JOURNAL OF STATISTICAL THEORY AND PRACTICE
2018, VOL. 12, NO. 4, 837–860
https://doi.org/10.1080/15598608.2018.1481468

Check for updates

The beta skew *t* distribution and its properties

Doaa Basalamah[a,b], Wei Ning[b] and Arjun Gupta[b]

[a]Mathematical Science Department, Umm Al Qura University, Mecca, Saudi Arabia; [b]Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, USA

ABSTRACT
In this article we introduce a new generalization of the skew *t* distribution based on the beta generalized distribution. The new class of distribution, which is called the beta skew *t* (*BST*), has the ability of fitting skewed and heavy-tailed data and is more general than the skew *t* distribution as it contains the skew *t* distribution as a special case. Related properties of the new distribution, such as moments and the order statistics, are derived. The proposed distribution is applied to real data to illustrate the fitting procedure using the maximum likelihood method and the L-moments method.

ARTICLE HISTORY
Received 13 August 2017
Accepted 23 May 2018

KEYWORDS
Beta skew *t* distribution; skew normal distribution; heavy-tailed data; maximal likelihood estimate; L-moments

AMS SUBJECT CLASSIFICATION
62E15; 62F10

1. Introduction

Azzalini (1985) introduced the univariate skew normal distribution as an extension of the normal distribution to accommodate asymmetry. Inspired by Azzalini's work, numerous work has been done on the applications of skewed distributions. Among all skewed distributions, the skew *t* distribution received special attention after the introduction of the skew multivariate normal distribution by Azzalini and Dalla Valle (1996). Gupta (2003) defined the skew multivariate *t* distribution using a pair of independent standard skew normal and chi-squared random variables. Azzalini and Capitanio (2003) 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 and Genton (2005) discussed generalized skew distributions in the multivariate setting, including the skew *t*. Huang and Chen (2006) studied generalized skew *t* distributions and used them in data analysis. Hasan (2013) presented a new approach to define the noncentral skew *t* distribution. Shafiei and Doostparast (2014) introduced the Balakrishnan skew *t* distribution and its associated statistical characteristics, to name a few. To provide a wide and flexible family to model data that accounts for skewness and heavy tail weight, Jones (2004) introduced the beta-generated distribution as a generalization of the distribution of order statistics of a random sample from a distribution F or by applying the inverse probability integral transformation to the beta distribution.

The beta normal distribution was introduced by Eugene, Lee, and Famoye (2002). In their work they studied the shape properties of the beta normal distribution as well as estimation of the parameters using the maximum likelihood method. Silva, Ortega, and Cordeiro (2010) proposed the modified Weibull distribution. Cordeiro and De Castro

CONTACT Wei Ning wning@bgsu.edu Department of Mathematics and Statistics, Bowling Green State University, 450 MSC, Bowling Green, OH 43403, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ujsp.

(2011) studied the beta Weibull geometric distribution and its properties. Cordeiro and Nadarajah et al. (2011) derived a closed form expression for moments of the class beta generalized distributions. Rêgo and Nadarajah (2011) provided more detailed properties of the beta normal distribution. As a generalization of the skew normal, Mameli and Musio (2013) introduced the beta skew normal distribution, to name a few.

In this article we introduce a new generalization of the skew t distribution based on the beta generalized distribution. The new class of distribution, which is called the beta skew t (BST), has the ability of fitting skewed and heavy tailed data and is more general than the skew t distribution, as it contains the skew t distribution as a special case. Related properties of the new distribution, such as moments and the order statistics, are derived. The proposed distribution is applied to real data to illustrate the fitting procedure using the maximum likelihood method and the L-moments method. Further, parameter estimation for simulated and real life data is conducted to illustrate the advantage of L-moments method over the maximum likelihood estimators (MLEs).

2. Density and distribution functions

For a continuous distribution F with the density function f and parameters $a>0$ and $b>0$, Jones (2004) defined the density of the beta-generated distribution g_F by

$$g_F(x;a,b) = \frac{1}{B(a,b)} f(x)F(x)^{a-1}(1-F(x))^{b-1}, \tag{1}$$

where $B(a,b)$ is the complete beta function, defined by

$$B(a,b) = \int_0^1 t^{(a-1)}(1-t)^{(b-1)}dt, \tag{2}$$

for $a,b \in \Re^+$. The distribution function G_F is given by

$$G_F(x;a,b) = I_{F(x)}(a,b), \tag{3}$$

where $I_{F(x)}(a,b)$ is the incomplete beta function ratio, defined by

$$I_{F(x)}(a,b) = \frac{B_{F(x)}(a,b)}{B(a,b)}, \quad 0 \le F(x) \le 1, \tag{4}$$

and $B_{F(x)}(a,b)$ is the incomplete beta function, defined by

$$B_{F(x)}(a,b) = \int_0^{F(x)} z^{a-1}(1-z)^{b-1}dz. \tag{5}$$

Thus, the distribution function G_F can be written as

$$G_F(x;a,b) = \frac{1}{B(a,b)} \int_0^{F(x)} z^{a-1}(1-z)^{b-1}dz, \quad 0 \le F(x) \le 1. \tag{6}$$

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

Replacing $F(x)$ by $F(x;\lambda,r)$ in Eq. (6), we defined the beta skew t distribution denoted by $BST(a,b,\lambda,r)$ as follows.

Definition 2.1. A random variable X is said to have the beta skew t distribution if it has the distribution function given by

$$G_F(x;a,b,\lambda,r) = \frac{1}{B(a,b)} \int_0^{F(x;\lambda,r)} z^{a-1}(1-z)^{b-1}dz, \tag{7}$$

and probability density function (pdf)

$$g_F(x;a,b,\lambda,r) = \frac{1}{B(a,b)} f(x;\lambda,r)F(x;\lambda,r)^{a-1}(1-F(x;\lambda,r))^{b-1}, \tag{8}$$

where $-\infty < x < \infty$, $a,b>0$, and $f(x;\lambda,r)$ and $F(x;\lambda,r)$ are the *pdf* and cumulative density function *cdf* of the skew-t distribution. The term $f(x;\lambda,r)$ is given by Azzalini and Capitanio (2014) as follows:

$$f(x;\lambda,r) = 2t(x;r)T(\lambda x\sqrt{\frac{r+1}{x^2+r}};r+1), \tag{9}$$

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

2.1. *Expansion of the density and distribution function*

Using Newton's binomial expansion for $b \in \Re^+$, the *pdf* of $BST(a,b,\lambda,r)$ in Eq. (8) can be rewritten as

$$g_F(x;a,b,\lambda,r) = \frac{1}{B(a,b)} \sum_{k=0}^{\infty} (-1)^k \binom{b-1}{k} f(x;\lambda,r)F(x;\lambda,r)^{a+k-1}. \tag{10}$$

If $b \in \mathbb{Z}^+$, then the index k in the sum in Eq. (10) stops at $b-1$.

In order statistics literature, Rohatgi and Ehsanes Saleh (1988) generalized Eq. (3) as follows:

$$G_{F(x)}(x;a,b) = \sum_{k=0}^{b-1} \binom{a+b-1}{k} F(x)^{a+b-k-1}(1-F(x))^k, \tag{11}$$

where $b \in \mathbb{Z}^+$ and $a \in \Re^+$, and

$$G_{F(x)}(x;a,b) = 1 - \sum_{k=0}^{a-1} \binom{a+b-1}{k} F(x)^k(1-F(x))^{a+b-k-1}, \tag{12}$$

where $a \in \mathbb{Z}^+$ and $b \in \Re^+$.

According to Gupta and Nadarajah (2004, 12), the integral representation for incomplete beta ratio $G_{F(x;\lambda,r)}(a,b)$ can be written as

$$G_{F(x;\lambda,r)}(a,b) = \frac{F(x;\lambda,r)^a}{aB(a,b)\,B(1-b,\,a+b)} \int_0^1 Z^{-b}\,(1-z)^{a+b-1}(1-zF(x;\lambda,r))^{-a}dz. \tag{13}$$

3. Properties and simulations

In this section we study some theoretical properties of the proposed distribution. Then we provide graphical illustrations of these properties. Finally, we discuss a classical approach to generate a random sample from *BST* distribution.

3.1. Properties

Proposition 3.1. *Let $X \sim BST(a,b,\lambda,r)$,then:*

(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 $X \sim beta-t_r(a,b)$.

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

(f) If $a=1$,then $X \sim KwST(1,b,\lambda,r)$.

(g) If $b=1$,then $X \sim KwST(a,1,\lambda,r)$.

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

(i) If $Y=1-F(x;\lambda,r)$,then $X \sim beta(b,a)$.

(j) $Y=(F(X;\lambda,r))^{1/a} \sim Kw(a,b)$.

(k) $Y=(1-F(X;\lambda,r))^{1/b} \sim Kw(b,a)$.

The proof of Proposition 3.1 follows directly from Eq. (8) and elementary properties of the skew *t* distribution. Note that in parts (d) and (e), the distribution function of $beta-t_r(a,b)$ and $beta-Cauchy(a,b,0,1)$ are given by substituting the $F(x)$ in Eq. (3) by the the distribution function of the Student *t* with degrees of freedom *r* and the distribution function of $Cauchy(0,1)$, respectively. The proofs of the following properties are given in the appendix.

Proposition 3.2. *Let $X \sim BST(a,b,\lambda,r)$ with pdf $g_F(x;a,b,\lambda,r)$ in Eq. (8), then:*

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

(b) As $r \to \infty$, $X \sim beta-SN(a,b,\lambda)$.

(c) As $\lambda \to \infty$, $X \sim beta-|t_r|(a,b,r)$.

