RESEARCH　　　　　　　　　　　　　　　　　　　**Open Access**

The multivariate slash and skew-slash student *t* distributions

Fei Tan[1*], Yuanyuan Tang[2] and Hanxiang Peng[1]

*Correspondence:
ftan@math.iupui.edu
[1] Department of Mathematical Sciences, Indiana University Purdue University Indianapolis, N. Blackford Street, IN 46202, Indianapolis, USA
Full list of author information is available at the end of the article

Abstract

In this article, we introduce the multivariate slash and skew-slash *t* distributions which provide alternative choices in simulating and fitting skewed and heavy tailed data. We study their relationships with other distributions and give the densities, stochastic representations, moments, marginal distributions, distributions of linear combinations and characteristic functions of the random vectors obeying these distributions. We characterize the skew *t*, the skew-slash normal and the skew-slash *t* distributions using both the hidden truncation or selective sampling model and the order statistics of the components of a bivariate normal or *t* variable. Density curves and contour plots are drawn to illustrate the skewness and tail behaviors. Maximum likelihood and Bayesian estimation of the parameters are discussed. The proposed distributions are compared with the skew-slash normal through simulations and applied to fit two real datasets. Our results indicated that the proposed skew-slash *t* fitting outperformed the skew-slash normal fitting and is a competitive candidate distribution in analyzing skewed and heavy tailed data.

Mathematics Subject Classification Primary 62E10; Secondary 62P10

Keywords: Heavy tail; Skew distribution; Skew-slash distribution; Slash distribution; Student *t* distribution

1 Introduction

Skewed and heavy tailed data occur frequently in real life and pose challenges to our usual way of thinking. Examples of such data include household incomes, loss data such as crop loss claims and hospital discharge bills, and files transferred through the Internet to name a few. Candidate distributions for simulating and fitting such data are not abundant. One can't simply take the normal or the *t* distributions or as such as substitutes. Even though Cauchy distribution can be used to simulate and fit such data, its sharp central peak and the fact that its first moment does not exist narrow its applications. Thus additional distributions are needed to study such skewed and heavy tailed data.

Kafadar (1988) introduced the univariate normal slash distribution as the resulting distribution of the ratio of a standard normal random variable (rv) and an independent uniform rv (hereafter referred to the distribution as the slash normal). Generalizing the standard normal by introducing a tail parameter, the slash normal has heavier tails than the standard normal, hence it could be used to simulate and fit heavy tailed data. Wang

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 2 of 22

and Genton (2006) generalized the univariate slash normal to the multivariate slash normal and investigated its properties. They also defined the multivariate skew-slash normal as the resulting distribution of the ratio of a skewed normal rv and an independent uniform rv (hereafter referred to the distribution as the skew-slash normal). They applied it to fit two real datasets.

In this article, we introduce the *slash (student) t distribution* and the *skew-slash (student) t distribution*. They could be used to simulate and fit skewed and heavy tailed data. The slash t distribution generalizes the slash normal distribution of Kafadar (1988) and the multivaraiate slash normal distribution of Wang and Genton (2006), and the skew-slash t distribution generalizes the skew-slash normal distribution of the latter two authors. In the skew-slash t there is one parameter to regulate the skewness of the distribution and another parameter to control the tail behavior. By setting the skewness parameter to zero, the skew-slash t reduces to the slash t. By letting the tail parameter to be infinity, the skew-slash t simplifies to the skew t of Azzalini and Capitanio (2003). As both the slash and skew t take the t and hence the normal as their special cases, so does the skew-slash t. To fit data, one can start with the skew-slash t. If the fitted value of the degrees of freedom is very large, then one takes the simpler skew-slash normal model. This idea of course can be used to perform the hypothesis testing of a skew-slash normal sub-model against a skew-slash t model. We have derived the formulas for the densities, moments, marginal distributions and linear combinations of these distributions. Thus it could be expected that they can be used to analyze skewed and heavy tailed data.

Compared to the slash and skew-slash normal distributions, an additional parameter, the degrees of freedom, is included in the slash and skew-slash t distributions. This parameter gives the latter distributions more flexibility in fitting data than the former. Even though there is a tail parameter in both slash and skew-slash normal and slash and skew-slash t distributions, the degrees of freedom in the t distribution may lend an additional hand to model heavy (fat) tails and joints with the tail parameter to better fit data. This could be used to explain why the skew-slash fitting to the real GAD data in our application outperformed the skew-slash normal fitting. See Figure 1 where the skew-slash t fitting was able to better capture the peak of the histogram, and Table 1 where the AIC values indicated that the skew-slash t fitting was better than that of the skew-slash normal fitting. In fact, one observes that the standard error (SE) of the MLE of the tail parameter q in the table is very big (83.194) for the skew-slash normal fitting and much smaller (2.19) for the skew-slash t fitting. Noticing also the fact that the estimate of the degrees of freedom r is reasonable with a small SE, we would comment that the joint endeavor of q and r had higher fitting capability than that of a single q. Our simulation results in Section 5.1 and in particular in Table 2 also exhibited the superior performance of the proposed skew-slash t distribution to the skew-slash normal.

Azzalini and Dalla Valle (1996) introduced the multivariate skew normal distribution that extends the normal distribution with an additional skewness parameter. It provides an alternative modeling distribution to skewed data that are often observed in many areas such as economics, computer science and life sciences. Many authors have investigated skew t distributions, see e.g. Azzalini and Dalla Valle (1996), Gupta (2003), and Sahu et al. (2003). Azzalini and Capitanio (2003) proposed the multivariate skew t distribution by allowing a skewness parameter in a multivariate t distribution.

Figure 1 Histogram and fitted density curves.
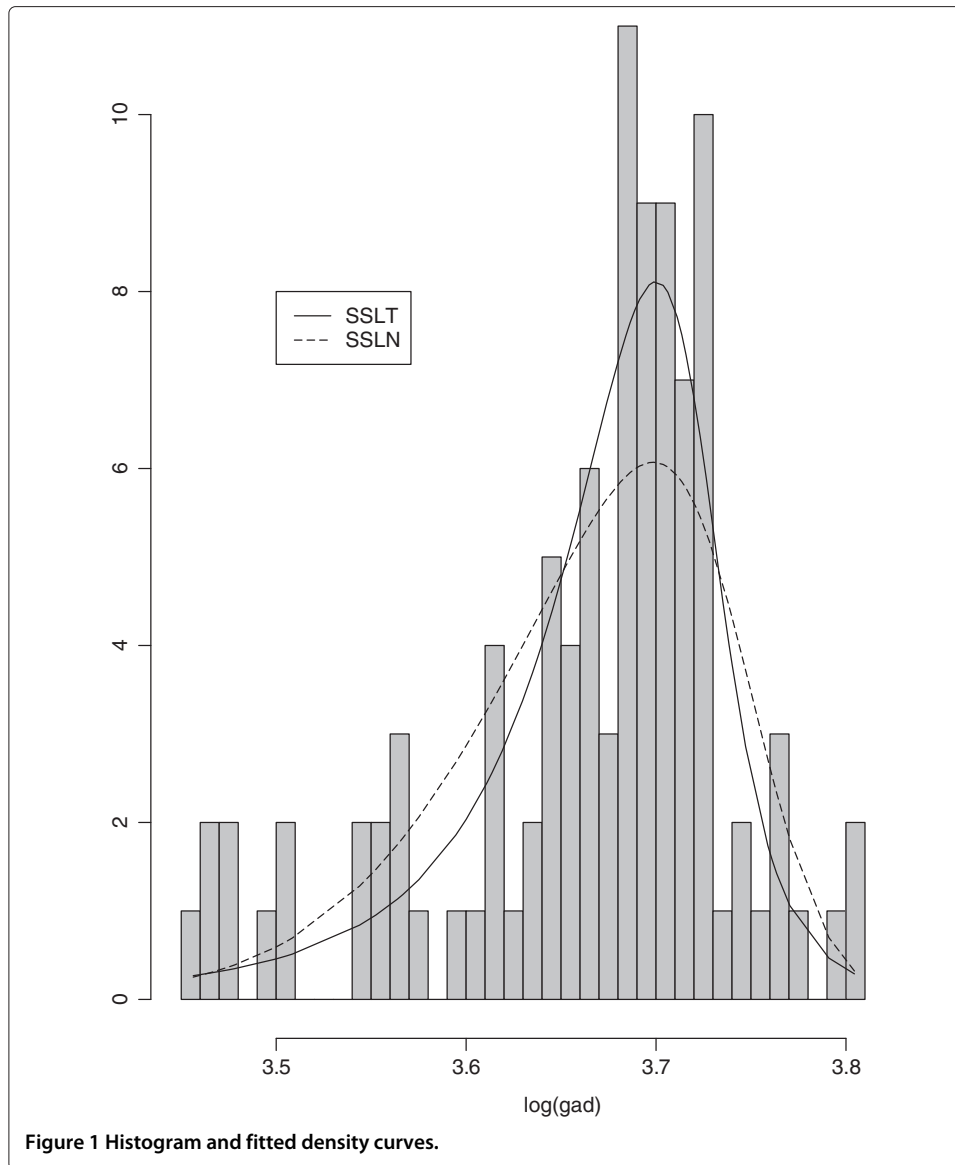
It is our belief that the proposed slash and skew-slash t distributions throw some additional light on this theory and contribute to the family of candidate distributions for modeling and simulating skewed and heavy tailed data.

The article is organized as follows. In Section 2, we introduce the multivariate slash t distribution, study its relationships with other distributions and derive the density

Table 1 The skew-slash normal and t fitting to the GAD data

	SSLT		SSLN	
	MLE	**SE**	**MLE**	**SE**
AIC	-242.318		-240.687	
μ	3.726	0.0070	3.748	0.036
σ^2	0.002	0.0004	0.009	0.038
λ	-2.056	0.6190	-3.199	3.204
q	3.812	2.1990	7.456	83.194
r	3.044	1.4080		

SSLN=Skew-Slash Normal and SSLT=Skew-Slash t.

