6-29-2017

The Kumaraswamy Skew-t Distribution and Its Related Properties

Wei Ning
Bowling Green State University, wning@bgsu.edu

Khamis K. Said

Doaa Basalamah

Arjun K. Gupta
Bowling Green State University, gupta@bgsu.edu

The Kumaraswamy Skew-t Distribution and Its Related Properties

Khamis K. Said[1], Doaa Basalamah[2,3], Wei Ning[3*], Arjun Gupta[3]

[1] School of Mathematics and Statistics

Beijing Institute of Technology, Beijing, China

[2]Mathematical Science Department

Umm Al Qura University, Mecca 24231, Saudi Arabia

[3]Department of Mathematics and Statistics

Bowling Green State University, Bowling Green, OH 43403, USA

Abstract: Skew normal distribution has been introduced by Azzalini (1985) as an alternative to the normal distribution to accommodate asymmetry. Since then extensive studies have been done on applying Azzalini's skewness mechanism to other well-known distributions, such as skew-t distribution which is more flexible and can better accommodate long tailed data than the skew normal one. Cordeiro and de Castro (2011) proposed a new class of distribution called the Kumaraswamy generalized distribution $(Kw - F)$ which is capable of fitting skewed data that cannot be fitted well by existing distributions. Since then, the $Kw - F$ distribution has been widely studied and various versions of generalization of this distribution family have been introduced. In this paper we introduce a new generalization of the skew-t distribution based on the Kumaraswamy generalized distribution. The new class of distribution which we call the Kumaraswamy skew-t $(KwST)$ has the ability of fitting skewed, long and heavy tailed data and is more flexible than the skew-t distribution as it contains the skew-t distribution as a special case. Related properties of this distribution family such as mathematical properties, moments, and order

*Corresponding author. Email: wning@bgsu.edu

statistics are discussed. The proposed distribution is applied to a real data set to illustrate the estimation procedure.

Keywords: Skew-t distribution; Kumaraswamy distribution; Kumaraswamy skew-t distribution; Maximum likelihood estimation.

1 Introduction

Kumaraswamy (1980) introduced a distribution on $(0, 1)$ called a distribution of double bounded random process (DB) which has been used widely in hydrological applications. The DB distribution shares many similarities with the beta distribution, while the DB distribution has the advantage of the tractability of its distribution function and its density function does not depend on some special function which makes the computation of the MLE's easier. Jones (2009) provided detailed survey of the similarities and differences between the beta distribution and the distribution of double bounded random process (DB). Based on the double bounded random process (DB) distribution, Cordeiro and de Castro (2011) proposed a new class of distribution which is called the Kumaraswamy generalized distribution denoted as $(Kw-F)$. They extended this class of distributions to the normal, Weibull, gamma, Gumbel, and inverse Gaussian distributions by choosing F to be the corresponding distribution functions of these distributions. A major benefit of the Kumaraswamy generalized distribution is its ability of fitting skewed data that can not be fitted well by existing distributions. Since then, the Kumaraswamy generalized distribution has been widely studied and many authors have developed various generalized versions based on this distribution. Nadarajah et al. (2011) studied some new properties of the Kumaraswamy generalized distribution including asymptotes, shapes, moments, moment generating function, and mean deviations. Cordeiro et al. (2015) studied moments for the various classes of Kumaraswamy generalized distributions such as Kumaraswamy normal, Kumaraswamy Student-t, Kumaraswamy beta, and Kumaraswamy snedecor F distribution. Mameli (2015) introduced the Kumaraswamy skew normal distribution, $KwSN$, and derived the moments, moments generating function, and the maximum likelihood estimators for special values of the parameters, to name a few.

Among all the skew distributions proposed in literature, the skew-t distribution receives special attention after the introduction of the skew multivariate normal distribution by Azzalini and Dalla Valle (1996). Gupta (2000) defined a skew multivariate t distribution using a pair of independent standard skew normal and chi-squared random variables. Azzalini and Capitanio (2002) defined a skew-t variate as a scale mixture of skew normal and chi-squared variables. Several authors studied possible extensions and generalizations of the skew-t distribution. Arellano-Valle et al. (2005) discussed generalized skew distributions in the multivariate setting including the skew-t one. Huang et al. (2007) studied generalized skew-t distributions and used it in data analysis. Hasan (2013) proposed a new approach to define the non-central skew-t distribution. Shafiei and Doostparast (2014) introduced the Balakrishnan skew-t distribution and its associated statistical characteristics, to name a few.

In this paper we introduce a new generalization of the skew-t distribution based on the Kumaraswamy generalized distribution. The new class of distribution which we call the Kumaraswamy skew-t ($KwST$) has the ability of fitting skewed and heavy tailed data and is more flexible than the skew-t distribution as it contains the skew-t distribution and other important distributions as special cases. Related properties of the $KwST$ such as mathematical properties, moments, and order statistics are discussed. Furthermore, the proposed distribution is applied to a real data to illustrate the fitting procedure.

The rest of the paper is organized as follows. The distribution function, the probability density function and some expansions of probability density function of the Kumaraswamy skew-t ($KwST$) are given in section 2. We establish some mathematical and distributional properties and provide several methods to simulate a sample of $KwST$ in section 3. We derive explicit expressions for the moments in section 4. A new representation for the density of order statistics of $KwST$ and Kumaraswamy generalized distributions are given in section 5. Maximum likelihood estimators (MLEs) of the $KwST$ parameters are given in section 6. In section 7, the proposed distribution is applied to a well-known data set to illustrate the fitting superiority as well as the comparison to other existing distributions to indicate its advantage. Some discussion is provided in section 8.

2 Distribution and Density Functions

Let $F(x)$, $f(x)$ be the *cdf* and *pdf* of a continuous random variable X. Cordeiro and de Castro (2011) proposed the Kumaraswamy generalized distribution denoted by $Kw - F(a,b)$ with *pdf* $g(x;a,b)$ and *cdf* $G(x;a,b)$ given by:

$$g(x;a,b) = abf(x)F(x)^{a-1}(1 - F(x)^{a-1})^{b-1}, \tag{1}$$

$$G(x;a,b) = 1 - \{1 - F(x)^a\}^b, \tag{2}$$

where $a > 0$ and $b > 0$ are parameters to control the skewness and tail weights. By taking $F(x)$ to be the cdf of the normal, Weibull, gamma, Gumbel, and inverse Gaussian distributions, Cordeiro and de Castro (2011) defined the Kw-normal, Kw-Weibull, Kw-gamma, Kw-Gumbel, and Kw-inverse Gaussian distributions, respectively.

