Pak.j.stat.oper.res. Vol.18 No.1 2022 pp 99-120 DOI: https://doi.org/10.18187/pjsor.v18i1.3908

Pakistan Journal of Statistics and Operation Research

Contributions Towards New Families of Distributions: An Investigation, Further Developments, Characterizations and Comparative Study

Zubair Ahmad[1], Eisa Mahmoudi[2*], Rasool Roozegarz[3], G.G. Hamedani[4], Nadeem Shafique Butt[5]

*Corresponding author

1. Department of Statistics, Yazd University, P.O. Box 89175-741, Yazd, Iran, z.ferry21@gmail.com
2. Department of Statistics, Yazd University, P.O. Box 89175-741, Yazd, Iran, emahmoudi@yazd.ac.ir
3. Department of Statistics, Yazd University, P.O. Box 89175-741, Yazd, Iran, rroozegar@yazd.ac.ir
4. Department of Mathematical and Statistical Sciences, Marquette University, WI 53201-1881, Milwaukee, USA, g.hamedani@mu.edu
5. Department of Family and Community Medicine, King Abdulaziz University, KSA, nshafique@kau.edu.sa

Abstract

In the past couple of years, statistical models have been extensively used in applied areas for analyzing real data sets. However, in numerous situations, the traditional distributions are not flexible enough to cater different aspects of the real phenomena. For example, (i) in the practice of reliability engineering and bio-medical analysis, some distributions provide the best fit to the data having monotonic failure rate function, but fails to provide the best fit to the data having non-monotonic failure rate function, (ii) some statistical distributions provide the best fit for small insurance losses, but fails to provide an adequate fit to large claim size data, and (iii) some distributions do not have closed forms causing difficulties in estimation process. To address the above issues, therefore, several methods have been suggested to improve the flexibility of the classical distributions. In this article, we investigate some of the former methods of generalizing the existing distributions. Further, we propose nineteen new methods of extending the classical distributions to obtain flexible models suitable for modeling data in applied fields. We also provide certain characterizations of the newly proposed families. Finally, we provide a comparative study of the newly proposed and some other existing well-known models via analyzing three real data sets from three different disciplines such as reliability engineering, medical and financial sciences.

Key Words: Weibull distribution; Families of distributions; Developments of new families; Characterizations; Monti Carlo simulation; Comparative study.

Mathematical Subject Classification: 60E05, 62E15.

1. Introduction

Speaking broadly, the modeling of lifetime testing data, bio-medical data and financial data etc. is an important research topic. Statistical distributions are utilized quite effectively for modeling such data sets. The distributions that are frequently used include: Exponential, Rayleigh, Weibull, Beta, Gamma, Lognormal, Pareto, Lomax and Burr, for more detail we refer to Zichuan et al. (2020). However, these traditional distributions are not flexible enough for countering complex form of data sets. For example, in lifetime testing and bio-medical analysis, the data sets are usually unimodal, modified unimodal or having bath-tub shape (U-shaped) failure rate, see Demicheli et al. (2004), Lai and Xie (2006), Zajicek (2011) and Almalki and Yuan (2013). Hence, in such cases, the utilization of the Exponential, Rayleigh, Weibull or Lomax distributions may not be a good choice to apply. Whereas, the Gamma, Beta and lognormal distributions do not have closed forms of cumulative distribution function (cdf) causing difficulties in estimating the model parameters.

Furthermore, in financial and actuarial risk management problems, the data sets are usually unimodal, skewed to the right and possess thick right tail, for some details see, Azzalini et al. (2002), Coorey and Ananda (2005), Vernic (2006) and Eling (2012), among others. The distributions that exhibit such characteristics can be used quite effectively to model insurance loss data to estimate the business risk level. The distributions commonly used in the literature include Pareto by Coorey and Ananada (2005), Lomax by Scollnik (2007), Burr by Nadarajah and Bakar (2014) and Weibull by Bakar et al. (2015) which are particularly appropriate for modeling of insurance losses, financial returns, file sizes on the network servers etc. for detail see Resnick (1997). Unfortunately, these distributions are subject of some sort of deficiencies. For example, the Pareto distribution, due to the monotonically decreasing shape of the density, does not provide a reasonable fit in many applications. Whereas, the Weibull model covers better the behavior of small losses, but fails to cover the behavior of large losses.

Hence, there are only few probability distributions which accommodate some aforementioned characteristics. None of them is flexible enough to provide greater accuracy in data fitting. Therefore, to bring flexibility to a model by introducing additional parameter(s) having closed form for cdf and capable of modeling heavy tailed data is an important research topic. In this regard, a serious attempt has been made and still growing rapidly to develop new statistical models having closed forms for the survival function (sf), hazard function (hf) and heavier tail than the exponential distribution.

In the recent era, the researchers have shown an increased interest in defining new families of distributions by incorporating one or more parameters. The new families have been introduced through many different approaches introducing, location, scale, shape and transmuted parameters, etc. In view of these, various modifications and extensions have been proposed by adding new parameter(s) to these classical distributions. These generalizations are mainly based on, but not limited to, the following approaches (i) transformation method, (ii) composition of two or more distributions, (iii) compounding of distributions, and (iv) finite mixture of the classical distributions.

In this article, we investigate some methods that have been used in proposing new distributions. Furthermore, we carry this branch of distribution theory and propose some new methods to introduce new distributions.

The rest of this article is organized as follows: in Section 2, some former methods of extending the existing distributions are investigated. Further developments in the field of distribution theory are provided in Section 3. Certain characterizations of the newly proposed family are provided in Section 4. A special sub-model of a newly proposed family is discussed in Section 5. Maximum likelihood estimation and simulation study are provided in Section 6. Comparative study is presented in Section 7. Finally, the article is concluded in Section 8.

2. An investigation of the former methods

In this section, we discuss some former approaches that have been used to modify the existing distributions to obtain more flexible models.

2.1. The power series approach

One of the most interesting methods of generalizing the distributions is the power series (PS) extension proposed by Noack (1950). The PS class of distribution is defined by

$$G(x;\theta,\xi)=\frac{C(\theta F(x;\xi))}{C(\theta)}, \qquad \theta\in(0,s),\ \xi>0,\ x\in\mathbb{R}, \tag{1}$$

where $C(\theta)=\sum_{n=1}^{\infty}a_n\theta^n,\quad a_n>0.$

2.2. The exponentiated family

Another prominent approach of adding a parameter to the existing distributions is exponentiation. The exponentiated family pioneer by Mudholkar and Srivastava (1993), is defined by the following cdf

$$G(x;a,\xi) = F(x;\xi)^a, \qquad a,\xi>0,\ x\in\mathbb{R}, \tag{2}$$

where a is the additional shape parameter.

2.3. The Marshall-Olkin family

Marshall and Olkin (1997) pioneered a new simple approach of introducing a single parameter to a family of distributions. The cdf of the Marshall-Olkin (MO) family is given by

$$G(x;\sigma,\xi) = \frac{F(x;\xi)}{\sigma+(1-\sigma)F(x;\xi)}, \qquad \sigma,\xi>0,\ x\in\mathbb{R}, \tag{3}$$

with additional parameter σ.

2.4. The odd log-logistic-G approach

Gleaton and Lynch (2006) proposed a new method of transformation, called Odd log-logistic-G (OLL-G) approach to add a single additional shape parameter a to extend distributions. The cdf of the OLL-G family is given by

$$G(x;a,\xi) = \frac{F(x;\xi)^a}{F(x;\xi)^a+\bar{F}(x;\xi)^a}, \qquad a,\xi>0,\ x\in\mathbb{R}, \tag{4}$$

where $\bar{F}(x;\xi) = 1 - F(x;\xi)$.

2.5. The transmuted family

Shaw and Buckley (2009) pioneered an interesting method of adding a parameter to a family of distributions. The distribution function of the transmuted family is given by

$$G(x;\lambda,\xi) = (1+\lambda)F(x;\xi) - \lambda F(x;\xi)^2, \quad \xi>0,\ |\lambda|\le 1, x\in\mathbb{R}, \tag{5}$$

with additional transmuted parameter λ.

2.6. The Kumaraswamy-G family

The Kumaraswamy distribution introduced by Kumaraswamy (1980), later on, Cordeiro and Castro proposed the Kumaraswamy-G family defined by

$$G(x;a,b,\xi) = 1-\left(1-F(x;\xi)^a\right)^b, \qquad a,b,\xi>0,\ x\in\mathbb{R}, \tag{6}$$

where a and b are the additional shape parameters.

2.7. The T-*X* family approach

Alzaatreh et al. (2013) proposed another prominent approach of generating families of continuous distributions, called T-*X* family. The cdf of the T-*X* family is obtained as

$$G(x,\eta,\xi) = \int_a^{W[F(x;\xi)]} v(t;\eta)\,dt, \tag{7}$$

where $W[F(x;\xi)]$ satisfy some conditions, for detail see Alzaatreh et al. (2013).

2.8. The alpha power transformation

Mahdavi and Kundu (2017) proposed a new method for introducing statistical distributions via the cdf given by

$$G\left(x;\alpha,\xi\right)=\frac{\alpha^{F(x;\xi)}-1}{\alpha-1},\qquad \alpha,\xi>0,\ \alpha\neq 1,\ x\in\mathbb{R}, \tag{8}$$

with additional parameter α.

2.9. The extended alpha power transformed family

Recently, Ahmad et al. (2018) proposed a new family of lifetime distributions, called the extended alpha power transformed (EAPT) family with the cdf

$$G\left(x;\alpha,\xi\right)=\frac{\alpha^{F(x;\xi)}-e^{F(x;\xi)}}{\alpha-e},\qquad \xi>0,\ \alpha\neq e,\alpha>e,\ x\in\mathbb{R}, \tag{9}$$

with additional parameter α.