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 4 of 22

Table 2 Skew-slash *t* and skew-slash normal comparison

	True	Estimated		True	Estimated	
	SSLT	SSLT	SSLN	SSLN	SSLT	SSLN
AIC		1524.939	1543.771		1416.485	1425.878
μ_1	0	-0.160	-0.467	0	-0.309	-0.690
μ_2	0	0.032	-0.225	0	0.025	-0.212
σ_{11}^2	1	0.827	0.809	1	0.587	0.651
σ_{12}	0	0.122	0.230	0	0.130	0.304
σ_{22}^2	1	1.154	1.527	1	1.019	1.314
λ_1	-5	-2.830	-1.581	-5	-1.989	-0.860
λ_2	3	1.688	1.422	3	1.363	1.127
q	5	6.383	4.553	5	6.697	5.151
r	8	7.610			7.506	

The comparison of the average AIC and MLE's of the parameters between the skew-slash *t* and skew-slash normal based on sample size 250 and repetitions 200. SSLT=Skew-Slash *t* and SSLN=Skew-Slash Normal. A smaller AIC indicates a better fit.

function. We investigate its tractable properties such as heavy tail behavior and closeness of marginal distributions and linear combinations. We give the stochastic representations, moments and characteristic function. We close this section with an example which graphically displays the densities. In Section 3, we define the skew-slash *t* distribution, derive the densities and characteristic functions, and give the moments and distributions of linear combinations of these distributions. We first define the standard skew-slash distribution in Subsection 3.1 and study their relationships with other distributions. In subsection 3.2, we characterize the skew *t*, skew-slash normal and skew-slash *t* distributions using hidden truncation or selective sampling model and the order statistics of the components of a bivariate normal or *t* variable. In Subsection 3.2, we define the general multivariate skew-slash distribution. An example is presented to illustrate the densities of the proposed distributions. Section 4 covers parameter estimation and statistical inference. Here we briefly discuss the maximum likelihood and Byesian approaches. Section 5 is devoted to simulations as well as applications of the proposed skew-slash *t* distribution to fit two real datasets. Finally, some concluding remarks are given in Section 6.

2   The multivariate slash *t* distribution

In this section, we define the multivarate slash *t* distribution, derive the density and study its tail behaviors and relationships with other distributions. We give the stochastic representations, moments, and characteristic function and discuss marginal distributions and linear combinations. We close this section with an example.

Let us first recall the multivariate *t* distribution. There are several variants of the definitions in the literature and we will adopt the following one. Details can be found in e.g. Kotz and Nadarajah (2004) or Johnson and Kotz (1972, p. 134). A continuous *k*-variate random vector $\mathbf{T}$ has a *t* distribution with degrees of freedom r, mean vector $\mathbf{m}$, and correlation matrix $\mathbf{R}$ (covariance matrix Σ), written $\mathbf{T} \sim t_k(r, \mathbf{m}, \mathbf{R})$, if it has the probability density function (pdf) given by

$$t_k(\mathbf{t}; r, \mathbf{m}, \mathbf{R}) = \frac{\Gamma\left(\frac{r+k}{2}\right)}{(r\pi)^{k/2}\Gamma\left(\frac{r}{2}\right)|\mathbf{R}|^{1/2}} \left(1 + \frac{(\mathbf{t} - \mathbf{m})^\top \mathbf{R}^{-1}(\mathbf{t} - \mathbf{m})}{r}\right)^{-(r+k)/2}, \quad \mathbf{t} \in \mathbb{R}^k,$$

where $|\mathbf{M}|$ denotes the determinant of a square matrix $\mathbf{M}$. If $\mathbf{m} = 0$ and $\mathbf{R} = \mathbf{I}_k$ where $\mathbf{I}_k$ denotes the $k \times k$ indentity matrix, it is referred to as the standard k-variate t distribution and denoted by $t_k(r)$. We now introduce the k-variate slash t distribution. Write $U \sim U(0, 1)$ for the rv uniformly distributed over $(0, 1)$.

Definition 1. A k-variate continuous random vector $\mathbf{X}$ is said to have a *slash t distribution* with tail parameter $q > 0$, degrees of freedom r, $\mathbf{m} \in \mathbb{R}^k$, and matrix $\mathbf{R}$, written $\mathbf{X} \sim SLT_k(q, r, \mathbf{m}, \mathbf{R})$, if it can be expressed as the ratio of two independent rv's $\mathbf{T} \sim t_k(r, \mathbf{m}, \mathbf{R})$ and $U \sim U(0, 1)$ as follows:

$$\mathbf{X} = \mathbf{T}/U^{1/q}.$$

When $\mathbf{m} = 0$ and $\mathbf{R} = \mathbf{I}_k$, it is referred to as the *standard (k-variate) slash t* and denoted by $SLT_k(q, r)$. It can be easily seen that the k-variate slash t distribution generalizes the k-variate t distribution as stated below.

Remark 1. *The limiting distribution of the slash t distribution $SLT_k(q, r)$, as $q \to \infty$, is the student t distribution $t_k(r)$.*

Let us now derive the density of the slash t distribution. Note that the joint density of $\mathbf{T}$ and U is

$$t_k(\mathbf{t}; r, m, R)\mathbf{1}_{[0,1]}(u), \quad \mathbf{t} \in \mathbb{R}^k, u \in [0, 1].$$

For the substitution $v = u^{1/q}, \mathbf{x} = \mathbf{t}/u^{1/q} = \mathbf{t}/v$, the Jacobian determinant is qv^{k+q-1}. Hence the joint density of $(\mathbf{X}, V)$ is given by

$$qv^{k+q-1}t_k(v\mathbf{x}; r, \mathbf{m}, \mathbf{R})\mathbf{1}_{[0,1]}(v), \quad \mathbf{x} \in \mathbb{R}^k, v \in [0, 1].$$

Integrating v out yields the density $f_k(\mathbf{x}; q, r, \mathbf{m}, \mathbf{R})$ of $\mathbf{X}$ as follows:

$$f_k(\mathbf{x}; q, r, \mathbf{m}, \mathbf{R}) = \int_0^1 qv^{k+q-1}t_k(\mathbf{x}v; r, \mathbf{m}, \mathbf{R})\, dv, \quad \mathbf{x} \in \mathbb{R}^k. \tag{2.1}$$

From this density it immediately follows that the standard k-variate slash t distribution $SLT_k(q, r)$ is symmetric about 0 as the standard k-variate t is so.

The heavy-tail behavior The cumulative distribution function (cdf) of the k-variate slash t is given by

$$F_k(\mathbf{x}; q, r, \mathbf{m}, \mathbf{R}) = \int_{-\infty}^{\mathbf{x}} f_k(\mathbf{y}; q, r, \mathbf{m}, \mathbf{R})\, d\mathbf{y} = \int_0^1 qv^{k+q-2}H_k(\mathbf{x}v; r, \mathbf{m}, \mathbf{R})\, dv,$$

for $\mathbf{x} \in \mathbb{R}^k$, where H_k is the cdf of the k-variate $t_k(r, \mathbf{m}, \mathbf{R})$. Denote by $\bar{H}_k = 1 - H_k$ the survival function. The survival function of the k-variate slash t is then given by

$$\bar{F}_k(\mathbf{x}; q, r, \mathbf{m}, \mathbf{R}) = \frac{k-1}{k+q-1} + \int_0^1 qv^{k+q-2}\bar{H}_k(\mathbf{x}v; r, \mathbf{m}, \mathbf{R})\, dv, \ \mathbf{x} \in \mathbb{R}^k. \tag{2.2}$$

Let us focus on the standard univariate case and write $f = f_1, F = F_1, \bar{F} = 1 - F$ and $H = H_1, \bar{H} = 1 - H$. It is well known that the univariate t distribution $t_1(r, 0, 1)$ is a heavy tailed distribution with tail index r. In other words, the survival function decays at the rate of the power function r:

$$\bar{H}(t; r, 0, 1) \propto t^{-r}, \quad t \to \infty,$$

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 6 of 22

where $a(t) \propto b(t)$ if $\lim \sup_{t \to \infty} a(t)/b(t) < \infty$. Let c be the constant in the density of the k-variate t. Then by L'Hopital' Rule we have

$$\lim_{x \to \infty} \frac{\bar{H}(x)}{x^{-r}} = c \lim_{x \to \infty} \frac{\int_x^\infty (1 + t^2/r)^{-(r+1)/2} \, dt}{x^{-r}}$$
$$= c \lim_{x \to \infty} \frac{-(1 + x^2/r)^{-(r+1)/2} \, dt}{-rx^{-r-1}} = cr^{(r-1)/2} < \infty.$$

This shows the above rate holds. Similarly, one can show that if $q > r$ then

$$\bar{F}(x; q, r, 0, 1) \propto x^{-r}, \quad x \to \infty.$$

This manifests that the standard univariate slash t is also heavy tailed.

Further, by (2.2) and in view of $\bar{H}(xv; r, 0, 1) \geq \bar{H}(x; r, 0, 1)$ for $v \in [0, 1]$ one derives

$$\bar{F}(x; q, r, 0, 1) \geq (k-1)/(k+q-1) + \bar{H}(x; r, 0, 1) \int_0^1 qv^{k+q-2} \, dv$$
$$\geq (k-1)/(k+q-1) + q/(k+q-1)\bar{H}(x; r, 0, 1)$$
$$\geq \bar{H}(x; r, 0, 1), \quad x \geq 0.$$

This shows that the standard univariate slash t has heavier tails than the standard univariate t. In fact, the last inequality also holds for k-variate slash t for $\mathbf{x} = (x_1, \ldots, x_k)^\top$ with $x_i \geq 0, i = 1, \ldots, k$ as

$$\bar{F}_k(\mathbf{x}; q, r, \mathbf{m}, \mathbf{R}) = P(X_1 > x_1, \ldots, X_k > x_k)$$
$$= \int_0^1 P(T_1 > u^{1/q} x_1, \ldots, T_k > u^{1/q} x_k) \, du$$
$$\geq \int_0^1 P(T_1 > x_1, \ldots, T_k > x_k) \, du = \bar{H}_k(\mathbf{x}; r, \mathbf{m}, \mathbf{R}).$$

Stochastic representations Stochastic Representations not only reveal the relations with other distributions but are very useful, for instance, in calculating moments and random generation. We provide two stochastic representations for the slash t distribution based on the two stochastic representations of the multivariate t distribution. According to Kafadar (1988), a continuous random variable ξ has a slash normal distribution, written $\xi \sim SLN(q, \mathbf{0}, \Sigma)$, if it can be expressed as $\xi = \mathbf{Z}/U^{1/q}$ where $\mathbf{Z} \sim N_k(0, \Sigma)$ and $U \sim U(0, 1)$ are independent.

Note that if a rv $\mathbf{T}$ has a k-variate t distribution $t_k(r, \mathbf{m}, \mathbf{R})$, then it has the stochastic representation

$$\mathbf{T} = S^{-1}\mathbf{Z} + \mathbf{m},$$

where $\mathbf{Z} \sim N_k(0, \mathbf{R})$, rS^2 has the Chi-square distribution χ_r^2 with r degrees of freedom, and $\mathbf{Z}$ and S are independent. This can be easily verified. Let $\mathbf{X} \sim SLT_k(q, r, \mathbf{m}, \mathbf{R})$. Then from the definition of the k-variate slash t it immediately follows that

$$\mathbf{X} = S^{-1}\xi + \eta, \tag{2.3}$$

where $\xi \sim SLN_k(q, 0, \mathbf{R})$, $\eta = \mathbf{m}U^{-1/q}$, and both ξ and η are independent of S.

