International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 23(1), March, 2023, pp 13-26

Analysis and Model Validation of Right Censored Survival Data with Complementary Geometric-Topp-Leone-G Family of Distributions

Laba Handique[1*], Khaoula Aidi[2], Subrata Chakraborty[3], I. Elbatal[4] and M. Masoom Ali[5]

[1*]Department of Statistics, Darrang College, Tezpur-784001, India
[2]Laboratory of Probability and Statistics LaPS, University Badji Mokhtar-Annaba, Algeria
[3]Department of Statistics, Dibrugarh University, Dibrugarh-786004, India
[4]Department of Mathematics and Statistics - College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU)-Saudi Arabia
[5]Department of Mathematical Sciences, Ball State University, Muncie, IN 47306, USA
[*]Correspondence should be addressed to Laba Handique
(Email: handiquelaba@gmail.com)

[Received February 25, 2023; Accepted March 07, 2023]

Abstract

A family of continuous probability distributions, namely the Complementary Geometric-Topp-Leone-G family constructed here, is proposed by using Topp-Leone-G distribution as the baseline distribution in the complementary geometric-G construction. This family has recently been formulated from the family from Marshall-Olkin-G. Here we investigate some new aspects of this family. A stochastic genesis of the family is provided, asymptotes and shapes studied analytically. The parameter estimation for right censored data by method of maximum likelihood and goodness-of-fit tests based on the proposed model is developed. The estimation and testing are validated through extensive simulation experiments. Real life modeling of right censored data is presented.

Keywords: Censored data, Topp-Leone distribution, Estimation, Simulation, Goodness-of-fit, Stochastic formulation.

Mathematics Subject Classification: 60E05, 62G05, 62G20.

1. Introduction

Afify *et al.* (2017) introduced a new family of continuous distributions called the Complementary Geometric-G (CG-G) family starting with a base line distribution with cumulative distribution function (cdf) and probability density function (pdf) $F(x;\phi)$ and $f(x;\phi)$ respectively. The cdf and pdf of the CG-G are given by

$$F^{CGG}(x;\beta)=\frac{\beta F(x:\varphi)}{1-\bar{\beta}F(x:\varphi)} \text{ and } f^{CGG}(x;\beta)=\frac{\beta\, f(x:\varphi)}{[1-\bar{\beta}F(x:\varphi)]^2}.$$

Ali *et al*., (2016) defined the cdf and pdf of the Topp-Leone-G (TL-G) family which is given $G^{TLG}(x;\lambda)=[1-\bar{G}(x)^2]^{\lambda}$ and $g^{TLG}(x;\lambda)=2\lambda\, g(x)\bar{G}(x)\;[1-\bar{G}(x)^2]^{\lambda-1}; x\in R,\ \lambda>0, \beta>0$, where $g(x)$ and $\bar{G}(x)=1-G(x)$ are the pdf and survival function(sf) of the baseline distribution, respectively.

Remark: In appendix A of Afify *et al.* (2017) the parameter $\beta \in (0,1)$, but this restriction is not essential for CGTL-G to be a proper distribution.

Now, by considering the Topp-Leone-G (TL-G) as the baseline will give us a new family with cdf and pdf respectively given by

$$F^{CGTLG}(x;\beta,\lambda)=\frac{\beta[1-\bar{G}(x;\varphi)^2]^{\lambda}}{1-\bar{\beta}[1-\bar{G}(x;\varphi)^2]^{\lambda}} \text{ and}$$

$$f^{CGTLG}(x;\beta,\lambda)=\frac{2\lambda\beta\, g(x;\varphi)\bar{G}(x;\varphi)\;[1-\bar{G}(x;\varphi)^2\;]^{\lambda-1}}{[1-\bar{\beta}\{1-\bar{G}(x;\varphi)^2\}^{\lambda}]^2}.$$

This new family is referred to as the $CG-TLG(\beta,\lambda)$ family of distribution.

$CG-TLG(\beta,\lambda)$ family can alternatively be formulated using Marshall and Olkin (1997) method as follows.

The sf and pdf of Marshall-Olkin (MO) that is Marshall and Olkin (1997) generated family of distributions is constructed with baseline distribution having sf $\bar{F}(x)$ and pdf $f(x)$ are given by

$$\bar{F}^{MO}(x;\alpha)=\frac{\alpha\,\bar{F}(x)}{1-\bar{\alpha}\,\bar{F}(x)} \text{ and } f^{MO}(x;\alpha)=\frac{\alpha\, f(x)}{[1-\bar{\alpha}\,\bar{F}(x)]^2} \quad ; \quad x\in R,\ \alpha>0,\ \bar{\alpha}=1-\alpha.$$

Consideration of the Topp-Leone-G (TL-G) as the baseline will give us a new family called $MOTL\text{-}G(\alpha,\lambda)$ with sf and pdf respectively given by

$$\bar{F}(x;\alpha,\lambda)=\frac{\alpha[1-\{1-\bar{G}(x)^2\}^{\lambda}]}{1-\bar{\alpha}[1-\{1-\bar{G}(x)^2\}^{\lambda}]} \qquad (1)$$

and $$f(x;\alpha,\lambda)=\frac{2\alpha\lambda\, g(x)\bar{G}(x)\,[1-\bar{G}(x)^2]^{\lambda-1}}{[1-\bar{\alpha}[1-\{1-\bar{G}(x)^2\}^{\lambda}]]^2}. \qquad (2)$$

It's easy to check that simple re-parameterisation of $\beta=1/\alpha$ will show that $CG-TLG(\beta,\lambda)\equiv MOTL\text{-}G(1/\beta,\lambda)$.

$MO-TLG$ family has very recently appeared in the literature (Khaleel *et al.*, 2020). The authors derived moments, moment generating functions, quantile function and provided three data fitting applications. In view of the flexibility of this family we are motivated to extend the investigation to derive additional properties not addressed in Khaleel *et al.* (2020) and more importantly we discuss application for this family to censored data which we think will help to enhance and establish the relevance of this family.

The rest of this article is organized in six more Sections. In Section 2, we start with a stochastic formulation of the family and asymptotes and shapes are analytically discussed. The maximum likelihood estimation (MLE) of the parameters for the proposed model in the case of right censored data is addressed in Section 3. In Section 4 goodness-of-fit tests based on the approach of Bagdonavicius and Nikulin (2011) is presented. The MLE and testing is validated through simulation experiments in Section 5. An illustrative modeling example with aright censored data is givenin Section 6. The article ends with a conclusion in Section 7.

