Journal of Biostatistics and Biometric Applications
Volume 9 | Issue 1
ISSN: 2455-765X

Research Article Open Access

Type1 Heavy-Tailed Mixture Cure Rate Survival Model Based on a Variant of T-X Family of Distribution as Baseline

Ligeiaziba Sylva*

[1]Department of Mathematics, Bayelsa Medical University, Yenagoa, Bayelsa State, Nigeria

***Corresponding Author:** Ligeiaziba Sylva, Department of Mathematics, Bayelsa Medical University, Yenagoa, Bayelsa State, Nigeria, Tel.: +2348143249582,E-mail: sylva.ligeiaziba@bmu.edu.ng

Citation: Ligeiaziba Sylva (2024) Type1 Heavy-Tailed Mixture Cure Rate Survival Model Based on a Variant of T-X Family of Dis tribution as Baseline, J Biostat Biometric App 9(1): 102

Received Date: April 03, 2024 **Accepted Date:** May 03, 2024 **Published Date:** May 08, 2024

Abstract

In this research work, a new class of heavy-tailed mixture cure rate models; Type 1 Heavy-Tailed Gamma (T1HT-G) mixture cure rate model was introduced using a new class of heavy-tailed distributions; Type 1 Heavy-Tailed Gamma (T1HT-G) distribution as baseline. The maximum likelihood parameters estimation approach was adopted for estimating the model parameters. Also, the Monte Carlo simulation approach was adopted to assess the performance of the maximum likelihood parameter estimation. Deviance information criteria such as AIC, BIC and CAIC were adopted to measure the models' performance. The simulation studies were conducted using three different sample sizes and 50 replications. Results from the model applications using real life biological and biomedical data. Comparative measures from the models TI-HTG mixture cure rate (AIC=51.60, BIC=61.55, CAIC=52.08) were smaller which showed the adequacy of the models to provide better fit for heavy-tailed data than the well--known standard distributions. These results have supported the fact that heavy tailed models provide better fits than the usual standard distributions when analyzing heavy-tailed data.

Keywords: Heavy-tail; Mixture Cure Rate Model and Baselines

Introduction

Many real-life applications have shown that many research areas, particularly in health science and finance, data are usually positive, and their distribution is usually a unimodal hump shaped and extreme values yield tails which are heavier than those of the standard well-known distributions [1]. Traditional classical distributions often fall short in accurately modeling such data, making it challenging to draw meaningful conclusions and potentially leading to the loss of crucial data characteristics. This is because the classical distributions are not flexible enough to handle such heavy-tailed data sets. Heavy-tailed distributions are more flexible and reliable to use.

To address this issue, a more robust and adaptable distribution is required, given the unique nature of the data being analyzed. T-X family of distribution techniques offers a better fit for data with heavy tails compared to classical distributions, variable transformations, and combinations of multiple distributions [2]. This was shown through a comparative analysis of the T-X method, conventional distributions, variable transformations, and combinations of distributions. In this context, a heavy-tailed distribution, as defined as one in which the probability of extreme values in the right tail is significantly higher than that of an exponential distribution [3].

That is, a distribution *f (x)* is said to be heavy tail if its survival function satisfies the condition;

$$\lim_{x\to\infty}\frac{1-F(x)}{e^{-\alpha x}}=\infty\forall\alpha>0$$

Literature holds records of many developments and applications of heavy-tailed distribution to real life data. Probability distributions are very important for parametric and semi-parametric inferences [4]. Literature holds records of the development of many methods used to develop probability distributions. They had a good understanding of the initial methods used to generate univariate continuous distributions. These methods included Johnson's (1949) translation procedures, Tukey's (1960) quantile function approaches, and Pearson's (1895) differential equation techniques. In recent decades, there has been a consistent effort to develop new methods for creating more versatile and innovative distributions. Lee et al. (2013) categorize most of the approaches developed after the 1980s as "combination" techniques, as they are based on the idea of combining two existing distributions or enhancing an existing distribution with additional features to create a new family of distributions. In this work, we have proposed a new robust and flexible mixture cure rate models for handling survival data with heavy-tails using two a variant of T-X family of distributions as basslines distribution.

Methodology

In the development of the new distribution, we have considered a family of distributions and a baseline function after preliminary analysis of the data. The family of distribution and the baseline adopted are discussed below.

Family of Distribution

The T-X family of distributions was adopted for the development of the new distribution. Zhao et al (2020) developed a new family of distributions called Type1 Heavy-tailed Family of Distributions. This family of distribution was adopted because to its flexibility and applicability.

Baseline Distribution: Gamma Distribution

Gamma distribution was adopted as the baseline distribution for the new heavy-tailed distributions. The Gamma distribution is one of the probabilistic tools used to evaluate the appropriate model for predicting continuous random variables. The gamma distribution stands apart from distributions with one, two, three, or four parameters due to its unique characteristic of having only a single parameter. This distribution becomes particularly relevant when one is interested in estimating the time required for a specific number of independent events to transpire.

