Austrian Journal of Statistics
January 2023, Volume 52, 116–140.
http://www.ajs.or.at/
doi:10.17713/ajs.v52i2.1407

A Flexible Probability Model for Proportion Data: Unit Gumbel Type-II Distribution, Development, Properties, Different Method of Estimations and Applications

Anum Shafiq
School of Mathematics and Statistics

Tabassum Naz Sindhu
Quaid-i-Azam University

Zawar Hussain
Cholistan University of Veterinary & Animal Sciences

Josmar Mazucheli
Universidade Estadual de Maringá

Bruna Alves
Universidade Estadual de Maringá

Abstract

In explaining and forecasting real life scenarios, statistical distributions are very helpful. It is very important to select the best fitting statistical distribution for modelling data. In analysis of real world phenomena like in reliability and economics, we may find distributions for bounded data observed as percentages, proportions or fractions (see, for example, Marshall and Olkin (2007)). In this context, in view of pertinent transformation on the Gumbel Type-II model, we suggest and study the unit Gumbel Type-II (UG-TII) model and explore few of its statistical characteristics. We also consider various methods of estimating the unknown parameters of UG-TII model from the frequentist perspective. Monte Carlo simulations are worked out in order to compare efficiency of suggested estimation methods for small as well as large samples. The efficiency of estimators is measured using simulated samples in terms of their bias and mean square error. In the end, two datasets have been examined in attempt to validate the realistic possibilities of new model. In comparison to the six severe competitors.

Keywords: UG-TII distribution, estimation methods, Cramér-von Mises estimates, right tail Anderson Darling estimates, likelihood..

1. Introduction

A typical challenge in applied statistics is concerned with the uncertainty dynamics found within the interval (0, 1). In real life, we frequently experience measurements such as proportion/fraction of a particular characteristic, ratings of certain ability evaluations, various indices and rates, in the interval, that lies between (0, 1) (see, for instance, Gupta and Nadarajah (2004); Hunger, Baumert, and Holle (2011); Kieschnick and McCullough (2003) among others studies). In this regard, Papke and Wooldridge (1996) noticed that variables within zero and one normally exist in many economic contexts, like the proportion of the overall fortnightly hours concentrated on jobs. The percentage of money spent on non-durable consumption, the membership rate of the pension scheme, the market share of the sector, television ratings, the fraction of the land allotted to cultivation, etc. There is finite support for the distribution of UG-TII, and in reliability and lifespan studies Often data sets are modelled using finite support distributions (e.g. Barlow and Proschan (1975)). In particular, the UG-TII distribution can be well applicable if reliability is evaluated as ratio of the number of effective trials to the number of total trials. These distributions have applications in many domains in stress-strength analysis. If Y denotes the radius of the base of a little mug and X denotes the radius of the circular depression in the centre of a plate then P denotes the chance of having the cup. Another example, a market research agency may like to compare two products' sales percentages with a separate promotional strategy on each one. In these scenarios, it is more appropriate to use a distribution of bounded support than a distribution with unbounded support.

In this article, we present a two parameter UG-TII distribution, derived from a transformation on the Gumbel distribution. The key benefit of the UG-TII distribution is that experts would have a novel, very flexible, unimodal two parameter model that has some critical characteristics that are not enjoyed by other distributions restricted to domain $(0, 1)$. The challenge of parameter estimation for several distributions is very vital. Common estimation methods like the MLE, MME, LSE and WLSE are often chosen for parameter estimation. Each one has its own positives and negatives however the most popular technique of estimation is the maximum likelihood estimation technique. In terms of determining the parameters of any distribution due to its desirable properties, maximum likelihood estimation (MLE) is normally a starting point. They are, for instance, asymptotically unbiased, consistent, and asymptotically normally distributed Lehmann (1999). However, there are other estimation techniques formulated over time for other distributions. In the last few years there are several ample articles on parameter estimation for different distributions in literature. Ramos and Louzada (2016) also developed a new distribution named as the generalized weighted lindley distribution and considered various techniques of assessment for this distribution. Tahir, Cordeiro, Ali, Dey, and Manzoor (2018) have examined the techniques of assessment of parameters for Nadarajah and Haghighi distribution. Dey, Mazucheli, and Nadarajah (2018) have analyzed estimation approaches for KSW distribution. Ramos, Louzada, Ramos, and Dey (2020) have viewed problem of estimation of parameters for Frechet distribution. Loganathan and Uma (2017) evaluated the MLE, LSE, WLSE and MME and outlined that WLSE yielded similar outcomes. The parameter estimation for the complementary Beta distribution considering the L-moments and maximum probability methods was studied by Mazucheli and Menezes (2019). The maximum likelihood and maximum product spacing methods were used by Almetwally and Almongy (2019) to estimate parameters of generalized power Weibull model. Sindhu, Shafiq, and Al-Mdallal (2021b) used MLE approach to estimate the new class of model. Sindhu, Shafiq, and Al-Mdallal (2021a) presented a new three parametric distribution known as Exponentiated transformation of Gumbel Type-II model and use MLE approach to estimate parameters of model.

In this study, we characterize a novel two-parameter lifetime model by incorporating Gumbel Type II distribution. The goal of current article is to incorporate different evaluation methodologies for the unknown UG-TII model indexing parameters and to analyze the behaviour of these estimators for different sample sizes and various parameter values. Apart from the above mentioned techniques, we study six methods to estimate the parameters of UG-TII model. In particular, we discuss ML, LS, WLS, CVM, AD and RTAD estimators. The ultimate motivation of the current study is to explain how various frequentist estimators of this model are used and execute for varying sample sizes and different parameter values and to provide a framework for selecting suitable estimation technique for UG-TII model. The present research is different in that no effort has been made in this direction to date. This research is focused to both its analytical and practical characteristics, with a concentration on the implemented side.

2. UG-TII specifications

Over the years, especially in extreme value analysis of extreme events, the Gumbel distribution, also named as the type 1 extreme value model, has got considerable research interest. For an examination of the latest developments of Gumbel model and its uses, see Pinheiro and Ferrari (2016). Its PDF and CDF are specified as

$$f(x;\Theta) = \nu\mu x^{-\mu-1}\exp\left(-\nu x^{-\mu}\right) \tag{1}$$

and

$$F(x;\Theta) = \exp\left(-\nu x^{-\mu}\right) \tag{2}$$

where $x > 0$, and $\Theta = (\nu, \mu)$. Here $\mu > 0$ is the shape parameter and $\nu > 0$. Observe that ν is neither a rate nor a scale parameter — $F(x;\nu,\mu) \neq F(\nu x;1,\mu)$ and $F(x;\nu,\mu) \neq F(\nu/x;1,\mu)$. From (1) we develop a new distribution utilizing the transformation $Z = X/(1+X)$ with unit-interval support. The PDF and CDF of extracted distribution are given, accordingly, by

$$g(z;\Theta) = \mu\nu\frac{z^{-\mu-1}}{(1-z)^{1-\mu}}\exp\left[-\nu\left(\frac{z}{1-z}\right)^{-\mu}\right] \tag{3}$$

and

$$G(z;\Theta) = \exp\left[-\nu\left(\frac{z}{1-z}\right)^{-\mu}\right]. \tag{4}$$

The HRF i.e. $h(z) = g(z;\ \Theta)/[1 - G(z;\ \Theta)]$ is an ideal mechanism in reliability study. Reliability or survival function $S(z)$ is an indicator of the capability of equipment to work without failure when placed into operation and it is a non-increasing function. Here, the $S(z)$ and $h(z)$ functions of the UG-TII (Θ) distribution are respectively,

$$S(z;\Theta) = 1 - \exp\left[-\nu\left(\frac{z}{1-z}\right)^{-\mu}\right] \tag{5}$$

and

$$h(z;\Theta) = \frac{\mu\nu\frac{z^{-\mu-1}}{(1-z)^{1-\mu}}\exp\left[-\nu\left(\frac{z}{1-z}\right)^{-\mu}\right]}{1-\exp\left[-\nu\left(\frac{z}{1-z}\right)^{-\mu}\right]}. \tag{6}$$

The CHRF is also called the integrated hrf. The CHRF is not a probability. It is also, however, a measure of risk: the higher the $H(t)$ value, the higher the risk of failure by t-time.

$$H(t) = \int_0^t h(z;\Theta)dz = -\log\left[S(t)\right]. \tag{7}$$

It is noted that

$$S(z) = e^{-H(z)} \text{ and } f(z) = h(z)\,e^{-H(z)}. \tag{8}$$

Therefore,

$$H(z) = -\log\left\{1 - \exp\left[-\nu\left(\frac{z}{1-z}\right)^{-\mu}\right]\right\}. \tag{9}$$

Just in Figs. 1 and 2 the above-mentioned PDF and HRF demonstrate how the parameters (Θ) affect the density of UG-TII(Θ) model. We would have to note that the values for

Θ parameters have indeed been chosen arbitrarily till we captured a wide range of shapes for the parameters concerned. We note that the PDF is right-skewed or reversed-J shaped, symmetrical and U formed. It is U-shaped $\mu = 0.2$ along with various choices of ν and inverted U-shaped when $\mu \geq 1$. Fig. 2 provides flexible hazard rate shapes such as increasing, inverted U-shaped and U shaped. Data that can be modelled by UG-TII(Θ) model are percentage data, rates, particle sizes and certain chemical processes.

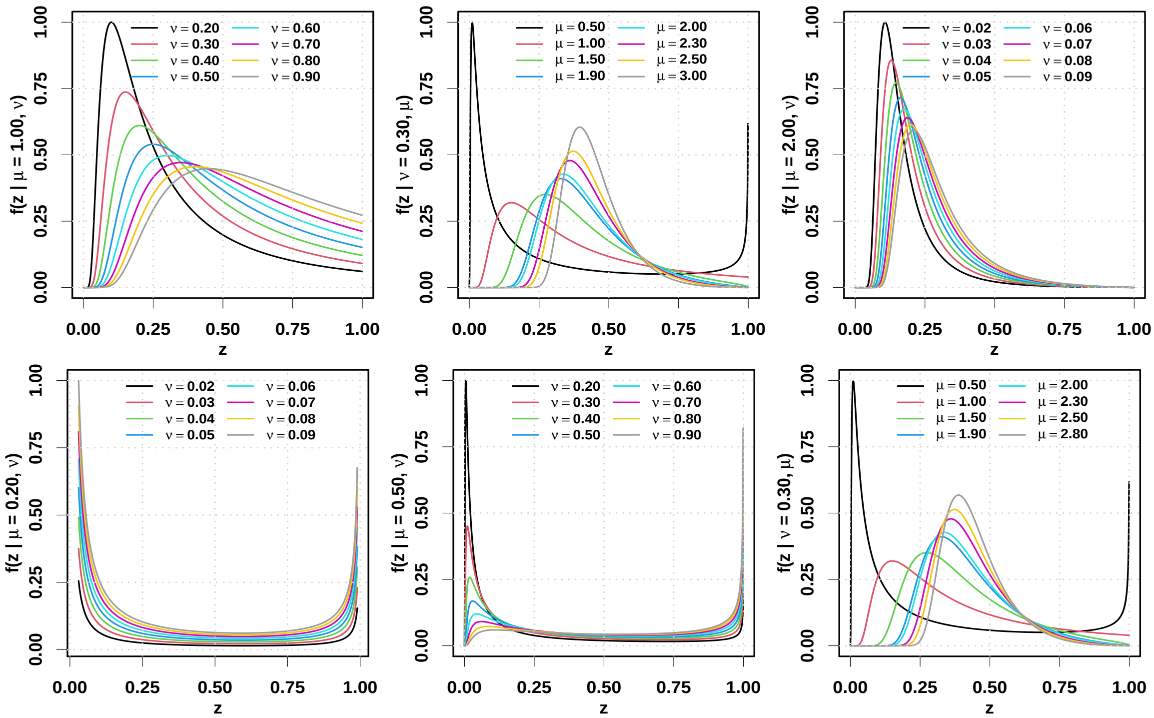

Figure 1: Variations of $g(z;\Theta)$ of UG-TII(Θ) along with μ and ν

These maps show that in modelling various data sets of different forms, the UG-TII(Θ) distribution is very useful.