Using another stochastic representation of the k-variate t distribution from page 7 of Kotz and Nadarajah (2004), we obtain the second stochastic representation for the k-variate slash t rv $\mathbf{X} \sim SLT_k(q, r, \mathbf{m}, \mathbf{R})$ as follows:

$$\mathbf{X} = \mathbf{V}^{-1/2}\xi + \eta, \tag{2.4}$$

where $\boldsymbol{\xi} \sim SLN_k(q, 0, r\mathbf{I}_k)$, $\boldsymbol{\eta} = \mathbf{m}U^{-1/q}$, and $\boldsymbol{\xi}, \boldsymbol{\eta}$ are independent of $\mathbf{V}$. Here $\mathbf{V}^{-1/2}$ is the inverse of the symmetric square root $\mathbf{V}^{1/2}$ of $\mathbf{V}$, where $\mathbf{V}$ has a k-variate Wishart distribution with degrees of freedom $r + k - 1$ and covariance matrix $\mathbf{R}^{-1}$.

The moments Let us now calculate the mean vector and covariance matrix of the k-variate slash t. For $\mathbf{X} \sim SLT_k(q, r, \mathbf{m}, \mathbf{R})$ with $\mathbf{R} = (R_{ij})$, by the independence of $\mathbf{T}$ and U we have

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{X}) = \mathbb{E}(\mathbf{T}/U^{1/q}) = \mathbb{E}(\mathbf{T})\mathbb{E}(U^{-1/q}).$$

It is easy to calculate

$$\mathbb{E}(U^{-1/q}) = q/(q - 1), \quad q > 1, \tag{2.5}$$

and

$$\mathbb{E}(U^{-2/q}) = q/(q - 2), \quad q > 2. \tag{2.6}$$

For $\mathbf{T} \sim t_k(r, \mathbf{m}, \mathbf{R})$, from page 11 of Kotz and Nadarajah (2004) it follows

$$\mathbb{E}(\mathbf{T}) = \mathbf{m}, \quad \text{Var}(\mathbf{T}) = \mathbf{R}r/(r - 2), \quad r > 2. \tag{2.7}$$

Hence by (2.5) and the first equality of (2.7) one has

$$\boldsymbol{\mu} = \mathbf{m}q/(q - 1). \tag{2.8}$$

To calculate $\text{Var}(\mathbf{X})$ we use the formula

$$\text{Var}(\mathbf{X}) = \text{Var}(\mathbb{E}(\mathbf{X}|U)) + \mathbb{E}(\text{Var}(\mathbf{X}|U)). \tag{2.9}$$

It is easy to see for $r > 2$,

$$\text{Var}(\mathbf{X}|U) = \text{Var}(\mathbf{T})/U^{2/q} = \mathbf{R}r/((r - 2)U^{2/q}),$$

hence by (2.6) one has

$$\mathbb{E}(\text{Var}(\mathbf{X}|U)) = \mathbf{R}rq/((r - 2)(q - 2)), \quad q > 2, r > 2. \tag{2.10}$$

Also since $\mathbb{E}(\mathbf{X}|U) = \mathbb{E}(\mathbf{T})/U^{1/q} = \mathbf{m}/U^{1/q}$ it follows from (2.6) that

$$\mathbb{E}((\mathbb{E}(\mathbf{X}|U))^{\otimes 2}) = \mathbf{m}^{\otimes 2}q/(q - 2), \quad \mathbb{E}(\mathbf{X}^{\otimes 2}) = \mathbf{m}^{\otimes 2}q^2/(q - 1)^2, \quad q > 2,$$

where $\mathbf{M}^{\otimes 2} = \mathbf{M}\mathbf{M}^{\top}$. Hence for $q > 2$,

$$\text{Var}(\mathbb{E}(\mathbf{X}|U)) = \mathbf{m}^{\otimes 2}q/(q - 2) - \mathbf{m}^{\otimes 2}q^2/(q - 1)^2$$
$$= \mathbf{m}^{\otimes 2}q/((q - 1)^2(q - 2)), \quad q > 2.$$

This, (2.10) and (2.9) yield the variance-covariance matrix $\mathbf{X}$ as follows:

$$\text{Var}(\mathbf{X}) = \frac{rq\mathbf{R}}{(r - 2)(q - 2)} + \frac{q\mathbf{m}^{\otimes 2}}{(q - 1)^2(q - 2)}, \quad q > 2, r > 2. \tag{2.11}$$

In particular, if $\mathbf{m} = 0$ one has

$$\text{Var}(\mathbf{X}) = \frac{rq\mathbf{R}}{(r - 2)(q - 2)}, \quad q > 2, r > 2. \tag{2.12}$$

Hence the k-variate slash t has the same correlation matrix $\mathbf{R}$ as the k-variate t.

In the case of the standard t, there are convenient formulae for the moments. Given non-negative integers $p_1, \ldots, p_k$ such that $p = p_1 + \ldots + p_k < r/2$. If any of the $p_1, \ldots, p_k$ is odd, then

$$\mathbb{E}(T_1^{p_1} \cdots T_k^{p_k}) = 0.$$

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 8 of 22

If all of them are even and $r > p$, then

$$\mathbb{E}\left(T_1^{p_1} \cdots T_k^{p_k}\right) = \frac{r^{p/2} \Pi_{j=1}^k \left[1 \cdot 3 \cdot 5 \cdots (2p_j - 1)\right]}{(r-2)(r-4) \cdots (r-p)} := c. \tag{2.13}$$

For details, see Kotz and Nadarajah (2004). Based on these formulae we have the following.

Theorem 1. *Let* $\mathbf{X} = (X_1, \ldots, X_k)^\top \sim SLT_k(q, r)$. *Assume* $p_1, \ldots, p_k$ *are nonnegative integers such that* $p = p_1 + \ldots + p_k < r/2$.

1) Suppose at leasts one of the $p_1, \ldots, p_k$ *is odd. If* $q > p$, *then*

$$\mathbb{E}\left(X_1^{p_1} \cdots X_k^{p_k}\right) = 0.$$

2) Suppose all $p_1, \ldots, p_k$ *are even. If* $q > p$ *then*

$$\mathbb{E}\left(X_1^{p_1} \cdots X_k^{p_k}\right) = cq/(q - p),$$

where c *is given in (2.13). Otherwise if* $q \le p$ *then* $\mathbb{E}(X_1^{p_1} \cdots X_k^{p_k})$ *diverges.*

PROOF. By the density formula (2.1) and using the substitution $\mathbf{t} = v\mathbf{x}$ we have

$$\mathbb{E}\left(X_1^{p_1} \cdots X_k^{p_k}\right) = \int x_1^{p_1} \cdots x_k^{p_k} \left\{ \int_0^1 qv^{k+q-1} t_k(vx_1, \ldots, vx_k; r) \, dv \right\} dx_1 \cdots dx_k$$

$$= \int_0^1 qv^{q-1-p} \, dv \left\{ \mathbb{E}\left(T_1^{p_1} \cdots T_k^{p_k}\right) \right\}.$$

Note that the integral $\int_0^1 qv^{q-1-p} \, dv$ converges to $q/(q - p)$ if $q - p > 0$ and diverges otherwise. These and (2.13) yield the desired results. $\square$

The marginal distributions Since the marginal distributions of a k-variate t are still t, the marginal distributions of a k-variate slash t are slash t.

Theorem 2. *The marginal distributions of a k-variate slash t distribution are still slash t.*

PROOF. It suffices to show without loss of generality that for every $0 \le s \le k$,

$$\int \cdots \int f_k(x_1, \ldots, x_s, x_{s+1}, \ldots, x_k) \, dx_{s+1} \cdots dx_k = f_s(x_1, \ldots, x_s), \quad x_1, \ldots, x_s \in \mathbb{R},$$

where $f_k(\mathbf{x}) = f_k(\mathbf{x}; q, r, \mathbf{m}, \mathbf{R})$. Substitution of the density (2.1) in the left hand of the above equality gives

$$\int \cdots \int f_k(\mathbf{x}) \, dx_{s+1} \cdots dx_k = \int_0^1 qv^{k+q-1} \int \cdots \int t_k(v\mathbf{x}) dx_{s+1} \cdots dx_k dv,$$

where $t_k(\mathbf{t}) = t_k(\mathbf{t}; r, \mathbf{m}, \mathbf{R})$. By substitution $y_{s+1} = vx_{s+1}, \ldots, y_k = vx_k$ one derives

$$\int \cdots \int t_k(v\mathbf{x}) dx_{s+1} \cdots dx_k = v^{s-k} \int \cdots \int t_k(vx_1, \ldots, vx_s, y_{s+1}, \ldots, y_k) \, dy_{s+1} \cdots dy_k.$$

Because the marginals of the k-variate t distribution are still t, we have

$$\int \cdots \int t_k(vx_1, \ldots, vx_s, y_{s+1}, \ldots, y_k) \, dy_{s+1} \cdots dy_k = t_s(vx_1, \ldots, vx_k; r, \mathbf{m}_1, \mathbf{R}_{11}),$$

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 9 of 22

where $\mathbf{m} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top)^\top$ with $\mathbf{m}_1 \in \mathbb{R}^s$ and $\mathbf{R}$ is partitioned into the 2×2 block matrix with $\mathbf{R}_{11}$ being the $s \times s$ matrix at the position of $(1, 1)$-block. See pages 15-16 of Kotz and Nadarajah (2004). Combining the last two equalities yields the desired equality. □

Linear combinations Since the distribution of a linear function of a k-variate t variable is still t, it immediatly yields the following.

Theorem 3. *Let $\mathbf{A}$ be a nonsingular nonrandom matrix. If $\mathbf{X} \sim SLT_k(q, r, \mathbf{m}, \mathbf{R})$, then $\mathbf{AX} \sim SLT_k(q, r, \mathbf{Am}, \mathbf{ARA}^\top)$.*

The characteristic function Several authors derived the formulas for the characteristic functions $\varphi_\mathbf{T}$ of the k-variate t rv $\mathbf{T}$, see e.g. Joarder and Ali (1996), Dreier and Kotz (2002). Based on these formulas we can obtain the characteristic functions of the k-variate slash t using the following formula. For $\mathbf{X} \sim SLT_k(q, r, \mathbf{m}, \mathbf{R})$, it can be expressed as the ratio $\mathbf{X} = \mathbf{T}/U^{1/q}$ of two independent rv's, so that its characteristic function $\varphi_\mathbf{X}$ can be written as

$$\varphi_\mathbf{X}(\mathbf{t}) = \mathbb{E}(\exp(i\mathbf{t}^\top \mathbf{X})) = \int_0^1 \mathbb{E}(\exp(i\mathbf{t}^\top \mathbf{T} u^{-1/q})) \, du = \int_0^1 \varphi_\mathbf{T}(\mathbf{t} u^{-1/q}) \, du \qquad (2.14)$$

for $\mathbf{t}$ in some neighborhood $\mathbb{N}$ of the origin in which the above integral converges.