Proposition 3.3. *Let $Y_1, Y_2, \ldots..Y_n$ be a random sample of size n from a $st_r(\lambda)$ with probability density function $f(x;\lambda,r)$ defined in (9) and distribution function $F(x;\lambda,r)$. Let $Y_{1:n} \le Y_{2:n} \le \ldots \le Y_{n:n}$ be the order statistics of the random sample. Then:*

(a) The ith order statistic is $Y_{i:n} \sim BST(i,n-i+1,\lambda,r)$,where $i=1,2,\ldots,n$.

(b) The largest order statistic is $Y_{n:n}(y)=max\{Y_1,\ldots,Y_n\} \sim BST(n,1,\lambda,r)$.

(c) The smallest order statistic is $Y_{1:n}(y)=min\{Y_1,\ldots,Y_n\} \sim BST(1,n,\lambda,r)$.

Proposition 3.4. *Let $X \sim BST(a,b,\lambda,r)$ be independent from $(Y_1,Y_2,\ldots..Y_n)$,which is a random sample of size n from $st_r(\lambda)$ with probability density function $f(x;\lambda,r)$ defined in Eq. (9) and distribution function $F(x;\lambda,r)$. Let $Y_{1:n} \le Y_{2:n} \le \ldots \le Y_{n:n}$ be the order statistics of the random sample. Then,*

(a) $W=(X|Y_{1:n} \ge X) \sim \mathrm{BST}(a,b+n,\lambda,r)$,

(b) $W^* = (X|Y_{n:n} \leq X) \sim \text{BST}(a+n, b, \lambda, r)$,
where $Y_{1:n} = min\{Y_1, Y_2, ...Y_n\}$ *and* $Y_{n:n} = max\{Y_1, Y_2, ...Y_n\}$.

Proposition 3.4 can be generalized to any $c, d \in \Re^+$ as follows.

Proposition 3.5. *Let* $X \sim BST(a, b, \lambda, r)$ *be independent from* $Y \sim BST(c, 1, \lambda, r)$ *and* $Z \sim BST(1, d, \lambda, r)$, *where* $c \in \Re^+$ *and* $d \in \Re^+$. *Then:*

(a) $(X|Y \leq X) \sim BST(a + c, b, \lambda, r)$.

(b) $(X|Z \geq X) \sim BST(a, b + d, \lambda, r)$.

3.2. *Graphical illustration*

To understand the effect of the parameters on the overall shape of the beta skew t probability density, we illustrate different shapes of the density curve by fixing five parameters and varying the sixth one in the following figures. For simplicity, we set up the location parameter μ to be zero and the scale parameter σ to be one. In Figure 1, we study the effect of the parameter a on the density shape by fixing the remaining parameters $(b = 3, \lambda = 1, r = 3)$ and we graph the density of BST for different values of a. Figure 1 shows that the left tail of the *BST* density curve gets lighter, as a increases.

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

Figure 3 illustrates the effect of the parameter λ on the shape of the *BST* density curve 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 $BST(5, 3, \lambda, 3)$ with the curve of $beta - |t_r|(a = 5, b = 3, r = 3)$. 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 *BST* density curve overlaps the $beta - |t_r|$ density curve, which graphically proves part (c) of proposition 3.2.

In Figure 4, we study the effect of the degrees of freedom r on the shape of the *BST* 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 $BST(5, 3, -1, r)$ density

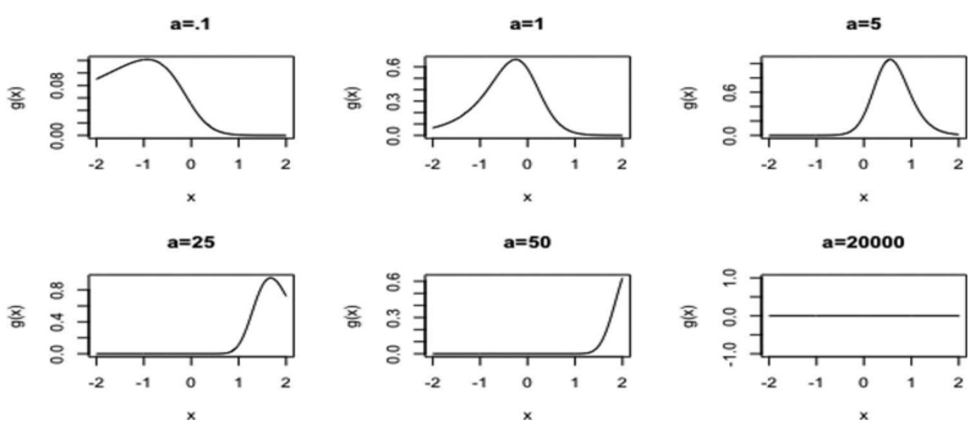


Figure 1. $BST(a, b = 5, \lambda = 1, r = 3)$ density curves as a varies.

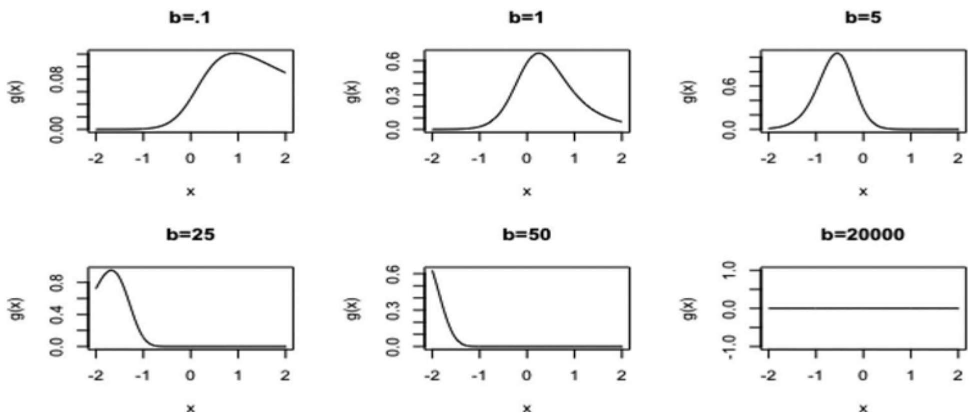


Figure 2. $BST(a = 5, b, \lambda = -1, r = 3)$ density curves as b varies.

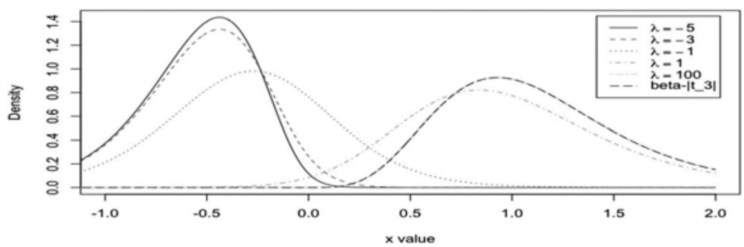


Figure 3. $BST(a = 5, b = 3, \lambda, r = 3)$ density curves as λ varies.

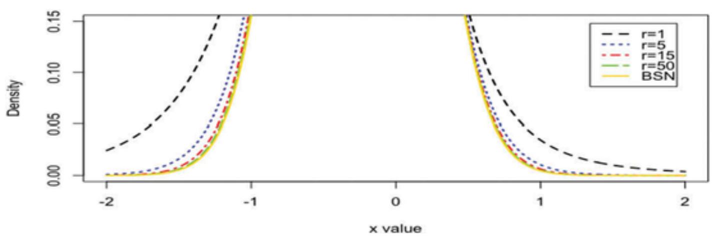
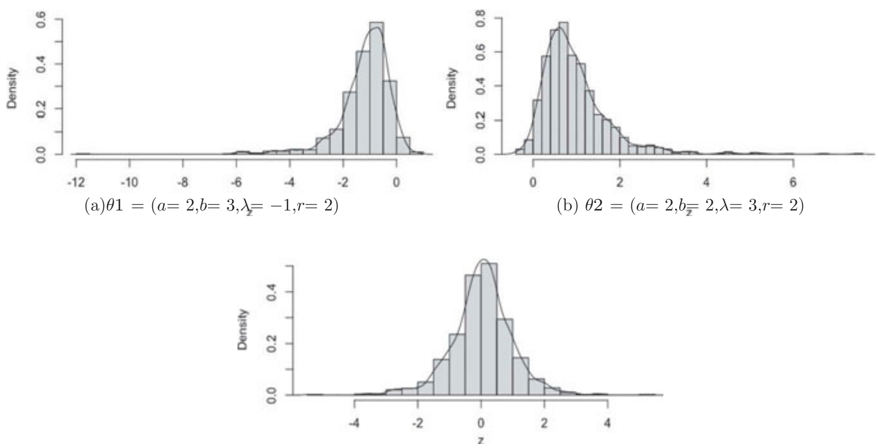


Figure 4. $BST(a = 5, b = 3, \lambda = -1, r)$ density curves as r varies.

gets closer to the one of the $BSN(5, 3, -1)$ as the degrees of freedom r increases, which agrees with part (b) of Proposition 3.2. The tail gets thicker as the degrees of freedom decrease. These two properties are inherited from the baseline skew t distribution. Furthermore, Figures 1 to 4 show that the BST inherits the unimodality from the baseline distribution.

3.3. *Simulations*

A random sample from BST can be generated using the classical inverse probability integral transform technique as follows:



(c) $\theta 3$ =(a= 2,b= 2,λ= 0,r= 2)

Figure 5. Histogram for random samples of size 1000 of *BST* distribution.