2.10. A new sine-G family approach

Mahmood and Chesneau (2019) proposed a new family of distribution by using a trigonometric function called, a new sine-G family define by

$$G\left(x;\xi\right)=\sin\left(\frac{\pi}{4}F\left(x;\xi\right)\left(F\left(x;\xi\right)+1\right)\right),\qquad \xi>0,\ x\in\mathbb{R}. \tag{10}$$

For further detail about the new developments in distribution theory, we refer the interested readers to Tahir and Nadarjah (2015), Tahir and Cordeiro (2016) and Ahmad et al. (2019).

3. Further developments

As we mentioned in Sections 1 and 2, the distribution theory has received serious consideration in the literature to propose novel methods of introducing new distributions. We carry further this branch of distribution theory and propose some new methods for generating new distributions.

3.1. The cosine-*X* family

In this sub-section, we introduce a new method of proposing distributions, called the cosine-*X* (C-*X*) family. If a random variable *X* follows the C-*X* family, its cdf is given by

$$G\left(x;\sigma,\xi\right)=\cos\left(\frac{\frac{\pi}{2}\left[1-F\left(x;\xi\right)\right]}{1-\bar{\sigma}F\left(x;\xi\right)}\right),\qquad \sigma,\xi>0,\ x\in\mathbb{R}, \tag{11}$$

where $\bar{\sigma}=(1-\sigma)$. When $\sigma=1$, then (11) is a special case of Sine Kumaraswamy-G (SKu-G) family of Chesneau and Jamal (2019). The cdf of the SKu-G family is given by

$$G\left(x;a,b,\xi\right)=\cos\left(\frac{\pi}{2}\left[1-F\left(x;\xi\right)^{a}\right]^{b}\right),\qquad a,b,\xi>0,\ x\in\mathbb{R}.$$

The pdf corresponding to (11) is given by

$$g\left(x;\sigma,\xi\right)=\frac{\sigma\pi}{2\left(1-\bar{\sigma}F\left(x;\xi\right)\right)^{2}}f\left(x;\xi\right)\sin\left(\frac{\frac{\pi}{2}\left[1-F\left(x;\xi\right)\right]}{1-\bar{\sigma}F\left(x;\xi\right)}\right),\quad x\in\mathbb{R}. \tag{12}$$

The pdf is easily obtained by simple differentiation of the cdf. Henceforth, for the other families, we just provide the distribution function.

3.2. The cosine exponentiated-*X* family

In this sub-section, we introduce another new method of proposing statistical distributions, called the cosine exponentiated-*X* (CEx-*X*) family. If a random variable *X* follows the CEx-*X* family, its cdf is given by

$$G\left(x;a,\sigma,\xi\right)=\cos\left(\frac{\frac{\pi}{2}\left[1-F\left(x;\xi\right)^{a}\right]}{1-\bar{\sigma}F\left(x;\xi\right)^{a}}\right),\quad a,\sigma,\xi>0,\ x\in\mathbb{R}. \tag{13}$$

For $a=1$, the expression (13) reduces to (11).

3.3. The extended cosine exponentiated-*X* family

Another extension of (11) can easily be obtained by adding another shape parameter to CEx-*X* family. The new extension of (11) may be named as the extended cosine exponentiated-*X* (ECEx-*X*) family. The cdf of the ECEx-*X* family is given by

$$G\left(x;a,b,\sigma,\xi\right)=\cos\left(\frac{\frac{\pi}{2}\left[1-F\left(x;\xi\right)^{a}\right]^{b}}{1-\bar{\sigma}F\left(x;\xi\right)^{a}}\right),\quad a,b,\sigma,\xi>0,\ x\in\mathbb{R}. \tag{14}$$

The expression (14) is very interesting and reduces to (i) CEx-*X* family for *b*=1 and (ii) SKu-G family for $\sigma=1$.

3.4. The new logistic cosine-*X* family

In this sub-section, we use the T-*X* family approach and define a new family, called the new logistic cosine-*X* (NLC-*X*) family. If *T* follows the logistic distribution with parameter $\lambda>0$, say logistic (λ), its cdf is given by $V_L\left(t\right)=\left(1+e^{-\lambda t}\right)^{-1},\quad -\infty<t<\infty$, and pdf $v_L\left(x;\lambda\right)=\lambda e^{-\lambda t}\left(1+e^{-\lambda t}\right)^{-2}$. Now, by taking $W\left[F\left(x\right)\right]=\log\left\{-\log\left(\frac{\cos\left(\frac{\pi}{2}F\left(x;\xi\right)\right)}{1-\left(1-\theta\right)F\left(x;\xi\right)}\right)\right\}$ and $v(t)$ follow the logistic distribution. Then, using (7), the cdf of the NLC-*X* family is given by

$$G\left(x;\theta,\lambda,\xi\right)=\left[1+\left\{-\log\left(\frac{\cos\left(\frac{\pi}{2}F\left(x;\xi\right)\right)}{1-\left(1-\theta\right)F\left(x;\xi\right)}\right)\right\}^{-\lambda}\right]^{-1}, \tag{15}$$

where $\theta,\lambda,\xi>0,\ x\in\mathbb{R}$.

3.5. The type-I cosine-*X* family

In this sub-section, we introduce a new prominent approach, called Type-I cosine-*X* (TIC-*X*) family. If a random variable *X* follows the TIC-*X* family, its cdf is given by

$$G\left(x;\xi\right)=\frac{\exp\left\{\cos\left(\frac{\pi}{2}\left[1-F\left(x;\xi\right)\right]\right)\right\}-1}{e-1},\quad \xi>0,\ x\in\mathbb{R}. \tag{16}$$

3.6. The type-I cosine exponentiated-*X* family

In this sub-section, we introduce an extension of the TIC-*X* family, called Type-I cosine exponentiated-*X* (TICEx-*X*) family. The cdf of the TICEx-*X* random variable is given by

$$G\left(x;a,\xi\right)=\frac{\exp\left\{\cos\left(\frac{\pi}{2}\left[1-F\left(x;\xi\right)^{a}\right]\right)\right\}-1}{e-1},\quad a,\xi>0,\ x\in\mathbb{R}. \tag{17}$$

For $a=1$, the expression (17) reduces to (16).

3.7. The alpha power transformed cosine-*X* family

In this sub-section, a new extended family of distribution is introduced. The newly proposed method is called the alpha power transformed cosine-*X* (APTC-*X*) family. The distribution function of the APTC-*X* family is defined by

$$G\left(x;\alpha,\xi\right)=\frac{\alpha^{\left\{\cos\left(\frac{\pi}{2}\left[1-F(x;\xi)\right]\right)\right\}}-1}{\alpha-1},\quad \alpha,\xi>0,\ \alpha\neq1,\ x\in\mathbb{R}. \tag{18}$$

3.8. The alpha power transformed cosine exponentiated-*X* family

In this sub-section, a generalized version of the (18) is proposed. The new generalized family is called the alpha power transformed cosine exponentiated-*X* (APTCEx-*X*) family. The distribution function of the APTCEx-*X* family is defined by

$$G\left(x;\alpha,a,\xi\right)=\frac{\alpha^{\left\{\cos\left(\frac{\pi}{2}\left[1-F(x;\xi)^{a}\right]\right)\right\}}-1}{\alpha-1},\qquad a,\alpha,\xi>0,\ \alpha\neq1,\ x\in\mathbb{R}. \tag{19}$$

3.9. The exponentiated cosine-*X* family

This sub-section introduces the exponentiated version of (11) called the exponentiated cosine-*X* (ExC-*X*) family. The cdf of the ExC-*X* family is given by

$$G\left(x;a,\sigma,\xi\right)=\left\{\cos\left(\frac{\frac{\pi}{2}\left[1-F\left(x;\xi\right)\right]}{1-\bar{\sigma}F\left(x;\xi\right)}\right)\right\}^{a},\quad a,\sigma,\xi>0,\ x\in\mathbb{R}. \tag{20}$$

For $a=1$, the expression (20) reduces to (11).

3.10. The Kumaraswamy cosine-*X* family

This sub-section introduces the Kumaraswamy version of (11) called the Kumaraswamy cosine-*X* (KuC-*X*) family of distributions. The distribution function of the KuC-*X* family is given by

$$G\left(x;a,b,\sigma,\xi\right)=1-\left[1-\left\{\cos\left(\frac{\frac{\pi}{2}\left[1-F\left(x;\xi\right)\right]}{1-\bar{\sigma}F\left(x;\xi\right)}\right)\right\}^{a}\right]^{b},\ \ a,b,\sigma,\xi>0,\ x\in\mathbb{R}. \tag{21}$$

For $a=b=1$, the expression (21) reduces to (11).

3.11. The Marshall-Olkin cosine-*X* family

In this sub-section, we propose the Marshall-Olkin version of the C-*X* family, called the Marshall-Olkin cosine-*X* (MOC-*X*) family given by

$$G\left(x;\sigma,\xi\right)=\frac{\cos\left(\frac{\pi}{2}\left[1-F\left(x;\xi\right)\right]\right)}{\sigma+\left(1-\sigma\right)\cos\left(\frac{\pi}{2}\left[1-F\left(x;\xi\right)\right]\right)},\quad \sigma,\xi>0,\ x\in\mathbb{R}. \tag{22}$$

3.12. The Marshall-Olkin cosine exponentiated-*X* family

In this sub-section, the extended version of (22), called the Marshall-Olkin cosine exponentiated-*X* (MOCEx-*X*) family is proposed. The cdf of the MOCEx-*X* family is given by

$$G\left(x;a,\sigma,\xi\right)=\frac{\cos\left(\frac{\pi}{2}\left[1-F\left(x;\xi\right)^{a}\right]\right)}{\sigma+\left(1-\sigma\right)\cos\left(\frac{\pi}{2}\left[1-F\left(x;\xi\right)^{a}\right]\right)},\quad a,\sigma,\xi>0,\ x\in\mathbb{R}. \tag{23}$$

For $a=1$, the expression (23) reduces to (22).