We take $F(x)$ in (2) to be the distribution function of the skew-t distribution and we introduce a new distribution called the Kumaraswamy skew-t distribution denoted by $KwST(a,b,\lambda,r)$ with *pdf* $g(x;a,b,\lambda,r)$ and *cdf* $G(x;a,b,\lambda,r)$ given by:

$$g(x;a,b,\lambda,r) = abf(x;\lambda,r)F(x;\lambda,r)^{a-1}(1 - F(x;\lambda,r)^a)^{b-1}, \tag{3}$$

$$G(x;a,b,\lambda,r) = 1 - \{1 - F(x;\lambda,r)^a\}^b, \tag{4}$$

where $a > 0$ and $b > 0$, $x \in \Re$, $f(x;\lambda,r)$ and $F(x;\lambda,r)$ are the *pdf* and *cdf* of the skew-t distribution given by Azzalini and Capitanio (2014) as follows:

$$f(x;\lambda,r) = 2t(x;r)T(\lambda x\sqrt{\frac{r+1}{x^2+r}};r+1),$$

where $T(x;r)$ and $t(x;r)$ denote the *cdf* and *pdf* of the Student-t distribution with degrees of freedom $r > 0$ and the shape parameter $\lambda \in \Re$.

The $KwST$ distribution can be extended to include location and scale parameters $\mu \in \Re$ and $\sigma > 0$. If $X \sim KwST(a,b,\lambda,r)$, then $Y = \mu + \sigma X$ leads to a six parameters $KwST$ distribution with the parameter vector $\xi = (a,b,\mu,\sigma,\lambda,r)$. We denote it by $Y \sim$

$KwST(a, b, \mu, \sigma, \lambda, r)$.

Throughout this paper we will denote t_r the Student-t distribution with *pdf* $t(x; r)$ and *cdf* $T(x; r)$, $st_r(\lambda)$ the skew-t with *pdf* $f(x; \lambda, r)$ and *cdf* $F(x; \lambda, r)$, and $KwST(a, b, \lambda, r)$ the Kumaraswamy skew-t distribution with *pdf* $g(x; a, b, \lambda, r)$ and *cdf* $G(x; a, b, \lambda, r)$.

2.1 Expansion of the Density Function

According to Cordeiro and Castro (2011), using the binomial expansion for $b \in \Re^+$, the *pdf* of $KwST$ (3) can be rewritten as

$$g(x; a, b, \lambda, r) = f(x; \lambda, r) \sum_{i=0}^{\infty} w_i F(x; \lambda, r)^{a(i+1)-1}, \qquad (5)$$

where the binomial coefficient w_i is defined for all real numbers with, $w_i = (-1)^i ab\binom{b-1}{i}$. If b is an integer, the index i in the sum of (5) stops at $b - 1$. If a is an integer, then (5) is the density of st_r multiplied by infinite weighted power series of the cdf of st_r. On the other hand, if a is not an integer, we can expand the term $F(x; \lambda, r)^{a(i+1)-1}$ as follows

$$\begin{aligned} F(x; \lambda, r)^{a(i+1)-1} &= [1 - (1 - F(x; \lambda, r))]^{a(i+1)-1} \\ &= \sum_{j=0}^{\infty} (-1)^j \binom{a(i+1)-1}{j} (1 - F(x; \lambda, r))^j \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{j} (-1)^{j+k} \binom{a(i+1)-1}{j} \binom{j}{k} F(x; \lambda, r)^k. \end{aligned}$$

Further, the density $g(x; \lambda, r, a, b)$ in (3) can be rewritten as

$$g(x; \lambda, r, a, b) = f(x; \lambda, r) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{j} w_{i,j,k} F(x; \lambda, r)^k, \qquad (6)$$

where the coefficient $w_{i,j,k}$ is defined for as,

$$w_{i,j,k} = (-1)^{i+j+k} \binom{b-1}{i} \binom{a(i+1)-1}{j} \binom{j}{k} ab.$$

3 Properties and Simulation

In this section we study some theoretical properties of $KwST$ distribution. Then we provide graphical illustrations of these properties. Finally, we present different approaches to generate a sample from $KwST$ distribution.

3.1 Properties

Property 3.1.1 Let $X \sim KwST(a, b, \lambda, r)$,

(a) If $a = b = 1$, then $X \sim st_r(\lambda)$.

(b) If $\lambda = 0$ and $a = b = 1$, then $X \sim t_r$.

(c) If $\lambda = 0$ and $a = b = r = 1$, then $X \sim Cauchy(0, 1)$.

(d) If $\lambda = 0$, then $X \sim Kw - t_r(a, b)$.

(e) If $\lambda = 0$ and $r = 1$, then $X \sim Kw - Cauchy(a, b)$.

(f) If $Y = F(x; \lambda, r)$, then $X \sim Kw(a, b)$.

The proof of Property 3.1.1 follows from (3) and from elementary properties of the skew-t distribution. Note that in part (d) and (e), the distribution function of $Kw - t_r(a, b)$ and $Kw - Cauchy(a, b)$ are given by substituting the $F(x)$ in (2) by the the distribution function of the Student-t with degrees of freedom r and the Cauchy (0,1) respectively. The following two properties are based on properties 5 and 6 in Mameli (2015).
Property 3.1.2 Let $X \sim KwST(a, b, \lambda, r)$ and $Y \sim KwST(a, d, \lambda, r)$ be two independent random variables. Then, $(X|Y \geq X) \sim KwST(a, b + d, \lambda, r)$, where $a, b,$ and $d > 0$.
Property 3.1.3 Let $X \sim KwST(a, 1, \lambda, r)$ and $Y \sim KwST(c, 1, \lambda, r)$ be two independent random variables. Then, $(X|Y \leq X) \sim KwST(a + c, 1, \lambda, r)$, where a and $c > 0$.

The following property studies the limiting distribution of $KwST(a, b, \lambda, r)$ as one of the parameters approaches ∞ while the others remain fixed.
Property 3.1.4 Let $X \sim KwST(a, b, \lambda, r)$ be a random variable with *pdf* $g(x; a, b, \lambda, r)$ defined in (3). Then,

(a) As $a \to \infty$ or $b \to \infty$, the probability distribution function $g(x; a, b, \lambda, r)$ degenerates to zero.

(b) As $r \to \infty$, $X \sim KwSN(a, b, \lambda)$.

(c) As $\lambda \to \infty$, $X \sim Kw\text{-}|t_r|(a, b, r)$.

Part (a) in property 3.1.4 can be generalized to the class of the Kumaraswamy generalized family $Kw - F(a, b)$ with *pdf* $g(x; a, b)$ defined in (1) as follows.

Property 3.1.5 Let $X \sim Kw - F(a, b)$. As $a \to \infty$ or $b \to \infty$, then the probability distribution function $g(x; a, b)$ degenerates to zero.

