Scandinavian Journal of Statistics
Theory and Applications

Scandinavian Journal of Statistics
doi: 10.1111/j.1467-9469.2011.00779.x

Objective Bayesian Analysis of Skew-*t* Distributions

MARCIA D'ELIA BRANCO

Departamento de Estatistica, Universidade de São Paulo

MARC G. GENTON

Department of Statistics, Texas A&M University

BRUNERO LISEO

MEMOTEF, Sapienza Università di Roma

ABSTRACT. We study the Jeffreys prior and its properties for the shape parameter of univariate skew-*t* distributions with linear and nonlinear Student's *t* skewing functions. In both cases, we show that the resulting priors for the shape parameter are symmetric around zero and proper. Moreover, we propose a Student's *t* approximation of the Jeffreys prior that makes an objective Bayesian analysis easy to perform. We carry out a Monte Carlo simulation study that demonstrates an overall better behaviour of the maximum a *posteriori* estimator compared with the maximum likelihood estimator. We also compare the frequentist coverage of the credible intervals based on the Jeffreys prior and its approximation and show that they are similar. We further discuss location-scale models under scale mixtures of skew-normal distributions and show some conditions for the existence of the posterior distribution and its moments. Finally, we present three numerical examples to illustrate the implications of our results on inference for skew-*t* distributions.

Key words: approximation, Jeffreys prior, maximum *a posteriori* estimator, maximum likelihood estimator, skew-normal, skew-symmetric, skew-*t*

1. Introduction

Practical applied statistics reveals that many datasets exhibit both heavy tail and skewness behaviour, hence flexible non-Gaussian distributions are needed. To this end, we consider a class of skew-symmetric distributions (Azzalini & Capitanio, 2003; Ma & Genton, 2004; Wang *et al.*, 2004) whose probability density functions, up to location and scale parameters, are of the form

$$f(z \mid \alpha) = 2f_0(z)G(\alpha z), \quad z \in \mathbb{R}, \tag{1}$$

where $f_0(\cdot)$ is a symmetric density in $\mathbb{R}$, that is, $f_0(-z) = f_0(z)$ for all $z \in \mathbb{R}$, $G(\cdot)$ is a symmetric absolutely continuous cumulative distribution function, that is, $G(-z) = 1 - G(z)$ for all $z \in \mathbb{R}$, with density $g(\cdot)$ and $\alpha \in \mathbb{R}$ is a shape parameter controlling skewness. When $\alpha = 0$, the symmetric base density $f_0(\cdot)$ is retrieved, hence $G(\alpha z)$ is called a skewing function. Parameters of location, $\xi \in \mathbb{R}$, and of scale, $\omega > 0$, can be introduced through $Y = \xi + \omega Z$, where Z is a random variable with density (1). The class (1) contains many interesting distributions, see the book edited by Genton (2004) and the review article by Azzalini (2005) for a detailed account. For instance, the choice $f_0(z) = \phi(z)$ and $G(z) = \Phi(z)$, the standard normal density and cumulative distribution function, respectively, yields the skew-normal distribution, SN(α), introduced by Azzalini (1985). Picking $f_0(z) = t(z \mid \nu)$ and $G(z) = T(z \mid \nu)$, the standard Student's t density and cumulative distribution function with ν degrees of freedom, respectively, yields the linear skew-t distribution generated by lemma 1 of Azzalini (1985). If $G(\alpha z)$ in (1) is replaced by $T\{\alpha z\sqrt{(\nu+1)/(\nu+z^2)}\,|\,\nu+1\}$, then the nonlinear skew-*t* distribution of Branco & Dey (2001) and Azzalini & Capitanio (2003) is obtained. Other choices include setting $f_0(z)=\phi(z)$ while letting G be any symmetric cumulative distribution function that is different from the normal distribution (Nadarajah & Kotz, 2003), or setting $G(z)=\Phi(z)$ while letting f_0 be any symmetric density that is different from the normal density (Gómez *et al.*, 2007). In all those cases, frequentist inference on the shape parameter α based on the maximum likelihood paradigm is problematic as we explain next.

First, consider the case of the skew-normal distribution. It is well-known in the literature that the maximum likelihood estimator (MLE) of α can take the boundary values $\pm\infty$ with positive and non-negligible probability in small to moderate sample sizes, especially when the absolute value of α is large. Even though this probability vanishes for large samples (typically when the size is larger than about 400), it remains a problematic aspect for many common datasets. There exist currently three approaches to deal with this issue. The first one, proposed by Azzalini & Capitanio (1999) and called the deviance approach, is based on the fact that when α is large, the form of the skew-normal density remains almost unchanged if α is further increased. Therefore, the MLE of α can be replaced, in samples where it occurs on the boundary, by the smallest value α_0 such that the null hypothesis $H_0:\alpha=\alpha_0$ is not rejected by a likelihood ratio test based on the χ^2_1 distribution at a fixed level. Although this procedure is fairly easy to implement, it depends on a somewhat arbitrary choice of the level of the test, which needs to be unusually large to provide some partial solution to the boundary estimate issue. A second approach due to Sartori (2006) consists in constructing a modified score function as an estimating equation. The resulting modified MLE of the shape parameter of the skew-normal distribution has been proved to be always finite. A third approach is based on an objective Bayesian point of view. Liseo & Loperfido (2006) have shown that the Jeffreys prior and the reference prior for the scalar skew-normal shape parameter are proper and, consequently, the posterior mode of its posterior distribution, also called maximum a *posteriori* (MAP) estimator, is always finite. Although there is no closed form expression for the Jeffreys prior, Bayes & Branco (2007) proved that a remarkable approximation to it can be achieved by the use of a Student's *t* density.

Consider now the case of the (linear or nonlinear) skew-*t* distribution. An additional difficulty arises from the fact that (1) now also involves a degrees of freedom parameter ν. For the symmetric Student's *t* distribution, it is known that the likelihood function tends to infinity when ν goes to zero, see Fernández & Steel (1999). Even if the parameter space is restricted to (ν_0,∞) for $\nu_0>0$, the supremum of the likelihood function may be achieved when ν goes to infinity. Fonseca *et al.* (2008) gave a condition for the existence of the MLE of ν in that case. For the skew-*t* distribution, the deviance approach has been implemented by Azzalini & Genton (2008), where now the replacement of the MLE of (α,ν) is based on the null hypothesis $H_0:(\alpha,\nu)=(\alpha_0,\nu_0)$ and on a χ^2_2 distribution. However, simulation results have shown that this procedure provides only a partial solution to the problem. Alternatively, the modified score function approach has been applied to the skew-*t* distribution by Sartori (2006), although no proof of the finiteness of the resulting shape estimator has been provided. Moreover, that method requires the degrees of freedom parameter ν to be fixed. Our goal in this article is to provide an objective Bayesian solution to this problem. To this end, we shall derive the Jeffreys prior of the shape parameter for the linear and nonlinear skew-*t* distributions.

The shape parameter α in the density (1) occurs only in the skewing function G, hence the log-likelihood function associated with one observation is proportional to $\log G(\alpha z)$. The Fisher information for the shape parameter is then

$$I(\alpha)=\int_{-\infty}^{+\infty} 2z^2\frac{g^2(\alpha z)}{G^2(\alpha z)}f_0(z)G(\alpha z)\,\mathrm{d}z=\int_0^{+\infty} 2z^2 f_0(z)\{\pi h(\alpha z)\}^2\,\mathrm{d}z, \tag{2}$$

where

$$h(z)=\frac{1}{\pi}\frac{g(z)}{\sqrt{G(z)\{1-G(z)\}}}. \tag{3}$$

Finally, the Jeffreys prior for α is $\pi^J(\alpha)\propto\{I(\alpha)\}^{1/2}$.

The rest of the article is organized as follows. In section 2.1 we present an approximation of the density $h(z)$ in (3). Then, in the rest of section 2, we discuss properties of the Jeffreys priors for the shape parameter of the linear and nonlinear skew-t distributions. In both cases, we show that the resulting priors for the shape parameter are proper. We also propose a Student's t approximation of the Jeffreys prior that makes an objective Bayesian analysis easy to perform. In section 3.1, we perform a Monte Carlo simulation study that demonstrates an overall better behaviour of the MAP estimator compared with the MLE. We also compare the frequentist coverage of the credible intervals based on the Jeffreys prior and its approximation in section 3.2. We extend the discussion to location-scale models in section 4 and consider scale mixtures of skew-normal distributions. Finally, we present three numerical examples in section 5 to illustrate the implications of our results on inference for skew-t distributions. We end the article with a discussion in section 6. The proofs of the theoretical results are deferred to the Appendix.

2. The Jeffreys prior for the shape parameter of skew-*t* distributions

2.1. A useful approximation

In this section, we study a useful approximation of the density $h(z)$ defined by (3). Consider a random variable $X\sim\text{Beta}(1/2,1/2)$ and let G be a cumulative distribution function of an absolutely continuous symmetric random variable, that is, $G(-z)=1-G(z)$ for all $z\in\mathbb{R}$. Then the random variable $Z=G^{-1}(X)$ has a probability density function given by (3). The function h is a valid density with corresponding cumulative distribution function H given by

$$H(z)=2\arcsin\left\{\sqrt{G(z)}\right\},\quad z\in\mathbb{R}.$$

A possible approximation of h in terms of g can be obtained by choosing the scale parameter σ such that $h(z)\approx g(z/\sigma)/\sigma$.

The quality of the approximation of course depends on the thickness of the tails of g. For illustration, we explore the case when $g(z)=t(z\mid\nu)$, the Student's t density with ν degrees of freedom. Figure 1 depicts the corresponding density $h(z)$ (solid curve) and its approximation (dashed curve) for $\nu=1,5,10,100$. The approximations could also be compared more systematically using quantiles. Alternatively, one could choose the value of σ that minimizes the Kullback–Leibler distance between $h(z)$ and $g(z/\sigma)/\sigma$. For the normal distribution, the best approximation is obtained at $\tilde{\sigma}=1.54$, not so much different from $\pi/2=1.571$, the scale used by Bayes & Branco (2007).