3.13. The transmuted cosine-*X* family

In this sub-section, we propose another new family, called the transmuted cosine-*X* (TC-*X*) family. The cdf of the TC-*X* family is given by

$$G(x;\lambda,\xi) = (1+\lambda)\cos\left(\frac{\pi}{2}(1-F(x;\xi))\right) - \lambda\left\{\cos\left(\frac{\pi}{2}(1-F(x;\xi))\right)\right\}^2, \tag{24}$$

where $\xi > 0,\ |\lambda| \leq 1,\ x \in \mathbb{R}$.

3.14. The transmuted cosine exponentiated-*X* family

In this sub-section, we introduce the extended version of (24), called the transmuted cosine exponentiated-*X* (TCEx-*X*) family. The cdf of the TCEx-*X* is given by

$$G(x;a,\lambda,\xi) = (1+\lambda)\cos\left(\frac{\pi}{2}(1-F(x;\xi)^a)\right) - \lambda\left\{\cos\left(\frac{\pi}{2}(1-F(x;\xi)^a)\right)\right\}^2, \tag{25}$$

where $a, \xi > 0,\ |\lambda| \leq 1,\ x \in \mathbb{R}$
For $a = 1$, the expression (25) reduces to (24).

3.15. The *arc*cosine-*X* family

In this sub-section, we propose another novel approach of proposing new distributions, called the *arc*cosine-*X* (AC-*X*) family. The cdf of the AC-*X* family is given by

$$G(x;\xi) = 1 - \frac{2}{\pi} arc\cos(F(x;\xi)), \qquad \xi > 0,\ x \in \mathbb{R}. \tag{26}$$

3.16. The *arc*cosine exponentiated-*X* family

In this sub-section, we introduce an extension of the AC-*X* family, called the *arc*cosine exponentiated-*X* (ACEx-*X*) family. The cdf of the ACEx-*X* family is given by

$$G(x;a,\xi) = 1 - \frac{2}{\pi} arc\cos(F(x;\xi)^a), \qquad a,\xi > 0,\ x \in \mathbb{R}. \tag{27}$$

For $a = 1$, the expression (27) reduces to (26).

3.17. The inverted type-I cosine-*X* family

Here, we introduce the inverted version of Type-I cosine-*X* (TIC-*X*) family of distributions, called the invereted Type-I cosine-*X* (ITIC-*X*) family. If a random variable *X* follows the ITIC-*X* family, its cdf is given by

$$G(x;\xi) = 1 - \frac{2}{\pi} arc\cos\log\{(e-1)F(x;\xi)+1\}, \quad \xi > 0,\ x \in \mathbb{R}. \tag{28}$$

3.18. The inverted type-I cosine exponentiated-*X* family

Here, we propose an extension of the invereted Type-I cosine-*X* (ITIC-*X*) family, called the invereted Type-I cosine exponentiated-*X* (ITICEx-*X*) family. If a random variable *X* follows the ITICEx-*X* family, its cdf is given by

$$G(x;a,\xi) = 1 - \frac{2}{\pi} arc\cos\log\{(e-1)F(x;\xi)^a+1\}, \quad a,\xi > 0,\ x \in \mathbb{R}. \tag{29}$$

For $a = 1$, the expression (29) reduces to (28).

3.19. The exponentiated *arc*cosine-*X* family

In this sub-section, we introduce the exponentiated version of the AC-*X* family, called the exponentiated *arc*cosine (ExAC-*X*) family. The cdf of the ExAC-*X* family is given by

$$G(x;a,\xi)=\left(1-\frac{2}{\pi}arc\cos\left(F(x;\xi)\right)\right)^{a},\qquad a,\xi>0,\ x\in\mathbb{R}. \tag{30}$$

For $a=1$, the expression (30) reduces to (26).

4. Characterizations

This section is devoted to the characterizations of the cosine-*X* distribution based on: (i) a simple relationship between two truncated moments and (ii) reverse hazard function. It should be mentioned that for the characterizations (i) the cdf is not required to have a closed form. We present our characterization (i) and (ii) in two subsections.

4.1. Characterizations based on two truncated moments

The first characterization result employs a theorem due to Glänzel (1987); see Theorem 4.1 below. Note that the result holds also when the interval *H* is not closed. As shown in Glänzel (1990), this characterization is stable in the sense of weak convergence. We like to mention that due to the nature of the cdf, our characterizations may be the only possible ones.

Theorem 4.1. Let $(\Omega,\mathcal{F},\mathbf{P})$ be a given probability space and let *H* =[*d; e*] be an interval for some *d*<*e*$(d=-\infty;\ e=\infty$ might as well be allowed). Let *X*: $\Omega\rightarrow H$ be a continuous random variable with the distribution function *F* and let q_1 and q_2 be two real functions defined on *H* such that

$$E\left(q_2\left(X\right)|X\geq x\right)=E\left(q_1\left(X\right)|X\geq x\right)\eta\left(x\right),\qquad x\in H,$$

is defined with some real function η. Assume that $q_1,q_2\in C^1(H),\eta\in C^2(H)$ and *F* is twice continuously differentiable and strictly monotone function on the set *H*. Finally, assume that the equation $\eta q_1=q_2$ has no real solution in the interior of *H*. Then *F* is uniquely determined by the functions$q_1,\ q_2$ and η particularly

$$F(x)=\int_a^x C\left|\frac{\eta^{/}(u)}{\eta(u)\,q_1(u)-q_2(u)}\right|\exp\left(-s\left(u\right)\right)du,$$

where the function s is a solution of the differential equation $s^{/}=\frac{\eta^{/}q_1}{\eta q_1-q_2}$ and *C* is the normalization constant, such that $\int_H dF=1$.

Remark 4.1. The goal in Theorem 4.1 is to have $\eta(x)$ as simple as possible.

Proposition 4.1. Let X: $\Omega\rightarrow\mathbb{R}$ be a continuous random variable and let $q_1(x)=\left\{\sin\left(\frac{\frac{\pi}{2}(1-F(x;\xi))}{1-\bar{\sigma}F(x;\xi)}\right)\right\}^{-1}$ and $q_2(x)=q_1(x)\left\{1-\bar{\sigma}F(x;\xi)\right\}^{-1}$ for $x\in\mathbb{R}$. The random variable *X* has pdf (12) if and only if the function η defined in Theorem 4.1 has the form

$$\eta(x)=\frac{1}{2}\left[\frac{1}{\sigma}-\left\{1+\bar{\sigma}F(x;\xi)\right\}^{-1}\right],\qquad x\in\mathbb{R}.$$

Proof. Let X be a random variable with pdf (12), then

$$\left(1-G\left(x;\sigma,\xi\right)\right)E\left(q_1\left(X\right)|X\geq x\right)=\frac{\sigma\pi}{2\bar{\sigma}}\left\{\frac{1}{\sigma}-\left(1+\bar{\sigma}F\left(x;\xi\right)\right)^{-1}\right\},\ x\in\mathbb{R},$$

and

$$\left(1-G\left(x;\sigma,\xi\right)\right)E\left(q_2\left(X\right)|X\geq x\right)=\frac{\sigma\pi}{4\bar{\sigma}}\left\{\frac{1}{\sigma^2}-\left(1+\bar{\sigma}F\left(x;\xi\right)\right)^{-2}\right\},\ x\in\mathbb{R},$$

and finally

$$\eta\left(x\right)q_{1}\left(x\right)-q_{2}\left(x\right)=\frac{q_{1}\left(x\right)}{2}\left\{\frac{1}{\sigma}-\left(1+\bar{\sigma}F\left(x;\xi\right)\right)^{-1}\right\}\neq 0,\qquad \text{for }\ x\in\mathbb{R}.$$

Conversely, if η is given as above

$$s^{/}\left(x\right)=\frac{\eta^{/}\left(x\right)q_{1}\left(x\right)}{\eta\left(x\right)q_{1}\left(x\right)-q_{2}\left(x\right)}=\frac{\bar{\sigma}\sigma f\left(x;\xi\right)}{\left(1+\bar{\sigma}F\left(x;\xi\right)\right)-\sigma},\qquad x\in\mathbb{R}.$$

and hence

$$s\left(x\right)=-\sigma\log\left[\left(1+\bar{\sigma}F\left(x;\xi\right)\right)-\sigma\right],\qquad x\in\mathbb{R}.$$

No, in view of Theorem 4.1, *X* has density (12).

Corollary 4.1. Let *X*: $\Omega\rightarrow\mathbb{R}$ be a continuous random variable and let $q_1\left(x\right)$ be as in Proposition 4.1. The random variable *X* has pdf (12) if and only if there exist functions $q_2\left(x\right)$ and $\eta\left(x\right)$ defined in Theorem 4.1 satisfying the following differential equation

$$\frac{\eta^{/}\left(x\right)q_{1}\left(x\right)}{\eta\left(x\right)q_{1}\left(x\right)-q_{2}\left(x\right)}=\frac{\sigma\bar{\sigma}f\left(x;\xi\right)}{\left(1+\bar{\sigma}F\left(x;\xi\right)\right)-\sigma},\qquad x\in\mathbb{R}.$$

Corollary 4.2. The general solution of the differential equation in Corollary 4.1 is

$$\eta\left(x\right)=\left[\left(1+\bar{\sigma}F\left(x;\xi\right)\right)-\sigma\right]^{-1}\left[-\int\sigma\bar{\sigma}f\left(x\right)\left(h\left(x\right)\right)^{-1}g\left(x\right)dx+D\right],$$

where *D* is a constant. Note that a set of functions satisfying the above differential equation is given in Proposition 4.1 with *D*=1/2. However, it should be also noted that there are other triplets $\left(q_1,q_2,\eta\right)$ satisfying the conditions of Theorem 4.1.