3.2 Graphs

To understand the effect of each parameter in determining the overall shape of the $KwST$ density, we present some graphs with five fixed parameters and the sixth one varying. For simplicity, we fix the location parameter μ to be zero and the scale parameter σ to be one in all graphs. In Figure 1 we fixed the parameters $(b = 3, \lambda = 1, r = 3)$ and we graph the density of $KwST(a, 3, 1, 3)$ density for different values of a. Figure 1 shows that as a increases the left tail of the $KwST$ density gets lighter.

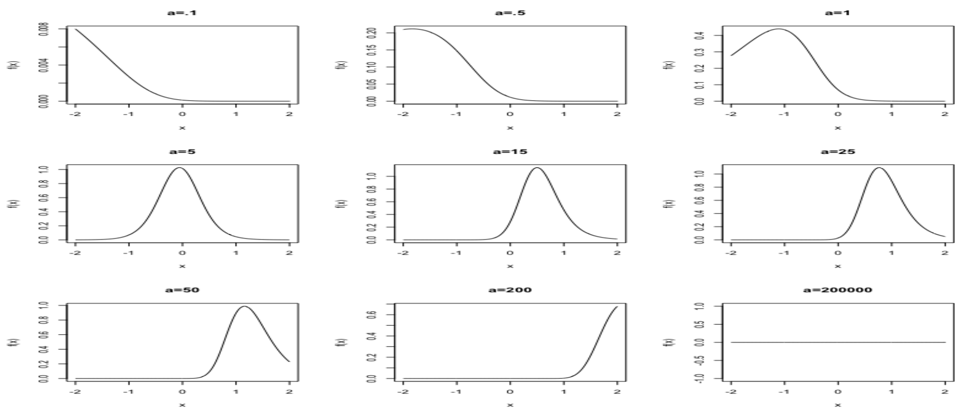


Figure 1: $KwST(a, 3, 1, 3)$ density as the parameter a varies.

On the other hand, we note that the parameter b controls the right tail weight of the $KwST$ density when b varies and all other parameters are fixed $(a = 5, \lambda = -1, r = 3)$ as shown in Figure 2. In addition, Figures 1 and 2 show that as a or b approach infinity the $KwST$ density degenerate to zero.

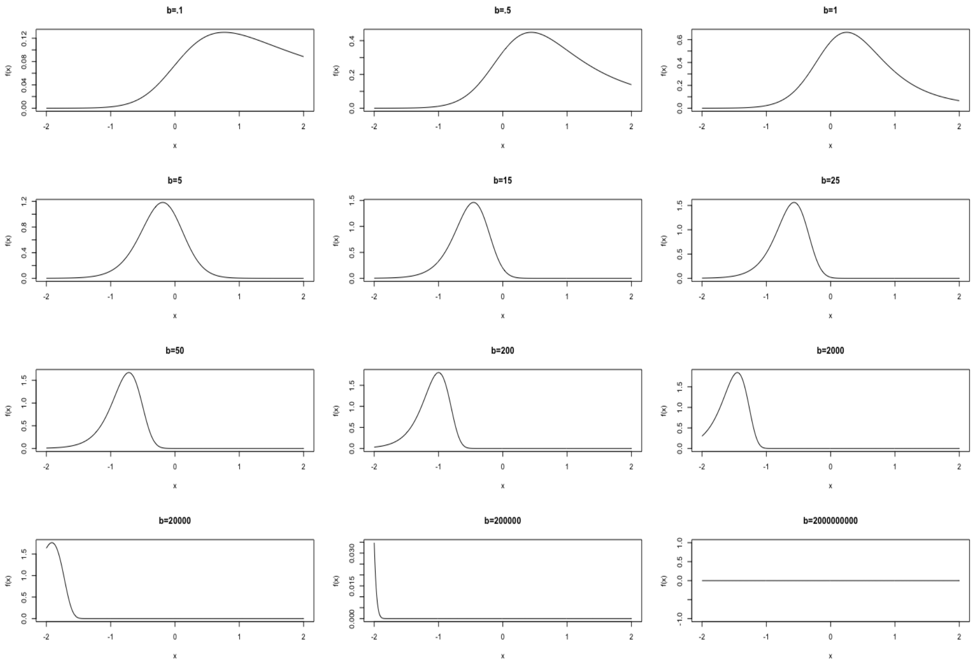


Figure 2: $KwST(5, b, -1, 3)$ density as the parameter b varies.

Figure 3 studies the effect of the parameter λ on the shape of the $KwST$ density by fixing the parameters ($a = 5, b = 3, r = 3$) and taking the parameter λ ranging from -5 to 100. Then, we compare the density curves of $KwST(5, 3, \lambda, 3)$ with the curve of $Kw - |t_r|$ ($a = 5, b = 3, r = 3$). As expected, the graph is skewed to the right for positive values of λ and skewed to the left for negative values of λ. Moreover, we observe that as λ increases the $KwST$ density curve overlaps the $Kw - |t_r|$ density curve which graphically proves part (c) of property 3.1.4.

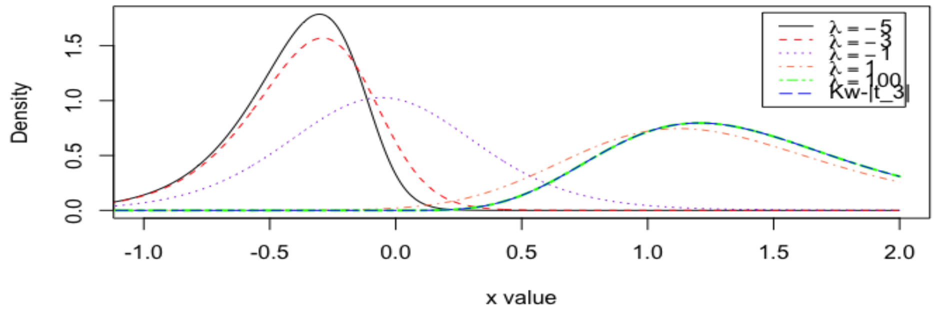


Figure 3: $KwST(5, 3, \lambda, 3)$ density as the parameter λ varies.

In Figure 4, we study the effect of the degrees of freedom r on the shape of the $KwST$ density by fixing the parameters $(a = 5, b = 3, \lambda = -1)$ and taking the degrees of freedom $r = 1, 5, 15$ and 50. We observe that the shape of the $KwST(5, 3, -1, r)$ density gets closer to the one of the $KwSN(5, 3, -1)$ as the degrees of freedom r increases, which agrees with the part (b) of property 3.1.4. The tail gets thicker as the decreases of the degrees of freedom. Properties in the last two graphs are inherited from the baseline skew-t distribution. Furthermore, Figures 1-4 show that the $KwST$ inherited the unimodality from its baseline distribution.