When $G=\Phi$, Bayes & Branco (2007) noticed that

$$\frac{\phi(z)}{\sqrt{\Phi(z)\{1-\Phi(z)\}}}\approx 2\phi(2z/\pi).$$

Using the aforementioned results, they obtained an approximated version of $I(\alpha)$, namely

$$I(\alpha)\approx\frac{2}{\pi}\left(1+\frac{2\alpha^2}{\pi^2/4}\right)^{-3/2}.$$

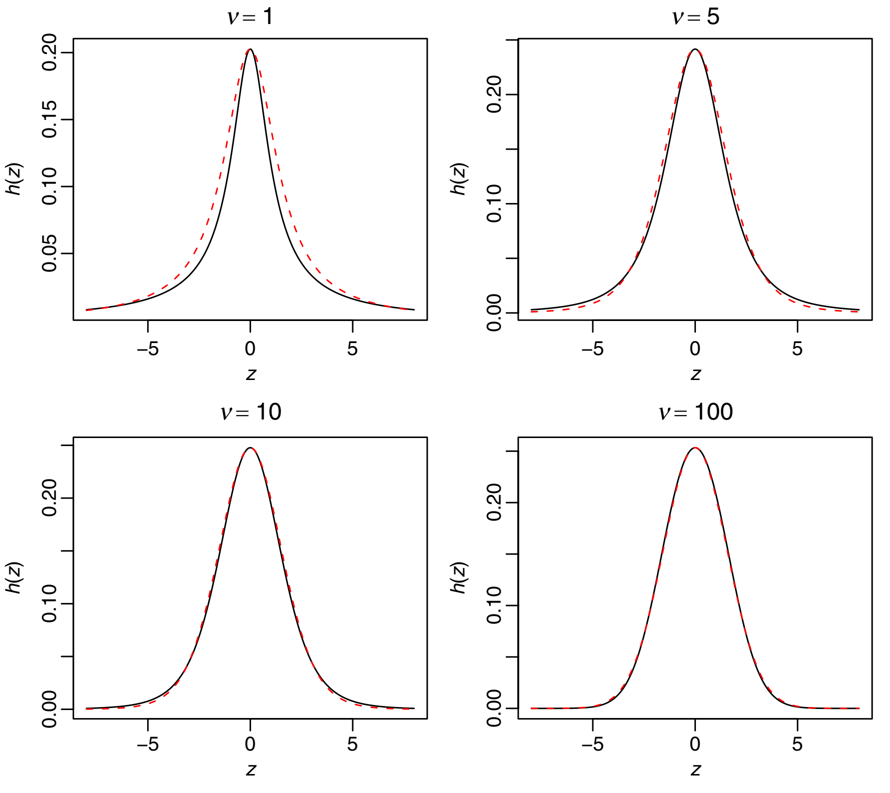

Fig. 1. Density $h(z)=g(z)/[\pi\sqrt{G(z)\{1-G(z)\}}]$ (solid curve) based on the Student's t density with v degrees of freedom, $g(z)=t(z\,|\,v)$, and its approximation, $g(z/\sigma)/\sigma$, with $\sigma=\pi/2$ (dashed curve) for $v=1, 5, 10, 100$.

One can then easily see that the Jeffreys prior for the shape parameter α, obtained by Liseo & Loperfido (2006), can be approximated by

$$\sqrt{\frac{2}{\pi}}\left(1+\frac{2\alpha^2}{\pi^2/4}\right)^{-3/4}, \tag{4}$$

which is the density of a Student's t distribution with mean zero, scale $\pi/2$ and degrees of freedom 1/2, denoted by $t(0, \pi/2, 1/2)$. Liseo & Loperfido (2006) showed that the Jeffreys prior for the shape parameter of the skew-normal distribution is symmetric around zero and the tails are of order $O(\alpha^{-3/2})$, hence it is a proper prior. We will extend this result to linear and nonlinear skew-t distributions in the next sections.

It is worth noting that an alternative approximation to the Jeffreys prior given in Liseo & Loperfido (2006) can be obtained using a different parameterization. Instead of using the natural shape parameter α, one can use the quantity $\delta=\alpha/\sqrt{1+\alpha^2}$ whose range is bounded to $[-1, 1]$. Denote $\beta=(\delta-1)/2$. It can be easily checked that, adopting a symmetric Beta prior for β, that is, $\beta\sim\text{Beta}(\tau,\tau)$, results in the following Student's t prior for α:

$$\tilde{\pi}(\alpha)=\frac{1}{2^{2\tau-1}}\frac{\Gamma(2\tau)}{\Gamma(\tau)\Gamma(\tau)}\frac{1}{(1+\alpha^2)^{\tau+1/2}}. \tag{5}$$

If $\tau=1/2$, then (5) results in a standard Cauchy prior for α. The Beta(1/2, 1/2) is the Jeffreys prior for the binomial model. The correct $O(\alpha^{-3/2})$ order for the tails is obtained for $\tau=1/4$.

In this case, the prior for α is a Student's t distribution, $t(0, \sqrt{2}, 1/2)$, quite similar to the aforegiven approximation (4).

2.2. Linear skew-t distributions

We consider the family of skew-symmetric distributions (1) with $f_0(z) = t(z \mid v)$ and a linear Student's t skewing function with v degrees of freedom, that is, the linear skew-t distribution:

$$f(z \mid \alpha, v) = 2t(z \mid v)T(\alpha z \mid v), \quad z \in \mathbb{R}. \tag{6}$$

Proposition 1. *The Jeffreys prior of the shape parameter α of the linear skew-t distribution with density* (6) *has the following properties:*

(i) *it is symmetric about zero;*
(ii) *it has tails of order $O(\alpha^{-3/2})$; and*
(iii) *it is proper and given by*

$$\pi^J(\alpha \mid v) = \frac{2}{C(v)} \left\{ \int_{-\infty}^{+\infty} z^2 t(z \mid v) \frac{t^2(\alpha z \mid v)}{T(\alpha z \mid v)} \, \mathrm{d}z \right\}^{1/2}, \tag{7}$$

where $C(v)$ is a normalization constant depending on v.

The proof is given in the Appendix.

The approximation given in the skew-normal case suggests that a Student's t density with tails $O(\alpha^{-3/2})$ may also be a good approximation for $\pi^J(\alpha \mid v)$. To obtain the same tail behaviour and symmetry around zero, we fix the location and degrees of freedom parameters of the Student's t distribution to be zero and 1/2, respectively, and then we optimize the choice with respect to the scale parameter. That is, we consider the density $t(\alpha \mid 0, \sigma, 1/2)$ as a potential approximation. We note that, evaluating $\pi^J(\alpha \mid v)$ at $\alpha = 0$, one obtains

$$\pi^J(0 \mid v) = \frac{2\Gamma\{(v+1)/2\}}{C(v)\Gamma\left(v/2\right)\sqrt{v\pi}} \left\{ \int_{-\infty}^{+\infty} z^2 t(z \mid v) \, \mathrm{d}z \right\}^{1/2}. \tag{8}$$

This implies that, for $v > 2$, the prior (8) is well defined at zero; the integral in (8) is simply the variance of a standard Student's t distribution, which is $v/(v-2)$, for $v > 2$. The density of a Student's t distribution with mean zero and degrees of freedom 1/2, evaluated at zero, is

$$t(0 \mid 0, \sigma, 1/2) = \frac{1}{\sigma} \frac{\Gamma(3/4)}{\Gamma(1/4)} \sqrt{\frac{2}{\pi}}. \tag{9}$$

By equating (8) and (9), we obtain

$$\sigma(v) = C(v) \frac{\Gamma(3/4)\Gamma(v/2)}{\Gamma(1/4)\Gamma\{(v+1)/2\}} \sqrt{\frac{v-2}{2}}.$$

The function $C(v)$ needs to be evaluated numerically. Table 1 shows some values of $\sigma(v)$ as a function of v. When $v \le 2$, $\pi^J(\alpha \mid v)$ has a pole at zero and the aforesaid approximation does not work. However, when $v = 1$ or 2, the cumulative distribution function of a standard Student's t distribution is available in closed form, and simpler and more accurate approximations are possible in these cases.

Figure S1 in the Supporting Information compares the Jeffreys prior (7) with its Student's t approximation for different values of v. The approximation clearly improves for

Table 1. *Some values of* $\sigma(v)$ *as a function of* v *for the Student's t approximation of the Jeffreys prior for the shape parameter of the linear skew-t distribution*

v	3	5	10	100	∞
$\sigma(v)$	0.883	1.184	1.376	1.540	$1.570=\pi/2$

large values of v. However, even for small values, the approximation is fairly good. Therefore, we suggest the use of the Student's t density $t(\alpha\,|\,0,\sigma(v),1/2)$ as an approximation of the Jeffreys prior of the shape parameter α for the linear skew-t distributions defined by (6).

2.3. *Reference priors for models with latent structure*

Reference priors have been proposed by Bernardo (1979) and further developed by Berger & Bernardo (1989, 1992) and Berger *et al.* (2009). They represent a generalization of the Jeffreys priors: in fact the reference prior explicitly depends upon the order of interest among the parameters of the model. If one is interested in the entire parameter vector, the two approaches would provide the same prior. Recently, Liseo *et al.* (2010) have proposed an alternative approach to produce approximated reference priors when the model at hand can be expressed in terms of a latent structure. Given a parametric model $\mathcal{P}=\{p(\cdot\,|\,\theta);\theta\in\Theta\}$, with $\Theta\subset\mathbb{R}^d$, it may happen that the expected Fisher information matrix $I(\theta)$, with

$$I_{rs}(\theta)=-E\left\{\frac{\partial^2}{\partial\theta_r\theta_s}\ell(\theta)\right\},\quad r,s=1,\ldots,d,$$

and where $\ell(\theta)$ is the log-likelihood function, is difficult to obtain in a closed form. In some of these cases, one can resort to the completion idea, already exploited in Bayesian computational strategies; see, for example, Robert (2001). This amounts to write our model, for appropriate choices of g and h, as

$$p(\cdot\,|\,\theta)=\int g(\cdot\,|\,v,\theta)h(v\,|\,\theta)\,\mathrm{d}v, \tag{10}$$

where v is a suitable vector of latent variables. Then one can pretend that the newly introduced vector v represents an additional parameter to be estimated. In this respect, a version of the reference prior algorithm in the presence of partial information, as illustrated in Sun & Berger (1998), can be used to derive the desired prior. If the function $h(v\,|\,\theta)$ in (10) actually depends on θ then one needs to calculate the marginal prior for θ. On the other hand, if $h(v\,|\,\theta)$ does not depend on θ, then the conditional prior for θ given v should be the object of interest. Liseo *et al.* (2010) have shown that the reference priors obtained following this route very often (but not always!) coincide with those obtained in the direct way. In the next section, we use this approach to derive an approximation to the reference prior in the nonlinear skew-t case.

