Journal of Data Science 251-276 , DOI:10.6339/JDS.201804_16(2).0003

A New Generalized Family of Distributions from Bounded Support

M. H. Tahir [*,1], Gauss M. Cordeiro [2], M. Mansoor [1,3], Ayman Alzaatreh [4], and M. Zubair [1,5]

[1] *Department of Statistics, The Islamia University of Bahawalpur*
[2] *Department of Statistics, Federal University of Pernambuco*
[3] *Department of Statistics, Government Degree College Liaquatpur*
[4] *Department of Mathematics and Statistics, American University of Sharjah*
[5] *Department of Statistics, Government S.E. College Bahawalpur*

Abstract: In this paper, we introduce a new generalized family of distri- butions from bounded support (0,1), namely, the Topp-Leone-G family.Some of mathematical properties of the proposed family have been studied. The new density function can be symmetrical, left-skewed, right-skewed or reverse-J shaped. Furthermore, the hazard rate function can be constant, in- creasing, decreasing, J or bathtub hazard rate shapes. Three special models are discussed. We obtain simple expressions for the ordinary and incomplete moments, quantile and generating functions, mean deviations and entropies. The method of maximum likelihood is used to estimate the model parame- ters. The flexibility of the new family is illustrated by means of three real data sets.

Key words: G-class, J-shaped distribution, moments, reliability function,T-X family, Topp-Leone distribution.

* Corresponding author.

1. Introduction

Several generators have been defined in the literature by introducing one or more parameters to a parent distribution in order to construct more flexible models. It is still debatable whether a

G-class (or distribution) for bounded support is more appropriate than the G-class (or distribution) with unbounded support. For bounded interval (0,1), few G-classes have been explored in the literature such as the beta-G by Eugene et al.(2002) and Jones(2004), Kumaraswamy-G (Kw-G) by Cordeiro and de Castro (2011) and McDonald-G (Mc-G) by Alexan- der et al. (2012). For more details about recent developments in distribution theory, see Alzaatreh et al. (2013) and Lee et al. (2013).

The Topp-Leone (TL) random variable (Topp and Leone, 1955) has bounded support of (0,1) and exhibits J-shaped density and bathtub-shaped hazard rate.The later characteristic is attractive and useful in reliability and lifetime data analysis. Since Nadarajah and Kotz (2003) paper, the TL model has attracted researchers to explore the distribution further. In contrast to the beta distribution, the TL model has a closed-form cumulative distribution function (cdf). If a random variable Z has the TL distribution with shape parameter $\alpha \in (0, 1)$, then its cdf and probability density function (pdf), for $x \in (0, 1)$, are respectively, given by

$$R_{TL}(X) = [1-(1-X)^2]^{\alpha} and\; rTL(X) = 2\alpha(1-X)[1-(1-X)^2]^{\alpha-1}. \quad (1)$$

We study a new class of distributions based on the TL random variable called the Topp-Leone generalized (TLG) family. The paper is organized as follows. In Section 2, we define the new family. Section 3 provides three special TLG distributions. In Section 4, some of its mathematical properties are derived including asymptotics, a useful linear representation for the density function, ordinary and incomplete moments, generating function, a power series for the quantile function (qf) and mean deviations. Two types of entropies are derived in Section 5. In Section 6, the model parameters are estimated by the maximum likelihood method. A simulation study is presented in Section 7. In Section 8, we prove the usefulness of the TLG family by means of three applications to real data sets. Finally, Section 9 offers some concluding remarks.

2. The New Generalized Family

Let T, R and Y be random variables with cdfs $F_T(x) = P(T \leq x)$, $G_R(x) = P(R \leq x)$ and $D_Y(x) = P(Y \leq x)$. The corresponding qfs are $Q_T(p), Q_R(p)$ and $Q_Y(p)$,where the qf is defined by $Q_Z(p) = inf\{ z : F_Z(z) \geq p\}, 0 < p < 1$.If the densities exist, we denote themby $f_T(x)$,g_R(x) and $d_Y(x)$.We assumethe random variables T∈(a,b) and Y∈(c,d), for -∞ $\leq a < b \leq \infty$ and -∞ $\leq c < d \leq \infty$. Aljarrah et al. (2014) (see also Alzaatreh et al., 2014) defined the cdf of the T-R{Y}family

$$by \;\; F_X(x) = \int_a^{Q_Y(G_R(x))} f_T(t)\,dt = F_T\left(Q_Y\left(G_R(x)\right)\right). \quad (2)$$

The pdf and hazard rate function (hrf) corresponding to (2) are, respectively, given by (Alzaatreh et al., 2014)

$$f_X(x) = G_R(x) \times \frac{f_T(Q_Y(G_R(x))))}{d_Y\left(Q_Y\left(G_R(x)\right)\right)} \;\; and \;\; h_X(x) \times \frac{h_T(Q_Y(G_R(x))))}{h_Y(Q_Y(G_R(x))))}$$

If T has the TL distribution and taking the qf of the standard uniform distribution, $Q_Y(x) = x$, the cdf $F\ (x) = F(x;\alpha,\xi)$corresponding to (2) becomes

$$F(x) \int_0^{G_R(x;\xi)} 2\alpha(1-t)[t(2-t)]^{\alpha-1}\,dt = \{1-[1-G_R(x;\boldsymbol{\xi})]^2\}^{\alpha}. \quad (3)$$

Then, the pdf $f\ (x) = f(x;\alpha,\xi)$ of X is given by

$$\mathrm{F}(x) = 2\alpha g_R(x;\xi)[1-G_R(x;\xi)]\{1-[1-G_R(x;\xi)]^2\}^{\alpha-1}. \quad (4)$$

where $g_R(x;\xi)$ is the baseline pdf. The class of distributions in (4) is called the TLG family. The parameter α controls the skewness and kurtosis of the generated family and its tail weights. Further, we can omit sometimes the dependence on the parameter vector ξ and write simply $G(x) = G_R(x;\xi)$ and $g(x) = g_R(x;\xi)$.

Equation (4) will be most tractable when $G(x)$ and $g(x)$ have simple analytic expressions. Hereafter, a random variable X with pdf (4) is denoted by $X \sim TLG(\alpha,\xi)$. The TLG family can be used in statistical communication theory as a model for the amplitude of a periodic signal in thermal noise, as the time allocation in project management and control systems and as the limiting spectral density function of a high-index-angle modulated carrier. Further, it can be applied for modeling economic data, variability of soil properties and proportions of the minerals in rocks in stratigraphy and the behavior of random variables limited to finite intervals in a wide variety of areas.

For any baseline pdf $g_R(x;\xi)$ with parameter vector ξ, the exponentiated-G (“exp-G”) model with power parameter d > 0, say exp-G(d), is defined by the cdf and pdf

$$V_d(x;\xi) = G(x;\xi)^d \quad and \quad v_d(x;\xi) = dg(x;\xi)G(x;\xi)^{d-1},$$

respectively.This transformation is also called the Lehmann type I class of distributions. Some structural properties of the exp-G distributions have been studied by Mudholkar and Srivastava (1993), Mudholkar et al. (1995), Mudholkar and Hutson (1996), Gupta et al. (1998), Gupta and Kundu (1999, 2001), Nadarajah and Kotz (2006), and Nadarajah (2011).