2.1. Another form of PDF and CDF

Here,we have alternate representations for the PDF and CDF of the UG-TII(Θ) distribution.

Proposition 1. *An alternate representation of PDF is as follows*

$$g(z;\ \Theta) = \sum_{j=k=0}^{+\infty} B_{jk} z^{k-\mu(1+j)-1}, \tag{10}$$

where $B_{jk} = \mu\nu^{j+1} (-1)^{j+k} \binom{\mu(1+j)-1}{k}$.

Proof. Using $|x| < 1$ and $s \in \mathsf{R}^+$, we have $(1-x)^{s-1} = \sum_{i=0}^{+\infty} (-1)^i \binom{s-1}{i} x^i$.

By using exponential series on $\exp\left[-\nu\left(\frac{z}{1-z}\right)^{-\mu}\right]$, we can write

$$\exp\left[-\nu\left(\frac{z}{1-z}\right)^{-\mu}\right] = \sum_{j=0}^{+\infty} (-1)^j \nu^j z^{-\mu j} (1-z)^{\mu j}. \tag{11}$$

Also $[1-z]^{\mu(1+j)-1} \in (0,1)$, therefore using the above expansion provides,

$$[1-z]^{\mu(1+j)-1} = \sum_{k=0}^{+\infty} (-1)^k \binom{\mu(1+j)-1}{k} z^k. \tag{12}$$

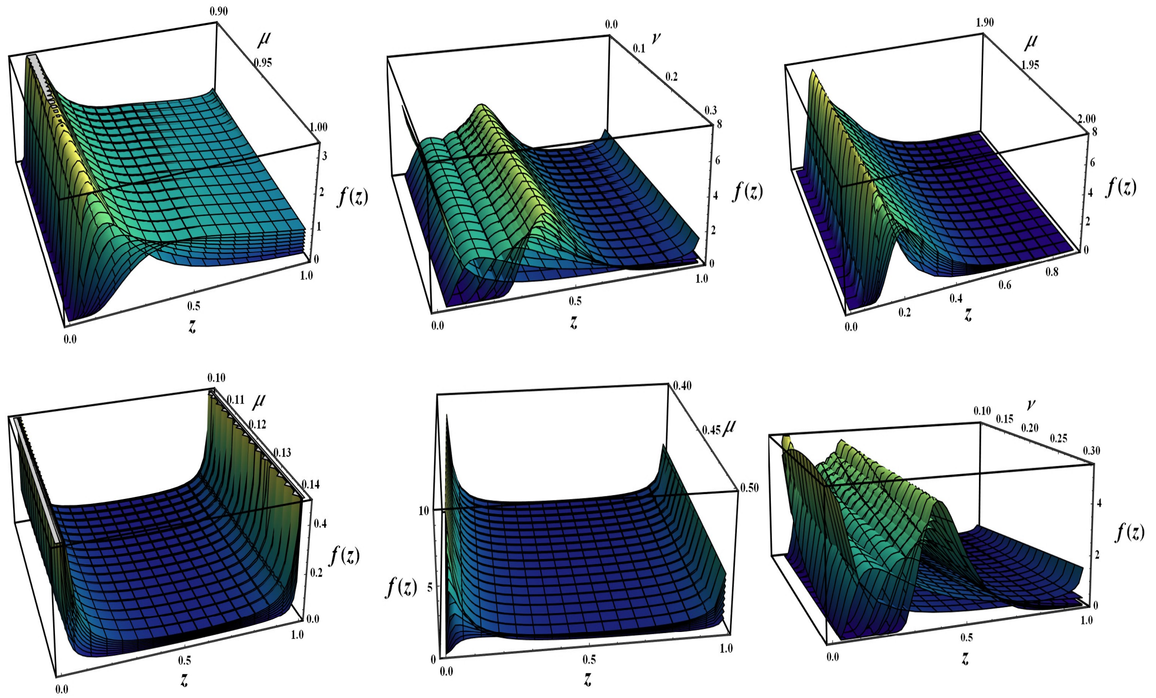

Figure 2: 3D fluctuations of $g(z;\Theta)$ of UG-TII(Θ) along with μ and ν.

By incorporating together the above equations, we have

$$g(z;\ \Theta) = \sum_{j=k=0}^{+\infty} B_{jk} z^{k-\mu(1+j)-1}.$$

This concludes the confirmation of Proposition 1. □

Remark 1. *An alternative form of CDF of the UG-TII*(Θ) *is represented by Eq.* (11).

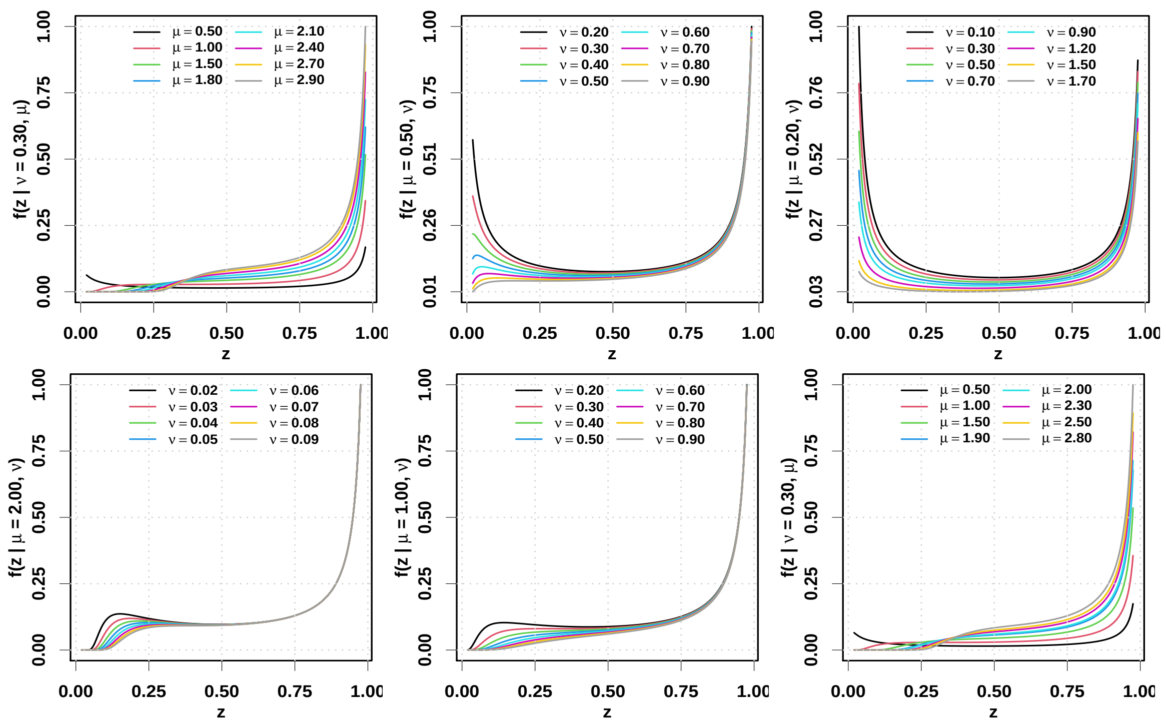

Figure 3: Fluctuations of $h(z;\Theta)$ of UG-TII(Θ) distribution along with μ and ν

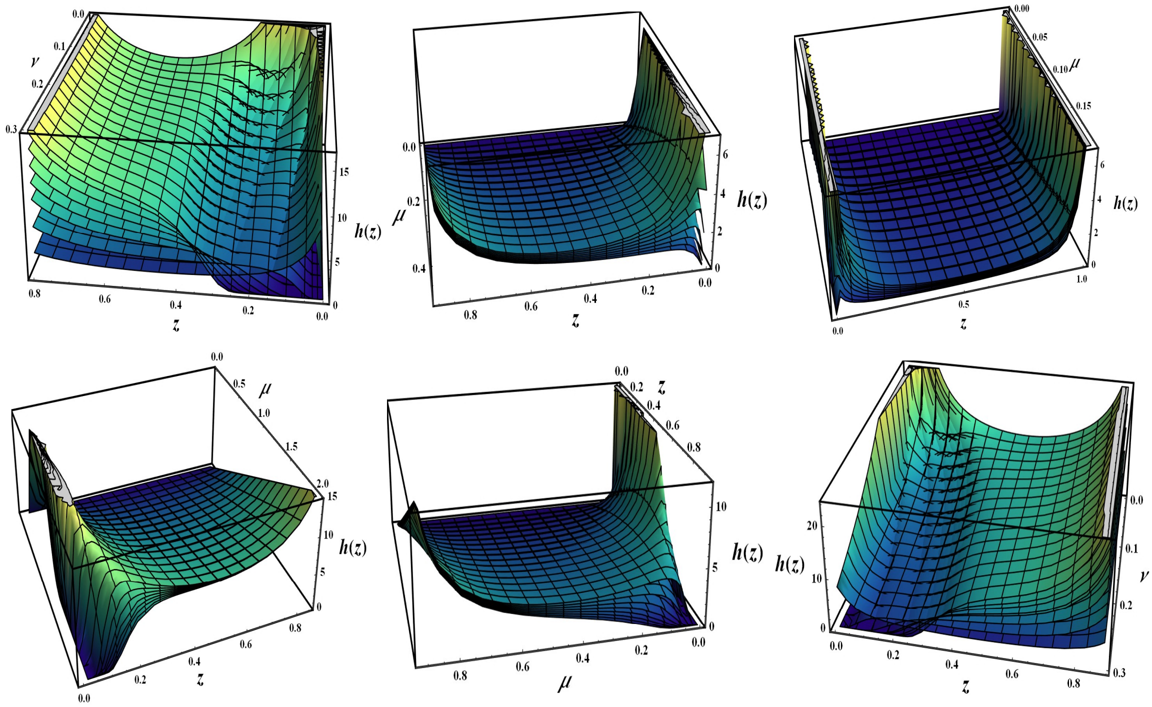

Figure 4: 3D fluctuations of $h(z;\Theta)$ of UG-TII(Θ) distribution along with μ and ν

Table 1: Mean, variance, $\acute{\mu}_2$, $\acute{\mu}_3$, $\acute{\mu}_4$, coefficient of skewness and kurtosis of UG-TII(Θ) distribution for arbitrary values of μ and ν

$(\mu,\nu)\downarrow$	Mean	Variance	$\acute{\mu}_2$	$\acute{\mu}_3$	$\acute{\mu}_4$	S_Θ	K_Θ
(2, 2)	0.64	0.09	0.42	0.29	0.21	0.07	1.19
(2, 3)	0.68	0.01	0.48	0.34	0.25	0.05	1.19
(2, 4)	0.71	0.01	0.52	0.38	0.29	0.04	1.19
(2, 6)	0.75	0.01	0.57	0.44	0.35	0.02	1.18
(3, 3)	0.63	0.03	0.40	0.27	0.18	0.09	1.23
(3, 4)	0.65	0.15	0.43	0.29	0.20	0.08	1.23
(3, 6)	0.68	0.01	0.47	0.33	0.23	0.07	1.22
(3, 8)	0.70	0.01	0.50	0.36	0.26	0.07	1.22
(4, 3)	0.60	0.04	0.37	0.23	0.14	0.10	1.25
(5, 3)	0.58	0.03	0.34	0.20	0.12	0.11	1.26
(6, 3)	0.57	0.02	0.33	0.19	0.11	0.11	1.27
(8, 3)	0.55	0.02	0.31	0.17	0.10	0.11	1.27

The values of mean, variance, $\acute{\mu}_2$, $\acute{\mu}_3$, $\acute{\mu}_4$, coefficient of skewness (S_Θ) and kurtosis (K_Θ) of UG-TII(Θ) distribution for arbitrary values of μ and ν are given in Table 1. For the fixed levels of μ, it can be observed that, with the increasing trend of the ν, the mean, variance, S_Θ and K_Θ decrease, also for the fixed levels of ν, these measures decrease with the few exceptions. In all cases of consideration platykurtic and positively skewed behavior of UG-TII(Θ) distribution are noticed.