4.2. Characterization in terms of the reverse hazard function

The reverse hazard function r_G of a twice differentiable distribution function, G, is defined as

$$r_{G}\left(x\right)=\frac{g\left(x\right)}{G\left(x\right)},\qquad x\in\text{support } of\ G.$$

In this subsection we present a characterization of the C-*X* distributions in terms of the reverse hazard function.

Proposition 4.2. Let *X*: $\Omega\rightarrow\mathbb{R}$ be a continuous random variable. The random variable *X* has pdf (12) if and only if its reverse hazard function r_G satisfies the following differential equation

$$r_{G}^{/}\left(x\right)=\frac{f^{/}\left(x;\xi\right)}{f\left(x;\xi\right)}r_{G}\left(x\right)=\frac{\pi\sigma f\left(x;\xi\right)}{2}\frac{d}{dx}\left(\frac{\tan\left(\frac{\frac{\pi}{2}\left(1-F\left(x;\xi\right)\right)}{\left(1-\bar{\sigma}F\left(x;\xi\right)\right)}\right)}{\left(1-\bar{\sigma}F\left(x;\xi\right)\right)}\right),\quad x\in\mathbb{R},$$

with boundary condition $\lim_{x\rightarrow\infty}r_G\left(x\right)=0$.

Proof. If X has pdf (12), then clearly the above differential equation holds. If the differential equation holds, then

$$\frac{d}{dx}\left(\left(f\left(x;\xi\right)\right)^{-1}r_{G}\left(x\right)\right)=\frac{\pi\sigma}{2}\frac{d}{dx}\left(\frac{\tan\left(\frac{\frac{\pi}{2}\left(1-F\left(x;\xi\right)\right)}{\left(1-\bar{\sigma}F\left(x;\xi\right)\right)}\right)}{\left(1-\bar{\sigma}F\left(x;\xi\right)\right)}\right),\quad x\in\mathbb{R},$$

or

$$\left(\left(f\left(x;\xi\right)\right)^{-1}r_{G}\left(x\right)\right)=\frac{\pi\sigma}{2}\left(\frac{\tan\left(\frac{\frac{\pi}{2}\left(1-F\left(x;\xi\right)\right)}{\left(1-\bar{\sigma}F\left(x;\xi\right)\right)}\right)}{\left(1-\bar{\sigma}F\left(x;\xi\right)\right)}\right),\qquad x\in\mathbb{R},$$

from which we arrive at the hazard function corresponding to the pdf (12).

5. A special sub-model of the cosine-*X* family

In this section, we introduce a special sub-model of the C-*X* family, called the cosine Weibull (C-W) distribution. Let $F(x;\xi)$ and $f(x;\xi)$ be the cdf and pdf of the two parameters Weibull distribution given by $F(x;\xi) = 1 - e^{-\gamma x^{\alpha}}, \quad \xi, x \geq 0$, and $f(x;\xi) = \alpha\gamma x^{\alpha-1} e^{-\gamma x^{\alpha}}$, respectively, where $\xi = (\alpha, \gamma)$. Then, the cdf of the C-W distribution has the following expression

$$G(x;\sigma,\xi) = \cos\left(\frac{\frac{\pi}{2} e^{-\gamma x^{\alpha}}}{1-\bar{\sigma}\left(1-e^{-\gamma x^{\alpha}}\right)}\right), \qquad \sigma, \xi > 0, \ x \geq 0. \tag{31}$$

The pdf and sf corresponding (31), are given, respectively by

$$g(x;\sigma,\xi) = \frac{\pi\alpha\sigma\gamma}{2\left(1-\bar{\sigma}\left(1-e^{-\gamma x^{\alpha}}\right)\right)^2} x^{\alpha-1} e^{-\gamma x^{\alpha}} \sin\left(\frac{\frac{\pi}{2} e^{-\gamma x^{\alpha}}}{1-\bar{\sigma}\left(1-e^{-\gamma x^{\alpha}}\right)}\right), \quad x > 0, \tag{32}$$

and

$$S(x;\sigma,\xi) = 1 - \cos\left(\frac{\frac{\pi}{2} e^{-\gamma x^{\alpha}}}{1-\bar{\sigma}\left(1-e^{-\gamma x^{\alpha}}\right)}\right), \qquad x \geq 0.$$

For selected values of the model parameters, some possible shapes for the density function of the C-W distribution are sketched in Figure 1.

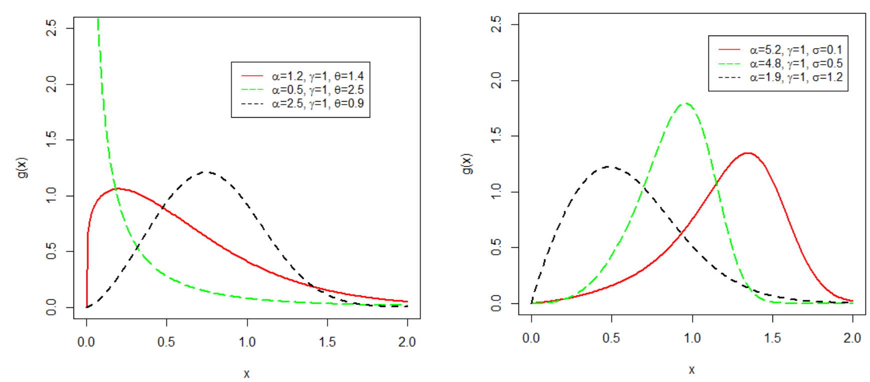

Figure 1: Different plots for the pdf of the C-W distribution

For different values of the model parameters, plots of the simulated histograms of C-W distribution are sketched in Figures 2-4.

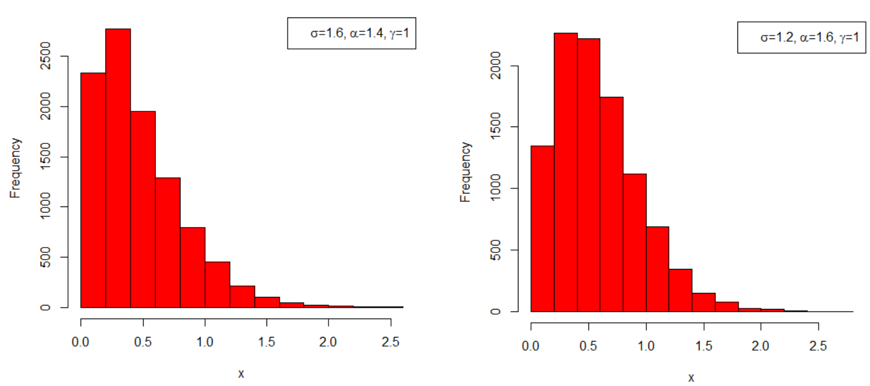

Figure 2: Different positively skewed plots for the simulated histogram of the C-Weibull distribution.

 DOI: https://doi.org/10.18187/pjsor.v18i1.3908

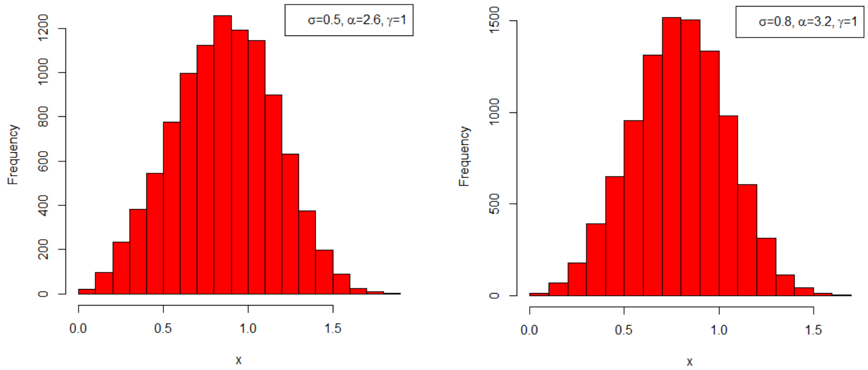

Figure 3: Different symmetrical plots for the simulated histogram of the C-Weibull distribution.

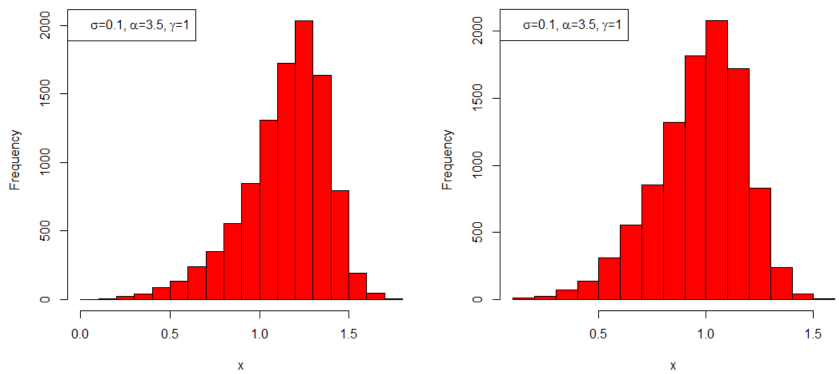

Figure 4: Different negatively skewed plots for the simulated histogram of the C-Weibull distribution.

6. Estimation and simulation study

In this section, we obtain the maximum likelihood estimates for the parameters of the C-X distribution and provide a simulation study to evaluate these estimators.

