Communications for Statistical Applications and Methods
2016, Vol. 23, No. 4, 321–334

http://dx.doi.org/10.5351/CSAM.2016.23.4.321
Print ISSN 2287-7843 / Online ISSN 2383-4757

Power *t* distribution

Jun Zhao[a], Hyoung-Moon Kim[1,a]

[a]Department of Applied Statistics, Konkuk University, Korea

Abstract

In this paper, we propose power *t* distribution based on *t* distribution. We also study the properties of and inferences for power *t* model in order to solve the problem of real data showing both skewness and heavy tails. The comparison of skew *t* and power *t* distributions is based on density plots, skewness and kurtosis. Note that, at the given degree of freedom, the kurtosis's range of the power *t* model surpasses that of the skew *t* model at all times. We draw inferences for two parameters of the power *t* distribution and four parameters of the location-scale extension of power *t* distribution via maximum likelihood. The Fisher information matrix derived is nonsingular on the whole parametric space; in addition we obtain the profile log-likelihood functions on two parameters. The response plots for different sample sizes provide strong evidence for the estimators' existence and unicity. An application of the power *t* distribution suggests that the model can be very useful for real data.

Keywords: skew-*t* distribution, Fisher information matrix, heavy tail, profile likelihood, skewness, maximum likelihood, four parameter distribution

1. Introduction

The error term is typically assumed to follow a normal distribution in regression analysis, but is not a realistic assumption in many applications. This restrictive assumption can also result in a lack of robustness against departure from normal distribution and invalid statistical inference, especially when the data show skewness and outlying observations. Some models have been developed to solve these problems.

A skew-normal distribution with density function is given by

$$\phi(z; \lambda) = 2\phi(z)\Phi(\lambda z), \qquad z \in \mathbb{R}, \tag{1.1}$$

where ϕ and Φ are the probability density function (pdf) and cumulative distribution function (cdf) of a standard normal distribution, respectively. This distribution was developed by Azzalini (1985, 1986) and has been studied in detail by Henze (1986), Arnold *et al.* (1993), and Pewsey (2000). However, the Fisher information matrix of skew-normal distribution is singular.

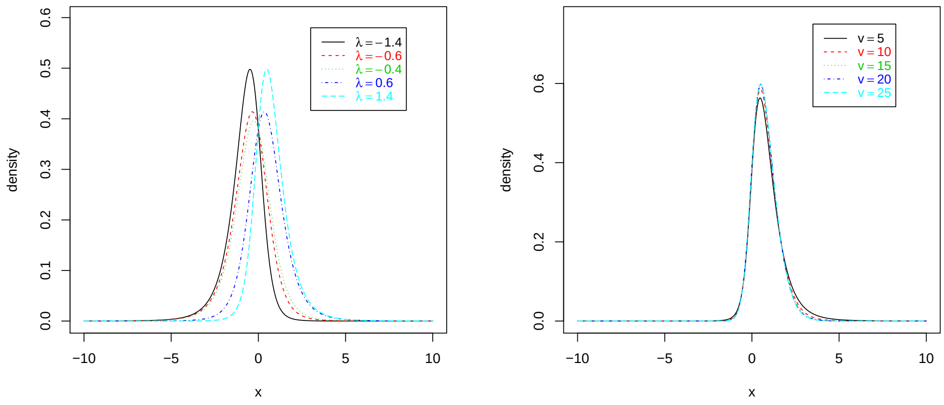
Based on a normal distribution, Durrans (1992) developed a power-normal distribution with density function

$$\phi(z; \alpha) = \alpha\phi(z)\{\Phi(z)\}^{\alpha-1}, \qquad z \in \mathbb{R},\ \alpha \in \mathbb{R}^+, \tag{1.2}$$

where ϕ and Φ are the same as before. Gupta and Gupta (2008) and Pewsey *et al.* (2012) also considered a power-normal distribution.

[1] Corresponding author: Department of Applied Statistics, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, Seoul 05029, Korea. E-mail: hmkim@konkuk.ac.kr

Published 31 July 2016 / journal homepage: http://csam.or.kr



(a) (b)

Figure 1: *Skew-t pdfs. (a)* $St(\lambda, v = 5)$ *for* $\lambda = -1.4$ *(solid line)*, -0.6 *(dashed line)*, -0.4 *(dotted line)*, 0.6 *(dotted-dashed line) and* 1.4 *(long-dashed line); (b)* $St(\lambda = 2.2, v)$ *for* $v = 5$ *(solid line)*, 10 *(dashed line)*, 15 *(dotted line)*, 20 *(dotted-dashed line), and* 25 *(long-dashed line).*

Another set of distributions for modeling skewed and heavy-tailed data is the skew extensions to the Student's t distribution; these distributions possess the nonsingular Fisher information matrix. Branco and Dey (2001) were the first to develop the skew-t distribution. Since then, several studies have examined different skew-t-type distributions such as Jones (2001), Azzalini and Capitanio (2003), Gupta (2003), Jones and Faddy (2003), Sahu *et al.* (2003), and Jones and Larsen (2004). In this paper, we mainly use skew-t distribution by Azzalini and Capitanio (2003).

In this study, we introduce a power t distribution whose Fisher information matrix is nonsingular and draw some inferences based on t distribution. In Section 2, we define the power t model and compare its behavior with skew-t distribution. For a given degree of freedom, the skewness interval of the power t model contains the skew-t model under some conditions; however, the kurtosis contains it all the time. We also provide some results that can be used to draw likelihood-based inferences for the distribution parameter and its location-scale extension. We also present the details of score equations and the observed and expected information matrices. We then apply the model to the data of total glycerol content in Grignolino wine in Section 3. Finally, Section 4 concludes the paper.