3. Mathematical and statistical features

3.1. Quantile function

The next result can be utilized to simulat values from the UG-TII(Θ) distribution. The QF of Z is given by

$$Q(q;\Theta)=\frac{\left(-\frac{1}{\nu}\log q\right)^{\frac{-1}{\mu}}}{\left[1+\left(-\frac{1}{\nu}\log q\right)^{\frac{-1}{\mu}}\right]},\ 0<q<1, \tag{13}$$

were $\tilde{Z}=Q(0.5;\Theta)$ gives the median of Z. The other partition values can be explained in a same way. In specific, by putting $q=(0.25,0.75)$ in Eq. (13), the first, and third quartiles are attained. The accompanying quantile density function is provided by the differentiation of

$$Q'(q;\Theta)=-\frac{\left(-\frac{1}{\nu}\log q\right)^{\frac{-1}{\mu}}}{q\mu(\log q)\left[1+\left(-\frac{1}{\nu}\log q\right)^{\frac{-1}{\mu}}\right]^2}. \tag{14}$$

Based on partition measures, the assessment of the variability of the skewness and kurtosis of Z can be studied. The Bowley skewness is

$$S_\Theta=\frac{Q\left(0.75;\Theta\right)-2Q\left(0.5;\Theta\right)+Q\left(0.25;\Theta\right)}{IQR}, \tag{15}$$

and the Moors' kurtosis is

$$K_\Theta=\frac{Q\left(0.875;\Theta\right)-Q\left(0.625;\Theta\right)+Q\left(0.375;\Theta\right)-Q\left(0.125;\Theta\right)}{IQR}. \tag{16}$$

At different values of the distribution parameters ν and μ, Fig. 5 provides maps of the median, skewness and kurtosis. The distribution is found to be positively skewed and leptokurtic to platykurtic in nature. As the higher inputs of the parameter ν and μ contribute the higher change in median curve. On the other hand significant change in the skewness behavior is noticed along ν for smaller values of μ. The variation of μ does not contribute any significant change in the behavior of kurtosis as does the parameter ν. Also median yields lower values when μ is less than 3, but as μ and ν increase, it ends up to approximately 1.

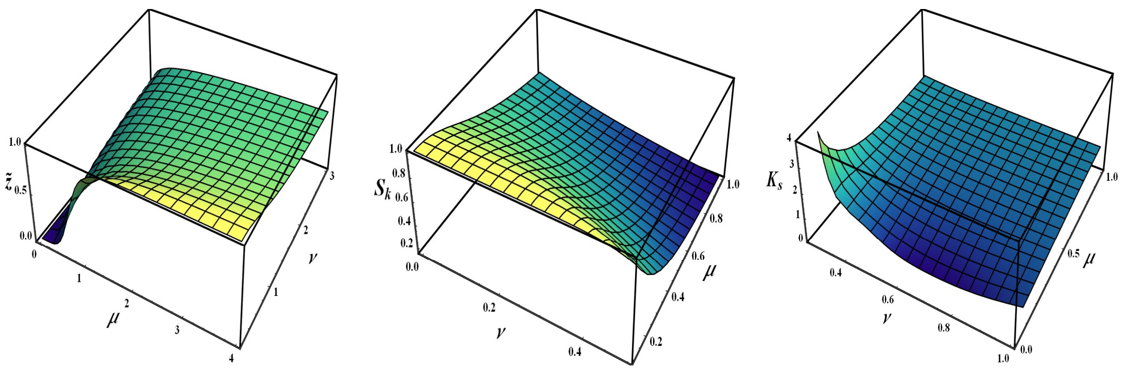

Figure 5: Fluctuation of $\tilde{Z}$, skewness and kurtosis of UG-TII(Θ) along with μ and ν

3.2. Ordinary moments

We hardly require to moments in any data analysis particularly in applied study. Some of the most essential features and characteristics of a model can be analyzed through moments including, tendency, dispersion, skewness and kurtosis. For $\Theta>0$, using the description of the r^{th} moment, $r = 1, 2, ...$, of the random variable Z and replacing Eq. (3), we obtain the UG-TII(Θ) moments as

$$\acute{\mu}_r = E\left(Z^r\right) = \int_0^1 z^r dF_Z(z;\ \Theta),$$

$$\acute{\mu}_r = \int_0^1 z^r \mu\nu \frac{z^{-\mu-1}}{(1-z)^{1-\mu}} \exp\left[-\nu\left(\frac{z}{1-z}\right)^{-\mu}\right] dz. \tag{17}$$

Let $\left(\frac{z}{1-z}\right)^{-\mu} = t$, then $-\mu\left(\frac{z^{-\mu-1}}{(1-z)^{1-\mu}}\right)dz = dt$ and after some algebraic manipulation, we have

$$\acute{\mu}_r = \nu \int_0^\infty \frac{t^{\frac{-r}{\mu}}}{\left(1+t^{\frac{-1}{\mu}}\right)^r} \exp\{-\nu t\}\, dt = \nu \int_0^\infty \frac{1}{\left(1+t^{\frac{1}{\mu}}\right)^r} \exp\{-\nu t\}\, dt.$$

After some algebra, we have

$$\acute{\mu}_r = \sum_{k=0}^{+\infty} (-1)^k \nu \binom{r+k-1}{k} \frac{\Gamma(1+k/\mu)}{\nu^{(1+k/\mu)}}, \text{ if } \operatorname{Re}\left(\frac{k}{\mu}\right) > -1, \quad \operatorname{Re}(\nu) > 0. \tag{18}$$

The mean of Z can be obtained by putting $r = 1$ in Eq. (18) i.e.,

$$\mu_z = \sum_{k=0}^{+\infty} (-1)^k \nu \frac{\Gamma(1+k/\mu)}{\nu^{(1+k/\mu)}}, \text{ if } \operatorname{Re}\left(\frac{k}{\mu}\right) > -1, \quad \operatorname{Re}(\nu) > 0.$$

The variance of Z $\left(\sigma_z^2\right)$ can be expressed as

$$\begin{aligned}\sigma_z^2 &= \sum_{j=k=0}^{+\infty} \sum_{k=0}^{+\infty} (-1)^k \nu \binom{k+1}{k} \frac{\Gamma(1+k/\mu)}{\nu^{(1+k/\mu)}} \\ &\quad - \left(\sum_{k=0}^{+\infty} (-1)^k \nu \frac{\Gamma(1+k/\mu)}{\nu^{(1+k/\mu)}}\right)^2, \text{ if } \operatorname{Re}\left(\frac{k}{\mu}\right) > -1, \quad \operatorname{Re}(\nu) > 0.\end{aligned} \tag{19}$$

Table 2: Real and approximated values of first four ordinary moments of UG-TII(Θ) distribution for arbitrary values of μ and ν

(μ, ν)	$\acute{\mu}1\|_{Real}$	$\acute{\mu}1\|_{Approx.}$	$\acute{\mu}2\|_{Real}$	$\acute{\mu}2\|_{Approx.}$	$\acute{\mu}3\|_{Real}$	$\acute{\mu}3\|_{Approx.}$	$\acute{\mu}4\|_{Real}$	$\acute{\mu}4\|_{Approx.}$
$(2, 0.5)$	0.48	0.48	0.25	0.25	0.14	0.14	0.09	0.09
$(3, 0.2)$	0.42	0.42	0.18	0.18	0.09	0.09	0.04	0.04
$(4, 0.9)$	0.53	0.53	0.28	0.28	0.16	0.16	0.09	0.09
$(5, 2.0)$	0.56	0.56	0.32	0.32	0.18	0.18	0.11	0.11
$(6, 3.0)$	0.57	0.57	0.33	0.33	0.19	0.19	0.11	0.11

In Table 2, for comparing the approximate expression of $\acute{\mu}_r$ given in Eq. (18) with the direct result of $\acute{\mu}_r$ in Eq. (17), we provided a table for multiple values of parameters and compute $\acute{\mu}_1, ..., \acute{\mu}_4$ using computational software Mathematica 12.1. These outcomes show that real values of moments are well coincide with the approximate results.

It is important to emphasize that incomplete moments, mean deviations, Bonferroni and Lorenz curves Pundir, Arora, and Jain (2005) are also expressed in infinite series, however they can be easily calculated numerically.

3.3. Moment generating function

The MGF is broadly utilized in characterisation of model. The MGF of UG-TII(Θ) using the Maclaurin series expansion of an exponential function is mentioned as

$$\begin{aligned} M(z;\Theta) &= E\left(e^{tz}\right) = \sum_{r=0}^{+\infty} \frac{t^r}{r!} \acute{\mu}_r = \sum_{r,k=0}^{+\infty} \frac{t^r}{r!} (-1)^k \nu \binom{r+k-1}{k} \frac{\Gamma(1+k/\mu)}{\nu^{(1+k/\mu)}}, \\ &\text{if } \operatorname{Re}\left(\frac{k}{\mu}\right) > -1, \quad \operatorname{Re}(\nu) > 0. \end{aligned} \tag{20}$$

3.4. The central moments

The r^{th} central moment of UG-TII(Θ) can be attained as

$$\mu_r = \sum_{l=0}^{r} \binom{r}{l} (-1)^l \mu_z^l \int_0^1 z^{r-l} dF_Z(z;\Theta) \tag{21}$$

replacing r by $r-l$ in Eq. (18), we have

$$\begin{aligned} \mu_r &= \sum_{l=0}^{r} \sum_{k=0}^{+\infty} \binom{r}{l} (-1)^l \mu_z^l (-1)^k \nu \binom{r-l+k-1}{k} \frac{\Gamma(1+k/\mu)}{\nu^{(1+k/\mu)}}, \\ &\text{if } \operatorname{Re}\left(\frac{k}{\mu}\right) > -1, \quad \operatorname{Re}(\nu) > 0. \end{aligned} \tag{22}$$

3.5. Characteristic function

The cf of Z can be evaluated as

$$\Phi(\tau z;\Theta) = \int_0^1 e^{i\tau z} dF_Z(z;\Theta). \tag{23}$$

Appplying Taylor expansion on $e^{i\tau z}$, we have

$$\Phi(\tau z;\Theta) = \sum_{r=0}^{+\infty} \frac{(i\tau)^r}{r!} \int_0^1 z^r dF_Z(z;\Theta). \tag{24}$$

