Bayesian modelling of skewness and kurtosis with two-piece scale and shape transformations

Francisco Javier Rubio and Mark F. J. Steel

University of Warwick

30. June 2014

Bayesian modelling of skewness and kurtosis with two-piece scale and shape transformations

F.J. Rubio and M.F.J. Steel[*]

Abstract

We introduce the family of univariate double two–piece distributions, obtained by using a density–based transformation of unimodal symmetric continuous distributions with a shape parameter. The resulting distributions contain five interpretable parameters that control the mode, as well as the scale and shape in each direction. Four-parameter subfamilies of this class of distributions that capture different types of asymmetry are presented. We propose interpretable scale and location-invariant benchmark priors and derive conditions for the existence of the corresponding posterior distribution. The prior structures used allow for meaningful comparisons through Bayes factors within flexible families of distributions. These distributions are applied to models in finance, internet traffic data, and medicine, comparing them with appropriate competitors.

keywords: model comparison; posterior existence; prior elicitation; scale mixtures of normals; unimodal continuous distributions

1 Introduction

In the theory of statistical distributions, skewness and kurtosis are features of interest since they provide information about the shape of a distribution. Definitions and quantitative measures of these features have been widely discussed in the statistical literature (see e.g. van Zwet, 1964; Groeneveld and Meeden, 1984; Critchley and Jones, 2008).

Distributions containing parameters that control skewness and/or kurtosis are attractive since they can lead to robust models. This sort of flexible distributions are typically obtained by adding parameters to a known symmetric distribution through a parametric transformation. General representations of density–based and variable–based parametric transformations have been proposed in Ferreira and Steel (2006) and Ley and Paindaveine (2010), respectively. Transformations that include a parameter that controls skewness are usually referred to as "skewing mechanisms" (Ferreira and Steel, 2006; Ley and Paindaveine, 2010) while those that add a kurtosis parameter have been called "elongations" (Fischer and Klein, 2004), due to the effect produced on the shoulders and the tails of the distributions. Some examples of skewing mechanisms can be found in Azzalini (1985) and Fernández and Steel (1998a). Examples of elongations can be found in Tukey (1977), Haynes et al. (1997), Fischer and Klein (2004), and

Klein and Fischer (2006). A third class of transformations consists of those that contain two parameters that are used for modelling skewness and kurtosis jointly. Some members of this class are the Johnson S_U family (Johnson, 1949), Tukey-type transformations such as the g-and-h transformation and the LambertW transformation (Tukey, 1977; Goerg, 2011), and the sinh-arcsinh transformation (Jones and Pewsey, 2009). These sorts of transformations are typically, but not exclusively, applied to the normal distribution. Alternatively, distributions that can account for skewness and kurtosis can be obtained by introducing skewness into a symmetric distribution that already contains a shape parameter. Examples of distributions obtained by this method are skew-t distributions (Hansen, 1994; Fernández and Steel, 1998a; Azzalini and Capitanio, 2003; Rosco et al., 2011), and skew-Exponential power distributions (Azzalini, 1986; Fernández et al., 1995). Other distributions containing shape and skewness parameters have been proposed in different contexts such as the hyperbolic distribution (Barndorff-Nielsen, 1977; Aas and Haff, 2006), the skew−t proposed in Jones and Faddy (2003), and the $\alpha-$stable family of distributions. With the exception of the so called "two–piece" transformation (Fernández and Steel, 1998a; Arellano-Valle et al., 2005), the aforementioned transformations produce distributions with different shapes and/or different tail behaviour in each direction. A good survey can be found in Jones (2014).

We introduce a generalisation of the two-piece transformation defined on the family of unimodal, continuous and symmetric univariate distributions that contain a shape parameter. This generalisation consists of using different scale and shape parameters either side of the mode. We call this the "Double two-piece" (DTP) transformation. The resulting distributions contain five interpretable parameters that control the mode and the scale and shape in each direction. Our proposed transformation contains the original two-piece transformation as a subclass as well as a new class of transformations that only vary the shape of the distribution on each side of the mode. These two subclasses of distributions capture different types of asymmetry, recently denoted as "main-body skewness" and "tail skewness", respectively, by Jones (2014). Although some particular members of the proposed DTP family have already been studied (Zhu and Galbraith, 2010, 2011), we formalise this idea and extend it to a wider family of distributions, analysing the types of asymmetry that these distributions can capture. In addition, we propose and implement Bayesian methods for DTP distributions that allow us to meaningfully compare different distributions in these very flexible families through the use of Bayes factors. This directly sheds light on important features of the data.

In Section 2, we introduce the DTP transformation and discuss some of its properties as well as two interesting subfamilies. We examine the nature of the asymmetry induced by these transformations and propose a useful reparameterisation. In Section 3 we present scale and location-invariant prior structures for the proposed models and derive conditions for the existence of the corresponding posterior distributions. Section 4 contains three examples using real data. The first two examples concern the fitting of internet traffic and financial data, and we show how DTP distributions can be used to better understand the asymmetry of these data. In a second type of application we study the use of DTP distributions to model the random effects in a Bayesian hierarchical model. We compare various flexible distributions in this context, using medical data. Proofs are provided in Appendix B.

2 Two-Piece Scale and Shape Transformations

Let $\mathcal{F}$ be the family of continuous, unimodal, symmetric densities $f(\cdot; \mu, \sigma, \delta)$ with support on $\mathbb{R}$ and with mode and location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}_+$, and shape parameter $\delta \in \Delta \subset \mathbb{R}$. A shape parameter is anything that is not a location or a scale parameter.