Example 1. For the standard univariate and bivariate slash t distributions $SLT_k(q, r), k = 1, 2$, their densities are given by

$$f_1(x; q, r) = \int_0^1 q v^q t_1(vx; r) \, dv, \quad x \in \mathbb{R},$$

and

$$f_2(\mathbf{x}; q, r) = \int_0^1 q v^{q+1} t_2(v\mathbf{x}; r) \, dv, \quad \mathbf{x} \in \mathbb{R}^2.$$
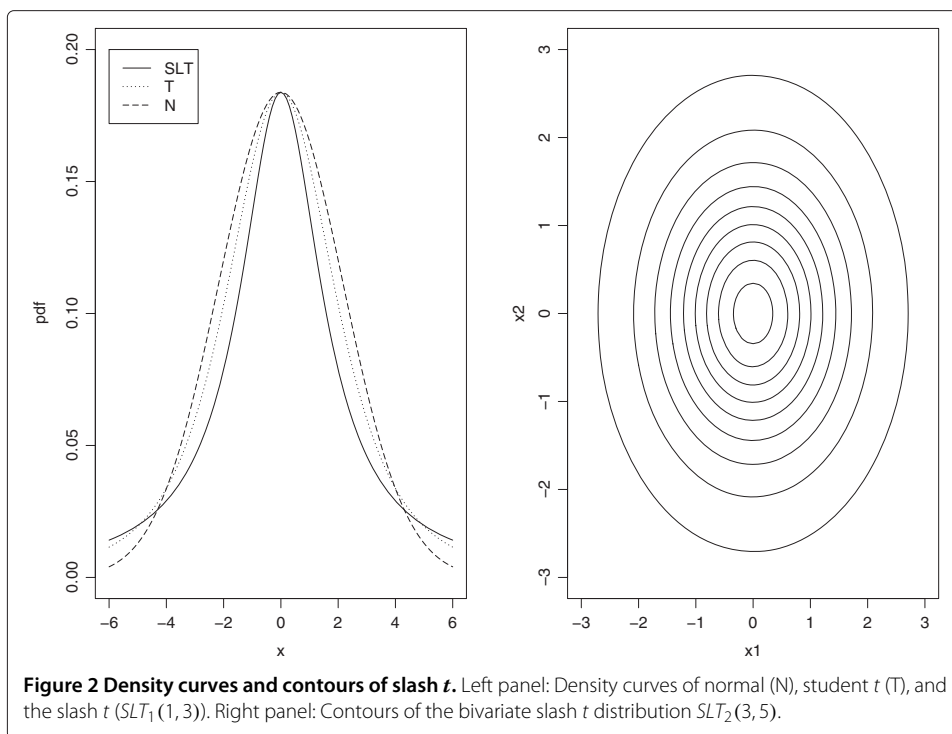
Displayed in Figure 2 are the density curves and contours. On the left panel are the density curves of the normal, t and the standard slash t with $q = 1$ and $r = 3$ degrees of freedom. The curves are calibrated so that they have the same height at the origin. Observe that the slash t has the fattest tail whereas the normal has the slimmest one. On the right panel are the contours of the bivariate slash t with $q = 3$ and $r = 5$. Clearly the contours are symmetric.

3 The multivariate skew-slash *t* distributions

In this section, we first recall the skew normal and skew t. In subsection 3.1, we define the standard skew-slash t distribution, study its relationships with other distributions and give the moments and characteristic function. In subsection 3.2, we use hidden truncation or selective sampling model and the order statistics to characterize the skew, slash and skew-slash normal and t distributions. In subsection 3.2, we define the general skew-slash t distribution, study its linear transformation and give an example in the end.

Azzalini and Dalla Valle (1996) introduced the skew normal distribution. A k-variate standard skew normal distribution has the density given by

$$2\phi_k(\mathbf{z})\Phi(\boldsymbol{\lambda}^\top \mathbf{z}), \quad \boldsymbol{\lambda}, \mathbf{z} \in \mathbb{R}^k,$$

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 10 of 22

Figure 2 Density curves and contours of slash *t*. Left panel: Density curves of normal (N), student *t* (T), and the slash *t* ($SLT_1(1,3)$). Right panel: Contours of the bivariate slash *t* distribution $SLT_2(3,5)$.

where ϕ_k is the pdf of the k-variate standard normal $N_k(\mathbf{0}, \mathbf{I})$ and Φ is the cdf of the standard normal $N(0,1)$. Denote it by $SKN_k(\boldsymbol{\lambda})$.

Kotz and Nadarajah (2004) wrote "... that the possibilities of constructing skewed multivariate *t* distributions are practically limitless". The two authors surveyed the definitions given by Gupta (2003), Sahu et al. (2003), Jones (2002) and Azzalini and Capitanio (2003).

Based on these definitions, we may define the skew-slash *t* distributions in different ways. In this article, however, we will take the following approach. First, we will define the standard skew-slash *t* based on the standard skew *t*, a common special case of Gupta (2003), Azzalini and Capitanio (2003), and others. We then introduce a general skew-slash *t* distribution by introducing location and scale parameters. Our definition may lose some nice interpretations. But we think that this definition is natural, concise and, in particular, convenient in applications.

According to Gupta (2003), a k-variate skew *t* distribution with parameters $\boldsymbol{\mu} \in \mathbb{R}^k$, Σ (correlation matrix $\mathbf{R}$), $\boldsymbol{\lambda} \in \mathbb{R}^k$ and $r > 0$ has the density

$$
h_k(\mathbf{t}; \boldsymbol{\lambda}, r, \Sigma) = 2t_k(\mathbf{t}; r, 0, \mathbf{R}) \Psi\left(\frac{\boldsymbol{\lambda}^\top \mathbf{t}}{\sqrt{1 + \mathbf{t}^\top \Sigma^{-1} \mathbf{t}/r}}; r + k\right), \quad \mathbf{t} \in \mathbb{R}^k, \tag{3.15}
$$

where Ψ is the cdf of the univariate standard *t* distribution with $r + k$ degrees of freedom. Denote this distribution by $SKT_k(\boldsymbol{\lambda}, r, \Sigma)$. This form of the density given here is slightly different from that of Gupta (2003). Several constant parameters appeared in his density formual are not explicitly expressed in our form of the density. We have incorporated them in the parameters in the above density. Accordingly parameters of the same names may have different values.

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 11 of 22

When Σ is the identity matrix $\mathbf{I}_k$, $SKT_k(\boldsymbol{\lambda}, r, \mathbf{I}_k)$ is referred to as the *standard skew t* and denoted by $SKT_k(\boldsymbol{\lambda}, r)$ with density $h_k(\mathbf{t}; \boldsymbol{\lambda}, r)$ given by

$$h_k(\mathbf{t}; \boldsymbol{\lambda}, r) = 2t_k(\mathbf{t}; r)\Psi\left(\frac{\boldsymbol{\lambda}^\top \mathbf{t}}{\sqrt{1 + \mathbf{t}^\top \mathbf{t}/r}}; r+k\right), \quad \mathbf{t} \in \mathbb{R}^k,$$

where $t_k(\mathbf{t}; r)$ is the density of the standard k-variate distribution $t_k(r)$ with degrees of freedom r. This is a common special case of the skew t shared by Gupta (2003), Azzalini and Capitanio (2003), and others.

3.1 The standard multivariate skew-slash *t* distribution

We begin with the definition, followed the moments and characteristic function.

Definition 2. A k-variate continuous random vector $\mathbf{W}_0$ is said to have a standard multivariate skew-slash t distribution with skewness parameter $\boldsymbol{\lambda}$, tail parameter q and degrees of freedom r, written $\mathbf{W}_0 \sim SSLT_k(\boldsymbol{\lambda}, q, r)$, if it can be written as the ratio of two independent rv's $\S \sim SKT_k(\boldsymbol{\lambda}, r)$ and $U \sim U(0, 1)$ as follows:

$$\mathbf{W}_0 = \mathbf{S}/U^{1/q}.$$

The standard skew-slash t generalizes the proposed standard k-variate slash t, the standard skew t of Gupta (2003) and Azzalini and Capitanio (2003), the standard slash normal of Kafadar (1988), and the standard skew-slash normal of Wang and Genton (2006). This is stated below.

Remark 2. *The limiting distribution of the standard skew-slash t distribution $SSLT_k$ $(\boldsymbol{\lambda}, q, r)$ is, as $q \to \infty$, the standard skew t distribution $SKT_k(\boldsymbol{\lambda}, r)$. The limiting distribution of $SSLT_k(\boldsymbol{\lambda}, q, r)$ is, as $r \to \infty$, the standard skew-slash normal $SSLN_k(\boldsymbol{\lambda}, q)$, which includes as special cases the standard skew normal $SKN_k(\boldsymbol{\lambda})$ ($q = \infty$) and the standard slash normal $SLN_k(q)$ ($\boldsymbol{\lambda} = 0$). As $\boldsymbol{\lambda} = 0$, $SSLT_k(\boldsymbol{\lambda}, q, r)$ reduces to the slash t distribution $SLT_k(q, r)$.*

In a similar way to the derivation of the standard slash t density in Section 2, we can obtain the density of the standard skew-slash t distribution $SSLT_k(\boldsymbol{\lambda}, q, r)$ as follows: for $\mathbf{w} \in \mathbb{R}^k$,

$$g_k(\mathbf{w}; \boldsymbol{\lambda}, q, r) = 2q\int_0^1 u^{k+q-1} t_k(u\mathbf{w}; r)\Psi\left(\frac{u\boldsymbol{\lambda}^\top \mathbf{w}}{\sqrt{1 + u^2\mathbf{w}^\top \mathbf{w}/r}}; k+r\right) du, \tag{3.16}$$

The moments Using the results of Azzalini and Dalla Valle (1996) or Kotz and Nadarajah (2004)(p.100-101) for the mean vector and covariance of the k-variate skew t distribution (i.e. by setting their $\boldsymbol{\alpha}$ equal to $\boldsymbol{\lambda}/\sqrt{1 + p/v}$ with their $p = k$ and $v = r$ here), we obtain the mean vector and covariance matrix of $\mathbf{W}_0 \sim SSLT_k(\boldsymbol{\lambda}, q, r)$ as follows:

$$\mathbb{E}(\mathbf{W}_0) = \mathbb{E}(\mathbf{S})\mathbb{E}(U^{-1/q}) = \frac{\sqrt{2}qr}{\sqrt{\pi}(q-1)(r-2)}\frac{\boldsymbol{\lambda}}{\sqrt{r+k+r\boldsymbol{\lambda}^\top \boldsymbol{\lambda}}}, \quad q > 1, r > 2,$$

$$\text{Var}(\mathbf{W}_0) = \frac{qr}{(q-2)(r-2)(r-4)}\left(\mathbf{I}_k - \frac{2(r+4)}{\pi(r-2)}\frac{r\boldsymbol{\lambda}\boldsymbol{\lambda}^\top}{r+k+r\boldsymbol{\lambda}^\top \boldsymbol{\lambda}}\right), \quad q > 2, r > 4.$$

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 12 of 22

The characteristic function As in deriving the characteristic function for the multivariate slash t in (2.14), one can obtain the characteristic function $\varphi_{\mathbf{W}_0}$ of $\mathbf{W}_0$ as follows:

$$\varphi_{\mathbf{W}_0}(\mathbf{t}) = \mathbb{E}(\exp(i\mathbf{t}^\top \mathbf{W}_0)) = \int_0^1 \varphi_{\mathbf{S}}(\mathbf{t}u^{-1/q})\, du, \quad \mathbf{t} \in \mathbb{N}$$

for some neighborhood $\mathbb{N}$ of the origin in which the above integral converges, where $\varphi_{\mathbf{S}}$ is the characteristic function of the standard skew t distribution $\mathbf{S} \sim SKT_k(\boldsymbol{\lambda}, r)$.