Note that there is a dual transformation, say exp-(1-G)(d), called the Lehmann type II class defined by the cdf $W_d(x;\xi) = 1 - \{1 - G(x;\xi)\}^d$.Thus, equation (3) corresponds to the two-stage construction exp-[exp-(1-G)(2)](α). In other words, the exponentiated transformation applied to the Lehmann type II transformation with parameter two thus generates the TLG family. Some properties of the new family may be facilitated by this construction.

If α is a positive integer, the TLG family shares a physical interpretation scheme. Consider a system made of α parallel independent components, where each component consists of a series of two independent sub-components identically distributed. The system fails if all α components fail and each component fails if at least one of the two sub-component fails. For j = 1, . . . , α, let Xj1 and Xj2 denote the lifetimes of the sub-components within the jth component having a common cdf tt(x; ξ). Let Xj denote the lifetime of the jth component and let X denote the lifetime of the system. Thus, the cdf of X is given by

$$P(X \le x) = \left[1 - P(X_j > x)\right]^\alpha = \left[1 - P(X_{j1} > x, X_{j2} > x)\right]^\alpha$$
$$= [1 - \{P(X_{j1} \le x)\}^2]^\alpha,$$

and then the lifetime of the system follows the TLG family.

Remark 1. If T and X have the TL and TLG distributions, it is easy to prove the following:

$(i) X \overset{d}{=} Q_R(T),$

$(ii) Q_x(P) = Q_R\big(Q_T(P)\big),$

$(iii) if \ T \overset{d}{=} Uniform, then\ X \overset{d}{=} R.$

The survival function (sf), $F(x;\alpha,\xi)$, and hrf, $h(x;\alpha,\xi)$, corresponding to the pdf (4) are given by

$$\bar{F}(x;\alpha,\xi) = 1 - \{1 - [1 - G_R(x;\xi)]^2\}^{\alpha}$$

and

$$h(x;\alpha,\xi) = \frac{2\alpha gR(x;\xi)\{1-G_R(x;\xi)\}\{1-[1-G_R(x;\xi)]^2\}^{\alpha-1}}{1-\{1-[1-G_R(x;\xi)]^2\}^{\alpha}} \qquad (5),$$

respectively.

3. Special models of the TLG family

In this section, we provide four special models of the TLG family, namely:the TL-Weibull, TL-Gamma, TL-log-logistic and TL-logistic distributions.

3.1 The TL-Weibull (TLW) distribution

If the parent distribution has the Weibull distribution with pdf and cdf $\mathrm{g(x) = abx^{b-1}e^{-(ax)^b}}$ and $G(x) = 1 - e^{-(ax)^b}$, x, a, b > 0, the TLW pdf can be expressed as

$$f_{TLW}(x,\alpha,a,b) = 2ab(ax)^{b-1}e^{-2(ax)^b}[1 - e^{-2(ax)^b}]^{\alpha-1}, x > 0$$

If b = 1, then the TLW distribution reduces to the TL-exponential (TLE) model.

If b = 2, it gives the TL-Rayleigh (TLR) distribution.

3.2 The TL-Gamma (TLGa) distribution

If the parent distribution has the gamma distribution with pdf $g(x;\ a,b) = \frac{b^a}{\Gamma(a)} x^{a-1}e^{-bx}$ and cdf $G(x;a,b) = \frac{\gamma(a;b\,x)}{\Gamma(a)}, x, a, b > 0$, where $\gamma(a,y) = \int_0^z y^{a-1}e^{-y}\,dy$ is the incomplete gamma function, the TLGa bdf is given by

$$f_{TLGa}(x;a,b) = \frac{2\alpha b^a}{\Gamma(a)}\, x^{a-1}e^{-bx}[1 - \frac{\gamma(a;b\,x)}{\Gamma(a)}][1 - \{1 - \frac{\gamma(a;b\,x)}{\Gamma(a)}\}^2]^{\alpha-1}.$$

3.3 The TL logistic (TLLc) distribution

If the parent distribution follows the logistic distribution with pdf $g(x;\lambda) = \lambda e^{-\lambda x}(1 + e^{-\lambda x})^{-2}$, $x \in$ R and cdf $G(x;\lambda) = (1 + e^{-\lambda x})^{-1}$, the TLLc pdf is given by

$$f_{TLLc}(x;\alpha,\lambda) = 2\alpha\lambda e^{-\lambda x}(1+e^{-\lambda x})^{-2}\left\{1-\left(1+e^{-\lambda x}\right)^{-1}\right\}\times\left[1-\left\{1-\left(1+e^{-\lambda x}\right)^{-1}\right\}^{2}\right]^{\alpha-1}.$$

3.4 The TL log-logistic (TLL) distribution

If the parent distribution has the log-logistic distribution with pdf $g(x;s,c) = cs^{-c}x^{c-1}[1 + (\frac{x}{s})^{c}]^{-2}$ and cdf $G(x;s,c) = 1 - [1 + (\frac{x}{s})^{c}]^{-1}, x, s, c > 0$, then the TLL pdf is given by

$$f_{TLL}(x;\alpha,s,c) = 2\alpha c s^{-c}x^{c-1}\left[1+\left(\frac{x}{s}\right)^{c}\right]^{-3}\left\{1-\left[1+\left(\frac{x}{s}\right)^{c}\right]^{-2}\right\}^{\alpha-1}.$$

In Figures 1–4, we display some plots of the pdf and hrf of the TLW, TLGa, TLLc and TLL distributions for fixed scale parameters and selected shape parameter values. The plots of Figures 1 and 2 indicate that the TLG family generates distributions with various shapes such as symmetrical, reversed-J, left- skewed and right-skewed. Also, the plots in Figures 3 and 4 reveal that this family can produce flexible hazard rate shapes such as constant, increasing, decreasing, upside-down bathtub and bathtub. These facts show that the TLG family can be very useful in reliability and life-testing for fitting different data sets with various shapes.

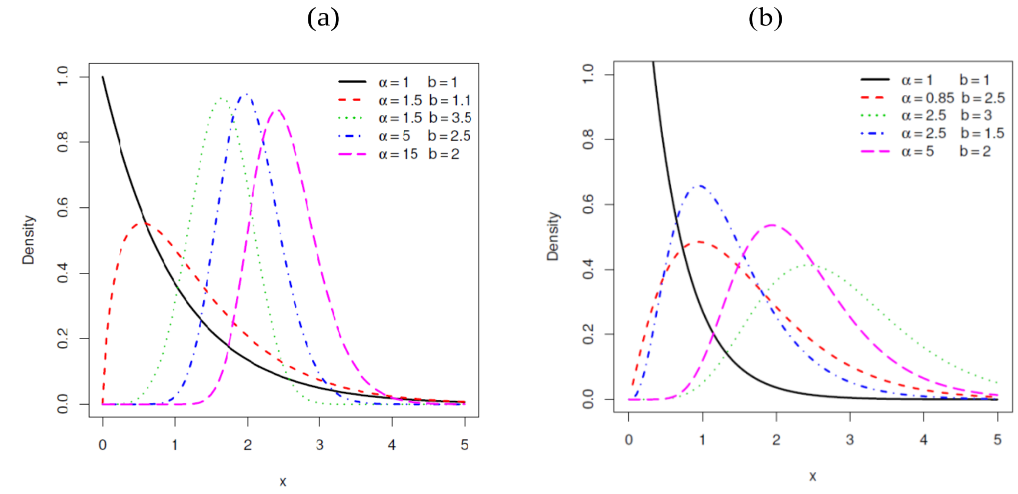


Figure 1: Plots of the (a) TLW and (b) TLGa densities for some parameter values.