(1) Generate a random sample $Y_1, Y_2, \ldots, Y_n$ from beta(a, b) distribution.
(2) Let $X_i = F^{-1}(Y_i; \lambda, r)$, where $F^{-1}(., \lambda, r)$ is the quantile function of the skew t distribution.
(3) $X_1, X_2, \ldots, X_n \sim BST(a, b, \lambda, r)$.

Figure 5 shows histograms of three random samples of size 1000 generated from $BST(\theta)$, $\theta = (a, b, \lambda, r)$, a distribution using the classical inverse probability integral transform technique with different parameter vectors $\theta_1 = (a = 2, b = 3, \lambda = -1, r = 2)$ as in Figure 5, $\theta_2 = (a = 2, b = 2, \lambda = 3, r = 2)$ as in Figure 5, and $\theta_3 = (a = 2, b = 2, \lambda = 0, r = 2)$ as in Figure 5.

4. Moments

In this section we derive an explicit form of the nth moment as a function of the nth moment of the baseline distribution, the skew t distribution. The proofs of Theorems 4.1 and 4.2 are given in the appendix.

Theorem 4.1. *Let $X \sim BST(a, b, \mu, \sigma, \lambda, r)$. Then the* n*th moment for integer $n \geq r$ is given by*

$$E(X^n) = \frac{\sigma^n}{B(a, b)} \sum_{j=0}^{\infty} \sum_{i=0}^{n} (-1)^j \binom{b-1}{j} \binom{n}{i} (\frac{\mu}{\sigma})^i E_Y(Y^{n-i})[F(y; \lambda, r)^{a+j-1} - 1], \tag{14}$$

where $Y \sim st_r(\lambda)$,and $\mu \neq 0$. If $b \in \mathbb{Z}^+$,then the index j stops at $b-1$.

According to Azzalini and Capitanio (2014), the nth 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 $Z \sim SN(0, 1, \lambda)$.

Proposition 4.1. *Let $X \sim BST(a,b,\lambda,r)$. Then the* nth *moment for integer $n \geq r$ is given by*

$$E(X^n) = \frac{1}{B(a,b)} \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} E_Y(Y^n)[F(y;\lambda,r)^{a+j-1} - 1], \tag{15}$$

where $Y \sim st_r(\lambda)$. If $b \in \mathbb{Z}^+$,then the index j stops at $b-1$.

Alternatively, the nth moment of $X \sim BST(a,b,\lambda,r)$ as a random variable with integers $a \geq 2$ and $b \geq 2$ can be expressed as a function of the nth moment of the baseline distribution $st_r(\lambda)$ multiplied by a constant as presented in the following theorem.

Theorem 4.3. *Let $X \sim BST(a,b,\lambda,r)$ with integers $a \geq 2$, $b \geq 2$, $n > 0$,and $r \geq n$:*

$$E(X^n) = c(a,b)E_Y(Y^n), \tag{16}$$

where

$$c(a,b) = \frac{1}{B(a,b)} \left[\sum_{i=0}^{b-2} \frac{(-1)^i}{B(i+1,b-i-1)} \left[\frac{1}{a+i} - \frac{(a-1)}{(a+i-1)(b-i-1)} \right] - (-1)^{b-1} \frac{(a-1)}{a+b-2} \right],$$

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

5. Order statistics

Order statistics make their appearance in many areas of statistical theory and practice. In this section we derive an explicit form of the probability density function of the *BST* order statistics. The proofs of Theorems 5.1 and 5.2 are given in the appendix.

Theorem 5.1. *Let $X_1, \ldots, X_n$ be a random sample from BST distribution with distribution function $G_F(x;a,b,\lambda,r)$ in Eq. (7) and probability density function $g_F(x;a,b,\lambda,r)$ in Eq. (8). Let $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ be the order statistics of the random sample. The density function of the ith order statistic, for $i = 1, \ldots, n$,is given by*

$$g_{i:n}(x) = \sum_{k=0}^{n-i} (-1)^k i \binom{n}{i} \binom{n-i}{k} g_F(x;a^*,b,\lambda,r) h^*(x) c_k(a,b), \tag{17}$$

where $a^* = a(k+i)$,

$$c_k(a,b) = \frac{B(a^*,b)}{B(a,b)[aB(a,b)B(1-b,a+b)]^{k+i-1}},$$

and

$$h^*(x) = [\int_0^1 z^{-b}(1-z)^{a+b-1}(1-zF(x;\lambda,r))^{-a} dz]^{k+i-1}.$$

According to Thukral (2014), the beta function $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ can be relaxed to include all real numbers a and b. Therefore, the preceding expression of the density function $g_{i:n}(x)$ of the ith order statistic is true for all real $a, b \in \Re$.

Using the density of the ith order statistic we derived in Theorem 5.1, we provide the expression of the largest and the smallest order statistics of a $BST(a, b, \lambda, r)$ random sample as follows.

Corollary 5.1. Let $X_1, ..., X_n$ be a random sample from $BST(a, b, \lambda, r)$ distribution. Then, for $b \in \mathbb{Z}^+$ an integer and $a \in \Re^+$:

(a) The density of the largest order statistic $X_{n:n}(x) = max\{x_1, ..., x_n\}$ is given by

$$g_{n:n}(x) = ng_F(x; na, b, \lambda, r)h(x)^{n-1} \frac{B(an, b)}{B(a, b)[aB(a, b)B(1 - b, a + b)]^{n-1}}. \tag{18}$$

(b) The density of the smallest order statistic $X_{1:n}(x) = min\{x_1, ..., x_n\}$ is given by

$$g_{1:n}(x) = \sum_{k=0}^{n-1} (-1)^k n \binom{n-1}{k} g_F(x; a^*, b, \lambda, r)h(x)^k \frac{B(a^*, b)}{B(a, b)[aB(a, b)B(1 - b, a + b)]^k}, \tag{19}$$

where $h(x) = \int_0^1 z^{-b}(1 - z)^{a+b-1}(1 - zF(x; \lambda, r))^{-a}dz$ and $a^* = a(k + 1)$.

Using expression (13) of the incomplete beta function and for integer $b>0$, the ith order statistics of $X \sim BST(a, b, \lambda, r)$ can be written as follows.

Theorem 5.2. *Let $X_1, ..., X_n$ be random variables from BST distribution with common distribution function $G_F(x; a, b, \lambda, r)$ in Eq. (7) and probability density function $g_F(x; a, b, \lambda, r)$ in Eq. (8). Let $X_{1:n} \le X_{2:n} \le ... \le X_{n:n}$ be the order statistics. The density of the ith order statistic, for $i = 1, ..., n$, is given by*

$$g_{i:n}(x) = \sum_{k=0}^{n-i} (-1)^k i \binom{n}{i} \binom{n-i}{k} g_F(x; a, b, \lambda, r) \left\{ \frac{1 - F(x; \lambda, r)}{f(x; \lambda, r)} \sum_{j=0}^{b-1} g_F(x; (a + b - j), j, \lambda, r) \right\}^{k+i-1}, \tag{20}$$

where $b \in \mathbb{Z}^+$ and $a \in \Re^+$.

6. Maximum likelihood estimation

In this section, the maximum likelihood estimators (MLEs) of the BST parameters are given. Let $x_1, x_2, ..., x_n$ be a random sample of size n from the $BST(a, b, \mu, \sigma, \lambda, r)$ distribution. The log-likelihood function $l(\theta)$ for the parameter vector of $\theta = (a, b, \mu, \sigma, \lambda, r)$ can be written as

$$l(\theta) = nlog(\Gamma(a + b)) - nlog(\Gamma(a)) - nlog(\Gamma(b)) - nlog(\sigma) + \sum_{i=1}^{n} logf(z_i; \mu, \sigma, \lambda, r)$$
$$+ (a - 1) \sum_{i=1}^{n} log(F(z_i; \mu, \sigma, \lambda, r)) + (b - 1) \sum_{i=1}^{n} log(1 - F(z_i; \mu, \sigma, \lambda, r)), \tag{21}$$

where $z_i = \frac{x_i - \mu}{\sigma}$. The log-likelihood can be maximized either directly by using the *optim* function in R or by solving the nonlinear likelihood equations obtained by differentiating Eq. (20). The components of the score vector $U(\theta)$ are given by

$$
\begin{aligned}
U_a(\theta) &= n\psi(a+b) - n\psi(a) + \sum_{i=0}^{n} log(F(z_i;\mu,\sigma,\lambda,r)),\\
U_b(\theta) &= n\psi(a+b) - n\psi(b) + \sum_{i=0}^{n} log(1-F(z_i;\mu,\sigma,\lambda,r)),\\
U_\mu(\theta) &= \sum_{i=0}^{n} \frac{-1}{\sigma f(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)} \frac{df(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)}{d\mu}\\
&\quad - \frac{(a-1)}{\sigma}\sum_{i=0}^{n} \frac{1}{F(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)} \frac{dF(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)}{d\mu}\\
&\quad + \frac{(b-1)}{\sigma}\sum_{i=0}^{n} \frac{1}{(1-F(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r))} \frac{d(1-F(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r))}{d\mu},\\
U_\sigma(\theta) &= -\frac{n}{\sigma} + \sum_{i=0}^{n} \frac{1}{\sigma f(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)} \frac{df(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)}{d\sigma}\\
&\quad - \frac{(a-1)}{\sigma}\sum_{i=0}^{n} \frac{1}{F(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)} \frac{dF(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)}{d\sigma}\\
&\quad + \frac{(b-1)}{\sigma}\sum_{i=0}^{n} \frac{1}{(1-F(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r))} \frac{d(1-F(\frac{x_i-\sigma}{\sigma};\mu,\sigma,\lambda,r))}{d\sigma},
\end{aligned}
$$

$$
\begin{aligned}
U_\lambda(\theta) &= \sum_{i=0}^{n} \frac{1}{f(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)} \frac{df(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)}{d\lambda}\\
&\quad +(a-1)\sum_{i=0}^{n} \frac{1}{F(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)} \frac{dF(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r)}{d\lambda}\\
&\quad +(b-1)\sum_{i=0}^{n} \frac{1}{(1-F(\frac{x_i-\mu}{\sigma};\mu,\sigma,\lambda,r))} \frac{d(1-F(\frac{x_i-\sigma}{\sigma};\mu,\sigma,\lambda,r))}{d\lambda},
\end{aligned}
$$

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

where $\psi(x)$ is the digamma function defined by $\frac{d}{dx}\log\Gamma(x)$.