Consider a $\alpha>0$ and $\beta>0$ as two parameters of a distribution. A Gamma function is denoted as $\Gamma(\alpha)$ and defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Dividing both sides by Γ(α)

$$1 = \int_0^\infty \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} dx = \int_0^\infty \frac{\beta^\alpha y^{\alpha-1} e^{-\beta x} dy}{\Gamma(\alpha)}$$

Substituting $x=\beta y$ and defined

$$f(x;\alpha,\beta) = \begin{cases} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, x \geq 0 \\ 0, x < 0 \end{cases} \quad (1.1)$$

The distribution denoted as $f(x;\alpha,\beta)$ is known as the Gamma distribution, with the parameter $\alpha > 0$ being referred to as the shape parameter. This parameter primarily affects the distribution's peak, as discussed by Stephenson et al. in 1999. When α is greater than 0, it is alternatively termed the scale parameter, as it primarily influences the dispersion of the distribution. The gamma distribution with two parameters is commonly employed for analyzing lifespan data [6]. Furthermore, it highlighted that the Gamma distribution is the most frequently used statistical distribution in the fields of reliability analysis and survival analysis (7). In essence, the Gamma distribution is the preferred choice for statistical analysis in these domains.

Probability Distribution Function (PDF) for Type 1 Heavy Tailed Family of Distributions

The PDF of the distribution developed by Zhao *et al* (2020) is defined as;

$$g(x,\theta,\mathcal{E}) = \frac{\theta^2 f(x,\mathcal{E}) \{1-F(x,\mathcal{E})\}^{\theta-1}}{\{1-(1-\theta)F(x,\mathcal{E})\}^{\theta+1}} \quad (1.2)$$

where,

θ is the additional parameter for the family of distribution

ε is the parameter space of the baseline distribution.

$f(x,\varepsilon)$ is the pdf of the baseline distribution

$F(x,\varepsilon)$ is the CDF of the baseline distribution

Cumulative Distribution Function (CDF) for Type 1 Heavy Tailed Family of Distributions

Zhao et al (2020) defined the cumulative distribution function of Type 1 Heavy tailed family of distribution as:

$$G(x,\theta,\mathcal{E}) = 1 - \left[\frac{1-F(x,\mathcal{E})}{1-(1-\theta)F(x,\mathcal{E})}\right]^\theta \quad (1.3)$$

where,

θ is the additional parameter for the family of distribution

ε is the parameter space of the baseline distribution.

Survival Function of the Type 1 Heavy Tailed Family of Distributions

Zhao et al (2020) defines the survival function of Type 1 Heavy tailed family of distribution as:

$$S\left(x\right)=1-G\left(x,\theta,\mathcal{E}\right)\quad(1.4)$$

where

G (x, θ, ε) is the cumulative distribution function for Type 1 Heavy Tailed family of distribution

Hazard Function of the Type 1 Heavy Tail Family of Distributions

The Hazard Function of the Type 1 Heavy Tail family of distribution is given as

$$h\left(x\right)=\frac{g\left(x\right)}{S\left(x\right)}\quad(1.5)$$

Rth Moment of the Proposed Model for Type 1 Heavy Tailed Family of Distributions

In this research, effort was made to derive the $\mu_r{'}$ for type 1 heavy tailed gamma distribution by deriving $K_{(r,i+j)}$ for the proposed model. Zhao et al (2020) derived the rth moment for type 1 heavy tailed family of distribution as

$$\mu_r^{'}=\theta^2\sum_{i=0}^{\infty}\sum_{j=0}^{\theta-1}\begin{pmatrix}\theta-1\\ j\end{pmatrix}\begin{pmatrix}\theta+i\\ \theta\end{pmatrix}(-1)^j(1-\theta)^i k_{r,i+j}\quad(1.6)$$

where

θ is the additional parameter for the family of distribution and

$$K_{r,i+j}=\int_{-\infty}^{\infty}x^r f\left(x\right)F(x)^{i+j}dx$$

Results and Discussion

Derivation of the PDF of Type1 Heavy-tailed Gamma(T1HT-G)

Denote the pdf of T1HT-G as $g_0(x,\theta,\varepsilon)$. then $g_0(x,\theta,\varepsilon)$ is gotten by substituting equation (1.1) into (1.2) and defined

$$F\left(x,\mathcal{E}\right)=\frac{\gamma\left(\alpha,\beta x\right)}{\Gamma\left(\alpha\right)}\quad(1.7)$$