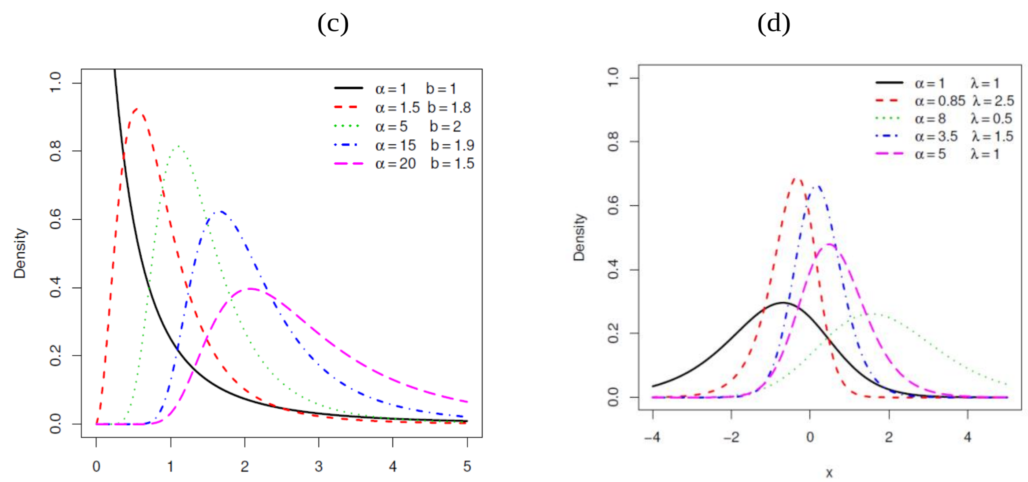


Figure 2: Plots of the (c) TLLc and (d) TLL densities for some parameter values.

4. Some Mathematical Properties of TLG Family of Distributions

In this section, we provide some general properties of the TLG family.

4.1 Asymptotics

Proposition 4.1 *The asymptotics of equations (3),(4) and (5) when* $G(x) \rightarrow 0$

(a) (b)

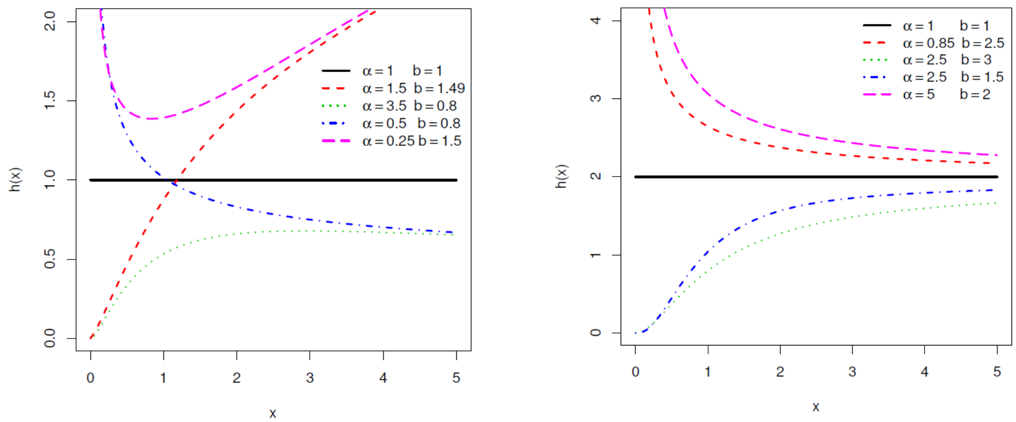


Figure 3: Plots of the (a) TLW and (b) TLGa hazard functions for some param- eter values.

(c) (d)

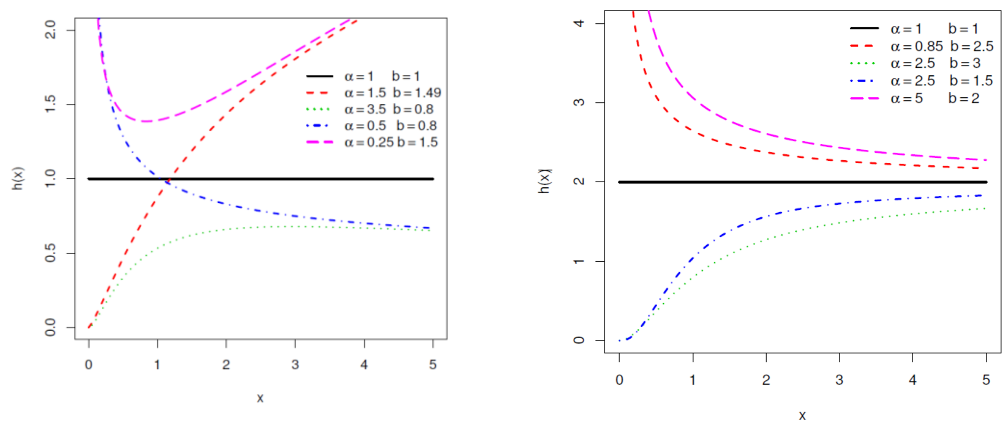


Figure 4: Plots of the (c) TLLc and (d) TLL hazard functions for some parameter values.

are given by

$$F(x) \sim [2G_R(x)]^{\alpha} \qquad as\ G(x) \to 0,$$
$$f(x) \sim 2\alpha gR(x)[2G_R(x)]^{\alpha-1} \qquad as\ G(x) \to 0,$$
$$h(x) \sim 2\alpha gR(x)[2G_R(x)]^{\alpha-1} \qquad as\ G(x) \to 0.$$

Proposition 4.2 *The asymptotics of equations (3), (4) and (5) when* $x \to \infty$ *are* given by

$$1 - F(x) \sim 2\alpha[1 - G_R(x)] \quad as \quad x \to \infty,$$
$$f(x) \sim 2\alpha gR(x) \quad as \quad x \to \infty,$$
$$\hbar(x) \sim \frac{gR(x)}{1 - G_R(x)} \quad as \quad x \to \infty.$$

4.2 Useful Expansions

We can demonstrate that the cdf (3) of X admits the expansion

$$F(x) = \sum_{k=0}^{\infty} b_k V_k(x) \qquad ,(6)$$

where

$$b_k = \sum_{i=i(k)}^{\infty} (-1)^{i+k} \binom{\alpha}{i}\binom{2i}{k} \qquad ,(7)$$

$I(k) = [\frac{k}{2}]$ when k is even and $i(k) = \left[\frac{k}{2}\right] + 1$ when k is odd, $\left[\frac{k}{2}\right]$ denotes the integer part of $\frac{k}{2}(for\ k \geq 0)$ and $V_k(x) = V_k(x;\xi) = G(x;\xi)^k$ represents the exp-G(k) cdf.

