Revista Colombiana de Estadística - Applied Statistics
July 2024, volume 47, issue 2, pp. 355–383
http://doi.org/10.15446/rce.v47n2.112929

Cancer Data Modelling: Application of the Gamma-Odd Topp-Leone-G Family of Distributions

Modelado de datos sobre el cáncer: aplicación de la familia de distribuciones GammaOdd Topp-Leone-G

Chipo Zidana[1,a], Whatmore Sengweni[2,b], Broderick O. Oluyede[1,c], Fastel Chipepa[1,d]

[1]Department of Mathematics and Statistical Sciences, Faculty of Science, Botswana International University of Science and Technology, Palapye, Botswana

[2]Department of Applied Mathematics and Statistics, Faculty of Science, Midlands State University, Gweru, Zimbabwe

Abstract

The study introduces a new generalised family of distributions for cancer data modelling using a generalisation of the gamma function and a Topp-Leone-G distribution called the Gamma-Odd Topp-Leone-G (GOTL-G). Cancer data is normally characterised by complex heterogeneous properties like skewness, kurtosis, and presence of extreme values which makes it difficult to model using classical distributions. We derived multiple statistical properties including the linear representation, Reńyi entropy, quantile functions, distribution of order statistics, and maximum likelihood estimates which normally guarantees a positive effect on the generalisability of cancer data. Interestingly, we observed that these derived statistical properties make it possible for the generalisation of different models which are useful in the analysis, control, insurance, and survival of cancer patients. Our results show that this new family of distributions can be applied to a variety of data sets such as bladder and breast cancer data which exhibited high level of skewness and kurtosis as well as symmetric attributes. Therefore, we can conclude that the GOTL-G family of distributions can be extremely useful in capturing distinct complex heterogeneous properties normally exhibited by cancer patients. We recommend that this new family of distributions can be useful in modelling complex real-life applications including cancer data.

***Key words*:** Exponentiated general distribution; Gamma function; Maximum likelihood estimation; Topp-Leone; Cancer modelling.

[a]Ph.D. E-mail: zidanac@biust.ac.bw
[b]Ph.D. E-mail: sengweniw@staff.msu.ac.zw
[c]Ph.D. E-mail: oluyedeb@biust.ac.bw
[d]Ph.D. E-mail: chipepaf@biust.ac.bw

Resumen

El estudio presenta una nueva familia generalizada de distribuciones para el modelado de datos sobre cáncer utilizando una generalización de la función gamma y una distribución Topp-Leone-G llamada Gamma-Odd Topp-Leone-G (GOTL-G). Los datos sobre el cáncer normalmente se caracterizan por propiedades heterogéneas complejas como asimetría, curtosis y presencia de valores extremos, lo que dificulta el modelado utilizando distribuciones clásicas. Derivamos múltiples propiedades estadísticas, incluida la representación lineal, la entropía de Re'nyi, funciones cuantiles, estadísticas de distribución de orden y estimaciones de máxima verosimilitud, que normalmente garantizan un efecto positivo en la generalización de los datos sobre el cáncer. Curiosamente, observamos que estas propiedades estadísticas derivadas permiten la generalización de diferentes modelos que son útiles en el análisis, control, seguro y supervivencia de pacientes con cáncer. Nuestros resultados muestran que esta nueva familia de distribuciones se puede aplicar a una variedad de conjuntos de datos, como datos de cáncer de vejiga y de mama, que mostraron un alto nivel de asimetría y curtosis, así como atributos simétricos. Por lo tanto, podemos concluir que la familia de distribuciones GOTL-G puede ser extremadamente útil para capturar distintas propiedades heterogéneas complejas que normalmente exhiben los pacientes con cáncer. Recomendamos que esta nueva familia de distribuciones pueda resultar útil para modelar aplicaciones complejas de la vida real, incluidos datos sobre cáncer.

***Palabras clave*:** Distribución general exponenciada; Función gamma; Estimación de máxima verosimilitud; Topp-Leone; Modelado de cáncer.

1. Introduction

Cancer modelling and survival analysis requires robust tools that captures and incorporates many statistical and structural aspects for cancer attributes with a high degree of accuracy and precision. In many cases cancer data is often characterised by unpleasant data properties including sparsity, skeweness, kurtosis, unbalanced labels among others which makes its generalisation ability difficult especially for gene expression data as alluded by Qi et al. (2019). It is paramount therefore, that more flexible mathematical tools including probability distributions for such complex scenarios are developed to comprehend the ever-developing sophisticated phenomena in cancer modelling. Several mathematical frameworks including dynamical systems have been used to model different cancer dynamics for different cancer types including an overview of cancer survival analysis by Pawar et al. (2022), metastatic analysis of cancer spread Franssen et al. (2019), and breast cancer immune dynamics Mufudza et al. (2012). The complexity of the disease has however seen limited distributions been developed to models its dynamics.

Although we have seen a surge in development of different probability distributions for many diseases including the recent COVID-19 by using Odd-Topp–Leone–Gompertz-G distribution by Musekwa & Makubate (2023), a lot more still needs to be done to improve the ever changing life complexities including cancer. The flexibility of both the gamma and beta functions have been very helpful in the inventions and improvements towards achieving mathematical tools and probability density function that incorporate variety of unique statistical properties. Among critical distributions in the literature are the beta-G family of distributions by Eugene & Famoye (2002), the Kumaraswamy-G family of distributions by Cordeiro & de Castro (2011) and the generalised odd Burr III family by Haq et al. (2019). These have been applied to various areas including engineering and science. Besides the beta function, the gamma function is also one of the most useful function due to its flexibility.

Thus, we have seen a range of gamma derived family of distributions with a variety of properties. Some notable gamma derived distributions include the odd gamma-G type 1 by Hosseini et al. (2018), Gamma Weibull-G by Oluyede et al. (2018), Gamma Uniform distribution by Torabi & Hedesh (2012) and the Gamma Odd Burr III-G by Peter et al. (2021) among others. Arshad et al. (2020) recently introduced the Gamma Kumaraswamy-G family of distributions which exhibited symmetric, left and right skewness, and bathtub shape on the hazard function just as some of the flexible features needed to model complex situations. Topp-Leone family of distributions by Al-Shomrani et al. (2016) is another distribution which has become a powerful tool in capturing unique complex attributes in real life datasets. A number of generations using the Topp-Leone family of distributions with other distributions have also been introduced including: Topp-Leone Exponential-G by Sanusi et al. (2020), Topp-Leone-Marshall-Olkin-G by Chipepa et al. (2020), and Topp-Leone Weibull by Tuoyo et al. (2021) among others.

We therefore, believe that with the complexities involved in cancer data combining the unique properties from both the gamma function and Topp-Leone distribution maybe be able to bring out more flexibility to capture the complex heterogeneous characteristics normally exhibited in various cancer types.

In this study, we derive a new family of distribution by incorporating the odd function from the Topp-Leone-G (TLG) by Al-Shomrani et al. (2016) into the gamma transformation as derived by Torabi & Hedesh (2012). The new developed family of distributions is then applied to bladder cancer and breast cancer data. The TL-G distribution has the probability density function (pdf) and cumulative density function (cdf) given by Equations (1) and (2) as

$$h(x; b, \boldsymbol{\xi}) = 2bg(x; \boldsymbol{\xi})\bar{G}(x; \boldsymbol{\xi}) \left[1 - \bar{G}^2(x; \boldsymbol{\xi})\right]^{b-1}, \tag{1}$$

and

$$H(x; b, \boldsymbol{\xi}) = \left[1 - \bar{G}^2(x; \boldsymbol{\xi})\right]^{b}, \tag{2}$$

for $x,\ b > 0,\ \bar{G}(x, \boldsymbol{\xi})) = (1 - G(x, \boldsymbol{\xi}))$ where $g(x, \boldsymbol{\xi})$, $G(x, \boldsymbol{\xi})$ represents the pdf and cdf of the baseline distribution, respectively. We let $h(x) = h(x; b, \boldsymbol{\xi})$ and $H(x) = H(x; b, \boldsymbol{\xi})$.

Suppose the gamma function $\gamma_1(\delta, z)$ is the regularised lower incomplete gamma function defined by $\gamma_1(\delta, z) = \gamma(\delta, z)/\Gamma(\delta)$, where $\gamma(\delta, z) = \int_0^z t^{\delta-1}e^{-t}dt$ and $\Gamma(\delta) = \int_0^\infty t^{\delta-1}e^{-t}dt$. The gamma-generated distribution as generalised by Torabi

& Hedesh (2012) has a continuous cdf and pdf given by Equation (3) and (4), respectively as

$$\begin{aligned} F(x) &= \frac{1}{\Gamma(\alpha)} \int_0^{G(x)/\bar{G}(x)} \omega^{\alpha-1} e^{-\omega} d\omega \\ &= \gamma_1\left(\alpha, G(x)/\bar{G}(x)\right), \end{aligned} \tag{3}$$

and

$$f(x) = \frac{1}{\Gamma(\alpha)} \frac{g(x) G^{\alpha-1}(x)}{\bar{G}^{\alpha+1}(x)} \exp\left(-\frac{G(x)}{\bar{G}(x)}\right), \quad \alpha > 0, \tag{4}$$

where $\gamma_1(\alpha, x)$, $g(x), G(x)$, is an incomplete gamma function, baseline pdf and cdf, respectively. If $G(x) = H(x)$ such that the baseline cdf is a TL-G family as given by Equation (2), then the cdf for the new family of distributions called the Gamma Odd Topp-Leone-G (GOTL-G) family of distributions is defined as follows:

$$F_{GOTL-G}(x) = \gamma_1\left(\delta, \frac{H(x)}{\bar{H}(x)}\right), \quad x \in R,$$

and the corresponding pdf for the GOTL-G family is given by

$$f_{GOTL-G}(x) = \frac{1}{\Gamma(\delta)} \frac{h(x)(H(x))^{\delta-1}}{(\bar{H}(x))^{\delta+1}} \exp\left(\frac{H(x)}{\bar{H}(x)}\right),$$

where $H(x)$ is the TL-G cdf.

We are going to therefore, focus on the proposed new GOTL-G family of distributions, its properties and usefulness via applications. Section 2 dwells on derivation of the pdf and cdf of the new distribution. The general theoretical and statistical characteristics are explored in Section 3. Special members of the GOTL-G family with varying baseline distributions including the Weibull, Burr III, and Uniform are discussed in Section 4. A simulation study for the new family of distributions is presented in Section 5. Usefulness of the distribution and applications from a breast cancer tumour texture and bladder cancer patients datasets (see Wolberg et al., 1995 and Lee & Wang, 2003 as given in Section 6).