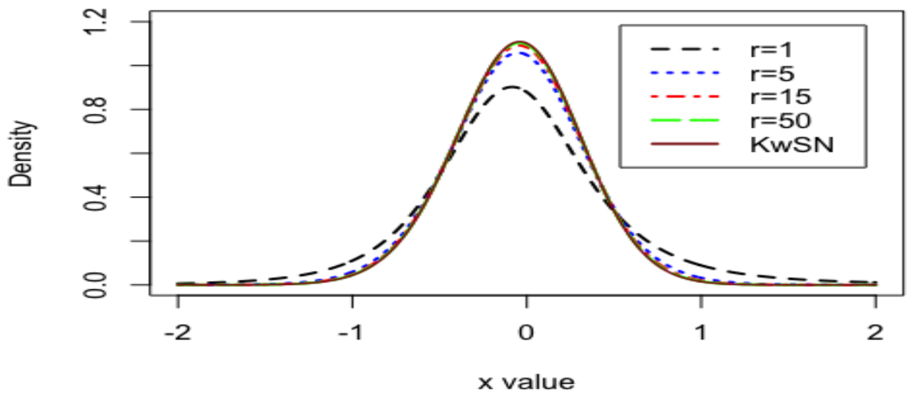


Figure 4: $KwST(5, 3, -1, r)$ density as the degrees of freedom r varies.

3.3 Simulation

In this section, we provide several methods to generate samples from $KwST(a, b, \lambda, r)$ distribution. The $KwST$ quantile function is obtained by inverting (4)

$$x = Q(u) = G^{-1}(u) = F^{-1}[1 - (1 - u)^{1/b}]^{1/a}, \tag{7}$$

where U is a uniform random variable on $(0, 1)$ and F^{-1} is the inverse function of the distribution function of $st_r(\lambda)$. Then, applying the inverse transformation technique we generate $KwST$ random sample using (7).

An alternative method to generate a $KwST(a, b, \lambda, r)$ random sample is to use the algorithm of the acceptance rejection method proposed by Nadarajah et al. (2012) as follows. Define a constant M by $M = \frac{a^b b(a-1)^{1-1/a}(b-1)^{b-1}}{(ab-1)^{b-1/a}}$ for given $a \geq 1$ and $b \geq 1$.

Then the following scheme holds for generating $KwST(a, b, \lambda, r)$ variable:

(a) Generate $X = x$ from the pdf of skew-t.

(b) Generate $Y = UMx$, where U is a uniform variate on (0,1).

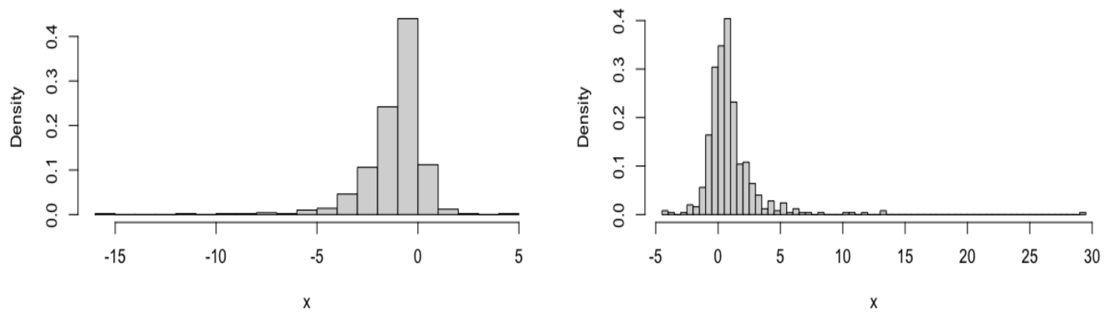
(c) Accept $X = x$ as $KwST$ variable if $Y < f(x, \lambda, r)$. Otherwise, return to step (b).

Additional method to generate $KwST(a, b, \lambda, r)$ random sample is to directly apply part (f) in property 3.1.1.

Figure 5 shows the histograms of two random samples with size 500 simulated from $KwST(\xi)$ distribution using the acceptance rejection method, with the parameter vectors $\xi_1 = (a = 5, b = 2, \mu = 0, \sigma = 1, \lambda = -2, r = 2)$ and $\xi_2 = (a = 2, b = 4, \mu = 0, \sigma = 1, \lambda = 1, r = 2)$ respectively.

4 Moments

In this section we derive explicit expressions for the moments of $KwST$ random variable using different techniques. Further, we present some numerical moments estimation of



(a) $\xi_1 = (a = 5, b = 2, \lambda = -2, r = 2)$ (b) $\xi_2 = (a = 2, b = 4, \lambda = 1, r = 2)$

Figure 5: Histogram for $KwST$ random samples of size 500.

the mean (μ_{KwST}), variance(σ^2_{KwST}), skewness(γ_1), and kurtosis(γ_2) of $KwST(a, b, \lambda, r)$ random variable for selected values of the parameters a, b, λ, and r.

Theorem 4.1.1 Let $X \sim KwST(a, b, \mu, \sigma, \lambda, r)$ where a, b and n are positive integers and $r \geq n$, then

$$E_X(X^n) = ab\mu^n \sum_{i=0}^{n} \binom{n}{i} (\frac{\sigma}{\mu})^i \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} E_Y(Y^i F(Y; \lambda, r)^{a(j+1)-1}), \tag{8}$$

where $Y \sim st_r(\lambda)$.

Further, the n^{th} moment of X with the *pdf* $g(x; a, b, \lambda, r)$ in (6) can be expressed in terms of the probability weighted moments of the baseline distribution $st_r(\lambda)$ as follows

Proposition 4.1.1 Let $X \sim KwST(a, b, \lambda, r)$ be a random variable where a and b are positive real numbers, n is a positive integer, $n \geq 1$ and $r \geq n$, then

$$E_X(X^n) = \sum_{i,j=0}^{\infty} \sum_{k=0}^{j} w_{i,j,k} E_Y(Y^n F(Y; \lambda, r)^k), \tag{9}$$

where $Y \sim st_r(\lambda)$ and $w_{i,j,k} = (-1)^{i+j+k} \binom{b-1}{i} \binom{a(i+1)-1}{j} \binom{j}{k} ab$.

If a is an integer, then we use the pdf (5) to derive the the n^{th} moments of $X \sim$

$KwST(\lambda, r, a, b)$ as follows.

$$E_X(X^n) = \sum_{i=0}^{\infty} w_i E_Y(Y^n F(Y;\lambda,r)^{a(i+1)-1}), \qquad (10)$$

where $Y \sim st_r(\lambda)$ and $w_i = (-1)^i \binom{b-1}{i} ab$.

If b is an integer, the index i in the first sum in (9) and the sum in (10) stop at $b-1$.

Alternatively, the n^{th} moment of $X \sim KwST(a, b, \lambda, r)$ random variable with integers $a \geq 2$ and $b \geq 2$ can be expressed in terms of the n^{th} moment of the $st_r(\lambda)$ multiplied by a constant as presented in the following proposition.

