\sigma_t \lambda{-}1 - t\Gamma(r+1)}{(a(t+1))^{r+1}}$$

and the incomplete moments $\Phi_y(k)$ is defined by

$$\Phi_y(k) = \int\limits_0^k x^y f(x)dx.$$

If (3.3) is substituted, then $\Phi_y(k)$ is written as follows

$$\sum_{t=0}^{\infty} \sigma_t \lambda^{-1-t} \int\limits_0^k x^y H^{(a(t+1)x)} dx$$

then for $H = e$,

$$\Phi_y(k) = \sum_{t=0}^{\infty} \sigma_t \lambda^{-1-t} \int\limits_0^k x^y e^{-(-a(t+1)x)} dx.$$

By using the lower incomplete gamma function, we obtain the following

$$\Phi_y(k) = \sum_{t=0}^{\infty} - \frac{\sigma_t \lambda^{-1-t} \upsilon(y+1, a(t+1)k)}{(a(t+1))^{y+1}}$$

where $\upsilon(y,k) = \int\limits_0^k x^{y-1} e^{-x} dx$ is the lower incomplete gamma function.

Also, the conditional moments, let $\xi_y(k)$, is given by

$$\xi_y(k) = \int_t^{\infty} x^y f(x)dx.$$

Therefore, using the pdf in (3.3), one obtains

$$\xi_y(k) = \sum_{t=0}^{\infty} \sigma_t \lambda^{-1-t} \int_t^{\infty} x^y H^{(a(t+1)x)} dx.$$

Then, for $H = e$,

$$\xi_y(k) = \sum_{t=0}^{\infty} \sigma_t \lambda^{-1-t} \int_t^{\infty} x^y e^{-(-a(t+1)x)} dx.$$

Finally, if we use the upper incomplete gamma function, we obtain

$$\xi_y(k) = \sum_{t=0}^{\infty} - \frac{\sigma_t \lambda^{-1-t} \Omega(y+1, a(t+1)k)}{(a(t+1))^{y+1}}$$

where $\Omega(y,k) = \int_k^{\infty} x^{y-1} e^{-x} dx$ is the upper incomplete gamma function.

3.4. **Residual life function.** The n^{th} moment of the reversed residual life of X is defined by

$$M_n(k) = \frac{1}{R(k)} \int_0^k (x-k)^n f(x)dx$$

If we apply the binomial expansion of $(x-k)^n$ to the formula above, we obtain the following equation

$$M_n(k) = \frac{1}{R(k)} \sum_{t=0}^{\infty} \sigma_t \lambda^{-1-t} \int_0^k (x-k)^n H^{(a(t+1)x)} dx.$$

Then, for $H = e$,

$$M_n(k) = \frac{1}{R(k)} \sum_{t=0}^{\infty} \sum_{d=0}^{n} (-1)^{k+1} \lambda^{-1-t} k^d \binom{n}{d} \frac{\sigma_t \upsilon(n-d+1, a(t+1)k}{(a(t+1))^{n-d+1}},$$

where $\upsilon(y,k)$ is the lower incomplete gamma function.

The n^{th} moment of the residual life of X is defined as

$$m_n(k) = \frac{1}{R(k)} \sum_{t=0}^{\infty} \sigma_t \lambda^{-1-t} \int_k^{\infty} (x-k)^n H^{(a(t+1)x)} dx$$

If we apply the binomial expansion of $(x-k)^n$ to the formula above, we obtain the following equation

$$m_n(k) = \frac{1}{R(k)} \sum_{t=0}^{\infty} \sum_{d=0}^{n} (-1)^{n+d+1} \lambda^{-1-t} (-k)^d \binom{n}{d} \frac{\sigma_t \Omega(n-d+1, a(t+1)k}{(a(t+1))^{n-d+1}}),$$

where $\Omega(y,k)$ is the upper incomplete gamma function.

3.5. **Rėnyi and q-entropies.** Entropy serves as a metric for the diversity or vulnerability of a random variable X. Among the prevalent measures of entropy are Rėnyi, q and Shannon. Rėnyi entropy, in particular, holds significant importance across various fields such as statistical inference, classification, problem identification in statistics, econometrics, and pattern recognition in computer sciences. The provided theorem furnishes an expression for Rėnyi entropy, further elucidating its significance in quantitative analysis and modeling.