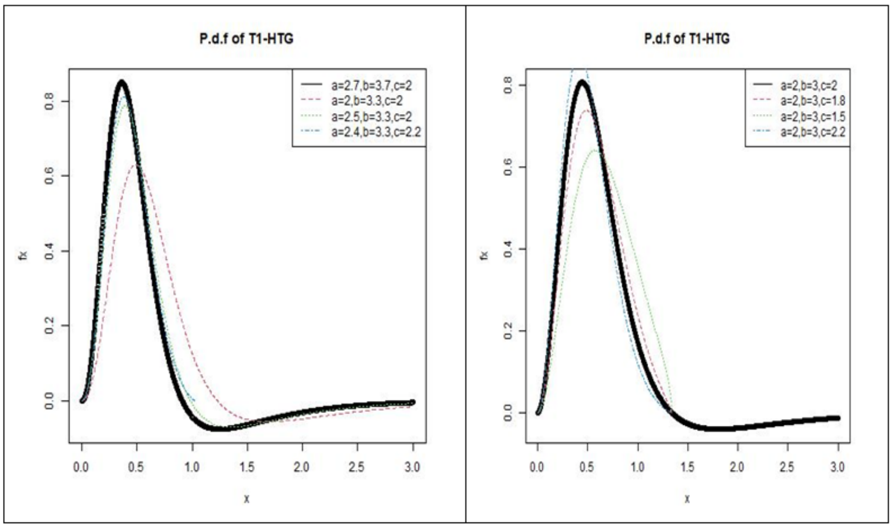
where γ (α, βx) is the lower incomplete gamma function and it is defined as

$$\gamma\left(\alpha,\beta x\right)=\int_{0}^{\beta x}x^{\alpha-1}e^{-x}dx\quad(1.8)$$

$$g_0(x,\theta,E)=\frac{\theta^2\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\left\{1-\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)}\right\}^{\theta-1}}{\left\{1-(1-\theta)\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)}\right\}^{\theta+1}}$$

$$g_0(x,\theta,\mathcal{E}) = \frac{\theta^2 \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \left\{1-\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)}\right\}^{\theta-1} / \Gamma(\alpha)^\theta}{\left\{1-(1-\theta)\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)} / \Gamma(\alpha)^\theta\right\}^{\theta+1}}$$

$$g_0\left(x,\theta,\mathcal{E}\right) = \frac{\theta^2 \beta^\alpha x^{\alpha-1} e^{-\beta x} \Gamma\left(\alpha\right) \left\{\Gamma\left(\alpha\right) - \gamma\left(\alpha,\beta x\right)\right\}^{\theta-1}}{\left\{\Gamma\left(\alpha\right) - \left(1-\theta\right)\gamma\left(\alpha,\beta x\right)\right\}^{\theta+1}} \quad (1.9)$$



Figures 1.1: PDF plot of T1HT-G

Figures 1.1 shows the probability density function plots of the T1HT-G for different parameter values. The plotted graphs showed that the curves are skewed to the right and also showing heavy tails. This establish the ability of the T1HT-G distribution to handle skewed and heavy tailed data.

Derivation of the Cumulative Distribution Function (CDF) for Type 1 Heavy-tailed Gamma

Cumulative Distribution function (CDF) proposed Zhao et al (2020) defined the CDF of the Type1 Heavy-tailed family of distribution as:

$$G\left(x,\theta,\mathcal{E}\right) = 1 - \left[\frac{1-F\left(x,\mathcal{E}\right)}{1-\left(1-\theta\right)F\left(x,\mathcal{E}\right)}\right]^\theta \quad (1.10)$$

Thus, the CDF of the T1HT-G is derived as follows

$$G(x,\theta,\mathcal{E}) = 1 - \left[\frac{1-\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)}}{1-(1-\theta)\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)}}\right]^\theta$$

$$G\left(x,\theta,\mathcal{E}\right) = 1 - \left[\frac{\Gamma\left(\alpha\right) - \gamma\left(\alpha,\beta x\right)/\Gamma\left(\alpha\right)}{\Gamma\left(\alpha\right) - \left(1-\theta\right)\gamma\left(\alpha,\beta x\right)/\Gamma\left(\alpha\right)}\right]^\theta$$

$$G\left(x,\theta,\mathcal{E}\right) = 1 - \left[\frac{\Gamma\left(\alpha\right) - \gamma\left(\alpha,\beta x\right)}{\Gamma\left(\alpha\right) - \left(1-\theta\right)\gamma\left(\alpha,\beta x\right)}\right]^\theta \quad (1.11)$$

Therefore, the CDF of the proposed model is given by (Eq.1.11)

Derivation of the Survival Function of Type 1 Heavy-Tailed Gamma Baseline

The Survival Function of the Proposed Model is given as

$$S_{T1HT}(x) = 1 - G(x, \theta, \mathcal{E})$$

$$S_{T1HT}(x) = 1 - 1 - \left[\frac{\Gamma(\alpha) - \gamma(\alpha, \beta x)}{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)}\right]^{\theta} \quad (1.12)$$

$$S_{T1HT}(x) = \left[\frac{\Gamma(\alpha) - \gamma(\alpha, \beta x)}{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)}\right]^{\theta} \quad (1.13)$$