2. Gamma Odd Topp-Leone-G Distribution

Substituting the odd Topp-Leone-G defined by

$$\frac{H(x)}{\bar{H}(x)} = \frac{[1 - \bar{G}^2(x; \boldsymbol{\xi})]^b}{1 - [1 - \bar{G}^2(x; \boldsymbol{\xi})]^b} = \left([1 - \bar{G}^2(x; \boldsymbol{\xi})]^{-b} - 1\right)^{-1}, \quad b > 0,$$

into the gamma function, results in the new GOTL-G family of distributions with a cdf given by

$$\begin{aligned} F_{GOTL-G}(x) &= \frac{1}{\Gamma(\delta)}\int_0^{\frac{H(x)}{\bar{H}(x)}} t^{\delta-1}e^{-t}dt \\ &= \frac{1}{\Gamma(\delta)}\int_0^{\left([1-\bar{G}^2(x;\boldsymbol{\xi})]^{-b}-1\right)^{-1}} t^{\delta-1}e^{-t}dt \\ &= \frac{\gamma\left(\delta,\left(\left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{-b}-1\right)^{-1}\right)}{\Gamma(\delta)}, \end{aligned} \tag{5}$$

where $\delta, b\ > 0$ and $\boldsymbol{\xi}$ is a vector of parameters from the baseline distribution G(.). The corresponding GOTL-G pdf is given by:

$$f_{GOTL-G}(x) = \frac{2bg(x;\boldsymbol{\xi})\bar{G}(x;\boldsymbol{\xi})\left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{\delta b-1}\exp\left\{-\frac{[1-\bar{G}^2(x;\boldsymbol{\xi})]^b}{1-[1-\bar{G}^2(x;\boldsymbol{\xi})]^b}\right\}}{\Gamma(\delta)\left[1-[1-\bar{G}^2(x;\boldsymbol{\xi})]^b\right]^{\delta+1}}, \tag{6}$$

where $G(x;\boldsymbol{\xi})$ is the baseline cdf and $b,\delta \geq 0$. The reliability function for the GOTL-G family of distributions is given by

$$\begin{aligned} S_{GOTL-G}(x) &= \bar{F}_{GOTL-G}(x;b,\delta,\boldsymbol{\xi}) \\ &= 1-\frac{\gamma\left(\delta,\left(\left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{-b}-1\right)^{-1}\right)}{\Gamma(\delta)}. \end{aligned} \tag{7}$$

Consequently, the hazard rate function (hrf) for the GOTL-G family of distributions is represented by

$$\begin{aligned} h_{GOTL-G}(x) &= \frac{f_{GOTL-G}(x)}{S_{GOTL-G}(x)} \\ &= \frac{2bg(x;\boldsymbol{\xi})\bar{G}(x;\boldsymbol{\xi})\left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{\delta b-1}\exp\left\{-\frac{[1-\bar{G}^2(x;\boldsymbol{\xi})]^b}{1-[1-\bar{G}^2(x;\boldsymbol{\xi})]^b}\right\}}{\left[1-[1-\bar{G}^2(x;\boldsymbol{\xi})]^b\right]^{\delta+1}\left[\Gamma(\delta)-\gamma\left(\delta,(\left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{-b}-1)^{-1}\right)\right]}, \end{aligned} \tag{8}$$

for $b,\delta > 0$ and parameter vector $\boldsymbol{\xi}$.

2.1. Linear Representation

In this section, we explore the expansion of the cdf and pdf of the GOTL-G family to an infinite linear representation of exponentiated-G (Exp-G) distributions. This generalisation is helpful in exploring the flexibility and simplicity that may help in interpreting theory of complexity in most cancer experimental data (Grizzi & Chiriva-Internati, 2006)

Theorem 1. *The pdf of GOTL-G family of distributions can be expressed as the following linear representation*

$$f_{GOTL-G}(x) = \sum_{q=0}^{\infty}\eta_{q+1}g^*_{q+1}(x;\boldsymbol{\xi}), \tag{9}$$

where

$$\eta_{q+1} = \sum_{k,l,p=0}^{\infty} \frac{(-1)^{k+p+q}\Gamma(\delta+k+l+1)}{\Gamma(\delta+k+l)(q+1)k!l!} \binom{b(\delta+k+l)-1}{p} \binom{2p+1}{q} \tag{10}$$

and $g^*_{q+1}(x;\boldsymbol{\xi}) = (q+1)g(x;\boldsymbol{\xi})G^q(x;\boldsymbol{\xi})$ *is an Exp-G density with power parameter* $q+1$.

Proof . Using the following generalised series expansions as given in Gradshteyn & Ryzhik (2000) we have: $(1-z)^{-\alpha} = \sum_{q=0}^{\infty} \frac{\Gamma(\alpha+q)}{\Gamma(\alpha)q!} z^q$ and $(1-z)^{\alpha} = \sum_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} z^k$, for $|z| < 1$, and $e^y = \sum_{p=0}^{\infty} \frac{y^p}{p!}$, the pdf of the GOTL-G family of distributions can be written as follows:

$$\begin{aligned}
f_{GOTL-G}(x) &= \frac{2b}{\Gamma(\delta)} g(x;\boldsymbol{\xi})\bar{G}(x;\boldsymbol{\xi}) \\
&\times \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{b(\delta+k)-1} \left[1-[1-\bar{G}^2(x;\boldsymbol{\xi})]^b\right]^{-(\delta+k+1)} \\
&= \frac{2b}{\Gamma(\delta)} g(x;\boldsymbol{\xi})\bar{G}(x;\boldsymbol{\xi}) \sum_{k,l=0}^{\infty} \frac{(-1)^k\Gamma(\delta+k+l+1)}{\Gamma(\delta+k+l)k!l!} \\
&\times \left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{b(\delta+k+l)-1} \\
&= \frac{2b}{\Gamma(\delta)} g(x;\boldsymbol{\xi}) \sum_{k,l,p=0}^{\infty} \frac{(-1)^k\Gamma(\delta+k+l+1)}{\Gamma(\delta+k+l)k!l!} \binom{b(\delta+k+l)-1}{p} \\
&\times \bar{G}^{2p+1}(x;\boldsymbol{\xi}) \\
&= \frac{2b}{\Gamma(\delta)} g(x;\boldsymbol{\xi}) \sum_{k,l,p=0}^{\infty} \frac{(-1)^k\Gamma(\delta+k+l+1)}{\Gamma(\delta+k+l)k!l!} \binom{b(\delta+k+l)-1}{p} \\
&\times \binom{2p+1}{q} G^q(x;\boldsymbol{\xi}) \\
&= \sum_{q=0}^{\infty} \eta_{q+1} g^*_{q+1}(x;\boldsymbol{\xi}).
\end{aligned}$$

□

Therefore, the pdf of the GOTL-G family of distributions can be expressed as an infinite sum of the Exp-G family of distribution with power parameter $q+1$. Consequently, the mathematical and statistical properties of the GOTL-G family of distributions follow directly from those of the Exponentiated-G family of distributions.

3. Some Statistical Properties

Some properties of the GOTL-G family of distributions including Reńyi entropy, moments, quantile function, distribution of order statistics and maximum likelihood estimates of model parameters are presented in this section.

3.1. Reńyi Entropy

In this section, we present Reńyi entropy for the GOTL-G family of distributions. Reńyi entropy can be very useful in many ways including quantifying the randomness, diversity and uncertainty for any given system as applied to ecology and dynamical systems. The Reńyi entropy is defined by the formula:

$$I_R(\epsilon) = (1-\epsilon)^{-1} \log \left[\int_0^\infty (f(x))^\epsilon \, dx, \right] \;\; \epsilon \neq 1, \; \epsilon > 0. \tag{11}$$

It should be noted that Reńyi entropy becomes Shannon entropy as $\epsilon \to 1$. Now expanding the power of $f_{GOTL-G}(x;\boldsymbol{\xi})$ using some generalised binomial expansions, we have:

$$\begin{aligned}
f^\epsilon(x) &= \frac{\left(\frac{2b}{\Gamma(\delta)}\right)^\epsilon g^\epsilon(x;\boldsymbol{\xi}) \bar{G}^\epsilon(x;\boldsymbol{\xi}) \left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{b\delta\epsilon-\epsilon} \exp\left\{\frac{-\epsilon(1-\bar{G}^2(x;\boldsymbol{\xi}))^b}{1-[1-\bar{G}^2(x;\boldsymbol{\xi})]^b}\right\}}{\left(1-(1-\bar{G}(x;\boldsymbol{\xi})^2)^b\right)^{-\epsilon(\delta+1)}} \\
&= \left(\frac{2b}{\Gamma(\delta)}\right)^\epsilon g^\epsilon(x;\boldsymbol{\xi}) \bar{G}^\epsilon(x;\boldsymbol{\xi}) \sum_{k=0}^{\infty} \frac{(-1)^k \epsilon^k \left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{bk+\delta b\epsilon-\epsilon}}{k! \left(1-(1-\bar{G}^2(x;\boldsymbol{\xi}))^b\right)^{k+\epsilon(\delta+1)}} \\
&= \left(\frac{2b}{\Gamma(\delta)}\right)^\epsilon g^\epsilon(x) \bar{G}^\epsilon(x;\boldsymbol{\xi}) \sum_{k=0}^{\infty} \frac{(-1)^k \epsilon^k}{k!} \left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^{b(k+\delta\epsilon)-\epsilon} \\
&\times \left(1-\left[1-\bar{G}^2(x;\boldsymbol{\xi})\right]^b\right)^{-(k+\epsilon(\delta+1))} \\
&= \left(\frac{2b}{\Gamma(\delta)}\right)^\epsilon g^\epsilon(x;\boldsymbol{\xi}) \bar{G}^\epsilon(x;\boldsymbol{\xi}) \sum_{k,l=0}^{\infty} \frac{(-1)^k \epsilon^k}{k!} \left[\frac{\Gamma(k+\epsilon(\delta+1)+l)}{\Gamma(k+\epsilon(\delta+1))l!}\right] \\
&\times \left(1-\bar{G}^2(x;\boldsymbol{\xi})\right)^{b(k+\delta\epsilon)-\epsilon+bl} \\
&= \left(\frac{2b}{\Gamma(\delta)}\right)^\epsilon g^\epsilon(x;\boldsymbol{\xi}) \sum_{k,l,p=0}^{\infty} \frac{(-1)^{k+p} \epsilon^k}{k!} \left[\frac{\Gamma(k+\epsilon(\delta+1)+l)}{\Gamma(k+\epsilon(\delta+1))l!}\right] \\
&\times \binom{b(k+\delta\epsilon+l)-\epsilon}{p} \bar{G}^{2p+\epsilon}(x;\boldsymbol{\xi}) \\
&= \left(\frac{2b}{\Gamma(\delta)}\right)^\epsilon g^\epsilon(x;\boldsymbol{\xi}) \sum_{k,l,p=0}^{\infty} \frac{(-1)^{k+p} \epsilon^k}{k!} \left[\frac{\Gamma(k+\epsilon(\delta+1)+l)}{\Gamma(k+\epsilon(\delta+1))l!}\right] \\
&\times \binom{b(k+\delta\epsilon+l)-\epsilon}{p} \binom{2p+\epsilon}{q} G^q(x;\boldsymbol{\xi}).
\end{aligned} \tag{12}$$