3.2 Hidden truncation and order-statistics characterization

In this subsection, we characterize the skew t, skew-slash normal and skew-slash t distributions using the hidden truncation or selective sampling model and the order statistics of the components of a bivariate normal or t variable.

Hidden truncation or selective sampling We first give a fact about conditional pdf. Let $\mathbf{X}$ be a continuous random vector that has pdf f. Let X_0 be a random variable with cdf F_0. Let a be a measurable function of $\mathbf{X}$ such that $P(A) > 0$ with $A = \{a(\mathbf{X}) \geq X_0\}$. Suppose $\mathbf{X}$ and X_0 are independent. Then for every $\mathbf{x}$,

$$P(\mathbf{X} \leq \mathbf{x}|a(\mathbf{X}) \geq X_0) = P(\mathbf{X} \leq \mathbf{x}, a(\mathbf{X}) \geq X_0)/P(A)$$

$$= E(\mathbf{1}[\,\mathbf{X} \leq \mathbf{x}]\, F_0(a(\mathbf{X})))/P(A) = \int_{-\infty}^{\mathbf{x}} f(\mathbf{y}) F_0(a(\mathbf{y}))\, d\mathbf{y}/P(A).$$

Hence the conditional pdf of $\mathbf{X}$ given A is

$$f(\mathbf{x}|A) = f(\mathbf{x}) F_0(a(\mathbf{x}))/P(A). \tag{3.17}$$

Using this we immediately derive the following results.

Proposition 1. *Suppose* $\mathbf{T} \sim t_k(r)$ *and* $T_0 \sim t_1(r+k)$ *are independent. Then the conditional pdf of* $\mathbf{T}$ *given A is*

$$f(\mathbf{t}|A) = t_k(\mathbf{t}; r) \Psi(a(\mathbf{t}); r+k)/P(A), \quad \mathbf{t} \in \mathbb{R}^k. \tag{3.18}$$

Consequently, if $U \sim U(0,1)$ *is independent of both* $\mathbf{T}$ *and* T_0 *then the conditional pdf of* $\mathbf{T}/U^{1/q}$ *for* $0 < q < 1$ *given A is*

$$g(\mathbf{t}|A) = 2q \int_0^1 u^{k+q-1} t(u\mathbf{t}; r) \Psi(a(u\mathbf{t}); r+k)\, du/2P(A), \quad \mathbf{t} \in \mathbb{R}^k. \tag{3.19}$$

In particular, both (3.18) and (3.19) hold for $a(\mathbf{t}) = (\tau_0 + \boldsymbol{\tau}^\top \mathbf{t})/\sqrt{1 + \mathbf{t}^\top \mathbf{t}/r}$ *where* $\tau_0, \boldsymbol{\tau}$ *are arbitrary constants. In this case,*

$$f(\mathbf{t}|A) = 2t_k(\mathbf{t}; r) \Psi\left(\frac{\tau_0 + \boldsymbol{\tau}^\top \mathbf{t}}{\sqrt{1 + \mathbf{t}^\top \mathbf{t}/r}}; r+k\right)/2P(A), \quad \mathbf{t} \in \mathbb{R}^k, \tag{3.20}$$

$$g(\mathbf{t}|A) = 2q \int_0^1 u^{k+q-1} t(u\mathbf{t}; r) \Psi\left(\frac{\tau_0 + u\boldsymbol{\tau}^\top \mathbf{t}}{\sqrt{1 + u^2 \mathbf{t}^\top \mathbf{t}/r}}; r+k\right) du/2P(A). \tag{3.21}$$

Remark 3. *The density function in (3.21) is the conditional pdf of the k-variate slash t rv* $\mathbf{X} = \mathbf{T}/U^{1/q}$ *given A. It is noteworthy that the hidden truncation model yields the pdf (3.20) and (3.21), the former is proportional to the pdf (3.15) of the skew t distribution and the latter is proportional to the pdf (3.16) of the proposed skew-slash t distribution.*

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 13 of 22

For more discussion see e.g. Chapter 6 of Genton (2004) and the references therein. The skew-slash normal of Wang and Genton (2006) is the special case of $r = \infty$.

Order statistics characterization Generalizing (c) of Theorem 1 in Arnold and Lin (2004), we give the following fact. Let (Y_1, Y_2) be a bivariate rv with pdf f and cdf F. Let $Y_{1,2} = \min(Y_1, Y_2)$ and $Y_{2,2} = \max(Y_1, Y_2)$ be the order statistics of the components of the bivariate random vector. Then

$$P\left(Y_{1,2} > y_1\right) = P\left(Y_2 \geq Y_1 > y_1\right) + P(Y_1 > Y_2 > y_1)$$

$$= \int_{y_1}^{\infty} dx_1 \int_{x_1}^{\infty} f(x_1, x_2)\, dx_2 + \int_{y_1}^{\infty} dx_2 \int_{x_2}^{\infty} f(x_1, x_2)\, dx_1.$$

Thus the pdf $f_{(1)}(y_1)$ of $Y_{1,2}$ is given by

$$f_{(1)}(y_1) = \int_{y_1}^{\infty} f(y_1, x_2)\, dx_2 + \int_{y_1}^{\infty} f(x_1, y_1)\, dx_1. \tag{3.22}$$

Let f_1, f_2 be the respective marginal pdf of Y_1, Y_2. Let

$$F_1(y_1|y_2) = P\left(Y_1 \leq y_1|Y_2 = y_2\right), \quad F_2(y_2|y_1) = P\left(Y_2 \leq y_2|Y_1 = y_1\right).$$

Then using $\int_{y_1}^{\infty} f(y_1, x_2)\, dx_2 = f_1(y_1) \int_{y_1}^{\infty} f(y_1, x_2)/f_1(y_1)\, dx_2 = f_1(y_1)\bar{F}_2(y_1)$ and (3.22) we derive

$$f_{(1)}(y_1) = f_1(y_1)\bar{F}_2(y_1|y_1) + f_2(y_1)\bar{F}_1(y_1|y_1). \tag{3.23}$$

Similarly we derive the pdf $f_{(2)}(y_2)$ of $Y_{2,2}$ below:

$$f_{(2)}(y_2) = f_2(y_2)F_1(y_2|y_2) + f_1(y_2)F_2(y_2|y_2). \tag{3.24}$$

In their Theorem 1, Arnold and Lin (2004) showed that the order statistics of the components of a random vector from a bivariate normal distribution obey the skew-normal law. Using (3.23) and (3.24) we can show that the order statistics of the components of a random vector from a bivariate t distribution obey the skew t law. Thus we extend their result from the normal to t distribution as stated below.

Proposition 2. *Let (T_1, T_2) have a bivariate t distribution $t_2(r, \mathbf{0}, \mathbf{R})$ with degrees of freedom r, mean zero vector and correlation matrix $\mathbf{R} = (1, \rho, \rho, 1)$ with $-1 < \rho < 1$. Define $T_{1,2} = \min(T_1, T_2)$ and $T_{2,2} = \max(T_1, T_2)$. Then the pdf $t_{(1)}, t_{(2)}$ of $T_{1,2}, T_{2,2}$ are given by*

$$t_{(1)}(t_1) = 2t_1(t_1; r)\Psi\left(\frac{-\lambda t_1}{\sqrt{1 + t_1^2/r}}; r+1\right), \quad t_1 \in \mathbb{R}, \tag{3.25}$$

$$t_{(2)}(t_2) = 2t_1(t_2; r)\Psi\left(\frac{\lambda t_2}{\sqrt{1 + t_2^2/r}}; r+1\right), \quad t_2 \in \mathbb{R}, \tag{3.26}$$

where $\lambda = \lambda(r, \rho) = \sqrt{(1 + 1/r)(1 - \rho)/(1 + \rho)}$.

Remark 4. *The density functions in (3.25) and (3.26) reduce to the result (c) of Theorem 1 of Arnold and Lin (2004) when the df $r = \infty$ as $\lambda(\infty, \rho) = \sqrt{(1 - \rho)/(1 + \rho)}$ is equal to their skewness parameter γ in the skew-normal distribution.*

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 14 of 22

PROOF OF PROPOSITION 2. Note first that T_1, T_2 have the same standard univariate t distribution $t_1(r)$. Using the formula for the conditional pdf of t distribution (see e.g. page 16 of Kotz and Nadarajah (2004)), the conditional cdf of T_2 given $T_1 = t_1$ can be written as

$$P(T_2 \le t_2 | T_1 = t_1) = \Psi(b(t_1, t_2; r, \rho); r + 1), \quad t_1, t_2 \in \mathbb{R},$$

where

$$b(t_1, t_2; r, \rho) = \frac{(t_2 - \rho t_1)\sqrt{1 + 1/r}}{\sqrt{(1 + t_1^2/r)(1 - \rho^2)}}.$$

Similarly,

$$P(T_1 \le t_1 | T_2 = t_2) = \Psi(b(t_2, t_1; r, \rho); r + 1), \quad t_1, t_2 \in \mathbb{R}.$$

We now apply (3.24), with both f_1 and f_2 equal to the pdf of $t_1(r)$ and both F_1 and F_2 equal to the cdf $\Psi(; r + 1)$ of the t distribution $t(r + 1)$, to obtain the pdf $t_{(2)}$ given in (3.26), noting in this case $b(t_2, t_2; r, \rho) = \lambda t_2/\sqrt{1 + t_2^2/r}$. Aanloguously we can prove (3.25) in view of the equality $\bar{\Psi}(t; r) = \Psi(-r; r)$ by the symmetry of the univariate t distribution. This completes the proof. $\qquad\square$

As a corollary of Proposition 2, we obtain a characterization of the skew-slash normal and t distributions through the order statistics of the components of a random vector from a bivariate t distribution as stated below.