Deviation of the Hazard Function of the Type 1 Heavy-Tailed Gamma

The Hazard Function of the proposed model is given as

$$h(x) = \frac{g(x)}{S(x)}$$

$$h(x) = \frac{\frac{\theta^2 \beta^\alpha x^{\alpha-1} e^{-\beta x} \Gamma(\alpha) \{\Gamma(\alpha) - \gamma(\alpha, \beta x)\}^{\theta-1}}{\{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)\}^{\theta+1}}}{\left[\frac{\Gamma(\alpha) - \gamma(\alpha, \beta x)}{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)}\right]^{\theta}}$$

$$h(x) = \frac{\theta^2 \beta^\alpha x^{\alpha-1} e^{-\beta x} \Gamma(\alpha) \{\Gamma(\alpha) - \gamma(\alpha, \beta x)\}^{-1}}{\{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)\}} \quad (1.14)$$

Some Statistical Properties for the Type1 Heavy-Tailed Gamma Distribution

R^{th} Moment of the for type 1 heavy Tailed Family of Distribution

Furthermore, Zhao et al (2020) derived the rth moment for type 1 heavy tailed family of distribution as

$$\mu_r^{'} = \theta^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\theta-1} \left(\begin{array}{c} \theta - 1 \\ j \end{array}\right) \left(\begin{array}{c} \theta + i \\ \theta \end{array}\right) (-1)^j (1-\theta)^i k_{r,i+j} \quad (1.15)$$

Where $K_{r,i+j} = \int_{-\infty}^{\infty} x^r f(x) F(x)^{i+j} dx$

In this research, effort was made to derive the μ_r' for type 1 heavy tailed gamma distribution by deriving $K_{(r,i+j)}$ for the proposed model

Recall that

$$F(x, \mathcal{E}) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$

The power series expansion of $\gamma(\alpha, \beta x)$ is

$$\gamma(\alpha, \beta x) = \sum_{k=0}^{\infty} \frac{(\beta x)^k (\beta x)^\alpha \Gamma(\alpha) e^{-\beta x}}{\Gamma(\alpha + k + 1)} = (\beta)^{k+\alpha} \Gamma(\alpha) \sum_{k=0}^{\infty} \frac{(x)^{k+\alpha} e^{-\beta x}}{\Gamma(\alpha + k + 1)}$$

$$K_{r,i+j} = \int_{-\infty}^{\infty} x^r \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \left[\sum_{k=0}^{\infty} \frac{(\beta)^{k+\alpha} (x)^{k+\alpha} e^{-\beta x}}{\Gamma(\alpha+k+1)} \right]^{i+j} dx$$

$$K_{r,i+j} = \sum_{k=0}^{\infty} \int_{-\infty}^{\infty} x^{r+\alpha-1+(i+j)(k+\alpha)} \frac{\beta^{\alpha+(i+j)(k+\alpha)}}{\Gamma(\alpha)} \frac{e^{-(\beta x+(i+j)\beta x)}}{\Gamma(\alpha+k+1)^{i+j}} dx$$

$$K_{r,i+j} = \sum_{k=0}^{\infty} \frac{\beta^{\alpha+(i+j)(k+\alpha)}}{\Gamma(\alpha)\Gamma(\alpha+k+1)^{i+j}} \int_{-\infty}^{\infty} x^{r+\alpha+(i+j)(k+\alpha)-1} e^{-(i+j+1)\beta x} dx$$

Let

$$y = (i+j+1)\beta x \Rightarrow x = \frac{y}{(i+j+1)\beta}$$

And

$$\frac{dx}{dy} = \frac{1}{(i+j+1)\beta}$$

$$K_{r,i+j} = \sum_{k=0}^{\infty} \frac{\beta^{\alpha+(i+j)(k+\alpha)}}{\Gamma(\alpha)\Gamma(\alpha+k+1)^{i+j}} \int_{-\infty}^{\infty} \left[\frac{y}{(i+j+1)\beta} \right]^{r+\alpha+(i+j)(k+\alpha)-1} e^{-y} \frac{1}{(i+j+1)\beta} dy$$

$$K_{r,i+j} = \frac{1}{\beta^r \Gamma(\alpha)} \sum_{k=0}^{\infty} \frac{\int_{-\infty}^{\infty} y^{r+\alpha+(i+j)(k+\alpha)-1} e^{-y} dy}{[(i+j+1)]^{r+\alpha+(i+j)(k+\alpha)} \Gamma(\alpha+k+1)^{i+j}}$$

$$K_{r,i+j} = \frac{1}{\beta^r \Gamma(\alpha)} \sum_{k=0}^{\infty} \frac{2\Gamma(r+\alpha+(i+j)(k+\alpha))}{[(i+j+1)]^{r+\alpha+(i+j)(k+\alpha)} \Gamma(\alpha+k+1)^{i+j}} \quad (1.16)$$

Equation (Eq.1.16) is the $K_{r,\,i+j}$ for type 1 heavy tailed gamma distribution and substituting it into (Eq.1.15) gives the Rth moment for the proposed model (Type 1 heavy tailed Gamma Distribution)

Some of the most important features and characteristics of a model can be obtain through its moments.