2. Power *t* distribution

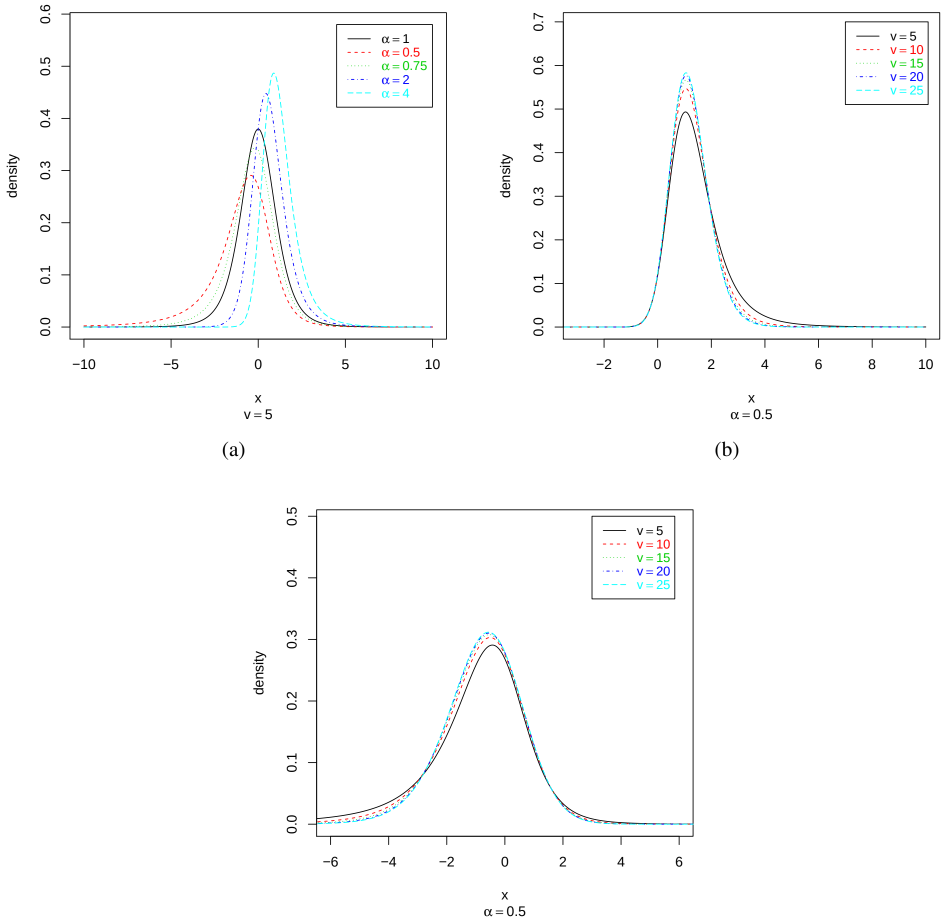
2.1. Definition

Random variable X is said to have the power t distribution with degrees of freedom v if its pdf is given by

$$\phi_{\alpha,v}(x) = \alpha\phi_v(x)\{\Phi_v(x)\}^{\alpha-1}, \qquad x \in \mathbb{R},\ \alpha \in \mathbb{R}^+, \tag{2.1}$$

where

$$\phi_v(x) = t(x; v),$$
$$\Phi_v(x) = \int_{-\infty}^{x} \phi_v(t)dt,$$



(a) (b) (c)

Figure 2: *Power t pdfs. (a)* $Pt(\alpha, v = 5)$ *for* $\alpha = 0.5$ *(dashed line)*, 0.75 *(dotted line)*, 1 *(solid line)*, 2 *(dotted-dashed line)*, 4 *(long-dashed line); (b)* $Pt(\alpha = 5, v)$ *for* $v = 5$ *(solid line)*, 10 *(dashed line)*, 15 *(dotted line)*, 20 *(dotted-dashed line)*, 25 *(long-dashed line); (c)* $Pt(\alpha = 0.5, v)$ *for* $v = 5$ *(solid line)*, 10 *(dashed line)*, 15 *(dotted line)*, 20 *(dotted-dashed line)*, 25 *(long-dashed line).*

$$t(x; v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}},$$

and we write $X \sim Pt(\alpha, v)$.

A special case of power t model occurs when $\alpha = 1$. That is, the t model $\phi_{\alpha=1,v}(x) = t(x; v)$ follows. Further, the power t model reduces to the power normal model (Kundu and Gupta, 2013) when the degree of freedom v goes to infinity. Figures 1 and 2 shown some pdfs of the skew-t and power t distributions.

Figures 1 and 2 depict the behavior of the skew-t and power t distributions for different values of λ, v and α, v respectively. The left plot of Figure 2 shows that parameter α affects both skewness and

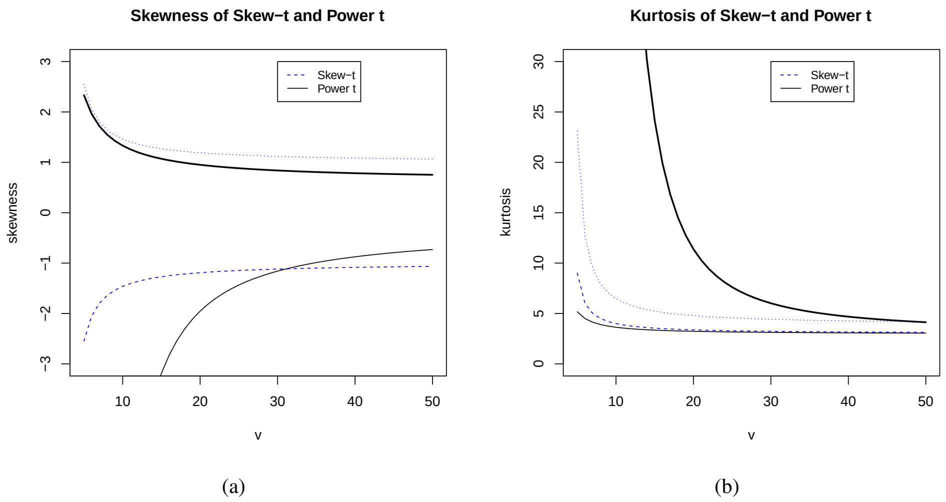


Figure 3: *Skewness and kurtosis of skew-t and power t models.*

kurtosis of the power t distribution.

For the moments of a random variable $X \sim Pt(\alpha, v)$, no closed form exists, but under a variable change, the r^{th} moment of the random variable X can be written as

$$E\left(X^r\right) = \alpha \int_0^1 \left[\Phi_v^{-1}(x)\right]^r x^{\alpha-1} dx, \tag{2.2}$$

where $\Phi_v^{-1}(x)$ is the inverse of the function $\Phi_v(x)$. This agrees with the expected value of the function $[\Phi_v^{-1}(x)]^r$ with random variable $X \sim \text{Beta}(\alpha, 1)$.

