Inference in Two-Piece Location-Scale models with Jeffreys Priors

F. J. Rubio[*] and M. F. J. Steel[†]

Abstract

This paper addresses the use of Jeffreys priors in the context of univariate three-parameter location-scale models, where skewness is introduced by differing scale parameters either side of the location. We focus on various commonly used parameterizations for these models. Jeffreys priors are shown not to allow for posterior inference in the wide and practically relevant class of distributions obtained by skewing scale mixtures of normals. Easily checked conditions under which independence Jeffreys priors can be used for valid inference are derived. We empirically investigate the posterior coverage for a number of Bayesian models, which are also used to conduct inference on real data.

Key Words: coverage; Bayesian inference; noninformative prior; posterior existence; skewness

1 Introduction

The use of skewed distributions is an attractive option for modeling data presenting departures from symmetry. Several mechanisms to obtain skewed distributions by appropriately modifying symmetric distributions have been presented in the literature (Azzalini, 1985; Fernández and Steel, 1998; Mudholkar and Hutson, 2000). We focus on

[*]University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK. E-mail: F.J.Rubio@warwick.ac.uk

[†]University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK. E-mail: M.F.Steel@stats.warwick.ac.uk

the simple univariate location-scale model where we induce skewness by the use of different scales both sides of the mode and only distinguish three scalar parameters. We investigate Bayesian inference using Jeffreys priors in this simple setting.

Firstly, we consider univariate (continuous) two-piece distributions with different scales both sides of the location parameter. Then, we will focus on the family of skewed distributions defined in Arellano-Valle et al. (2005), where the scales are reparameterized in terms of a common scale and two functions, $a(\gamma)$ and $b(\gamma)$, depending on a single skewness parameter γ. By appropriately choosing these functions, this family covers the models presented in Fernández and Steel (1998) and Mudholkar and Hutson (2000), among others. As shown in Jones (2006), all the members of this family are merely reparameterizations of each other. However, we will see that inferential properties can vary for different parameterizations.

Whereas we discuss orthogonality of parameterizations, which is of direct interest for likelihood-based frequentist inference, we will mostly focus on Bayesian inference in this paper. A commonly used prior structure to reflect an absence of prior information is the Jeffreys (or "Jeffreys-rule") prior, which is the reference prior (Berger et al., 2009) in the case of a scalar parameter under asymptotic posterior normality. Under these conditions, Clarke and Barron (1994) showed that this prior asymptotically maximizes the expected information from repeated sampling. The Jeffreys prior is an interesting choice because no subjective parameters have to be elicited and it is invariant under reparameterizations (Jeffreys, 1941).

However, in our two-piece location-scale framework (or its reparameterizations), we show that Jeffreys prior does not lead to a posterior in the wide and empirically interesting class of distributions obtained by skewing scale mixtures of normals. Thus, we consider models that do not imply a reparameterization of the two-piece model, and give an example where this is induced by truncation of the range of γ. In addition, we consider the independence Jeffreys prior, which is shown to lead to a valid posterior in some cases. Simple conditions regarding posterior existence with the independence Jeffreys prior are derived. For example, it is shown that neither the Jeffreys nor the independence Jeffreys prior can be used for Bayesian inference with the skewness transformation of Fernández and Steel (1998) applied to the entire class of scale mixtures of normals.

The structure of this document is as follows: in Section 2 we present the two-piece model and the family of skewed distributions defined in Arellano-Valle et al. (2005). We also derive the Fisher information matrix for these models as well as the Jeffreys and independence Jeffreys priors. In Section 3 we examine posterior existence with these priors in the context of a scale mixture of normals for the underlying symmetric distribution. In Section 4 we conduct a numerical coverage analysis of the 95% credible intervals for various models, which are also applied to a real data set. The final section contains concluding remarks. Proofs of all theorems are given in the Appendix.

2 Fisher information matrix and Jeffreys priors

2.1 Two-piece location-scale models

Let $f(y|\mu,\sigma)$ be an absolutely continuous density with support on $\mathbb{R}$, location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma \in \mathbb{R}^+$, and denote $f\left(\frac{y-\mu}{\sigma}|0,1\right) = f\left(\frac{y-\mu}{\sigma}\right)$. Consider the following "two-piece" density constructed of $f\left(\frac{y-\mu}{\sigma_1}\right)$ truncated to $(-\infty,\mu)$ and $f\left(\frac{y-\mu}{\sigma_2}\right)$ truncated to $[\mu,\infty)$:

$$g(y|\mu,\sigma_1,\sigma_2,\varepsilon) = \frac{2\varepsilon}{\sigma_1} f\left(\frac{y-\mu}{\sigma_1}\right) I_{(-\infty,\mu)}(y) + \frac{2(1-\varepsilon)}{\sigma_2} f\left(\frac{y-\mu}{\sigma_2}\right) I_{[\mu,\infty)}(y), \qquad (1)$$

where $\sigma_1 \in \mathbb{R}^+$ and $\sigma_2 \in \mathbb{R}^+$ are separate scale parameters and $0 < \varepsilon < 1$. In order to get a continuous density, we need to consider the special case where $\varepsilon = \sigma_1/(\sigma_1+\sigma_2)$, so that

$$s(y|\mu,\sigma_1,\sigma_2) = \frac{2}{\sigma_1+\sigma_2}\left[f\left(\frac{y-\mu}{\sigma_1}\right) I_{(-\infty,\mu)}(y) + f\left(\frac{y-\mu}{\sigma_2}\right) I_{[\mu,\infty)}(y)\right]. \qquad (2)$$

Typically, f will be a symmetric density function. In this paper, we will assume f to be symmetric with a single mode at zero, which means that μ is the mode of the density in (2). If we choose f to be normal and Student densities, the distribution in (2) corresponds to split-normal and split-t distributions, respectively, as defined in Geweke (1989). In earlier work, the case with normal f was termed joined half-Gaussian by Gibbons and Mylroie (1973) and two-piece normal by John (1982). We shall denote the model in (2) as the two-piece model. Note that

$$\int_{-\infty}^{\mu} s(y|\mu,\sigma_1,\sigma_2)\,dy = \frac{\sigma_1}{\sigma_1+\sigma_2}, \qquad (3)$$

so that s is skewed about μ if $\sigma_1 \neq \sigma_2$ and the ratio σ_1/σ_2 controls the allocation of mass to each side of μ.

We are mainly interested in the inferential properties of these skewed distributions under the popular Jeffreys priors, but will also briefly discuss orthogonality of their parameters. We use the concept of orthogonality in Cox and Reid (1987), which relates to zeros in the Fisher information matrix of the model. If θ_1 is orthogonal to θ_2, we will denote this as $\theta_1 \perp \theta_2$.

We first calculate the Fisher information matrix and characterize, in terms of the symmetric density f, the cases where this matrix is well defined:

Theorem 1 *Let $s(y|\mu, \sigma_1, \sigma_2)$ be as in (2) and suppose that the following conditions hold*

(i) $\int_0^\infty \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt < \infty,$

(ii) $\int_0^\infty t^2 \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt < \infty,$

(iii) $\lim_{t\to\infty} tf(t) = 0$ *or* $\int_0^\infty tf'(t)dt = -\frac{1}{2}$*, which means that* $f(t)$ *is* $o\left(\frac{1}{t}\right)$.

Then the Fisher information matrix $I(\mu, \sigma_1, \sigma_2)$ *is*

$$\begin{pmatrix} \frac{2\alpha_1}{\sigma_1\sigma_2} & -\frac{2\alpha_3}{\sigma_1(\sigma_1+\sigma_2)} & \frac{2\alpha_3}{\sigma_2(\sigma_1+\sigma_2)} \\ -\frac{2\alpha_3}{\sigma_1(\sigma_1+\sigma_2)} & \frac{\alpha_2}{\sigma_1(\sigma_1+\sigma_2)} + \frac{\sigma_2}{\sigma_1(\sigma_1+\sigma_2)^2} & -\frac{1}{(\sigma_1+\sigma_2)^2} \\ \frac{2\alpha_3}{\sigma_2(\sigma_1+\sigma_2)} & -\frac{1}{(\sigma_1+\sigma_2)^2} & \frac{\alpha_2}{\sigma_2(\sigma_1+\sigma_2)} + \frac{\sigma_1}{\sigma_2(\sigma_1+\sigma_2)^2} \end{pmatrix}, \tag{4}$$

where

$$\begin{aligned} \alpha_1 &= \int_0^\infty \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt, \\ \alpha_2 &= 2\int_0^\infty \left[1 + t\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt = -1 + 2\int_0^\infty t^2 \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt, \\ \alpha_3 &= \int_0^\infty t \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt. \end{aligned}$$

Conditions (i) and (ii) are required for the existence of the expression in (4). Condition (iii) is useful to simplify some expressions and is satisfied by many models of interest. As examples, normal, Student t, logistic, Cauchy, Laplace and exponential power distributions all satisfy $(i) - (iii)$. Note that if α_1, α_2 and α_3 are positive, none of the entries of the Fisher information matrix are zero. Therefore this is a non-orthogonal parameterization.