Corollary 1. *Let (T_1, T_2) be given in Proposition 2. Assume $U \sim U(0, 1)$ is independent of (T_1, T_2). Then the pdf $g_{(1)}, g_{(2)}$ of $T_{1,2}/U^{1/q}, T_{2,2}/U^{1/q}$ for $0 < q < 1$ are given by*

$$g_{(1)}(t_1) = 2q \int_0^1 u^q t_1(ut_1; r) \Psi\left(\frac{-u\lambda t_1}{\sqrt{1 + u^2 t_1^2/r}}; r + 1\right) du, \quad t_1 \in \mathbb{R}, \qquad (3.27)$$

$$g_{(2)}(t_2) = 2q \int_0^1 u^q t_1(ut_2; r) \Psi\left(\frac{u\lambda t_2}{\sqrt{1 + u^2 t_2^2/r}}; r + 1\right) du, \quad t_2 \in \mathbb{R}. \qquad (3.28)$$

Remark 5. *The density functions in (3.27) and (3.28) are (i) the pdf of the order statistics of the components of the random vector $(T_1, T_2)/U^{1/q}$ from the bivariate slash t distribution $t_2(r, q)$, and (ii) reduce to the case of the skew-slash normal of Wang and Genton (2006) when the df $r = \infty$.*

PROOF OF COROLLARY 1. The desired (3.27) follows from (3.25) and the equalities

$$P(T_{1,2}/U^{1/q} \le t_1) = \int_0^1 P(T_{1,2} \le u^{1/q} t_1) \, du = 2q \int_{-\infty}^{t_1} \int_0^1 v^q t_{(1)}(vs) \, dv \, ds,$$

where the independence of $T_{1,2}$ and U is used to claim the first equality while the second equality follows from a change of variables. Similarly (3.28) can be proved and this finishes the proof. $\qquad\square$

The multivariate skew-slash t distributions We now introduce a general multivariate skew-slash t distribution by incorporating location and scale parameters.

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 15 of 22

Definition 3. A continuous k-variate rv $\mathbf{W}$ has a *multivariate skew-slash t distribution* with location $\boldsymbol{\mu}$, scale Σ, skewness parameter $\boldsymbol{\lambda} \in \mathbb{R}^k$, tail parameter q and degrees of freedom r, written $\mathbf{W} \sim SSLT_k(\boldsymbol{\lambda}, q, r, \boldsymbol{\mu}, \Sigma)$, if it can be represented as a linear transformation of the standard multivariate skew-slash t rv $\mathbf{W}_0 \sim SSLT_k(\boldsymbol{\lambda}, q, r)$ as follows:

$$\mathbf{W} = \boldsymbol{\mu} + \Sigma^{1/2} \mathbf{W}_0,$$

where $\Sigma^{1/2}$ is the the choleski decomposition of the positive definite covariance matrix Σ.

By a change of variables in (3.16), we derive the density of $\mathbf{W}$ given by

$$
\begin{aligned}
g_k(\mathbf{w}; \boldsymbol{\lambda}, q, r, \boldsymbol{\mu}, \Sigma) = {} & 2q|\Sigma|^{-1/2} \int_0^1 u^{k+q-1} t_k(u\Sigma^{-1/2}(\mathbf{w} - \boldsymbol{\mu}); r) \\
& \times \Psi\left(\frac{u\boldsymbol{\lambda}^\top \Sigma^{-1/2}(\mathbf{w} - \boldsymbol{\mu})}{\sqrt{1 + u^2 Q(\mathbf{w}; \boldsymbol{\mu}, \Sigma)/r}}; k + r \right) du, \quad \mathbf{w} \in \mathbb{R}^k,
\end{aligned}
\tag{3.29}
$$

where $Q(\mathbf{w}; \boldsymbol{\mu}, \Sigma) = (\mathbf{w} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{w} - \boldsymbol{\mu})$.

As in the case of the standard skew-slash t, one notices that the skew-slash t generalizes the slash t, the skew t of Azzalini and Capitanio, the slash normal of Kafadar, the skew normal of Azzalini and Dalla Valle and the skew-slash normal of Wang and Genton. This is stated below.

Remark 6. *The limiting distribution of the multivariate skew-slash t distribution $SSLT_k(\boldsymbol{\lambda}, q, r, \boldsymbol{\mu}, \Sigma)$, as q tends to infinity, is the skew t distribution $SKT_k(\boldsymbol{\lambda}, r, \boldsymbol{\mu}, \Sigma)$. The limiting distribution of $SSLT_k(\boldsymbol{\lambda}, q, r, \boldsymbol{\mu}, \Sigma)$ is, as r tends to infinity, the skew-slash normal $SSLN_k(\boldsymbol{\lambda}, q, \boldsymbol{\mu}, \Sigma)$, which include as special cases the k-variate skew normal $SKN_k(\boldsymbol{\lambda}, \boldsymbol{\mu}, \Sigma)$ ($q = \infty$) and the k-variate slash normal $SLN_k(q; \boldsymbol{\mu}, \Sigma)$ ($\boldsymbol{\lambda} = 0$). As $\boldsymbol{\lambda} = 0$, $SSLT_k(\boldsymbol{\lambda}, q, r, \boldsymbol{\mu}, \Sigma)$ simplifies to the k-variate slash t distribution $SLT_k(q, r, \boldsymbol{\mu}, \Sigma)$.*

Linear combinations Since the distribution of a linear function of a k-variate skew t variable is still skew t (see e.g. Section 5.9 of Kotz and Nadarajah (2004)), it immediatly yields the following result. Note that the relationship between our skewness parameter $\boldsymbol{\lambda}$ and their $\boldsymbol{\alpha}$ is $\boldsymbol{\lambda} = \sqrt{1 + k/r} \boldsymbol{\alpha}$. Let $\mathbf{D} = \mathrm{diag}(\sigma_{1,1}, \ldots, \sigma_{k,k})$ denote the diagonal matrix consisting of the diagonal entries of $\Sigma = (\sigma_{i,j})$ and $\mathbf{R} = \mathbf{D}^{-1/2} \Sigma \mathbf{D}^{-\top/2}$ be the correlation matrix.

Theorem 4. *Let $\mathbf{A}$ be a nonsingular matrix. If $\mathbf{W} \sim SSLT_k(\boldsymbol{\lambda}, q, r, \boldsymbol{\mu}, \Sigma)$, then $\mathbf{A}\mathbf{W} \sim SSLT_k(\tilde{\boldsymbol{\lambda}}, q, r, \mathbf{A}\boldsymbol{\mu}, \widetilde{\Sigma})$ where $\widetilde{\Sigma} = \mathbf{A}\Sigma\mathbf{A}^\top$ and*

$$
\tilde{\boldsymbol{\lambda}} = \frac{\widetilde{\Sigma}^{-1/2} \mathbf{B}^\top \boldsymbol{\lambda}}{\sqrt{1 + (1 + k/r)^{-1} \boldsymbol{\lambda}^\top (\mathbf{R} - \mathbf{B}\widetilde{\Sigma}^{-1}\mathbf{B}^\top)\boldsymbol{\lambda}}}, \quad \mathbf{B} = \mathbf{D}^{-1/2} \Sigma \mathbf{A}.
$$

To give graphical view of the skewness and tail behaviors of the skew-slash t distributions, we plot the density curves of the univariate standard skew-slash t and contours of the bivariate standard skew-slash t below.

Example 2. For the univariate and bivariate standard skew-slash t distributions $SSLT_1(\lambda, q, r)$ and $SSLT_2(\lambda_1, \lambda_2, q, r)$, the densities are given by
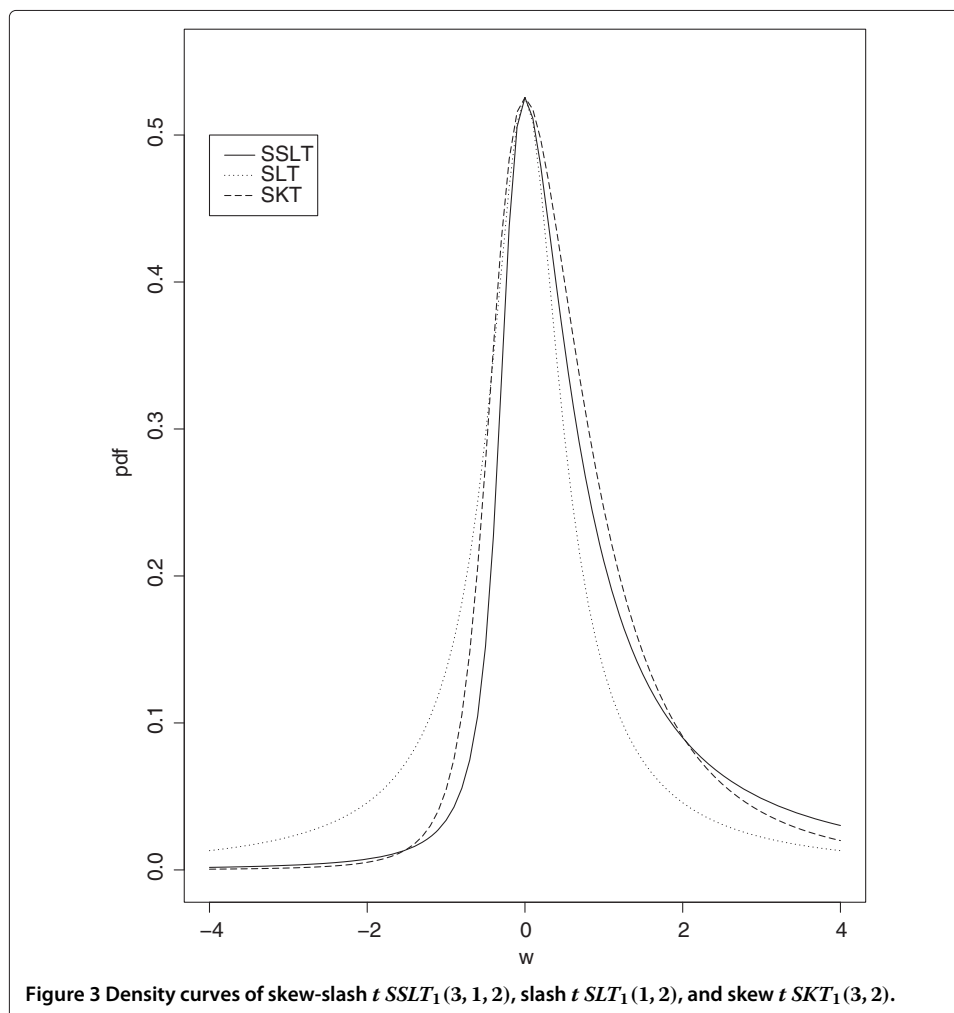
$$g_1(w; \lambda, q, r) = 2q \int_0^1 u^q t_1(uw; r) \Psi\left(\frac{u\lambda w}{\sqrt{1 + u^2 w^2/r}}; 1 + r\right) du, \quad w \in \mathbb{R}, \qquad (3.30)$$
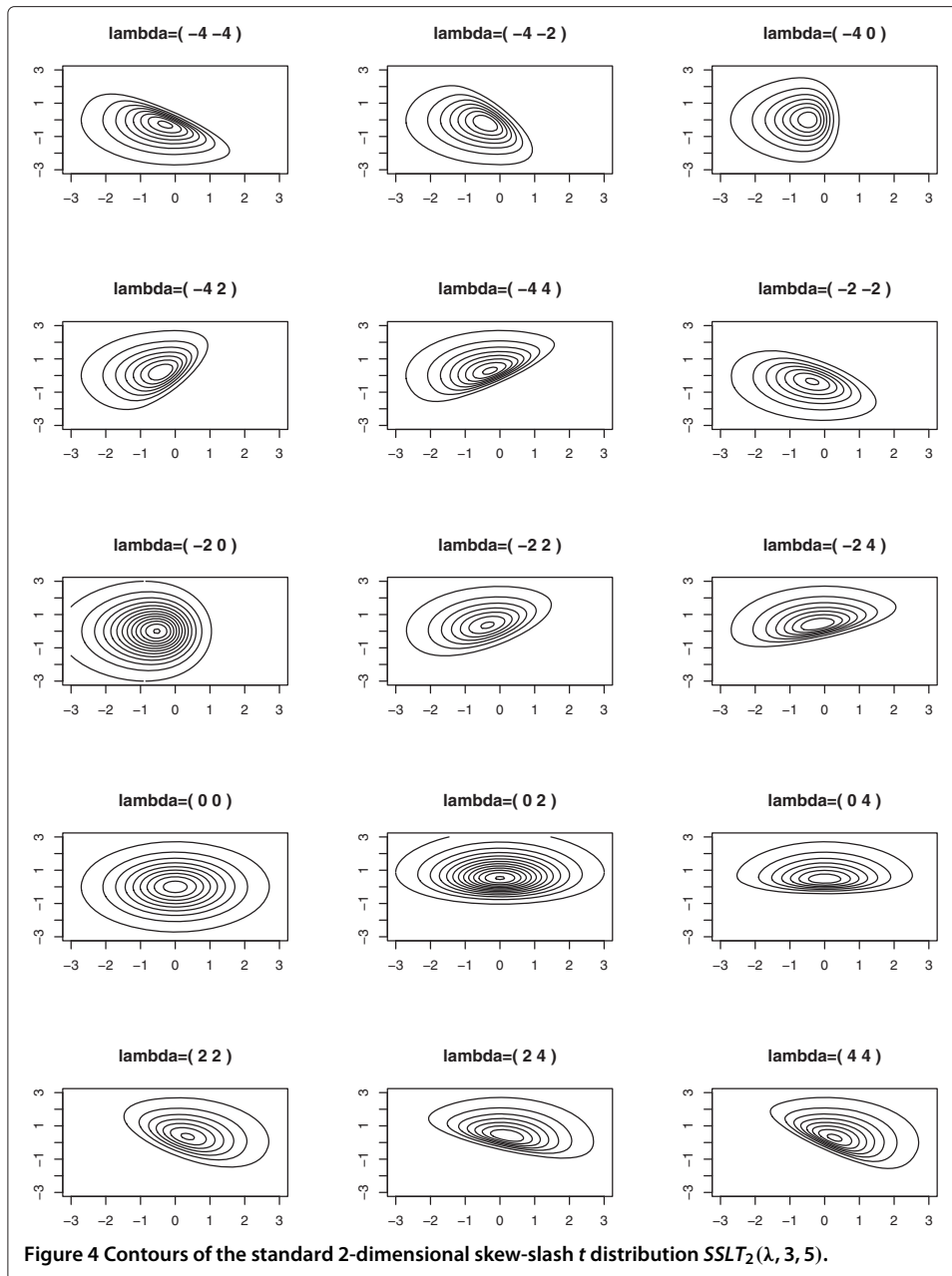
and, with $\mathbf{w} = (w_1, w_2)^\top \in \mathbb{R}^2$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$,