Denote $f(x; \mu, \sigma, \delta) = \dfrac{1}{\sigma} f\left(\dfrac{x-\mu}{\sigma}; 0, 1, \delta\right) \equiv \dfrac{1}{\sigma} f\left(\dfrac{x-\mu}{\sigma}; \delta\right)$. Distribution functions are denoted by the corresponding uppercase letters. We define the two-piece probability density function constructed of $f(x; \mu, \sigma_1, \delta_1)$ truncated to $(-\infty, \mu)$ and $f(x; \mu, \sigma_2, \delta_2)$ truncated to $[\mu, \infty)$:

$$s(x; \mu, \sigma_1, \sigma_2, \delta_1, \delta_2) = \frac{2\varepsilon}{\sigma_1} f\left(\frac{x-\mu}{\sigma_1}; \delta_1\right) I(x < \mu) + \frac{2(1-\varepsilon)}{\sigma_2} f\left(\frac{x-\mu}{\sigma_2}; \delta_2\right) I(x \geq \mu), \quad (1)$$

where we achieve a continuous density function if we choose

$$\varepsilon = \frac{\sigma_1 f(0; \delta_2)}{\sigma_1 f(0; \delta_2) + \sigma_2 f(0; \delta_1)}. \qquad (2)$$

We denote the family defined by (1) and (2) as the Double Two-Piece (DTP) family. The distributions obtained by means of this transformation will be denoted as DTP distributions. The corresponding cumulative distribution function is then given by

$$\begin{aligned}
S(x; \mu, \sigma_1, \sigma_2, \delta_1, \delta_2) &= 2\varepsilon F\left(\frac{x-\mu}{\sigma_1}; \delta_1\right) I(x < \mu) \\
&+ \left\{\varepsilon + (1-\varepsilon)\left[2F\left(\frac{x-\mu}{\sigma_2}; \delta_2\right) - 1\right]\right\} I(x \geq \mu). \qquad (3)
\end{aligned}$$

The quantile function can be obtained by inverting (3). By construction, the density (1) is continuous, unimodal with mode at μ, and the amount of mass to the left of its mode is given by $S(\mu; \mu, \sigma_1, \sigma_2, \delta_1, \delta_2) = \varepsilon$. This transformation preserves the ease of use of the original distribution f and allows s to have different shapes in each direction, dictated by δ_1 and δ_2. In addition, by varying the ratio σ_1/σ_2, we control the allocation of mass on either side of the mode.

The family $\mathcal{F}$, on which the proposed transformation is defined, can be chosen to be, for example, the symmetric Johnson-S_U distribution (Johnson, 1949), the symmetric sinh-arcsinh distribution (Jones and Pewsey, 2009), or the family of scale mixtures of normals, for which the density f with shape parameter δ can be written as $f(x_j; \delta) = \int_0^\infty \tau_j^{1/2} \phi(\tau_j^{1/2} x_j) dP_{\tau_j|\delta}$ for observation x_j, where ϕ is the standard normal density and $P_{\tau_j|\delta}$ is a mixing distribution on $\mathbb{R}_+$. This is a broad class of distributions that includes, *i.a.* the Student-t distribution, the symmetric $\alpha-$stable, the exponential power distribution ($1 \leq \delta \leq 2$), the symmetric hyperbolic distribution (Barndorff-Nielsen, 1977), and the symmetric $\alpha-$stable family (see Fernández and Steel, 2000 for a more complete overview). Here we also introduce the case where the mixing distribution is a Birnbaum-Saunders(δ, δ) distribution, leading to what we call the SMN-BS distribution. Expressions for the density of the SMN-BS and some other less common distributions are presented in Appendix A. The shape parameter, $\delta > 0$, in all these models can be interpreted as a kurtosis parameter. Figure 1 illustrates the variety of shapes that we can obtain by applying the DTP transformation in (1) to the symmetric sinh-arcsinh distribution.

The DTP transformation preserves the existence of moments, if and only if they exist for both δ_1 and δ_2, since

$$\int_{\mathbb{R}} x^r s(x; \mu, \sigma_1, \sigma_2, \delta_1, \delta_2) dx = 2\varepsilon \int_{-\infty}^{\mu} x^r f(x; \mu, \sigma_1, \delta_1) dx + 2(1-\varepsilon) \int_{\mu}^{\infty} x^r f(x; \mu, \sigma_2, \delta_2) dx.$$
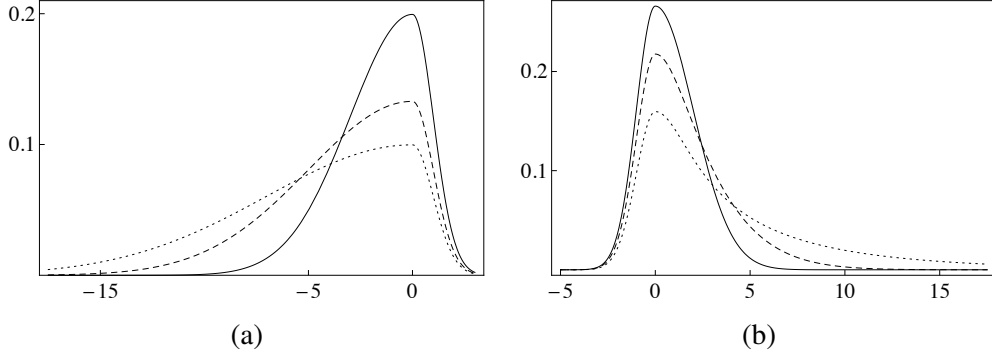
Figure 1: DTP sinh-arcsinh (DTP SAS) distribution with $\mu = 0$ and: (a) $\sigma_1 = 3, 5, 7$, $\sigma_2 = \delta_1 = \delta_2 = 1$; (b) $\sigma_1 = 1, \sigma_2 = 2, \delta_1 = 1, \delta_2 = 1, 0.75, 0.5$.

For example, if f in (1) is the Student-t density with δ degrees of freedom, then the rth moment of s exists if and only if both $\delta_1, \delta_2 > r$.

2.1 Subfamilies with 4 Parameters

Two-Piece Scale (TPSC) Transformation

The DTP family of transformations naturally includes the original two–piece transformation by setting the condition $\delta_1 = \delta_2 = \delta$ in (1), leading to

$$s(x; \mu, \sigma_1, \sigma_2, \delta) = \frac{2}{\sigma_1 + \sigma_2} \left[f\left(\frac{x-\mu}{\sigma_1}; \delta\right) I(x < \mu) + f\left(\frac{x-\mu}{\sigma_2}; \delta\right) I(x \geq \mu) \right]. \qquad (4)$$

The cases where $f(\cdot; \delta)$ is a Student-t distribution or an exponential power distribution have already been analysed in some detail (Fernández et al., 1995; Fernández and Steel, 1998a).