The Jeffreys prior, proposed by Jeffreys (1941), is defined as the square root of the determinant of the Fisher information matrix. In contrast, the independence Jeffreys prior is defined as the product of the Jeffreys priors for each parameter independently, while treating the others parameters as fixed.

Corollary 1 *If the Fisher information matrix in (4) is non-singular, then the Jeffreys prior for the parameters in (2) is*

$$\pi_J(\mu,\sigma_1,\sigma_2) \quad \propto \quad \frac{1}{\sigma_1\sigma_2(\sigma_1+\sigma_2)}. \tag{5}$$

The independence Jeffreys prior is

$$\pi_I(\mu,\sigma_1,\sigma_2) \quad \propto \quad \frac{\sqrt{[\sigma_1+\alpha_2(\sigma_1+\sigma_2)][\sigma_2+\alpha_2(\sigma_1+\sigma_2)]}}{\sqrt{\sigma_1\sigma_2}(\sigma_1+\sigma_2)^2}. \tag{6}$$

The Jeffreys prior is defined only in the cases when the Fisher information matrix is non-singular. Note that the determinant of the Fisher information matrix can be factored into two terms, one dependent on the parameters and the other dependent on the constants $(\alpha_1,\alpha_2,\alpha_3)$. The former is always positive. The following result gives conditions on the density f that ensure that the second factor does not vanish and the Fisher information matrix is thus non-singular.

Theorem 2 *If the conditions of Theorem* 1 *are satisfied and* $f'(t) \neq 0$ $a.e.$, *then the Fisher information matrix is non-singular.*

In particular, the Fisher information matrix (4) is non-singular if f corresponds to a normal, Laplace, exponential power, logistic, Cauchy or Student t distribution. The structure of the independence Jeffreys prior in (6) assumes that $\alpha_2 > 0$, which will always be the case (see the proof of Theorem 2).

2.2 Reparameterizations of the two-piece model

In order to link the two-piece model in (2) with the family defined in Arellano-Valle et al. (2005), consider the following reparameterization (one-to-one transformation)

$$\begin{aligned}(\mu,\sigma_1,\sigma_2) &\leftrightarrow (\mu,\sigma,\gamma), \\ \mu &= \mu, \\ \sigma_1 &= \sigma b(\gamma), \\ \sigma_2 &= \sigma a(\gamma),\end{aligned} \tag{7}$$

where $\gamma \in \Gamma$, $\sigma > 0$ and $a(\gamma) > 0$ and $b(\gamma) > 0$ are differentiable functions such that

$$0 < |\lambda(\gamma)| < \infty, \text{with } \lambda(\gamma) \equiv \frac{d}{d\gamma}\log\left[\frac{a(\gamma)}{b(\gamma)}\right]. \tag{8}$$

The condition in (8) implies that (7) is a non-singular mapping and is thus necessary for it to be a one-to-one transformation. Then we get the following reparameterized density from (2)

$$s(y|\mu,\sigma,\gamma)=\frac{2}{\sigma[a(\gamma)+b(\gamma)]}\left[f\left(\frac{y-\mu}{\sigma b(\gamma)}\right)I_{(-\infty,\mu)}(y)+f\left(\frac{y-\mu}{\sigma a(\gamma)}\right)I_{[\mu,\infty)}(y)\right]. \quad (9)$$

This expression was presented by Arellano-Valle et al. (2005) as a general class of asymmetric distributions, which includes various skewed distributions presented in the literature. Like Jones (2006), we view (9) with a given choice of f not as a class of densities but as a class of reparameterizations of the same density.

Two parameterizations in terms of the functions $\{a(\gamma),b(\gamma)\}$ have been widely studied: the inverse scale factors (ISF) model (Fernández and Steel, 1998), corresponding to $\{a(\gamma),b(\gamma)\}=\{\gamma,1/\gamma\}$ for $\gamma\in\mathbb{R}^+$ and the ϵ-skew model (Mudholkar and Hutson, 2000), which chooses $\{a(\gamma),b(\gamma)\}=\{1+\gamma,1-\gamma\}$ for $\gamma\in(-1,1)$.

The Fisher information matrix for the reparameterized model in (9) is as follows:

Theorem 3 *Let $f(y|\mu,\sigma)$ be as in Theorem 1. Then the Fisher information matrix $I(\mu,\sigma,\gamma)$ for model (9) is*

$$\begin{pmatrix} \frac{2\alpha_1}{a(\gamma)b(\gamma)\sigma^2} & 0 & \frac{2\alpha_3}{\sigma[a(\gamma)+b(\gamma)]}\left[\frac{a'(\gamma)}{a(\gamma)}-\frac{b'(\gamma)}{b(\gamma)}\right] \\ 0 & \frac{\alpha_2}{\sigma^2} & \frac{\alpha_2}{\sigma}\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right] \\ \frac{2\alpha_3}{\sigma[a(\gamma)+b(\gamma)]}\left[\frac{a'(\gamma)}{a(\gamma)}-\frac{b'(\gamma)}{b(\gamma)}\right] & \frac{\alpha_2}{\sigma}\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right] & \frac{\alpha_2+1}{a(\gamma)+b(\gamma)}\left[\frac{b'(\gamma)^2}{b(\gamma)}+\frac{a'(\gamma)^2}{a(\gamma)}\right]-\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right]^2 \end{pmatrix}.$$

The fact that the elements I_{12} and I_{21} are zero indicates that this reparameterization is interesting because it induces orthogonality between the parameters μ and σ for any choice of $\{a(\gamma),b(\gamma)\}$. In addition, by appropriately choosing the pair of functions $\{a(\gamma),b(\gamma)\}$ we can generate more zero entries in the Fisher information matrix, as shown in the following corollary.

Corollary 2 *If $\frac{d}{d\gamma}\log[a(\gamma)+b(\gamma)]=0$, then $I_{23}=I_{32}=0$. In particular if $a(\gamma)+b(\gamma)$ is constant, then $I_{23}=I_{32}=0$.*

Note that if $\alpha_3>0$, then $I_{13}=I_{31}=0$ only if $a(\gamma)\propto b(\gamma)$ which does not satisfy (8). Jones and Anaya-Izquierdo (2010) analysed the zeroes of the expectation of the Hessian matrix of (μ,σ,γ) in model (9) augmented with an extra parameter to model the properties of f. They also found that $\mu\perp\sigma$ and if $a(\gamma)+b(\gamma)$ is constant then $\sigma\perp\gamma$ as in Corollary 2.

The corresponding Jeffreys prior and independence Jeffreys prior for the parameterization in (7) are given in the following result.

Corollary 3 *If the Fisher information matrix is non-singular, then the Jeffreys prior for the parameters in (9) is*

$$\pi_J(\mu,\sigma,\gamma) \propto \frac{|a'(\gamma)b(\gamma)-a(\gamma)b'(\gamma)|}{\sigma^2 a(\gamma)b(\gamma)[a(\gamma)+b(\gamma)]} = \frac{|\lambda(\gamma)|}{\sigma^2[a(\gamma)+b(\gamma)]}, \tag{10}$$

where $\lambda(\gamma)$ *was defined in (8). The independence Jeffreys prior is*

$$\pi_I(\mu,\sigma,\gamma) \quad \propto \quad \frac{1}{\sigma}\sqrt{\frac{\alpha_2+1}{a(\gamma)+b(\gamma)}\left[\frac{b'(\gamma)^2}{b(\gamma)}+\frac{a'(\gamma)^2}{a(\gamma)}\right]-\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right]^2}. \tag{11}$$

Conditions to ensure non-singularity of the Fisher information matrix for the parameterization in (9) are similar to those obtained for the two-piece model (2) in Theorem 2. The only difference is that in this case we have to choose a pair of functions $\{a(\gamma), b(\gamma)\}$ such that (7) corresponds to a non-singular transformation:

Corollary 4 *If the conditions of Theorem* 1 *are satisfied,* $f'(t) \neq 0$ *a.e., and (8) holds, then the Fisher information matrix corresponding to model (9) is non-singular.*

Due to the invariance property of the Jeffreys prior there is a one-to-one relationship between (5) and (10). On the other hand, the independence Jeffreys prior is not invariant under reparameterizations, so the properties of this prior are dependent on the choice of $\{a(\gamma), b(\gamma)\}$.