6.1. Maximum likelihood estimation

In this subsection, we determine the maximum likelihood (ML) estimates of the parameters of the C-X family. Let $x_1, x_2, \cdots, x_n$ be the observed values taken from the C-X distribution with parameters σ and ξ. The total log-likelihood function corresponding to (12) is given by

$$\begin{aligned}\log L\left(x_i;\sigma,\xi\right) = n\log\left(\tfrac{\pi}{2}\right) + n\log\sigma + \sum_{i=1}^{n}\log f\left(x_i;\xi\right) - 2\sum_{i=1}^{n}\log\left\{1-\bar{\sigma}F\left(x_i;\xi\right)\right\} \\ + \sum_{i=1}^{n}\log\left\{\sin\left(\tfrac{\frac{\pi}{2}\left[1-F\left(x_i;\xi\right)\right]}{1-\bar{\sigma}F\left(x_i;\xi\right)}\right)\right\}.\end{aligned} \quad (33)$$

The partial derivatives of (33) are

$$\frac{\partial}{\partial\sigma}\log L\left(x_i;\sigma,\xi\right) = \frac{n}{\sigma} - 2\sum_{i=1}^{n}\frac{F\left(x_i;\xi\right)}{B_i} - \sum_{i=1}^{n}\cot\left(\frac{\frac{\pi}{2}A_i}{B_i^2}\right)\frac{\frac{\pi}{2}A_i}{B_i^2}F\left(x_i;\xi\right),$$

and

$$\begin{aligned}\tfrac{\partial}{\partial\xi}\log L\left(x_i;\sigma,\xi\right) = \sum_{i=1}^{n}\tfrac{\partial f(x_i;\xi)/\partial\xi}{f(x_i;\xi)} + \tfrac{\pi}{2}\sum_{i=1}^{n}\cot\left(\tfrac{\frac{\pi}{2}A_i}{B_i^2}\right)\tfrac{\partial F(x_i;\xi)/\partial\xi\{\bar{\sigma}A_i-B_i\}}{B_i^2} \\ + 2\bar{\sigma}\sum_{i=1}^{n}\tfrac{\partial F(x_i;\xi)/\partial\xi}{B_i},\end{aligned}$$

where $A_i = 1 - F(x_i;\xi)$ and $B_i = 1 - \bar{\sigma}F(x_i;\xi)$. Setting $\frac{\partial}{\partial\sigma}\log L(x_i\sigma,\xi)$ and $\frac{\partial}{\partial\xi}\log L(x_i;\sigma,\xi)$ equal to zero and solving numerically these expressions simultaneously yield the maximum likelihood estimators (MLEs) of (σ,ξ).

6.2. Simulation Study

In this sub-section, we assess the behavior of the ML estimators for a finite sample of size n. Simulation study based on the C-W distribution is carried out. The random numbers are generated via quantile technique from the C-W distribution by means of R software. The simulation study is based on the following steps:

1. We generate N=500 samples of sizes n = 25, 50,..., 500 from the C-W distribution.
2. Compute the maximum likelihood estimates for the model parameters.
3. Compute the MSEs and biases given by

$MSE(n) = \frac{1}{500}\sum_{i=1}^{500}(\hat{w}-w)^2$and $Bias(n) = \frac{1}{500}\sum_{i=1}^{500}(\hat{w}-w)$ for $w = (\alpha,\sigma,\gamma)$, respectively.
The simulation results are provided in Tables 1-3. Corresponding to each Table, the graphical representation of the simulation results is also provided.

Table 1: Simulation results for different combination of parameter values of C-W distribution.

Set 1: $\alpha = 0.5, \sigma = 1.2, \gamma = 1$				
n	parameters	MLE	Bias	MSE
	$\hat{\alpha}$	0.549820	0.04982020	0.01454605
25	$\hat{\sigma}$	1.837055	0.63705522	3.11933133
	$\hat{\gamma}$	1.134269	0.13426960	0.59259310
	$\hat{\alpha}$	0.525388	0.02538884	0.00326645
100	$\hat{\sigma}$	1.490360	0.29036014	0.98324671
	$\hat{\gamma}$	0.943593	-0.0564069	0.08753167
	$\hat{\alpha}$	0.504984	0.00498437	0.00042274
300	$\hat{\sigma}$	1.241815	0.04181549	0.08711459
	$\hat{\gamma}$	0.984456	-0.0155432	0.00989175
	$\hat{\alpha}$	0.501852	0.00185200	0.00016835
500	$\hat{\sigma}$	1.214006	0.01400643	0.02613394
	$\hat{\gamma}$	0.994697	-0.0053029	0.00214465

In support of Table 1, the simulation results are displayed graphically in Figures 5 and 6.

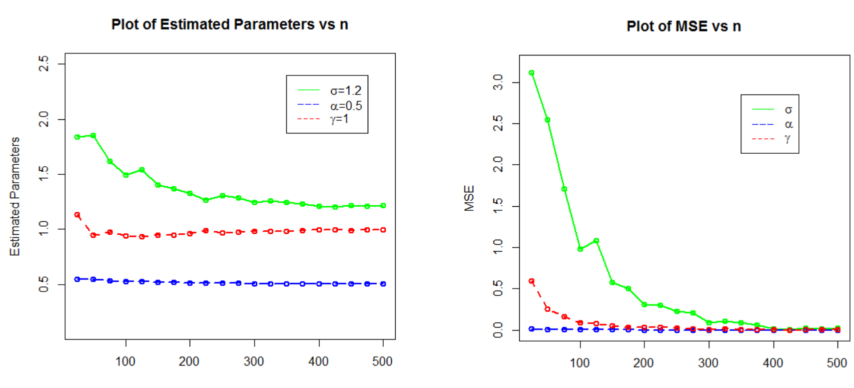

Figure 5: Plots of the estimated parameters and MSEs of the C-W distribution for Table 1.

In support of Table 2, the simulation results are displayed graphically in Figures 7 and 8.
In support of Table 3, the simulation results are displayed graphically in Figures 9 and 10.

7. Comparative Study

As we have mentioned earlier, that in the recent trend, the researchers have been developing new distributions to provide the best fit to heavy tailed data in applied areas such as reliability engineering, medical and particularly in

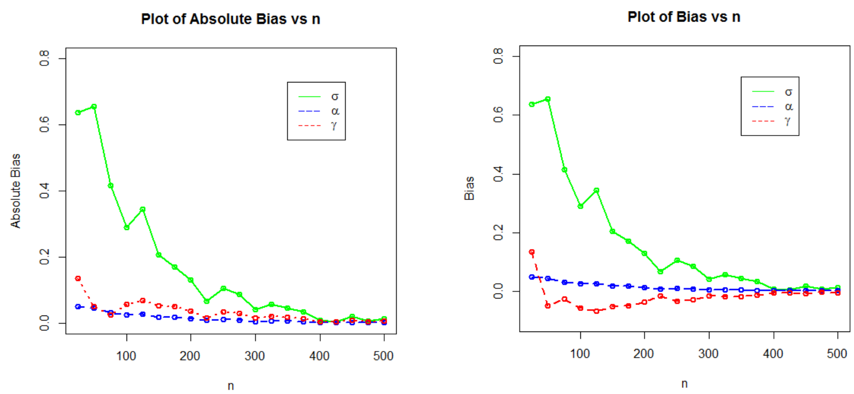

Figure 6: Plots of Biases and absolute Biases of the C-W distribution for Table 1.

Table 2: Simulation results for different combination of parameter values of C-W distribution.

Set 1: $\alpha = 1.5, \sigma = 0.8, \gamma = 1$				
n	parameters	MLE	Biases	MSE
25	$\hat{\alpha}$	1.617146	0.1171462	0.2414750
	$\hat{\sigma}$	1.446904	0.6469040	3.2687380
	$\hat{\gamma}$	1.320298	0.3202975	1.2113075
100	$\hat{\alpha}$	1.508538	0.0085384	0.1026113
	$\hat{\sigma}$	0.912081	0.1120810	0.8259544
	$\hat{\gamma}$	1.255987	0.2559871	0.8137743
300	$\hat{\alpha}$	1.524138	0.0241378	0.0379107
	$\hat{\sigma}$	0.772667	-0.0273327	0.2292647
	$\hat{\gamma}$	1.076789	0.0767885	0.1953002
500	$\hat{\alpha}$	1.540067	0.0400673	0.0253885
	$\hat{\sigma}$	0.795820	-0.034179	0.1790604
	$\hat{\gamma}$	1.034145	0.0341449	0.1377070

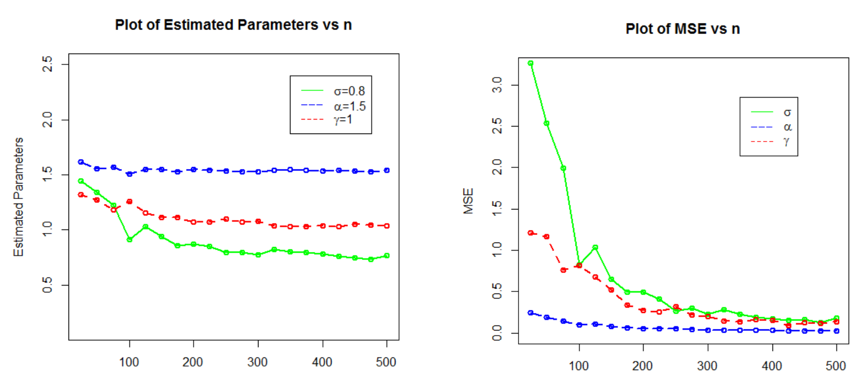

Figure 7: Plots of the estimated parameters and MSEs of the C-W distribution for Table 2.

actuarial and financial sciences. The key motivations of the heavy tail distributions are the adequately best fitting to the heavy tailed data. Therefore, in this section, we consider three real life applications of heavy tailed data from three different discipline of applied areas including medical, engineering and financial sciences. For each data set, the C-W distribution is compared with different well-known distributions and we observed that the proposed distribution outclass other competitors.