Two-Piece Shape (TPSH) Transformation

An alternative subfamily can be obtained by fixing $\sigma_1 = \sigma_2 = \sigma$ in (1), implying

$$s(x; \mu, \sigma, \delta_1, \delta_2) = \frac{2\varepsilon}{\sigma} f\left(\frac{x-\mu}{\sigma}; \delta_1\right) I(x < \mu) + \frac{2(1-\varepsilon)}{\sigma} f\left(\frac{x-\mu}{\sigma}; \delta_2\right) I(x \geq \mu), \qquad (5)$$

where $\varepsilon = \dfrac{f(0; \delta_2)}{f(0; \delta_1) + f(0; \delta_2)}$. This transformation, which has not yet been studied in general, produces distributions with different shape parameters in each direction. Note also that ε, the mass cumulated to the left of the mode, differs from $1/2$ whenever $f(0; \delta_1) \neq f(0; \delta_2)$. In the TPSH subclass skewness can only be introduced if the shape parameters differ in each direction. Other distributions with parameters that can control the tail behaviour in each direction have been proposed, for instance, in Jones and Faddy (2003), Aas and Haff (2006), and Jones and Pewsey (2009). Figure 2 shows two examples of distributions obtained with the TPSH transformation. Interchanging δ_1 and δ_2 reflects the density function around the mode.
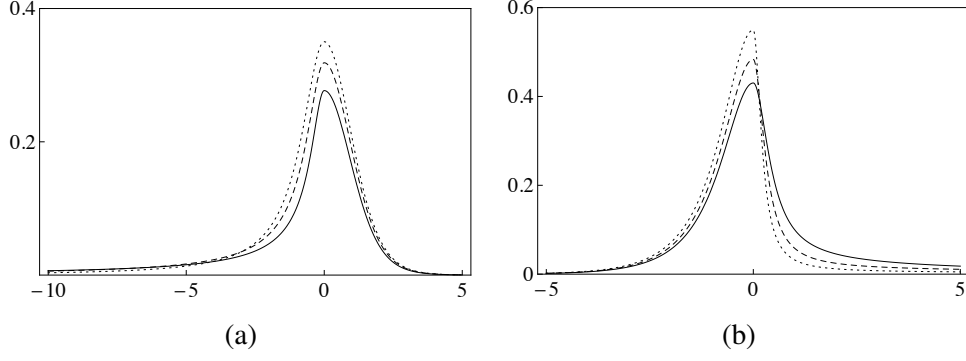
4

Figure 2: TPSH densities with $(\mu, \sigma) = (0, 1)$: (a) TPSH Student-t, $\delta_1 = 0.25, 0.5, 1$, $\delta_2 = 10$; (b) TPSH SMN-BS, $\delta_1 = 1, \delta_2 = 5, 10, 20$.

2.2 Understanding the Skewing Mechanism Induced by the Proposed Transformations

In order to provide more insight into the DTP transformation, we analyse the TPSC and TPSH families of transformations separately. For this purpose we employ two measures of asymmetry defined for continuous unimodal distributions, the Critchley-Jones (CJ) functional asymmetry measure (Critchley and Jones, 2008) and the Arnold-Groeneveld (AG) scalar measure of skewness (Arnold and Groeneveld, 1995). The CJ functional measures discrepancies between points located on each side of the mode $(x_L(p), x_R(p))$ of the density g such that $g(x_L(p)) = g(x_R(p)) = pg(\text{mode})$, $p \in (0, 1)$. It is defined as follows

$$\text{CJ}(p) = \frac{x_R(p) - 2 \times \text{mode} + x_L(p)}{x_R(p) - x_L(p)}. \tag{6}$$

Note that this measure takes values in $(-1, 1)$; negative values of $\text{CJ}(p)$ indicate that the values $x_L(p)$ are further from the mode than the values $x_R(p)$. An analogous interpretation applies to positive values. The AG measure of skewness is defined as $1 - 2G(\text{mode})$, where G is the distribution function associated with g. This measure also takes values in $(-1, 1)$; negative values of AG are associated with left skewness and positive values correspond to right skewness. For our DTP family in (1) these quantities are easy to calculate since $\text{AG} = 1 - 2\varepsilon$, and

$$\text{CJ}(p) = \frac{\sigma_2 f_R^{-1}(pf(0; \delta_2); \delta_2) + \sigma_1 f_L^{-1}(pf(0; \delta_1); \delta_1)}{\sigma_2 f_R^{-1}(pf(0; \delta_2); \delta_2) - \sigma_1 f_L^{-1}(pf(0; \delta_1); \delta_1)}, \tag{7}$$

where $f_L^{-1}(\cdot; \delta)$ and $f_R^{-1}(\cdot; \delta)$ represent the negative and positive inverse of $f(\cdot; \delta)$, respectively. Note also that $\text{CJ}(p) = \text{AG}$ when $\delta_1 = \delta_2$ for every $p \in (0, 1)$. This means that for the TPSC family both measures coincide. In general, the AG measure of skewness can be seen as an average of the asymmetry function CJ (Critchley and Jones, 2008). In the TPSC family, asymmetry is produced by varying the scale parameters on each side of the mode. This simply reallocates the mass of the distribution while preserving the tail behaviour and the shape in each direction. Since the nature of the asymmetry induced by the TPSC transformation is intuitively rather straightforward and has been discussed in *e.g.* Fernández and Steel (1998a), we now focus on the study of TPSH transformations.

Figure 3 shows some examples of (7) with distributions obtained using the TPSH transformation with parameters and AG as in Table 1. Figures 3(a) and 3(b) show examples where $\text{CJ}(p)$ changes sign