Now we will briefly discuss the inverse scale factors and ϵ-skew models.

2.2.1 Inverse scale factors model

The ISF model corresponds to choosing $\{a(\gamma)=\gamma, b(\gamma)=1/\gamma\}$, $\gamma \in \mathbb{R}^+$ in (9), so that from Theorem 3 the Fisher information matrix of the parameters (μ,σ,γ) is

$$I(\mu,\sigma,\gamma) = \begin{pmatrix} \frac{2\alpha_1}{\sigma^2} & 0 & \frac{4\alpha_3}{\sigma(\gamma^2+1)} \\ 0 & \frac{\alpha_2}{\sigma^2} & \frac{\alpha_2\left(\gamma^2-1\right)}{\sigma(\gamma^3+\gamma)} \\ \frac{4\alpha_3}{\sigma(\gamma^2+1)} & \frac{\alpha_2\left(\gamma^2-1\right)}{\sigma(\gamma^3+\gamma)} & \frac{\alpha_2}{\gamma^2}+\frac{4}{\left(\gamma^2+1\right)^2} \end{pmatrix}. \tag{12}$$

If the Fisher information matrix in (12) is non-singular, then the Jeffreys prior for the ISF model is

$$\pi_J(\mu,\sigma,\gamma) \quad \propto \quad \frac{1}{\sigma^2\left(1+\gamma^2\right)}, \tag{13}$$

which has a finite integral over $\gamma \in \mathbb{R}^+$, but is improper in terms of μ and σ. The independence Jeffreys prior is

$$\pi_I(\mu, \sigma, \gamma) \quad \propto \quad \frac{1}{\sigma}\sqrt{\frac{\alpha_2}{\gamma^2} + \frac{4}{(\gamma^2+1)^2}}, \tag{14}$$

which is not integrable in any of the parameters.

2.2.2 ϵ-skew model

For the ϵ-skew model we choose $\{a(\gamma) = 1-\gamma, b(\gamma) = 1+\gamma\}$ in (9), where $\gamma \in (-1, 1)$, leading to the Fisher information matrix

$$I(\mu, \sigma, \gamma) = \begin{pmatrix} \frac{2\alpha_1}{\sigma^2(1-\gamma^2)} & 0 & -\frac{2\alpha_3}{\sigma(1-\gamma^2)} \\ 0 & \frac{\alpha_2}{\sigma^2} & 0 \\ -\frac{2\alpha_3}{\sigma(1-\gamma^2)} & 0 & \frac{\alpha_2+1}{1-\gamma^2} \end{pmatrix}. \tag{15}$$

The ϵ-skew parameterization satisfies the condition in Corollary 2 and thus its Fisher information matrix has four zeroes. This feature simplifies classical inference. For example, in the cases where f is normal or Laplace, the corresponding ϵ-skew model leads to maximum likelihood estimators in closed form (Mudholkar and Hutson, 2000; Arellano-Valle et al., 2005).

Provided the Fisher information matrix in (15) is non-singular, the Jeffreys prior for the ϵ-skew model is

$$\pi_J(\mu, \sigma, \gamma) \quad \propto \quad \frac{1}{\sigma^2(1-\gamma^2)}, \tag{16}$$

which is not integrable in any of the parameters. The independence Jeffreys prior is

$$\pi_I(\mu, \sigma, \gamma) \quad \propto \quad \frac{1}{\sigma\sqrt{1-\gamma^2}}, \tag{17}$$

which has a finite integral over $\gamma \in (-1, 1)$, but does not integrate in μ and σ. Note that for this model the independence Jeffreys prior does not depend on f (through α_2), in contrast with the priors for the two-piece model in (6) and the ISF model in (14).

In the different models mentioned above, the skewness parameter γ does not have the same interpretation. This makes it particularly difficult to compare models and to propose compatible priors on γ. It is therefore helpful to introduce a measure of skewness which has a common meaning for all models. In particular, we use the skewness measure with respect to the mode from Arnold and Groeneveld (1995), defined as

Definition 1 *The Arnold-Groeneveld measure of skewness for a distribution function S corresponding to a unimodal density with the mode at M is defined as*

$$AG = 1 - 2S(M).$$

The AG measure takes values in $(-1, 1)$ and provides information about the allocation of mass to each side of the mode. Positive values of AG indicate right skewness while negative values indicate left skewness. From (3) it is immediate that for the two-piece model $AG = (\sigma_2 - \sigma_1)/(\sigma_1 + \sigma_2)$, which only depends on the two scales and not on the properties of f. Similarly, for the parameterization in Arellano-Valle et al. (2005) in (9) the AG skewness measure has a closed form which only depends on γ:

$$AG(\gamma) = \frac{a(\gamma) - b(\gamma)}{a(\gamma) + b(\gamma)}.$$

For the special case of the ISF model in Subsection 2.2.1, this reduces to

$$AG(\gamma) = \frac{\gamma^2 - 1}{\gamma^2 + 1},$$

while for the ϵ-skew model in Subsection 2.2.2 we obtain $AG(\gamma) = -\gamma$.

In both examples above, the AG skewness measure is a monotonic function of γ, so we can meaningfully interpret γ as a skewness parameter. In general, we will be mostly interested in parameterizations such that this is the case, which can be characterized as follows:

Theorem 4 *Let s, $a(\gamma)$ and $b(\gamma)$ be as in (9), then for any unimodal density f*

- *$AG(\gamma)$ is increasing if and only if $\lambda(\gamma) > 0$.*
- *$AG(\gamma)$ is decreasing if and only if $\lambda(\gamma) < 0$.*

3 Inference

In this section we will present necessary and/or sufficient conditions for the properness of the posterior distribution of the parameters of the two-piece models considered when using the priors presented in the last section. Throughout this section we will assume that we have observed a sample of n independent replications from either (2) or (9) and that all the observations are different, as we are dealing with continuous sampling distributions. Most of the results in this section are for the case where the underlying symmetric distribution (with density f) belongs to the wide class of scale mixtures of normals. For those (rare) cases where such an f does not lead to a nonsingular information matrix (see Theorem 2 and Corollary 4) or a well-defined independence Jeffreys prior, we could either implicitly impose any necessary restrictions upon the class, or we could simply consider the results as valid for the entire class of scale mixtures of normals but with a prior structure that is not strictly the (independence) Jeffreys prior (but certainly inspired by the latter). However, most cases of practical interest will correspond to an f that allows for a straightforward interpretation of the results in this section.

3.1 Independence Jeffreys prior

The independence Jeffreys prior is not invariant under reparameterizations. Therefore if we consider one-to-one transformations as in (7), we need to analyse the properness of the posterior distribution of (μ, σ, γ) for each specific choice of $\{a(\gamma), b(\gamma)\}$.

Theorem 5 *The posterior distribution of the parameters* $(\mu, \sigma_1, \sigma_2)$ *of model (2) is proper using the independence Jeffreys prior (6) if* f *is a scale mixture of normals and the number of observations* $n \geq 2$*.*

Scale mixtures of normals contain some important distributions, such as the normal, Student t with ν degrees of freedom, logistic, Laplace, Cauchy and the exponential power family with power $1 \leq q < 2$. Thus, for this wide and practically important class of distributions the two-piece model in (2) with the independence Jeffreys prior leads to valid inference in any sample of two or more observations.

We can derive a similar existence result for the model in (9) within a class of prior distributions:

Theorem 6 *If* f *is a scale mixture of normals in the model (9), then for any parameterization* $\{a(\gamma), b(\gamma)\}$ *the posterior distribution of* (μ, σ, γ) *is proper for* $n \geq 2$ *under any prior structure of the form* $\pi(\mu, \sigma, \gamma) \propto \sigma^{-1}\pi(\gamma)$*, where* $\pi(\gamma)$ *is proper.*

This Theorem implies that a posterior will exist for the ϵ-skew model under the independence Jeffreys prior in (17), as this prior is a member of the class in Theorem 6.