2.4. *Nonlinear skew-t distributions*

We consider the family of skew-symmetric distributions (1) with $f_0(z)=t(z\,|\,v)$ and a nonlinear Student's t skewing function with v degrees of freedom, that is, the nonlinear skew-t distribution:

$$f(z\,|\,\alpha,v)=2t(z\,|\,v)T\left(\alpha z\sqrt{\frac{v+1}{v+z^2}}\,\middle|\,v+1\right),\quad z\in\mathbb{R}. \tag{11}$$

Proposition 2. *For any fixed and known positive value of v, the Jeffreys prior of the shape parameter α of the nonlinear skew-t distribution with density* (11) *has the following properties:*

(i) *it is symmetric about zero;*
(ii) *it has tails of order $O(\alpha^{-3/2})$; and*
(iii) *it is proper and given by*

$$\pi^J(\alpha \mid v) \propto \left[\int_0^\infty z^2 t(z \mid v) \frac{t^2(\alpha z \mid v)}{T\left(\alpha z \sqrt{\frac{v+1}{v+z^2}} \,\middle|\, v+1\right) \left\{1 - T\left(\alpha z \sqrt{\frac{v+1}{v+z^2}} \,\middle|\, v+1\right)\right\}} \mathrm{d}z \right]^{1/2}. \quad (12)$$

The proof is given in the Appendix.

When v is also unknown we propose two different routes: the first one is to adopt the Jeffreys prior (12) as the conditional prior for $\alpha \mid v$ and then use the Jeffreys prior for v derived by Fonseca *et al.* (2008) in the symmetric Student's t case. Alternatively, one can follow the approach described in section 2.3 to find the approximated reference prior for the shape parameter of the distribution given by (11). The direct calculation of the reference prior from Fisher's information of the model is quite difficult. We then assume to observe a random sample $Z_1, \ldots, Z_n$ from a scale mixture of skew-normal distributions. Then, for given values of α and V_j, it can be easily seen that Z_j has a skew-normal density with scale parameter $V_j^{-1/2}$ and shape $\alpha, j=1,\ldots,n$. Therefore, the log-likelihood function for α and the vector $v=(v_1,\ldots,v_n)$ is

$$\ell(\alpha, v \mid z) \propto \sum_{j=1}^{n} \left\{ \log \phi(v_j^{1/2} z_j) + \frac{1}{2} \log v_j + \log \Phi(\alpha v_j^{1/2} z_j) \right\}. \quad (13)$$

In this model the function $h(v \mid \theta)$ in (10) does not depend on θ (i.e. α in our particular case) so we need to find the conditional reference prior for α given v. To apply the algorithm proposed in Liseo *et al.* (2010) one needs to calculate the element $I_{\alpha,\alpha}$ of the Fisher information matrix. In the Appendix, we show that $I_{\alpha,\alpha}$ can be approximated by $(1+8\alpha^2/\pi^2)^{-3/2}$, the square root of which is proportional to the conditional prior for $\alpha \mid v$. We recognize the kernel of a Student's t distribution with mean zero, scale $\pi/2$ and degrees of freedom 1/2. This is exactly the same prior approximation as obtained in section 2.1 for the linear skew-normal distribution. This prior has many noteworthy characteristics: it does not depend on the latent vector v, hence it is also the marginal reference prior for α. This remark in turn implies that the marginal prior for α does not depend on the value of v. Since the closed form expression of the Fisher information matrix is too complicated in this case, we will use this prior as an approximation of the true Jeffreys prior for the shape parameter for any skew-t model with fixed and known value of v. Figure S2 in the Supporting Information depicts the Jeffreys prior and its Student's t approximation for the nonlinear skew-t model with $v=2, 4, 10$.

3. Monte Carlo simulation study

3.1. Comparisons with MLEs

In this section we perform a systematic comparison between maximum likelihood and objective Bayesian estimation strategies for the nonlinear skew-t distribution via Monte Carlo simulations. We fix the location and scale parameters to $\xi=0$ and $\omega=1$, and consider several values of the shape parameter α and degrees of freedom v, and three different sample sizes, namely $n=50, 100$ and 500. For each sample, we calculate the MLE and the MAP estimator. The posterior density is obtained by considering independence between α and v and the Student's t approximation to the Jeffreys prior for α

$$\alpha \sim t(0, \pi/2, 1/2). \tag{14}$$

For v, we use the Jeffreys prior obtained by Fonseca *et al.* (2008) for the symmetric Student's t distribution

$$\pi^J(v) \propto \left(\frac{v}{v+3}\right)^{1/2} \left\{\psi_3\left(\frac{v}{2}\right) - \psi_3\left(\frac{v+1}{2}\right) - \frac{2(v+3)}{v(v+1)^2}\right\}^{1/2},$$

where ψ_3 is the trigamma function, the second derivative of the logarithm of the Gamma function. We use $R=500$ replicates for each combination of parameter values and sample size. To evaluate the simulation results, we compute the mean bias (MB)

$$\text{MB} = \frac{1}{R}\sum_{i=1}^{R}(\hat{\alpha}_i - \alpha),$$

and the root mean squared error (RMSE)

$$\text{RMSE} = \left\{\frac{1}{R}\sum_{i=1}^{R}(\hat{\alpha}_i - \alpha)^2\right\}^{1/2}.$$

Moreover, for the degrees of freedom parameter we consider the relative mean bias (RELMB)

$$\text{RELMB} = \frac{1}{R}\sum_{i=1}^{R}\frac{\hat{v}_i - v}{v},$$

and the root relative mean squared error (RRELMSE)

$$\text{RRELMSE} = \left\{\frac{1}{R}\sum_{i=1}^{R}\frac{(\hat{v}_i - v)^2}{v^2}\right\}^{1/2}.$$

Tables 2 and S1 report the MB and the RMSE, respectively, of the MLE and MAP estimators of the shape parameter α for all samples (i.e. including infinite MLEs). These results show that the MAP is superior to the MLE in terms of those measures. The proportions of infinite MLEs are reported in parentheses. They decrease when the sample size becomes larger. Figure 2 depicts the RMSE of the MLE and MAP estimators of α only for the samples with finite MLEs when $n=50$. Here too, we see a better behaviour of the MAP estimator compared with the MLE.

Tables 3 and S2 report the RELMB and the RRELMSE, respectively, of the MLE and MAP estimators of the degrees of freedom parameter v for samples with finite MLEs. These results show that the MAP is superior to the MLE in terms of those measures. Moreover, it is interesting to note the sign of the bias: whereas it is most of the time positive for the MLE, in many cases it is negative for the MAP. This can be justified by the influence of the prior distribution of v, which is concentrated around small values of the parameter. In addition, the RELMB and RRELMSE of the MAP estimator are fairly constant across the values of α. Figure 2 also depicts the RRELMSE of the MLE and MAP estimators of v only for the samples with finite MLEs when $n=50$. Here too, we see a generally better behaviour of the MAP estimator compared with the MLE. However, for $v=30$, the MAP and MLE estimators are somewhat similar.

Table 2. *Mean bias (MB) for the maximum likelihood estimator(MLE) and maximum* a posteriori *(MAP) estimator of the shape parameter α (all samples) of a nonlinear skew-t distribution. Percentages next to ∞ indicate proportion of infinite MLEs*

		α				
ν	Estimator	0	1	3	5	7
$n=50$						
2	MLE	0.0086	0.0752	0.5173	∞ (3.6%)	∞ (9.8%)
	MAP	0.0084	0.0166	−0.0944	−0.3118	−0.9125
8	MLE	0.0090	0.0445	0.5293	∞ (4.4%)	∞ (8.4%)
	MAP	0.0083	0.0368	0.0483	0.00817	−0.6523
15	MLE	0.0028	0.0678	0.6801	∞ (4.6%)	∞ (11.4%)
	MAP	0.0041	0.0662	0.1621	0.0447	−0.3443
30	MLE	−0.0142	0.0954	0.5082	∞ (3.0%)	∞ (11.0%)
	MAP	−0.0145	0.0969	0.0889	0.0316	−0.4228
$n=100$						
2	MLE	0.0002	0.0378	0.2303	0.7536	1.3662
	MAP	0.0003	0.0116	−0.0282	−0.1108	−0.3640
8	MLE	0.0094	0.0292	0.2114	0.7657	1.5935
	MAP	−0.0095	0.0355	0.0681	0.1160	−0.0212
15	MLE	−0.0140	0.0237	0.2684	0.6627	∞ (1.2%)
	MAP	−0.0148	0.0411	0.1461	0.1157	0.1276
30	MLE	−0.0052	0.0396	0.2759	0.6361	∞ (1.2%)
	MAP	−0.0053	0.0572	0.1510	0.1257	0.4828
$n=500$						
2	MLE	−0.0026	−0.0062	0.0508	−0.2588	0.1893
	MAP	−0.0026	−0.0110	0.0172	−0.0541	−0.1486
8	MLE	−0.0021	0.0094	0.0755	0.0754	0.2798
	MAP	−0.0021	0.0122	0.0598	0.0054	0.0830
15	MLE	0.0006	0.0069	0.0277	0.1098	0.2680
	MAP	0.0006	0.0155	0.0332	0.0469	0.1215
30	MLE	0.0018	0.0098	0.0551	0.0974	0.2515
	MAP	−0.0018	0.0227	0.0563	0.0535	0.1389

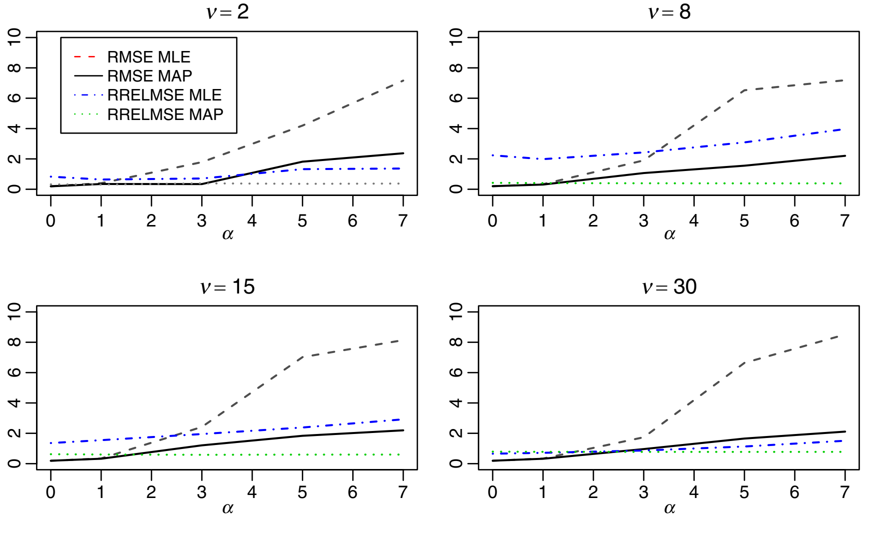

Fig. 2. Root mean squared error (RMSE) of maximum likelihood estimator (MLE) and maximum *a posteriori* (MAP) estimator of α and root relative mean squared error (RRELMSE) of MLE and MAP estimator of ν for the nonlinear skew-*t* distribution only for the samples with finite MLEs when $n=50$.