5

in cases where AG is nonzero. This means that the relative distance of the points $(x_L(p), x_R(p))$ to the mode varies from the tails to the mode of the density as a consequence of the different shapes and clearly the TPSH transformation is quite different from the TPSC one (for which CJ is constant). Figure 3(c) corresponds to densities where $\mathrm{CJ}(p)$ changes sign for some combinations of the parameters (δ_1, δ_2) while retaining the same sign for others. Finally, in Figure 3(d) $\mathrm{CJ}(p)$ retains the same sign for each p. Note that CJ for the SMN-BS distribution does not vary much with p, which means that TPSH and TPSC transformations are not that different. For the Student-t and exponential power distributions (see Figures 3(a) and 3(b)) changing scale and shape parameters has very different consequences: skewness (as measured by AG) is only induced for extremely low values of one of the shape parameters and the link between shape parameters and skewness (as measured by $\mathrm{CJ}(p)$) does not have a well-defined sign. Thus, the Student-t and the exponential power distributions are two interesting choices for f in our DTP class as the roles of the shape and scale parameters are clearly separated: the TPSH transformation leads to differential tail behaviour (and leaves skewness virtually unchanged for these particular choices of f) while the TPSC transformation induces skewness (and does not affect the tail behaviour, as is the case for any f).
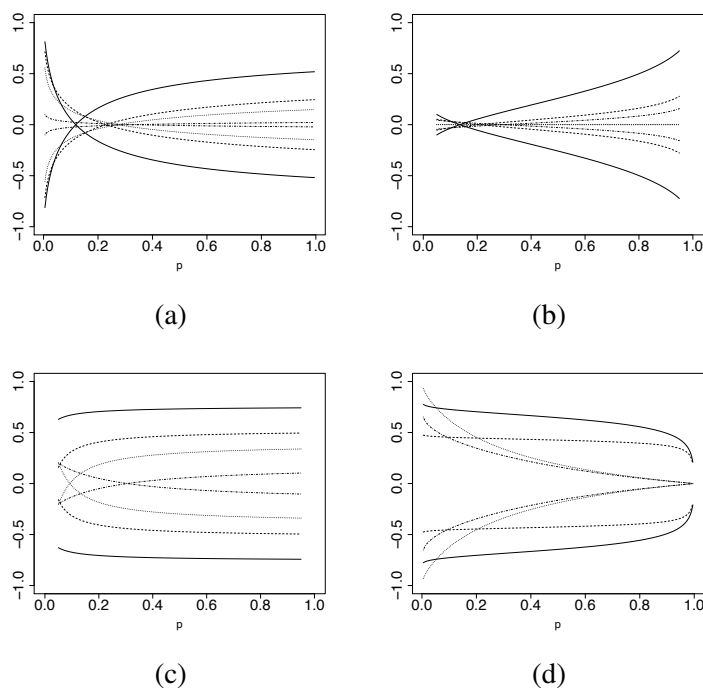

(a)   (b)

(c)   (d)

Figure 3: Asymmetry functional CJ for: (a) TPSH Student t; (b) TPSH exponential power; (c) TPSH SMN-BS; (d) TPSH sinh-arcsinh distribution. Lines correspond to δ_1 and δ_2 as in Table 1 and those values reversed.

2.3 Reparameterisations

For the TPSC family (4), Arellano-Valle et al. (2005) propose the reparameterisation $(\mu, \sigma_1, \sigma_2, \delta) \leftrightarrow (\mu, \sigma, \gamma, \delta)$ using the transformation $\sigma_1 = \sigma b(\gamma)$, $\sigma_2 = \sigma a(\gamma)$, where $\{a(\cdot), b(\cdot)\}$ are positive differentiable functions, $\gamma \in \Gamma \subset \mathbb{R}$, and the parameter space Γ depends on the choice of $\{a(\cdot), b(\cdot)\}$. The most common choices for $a(\cdot)$ and $b(\cdot)$ correspond to the *inverse scale factors* parameterisation

TPSH Student-t			TPSH sinh-arcsinh			TPSH SMN-BS			TPSH exp. power		
δ_1	δ_2	AG	δ_1	δ_2	AG	δ_1	δ_2	AG	δ_1	δ_2	AG
1/10	10	-0.45	5	1	2/3	1	50	-0.44	1	2	0.11
1/2	10	-0.18	5	2	0.43	1	10	-0.09	1.5	2	0.03
1	10	-0.1	1	1/4	3/5	1	5	0.03	2	2	0
5	10	-0.01	1	1/2	1/3	2	1	-0.07	2.5	2	-0.01

Table 1: Parameters used to obtain the functionals in Figure 3.

$\{a(\gamma), b(\gamma)\} = \{\gamma, 1/\gamma\}$, $\gamma \in \mathbb{R}_+$ (Fernández and Steel, 1998a), and the $\epsilon-skew$ parameterisation $\{a(\gamma), b(\gamma)\} = \{1 - \gamma, 1 + \gamma\}$, $\gamma \in (-1, 1)$ (Mudholkar and Hutson, 2000). Jones and Anaya-Izquierdo (2010) and Rubio and Steel (2014) show that choosing $a(\gamma) + b(\gamma)$ to be constant induces orthogonality between σ and γ. This reparameterisation is also appealing because the scalar γ can be interpreted as a skewness parameter since the AG measure of skewness depends only on this parameter. In particular, we obtain

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

This reparameterisation can also be used in DTP distributions for inducing orthogonality between σ and γ through parameterisations that verify $a(\gamma) + b(\gamma) = $ constant. Under this reparameterisation, density (1) becomes

$$s(x; \mu, \sigma, \gamma, \delta_1, \delta_2) = \frac{2}{\sigma c(\gamma, \delta_1, \delta_2)} \left[f(0; \delta_2) f\left(\frac{x - \mu}{\sigma b(\gamma)}; \delta_1\right) I(x < \mu) + f(0; \delta_1) f\left(\frac{x - \mu}{\sigma a(\gamma)}; \delta_2\right) I(x \geq \mu) \right],$$

(8)

where $c(\gamma, \delta_1, \delta_2) = b(\gamma) f(0; \delta_2) + a(\gamma) f(0; \delta_1)$. The interpretation of γ in the wider DTP family is slightly different since the cumulation of mass (and thus AG) depends also on the shape parameters (δ_1, δ_2). However, the parameter γ does not modify the shape of s.

Using this reparameterisation we can obtain the "generalized asymmetric Student-t distribution" proposed in Zhu and Galbraith (2010) by taking f to be a Student-t density and $\{a(\gamma), b(\gamma)\} = \{\gamma, 1 - \gamma\}$, $\gamma \in (0, 1)$. Under the same parameterisation, the "generalized asymmetric exponential power distribution" proposed in Zhu and Galbraith (2011) corresponds to an exponential power density for f.

For the TPSH family (5) there seems to be no obvious reparameterisation that induces parameter orthogonality. However, we can employ the reparameterisation $\delta_1 = \delta b^*(\zeta)$, $\delta_2 = \delta a^*(\zeta)$, with $\{a^*(\cdot), b^*(\cdot)\}$ positive differentiable functions. This helps to separate the roles of the shape parameters, since δ can be interpreted as in the underlying symmetric model, while ζ explains the difference between the shapes either side of the mode. The latter follows by noting that $\delta_1/\delta_2 = b^*(\zeta)/a^*(\zeta)$. This reparameterisation can also be applied to the DTP family, leading to the following density

$$s(x; \mu, \sigma, \gamma, \delta, \zeta) = \frac{2}{\sigma c(\gamma, \delta, \zeta)} \left[f(0; \delta a^*(\zeta)) f\left(\frac{x - \mu}{\sigma b(\gamma)}; \delta b^*(\zeta)\right) I(x < \mu) \right.$$

$$\left. + f(0; \delta b^*(\zeta)) f\left(\frac{x - \mu}{\sigma a(\gamma)}; \delta a^*(\zeta)\right) I(x \geq \mu) \right],$$

(9)

where $c(\gamma, \delta, \zeta) = b(\gamma)f(0; \delta a^*(\zeta)) + a(\gamma)f(0; \delta b^*(\zeta))$.