However, for the ISF model the independence Jeffreys prior does not integrate in γ and we can show that a posterior does not exist in this case:

Theorem 7 *If f is a scale mixture of normals in (9) and $\{a(\gamma), b(\gamma)\}$ are as in the inverse scale factors model, then the posterior distribution of (μ, σ, γ) is improper under the independence Jeffreys prior (14).*

Theorems 6 and 7 emphasize the relevance of the choice of the functions $\{a(\gamma), b(\gamma)\}$ for the properness of the posterior distribution of (μ, σ, γ) when using the independence Jeffreys prior. The fact that the ISF model does not allow for inference with the independence Jeffreys prior is rather surprising since this prior almost always leads to proper posteriors, and the ISF model is quite a straightforward extension of the usual location-scale model. Subsection 3.3 will shed more light on this.

3.2 Jeffreys prior

If we consider functions f, $a(\gamma)$ and $b(\gamma)$ such that the Fisher information matrix is non-singular (see Theorem 2 and Corollary 4) we can think of making inference using the Jeffreys prior. We now study the properness of the posterior distribution of the parameters (μ, σ, γ) when we choose this prior. An important feature of this prior is the invariance under one-to-one reparameterizations. Therefore, the results regarding the properness of the posterior of (μ, σ, γ) for any choice of $\{a(\gamma), b(\gamma)\}$ in model (9) that corresponds to a one-to-one transformation in (7) are the same and also applicable to the posterior of $(\mu, \sigma_1, \sigma_2)$ in model (2).

Theorem 8 *Let s be as in (9), assume that f is a scale mixture of normals and consider the Jeffreys prior (10) for the parameters of this model. Then, for $n \geq 2$, a necessary condition for the properness of the posterior distribution of (μ, σ, γ) is*

$$\int_{\Gamma} \left[\frac{a(\gamma)}{a(\gamma) + b(\gamma)} \right]^{n+1} |\lambda(\gamma)| \, d\gamma < \infty, \tag{18}$$

with $\lambda(\gamma)$ defined as in (8).

Corollary 5 *Consider sampling from (9) with f a scale mixture of normals and $\{a(\gamma), b(\gamma)\}$ as in the inverse scale factors model, then the posterior distribution of (μ, σ, γ) is improper using the Jeffreys prior (10). As a consequence, for any pair of functions $\{a(\gamma), b(\gamma)\}$ such that the mapping $(\mu, \sigma_1, \sigma_2) \leftrightarrow (\mu, \sigma, \gamma)$ is one-to-one, the posterior distribution of (μ, σ, γ) is improper using the Jeffreys prior (10).*

Proof. *We can verify that the necessary condition (18) is not satisfied for these functions.*

This corollary implies that we can not conduct Bayesian inference for the parameters of this type of skewed distributions using the Jeffreys prior. It is rather rare to find that the Jeffreys prior does not lead to a proper posterior, and it is somewhat surprising to find that we can not use this prior in these rather simple classes of two-piece distributions with only three parameters.

Because the Jeffreys prior is invariant under reparameterization, its use is thus prohibited in any one-to-one reparameterization of the two-piece models in (2) or (9). However, one way to get around this problem is to choose functions $\{a(\gamma), b(\gamma)\}$ such that the mapping $(\mu, \sigma, \gamma) \mapsto (\mu, \sigma_1, \sigma_2)$ is not one-to-one, but hopefully still of some interest for modelling. Another way to produce a proper posterior distribution when using the Jeffreys prior is to restrict Γ such that $\lambda(\gamma)$ is absolutely integrable.

Theorem 9 *Let s be as in (9) where f is normal or Laplace. Consider the Jeffreys prior (10) for the parameters of this model. Let $\{a(\gamma), b(\gamma)\}$ be continuously differentiable functions for $\gamma \in \Gamma$ such that*

$$0 < \int_{\Gamma} |\lambda(\gamma)| \, d\gamma < \infty. \tag{19}$$

Then we have the following results

(i) The posterior distribution of (μ, σ, γ) is proper for $n \geq 2$.

(ii) The mapping $(\mu, \sigma, \gamma) \mapsto (\mu, \sigma_1, \sigma_2)$ is not one-to-one.

(iii) If Γ is an interval (not necessarily bounded) and $AG(\gamma)$ is monotonic, then $AG(\gamma)$ is not surjective.

First, we considered forcing existence of the posterior through the choice of the functions $\{a(\gamma), b(\gamma)\}$, in particular such that the ratio $a(\gamma)/b(\gamma)$ is bounded, which excludes a one-to-one reparameterization in (7). However, the examples we generated in this way did not lead to implied priors on AG that could be of interest to practitioners.

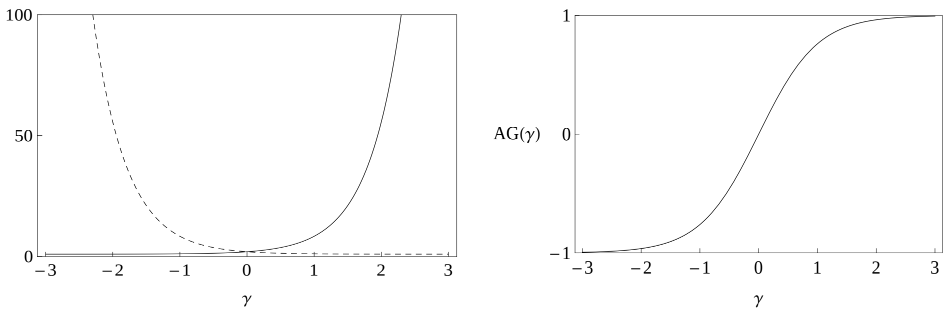
It is actually easier to generate examples of practical relevance if we consider restricting the parameter space of γ in the context of functions $\{a(\gamma), b(\gamma)\}$ that would not lead to a posterior with unrestricted γ. The following is such an example.

Example 1 (Logistic AG) *Consider* $a(\gamma) = 1 + \exp(2\gamma)$, $b(\gamma) = 1 + \exp(-2\gamma)$ *for* $\gamma \in \mathbb{R}$, *then*

$$\begin{aligned} AG(\gamma) &= \tanh(\gamma), \\ \lambda(\gamma) &= 2 \\ \pi_J(\mu, \sigma, \gamma) &\propto \frac{1}{\sigma^2} \operatorname{sech}(\gamma)^2. \end{aligned} \tag{20}$$

In addition, the functions $a(\gamma)$, $b(\gamma)$ *and* $AG(\gamma)$ *are monotonic* $\forall\ \gamma \in \mathbb{R}$, *the Jeffreys prior in (20) implies that* $AG \sim \mathrm{Unif}(-1, 1)$ *and* $AG : \mathbb{R} \mapsto (-1, 1)$. *Clearly,* $\lambda(\gamma)$ *is not integrable on* $\mathbb{R}$, *but if we restrict* $\gamma \in [-B, B]$ *for some* $0 < B < \infty$, *then we can use the Jeffreys prior* (20) *for making inference on* (μ, σ, γ) *for normal of Laplace* f *and* $AG : \mathbb{R} \mapsto [\tanh(-B), \tanh(B)]$. *Figure* 1 *presents the functions* $a(\gamma)$, $b(\gamma)$ *and* $AG(\gamma)$ *and Figure* 2 *shows the factor depending on* γ *in the Jeffreys prior for* $B = 3$. *The induced prior on* AG *is a Uniform over the range* $[\tanh(-B), \tanh(B)] = [-0.995, 0.995]$.

We will call the model in Example 1 the "logistic AG model" as AG(γ) is a logistic function of γ transformed to take values in the interval (-1,1) for $\gamma \in \mathbb{R}$. The choice of $a(\gamma)$ and $b(\gamma)$ does lead to a one-to-one transformation in (7) when $\gamma \in \mathbb{R}$, but not if γ is restricted to a bounded interval: then the ratio $a(\gamma)/b(\gamma)$ is also bounded and this precludes a one-to-one mapping. Note that $a(\gamma)$ and $b(\gamma)$ satisfy the condition $a(\gamma) + b(\gamma) = a(\gamma)b(\gamma)$, which induces a really interesting structure on the Jeffreys prior, namely that it implies a uniform prior in terms of the AG measure. This might be an attractive prior for practitioners to use in the absence of strong prior information.



(a) (b)

Figure 1: (a) $a(\gamma)$ (solid line) and $b(\gamma)$ (dashed line); (b) $AG(\gamma)$.