Substituting (12) into (11), we have

$$I_R(\epsilon) = (1-\epsilon)^{-1} \log \left[\int_0^\infty \sum_{q=0}^\infty \Phi_{q+1} e^{(1-\epsilon) I_{REG}} \right], \tag{13}$$

where

$$\begin{aligned} \Phi_{q+1} &= \left(\frac{2b}{\Gamma(\delta)}\right)^\epsilon \sum_{k,l,p=0}^\infty \frac{(-1)^{k+p}\epsilon^k}{k!} \left[\frac{\Gamma(k+\epsilon(\delta+1)+l)}{\Gamma(k+\epsilon(\delta+1))l!}\right] \\ &\times \binom{b(k+\delta\epsilon+l)-\epsilon}{p}\binom{2p+\epsilon}{q}\left(\frac{q}{\epsilon}+1\right)^{-1} \end{aligned} \tag{14}$$

and $I_{REG} = \int_0^\infty \left[\left(\frac{q}{\epsilon}+1\right) g(x,\boldsymbol{\xi}) G(x;\boldsymbol{\xi})^{q/\epsilon}\right]^\epsilon dx$ is the Rényi entropy of the Exp-G distribution with power parameter $(q/\epsilon)+1$.

3.2. Moments and Quantile Function

Given the fact that from Equation (9), the GOTL-G density function is actually a weighted sum of the exponentiated-G distribution, it follows that the moments of GOTL-G family of distributions can be obtained from the exponentiated-G family of distributions. If Y_{q+1}, denote the Exp-G random variable with power parameter $q+1$. Then the r^{th} raw moment of X, say μ'_r follows from Equation (9) as

$$\mu'_r = E(X^r) = \sum_{q=0}^\infty \eta_{q+1} E(Y_{q+1}^r), \tag{15}$$

where η_{q+1} is as given in Equation (10). The moment generating function (mgf) $M_X(t) = E(e^{tX})$ of X can be derived from Equation (9) and is given by

$$M_X(t) = \sum_{q=0}^\infty \eta_{q+1} M_q(t),$$

where $M_q(t)$ is the mgf of Y_{q+1}. Hence, $M_X(t)$ can be determined from the Exp-G generating function. Beside the moments, another important function for GOTL-G family is the quantile function. This function can be used to generate random data for GOTL-G family of distributions. The quantile function is obtained by solving the nonlinear equation:

$$F_{GOTL-G}(Q(u)) = \frac{\gamma\left(\delta, \left(1-\left[1-\bar{G}^2(Q(u))\right]^{-b}-1\right)^{-1}\right)}{\Gamma(\delta)} = u, \tag{16}$$

for $0 \leq u \leq 1$. That is,

$$\gamma^{-1}(\delta, u\Gamma(\delta)) = \frac{[1-\bar{G}^2(Q(u))]^b}{1-[1-\bar{G}^2(Q(u))]^b},$$

and

$$G(Q(u)) \quad = \quad 1-\left\{1-\left[\frac{\gamma^{-1}(\delta, u\Gamma(\delta))}{1+\gamma^{-1}(\delta, u\Gamma(\delta))}\right]^{\frac{1}{b}}\right\}^{\frac{1}{2}}.$$

Consequently, the quantile function can be written as

$$Q_{GOTL-G}(u) = G^{-1}\left(1-\left\{1-\left[\frac{\gamma^{-1}(\delta, u\Gamma(\delta))}{1+\gamma^{-1}(\delta, u\Gamma(\delta))}\right]^{\frac{1}{b}}\right\}^{\frac{1}{2}}\right). \tag{17}$$

Based on Equation (17), it can be seen that we can use $Q_{GOTL-G}(u)$ to generate random variables having a GOTL-G family of distributions cdf as the distribution function. Therefore, a quantile function can serve as random sampling generator from the GOTL-G family of distributions. Moreover, other important measure of spread which are less affected by outliers or skewness like the Interquartile range, can be derived using the quantile function.

3.3. Order Statistics

This section will involve the determination of the distribution of the order statistic from the GOTL-G family of distributions. Order statistics is naturally helpful in survival analysis, reliability, and applications. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the GOTL-G family of distributions and $X_{i:n}$ be the ith order statistic, then the pdf $f_{i:n}(x)$ of $X_{i:n}$ from the GOTL-G family of distributions is given by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} f_{GOTL-G}(x)[F_{GOTL-G}(x)]^{i-1}[1-F_{GOTL-G}(x)]^{n-i}. \tag{18}$$

Using Equation (5) and (6) the pdf of the ith order statistic will be given by

$$\begin{aligned} f_{i:n}(x) &= \sum_{m=0}^{n-i} \frac{n!(m+i)(-1)^m}{(i-1)!(n-i)!(m+i)} \binom{n-i}{m} [F(x)]^{m+i-1} f(x) \\ &= \sum_{m=0}^{n-i} \Psi_{(i,m)} f_{m+i}(x; b, \delta, \boldsymbol{\xi}), \end{aligned} \tag{19}$$

where $f_{m+i}(x; b, \delta, \boldsymbol{\xi})$ is the GOTL-G pdf with exponentiated parameter $m+i > 0$ and weights $\Psi_{(i,m)}$ given by

$$\begin{aligned} \Psi_{(i,m)} &= \frac{1}{B(i, n-i+1)} \frac{(-1)^m}{m+i} \binom{n-i}{m} \\ &= (-1)^m \binom{n}{m+i} \binom{m+i-1}{m}. \end{aligned} \tag{20}$$

Thus, the pdf of ith order statistic, $X_{i:n}$ can be expressed as a linear combination of pdfs of the exponentiated-G family of distributions.

4. Some Special Cases of the GOTL-G Distribution

This section explores a variety of the GOTL-G special cases by incorporating a range of baseline distributions. Characteristics of these special cases are also analysed using distribution attributes like skewness and kurtosis for the special distributions.

4.1. Gamma Odd Topp-Leone-Weibull Distribution

Suppose, we incorporate Weibull distribution as a baseline distribution that is, $G(x;\lambda) = 1 - e^{-x^{\lambda}}$, for $x, \lambda > 0$. This results in a GOTL-Weibull (GOTL-W) distribution, which can be represented by

$$F_{GOTL-W}(x) = \frac{\gamma\left(\delta, \left([1 - e^{-2x^{\lambda}}]^{-b} - 1\right)^{-1}\right)}{\Gamma(\delta)}. \tag{21}$$

The corresponding pdf of the GOTL-W distribution is

$$f_{GOTL-W}(x) = \frac{2b\lambda x^{\lambda-1} e^{-2x^{\lambda}} \left[1 - e^{-2x^{\lambda}}\right]^{\delta b - 1} \exp\left\{-\frac{[1-e^{-2x^{\lambda}}]^{b}}{1-[1-e^{-2x^{\lambda}}]^{b}}\right\}}{\Gamma(\delta)\left[1 - (1 - e^{-2x^{\lambda}})^{b}\right]^{\delta+1}}, \tag{22}$$

and the hazard rate function is given by

$$h_{GOTL-W}(x) = \frac{2b\lambda x^{\lambda-1} e^{-2x^{\lambda}} \left[1 - e^{-2x^{\lambda}}\right]^{\delta b - 1} \exp\left\{-\frac{[1-e^{-2x^{\lambda}}]^{b}}{1-[1-e^{-2x^{\lambda}}]^{b}}\right\}}{\left[1 - (1 - e^{-2x^{\lambda}})^{b}\right]^{\delta+1}\left[\Gamma(\delta) - \gamma(\delta, [(1 - e^{-2x^{\lambda}})^{-b} - 1]^{-1})\right]}. \tag{23}$$

4.2. Gamma Odd Topp-Leone-Burr III Distribution

Suppose, we incorporate Burr III distribution as a baseline distribution that is, $G(x;\alpha,\beta) = (1 + x^{-\alpha})^{-\beta}$, then we can obtain the cdf of the GOTL-Burr III distribution, as represented by

$$F_{GOTL-BIII}(x) = \frac{\gamma\left(\delta, \left([1 - [1 - (1 + x^{-\alpha})^{-\beta}]^{2}]^{-b} - 1\right)^{-1}\right)}{\Gamma(\delta)}. \tag{24}$$

The corresponding pdf of the GOTL-Burr III distribution is:

$$f_{GOTL-BIII}(x) = \left\{ \frac{2b\alpha\beta x^{-(\alpha+1)}(1+x^{-\alpha})^{-(\beta+1)}(1-(1+x^{-\alpha})^{-\beta})}{\Gamma(\delta)\left[1-(1-[1-(1+x^{-\alpha})^{-\beta}]^2)^b\right]^{\delta+1}} \right\}$$
$$\times \frac{\left[[1-[1-(1+x^{-\alpha})^{-\beta}]^2]\right]^{\delta b-1}}{\Gamma(\delta)\left[1-(1-[1-(1+x^{-\alpha})^{-\beta}]^2)^b\right]^{\delta+1}}$$
$$\times \exp\left\{-\frac{[1-[1-(1+x^{-\alpha})^{-\beta}]^2]^b}{1-[[1-[1-(1+x^{-\alpha})^{-\beta}]^2]]^b}\right\}, \tag{25}$$

whilst the hazard rate function will be given by

$$h_{GOTL_{BIII}}(x) = \frac{2b\alpha\beta x^{-(\alpha+1)}(1+x^{-\alpha})^{-(\beta+1)}(1-(1+x^{-\alpha})^{-\beta})}{[1-([1-[1-(1+x^{-\alpha})^{-\beta}]^2])^b]^{\delta+1}}$$
$$\times \frac{\left[[1-[1-(1+x^{-\alpha})^{-\beta}]^2]\right]^{\delta b-1}}{\Gamma(\delta)-\gamma\left(\delta,([1-[1-(1+x^{-\alpha})^{-\beta}]^2]^{-b}-1)^{-1}\right)}$$
$$\times \exp\left\{-\frac{[1-[1-(1+x^{-\alpha})^{-\beta}]^2]^b}{1-[[1-[1-(1+x^{-\alpha})^{-\beta}]^2]]^b}\right\}. \tag{26}$$