Derivation for the Mean of T1HT-G

The first raw moment also called the mean is obtained by substitution r=1 in equation (1.16).

$$\mu_r' = \theta^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\theta-1} \binom{\theta-1}{j} \binom{\theta+i}{\theta} (-1)^j (1-\theta)^i k_{r,i+j} \quad (1.17)$$

where

$$K_{1,i+j} = \frac{1}{\beta^r \Gamma(\alpha)} \sum_{k=0}^{\infty} \frac{2\Gamma(1+\alpha+(i+j)(k+\alpha))}{[(i+j+1)]^{1+\alpha+(i+j)(k+\alpha)} \Gamma(\alpha+k+1)^{i+j}}$$

Derivation for the Variance of T1HT-G

The variance denoted as μ_2 is obtained by

$$\mu_2 = \mu_2' - (\mu_{r'})^2 \quad (1.18)$$

Where μ_2' is the second raw moment obtained by

$$\mu_r' = \theta^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\theta-1} \binom{\theta-1}{j} \binom{\theta+i}{\theta} (-1)^j (1-\theta)^i k_{2,i+j} \quad (1.19)$$

where

$$K_{2,i+j} = \frac{1}{\beta^r \Gamma(\alpha)} \sum_{k=0}^{\infty} = \frac{2\Gamma(2+\alpha+(i+j)(k+\alpha))}{[(i+j+1)]^{2+\alpha+(i+j)(k+\alpha)} \Gamma(\alpha+k+1)^{i+j}}$$

Derivation of the Quantile Function

The quantile function denoted by *Q(p)* is the inverse of equation. (1.11) where $p \in (0,1)$.That is solve for x in the equation (1.17)

$$1 - \left[\frac{\Gamma(\alpha) - \gamma(\alpha, \beta x)}{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)}\right]^{\theta} = p \quad (1.17)$$

$$\frac{\Gamma(\alpha) - \gamma(\alpha, \beta x)}{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)} = (1-p)^{\frac{1}{\theta}}$$

$$\Gamma(\alpha) - \gamma(\alpha, \beta x) = (1-p)^{\frac{1}{\theta}} \Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)$$

$$\gamma(\alpha, \beta x) = \frac{\Gamma(\alpha)\left[1 - (1-p)^{\frac{1}{\theta}}\right]}{\left[1 - (1-p)^{\frac{1}{\theta}}(1-\theta)\right]}$$

Equation (1.18) is the quantile function of T1HT-G

The quantile function of T1HT-G 1st, 2nd and 3rd quantiles are obtained by substituting p = 0.25, 0.5 and 0.75 in equation (1.18), respectively

Some Special Cases of the T1HT-G Distribution

Different values for the parameters reduce the T1HT-G distribution to other forms. Some of which are listed below

1. θ =1, T1HT-G reduces to a Gamma distribution with parameters α and β

2. If $\alpha = \theta$ =1, T1-HT-G reduces T1HT-E (Type 1 Heavy-tailed Exponential distribution) with parameter $\beta>0$

3. If β = 1, T1HT-G reduces to a one parameter T1HT-G.

4. If β = 2 and $\alpha = \tau/2$ where τ is an integer , then T1HT-G reduces to a Type 1 Heavy tailed Chi Square (T1HT-CS) distribution (New distribution)

Log-Likelihood function for Type 1 Heavy Tailed Gamma Distribution for Complete Data Set

By definition

$$L(\Theta) = \prod_{i=1}^{n} g(x_1; \theta, \alpha, \beta) = \prod_{i=1}^{n} \frac{\theta^2 \beta^\alpha x^{\alpha-1} e^{-\beta x} \Gamma(\alpha) \{\Gamma(\alpha) - \gamma(\alpha, \beta x)\}^{\theta-1}}{\{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)\}^{\theta+1}} (1.19)$$

$$L(\Theta) = \frac{\theta^{2n} \beta^{n\alpha} x^{n(\alpha-1)} e^{-n\beta x} (\Gamma(\alpha))^n \{\Gamma(\alpha) - \gamma(\alpha, \beta x)\}^{n(\theta-1)}}{\{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)\}^{n(\theta+1)}} \quad (1.20)$$

Taking log on both sides and applying the logarithmic laws gives

$$l(\Theta) = 2n\log\theta + n\alpha\log\beta + (\alpha-1)\Sigma\log x - \beta\Sigma x + n\log\Gamma(\alpha) + n(\theta-1)\log\left[\Gamma(\alpha) - \gamma(\alpha,\beta x)\right] - n(\theta+1)\log\left[\Gamma(\alpha) - (1-\theta)\gamma(\alpha,\beta x)\right] \quad (1.21)$$

Maximum Likelihood Estimated for T1-HT-G Parameter for Complete Data Set

Maximizing l(θ) with respect to the parameters will give us the MLE estimation. solving equations below will yield the MLE for the parameter α ,β and θ

$$\frac{\partial\log l(x;\alpha,\beta,\theta)}{\partial\alpha} = 0 \quad (1.22a)$$
$$\frac{\partial\log l(x;\alpha,\beta,\theta)}{\partial\beta} = 0 \quad (1.22b)$$
$$\frac{\partial\log l(x;\alpha,\beta,\theta)}{\partial\theta} = 0 \quad (1.22c)$$

But the partial derivatives are not tractable, therefore we will estimate it numerically using a computer software.