For a comparison of the coefficients of skewness and kurtosis between the skew-t and power t distribution under the v-values (all natural numbers from 5 to 50), α-values (ranging from 0.2 to 100), and λ-values (ranging between -50 and 50), see Figure 3.

Figure 3 shows that on the negative part of skewness, the skewness of the power t model surpasses that of skew-t distribution when v is less than 31 at the same degree of freedom. While on the positive part of skewness, the skew-t one surpasses that of power t model at the same degree of freedom. However, when v increases from 31 to 50, the skewness interval of the skew-t model contains those intervals for the power t model at the same degree of freedom. The kurtosis interval of the power t model surpasses the corresponding interval of the skew-t model for a given degree of freedom. The kurtosis intervals of the two models are almost the same when v is large; consequently, the power t model is more flexible than the skew-t model on kurtosis with the same degree of freedom.

2.2. Inference

Given a random sample $X_1, X_2, \ldots, X_n$ with $X_i \sim Pt(\alpha, v)$, the likelihood function for $\theta = (\alpha, v)'$ is

$$L(\theta, X) = \alpha^n (\pi v)^{-\frac{n}{2}} \left\{ \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \right\}^n \prod_{i=1}^{n} \left(1 + \frac{x_i^2}{v}\right)^{-\frac{v+1}{2}} \{\Phi_v(x_i)\}^{\alpha-1}$$

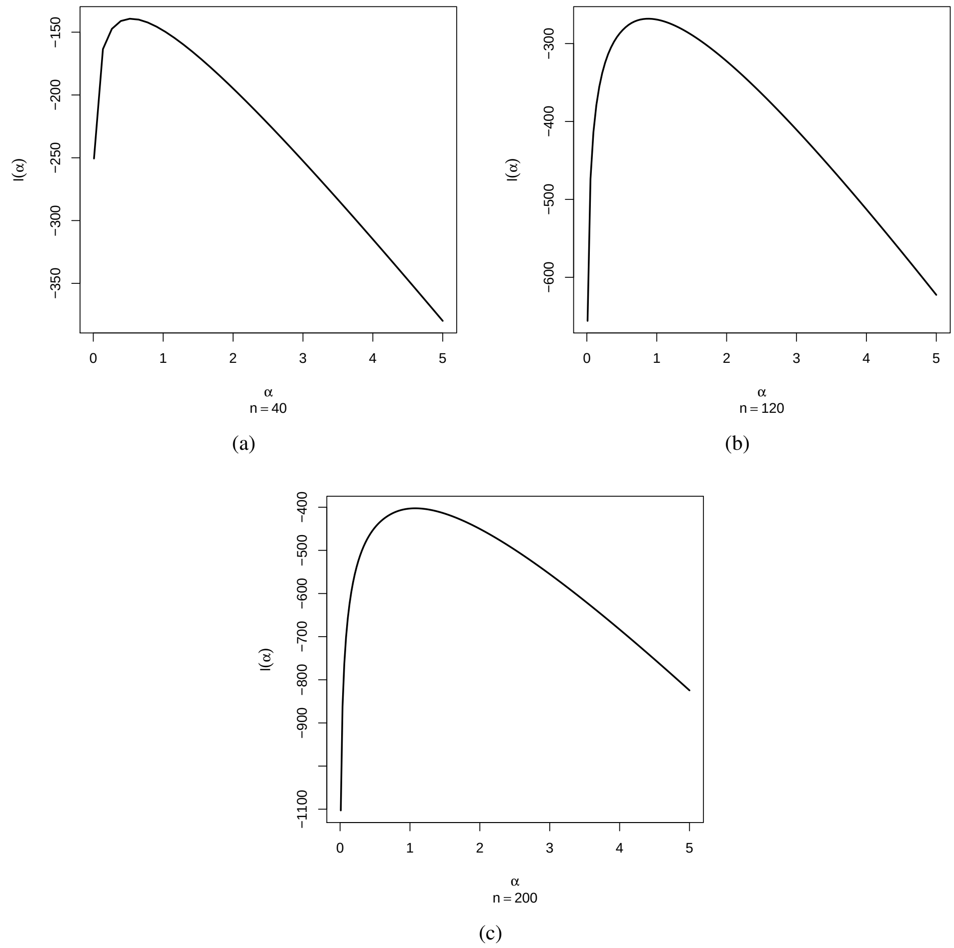


Figure 4: *Profile log-likelihood function* $l_p(\alpha)$*. Sample sizes are (a)* $n = 40$*, (b)* $n = 120$*, and (c)* $n = 200$*.*

and the log-likelihood function for $\theta = (\alpha, v)'$ is

$$l(\theta, X) = \text{constant} + n\log\alpha - \frac{n}{2}\log v + n\log\Gamma\left(\frac{v+1}{2}\right) - n\log\Gamma\left(\frac{v}{2}\right)$$
$$- \frac{v+1}{2}\sum_{i=1}^{n}\log\left(1 + \frac{x_i^2}{v}\right) + (\alpha - 1)\sum_{i=1}^{n}\log\Phi_v(x_i).$$

Figure 4 depicts the profile log-likelihood function of α ranging from 0.01 to 5 for the samples simulated from the distribution $Pt(\alpha, v = 5)$; the sample sizes are 40, 120, and 200, respectively. Obviously v is estimated to get the profile log-likelihood function of α.

Figure 5 presents the profile log-likelihood function of the degree of freedom v between 0.5 and 30 for the samples simulated from the distribution $Pt(\alpha = 3, v)$ and the sample sizes are the same as above. Also α is estimated to obtain Figure 5. These figures illustrate that the profiles are quite regular, providing strong evidence for the existence and unicity of estimators.