3 Bayesian Inference

3.1 Improper priors

In this section we propose a class of "benchmark" priors for the models studied in Section 2 with the parameterisations in (8) or (9). The proposed prior structure is inspired by the independence Jeffreys prior for the symmetric model, producing a scale and location-invariant prior.

The following result shows that the use of improper priors on the shape parameters of DTP models often leads to improper posteriors.

Theorem 1 *Let* $\mathbf{x} = (x_1, ..., x_n)$ *be an independent sample from* (8) *and consider the prior structure*

$$p(\mu, \sigma, \gamma, \delta_1, \delta_2) \propto p(\mu)p(\sigma)p(\gamma)p(\delta_1)p(\delta_2), \tag{10}$$

where $p(\delta_1)$ *and/or* $p(\delta_2)$ *are improper priors. It follows that*

(i) If $f(0; \delta)$ *does not depend upon* δ*, then the posterior is improper.*

(ii) If $f(0; \delta)$ *is bounded from above, then a necessary condition for posterior propriety is*

$$\int_\Delta f(0; \delta_i)^n p(\delta_i) d\delta_i < \infty, \;\; i = 1, 2. \tag{11}$$

(iii) If $f(0; \delta)$ *is a continuous and monotonic function of* δ*, then there exists* $M > 0$ *such that a necessary condition for the propriety of the posterior is*

$$\int_\Delta \frac{f(0; \delta_i)^n}{[f(0; \delta_i) + M]^n} p(\delta_i) d\delta_i < \infty, \;\; i = 1, 2. \tag{12}$$

This theorem provides a red flag for the use of improper priors on the shape parameters of DTP models. For instance, (i), (ii) and (iii) imply, respectively, that the use of improper priors on the shape parameters (δ_1, δ_2) of DTP exponential power (with the parameterisation in Zhu and Zinde-Walsh, 2009), DTP Student–t, and DTP sinh–arcsinh distributions leads to improper posteriors.

In DTP model (8) the parameters γ and (δ_1, δ_2) control the difference in the scale and the shapes either side of the mode, respectively. So we adopt a product prior structure $p(\gamma)p(\delta_1, \delta_2)$, allowing for prior dependence between δ_1 and δ_2. The following result provides conditions for the existence of the corresponding posterior distribution when f is a scale mixture of normals. The case where the sample contains repeated observations is covered as well.

Theorem 2 *Let* $\mathbf{x} = (x_1, ..., x_n)$ *be an independent sample from* (8)*. Let* f *be a scale mixture of normals and consider the prior structure*

$$p(\mu, \sigma, \gamma, \delta_1, \delta_2) \propto \frac{1}{\sigma} p(\gamma)p(\delta_1, \delta_2), \tag{13}$$

where $p(\gamma)$ *and* $p(\delta_1, \delta_2)$ *are proper. It follows that*

8

(i) The posterior distribution of $(\mu, \sigma, \gamma, \delta_1, \delta_2)$ is proper if $n \geq 2$ and all the observations are different.

(ii) If $\mathbf{x}$ contains repeated observations, let k be the largest number of observations with the same value in $\mathbf{x}$ and $1 < k < n$, then the posterior of $(\mu, \sigma, \gamma, \delta_1, \delta_2)$ is proper if and only if the mixing distribution of f satisfies for $i = 1, 2$ and j the observation index

$$\int_{0<\tau_1 \leq \cdots \leq \tau_n < \infty} \tau_{n-k}^{-(n-2)/2} \prod_{j \neq n-k, n} \tau_j^{1/2} dP_{(\tau_1, \ldots, \tau_n | \delta_i)} d\delta_i < \infty. \tag{14}$$

In the case of a two-piece Student-t sampling model, (14) is equivalent to

$$\int_{(k-1)/(n-k)}^{(k-1)/(n-k)+\xi} \frac{p(\delta_i)}{(n-k)\delta_i - (k-1)} d\delta_i < \infty \quad \text{and} \quad \int_0^{(k-1)/(n-k)} p(\delta_i) d\delta_i = 0, \tag{15}$$

for all $\xi > 0$ and $i = 1, 2$.

For the reparameterisation (9), the parameters (γ, δ, ζ) have separate roles: γ controls the difference in the scale either side of the mode, δ represents the shape parameter of the underlying symmetric density, and ζ controls the difference in the shape either side of the mode. For this reason, it is reasonable to adopt an independent prior structure on these parameters. The following result provides conditions for the existence of the posterior distribution.

Remark 1 *Let $\mathbf{x} = (x_1, \ldots, x_n)$ be an independent sample from (9). Let f be a scale mixture of normals and consider the prior structure*

$$p(\mu, \sigma, \gamma, \delta, \zeta) \propto \frac{1}{\sigma} p(\gamma) p(\delta) p(\zeta), \tag{16}$$

where $p(\gamma)$, $p(\delta)$, and $p(\zeta)$ are proper. The posterior distribution of $(\mu, \sigma, \gamma, \delta, \zeta)$ is proper if $n \geq 2$ and all the observations are different. If the sample contains repeated observations, We need to check that the induced prior on (δ_1, δ_2), for the parameterisation (8), satisfies (14).

Proof. *The results follows by a change of variable from (δ_1, δ_2) to (δ, ζ).*

As discussed in previous sections, the parameters of a distribution obtained through the TPSC transformation, $(\mu, \sigma, \gamma, \delta)$, can be interpreted as location, scale, skewness and shape, respectively. For this reason we adopt the product prior structure $p(\mu, \sigma, \gamma, \delta) \propto \frac{1}{\sigma} p(\gamma) p(\delta)$ for this family. In TPSH models the shape parameters (δ_1, δ_2) control the mass cumulated on each side of the mode as well as the shape. In addition, these parameters are not orthogonal in general. We therefore adopt the product prior structure $p(\mu, \sigma, \delta_1, \delta_2) \propto \frac{1}{\sigma} p(\delta_1, \delta_2)$ in this family, where $p(\delta_1, \delta_2)$ denotes a proper joint distribution which allows for prior dependence between δ_1 and δ_2. Theorem 2 covers the propriety of the posterior under these priors for TPSC and TPSH sampling models. For TPSH models with the parameterisation (9), Remark 1 provides conditions for the existence of the posterior distribution under the prior $p(\mu, \sigma, \delta, \zeta) \propto \frac{1}{\sigma} p(\delta) p(\zeta)$.