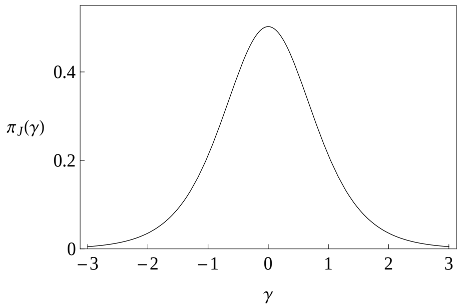


Figure 2: Jeffreys prior (20) as a function of γ.

3.3 Intuitive explanation

As mentioned before, the lack of a posterior under a commonly used prior in what is essentially a very simply generalisation of a standard location-scale model can be considered surprising. Thus, we offer a few explanatory comments in this subsection. These are not meant to be formal proofs (they can be found in the Appendix), but merely intuitive ideas that help us understand what drives the main results we have found in the previous subsections.

In the context of the two-piece model in (2), it is easy to see that as σ_1 tends to zero, the sampling density tends to the half density on $[\mu, \infty)$ with scale σ_2. Thus, the likelihood will be constant in σ_1 in the neighbourhood of zero. This means the prior needs to integrate in that neighbourhood for a posterior to exist. If we consider the independent Jeffreys prior in (6) it behaves like $\sigma_1^{-1/2}$ for small σ_1 and this integrates close to zero. Indeed, we have a posterior in this case. However, the Jeffreys prior in (5) behaves like $1/\sigma_1$ for small σ_1 and this does not integrate, thus precluding a posterior. Of course, similar arguments hold in the case of small σ_2.

In the case of the reparameterized model in (9), we have a potential problem if one of the scales, say, $\sigma a(\gamma)$ goes to zero. If then the ratio $b(\gamma)/a(\gamma)$ has an upper bound, this will necessarily imply that both scales tend to zero, so the model behaves like a standard location-scale model which leads to a proper posterior under the Jeffreys prior. This is the case explored in Theorem 9 and Example 1. If, however, the ratio between the functions $a(\gamma)$ and $b(\gamma)$ is not bounded and (7) defines a one-to-one mapping, we will have no posterior with the Jeffreys prior due to the invariance of this prior under reparameterization, and it depends on the particular choice of functions $\{a(\gamma), b(\gamma)\}$ whether the independence Jeffreys prior will lead to a posterior. It is helpful to transform the parameters back to those of the two-piece model in (2). Then, for the ϵ-skew model the independence Jeffreys prior in (17) can be shown to behave like $\sigma_i^{-1/2}$ for small $\sigma_i, i = 1, 2$, which is integrable close to zero, and the posterior is well-defined. On the other hand, the independence Jeffreys prior for the ISF model in (14) behaves like $1/\sigma_i$ for small $\sigma_i, i = 1, 2$, which does not integrate in a neighbourhood of zero and precludes posterior existence.

4 Numerical results

4.1 Simulation study

In this section we investigate the empirical coverage of the 95% posterior credible intervals, defined by the 2.5th and 97.5th percentiles. We simulate $N = 10,000$ datasets of size $n = 30, 100$ and 1000 from various sampling models where we take f to be a normal distribution throughout, and analyse these data using the corresponding Bayesian model. Model 1 consists of the two-piece model (2) and the independence Jeffreys prior (6). Model 2 corresponds to (9) using $\{a(\gamma), b(\gamma)\}$ of the ϵ-skew model under the independence Jeffreys prior. Model 3 is the logistic AG model of Example 1 for $\gamma \in [-B, B]$ with the Jeffreys prior in (20). For each of these N datasets, a sample of size $3,000$ was obtained from the posterior distribution using a Markov chain Monte Carlo sampler after a burn-in period of $5,000$ iterations and thinned to every 50th iteration. Finally, the proportion of 95% credible intervals that include the true value of the parameter was calculated. Results are presented in Tables 1-4. For Model 3 we know that the truncation to a finite interval is what makes the posterior well-defined. To investigate how sensitive the results are to the particular value chosen for B, we have used various values.

Sample size	$n = 30$		$n = 100$		$n = 1000$	
Parameters	$\sigma_1 = 2.0$ $\sigma_2 = 0.5$	$\sigma_1 = 0.66$ $\sigma_2 = 1.50$	$\sigma_1 = 2.0$ $\sigma_2 = 0.5$	$\sigma_1 = 0.66$ $\sigma_2 = 1.50$	$\sigma_1 = 2.0$ $\sigma_2 = 0.5$	$\sigma_1 = 0.66$ $\sigma_2 = 1.50$
μ	0.9761	0.9672	0.9711	0.9559	0.9482	0.9534
σ_1	0.9606	0.9513	0.9741	0.9581	0.9473	0.9492
σ_2	0.9748	0.9711	0.9606	0.9512	0.9485	0.9505

Table 1: Coverage proportions. Mixture model with independence Jeffreys prior (Model 1)

All models lead to coverage probabilities above the nominal level for samples of size $n = 30$, especially in the case of σ for Model 3. Once we increase the sample size

Sample size	$n = 30$		$n = 100$		$n = 1000$	
Parameter	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$
μ	0.9710	0.9699	0.9543	0.9552	0.9469	0.9485
σ	0.9591	0.9602	0.9475	0.9452	0.9527	0.9541
γ	0.9707	0.9691	0.9580	0.9575	0.9484	0.9519

Table 2: Coverage proportions. ϵ-skew model with independence Jeffreys prior (Model 2)

Sample size	$n = 30$		$n = 100$		$n = 1000$	
Parameter	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$
μ	0.9673	0.9641	0.9493	0.9530	0.9481	0.9493
σ	0.9949	0.9908	0.9522	0.9600	0.9480	0.9473
γ	0.9640	0.9654	0.9488	0.9520	0.9477	0.9469

Table 3: Coverage proportions. Logistic AG model with Jeffreys prior (Model 3) and $B = 3$

Sample size	$n = 30$		$n = 100$		$n = 1000$	
Parameter	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$
μ	0.9680	0.9652	0.9486	0.9488	0.9494	0.9450
σ	0.9905	0.9916	0.9575	0.9529	0.9504	0.9417
γ	0.9659	0.9641	0.9517	0.9517	0.9525	0.9447

Table 4: Coverage proportions. Logistic AG model with Jeffreys prior (Model 3) and $B = 30$

to $n = 100$, the coverage is quite close to the nominal value, except for one setting for Model 1, where the coverage is still a bit high. As we further increase to samples of 1000 observations, all cases lead to coverage very close to 95%, as we would expect. For Model 3, the choice of B (among reasonable values) does not seem to have any noticeable effect. Overall, the frequentist coverage properties of the models examined are pretty good, with perhaps Model 2 displaying the best performance.

4.2 Application to real data

Consider the data set presented in Mudholkar and Hutson (2000) which contains the heights of 219 of the world's volcanoes. We use Models 1 to 3 described in the previous subsection as well as the skew-normal model of (Azzalini, 1985), which will be denoted as Model 4, given by

$$s(y|\mu, \sigma, \lambda) = \frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right),$$

using the prior

$$\pi(\mu, \sigma, \lambda) \propto \sigma^{-1}\pi(\lambda). \tag{21}$$

The structure of this prior, using the Jeffreys prior of λ derived in the model without location and scale parameters for $\pi(\lambda)$, was proposed in Liseo and Loperfido (2006), who also prove existence of the posterior under this prior. Bayes and Branco (2007) show that the Jeffreys prior of λ can be approximated by a Student t distribution with $1/2$ degrees of freedom, which is what was used for our calculations.

A sample of size $10,000$ was drawn from the posterior distribution after a burn-in period of $50,000$ iterations with a thinning of 100 iterations for all models.

Figure 3 shows the fit of the predictive densities of the various models overplotted with the data histogram. Models 1-3 lead to almost overlapping predictives, but the one for Model 4 is slightly different.

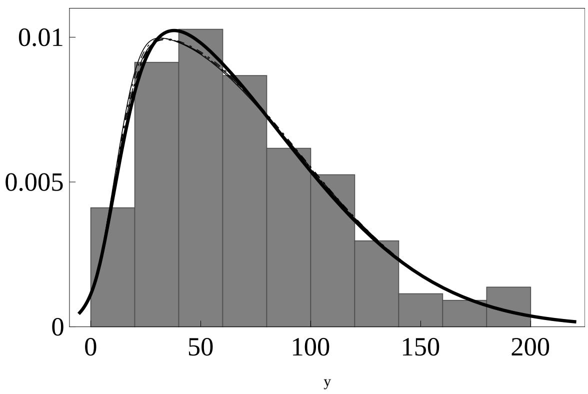


Figure 3: Predictive distributions and data histogram: Model 1 (continuous line); Model 2 (long-dashed line); Model 3, $B = 3$ (dashed line); Model 3, $B = 10$ (dotted line); Model 3, $B = 30$ (dotted-dashed line); Model 4 (bold line).