The Rėnyi entropy can be derived using the below relation

$$H_\omega(X) = \frac{1}{1-\omega}\log\int\limits_{-\infty}^{\infty}[f(x)]^\omega\,dx, \quad \omega > 0 \text{ and } \omega \neq 1)$$

By applying the binomial formula (3.2) in the pdf 2.1, then the pdf $[f(x)]^\omega$ can be expressed as follows

$$[f(x)]^\omega = \sum_{t=0}^{\infty}\rho_t\lambda^{-1-t}H^{-(-a(t+\omega)x)}$$

where $\rho_t = (-1)^t(a\ln(H))^\omega\binom{2\omega+t-1}{t}$.

Thus, the Rėnyi entropy of TPLT distribution is given by

$$H_\omega(X) = \frac{1}{1-\omega}\log\left[\sum_{t=0}^{\infty}\rho_t\int\limits_{0}^{\infty}H^{-(-a(t+\omega)x)}dx\right],$$

that is,

$$H_\omega(X) = \frac{1}{1-\omega}\log\left[\sum_{t=0}^{\infty}-\frac{\rho_t\lambda^{-1-t}}{a(t+w)x\ln H}\right].$$

The q - entropy is defined by following relation

$$H_\omega(X) = \frac{1}{1-q}\log\left(1-\int\limits_{-\infty}^{\infty}[f(x)]^q\,dx\right), \quad q > 0 \text{ and } q \neq 1$$

thus, the q - entropy of TPLT distribution is given by

$$H_\omega(X) = \frac{1}{1-q}\log\left(1-\left[\sum_{t=0}^{\infty}-\frac{\rho_t\lambda^{-1-t}}{a(t+q)x\ln H}\right]\right).$$

3.6. **Maximum likelihood estimation.** Maximum likelihood estimates (MLE) of the parameters for the three - parameter logistic-type distribution (TPLT distribution) are derived using complete samples. Suppose X_1, X_2, ..., X_n are denote observed values from the TPLT distribution with the parameter vector $\gamma = (a,H,\lambda)^T$. The log - likelihood function for γ can be expressed

$$\ln L(\gamma) = n\ln\lambda + n\ln a + n\ln(\ln H) + a\ln H\sum_{i=1}^{n}x_i - 2\sum_{i=1}^{n}\ln(\lambda + H^{ax_i}).$$

where $x_0 = 0$.

The components of the score function $U(\gamma) = (U_a, U_H, U_\lambda)$ are given by

$$\begin{aligned}
\mathbf{U}_a &= \frac{n}{a} + \ln H \sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} \frac{x_i H^{ax_i} \ln H}{\lambda + H^{ax_i}} \\
\mathbf{U}_H &= \frac{n}{H \ln H} + \frac{a}{H} \sum_{i=1}^{n} x_i - 2a \sum_{i=1}^{n} \frac{x_i H^{ax_i - 1}}{\lambda + H^{ax_i}} \\
\mathbf{U}_\lambda &= \frac{n}{\lambda} - 2\sum_{i=1}^{n} \frac{1}{\lambda + H^{ax_i}}
\end{aligned}$$

The maximum likelihood estimates (MLE) of the parameters are obtained by setting the last two equations to zero and solving them. Clearly, solving these equations is challenging, hence the Newton-Raphson iterative method is employed using Python to perform the maximum likelihood estimation for the parameter a. The Newton-Raphson method is a fast and efficient iterative technique used to find the roots of a function. In this method, the root is iteratively found using the derivative based on the initial value. Below, the MLE estimation is performed first for simulation data and then for breast cancer data. Maximum likelihood estimation (MLE) is used to find the values of parameters that maximize the likelihood of the observed data. This method is widely used to determine the most suitable parameters for parametric models. Obviously, it is difficult to solve them, therefore, by applying Newton Raphson's iteration method and using Python, the maximum likelihood estimation with respect to parameter a was performed. In the following, MLE estimation is performed first for the simulation data and then for the breast cancer data.