Another context of practical interest is when the sample consists of set observations. A set observation S is simply defined as a set of positive probability under the sampling model, *i.e.* $\mathbb{P}[\text{Observing } S] >$

0. In practice, this corresponds to any observation recorded with finite precision, as well as left, right and interval censoring. When the quantitative effect of censoring is not negligible, this must be formally taken into account. The following corollary provides conditions for the existence of the posterior from set observations with DTP sampling models.

Corollary 1 *Let* $\mathbf{x} = (S_1, ..., S_n)$ *be an independent sample of set observations from* (8). *Let* f *be a scale mixture of normals and consider the prior structure* (13). *Then, the posterior distribution of* $(\mu, \sigma, \gamma, \delta_1, \delta_2)$ *is proper if* $n \geq 2$ *and there exists a pair of sets, say* (S_i, S_j), *such that*

$$\inf_{x_i \in S_i, x_j \in S_j} |x_i - x_j| \quad > \quad 0. \tag{17}$$

Thus, whenever each sample of set observations contains at least two intervals that do not overlap, the posterior distribution of $(\mu, \sigma, \gamma, \delta_1, \delta_2)$ is proper. This result also applies to the parameterisation (9) with prior (16).

3.2 Choice of the prior on (γ, δ, ζ)

We now propose specific priors for the parameters (γ, δ, ζ) in (16) for a general choice of f in (9), and its corresponding subfamilies. We employ the parameterisation $\{a(\gamma), b(\gamma)\} = \{1 - \gamma, 1 + \gamma\}$, $\gamma \in (-1, 1)$ (so that σ and γ are orthogonal), and $\{a^*(\zeta), b^*(\zeta)\} = \{1 - \zeta, 1 + \zeta\}$, $\zeta \in (-1, 1)$. The shape parameter δ typically controls the peakedness and the heaviness of tails of the density function. As mentioned earlier, the parameters γ and ζ control the difference in scale and shape either side of μ. This interpretability of the parameters facilitates the choice of hyperparameters. In particular, reasonable priors to reflect vague prior beliefs are that $\gamma \sim \text{Unif}(-1, 1)$ and $\zeta \sim \text{Unif}(-1, 1)$. The elicitation of the prior on the parameter δ is more delicate, given that this parameter has different interpretations for different models. However, in all the models of interest, δ can be interpreted as a kurtosis parameter. Therefore, in order to come up with a more general elicitation strategy we propose basing this choice on a prior for a bounded kurtosis measure, which is common to all models and is an injective function of δ, say $\kappa = \kappa(\delta)$. The boundedness assumption on κ allows us to assign a proper uniform prior on this quantity, while the injectivity is required for obtaining the induced prior on the parameter δ by inverting this function. See Critchley and Jones (2008) for a good survey on kurtosis measures.

We propose to adopt the scalar kurtosis measure $\kappa = 2\dfrac{f(\pi_R)}{f(\text{mode})} - 1$ from Critchley and Jones (2008), where π_R represents the positive mode of $-f'$ (the inflection point). This measure κ takes values in $K \subset (-1, 1)$, assigning the value $\kappa = 0.213$ to the normal distribution. Numerically, we have found that κ is an injective function of δ for many distributions f, such as the Student-t, the symmetric sinh-arcsinh, the symmetric Johnson-S_U, the exponential power with $\delta > 1$, the symmetric hyperbolic, the SMN-BS with $\delta < 2.65$ and the Meixner distribution. Another appealing feature of this measure of kurtosis is that both the AG skewness measure and κ can be interpreted as the average of certain functional measures of asymmetry and kurtosis using the same weight function (see Critchley and Jones, 2008). Algorithm 1 shows how to approximate the priors induced on δ by a uniform prior on the appropriate range for κ. Figure 4 shows the induced priors on δ for the Student-t and symmetric sinh-arcsinh distributions.

Algorithm 1 Construction of the prior on δ.

1: Identify the range K covered by varying δ in the model of interest.

2: Simulate $\mathbf{u} = (u_1, \ldots, u_N) \sim \text{Unif}(K)$.

3: Calculate $\mathbf{d} = \kappa^{-1}(\mathbf{u})$.

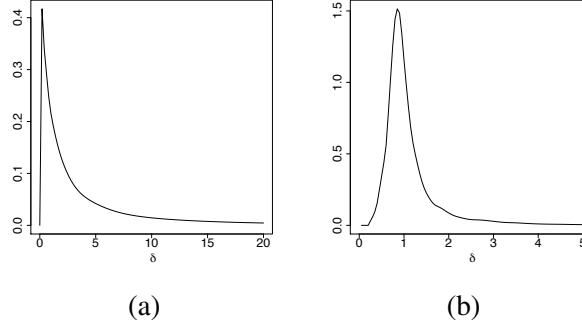4: Approximate the distribution of $\mathbf{d}$ using a kernel density estimator.

Figure 4: Priors for δ: (a) Student-t distribution; (b) Sinh-arcsinh distribution.

3.3 Weakly informative priors

We may prefer to use a "vague" proper prior which is not very influential on the posterior inference. In the previous section we provided weakly informative priors for the shape parameters (γ, δ, ζ). We can combine that with independent vague proper priors on the location and scale parameters (μ, σ). For the location parameter we propose a uniform prior on an appropriate bounded interval $\mathcal{D}$, while for the scale parameter we employ a Half-Cauchy distribution with location 0 and scale s (Polson and Scott, 2012). Unfortunately, general choices for $\mathcal{D}$ and s are not available, given that these values depend on the units of measurement. We recommend conducting sensitivity analyses with respect to $\mathcal{D}$ and s. Note that the structure of this prior resembles that of the improper benchmark priors discussed in the previous sections.

This prior structure is also useful for choices of f that do not belong to the family of scale mixtures of normals and, consequently, the existence of the posterior under improper priors is not covered by the results in Subsection 3.1.

4 Applications

We present three examples with real data to illustrate the use of DTP, TPSC and TPSH distributions. We adopt the $\epsilon-$skew parameterisation for DTP and TPSC models. In the first two examples, simulations of the posterior distributions are obtained using the t-walk algorithm (Christen and Fox, 2010). Given the hierarchical nature of the third example, we use the adaptive Metropolis within Gibbs sampler implemented in the R package 'spBayes' (Finley and Banerjee, 2013). R codes used here and the R-package 'DTP', which implements basic functions related to the proposed models, are available on request.