Bayes factors can be computed between Models 1, 2 and 4 despite the arbitrary integrating constant (improperness) of the prior, since the prior has a product structure with an improper factor (in σ and μ) which is common to all models, and the factor corresponding to the skewness parameter is integrable and thus properly normalised. As Model 3 does not share the same factor in σ it can not be compared with the other models through Bayes factors. The marginal likelihoods needed in the calculation of

Bayes factors are estimated using the generalised harmonic mean estimator (Chopin and Robert, 2010), with an importance function chosen to resemble the posterior but with thinner tails. The resulting Bayes factors are close to unity.

5 Concluding Remarks

We consider the class of univariate continuous two-piece distributions, which are often used to modify a symmetric location-scale model to allow for skewness, and its reparameterized versions as presented in Arellano-Valle et al. (2005), where we can identify a location, a scale and a skewness parameter. A number of well-known distributions correspond to particular choices of this parameterization. In particular, we focus on Bayesian inference in these models using Jeffreys or the independence Jeffreys prior. We prove that these models do not lead to valid posterior inference under Jeffreys prior for any underlying symmetric distribution in the class of scale mixture of normals. As an ad-hoc fix, we show that modifying Jeffreys prior by truncating the support of the skewness parameter can lead to posterior existence. A more fundamental solution is to use the independence Jeffreys prior instead, which is shown to lead to a valid posterior for some parameterizations of these sampling models. For a number of models that lead to valid inference, we compute empirical coverage probabilities of the posterior credible intervals. This reveals a mostly satisfactory behaviour of these models. Finally, we apply the models, as well as an alternative skewed distribution due to Azzalini (1985), to some real data.

It is important to stress that the three-parameter sampling models examined here are quite simple modifications of the standard location-scale model, and that the Jeffreys prior is a very commonly used prior in the absence of subjective prior information. The fact that the combination of these sampling models with a Jeffreys prior does not lead to a posterior is somewhat surprising and definitely relevant for statistical practice, as these models seem attractive options to deal with skewed data. The better properties of the independence Jeffreys prior are in line with statistical folklore: Jeffreys (1961, p. 182) himself preferred this prior for location-scale problems, and in the univariate normal case, the independence Jeffreys is a matching prior (Berger and Sun, 2008). Even with this prior, however, problems of posterior existence can occur, depending on which parameterization we choose. Similar problems of nonexistence under the independence Jeffreys prior also occur for Birnbaum-Sanders distributions (see Xu and Tang, 2011). Ongoing research examines other "non-subjective" prior structures for use with two-piece distributions which can be attractive due to their mathematical properties and their practicality.

Appendix: Proofs

Proof of Theorem 1

The first partial derivatives of $\log[s(y|,\mu,\sigma,\gamma)]$ are given by

$$\begin{aligned}
\frac{\partial}{\partial\mu}\log[s(y|,\mu,\sigma_1,\sigma_2)] &= -\frac{1}{\sigma_1}\frac{f'\left(\frac{y-\mu}{\sigma_1}\right)}{f\left(\frac{y-\mu}{\sigma_1}\right)}I_{(-\infty,\mu)}(y) - \frac{1}{\sigma_2}\frac{f'\left(\frac{y-\mu}{\sigma_2}\right)}{f\left(\frac{y-\mu}{\sigma_2}\right)}I_{[\mu,\infty)}(y),\\
\frac{\partial}{\partial\sigma_1}\log[s(y|,\mu,\sigma_1,\sigma_2)] &= -\frac{1}{\sigma_1+\sigma_2} - \frac{y-\mu}{\sigma_1^2}\frac{f'\left(\frac{y-\mu}{\sigma_1}\right)}{f\left(\frac{y-\mu}{\sigma_1}\right)}I_{(-\infty,\mu)}(y),\\
\frac{\partial}{\partial\sigma_2}\log[s(y|,\mu,\sigma_1,\sigma_2)] &= -\frac{1}{\sigma_1+\sigma_2} - \frac{y-\mu}{\sigma_2^2}\frac{f'\left(\frac{y-\mu}{\sigma_2}\right)}{f\left(\frac{y-\mu}{\sigma_2}\right)}I_{[\mu,\infty)}(y).
\end{aligned}$$

Then the entries of the Fisher information matrix of (μ,σ_1,σ_2) are given by

$$\begin{aligned}
I_{11} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|,\mu,\sigma_1,\sigma_2)]\right)^2\right] = \frac{2\alpha_1}{\sigma_1\sigma_2},\\
I_{22} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma_1}\log[s(y|,\mu,\sigma_1,\sigma_2)]\right)^2\right] = \frac{\alpha_2}{\sigma_1(\sigma_1+\sigma_2)} + \frac{\sigma_2}{\sigma_1(\sigma_1+\sigma_2)^2},\\
I_{33} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma_2}\log[s(y|,\mu,\sigma_1,\sigma_2)]\right)^2\right] = \frac{\alpha_2}{\sigma_2(\sigma_1+\sigma_2)} + \frac{\sigma_1}{\sigma_2(\sigma_1+\sigma_2)^2},\\
I_{12} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|,\mu,\sigma_1,\sigma_2)]\right)\left(\frac{\partial}{\partial\sigma_1}\log[s(y|,\mu,\sigma_1,\sigma_2)\right)\right] = -\frac{2\alpha_3}{\sigma_1(\sigma_1+\sigma_2)},\\
I_{13} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|,\mu,\sigma_1,\sigma_2)]\right)\left(\frac{\partial}{\partial\sigma_2}\log[s(y|,\mu,\sigma_1,\sigma_2)\right)\right] = \frac{2\alpha_3}{\sigma_2(\sigma_1+\sigma_2)},\\
I_{23} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma_1}\log[s(y|,\mu,\sigma_1,\sigma_2)]\right)\left(\frac{\partial}{\partial\sigma_2}\log[s(y|,\mu,\sigma_1,\sigma_2)\right)\right] = -\frac{1}{(\sigma_1+\sigma_2)^2}.
\end{aligned}$$

□

Proof of Theorem 2

The determinant of the Fisher information matrix is

$$|I(\mu, \sigma_1, \sigma_2)| = \frac{2\alpha_2\left(\alpha_1 + \alpha_1\alpha_2 - 2\alpha_3^2\right)}{\sigma_1^2\sigma_2^2(\sigma_1+\sigma_2)^2}.$$

We will first prove that $\alpha_2 > 0$. From the definition of α_2 it can only be zero if $1 + tf'(t)/f(t) = 0$ whenever $f(t) > 0$. This means that $f(t) = -tf'(t)$ and this only happens if $f(t) = K/t$ for any positive K. The latter, however, is not a probability density function on $\mathbb{R}$. Thus, α_2 can not be zero.

Next, we will prove that $\alpha_1(1+\alpha_2) > 2\alpha_3^2$. Applying the Cauchy-Schwarz inequality we have $\alpha_1(1+\alpha_2) \geq 2\alpha_3^2$. We will show that this is a strict inequality. The condition in Theorem 2 implies that

$$0 < \int_0^\infty t\left[\frac{f'(t)}{f(t)}\right]^2 f(t)\,dt.$$

Let

$$\phi(t) = \left|\frac{f'(t)}{\sqrt{f(t)}}\right| > 0 \text{ a.e. and } \psi(t) = t\left|\frac{f'(t)}{\sqrt{f(t)}}\right| > 0 \text{ a.e.}$$

Note that $[\beta\phi(t)+\psi(t)]^2 > 0$ a.e. for any $\beta \in \mathbb{R}$, and thus

$$0 < \int_0^\infty [\beta\phi(t)+\psi(t)]^2\,dt = \beta^2\int_0^\infty \phi^2(t)\,dt + 2\beta\int_0^\infty \phi(t)\psi(t)\,dt + \int_0^\infty \psi^2(t)\,dt.$$

This is a polynomial of degree 2 in β with positive coefficients and no real roots, implying that the discriminant is negative, so that

$$\left[\int_0^\infty t\left[\frac{f'(t)}{f(t)}\right]^2 f(t)\,dt\right]^2 < \left[\int_0^\infty t^2\left[\frac{f'(t)}{f(t)}\right]^2 f(t)\,dt\right]\left[\int_0^\infty \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\,dt\right].$$