6.1. *Illustrative examples*

We illustrate the superiority of the *BST* distributions proposed here by comparing with some of its submodels such as the beta *t* distribution Bt_r and the *t* distribution t_r using the Akaike information criterion (AIC) and Schwarz information criterion (SIC). We give an application using well-known data sets to demonstrate the applicability of the proposed model. Tables are used to display the six parameters $\theta = (\mu,\sigma,\lambda,r,a,b)$ estimate for each model with 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). This data contains 1500 general liability claims giving for each the indemnity payment, denoted by "loss." For the purposes of scaling, we divide the data set by 1000.

The U.S. indemnity losses data are available in the *R* packages *copula* and *evd*. Descriptive statistics of the data are given in Table 1.

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. Descriptive statistics of the U.S. indemnity losses data set are given in Table 1.

From Table 2, we observe that the *BST* model has the smallest SIC value among all other models, which indicates that it provides the best fit.

Figure 7 presents a graphical display of the density curves fitted to the histogram of the U.S. indemnity losses data where the solid line presents the *BST* density curve, the dashed line presents the Bt_r density curve, and the dotted line presents the t_r density curve. Figure 8 presents a closer look at the fitted density curves.

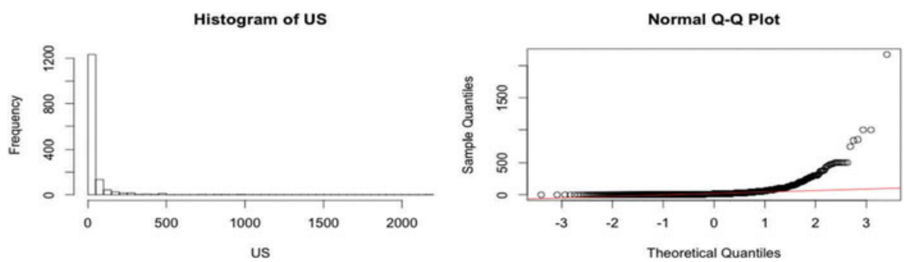


Figure 6. Histogram and Q–Q plot for U.S. indemnity losses data set.

Table 1. Summary description of the U.S. indemnity losses data set.

Minimum	Median	Mean	SD	Maximum	Skewness	Kurtosis
0.01	12.00	41.21	102.74	2174.00	9.154	141.978

Table 2. Parameter estimations for the U.S. indemnity losses data set.

Distribution	μ	σ	λ	r	a	b	$log(\theta)$	AIC	SIC
BST	0.532	1.548	0.644	0.383	8.554	2.764	6596.379	13204.76	13236.64
Bt_r	1.539	2.533		0.238	7.429	3.363	6722.746	13455.49	13482.06
t_r	7.383	7.317		0.788			7243.32	14492.64	14508.58

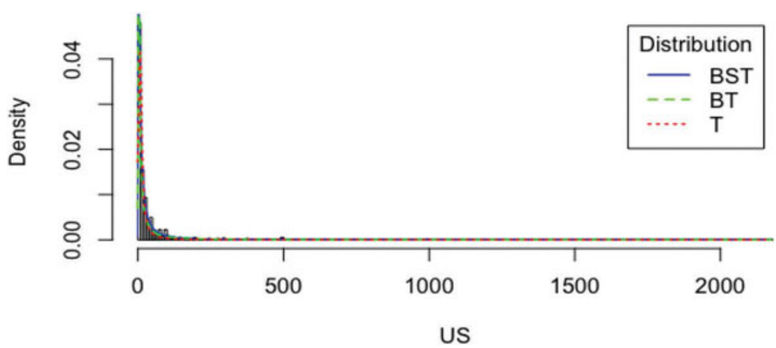


Figure 7. Histogram and density curves fitted to the U.S. indemnity losses data.

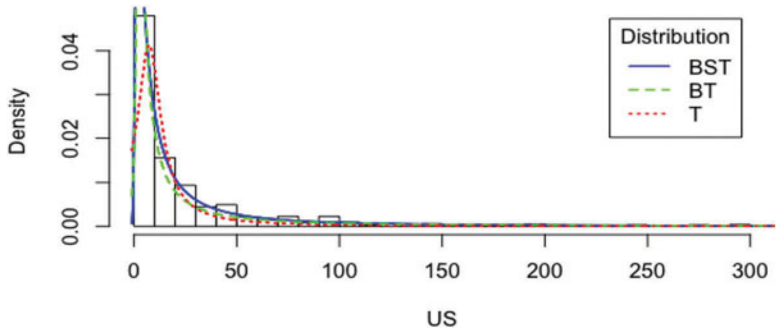


Figure 8. Closer look at the histogram and density curves fitted to the U.S. indemnity losses data.

Table 3. Parameter estimations for the U.S. indemnity losses data set.

Distribution	μ	σ	λ	r	a	b	$-log(\theta)$	AIC	SIC
BST	0.532	1.548	0.644	0.383	8.554	2.764	6596.379	13204.76	13236.64
st_r	0.0096	10.687	80448.45	0.859			6594.952	13197.9	13219.16

Finally, in Table 3 we compare the fitting superiority of the *BST* distribution with the baseline distribution st_r. We observe that the *BST* 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 skew t distribution. Further, note that for the st_r distribution the estimated skewness parameter λ is very large while the *BST* distribution produced a reasonable estimated value of the parameter λ. Therefore, we suggest using the *BST* distribution to fit this data set. Figure 9 shows the graphical display of the fitted density curves to the histogram of the U.S. indemnity losses data, while a closer look to demonstrate the tail fitting for both distributions is presented in Figure 10. From the fitting results we conclude that the *BST* distribution is very promising distribution that has the ability to fit very skewed and heavy tailed data.

7. L-moments estimation

The L-moments are defined as linear combinations of expectations of order statistics that exist for any random variable with a finite mean. L-moments are useful in fitting

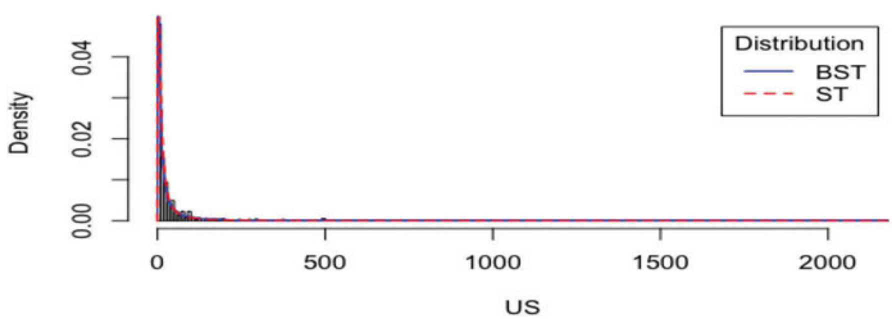


Figure 9. *BST* versus st_r MLE fitting to the U.S. indemnity losses data set.

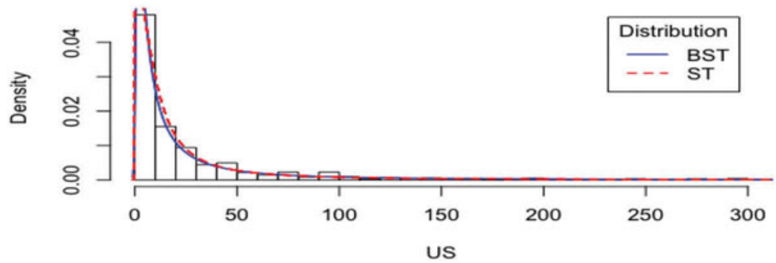


Figure 10. Closer look at *BST* versus st_r MLE fitting to the U.S. indemnity losses data set.

distributions because they specify location, scale, skewness, and kurtosis. There are many advantage of L-moments over the ordinary moments. Unlike the ordinary moments, L-moments exist whenever the underlying random variable has a finite mean. In addition, when dealing with data that has large variation, large skewness, and heavy tails, L-moments have the advantage of natural unbiasedness, robustness, and often smaller sampling variances than other estimators.

In this section, following the definition of L-moments by Hosking (1990), we derive the first seven theoretical L-moments of the proposed *BST* distribution. Then, we estimate the first four L-moments and the first two L-moments ratios by varying one parameter while fixing other parameters. Further, we conduct some parameter estimation for simulated and real life data using L-moments method. Finally, we illustrate the fitting superiority of L-moments parameters estimation and compare it with the classical ML estimators by the AIC and SIC values.