Model comparison within the DTP family is conducted via Bayes factors which are obtained using the Savage–Dickey ratio for nested models, and through importance sampling when we compare non-

nested choices for f. We also compare the DTP model and its submodels with other distributions used in the literature. For a fair model comparison, we include appropriate competitors in each example, matched to the features of the data. A meaningful Bayesian comparison with these other models would require the specification of priors for the parameters in these other distributions that are comparable (matched) to our models, and to compute Bayes factors we would need to use proper priors for all model-specific parameters. This would be a nontrivial undertaking and would risk diluting the main message of the paper. We choose instead to compare with these other classes of distributions through classical information criteria based on maximum likelihood estimates (MLE). We aim to show that our DTP families are flexible enough and then we can use formal Bayesian methods to select (or average) models within these families.

Given that DTP, TPSC, and TPSH distributions capture different sorts of asymmetry, conducting model comparison between these distributions not only provides information about which model fits the data better but it also indicates what kind of asymmetry is favoured by the data. In addition, the DTP family provides important advantages in terms of interpretability of parameters (and, thus, prior elicitation) and inferential properties.

4.1 Internet traffic data

In this example we analyse the teletraffic data set studied in Ramirez-Cobo et al. (2010), which contains $n = 3143$ observations, representing transferred bytes/sec within consecutive seconds. Ramirez-Cobo et al. (2010) propose the use of a Normal Laplace distribution to model these data after a logarithmic transformation. The Normal Laplace distribution is obtained as the convolution of a Normal distribution and a two–piece Laplace distribution with location 0 and two parameters (α, β) that jointly control the scale and the skewness. The Normal Laplace distribution has tails heavier than those of the normal distribution (Reed and Jorgensen, 2004). We also use the sinh-arcsinh distribution of Jones and Pewsey (2009), indicated by s_{JP} and the skew-t of Azzalini and Capitanio (2003), denoted by s_{AC} (see Appendix A). Here, we explore the performance of the DTP sinh–arcsinh distribution (DTP SAS). This distribution allows for all moments to exist and accommodates both heavier and lighter tails than the normal distribution, which is a submodel of the DTP SAS ($\delta_1 = \delta_2 = 1$, $\gamma = 0$). We use the priors of Subsection 3.3: $\mu \sim \text{Unif}(0, 25), \sigma \sim \text{HalfCauchy}(0, s), \gamma \sim \text{Unif}(-1, 1), \zeta \sim \text{Unif}(-1, 1)$, where $s = 1/5, 1, 5$ and for δ we adopt the prior in Figure 4. The results were not sensitive to the choice of s. Table 2 shows the MLE and the classical model comparison criteria for all models considered. The DTP SAS results indicate that the right tail is much lighter than that of the normal distribution, a feature that cannot be captured by the Normal Laplace distribution used in Ramirez-Cobo et al. (2010). In addition, there is strong evidence of "main-body" skewness, captured by different scales. Both features of the models are clearly important for these data and the DTP SAS model is strongly favoured by AIC and BIC. Bayes factors within the DTP SAS family also strongly support the most complete model, versus the possible submodels in Table 3. Posterior predictive densities shown in Figure 5 illustrate how the DTP SAS model differs from the others in mode and tail behaviour (see the right panel).

Model	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\gamma}$	$\widehat{\delta}$	$\widehat{\zeta}$	AIC	BIC
DTP SAS	11.16	11.39	-0.98	10.795	-0.94	**5849.14**	**5879.40**
TPSC SAS	11.80	0.85	0.14	1.26	–	5884.95	5909.16
TPSH SAS	11.75	0.87	–	1.30	-0.08	5880.20	5904.41
s_{JP}	11.78	0.84	$(\widehat{\varepsilon})$ -0.16	1.25	–	5886.84	5911.05
Normal Laplace	11.77	8.39	$(\widehat{\alpha})$ 4.09	$(\widehat{\beta})$ 0.56	–	5922.73	5946.94
s_{AC}	12.07	0.75	$(\widehat{\lambda})$ -0.98	1057.40	–	5919.52	5943.73

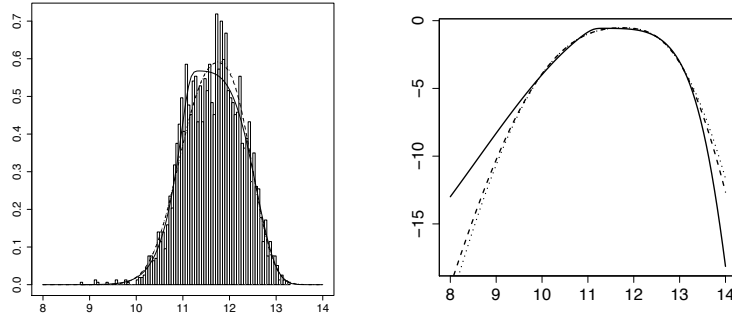Table 2: Internet traffic data: Maximum likelihood estimates, AIC and BIC (best values in bold).

Figure 5: Internet traffic data (in logarithms; histogram) with (a) Predictive densities and (b) Log-predictive densities: DTP (continuous line); TPSH (dashed line); TPSC (dotted line).

Model	DTP SAS	TPSH SAS	TPSC SAS	TPSC normal	Symm. SAS	normal
BF	1	≈ 0	≈ 0	≈ 0	≈ 0	≈ 0

Table 3: Internet traffic data. Bayes factors of submodels vs. DTP SAS model (entries $< 10^{-100}$).

4.2 Actuarial Application

In this application we analyse the claim sizes reported in Berlaint et al. (2004) which can be found in `http://lstat.kuleuven.be/Wiley/`. This data set contains $n = 1823$ observations provided by the reinsurance brokers Aon Re Belgium. Such data typically contain extreme observations, and the logarithmic transformation is often used to reduce the effect of these extreme values (Ramirez-Cobo et al., 2010). A quantity of interest in this context is the probability that the claims exceed a certain bound (Venturini et al., 2008). This is often used for budgetary planning, which emphasises the importance of properly modelling the tails of the distribution.