For the rest of the article we will refer to (2) as $\mathrm{CGTL-G}(\alpha,\lambda)$. We provide plots of a distribution belonging to the $\mathrm{CGTL-G}(\alpha,\lambda)$ family obtained by taking $\mathrm{Exponential}(\delta)$ as the base line distribution G to reveal the flexibility in shape of the pdf and hazard function. The pdf and cdf of $\mathrm{CGTL\text{-}E}(\alpha,\lambda,\delta)$ are respectively given by

$$f^{CGTLE}(x;\alpha,\lambda,\delta) = \frac{2\alpha\lambda\delta e^{-2\delta x}(1-e^{-2\delta x})^{\lambda-1}}{[1-\bar{\alpha}\{1-(1-e^{-2\delta x})^{\lambda}\}]^{2}} \text{ and}$$

$$F^{CGTLE}(x;\alpha,\lambda,\delta) = \frac{(1-e^{-2\delta x})^{\lambda}}{1-\bar{\alpha}[1-(1-e^{-2\delta x})^{\lambda}]} .$$

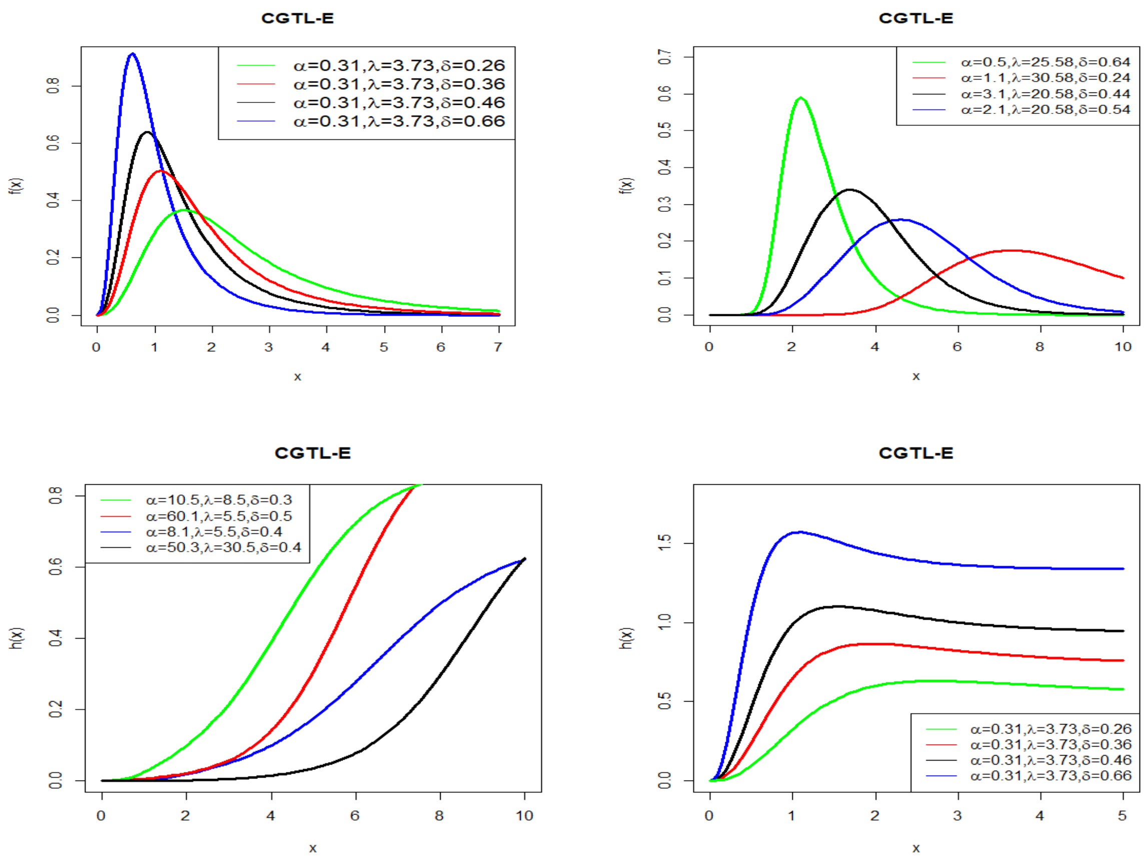

Figure 1: PDF and HRF plots of the $\mathrm{CGTL-E}(\alpha,\lambda,\delta)$

2. Properties

2.1 Stochastic formulation

Theorem1: Let $X_i \ (i = 1, 2, ..., N)$ be a sequence of *i.i.d.* random variables with survival function $\bar{G}^{TLG}(x;\lambda) = 1 - [1 - \bar{G}(x)^2]^{\lambda}$, then

i. if N has a geometric distribution with parameter $\alpha \ (0 < \alpha \leq 1)$ independent of X_i's, then $U = \min(X_1, X_2, ..., X_N)$ is distributed as $\text{CGTL} - \text{G}(\beta, \lambda)$, $\beta (= 1/\alpha) > 1$ or

ii. if$N$ has a geometric distribution with parameter $1/\alpha \ (\alpha > 1)$ independent of X_i's, then $V = \max(X_1, X_2, ..., X_N)$ is distributed as $\text{CGTL} - \text{G}(\beta, \lambda)$, $\beta (= 1/\alpha) < 1$.

Proof: For given N

(i) For $0 < \alpha < 1$, $P[U > x] = P[\min(x_1, x_2, \ldots x_N) > x] = \prod^{N} P(X_i > x) = \prod^{N} [\bar{F}_X(x)]^N$.