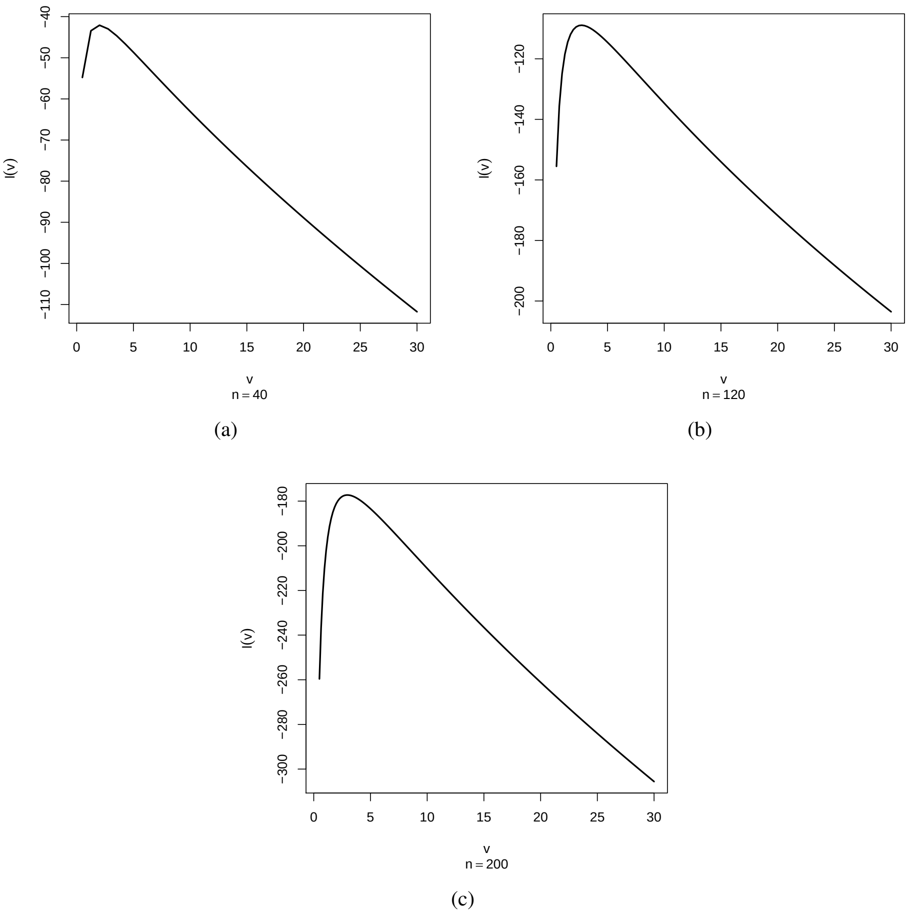


Figure 5: *Profile log-likelihood function $l_p(v)$. Sample sizes are (a) $n = 40$, (b) $n = 120$, and (c) $n = 200$.*

The components of the score function are:

$$U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n} \log \Phi_v(x_i),$$

$$U(v) = -\frac{n\alpha}{2v} + \frac{n\alpha}{2}\Psi\left(\frac{v+1}{2}\right) - \frac{n\alpha}{2}\Psi\left(\frac{v}{2}\right) - \frac{1}{2}\sum_{i=1}^{n}\log\left(1 + \frac{x_i^2}{v}\right)$$
$$+ \frac{v+1}{2v}\sum_{i=1}^{n}\frac{x_i^2}{v + x_i^2} + \frac{\alpha - 1}{2}\sum_{i=1}^{n}\frac{g(x_i, v)}{\Phi_v(x_i)},$$

where

$$\Psi(x) = \frac{d}{dx}\log\Gamma(x),$$
$$g(x_i, v) = \int_{-\infty}^{x_i}\left\{\frac{(v+1)m^2}{v(v+m^2)} - \ln\left(1 + \frac{m^2}{v}\right)\right\} t(m; v)dm.$$

By using some iterative procedures, we can obtain the maximum likelihood estimators for $\theta = (\alpha, v)'$. Besides, the elements of Hessian matrix are given by

$$\frac{\partial^2}{\partial\alpha^2} l(\theta, X) = -\frac{n}{\alpha^2},$$

$$\frac{\partial^2}{\partial\alpha\partial v} l(\theta, X) = \frac{1}{2}\left\{\Psi\left(\frac{v+1}{2}\right) - \frac{1}{v} - \Psi\left(\frac{v}{2}\right)\right\}\sum_{i=1}^{n}\Phi_v(x_i) + \frac{1}{2}\sum_{i=1}^{n} g(x_i; v),$$

$$\frac{\partial^2}{\partial v^2} l(\theta, X) = \frac{n\alpha}{2v^2} + \frac{n\alpha}{4}\Psi_1\left(\frac{v+1}{2}\right) - \frac{n\alpha}{4}\Psi_1\left(\frac{v}{2}\right) + \frac{v-1}{2v^2}\sum_{i=1}^{n}\frac{x_i^2}{v+x_i^2} - \frac{v+1}{2v}\sum_{i=1}^{n}\frac{x_i^2}{\left(v+x_i^2\right)^2} + \frac{\alpha-1}{2}\sum_{i=1}^{n}\frac{d}{dv}\frac{g(x_i, v)}{\Phi_v(x_i)},$$

where

$$\Psi_1(x) = \frac{d}{dv}\Psi(x),$$

$$\frac{d}{dv}\frac{g(x_i, v)}{\Phi(x_i)} = \frac{1}{\Phi_v(x_i)}\int_{-\infty}^{x_i}\frac{m^2\left(vm^2 - 2v - m^2\right)}{v^2\left(v+m^2\right)^2}t(m; v)dm - \frac{1}{2}\frac{g^2(x_i, v)}{\Phi_v^2(x_i)} + \frac{1}{2\Phi_v(x_i)}\int_{-\infty}^{x_i}\left\{\frac{(v+1)m^2}{v\left(v+m^2\right)} - \ln\left(1+\frac{m^2}{v}\right)\right\}^2 t(m; v)dm.$$

Since $I(\theta) = -H$, the observed information matrix can be obtained directly from the Hessian matrix. For $n = 1$, through some algebraic manipulations, the elements of the expected (Fisher) information matrix can be obtained as:

$$i_{\alpha\alpha} = \frac{1}{\alpha^2},$$

$$i_{\alpha v} = -\frac{1}{2}\left\{\Psi\left(\frac{v+1}{2}\right) - \frac{1}{v} - \Psi\left(\frac{v}{2}\right)\right\}E[\Phi_v(x)] - \frac{1}{2}E[g(x; v)],$$

$$i_{vv} = -\frac{\alpha}{2v^2} - \frac{\alpha}{4}\Psi_1\left(\frac{v+1}{2}\right) + \frac{\alpha}{4}\Psi_1\left(\frac{v}{2}\right) - \frac{v-1}{2v^2}E\left(\frac{x^2}{v+x^2}\right) + \frac{v+1}{2v}E\left(\frac{x^2}{\left(v+x^2\right)^2}\right) - \frac{\alpha-1}{2}E\left[\frac{d}{dv}\frac{g(x, v)}{\Phi_v(x)}\right].$$

Note that the observed and expected (Fisher) information matrices are nonsingular since no linear relationship exist between rows (columns). For the special case $\alpha = 1$ and $v = 5$, as an example, the expected (Fisher) information matrix can be given by

$$I_F(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & 0.00305754 \end{pmatrix}$$

and the determinant is

$$|I_F(\theta)| = 0.00305754 > 0.$$

Therefore, the expected information matrix is nonsingular.