Proposition 4.1.2 Let $X \sim KwST(a, b, \lambda, r)$ with integers n, a and b where a and $b \geq 2$ $n \geq 1$ and $r \geq n$.

$$E_X(X^n) = E_Y(Y^n)c(a, b), \qquad (11)$$

where $c(a,b) = ab\left[\sum_{i=0}^{b-2} \frac{(-1)^i}{B(i+1,b-i-1)}\left[\frac{a}{a(2+i)-1} - \frac{(a-1)}{(b-i-1)(a(1+i)-1)}\right] - (-1)^{b-1}\frac{(a-1)}{ab-1}\right]$,
$B(a, b)$ is the complete beta function and $Y \sim st_r(\lambda)$.

According to Azzalini and Capitanio (2014), the n^{th} moment of $Y \sim st_r(\lambda)$ is given by

$$\begin{aligned} E_Y(Y^n) &= E_V(V^{n/2})E_Z(Z^n) \\ &= \frac{(r/2)^{n/2}\Gamma(\frac{r-n}{2})}{\Gamma(\frac{r}{2})}E(Z^n), \end{aligned}$$

where $Z \sim SN(0, 1, \lambda)$.

The mean, variance, skewness and kurtosis measures can be computed numerically using existing softwares. Table 1 shows numerical estimations of these measures by computing the first four moments for various values of the parameters a, b, λ, and r with fixed $\mu = 0$ and $\sigma = 1$, where Table 1(a) presents the numerical estimations of $KwST(a, b, \lambda, r)$ random variable for different values of a, b, and λ and fixed degrees of freedom $r = 5$, while in Table 1(b) the parameter $\lambda = 2$ is fixed and the parameters a, b, and the degrees of freedom r vary. Skewness and kurtosis are calculated using the well-known relations

$$\gamma_1(X) = E_X\left[\left(\frac{X - E(X)}{Var(X)^{1/2}}\right)^3\right],$$

and

$$\gamma_2(X) = E_X\Big[\Big(\frac{X - E(X)}{Var(X)^{1/2}}\Big)^4\Big].$$

Table 1(a): moments estimation of the mean (μ_{KwST}), variance(σ^2_{KwST}), skewness(γ_1), and kurtosis(γ_2) of $KwST(a, b, \lambda, r)$ random variable for different values of a, b, and λ

	a	b	λ	r	μ_{KwST}	σ^2_{KwST}	γ_1	γ_2
$KwST$	1	.50	-10	5	-0.563	0.616	-2.419	28.051
			-1		0.119	2.842	4.919	119.904
			0		1.107	6.303	5.757	97.085
			1		1.753	6.804	6.674	108.903
			10		1.927	6.321	7.448	125.575
			50		1.929	6.312	7.463	125.904
$Kw - \lvert t_r\rvert$	1	.50	-	5	1.929	6.311	7.465	125.943
$KwST$	5	3	-1	5	-0.055	0.157	-0.005	3.465
			0		0.676	0.281	0.298	3.623
			1		1.197	0.251	0.542	3.934
			10		1.288	0.213	0.761	4.248
			50		1.288	0.213	0.761	4.247
$Kw - \lvert t_r\rvert$	5	3	-	5	1.288	0.213	0.761	4.247
$KwST$	10	1	-1	5	0.850	0.463	2.161	19.937
			0		2.003	1.270	2.462	25.524
			1		2.545	1.477	2.737	30.106
			10		2.591	1.457	2.812	31.254
			50		2.591	1.457	2.812	31.655
$Kw - \lvert t_r\rvert$	10	1	-	5	2.591	1.457	2.812	31.655
$KwST$	10	2000	-10	5	0	0	NA	NA
			0		0	0	NA	NA
			10		0	0	NA	NA

From the numerical results in Table 1, we observe that the $KwST(a, b, \lambda, r)$ distribution degenerates to zero as the increase of a or b. Thus, the skewness and kurtosis do not exist, and their values are replaced by NA. Further, the $KwST(a, b, \lambda, r)$ moments estimates get closer to the $KwSN(a, b, \lambda)$ ones as the degrees of freedom r increases and to the $Kw - |t_r|(a, b, r)$ as the parameter λ increases, where the numerical estimations of the $KwSN$ and $Kw - |t_r|$ are presented on the last line of each block. Numerical results in Table 1 agree with the property 3.1.4.

Table 1(b): moments estimation of the mean (μ_{KwST}), variance(σ^2_{KwST}), skewness(γ_1), and kurtosis(γ_2) of $KwST(a, b, \lambda, r)$ random variable for different values of a, b, and r

	a	b	λ	r	μ_{KwST}	σ^2_{KwST}	γ_1	γ_2
$KwST$	1	1	2	5	0.849	0.946	1.791	16.428
				10	0.773	0.652	0.865	5.042
				50	0.767	0.453	0.661	4.053
				300	0.758	0.412	0.873	3.547
$KwSN$	1	1	2	-	0.714	0.491	0.453	3.301
$KwST$	10	1	2	5	2.587	1.461	2.800	31.086
				10	2.176	0.574	1.283	6.703
				50	1.932	0.304	0.683	3.856
				300	1.889	0.269	0.588	3.587
$KwSN$	10	1	2	-	0.714	0.491	0.453	3.301
$KwST$	2	5	2	5	0.447	0.137	0.158	3.495
				10	0.432	0.124	0.090	3.257
				50	0.422	0.115	0.040	3.109
				200	0.420	0.113	0.031	3.109
$KwSN$	2	5	2	-	0.419	0.113	0.028	3.073
$KwST$	2000	15	2	5	0	0	NA	NA
				10	0	0	NA	NA
				20	0	0	NA	NA

5 Order Statistics

Order statistics make their appearance in many areas of statistical theory and practice. Cordeiro and de Castro (2011) derived the density of order statistics of the $Kw - F$ distribution as a function of the baseline density multiplied by infinite weighted sums of powers of the distribution function $F(x)$ as given by

$$g_{i:n}(x) = \frac{f(x)}{B(i, n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \sum_{r,u,v=0}^{\infty} \sum_{t=0}^{v} w_{u,v,t} p_{r,i+j-1}(a, b) F(x)^{r+t},$$

where a is a positive real number,

$$w_{u,v,t} = w_{u,v,t}(a, b) = (-1)^{u+v+t} ab \binom{a(u+1)-1}{v} \binom{b-1}{u} \binom{v}{t},$$

and

$$p_{r,i+j-1}(a,b) = \sum_{k=0}^{i+j-1} \binom{i+j-1}{k} (-1)^k \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} (-1)^{mr+l} \binom{kb}{m} \binom{ma}{l} \binom{l}{r}.$$

If a is a positive integer, then density of order statistics of the $Kw - F$ distribution is given by

$$g_{i:n}(x) = \frac{f(x)}{B(i, n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \sum_{r,u=0}^{\infty} w_u p_{r,i+j-1}(a,b) F(x)^{a(u+1)+r-1},$$

where $w_u = w_u(a,b) = (-1)^u ab\binom{b-1}{u}$.