Let $N \sim geo(\alpha)$ then $P(N = u) = (1-\alpha)^{n-1}\alpha, \quad n = 1, 2, ...$ then

$$E[\bar{F}_X(x)]^N = \sum_{n\geq 1}^{\infty} \bar{F}_X(x)^n (1-\alpha)^{n-1}\alpha = \alpha \bar{F}_X(x) \sum_{n\geq 1}^{\infty} \bar{F}_X(x)^{n-1}(1-\alpha)^{n-1} = \frac{\alpha \ \bar{F}_X(x)}{1 - \bar{\alpha} \ \bar{F}_X(x)}$$

Now $\bar{F}_X(x) = 1 - [1 - \bar{G}(x)^2]^{\lambda}$ then $\bar{F}_U(x) = \dfrac{\alpha[1 - \{1 - \bar{G}(x)^2\}^{\lambda}]}{1 - \bar{\alpha}[1 - \{1 - \bar{G}(x)^2\}^{\lambda}]}$ and hence

$$F_U(x) = 1 - \frac{\alpha[1 - \{1 - \bar{G}(x)^2\}^{\lambda}]}{1 - \bar{\alpha}[1 - \{1 - \bar{G}(x)^2\}^{\lambda}]} = 1 - \frac{\alpha - \alpha\{1 - \bar{G}(x)^2\}^{\lambda}}{\alpha + \bar{\alpha}\{1 - \bar{G}(x)^2\}^{\lambda}} = \frac{\{1 - \bar{G}(x)^2\}^{\lambda}}{\alpha + \bar{\alpha}\{1 - \bar{G}(x)^2\}^{\lambda}}$$

$$= \frac{(1/\alpha)\{1 - \bar{G}(x)^2\}^{\lambda}}{1 + (\bar{\alpha}/\alpha)\{1 - \bar{G}(x)^2\}^{\lambda}} = \frac{\beta\{1 - \bar{G}(x)^2\}^{\lambda}}{1 - \bar{\beta}\{1 - \bar{G}(x)^2\}^{\lambda}}$$

which is the survival function of $\text{CGTL} - \text{G}(\beta, \lambda)$, $\beta > 1$.

(ii) For, $\alpha > 1$, $F_V(x) = P(V \leq x) = P[X_i \leq x, \ i = 1, 2, .., N] = F_X(x)^N$

Let $N \sim geo(1/\alpha)$ then $P(N = u) = (1 - 1/\alpha)^{n-1} 1/\alpha, \quad n = 1, 2, ...$

$$F_V(x) = E_N[F_X(x)]^N = \sum_{n\geq 1}^{\infty} F_X(x)^n (1 - 1/\alpha)^{n-1} 1/\alpha$$

$$= 1/\alpha \ F_X(x) \sum_{n\geq 1}^{\infty} F_X(x)^{n-1}(1 - 1/\alpha)^{n-1} = \frac{1/\alpha \ F_X(x)}{1 - (1 - 1/\alpha) \ F_X(x)} = \frac{\beta\{1 - \bar{G}(x)^2\}^{\lambda}}{1 - \bar{\beta}\{1 - \bar{G}(x)^2\}^{\lambda}}$$

which is the cdf of $\text{CGTL} - \text{G}(\beta, \lambda)$, $0 < \beta < 1$.

2.2 Infinite mixture of the exponentiated-G (EG)

Here we express pdf of $\mathrm{CGTL-G}(\beta,\lambda)$ an infinite mixture of the exponentiated-G (EG) densities. For $|z|<1$ and $k>0$, considering the infinite series expansion of $(1-z)^{-k}$ when $\alpha\in(0,1)$ we obtain after some patch work

$$\bar{F}^{CGTLG}(x;\alpha,\lambda)=\frac{\alpha\bar{G}^{TLG}(x;\lambda)}{1-\bar{\alpha}\bar{G}^{TLG}(x;\lambda)}=\alpha\bar{G}^{TLG}(x;\lambda)[1-\bar{\alpha}\bar{G}^{TLG}(x;\lambda)]^{-1}$$

$=\sum_{i=0}^{\infty}\alpha(1-\alpha)^{j}\{\bar{G}^{TLG}(x;\lambda)\}^{i+1}$ which on differentiating gives

$$f^{CGTLG}(x;\alpha,\lambda)=-\sum_{i=0}^{\infty}\alpha(1-\alpha)^{i}\frac{d}{dx}[\bar{G}^{TLG}(x;\lambda)]^{i+1}$$

$$=\alpha g^{TLG}(x;\lambda)\sum_{i=0}^{\infty}(1-\alpha)^{i}(i+1)[\bar{G}^{TLG}(x;\lambda)]^{i}$$

Next inserting the pdf and sf of TL-G we get

$$f^{CGTLG}(x;\alpha,\lambda)=2\alpha\lambda\sum_{i=0}^{\infty}\bar{\alpha}^{i}(i+1)\,g(x)\bar{G}(x)\,[1-\bar{G}(x)^{2}]^{\lambda-1}[1-\{1-\bar{G}(x)^{2}\}^{\lambda}]^{i}.$$

Since $[1-\{1-\bar{G}(x)^{2}\}^{\lambda}]^{i}=\sum_{j=0}^{i}(-i)^{j}\binom{i}{j}\{1-\bar{G}(x)^{2}\}^{\lambda j}$

then $f^{CGTLG}(x;\alpha,\lambda)=2\alpha\lambda\sum_{i=0}^{\infty}\bar{\alpha}^{i}(i+1)\,g(x)\bar{G}(x)\sum_{j=0}^{i}(-i)^{j}\binom{i}{j}[1-\bar{G}(x)^{2}]^{\lambda(j+1)-1}$

$$=2\alpha\lambda\sum_{i=0}^{\infty}\sum_{j=0}^{i}(-i)^{j}\binom{i}{j}\bar{\alpha}^{i}(i+1)\,g(x)\bar{G}(x)\,[1-\bar{G}(x)^{2}]^{\lambda(j+1)-1}\qquad(3)$$

For $\lambda(j+1)-1>0$ we have the binomial expansion

$$[1-\bar{G}(x)^{2}]^{\lambda(j+1)-1}=\sum_{k=0}^{\lambda(j+1)-1}(-i)^{k}\binom{\lambda(j+1)-1}{k}\{\bar{G}(x)^{2}\}^{k}\qquad(4)$$

Using equation (4) in equation (3), we get

$$f^{CGTLG}(x;\alpha,\lambda)$$
$$=2\alpha\lambda\sum_{i=0}^{\infty}\sum_{j=0}^{i}(-i)^{j}\binom{i}{j}\bar{\alpha}^{i}(i+1)\,g(x)\sum_{k=0}^{\lambda(j+1)-1}(-i)^{k}\binom{\lambda(j+1)-1}{k}\bar{G}(x)^{2k+1}$$

But $\bar{G}(x)^{2k+1} = [1-G(x)]^{2k+1} = \sum_{l=0}^{2k+1} (-1)^l \binom{2k+1}{l} G(x)^l$.