4.3. Gamma Odd Topp-Leone-Uniform Distribution

In this sub-section, we introduce a GOTL-Uniform (GOTL-U) distribution by using the uniform distribution as baseline distribution with cdf $G(x;\theta) = x/\theta,\ 0 \leq x \leq \theta$. The cdf of the GOTL-U distribution is given as

$$F_{GOTL-U}(x) = \frac{\gamma\left(\delta,\left([1-(1-x/\theta)^2]^{-b}-1\right)^{-1}\right)}{\Gamma(\delta)}, \tag{27}$$

whilst the pdf is given by

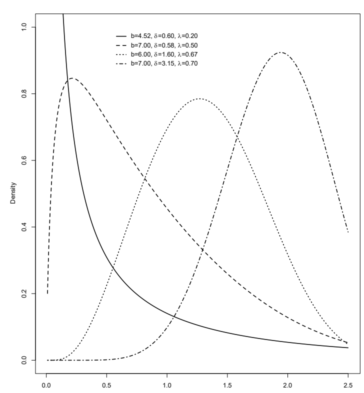
$$f_{GOTL-U}(x) = \frac{2b(1/\theta)(1-x/\theta)(1-(1-x/\theta)^2)^{\delta b-1}\exp\left\{-\frac{[1-(1-x/\theta)^2]^b}{1-[1-(1-(x/\theta))^2]^b}\right\}}{\Gamma(\delta)\left[1-(1-(1-x/\theta)^2)^b\right]^{\delta+1}}. \tag{28}$$

The hazard rate function is

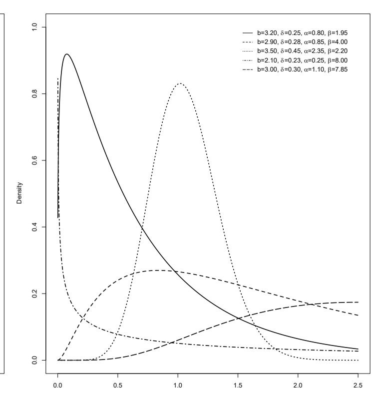
$$h_{GOTL-U}(x) = \frac{2b(1/\theta)(1-x/\theta)(1-(1-x/\theta)^2)^{\delta b-1}\exp\left\{-\frac{[1-(1-x/\theta)^2]^b}{1-[1-(1-(x/\theta))^2]^b}\right\}}{\left[1-(1-(1-x/\theta)^2)^b\right]^{\delta+1}\left(\Gamma(\delta)-\gamma(\delta,[(1-(1-x/\theta)^2)^{-b}-1]^{-1})\right)}. \tag{29}$$

4.4. Special Densities and Hazard Rate

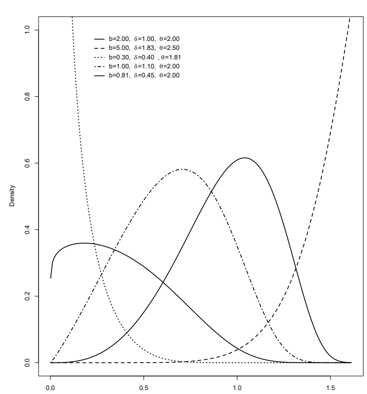
Figure 1(a-c) shows the probabilities density function for the GOTL-W, GOTL-Burr III and the GOTL-Uniform distribution, respectively for varying parameter values. They exhibits a variety of desirable properties normally desirable to capture heterogeneous characteristics in cancer modelling. The GOTL-W shows symmetrical, left and right skewness and decreasing L-shape as shown by the GOTL-W whilst the GOTL-Burr III distribution exhibits captures decreasing, almost symmetric, and skewed shapes and lastly the GOTL-U pdf captures almost symmetric, increasing, decreasing and both left and right-skewed shapes. The hazard rate function for the GOTL-W distribution shows flexibility of the distribution by capturing various shapes including bathtub, upside down U shape function, reverse-J shape, upside down bathtub the decreasing shape and a increasing hazard rate function among others as given by Figure 1(d) on the other hand the GOTL-Burr III distribution shows that indeed the model can be applied in various for data which exhibits J and reverse-J shape, increasing then decreasing, and increasing among others shapes as given by Figure 1(b). Figure 1(e) shows the hazard rate function from the GOTL-U which models a variety of shapes including bathtub, increasing, decreasing J and reverse-J shapes, an indication that the GOTL distribution captures most desirable properties in modelling lifetime distributions.



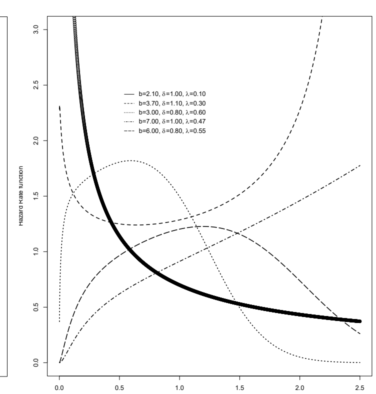
(a)



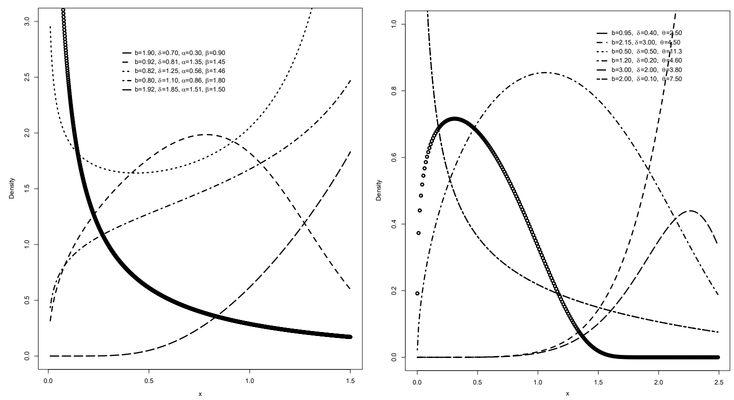
(b)



(c)



(d)



(e) (f)

FIGURE 1: GOTL densities and hazard rate functions a) GOTL-Weibull pdf b) GOTL-Burr III pdf c)GOTL-Uniform pdf d) GOTL-Weibull hazard rate e) GOTL-Burr III hazard rate f) GOTL-Uniform hazard rate

4.5. Moments and Quantiles for GOTL-Weibull Distribution

We explored the moments and quantiles for some selected parameter values of the GOTL-Weibull distribution. Selected moments for the GOTL-Weibull distribution are presented in Table 1. It is clear that for different parameter values, unique measures of central tendency, dispersion, and skewness can be explored using the GOTL-Weibull distribution.

TABLE 1: Moments for the GOTL-Weibull distribution

	Parameter Values				
Moment	(1.1, 0.5, 3.5)	(1.2, 0.9, 2.5)	(0.3, 0.8, 4.5)	(0.1, 1.2, 3)	(0.2, 0.4, 2)
E(X)	0.5317889	0.5901234	0.37786964	0.1045283128	4.261136e-028
$E(X^2)$	0.3276143	0.3869858	0.18427787	0.0235402804	8.159301e-03
$E(X^3)$	0.2196903	0.2712380	0.10115997	0.0068971758	2.233264e-03
$E(X^4)$	0.1559241	0.1994421	0.05970739	0.0023487821	7.394439e-04
$E(X^5)$	0.1153430	0.1521157	0.03707230	0.0008857265	2.774581e-04
SD	0.2116953	0.1968253	0.20369683	0.1123125649	7.964655e-02
CV	0.3980814	0.3335325	0.53906641	1.0744702742	1.869139e+00
CS	-0.2314194	-0.3743419	0.02005738	1.2701747999	2.662019e+00
CK	2.2789640	2.6373739	2.04249468	4.0855048186	1.087933e+01

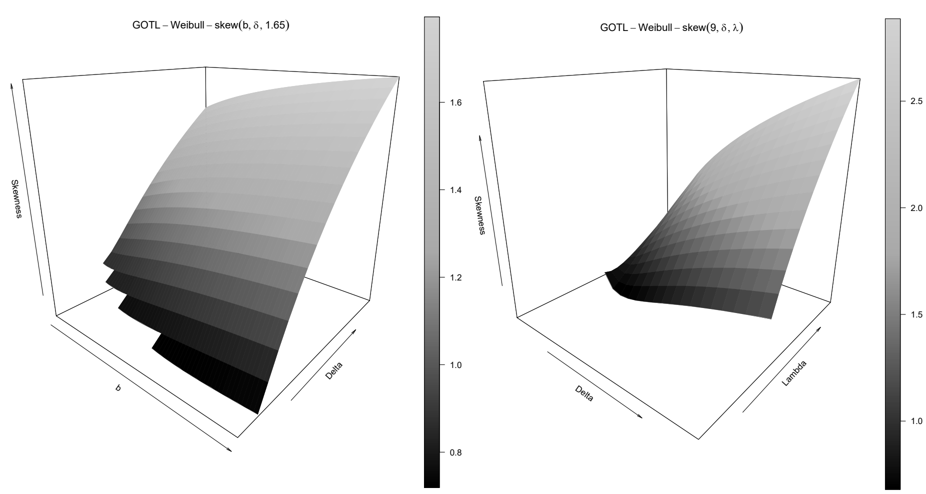
Table 2 shows the Quantile values explored using different parameter values of the GOTL-Weibull distribution. It is therefore, evident that for different parameter values the measure of spread like interquartile range changes. Thus the GOTL-Weibull distribution can capture data with a range of dispersion parameters and characteristics.