2.3. Location-scale version

Let $X \sim Pt(\alpha, \nu)$ and take the transformation $Y = \xi + \eta X$, where $\xi \in \mathbb{R}$ and $\eta \in \mathbb{R}^+$, a location-scale extension is obtained. The pdf of Y can be written as

$$\phi_{\alpha,\nu}(y;\xi,\eta) = \alpha\phi_\nu(y;\xi,\eta)\left\{\Phi_\nu(y;\xi,\eta)\right\}^{\alpha-1}, \tag{2.3}$$

where

$$\phi_\nu(y;\xi,\eta) = \frac{1}{\eta}t\left(\frac{y-\xi}{\eta};\nu\right),$$
$$\Phi_\nu(y;\xi,\eta) = \int_{-\infty}^{y}\phi_\nu(m;\xi,\eta)dm,$$

and we write $Y \sim Pt(\xi,\eta,\alpha,\nu)$.

Given a random sample $Y_1, Y_2, \ldots, Y_n$ with $Y_i \sim Pt(\xi,\eta,\alpha,\nu)$, the likelihood function for $\theta = (\xi,\eta,\alpha,\nu)'$ is

$$L(\theta, Y) = \left(\frac{\alpha}{\eta}\right)^n (\pi\nu)^{-\frac{n}{2}}\left\{\Gamma\left(\frac{\nu+1}{2}\right)\right\}^n\left\{\Gamma\left(\frac{\nu}{2}\right)\right\}^{-n}\prod_{i=1}^{n}\left(1+\frac{x_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}\left\{\Phi_\nu(x_i)\right\}^{\alpha-1},$$

and the log-likelihood function $\theta = (\xi,\eta,\alpha,\nu)'$ is

$$l(\theta, Y) = \text{constant} + n\log\alpha - n\log\eta - \frac{n}{2}\log\nu + n\log\Gamma\left(\frac{\nu+1}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right)$$
$$- \frac{\nu+1}{2}\sum_{i=1}^{n}\log\left(1+\frac{x_i^2}{\nu}\right) + (\alpha-1)\sum_{i=1}^{n}\log\Phi_\nu(x_i),$$

where $x_i = (y_i - \xi)/\eta$.

We compute the first derivatives of the log-likelihood function, to obtain the elements of the score function as:

$$U(\xi) = \frac{\nu+1}{\eta}\sum_{i=1}^{n}\frac{x_i}{\nu+x_i^2} - \frac{\alpha-1}{\eta}\sum_{i=1}^{n}\frac{t(x_i;\nu)}{\Phi_\nu(x_i)},$$
$$U(\eta) = -\frac{n}{\eta} + \frac{\nu+1}{\eta}\sum_{i=1}^{n}\frac{x_i^2}{\nu+x_i^2} - \frac{\alpha-1}{\eta}\sum_{i=1}^{n}\frac{x_i t(x_i;\nu)}{\Phi_\nu(x_i)},$$
$$U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n}\log\Phi_\nu(x_i),$$
$$U(\nu) = -\frac{n\alpha}{2\nu} + \frac{n\alpha}{2}\Psi\left(\frac{\nu+1}{2}\right) - \frac{n\alpha}{2}\Psi\left(\frac{\nu}{2}\right) - \frac{1}{2}\sum_{i=1}^{n}\log\left(1+\frac{x_i^2}{\nu}\right) + \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\frac{x_i^2}{\nu+x_i^2} + \frac{\alpha-1}{2}\sum_{i=1}^{n}\frac{g(x_i,\nu)}{\Phi_\nu(x_i)},$$

where

$$\Psi(x) = \frac{d}{dx}\log\Gamma(x),$$
$$g(x_i,\nu) = \int_{-\infty}^{x_i}\left\{\frac{(\nu+1)m^2}{\nu(\nu+m^2)} - \ln\left(1+\frac{m^2}{\nu}\right)\right\}t(m;\nu)\,dm.$$

By equating the score functions to zero and solving the resulting equations, the maximum likelihood estimators for $\theta=(\xi,\eta,\alpha,\nu)'$ can be obtained. The elements of Hessian matrix are given by

$$\frac{\partial^2}{\partial\xi^2}l(\theta,Y)=-\frac{\nu+1}{\eta^2}\sum_{i=1}^{n}\frac{\nu-x_i^2}{\left(\nu+x_i^2\right)^2}-\frac{\alpha-1}{\eta^2}\sum_{i=1}^{n}\left\{\frac{x_i(\nu+1)}{\nu+x_i^2}\frac{t(x_i;\nu)}{\Phi_\nu(x_i)}+\frac{t^2(x_i;\nu)}{\Phi_\nu^2(x_i)}\right\},$$

$$\frac{\partial^2}{\partial\xi\partial\eta}l(\theta,Y)=-\frac{\nu+1}{\eta^2}\sum_{i=1}^{n}\frac{x_i}{\nu+x_i^2}-\frac{\nu+1}{\eta^2}\sum_{i=1}^{n}\frac{x_i\left(\nu-x_i^2\right)}{\left(\nu+x_i^2\right)^2}+\frac{\alpha-1}{\eta^2}\sum_{i=1}^{n}\frac{t(x_i;\nu)}{\Phi_\nu(x_i)}$$
$$-\frac{\alpha-1}{\eta^2}\sum_{i=1}^{n}x_i\left\{\frac{x_i(\nu+1)}{\nu+x_i^2}\frac{t(x_i;\nu)}{\Phi_\nu(x_i)}+\frac{t^2(x_i;\nu)}{\Phi_\nu^2(x_i)}\right\},$$

$$\frac{\partial^2}{\partial\xi\partial\alpha}l(\theta,Y)=-\frac{1}{\eta}\sum_{i=1}^{n}\frac{t(x_i;\nu)}{\Phi_\nu(x_i)},$$