Therefore $f^{CGTLG}(x;\alpha,\lambda)$

$$= 2\alpha\lambda \sum_{i=0}^{\infty} \sum_{j=0}^{i} \sum_{k=0}^{\lambda(j+1)-1} (-1)^{j+k} \binom{i}{j} \bar{\alpha}^i (i+1) g(x) \binom{\lambda(i+1)-1}{j} \sum_{l=0}^{2k+1} (-1)^l \binom{2k+1}{l} G(x)^l$$

$$= \sum_{i=0}^{\infty} \left\{ \sum_{j=0}^{i} \sum_{k=0}^{\lambda(j+1)-1} \sum_{l=0}^{2k+1} 2\alpha\lambda \frac{(-i)^{j+k+l}}{l+1} \binom{i}{j} \binom{\lambda(j+1)-1}{k} \binom{2k+1}{l} \bar{\alpha}^i (i+1) \right\} \{(l+1) g(x) \; G(x)^l \}$$

$$= \sum_{l=0}^{\infty} \mu_l \, r_{l+1}(x) \text{ ,where}$$

$$\mu_l = \sum_{k=\max\{0,\lfloor (l-1)/2 \rfloor\}}^{\infty} \; \sum_{j=\max\{0,\lfloor (k+1-\lambda)/\lambda \rfloor\}}^{\infty} \; \sum_{i=j}^{\infty} 2\alpha\lambda \frac{(-i)^{j+k+l}}{l+1} \binom{i}{j} \binom{\lambda(j+1)-1}{k} \binom{2k+1}{l} \bar{\alpha}^i (i+1) \,,$$

and $r_l(x) = l g(x) G(x)^{l-1}$.

Note that the $f^{CGTLG}(x;\alpha,\lambda)$ expression $\lfloor a \rfloor$ means integer part of *a*. Since $\binom{n}{m} = 0, \;\; for\; m > n$ the $f^{CGTLG}(x;\alpha,\lambda)$ reduces to

$$= \sum_{l=0}^{\infty} \left\{ \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} 2\alpha\lambda \frac{(-i)^{j+k+l}}{l+1} \binom{i}{j} \binom{\lambda(j+1)-1}{k} \binom{2k+1}{l} \bar{\alpha}^i (i+1) \right\} \{(l+1) g(x) \; G(x)^l \} \cdot$$

Thus the $\text{CGTL}-\text{G}(\alpha,\lambda)$ family can be written as an infinite mixture of the exponentiated-G (EG) densities with exponentiation parameter (*k* + 1), and, therefore, the $\text{CGTL}-\text{G}(\alpha,\lambda)$ density is a linear combination of the EG densities. A lot of the mathematical and statistical properties of the $\text{CGTL}-\text{G}(\alpha,\lambda)$ family of distributions can therefore be derived from those of EG distribution. Similarly, the cdf of the $\text{CGTL}-\text{G}(\alpha,\lambda)$ family can also be expressed as alinear combination of EG cdfs given by $F^{CGTLG}(x;\alpha,\lambda) = \sum_{l=0}^{\infty} \mu_l \, R_{l+1}(x)$,

where $R_l(x) = G(x)^l$ is the cdf of exp-G distribution with exponentiation parameter (*l*+ 1).

2.3 Asymptotes and shapes

Two propositions regarding asymptotes of the proposed family are discussed here.

Proposition 1: The asymptotes of pdf, sf and hrf of $\text{CGTL}-\text{G}(\alpha,\lambda)$ as $x \to 0$ are given by

$f^{CGTLG}(x) \sim 2\alpha^{-1}\lambda\, g(x)\,[1-\bar{G}(x)^2\,]^{\lambda-1}$, $\bar{F}^{CGTLG}(x) \sim 1$ and $h^{CGTLG}(x)$

$\sim 2\alpha^{-1}\lambda\, g(x)\,[1-\bar{G}(x)^2\,]^{\lambda-1}$

Proposition 2: The asymptotes of pdf, sf and hrf of $\mathrm{CGTL-G}(\alpha,\lambda)$ as $x \to \infty$ are given by

$f^{CGTLG}(x) \sim 2\,\alpha\,\lambda\, g(x)\bar{G}(x)$, $\bar{F}^{CGTLG}(x) \sim \alpha[1-\{1-\bar{G}(x)^2\,\}^{\lambda}\,]$and

$h^{CGTLG}(x) \sim 2\,\lambda\, g(x)\bar{G}(x)\,[1-\{1-\bar{G}(x)^2\}^{\lambda}\,]^{-1}$.

A visual demonstration of the above can be shown by plotting the original and corresponding asymptotic in the same figure. As an illustration in figure 2 we plotted the pdf of $\mathrm{CGTL-E}(\alpha,\lambda,\delta)$ from where it is easy to verify it's asymptotic.