7.1. *Theoretical and sample l-moments*

Denote the theoretical L-moments by $L_1, L_2, \ldots$ throughout this article. From the expectations of order statistics, Hosking (1990) defined the theoretical L-moments for a real valued random variable X as follows:

$$L_m = \frac{1}{m}\sum_{k=0}^{m-1}(-1)^k\binom{k}{i-1}E[X_{m-k:m}], \quad \text{for} \quad \mathrm{m} = 1, 2, \ldots. \tag{22}$$

where $E[X_{m-k:m}]$ is the expectation of the $m-k$ order statistic of a sample of size m. The first four theoretical L-moments are expressed by

$$\begin{aligned} L_1 &= E[X], \\ L_2 &= \tfrac{1}{2}E[X_{2:2} - X_{1:2}], \\ L_3 &= \tfrac{1}{3}E[X_{3:3} - 2X_{2:3} + X_{1:3}], \\ L_4 &= \tfrac{1}{4}E[X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}]. \end{aligned}$$

The L-moments ratio are independent of the units of measurement of X and are defined for higher moments, $m \geq 3$, as

$$\tau_m = \frac{L_m}{L_2}, m = 3, 4, \ldots \quad . \tag{23}$$

It is clear that L_1 is the mean of X and hence is a measure of location, as known as L-location. L_2 is known as L-scale, and the L-moments ratios τ_3 and τ_4 are the L-skewness and L-kurtosis, respectively. Based on the definition of the theoretical L-moments Hosking (1990), we derive the theoretical L-moments for the BST distribution as follows.

Theorem 7.1. *The theoretical L-moments for a BST random variable X with distribution function $G(X; a, b, \lambda, r)$,provided in Eq. (7), are defined as*

$$L_m = \sum_{k=0}^{m-1}\sum_{j=0}^{k}(-1)^{k+j}\binom{m-1}{k}^2\binom{k}{j}E[XG(X; a, b, \lambda, r)^{m-k+j-1}]. \tag{24}$$

Corollary 7.1. The first seven BST theoretical L-moments are expressed by

$$\begin{aligned}
L_1 &= E[XG(X; a, b, \lambda, r)],\\
L_2 &= -1E[XG(X; a, b, \lambda, r)] + 2E[XG(X; a, b, \lambda, r)^2],\\
L_3 &= E[XG(X; a, b, \lambda, r)] - 6E[XG(X; a, b, \lambda, r)^2]] + 6E[XG(X; a, b, \lambda, r)^3],\\
L_4 &= -E[XG(X; a, b, \lambda, r)] + 12E[XG(X; a, b, \lambda, r)^2] - 30E[XG(X; a, b, \lambda, r)^3]\\
&\quad +20E[XG(X; a, b, \lambda, r)^4],\\
L_5 &= E[XG(X; a, b, \lambda, r)] - 20E[XG(X; a, b, \lambda, r)^2] + 90E[XG(X; a, b, \lambda, r)^3]\\
&\quad -140E[XG(X; a, b, \lambda, r)^4] + 70E[XG(X; a, b, \lambda, r)^5],\\
L_6 &= -E[XG(X; a, b, \lambda, r)] + 30E[XG(X; a, b, \lambda, r)^2] - 210E[XG(X; a, b, \lambda, r)^3]\\
&\quad +560E[XG(X; a, b, \lambda, r)^4] - 630E[XG(X; a, b, \lambda, r)^5] + 252E[XG(X; a, b, \lambda, r)^6],\\
L_7 &= E[XG(X; a, b, \lambda, r)] - 42E[XG(X; a, b, \lambda, r)^2] + 420E[XG(X; a, b, \lambda, r)^3]\\
&\quad -1680E[XG(X; a, b, \lambda, r)^4] + 3150E[XG(X; a, b, \lambda, r)^5] - 2772E[XG(X; a, b, \lambda, r)^6]\\
&\quad +924E[XG(X; a, b, \lambda, r)^7].
\end{aligned}$$

The L-location (L_1), L-scale (L_2), L-skewness (τ_3), and L-kurtosis (τ_4) measures of $X \sim BST(a, b, \mu, \sigma, \lambda, r)$ can be computed numerically using existing software. Table 4 shows numerical estimations of these measures by computing the first four L-moments for various values of the parameters $a, b, \lambda,$ and r with fixed $\mu = 0$ and $\sigma = 1$, where Table 4a presents the numerical estimations of $BST(a, b, \lambda, r)$ random variable for different values of $a, b,$ and λ and fixed degrees of freedom $r = 5$, while in Table 4b the parameter $\lambda = 2$ is fixed and $a, b,$ and the degrees of freedom r vary.

Since the theoretical L-moments (L_m) are defined as linear functions of the expected order statistics of a sample of size m. The sample L-moments are computed from the sample of size n of order statistics $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$ as follows:

$$l_m = \frac{1}{m\binom{n}{m}}\sum_{i=1}^{n} x_{i:n}\left[\sum_{j=0}^{m-1}(-1)^j\binom{m-1}{j}\binom{i-1}{m-j-1}\binom{n-i}{j}\right]. \tag{25}$$

The sample L-moments ratio denoted as $\hat{\tau}_m, m \geq 3$ are defined as

$$\hat{\tau}_m = \frac{l_m}{l_2}, m = 3, 4, \ldots \tag{26}$$

Table 4a. Estimation of the L-location (L_1), L-scale (L_2), L-skewness (τ_3), and L-kurtosis (τ_4) of $BST(a, b, \lambda, r)$ random variable for different values of $a, b,$ and λ.

	a	b	λ	r	L_1	L_2	τ_3	τ_4
BST	1	1	−5	5	−0.931	0.451	−0.258	0.182
			−1		−0.671	0.585	−0.097	0.194
			0		0.000	0.692	0.000 0.	0.194
			1		0.671	0.585	0.097	0.194
			5		0.931	0.451	0.258	0.182
			50		0.949	0.435	0.295	0.169
			500		0.949	0.435	0.295	0.169
BHT			-		0.949	0.435	0.295	0.169
BST	5	3	−5	5	−0.535	0.157	−0.120	0.132
			−1		−0.277	0.230	−0.006	0.138
			0		0.388	0.293	0.039	0.139
			1		0.932	0.264	0.080	0.141
			5		1.051	0.232	0.130	0.137
			50		1.052	0.231	0.132	0.136
			500		1.052	0.231	0.132	0.136
BHT			-		1.052	0.231	0.132	0.136
BST	2	20	−5	5	−2.307	0.351	−0.166	0.156
			−1		−2.261	0.357	−0.160	0.156
			0		−1.747	0.341	−0.142	0.153
			1		−0.691	0.220	−0.114	0.149
			5		0.038	0.077	−0.007	0.145
			50		0.120	0.044	0.202	0.134
			500		0.121	0.043	0.210	0.130
BHT			-		0.121	0.043	0.210	0.130

Table 4b. Estimation of the L-mean (L_1), L-variance (L_2), L-skewness (τ_3), and L-kurtosis (τ_4) of $BST(a, b, \lambda, r)$ random variable for different values of $a, b,$ and r.

	a	b	λ	r	L_1	L_2	τ_3	τ_4
KwST	1	1	2	1	43.012	47.411	0.876	0.975
				5	0.849	0.505	0.174	0.193
				50	0.725	0.401	0.086	0.133
				300	0.715	0.394	0.079	0.128
				500	0.715	0.393	0.078	0.128
BSN				-	0.714	0.392	0.078	0.128
BST	5	3	2	1	1.877	0.727	0.382	0.278
				5	1.029	0.241	0.109	0.140
				50	0.930	0.198	0.058	0.126
				300	0.922	0.195	0.053	0.125
				500	0.921	0.195	0.053	0.125
BSN				-	0.920	0.194	0.052	0.125
BST	2	20	2	1	−0.565	0.395	−0.455	0.374
				5	−0.245	0.144	−0.084	0.146
				50	−0.217	0.125	−0.034	0.128
				300	−0.215	0.124	−0.030	0.127
				500	−0.215	0.123	−0.029	0.126
BSN				-	−0.214	0.123	−0.029	0.127

7.2. *L-moments parameter estimation*

To obtain L-moments of the parameters, Hosking (1990) suggested equating the first seven sample L-moments to the corresponding population quantities. Therefore, we obtain parameter estimation of the proposed distribution $BST(a, b, \mu, \sigma, \lambda, r)$ using the L-moments method numerically by minimizing the combined Pythagorean distance between the combined square errors as given by

$$(L_1 - l_1)^2 + (L_2 - l_2)^2 + (\tau_3 - \hat{\tau}_3)^2 + (\tau_4 - \hat{\tau}_4)^2 + (\tau_5 - \hat{\tau}_5)^2 + (\tau_6 - \hat{\tau}_6)^2 + (\tau_7 - \hat{\tau}_7)^2,$$

where L_i and τ_i are the theoretical L-moment and L-ratio of the BST distribution, l_i and $\hat{\tau}_i$ are the sample L-moment and L-ratio, respectively. This technique is implemented using the *optim* function in R for minimization.

7.3. Illustrative examples

To demonstrate the performance of the L-moments method compared with the maximum likelihood method, we conduct parameters estimation and data fitting using simulated data from skew *t* distribution and the Danish fire losses data set. The Danish fire losses data set consists of 2156 fire losses of more than 1 million Danish Kroner (DKK) from the year 1980 to 1990. The fire losses reported in the data set correspond to the damage to buildings, furnishings, and personal property, as well as loss of profits. This data set has been previously studied in the literature by many authors, such as McNeil (1997), Resnick (1997), Cooray and Ananda (2005), Ahn, Kim, and Ramaswami (2012), and Farias, Montoril, and Andrade (2016), to name a few. We conduct parameter estimate for the *BST* model. Then we compare the performance of the methods using the MLE method and the L-moments method using the information criteria AIC and SIC.

The following is a *BST* parameter estimation using L-moments to a random sample of size 100 generated from the skew *t* distribution with parameter vector $(\mu = 2, \sigma = 1, \lambda = 2, r = 3)$. Table 5 shows parameter estimate for the *BST* using L-moments and MLE methods. Based on the AIC and SIC criteria, the method of L-moments provides a good alternative estimation to the method of MLE.