Using Eq. (18) we obtain the characteristic function of the UG-TII(Θ) as

$$\begin{aligned} \Phi(\tau z;\Theta) &= \sum_{k,r=0}^{+\infty} \frac{(i\tau)^r}{r!} (-1)^k \nu \binom{r+k-1}{k} \frac{\Gamma(1+k/\mu)}{\nu^{(1+k/\mu)}}, \\ &\text{if } \operatorname{Re}\left(\frac{k}{\mu}\right) > -1, \quad \operatorname{Re}(\nu) > 0. \end{aligned} \tag{25}$$

3.6. Factorial generating function

The FGF of UG-TII(Θ) is obtained as

$$\Psi(\tau z;\Theta) = \int_0^1 e^{\log(1+\tau)^z} dF_Z(z;\Theta) = \sum_{r=0}^{+\infty} \frac{\{\log(1+\tau)\}^r}{r!} \int_0^1 z^r dF_Z(z;\Theta). \tag{26}$$

Including the result given in Eq. (18), The FGF of UG-TII(Θ) is taken in the form

$$\Psi(\tau z;\Theta) = \sum_{k,r=0}^{+\infty} \frac{\{\log(1+\tau)\}^r}{r!}(-1)^k \nu \binom{r+k-1}{k} \frac{\Gamma(1+k/\mu)}{\nu^{(1+k/\mu)}},$$
$$\text{if } \operatorname{Re}\left(\frac{k}{\mu}\right) > -1, \quad \operatorname{Re}(\nu) > 0. \tag{27}$$

3.7. Order statistics

In the perspective of the standard normal distibution, order statistics was first discussed in Tippett (1925). It continuously grow in a more broad manner for modelling a huge range of phenomenas, generally in life testing and reliability analysis. Here, we have some useful findings concerning the order statistics of the UG-TII(Θ) distribution. Let $Z_{(1)} \leq Z_{(1)} ... \leq Z_{(n)}$ be order statistics of a random sample of size n from the model UG-TII(Θ) . Thus, for $m = 1, 2, ..., n$, and $i = 1, 2, 3$,the PDF of m^{th} order statistic, $Z_{(m)}$ is

$$\begin{aligned} g_{(m)}(z;\Theta) &= \tilde{K} \quad G(z;\Theta)^{m-1} \{1 - G(z;\Theta)\}^{n-m} g(z;\Theta), \\ &= \tilde{K} \sum_{l=0}^{n-m} (-1)^l \binom{n-m}{l} g(z;\Theta) G(z;\Theta)^{m+l-1} \end{aligned} \tag{28}$$

where $\tilde{K} = \frac{n!}{(n-m)!(m-1)!}$. Thus the PDF of m^{th} order statistic is obtained by replacing Eq. (3), Eq. (4) and above Eq. (28), we have

$$g_{(m)}(z;\phi,\ \Theta) = \hat{\Psi}_m \sum_{l=0}^{n-m} \frac{z^{-\mu-1}}{(1-z)^{1-\mu}} \exp\left[-\nu(m+l)\left(\frac{z}{1-z}\right)^{-\mu}\right], \tag{29}$$

where $\hat{\Psi}_m = \tilde{K}(-1)^l \binom{n-m}{l} \mu\nu$. The CDF of $Z_{(m)}$ is given by

$$\begin{aligned} G_{(m)}(z;\Theta) &= \sum_{j=m}^{n} \binom{n}{j} G(z;\phi,\ \Theta)^j \{1 - G(z;\phi,\ \Theta)\}^{n-j}, \\ &= \sum_{j=m}^{n} \sum_{p=0}^{n-j} \binom{n}{j} \binom{n-j}{p} (-1)^p G^{j+p}(z;\Theta), \end{aligned} \tag{30}$$

then the CDF of the m^{th} order statistic, $Z_{(m)}$ of UG-TII(Θ) model is

$$G_{(m)}(z;\Theta) = \sum_{j=m}^{n} \sum_{p=0}^{n-j} \binom{n}{j} \binom{n-j}{p} (-1)^p \exp\left[-\nu(j+p)\left(\frac{z}{1-z}\right)^{-\mu}\right]. \tag{31}$$

Specifically, the CDFs of $Z_{(n)}$ and $Z_{(1)}$ are attain, respectively by

$$G_{(n)}(z;\Theta) = G^n(z;\Theta), \qquad G_{(1)}(z;\Theta) = 1 - [1 - G(z;\Theta)]^n. \tag{32}$$

For i.i.d. RV, it is practicable to obtain the equation for the s^{th} ordinary moment of the order statistics for $\acute{\mu}_s < \infty$. Therefore, we can define the s^{th} moment of the m^{th} order statistic $Z_{(m)}$ as (see Silva, Ortega, and Cordeiro (2010))

$$\mu^s_{(m)} = E\left\{Z^s_{(m)}\right\} = \sum_{j=n-m+1}^{n} \binom{j-1}{n-m} \binom{n}{j} (-1)^{j-n+m-1} \mathrm{kI}_j(s), \tag{33}$$

where $\mathrm{kI}_j(s) = s \int_0^1 z^{s-1} [1 - G(z;\ \Theta)]^j \, dz$.

In particular, for the UG-TII(Θ) model, we obtain

$$\mathrm{kI}_j(s) = s \sum_{j=n-m+1}^{n} \binom{j-1}{n-m} \binom{n}{j} (-1)^{j-n+m-1} \int_0^1 z^{s-1} \left[1 - G(z;\Theta)\right]^j dz. \tag{34}$$

where the last integral can be evaluated numerically.

4. Inference with simulation

Several statistical characteristics of the UG-TII distribution are contributed to this section, considering that Θ are unknown. The assessment of Θ is carried out by the variety of known estimation methods. From now, $z_1, z_2, ..., z_n$ represent n observed values from Z and their ascending odering values $z_{(1)} \leq z_{(2)} \leq ... \leq z_{(n)}$.

4.1. MLE approach

There are many techniques for calculating parameters, but the most widely used is the maximum likelihood method. The MLE's have beneficial properties, like developing confidence intervals for the parameters of the model. For these estimates, large sample theory provides straightforward approximations that perform well in finite samples. Let $\Theta = (\nu, \mu)^\tau$ be the 2×1 parameter vector. The assesments of MLEs of Θ can be provided by optimizing the likelihood function with respect to ν and μ given by $L(\mathbf{z};\Theta) = \prod_{i=1}^{n} g(z_i;\ \Theta)$ or likewise the log-likelihood function for ν and μ given by

$$l(\mathbf{z};\Theta) = n\log\mu + n\log\nu - (\mu+1)\sum_{i=1}^{n}\log z_i + (\mu-1)\sum_{i=1}^{n}\log(1-z_i) - \nu\sum_{i=1}^{n}\left(\frac{z_i}{1-z_i}\right)^{-\mu}. \tag{35}$$

The MLEs are therefore derived by concurrently solving the following equations $\partial l(\mathbf{z};\Theta)/\partial\nu = 0$, and $\partial l(\mathbf{z};\Theta)/\partial\mu = 0$, where

$$\frac{\partial l(\mathbf{z};\Theta)}{\partial\nu} = \frac{n}{\nu} - \sum_{i=1}^{n}\left(\frac{z_i}{1-z_i}\right)^{-\mu}, \tag{36}$$

$$\frac{\partial l(\mathbf{z};\Theta)}{\partial\mu} = \frac{n}{\mu} - \sum_{i=1}^{n}\log z_i + \sum_{i=1}^{n}\log(1-z_i) + \nu\sum_{i=1}^{n}\left(\frac{z_i}{1-z_i}\right)^{-\mu}\log\left(\frac{z_i}{1-z_i}\right). \tag{37}$$

Although these equations can not be analytically solved, we use statistical software through iterative techniques such as a Newton-Raphson technique to address them numerically.

4.2. The other estimation approaches

There are many ways to evaluate the parameters of distributions that each of them has its distinctive features and strengths. Five of those strategies are presented momentarily in this subsection and will be numerically investigated in the simulation study. Here, G is the distribution function of the UG-TII(Θ) distribution.

4.3. LS approach

Swain, Venkatraman, and Wilson (1988) introduced the least square estimators and the weighted least square estimators to estimate the parameters of beta distributions. The LSEs,

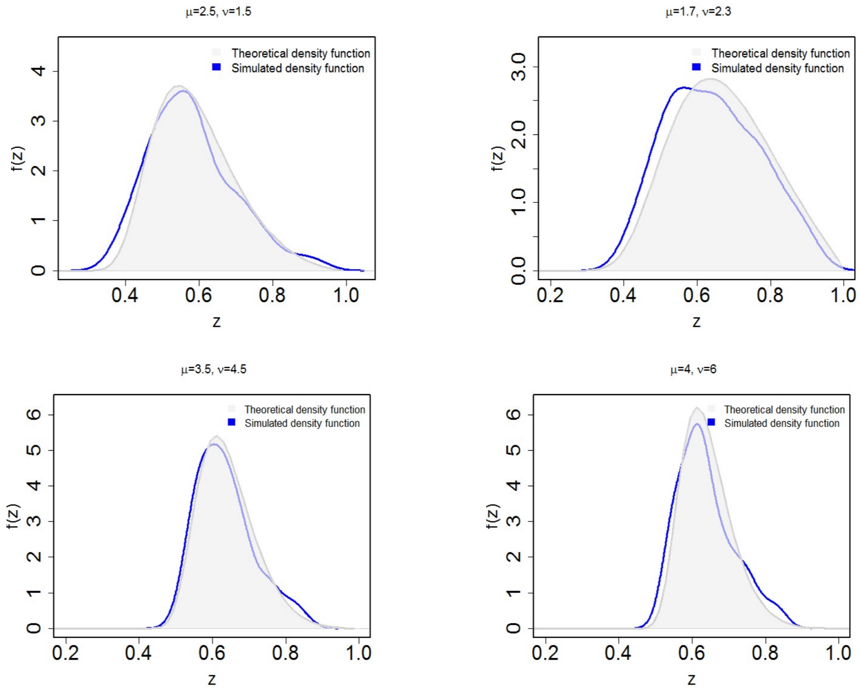

Figure 6: Fluctuation of theoretical and simulated density function under different parametric values

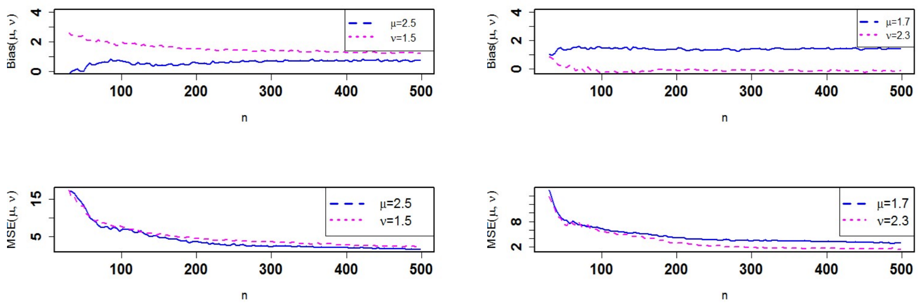

Figure 7: Fluctuations of Bias and MSE of estimations for parameter values (i) $\mu = 2.5$, $\nu = 1.5$ (ii) $\mu = 1.7$, $\nu = 2.3$