Simulation Results for Complete Data

Table 1.1: Simulation Result for Parameter Estimation for T1HT-G and EOW-G using sample size of n=N T=50

N	Parameter	AE	SE	MSE
10	θ	2.38635	0.17006	0.036619
	α	2.55101	0.006442	1.64E-05
	β	3.77481	0.321439	0.01177
50	θ	1.85286	0.09345	0.008819
	α	0.94053	0.00211	1.01E-06
	β	2.97632	0.001296	0.01126
100	θ	1.45040	0.09772	0.009549
	α	0.48827	0.002293	5.23353E-06
	β	2.1095	0.001712	0.001611

Note: (*T1HT-G*: $\theta = 1.4$, $\beta=2.0$ and $\alpha = 0.5$)

Based on the simulation results presented in Tables 1.1, the estimations of the TIHT-G parameters exhibited favorable performance, displaying minimal bias and respectable Mean Square Errors (MSEs) across all tested scenarios. These findings emerged from the research, implying that with increasing sample size, these estimates become progressively more accurate and reliable. Furthermore, the evidence supporting the asymptotic unbiasedness of these estimators becomes apparent as the biases approach zero with the growth of the sample size. Additionally, these estimators demonstrate their validity by reducing the MSEs when applied to the TIHT-G parameters as the sample size increases.

Application to Real Data Set

To examine the usefulness and performance of the distributions we have applied it to real life complete data, it shows how to fitting the model to real life dada. We also carried out some comparative procedures using Akaike Information Criteria (AIC), Bayesian Information Criteria and Corrected Akaike Information Criteria (CAIC) to examine the performance of the model when compared to other sub-models

Application of T1HT-G to complete Data Set

The data is made up 72 Guinea Pigs survival times collected from research work by Elbatal **et al**. (2013). The pigs were infected and their survival times were measured and recorded. The data set is listed below:

0.1	0.33	0.44	0.56	0.59	0.72	0.74	0.77	0.92
0.93	0.96	1	1	1.02	1.05	1.07	1.07	1.08
1.08	1.08	1.09	1.12	1.13	1.15	1.16	1.2	1.21
1.22	1.22	1.24	1.3	1.34	1.36	1.39	1.44	1.46
1.53	1.59	1.6	1.63	1.63	1.68	1.71	1.72	1.76
1.83	1.95	1.96	1.97	2.02	2.13	2.15	2.16	2.22
2.3	2.31	2.4	2.45	2.51	2.53	2.54	2.54	2.78
2.93	3.27	3.42	3.47	3.61	4.02	4.32	4.58	5.55

Table 1.2: T1HT-G results for Guinea Pigs survival time

Model	Parameter	Estimates	loglik	AIC	BIC	CAIC
T1HT-G	θ	0.009	-70.38	146.76	156.29	147.29
	α	1.53				
	β	24.40				
	α	1.47				
	β	22.10				
	γ	2.24				
Gamma	α	2.180	-76.27	156.54	160.37	156.80
	β	1.105				
Weibull	α	0.02	-76.22	156.44	160.265614	156.69
	β	0.92				
	γ	0.24				
Exponential	α	0.50				
	β	0.49	-83.45	166.51	168.40	166.60

Table 1.2 showed the result of the analysis. The T1HT-G showed smaller values for all three informstion criteria. Which means T1HT-G performed better than the well known classical distribution.

Derivation of the T1HT-G mixture Model

A mixture model for lifetime data sets assumes that the probability of the time-to-event to be greater than a specified time t and defined the survival function as

$$S(t) = p + (1-p)\, S_{T1HT}(t) \quad (1.21)$$

$$g(t) = \frac{dF(t)}{dt} = (1-p)\, g_0(t) \quad (1.22)$$

Substituting

$$g_0(t) = \frac{\theta^2 \beta^\alpha t^{\alpha-1} e^{-\beta t} \Gamma(\alpha) \{\Gamma(\alpha) - \gamma(\alpha, \beta t)\}^{\theta-1}}{\{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta t)\}^{\theta+1}} \quad (1.23)$$

Into Eq. 4.24 gives

$$g(t) = (1-p) \frac{\theta^2 \beta^\alpha t^{\alpha-1} e^{-\beta t} \Gamma(\alpha) \{\Gamma(\alpha) - \gamma(\alpha, \beta t)\}^{\theta-1}}{\{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta t)\}^{\theta+1}} \quad (1.24)$$

Eq.4.26 is the T1HT-G density function for the lifetime T.