Figure 11 shows the fitted density of $BST(a, b, \mu, \sigma, \lambda, r)$ model using both estimation procedures, where the solid line presents *BST* fitted density curve using the L-moments

Table 5. Parameters estimation of $BST(a, b, \mu, \sigma, \lambda, r)$ using the method of L-moments and MLE.

	a	b	μ	σ	λ	r	*AIC*	*SIC*
L-moments	1.975	1.270	1.976	0.799	0.525	1.924	276.0901	291.7211
MLE	1.753	1.153	1.570	1.0101	1.986	3.156	273.5269	289.1579

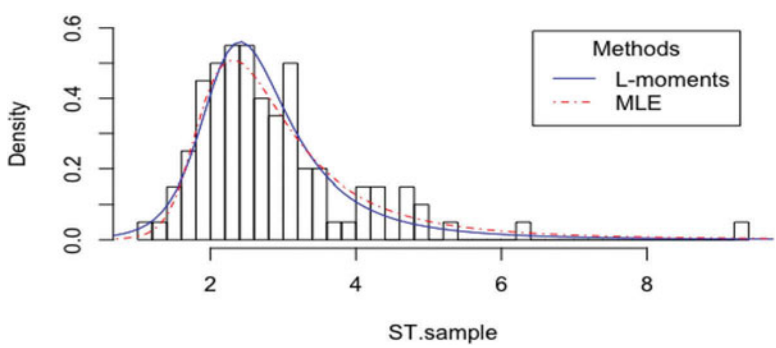


Figure 11. Fitted density of $BST(a, b, \mu, \sigma, \lambda, r)$.

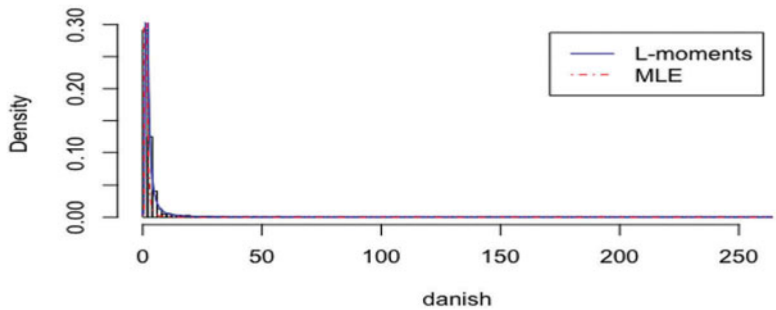


Figure 12. Fitted density of $BST(a, b, \mu, \sigma, \lambda, r)$ to the Danish fire losses data set.

estimated parameters and the dashed line presents the MLE ones. We observe that the L-moments method captured the density peak better than the ML method.

Table 6 presents the *BST* parameter estimation using L-moments and the MLE estimation method for the Danish fire losses data set. Similarly, the L-moments provide a good alternative estimation method to the MLEs based on the AIC and SIC criteria.

Figure 12 shows the fitted density of $BST(a, b, \mu, \sigma, \lambda, r)$ to the Danish fire losses data. Solid line presents BST fitted density curve using the L-moments estimated parameters, and the dashed line presents the MLE ones. Figure 13 presents a closeup look at the fitted density curves of $BST(a, b, \mu, \sigma, \lambda, r)$ using both estimation procedures, where the solid line presents BST fitted density curve using the L-moments estimated parameters and the dashed line presents the MLEs. In comparison with the MLE method, we note that the L-moments method provided a very good fit to the data set.

Table 6. $BST(a, b, \mu, \sigma, \lambda, r)$ parameters estimation of Danish fire losses data.

	a	b	μ	σ	λ	r	*AIC*	*SIC*
L-moments	4.641	7.470	0.945	0.908	1.872	0.235	6959.424	6999.192
MLE	2.680	4.343	0.9573	0.5925	4.7408	0.2730	6855.387	6889.473

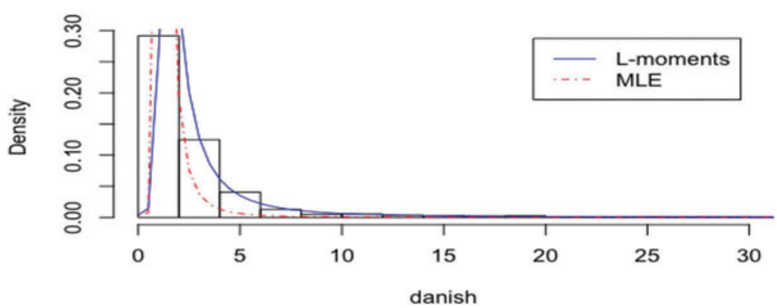


Figure 13. Closer look at the fitted density of $BST(a, b, \mu, \sigma, \lambda, r)$ to the Danish fire losses data set.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Ahn, S., J. H. Kim, and V. Ramaswami. 2012. A new class of models for heavy tailed distributions in finance and insurance risk. *Insurance: Mathematics and Economics* 51(1):43–52.

Arellano-Valle, R. B., and M. G. Genton. 2005. On fundamental skew distributions. *Journal of Multivariate Analysis* 96(1):93–116. doi:10.1016/j.jmva.2004.10.002.

Azzalini, A. 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12(2):171–178.

Azzalini, A., and A. Capitanio. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(2):367–89. doi:10.1111/rssb.2003.65.issue-2.

Azzalini, A., and A. Capitanio. 2014. *The Skew-Normal and Related Families*, Vol. 3. Cambridge, UK: Cambridge University Press.

Azzalini, A., and A. Dalla Valle. 1996. The multivariate skew-normal distribution. *Biometrika* 83 (4):715–26. doi:10.1093/biomet/83.4.715.

Cooray, K., and M. M. Ananda. 2005. Modeling actuarial data with a composite lognormal-pareto model. *Scandinavian Actuarial Journal* (5):321–34. doi:10.1080/03461230510009763.

Cordeiro, G. M., and M. De Castro. 2011. A new family of generalized distributions. *Journal of Statistical Computation and Simulation* 81(7):883–98. doi:10.1080/00949650903530745.

Cordeiro, G. M., S. Nadarajah, et al. 2011. Closed-form expressions for moments of a class of beta generalized distributions. *Brazilian Journal of Probability and Statistics*. 25(1):14–33. doi:10.1214/09-BJPS109.

Eling, M. 2012. Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models?. *Insurance: Mathematics and Economics* 51(2):239–48.

Eugene, N., C. Lee, and F. Famoye. 2002. Beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods* 31(4):497–512. doi:10.1081/STA-120003130.

Farias, R. B., M. H. Montoril, and J. A. Andrade. 2016. Bayesian inference for extreme quantiles of heavy tailed distributions. *Statistics & Probability Letters* 113:103–07. doi:10.1016/j.spl.2016.02.020.

Frees, E. W., and E. A. Valdez. 1998. Understanding relationships using copulas. *North American Actuarial Journal* 2(1):1–25. doi:10.1080/10920277.1998.10595667.

Gupta, A. K. 2003. Multivariate skew t-distribution. *Statistics: A Journal of Theoretical and Applied Statistics* 37(4):359–63. doi:10.1080/715019247.

Gupta, A. K., and S. Nadarajah. 2004. *Handbook of Beta Distribution and its Applications*. Boca Raton, FL: CRC press.

Hasan, A. M. (2013). *A Study of Non-Central Skew t Distributions and Their Applications in Data Analysis and Change Point Detection*. PhD thesis, Bowling Green State University.

Hosking, J. R. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* 52(1):105–124.

Huang, W.-J., and Y.-H. Chen. 2006. Quadratic forms of multivariate skew normalsymmetric distributions. *Statistics & Probability Letters* 76(9):871–79. doi:10.1016/j.spl.2005.10.018.

Jones, M. 2004. Families of distributions arising from distributions of order statistics. *Test* 13(1):1–43. doi:10.1007/BF02602999.

Mameli, V., and M. Musio. 2013. A generalization of the skew-normal distribution: The beta skew-normal. *Communications in Statistics-Theory and Methods* 42(12):2229–44. doi:10.1080/03610926.2011.607530.

McNeil, A. J. 1997. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* 27(01):117–37. doi:10.2143/AST.27.1.563210.

Rêgo, G., and S. Nadarajah. 2011. On some properties of beta normal distribution. *Communications in Statistics Theory and Methods* 41(20):3722–38. doi:10.1080/03610926.2011.568156.

Resnick, S. I. 1997. Discussion of the danish data on large fire insurance losses. *ASTIN Bulletin* 27 (01):139–51. doi:10.2143/AST.27.1.563211.

Rohatgi, V. K., and A. M. Ehsanes Saleh. 1988. A class of distributions connected to order statistics with nonintegral sample size. *Communications in Statistics-Theory and Methods* 17(6):2005–12. doi:10.1080/03610928808829728.

Shafiei, S., and M. Doostparast. 2014. Balakrishnan skew-t distribution and associated statistical characteristics. *Communications in Statistics-Theory and Methods* 43(19):4109–22. doi:10.1080/03610926.2012.701697.

Silva, G. O., E. M. Ortega, and G. M. Cordeiro. 2010. The beta modified Weibull distribution. *Lifetime Data Analysis* 16(3):409–30. doi:10.1007/s10985-010-9161-1.

Thukral, A. K. 2014. Factorials of real negative and imaginary numbers-a new perspective. *SpringerPlus* 3(1):658. doi:10.1186/2193-1801-3-658.