Table 2: Quantiles for the GOTL-Weibull distribution

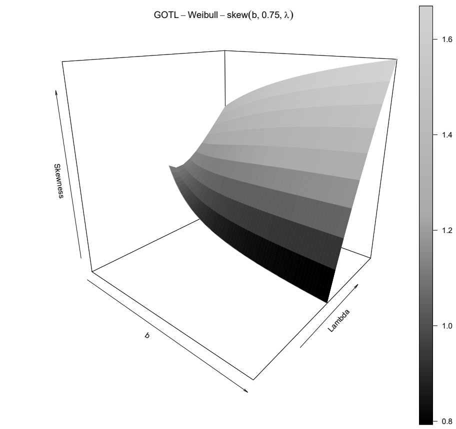
	Parameter Values				
Quantile	(1.1, 0.5, 3.5)	(1.2, 0.9, 2.5)	(0.3, 0.8, 4.5)	(0.1, 1.2, 3)	(0.2, 0.4, 5)
0.1	0.000000e+00	0.003898499	0.09362547	0.001305866	0.00202368
0.2	7.934469e-05	0.016859268	0.17409935	0.007492871	0.01149901
0.3	4.080079e-04	0.038659599	0.24780732	0.019821506	0.03150899
0.4	1.976592e-03	0.068569837	0.31648667	0.038619306	0.06401001
0.5	6.646832e-03	0.106006172	0.38150834	0.064236774	0.11029708
0.6	1.777705e-02	0.151048405	0.44440320	0.097300245	0.17117988
0.7	4.068183e-02	0.204852872	0.50708166	0.139400570	0.24755394
0.8	8.427332e-02	0.271039587	0.57296060	0.194354707	0.34185296
0.9	1.688382e-01	0.361666744	0.65053222	0.274043289	0.46381451

4.6. Skewness and Kurtosis for GOTL-Weibull Distribution

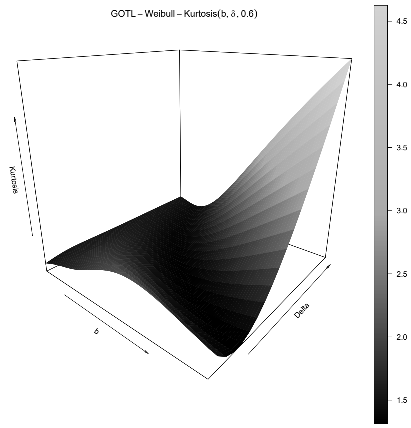
The heterogeneous behaviours captured by the GOTL-G family of distributions were investigated using skewness and kurtosis properties for the GOTL-W and GOTL-U distributions. The skewness and kurtosis properties for the GOTL-W distribution as represented by Figure 2. Figure 2(a-c) shows a variety shapes in the skewness whilst varying kurtosis shapes from the GOTL-Weibull distribution are given by Figure 2(d-f). It can be seen that for unique values b, δ and λ of the GOTL-W distribution results in unique skewness or kurtosis shapes. We can therefore conclude that the GOTL-W is flexible due to its ability to capture complex heterogeneous properties including outliers, kurtosis, and skewness among others.



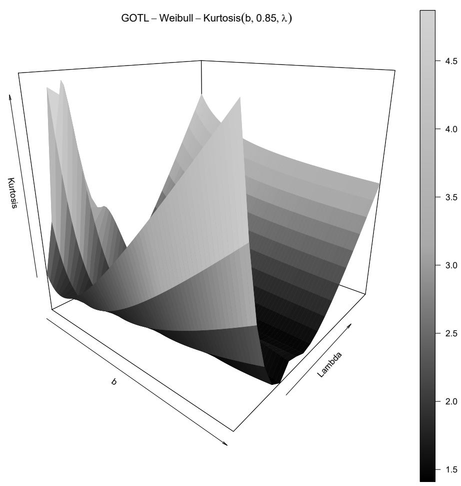
(a) (b)



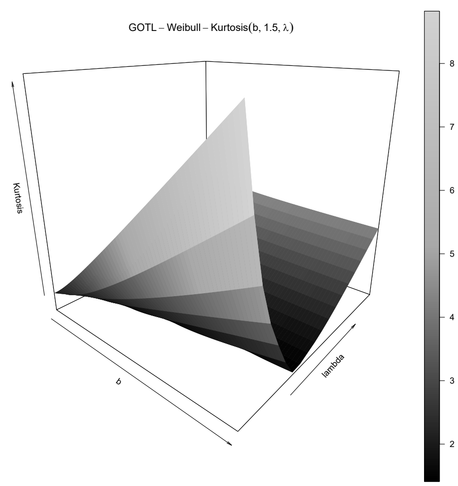
(c)



(d)



(e)



(f)

FIGURE 2: GOTL-Weibull: Skewness shapes for various parameters (a-c) and kurtosis shapes for various parameters (d-f)

4.7. Maximum Likelihood Estimation

Maximum likelihood estimates (MLE) are obtained by minimising the log likelihood function of the joint pdf of the GOTL-G family of distributions. Suppose $x_1, \ldots, x_n$ be a random sample from the GOTL-G distribution and $\boldsymbol{\Delta} = (b, \delta, \boldsymbol{\xi})^T$ be the parameter vector. To determine the MLE of Δ, we need the log of likelihood function for the joint pdf of the GOTL-G family obtained from Equation (6) and given by

$$\begin{aligned} L &= n\ln(2) + n\ln(b) + \sum_{i=1}^{n} \ln(g(x,\boldsymbol{\xi})) + \sum_{i=1}^{n} \ln(\bar{G}(x,\boldsymbol{\xi})) \\ &+ (\delta b - 1)\sum_{i=1}^{n} \ln(1-\bar{G}^2(x,\boldsymbol{\xi})) - n\ln(\Gamma(\delta)) - (\delta+1)\sum_{i=1}^{n} \ln\left(1-[1-\bar{G}^2(x,\boldsymbol{\xi})]^b\right) \\ &- \sum_{i=1}^{n} \left\{ \frac{[1-\bar{G}^2(x,\boldsymbol{\xi})]^b}{1-\left(1-\bar{G}^2(x,\boldsymbol{\xi})\right)^b} \right\}^n . \end{aligned} \tag{30}$$

The MLE $\hat{\Delta} = (\hat{b}, \hat{\delta}, \hat{\boldsymbol{\xi}})^T$ are obtained by equating the first partial derivatives with respect to each of the parameters to zero and solve them simultaneously, that is

$$\frac{\partial L}{\partial \delta} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \boldsymbol{\xi}} = 0.$$

These equations normally result in a non-linear system which can be solved by numerical methods like the quasi-Newton algorithm to numerically maximize L. In this work we used a GOTL- Weibull distribution to determine L where $g(x, \boldsymbol{\xi})$ is pdf and $G(x, \boldsymbol{\xi})$ is the cdf given by Equations (22) and (21), respectively. Interval estimation is from the observed information matrix $J(\boldsymbol{\Delta}) = \{\partial^2 L/\partial r \partial s\}$ for $r, s = b, \delta, \boldsymbol{\xi}$, whose elements were computed numerically using the limited memory algorithm for bound constrained optimization technique (`L-BFGS-B`) as explained by Byrd et al. (1995) and implemented in R through the bblme package by Bolker (2014) in R.

5. Simulation Study

A simulation study on the parameter values was conducted shown in Appendix A.2. The simulation results for the different true parameter values were computed for a sample of $N = 3000$ as given by Tables A2 and A3. The tables lists mean MLEs of the model parameters $\hat{b}$, $\hat{\delta}$, $\hat{\lambda}$ together with the respective average bias (ABIAS) and root mean squared errors (RMSEs). In general, the average bias (ABIAS) and RMSE for the estimated parameter, say, $\hat{\theta}$, are given by: $ABIAS(\hat{\theta}) = \frac{\sum_{i=1}^{N} \hat{\theta}_i}{N} - \theta$ and $RMSE(\hat{\theta}) = \sqrt{\frac{\sum_{i=1}^{N} (\hat{\theta}_i - \theta)^2}{N}}$. Simulation results in both Tables A2 and A3 show that as $n \to \infty$, both RMSE and Bias reduces towards zero whilst the mean estimates of the parameters tend closer to the true parameter values ($\hat{\boldsymbol{\Delta}} \to \boldsymbol{\Delta}$). Thus, we can conclude that the parameters generated by our new distribution are consistent and efficient.

6. Applications

In this section, we assess the usefulness and applicability of the proposed new GOTL-G family of distributions on different types of cancer data using the Weibull distribution as baseline distribution. A range of methods were used to assess the goodness-of-fit for the model. These goodness-of-fit statistics includes -2 log likelihood (-2 log L), Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (AICC), Bayesian Information Criterion (BIC), Cramer-von Mises (W^*) and Andersen-Darling (A^*) and the p-value (PV) for Kolmogorov-Smirnov (K-S) statistic, as described by Chen & Balakrishnan (1995) were presented in this study. These statistics are used to verify which model fits the given dataset by checking if the model had smallest values for the GOTL-W distribution compared to other models. Smaller values for the goodness-of-fit statistics indicates better fit whilst a high p-value for the KS statistic, is an indication of better fit. In this study, we used two datasets over a range of competing and comparable models. We fitted the data to our new model (GOTL-W) and compared it to competing model like: two parameter weibull (Weibull), three parameter gamma (Gamma3), Weibull Lomax (WLx) by Tahir et al. (2015), Weibull Exponential (WE) by Bilal et al. (2021), Generalised Odd Log Logistic (GOLLE) by Cordeiro et al. (2017), Type I Half Logistic Weibull (TIHLW) distribution by Kumar et al. (2015), and New Modified Burr III by Jamal et al. (2021). The pdfs of these distributions are given in the Appendix A.1.