Table 3. *Relative mean bias (RELMB) for the maximum likelihood estimator (MLE) and maximum* a posteriori *(MAP) estimator of the degrees of freedom parameter ν (only samples with finite MLE) of a nonlinear skew-t distribution*

		α				
ν	Estimator	0	1	3	5	7
$n=50$						
2	MLE	0.1793	0.1820	0.2000	0.2952	0.2970
	MAP	−0.0168	0.0125	0.0405	0.0556	0.0370
8	MLE	1.0955	0.9187	1.1785	1.4064	1.8737
	MAP	−0.3786	−0.3667	−0.3408	−0.3452	−0.3324
15	MLE	0.5904	0.6869	1.0125	1.1874	1.5099
	MAP	−0.6152	−0.5930	−0.5724	−0.5856	−0.5891
30	MLE	−0.0307	0.0351	0.1583	0.3580	0.6431
	MAP	−0.7891	−0.7775	−0.7736	−0.7769	−0.7749
$n=100$						
2	MLE	0.0777	0.0718	0.0465	0.0897	0.1044
	MAP	0.0092	0.0130	−0.0027	0.0316	0.0159
8	MLE	0.9348	0.7902	0.7328	0.9318	0.8471
	MAP	−0.2125	−0.1757	−0.1755	−0.1803	−0.2032
15	MLE	0.8576	0.7322	0.7956	0.8442	1.1217
	MAP	−0.4743	−0.4358	−0.4358	−0.4305	−0.4416
30	MLE	0.4246	0.2181	0.1215	0.2894	0.5397
	MAP	−0.6916	−0.6688	−0.6759	−0.6739	−0.6753
$n=500$						
2	MLE	0.0141	0.0124	0.0155	−0.0561	0.0021
	MAP	0.0028	0.0025	0.0052	0.0016	−0.0035
8	MLE	0.0770	0.0932	0.1132	0.0778	0.0967
	MAP	−0.0404	−0.0190	−0.0119	−0.0330	−0.0372
15	MLE	0.3028	0.2563	0.2225	0.1837	0.2728
	MAP	−0.1324	−0.1281	−0.1303	−0.1458	−0.1388
30	MLE	0.2597	0.3381	0.2058	0.1326	0.1783
	MAP	−0.3854	−0.3617	−0.3639	−0.3863	−0.3923

3.2. Frequentist evaluation of the Jeffreys prior

The Jeffreys prior is probably the most well-known objective prior. In the one parameter case it is a second-order matching prior, that is, the one tail posterior credible sets constructed via the use of the Jeffreys prior have the exact frequentist coverage up to a $o(n^{-1})$ order. This optimality property is not guaranteed for multiparameter models but the frequentist behaviour is typically more than satisfactory. In this section, we explore the frequentist behaviour of the Bayesian procedure based on the Jeffreys prior in the case of the nonlinear skew-*t* distribution. In particular, we study, through simulations, the frequentist coverage probability of the one tail credible sets based on the use of the Jeffreys prior (12) and on its Student's *t* approximation (14). Suppose the model contains a parameter of interest θ and a nuisance parameter λ. The frequentist coverage probability at level 100γ is defined as $P(\theta \leq \theta_\gamma; \theta, \lambda)$, that is, the probability that θ_γ, the posterior quantile obtained with a specific prior, is larger than the actual θ. A prior with good frequentist properties should have a value of the frequentist coverage very close to γ. Table 4 shows the results of a simulation for several values of α and ν and $n=30$. Although in the previous simulations we have used larger values of the sample size, a comparison among priors should be made with a small value of n otherwise the effect of the prior tends to be negligible. Each entry in the table is the result of 1000 simulated samples. For each sample a Metropolis Hastings algorithm has been used to produce a sample from the joint posterior distribution of (α, ν). We have used

Table 4. *Frequentist coverage probabilities of* 0.05 *and* 0.95 *posterior quantiles using the Jeffreys prior,* π^J*, and its approximation,* $\tilde{\pi}^J$*, for the shape parameter* α *of the nonlinear skew-t distribution and sample size* $n=30$

	$\gamma=0.05$		$\gamma=0.95$	
(α, ν)	π^J	$\tilde{\pi}^J$	π^J	$\tilde{\pi}^J$
(0, 2)	0.044	0.041	0.941	0.944
(0, 8)	0.041	0.039	0.943	0.943
(0, 15)	0.041	0.040	0.943	0.940
(3, 2)	0.040	0.040	0.935	0.934
(3, 8)	0.048	0.046	0.960	0.961
(3, 15)	0.049	0.045	0.952	0.954
(5, 2)	0.040	0.039	0.931	0.929
(5, 8)	0.041	0.041	0.953	0.954
(5, 15)	0.039	0.040	0.944	0.944
(7, 2)	0.050	0.051	0.911	0.912
(7, 5)	0.051	0.053	0.928	0.929
(7, 15)	0.052	0.054	0.949	0.949

single chains of length 5×10^4 and we discarded the first 4×10^4 draws. The algorithm was based on a random walk Student's t proposal for α and on an independent proposal for ν (the proposal distribution was proportional to ν^{-2}). The proposal for α has been tuned up in terms of the scale parameter and the degrees of freedom to get an acceptance rate of about 25–40 per cent. The standard error of each entry is about $p(1-p)/1000$, where p is the entry. The results of the simulations are reassuring: there are no substantial differences between the two priors. Frequentist coverage is almost always satisfactory, the worst case being the one with large skewness and fat tails.

4. Location-scale models

The densities in (6) and (11) are only standard versions of a more general location-scale family of skew-t densities. If we consider $Y=\xi+\omega Z$, then this new random variable has a skew-t distribution with location $\xi\in\mathbb{R}$, scale $\omega>0$ and shape parameter $\alpha\in\mathbb{R}$. Now the parameter of interest is the vector $\theta=(\xi,\omega,\alpha)$. To simplify, suppose prior independence between the new set of parameters (ξ,ω) and α, that is,

$$\pi(\theta)=\pi(\xi,\omega)\pi(\alpha).$$

It is well-known that, in the presence of location and scale parameters (ξ,ω), the Jeffreys rule suggests considering these parameters separately and this leads to the use of the usual improper prior $\pi(\xi,\omega)\propto\omega^{-1}$. In the next propositions we study the existence of the posterior distribution under the Jeffreys prior.

To this end, we consider a scale mixture of skew-normal distributions with density given by

$$f(y)=\int_0^\infty 2\frac{\sqrt{v}}{\omega}\phi\left\{(y-\xi)\sqrt{v}/\omega\right\}\Phi\left\{\alpha(y-\xi)\sqrt{v}/\omega\right\}p(v)\,\mathrm{d}v, \tag{15}$$

where $p(v)$ is a density of a positive random variable. For a random variable Y with density (15), we use the notation $Y\sim\mathrm{SMSN}(\xi,\omega,\alpha,p)$. When $p(v)$ is the Gamma density with shape and scale parameters equal to $\nu/2$, the density (15) reduces to the nonlinear skew-t distribution denoted by $\mathrm{ST}(\xi,\omega,\alpha,\nu)$ and considered in section 2.4. We have the following general result.

Proposition 3. *Let $y_1, \ldots, y_n$ be independent observations from $Y \sim \mathrm{SMSN}(\xi, \omega, \alpha, p)$ and consider the prior $\pi(\xi, \omega, \alpha) \propto \omega^{-1}\pi(\alpha)$, with $\pi(\alpha)$ proper. Suppose that the density p is such that, for all n,* if $V_1, V_2, \ldots, V_n \overset{\text{i.i.d.}}{\sim} p(\cdot)$, *then*

$$E\left\{\frac{\left(V_1 \cdot V_2 \cdots V_n\right)^{1/2}}{\left(V_1 + \cdots + V_n\right)^{n/2}}\right\} < \infty.$$

Then, the posterior distribution of (ξ, ω, α) is proper.

The proof is given in the Appendix.

One can easily check, as a special case, that for a fixed number of degrees of freedom ν, under the nonlinear skew-*t* model given in section 2.4 and the prior $\pi(\xi, \omega, \alpha \mid \nu) \propto \omega^{-1}\pi(\alpha \mid \nu)$, with $\pi(\alpha \mid \nu)$ proper, the posterior distribution of (ξ, ω, α) is proper.

This result can be extended to the case of unknown degrees of freedom, provided that one adopts a proper prior for ν, such as the one proposed by Fonseca *et al.* (2008). We state the result as a corollary, without proof.

Corollary 1. *Let $y_1, \ldots, y_n$ be independent observations from $Y \sim$ ST $(\xi, \omega, \alpha, \nu)$ and consider the prior $\pi(\xi, \omega, \alpha, \nu) \propto \omega^{-1}\pi(\alpha \mid \nu)\pi(\nu)$, where $\pi(\alpha \mid \nu)$ and $\pi(\nu)$ are proper. Then, for $n \geq 2$*

(i) *the posterior distribution of $(\xi, \omega, \alpha, \nu)$ is proper;*
(ii) *the posterior moments $E(\alpha^k \mid \nu, y)$ exist if the prior moments $E(\alpha^k \mid \nu)$ exist; and*
(iii) *the posterior moments $E(\alpha^k \mid y)$ exist if $\int_0^\infty E(\alpha^k \mid \nu)\pi(\nu)\,\mathrm{d}\nu < \infty$.*

One can notice that the Student's *t* approximation of the Jeffreys prior for the shape parameter α has no moments. Consequently, there is no guarantee that the resulting marginal posterior distribution of α will have finite moments of some order. For this reason, one cannot routinely use the posterior mean as a Bayesian estimator: marginal posterior median or mode should be considered. In this article, mainly for comparisons with the likelihood approach, we propose the use of the posterior mode.