The density function of X can be expressed as

$$f(x;\alpha,\xi) = \sum_{k=0}^{\infty} b_{k+1}\, v_{k+1}(x) \qquad ,(8)$$

where $v_{k+1}(x) = dV_{k+1}(x)/dx$. Equation (8) reveals that the TLG density function is a linear combination of exp-G density functions. Thus, some mathematical properties of the new model can be derived from those properties of the exp-G distribution.

4.3 Moments

Let Y be a random variable with the baseline G distribution and let Y_{k+1} have the exp-G distribution with power parameter $k+1, i.e. Y_{k+1} \sim exp-G(k+1)$. So, $Y_1 = Y \sim G$. We define the (n, r)th probability weighted moment (PWM) of Y by $\tau_{n,r} = E[Y^n G(Y)^r]$, for n, r = 0, 1,…. The PWMs are usually evaluated numerically since they are available in closed-form for few distributions. The nth moment of X can be obtained from equation (8) and the exp-G moments as

$$\mu' = E(X^n) = \sum_{k=0}^{\infty} b_{k+1} E(Y_{K+1}^n) = \sum_{k=0}^{\infty} (k+1) b_{k+1}\, \tau_{n,k}. \qquad (9)$$

Equations (9) is the main result of this section and can be used to derive several TLG moments.

Then,the moments of the TLG family can also be determined as infinite weighted linear combinations of the PWMs of the G distribution.

We provide two simple examples. First, we take the Weibull baseline distribution with scale parameter a > 0 and shape parameter b > 0.The moments of the TLW model (presented in Section 3.1) follow easily from (9) and the moments of the exp-Weibull given by

$$E(Y_{k+1}^{n}) = \frac{(k+1)}{a^{n}}\Gamma(\frac{n+b}{b})\sum_{i=0}^{\infty}\frac{(-k)_{i}}{i!\,(i+1)^{(n+b)/b'}}$$

where $(-k)_i = (-1)^i k(k+1)\dots(k+1-i)$.

For the second example, we consider the TL-standard logistic (TLSL),where $G(x) = (1 + e^{-x}) - 1$. We can obtain $\tau_{n,k}(for\ t < 1)$ using Mathematica and then

$$\mu_n' = \sum_{k=0}^{\infty}(k+1)b_{k+1}(\frac{\partial}{\partial t})^{n}\ B(t+k+1,1-t)|_{t=0},$$

where $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$ is the beta fumction.

The central moments (μ_n) and cumulants (K_n) of X can be determined from (9) as

$$\mu_n = \sum_{k=0}^{n}(-1)^k\binom{n}{k}\mu_1'^{n}\mu_{n-k}' \quad and\ \ K_n = \mu_n' - \sum_{k+1}^{n-1}\binom{n-1}{k-1}K_k\mu_{n-k}'$$

respectively, where $K_1 = \mu_1'$. The skewness $\gamma_1 = K_3/K_2^{3/2}$ and kurtosis $\gamma_2 = K_4/K_2^2$ can be evaluated from the third and fourth standardized cumulants.

4.4 Generating function

Here, we obtain the moment generating function $(mgf)\ M(t) = E(e^{tx})$ of X.We can write M (t) from (8) as

$$M(t) = \sum_{k=0}^{\infty} b_{k+1}M_{K+1}(t) = \sum_{k=0}^{\infty}(k+1)b_{k+1}\rho(t,k) \qquad ,(10)$$

Where $M_{K+1}(t)$ is the mgf of Y_{k+1} and

$$\rho(t,k)\int_0^1 exp[tQ_G(u)]\,u^k\,du.$$

Hence, M (t) can be immediately determined from the exp-G generating function.

Next, we provide three application of equations (10) for special TLG distributions. The mgfs of the TL-exponential (TLE) (with rate parameter λ and t < λ^{-1}) and TLSL (defined in Section 4.3, for t < 1) distributions are given by

$$M(t) = \sum_{k=0}^{\infty}(k+1)b_{k+1}B(k+1,1-\lambda t)$$

and

$$M(t) = \sum_{k,j=0}^{\infty}(k+1)b_{k+1}B(t+1+k,1-t),$$

respectively. As a third example, we consider the cdf $G(x) = (1-e^{-x})^{\alpha}$ of the exponentiated unit exponential (EE) with power parameter $\alpha > 0$,and mgf given by $M_Y(t) = \Gamma(\alpha+1)\,\Gamma(1-t)/\Gamma(\alpha-t+1)\ for\ -1 < t < 1$. Then, the mgf of the TL-EE with unit parameter is given by

$$M(t) = \sum_{k=0}^{\infty} b_{k+1}\frac{(k+1)!\Gamma(1-t)}{\Gamma(k-t+2)}, \qquad -1<t<1.$$

4.5 Quantile Power Series

The qf is very useful to obtain various mathematical properties of a distribution and it is widely used in Statistics to generate values of a random variable having F (x) as its distribution function. By inverting F (x) = u in equation (3) with respect to x for some fixed u ∈ (0, 1), the qf of the TLG family is given by

$$Q_X(u) = Q_G[1-(1-u^{\frac{1}{\alpha}})^{\frac{1}{2}}], \qquad (11)$$

where $Q_G(u) = G^{-1}(u)$ denotes the qf of the baseline model.

We can derive some properties of the TLG family based on a power series for (11), which requires an expansion for the argument of $Q_G(\cdot)$, namely $z(u) = 1-\left(1-u^{\frac{1}{\alpha}}\right)^{\frac{1}{2}}$.

By using the generalized binomial expansion three times, we can write

$$z(u) = \sum_{k=0}^{\infty} \delta_k u^k$$

Where $\boldsymbol{\delta_0} = \mathbf{1} - \sum_{i,j=0}^{\infty} (-\mathbf{1})^{i+j} \binom{\frac{1}{2}}{i} \binom{\frac{1}{\alpha}}{j}$ and, for $k \geq 1$,

$$\delta_k = \sum_{i=0}^{\infty} \sum_{j=k}^{\infty} (-1)^{i+j+k+1} \binom{\frac{1}{2}}{i} \binom{\frac{1}{\alpha}}{j} \binom{j}{k}.$$

Then, the qf of X can be expressed as

$$Q_X(u) = Q_G\left(\sum_{t=0}^{\infty} \delta_k u^k\right). \qquad (12)$$

If Q_G(u) has a closed-form expression, we can derive a power series for Q_X(u). However, if Q_G(u) does not have a closed-form expression, it can be written in terms of a power series

$$Q_G(u) = \sum_{i=0}^{\infty} a_i u^i, \qquad (13)$$

where the coefficients $a_i's$ are functions of the parameters of the G distribution. For several important distributions, such as the normal, Student t, gamma and beta distributions, Q_G(u) does not have explicit expressions but it can be expanded as in equation (13).