6.1. Breast Cancer Tumour Data

Our new GOTL-W model was fitted to breast cancer tumour texture data by Wolberg et al. (1995) which represents the texture of breast cancer tumour data on UCI. Table 3 shows the ML estimates (and their standard errors) and goodness-of-fit statistics for GOTL-W distribution together with other models. It is evident based on Table 3 statistics for the breast cancer tumour texture data, that the GOTL-W distribution has the lowest AIC, AICC, BIC, W^*, A^* and highest p-value for the K-S statistic. The estimated variance-covariance matrix for the GOTL-W model on the breast cancer tumour textures data is given by,

$$\begin{bmatrix} 4.661 \times 10^{-5} & 1.735 \times 10^{-7} & 2.1525 \times 10^{-5} \\ 1.735 \times 10^{-7} & 6.458 \times 10^{-10} & 8.022 \times 10^{-8} \\ 2.1525 \times 10^{-5} & 8.022 \times 10^{-8} & 1.014 \times 10^{5} \end{bmatrix}$$

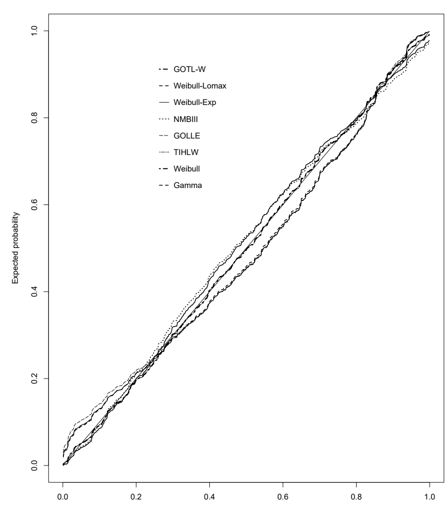
The 95% confidence interval for $b,\ \delta,\ \lambda$ are given by $0.225 \pm 0.013,\ 96.924 \pm 9.761 \times 10^{-5}$ and 0.150 ± 0.006, respectively.

Table 3: Breast cancer Tumour data: ML estimates for various fitted models and Goodness-of-Fit measures

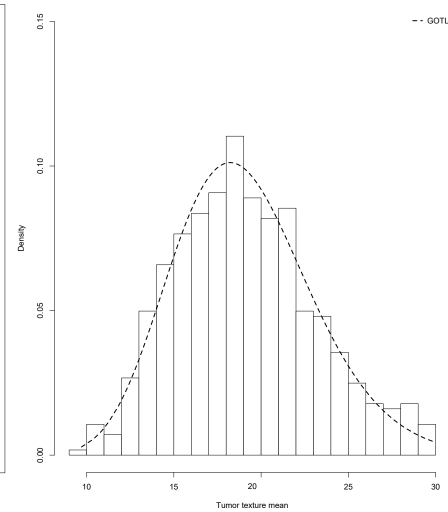
Model	Estimates			Statistics							
				$2\log L$	AIC	AICC	BIC	W^*	A^*	KS	PV
	b	δ	λ								
GOTL-W	0.22525	96.924	0.15019	3143.448	3149.448	3149.491	3162.443	0.02483469	0.2612525	0.01848	0.9928
	(6.8271×10^{-3})	(2.5413×10^{-5})	(3.184×10^{-3})								
	k	λ									
Weibull	5.04308925	0.04816469	-	3191.023	7839.645	7839.666	7848.308	0.04613467	0.3650638	0.67061	<0.001
	(0.15806224)	(0.00042685)									
	a	τ	p								
Gamma3	0.499033	24.785614	0.907647	3143.453	3149.453	3149.516	3162.457	0.02491597	0.2623263	0.018197	0.9903
	(0.141540)	(0.018920)	(0.054155)								
	α	β	δ								
WLx	0.572384	0.113771	6.222998	3183.851	3189.851	3189.894	3202.846	0.5179289	3.527784	0.052407	0.09126
	(0.078635)	(0.027016)	(0.455755)								
	c	α	λ								
WE	5.0431	2.7166	0.01773	3191.023	3197.023	3197.066	3210.018	0.6144914	4.145029	0.056857	0.05285
	(0.15806)	(1.0254×10^{-6})	(1.5712×10^{-4})								
	α	δ	λ								
NMBIII	0.0001	125.080	0.2806	3166.465	3172.465	3172.508	3185.46	0.2604548	1.660905	0.041566	0.286
	(0.07977)	(3.613×10^{-4})	(0.013415)								
	α	θ	λ								
GOLLE	1.1906	67.271	24.692	3168.887	3174.887	3174.93	3187.882	0.1952476	1.276323	0.033922	0.5374
	(0.041866)	(1.8287×10^{-4})	(2.242×10^{-3})								
	λ	δ	γ								
TIHLW	4.1029×10^{-5}	0.09429	4.2398	3209.818	232.3835	232.8198	238.6161	0.1030	0.5809	0.0963	0.6101
	(1.4665×10^{-6})	(6.3807×10^{-10})	(1.8844×10^{-10})								

6.1.1. Goodness-of-Fit Plots for the Breast Cancer Texture Data

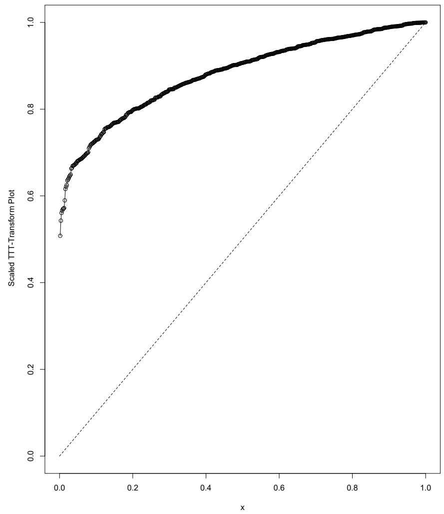
The goodness-of-fit plot for the GOTL-Weibull distribution using the breast cancer tumour texture data is given by Figure 3. The plots include the probability plot in Figure 3(a), the histogram and fitted densities in Figure 3(b), total time on test transform (TTT) on hazard rate function as shown by Figure 3(c), the Kaplan-Meier plot of survival function as in Figure 3(d) and the empirical cdf with fitted cdf given as Figure 3(e). Figure 3(b) confirms the GOTL-Weibull distribution as a better fit compared to the other non-nested models. The TTT plot has a concave shape an indication that the empirical hazard rate function is an increasing function. The pdf, ecdf and Kaplan-Meier plot of survival graphs shows that the GOTL-Weibull fit well on breast tumour texture data and the pdf is distributed across the data. Thus, we can conclude that this GOTL-Weibull distribution fit very well on a unimodal data set like breast cancer tumour texture data. Therefore, from the goodness-of-fit statistics values, p-value of the K-S statistic in Table 3 and the plots in Figure 3, it can be concluded that the GOTL-W distribution does not only fit well on unimodal data but provide a better fit on the breast cancer tumour texture data compared to the other non-nested model being considered.



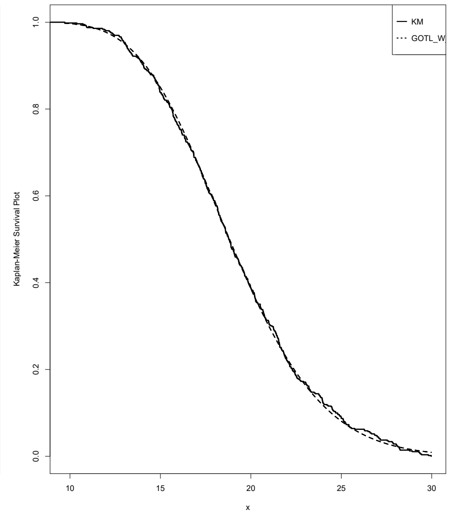
(a)



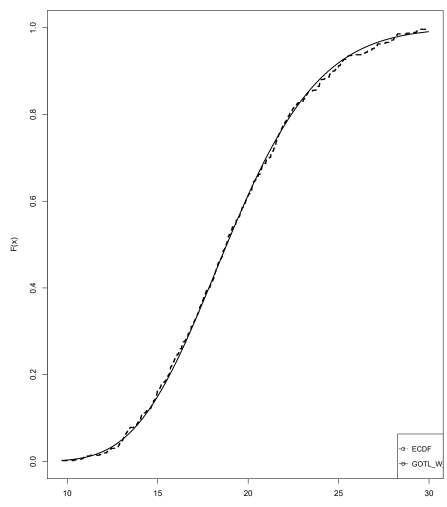
(b)



(c)



(d)



(e)

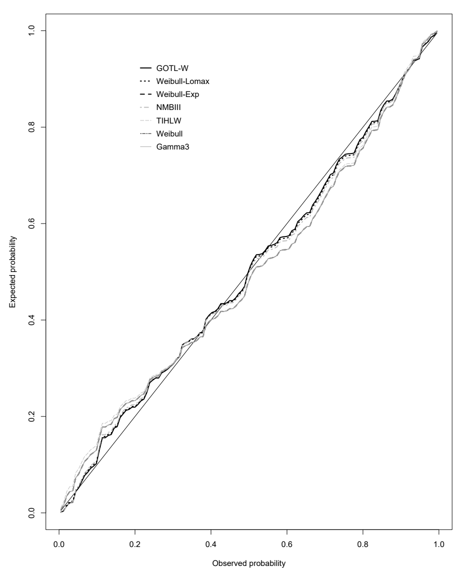
FIGURE 3: Goodness-of-fit plots for the breast cancer texture mean data: a) Probability plot, b) Fitted densities, c) Total time on test plot, d) Kaplan-Meier curve, e) ecdf

6.2. Bladder Cancer Data

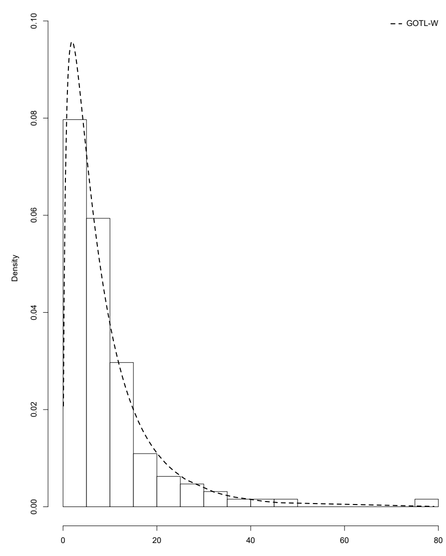
The GOTL-W distribution was also fitted and compared to different models using measures on the uncensored dataset for the remission times in months on 128 random bladder cancer patients as given by Lee & Wang (2003). The results after fitting the different models to the data are presented in Table 4. The goodness-of -fit statistics considered for the bladder cancer data were the same as in Table 3. Based on these results, we can observe that the GOTL-W distribution performs better compared to the other competing models due to their lowest AIC, AICC, BIC, W^*, A^* values and a higher p-value for the K-S statistic. Consistent with these goodness-of-fit results are the results for the probability plot in Figure 4(a) and histogram with fitted densities as given by Figure 4(b), respectively. Both graphs in Figure 4(a-b) shows that GOTL-W distribution has a better fit to the bladder cancer data and hence, we can concluded that the GOTL-W distribution out performs the other models when fitted to the bladder bladder cancer data. The estimated variance-covariance matrix for the cancer patient data is given by,

$$\begin{bmatrix} 0.625 & -0.920 & 0.028 \\ -9.202 & 5.715 & -0.405 \\ 0.028 & -0.405 & 0.001 \end{bmatrix}.$$

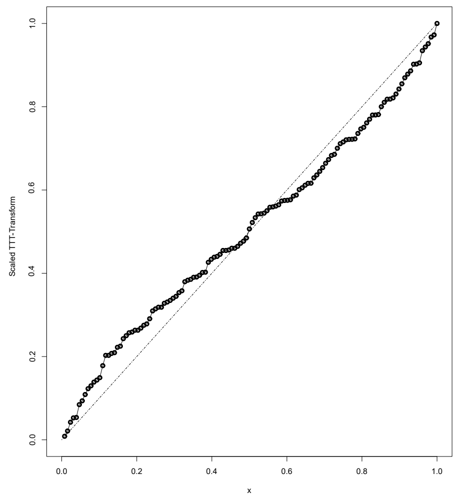
The 95% confidence interval for $b,\ \delta,\ \lambda$ are given by $0.776 \pm 0.010,\ 13.569 \pm 0.0003$, and 0.101 ± 0.007, respectively.



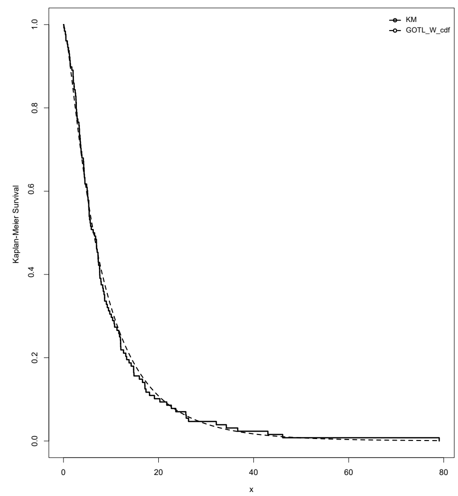
(a)



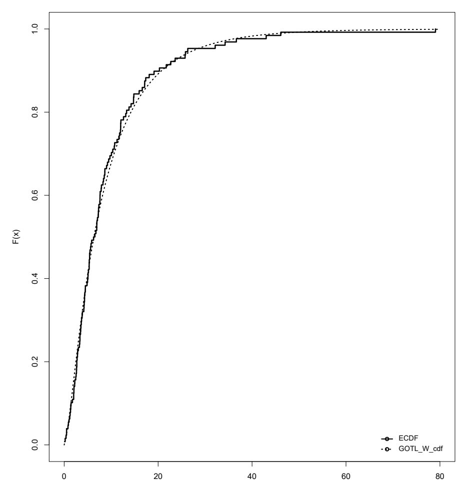
(b)



(c)



(d)



(e)

Figure 4: Goodness-of- fit plots for the bladder cancer data: a) Probability Plot b) Fitted Densities c) TTT plot d) Kaplan-Meier e) ecdf

Table 4: Bladder Cancer Data: ML Estimates for Various Fitted Models and Goodness-of-Fit Measures

Model	Estimates			Statistics							
				$-2\log$ L	AIC	$AICC$	BIC	W^*	A^*	KS	PV
	b	δ	λ								
GOTL−W	0.7756	13.5686	0.1015	821.6687	827.6687	827.8623	836.2248	0.0473	0.3109	0.0464	0.9449
	(0.0790)	(1.6403×10^{-2})	(0.0352)								
	k	λ									
Weibull	1.0478	0.1046	–	828.17386	832.1738	832.2698	837.8778	0.1314	0.7865	0.0700	0.5570
	(0.0676)	(0.0093)									
	a	τ	p								
Gamma3	0.5951	1.9493	0.5201	821.7082	827.7082	827.9018	836.2643	0.0479	0.3143	0.0468	0.9414
	(1.4191)	(0.6564)	(0.1952)								
	α	β	λ								
WLx	0.3926	0.5409	1.6121	821.7901	827.7901	827.9837	836.3462	0.0505	0.3252	0.0491	0.9170
	(0.1520)	(0.4578)	(0.3907)								
	a	b	α								
WE	1.0478	2.7650	0.0378	828.1738	834.1738	834.3673	842.7298	0.1314	0.7864	0.0700	0.5570
	(6.7577×10^{-2})	(4.6165×10^{-5})	(3.3740×10^{-3})								
	α	δ	λ								
NMBIII	0.6296	4.2039	0.0847	824.4131	830.4131	830.6067	838.9692	0.0799	0.5091	0.0529	0.8669
	(0.0984)	(0.4190)	(0.0292)								
	c	α	λ								
GOLLE	0.4000	6.6802	0.3452	866.8642	872.8642	873.0578	881.4203	0.2121	1.2558	0.0762	0.4460
	(0.0596)	(1.4571)	(0.0433)								
	λ	δ	γ								
TIHLW	5.1351	0.0392	0.8880	830.1918	836.1918	836.3854	844.7479	0.1660	0.9780	0.0765	0.4419
	$(5.3918 \times10^{-5}$ R	(7.0592×10^{-3})	(6.0335×10^{-2})								

6.2.1. Goodness-of-Fit Plots for the Bladder Cancer Data

Figure 4(c-d) shows the goodness-of-fit plot for the GOTL-Weibull distribution using the bladder cancer dataset. These include TTT plot on hazard rate function (Figure 4(c)), the Kaplan-Meier plot of survival function (Figure 4(d)) and the empirical cdf with fitted cdf (Figure 4(e)). All graphs confirms that the GOTL-Weibull distribution fit very well on bladder cancer data which right skewed. Figure 4(c) shows that the TTT plot represents a hazard rate function that is a bathtub function. We presented the same goodness-of-fit graphs on a unimodal data using the breast cancer tumour texture dataset and it was shown that GOTL-Weibull distribution can fit an increasing hazard rate function. Thus, the new GOTL-W distribution was found to be an adequate fit for the two datasets with various characteristics.

7. Discussion

Our study developed the new Gamma Odd Topp-Leone-G family of distributions, its behaviour and statistical properties used in real life applications. It was interesting to note that this new family of distributions can cater for different density shapes and hazard rate functions. The GOTL-G family of distributions captured increasing, decreasing and bathtub hazard rates functions among others. Thus, the GOTL-G family has the capacity to capture and model unique heterogeneous attributes in different formats as demonstrated in Section 4.4. This ability of the new developed family of distribution to capture various shapes and attributes makes it ideal for cancer data modelling which normally exhibits unique but complex attributes. Moreover, GOTL-G family of distributions can model complex kurtosis and skewness shapes as demonstrated in Section 4.6. Applicability of any distribution function to different real-life scenarios is very important in statistics. In this work we demonstrated the applicability of our new family of distributions by deriving various expressions of the GOTL-G family in Sections 2.1 and 3, which are useful and relevant in the applicability of cancer modelling in different disciplines or industries. These derived expressions which can be applied to various cancer specialities and research as they include the Rényi entropy, linear representation, moments, and distribution of order statistics. In general, we also expect that parameter estimates should be efficient and consistent. Our newly formed family of distributions satisfy this requirement exceptionally well as given in simulations results Appendix A.2. Our study investigated the performance of the new GOTL-G family of distribution using GOTL-Weibull, a special case of the new family of distributions to different type of cancer data. We fitted the GOTL-Weibull distribution to real life datasets and compared it to other competing models, thus exploring the applicability and usefulness of the new model. Results from Section 6 shows that GOTL-W distribution outperforms other competing distributions like the Weibull, Gamma, GoLLE among others when fitted on both skewed (bladder cancer data) and almost symmetric dataset (breast cancer dataset). This applicability, usefulness, and flexibility advantage makes our new GOTL-G a better model which even fit better to data as confirmed by the goodness-of-fit of plots in Figures 3 and 4. These same plots also shows that the new family of distribution is capable of modelling cancer data exhibiting both an L and S shapes. The GOTL-G family of distribution characteristics possess several advantages when modelling data compared to other competing distributions.

8. Conclusion

Considering the flexibility advantages shown through the derived statistical properties of GOTL-G family of distributions, we can conclude that this new family of distributions is applicable to a variety of data including cancer data. The applicability of this new family of distribution makes it possible to model data which exhibiting simple to complex attributes. We therefore, recommend that this flexibility advantage can be useful in various statistical models used in survival analysis, reliability, finance and many more.

[Received: February 2024 — Accepted: May 2024]

References

Al-Shomrani, A., Arif, O., Shawky, A., Hanif, S. & Shahbaz, M. Q. (2016), 'Topp–Leone family of distributions: Some properties and applications', *Pakistan Journal of Statistics and Operation Research* **12**, 443–451.

Arshad, R. M. I., Tahir, M. H., Chesneau, C. & Jamal, F. (2020), 'The Gamma Kumaraswamy-G family of distributions: Theory, inference and applications', *Statistics in Transition New Series* **21**, 17–40.

Bilal, M., Mohsin, M. & Aslam, M. (2021), 'Weibull-Exponential distribution and its application in monitoring industrial process', *Mathematical Problems in Engineering* **2021**, 6650237.

Bolker, B. (2014), 'Maximum likelihood estimation and analysis with the bbmle package', *R package version* **1**, 17.

Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. (1995), 'A limited memory algorithm for bound constrained optimization', *SIAM Journal on scientific computing* **16**(5), 1190–1208.

Chen, G. & Balakrishnan, N. (1995), 'A general purpose approximate goodness-of-fit test', *Journal of Quality Technology* **27**, 154–161.

Chipepa, F., Oluyede, B. & Makubate, B. (2020), 'The Topp-Leone Marshall-Olkin-G family of distributions with applications', *International Journal of Statistics and Probability* **9**, 1–15.

Cordeiro, G. M., Alizadeh, M., Ozel, G., Hosseini, B., Ortega, E. M. M. & Altun, E. (2017), 'The generalised Odd Log-logistic family of distributions: Properties, regression models and applications', *Journal of Statistical Computation and Simulation* **87**, 908–932.

Cordeiro, G. M. & de Castro, M. (2011), 'A new family of generalised distributions', *Journal of Statistical Computation and Simulation* **81**, 883–898.

Eugene, N.and Lee, C. & Famoye, F. (2002), 'Beta-normal distribution and its applicationsl', *Communications in Statistics-Theory and Methods* **31**, 497–512.

Franssen, L. C., Lorenzi, T., Burgess, A. E. & Chaplain, M. A. (2019), 'A mathematical framework for modelling the metastatic spread of cancer', *Bulletin of mathematical biology* **81**, 1965–2010.

Gradshteyn, I. S. & Ryzhik, I. M. (2000), *Table of Integrals, Series, and Products*, 6th edn, San Diego: Academic, San Diego.

Grizzi, F. & Chiriva-Internati, M. (2006), 'Cancer: looking for simplicity and finding complexity', *Cancer Cell International* **6**(1), 1–7.

Haq, M. A. U., Elgarhy, M. & Hashmi, S. (2019), 'The generalised odd Burr III family of distributions: Properties, applications and characterizations', *Journal of Taibah University for Science* **13**, 961–971.

Hosseini, B., Afshari, M. & Alizadeh, M. (2018), 'The generalised odd Gamma-G family of distributions: Properties and applications', *Austrian Journal of Statistics* **47**, 69–89.

Jamal, F., Abuzaid, A. H., Tahir, M. H., Nasir, M. A., Khan, S. & Mashwani, W. K. (2021), 'New modified Burr III distribution, properties and applications', *Mathematical and Computational Applications* **26**, 82.

Kumar, D., Jain, N. & Gupta, S. (2015), 'The type i generalised half-logistic distribution based on upper record values', *Journal of Probability and Statistics* **2015**.

Lee, E. & Wang, J. (2003), *Statistical Methods for Survival Data Analysis*, 3rd edn, John Wiley & Sons, Hoboken, New Jersey.

Mufudza, C., Sorofa, W. & Chiyaka, E. (2012), 'Assessing the effects of estrogen on the dynamics of breast cancer', *Computational and Mathematical Methods in Medicine* **2012**, 473572. https://doi.org/10.1155/2012/473572

Musekwa, R. R. & Makubate, B. (2023), 'Statistical analysis of Saudi Arabia and UK Covid-19 data using a new generalized distribution', *Scientific African* **22**, e01958.

Oluyede, B., Pu, S., Makubate, B. & Qiu, Y. (2018), 'The Gamma-Weibull-G family of distributions', *Austrian Journal of Statistics* **47**, 45–76.

Pawar, A., Chowdhury, O. R. & Salvi, O. (2022), 'A narrative review of survival analysis in oncology using R', *Cancer Research, Statistics, and Treatment* **5**(3), 554–561.

Peter, P. O., Oluyede, B., Bindele, H. F., Ndwapi, N. & Mabikwa, O. (2021), 'The gamma odd Burr III-G family of distributions: Mmodel, properties and applications', *Revista Colombiana de Estadística* **44**, 331–368.

Qi, X., Chen, G., Li, Y., Cheng, X. & Li, C. (2019), 'Applying neural-network-based machine learning to additive manufacturing: current applications, challenges, and future perspectives', *Engineering* **5**(4), 721–729.

Sanusi, A., Doguwa, S., Audu, I. & Baraya, Y. (2020), 'Topp-Leone Exponential-G family of distributions: Properties and applications', *Journal of Science and Technology Research* **2**, 83–96.

Tahir, M. H., Cordeiro, G. M., Mansoor, M. & Zubaír, M. (2015), 'The Weibull-Lomax distribution: Properties and applications', *Hacettepe Journal of Mathematics and Statistics* **44**, 455–474.

Torabi, H. & Hedesh, N. M. (2012), 'The gamma-uniform distribution and its applications', *Kybernetika* **48**, 16–30.

Tuoyo, D. O., Opone, F. C. & Ekhosuehi, N. (2021), 'The Topp-Leone weibull distribution: Its properties and application', *Earthline Journal of Mathematical Sciences* **7**, 381–401.

Wolberg, W.and Mangasarian, O., Street, N. & Street, W. (1995), *Breast Cancer Wisconsin (Diagnostic).* DOI: https://doi.org/10.24432/C5DW2B.

Appendix A. Appendix

Appendix A.1. Non-nested Models

TABLE A1: Comparison models

Distribution	Density function	Parameters
Weibull	$\frac{\lambda x^{\lambda-1}}{k^{\lambda}} exp^{-(x/k)^{\lambda}}$, $x > 0$	$k,\ \lambda > 0$
Weibull Lomax	$\alpha\beta\delta(1+\beta x)^{\delta-1}\left(1-(1+\beta x)^{-\alpha}\right)^{\delta-1}\exp\left(\frac{1-(1+\beta x)^{-\alpha}}{((1+\beta x)^{-\alpha})^{\delta}}\right)$, $x > 0$	$\alpha, \beta, \delta > 0$
Weibull Exponential	$c(\alpha\lambda)^{c}x^{c-1}\exp\left(-(\alpha\lambda x)^{c}\right)$, $x > 0$	$c, \alpha, \lambda > 0$
Gamma	$\frac{p/a^{d}}{\gamma(d/p)} x^{d-1} exp^{-(x/a)^{P}}$, $x > 0$	$a,\ d,\ p > 0$
New Modified Burr III	$\frac{k(\lambda + c/x)x^{-c}exp^{-\lambda x}}{[(1+x^{-c})exp^{-\lambda x}]^{k+1}}$, $x > 0$	$c, k,\ \lambda > 0$
Generalised Odd Log Logistic	$\frac{\alpha\theta\lambda(1-exp^{-\lambda x})^{\alpha\theta-1}(1-(1-exp^{-\lambda x})^{\theta})^{\alpha-1}}{[(1-exp^{-\lambda x})^{\alpha\theta-1}+(1-(1-exp^{-\lambda x})^{\theta})^{\alpha}]^{2}}$, $x > 0$	$\lambda, \alpha, \theta > 0$
Type I Half Logistic Weibull	$\frac{2\lambda\delta\gamma x^{\gamma-1}exp^{-\lambda\delta x^{\gamma}}}{[1+exp^{-\lambda\delta x^{\gamma}}]^{2}}$, $x > 0$	$\lambda, \delta, \gamma > 0$

Appendix A.2. Simulations

Table A2: Monte-Carlo simulations for the GOTL-Weibull distribution

		$b = 2.00,\ \delta = 2.00,\ \lambda = 2.00$			$b = 2.00,\ \delta = 2.00,\ \lambda = 0.50$		
	n	Mean	RMSE	ABias	Mean	RMSE	ABias
	50	4.053763	5.6298745	2.05376320	4.2211552	5.90579508	2.221155238
	100	3.404910	4.6279535	1.40491044	3.5997635	5.06360070	1.599763489
	200	3.313668	3.9253077	1.31366828	3.2699330	3.79290600	1.269932992
b	400	3.094151	3.6970876	1.09415065	3.1246232	3.73735283	1.124623153
	800	2.596335	2.4999765	0.59633540	2.5467447	2.15454156	0.546744665
	1000	2.436200	1.9013593	0.43619987	2.5240734	2.30548664	0.524073350
	1200	2.220524	1.1025016	0.22052378	2.2596466	1.19736023	0.259646554
	50	4.935156	9.5752003	2.93515604	5.0847653	10.30873778	3.084765299
	100	3.446100	5.4494132	1.44609958	3.3692249	5.38401973	1.369224865
	200	2.377072	2.2342499	0.37707211	2.3824646	2.30892370	0.382464604
δ	400	2.130706	1.4052288	0.13070560	2.1291709	1.42645410	0.129170914
	800	2.015972	0.9532594	0.01597189	2.0060039	0.95540890	0.006003876
	1000	2.010358	0.8310740	0.01035838	1.9737304	0.81866788	-0.026269578
	1200	2.055720	0.7688438	0.05571967	2.0194396	0.75607763	0.019439603
	50	2.335070	1.2693872	0.33506966	0.5912876	0.32474097	0.091287556
	100	2.200046	0.9608486	0.20004602	0.5582071	0.24813662	0.058207127
	200	2.243183	0.8291074	0.24318328	0.5600475	0.20407808	0.060047463
λ	400	2.191914	0.7030000	0.19191445	0.5487144	0.17759241	0.048714372
	800	2.120520	0.5000153	0.12051957	0.5291628	0.12040551	0.029162828
	1000	2.095594	0.4398281	0.09559423	0.5271335	0.11457498	0.027133491
	1200	2.049524	0.3475503	0.04952394	0.5149737	0.08842581	0.014973737

Table A3: Monte-Carlo Simulations for the GOTL-Weibull Distribution

		($b=0.60$ $\delta=1.80$, $\lambda=1.80$)			$b=1.80$, $\delta=1.80$, $\lambda=0.60$		
	n	Mean	RMSE	ABias	Mean	RMSE	ABias
	50	0.4741299	0.23580362	-0.125870127	3.5571789	4.7100126	1.75717895
	100	0.5328153	0.16044065	-0.067184667	3.4688522	4.4978831	1.66885216
	200	0.5625865	0.10513903	-0.037413515	3.2007423	4.2125744	1.40074231
b	400	0.5793613	0.07074384	-0.020638706	2.8328128	3.6752462	1.03281282
	800	0.5913271	0.04736305	-0.009672855	2.3398442	2.5385020	0.53984417
	1000	0.5913450	0.04226301	-0.008654952	2.1031555	1.6590032	0.30315554
	1200	0.5937524	0.03806235	-0.006247620	2.0374616	1.3829566	0.23746160
	50	2.9056662	3.71860957	1.105666204	4.5437001	9.0142969	2.74370005
	100	2.1619308	1.36979952	0.361930849	2.6489368	3.5777992	0.84893680
	200	1.9642298	0.74082291	0.164229787	2.1208519	1.8907192	0.32085188
δ	400	1.8676227	0.46397905	0.067622661	1.8854597	1.1986969	0.08545969
	800	1.8390672	0.30778760	0.039067198	1.8276408	0.8310893	0.02764084
	1000	1.8360048	0.27164756	0.036004819	1.8235521	0.7182960	0.02355209
	1200	1.8255301	0.25666936	0.025530084	1.8203443	0.6682909	0.02034432
	50	2.7479429	2.68941759	0.947942865	0.7681703	0.5268661	0.16817029
	100	2.2236140	1.42110137	0.423613963	0.7344154	0.4036371	0.13441540
	200	1.9770152	0.76566042	0.177015157	0.7041163	0.3318745	0.10411635
λ	400	1.8919489	0.45449199	0.091948865	0.6752902	0.2594751	0.07529017
	800	1.8341767	0.29843411	0.034176680	0.6416348	0.1826325	0.04163477
	1000	1.8278455	0.26400177	0.027845458	0.6285925	0.1417426	0.02859252
	1200	1.8248938	0.24128522	0.024893847	0.6229991	0.1256737	0.0229991