5. Numerical examples

5.1. A dataset from the Monte Carlo simulation study

As a first numerical example, we consider one of the samples from our Monte Carlo simulation study. The sample size is $n = 50$ from a nonlinear skew-*t* distribution with fixed $\xi = 0$ and $\omega = 1$. The shape is $\alpha = 5$ and the degrees of freedom are $\nu = 8$. A histogram of the data is plotted in Fig. 3 (top panel). As all data values are positive, the MLE of the shape parameter is infinite. Hence, we obtain $\hat{\alpha}_{\mathrm{MLE}} = \infty$ and $\hat{\nu}_{\mathrm{MLE}} = 44.0$. The MAP estimators are $\hat{\alpha}_{\mathrm{MAP}} = 8.4$ and $\hat{\nu}_{\mathrm{MAP}} = 6.7$. The corresponding fitted nonlinear skew-*t* densities are plotted in Fig. 3 (top panel) for the MLE (dotted curve) and the MAP (dashed curve). The true underlying density is plotted as well (thick solid curve). The density fitted by maximum likelihood is a half-skew-*t*, not far from a half-normal distribution. The density obtained with the MAP estimator is closer to the underlying nonlinear skew-*t* distribution.

5.2. A dataset from Sartori

While discussing bias correction of MLEs of the shape of the skew-normal and skew-*t* distributions, Sartori (2006) simulated a dataset of size $n = 30$ from a nonlinear skew-*t* distribution with fixed $\xi = 0$, $\omega = 1$ and $\nu = 3$. The shape is $\alpha = 5$. A histogram of the data is plotted in

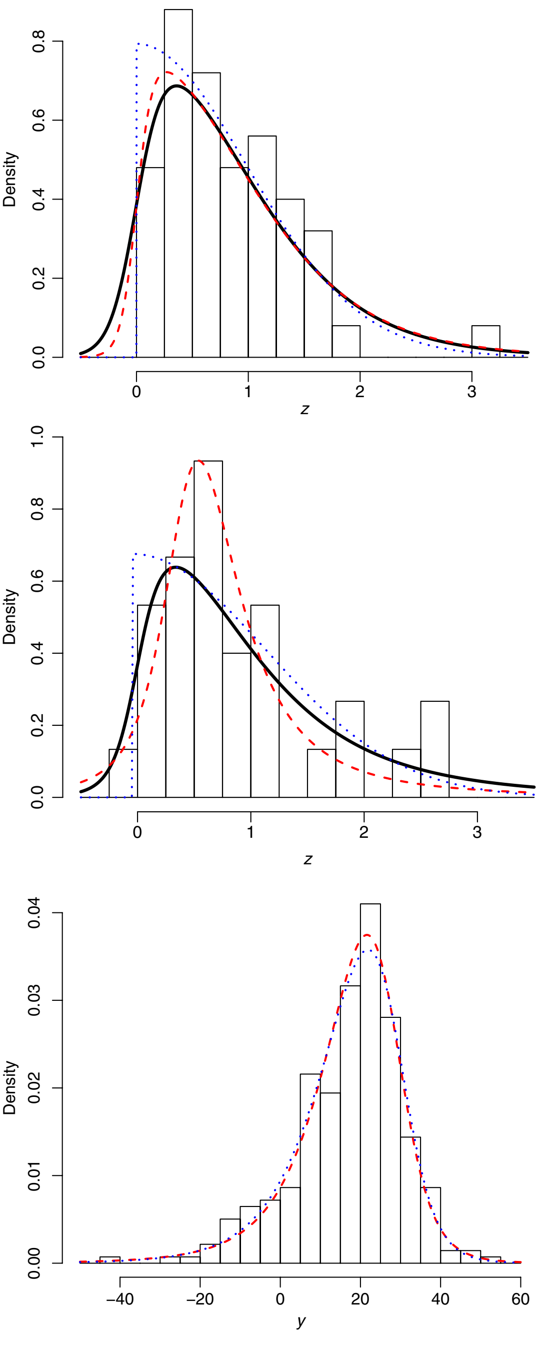

Fig. 3. Top: Histogram of a simulated dataset of size $n=50$ with fitted nonlinear skew-t densities from the maximum likelihood estimator (MLE; dotted curve) and the maximum *a posteriori* (MAP) estimator (dashed curve), and true underlying density (thick solid curve). *Middle:* Histogram of a dataset of size $n=30$ simulated by Sartori (2006) with fitted nonlinear skew-t densities from the MLE (dotted curve) and the MAP estimator (dashed curve), and true underlying density (thick solid curve). *Bottom:* Histogram of a wind velocity dataset of size $n=278$ with fitted nonlinear skew-t densities from the MLE (dotted curve) and the MAP estimator (dashed curve).

Fig. 3 (middle panel). In this case, $\hat{\alpha}_{\mathrm{MLE}}=13.5$ while Sartori's modified estimate is $\hat{\alpha}_{\mathrm{SAR}}=6.6$. We obtain $\hat{\alpha}_{\mathrm{MAP}}=5.7$, yet a bit closer to the true value $\alpha=5$. Sartori (2006) conjectured that $\hat{\alpha}_{\mathrm{SAR}}<\infty$ whenever $\hat{\alpha}_{\mathrm{MLE}}=\infty$. We can actually prove that $\hat{\alpha}_{\mathrm{MAP}}<\infty$ whenever $\hat{\alpha}_{\mathrm{MLE}}=\infty$. A rigorous proof of this result is too long to report: we give here a sketch of it. Let $L(\alpha,\omega,v)$ be the likelihood function associated to an i.i.d. sample from the nonlinear skew-t model and suppose that, without loss of generality, for a given sample, $\hat{\alpha}_{\mathrm{MLE}}=\infty$. This implies that the integrated likelihood,

$$\tilde{L}(\alpha)=\int_v\int_\omega L(\alpha,\omega,v)\frac{1}{\omega^2}\pi(v)\,\mathrm{d}\omega\,\mathrm{d}v$$

is, for $\alpha>\alpha_0$, non-decreasing and bounded by a constant, where α_0 depends on the dataset. This also implies that the right tail of the marginal posterior is entirely driven by the tail order of the prior: then the marginal posterior vanishes as $\alpha\to\infty$ and $\hat{\alpha}_{\mathrm{MAP}}<\infty$.

In practice though, the location, scale and degrees of freedom parameters are generally unknown. We fit the nonlinear skew-t distribution to Sartori's simulated data by maximum likelihood and obtain $\hat{\xi}_{\mathrm{MLE}}=-0.05$, $\hat{\omega}_{\mathrm{MLE}}=1.2$, $\hat{\alpha}_{\mathrm{MLE}}=\infty$ and $\hat{v}_{\mathrm{MLE}}=\infty$. On the other hand, using our objective Bayesian approach, we obtain $\hat{\xi}_{\mathrm{MAP}}=0.4$, $\hat{\omega}_{\mathrm{MAP}}=0.4$, $\hat{\alpha}_{\mathrm{MAP}}=1.0$ and $\hat{v}_{\mathrm{MAP}}=1.8$. Figure 3 (middle panel) depicts those two fitted nonlinear skew-t densities from the MLE (dotted curve) and the MAP (dashed curve), and the true underlying density (thick solid curve). The maximum likelihood fit is a half-normal distribution which is clearly not appropriate. Our fit with the MAP estimator is somewhat different from the true underlying density but is a fairly good fit to the generated data.

5.3. *Wind data*

This dataset is part of a study of wind energy by Hering & Genton (2010). It consists of hourly average wind speed collected at the Vansycle meteorological tower in Oregon. We consider the wind velocity (i.e. wind speed with a positive sign in the east direction and negative sign in the west direction) from 25 February to 30 November 2003 recorded at midnight, a time when wind speeds tend to peak. In this region, wind patterns are mostly westerly, hence we expect to see right skewness in the distribution of the data. This is confirmed by a histogram of the $n=278$ observations in Fig. 3 (bottom panel).

We fit a nonlinear skew-t distribution to the wind velocity data by maximum likelihood and obtain $\hat{\xi}_{\mathrm{MLE}}=29.4$, $\hat{\omega}_{\mathrm{MLE}}=15.4$, $\hat{\alpha}_{\mathrm{MLE}}=-1.9$ and $\hat{v}_{\mathrm{MLE}}=6.3$. As the sample size is not small, we expect to obtain fairly similar results with our objective Bayesian approach. Indeed, we obtain $\hat{\xi}_{\mathrm{MAP}}=28.1$, $\hat{\omega}_{\mathrm{MAP}}=13.5$, $\hat{\alpha}_{\mathrm{MAP}}=-1.5$ and $\hat{v}_{\mathrm{MAP}}=4.4$. Figure 3 (bottom panel) depicts those two fitted nonlinear skew-t densities from the MLE (dotted curve) and the MAP (dashed curve). As expected, the two fitted densities are very similar.

6. Discussion

We have studied the Jeffreys prior and its properties for the shape parameter of univariate linear and nonlinear skew-t distributions. In both cases, we have shown that the resulting priors for the shape parameter are symmetric around zero and proper. Moreover, we have proposed a Student's t approximation of the Jeffreys prior that makes an objective Bayesian analysis easy to perform. We have performed a Monte Carlo simulation study that demonstrated an overall better behaviour of the MAP estimator compared with the MLE. We also compared the frequentist coverage of the credible intervals based on the Jeffreys prior and its approximation and showed that they were similar. The results have been extended to location-scale models under the usual Jeffreys prior and we have shown the existence of the posterior distribution and its moments in that case. We have also presented three numerical examples to illustrate the implications of our results on inference for skew-t distributions.

In the framework of skew-t distributions, it is important to have a proper prior for the shape parameter. Indeed, when the likelihood function is unbounded for some samples, a constant prior for the shape parameter leads to an improper posterior. Although we have shown the existence and finiteness of the MAP estimator, its uniqueness is not guaranteed. However, it seems very unlikely that the posterior distribution of the shape parameter would be bimodal or multimodal with several modes having exactly the same mass. In all our simulations and examples, the posterior distribution of the shape parameter has been unimodal, but a formal proof of this property appears to be difficult.

We have focused our results on skew-t distributions, that is, $f_0(z)=t(z\,|\,v)$ and $G=T$ in (1). However, for other choices of the base density $f_0(z)$ and of G, one could still study the propriety of the Jeffreys prior resulting from (2) and use the approximation of the density $h(z)$ in (3) as explained in section 2.1 to obtain an approximation of the Jeffreys prior.

We have restricted ourselves to the setting of univariate distributions. An open question is to extend our results to multivariate cases. However, even for the simple bivariate skew-normal distribution, this seems to be a non-trivial project. For this problem, we have only partial results: the Jeffreys prior for the shape vector of parameters turns out to be improper, while the reference prior algorithm does produce proper priors, at least for some specific parameter ordering.