$\hat{\nu}_{LS}$, and $\hat{\mu}_{LS}$, can be achieved by minimizing

$$LS(\Theta) = \sum_{i=1}^{n} \left(G(z_{(i)}; \Theta) - \frac{i}{n+1} \right)^2$$

$$LS(\Theta) = \sum_{i=1}^{n} \left\{ \exp\left[-\nu \left(\frac{z_{(i)}}{1 - z_{(i)}} \right)^{-\mu} \right] - \frac{i}{n+1} \right\}^2 . \tag{38}$$

with respect to ν and μ. For more details see Erto (1989). These can be extracted equivalently by solving $\partial LS(\Theta)/\partial\nu = 0$, and $\partial LS(\Theta)/\partial\mu = 0$ where

$$\frac{\partial LS(\Theta)}{\partial \nu} = 2 \sum_{i=1}^{n} \xi_i^1(\Theta) \left\{ \exp\left[-\nu \left(\frac{z_{(i)}}{1 - z_{(i)}} \right)^{-\mu} \right] - \frac{i}{n+1} \right\}, \tag{39}$$

$$\frac{\partial LS(\Theta)}{\partial \mu} = 2 \sum_{i=1}^{n} \xi_i^2(\Theta) \left\{ \exp\left[-\nu \left(\frac{z_{(i)}}{1 - z_{(i)}} \right)^{-\mu} \right] - \frac{i}{n+1} \right\}, \tag{40}$$

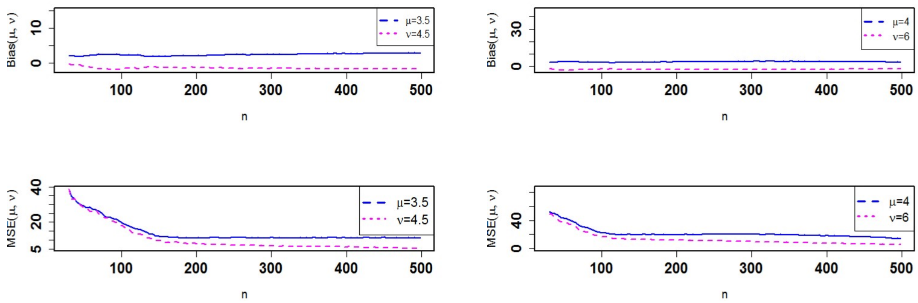

Figure 8: Fluctuations of Bias and MSE of estimations for parameter values (i) $\mu = 2.5$, $\nu = 1.5$ (ii) $\mu = 1.7$, $\nu = 2.3$

and

$$\begin{aligned}
\xi_i^1(\Theta) &= \left(\frac{z_{(i)}}{1-z_{(i)}}\right)^{-\mu} e^{-\nu\left(\frac{z_{(i)}}{1-z_{(i)}}\right)^{-\mu}}, \\
\xi_i^2(\Theta) &= \exp\left[-\nu\left(\frac{z_{(i)}}{1-z_{(i)}}\right)^{-\mu}\right] \nu \left(\frac{z_{(i)}}{1-z_{(i)}}\right)^{-\mu} \log\left(\frac{z_{(i)}}{1-z_{(i)}}\right)
\end{aligned} \tag{41}$$

4.4. WLS approach

The WLSEs, $\hat{\nu}_{WLS}$ and $\hat{\mu}_{WLS}$, can be determined by minimizing, with respect to ν and μ, the following function

$$\begin{aligned}
WLS(\Theta) &= \sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \left[G(z_{(i)};\Theta) - \frac{i}{n+1}\right]^2 \\
&= \sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \sum_{i=1}^{n} \left\{\exp\left[-\nu\left(\frac{z_{(i)}}{1-z_{(i)}}\right)^{-\mu}\right] - \frac{i}{n+1}\right\}^2.
\end{aligned} \tag{42}$$

We can also obtain these estimators by solving $\partial WLS(\Theta)/\partial\nu = 0$, and $\partial WLS(\Theta)/\partial\mu = 0$ where

$$\frac{\partial LS(\Theta)}{\partial \nu} = 2\sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \xi_i^1(\Theta) \left\{\exp\left[-\nu\left(\frac{z_{(i)}}{1-z_{(i)}}\right)^{-\mu}\right] - \frac{i}{n+1}\right\}, \tag{43}$$

$$\frac{\partial LS(\Theta)}{\partial \mu} = 2\sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \xi_i^2(\Theta) \left\{\exp\left[-\nu\left(\frac{z_{(i)}}{1-z_{(i)}}\right)^{-\mu}\right] - \frac{i}{n+1}\right\}, \tag{44}$$

where $\xi_i^j(\Theta)$, $j = 1, 2$. are given in Eq. (41).

4.5. CVM approach

Minimum distance estimators of the Cramér-von Mises type are focused on minimising the distance between the theoretical and empirical cumulative distribution functions. Macdonald (1971) presented empirical evidence that these estimators' bias is smaller than the bias of other estimators of minimum distance.

$$\begin{aligned}
C(\Theta) &= \frac{1}{12n} + \sum_{i=1}^{n} \left[G(z_{(i)};\ \Theta) - \frac{2i-1}{2n}\right]^2, \\
&= \frac{1}{12n} + \sum_{i=1}^{n} \left\{\exp\left[-\nu\left(\frac{z_{(i)}}{1-z_{(i)}}\right)^{-\mu}\right] - \frac{2i-1}{2n}\right\}^2.
\end{aligned} \tag{45}$$

Thus, the CVEs, $\hat{\nu}_{CVM}$, and $\hat{\mu}_{CVM}$, are obtained by solving the following equations simultaneously $\partial C(\Theta)/\partial\nu = 0$, and $\partial C(\Theta)/\partial\mu = 0$ where

$$\frac{\partial C(\Theta)}{\partial \nu} = 2\sum_{i=1}^{n} \xi_i^1(\Theta) \left\{ \exp\left[-\nu \left(\frac{z_{(i)}}{1 - z_{(i)}} \right)^{-\mu} \right] - \frac{2i-1}{2n} \right\}, \tag{46}$$

$$\frac{\partial C(\Theta)}{\partial \mu} = 2\sum_{i=1}^{n} \xi_i^2(\Theta) \left\{ \exp\left[-\nu \left(\frac{z_{(i)}}{1 - z_{(i)}} \right)^{-\mu} \right] - \frac{2i-1}{2n} \right\}, \tag{47}$$

where $\xi_i^j(\Theta)$ and $j = 1, 2$ are defined in Eq.(41)

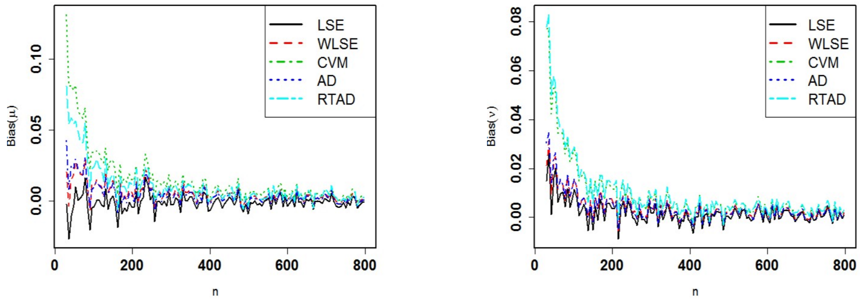

Figure 9: Fluctuations of bias of estimations for parameter values $\mu = 2.5$, and $\nu = 1.5$

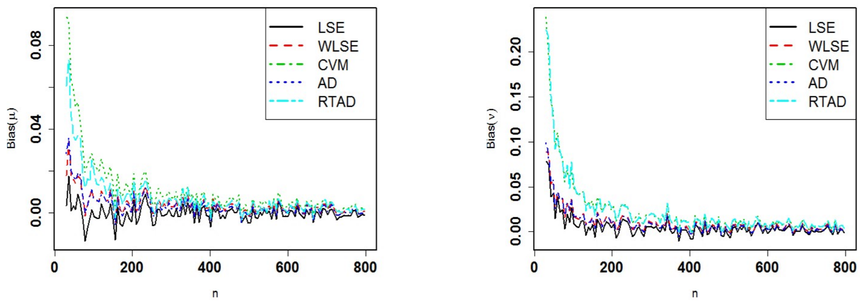

Figure 10: Fluctuations of bias of estimations for parameter values $\mu = 1.7$, and $\nu = 2.3$

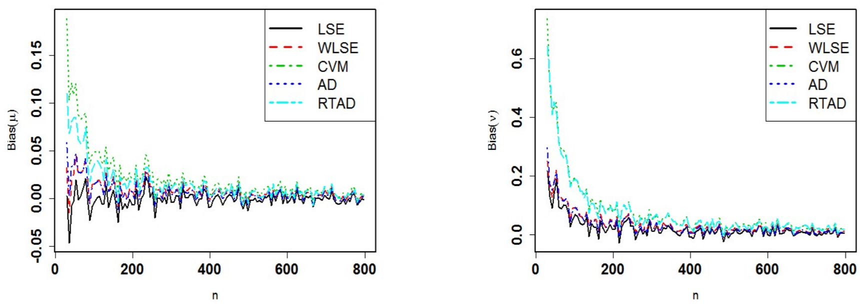

Figure 11: Fluctuations of bias of estimations for parameter values $\mu = 3.5$, and $\nu = 4.5$

4.6. AD approach

Anderson and Darling (1952) introduced the AD test as an alternative to statistical tests for identifying sample distributions' the deviation from normality. Notably Anderson–Darling

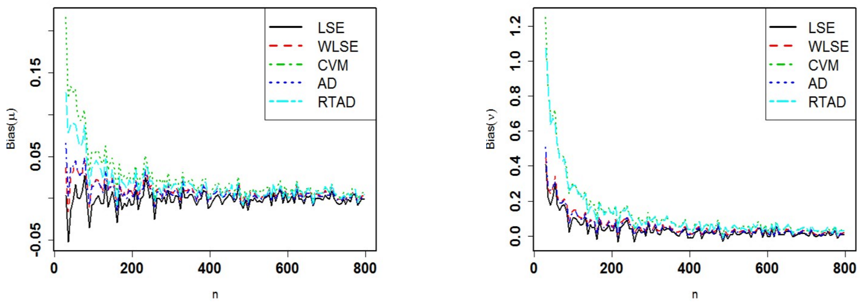

Figure 12: Fluctuations of bias of estimations for parameter values $\mu = 4.0$, and $\nu = 6.0$

test converges quite rapidly towards the asymptote (Anderson and Darling (1954); Pettitt (1976); Stephens (1974)). Boos (1981) also discussed the properties of the AD estimators.

$$A(\Theta) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log(G(z_{(i)};\ \Theta)] + \log(\bar{G}(z_{(n+1-i)};\ \Theta))\right], \tag{48}$$

where $\bar{G}(.) = 1 - G(.)$. Thus, the ADs, $\hat{\nu}_{AD}$, and $\hat{\mu}_{AD}$, are obtained by solving the following equations simultaneously $\partial AD(\Theta)/\partial\nu = 0$, and $\partial AD(\Theta)/\partial\mu = 0$ where

$$\frac{\partial A(\Theta)}{\partial \nu} = -\frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\frac{\xi_i^1(\Theta)}{G(z_{(i)};\ \Theta)} - \frac{\xi_{(n+1-i)}^1(\Theta)}{\bar{G}(z_{(n+1-i)};\ \Theta)}\right], \tag{49}$$

$$\frac{\partial A(\phi,\Theta)}{\partial \mu} = -\frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\frac{\xi_i^2(\Theta)}{G(z_{(i)};\ \Theta)} - \frac{\xi_{(n+1-i)}^2(\Theta)}{\bar{G}(z_{(n+1-i)};\ \Theta)}\right], \tag{50}$$

where $\xi_i^j(\Theta)$ and $j = 1, 2$ are given in Eq. (40) respectively.