To decide about the goodness of fit among the applied distributions, we consider certain analytical measures. In this regard, we consider four discrimination measures such as the Akaike information criterion (AIC) introduced by Akaike (1974), Bayesian information criterion (BIC) of Schwarz (1978), Hannan-Quinn information criterion (HQIC) proposed by Hannan and Quinn (1979) and Consistent Akaike Information Criterion (CAIC) introduce by Bozdogan (1987). These measures are given by

- The AIC is given by

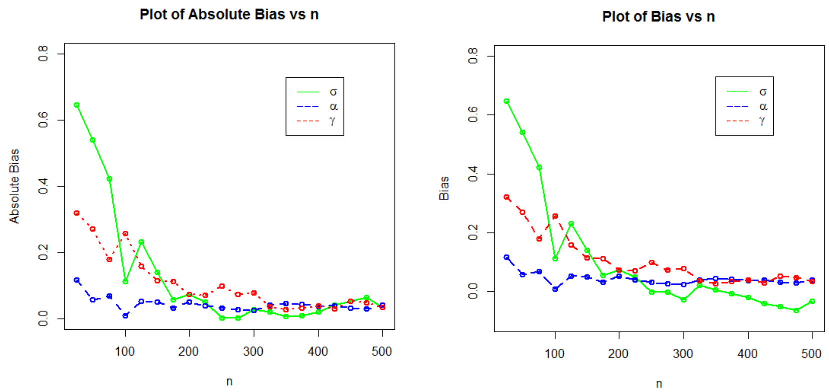

Figure 8: Plots of Biases and absolute Biases of the C-W distribution for Table 2.

Table 3: Simulation results for different combination of parameter values of C-W distribution.

Set 1: $\alpha = 1.8, \sigma = 1.4, \gamma = 1$				
n	parameters	MLE	Biases	MSE
	$\hat{\alpha}$	1.909853	0.10985320	0.2894614
25	$\hat{\sigma}$	1.941777	0.54177654	3.5950269
	$\hat{\gamma}$	1.417527	0.41752663	1.4746826
	$\hat{\alpha}$	1.832153	0.03215338	0.0939641
100	$\hat{\sigma}$	1.586087	0.18608719	1.8058100
	$\hat{\gamma}$	1.201920	0.20192041	0.5969180
	$\hat{\alpha}$	1.847510	0.04751016	0.0399412
300	$\hat{\sigma}$	1.467143	-0.0328568	0.8267543
	$\hat{\gamma}$	1.075459	0.07545884	0.2010884
	$\hat{\alpha}$	1.847284	0.04728359	0.0298447
500	$\hat{\sigma}$	1.382797	-0.0972030	0.5847341
	$\hat{\gamma}$	1.055395	0.05539532	0.1327323

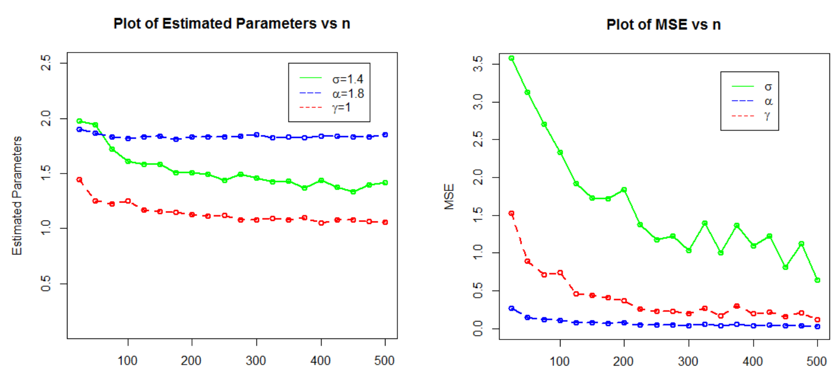

Figure 9: Plots of the estimated parameters and MSEs of the C-W distribution for Table 3.

$$AIC = 2k - 2l.$$

- The BIC is given by

$$BIC = k\log(n) - 2l.$$

- The HQIC is given by

$$HQIC = 2k\log(\log(n)) - 2l.$$

- The CAIC is given by

$$CAIC = \frac{2nk}{n-k-1} - 2l,$$

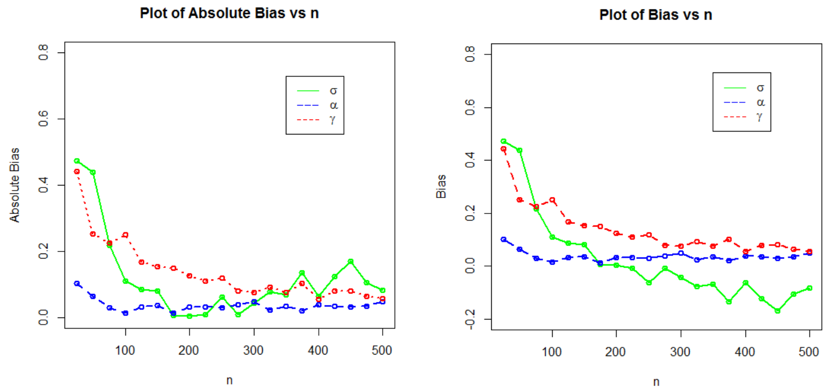

Figure 10: Plots of Biases and absolute Biases of the C-W distribution for Table 3.

where l denotes the log-likelihood function evaluated at the MLEs, k is the number of model parameters and n is the sample size. In addition to the discrimination measures, we further considered other goodness of fit measures such as Anderson Darling (AD) test statistic, Cramer-Von-Messes (CM) test statistic and Kolmogorov Simonrove (KS) test statistics with corresponding p-values. These measures are given by

- The AD test statistic

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log G\left(x_i\right) + \log\left\{1 - G\left(x_{n-i+1}\right)\right\}\right],$$

where:
$n =$ the sample size,
$i =$ the i^{th} sample, calculated when the data is sorted in ascending order.

- The CM test statistic

$$CM = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{2i-1}{2n} - G\left(x_i\right)\right]^2.$$

- The KS test statistic is given by

$$KS = sup_x\left[G_n\left(x\right) - G\left(x\right)\right],$$

where $G_n\left(x\right)$is the empirical cdf and $\sup_x$ is the supremum of the set of distances.
A distribution with lower values of these analytical measures is considered to be a good candidate model among the applied distributions for the underlying data sets. By considering these statistical tools, we observed that the C-W distribution provides the best fit compare to other distributions because the values of all selected criteria of goodness of fit are significantly small for the proposed distribution.

7.1. A real life application of bio-medical analysis

The first data set representing the remission times (in months) of a random sample of 128 bladder cancer patients, see Lee and Wang(2003). The C-W model is applied to this data set in comparison with other competitors including Weibull, flexible Weibull extended (FWE), alpha power transformed Weibull (APTW), Marshall-Olkin Weibull (MOW), transmuted Weibull (TW) and modified Weibull (MW) distributions. The distribution functions of the competitive models are:

1. Weibull

$$G\left(x;\alpha,\gamma\right) = 1 - e^{-\gamma x^{\alpha}}, \qquad x \geq 0,\ \alpha,\gamma > 0.$$

2. FWE distribution

$$G\left(x;\alpha,\sigma,\gamma\right) = 1 - \exp\left\{-e^{\sigma x^2 - \frac{\gamma}{x^{\alpha}}}\right\}, \qquad x \geq 0,\ \alpha,\sigma,\gamma > 0.$$

3. APTW distribution

$$G\left(x;\alpha_1,\alpha,\gamma\right)=\frac{\alpha_1^{\left(1-e^{-\gamma x^{\alpha}}\right)}-1}{\alpha_1-1},\qquad x\geq 0,\ \alpha_1\neq 1,\ \alpha,\gamma>0.$$

4. Marshall-Olkin Weibull (MOW) distribution

$$G\left(x;\alpha,\gamma,\sigma\right)=\frac{\left(1-e^{-\gamma x^{\alpha}}\right)}{\sigma+\left(1-\sigma\right)\left(1-e^{-\gamma x^{\alpha}}\right)},\qquad x\geq 0,\ \alpha,\gamma,\sigma>0.$$

5. TW distribution

$$G\left(x;\alpha,\gamma,\lambda\right)=\left(1+\lambda\right)\left(1-e^{-\gamma x^{\alpha}}\right)-\lambda\left(1-e^{-\gamma x^{\alpha}}\right)^2,\quad x\geq 0,$$

$\alpha,\gamma>0,\quad |\lambda|\leq 1.$

6. MW distribution

$$G\left(x;\alpha,\gamma,\theta\right)=1-e^{-\theta x-\gamma x^{\alpha}},\qquad x\geq 0,\ \alpha,\gamma,\theta>0.$$

The maximum likelihood estimates of the model for the analyzed data are presented in Table 4. The discrimination measures are provided in Table 5 and the other goodness of fit measures of the proposed and other competitive models are provided in Table 6. Form Tables 5 and 6, it is clear that the proposed distribution has lower values of these measures than the other models applied in comparison. The fitted cdf and Kaplan-Meier survival plots of the proposed model for the analyzed data set are plotted in Figure 11. Whereas, the PP plot of the proposed model and box plot of the data set are sketched in Figure 12. From 11, it is clear that the proposed model fit the estimated cdf and Kaplan Meier survival plots very closely. From 12, we can easily detect that the data has a heavy tail skewed to the right (box plot) and proposed model is closely followed the PP-plot

Table 4: Estimated values with standard error (in parenthesis) of the proposed and other competitive models for data 1.