4. Simulation data

Comparing the theoretical performances of various estimators (MLE) for the distribution of three-parameter logistic-type function (TPLT distribution) is particularly challenging. Therefore, a simulation study is conducted to assess the performance of different estimation methods, focusing primarily on their biases, mean square errors (MSEs), and for varying sample sizes. The numerical study is implemented using Python, considering different sample sizes: $n = 20, 30, 50, 100, 200$ and 300. Additionally, various values of parameters a, H and λ are examined. In each iteration, parameter estimates are obtained using maximum likelihood estimation methods. In Table 1, it is observed that as n increases, we approach the initial values of the parameter a, that is, the bias becomes smaller. In Table 2 shows that as parameter H increases, better results are obtained for the same a values. Finally, it is seen in Table 3 that when the value of the parameter λ is increased, the a value is better approximated to the initial value.

Set1 : $H = 2.5$; $\lambda = 1$

n	Par	$Init$	MLE	$Bias$	MSE
20	a	0.8	1.035	0.235	0.055
		0.6	0.623	0.023	0.000529
30	a	0.8	0.903	0.103	0.010
		0.6	0.614	0.014	0.000196
50	a	0.8	0.862	0.062	0.003
		0.6	0.609	0.009	0.000081
100	a	0.8	0.831	0.031	0.0009
		0.6	0.605	0.005	0.00025
200	a	0.8	0.813	0.013	0.000169
		0.6	0.602	0.002	0.000001
300	a	0.8	0.808	0.008	0.000064
		0.6	0.601	0.001	0.000001

Table 1: The parameter estimation from TPLT distribution using MLE

Set2 : $H = 3.5$; $\lambda = 1$

n	Par	$Init$	MLE	$Bias$	MSE
20	a	0.8	0.870	0.070	0.004900
		0.6	0.6056	0.023	0.000031
30	a	0.8	0.839	0.039	0.001521
		0.6	0.6034	0.034	0.000011
50	a	0.8	0.826	0.026	0.000676
		0.6	0.6026	0.009	0.0000067
100	a	0.8	0.814	0.031	0.000196
		0.6	0.605	0.005	0.00025
200	a	0.8	0.806	0.06	0.00036
		0.6	0.602	0.002	0.000001
300	a	0.8	0.804	0.004	0.000016
		0.6	0.600	0.001	0.000001

Table 2: The parameter estimation from TPLT distribution using MLE

Set3 : $H = 2.5$; $\lambda = 10$

n	Par	$Init$	MLE	$Bias$	MSE
20	a	0.8	1.004	0.204	0.041616
		0.6	0.6016	0.0016	0.00000256
30	a	0.8	0.892	0.092	0.008464
		0.6	0.609	0.0090	0.000081
50	a	0.8	0.854	0.054	0.002916
		0.6	0.605	0.0050	0.000025
100	a	0.8	0.824	0.024	0.000576
		0.6	0.6001	0.0001	0.00000001
200	a	0.8	0.811	0.011	0.000121
		0.6	0.600006	0.000006	0.000000000000036
300	a	0.8	0.807	0.007	0.000049
		0.6	0.6000004	0.000004	0.000000000000016

Table 3: The parameter estimation from TPLT distribution using MLE

In this simulation data, it can be seen that the larger H or the smaller the value of a, the closer the MLE calculation is to the given initial value. For these results, much smaller biases and MSEs can be obtained with different tables. In Table 3, it is also observed that as the parameter λ increases, Bias and MSE become smaller.