We derive a new and simpler representation for the density of the order statistics of $KwST$ random sample and we generalize the result to the order statistics of the Kumaraswamy generalized family $Kw - F$.

Theorem 5.1 Let $X_1, ..., X_n$ be a random sample from a $KwST$ distribution with the defined probability density function $g(x; a, b, \lambda, r)$ in (3) and distribution function $G(x; a, b, \lambda, r)$ in (4). Let $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ be the order statistics of the random sample. The density function $g_{i:n}(x; a, b, \lambda, r)$ of the i^{th} order statistic $X_{i:n}$, for $i = 1, ..., n$, is given by

$$g_{i:n}(x; a, b, \lambda, r) = \sum_{k=0}^{i-1} s_{i,n,k}\, g(x; a, b(n-i+k+1), \lambda, r), \qquad (13)$$

where $s_{i,n,k} = (-1)^k \binom{n-i+k}{k} \binom{n}{i-k-1}$. Formula (13) immediately yields the density of order statistics of the $KwST$ distribution as a function of finite weighted sums of the density of the same class of $KwST$ distribution with a new parameter $b^* = b(n-i+k+1)$, which is written as a function of the sample size n, the order i and a constant k, where $0 \leq k \leq i-1$. Hence, the ordinary moments of order statistics of the $KwST$ distribution can be written as finite weighted sums of moments of the $KwST$ distribution with a new parameter b^*.

From Theorem 5.3, the density of the smallest and largest order statistic are as follow.

Proposition 5.1

(a) The density of the largest order statistic $X_{n:n}(x)$ is given by

$$g_{n:n}(x;a,b,\lambda,r)=\sum_{k=0}^{n-1}s_k g(y;a,b(k+1)\lambda,r), \tag{14}$$

where $s_k=\frac{(-1)^k}{B(k+1,n-k)}$.

(b) The density of the smallest order statistic $X_{1:n}(x)$ is simply $g(y;a,nb\lambda,r)$ which is the density of $Y\sim KwST(a,bn,\lambda,r)$.

The result in theorem 5.1 can be generalized to the class of the Kumaraswamy generalized family $Kw-F$, defined in (1).

Theorem 5.2 Let $X_1,...,X_n$ be a random sample from a Kumaraswamy generalized family $Kw-F$ distribution with the defined probability density function $g(x;a,b)$ in (1) and distribution function $G(x;a,b)$ in (2). Let $X_{1:n}\le X_{2:n}\le ...\le X_{n:n}$ be the order statistics of the random sample. The density function $g_{i:n}(x;a,b)$ of the i^{th} order statistic $X_{i:n}$, for $i=1,...,n$, is given by

$$g_{i:n}(x;a,b)=\sum_{k=0}^{i-1}s_{i,n,k}g(x;a,b(n-i+k+1)), \tag{15}$$

where $s_{i,n,k}=(-1)^k\binom{n-i+k}{k}\binom{n}{i-k-1}$.

6 Maximum Likelihood Estimation

The likelihood-based inference is a primary approach to statistical methodology. The maximum likelihood inference is a well-known concept with a quite standard notation. In this section, the maximum likelihood estimators (MLE's) of the $KwST$ parameters are given.

Consider a sample $x_1,x_2,.....,x_n$ from the $KwST(a,b,\mu,\sigma,\lambda,r)$ distribution. The log-likelihood function $l(\xi)$ for the parameter vector of $\xi = (a, b, \mu, \sigma, \lambda, r)$ is

$$l(\xi) = n\ log(a) + n\ log(b) - n\ log(\sigma) + \sum_{i=1}^{n} log f(z_i; \mu, \sigma, \lambda, r) + (a-1)\sum_{i=1}^{n} log(F(z_i; \mu, \sigma, \lambda, r))$$
$$+ (b-1)\sum_{i=1}^{n} log(1 - F(z_i; \mu, \sigma, \lambda, r)^a), \qquad (16)$$

where $z_i = \frac{x_i - \mu}{\sigma}$. The components of the score vector $U(\xi)$ are given by

$$U_a(\xi) = \frac{n}{a} + \sum_{i=0}^{n} log(F(z_i; \mu, \sigma, \lambda, r)) - (b-1)\sum_{i=1}^{n} \frac{F(z_i; \mu, \sigma, \lambda, r)^a log(F(z_i; \mu, \sigma, \lambda, r))}{1 - F(z_i; \mu, \sigma, \lambda, r)^a}.$$
$$U_b(\xi) = \frac{n}{b} + \sum_{i=0}^{n} log(1 - F(z_i; \mu, \sigma, \lambda, r)^a).$$
$$U_\mu(\xi) = \sum_{i=0}^{n} \frac{-1}{\sigma f(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)} \frac{df(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)}{d\mu}$$
$$- \frac{(a-1)}{\sigma}\sum_{i=0}^{n} \frac{1}{F(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)} \frac{dF(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)}{d\mu}$$
$$+ \frac{(b-1)}{\sigma}\sum_{i=0}^{n} \frac{1}{(1 - F(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)^a)} \frac{d(1 - F(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)^a)}{d\mu}.$$
$$U_\sigma(\xi) = -\frac{n}{\sigma} + \sum_{i=0}^{n} \frac{1}{\sigma f(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)} \frac{df(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)}{d\sigma}$$
$$- \frac{(a-1)}{\sigma}\sum_{i=0}^{n} \frac{1}{F(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)} \frac{dF(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)}{d\sigma}$$
$$+ \frac{(b-1)}{\sigma}\sum_{i=0}^{n} \frac{1}{(1 - F(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)^a)} \frac{d(1 - F(\frac{x_i-\sigma}{\sigma}; \mu, \sigma, \lambda, r)^a)}{d\sigma}.$$
$$U_\lambda(\xi) = \sum_{i=0}^{n} \frac{1}{f(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)} \frac{df(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)}{d\lambda}$$
$$+ (a-1)\sum_{i=0}^{n} \frac{1}{F(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)} \frac{dF(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)}{d\lambda}$$
$$+ (b-1)\sum_{i=0}^{n} \frac{1}{(1 - F(\frac{x_i-\mu}{\sigma}; \mu, \sigma, \lambda, r)^a)} \frac{d(1 - F(\frac{x_i-\sigma}{\sigma}; \mu, \sigma, \lambda, r)^a)}{d\lambda}.$$

$$U_r(\xi) = \sum_{i=0}^{n} \frac{1}{f(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)} \frac{df(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)}{dr}$$
$$+ (a-1)\sum_{i=0}^{n} \frac{1}{F(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)} \frac{dF(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)}{dr}$$
$$+ (b-1)\sum_{i=0}^{n} \frac{1}{(1-F(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)^a)} \frac{d(1-F(\frac{x_i-\sigma}{\sigma};\mu,\sigma,\lambda,r)^a)}{dr}.$$

Solving the components of the score vector simultaneously yield the maximum likelihood estimates (MLEs) of the six parameters. Estimation of each parameter can be carried out using one of the numerical procedures available on computational software. We used the *optim* function which is available in *R* software to do so.