$Dist.$	α	γ	σ	λ	α_1	θ
C-W	1.3269	0.003	12.978			
	(0.0815)	(0.0005)	(3.4361)			
Weibull	1.047	0.093				
	(0.0675)	(0.0190)				
FWE	4.332	0.720	0.541			
	(3.5347)	(0.5492)	(0.1883)			
APTW	0.014	0.016			0.014	
	(0.0865)	(0.0064)			(0.0216)	
MOW	1.268	0.877	11.829			
	(0.1308)	(0.5205)	(11.2869)			
TW	1.133	0.047		0.744		
	(0.0753)	(0.0113)		(0.2021)		
MW	1.007	0.951				0.863
	(0.0313)	(4.2501)				(4.2551)

Table 5: Discrimination measures of the proposed and other competitive models for data 1.

$Dist.$	AIC	BIC	CAIC	HQIC
C-W	825.568	834.124	825.762	829.044
Weibull	832.173	837.877	832.269	834.491
FWE	829.219	837.775	829.413	832.695
APTW	826.378	836.934	827.471	830.754
MOW	834.988	843.544	835.182	843.544
TW	829.916	838.472	827.955	836.512
MW	833.969	842.525	834.162	837.445

Table 6: Goodness of fit measures of the proposed and other competitive models for data 1.

Dist.	CM	AD	KS	p-value
NHT-W	0.026	0.168	0.041	0.981
Weibull	0.131	0.786	0.069	0.558
FWE	0.051	0.329	0.049	0.910
APTW	0.042	0.255	0.045	0.949
MOW	0.150	0.884	0.075	0.451
TW	0.086	0.516	0.058	0.768
MW	0.133	0.797	0.073	0.494

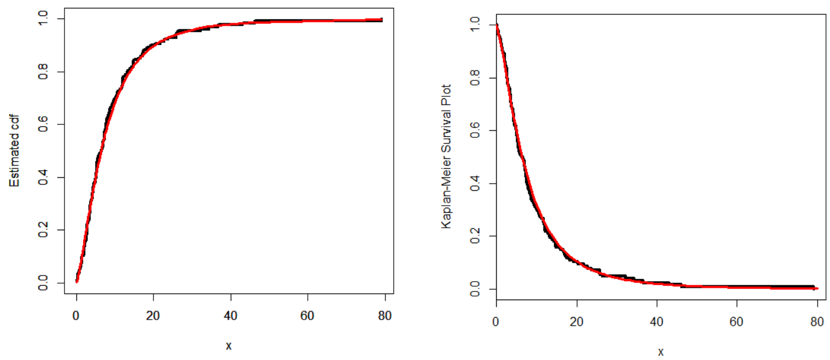

Figure 11: Estimated cdf and Kaplan-Meier survival plots of the C-W distribution for the first data set.

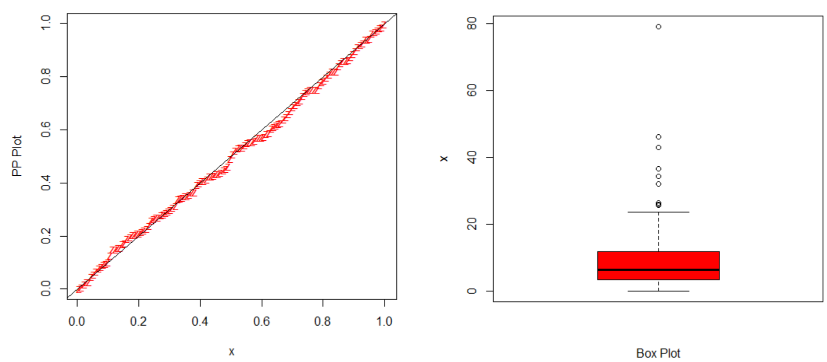

Figure 12: PP plot of the C-W distribution and box plot for the first data set.

7.2. A Real Life Application from Reliability Engineering

Here, we illustrate the C-W distribution via analyzing a heavy tailed data from reliability engineering. The second data set is taken from Algamal (2008) representing the failure time of coating machine. To show the potentiality of the proposed method, the proposed model and other competitive distributions are applied this data set and it is observed that the C-W model again outclass some well-known distributions including the extended alpha power transformed Weibull (Ex-APTW), Kumaraswamy Weibull (Ku-W) and beta Weibull (BW) distributions. The distribution functions of the competitive distributions are

1. Ex-APTW distribution

$$G(x;\alpha_1,\alpha,\gamma)=\frac{\alpha_1^{\left(1-e^{-\gamma x^{\alpha}}\right)}-e^{\left(1-e^{-\gamma x^{\alpha}}\right)}}{\alpha_1-e},\qquad x\geq 0,\ \alpha_1,\alpha,\gamma>0.$$

2. Ku-W distribution

$$G(x;\alpha,\gamma,a,b)=1-\left[1-\left(1-e^{-\gamma x^{\alpha}}\right)^{a}\right]^{b},\quad x\geq 0,\ \alpha,\gamma,a,b>0.$$

3. BW distribution

$$G\left(x;\alpha,\gamma,a,b\right)=I_{\left(1-e^{-\gamma x^{\alpha}}\right)}\left(a,b\right),\quad x\geq 0,\ \alpha,\gamma,a,b>0.$$

The values of the model parameters are reported in Table 7. The discrimination measures are provided in Table 8 and the other goodness of fit measures of the proposed and other competitive models are provided in Table 9. The estimated cdf and Kaplan-Meier survival plots are sketched in Figure 13, which shows that proposed distribution fit the estimated cdf and Kaplan-Meier survival plots very closely. The PP-Plot and Box plot are sketched in Figure 14. From box plot of the second data set is also clear that the data set has Heavier tail.

Table 7: Estimated values with standard error (in parenthesis) of the proposed and other competitive models for data 2.

$Dist.$	α	γ	σ	θ	a	b
C-W	0.814	0.003	6.024			
	(0.0927)	(0.0012)	(3.5364)			
Ex-APTW	0.510	0.172		5.425		
	(0.5094)	(0.6258)		(7.0766)		
Ku-W	0.620	0.501			0.702	0.118
	(0.3093)	(1.0970)			(3.2715)	(2.0964)
BW	0.478	0.502			2.797	0.344
	(0.2696)	(0.5522)			(3.1595)	(0.6646)

Table 8: Discrimination measures of the proposed and other competitive models for data 2.

$Dist.$	AIC	BIC	CAIC	HQIC
C-W	333.857	337.959	334.817	335.142
Ex-APTW	335.071	339.172	336.065	336.979
Ku-W	337.750	343.220	337.254	337.300
BW	335.457	340.926	337.124	337.170

Table 9: Goodness of fit measures of the proposed and other competitive models for data 2.

$Dist.$	CM	AD	KS	p-value
C-W	0.071	0.385	0.131	0.694
Ex-APTW	0.093	0.491	0.142	0.598
Ku-W	0.091	0.546	0.146	0.488
BW	NaN	NaN	0.144	0.603

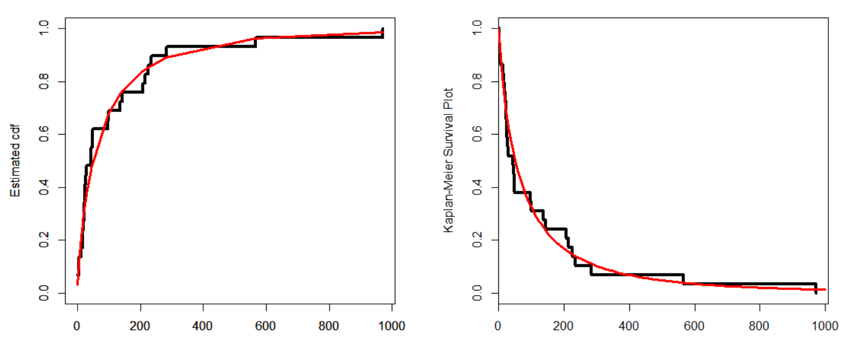

Figure 13: Estimated cdf and Kaplan-Meier survival plots of the C-W distribution for the second data set.

7.3. A real life application from insurance sciences

Here, we analyze another heavy tailed data set taken from insurance sciences available at: httpsinstruction.bus.wisc.edujfreesjfreesbooksRegression. To illustrate the potentiality of the proposed method, the C-W model and other competitive distributions including Lomax, Burr and exponentiated Weibull (EW) distributions are applied to this data set and it is observed that the proposed model provide best fit to the heavy tailed insurance data. The distribution functions of the competitive distributions are

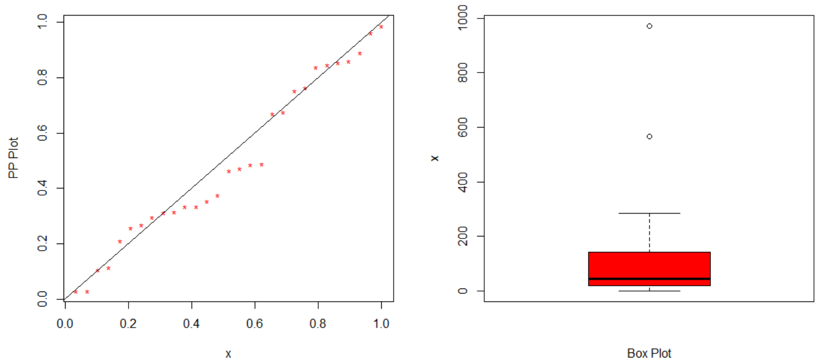

Figure 14: PP plot of the C-W distribution and box plot for the second data set.