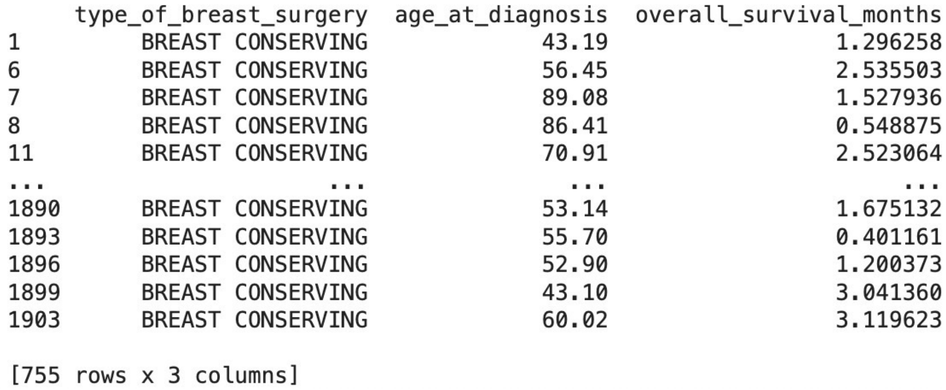
5. Real life applications

In this section, two real data sets are used to compare the fits of the TPLT distribution with those of the TIHLE and exponential (E) distributions. For both data sets, parameters are estimated using the maximum likelihood method. Tables 4 and 6 list the maximum likelihood estimates (MLEs) of the parameters for data sets 1 and 2, respectively. The numerical values of AIC, CAIC, BIC, HQIC, and related statistics are presented in Tables 4 and 6 for data sets 1 and 2, respectively. We consider criteria such as the Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC). Generally, lower values of these criteria indicate a better fit to the data. The formulas for these criteria are as follows:

$$AIC = 2k - 2\ln L \qquad BIC = k\ln(n) - 2\ln L$$
$$CAIC = AIC + \frac{2k(k+1)}{n-k-1} \qquad HQIC = 2k\ln(\ln(n)] - 2\ln L$$

This paper discusses a data set collected by Professor Carlos Caldas from the Cambridge Research Institute and Professor Sam Aparicio from the British Columbia Cancer Centre in Canada. The data set includes clinical data from 1904 individuals with breast cancer, comprising a total of 1904 rows and 693 columns [15]. For our analysis, we focus primarily on three columns: "type_of_breast_surgery" (type of breast surgery), "age_at_diagnosis" (age at diagnosis), and "overall_survival_months" (overall survival time in months). The "type_of_breast_surgery" column indicates the type of surgery undergone by patients, with two different types identified. The "age_at_diagnosis" column specifies the age of patients at the time of diagnosis.

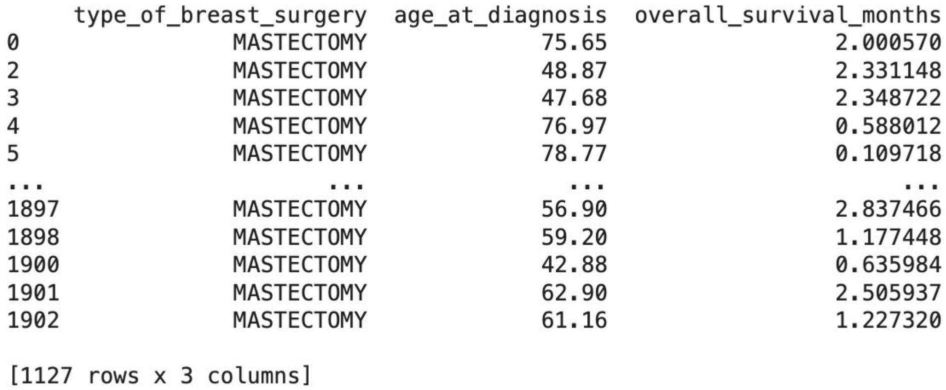
Lastly, the ”overall_survival_months” column denotes the duration from surgery until death (in months) for individuals with breast cancer. We divide the dataset into two distinct groups: the first group consists of patients who underwent BREAST CONSERVING surgery, while the second group comprises those who underwent MASTECTOMY. Subsequently, we apply Min-max normalization to the values in the ”overall_survival_months” column, transforming each value into a range of $[0, 5]$.



```
     type_of_breast_surgery  age_at_diagnosis  overall_survival_months
1          BREAST CONSERVING             43.19                 1.296258
6          BREAST CONSERVING             56.45                 2.535503
7          BREAST CONSERVING             89.08                 1.527936
8          BREAST CONSERVING             86.41                 0.548875
11         BREAST CONSERVING             70.91                 2.523064
...                      ...               ...                      ...
1890       BREAST CONSERVING             53.14                 1.675132
1893       BREAST CONSERVING             55.70                 0.401161
1896       BREAST CONSERVING             52.90                 1.200373
1899       BREAST CONSERVING             43.10                 3.041360
1903       BREAST CONSERVING             60.02                 3.119623

[755 rows x 3 columns]
```

Figure 4. Dataset 1.



```
     type_of_breast_surgery  age_at_diagnosis  overall_survival_months
0                MASTECTOMY             75.65                 2.000570
2                MASTECTOMY             48.87                 2.331148
3                MASTECTOMY             47.68                 2.348722
4                MASTECTOMY             76.97                 0.588012
5                MASTECTOMY             78.77                 0.109718
...                     ...               ...                      ...
1897             MASTECTOMY             56.90                 2.837466
1898             MASTECTOMY             59.20                 1.177448
1900             MASTECTOMY             42.88                 0.635984
1901             MASTECTOMY             62.90                 2.505937
1902             MASTECTOMY             61.16                 1.227320

[1127 rows x 3 columns]
```

Figure 5. Dataset 2.