Henceforth, we use a result by Gradshteyn and Ryzhik (2000) for a power series raised to a positive integer n (for n ≥ 1)

$$Q_G(u)^n = \left(\sum_{i=0}^{\infty} a_i u^i\right)^n = \sum_{i=0}^{\infty} c_{n,i}\, u^i, \qquad (14)$$

where the coefficients $c_{n,i}$ (for i = 1, 2, . . .) are obtained from the recurrence equation (with $c_{n,0} = a_0^n$)

$$c_{n,i} = (ia_0)^{-1} \sum_{m=1}^{i} [m(n+1) - i] a_m c_{n,i-m}. \qquad (15)$$

Clearly, the quantity $c_{n,i}$ can be determined from c_n,0,…, $c_{n,i}$ and then from the quantities $a_0, \ldots, a_i$.

For any baseline G distribution, we combine (12) and (13) to obtain

$$Q_X(u) = \sum_{i=0}^{\infty} a_i \left(\sum_{t=0}^{\infty} \delta_k u^k\right)^i,$$

and then using (14) and (15)

$$Q_X(u) = \sum_{t=0}^{\infty} e_t u^t, \qquad (16)$$

Where $e_0 = \sum_{i=0}^{\infty} a_i$, $e_t = \sum_{i=0}^{\infty} a_i d_{i,t} for\ t \geq 1, d_{i,0} = \delta_0^i$ and, for $t > 1, d_{i,t} = (t\delta_0)^{-1} \sum_{m=1}^{t} [m(i+1) - t]\delta_m\, d_{i,t-m}$.

Equation (16) is the main result of this section. It allows to derive various mathematical quantities for the TLG family. Let W (·) be any integrable function in a real line. We can write

$$\int_{-\infty}^{\infty} W(x) f(x; \alpha, \xi) dx = \int_0^1 W[Q_X(u)] du. \quad (17)$$

Thus, several mathematical quantities for special models of the TLG family can be reduced to integrals over (0,1) by combining (16) and (17).

4.6 Incomplete Moments

The nth incomplete moment of X is defined as $m_n(y) = E(X^n | X < y) = \int_{-\infty}^{y} x^r f(x) dx$. Here, we propose two methods to calculate the incomplete moments of the new family.First, the nth incomplete moment of X can be expressed as

$$m_n(y) = \sum_{k=0}^{\infty} b_{k+1} \int_0^{G(y;\xi)} Q_G(u)^n u^k du. \qquad (18)$$

The integral in (18) can be computed at least numerically for most baseline distributions. A second method to obtain the incomplete moments of X follows from (14), (15) and (18). We can write

$$m_n(y) = \sum_{k.m=0}^{\infty} \frac{(k+1) b_{k+1} c_{n,m}}{m+k+1} G(y; \xi)^{m+k+1}, \quad (19)$$

where the coefficients $c_{n,m}$ can be obtained recursively from (15).

4.7 Mean deviations

The amount of scatter of the TLG family can be measured by the totality of deviations from the mean and the median. The mean deviations about the mean (δ_1) and about the median (δ_2) of X

are given by

$$\delta_1 = E(|X - \mu_1'|) = 2\mu_1' F(\mu_1') - 2m_1(\mu_1')$$

and

$$\delta_2 = E(|X - M|) = \mu_1' - 2m_1(M),$$

respectively, where μ_1'= E(X), F(μ_1') is obtained from (3),M = Median(X) denotes the median evaluated from the nonlinear equation F (M) = 1/2, and $m_1(z) = \int_{-\infty}^{z} x\, f(x) dx$ is the basic quantity to determine δ_1 and δ_2 above. Setting u = G (x;ξ)∞in the linear representation (8) gives

$$T(z) = \sum_{r=0}^{\infty}(k+1)b_{k+1}T_k(z), \qquad (20)$$

where

$$T_k(z) = \int_0^{G(Z;\xi)} u^k Q_G(u)du.$$

Equation (20) is useful to obtain the Bonferroni and Lorenz curves defined (for a given probability π) by $B(\pi) = T(q)/(\pi\mu_1')$ and $L(\pi) = T(q)/\mu_1'$, respectively, where $\mu_1' = E(X)$ and $q = F^{-1}(\pi)$ is the qf of X given in Section 4.5.

5. Entropies

An entropy is a measure of variation or uncertainty of a random variable X.Two popular entropy measures are the R´enyi and Shannon entropies (R´enyi,1961;Shannon,1948).The R´enyi entropy of a random variable with pdf f (x) is defined as

$$I_R(\gamma) = \frac{1}{1-\gamma} log\left(\int_0^{\infty} f^{\gamma}(x)dx\right),$$

for $\gamma > 0$ and $\gamma \neq 1$. The Shannon entropy of a random variable X is defined by $\eta x = -E\{log[f(X)]\}$. It is the special case of the R´enyi entropy when $\gamma \uparrow 1$.

Lemma 4.1 Let X be a random variable with pdf (4). Then,

$$E\{log[1-G(X)]\} = \frac{\alpha}{2}\sum_{i=0}^{\infty}\frac{(-1)^{i+1}}{(i+1)^2}\binom{\alpha-1}{i},$$

$$E\{log\{1-[1-G(x;\xi]^2\}\} = \alpha\sum_{i=0}^{\infty}\frac{(-1)^i}{i+1}\frac{\partial}{\partial t}\binom{\alpha+t-1}{i}\Big|_{t=0}.$$

Proposition 4.3 If X follows the TLG family, then the Shannon entropy of X is given by

$$\eta x = -\log(\alpha) - \eta_g - \frac{\alpha}{2}\sum_{i=0}^{\infty}\frac{(-1)^{i+1}}{(i+1)^2}\binom{\alpha-1}{i} + \alpha(1-\alpha)\sum_{i=0}^{\infty}\frac{(-1)^i}{i+1}\frac{\partial}{\partial t}\binom{\alpha+t-1}{i}\Big|_{t=0},$$

where η_g is the parent Shannon entropy.

Proof. Direct calculation yields

$$\boldsymbol{\eta x = -\log(2\alpha) - \eta_g - E\{\log g[1-G(x;\xi)]\} + (1-\alpha)E\left[\log\{1-[1-G(x;\xi)]^2\}\right]}.$$

The rest of the proof follows from Lemma 4.1.