6.1 Illustrative examples

We illustrate the superiority of the *KwST* distribution proposed here as compared with some of its sub-models using the Akaike Information Criterion (AIC) and Schwarz information criterion (SIC). We give an application using well-known data set to demonstrate the applicability of the proposed model. Table is used to display the six parameters $\xi = (\mu, \sigma, \lambda, r, a, b)$ estimate for each model with the stander error of estimation in parenthesis as will as the negative log-likelihood, the AIC and the SIC values.

The data set used here is the U.S. indemnity losses used in Frees and Valdez (1998) and Eling (2012), to name a few. This data set contains 1500 general liability claims giving for each the indemnity payment, denoted by "loss". For scaling purposes, we divide the data set by 1000. The U.S. indemnity losses data set can be found in the *R* packages *copula* and *evd*.

Figure 6 presents the histogram for the U.S. indemnity losses data set, as well as the corresponding normal Q-Q plot. The histogram shows that we have a large number of small losses and a lower number of very large losses which is a typical feature of insurance claims data.

Table 2: Summary description of the U.S. indemnity losses data set.

N	Min.	Median	Mean	sd	Max.	skewness	kurtosis
1500	0.01	12.00	41.21	102.74	2174.00	9.154	141.978

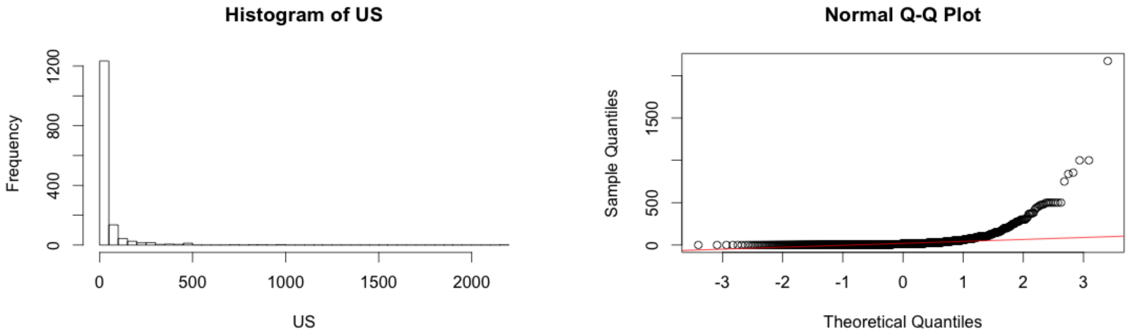


Figure 6: Histogram and Q-Q plot for U.S. indemnity losses data set.

Table 2 shows descriptive statistics for the U.S. indemnity losses data set. It presents the number of observations, indicators for the first four moments (mean, standard deviation, skewness, excess kurtosis), and the minimum and the maximum observation of the data. The descriptive statistics indicate that the U.S. indemnity losses data set is significantly skewed to the right and exhibit high kurtosis.

To illustrate the fitting superiority of the $KwST$ distribution as well as the comparison to other existing distributions to indicate its advantage, the Akaike Information Criterion (AIC) and Schwarz information criterion (SIC) are used. The calculation results are presented in Table 3. We observe that the $KwST$ distribution is a competitive candidate to fit the data as its AIC and SIC values are very close to the AIC and SIC of the St_r distribution. Further, note that for the St_r distribution the estimated skewness parameter λ is very large while the $KwST$ distribution produced a reasonable estimated values of its parameters. On the other hand, we note that the SN distribution fails to fit this data as it has the largest AIC and SIC values and the estimated skewness parameter is out of the range (-20,20) suggested by Azzalini (1986). Therefore, we suggest using the $KwST$ distribution to fit this data set.

The following figures are graphical display of the fitted density curves to the histogram of the U.S. indemnity losses data where the (———) line presents the $KwST$ model, the (- - - - -) line presents the skew-t model and the (..........) line presents the skew normal one.

[1]The stander error of estimations are reported in parenthesis.

Table 3: Parameter estimations for the U.S. indemnity losses data set. [1]

Dist.	μ	σ	λ	r	a	b	$-log(\xi)$	AIC	SIC
$KwST$	-0.0652	4.936	5.337	0.078	2.754	38.776	6595.183	13202.37	13234.24
	(0.00005)	(0.0001)	(0.0001)	(2.296x10^{-07})	(0.0067)	(3.358)			
St_r	0.0096	10.687	80448.45	0.859	-	-	6594.952	13197.99	13219.16
	(0.0418)	(0.5583)	(17.263)	(0.0486)					
SN	9.53x10^{-03}	1.1064x10^{02}	8.378x10^{05}	-	-	-	8148.48	16302.98	16318.92
	(0.00002)	(2.020)	(1.482x10^{3})						

We take a closer look to the fitting density curves in Figure 8 to show the advantage of our new model.

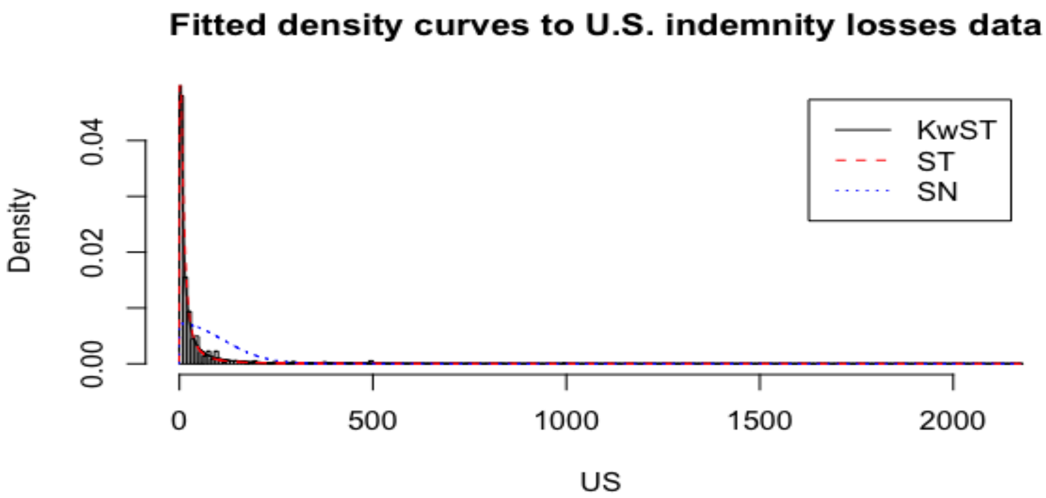


Figure 7: Histogram and fitted density curves to the U.S. indemnity losses data set.