$$\frac{\partial^2}{\partial\xi\partial\nu}l(\theta,Y)=\frac{1}{\eta}\sum_{i=1}^{n}\frac{x_i}{\nu+x_i^2}-\frac{\nu+1}{\eta}\sum_{i=1}^{n}\frac{x_i}{\left(\nu+x_i^2\right)^2}$$
$$-\frac{\alpha-1}{2\eta}\sum_{i=1}^{n}\left\{\left[\frac{(\nu+1)x_i^2}{\nu\left(\nu+x_i^2\right)}-\ln\left(1+\frac{x_i^2}{\nu}\right)\right]\frac{t(x_i;\nu)}{\Phi_\nu(x_i)}-\frac{t(x_i;\nu)g(x_i,\nu)}{\Phi_\nu^2(x_i)}\right\},$$

$$\frac{\partial^2}{\partial\eta^2}l(\theta,Y)=\frac{n}{\eta^2}-\frac{\nu+1}{\eta^2}\sum_{i=1}^{n}\frac{x_i^2}{\nu+x_i^2}-\frac{\nu+1}{\eta^2}\sum_{i=1}^{n}\frac{2\nu x_i^2}{\left(\nu+x_i^2\right)^2}+\frac{2(\alpha-1)}{\eta^2}\sum_{i=1}^{n}\frac{x_it(x_i;\nu)}{\Phi_\nu(x_i)}$$
$$-\frac{\alpha-1}{\eta^2}\sum_{i=1}^{n}x_i^2\left\{\frac{x_i(\nu+1)}{\nu+x_i^2}\frac{t(x_i;\nu)}{\Phi_\nu(x_i)}+\frac{t^2(x_i;\nu)}{\Phi_\nu^2(x_i)}\right\},$$

$$\frac{\partial^2}{\partial\eta\partial\alpha}l(\theta,Y)=-\frac{1}{\eta}\sum_{i=1}^{n}\frac{x_it(x_i;\nu)}{\Phi_\nu(x_i)},$$

$$\frac{\partial^2}{\partial\eta\partial\nu}l(\theta,Y)=\frac{1}{\eta}\sum_{i=1}^{n}\frac{x_i^2}{\nu+x_i^2}-\frac{\nu+1}{\eta}\sum_{i=1}^{n}\frac{x_i^2}{\left(\nu+x_i^2\right)^2}$$
$$-\frac{\alpha-1}{2\eta}\sum_{i=1}^{n}x_i\left\{\left[\frac{(\nu+1)x_i^2}{\nu\left(\nu+x_i^2\right)}-\ln\left(1+\frac{x_i^2}{\nu}\right)\right]\frac{t(x_i;\nu)}{\Phi_\nu(x_i)}-\frac{t(x_i;\nu)g(x_i,\nu)}{\Phi_\nu^2(x_i)}\right\},$$

$$\frac{\partial^2}{\partial\alpha^2}l(\theta,Y)=-\frac{n}{\alpha^2},$$

$$\frac{\partial^2}{\partial\alpha\partial\nu}l(\theta,Y)=\frac{1}{2}\left\{\Psi\left(\frac{\nu+1}{2}\right)-\frac{1}{\nu}-\Psi\left(\frac{\nu}{2}\right)\right\}\sum_{i=1}^{n}\Phi_\nu(x_i)+\frac{1}{2}\sum_{i=1}^{n}g(x_i;\nu),$$

$$\frac{\partial^2}{\partial\nu^2}l(\theta,Y)=\frac{n\alpha}{2\nu^2}+\frac{n\alpha}{4}\Psi_1\left(\frac{\nu+1}{2}\right)-\frac{n\alpha}{4}\Psi_1\left(\frac{\nu}{2}\right)+\frac{\nu-1}{2\nu^2}\sum_{i=1}^{n}\frac{x_i^2}{\nu+x_i^2}$$
$$-\frac{\nu+1}{2\nu}\sum_{i=1}^{n}\frac{x_i^2}{\left(\nu+x_i^2\right)^2}+\frac{\alpha-1}{2}\sum_{i=1}^{n}\frac{d}{d\nu}\frac{g(x_i,\nu)}{\Phi_\nu(x_i)},$$

where

$$\Psi_1(x) = \frac{d}{dv}\Psi(x),$$
$$\frac{d}{dv}\frac{g(x_i, v)}{\Phi(x_i)} = \frac{1}{\Phi_v(x_i)}\int_{-\infty}^{x_i}\frac{m^2\left(vm^2 - 2v - m^2\right)}{v^2\left(v + m^2\right)^2}t(m; v)dm - \frac{1}{2}\frac{g^2(x_i, v)}{\Phi_v^2(x_i)}$$
$$+ \frac{1}{2\Phi_v(x_i)}\int_{-\infty}^{x_i}\left\{\frac{(v+1)m^2}{v(v+m^2)} - \ln\left(1 + \frac{m^2}{v}\right)\right\}^2 t(m; v)dm.$$

The elements of the observed information matrix follows directly from the Hessian matrix, that is, $I(\theta) = -H$. For $n = 1$ the expected (Fisher) information matrix can be written as