Proposition 4.4 If X follows the TLG family, then:

1.The R´enyi entropy can be written as

$$I_R(\gamma) = \frac{1}{1-\gamma}\log(2\alpha) + \frac{1}{1-\gamma}\log\left[\sum_{i,j=0}^{\infty} w_{i,j}A(\gamma,j)\right],$$

Where $w_{i,j} = (-1)^{i+j}\binom{\gamma(\alpha-1)}{i}\binom{\gamma+2i}{j}$ and $A(\gamma,j) = \int_0^{\infty} g(x)^{\gamma}G(x)^j\,dx.$

2.The q-entropy, $H_q(f) = \frac{1}{q-1}\log\left[1-\int_0^{\infty} f^q(x)dx\right]$ *, can be written as*

$$H_q(f) = \frac{1}{q-1}\left[1-\sum_{i,j=0}^{\infty} q_{i,j}A(q,j)\right],$$

Where $q_{i,j} = (2\alpha)^q(-1)^{i+j}\binom{q(\alpha-1)}{i}\binom{q+2i}{j}$.

6. Estimation

Several approaches for parameter estimation have been proposed in the literature but the maximum likelihood method is the most commonly employed.The maximum likelihood estimators (MLEs) enjoy desirable properties and can be used to construct confidence intervals for the model parameters and also in test statistics.The normal approximation for these estimators in large sample theory is easily handled either analytically or numerically.So,we consider the estimation of the unknown parameters for the new distribution from complete samples only by maximum likelihood.Let $x_1,\dots,x_n$ be the observed values from the TLG distribution with parameters α and ξ. Let $\Theta = (\alpha,\xi^T)^T$ be the $r \times 1$ parameter vector.The total log-likelihood function for Θ is given by

$$l_n = l_n(\Theta) = nlog(2\alpha) + \sum_{i=1}^{n} log[gR(x_i;\xi)] + \sum_{i=1}^{n} log[1 - G_R(x_i;\xi)]$$

$$+(\alpha-1)\sum_{i=1}^{n} log\{1-[1-G_R(x_i;\xi)]^2\}. (21)$$

The log-likelihood (21) can be maximized either directly by using the NMaximize command in Mathematica,R(optim function),SAS(PROC NLMIXED), Ox program (sub-routine MaxBFGS), or by solving the nonlinear likelihood equations obtained by differentiating (21).

The components of the score function $Un(\Theta) = (\partial l_n/\partial\alpha, \partial l_n/\partial\xi^T)^T$ are

$$\frac{\partial l_n}{\partial\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} log\{1-[1-G_R(x_i;\xi)]^2\}$$

And

$$\frac{\partial l_n}{\partial\xi} = \sum_{i=1}^{n}\frac{g_R^{(\xi)}(x_i,\xi)}{gR(x_i,\xi)} - \sum_{i=1}^{n}\frac{G_R^{(\xi)}(x_i,\xi)}{G_R(x_i,\xi)} + 2(\alpha-1)\sum_{i=1}^{n}\frac{G_R^{(\xi)}(x_i,\xi)[1-G_R(x_i,\xi)]}{1-[1-G_R(x_i;\xi)]^2} ,$$

where $h_R^{(\xi)}(\cdot)$ means the derivative vector of the function h with respect to ξ.

Then, the MLEs of the model parameters can be determined iteratively as the solution of the nonlinear equations U (Θ) = 0 leading to the maximum value for (21).

7. Simulations

We evaluate the performance of the maximum likelihood method for estimating the parameters of the new family by using Monte Carlo simulations. We choose the TLW model for this purpose and select a total of twelve parameter combinations. We fix four sample sizes n=20, 50, 100 and 300. The process is repeated 1,000 times and the biases (estimate minus true value) and the mean square errors (MSEs) of the parameter estimates are reported in Table 1. The values of the biases and MSEs decrease when the sample size n increases in agreement with first-order asymptotic theory. These results indicate that the maximum likelihood method performs quite well to estimate the parameters in Θ.

Table 1: Biases and MSEs for various parameter values.

Sample size	True value			Bias			MSE		
n	α	b	a	$\tilde{\alpha}$	$\tilde{b}$	$\tilde{a}$	$\tilde{\alpha}$	$\tilde{b}$	$\tilde{a}$
20	0.5	0.5	1.0	0.439	0.355	1.799	2.745	0.551	2.023
	0.5	1.5	2.0	0.637	0.963	-0.161	2.283	4.759	1.834
	1.5	0.5	1.0	1.031	0.282	0.122	2.618	0.825	2.477
	1.5	1.5	2.0	3.633	0.823	2.170	2.100	0.246	2.177
50	0.5	0.5	1.0	0.029	0.187	2.390	0.106	0.223	3.502
	0.5	1.5	2.0	0.116	0.392	-1.134	0.304	1.260	1.393
	1.5	0.5	1.0	1.228	0.106	3.508	1.611	0.179	2.310
	1.5	0.5	1.0	1.228	0.106	3.508	1.611	0.179	2.310
100	0.5	0.5	1.0	0.031	0.072	2.142	0.069	0.054	0.632
	0.5	1.5	2.0	0.033	0.239	-1.202	0.090	0.557	1.474
	1.5	0.5	1.0	0.343	0.046	1.629	1.947	0.037	1.323
	1.5	1.5	2.0	0.283	0.119	-1.190	1.552	0.227	1.454
300	0.5	0.5	1.0	0.007	0.018	0.306	0.016	0.009	0.476
	0.5	1.5	2.0	0.001	0.064	-0.231	0.015	0.079	0.520
	1.5	0.5	1.0	0.097	0.010	0.450	0.285	0.007	0.851
	1,5	1.5	2.0	0.127	0.016	-0.214	0.311	0.060	0.484

8. Applications

In this section, we provide applications of the TLG family by fitting some of its members to three real life data sets.

Data set 1: Carbon ftbres. The first data set (Crowder et al., 1991) refers to the failure stresses of single carbon fibres (length 1mm).The data are : 2.247, 2.64, 2.842, 2.908, 3.099, 3.126, 3.245, 3.328, 3.355, 3.383, 3.572, 3.581, 3.681,3.726, 3.727, 3.728, 3.783, 3.785, 3.786, 3.896, 3.912, 3.964, 4.05, 4.063, 4.082,4.111, 4.118, 4.141, 4.216, 4.251, 4.262, 4.326, 4.402, 4.457, 4.466, 4.519, 4.542,4.555, 4.614,4.632, 4.634, 4.636, 4.678, 4.698, 4.738, 4.832, 4.924, 5.043, 5.099, 5.134, 5.359, 5.473, 5.571, 5.684, 5.721, 5.998, 6.06. A summary of these data is: n = 57, $\bar{x}$ =4.2350, s = 0.8352, skewness=0.0710, kurtosis=2.7098.

Data set 2: Guinea pigs. The second data set consists of the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli reported by Bjerkedal (1960).The data are: 12, 15, 22, 24, 24, 32, 32, 33, 34, 38, 38, 43, 44, 48,52, 53, 54, 54, 55, 56, 57, 58, 58, 59, 60, 60, 60,60, 61, 62, 63, 65, 65, 67, 68,70, 70, 72, 73, 75, 76, 76, 81, 83, 84, 85, 87, 91, 95, 96, 98, 99, 109, 110, 121, 127,129, 131, 143, 146, 146, 175, 175,211, 233, 258, 258, 263, 297, 341, 341, 376.A summary of these data is: n = 72, $\bar{x}$ = 99.8200, s=81.1180, skewness=1.7590, kurtosis=5.4596.