Acknowledgements

Branco's research was partially supported by FAPESP grant 04/15304-6. Genton's research was partially supported by NSF grant DMS-1007504. This publication is based in part on work supported by award no. KUS-C1-016-04 made by King Abdullah University of Science and Technology (KAUST). The authors thank the editor, an associate editor and a referee for their helpful comments and suggestions.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. The Jeffreys prior (solid curve) and its Student's t approximation (dashed curve), both denoted by $p(\alpha)$, for the shape parameter α of the linear skew-t model with $v=3,5,10,100$.
Figure S2. The Jeffreys prior (solid curve) and its Student's t approximation (dashed, dotted, dashed-dotted curves), both denoted by $p(\alpha)$, for the shape parameter α of the nonlinear skew-t model with $v=2,4,10$.
Table S1. Root mean squared error (*RMSE*) for the MLE and MAP estimator of the shape parameter α (all samples) of a nonlinear skew-t distribution.
Table S2. Root relative mean squared error (*RRELMSE*) for the MLE and MAP estimator of the degrees of freedom parameter v (only samples with finite MLE) of a nonlinear skew-t distribution.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–178.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.* **32**, 159–200.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. Roy. Statist. Soc. Ser. B* **61**, 579–602.

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. Roy. Statist. Soc. Ser. B* **65**, 367–389.

Azzalini, A. & Genton, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions. *Int. Stat. Rev.* **76**, 106–129.

Bayes, C. L. & Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Braz. J. Probab. Stat.* **21**, 141–163.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **41**, 113–147.

Berger, J. O. & Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84**, 200–207.

Berger, J. O. & Bernardo, J. M. (1992). On the development of reference priors. In *Bayesian statistics 4* (eds J. M. Bernardo, J. O. Berger, D. V. Lindley & A. F. M. Smith), 35–60 (with discussion). Oxford University Press, Oxford.

Berger, J. O., Bernardo, J. M. & Sun, D. (2009). The formal definition of reference priors. *Ann. Statist.* **37**, 905–938.

Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.* **79**, 99–113.

Feller, W. (1971). *An introduction to probability theory and its applications*, Vol. 2. Wiley, New York.

Fernández, C. & Steel, M. F. J. (1999). Multivariate Student t regression models: pitfalls and inference. *Biometrika* **86**, 153–167.

Fisher, R. A. (1925). Expansion of "Student's" integral in powers of n^{-1}. *Metron* **5**, 109–112.

Fonseca, T. C., Ferreira, M. A. R. & Migon, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika* **95**, 325–333.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications: a journey beyond normality*, edited volume. Chapman & Hall/CRC, Boca Raton, FL.

Gómez, H. W., Venegas, O. & Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics* **18**, 395–407.

Hering, A. S. & Genton, M. G. (2010). Powering up with space-time wind forecasting. *J. Amer. Statist. Assoc.* **105**, 92–104.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous univariate distributions 2*. Wiley, New York.

Liseo, B. & Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *J. Statist. Plann. Inference* **136**, 373–389.

Liseo, B., Tancredi, A. & Barbieri, M. M. (2010). Approximated reference priors in the presence of latent structure. In *Frontiers of statistical decision making and bayesian analysis – in honor of James O. Berger*, (eds M. Chen, D. K. Dey, P. Mueller, D. Sun & K. Ye), 23–42, Springer, New York.

Ma, Y. & Genton, M. G. (2004). A flexible class of skew-symmetric distributions. *Scand. J. Statist.* **31**, 459–468.

Nadarajah, S. & Kotz, S. (2003). Skewed distributions generated by the normal kernel. *Statist. Probab. Lett.* **65**, 269–277.

Resnick, S. (1992). *Adventures in stochastic processes*. Birkhauser, Boston.

Robert, C. (2001). *The Bayesian choice: From decision-theoretic foundations to computational implementation*, 2nd edn. Springer, New York.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *J. Statist. Plan. Inference* **136**, 4259–4275.

Sun, D. & Berger, J. (1998). Reference priors with partial information. *Biometrika* **85**, 55–71.

Wang, J., Boyer, J. & Genton, M. G. (2004). A skew-symmetric representation of multivariate distributions. *Statist. Sinica* **14**, 1259–1270.

Received October 2010, in final form October 2011

Marc G. Genton, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA.
E-mail: genton@stat.tamu.edu

Appendix

Preliminary results

Following Johnson *et al.* (1995), the Student's t density can be approximated by a Taylor expansion around the reciprocal of the degrees of freedom, that is,

$$t(z \mid v) \approx \phi(z)\left\{1+\frac{1}{4}(z^4-2z^2-1)v^{-1}+\frac{1}{96}(3z^8-28z^6+30z^4+12z^2+3)v^{-2}+\cdots\right\}, \qquad (16)$$

where ϕ is the standard normal density. The approximation (16) has been proposed by Fisher (1925) and is used to prove the next two lemmas that will be essential in the proof of proposition 1.

Lemma 1. *For all* $v=1,2,\ldots,$ *the function of* α:

$$\int_0^{+\infty} z^2\phi(z)\left\{1+\frac{1}{4v}(z^4-2z^2-1)\right\}\phi^2(b\alpha z)\left\{1+\frac{1}{4v}(b^4\alpha^4z^4-2b^2\alpha^2z^2-1)\right\}^2 \mathrm{d}z, \qquad (17)$$

where $-\infty<\alpha<\infty$ *and* $b>0$, *has tails of order* $O(\alpha^{-3})$.

Proof. The integral in (17) can be rearranged as

$$\int_0^{+\infty} \frac{(z^2)^{3/2-1}}{2(2\pi)^{3/2}}\, \mathrm{e}^{-(1/2+b^2\alpha^2)z^2} A(z)2z\, \mathrm{d}z,$$

where $A(z)$ includes the polynomial terms in z. Setting $u=z^2$, it follows that (17) is

$$\int_0^{+\infty} \frac{u^{3/2-1}}{2(2\pi)^{3/2}}\, \mathrm{e}^{-(1/2+b^2\alpha^2)u} A(u)\, \mathrm{d}u.$$

The last expression can be formally seen as an expected value

$$\frac{(1/2+b^2\alpha^2)^{-3/2}}{2^{7/2}\pi} E_U\{A(U)\}, \qquad (18)$$

where the expectation is taken with respect to a Gamma$(3/2, 1/2+b^2\alpha^2)$ distribution, and $A(u)$ is a polynomial in u of order 6 given by

$$A(u)=a_0+a_1(\alpha^2)u+a_2(\alpha^4)u^2+a_3(\alpha^6)u^3+a_4(\alpha^8)u^4+a_5(\alpha^8)u^5+a_6(\alpha^8)u^6,$$

where $a_i(\alpha^j)$, $i=1,\ldots,6, j=2,4,6,8,$ are polynomials in α of order j. Then

$$E\{A(U)\}=a_0+a_1(\alpha^2)E(U)+a_2(\alpha^4)E(U^2)+a_3(\alpha^6)E(U^3) + a_4(\alpha^8)E(U^4)+a_5(\alpha^8)E(U^5)+a_6(\alpha^8)E(U^6).$$

Now, using the following expression for the moments of the Gamma density

$$E(U^k)=\frac{\Gamma(3/2+k)}{\Gamma(3/2)}(1/2+b^2\alpha^2)^{-k}, \quad k=1,2,\ldots,$$

we get that $E\{A(U)\}=a_0+O(1)+O(1)+O(1)+O(1)+O(\alpha^{-2})+O(\alpha^{-4})=O(1)$ for $\alpha>1$. From (18) it follows that (17) is $O(\alpha^{-3})$. This ends the proof of lemma 1.

Lemma 2. *The function* $B(\alpha)=\int_0^{+\infty} z^2t(z \mid v)t^2(\alpha z \mid v)\, \mathrm{d}z$ *is of order* $O(\alpha^{-3})$.

Proof. Using lemma 1 with $b=1$, we see that an approximation for $B(\alpha)$ using only the first term of (16) is of order $O(\alpha^{-3})$. If one adds a new term in the approximation it can be noted that

$$B(\alpha) = \frac{(1/2 + b^2\alpha^2)^{-3/2}}{2^{7/2}\pi} E\{A^*(U)\},$$

where $A^*(U)$ is similar to $A(U)$ in lemma 1, that is, it is a polynomial function in the Gamma random variable U. Hence, $E\{A^*(U)\}$ is a sum of terms of the form $a(\alpha^d)E(U^b)$, where $d \leq 2b$. Therefore, again $E\{A^*(U)\} = O(1)$ and $B(\alpha) = O(\alpha^{-3})$. This ends the proof of lemma 2.

The following lemmas will be used in the proof of proposition 2.

Lemma 3. *(Resnick,* 1992*, lemma* 6.2.2*, p.* 487*). As* $x \to \infty$*,*

$$\frac{x}{1+x^2} \, e^{-x^2/2} < \int_x^\infty e^{-u^2/2} \, du < \frac{1}{x} \, e^{-x^2/2}.$$

Lemma 4. *For* $z < \sqrt{v}$ *and for all positive v,*

$$\{1 - T(z \mid v)\}^{-1} < \frac{t}{v} \mathrm{Be}(1/2, v/2) \left(1 - t^2/v\right)^{-(v+1)/2} \left(v - 2 + v/t^2\right),$$

where Be *represents Euler's Beta function* $\mathrm{Be}(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

Proof. The proof is based on the use of Jensen's inequality, on lemma 3 and on the representation of a Student's t cumulative distribution function as a scale mixture of normal densities.