Appendix

Proof of Proposition 3.2

Proof. (a) For fixed x, λ, r, b, and as $a \to \infty$

$$\begin{aligned}\lim_{a\to\infty} g_F(x;a,b,\lambda,r) &= \lim_{a\to\infty} \frac{f(x;\lambda,r)}{B(a,b)} F(x;\lambda,r)^{a-1}(1-F(x;\lambda,r))^{b-1} \\ &= f(x;\lambda,r)(1-F(x;\lambda,r))^{b-1} \lim_{a\to\infty} \frac{\Gamma(a+b)}{\Gamma(a)} F(x;\lambda,r)^{a-1} = 0.\end{aligned}$$

Similarly, For fixed x, λ, r, a, and as $b \to \infty$

$$\lim_{b\to\infty} g_F(x;a,b,\lambda,r) = 0.$$

This completes the proof of (a).

(b) Recall that the definition of skew t random variable X with *pdf* $f(x;\lambda,r)$ by Azzalini and Capitanio (2003) is constructed as a scale mixture of skew normal distribution using the following transformation:

$$X \overset{D}{=} \frac{Y}{\sqrt{\frac{Z}{r}}},$$

where $Y \sim SN(\lambda)$ and $Z \sim \chi^2_r$ are independent random variables. By the strong law of large number (SLLN),

$$\lim_{r\to\infty} \sqrt{\frac{Z}{r}} = \sqrt{\lim_{r\to\infty} \frac{Z}{r}} = 1.$$

Thus,

$$\lim_{r\to\infty} X \overset{D}{=} \lim_{r\to\infty} \frac{Y}{\sqrt{\frac{Z}{r}}} \sim SN(\lambda).$$

Also,

$$\lim_{r\to\infty} f(x;\lambda,r) \overset{D}{=} \phi(x;\lambda),$$

where $\phi(.;\lambda)$ and $\Phi(.;\lambda)$ are the *pdf* and *cdf* of the skew normal distribution, respectively. Thus, when $X \sim BST(a,b,\lambda,r)$ with *pdf* $g_F(x;a,b,\lambda,r)$ defined in Eq. (8), for fixed x,a,b,λ and as $r \to \infty$,

$$\lim_{r\to\infty} g_F(x;a,b,\lambda,r) \stackrel{D}{=} \frac{1}{B(a,b)} \phi(x;\lambda)\Phi(x;\lambda)^{a-1}(1-\Phi(x;\lambda))^{b-1}.$$

That is,

$$X \sim BSN(a,b,\lambda).$$

This completes the proof of (b).

(c) Let X be a skew t distributed random variable. We have

$$\begin{aligned}\lim_{\lambda\to\infty} \phi(y;\lambda) &= \lim_{\lambda\to\infty} 2\phi(y)\Phi(\lambda y)\\ &= 2\phi(y)\lim_{\lambda\to\infty}\Phi(\lambda y)\\ &= 2\phi(y)I_{[0,\infty]}(y),\end{aligned}$$

where $\phi(.)$ and $\Phi(.)$ are the *pdf* and *cdf* of the normal distribution, respectively. This indicates $\lim_{\lambda\to\infty} Y \stackrel{D}{=} |W|$, where $W \sim N(0,1)$. Then,

$$\lim_{\lambda\to\infty} X \stackrel{D}{=} \lim_{\lambda\to\infty} \frac{Y}{\sqrt{\frac{Z}{r}}} \stackrel{D}{=} \frac{\lim_{\lambda\to\infty} Y}{\sqrt{\frac{Z}{r}}} \stackrel{D}{=} \frac{|W|}{\sqrt{\frac{Z}{r}}} \stackrel{D}{=} |t_r|.$$

Thus, when $X \sim BST(a,b,\lambda,r)$ with *pdf* $g_F(x;a,b,\lambda,r)$ defined in Eq. (8), for fixed x,a,b,r, and as $\lambda \to \infty$,

$$\lim_{\lambda\to\infty} g_F(x;a,b,\lambda,r) \stackrel{D}{=} \frac{1}{B(a,b)} h(x;r)H(x;r)^a(1-H(x;r))^{b-1},$$

where $h(x;r)$ and $H(x;r)$ are the *pdf* and *cdf* of the half t distribution, respectively, with the degrees of freedom r. This completes the proof of (c). □

Proof of Proposition 3.3

Proof. This follows directly from the definition of the probability density function of order statistics. □

Proof of Proposition 3.4

Proof. (a)

$$P(Y_{1:n} \geq X) = \int_{\Re} \int_x^{\infty} g_{Y_{1:n}}(y_{1:n};a,d,\lambda,r) g_X(x;a,b,\lambda,r) dy_{1:n} dx.$$

$$\begin{aligned}
\int_x^\infty g_{Y_{1:n}}(y_{1:n};\lambda,r,a,d)dy_{1:n} &= \int_x^\infty nf(y_{1:n};\lambda,r)F(y_{1:n};\lambda,r)^{a-1}(1-F(y_{1:n};\lambda,r))^{d-1}dy_{1:n} \\
&= \int_{F(x;\lambda,r)}^\infty (1-F(y_{1:n};\lambda,r))^{d-1}nF(y_{1:n};\lambda,r)^{a-1}dF(y_{1:n};\lambda,r) \\
&= n\int_{F(x;\lambda,r)}^1 (1-s)^{n-1}ds = n\tfrac{(1-s)^n}{n}|_{F(x;\lambda,r)}^1 = (1-F(x;\lambda,r))^n,
\end{aligned}$$

where $s = F(y_{1:n};\lambda,r)$. Thus,

$$\begin{aligned}
P(Y_{1:n} \geq X) &= \int_\Re \tfrac{f(x;\lambda,r)}{B(a,b)}F(x;\lambda,r)^{a-1}(1-F(x;\lambda,r))^{n+b-1}dx \\
&= \tfrac{1}{B(a,b)}\int_0^1 F(x;\lambda,r)^{a-1}(1-F(x;\lambda,r))^{n+b-1}dF(x;\lambda,r) \\
&= \tfrac{1}{B(a,b)}\int_0^1 t^{a-1}(1-t)^{n+b-1}dt = \tfrac{B(a,n+b)}{B(a,b)},
\end{aligned}$$

where $t = F(x;\lambda,r)$. Then,

$$\begin{aligned}
P(W \leq w) &= \frac{\int_{-\infty}^w \tfrac{1}{B(a,b)}f(x;\lambda,r)F(x;\lambda,r)^{a-1}(1-F(x;\lambda,r))^{n+b-1}dx}{\tfrac{B(a,n+b)}{B(a,b)}} \\
&= \frac{1}{B(a,n+b)}f(w;\lambda,r)F(w;\lambda,r)^{a-1}(1-F(w;\lambda,r))^{n+b-1},
\end{aligned}$$

which is the *pdf* of $W\sim BST(a,n+b,\lambda,r)$. □

(b) Similar to the proof of (a).

Proof of Theorem 4.1

Proof.

$$\begin{aligned}
E(X^n) &= \int_\Re \tfrac{x^n}{B(a,b)\sigma}f(\tfrac{x-\mu}{\sigma};\lambda,r)F(\tfrac{x-\mu}{\sigma};\lambda,r)^{a-1}(1-F(\tfrac{x-\mu}{\sigma};\lambda,r))^{b-1}dx \\
&= \int_\Re \tfrac{x^n}{B(a,b)\sigma}\sum_{j=0}^\infty(-1)^j\binom{b-1}{j}f(\tfrac{x-\mu}{\sigma};\lambda,r)F(\tfrac{x-\mu}{\sigma};\lambda,r)^{a+j-1}dx \\
&= \tfrac{1}{B(a,b)\sigma}\sum_{j=0}^\infty(-1)^j\binom{b-1}{j}\int_\Re x^n f(\tfrac{x-\mu}{\sigma};\lambda,r)F(\tfrac{x-\mu}{\sigma};\lambda,r)^{a+j-1}dx.
\end{aligned}$$

Substituting $z = \frac{x-\mu}{\sigma}$ and using the binomial expansion, we obtain

$$\begin{aligned}
\int_\Re x^n f(\tfrac{x-\mu}{\sigma};\lambda,r)F(\tfrac{x-\mu}{\sigma};\lambda,r)^{a+j-1}dx &= \int_\Re(\mu+\sigma z)^n f(z;\lambda,r)F(z;\lambda,r)^{a+j-1}dz \\
&= \int_\Re\sum_{i=0}^n\binom{n}{i}(z\sigma)^{n-i}\mu^i f(z;\lambda,r)F(z;\lambda,r)^{a+j-1}dz \\
&= \sigma^n\sum_{i=0}^n\binom{n}{i}(\tfrac{\mu}{\sigma})^i\int_\Re z^{n-i}f(z;\lambda,r)F(z;\lambda,r)^{a+j-1}dz.
\end{aligned}$$

Applying integration by part for the quantity $\int_\Re z^{n-i}f(z;\lambda,r)F(z;\lambda,r)^{a+j-1}dz$, we let

$$u = F(z;\lambda,r)^{a+j-1},$$

and

$$dv = z^{n-i}f(z;\lambda,r)dz.$$

Then

$$du = (a+j-1)f(z;\lambda,r)F(z;\lambda,r)^{a+j-2},$$

and

$$v = \int_{\Re} z^{n-i}f(z;\lambda,r)dz = E(Y^{n-i}),$$

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

$$\begin{aligned}\int_{\Re} y^{n-i}f(y;\lambda,r)F(y;\lambda,r)^{a+j-1}dy \quad &= E_Y(Y^{n-i})F(y;\lambda,r)^{a+j-1} - E_Y(Y^{n-i})(a+j-1) \\ &\int_{\Re} f(y;\lambda,r)F(x;\lambda,r)^{a+j-1}dy, \\ &= E_Y(Y^{n-i})[F(y;\lambda,r)^{a+j-1} - (a+j-1) \\ &\int_{\Re} F(y;\lambda,r)^{a+j-1}dF(y;\lambda,r)], \\ &= E_Y(Y^{n-i})[F(y;\lambda,r)^{a+j-1} - (a+j-1)\tfrac{1}{(a+j-1)}], \\ &= E_Y(Y^{n-i})[F(y;\lambda,r)^{a+j-1} - 1].\end{aligned}$$