Data set 3: Wheaton river. The third data set refers to the 72 exceedances for the years 1958–1984 (rounded to one decimal place) of flood peaks (in m^3/s) of the Wheaton River near Carcross in Yukon Territory, Canada.Bourguignon et al. (2014) have recently analyzed these data. The data are: 1.7, 2.2, 14.4,1.1, 0.4, 20.6, 5.3, 0.7, 1.9, 13.0, 12.0, 9.3, 1.4, 18.7, 8.5, 25.5, 11.6, 14.1, 22.1, 1.1,2.5, 14.4, 1.7, 37.6, 0.6, 2.2, 39.0, 0.3, 15.0, 11.0, 7.3, 22.9, 1.7, 0.1, 1.1, 0.6, 9.0,1.7, 7.0,20.1,0.4, 2.8, 14.1,9.9,10.4,10.7,30.0,3.6,5.6,30.8,13.3, 4.2, 25.5,3.4, 11.9, 21.5, 27.6, 36.4, 2.7, 64.0, 1.5, 2.5, 27.4, 1.0, 27.1, 20.2, 16.8, 5.3, 9.7,27.5, 2.5, 27.0. A summary of these data is:n = 72, $\bar{x}$=12.2041, s = 12.2972, skewness=1.4725, kurtosis=2.8895.

The Akaike information criterion(AIC),Bayesian information criterion(BIC), Anderson-Darling (A^*) and Kolmogrov-Smirnov (K-S) statistics are used to compare the fitted models. In general, the smaller the values of these statistics, the better the fit to the data.The required computations are carried out using the R-software.

Table 2: MLEs and their standard errors (in parentheses) for the carbon fibres data.

Distribution	α	b	a	c	s
TLW	2.6939	3.4614	0.2171	-	-
	(0.1737)	(0.2874)	(0.0261)	-	-
TLGa	0.4873	33.6042	0.1555	-	-
	(0.5566)	(3.8248)	(0.1324)	-	-
TLL	0.9655	-	-	7.4445	4.7819
	(0.5101)	-	-	(1.9498)	(0.3321)

Table 3: The statistics $\hat{l}$, AIC, BIC, A^* and K-S for the fitted models to the carbon fibres data.

Distribution	$\hat{l}$	AIC	BIC	A^*	K-S	p-value
TLW	−70.049	146.098	152.228	0.1580	0.0591	0.9817
TLGa	−70.066	146.133	152.263	0.1686	0.0616	0.9727
TLL	−70.349	146.698	152.827	0.1351	0.0490	0.9981

Table 4: MLEs and their standard errors (in parentheses) for guinea pigs data.

Distribution	α	b	a	c	s
TLW	3.4343	1.0381	0.0054	-	-
	(1.5162)	(0.2057)	(0.0011)	-	-
TLGa	0.1723	12.6371	29.2307	-	-
	(0.020)	(0.0041)	(0.0041)	-	-
TLL	1.0460	-	-	2.4305	218.9469
	(0.4285)	-	-	(0.4779)	(6.6943)

Table 5: The statistics $\hat{l}$, AIC, BIC, A^* and K-S for the fitted models to guinea pigs data.

Distribution	$\hat{l}$	AIC	BIC	A^*	K-S	p-value
TLW	−425.760	857.519	864.349	0.5254	0.0890	0.618
TLGa	−428.575	863.151	869.981	1.2137	0.0986	0.4854
TLL	−425.144	856.288	863.118	0.4361	0.0795	0.7533

Table 6: MLEs and their standard errors (in parentheses) for the Wheaton river data.

Distribution	α	b	a	c	s
TLW	0.4961	1.4337	0.0308	-	-
	(0.2717)	(0.5568)	(0.0066)	-	-
TLGa	0.1065	6.8109	6.2596	-	-
	(0.0125)	(0.0025)	(0.1091)	-	-
TLL	0.1997	-	-	3.1532	34.5107
	(0.0931)	-	-	(1.1828)	(4.9361)

Table 7: The statistics $\hat{l}$, AIC, BIC, A^* and K-S for the fitted models to the Wheaton river data.

Distribution	$\hat{l}$	AIC	BIC	A^*	K-S	p-value
TLW	−251.055	508.109	514.939	0.6424	0.1039	0.4190
TLGa	−250.521	507.042	513.872	0.5641	0.1054	0.4002
TLL	−251.187	508.373	515.203	0.6434	0.1112	0.3354

Tables 2, 4 and 6 give the MLEs and their corresponding standard errors (in parentheses) of the model parameters.The values of the statistics $\hat{l}$, AIC, BIC, A^*, K-S and K-S p-values are listed in Table 3, 5 and 7. For the first data set, we note that the three fitted models provide good fits. For the second data set, the TLW and TLL models provide the best fits.Finally, for the third data set, the TLW and TLGa models provide more adequate fits. The histogram and estimated cdfs are displayed in Figure 3, 4 and 5, which also support the results in Table 3, 5 and 7.

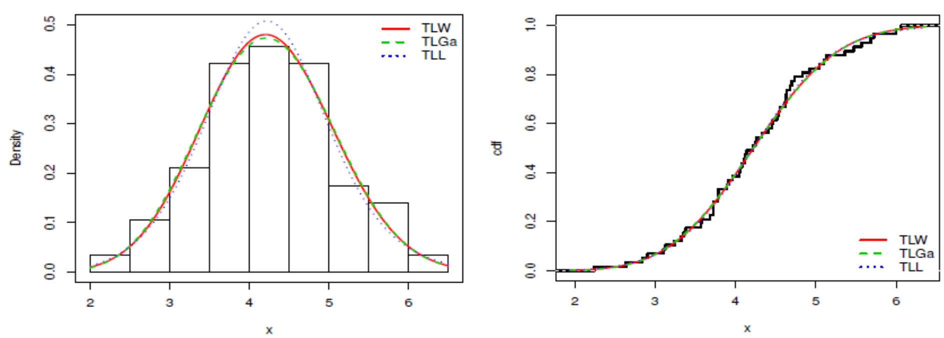
9. Concluding Remarks

In this paper, we propose the Topp-Leone-G family. We study some mathematical properties of the new family including a linear representation for the density function, explicit expressions for the ordinary and incomplete moments, generating and quantile functions, mean deviations and entropies. The maximum likelihood method is employed to estimate the model parameters. We fit three members of the family to real data sets to demonstrate the usefulness of the new family. These special models provide adequate fits to the data sets.