Parameter estimates for parameters a and H were made according to data Set 1 and data Set 2 in Table 4 and Table 6, respectively. Based on these estimates, it was observed that the AIC, CAIC, BIC, HQIC values of the TPLT distribution were more compatible with the breast cancer data set than those of the TIHLE and E distributions

	Estimated Parameters			
Distribution	a	H	λ	α
TPLT ($\lambda = 1, x_0 = 0$)	0.437	2.716	-	-
TIHLE	-	-	0.605	1.148
E	-	-	0.482	-

Table 4: ML estimates of the model parameters for first data set.

Model	AIC	CAIC	BIC	HQIC
TPLT	**2070.8927**	**2070.8980**	**2077.5194**	**2074.6749**
TIHLE	2468.2881	2468.2934	2474.9148	2472.0703
E	2614.7423	2614.7476	2621.3690	2618.5245

Table 5: Goodness of fit measures for estimates for the first data set.

	Estimated Parameters			
1	—l— TPLT ($\lambda = 1$, $x_0 = 0$)	0.451	3.841	-
1	—l— E	-	-	0.654

Table 6: ML estimates of the model parameters for second data set

Model	AIC	CAIC	BIC	HQIC
TPLT	**2202.5870**	**2202.5905**	**2209.6143**	**2206.4866**
TIHLE	3311.5381	3311.5416	3318.5654	3315.4377
E	3473.3245	3473.3280	3506.9712	3477.2241

Table 7: Goodness of fit measures for estimates for the second data set.

Additionally, the fitted densities for the first and second data sets are depicted in Figures 4 and 5, along with the corresponding data histograms. These results illustrate the potential of the TPLT distribution compared to the exponential and TIHLE distributions.

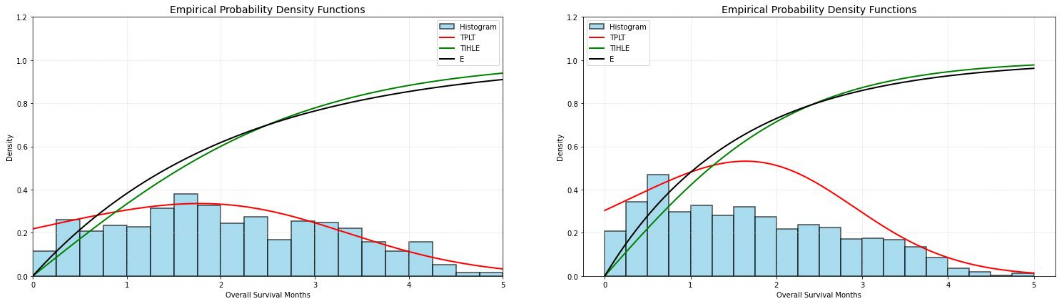


FIGURE 6. Estimates of the density function for the first and second data sets.

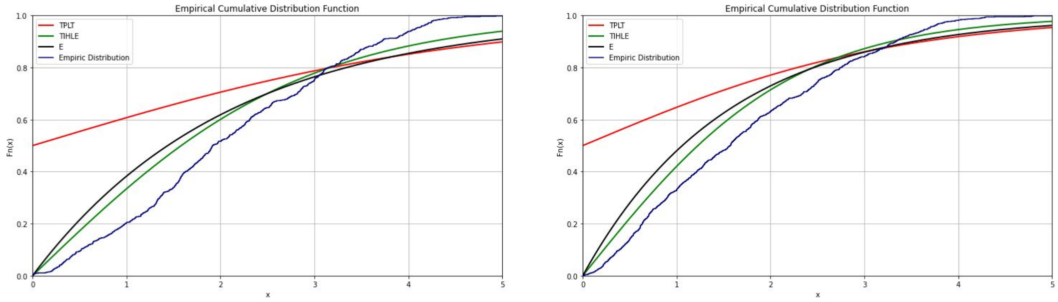


Figure 7. Estimates of the distribution function and empirical distribution for the first and second data sets.