4.7. RTAD approach

Similarly, the right-tail Anderson–Darling estimates (RTADEs) of $\hat{\nu}_{RTAD}$ and $\hat{\mu}_{RTAD}$ can be obtained by minimizing, with respect to μ and ν, the following function

$$R(\Theta) = \frac{n}{2} - 2\sum_{i=1}^{n} G(z_{(i)};\ \Theta) - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\log[\bar{G}(z_{(n+1-i)};\ \Theta)]. \tag{51}$$

The RTADEs can be evaluated by concurrently solving the below equations $\partial R\left(\mathbf{z}|\,\Theta\right)/\partial\nu = 0$ and $\partial R\left(\mathbf{z}|\,\Theta\right)/\partial\mu = 0$, where

$$\frac{\partial R(\Theta)}{\partial \nu} = -2\sum_{i=1}^{n}\frac{\xi_i^1(\Theta)}{G(z_{(i)};\ \Theta)} + \frac{1}{n}\sum_{i=1}^{n}(2i-1)\frac{\xi_{n+1-i}^1(\Theta)}{\bar{G}(z_{(i)};\ \Theta)}, \tag{52}$$

$$\frac{\partial R(\Theta)}{\partial \mu} = -2\sum_{i=1}^{n}\frac{\xi_i^2(\Theta)}{G(z_{(i)};\ \Theta)} + \frac{1}{n}\sum_{i=1}^{n}(2i-1)\frac{\xi_{n+1-i}^2(\Theta)}{\bar{G}(z_{(i)};\ \Theta)}. \tag{53}$$

5. Simulation study

We consider the one model that has been used in this section to explore the estimators introduced above, and analyse the bias and MSE of such estimators for various samples. The estimators of the parameters of proposed distribution have been evaluated by simulating

$(\mu, \nu) = \{(2.5, 1.5), (1.7, 2.3), (3.5, 4.5), (4, 6)\}$. The density functions of the UG-TII(Θ) model for these choices are shown in Fig. 5. The bias and the MSE of estimators have been used to evaluate the validity of the estimators. Consideration is given to the efficiency of each parameter estimation method for the UG-TII(Θ) model with respect to sample size n. Simulation study is executed for this purpose on the basis of the following steps

1. Generate two thousand samples of size n from Eq. (3). This work is carried out simply by QF and obtained data from uniform distribution.

2. Calculate the estimates for the 2000 samples, say $(\hat{\mu}_i, \hat{\nu}_i)$ for $i = 1, 2, ..., 2000$.

3. Calculate the biases and MSEs. These objectives are obtained with the help of the following formulas

$$Bias_{\Theta}(n) = \frac{1}{2000}\sum_{i=1}^{2000}\left(\hat{\Theta}_i - \Theta\right), \quad (54)$$

$$MSE_{\Theta}(n) = \frac{1}{2000}\sum_{i=1}^{2000}\left(\hat{\Theta}_i - \Theta\right)^2, \quad (55)$$

   where $\Theta = \mu, \nu$.

4. These steps were repeated for $n = 30, 34, ..., 500$, with the mentioned parameters for MLEs, and for other methods of estimation n is taken upto 800. So, for $\Theta = \mu, \nu$, and $n = 30, 34, ..., 500$, bias$_{\Theta}(n)$ and MSE$_{\Theta}(n)$ have been computed. We have used the optima function in R to obtain the value of the estimators. The outcome of the simulations of this subsection is indicated in Figs. 7-16.

In particular, we can infer that the estimators have the property of asymptotic unbiasedness, because as n grows, the bias tends to zero, whereas the pattern in the mean squared error indicates consistency, because the errors tend to zero when the value of n increases. From Figs. 7-16, the following observations can be extracted.

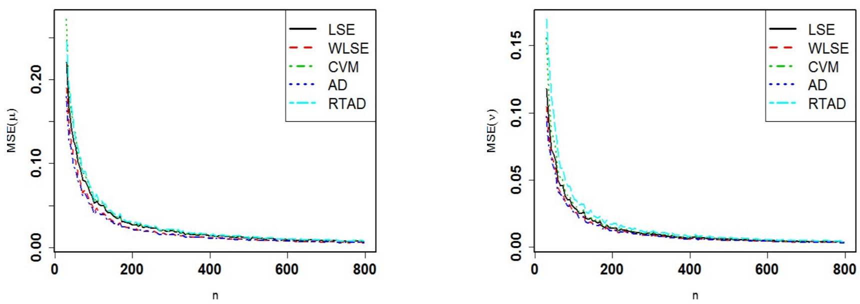

Figure 13: Fluctuations of MSE of estimations for parameter values $\mu = 2.5$, and $\nu = 1.5$

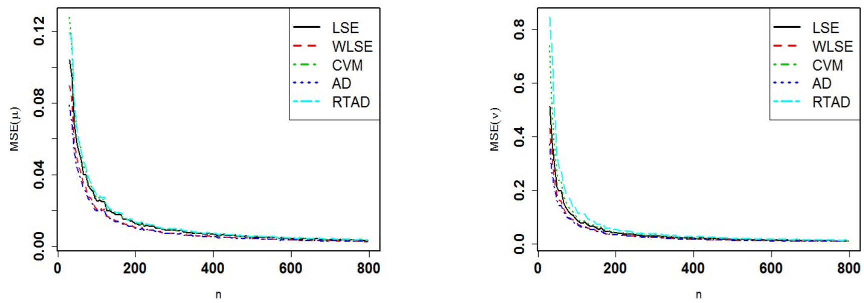

Figure 14: Fluctuations of MSE of estimations for parameter values $\mu = 1.7$, and $\nu = 2.3$

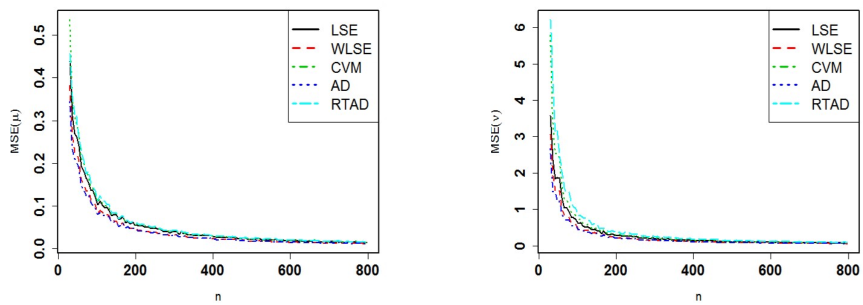

Figure 15: Fluctuations of MSE of estimations for parameter values $\mu = 3.5$, and $\nu = 4.5$

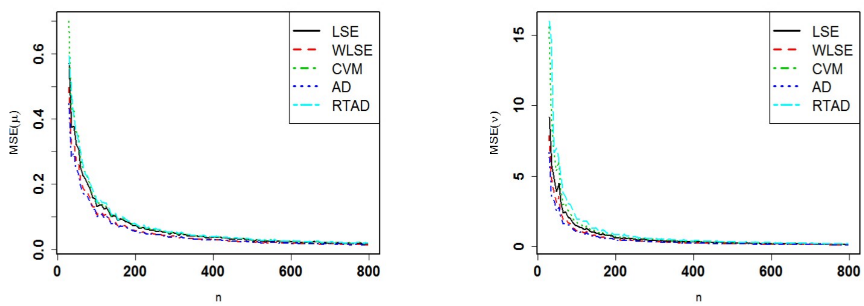

Figure 16: Fluctuations of MSE of estimations for parameter values $\mu = 4.0$, and $\nu = 6.0$

- For all estimation methods, the bias of $\hat{\mu}$ reduces as n increases.
- For all estimation methods, the bias of $\hat{\nu}$ reduces as n increases.
- For the MLE, the biases of $\hat{\mu}$ and $\hat{\nu}$ are generally positive.
- The bias of parameter μ is higher than parameter ν, when $\mu < \nu$ and it is related inversely accordingly $\mu > \nu$ in ML estimation approache (Figs. 7-8).
- The MSE of parameter μ is higher than parameter ν, when $\mu < \nu$ and it is related inversely accordingly $\mu > \nu$ in ML estimation approache (Figs.7-8).
- For the CVM, and the RTAD, the biases and MSEs of $\hat{\mu}$ and $\hat{\nu}$ seem larger (Figs.9-16).
- For the LSE, CVM, WLSE, AD and RTAD, the biases of $\hat{\mu}$ and $\hat{\nu}$ are generally positive, but for the LSE, the negative biases of $\hat{\mu}$ and $\hat{\nu}$ are also noticed (Figs.9-12).

- In case of other estimation methods, LS estimation is stronger than other estimation approach for the all chosen parameter values in terms of bias, approximately for both parameters, when sample volume tends towards infinity (Figs. 9-12).

- While considering other estimation approaches, WLS and AD estimation are stronger than other estimation approach for the all chosen parameter values in terms of MSE, approximately for both parameters, But when sample volume $n > 200$, each method of parameters estimations work approximately equally (Figs. 9-16).

A generic consequences about above figures are that bias and MSE graphs for two parameters with the increase in the volume of the sample all techniques will approach zero. This verifies the validity of these techniques of estimation and of the numerical calculations for the UG-TII distribution parameters.

6. Demonstrative example

Real data are now included in this section to illustrate the versatility and applicability of the proposed new model. The object of the UG-TII(Θ) distribution is to provide an alternate distribution to fit the unit interval data with other distributions that are present in the literature. We present two applications by fitting the UG-TII(Θ) model as well as some competitor models. For these two examples, the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Cram-von Mises (W^*), Anderson-Darling (A^*), Kolmogorov-Smirnov (K.S) and the P-Value of K.S test have been specified to compare models. The last two discussed in Chen and Balakrishnan (1995). In general, the lower the values of these statistics, the stronger the fit to the data. The Unit Gompertz Mazucheli, Menezes, and Dey (2019), Kumaraswamy (KSW), Size biased Kumaraswamy (SBKSW) distribution, Unit Weibull distribution Mazucheli, Menezes, Fernandes, de Oliveira, and Ghitany (2020) Beta distribution, Kumaraswamybeta (KSWbeta) distributions have been selected for comparison. The MLE approach has been used to estimate the parameters of models. The maximum likelihood estimation of the parameters are presented in Tables 3 and 4.

Table 3: MLEs and SEs of the parameters of considered distribution for data set I

Distributions	MLEs	Standard errors
Unit Gompertz (η, b)	(2.91, 0.01)	(0.23, 0.00)
KSW$(\alpha,\ \beta)$	(2.72, 44.81)	(0.29, 17.68)
SBKSW$(\lambda,\ \delta)$	(2.19, 32.76)	(0.30, 12.22)
Unit Weibull$(\alpha,\ \beta)$	(0.06, 5.14)	(0.02, 0.58)
Beta$(\alpha,\ \beta)$	(5.93, 21.18)	(1.18, 4.34)
UG-TII$(\mu,\ \nu)$	(2.47, 0.02)	(0.26, 0.01)
KSWbeta$(\alpha, \beta, \lambda,\ \kappa)$	(25.17, 30.99, 1.24, 0.39)	(35.37, 130.14, 4.70, 0.43)

The first dataset refers to twelve core specimens from four cross-sections of petroleum wells sampled, and there are 48 values. For permeability, each core sample was measured and each cross-section has the following variables: the total area of pores, the total perimeter of pores and shape. For further detail see in R Core Team (2017). The next data set refers to 20 overall flood level observations for the Susquehanna River at Harrisburg, Pennsylvania (in millions of cubic feet per second). In addition, we notice that Mazucheli, Menezes, and Dey (2018) analysed these datasets to demonstrate the suitability of the unit-Gamma distribution in order to compare second-order bias corrections MLE. To conclude, for the two datasets, the UG-TII(Θ) model shows itself to be the most suitable model, demonstrating its applicability in a realistic environment.