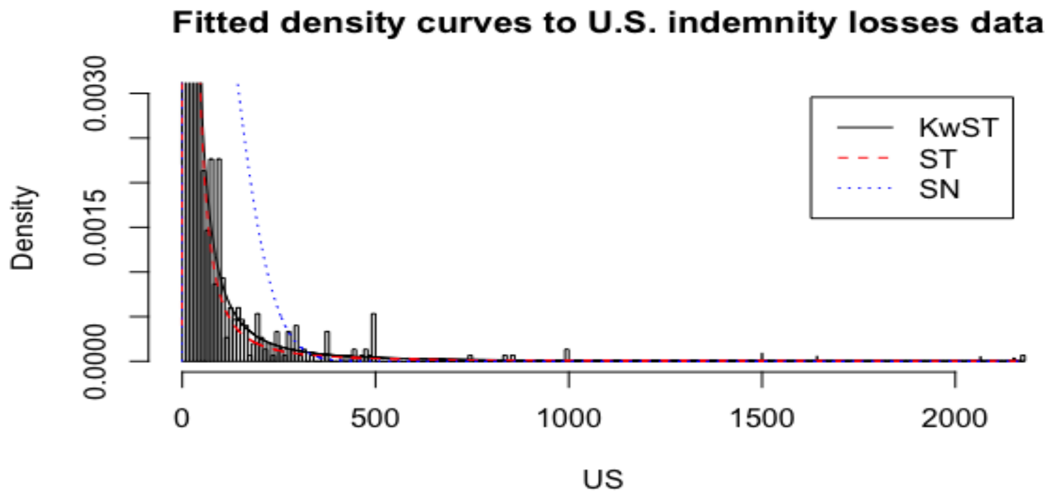


Figure 8: Closer look for fitted density curves to the U.S. indemnity losses data set.

7 Final remarks

A new statistical distribution, The Kumaraswamy skew- t distribution, has been introduced and denoted as $KwST$ with some structural properties. The $KwST$ distribution provides flexibility in modeling heavy-tailed and skewed data and it is more general than the skew-t distribution as it includes the t_r, $st_r(\lambda)$, $Kw - t$, $KwSN$ and some other important distributions as special cases of its parameters. The n^{th} moment of $X \sim KwST(a, b, \lambda, r)$ random variable with integers $a \geq 2$ and $b \geq 2$ can be expressed in terms of the n^{th} moment of the $st_r(\lambda)$ multiplied by a constant. The density of the order statistics of the $KwST$ can be written as a function of finite weighted sum of the density of the same $KwST$ distribution with parameter b^* which is written as a function of the sample size$n$, the order i and a positive constant k, where $0 \leq k \leq i-1$. The density of the smallest order statistic is nothing but the pdf of $Y \sim KwST(a, bn, \lambda, r)$. We provide an application using a well-known data set to demonstrate the applicability of the proposed model by comparing it with some of its sub-models using the Akaike Information Criterion (AIC) and Schwarz information criterion (SIC). We conclude that the $KwST$ distribution is a promising model when modeling skewed and heavy tailed data.

Acknowledgement

The authors are grateful to the editors and one anonymous referee for their valuable comments and suggestions, which substantially improved the quality of the paper.

References

[1] Abtahi, A, Towhidi,M. and Behboodian, J. (2011). An appropriate empirical version of skew-normal density. Statist. Papers, 52, 469-489.

[2] Arellano-Valle, R. B. and Genton, M. G. (2005). On fundamental skew distributions. J. of Multiv. Anal., 96, 93-116.

[3] Arnold, B, Balakrishnan, N and Nagaraja,H. (2008). A Firist Course in Order Statistics. Philadelphia, PA : SIAM

[4] Azzalini, A. (1985) A class of distributions which includes the normal ones. Scand. J.Statist. 12, 171-178.

[5] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. Statistica, 46, 199-208.

[6] Azzalini, A. (2014). **sn**. R package version 1.4-0, URL https://cran.r-project.org/web/packages/sn/index.html.

[7] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. J. Roy. Statist. Soc., B, 65, 367-389.

[8] Azzalini, A and Capitanio, A. (2014). The skew-normal and related families. Cambridge University Press.

[9] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew normal distribution. Biometrika, 83, 715-726.

[10] Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. J. Stat. Comput. Simul. 81, 883-898.

[11] Cordeiro, G. Ortega, E. M. M. and Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data. J. Franklin Inst., 347, 1399-1429.

[12] Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models?. Insurance: Math. and Eco., 51, 2, 239-248.

[13] Ferreira, J.T.A.S.,and Steel, M.F.J. (2006). A constructive representation of univariate skewed distributions. J. Amer. Statist. Assoc., 101 (474), 823-829.

[14] Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. North American Actuarial Journal, 2(1):1-25.

[15] Gupta, A. K., Chang, F. C. and Huang, W. J. (2002). Some skew symmetric models. Random Oper. and Stoch. Eqn., 10, 133-140.

[16] Gupta, A. K., Nadarajah, S. (2005). On the moments of the beta-normal distribution. Comm. Statist. Theory Methods, 33, 1-13.

[17] Gupta, A. K. and Nagar, D. K. (2000). Matrix variate distributions. Chapman and Hall Ltd.: New York.

[18] Hasan, A. (2013). A Study of non-central skew-t Distributions and their Applications in Data Analysis and Change Point Detection, Ph.D. dissertation, Bowling Green State University, USA.

[19] Huang, W. J. and Chen, Y. H. (2006). Quadratic forms of multivariate skew normal-symmetric distributions. Statist. and Prob. Lett., 76, 871-879.

[20] Jones, M.C. (2004). Families of distributions of order statistics. Statistical Methodology. 6, 70-91.

[21] Jones, M.C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. TEST, 13, 1-43.

[22] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. J. Hydrology. 46, 79-88.

[23] Mameli, V. (2015). The Kumaraswamy skew-normal distribution. Statist. and Prob. Letters, 104, 75-81.

[24] Nadarajah, S., Cordeiro, G.M. and Ortega, E.M.M (2012). General results for the Kumaraswamy-G distribution. J. Statist. Comput. Simul., 82 (7), 951-979.

[25] Shafiei, S and Doostparast, M. (2014). Balakrishnan skew-t distribution and associated statistical characteristics. Comm. Statist. Theory Methods, 43, 4109-4122.