The Parameter Estimation of T1HT-G Mixture Cure Rate Model

The likelihood function of the mixture form of the cure rate model is

$$L = \prod_{i=1}^{n} [(1-p) g_0(x)]^{\delta_i} [p + (1-p) S_0(t)]^{(1-\delta_i)} \quad (1.25)$$

Substituting

$$g_0(x) = \frac{\theta^2 \beta^\alpha x^{\alpha-1} e^{-\beta x} \Gamma(\alpha) \{\Gamma(\alpha) - \gamma(\alpha, \beta x)\}^{\theta-1}}{\{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)\}^{\theta+1}}$$

And

$$S_0(x) = \left[\frac{\Gamma(\alpha) - \gamma(\alpha, \beta x)}{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)}\right]^\theta$$

Into Eq.4.27 gives

$$L = \prod_{i=1}^{n} \left[(1-p) \frac{\theta^2 \beta^\alpha x^{\alpha-1} e^{-\beta x} \Gamma(\alpha) \{\Gamma(\alpha) - \gamma(\alpha, \beta x)\}^{\theta-1}}{\{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)\}^{\theta+1}}\right]^{\delta_i} \left[p + (1-p)\left[\frac{\Gamma(\alpha) - \gamma(\alpha, \beta x)}{\Gamma(\alpha) - (1-\theta)\gamma(\alpha, \beta x)}\right]^\theta\right]$$

Taking log on both sides gives

$$\begin{aligned} l &= \log L = \log(1-P) \sum \delta_i \\ &+ \log\theta \sum^2 \delta_i \\ &+ \delta_i(\alpha-1) \sum \log x + \log\Gamma(\alpha) \sum^{\delta_i} + \log[\Gamma(\alpha) - \gamma(\alpha, \beta x)] \sum^{(\theta+1)} \delta_i \end{aligned}$$

Equating the partial derivative of l with respect to each parameter to zero and solving the equation simultaneously will give the maximum likelihood estimates. This will be done using a computer software.

Table 1.3: Simulation Result for Parameter Estimation for T1HT-G using sample size of n=N T=50 with 20% Censoring

Sample size	Parameter	AE	SE	MSE
50	θ	2.384253	0.16006	0.025619
	α	0.551016	0.004402	1.94E-05
	β	0.774811	0.121439	0.014747
100	θ	1.85286	0.099092	0.009819

	α	0.54053	0.00318	1.01E-05
	β	0.276831	22.91346	525.0267
200	θ	1.450402	0.09772	0.009549
	α	0.528271	0.002295	5.26553E-06
	β	0.180219	0.039718	0.001578
500	θ	0.841998	0.038391	0.001474
	α	0.524191	0.001592	2.54E-06
	β	0.165629	0.035725	0.001276

Note: T1HT-G:θ=0.2, β=2.0 and α=0.5

Table 1.4: Simulation Result for Parameter Estimation for T1HT-G using sample size of n=N T=50 with 60% Censoring

Sample size	Parameter	AE	SE	MSE
50	θ	6.229661	2.015881	4.063778
	α	0.634267	0.008224	0.0000067
	β	0.353889	0.135219	0.1828497
100	θ	3.429987	1.137128	1.293061
	α	0.59887	0.005301	0.000062
	β	0.18585	0.079973	0.063958
200	θ	2.711301	1.011161	1.022447
	α	0.574444	0.00375	0.000014
	β	0.273157	0.001794	0.032325
500	θ	1.446203	0.065941	0.004348
	α	0.562754	0.00261	0.000006
	β	0.599708	0.004987	2487.695

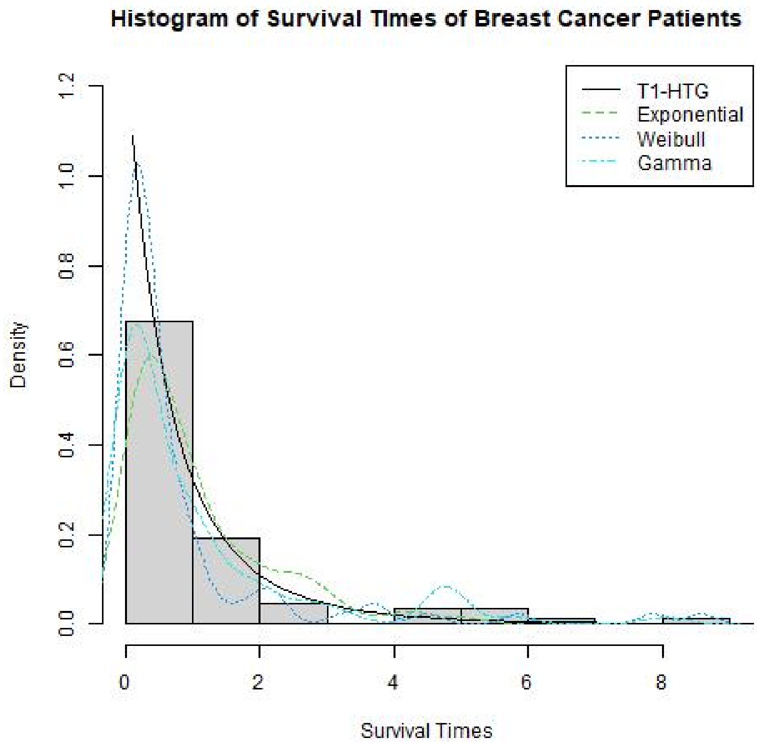
Note: (T1HT-G:θ=1.5,β=2.5 and α=0.5)

Tables 1.3 and 1.4 provide an insightful overview of the simulation results, demonstrating the strong performance of model parameter estimations. These estimations exhibit minimal bias and reasonable Mean Square Errors (MSEs), pointing to their reliability and precision, especially as sample size increases. This suggests that, with a growing sample size, the estimates become more accurate and dependable, aligning with the concept of asymptotically unbiased estimators, where biases approach zero as the sample size expands.