Table 4: MLEs and SEs of the parameters of considered distribution for data set II

Distributions	MLEs	Standard errors
Unit Gompertz (η, b)	(4.13, 0.02)	(0.75, 0.01)
KSW$(\alpha,\ \beta)$	(3.36, 11.79)	(0.60, 5.36)
SBKSW$(\lambda,\ \delta)$	(2.78, 10.57)	(0.62, 4.62)
Unit Weibull$(\alpha,\ \beta)$	(1.03, 3.89)	(0.24, 0.71)
Beta$(\alpha,\ \beta)$	(6.76, 9.11)	(2.10, 2.85)
UG-TII$(\mu,\ \nu)$	(2.56, 0.24)	(0.46, 0.10)
KSWbeta$(\alpha, \beta, \lambda,\ \kappa)$	(25.94, 24.65, 4.93, 0.19)	(0.73, 14.07, 0.45, 0.05)

Table 5: Values of the considered goodness-of-fit measures for data I

Distribution	AIC	BIC	$W*$	$A*$	$K-S$	p-value
Unit Gompertz (η, b)	-109.06	-105.32	0.04	0.31	0.08	0.90
KSW$(\alpha,\ \beta)$	-100.95	-97.20	0.21	1.28	0.16	0.20
SBKSW$(\lambda,\ \delta)$	-103.17	-99.43	0.05	0.45	0.99	0.00
Unit Weibull$(\alpha,\ \beta)$	-112.68	-108.94	0.03	0.20	0.09	0.88
Beta$(\alpha,\ \beta)$	-107.14	-103.39	0.10	0.78	1.00	0.00
UG-TII$(\mu,\ \nu)$	-111.24	-107.50	0.03	0.23	0.08	0.95
KSWbeta$(\alpha,\ \beta,\ \lambda,\ \kappa)$	-108.71	-101.23	0.04	0.23	0.09	0.81

Table 6: Values of the considered goodness-of-fit measures for data II

Distribution	AIC	BIC	$W*$	$A*$	$K-S$	p-value
Unit Gompertz (η, b)	-28.72	-26.73	0.05	0.30	0.15	0.78
KSW$(\alpha,\ \beta)$	-21.73	-19.74	0.16	0.97	0.21	0.34
SBKSW$(\lambda,\ \delta)$	-22.46	-20.47	0.05	0.30	0.97	0.00
Unit Weibull$(\alpha,\ \beta)$	-28.16	-26.17	0.06	0.34	0.14	0.85
Beta$(\alpha,\ \beta)$	-24.13	-22.13	0.07	0.44	1.00	0.00
UG-TII$(\mu,\ \nu)$	-28.90	-26.91	0.05	0.29	0.12	0.93
KSWbeta$(\alpha,\ \beta,\ \lambda,\ \kappa)$	-24.60	-20.61	0.06	0.34	0.15	0.80

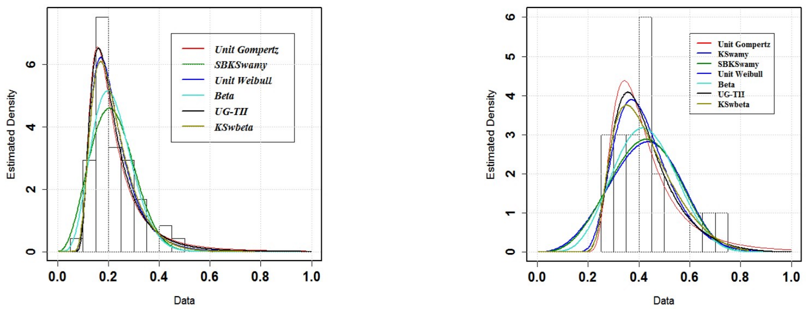

Figure 17: Histograms and fitted densities for data set I and II

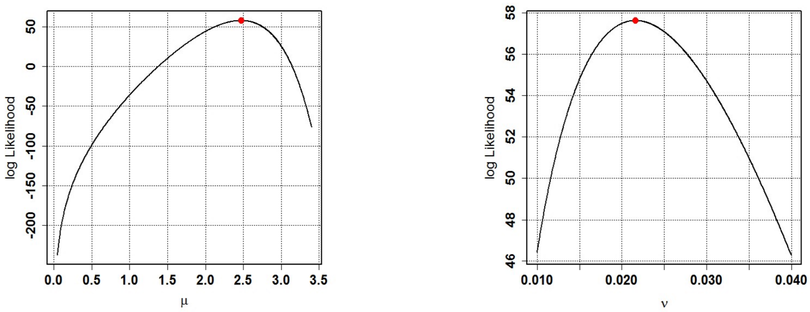

Figure 18: Uni-modal profile likelihood functions of parameters of UG-TII(Θ) for data I

Fig. 17 reveals the fitted pdf on histogram of data set I and II. The criteria mentioned are presented in Tables 2 and 3. As we can observe from Tables 5-6, for the UG-TII(Θ) distribution, the values of AIC, BIC, $W*$, $A*$ and $K-S$ are smaller than the Unit Gompertz, KSW, SBKSW, Beta, Weibull and Kumaraswamybeta distributions understudy for both data sets. P-Value even has the maximum value. So we infer that the suggested density works better than other distributions. This result is also verified by Fig. 17. We demonstrate the adequacy of the models by the use of probability-probability (PP) maps in Fig. 20, for Data sets I and II, respectively, in attempt to provide another point of view. For Dataset I in particular, in view of the perfect aligning the scatter plot by the PP line, it is obvious that comparison to the other distributions, the UG-TII(Θ) distribution has a better match. The uni-modal profile likelihood functions of parameters of UG-TII(Θ) are mapped in Figs. 18 and 19 for Datasets I and II, respectively, to demonstrate the uniqueness of the MLEs of μ and ν.

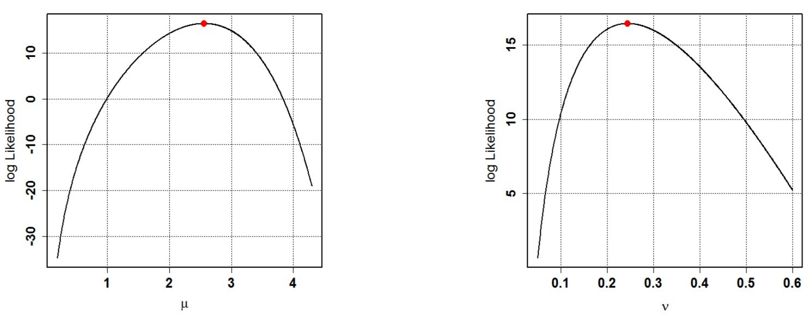

Figure 19: Uni-modal profile likelihood functions of parameters of UG-TII(Θ) for data II

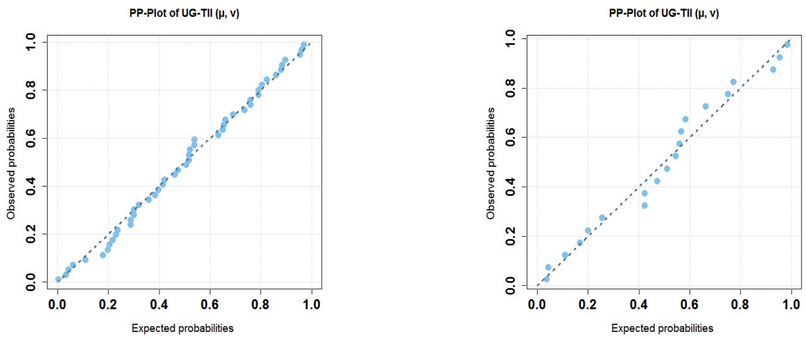

Figure 20: Fig. 20: Estimated and empirical CDFs of UG-TII(Θ) distribution for data sets I and II

7. Conclusion

In current article, we presented a novel two-parameter model named the unit Gumbel distribution with various forms of failure rate. We derive many characteristics of the new model like explicit expressions for the quantile, moments, moment-generating functions and distribution of order statistics. The parameters of the model are estimated by various estimation methods, including the MLE, LS, WLS, CVM, AD and RTAD estimation approaches. A simulation analysis is conducted with various samples and different parametric sets considering $\mu > \nu$, and $\mu < \nu$, to assess the results of these approaches. Through considering two practical datasets, the applicability of the unit Gumbel distribution has been demonstrated. The UG-TII model has been shown to be a serious competitor to other ones. The construction of various regression models, Bayesian parameter estimation, and new dataset analysis will be included in future work. Thanks to its particular features, we assume that UG-TII model can be valuable for the practitioner, for statistical analysis outwith framework of this article.

A. Abbreviations and nomenclature

Abbreviations

PDF	probability density function	RV	Random Variable
UG-TII	unit Gumbel type-II	QF	quantile function
MLE	maximum likelihood estimation	CHRF	cumulative hazard rate function
SF	survival function	MSE	mean square error
i.i.d	independently identically distributed	ADEs	Anderson-Darling estimates
CDF	cumulative distribution function	LSEs	least square estimates
WLSEs	weighted least square estimates	MGF	moment generating function
CVMEs	Cramér-von Mises estimates	KSW	Kumaraswamy
RTADEs	right tail Anderson Darling estimates	SBKSW	size biased Kumaraswamy
HRF	hazard rate function	FGF	factorial generating function
MMEs	Method of Moments Estimates	SEs	standard errors%

Nomenclature

$g(\Theta)$	PDF	$G(\Theta)$	CDF
$S(\Theta)$	SF	$h(\Theta)$	HRF

B. Quantile regression

It is important to point out that neither μ or ν has a direct interpretation in terms of the observed data. For example, μ is no longer a shape parameter as in the distribution of

Z. In order to assess the effect of covariates on the quantile of distribution of the response variable, in this appendix we give some directions on quantile regression model taken UG-TII distribution as baseline. Following, (for example, Mazucheli, Leiva, Alves, and Menezes (2021), Mazucheli *et al.* (2020), Korkmaz, Emrah, Chesneau, and Yousof (2020)), for quantile regression, ν in (13) must be re-parametrized as $\nu = g^{-1}(\sigma) = -\log(\tau)\left(\frac{\sigma}{1-\sigma}\right)^{\mu}$ such that σ is, for a fixed and known value τ, the τ-th quantile of the distribution of Z.
Considering the re-parameterization in ν the corresponding PDF and CDF are as follow

$$g(z;\mu,\sigma,\tau) = \left[-\log(\tau)\left(\frac{\sigma}{1-\sigma}\right)^{\mu}\right]\mu\frac{z^{-\mu-1}}{(1-z)^{1-\mu}}\exp\left\{\left[\log(\tau)\left(\frac{\sigma}{1-\sigma}\right)^{\mu}\right]\left(\frac{z}{1-z}\right)^{-\mu}\right\} \tag{56}$$

and

$$G(z;\mu,\sigma,\tau) = \exp\left\{\left[\log(\tau)\left(\frac{\sigma}{1-\sigma}\right)^{\mu}\right]\left(\frac{z}{1-z}\right)^{-\mu}\right\}. \tag{57}$$