$$g_2(\mathbf{w}; \boldsymbol{\lambda}, q, r) = 2q \int_0^1 u^{q+1} t_2(u\mathbf{w}; r) \Psi\left(\frac{u(\lambda_1 w_1 + \lambda_2 w_2)}{\sqrt{1 + u^2(w_1^2 + w_2^2)/r}}; 2 + r\right) du. \qquad (3.31)$$

Displayed in Figure 3 are the density curves of the standard skew-slash t distribution $SSLT_1(3, 1, 2)$, the skew t distribution $SKT_1(3, 2)$, and the slash t distribution $SLT_1(1, 2)$ distributions. The curves are shifted and rescaled for comparison. Observe that the skew-slash t has the fattest tail and is skewed to the right most.

Displayed in Figure 4 are the contour plots of the bivariate standard skew-slash t distribution $SSLT_2(\boldsymbol{\lambda}, 3, 5)$ for different values of the skewness parameter vector $\boldsymbol{\lambda}$. Clearly the contours are more skewed as the skewness parameter vector $\boldsymbol{\lambda}$ gets longer.

Figure 3 Density curves of skew-slash t $SSLT_1(3, 1, 2)$, slash t $SLT_1(1, 2)$, and skew t $SKT_1(3, 2)$.

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 17 of 22

Figure 4 Contours of the standard 2-dimensional skew-slash *t* distribution *SSLT*$_2$(λ, 3, 5).

4   Statistical inference

In this section, we discuss maximum likelihood estimation and the Bayesian method and provide the approximate sampling distribution of the estimates.

The likelihood approach   Let $p(\mathbf{z}; \boldsymbol{\theta})$ denote either the slash *t* density in (2.1) or the skew-slash *t* density in (3.29), where $\boldsymbol{\theta}$ denotes the corresponding parameter vector, i.e. $\boldsymbol{\theta} = (q, r, \mathbf{m}, \mathbf{R})$ or $\boldsymbol{\theta} = (\lambda, q, r, \boldsymbol{\mu}, \Sigma)$. As the degrees of freedom *r* is unknown, we estimate it by the MLE treating it as a positive real number. Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be a random sample from the density *p*. Based on the sample, the parameter $\boldsymbol{\theta}$ can be estimated by the MLE

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 18 of 22

$\hat{\boldsymbol{\theta}}$. This is the parameter value $\hat{\boldsymbol{\theta}}$ which maximizes $L(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\log p(\mathbf{Z}_i; \boldsymbol{\theta})$ of the log likelihood functions of the sample over the parameter space Θ, that is,

$$L(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \Theta} \frac{1}{n}\sum_{i=1}^{n}\log p(\mathbf{Z}_i; \boldsymbol{\theta}). \tag{4.32}$$

Let $\mathbf{S}_i(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}}\log p(\mathbf{Z}_i; \boldsymbol{\theta})$ be the score for the observation $\mathbf{Z}_i$. Under suitable conditions, the MLE $\hat{\boldsymbol{\theta}}$ is the solution to the score equation

$$\sum_{i=1}^{n}\mathbf{S}_i(\boldsymbol{\theta}) = 0. \tag{4.33}$$

Let $\mathbf{J}(\boldsymbol{\theta}) = E(\mathbf{S}_1(\boldsymbol{\theta})\mathbf{S}_1(\boldsymbol{\theta})^{\top})$ be the information matrix. Clearly the information matrix can be estimated by

$$\hat{\mathbf{J}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{S}_i(\hat{\boldsymbol{\theta}})\mathbf{S}_i(\hat{\boldsymbol{\theta}})^{\top}. \tag{4.34}$$

Under suitable conditions, the sampling distribution of $\hat{\boldsymbol{\theta}}$ can be approximated by the normal distribution with mean vector $\boldsymbol{\theta}$ and variance-covariance matrix $\hat{\mathbf{J}}^{-1}$, that is, approximately

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, n^{-1}\hat{\mathbf{J}}^{-1}). \tag{4.35}$$

Based on this normal approximation, one can perform hypothesis testing, construct confidence intervals and, in particular, calculate the standard error (SE) of each component $\hat{\theta}_j$ of the MLE $\hat{\boldsymbol{\theta}}$ as follows:

$$SE(\hat{\theta}_j) = \sqrt{n^{-1}\hat{J}^{jj}}, \quad j = 1, \ldots, k, \tag{4.36}$$

where $\hat{J}^{jj}$ is the (j, j)- entry of the estimated inverse information matrix $\hat{\mathbf{J}}^{-1}$.

The numerical value of the MLE $\hat{\boldsymbol{\theta}}$ can be found by solving the score equation (4.33) using the newton's method. Alternatively, one can directly search the solution of the maximization problem (4.32), for example, using the subroutine *optim* in the R package.

As for initial values of the newton's algorithm, one can use the moment estimates of the parameters or other available consistent estimates. One technical issue here is that the estimate $\hat{\Sigma}$ of Σ must be positive definite. What we did in our applications was that we estimated the entries of Σ and then verified the positive definiteness of $\hat{\Sigma}$.

The Bayesian approach Given observed data $\mathbf{D}$, the likelihood function $L(\mathbf{D}|\boldsymbol{\theta})$ can be obtained from the proposed multivariate slash or skew-slash t distribution with parameter vector $\boldsymbol{\theta}$. The posterior density then satisfies $p(\boldsymbol{\theta}|\boldsymbol{D}) \propto L(\boldsymbol{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$, where $\pi(\boldsymbol{\theta})$ is the joint prior density of $\boldsymbol{\theta}$ based on the available prior information on it. We choose a prior density for each component of $\boldsymbol{\theta}$ and take the joint prior density of $\boldsymbol{\theta}$ to be equal to the product of the marginal prior densities. The resulting full Bayesian model has the hierarchical structure with the conditional density of $\boldsymbol{D}|\boldsymbol{\theta}$ and the prior distribution $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$ in the proposed model. One can obtain a random sample from the joint posterior density by the Markov Chain Monte Carlo (MCMC) method, and a parametric Baysian analysis of the model can be implemented using the Gibbs sampling method in R or JAGS.

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 19 of 22

5 Simulations and applications

In this section, we use simulations to compare the performance of the skew-slash t and the skew-slash normal distributions. Simulations are also used to demonstrate parameter estimation using both the maximum likelihood criterion and Bayesian paradigm. Then the skew-slash t, with the density given in (3.29), and the skew-slash normal, with the density given below, are applied to fit two real datasets.

$$
\eta_k(\mathbf{w}; \boldsymbol{\lambda}, q, \boldsymbol{\mu}, \Sigma) = 2q \int_0^1 u^{k+q-1} \phi_k(u\mathbf{w}; u\boldsymbol{\mu}, \Sigma)
$$
$$
\times \Phi\left(u\boldsymbol{\lambda}^\top \Sigma^{-1/2}(\mathbf{w} - \boldsymbol{\mu})\right) du, \quad \mathbf{w} \in \mathbb{R}^k,
\tag{5.37}
$$

where ϕ_k is the density of the k-variate normal distribution $N_k(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

5.1 Simulation study

Comparison between the skew-slash t and skew-slash normal To compare the two distributions, we first generated data from the 2-variate skew-slash t then fitted it with both 2-variate skew-slash t and skew-slash normal, and vice versa (i.e. generated data from the 2-variate skew-slash normal then fitted it with the two distributions). Reported in Table 2 are the average AIC values and average MLE's of the parameters based on the sample size $n = 250$ and repetitions $M = 200$.

Notice that for data generated from both the skew-slash t and skew-slash normal, the average AIC values of the skew-slash t fitting were lower than those of the skew-slash normal, indicating a better overall model fitting of the former to the data than the latter.