□

Proof of Theorem 3

The first partial derivatives of $\log[s(y|,\mu,\sigma,\gamma)]$ are given by

$$\begin{aligned}
\frac{\partial}{\partial\mu}\log[s(y|,\mu,\sigma,\gamma)] &= -\frac{1}{\sigma b(\gamma)}\frac{f'\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}I_{(-\infty,\mu)}(y) - \frac{1}{\sigma a(\gamma)}\frac{f'\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}I_{[\mu,\infty)}(y),\\
\frac{\partial}{\partial\sigma}\log[s(y|,\mu,\sigma,\gamma)] &= -\frac{1}{\sigma} - \frac{y-\mu}{\sigma^2 b(\gamma)}\frac{f'\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}I_{(-\infty,\mu)}(y) - \frac{y-\mu}{\sigma^2 a(\gamma)}\frac{f'\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}I_{[\mu,\infty)}(y),\\
\frac{\partial}{\partial\gamma}\log[s(y|,\mu,\sigma,\gamma)] &= -\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)} - \frac{y-\mu}{\sigma}\frac{b'(\gamma)}{b(\gamma)^2}\frac{f'\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}I_{(-\infty,\mu)}(y)\\
&- \frac{y-\mu}{\sigma}\frac{a'(\gamma)}{a(\gamma)^2}\frac{f'\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}I_{[\mu,\infty)}(y).
\end{aligned}$$

Thus, the entries of the Fisher information matrix of (μ,σ,γ) are

$$\begin{aligned}
I_{11} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|,\mu,\sigma,\gamma)]\right)^2\right] = \frac{2\alpha_1}{a(\gamma)b(\gamma)\sigma^2},\\
I_{22} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma}\log[s(y|,\mu,\sigma,\gamma)]\right)^2\right] = \frac{\alpha_2}{\sigma^2},\\
I_{33} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\gamma}\log[s(y|,\mu,\sigma,\gamma)]\right)^2\right] = \frac{\alpha_2+1}{a(\gamma)+b(\gamma)}\left[\frac{b'(\gamma)^2}{b(\gamma)}+\frac{a'(\gamma)^2}{a(\gamma)}\right] - \left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right]^2,\\
I_{12} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|,\mu,\sigma,\gamma)]\right)\left(\frac{\partial}{\partial\sigma}\log[s(y|,\mu,\sigma,\gamma)\right)\right] = 0,\\
I_{13} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|,\mu,\sigma,\gamma)]\right)\left(\frac{\partial}{\partial\gamma}\log[s(y|,\mu,\sigma,\gamma)\right)\right]\\
&= \frac{2\alpha_3}{\sigma[a(\gamma)+b(\gamma)]}\left[\frac{a'(\gamma)}{a(\gamma)}-\frac{b'(\gamma)}{b(\gamma)}\right],\\
I_{23} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma}\log[s(y|,\mu,\sigma,\gamma)]\right)\left(\frac{\partial}{\partial\gamma}\log[s(y|,\mu,\sigma,\gamma)\right)\right]\\
&= \frac{\alpha_2}{\sigma}\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right].
\end{aligned}$$

□

Proof of Theorem 4

Note that

$$\frac{d}{d\gamma}AG(\gamma) = 2\frac{a'(\gamma)b(\gamma) - a(\gamma)b'(\gamma)}{[a(\gamma)+b(\gamma)]^2} = 2\frac{a(\gamma)b(\gamma)\lambda(\gamma)}{[a(\gamma)+b(\gamma)]^2},$$

so that

$$\frac{dAG(\gamma)}{d\gamma} > 0 \Leftrightarrow \frac{d\lambda(\gamma)}{d\gamma} > 0 \text{ and } \frac{dAG(\gamma)}{d\gamma} < 0 \Leftrightarrow \frac{d\lambda(\gamma)}{d\gamma} < 0.$$

□

Proof of Theorem 5

Consider the independence Jeffreys prior (6) and the change of variable (7), then

$$\begin{aligned}\pi_I(\mu,\sigma,\gamma) &\propto \frac{|a'(\gamma)b(\gamma) - a(\gamma)b'(\gamma)|\sqrt{[b(\gamma)+\alpha_2[a(\gamma)+b(\gamma)]][a(\gamma)+\alpha_2[a(\gamma)+b(\gamma)]]}}{\sigma\sqrt{a(\gamma)b(\gamma)}[a(\gamma)+b(\gamma)]^2} \\ &\leq \frac{(\alpha_2+1)|a'(\gamma)b(\gamma) - a(\gamma)b'(\gamma)|}{\sigma\sqrt{a(\gamma)b(\gamma)}[a(\gamma)+b(\gamma)]}.\end{aligned}$$

For the particular choice $\{a(\gamma), b(\gamma)\} = \{\gamma, 1/\gamma\}$, the upper bound of $\pi_I(\mu,\sigma,\gamma)$ is proportional to $[\sigma(1+\gamma^2)]^{-1}$.

Applying Theorem 1 from Fernández and Steel (1998) and using this upper bound we can derive the properness of the posterior distribution of (μ,σ,γ). Now, since the mapping $(\mu,\sigma,\gamma) \leftrightarrow (\mu,\sigma_1,\sigma_2)$ is one-to-one, it follows that the posterior distribution of (μ,σ_1,σ_2) is proper. □

Proof of Theorem 6

Let f be a scale mixture of normals with λ_j the mixing variable associated with y_j and where the λ_js are independent random variables defined on $\mathbb{R}^+$ with distribution P_{λ_j}. We get an upper bound for the marginal distribution of $(y_1, ..., y_n)$ proportional to

$$\int_{\mathbb{R}_n^+}\int_\Gamma\int_0^\infty\int_{-\infty}^\infty \left(\prod_{j=1}^n \lambda_j^{\frac{1}{2}}\right)\frac{\sigma^{-(n+1)}}{[a(\gamma)+b(\gamma)]^n}\exp\left[-\frac{1}{2\sigma^2 h(\gamma)^2}\sum_{j=1}^n \lambda_j(y_j-\mu)^2\right]$$
$$\times \quad \pi(\gamma)\, d\mu d\sigma d\gamma dP_{(\lambda_1,\ldots,\lambda_n)},$$

where $h(\gamma) = \max\{a(\gamma), b(\gamma)\}$. Consider the change of variable $\vartheta = \sigma h(\gamma)$ and rewrite the upper bound as follows

$$\int_\Gamma \left[\frac{h(\gamma)}{a(\gamma)+b(\gamma)}\right]^n \pi(\gamma)\, d\gamma \int_{\mathbb{R}_n^+} \int_0^\infty \int_{-\infty}^\infty \left(\prod_{j=1}^n \lambda_j^{\frac{1}{2}}\right) \vartheta^{-(n+1)}$$
$$\times \quad \exp\left[-\frac{1}{2\vartheta^2}\sum_{j=1}^n \lambda_j (y_j-\mu)^2\right] d\mu d\vartheta dP_{(\lambda_1,\ldots,\lambda_n)}.$$

Fernández and Steel (2000, Th. 1) show that the integral in $\mu, \vartheta, \lambda_1, \ldots, \lambda_n$ is finite if $n \geq 2$. Then the existence of the integral in γ is a sufficient condition for the properness of the posterior distribution of (μ, σ, γ). The result then follows from

$$\int_\Gamma \left[\frac{h(\gamma)}{a(\gamma)+b(\gamma)}\right]^n \pi(\gamma)\, d\gamma \leq \int_\Gamma \pi(\gamma)\, d\gamma.$$

□

Proof of Theorem 7

Assume f is a scale mixture of normals. With the independence Jeffreys prior we get a lower bound for the marginal density of $(y_1, \ldots, y_n)$ which is proportional to

$$\int_{\mathbb{R}_n^+} \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \left(\prod_{j=1}^n \lambda_j^{\frac{1}{2}}\right) \frac{\sigma^{-(n+1)}\gamma^n}{(1+\gamma^2)^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^n \lambda_j \gamma^{-2\,\mathrm{sign}(y_j-\mu)}(y_j-\mu)^2\right]$$
$$\times \quad \sqrt{\frac{\alpha_2}{\gamma^2} + \frac{4}{(\gamma^2+1)^2}}\, d\mu d\sigma d\gamma dP_{(\lambda_1,\ldots,\lambda_n)}$$
$$\geq \quad \sqrt{\alpha_2} \int_{\mathbb{R}_n^+} \int_0^\infty \int_0^\infty \int_{-\infty}^{y_{(1)}} \left(\prod_{j=1}^n \lambda_j^{\frac{1}{2}}\right) \frac{\sigma^{-(n+1)}\gamma^{n-1}}{(1+\gamma^2)^n} \exp\left[-\frac{1}{2\sigma^2\gamma^2}\sum_{j=1}^n \lambda_j (y_j-\mu)^2\right]$$
$$\times \quad d\mu d\sigma d\gamma dP_{(\lambda_1,\ldots,\lambda_n)}.$$