Proof of Proposition 1. The proof of proposition 1 is similar in spirit to that presented in Liseo & Loperfido (2006). To prove (i), it is enough to notice from (2) that $I(\alpha) = I(-\alpha)$. To prove (ii), we notice from part (i) that it is enough to study the right tail. Consider then $\alpha > 0$. We have

$$I(\alpha) = \int_0^\infty 2z^2 t(z \mid v) \frac{t^2(\alpha z \mid v)}{T(\alpha z \mid v)} \, dz + \int_0^\infty 2z^2 t(z \mid v) \frac{t^2(\alpha z \mid v)}{\{1 - T(\alpha z \mid v)\}} \, dz.$$

Consider the first integral. Since $\alpha > 0$ and $z > 0$, then $1 < T(\alpha z \mid v)^{-1} < 2$, and

$$B(\alpha) < \int_0^\infty z^2 t(z \mid v) \frac{t^2(\alpha z \mid v)}{T(\alpha z \mid v)} \, dz < 2B(\alpha),$$

where $B(\alpha) = \int_0^\infty z^2 t(z \mid v) t^2(\alpha z \mid v) \, dz$ which, from lemma 2, is of order $O(\alpha^{-3})$. The second integral can be split into two parts

$$D_1(\alpha) + D_2(\alpha) = \int_0^{1/\alpha} 2z^2 t(z \mid v) \frac{t^2(\alpha z \mid v)}{\{1 - T(\alpha z \mid v)\}} \, dz + \int_{1/\alpha}^\infty 2z^2 t(z \mid v) \frac{t^2(\alpha z \mid v)}{\{1 - T(\alpha z \mid v)\}} \, dz.$$

For $0 < z < \alpha^{-1}$, we consider $2 < \{1 - \Phi(\alpha z)\}^{-1} < c = \{1 - \Phi(1)\}^{-1}$. Then $2D_1^*(\alpha) < D_1(\alpha) < cD_1^*(\alpha)$, where

$$D_1^*(\alpha) = \int_0^{1/\alpha} 2z^2 t(z \mid v) t^2(\alpha z \mid v) \, dz \leq \int_0^\infty 2z^2 t(z \mid v) t^2(\alpha z \mid v) \, dz,$$

and the last term is of order $O(\alpha^{-3})$. Finally, observing that $1 - T(\alpha z \mid v) > 1 - \Phi(\alpha z)$ when $z > 1/\alpha$, we have

$$D_2(\alpha) \leq \int_{1/\alpha}^\infty 2z^2 t(z \mid v) \frac{t^2(\alpha z \mid v)}{1 - \Phi(\alpha z)} \, dz.$$

Now, using the following inequality (Feller, 1971, p. 175)

$$\frac{1}{1-\Phi(\alpha z)} \le \frac{\alpha^3 z^3}{(\alpha^2 z^2 - 1)\phi(\alpha z)},$$

we obtain

$$D_2(\alpha) \le \int_{1/\alpha}^{\infty} 2\alpha^3 z^5 t(z \mid v) \frac{t^2(\alpha z \mid v)}{(\alpha^2 z^2 - 1)\phi(\alpha z)}\,\mathrm{d}z \le \alpha^3 \int_0^{\infty} 2z^5 \phi(z)\phi(\alpha z) A^*(z)\,\mathrm{d}z = \frac{8\alpha^3}{(1+\alpha^2)^3} \mathrm{E}_U\{A^*(U)\},$$

where U has a Gamma(3, $\frac{1+\alpha^2}{2}$) distribution and $A^*(u)$ is a polynomial in u, as in lemma 2. It follows that $E_U\{A^*(U)\}$ is of order $O(1)$. Then $D_2(\alpha) \le O(\alpha^{-3})$ and consequently $I(\alpha) = O(\alpha^{-3})$.

Proof of Proposition 2. To prove part (i), first we note that $I(\alpha)$ can be expressed as:

$$I(\alpha) = 2\int_0^{\infty} \frac{z^2 u(v,z)^2 t(z \mid v) t^2\{\alpha z u(v,z) \mid v+1\}}{T\{\alpha z u(v,z) \mid v+1\}[1 - T\{\alpha z u(v,z) \mid v+1\}]}\,\mathrm{d}z,$$

where $u(v,z) = \{(v+1)/(v+z^2)\}^{1/2}$. To show that $I(\alpha)$ is symmetric, it is enough to notice that, for all z and for all positive v, $t(z \mid v) = t(-z \mid v)$ and $T(-z \mid v) = 1 - T(z|v)$.

To show that $I(\alpha)$ is monotonically decreasing it is sufficient, although not necessary, to show that

$$G_\xi(y) = \frac{t^2(y \mid \xi)}{T(y \mid \xi) T(-y \mid \xi)}$$

is decreasing in y, for all positive ξ. Set $g_\xi(y) = \log G_\xi(y)$. Then

$$g'_\xi(y) = 2\frac{t'(y \mid \xi)}{t(y \mid \xi)} - t(y \mid \xi)\left\{\frac{1}{T(y \mid \xi)} - \frac{1}{T(-y \mid \xi)}\right\} = 2\frac{t'(y \mid \xi)}{t(y \mid \xi)} - t(y \mid \xi)\frac{1 - 2T(y \mid \xi)}{T(y \mid \xi) T(-y \mid \xi)},$$

where

$$t'(y \mid \xi) = -c(\xi)\left(1 + y^2/\xi\right)^{-(\xi+3)/2} \frac{\xi+1}{\xi}.$$

Then

$$2\frac{t'(y \mid \xi)}{t(y \mid \xi)} = -2y\frac{\xi+1}{\xi}\left(1 + y^2/\xi\right)^{-1},$$

and implies that

$$g'_\xi(y) = -2\alpha\frac{\xi+1}{\xi}\left(1 + y^2/\xi\right)^{-1} - t(y \mid \xi)\frac{1 - 2T(y \mid \xi)}{T(y \mid \xi) T(-y \mid \xi)},$$

which is negative for all positive y and all positive ξ.

To prove part (ii), we need to study the behaviour of the tails of $I(\alpha)$. Since this quantity is symmetric around zero, we consider the right tail only. Set

$$\begin{aligned} I(\alpha) &= A(\alpha) + B(\alpha) \\ &= 2\int_0^{\infty} z^2 u(v,z)^2 t(z \mid v) t^2\{\alpha z u(v,z) \mid v+1\} \frac{1}{T\{\alpha z u(v,z) \mid v+1\}}\,\mathrm{d}z \\ &\quad + 2\int_0^{\infty} z^2 u(v,z)^2 t(z \mid v) t^2\{\alpha z u(v,z) \mid v+1\} \frac{1}{1 - T\{\alpha z u(v,z) \mid v+1\}}\,\mathrm{d}z. \end{aligned}$$

For all positive α, v and z, $1 < 1/T\{\alpha z u(v,z) \mid v+1\} < 2$. Then $A^*(\alpha) < A(\alpha) < 2A^*(\alpha)$, with

$$A^*(\alpha) \propto \int_0^{\infty} \frac{z^2}{v+z^2} \frac{1}{(1+z^2/v)^{(v+1)/2}} \frac{1}{\{1 + \alpha^2 z^2/(v+z^2)\}^{v+2}}.$$

Simple algebra shows that $A^*(\alpha) \propto (1+\alpha^2)^{\nu+2} W(\alpha, \nu)$, with

$$W(\alpha, \nu) = \int_0^\infty z^2(\nu+z^2)^{(\nu+1)/2} / \{z^2 + \nu/(1+\alpha^2)\}\, \mathrm{d}z.$$

It is easy to see that $W(\alpha, \nu)$ is increasing in α. Then, there are two positive values $m < M$, such that

$$m < W(0, \nu) < W(\alpha, \nu) < \lim_{\alpha \to \infty} W(\alpha, \nu) < M.$$

This amounts to say that $A(\alpha) = O\left(\alpha^{2(\nu+2)}\right)$. Consider now $B(\alpha)$. First, we make a change of variable $x = z\sqrt{\nu+1}/\sqrt{\nu+z^2}$ so that

$$B(\alpha) \propto \int_0^{\sqrt{\nu+1}} s(x \mid \nu, \alpha)\, \mathrm{d}x,$$

where

$$s(x \mid \nu, \alpha) \propto \frac{x^2 t(x\sqrt{\nu}/\sqrt{\nu+1-x^2} \mid \nu) t^2(\alpha x \mid \nu+1)}{(\nu+1-x^2)^{3/2} T(-\alpha x \mid \nu+1)}.$$

Set

$$B(\alpha) = B_1(\alpha) + B_2(\alpha) = \int_0^{1/\alpha} s(x \mid \nu, \alpha)\, \mathrm{d}x + \int_{1/\alpha}^{\sqrt{\nu+1}} s(x \mid \nu, \alpha)\, \mathrm{d}x.$$

For $0 \le x \le 1/\alpha$,

$$2 \le \frac{1}{T(-\alpha x \mid \nu+1)} \le c_1(\nu),$$

where $c_1(\nu)$ is a finite value depending on ν. This implies that $B_1(\alpha) \in \left[2B_1^*(\alpha), c_1(\nu) B_1^*(\alpha)\right]$, with

$$B_1^*(\alpha) = \int_0^{1/\alpha} \frac{x^2 t(x\sqrt{\nu}/\sqrt{\nu+1-x^2} \mid \nu) t^2(\alpha x \mid \nu+1)}{(\nu+1-x^2)^{3/2}}\, \mathrm{d}x.$$

With an argument similar to the one used for $A(\alpha)$, one can prove that $B_1(\alpha) = O(\alpha^{-2(\nu+2)})$.

From lemma 3:

$$\begin{aligned}
B_2(\alpha) &= \int_{1/\alpha}^{\sqrt{\nu+1}} \frac{x^2 t(x\sqrt{\nu}/\sqrt{\nu+1-x^2} \mid \nu) t^2(\alpha x \mid \nu+1)}{(\nu+1-x^2)^{3/2} T(-\alpha x \mid \nu+1)}\, \mathrm{d}x \\
&= \int_{1/\alpha}^{\sqrt{\nu+1}} \frac{x^2 \left\{1 + x^2/(\nu+1-x^2)\right\}^{-(\nu+1)/2} \left\{1 + \alpha^2 x^2/(\nu+1)\right\}^{-(\nu+1)}}{(\nu+1-x^2)^{3/2} T(-\alpha x \mid \nu+1)}\, \mathrm{d}x \\
&\le c'(\nu)\alpha \int_{1/\alpha}^{\sqrt{\nu+1}} \frac{x^3 \left\{\nu - 1 + \nu + 1/(\alpha^2 x^2)\right\} \left\{1 + x^2/(\nu+1-x^2)\right\}^{-(\nu+1)/2}}{\left\{1 - \alpha^2 x^2/(\nu+1)\right\}^{(\nu+2)/2} \left\{1 + \alpha^2 x^2/(\nu+1)\right\}^{\nu+1}}\, \mathrm{d}x \\
&= c''(\nu)\alpha \int_{1/\alpha\sqrt{\nu+1}}^{1} \frac{t^3 \left(1-t^2\right)^{(\nu-2)/2}}{\left(1-\alpha^2 t^2\right)^{(\nu+2)/2} (1+\alpha^2 t^2)^{\nu+1}} \left\{\nu - 1 + 1/(\alpha^2 t^2)\right\} \mathrm{d}t \\
&= c''(\nu) \left\{\alpha g_3(\alpha) + \frac{1}{\alpha} g_1(\alpha)\right\},
\end{aligned}$$

where c' and c'' are constants, not depending on t and α and

$$g_k(\alpha) = \int_{1/\alpha\sqrt{\nu+1}}^{1} \frac{t^k (1-t^2)^{\nu/2-1}}{(1-\alpha^2 t^2)^{\nu/2+1} (1+\alpha^2 t^2)^{\nu+1}}\, \mathrm{d}t.$$

Then,

$$g_k(\alpha)=\frac{1}{\alpha^{3v+4}}\int_{1/\alpha\sqrt{v+1}}^{1}\frac{t^k(1-t^2)^{v/2-1}}{(1/\alpha^4-t^4)^{v/2+1}(1/\alpha^2+t^2)^{v/2}}\mathrm{d}t.$$

For $(\alpha\sqrt{v+1})^{-1}\leq t\leq 1$,

$$\left(\alpha^{-4}-t^4\right)^{-1}\leq\frac{\alpha^4}{(1-\alpha^4)}$$

and

$$\left(\alpha^{-2}+t^2\right)^{-1}\leq\frac{\alpha^2(v+1)}{v+2}.$$

Now,

$$g_k(\alpha)\leq c(v)\frac{1}{\alpha^{3v+4}}\left(\frac{\alpha^4}{1-\alpha^4}\right)^{v/2+1}\alpha^{2v}\int_0^1 t^k(1-t^2)^{v/2-1}\,\mathrm{d}t=\frac{1}{(1+\alpha^4)^{v/2+1}}c'(v)\mathrm{Be}\left(\frac{k+1}{2},\frac{v}{2}\right).$$

Finally,

$$B_2(\alpha)<c''(v)\frac{1}{(1+\alpha^4)^{v/2+1}}\left(\alpha+\frac{1}{\alpha}\right).$$

Then, $B_2(\alpha)$ is bounded by an integrable function for large α.