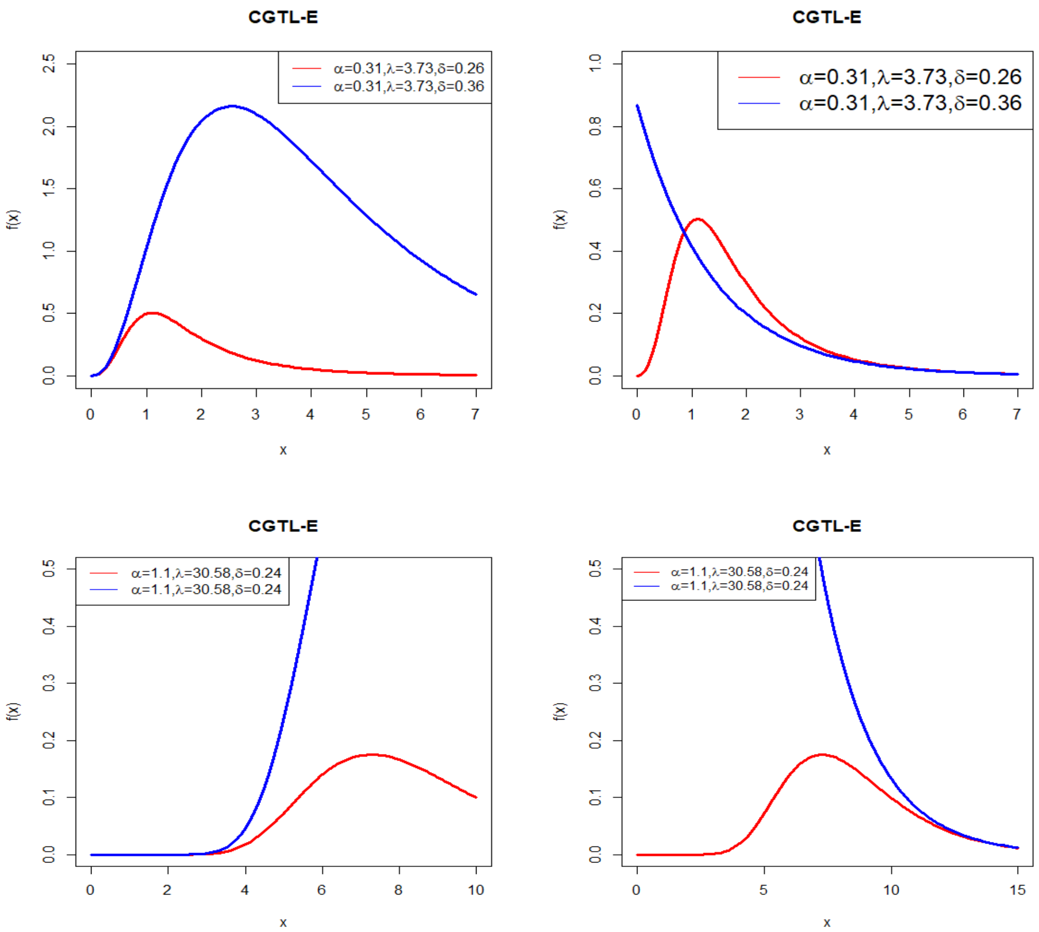

Figure 2: Plots of the pdf (red) and its asymptotic(blue) for same value of the parameters as $x \to 0$ on left and $x \to \infty$ on right.

The shapes of the density and hazard rate function can be described analytically. The critical points of the $\text{CGTL}-\text{G}(\alpha,\lambda)$ family density function are the roots of the equation

$$\frac{g'(x)}{g(x)}-\frac{g(x)}{\overline{G}(x)}+(\lambda-1)\frac{g(x)}{G(x)}-(\lambda-1)\frac{g(x)}{2-G(x)}-\frac{4\overline{\alpha}\,\lambda\, g(x)\overline{G}(x)G(x)^{\lambda-1}\,[2-G(x)]^{\lambda-1}}{1-\overline{\alpha}[1-G(x)^{\lambda}\,\{2-G(x)\}^{\lambda}]}=0\,. \quad (5)$$

The critical point of $\text{MOTL}-\text{G}(\alpha,\lambda)$ family hazard rate are the roots of the equation

$$\frac{g'(x)}{g(x)}-\frac{g(x)}{\overline{G}(x)}+(\lambda-1)\frac{g(x)}{G(x)}-(\lambda-1)\frac{g(x)}{2-G(x)}-\frac{2\overline{\alpha}\,\lambda\, g(x)\overline{G}(x)G(x)^{\lambda-1}\,[2-G(x)]^{\lambda-1}}{1-\overline{\alpha}[1-G(x)^{\lambda}\,\{2-G(x)\}^{\lambda}]}$$

$$+\frac{2\lambda\, g(x)\overline{G}(x)G(x)^{\lambda-1}\,[2-G(x)]^{\lambda-1}}{1-G(x)^{\lambda}\,[2-G(x)]^{\lambda}}=0\,. \quad (6)$$

There may be more than one root of (5) and (6). If $x=x_0$ is a root then it is a local maximum, or a local minimum or a point of inflexion if $\psi(x_0)<0, \psi(x_0)>0$ *or* $\psi(x_0)=0$ and for (6) if $\omega(x_0)<0,\ \omega(x_0)>0$ *or* $\omega(x_0)=0$ where $\psi(x)=(d^2/dx^2)\log[f(x)]$ and $\omega(x)=(d^2/dx^2)\log[h(x)]$.

3. Maximum likelihood estimation for right censored data

In reliability studies and survival analysis, data are often censored. If $X_1, X_2, \ldots, X_n$ is a censored sample from the $\text{CGTL}-\text{G}(\alpha,\lambda)$ distribution, each observation can be written as $x_i = min(X_i, C_i)$ for $i=1,\ldots,n$ where X_i are failure times and C_i censoring times. Censoring is considered to be non-informative, so the likelihood function can be written as

$$L(x,\Theta)=\prod_{i=1}^{n} f^{\Delta_i}(x_i)\bar{F}^{1-\Delta_i}(x_i), \qquad \Delta_i = 1_{\{X_i<C_i\}}.$$

In this case, the total log-likelihood function of $\Theta=(\alpha,\lambda,\xi)^T$ is obtained as follow

$$\begin{aligned}L_n(\Theta) &= \sum_{i=1}^{n}\Delta_i \ln(f(x_i) + \sum_{i=1}^{n}(1-\Delta_i) ln\bar{F}(x_i)\\ &= \sum_{i=1}^{n}\Delta_i\begin{bmatrix} ln(2\alpha\lambda)+\ln\big(g(x_i,\xi)\big)+\ln\big(\bar{G}(x_i,\xi)\big)+(\lambda-1)\ln\big(G(x_i,\xi)\big)\\ +(\lambda-1)\ln\big(2-G(x_i,\xi)\big)-2\ln\Big(1-\bar{\alpha}\big(1-G^{\lambda}(x_i,\xi)\{2-G(x_i,\xi)\}^{\lambda}\big)\Big)\end{bmatrix}\\ &+\sum_{i=1}^{n}(1-\Delta_i)\begin{bmatrix}\ln(\alpha)+\ln\big(1-G^{\lambda}(x_i,\xi)\{2-G(x_i,\xi)\}^{\lambda}\big)\\ -\ln\Big(1-\bar{\alpha}\big(1-G^{\lambda}(x_i,\xi)\{2-G(x_i,\xi)\}^{\lambda}\big)\Big)\end{bmatrix}.\end{aligned}$$

The components of the score functions are

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} &= \sum_{i=1}^{n}\Delta_i\left[\frac{1}{\alpha}-2\frac{1-G^{\lambda}(x_i,\xi)\{2-G(x_i,\xi)\}^{\lambda}}{1-\bar{\alpha}(1-G^{\lambda}(x_i,\xi)\{2-G(x_i,\xi)\}^{\lambda})}\right]\\ &\quad+\sum_{i=1}^{n}(1-\Delta_i)\left[\frac{1}{\alpha}-\frac{1-G^{\lambda}(x_i,\xi)\{2-G(x_i,\xi)\}^{\lambda}}{1-\bar{\alpha}(1-G^{\lambda}(x_i,\xi)\{2-G(x_i,\xi)\}^{\lambda})}\right]\end{aligned}$$