6. Conclusions

In this paper, the cumulative distribution function (cdf) and probability density function of the TPLT distribution were obtained. Some statistical properties were found using this pdf. MLE is calculated according to the a parameter, and Table 1, Table 2 and Table 3 show that as the a value decreases and the H, λ value increases, results close to the initial value are obtained. Finally, MLE calculations were made for two real breast cancer data sets and AIC, CAIC, BIC, HQIC were compared for TIHLE distribution and exponential distribution. Finally, the fitted densities for the first and second data sets are shown in Figures 4 and 5 (along with the data histogram), respectively. These results show the potential of the TPLT distribution over the TIHLE distribution and the exponential distribution.

Data availability

The datasets used and/or analyzed during the current study are publicly available on Kaggle at `https://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric`

Competing interests

The authors declare that they have no competing interests.

References

[1] N. Eugene, C. Lee, F. Famoye, *The beta-normal distribution and its applications*, Commun. Stat. Theory Methods, 31 (2002), 497–512.

[2] K. Zografos, N. Balakrishnan, *On families of beta- and generalized gamma-generated distributions and associated inference*, Statistical Methodology, 6 (2009), 344–362.

[3] G. M. Cordeiro, M. de Castro, *A new family of generalized distributions*, J. Stat. Comput. Simul., 81 (2011), 883–893.

[4] M. M. Ristic, N. Balakrishnan, *The gamma-exponentiated exponential distribution*, J. Stat. Comput. Simul., 82 (2012),1191–1206.

[5] G. M. Cordeiro, M. Alizadeh, E. M. M. Ortega, *The exponentiated half-logistic family of distributions: Properties and applications*, J. Probab. Stat., 2014 (2014), 21 pages.

[6] G. M. Cordeiro, M. Alizadeh, P. R. D. Marinho, *The type I half-logistic family of distributions*, J. Stat. Comput. Simul.,86 (2015), 707–728.

[7] M. Elgarhy, A. S. Hassan, M. Rashed, *Garhy-generated family of distributions with application*, Math. Theory Model.,6 (2016), 1–15.

[8] A. S. Hassan, M. Elgarhy, *A New family of exponentiated Weibull-generated distributions*, Int. J. Math. Appl., 4 (2016),135–148.

[9] A. S. Hassan, M. Elgarhy, *Kumaraswamy Weibull-generated family of distributions with applications*, Adv. Appl. Stat.,48 (2016), 205–239.

[10] A. S. Hassan, M. Elgarhy, M. Shakil, *Type II half Logistic family of distributions with applications*, Pak. J. Stat. Oper.Res., 13 (2017), 245–264.

[11] M. Elgarhy, M. Haq, G. Ozel, *A new exponentiated extended family of distributions with Applications*, Gazi University J. Sci., 30 (2017), 101–115.

[12] M. Haq, M. Elgarhy, *The odd Frechet-G family of probability distributions*, J. Stat. Appl. Prob., 7 (2018), 185–201.

[13] G. A. Anastassiou, *Banach Space Valued Multivariate Multi Layer Neural Network Approximation Based on q-Deformed and -Parametrized A-Generalized Logistic Function.* In: Parametrized, Deformed and General Neural Networks, Studies in Computational Intelligence, vol 1116. Springer, Cham., 2023, 365–394.

[14] A. M. Almarashi, M. Elgarhy, M. M. Elsehetry, B. M. G. Kibria, A. Algarni, *A new extension of exponential distribution with statistical properties and applications*, J. Nonlinear Sci. Appl., 12 (2019), 135–145.

[15] https://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric

Received 25 July 2025

[1,2] Ankara University, Graduate School of Natural and Applied Sciences, Ankara, Turkey

[3] Istanbul Atlas University, Faculty of Engineering and Natural Sciences, Department of Software Engineering, 34408, Istanbul, Turkey

[4] Ankara University, Faculty of Science, Department of Mathematics, Ankara, Turkey

Email address: [1] `matozan152@gmail.com`, [2] `huyan@ankara.edu.tr`, [3] `seda.karateke@atlas.edu.tr`, [4] `ibuyukyazici@gmail.com`