$$i_{\xi\xi} = \frac{v+1}{\eta^2}E\left(\frac{v - x^2}{(v + x^2)^2}\right) + \frac{\alpha - 1}{\eta^2}E\left\{\frac{x(v+1)}{v + x^2}\frac{t(x; v)}{\Phi_v(x)} + \frac{t^2(x; v)}{\Phi_v^2(x)}\right\},$$
$$i_{\xi\eta} = \frac{v+1}{\eta^2}E\left(\frac{x}{v+x^2}\right) + \frac{v+1}{\eta^2}E\left[\frac{x\left(v - x^2\right)}{(v+x^2)^2}\right] - \frac{\alpha-1}{\eta^2}E\left[\frac{t(x; v)}{\Phi_v(x)}\right] + \frac{\alpha-1}{\eta^2}E\left\{x\left[\frac{x(v+1)}{v+x^2}\frac{t(x; v)}{\Phi_v(x)} + \frac{t^2(x; v)}{\Phi_v^2(x)}\right]\right\},$$
$$i_{\xi\alpha} = \frac{1}{\eta}E\left[\frac{t(x; v)}{\Phi_v(x)}\right],$$
$$i_{\xi v} = -\frac{1}{\eta}E\left(\frac{x}{v+x^2}\right) + \frac{v+1}{\eta}E\left[\frac{x}{(v+x^2)^2}\right] + \frac{\alpha-1}{2\eta}E\left\{\left[\frac{(v+1)x^2}{v\,(v+x^2)} - \ln\left(1 + \frac{x^2}{v}\right)\right]\frac{t(x; v)}{\Phi_v(x)} - \frac{t(x; v)g(x, v)}{\Phi_v^2(x)}\right\},$$
$$i_{\eta\eta} = -\frac{1}{\eta^2} + \frac{v+1}{\eta^2}E\left(\frac{x^2}{v + x^2}\right) + \frac{v+1}{\eta^2}E\left(\frac{2vx^2}{(v + x^2)^2}\right) - \frac{2(\alpha-1)}{\eta^2}E\left[\frac{xt(x; v)}{\Phi_v(x)}\right]$$
$$+ \frac{\alpha - 1}{\eta^2}E\left\{x^2\left[\frac{x(v+1)}{v + x^2}\frac{t(x; v)}{\Phi_v(x)} + \frac{t^2(x; v)}{\Phi_v^2(x)}\right]\right\},$$
$$i_{\eta\alpha} = \frac{1}{\eta}E\left[\frac{xt(x; v)}{\Phi_v(x)}\right],$$
$$i_{\eta v} = -\frac{1}{\eta}E\left(\frac{x^2}{v + x^2}\right)$$
$$+ \frac{v+1}{\eta}E\left[\frac{x^2}{(v + x^2)^2}\right] + \frac{\alpha - 1}{2\eta}E\left\{\left[\frac{(v+1)x^2}{v(v + x^2)} - \ln\left(1 + \frac{x^2}{v}\right)\right]\frac{t(x; v)x}{\Phi_v(x)} - \frac{t(x; v)g(x, v)x}{\Phi_v^2(x)}\right\},$$
$$i_{\alpha\alpha} = \frac{1}{\alpha^2},$$
$$i_{\alpha v} = -\frac{1}{2}\left\{\Psi\left(\frac{v+1}{2}\right) - \frac{1}{v} - \Psi\left(\frac{v}{2}\right)\right\}E[\Phi_v(x)] - \frac{1}{2}E[g(x; v)],$$
$$i_{vv} = -\frac{\alpha}{2v^2} - \frac{\alpha}{4}\Psi_1\left(\frac{v+1}{2}\right) + \frac{\alpha}{4}\Psi_1\left(\frac{v}{2}\right) - \frac{v-1}{2v^2}E\left(\frac{x^2}{v+x^2}\right) + \frac{v+1}{2v}E\left(\frac{x^2}{(v+x^2)^2}\right) - \frac{\alpha-1}{2}E\left[\frac{d}{dv}\frac{g(x, v)}{\Phi_v(x)}\right].$$

Since no linear relationship exists between rows (columns), the observed and expected (Fisher) information matrices are nonsingular. For example, with $\alpha = 1$, $v = 5$, the expected (Fisher) information

Table 1: Summary statistics for dataset

n	Mean	Variance	Skewness	Kurtosis
71	7.806	1.605	1.238	7.916

Table 2: Parameter estimates for skew-*t* and power *t* distributions

Parameter	Skew-*t* model estimate	Power *t* model estimate
Log-lik	−111.017	−110.942
AIC	230.034	229.884
CAIC	243.085	242.935
ξ	7.6197 (0.609)	7.61488 (0.134)
η	0.8947 (0.183)	0.94220 (0.104)
ν	3.9722 (2.141)	4.04600 (0.271)
λ	0.1765 (0.768)	-
α	-	1.16956 (0.159)

AIC = Akaike Information Criterion; CAIC = consistent Akaike Information Criterion; Log-lik = Log-likelihood.

matrix can be written as

$$I_F(\theta) = \begin{pmatrix} 0.7504\dfrac{1}{\eta^2} & 0 & 0.72832\dfrac{1}{\eta} & 0 \\ 0 & 1.2480\dfrac{1}{\eta^2} & -0.45634\dfrac{1}{\eta} & -0.04124\dfrac{1}{\eta} \\ 0.72832\dfrac{1}{\eta} & -0.45634\dfrac{1}{\eta} & 1 & 0 \\ 0 & -0.04124\dfrac{1}{\eta} & 0 & 0.00306 \end{pmatrix}$$

for which the determinant is given by $|I_F(\theta)| = 0.000268/\eta^4 > 0$. Thus, the expected Fisher information matrix is nonsingular.

3. Application

An application is reported for the data set consisting of 71 total glycerol contents in Grignolino wine. The data are available by the R package “sn”; the summary statistics for the dataset is presented in Table 1.

The values of skewness and kurtosis imply a strong indication of asymmetry and heavy tail; therefore, an asymmetric and heavy-tailed model may provide a better fit for the data under study. Consequently, the skew-*t* and power *t* models are fitted to the dataset. Optimization of the log-likelihood function for the power *t* model is implemented using the Newton-Raphson method. Table 2 shows the maximum-likelihood estimates for each model. The initial estimate values of ξ, η, ν and α are mean, variance, kurtosis and skewness coefficients of data, respectively. The standard errors for the estimates of power *t* model in parenthesis are obtained by the inverse square root of the main diagonal elements of the observed information matrix. For the skew-*t* model, the estimates are obtained from the “selm” in R package “sn”. Figure 6 presents graphs for the fitted models; in addition, Figure 7 also shows further model fitting assessments provided by the P-P plot.