$$\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^{n} \Delta_i \left[\frac{1}{\lambda} + \ln\big(G(x_i,\xi)\big) + \ln\big(2 - G(x_i,\xi)\big) + 2\bar{\alpha}\frac{G^{\lambda}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda}\{\ln\big(G(x_i,\xi)\big) + \ln\big(2 - G(x_i,\xi)\big)\}}{1 - \bar{\alpha}(1 - G^{\lambda}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda})}\right]$$
$$+ \sum_{i=1}^{n}(1 - \Delta_i)\left[-\frac{G^{\lambda}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda}\{\ln\big(G(x_i,\xi)\big) + \ln\big(2 - G(x_i,\xi)\big)\}}{1 - G^{\lambda}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda}} + \frac{\bar{\alpha}G^{\lambda}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda}\{\ln\big(G(x_i,\xi)\big) + \ln\big(2 - G(x_i,\xi)\big)\}}{1 - \bar{\alpha}(1 - G^{\lambda}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda})}\right]$$

$$\frac{\partial \ell}{\partial \xi} = \sum_{i=1}^{n} \Delta_i \left[\frac{g_k{}'(x_i,\xi)}{g(x_i,\xi)} - \frac{G_k{}'(x_i,\xi)}{\bar{G}(x_i,\xi)} + (\lambda - 1)\frac{G'_k(x_i,\xi)}{G(x_i,\xi)} + 4\bar{\alpha}\lambda\frac{G^{\lambda-1}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda-1}G'(x_i,\xi)\big(1 - G(x_i,\xi)\big)}{1 - \bar{\alpha}(1 - G^{\lambda}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda})}\right]$$
$$+ \sum_{i=1}^{n}(1 - \Delta_i)\left[-2\lambda\frac{G^{\lambda-1}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda-1}G'(x_i,\xi)\big(1 - G(x_i,\xi)\big)}{1 - G^{\lambda}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda}} + 2\bar{\alpha}\lambda\frac{G^{\lambda-1}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda-1}G'(x_i,\xi)\big(1 - G(x_i,\xi)\big)}{1 - \bar{\alpha}(1 - G^{\lambda}(x_i,\xi)\{2 - G(x_i,\xi)\}^{\lambda})}\right].$$

As in complete data case, the maximum likelihood estimators cannot be obtained in their explicit form, so numerical methods are required.

4. Test statistic for right censored data

Let $X_1,\ldots,X_n$ be n *i.i.d.* random variables grouped into r classes I_i. To assess the adequacy of a parametric model F_0

$$H_0: \quad P(X_i \le x \mid H_0) = F_0(x;\Theta), x \ge 0, \qquad \Theta = (\Theta_1,\ldots,\Theta_s)^T \in \Theta \subset R^s$$

when data are right censored and the parameter vector Θ is unknown, Bagdonavicius and Nikulin (2011) proposed a statistic test Y^2 based on the vector

$$Z_j = \frac{1}{\sqrt{n}}(U_j - e_j), \quad j = 1,2,\ldots,r \ , \qquad with\, r \succ s.$$

This one represents the differences between observed and expected numbers of failures (U_j and e_j) to fall into these grouping intervals $I_j = (p_{j-1}, p_j]$ with $p_0 = 0, p_r = \tau$, where τ is a finite time. We considered p_j as random data functions such as the r intervals chosen to have equal expected numbers of failurese_j.

The statistic test Y^2 is defined by

$$Y^2 = Z^T\hat{\Sigma}^{-}Z = \sum_{i=1}^{r}\frac{\left(U_j - e_j\right)^2}{U_j} + Q$$

where $Z = (Z_1,\ldots,Z_k)^T$ and $\hat{\Sigma}^{-}$ is a generalized inverse of the covariance matrix $\hat{\Sigma}$ and

$$Q = W^T \hat{G}^- W, \widehat{A_j} = \frac{U_j}{n}, U_j = \sum_{i:X_i \in I_j} \Delta_i$$

$$W = (W_1, \dots, W_s)^T, \hat{G} = [\hat{g}_{ll'}]_{s\times s}, \hat{g}_{ll'} = \hat{\imath}_{ll'} - \sum_{j=1}^{r} \hat{C}_{lj} \hat{G}_{l'j} \hat{A}_j^{-1}$$

$$\hat{C}_{lj} = \frac{1}{n} \sum_{i:X_i \in I_j} \Delta_i \frac{\partial lnh(x_i, \hat{\Theta})}{\partial \Theta}, \hat{\imath}_{ll'} = \frac{1}{n} \sum_{i=1}^{n} \Delta_i \frac{\partial lnh(x_i, \hat{\Theta})}{\partial \Theta_l} \frac{\partial lnh(x_i, \hat{\Theta})}{\partial \Theta_{l'}}$$

$$\widehat{W}_l = \sum_{j=1}^{r} \hat{C}_{lJ} \hat{A}_J^{-1} Z_j \; l, l' = 1, \dots, s$$

Where $h(x_i, \hat{\Theta})$ is the hazard function and $\hat{\Theta}$ is the maximum likelihood estimator of Θ on initial non-grouped data.

Under the null hypothesis H_0, the limit distribution of the statistic Y^2 is a chi-square with $r = rank(\Sigma)$ degrees of freedom. The description and applications of modified chi-square tests are discussed in Voinov*et al.* (2013).

The interval limits p_j for grouping data into j classes I_j are considered as data functions and defined by

$$\hat{p}_j = H^{-1}\left(\frac{E_j - \sum_{l=1}^{i-1} H(x_l, \hat{\Theta})}{n - i + 1}, \hat{\Theta}\right), \quad \hat{p}_j = \max(X_{(n)}, \tau)$$

such as the expected failure times e_j to fall into these intervals are $e_j = \frac{E_r}{r}$for any j, with $E_r = \sum_{i=1}^{n} H(x_i, \Theta)$. The distribution of this statistic test $Y_n^2$is chi-square (see Voinov et al., 2013).