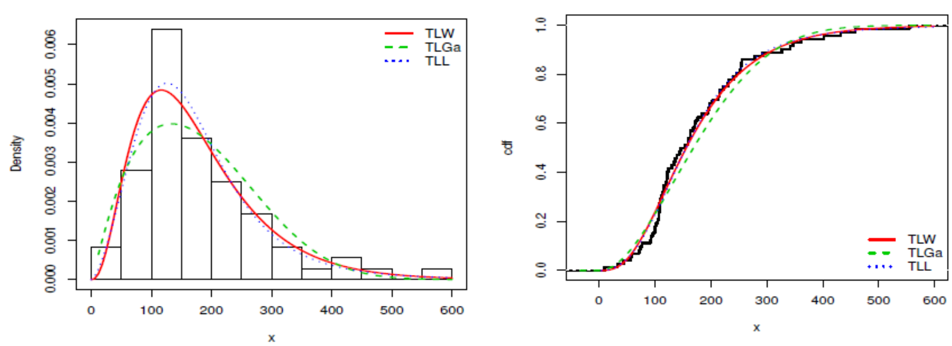
Acknowledgments

The authors are grateful to the Editor and the reviewer for the comments and suggestions that have helped in improving the paper.

(a) Estimated pdfs for data set 1 (b) Estimated cdfs for data set 1.



(c) Estimated pdfs for data set 2 (d) Estimated cdfs for data set 2.



(e) Estimated pdfs for data set 3 (f) Estimated cdfs for data set 3.

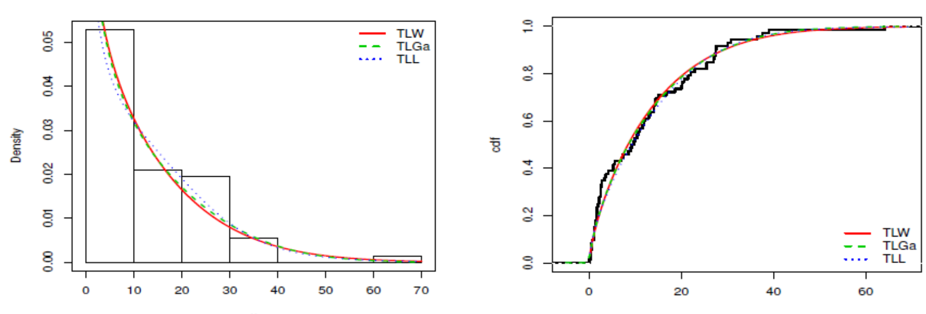


Figure 5: Plots of the estimated pdfs and cdfs of the TLW, TLGa, and TLL models for data sets 1, 2 & 3.

References

[1] Aljarrah, M.A., Lee, C. and Famoye, F. (2014).On generating T-X family of distributions using quantile functions. Journal of Statistical Distributions and Applications 1, Article 2.

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

[3] Alzaatreh, A., Lee, C. and Famoye, F. (2014). T-normal family of distributions:A new approach to generalize the normal distribution. Journal of Statistical Distributions and Applications 1, Article 16.

[4] Alexander, C. Cordeiro, G.M., Ortega, E.M.M. and Sarabia, J.M. (2012). Gen-eralized beta-generated distributions. Computational Statistics and Data Analysis 56, 1880–1897.

[5] Bjerkedal, T. (1960). Acquisition of resistance in Guinea pigs infected with different doses of virulent tubercle bacilli. American Journal of Hygiene 72, 130–148.

[6] Bourguignon, M., Silva, R.B. and Cordeiro, G.M. (2014). The Weibull–G family of probability distributions. Journal of Data Science 12, 53–68.

[7] Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distri- butions. Journal of Statistical Computation and Simulation 81, 883–893.

[8] Crowder, M.J., Kimber, A.C., Smith, R.L. and Sweeting, T.J. (1991). The Statistical Analysis of Reliability Data. Chapman and Hall, London.

[9] Eugene, N., Lee, C. and Famoye, F.(2002). Beta-normal distribution and its applications. Communications in Statistics–Theory and Methods 31, 497– 512.

[10] Gradshteyn, I.S. and Ryzhik, I.M. (2000). Table of Integrals, Series, and Products.Sixth edition, Academic Press, San Diego.

[11] Gupta, R.C., Gupta, P.I. and Gupta, R.D. (1998). Modeling failure time data by Lehmann alternatives. Communications in Statistics–Theory and Methods 27, 887–904.

[12] Gupta, R.D. and Kundu, D. (1999). Generalized exponential distribution. Australian and New Zealand Journal of Statistics 41, 173–188.

[13] Gupta, R.D. and Kundu, D. (2001). Generalized exponential distribution: Dif- ferent methods of estimations.Journal of Statistical Computation and Simulation 69, 315–337.

[14] Jones, M.C. (2004). Families of distributions arising from the distributions of order statistics. Test 13, 1–43.

[15] Lee, C., Famoye, F. and Alzaatreh, A. (2013). Methods for generating families of continuous distribution in the recent decades. WIRs: Computational Statistics 5, 219–238.

[16] Mudholkar, G.S. and Hutson, A.D. (1996).The exponentiated Weibull family:Some properties and a flood data application. Communications in Statistics–Theory and Methods 25, 3059–3083.

[17] Mudholkar, G.S. and Srivastava,D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure data. IEEE Transactions on Reliability 42, 299–302.

[18] Mudholkar, G.S., Srivastava, D.K. and Freimer, M. (1996). The exponentiated Weibull family: A reanalysis of the bus-motor failure data. Technometrics 37, 436–445.

[19] Nadarajah, S. (2011). The exponentiated exponential distribution: a survey.AStA Advances in Statistical Analysis 95, 219–251.

[20] Nadarajah, S., Cordeiro, G.M. and Ortega, E.M.M. (2015).The Zografos-Balakrishnan–G family of distributions: Mathematical properties and ap- plications.Communications in Statistics–Theory and Methods 44, 186–215.

[21] Nadarajah, S. and Kotz, S. (2003). Moments of some J-shaped distributions.Journal of Applied Statistics 30, 311–317.

[22] Nadarajah, S. and Kotz, S.(2006). The exponentiated type distributions. Acta Applicanda Mathematica 92, 97–111.

[23] R´enyi, A. (1961). On measures of entropy and information. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability- I, University of California Press, Berkeley, Volume 1, pp. 547–561.

[24] Shannon, C.E (1948). A mathematical theory of communication. Bell Systems Technical Journal 27, 379–432.

[25] Topp, C.W. and Leone, F.C. (1955). A family of J-shaped frequency functions.Journal of the American Statistical Association 50, 209–219.

Received February 10, 2016; accepted ???? ??, ????.

M. H. Tahir
Department of Statistics, Baghdad-ul-Jadeed Campus
The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
email: mtahir.stat@gmail.com

Gauss M. Cordeiro Department of Statistics
Federal University of Pernambuco, 50740-540, Recife, PE, Brazil

email: gausscordeiro@gmail.com

Muhammad Mansoor
Department of Statistics, Baghdad-ul-Jadeed Campus
The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
email: mansoor.abbasi143@gmail.com
& Department of Statistics Government Degree College Liaquat Pur, Pakistan

Ayman Alzaatreh
Department of Mathematics and Statistics American University of Sharjah
PO Box 26666, Sharjah United Arab Emirates
email: aalzaatreh@aus.edu

Muhammad Zubair Department of Statistics Government S.E. College Bahawalpur-63100
Pakistan
email: zubair.stat@yahoo.com