Thus,

$$\begin{aligned}E(X^n) \quad &= \tfrac{1}{B(a,b)}\sum_{j=0}^{\infty}(-1)^j\binom{b-1}{j}\sigma^n\sum_{i=0}^{n}\binom{n}{i}(\tfrac{\mu}{\sigma})^i E_Y(Y^{n-i})[F(y;\lambda,r)^{a+j-1} - 1] \\ &= \tfrac{\sigma^n}{B(a,b)}\sum_{j=0}^{\infty}\sum_{i=0}^{n}(-1)^j\binom{b-1}{j}\binom{n}{i}(\tfrac{\mu}{\sigma})^i E_Y(Y^{n-i})[F(y;\lambda,r)^{a+j-1} - 1].\end{aligned}$$

□

Proof of Theorem 4.2

Proof. By applying the integration by part, we have

$$E(X^n) = \frac{1}{B(a,b)}\int_{\Re} x^n f(x;\lambda,r)F(x;\lambda,r)^{a-1}(1-F(x;\lambda,r))^{b-1}dx.$$

Let

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

and

$$dv = x^n f(x;\lambda,r)dx.$$

Then

$$\begin{aligned} du = \quad & (a-1)f(x;\lambda,r)F(x;\lambda,r)^{a-2}(1-F(x;\lambda,r))^{b-1}dx+ \\ & (b-1)(1-F(x;\lambda,r))^{b-2}(-f(x;\lambda,r))F(x;\lambda,r)^{a-1}dx, \end{aligned}$$

and

$$v=\int_{\Re} x^n f(x;\lambda,r)dx.$$

Note that v is the n^{th} moment of a $st_r(\lambda)$ random variable. Then,

$$\begin{aligned} E(X^n) \quad & = \tfrac{1}{B(a,b)}[vF(X;\lambda,r)^{a-1}(1-F(X;\lambda,r))^{b-1}|_{-\infty}^{\infty} \\ & -\int_{\Re} v(a-1)f(x;\lambda,r)F(x;\lambda,r)^{a-2}(1-F(x;\lambda,r))^{b-1}dx \\ & +\int_{\Re} v(b-1)f(x;\lambda,r)F(x;\lambda,r)^{a-1}(1-F(x;\lambda,r))^{b-2}dx] \\ & = \tfrac{v}{B(a,b)}[\int_{-\infty}^{\infty}(b-1)f(x;\lambda,r)F(x;\lambda,r)^{a-1}(1-F(x;\lambda,r))^{b-2}dx \\ & -\int_{\Re}(a-1)f(x;\lambda,r)F(x;\lambda,r)^{a-2}(1-F(x;\lambda,r))^{b-1}dx]. \quad (*) \end{aligned}$$

Note that

$$\begin{aligned} & \int_{\Re}(b-1)f(x;\lambda,r)F(x;\lambda,r)^{a-1}(1-F(x;\lambda,r))^{b-2}dx \\ & = \int_{\Re}(b-1)f(x;\lambda,r)F(x;\lambda,r)^{a-1}\sum_{i=0}^{b-2}(-1)^i b-2\,{}_i F(x;\lambda,r)^i dx \\ & = \sum_{i=0}^{b-2}(-1)^i\binom{b-2}{i}(b-1)\int_{\Re} f(x;\lambda,r)F(x;\lambda,r)^{a+i-1}dx \\ & = \sum_{i=0}^{b-2}(-1)^i\binom{b-2}{i}(b-1)\tfrac{F(x;\lambda,r)^{a+i}}{a+i}|_{-\infty}^{\infty} \\ & = \sum_{i=0}^{b-2}\tfrac{(-1)^i}{B(i+1,b-i-1)(a+i)}. \quad (**) \end{aligned}$$

Similarly,

$$\int_{\Re}(a-1)f(x;\lambda,r)F(x;\lambda,r)^{a-2}(1-F(x;\lambda,r))^{b-1}dx=\sum_{i=0}^{b-1}\frac{(-1)^i\binom{b-1}{i}(a-1)}{a+i-2}. \quad (***)$$

Substituting Eqs. (**) and (***) into Eq. (*), we obtain

$$\begin{aligned} E(X^n) \quad & = \tfrac{v}{B(a,b)}[\sum_{i=0}^{b-2}\tfrac{(-1)^i}{B(i+1,b-i-1)(a+i)}-\sum_{i=0}^{b-1}\tfrac{(-1)^i\binom{b-1}{i}(a-1)}{a+i-2}] \\ & = \tfrac{v}{B(a,b)}[\sum_{i=0}^{b-2}\tfrac{(-1)^i}{B(i+1,b-i-1)}[\tfrac{1}{a+i}-\tfrac{(a-1)}{(a+i-1)(b-i-1)}]- \\ & (-1)^{b-1}\tfrac{(a-1)}{a+b-3}]. \end{aligned}$$

Proof of Theorem 5.1

Proof. We use expression (13) of the incomplete beta function.

Let $h(x) = \int_0^1 z^{-b}(1-z)^{a+b-1}(1-zF(x;\lambda,r))^{-a}dz$, so we have

$$
\begin{aligned}
g_{i:n}(x) &= \tfrac{n!}{(i-1)!(n-i)!} G_F(x;a,b,\lambda,r)^{i-1}(1-G_F(x;a,b,\lambda,r))^{n-i} g_F(x;a,b,\lambda,r)\\
&= \tfrac{n!}{(i-1)!(n-i)!} g_F(x;a,b,\lambda,r) G_F(x;a,b,\lambda,r)^{i-1} \sum_{k=0}^{n-i} (-1)^k n-i_{\,k}\, G_F(x;a,b,\lambda,r)^k\\
&= \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} i \binom{n}{i} g_F(x;a,b,\lambda,r) G_F(x;a,b,\lambda,r)^{k+i-1}\\
&= \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} i \binom{n}{i} g_F(x;a,b,\lambda,r) \left\{\tfrac{F(x;\lambda,r)^a h(x)}{aB(a,b)B(1-b,a+b)}\right\}^{k+i-1}\\
&= \sum_{k=0}^{n-i} \binom{n-i}{k} i \binom{n}{i} f(x;\lambda,r) F(x;\lambda,r)^{a(k+i)-1}[1-F(x;\lambda,r)]^{b-1} \left\{\tfrac{h(x)}{aB(a,b)B(1-b,a+b)}\right\}^{k+i-1}\\
&= \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} i \binom{n}{i} g_F(a(k+i),b,\lambda,r) \tfrac{B(a(k+i),b)}{B(a,b)} \left\{\tfrac{h(x)}{aB(a,b)B(1-b,a+b)}\right\}^{k+i-1}\\
&= \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} i \binom{n}{i} g_F(a(k+i),b,\lambda,r) \tfrac{B(a(k+i),b)}{B(a,b)} \tfrac{h(x)^*}{[aB(a,b)B(1-b,a+b)]^{k+i-1}}.
\end{aligned}
$$

where $h(x)^* = h(x)^{k+i-1}$

Proof of Theorem 5.2

Proof. By the definition of order statistics, we have

$$
\begin{aligned}
g_{i:n}(x) &= \tfrac{n!}{(i-1)!(n-i)!} G_F(x;a,b,\lambda,r)^{i-1}(1-G_F(x;a,b,\lambda,r))^{n-i} g_F(x;a,b,\lambda,r)\\
&= \tfrac{n!}{(i-1)!(n-i)!} g_F(x;a,b,\lambda,r) G_F(x;a,b,\lambda,r)^{i-1} \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} G_F(x;a,b,\lambda,r)^k\\
&= \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} i \binom{n}{i} g_F(x;a,b,\lambda,r) G_F(x;a,b,\lambda,r)^{k+i-1},
\end{aligned}
$$

If $b \in \mathbb{Z}$, by Eq. (11) we obtain

$$
\begin{aligned}
g_{i:n}(x) &= \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} i \binom{n}{i} g_F(x;a,b,\lambda,r)\\
&\quad \left\{ \sum_{j=0}^{b-1} \binom{a+b-1}{j} F(x;\lambda,r)^{a+b-j-1}(1-F(x;\lambda,r))^j \right\}^{k+i-1}\\
&= \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} i \binom{n}{i} g_F(x;a,b,\lambda,r)\\
&\quad \left\{ \tfrac{1-F(x;\lambda,r)}{f(x;\lambda,r)} \sum_{j=0}^{b-1} \tfrac{f(x;\lambda,r)F(x;\lambda,r)^{a+b-j-1}(1-F(x;\lambda,r))^{j-1}}{B(a+b-j,j)} \right\}^{k+i-1}\\
&= \sum_{k=0}^{n-i} (-1)^k i \binom{n}{i} \binom{n-i}{k} g_F(x;a,b,\lambda,r) \left\{ \tfrac{1-F(x;\lambda,r)}{f(x;\lambda,r)} \sum_{j=0}^{b-1} g_F(x;(a+b-j),j,\lambda,r) \right\}^{k+i-1}.
\end{aligned}
$$

□