All standard errors of power *t* model are smaller than those of skew-*t* distribution in this dataset. Besides, the best model is the one with the smallest Akaike Information Criterion (AIC) or Consistent Akaike Information Criterion (CAIC) values. From the AIC and CAIC criteria, the power *t* model

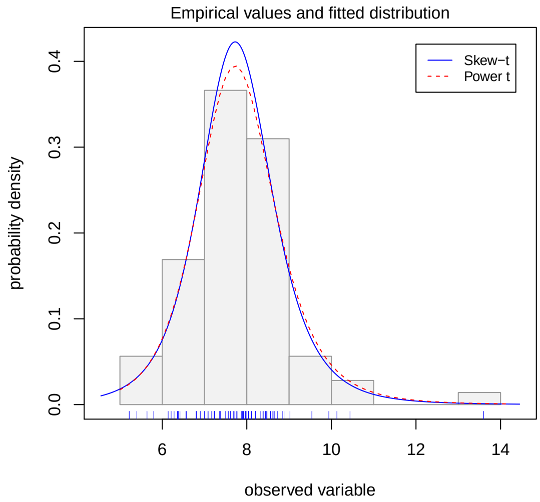


Figure 6: *Glycerol Contents in Grignolino.*

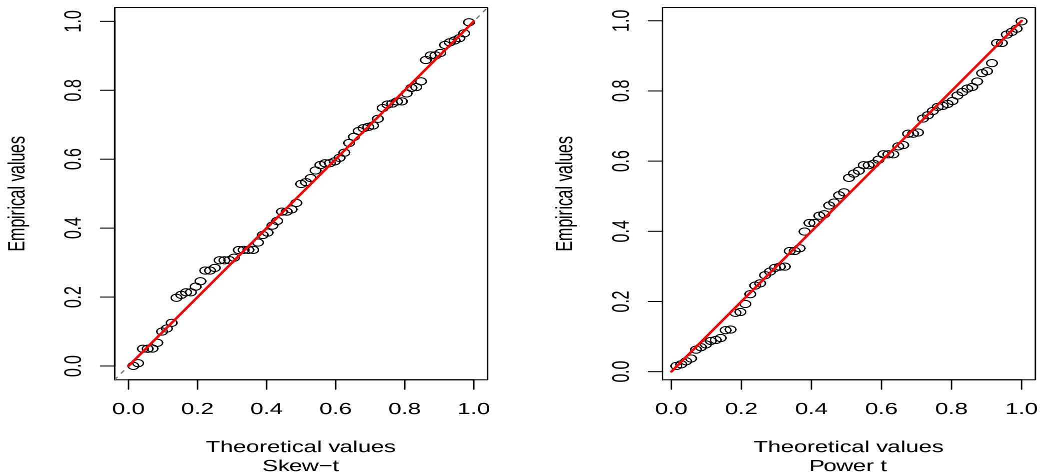


Figure 7: *P-P plots for skew-t and power t models.*

fits better for the dataset than the skew-t model, showing that the power t model can be quite useful in practice when datasets deviate from normality. We can say that the power t model is an alternative model to skew-t model in this data set because the AIC and CAIC values are slightly smaller at the power t distribution.

We also test the power t model against the t model. The chi-square limiting distribution for the likelihood ratio statistics is used since the Fisher information matrix was shown to be nonsingular. Formally, we have the following hypothesis:

$$H_0 : \alpha = 1 \quad \text{vs.} \quad H_1 : \alpha \neq 1;$$

this can be tested by using the statistics $\Lambda = l_t(\theta)/l_{pt}(\theta)$. After numerical evaluations, we obtain

$$-2\log(\Lambda) = 17.35675.$$

This is greater than the 5% chi-square critical value with one degree of freedom, $\chi^2_{1,5\%} = 3.8414$. Thus, the null hypothesis is rejected and the t model is not suitable for the above dataset.

4. Conclusion

This paper introduces a new asymmetric and heavy-tailed model as a generalization of the t distribution. At the same degree of freedom, the power t model presents a bigger range of skewness under some conditions and greater range of kurtosis compared to the skew-t distribution. The power t model can therefore handle more heavy-tailed data than skew-t model at the same degree of freedom. The Fisher information matrix derived is shown to be nonsingular and therefore guarantees valid large sample results for the likelihood ratio statistics. An application on a real dataset reveals that the model can be useful.

It is possible to develop the multivariate power t model based on the methods used in Martínez-Flórez *et al.* (2013) since power t model is a subclass of alpha-power model. We plan to do a future study on the identifiability of power t model referencing a previous paper by Otiniano *et al.* (2015) that analyzed the identifiability of finite mixture of skew-normal and skew-t distributions.

Acknowledgements

This work was supported by Konkuk University in 2015.

References

Arnold BC, Beaver RJ, Groeneveld RA, and Meeker WQ (1993). The nontruncated marginal of a truncated bivariate normal distribution, *Psychometrika*, **58**, 471–488.

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

Azzalini A (1986). Further results on a class of distributions which includes the normal ones, *Statistica*, **46**, 199–208.

Azzalini A and Capitanio A (2003). Distribution generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **65**, 367–389.

Branco MD and Dey DK (2001). A general class of multivariate skew-elliptical distributions, *Journal of Multivariate Analysis*, **79**, 93–113.

Durrans SR (1992). Distributions of fractional order statistics in hydrology, *Water Resources Research*, **28**, 1649–1655.

Gupta AK (2003). Multivariate skew t-distribution, *Statistics*, **37**, 359–363.

Gupta RD and Gupta RC (2008). Analyzing skewed data by power normal model, *Test*, **17**, 197–210.

Henze N (1986). A probabilistic representation of the 'skew-normal' distribution, *Scandinavian Journal of Statistics*, **13**, 271–275.

Jones MC (2001). A skew t distribution. In CA Charalambides et al. (Eds), *Probability and Statistical Models with Applications* (pp. 269–277), Chapman & Hall/CRC, Boca Raton, FL

Jones MC and Faddy MJ (2003). A skew extension of the t distribution, with application, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **65**, 159–174.

Jones MC and Larsen PV (2004). Multivariate distributions with support above the diagonal, *Biometrika*, **91**, 975–986.

Kundu D and Gupta RD (2013). Power-normal distribution, *Statistics*, **47**, 110–125.

Martínez-Flórez G, Arnold BC, Bolfarine H, and Gómez HW (2013). The multivariate alpha-power model, *Journal of Statistical Planning and Inference*, **143**, 1244–1255.

Otiniano CEG, Rathie PN and Ozelim LCSM (2015). On the identifiability of finite mixture of skew-normal and skew-*t* distributions, *Statistics and Probability Letters*, **106**, 103–108.

Pewsey A (2000). Problems of inference for Azzalini's skew-normal distribution, *Journal of Applied Statistics*, **27**, 859–870.

Pewsey A, Gómez HW, and Bolfarine H (2012). Likelihood-based inference for power distributions, *Test*, **21**, 775–789.

Sahu K, Dey DK, and Branco MD (2003). A new class of multivariate skew distributions with applications to Bayesian regression models, *Canadian Journal of Statistics*, **31**, 129–150.

Received May 24, 2016; Revised July 14, 2016; Accepted July 15, 2016