Consider the change of variable $\vartheta = \sigma\gamma$. Then we can rewrite this lower bound as follows

$$\int_0^\infty \frac{\gamma^{2n-1}}{(1+\gamma^2)^n} d\gamma \int_{\mathbb{R}_n^+} \int_0^\infty \int_{-\infty}^{y_{(1)}} \left(\prod_{j=1}^n \lambda_j^{\frac{1}{2}}\right) \vartheta^{-(n+1)} \exp\left[-\frac{1}{2\vartheta^2}\sum_{j=1}^n \lambda_j (y_j-\mu)^2\right]$$
$$\times \quad d\mu d\vartheta dP_{(\lambda_1,\ldots,\lambda_n)}.$$

The first integral is infinite for any value of n which implies the improperness of the posterior distribution. □

Proof of Theorem 8

If f is a scale mixture of normals, then integrating over a subspace with respect to μ we get a lower bound for the marginal distribution of $(y_1, ..., y_n)$ which is proportional to

$$\int_{\mathbb{R}_+^n}\int_\Gamma\int_0^\infty\int_{-\infty}^{y_{(1)}}\left(\prod_{j=1}^n \lambda_j^{\frac{1}{2}}\right)\frac{\sigma^{-(n+2)}}{[a(\gamma)+b(\gamma)]^n}\exp\left[-\frac{1}{2\sigma^2 a(\gamma)^2}\sum_{j=1}^n \lambda_j(y_j-\mu)^2\right]$$
$$\times\frac{|\lambda(\gamma)|}{a(\gamma)+b(\gamma)}\,d\mu d\sigma d\gamma dP_{(\lambda_1,...,\lambda_n)}.$$

Consider the change of variable $\vartheta = \sigma a(\gamma)$. Then we can rewrite this lower bound as follows

$$\int_\Gamma\left[\frac{a(\gamma)}{a(\gamma)+b(\gamma)}\right]^{n+1}|\lambda(\gamma)|\,d\gamma\int_{\mathbb{R}_+^n}\int_0^\infty\int_{-\infty}^{y_{(1)}}\left(\prod_{j=1}^n \lambda_j^{\frac{1}{2}}\right)\vartheta^{-(n+2)}$$
$$\times\exp\left[-\frac{1}{2\vartheta^2}\sum_{j=1}^n \lambda_j(y_j-\mu)^2\right]\,d\mu d\vartheta dP_{(\lambda_1,...,\lambda_n)}.$$

Therefore, the existence of the first integral is a necessary condition for the properness of the posterior distribution of (μ, σ, γ). □

Proof of Theorem 9

The proof of (i) is as follows. If f is normal, defining $h(\gamma) = \max\{a(\gamma), b(\gamma)\}$ we get an upper bound for the marginal distribution of $(y_1, ..., y_n)$ which is proportional to

$$\int_{-\infty}^\infty\int_\Gamma\int_0^\infty\frac{\pi_J(\mu,\sigma,\gamma)}{[a(\gamma)+b(\gamma)]^n\sigma^n}\exp\left[-\frac{1}{2\sigma^2 h(\gamma)^2}\sum_{j=1}^n(y_j-\mu)^2\right]d\sigma d\gamma d\mu$$
$$\propto\int_{-\infty}^\infty\left[\sum_{j=1}^n(y_j-\mu)^2\right]^{-\frac{n+1}{2}}d\mu\int_\Gamma\frac{h(\gamma)^{n+1}}{[a(\gamma)+b(\gamma)]^{n+1}}|\lambda(\gamma)|d\gamma.$$

The first integral exists if $n \geq 2$. Then the existence of the second integral is a sufficient condition for the existence of the posterior distribution. For the second integral we use that

$$\int_\Gamma\frac{h(\gamma)^{n+1}}{[a(\gamma)+b(\gamma)]^{n+1}}|\lambda(\gamma)|d\gamma \leq \int_\Gamma|\lambda(\gamma)|d\gamma,$$

which is finite by assumption. If f is Laplace, analogously to the normal case we get an upper bound for the marginal distribution of $(y_1, ..., y_n)$ which is proportional to

$$\int_{-\infty}^{\infty}\int_{\Gamma}\int_{0}^{\infty}\frac{\pi_J(\mu,\sigma,\gamma)}{[a(\gamma)+b(\gamma)]^n\sigma^n}\exp\left[-\frac{1}{\sigma h(\gamma)}\sum_{j=1}^{n}|y_j-\mu|\right]d\sigma d\gamma d\mu$$

$$\propto \int_{-\infty}^{\infty}\left[\sum_{j=1}^{n}|y_j-\mu|\right]^{-(n+1)}d\mu\int_{\Gamma}\frac{h(\gamma)^{n+1}}{[a(\gamma)+b(\gamma)]^{n+1}}|\lambda(\gamma)|d\gamma,$$

and the same argument leads to the result.

Result (ii) follows immediately from Corollary 5.

For (iii) let us assume, without loss of generality, that $AG(\gamma)$ is an increasing function and $\Gamma = (\underline{\gamma}, \overline{\gamma})$. First, note that we can rewrite $AG(\gamma)$ as follows

$$AG(\gamma) = \tanh\left\{\frac{1}{2}\log\left[\frac{a(\gamma)}{b(\gamma)}\right]\right\}.$$

Then

$$\lim_{\gamma\to\overline{\gamma}} AG(\gamma) = 1 \quad\Leftrightarrow\quad \lim_{\gamma\to\overline{\gamma}}\log\left[\frac{a(\gamma)}{b(\gamma)}\right] = \infty$$
$$\lim_{\gamma\to\underline{\gamma}} AG(\gamma) = -1 \quad\Leftrightarrow\quad \lim_{\gamma\to\underline{\gamma}}\log\left[\frac{a(\gamma)}{b(\gamma)}\right] = -\infty,$$

which contradicts the assumption that $\lambda(\gamma)$ is absolutely integrable. The result is analogous if AG is decreasing. □

References

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *The American Statistician* 49: 34–38.

Arellano-Valle, R. B., Gómez, H. W. and Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference* 128: 427–443.

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

Bayes, C. L. and Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics* 21: 141–163.

Berger, J. O., Bernardo, J. M. and Sun, D. (2009). The formal definition of reference priors. *Annals of Statistics* 37: 905–938.

Berger, J. O. and Sun, D. (2008). Objective priors for the bivariate normal model. *Annals of Statistics* 36: 963–982.

Chopin, N. and Robert, C. P. (2010). Properties of nested sampling. *Biometrika* 97: 741–745.

Clarke, B. and Barron, A. R. (1994). Jeffreys' prior is asymptotically least favorable under entropy risk. *Journal of Statistical Planning and Inference* 41: 37–60.

Cox, D. R. and Reid, N. (1987). Orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society*, Series B 49: 1–39.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93: 359–371.

Fernández, C. and Steel, M. F. J. (2000). Bayesian regression analysis with scale mixtures of normals. *Econometric Theory* 16: 80–101.

Frühwirth-Schnatter, S. (2004). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics: New York.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57: 1317-1339.

Gibbons, J. F. and Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions. *Applied Physics Letters* 22: 568–569.

Jeffreys, H. (1941). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London*. Series A, 183: 453–461.

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.) Oxford: Clarendon.

John, S. (1982). The three-parameter two-piece normal family of distributions and its fitting. *Communications in Statistics - Theory and Methods* 11: 879-885.

Jones, M. C. (2006). A note on rescalings, reparametrizations and classes of distributions. *Journal of Statistical Planning and Inference* 136: 3730-3733.

Jones, M. C. and Anaya-Izquierdo K. (2010). On parameter orthogonality in symmetric and skew models. *Journal of Statistical Planning and Inference* 141: 758–770.

Liseo, B. and Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* 136: 373-389.

Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference* 83: 291–309.

Xu, A. and Tang, Y. (2011). Bayesian analysis of Birnbaum-Saunders distribution with partial information. *Computational Statistics and Data Analysis*, 55: 2324–2333.