Application of the T1HT-G Mixture Cure Rate to the Breast Cancer Data Set

The data used for this analysis was sourced from the publication by Ilori and Awodutire (2017). It's worth noting that this data originated from the breast cancer case records of patients at the Ladoke Akintola University of Technology's Teaching Hospital in Osogbo. Survival time was calculated as the difference between the last day patients were seen and their initial treatment reporting date. Additionally, it's important to acknowledge that data access will be restricted after one year.

This analysis considered several prognostic factors, including years of nursing experience (in years), age at menarche, body mass index, stage at presentation, the availability of neoadjuvant treatment, and the use of contraception, all of which played a role in assessing and understanding breast cancer-related variables.



Figures 4.3: Histogram Plot of survial time of Breast Cancer Data

Figure 4.3 shows that the histogram of the survival times of breast cancer patients. It reveals that the data is rightly skewed which from (Ilori & Awodutire, 2017). had shown that the Type 1 Heavy-Tailed Gamma distribution can give a good fit. Figure 4.17 shows that the distribution fits the survival data well.

Table 1.5: T1HT-G Mixture Cure Rate results for Breast Cancer Data

Model	Parameter	Estimates	Loglik	AIC	BIC	CAIC
T1HT-G	θ	1.47	-21.80	51.60	61.55	52.08
	α	0.62				
	β	0.51				
Gamma	α	0.27	-43.97	93.95	101.42	94.23
	β	0.41				
Weibull	α	2.13	-39.65	85.31	92.78	85.60
	β	1.11				
	γ	1.69				
Exponential	α	0.33	-44.29	92.59	97.57	92.73
	β	0.241				

Table 1.5 revealed the result of the analysis. We can deduce that the new model performed better than the usual classical distribution as the CAIC, BIC, and AIC values are small.

Conclusion

In this study, we have undertaken the development and analysis of the T1HT-G distribution. Our investigation encompassed the derivation of key statistical properties such as the survival function, hazard rate function, probability distribution function, and R^{th} moment for these models. Furthermore, we employed the maximum likelihood approach to estimate the parameters of these distributions and models.

To evaluate the performance and reliability of these estimators, we employed a range of statistical criteria, including the Mean Square Error (MSE), Bayesian Information Criterion (BIC), Corrected Akaike Information Criterion (CAIC), and Akaike Information Criterion (AIC). The findings provided a comprehensive assessment of the accuracy and fit of our models to the data.

Notably, the consistency of the Maximum Likelihood Estimates (MLE) was demonstrated through a decrease in MSE as the sample size increased, further validating the robustness of our parameter estimates.

Future Works

Future work includes

1. A bivariate and multivariate extension of the T1HT-Gamma should be studied

2. Regression problems with covariates should also be considered for the model

Acknowledgement

The Author wishes to thank Tertiary Education Trust Fund (TETFund) Nigeria for funding this research through Institution Based Research (IBR) intervention

Conflict of Interest

There is no conflict of interest

References

1. Zhao W, Khosa SK, Ahmad Z, Aslam M, Afify AZ (2020) Type-I heavy tailed family with applications in medicine, engineering and insurance. PLoS ONE 15: e0237462.

2. Alzaatreh A, Lee C, Famoye F (2013) A new method for generating families of continuous distributions. Metron 71: 63-79.

3. Beirlant J, Matthys G, Dierckx G (2001) Heavy-Tailed Distributions and Rating. ASTIN Bulletin, 31: 37-58.

4. Aljarrah et al. (2014) On generating T-X family of distributions using quantile functions. Journal of Statistical Distributions and Applications, 1:2.

5. Lee C, Famoye F, Alzaatreh A (2013) Methods for generating families of univariate continuous distributions in the recent decades. WIREs Computational Statistics 5, 219–38.

6. Rameshwar D (2001) Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions. Biometrical Journal, 43: 117-30.

7. Saralees N (2008) Some gamma distributions, Statistics: A Journal of Theoretical and Applied Schober, Patrick & Vetter, Thomas. (2018). Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. Anesthesia & Analgesia. 127. 1. 10.1213/ANE.0000000000003653. Statistics, 42: 77-94.

8. Elbatal I, Merovci F, Elgarhy M (2013) A new generalized Lindley distribution, Mathematical Theory and Modeling, 3: 30-47.

9. Maller RA, Zhou X (1996) Survival analysis with long-term survivors. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd.