

Double Weighted XGamma Distribution for Cancer Survival Data

D. V. Athira^{1,2}, M. Vijayakumar¹

¹Department of Statistics, Annamalai University, Annamalai Nagar, Tamil Nadu, India,
e-mail: athiramanuphd@gmail.com

²Department of Statistics, Christ Nager College, Maranalloor, Kerala, India

SUMMARY

This article introduces a novel statistical model, the double weighted Xgamma distribution (DWXGD), for analysing cancer survival data. This study investigate its theoretical properties and illustrate its versatility through graphical representations of the probability density and cumulative distribution function for different values of parameters. Parameter estimation is done using the maximum likelihood estimation method. The practical utility of the proposed distribution is illustrated through the use of a real-world leukaemia dataset, in which it consistently outperforms existing models according to $-\log L$, AIC, BIC, AICC, HQIC and CAIC criteria. These findings demonstrate the effectiveness and responsiveness of the DWXGD in modelling skewed or heavily-tailed survival periods, furnishing better forecasts and insightful information for clinical decision-making.

Key words: double weighted Xgamma distribution, reliability analysis, maximum likelihood estimation, model selection technique: AIC, BIC, AICC, HQIC and CAIC

1. Introduction

One of the most significant uses of statistical analysis is in research and applications in the medical field. In order to establish the best model for fitting survival data, cancer research frequently involves further statistical analyses. Survival analysis is a set of statistical procedures for data analysis with the outcome variable of interest being the time until an event occurs. Weighted distributions are extensively used to address challenges in model specification and survival data analysis, particularly when observed data are prejudiced or non-

random. Statistical modelling for lifetime and survival data matters a lot in medical fields, especially in cancer research. Back in the 1930s, Ronald Fisher laid some groundwork with his methods that helped shape later stochastic modelling. In the 1970s David Cox introduced renewal theory, which significantly developed survival analysis and addressed reliability. Weighted distributions were introduced by Patil and Rao around 1978; they discussed how these handle biased data from size-biased sampling. Gupta and Kirmani expanded on this in 1990 with a solid framework for applications. Oluyede made further progress in 1999 with regard to reliability, showing how the distributions fit lifetime data with a high degree of accuracy. In survival analysis, Susarla and Van Ryzin produced an empirical Bayes model for right-censored data in 1978, especially useful for cancer studies. Zelen worked on tumour growth modelling in 1974, achieving good probabilistic models in oncology. Recent work pushes flexible weighted distributions to better capture skewness and variability. Al-Khadim and Hantoosh suggested double weighted exponential and inverse Weibull distributions in 2013 and 2014, detailing their properties and estimation performance. These demonstrated improved performance for lifetime data where hazard rates change. Rishwan added the double weighted Rayleigh distribution in 2013, and gave its characterization. Extensions are constantly being proposed for classical distributions. Priyadarshani constructed a new weighted generalized gamma class in 2011 to handle complex structures more flexibly. Atikankul and others proposed a length-biased weighted Lindley distribution in 2020, showing good performance for biased sampling and real datasets. Applications to cancer data have recently been developed at a fast rate. Alam and others used an arcsine-Gaussian distribution for liver cancer and leukaemia in 2021, showing that it fit better than old models. That shows that there is a need for more flexible distributions in biomedicine. The literature points to weighted and generalized distributions being key for survival and cancer data, dealing with skewness, censoring, and heterogeneity. However, we still appear to need better, more efficient models for the methodologically challenging aspects of cancer survival. This causes a push for new distributions to improve modelling and analysis of those datasets, although some gaps remain open.

Modelling cancer survival data becomes problematic because of the skewness and heavy tails found in the data, in addition to heterogeneity and the censoring issue, whereby some data cuts off early. The traditional setups like the exponential or Weibull distributions are insufficient for capturing such complexities. The Xgamma distribution seems somewhat more flexible, but still struggles with the highly irregular hazard rates found in cancer datasets. A gap therefore exists for a model that can handle statistical outliers in cancer survival data a lot better. This study aims to fill this gap with what will be called the double weighted Xgamma distribution. The goal is to make modelling more accurate and to obtain stronger results from inferences. Cancer survival analysis has significant implications for medical tasks, such as researching and planning treatments or assessing risks. Determining survival times accurately helps to work out how the disease progresses, and makes decisions easier. However, actual cancer data from the real world shows that hazard rates are not at all monotonic, and exhibits high skewness along with kurtosis and the problem of censored observations. Existing distributions often fail to cope with these features. It appears that weighted distributions may perform better, especially double weighted versions. These adjust the base probability structure in ways that add flexibility.

Formally, a double weighted distribution is defined as

$$f_{DW}(x; c) = \frac{w(x)f(x)F(cx)}{\int_0^\infty w(t)f(t)F(ct)dt}, \quad x \geq 0, c > 0,$$

where $f(x)$ and $F(x)$ denote the pdf and cdf of the baseline distribution, respectively, and $w(x)$ is a non-negative weight function.

2. Double Weighted XGamma Distribution (DWXGD)

The probability density function (pdf) and cumulative distribution function (cdf) of the baseline Xgamma distribution is given by

$$f(x) = \frac{\vartheta^2}{(1+\vartheta)} \left(1 + \frac{\vartheta}{2}x^2\right) e^{-\vartheta x} \quad x > 0, \vartheta > 0 \quad (2.1)$$

and

$$F(X) = 1 - \frac{(1+\vartheta+\vartheta x + \frac{\vartheta^2 x^2}{2})}{(1+\vartheta)} e^{-\vartheta x} \quad x > 0, \vartheta > 0$$

Then, the double weighted distribution is

$$f(x) = \frac{w(x)f(x)F(cx)}{W_D}, \quad x \geq 0, c > 0, \quad (2.2)$$

$$\text{where } W_D = \int_0^\infty w(x)f(x)F(cx)dx$$

$$\begin{aligned} &= \int_0^\infty x \frac{\vartheta^2}{(1+\vartheta)} \left(1 + \frac{\vartheta}{2}x^2\right) e^{-\vartheta x} \left(1 - \frac{(1+\vartheta+\vartheta cx + \frac{\vartheta^2 c^2 x^2}{2})}{(1+\vartheta)}\right) e^{-\vartheta cx} dx \\ &= -\frac{\vartheta^3}{(1+\vartheta)^2} \left[\frac{2c}{\vartheta^3(1+c)^3} + \frac{3c^2}{\vartheta^3(1+c)^4} + \frac{12c}{\vartheta^4(1+c)^5} + \frac{60c^2}{\vartheta^4(1+c)^6} \right] \end{aligned} \quad (2.3)$$

Consider the weight function $w(x) = x$.

Substitute equation (2.1), (2.3), $F(cx)$ and $w(x)$ in equation (2.2):

$$f(x) = \frac{x \frac{\vartheta^2}{(1+\vartheta)} \left(1 + \frac{\vartheta}{2}x^2\right) e^{-\vartheta x} \left(1 - \frac{(1+\vartheta+\vartheta cx + \frac{\vartheta^2 c^2 x^2}{2})}{(1+\vartheta)} e^{-\vartheta cx}\right)}{-\frac{\vartheta^3}{(1+\vartheta)^2} \left[\frac{2c}{\vartheta^3(1+c)^3} + \frac{3c^2}{\vartheta^3(1+c)^4} + \frac{12c}{\vartheta^4(1+c)^5} + \frac{60c^2}{\vartheta^4(1+c)^6} \right]}$$

Definition 1: A non-negative continuous random variable X is said to follow a double weighted Xgamma distribution with parameters ϑ and c if its pdf is given by

$$\begin{aligned} &f(x; \vartheta, c) \\ &= \frac{x^2(1 + \frac{\vartheta c}{2}x + \frac{\vartheta}{2}x^2 + \frac{\vartheta^2 c}{2}x^3)e^{-x\vartheta(1+c)}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6})} ; \vartheta, c, x \\ &> 0 \end{aligned} \quad (2.4)$$

This is denoted by $X \sim \text{DWXGD}(\vartheta, c)$.

The expression in equation (2.4) is the probability density function of the double weighted Xgamma distribution, and is denoted by $f(x, \vartheta, c) \forall x > 0$.

Note:

$$f(x; \vartheta, c) = \int_0^\infty \frac{x^2(1 + \frac{\vartheta c}{2}x + \frac{\vartheta}{2}x^2 + \frac{\vartheta^2 c}{2}x^3)e^{-x\vartheta(1+c)}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6})} dx \quad \dots \quad (2.5)$$

using the integral

$$\int_0^\infty x^n e^{-ax} dx = \frac{n!}{a^{n+1}} \quad \forall n > 0 \quad (2.6)$$

Using the integral of (2.6) in equation (2.5):

$$\begin{aligned} f(x; \vartheta, c) &= \frac{\int_0^\infty x^2 e^{-x\vartheta(1+c)} dx + \frac{\vartheta c}{2} \int_0^\infty x^3 e^{-x\vartheta(1+c)} dx + \frac{\vartheta}{2} \int_0^\infty x^4 e^{-x\vartheta(1+c)} dx + \frac{\vartheta^2 c}{2} \int_0^\infty x^5 e^{-x\vartheta(1+c)} dx}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6})} \\ f(x; \vartheta, c) &= \frac{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + \frac{3c}{\vartheta^3(1+c)^4} + \frac{60c}{\vartheta^4(1+c)^6}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6})} \end{aligned}$$

$$f(x; \vartheta, c) = 1,$$

Then the cumulative distribution function of the DWXGD is defined by

$$\begin{aligned} F(x; \vartheta, c) &= \int_0^x f(t; \vartheta, c) dt \\ &= \int_0^x \frac{t^2(1 + \frac{\vartheta c}{2}t + \frac{\vartheta}{2}t^2 + \frac{\vartheta^2 c}{2}t^3)e^{-x\vartheta(1+c)}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6})} dt \end{aligned}$$

Let us define the denominator as a normalizing constant k :

$$k = \frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6})$$

$$\begin{aligned} \text{Then } F(x; \vartheta, c) &= \frac{1}{k} \int_0^x t^2 \left(1 + \frac{\vartheta c}{2}t + \frac{\vartheta}{2}t^2 + \frac{\vartheta^2 c}{2}t^3 \right) e^{-x\vartheta(1+c)} dt \\ &= \frac{1}{k} \int_0^x \left(t^2 + \frac{\vartheta c}{2}t^3 + \frac{\vartheta}{2}t^4 + \frac{\vartheta^2 c}{2}t^5 \right) e^{-x\vartheta(1+c)} dt \\ &= \frac{1}{k} \left\{ \int_0^x t^2 e^{-x\vartheta(1+c)} dt + \int_0^x \frac{\vartheta c}{2} t^3 e^{-x\vartheta(1+c)} dt + \int_0^x \frac{\vartheta}{2} t^4 e^{-x\vartheta(1+c)} dt \right. \\ &\quad \left. + \int_0^x \frac{\vartheta^2 c}{2} t^5 e^{-x\vartheta(1+c)} dt \right\} \end{aligned}$$

Let

$$\int_0^x t^n e^{-at} dt = \frac{n!}{a^{n+1}} \left[1 - e^{-at} \sum_{j=0}^n \frac{(at)^j}{j!} \right] \quad \forall n > 0 \quad (2.7)$$

Using the integral (2.7)

$$F(x; \vartheta, c) = \frac{1}{k} \left\{ \frac{2}{\vartheta^3(1+c)^3} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^3 \frac{(\vartheta(1+c)x)^j}{j!} \right] + \frac{3c}{\vartheta^3(1+c)^4} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^3 \frac{(\vartheta(1+c)x)^j}{j!} \right] + \frac{12}{\vartheta^4(1+c)^5} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^4 \frac{(\vartheta(1+c)x)^j}{j!} \right] + \frac{60c}{\vartheta^4(1+c)^6} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^5 \frac{(\vartheta(1+c)x)^j}{j!} \right] \right\}; x \geq 0, \vartheta c > 0 \quad (2.8)$$

$$\text{where } k = \frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c \left(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6} \right)$$

At $x \rightarrow 0$

$$e^{-\vartheta(1+c)x} \rightarrow 1 \text{ and } (\vartheta(1+c)x)^j \rightarrow 0$$

$$\lim_{x \rightarrow 0} F(x; \vartheta, c) = 0$$

$$\text{At } x \rightarrow \infty$$

$$e^{-\vartheta(1+c)x} \rightarrow 0$$

$$\lim_{x \rightarrow \infty} F(x; \vartheta, c) = 1$$

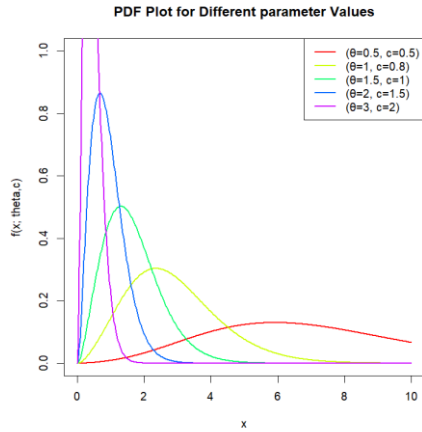


Figure 1: DWXGD pdf curves

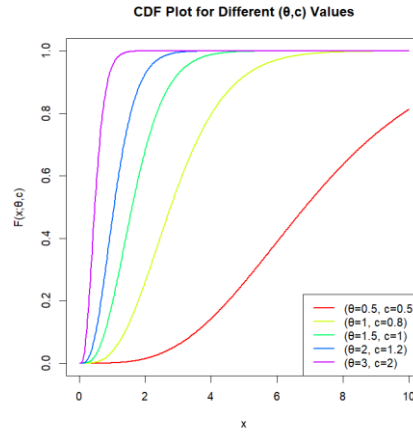


Figure 2: DWXGD distribution curves

Figure 1 shows the density functions of double weighted Xgamma distributions for different values of ϑ and c . Figure 2 shows the distribution functions of double weighted Xgamma distributions for different values of ϑ and c .

3. Reliability function

Assume that x is a continuous random variable from a double weighted Xgamma distribution with pdf $f(x)$. Then the general formulas for the survival and hazard functions are $s(x) = 1 - F(x)$ and $h(x) = \frac{f(x)}{s(x)}$ respectively.

Survival function

The survival function is also known as the reliability function. The survival function of the double weighted Xgamma distribution is given by

$$s(x, \vartheta, c) = 1 - F(x; \vartheta, c) \quad (3.1)$$

Substituting equation (2.8) in (3.1),

$$s(x, \vartheta, c) = 1 - \frac{1}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c\left(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6}\right)} \left\{ \frac{2}{\vartheta^3(1+c)^3} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^3 \frac{(\vartheta(1+c)x)^j}{j!} \right] + \frac{3c}{\vartheta^3(1+c)^4} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^3 \frac{(\vartheta(1+c)x)^j}{j!} \right] + \frac{12}{\vartheta^4(1+c)^5} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^4 \frac{(\vartheta(1+c)x)^j}{j!} \right] + \frac{60c}{\vartheta^4(1+c)^6} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^5 \frac{(\vartheta(1+c)x)^j}{j!} \right] \right\}$$

Hazard function

The hazard function of a double weighted Xgamma distribution is expressed as

$$h(x; \vartheta, c) = \frac{f(x; \vartheta, c)}{s(x; \vartheta, c)}$$

=

$$\frac{\frac{x^2(1+\frac{\vartheta c}{2}x+\frac{\vartheta}{2}x^2+\frac{\vartheta^2 c}{2}x^3)e^{-x\vartheta(1+c)}}{\frac{2}{\vartheta^3(1+c)^3}+\frac{12}{\vartheta^4(1+c)^5}+\frac{60}{\vartheta^3(1+c)^4}+\frac{60}{\vartheta^4(1+c)^6}}}{1-\frac{2}{\vartheta^3(1+c)^3}+\frac{12}{\vartheta^4(1+c)^5}+\frac{60}{\vartheta^3(1+c)^4}+\frac{60}{\vartheta^4(1+c)^6}}\left\{\left[1-e^{-\vartheta(1+c)x}\sum_{j=0}^3\frac{(\vartheta(1+c)x)^j}{j!}\right]+\frac{3c}{\vartheta^3(1+c)^4}\left[1-e^{-\vartheta(1+c)x}\sum_{j=0}^3\frac{(\vartheta(1+c)x)^j}{j!}\right]\right. \\ \left.+\frac{12}{\vartheta^4(1+c)^5}\left[1-e^{-\vartheta(1+c)x}\sum_{j=0}^4\frac{(\vartheta(1+c)x)^j}{j!}\right]+\frac{60c}{\vartheta^4(1+c)^6}\left[1-e^{-\vartheta(1+c)x}\sum_{j=0}^5\frac{(\vartheta(1+c)x)^j}{j!}\right]\right\}$$

$$h(x; \vartheta, c) = [x^2 \left(1 + \frac{\vartheta c}{2}x + \frac{\vartheta}{2}x^2 + \frac{\vartheta^2 c}{2}x^3\right) e^{-x\vartheta(1+c)}] / \left[\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c \left(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6}\right) - \left\{\frac{2}{\vartheta^3(1+c)^3} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^3 \frac{(\vartheta(1+c)x)^j}{j!}\right] + \frac{3c}{\vartheta^3(1+c)^4} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^3 \frac{(\vartheta(1+c)x)^j}{j!}\right] + \frac{12}{\vartheta^4(1+c)^5} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^4 \frac{(\vartheta(1+c)x)^j}{j!}\right] + \frac{60c}{\vartheta^4(1+c)^6} \left[1 - e^{-\vartheta(1+c)x} \sum_{j=0}^5 \frac{(\vartheta(1+c)x)^j}{j!}\right]\right\}\right]$$

The hazard function behaves as follows:

1. As $x \rightarrow 0$, $h(x; \vartheta, c) \rightarrow 0$
2. As $x \rightarrow \infty$, $h(x; \vartheta, c) = \vartheta(1+c)$

Hence the hazard function is increasing.

The survival and hazard function plots are shown in Figures 3 and 4 respectively.

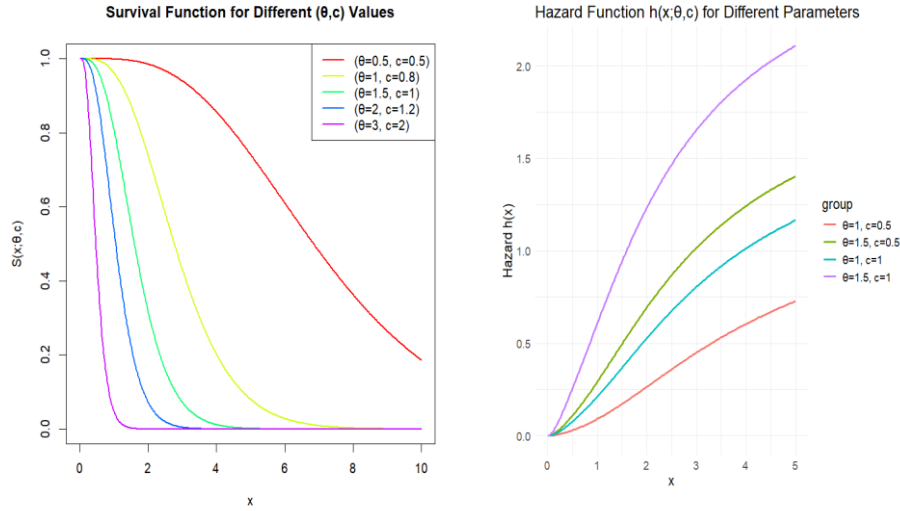


Figure 3: Survival function of DWXGD **Figure 4:** Hazard function of DWXGD

4. Statistical properties

In this section, we discuss statistical properties of the DWXGD, namely moments, moment generating function, and characteristic function.

Moments

The q^{th} order moment of the DWXGD is given by

$$\begin{aligned} \mu'_q &= \frac{\frac{\Gamma(q+3)}{\vartheta^{q+3}(1+c)^{q+3}} + \frac{\vartheta c}{2} \frac{\Gamma(q+4)}{\vartheta^{q+4}(1+c)^{q+4}} + \frac{\vartheta}{2} \frac{\Gamma(q+5)}{\vartheta^{q+5}(1+c)^{q+5}} + \frac{\vartheta^2 c}{2} \frac{\Gamma(q+6)}{\vartheta^{q+6}(1+c)^{q+6}}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6})} \end{aligned}$$

Proof:

The q^{th} moment is defined by

$$\begin{aligned} \mu'_q &= E(X^q) = \int_0^\infty x^q f(x; \vartheta, c) d(x) \\ &= \int_0^\infty x^q \frac{x^2(1 + \frac{\vartheta c}{2}x + \frac{\vartheta}{2}x^2 + \frac{\vartheta^2 c}{2}x^3)e^{-x\vartheta(1+c)}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6})} dx \end{aligned}$$

Let $a = \vartheta(1+c)$ with $\vartheta > 0$, $(1+c) > 0$.

$$\begin{aligned} k &= \frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c\left(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6}\right) \\ \mu'_q &= \int_0^\infty \frac{x^{2+q}(1 + \frac{\vartheta c}{2}x + \frac{\vartheta}{2}x^2 + \frac{\vartheta^2 c}{2}x^3)e^{-xa}}{k} dx \\ &= \frac{1}{k} \left\{ \int_0^\infty x^{q+2} e^{-xa} dx + \int_0^\infty x^{q+3} \frac{\vartheta c}{2} e^{-xa} dx + \int_0^\infty x^{q+4} \frac{\vartheta}{2} e^{-xa} dx \right. \\ &\quad \left. + \int_0^\infty x^{q+5} \frac{\vartheta^2 c}{2} e^{-xa} dx \right\} \end{aligned}$$

Using the standard gamma integral $\int_0^\infty x^v e^{-ax} dx = \frac{\Gamma(v+1)}{a^{v+1}}$

$$\mu'_q = \frac{\frac{\Gamma(q+3)}{a^{q+3}} + \frac{\vartheta c}{2} \frac{\Gamma(q+4)}{a^{q+4}} + \frac{\vartheta}{2} \frac{\Gamma(q+5)}{a^{q+5}} + \frac{\vartheta^2 c}{2} \frac{\Gamma(q+6)}{a^{q+6}}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c \left(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6} \right)}$$

where for $q > -3$ the q^{th} moment is

$$\mu'_q = \frac{\frac{\Gamma(q+3)}{\vartheta^{q+3}(1+c)^{q+3}} + \frac{\vartheta c}{2} \frac{\Gamma(q+4)}{\vartheta^{q+4}(1+c)^{q+4}} + \frac{\vartheta}{2} \frac{\Gamma(q+5)}{\vartheta^{q+5}(1+c)^{q+5}} + \frac{\vartheta^2 c}{2} \frac{\Gamma(q+6)}{\vartheta^{q+6}(1+c)^{q+6}}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c \left(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6} \right)} \quad (4.1)$$

Mean:

Put $q = 1$ in equation (4.1):

$$\begin{aligned} \mu'_1 &= \frac{\frac{\Gamma(4)}{\vartheta^4(1+c)^4} + \frac{\vartheta c}{2} \frac{\Gamma(5)}{\vartheta^5(1+c)^5} + \frac{\vartheta}{2} \frac{\Gamma(6)}{\vartheta^6(1+c)^6} + \frac{\vartheta^2 c}{2} \frac{\Gamma(7)}{\vartheta^7(1+c)^7}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c \left(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6} \right)} \\ &= \frac{\frac{6}{\vartheta^4(1+c)^4} + \frac{12 \vartheta c}{\vartheta^5(1+c)^5} + \frac{60 \vartheta}{\vartheta^6(1+c)^6} + \frac{360 \vartheta^2 c}{\vartheta^7(1+c)^7}}{\frac{2}{\vartheta^3(1+c)^3} + \frac{12}{\vartheta^4(1+c)^5} + c \left(\frac{3}{\vartheta^3(1+c)^4} + \frac{60}{\vartheta^4(1+c)^6} \right)} \\ &= \frac{6\vartheta^3(1+c)^3 + 12c\vartheta^3(1+c)^2 + 60\vartheta^2(1+c) + 360\vartheta^2c}{\vartheta(1+c)[2\vartheta^2(1+c)^3 + 3c\vartheta^2(1+c)^2 + 12\vartheta(1+c) + 60\vartheta c]} \end{aligned}$$

$$MEAN = \mu = \frac{6\vartheta(1+c)^3 + 12c\vartheta(1+c)^2 + 60(1+c) + 360c}{(1+c)[2\vartheta^2(1+c)^3 + 3c\vartheta^2(1+c)^2 + 12\vartheta(1+c) + 60\vartheta c]}$$

Variance: $v(x) = \mu'_2 - (\mu'_1)^2$

$$\mu'_2 = \frac{24\vartheta(1+c)^3 + 60c\vartheta(1+c)^2 + 360(1+c) + 2520c}{(1+c)^2[2\vartheta^3(1+c)^3 + 3c\vartheta^3(1+c)^2 + 12\vartheta^2(1+c) + 60\vartheta^2c]}$$

$$\begin{aligned} v(x) &= \frac{24\vartheta(1+c)^3 + 60c\vartheta(1+c)^2 + 360(1+c) + 2520c}{(1+c)^2[2\vartheta^3(1+c)^3 + 3c\vartheta^3(1+c)^2 + 12\vartheta^2(1+c) + 60\vartheta^2c]} \\ &\quad - \left\{ \frac{6\vartheta(1+c)^3 + 12c\vartheta(1+c)^2 + 60(1+c) + 360c}{(1+c)[2\vartheta^2(1+c)^3 + 3c\vartheta^2(1+c)^2 + 12\vartheta(1+c) + 60\vartheta c]} \right\}^2 \end{aligned}$$

$$\begin{aligned} v(x) &= \frac{12\{\vartheta^6(1+c)^6 + \vartheta^6c(1+c)^5 + 3\vartheta^6c^2(1+c)^4 + 24\vartheta^5(1+c)^4 + 210\vartheta^5(1+c)^2c^2 \\ &\quad + 60\vartheta^4(1+c)^2 + 720\vartheta^4(1+c)c + 1800c^2\vartheta^4\}}{[2\vartheta^3(1+c)^3 + 3c\vartheta^2(1+c)^2 + 12\vartheta^2(1+c) + 60\vartheta^2(1+c)^2c]^2} \end{aligned}$$

5. Estimation of parameters

This section considers parameter estimation for the double weighted Xgamma distribution. The maximum likelihood estimation method is used to estimate unknown parameters.

Let $x_1, x_2, \dots, x_n$ be a random sample of size n taken from a double weighted Xgamma distribution with parameters ϑ, c . Then the likelihood function is

$$L(x) = \prod_{i=1}^n f(x, \vartheta, c)$$

$$L(x) = \prod_{i=1}^n \left(\frac{x_i^2 (1 + \frac{\vartheta c}{2} x_i + \frac{\vartheta}{2} x_i^2 + \frac{\vartheta^2 c}{2} x_i^3) e^{-x_i \vartheta (1+c)}}{\frac{2}{\vartheta^3 (1+c)^3} + \frac{12}{\vartheta^4 (1+c)^5} + c(\frac{3}{\vartheta^3 (1+c)^4} + \frac{60}{\vartheta^4 (1+c)^6})} \right)$$

$$L(x) = \left(\frac{1}{\frac{2}{\vartheta^3 (1+c)^3} + \frac{12}{\vartheta^4 (1+c)^5} + c(\frac{3}{\vartheta^3 (1+c)^4} + \frac{60}{\vartheta^4 (1+c)^6})} \right)^n \prod_{i=1}^n \left(x_i^2 (1 + \frac{\vartheta c}{2} x_i + \frac{\vartheta}{2} x_i^2 + \frac{\vartheta^2 c}{2} x_i^3) e^{-x_i \vartheta (1+c)} \right)$$

The log likelihood function is given by (5.1):

$$\ln L = \sum_{i=1}^n \left\{ \ln x_i^2 + \ln \left(1 + \frac{\vartheta c}{2} x_i + \frac{\vartheta}{2} x_i^2 + \frac{\vartheta^2 c}{2} x_i^3 \right) - \vartheta (1+c) x_i - \right. \\ \left. n \ln k \right\} \quad (5.1)$$

$$\text{where } k = \frac{2}{\vartheta^3 (1+c)^3} + \frac{12}{\vartheta^4 (1+c)^5} + c \left(\frac{3}{\vartheta^3 (1+c)^4} + \frac{60}{\vartheta^4 (1+c)^6} \right)$$

From the log likelihood function given in (5.1), by taking the partial derivative and equating it to zero,

$$\frac{\partial \ln L}{\partial \vartheta} = \sum_{i=1}^n \frac{\frac{c x_i}{2} + \frac{x_i}{2} + \vartheta c x_i^3}{1 + \frac{\vartheta c}{2} x_i + \frac{\vartheta}{2} x_i^2 + \frac{\vartheta^2 c}{2} x_i^3} - \sum_{i=1}^n x_i (1+c) + \frac{n \{ 6\vartheta (1+c)^3 + 48(1+c) + 9c\vartheta (1+c)^2 + 240c \}}{\vartheta (2\vartheta (1+c)^2 + 12)(1+c) + 3c\vartheta^2 (1+c)^2 + 60c\vartheta} = 0 \quad (5.2)$$

$$\begin{aligned}
& \frac{\partial \ln L}{\partial c} \\
&= \sum_{i=1}^n \frac{\frac{\vartheta x_i}{2} + \frac{\vartheta^2 x_i^3}{2}}{1 + \frac{\vartheta c}{2} x_i + \frac{\vartheta}{2} x_i^2 + \frac{\vartheta^2 c}{2} x_i^3} \\
& - \sum_{i=1}^n x_i \vartheta + \frac{n\{12c^2\vartheta + 12c\vartheta + 6\}}{(2\vartheta(1+c)^2 + 12)(1+c) + 3c\vartheta(1+c)^3 + 60c(1+c)} \\
&= 0
\end{aligned} \tag{5.3}$$

The maximum likelihood estimate of the parameters of the double weighted Xgamma distribution is given by equations (5.2) and (5.3). These equations are generally difficult to solve; they involve both linear and nonlinear terms. The equations, however, can be solved numerically; we used the optimx package from R statistical software (Chaussé, P. 2010) to obtain the maximum likelihood estimator of the double Xgamma distribution.

6. Results and discussion

Model selection criteria – namely the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Corrected Akaike Information Criterion (AICC), Hannan–Quinn Information Criterion (HQIC), Consistent Akaike Information Criterion (CAIC) and $-\log L$ – are employed to evaluate and compare the performance of the proposed model.

Data set: the data set consists of 39 patients suffering from blood cancer (leukaemia in year) reported from one of the ministry of health's hospitals in Saudi Arabia.

0.315, 0.496, 0.616, 1.145, 1.208, 1.263, 1.414, 2.025, 2.036, 2.162, 2.211, 2.37, 2.532, 2.693, 3.858, 2.805, 2.91, 2.912, 3.192, 3.263, 3.348, 3.348, 3.427, 3.499, 3.534, 3.767, 3.751, 3.986, 4.049, 4.244, 4.323, 4.381, 4.392, 4.397, 4.753, 4.929, 4.973, 5.074, 5.381

Table 1: Measures (MLEs, model selection criteria) for blood cancer data

Distribution	Double weighted Xgamma	Double weighted exponential	Modified double weighted exponential
MLE	$\hat{\theta}=0.638817$ $\hat{c}=1.609108$	$\hat{\theta}=0.9670858$ $\hat{c}=0.0000010$	$\hat{\theta}=3.644723$ $\hat{c}=0.000001$
–logL	-69.46287	-71.56027	-74.04267
AIC	142.9257	147.1205	152.0853
BIC	146.2529	150.4476	155.4125
AICC	143.2591	144.1195	148.2529
HQIC	147.4538	148.3142	152.4476
CAIC	152.4187	153.2791	157.4125

Table 2: Goodness-of-fit tests for blood cancer data

Distribution	K-S (p-value)	A-D (p-value)	χ^2 (p-value)
Modified double weighted exponential	0.195661 (0.100956)	2.35556 (0.059365)	15.61538 (0.00136)
Double weighted exponential	0.160011 (0.2702)	1.5298 (0.1696)	8.3636 (0.039064)
Double weighted Xgamma	0.13794 (0.4481)	1.0453 (0.3339)	5.4200 (0.14350)

Survival data for 39 leukaemia patients were analysed using double weighted Xgamma, double weighted exponential, and modified double weighted exponential distributions. Comparing both datasets, the **double weighted Xgamma distribution** has the lowest values of –logL, AIC, BIC, AICC, HQIC and CAIC, indicating a good balance between model fit and complexity. Other distributions, including the Weibull and modified exponential models, gave higher information criteria values, reflecting a poorer fit to the observed survival times.

Model adequacy was verified using the Kolmogorov–Smirnov (K-S), Anderson–Darling (A-D), and chi-square goodness-of-fit tests, where higher p-values ($p > 0.05$) and lower test statistics indicate a superior fit. As summarized in Table 2, the modified double weighted exponential distribution exhibited the poorest fit. The double weighted Xgamma distribution demonstrated a consistently

superior fit relative to both baseline models, having the lowest test statistics (K-S = 0.13794, A-D = 1.0453, $\chi^2 = 5.4200$) and highest p-values (0.4481, 0.3339, 0.1435).

These results suggest that the double weighted Xgamma distribution is a flexible and reliable model for survival analysis, particularly for medical datasets with skewed or heavy-tailed characteristics. Practically, this model can improve predictions of patient survival times and support more informed clinical decision-making. Overall, the findings highlight the importance of using multiple information criteria in model selection to identify the distribution that most accurately represents the data.

7. Conclusion

The **double weighted Xgamma distribution** is a flexible and effective model for analysing cancer survival data. Application to leukaemia datasets showed that it provides the best fit compared with existing distributions, based on $-\log L$, AIC, BIC, AICC, HQIC and CAIC values and goodness-of-fit metrics. These results demonstrate its practical applicability for modelling skewed or heavy-tailed survival times, and highlight its potential to improve predictions and support clinical decision-making in cancer studies. **Future work may focus on extending this distribution to handle censored data, incorporating covariates through regression models.**

REFERENCES

- Alam F., Alrajhi S., Nassar M., Afify A.Z. (2021): Modeling liver cancer and leukemia data using arcsine-Gaussian distribution. *Computers, Materials & Continua* 67(2): 2185–2202. <https://doi.org/10.32604/cmc.2021.015089>
- Al-Khadim F., Hantoosh S. (2013): Double weighted exponential distribution: Properties and applications. *Journal of Applied Statistics*, 40(5): 1100–1115.
- Al-Khadim F., Hantoosh S. (2014): Double weighted inverse Weibull distribution: Statistical properties and estimation. *Communications in Statistics – Theory and Methods*, 43(12), 2678–2695.

- Atikankul Y., Thongteeraparp A., Bodhisuwan W., Volodin A. (2020): The length-biased weighted Lindley distribution with applications. *Lobachevskii Journal of Mathematics*, 41: 308–319.
- Chaussé P. (2010): Computing generalized method of moments and generalized empirical likelihood with R. *Journal of Statistical Software*, 34(11): 1–35.
- Cox D.R. (1972): *Renewal theory*. London: Methuen.
- Fisher R.A. (1934): *Statistical methods for research workers* (4th ed.). Edinburgh: Oliver and Boyd.
- Gupta R.C., Kirmani S.N.U.A. (1990): The role of weighted distributions in stochastic modeling. *Communications in Statistics – Theory and Methods*, 19, 3147–3162.
- Oluyede B.O. (1999): On inequalities and selection models for length-biased distributions. *Probability in the Engineering and Informational Sciences*, 13, 169–185.
- Patil G.P., Rao C.R. (1978): Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrika*, 65(2), 229–236.
- Priyadarshani P. (2011): A new class of weighted generalized gamma distributions and related distributions. *Journal of Statistical Theory and Practice*, 5(3), 243–256.
- Rao C.R. (1965): On discrete distributions arising out of methods of ascertainment. *Sankhyā: The Indian Journal of Statistics, Series A*, 27(2), 311–324.
- Rishwan A. (2013): Characterization and estimation of double weighted Rayleigh distribution. *Journal of Statistical Computation and Simulation*, 83(12), 2281–2296.
- Susarla V., Van Ryzin, J. (1978): Empirical Bayes estimation of a distribution (survival) function from right censored observations. *The Annals of Statistics*, 6(4), 740–754.
- Zelen M. (1974): On statistical methods for tumor growth. *Biometrics*, 30(3), 637–645.