Parameter estimation by the Bayesian method We now conducted simulations to study the behaviors of the MLE's and Bayesian estimates of the parameters in the data-generation models for the standard univariate and bivariate slash and skew-slash t distributions. Here the prior distributions are the standard normal $N(0, 1)$ for the skewness parameter λ (or the components of the parameter vector $\boldsymbol{\lambda}$), the exponential distribution with rate 0.1 truncated at 2 for the tail parameter q and truncated at 4 for the degrees of freedom r. It follows from Sahu et al. (2003) and Fernandez and Steel (1998) that the resulting distributions have finite variances. We generated 100 random samples of size 200 from the standard univariate slash and skew-slash t and the standard bivariate slash and skew-slash t. Reported in Table 3 are the average estimates of the parameters based on the maximum likelihood criterion and Bayesian paradigm. Observe that the two types of estimates are close for all the simulation setups.

5.2 Applications

Model fitting to the GAD Data Gestational age at delivery (GAD) is a variable widely studied in epidemiology, see, for example, Longnecker et al. (2001). We applied the skew-slash t and skew-slash normal distributions to fit the log transformed GAD of $n = 100$ observations. Figure 1 is the histogram superimposed with the fitted density curves, while Table 1 reports the MLE's, the standard errors (SE) of the parameter estimates and the AIC. We can see from Figure 1 that the skew-slash t distribution was able to better capture

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 20 of 22

Table 3 Maximum likelihood and Bayesian estimates comparison

Parameter	True Value	SLT MLE	SLT Bayes	SSLT MLE	SSLT Bayes
		Univariate			
λ	3			2.496	2.511
q	5	6.196	6.236	6.689	6.479
r	8	8.015	7.084	7.650	7.194
		Bivariate			
λ_1	-5			-5.543	-3.288
λ_2	3			3.333	1.967
q	5	6.410	4.113	5.899	4.118
r	8	7.480	8.468	7.726	8.609

The average maximum likelihood and Bayesian estimates based on 100 samples of size 200. SLT=Slash *t* and SSLT=Skew-Slash *t*.

the peak of the histogram, giving a better estimation of the density to the majority of data points. In the mean time, the AIC values in Table 1 indicated that the skew-slash *t* fitting was better than the skew-slash normal.

Model fitting to the AIS data Azzalini and Dalla Valle (1996) used their skew normal distribution to fit (LBM-lean body mass, BMI-body mass index) pairs of the athletes from Australian Institute of Sport (AIS), where the data of $n = 202$ observations were reported in Cook and Weisberg (1994). Wang and Genton (2006) used their skew-slash normal distribution to re-fit the data. Here we applied the proposed skew-slash *t* distribution to re-fit the (LBM, BMI) pairs in the AIS data. Before fitting we standardized the variables.

Figure 5 is the scatter plot superimposed with the fitted skew-slash *t* and skew-slash normal contours in the scale of LBM and BMI. The skew-slash normal contour is similar
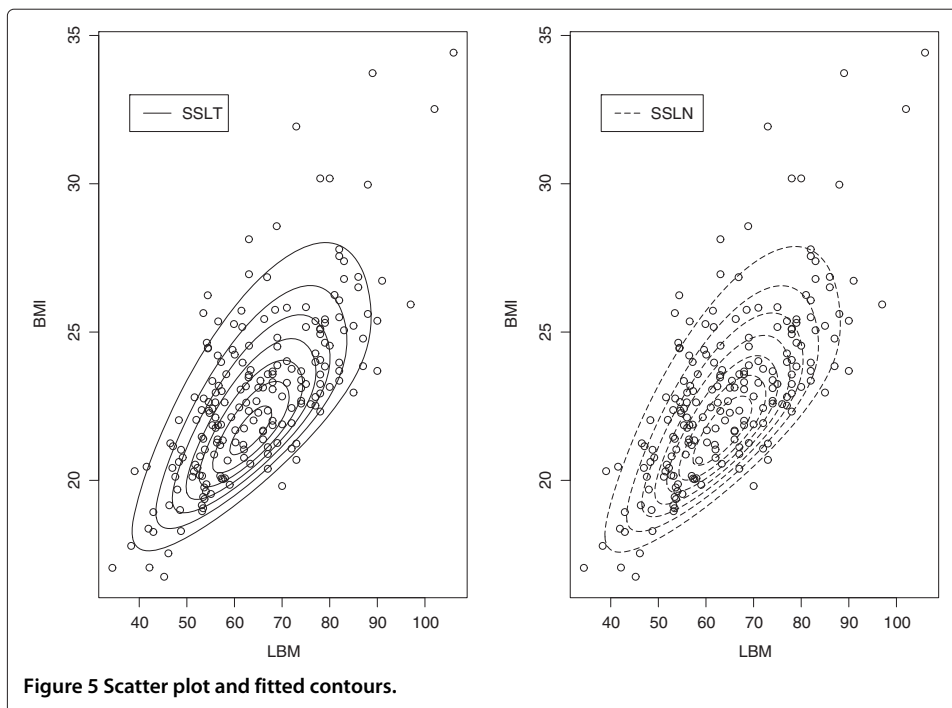

Figure 5 Scatter plot and fitted contours.

Tan *et al. Journal of Statistical Distributions and Applications* (2015) 2:3

Page 21 of 22

to what was reported in Wang and Genton (2006). The comparison between the two contours seemed to indicate that the fittings based on two models were close.

Reported in Table 4 are the MLE's, standard errors and AIC. A smaller AIC value of the skew-slash normal fitting than the skew-slash t fitting indicated that the former was a better fit to this data. When the skew-slash normal was used to fit the data we were able to obtain the MLE's. But the row and column corresponding to parameter q in the Hessian matrix were zero, suggesting a simpler skew normal fitting to this data. The skew-slash normal fitting by Wang and Genton (2006) led to the same conclusion though they did not report the standard errors of the parameter estimates.

In conclusion, the proposed slash and skew-slash t are competitive candidate models for fitting skewed and heavy tailed data. The parameters can be estimated under either the frequentist method or Bayesian paradigm. Although for a particular dataset the skew-slash t may not be the final model, it is a good choice to start with in model selection due to its flexibility and the fact that it takes the skew normal, skew t and hence the usual normal and t as its submodels.

6 Concluding remarks

In this article, we defined the multivariate slash and skew-slash distributions in a pursuit of providing additional distributions to simulate and fit skewed and heavy tailed data. We investigated the heavy tail behaviors and tractable properties of these distributions which are useful in simulations and applications to real data. We characterized the skew t, the skew-slash normal and the skew-slash t distributions using both the hidden truncation or selective sampling model and the order statistics of the components of a bivariate normal or t variable. We demonstrated that the proposed skew-slash t model takes as sub-models the slash t, the slash normal, the skew-slash normal, the skew normal, the skew t and hence the usual normal and t. This nested property can be used in hypothesis testing.

Our simulations and applications to real data indicated that the proposed skew-slash t fitting outperformed the skew-slash normal fitting. Even though the skew-slash normal contains a tail parameter q, the fitting with it to the GAD data was unsatisfactory as the SE of the MLE of the tail parameter q was large, see Table 1. This suggests that not all heavy-tail properties in data can be explained by the tail parameter q in the slash distributions. Thus it makes sense for us to further search for distributions which can be used to fit

Table 4 The skew-slash normal and *t* fitting to the AIS data

	SSLT		SSLN
	MLE	SE	MLE
AIC		981.018	978.426
μ_1	-0.210	0.151	-0.210
μ_2	-0.936	0.110	-0.954
σ_{11}^2	0.566	0.097	0.831
σ_{12}	0.465	0.086	0.708
σ_{22}^2	2.042	0.454	1.486
λ_1	-1.745	0.531	-2.225
λ_2	3.771	1.016	2.091
q	121.021	660.502	22.534
r	25.124	17.756	

SSLN=Skew-Slash Normal and SSLT=Skew-Slash *t*.

heavy tailed data. Our proposed slash and skew-slash t distributions can be considered as an example in this attempt. We complete our remarks by pointing out that the degrees-of-freedom parameter r and the tail parameter q would explain different types of fat tail behaviors existed in data.

Authors' contributions
FT contributed to the entire work, HP to the theoretical part, and YT to the inference, simulations and applications. All authors read and approved the final manuscript.

Author details
[1]Department of Mathematical Sciences, Indiana University Purdue University Indianapolis, N. Blackford Street, IN 46202, Indianapolis, USA. [2]Department of Statistics, Florida State University, N. Woodward Ave., FL 32306, Tallahassee, USA.

References
Arnold, BC, Lin, GD: Characterization of the Skew-normal and Generalized Chi Distributions. Sankhyā. **66**, 593–606 (2004)
Azzalini, A, Capitanio, A: Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. J. R. Stat. Soc. Ser. B. **65**, 367–389 (2003)
Azzalini, A, Dalla Valle, A: The multivariate skew-normal distribution. Biometrika. **83**, 715–726 (1996)
Cook, RD, Weisberg, S: An Introduction to Regression Graphics. Wiley, New York (1994)
Dreier, I, Kotz, S: A note on the characteristic function of the t-distribution. Statist. & Probabil. Lett. **57**, 221–224 (2002)
Fernandez, C, Steel, MFJ: On Bayesian Modeling of Fat Tails and Skewness. J. Am. Statist. Assoc. **93**, 359–371 (1998)
Genton, MG (ed.): Skew-elliptical distributions and their applications: a journey beyond normality. Chapman and Hall/CRC, Boca Raton (2004)
Gupta, AK: Multivariate skew t-distribution, Statistics: A Journal of Theoretical and Applied Statistics. **37**, 359–363 (2003). doi:10.1080/715019247
Joarder, AH, Ali, MM: On the characteristic function of the multivariate student t distribution. Pak. J. Statist. **12**, 55–62 (1996)
Johnson, N, Kotz, S: Distributions in Statistics: Continuous Multivariate Distributions. John Wiley & Sons, New York (1972)
Jones, MC: Marginal replacement in multivariate densities, with application to skewing spherically symmetric distributions. J. Multivar. Anal. **81**, 85–99 (2002)
Kafadar, K: Slash Distribution. In: Johnson, NL, Kotz, S, Read, C (eds.) Encyclopedia of Statistical Sciences. vol. 8, pp. 510–511. Wiley, New York, (1988)
Kotz, S, Nadarajah, S: Multivariate t distributions and their applications. Cambridge University Press, Cambridge (2004)
Longnecker, MP, Klebanoff, MA, Zhou, H, Brock, JW: Association between maternal serum concentration of the ddt metabolite dde and preterm and small-for-gestational-age babies at birth. Lancet. **358**, 110–114 (2001)
Sahu, SK, Dey, DK, Branco, MD: A new class of multivariate skew distributions with applications to Bayesian regression models. Can. J. Stat. **31**, 129–150 (2003)
Wang, J, Genton, M: The multivariate skew-slash distributions. J. Statist. Plann. Inferr. **136**, 209–220 (2006)