4.1 Test Criteria

Here we consider a particular case $\text{CGTLW}(\alpha, \lambda, \delta, \theta)$ for our investigation. This distribution can be obtained by taking the Weibull distribution (Weibull, 1951) with parameters $\delta > 0$ and $\theta > 0$ having pdf and cdf $g(x) = \delta \theta x^{\theta-1} e^{-\delta x^\theta}$ and $G(x) = 1 - e^{-\delta x^\theta}$, $x > 0$ respectively. The pdf and cdf of $\text{CGTLW}(\alpha, \lambda, \delta, \theta)$ distribution respectively as

$$F^{CGTLW}(x) = \frac{[1 - e^{-2\delta x^\theta}]^\lambda}{1 - \bar{\alpha}[1 - \{1 - e^{-2\delta x^\theta}\}^\lambda]} \text{ and } f^{CGTLW}(x) = \frac{2\alpha \lambda \delta \theta x^{\theta-1} e^{-2\delta x^\theta} \; [1 - e^{-2\delta x^\theta}]^{\lambda-1}}{[1 - \bar{\alpha}[1 - \{1 - e^{-2\delta x^\theta}\}^\lambda]]^2} .$$

For testing the null hypothesis H_0 that data belong to the $\text{CGTLW}(\alpha, \lambda, \delta, \theta)$model, we construct a modified chi-squared type goodness-of-fit test based on the statistic Y^2. Suppose that τ is a finite time, and observed data are grouped into $r > s$ sub-intervals$I_j = (p_{j-1}, p_j]$of $[0, \tau]$. The limit intervals p_j are considered as random variables such that the expected numbers of failures in each interval I_j are the same, so the expected numbers of failures e_j are obtained as

$$E_j = -\frac{j}{r-1}\sum_{i=1}^{n} ln\left[\frac{(1-\bar{\alpha})\left\{1-\left(1-e^{-2\delta x_i^{\theta}}\right)^{\lambda}\right\}}{1-\bar{\alpha}\left\{1-\left(1-e^{-2\delta x_i^{\theta}}\right)^{\lambda}\right\}}\right] \quad j = 1,\ldots,r-1.$$

➢ **Estimated matrix $\widehat{\boldsymbol{W}}$:**

The components of the estimated matrix $\widehat{\boldsymbol{W}}$ are derived from the estimated matrix $\widehat{\boldsymbol{C}}$ which is given by:

$\hat{C}_{1j} = \frac{-1}{n}\sum_{i:x_i\in I_j}^{n} \Delta_i \left[\frac{u_i^{\lambda}}{1-\bar{\alpha}u_i^{\lambda}}\right]$, $\hat{C}_{2j} = \frac{1}{n}\sum_{i:x_i\in I_j}^{n} \Delta_i \left[\frac{1}{\lambda} + \frac{ln(u_i)}{1-u_i^{\lambda}} + \frac{\bar{\alpha}u_i^{\lambda}\ln(u_i)}{1-\bar{\alpha}u_i^{\lambda}}\right]$

$$\hat{C}_{3j} = \frac{1}{n}\sum_{i:x_i\in I_j}^{n} \Delta_i \left[\frac{1}{\delta} - 2x_i^{\theta} + \frac{2(\lambda-1)x_i^{\theta}e^{-2\delta x_i^{\theta}}}{1-e^{-2\delta x_i^{\theta}}} + \frac{2\lambda x_i^{\theta}e^{-2\delta x_i^{\theta}}u_i^{\lambda-1}}{1-u_i^{\lambda}} + \frac{2\bar{\alpha}\lambda x_i^{\theta}e^{-2\delta x_i^{\theta}}u_i^{\lambda-1}}{1-\bar{\alpha}u_i^{\lambda}}\right]$$

$$\hat{C}_{4j} = \frac{1}{n}\sum_{i:x_i\in I_j}^{n} \Delta_i \left[\frac{1}{\theta} + \ln(x_i) - 2\delta x_i^{\theta}\ln(x_i) + \frac{2(\lambda-1)\delta x_i^{\theta}\ln(x_i)e^{-2\delta x_i^{\theta}}}{1-e^{-2\delta x_i^{\theta}}} + \frac{2\lambda\delta x_i^{\theta}\ln(x_i)e^{-2\delta x_i^{\theta}}u_i^{\lambda-1}}{1-u_i^{\lambda}} + \frac{2\bar{\alpha}\lambda\delta x_i^{\theta}\ln(x_i)e^{-2\delta x_i^{\theta}}u_i^{\lambda-1}}{1-\bar{\alpha}u_i^{\lambda}}\right]$$

and $\widehat{W}_l = \sum_{j=1}^{r} \hat{C}_{lJ}\hat{A}_J^{-1}Z_j,\ l, l' = 1,2,3,4 \quad j = 1,\ldots,r.$

➢ **Estimated Matrix $\widehat{\boldsymbol{G}}$:**

The estimated matrix $\hat{G} = [\hat{g}_{ll'}]_{4\times 4}$is defined by $\hat{g}_{ll'} = \hat{\imath}_{ll'} - \sum_{j=1}^{r}\hat{C}_{lJ}\hat{G}_{l'J}\hat{A}_J^{-1}$,

where $\hat{\imath}_{ll'} = \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{\partial}{\partial\Theta_l}\ln h(x_i,\hat{\Theta}) \frac{\partial}{\partial\Theta_{l'}}\ln h(x_i,\hat{\Theta})$, $\quad l,\ l' = 1,2,3,4.$

Therefore the quadratic form of the test statistic can be obtained easily:

$$Y_n^2(\hat{\theta}) = \sum_{j=1}^{r}\frac{\left(U_j - e_j\right)^2}{U_j} + \widehat{W}^T\left[\hat{\imath}_{ll'} - \sum_{j=1}^{r}\hat{C}_{lJ}\hat{G}_{l'J}\hat{A}_J^{-1}\right]^{-1}\widehat{W}.$$

5. Simulations

We generated $N = 10{,}000$ right censored samples with different sizes ($n = 25, 50, 130, 350, 500$) from the $\mathrm{CGTLW}(\alpha,\lambda,\delta,\theta)$ model with parameters $\alpha = 2$, $\lambda = 1.8$, $\delta = 2.5$ and $\theta = 3$.

To generate the random sample from $\mathrm{CGTL-G}(\alpha,\lambda)$ family we use quantile function given by

$Q(u) = G^{-1}\left[1-\sqrt{1-\lambda\sqrt{\alpha u/(1-u\bar{\alpha})}}\right]$; $0<u<1$, where G is the cdf of base line distribution.

Using statistical software R and the Barzilai-Borwein (BB) algorithm (Ravi, 2009), we calculate the maximum likelihood estimates of the unknown parameters and their mean squared errors (MSE). The maximum likelihood estimated parameter values, presented in Table 1, agree closely with the true parameter values while the MSE decreases rapidly with the increase in the *n* values.