Approximated derivation of the reference prior

Here we show how to derive an approximation to the Jeffreys prior following the approach described in section 2.3. In particular, we show that the conditional prior for α given v can be approximated by a Student's t distribution with mean zero, scale $\pi/2$ and degrees of freedom 1/2.

We start by computing $I_{\alpha,\alpha}$. The first derivative of (13) is

$$\frac{\partial\ell}{\partial\alpha}(\alpha,v\,|\,z)=\sum_{j=1}^{n}\left\{\frac{\partial}{\partial\alpha}\log\Phi\left(\alpha v^{1/2}z_j\right)\right\}=\sum_{j=1}^{n}\left\{z_j v_j^{1/2}\frac{\phi(\alpha v_j^{1/2}z_j)}{\Phi(\alpha v_j^{1/2}z_j)}\right\},$$

and the second derivative

$$\frac{\partial^2\ell}{\partial\alpha^2}(\alpha,v\,|\,z)=\sum_{j=1}^{n}\left\{z_j^2 v_j\frac{\phi^2(\alpha v_j^{1/2}z_j)+\alpha v_j^{1/2}z_j\phi(\alpha v_j^{1/2}z_j)\Phi(\alpha v_j^{1/2}z_j)}{\Phi^2(\alpha v_j^{1/2}z_j)}\right\}. \quad (19)$$

The expected Fisher information must then be computed with respect to the n-dimensional skew-normal sampling distribution. As each term of the sum in (19) depends on just one component of the vector z, we get

$$I_{\alpha,\alpha}=\sum_{j=1}^{n}\left\{v_j^{3/2}\int z_j^2\frac{\phi^2(\alpha v_j^{1/2}z_j)\phi(v_j^{1/2}z_j)}{\Phi(\alpha v_j^{1/2}z_j)}\,\mathrm{d}z_j\right\}+\sum_{j=1}^{n}\left\{v_j^2\int z_j^3\phi(\alpha v_j^{1/2}z_j)\phi(v_j^{1/2}z_j)\,\mathrm{d}z_j\right\}.$$

The second term of the aforegiven quantity is the sum of integrals of odd functions, hence it is zero. Then, setting $w_j=z_j v_j^{1/2}$, $j=1,\ldots,n$, we have

$$\begin{aligned}I_{\alpha,\alpha}&=\sum_{j=1}^{n}\left\{\int_{-\infty}^{\infty}w_j^2\phi(w_j)\frac{\phi^2(\alpha w_j)}{\Phi(\alpha w_j)}\,\mathrm{d}w_j\right\}\\&\propto\sum_{j=1}^{n}\left[\int_0^{\infty}w_j^2\phi(w_j)\frac{\phi^2(\alpha w_j)}{\Phi(\alpha w_j)\{1-\Phi(\alpha w_j)\}}\,\mathrm{d}w_j\right].\end{aligned}$$

Using the approximation proposed in Bayes & Branco (2007), we obtain

$$I_{\alpha,\alpha} \approx \sum_{j=1}^{n} \int_0^{\infty} w_j^2 \phi(w_j)\phi^2(\alpha w_j \pi/2)\,\mathrm{d}w_j \propto \sum_{j=1}^{n} (1+8\alpha^2/\pi^2/2)^{-3/2} \propto (1+8\alpha^2/\pi^2)^{-3/2}.$$

Then the conditional prior for α given the vector v does not depend on v and we obtain an approximation of the Jeffreys prior proportional to $(1+8\alpha^2\pi^2)^{-3/4}$.

Proof of Proposition 3. To make calculations easier, we prefer to use the parameter ω^2 instead of ω. As a result of the invariance of the Jeffreys prior, the prior distribution will be $\pi(\xi, \omega^2, \alpha) \propto \omega^{-2}\pi(\alpha)$. Let the likelihood function be

$$L(\xi, \omega^2, \alpha) = \prod_{i=1}^{n} \left[\int_0^{\infty} \frac{2\sqrt{v_i}}{\omega} \phi\left\{(y_i - \xi)\sqrt{v_i}/\omega\right\} \Phi\left\{\alpha(y_i - \xi)\sqrt{v}/\omega\right\} p(v_i)\,\mathrm{d}v_i \right].$$

We introduce the symbols $\underline{v} = (v_1, \ldots, v_n)$ and $V^{(n)} = [0, \infty)^n$; also, $p(\underline{v})$ denotes $\prod_{i=1}^{n} p(v_i)$. Since $\Phi(\cdot) < 1$,

$$L(\xi, \omega^2, \alpha) \le \int_{V^{(n)}} \prod_{i=1}^{n} v_i \frac{1}{(2\pi\omega^2)^{n/2}} \exp\left[-\frac{1}{2\omega^2} \sum_{i=1}^{n} \left\{ v_i(\xi - y_i)^2 \right\} \right] p(\underline{v})\,\mathrm{d}\underline{v}.$$

After some simple algebra, one can write

$$L(\xi, \omega^2, \alpha) \le \int_{V^{(n)}} \prod_{i=1}^{n} v_i \frac{1}{(2\pi\omega^2)^{n/2}} \exp\left\{ -\frac{1}{2\omega^2} \sum v_i \left(\xi - \sum v_i y_i / \sum v_i \right)^2 \right\}$$
$$\times \exp\left[-\frac{1}{2\omega^2} \left\{ \sum v_i y_i^2 - \left(\sum y_i v_i \right)^2 / \sum v_i \right\} \right] p(\underline{v})\,\mathrm{d}\underline{v}.$$

The posterior distribution $\pi(\xi, \omega^2, \alpha \mid y)$ is proportional to $\omega^{-2}\pi(\alpha)L(\xi, \omega^2, \alpha)$. Since $\pi(\alpha)$ is proper and using the above upper bound for the likelihood function, one immediately obtains

$$\int_{\alpha} \pi(\xi, \omega^2, \alpha \mid y)\,\mathrm{d}\alpha \le \int_{V^{(n)}} \prod_{i=1}^{n} v_i \frac{1}{(2\pi)^{n/2}} \left(\frac{1}{\omega^2} \right)^{n/2+1} \exp\left\{ -\frac{1}{2\omega^2} \sum v_i \left(\xi - \sum v_i y_i \Big/ \sum v_i \right)^2 \right\}$$
$$\times \exp\left[-\frac{1}{2\omega^2} \left\{ \sum v_i y_i^2 - \left(\sum y_i v_i \right)^2 \Big/ \sum v_i \right\} \right] p(\underline{v})\,\mathrm{d}\underline{v}.$$

Also,

$$\int_{\xi} \int_{\alpha} \pi(\xi, \omega^2, \alpha \mid y)\,\mathrm{d}\xi\,\mathrm{d}\alpha \le \int_{V^{(n)}} \prod_{i=1}^{n} v_i \frac{1}{(2\pi)^{(n-1)/2}} \left(\frac{1}{\omega^2} \right)^{(n+1)/2} \frac{1}{\sum v_i}$$
$$\times \exp\left[-\frac{1}{2\omega^2} \left\{ \sum v_i y_i^2 - \left(\sum y_i v_i \right)^2 \Big/ \sum v_i \right\} \right] p(\underline{v})\,\mathrm{d}\underline{v},$$

and

$$\int_{\xi} \int_{\omega^2} \int_{\alpha} \pi(\xi, \omega^2, \alpha \mid y)\,\mathrm{d}\xi\,\mathrm{d}\omega^2\,\mathrm{d}\alpha \le \int_{V^{(n)}} \prod_{i=1}^{n} v_i \frac{2^{(n-1)/2}}{(2\pi)^{(n-1)/2}} \Gamma\left(\frac{n-1}{2} \right)$$
$$\times \left\{ \sum v_i y_i^2 - \left(\sum y_i v_i \right)^2 \Big/ \sum v_i \right\}^{-(n-1)/2} p(\underline{v})\,\mathrm{d}\underline{v}.$$

It follows that the posterior distribution is proper when

$$\int_{V^{(n)}} \left(v_1 \cdot v_2 \cdots v_n \right)^{1/2} \frac{\left(\sum v_i \right)^{(n-1)/2}}{\left\{ \sum v_i y_i^2 \sum v_i - \left(\sum y_i v_i \right)^2 \right\}^{(n-1)/2}} p(\underline{v})\,\mathrm{d}\underline{v} \tag{20}$$

is finite. Denote

$$y^{(1)} = \min\left(y_1, \ldots, y_n\right) \quad \text{and} \quad y^{(n)} = \max\left(y_1, \ldots, y_n\right).$$

It is easy to see that

$$0 \leq \sum v_i y_i^2 \sum v_i - \left(\sum y_i v_i\right)^2 = \left| \sum v_i y_i^2 \sum v_i - \left(\sum y_i v_i\right)^2 \right| \geq \left| y^{(1)} - y^{(n)} \right| \left(\sum v_i\right)^2.$$

This implies that the right-hand term of (20) is less than or equal to

$$\int_{V^{(n)}} \frac{\left(v_1 \cdot v_2 \cdots v_n\right)^{1/2}}{\left(\sum v_i\right)^{-n/2}} p(\underline{v})\, \mathrm{d}\underline{v},$$

and this proves the proposition.