1. Lomax distribution

$$G\left(x;\alpha,\gamma\right)=1-\left(1+\lambda x\right)^{-\alpha}, \qquad x\geq 0,\ \alpha,\gamma>0.$$

2. Burr distribution

$$G\left(x;\alpha,\gamma\right)=1-\left(1+x^{\gamma}\right)^{-\alpha}, \qquad x\geq 0,\ \alpha,\gamma>0.$$

3. EW distribution

$$G\left(x,a,\alpha,\gamma\right)=\left(1-e^{-\gamma x^{\alpha}}\right)^{a}, \qquad x\geq 0,\ \alpha,\gamma,a>0.$$

For the third data set, parameters values are reported in Table 10. The discrimination measures are provided in Table 11 and the analytical measures are presented in Table 12. The estimated cdf and Kaplan-Meier survival plots are sketched in Figure 15. The PP-Plot and Box plot are sketched in Figure 16. From Figures 15 and 16, it is clear that the data set has a heavier tail and proposed model fit the estimated cdf and Kaplan-Meier survival plots very well.

Table 10: Estimated values with standard error (in parenthesis) of the proposed and other competitive models for data 3.

$Dist.$	α	γ	σ	a
C-W	0.745 (0.0440)	2.861 (1.3051)	0.515 (0.1814)	
Lomax	1.569 (0.2076)	0.365 (0.0747)		
Burr	2.885 (0.1298)	0.813 (0.0266)		
EW	1.944 (0.4596)	2.564 (0.2619)		0.487 (0.0633)

Table 11: Discrimination measures of the proposed and other competitive models for data 3.

$Dist.$	AIC	BIC	CAIC	HQIC
C-W	169.236	182.032	169.282	174.246
Lomax	194.249	202.779	194.271	197.589
Burr	171.169	184.699	172.191	177.191
EW	170.908	183.709	171.339	176.303

Table 12: Goodness of fit measures of the proposed and other competitive models for data 3.

$Dist.$	CM	AD	KS	p-value
C-W	0.155	0.983	0.041	0.327
Lomax	0.464	2.623	0.078	0.002
Burr	0.250	1.426	0.045	0.229
EW	0.168	0.997	0.049	0.150

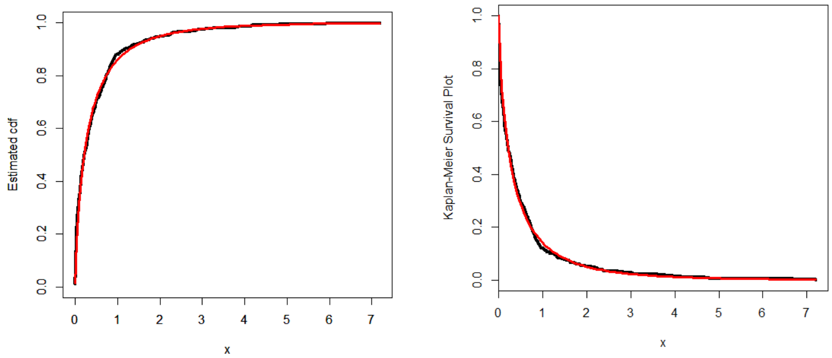

Figure 15: Estimated cdf and Kaplan-Meier survival plots of the C-W distribution for the third data set.

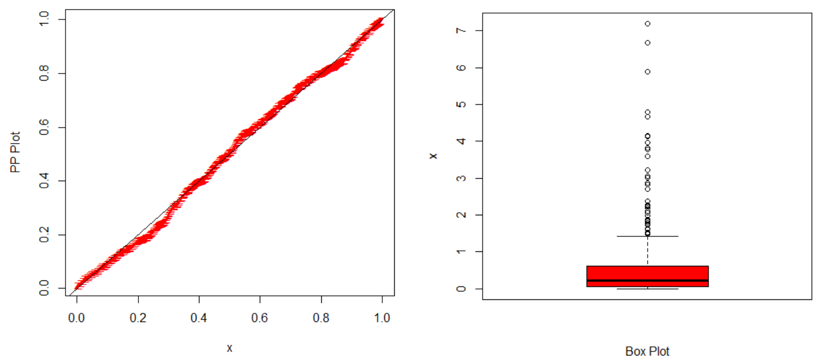

Figure 16: PP plot of the C-W distribution and box plot for the third data set.

8. Concluding Remarks

The importance of the extended distributions first realized in financial sciences and later in other applied fields such as engineering and medical sciences. To cater data in those fields, a number of methods of proposing new distributions have been introduced. In this regard, we further carried this branch of distribution theory and provide a new platform to the new researchers by developing nineteen new families of distributions. To prove the potential of the newly proposed methods, we consider a sub-case called the cosine-Weibull distribution. We applied the proposed distribution in three different disciplines and compared its goodness of fit withe other well-known distributions. The proposed distribution outclass the competitive models in all three disciplines. We hope that the new development will attract wider applications in the filed and would be quite helpful for the new comers in distribution theory.

Data Availability

This work is mainly a methodological development and has been applied on secondary data related to the insurance science data, but if required, data will be provided.

References

1. Ahmad, Z., Hamedani, G. G., & Butt, N. S. (2019). Recent Developments in Distribution Theory: A Brief Survey and Some New Generalized Classes of distributions. Pakistan Journal of Statistics and Operation Research, 15(1), 87-110.

2. Ahmad, Z., Ilyas, M. & Hamedani, G.G. (2019). The Extended Alpha Power Transformed Family of Distributions: Properties and Applications. Journal of Data Science (Accepted).

3. Akaike, H. (1974). A new look at the statistical model identification. In Selected Papers of Hirotugu Akaike (pp. 215-222). Springer, New York, NY.

4. Algamal, Z. Y. (2008). Exponentiated exponential distribution as a failure time distribution. IRAQI

Journal of Statistical science, 14, 63-75.

5. Almalki, S. J., & Yuan, J. (2013). A new modified Weibull distribution. Reliability Engineering and System Safety, 111, 164-170.

6. Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. Metron, 71(1), 63-79.

7. Azzalini, A., Del Cappello, T., & Kotz, S. (2002). Log-skew-normal and log-skew-t distributions as models for family income data. Journal of Income Distribution, 11(3-4), 2-2.

8. Bakar, S. A., Hamzah, N. A., Maghsoudi, M., & Nadarajah, S. (2015). Modeling loss data using composite models. Insurance: Mathematics and Economics, 61, 146-154.

9. Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, 52(3), 345-370.

10. Chesneau, C., & Jamal, F. (2019). THE SINE KUMARASWAMY-G FAMILY OF DISTRIBUTIONS.

11. Cooray, K., & Ananda, M. M. (2005). Modeling actuarial data with a composite lognormal-Pareto model. Scandinavian Actuarial Journal, 2005(5), 321-334.

12. Cordeiro, G. M. & Castro, M. de (2010). A new family of generalized distributions. Journal of statistical Computation and Simulation. DOI: 10.1080/0094965YY

13. Demicheli, R., Bonadonna, G., Hrushesky, W. J., Retsky, M. W., & Valagussa, P. (2004). Menopausal status dependence of the timing of breast cancer recurrence after surgical removal of the primary tumour. Breast Cancer Research, 6(6), R689.

14. Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models?. Insurance: Mathematics and Economics, 51(2), 239-248.

15. Glänzel, W. (1987). A characterization theorem based on truncated moments and its application to some distribution families. In Mathematical statistics and probability theory (pp. 75-84). Springer, Dordrecht.

16. Glanzel, W. (1990). Some consequences of a characterization theorem based on truncated moments. Statistics, 21(4), 613-618.

17. Gleaton, J. U., & Lynch, J. D. (2006). Properties of generalized log-logistic families of lifetime distributions. Journal of Probability and Statistical Science, 4(1), 51-64.

18. Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B (Methodological), 41(2), 190-195.

19. Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. Journal of Hydrology, 46(1-2), 79-88.

20. Lai, C. D., & Xie, M. (2006). Stochastic ageing and dependence for reliability. Springer Science and Business Media.

21. Lee, E. T., & Wang, J. W. (2003). Statistical Methods for Survival Data Analysis (3rd ed.). New York: Wiley.

22. Mahdavi, A., & Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. Communications in StatisticsTheory and Methods, 46(13), 6543-6557.

23. Mahmood, Z., & Chesneau, C. (2019). A New Sine-G Family of Distributions: Properties and Applications.

24. Marshall, A. W., & Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika, 84(3), 641-652.

25. Mudholkar, G. S., & Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE transactions on reliability, 42(2), 299-302.

26. Nadarajah, S., & Bakar, S. A. (2014). New composite models for the Danish fire insurance data. Scandinavian Actuarial Journal, 2014(2), 180-187.

27. Noack, A. (1950). A class of random variables with discrete distributions. The Annals of Mathematical Statistics 21(1):127–132.

28. Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. ASTIN Bulletin: The Journal of the IAA, 27(1), 139-151.

29. Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 6(2), 461-464. 30. Scollnik, D. P. (2007). On composite lognormal-Pareto models. Scandinavian Actuarial Journal, 2007(1), 20-33.

31. Shaw, W. T., & Buckley, I. R. (2009). The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. arXiv preprint arXiv:0901.0434.

32. Tahir, M. H.,& Cordeiro, G. M. (2016). Compounding of distributions: a survey and new generalized classes. Journal of Statistical Distributions and Applications, 3(1), 13.

33. Tahir, M. H., & Nadarajah, S. (2015). Parameter induction in continuous univariate distributions: Well-established G families. Anais da Academia Brasileira de Ciências, 87(2), 539-568.

34. Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. Insurance: Mathematics and economics, 38(2), 413-426.

35. Zajicek, G. (2011). A new kind of medical cancer epidemiology. Available at: www.what-is-cancer.com/papers/newmedicine/epidemiologyFrame.htm.

36. Zichuan, M., Hussain, S., Iftikhar, A., Ilyas, M., Ahmad, Z., Khan, D. M., & Manzoor, S. (2020). A New Extended-Family of Distributions: Properties and Applications. Computational and Mathematical Methods in Medicine, 2020.