Table 1: Mean simulated values of MLEs $\hat{\Theta}$ their corresponding MSEs

$N = 10000$	$n = 25$	$n = 50$	$n = 130$	$n = 350$	$n = 500$
$\hat{\alpha}$	1.9678	1.9787	1.9829	1.9976	1.9998
M.S.E	0.0093	0.0082	0.0054	0.0032	0.0022
$\hat{\lambda}$	1.8214	1.8195	1.8137	1.8086	1.8018
M.S.E	0.0087	0.0075	0.0057	0.0035	0.0025
$\hat{\theta}$	2.5294	2.5216	2.5186	2.5129	2.5086
M.S.E	0.0100	0.0093	0.0082	0.0065	0.0032
$\hat{\delta}$	2.9456	2.9527	2.9685	2.9785	2.9984
M.S.E	0.0078	0.0066	0.0042	0.0028	0.0019

5.1 Criteria test Y_n^2

For testing the null hypothesis H_0 that right censored data came from CGTLW$(\alpha, \lambda, \delta, \theta)$ model, we compute the criteria statistic $Y_n^2(\Theta)$ as defined above for $N = 10000$ simulated samples from the hypothesised distribution with different sizes (30, 50,150, 350, 500).Then, we calculate empirical levels of significance, when Y² > χ_ε^2 (r), corresponding to theoretical levels of significance (ε=0.10, ε=0.05, ε=0.01), We choose $r = 5$. The results are reported in Table 2.

Table 2: Simulated levels of significance for $Y_n^2(\Theta)$test for Complementary geometric Topp-Leone Weibullmodel against their theoretical values ($\varepsilon = 0.01, 0.05, 0.10$).

	$n = 25$	$n = 50$	$n = 130$	$n = 350$	$n = 500$
ε=0.01	0.0069	0.0073	0.0081	0.0090	0.0099
ε =0.05	0.0448	0.0465	0.0475	0.0481	0.0494
ε =0.10	0.0909	0.0953	0.0972	0.0989	0.1002

The null hypothesis H_0 for which simulated samples are fitted by CGTLW$(\alpha, \lambda, \delta, \theta)$ distribution is widely validated for the different levels of significance. Therefore, the test proposed in this work can be used to fit data from this new distribution.

6. Right censored data analysis with CGTLW$(\alpha, \lambda, \delta, \theta)$

Woolson (1981) has reported survival data on 26 psychiatric in-patients admitted to the university of Iowa hospitals during the years 1935-1948. This data for each patient consists of age at 1st admission to the hospital, sex, number of years of follow-up (years from admission to death or censoring) and patient status at the follow up time. The data set is {1, 1, 2, 11, 14, 22, 22, 24, 25, 26, 28, 30*, 30*, 31*, 31*, 32, 33*, 33*, 34*, 35, 35*, 35*, 36*, 37*, 39*, 40}. (* indicates the censorship).

We use the statistic test provided above to verify if these data are modelled by $\mathrm{CGTLW}(\alpha,\lambda,\delta,\theta)$, and to that end, we first calculate the maximum likelihood estimates of the unknown parameters

$$\Theta = (\alpha,\lambda,\delta,\theta)^T = (4.5362, 0.9846, 1.0936, 2.846)^T.$$

Data are grouped into $r = 5$ intervals I_j. We give the necessary calculus in the following Table 3.

Table 3: Values of $p_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}, \hat{C}_{3j}, \hat{C}_{4j}$.

p_j	20.5	27.75	32.767	35.781	40
U_j	5	5	6	6	4
e_j	2.7956	2.7956	2.7956	2.7956	2.7956
$\hat{C}_{1j}$	-1.8634	-2.2384	-1.5326	-3.4315	-2.3548
$\hat{C}_{2j}$	1.7154	0.9215	1. 8164	2.9653	0.8965
$\hat{C}_{3j}$	1.8563	2.8236	1.9364	0.8204	1.0960
$\hat{C}_{4j}$	0.9536	0.8137	1.0635	1.0274	0.9456

Then we obtain the value of the statistic test Y_n^2:

$$Y_n^2 = X^2 + Q = 5.6392 + 3.562 = 9.2012$$

For significance level ε = 0.05, the critical value $\chi_5^2 = 11.0705$ is greater than the value of $Y_n^2 = 9.2012$, so we can say that the proposed $\mathrm{CGTLW}(\alpha,\lambda,\delta,\theta)$ model is a good fit to this data.

7. Concluding remark

Complementary geometric-Topp-Leone-G construction is investigated. Some new characteristics of the family are derived. Maximum likelihood estimates and goodness-of-fit tests for the proposed model are presented for right censored survival data. Estimation and validation is assessed through extensive simulations. The advantage of the proposed model in data modelling is illustrated by considering a real life right censored data set using the proposed tests which yielded favourable results in support of the model.

Conflict of interest: Authors have no conflict of interest.

References

[1] Afify, A. Z., Cordeiro, G. M., Nadarajah, S., Haitham, M. Y., Ozel, G., Nofal, Z. M. and Emrah, A. (2017). The complementary geometric transmuted-G family of distributions: Model, properties and application. Hacettepe Journal of Mathematics and Statistics, https://doi.org/10.15672/HJMS.2017.439.

[2] Ali, A. S., Arif, O., Shawky, A., Hanif, S. and Shahbaz, M. Q. (2016). Topp-Leone Family of Distributions: Some Properties and Application. Pakistan Journal of Statistics and operation Research 3: 443-452.

[3] Bagdonavicius, V. and Nikulin, M. (2011). Chi-squared goodness-of-fit test for right censored data. International Journal of Applied Mathematics and Statistics 24: 30-50.

[4] Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. Water Resour Res 15: 1049-1054.

[5] Khaleel, M. A., Pelumi, E. O., AlAbbasi, J. N., Ibrahim, N. A. and Abujarad, M. A. (2020). The Marshall-OlkinTopp Leone-G family of distributions: A family for generalizing probability models. Scientific African, https://doi.org/10.1016/j.sciaf.2020.e00470.

[6] Marshall, A. And Olkin, I. (1997). A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families. Biometrika 84: 641-652.

[7] Ravi, V. and Gilbert, P. D. (2009). BB An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function.J. Statist. Software, 32: 1-26.

[8] Song, K. S. (2001). Rényi information, log likelihood and an intrinsic distribution measure. J Stat Plan Infer, 93: 51-69.

[9] Voinov, V. Nikulin and M., Balakrishnan, N. (2013). Chi-Squared Goodness of Fit Tests with Applications, Academic Press, Elsevier.

[10] Weibull, W. (1951). A statistical distribution functions of wide applicability. J. Appl. Mech.-Trans. ASME18: 293-297.

[11] Woolson (1981). Survival Analysis: Techniques for Censored and Truncated Data, Statistics for Biology and Health.