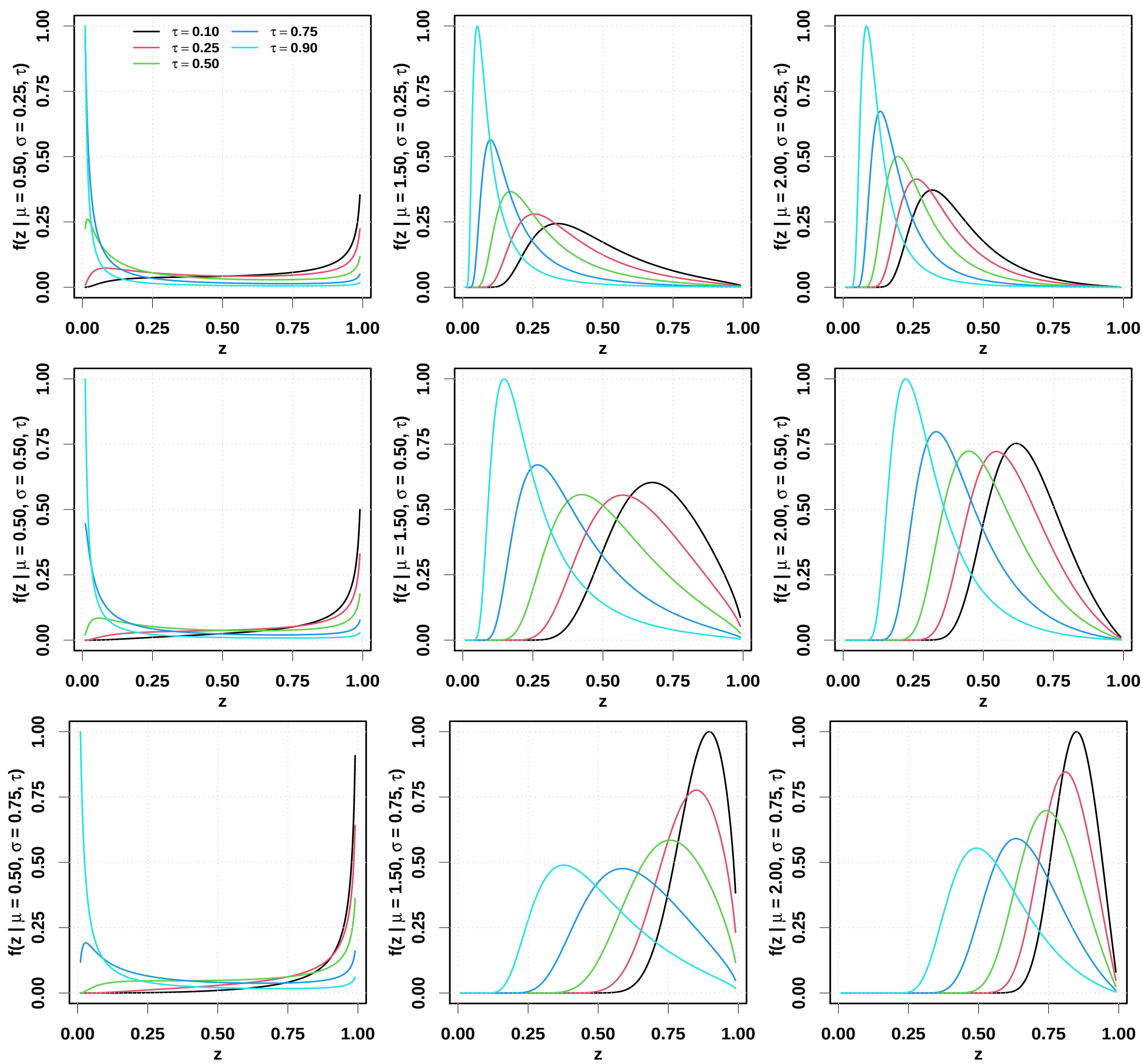

Figure 21: Plots of the re-parameterized PDF (3) for indicated values of σ, μ and τ

If concomitant with z_i, $i = 1, \ldots, n$ we also observe covariate vectors $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ we may be interested in evaluating the effects of these covariates on (μ, σ). For ML estimation, observed $\boldsymbol{z} = (z_1, \ldots, z_n)$ from n independent random variables $Z_1, \ldots, Z_n$ we may assume the following equations

$$h_1(\sigma_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

and

$$h_2(\mu_i) = \zeta_i = \delta_0 + \delta_1 y_{i1} + \cdots + \delta_p y_{iq}$$

linking both η_i and ζ_i with a linear combination of the explanatory variables $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{ip})$ and $\boldsymbol{y}_i = (1, y_{i1}, \ldots, y_{iq})$, respectively. Where $h_1(\cdot)$ and $h_2(\cdot)$ are strictly monotonic, twice differentiable functions. Applications of quantile regression using the UGT-II distribution as baseline are being studied by the authors and further details are not presented here.

References

Almetwally EM, Almongy HM (2019). "Maximum Product Spacing and Bayesian Method for Parameter Estimation for Generalized Power Weibull Distribution under Censoring Scheme." *Journal of Data Science*, **17**(2), 407–444.

Anderson TW, Darling DA (1952). "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes." *The Annals of Mathematical Statistics*, **23**(2), 193–212.

Anderson TW, Darling DA (1954). "A Test of Goodness of Fit." *Journal of the American Statistical Association*, **49**(268), 765–769.

Barlow RE, Proschan F (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart and Winston, New York.

Boos DD (1981). "Minimum Distance Estimators for Location and Goodness of Fit." *Journal of the American Statistical Association*, **76**(375), 663–670.

Chen G, Balakrishnan N (1995). "A General Purpose Approximate Goodness-of-fit Test." *Journal of Quality Technology*, **27**(2), 154–161.

Dey S, Mazucheli J, Nadarajah S (2018). "Kumaraswamy Distribution: Different Methods of Estimation." *Computational and Applied Mathematics*, **37**(2), 2094–2111.

Erto P (1989). "Genesis, Properties and Identification of the Inverse Weibull Lifetime Model." *Statistica Applicata*, **1**(4), 117–128.

Gupta AK, Nadarajah S (2004). *Handbook of Beta Distribution and Its Applications*. Marcel Dekker, New York.

Hunger M, Baumert J, Holle R (2011). "Analysis of SF-6D Index Data: Is Beta Regression Appropriate?" *Value in Health*, **14**(5), 759–767.

Kieschnick R, McCullough BD (2003). "Regression Analysis of Variates Observed on $(0,1)$: Percentages, Proportions and Fractions." *Statistical Modelling*, **3**(3), 193–213.

Korkmaz MÇ, Emrah A, Chesneau C, Yousof HM (2020). "On the Unit-Chen Distribution with Associated Quantile Regression and Applications." *Journal of Applied Statistics*, **44**(6), 1–22.

Lehmann EL (1999). *Elements of Large-Sample Theory*. Springer-Verlag, New York.

Loganathan A, Uma A (2017). "Comparison of Estimation Methods for Inverse Weibull Parameters." *Global and Stochastic Analysis*, **4**(1), 83–93.

Macdonald PDM (1971). "Comments and Queries Comment on "An Estimation Procedure for Mixtures of Distributions" by Choi and Bulgren." *Journal of the Royal Statistical Society: Series B (Methodological)*, **33**(2), 326–329.

Marshall AW, Olkin I (2007). *Life Distributions*. Springer Series in Statistics. Springer, New York. Structure of nonparametric, semiparametric, and parametric families.

Mazucheli J, Leiva V, Alves B, Menezes AFB (2021). "A New Quantile Regression for Modeling Bounded Data under a Unit Birnbaum–Saunders Distribution with Applications in Medicine and Politics." *Symmetry*, **13**(4), 1–21.

Mazucheli J, Menezes AF, Dey S (2019). "Unit-Gompertz Distribution with Applications." *Statistica*, **79**(1), 25–43.

Mazucheli J, Menezes AFB (2019). "L-Moments and Maximum Likelihood Estimation for the Complementary Beta Distribution with Applications on Temperature Extremes." *Journal of Data Science*, **17**(2), 391–406.

Mazucheli J, Menezes AFB, Dey S (2018). "Improved Maximum-likelihood Estimators for the Parameters of the Unit-Gamma Distribution." *Communications in Statistics-Theory and Methods*, **47**(15), 3767–3778.

Mazucheli J, Menezes AFB, Fernandes LB, de Oliveira RP, Ghitany ME (2020). "The Unit-Weibull Distribution as an Alternative to the Kumaraswamy Distribution for the Modeling of Quantiles Conditional on Covariates." *Journal of Applied Statistics*, **47**(6), 954–974.

Papke LE, Wooldridge JM (1996). "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics*, **11**(6), 619–632.

Pettitt AN (1976). "A Two-sample Anderson-Darling Rank Statistic." *Biometrika*, **63**(1), 161–168.

Pinheiro EC, Ferrari SLP (2016). "A Comparative Review of Generalizations of the Gumbel Extreme Value Distribution with an Application to Wind Speed Data." *Journal of Statistical Computation and Simulation*, **86**(11), 2241–2261.

Pundir S, Arora S, Jain K (2005). "Bonferroni Curve and the Related Statistical Inference." *Statistics & Probability Letters*, **75**(2), 140–150.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramos PL, Louzada F (2016). "The Generalized Weighted Lindley Distribution: Properties, Estimation, and Applications." *Cogent Mathematics*, **3**(1), 1–18.

Ramos PL, Louzada F, Ramos E, Dey S (2020). "The Fréchet Distribution: Estimation and Application - An Overview." *Journal of Statistics and Management Systems*, **23**(3), 549–578.

Silva GO, Ortega EMM, Cordeiro GM (2010). "The Beta Modified Weibull Distribution." *Lifetime Data Analysis*, **16**(3), 409–430.

Sindhu TN, Shafiq A, Al-Mdallal QM (2021a). "Exponentiated Transformation of Gumbel Type-II Distribution for Modeling COVID-19 Data." *Alexandria Engineering Journal*, **60**(1), 671–689.

Sindhu TN, Shafiq A, Al-Mdallal QM (2021b). "On the Analysis of Number of Deaths due to Covid- 19 Outbreak Data Using a New Class of Distributions." *Results in Physics*, **21**, 1–17.

Stephens MA (1974). "Components of Goodness-of-fit Statistics." In *Annales de l'IHP Probabilités et Statistiques*, pp. 37–54.

Swain JJ, Venkatraman S, Wilson JR (1988). "Least-squares Estimation of Distribution Functions in Johnson's Translation System." *Journal of Statistical Computation and Simulation*, **29**(4), 271–297.

Tahir MH, Cordeiro GM, Ali S, Dey S, Manzoor A (2018). "The Inverted Nadarajah–Haghighi Distribution: Estimation Methods and Applications." *Journal of Statistical Computation and Simulation*, **88**(14), 2775–2798.

Tippett LHC (1925). "On the Extreme Individuals and the Range of Samples Taken from a Normal Population." *Biometrika*, **17**(3), 364–387.

Affiliation:

Bruna Alves
Universidade Estadual de Maringá
Departament of Statistics
Maringá 87020-900, Brazil
E-mail: pg402900@uem.br

Submitted: 2021-08-25
Accepted: 2021-12-17