We explore two choices for f in (1): a Student-t distribution and an SMN-BS distribution (see Appendix A). We adopt the product prior structure (16) with uniform priors on γ and ζ. In order to produce matched priors on δ for these two models, we follow the strategy in Subsection 3.2. The measure of kurtosis $\kappa \in (0.213, 0.633)$ for the Student-t model and $\kappa \in (0.213, 0.560)$ for the SMN-BS model. Uniform priors for κ induce compatible priors for δ in both models. Given that the data set contains a maximum number of $k = 30$ repeated observations, we need to restrict the priors for (δ, ζ):

for the Student-t model we truncate $\delta > 2$ and restrict $\zeta \in (-0.99, 0.99)$. This truncation guarantees that condition (15) is satisfied since it implies that $\delta_1, \delta_2 > (k-1)/(n-k) \approx 0.02$. For the SMN-BS model, the κ measure is injective only on the interval $\delta \in (0, 2.65)$, which covers the range $\kappa \in (0.213, 0.560)$. In addition, for this model we can check that condition (14) is satisfied if we truncate the δ_i's away from zero, *e.g.* by imposing $\delta > 1 \times 10^{-6}$ and taking $\zeta \in (-0.999, 0.999)$. Thus, we restrict the prior for δ in the SMN-BS model to $(1 \times 10^{-6}, 2.65)$. The posterior distributions are proper by Remark 1.

We also use the skew-t distributions in Azzalini and Capitanio (2003) (s_{AC}) and Jones and Faddy (2003), denoted by s_{JF} (see Appendix A). Table 4 shows the MLE and the AIC and BIC criteria, which favour the TPSC SMN-BS model overall. The Bayes factors, reported in Table 5, favour the TPSC model for both underlying choices of f and favours the TPSC SMN-BS model overall, which agrees with the conclusion from AIC and BIC. However, there is no conclusive message from the SMN-BS models about which type of asymmetry is best for the data. The TSPH variant does almost as well. This is in line with the fact that the SMN-BS model does not distinguish clearly between TPSH and TPSC transformations, as discussed in Subsection 2.2. In contrast, the Student-t models, for which both transformations are very distinct, unambiguously indicate that the asymmetry is in the main body of the data and not in the tails: the TPSH t model does very badly indeed, using both classical and Bayesian methods. Figure 6 shows the corresponding predictive densities and illustrates the poor fit of the TPSH t model which clearly affects the estimation of the right-tail probabilities shown in Figure 6(b): this model produces a predictive probability of 0.01 for the event $x > 17$, while the other models lead to a predictive probability of less than 0.004. Unlike in the previous application, where right "main-body" skewness is combined with a heavier left tail (both γ and ζ are estimated to be highly negative), the skew-t by Azzalini and Capitanio (2003) does well here, as these data combine right skewness in the main body with a fatter right tail. This is a feature that the s_{AC} imposes (for positive λ with both asymmetries in the opposite direction for $\lambda < 0$). It is important to point out that the DTP families are not restricted in this way, as evidenced by the superiority of the DTP model in the previous application.

Model	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\zeta}$	AIC	BIC
DTP t	7.93	1.61	-0.57	13.33	0.26	7283.1	7310.7
TPSC t	7.90	1.62	-0.59	10.98	–	7281.6	7303.6
TPSH t	9.13	1.46	–	9998.80	0.99	7434.4	7456.5
DTP SMN-BS	7.96	2.38	-0.48	0.46	-0.23	7280.6	7308.2
TPSC SMN-BS	7.90	2.36	-0.58	0.51	–	**7279.4**	**7301.5**
TPSH SMN-BS	8.03	3.43	–	0.31	-0.83	7280.1	7302.2
s_{JF}	1.56	0.02	–	($\hat{a}$) 1560.6	($\hat{b}$) 5.07	7302.1	7324.1
s_{AC}	7.17	2.84	($\hat{\lambda}$) 4.90	13.75	–	7280.7	7302.7

Table 4: Aon data: Maximum likelihood estimates, AIC and BIC (best values in bold).

Model	DTP	TPSH	TPSC
Student-t	1	5.00×10^{-65}	2.05
SMN-BS	4.50	1.61	9.02

Table 5: Aon data: Bayes factors with respect to the DTP-t model.

4.3 Hierarchical Bayesian Models in Meta–Analysis

Bayesian hierarchical models are used in a variety of applied contexts to tackle parameter heterogeneity. A common example of this is the two–level normal model:

$$
\begin{aligned}
y_j | \theta_j &\sim N(\theta_j, \sigma_j), \ \ j = 1 \ldots n, \\
\theta_j &\sim N(\mu, \sigma).
\end{aligned} \tag{18}
$$

A natural question is whether the assumption of normality of the random effects is appropriate: the implications of departures from this assumption are discussed in Zhang and Davidian (2001), Thompson and Lee (2008) and McCulloch and Neuhaus (2011).

In order to produce models that are robust to departures from normality of θ_j, several generalisations of (18) have been proposed. For example, Doss and Hobert (2010) employ a Student–t distribution, Thompson and Lee (2008) use a TPSC t distribution with $\delta > 2$ degrees of freedom, while Dunson (2010) follows a Bayesian nonparametric approach. The use of non–normal distributional assumptions in this hierarchical model typically requires more sophisticated MCMC methods as discussed in Roberts and Rosental (2009).

4.3.1 Fluoride Meta–analysis

In this example we analyse the data set presented in Marinho et al. (2003) and used in Thompson and Lee (2008), which contains $n = 70$ trials assessing the effectiveness of fluoride toothpaste compared to a placebo conducted between 1954 and 1994. The treatment effect is the "prevented fraction", defined as the mean increment in the controls minus the mean increment in the treated group, divided by the mean increment in the controls. Thompson and Lee (2008) then propose the model

$$
\begin{aligned}
y_j | \theta_j &\sim N(\theta_j, \sigma_j), \\
\theta_j &\sim P,
\end{aligned} \tag{19}
$$

where y_j is the estimate of the treatment effect in study j, θ_j is the true treatment effect in study j, and the parameters σ_j are estimated from the data and assumed known. They compare the conclusions obtained for the true treatment effect for the following choices for P: (i) a TPSC t distribution with $\delta > 2$ degrees of freedom, (ii) a symmetric Student t distribution with $\delta > 2$ degrees of freedom, (iii) a TPSC normal distribution, and (iv) a normal distribution.

Here, we study six choices for P: (i) a normal distribution, (ii) a symmetric sinh–arcsinh (SAS) distribution, (iii) a TPSC normal distribution, (iv) a TPSC SAS distribution, (v) a TPSH SAS distribution and (vi) a DTP SAS distribution. For the DTP model, we adopt the prior structure as in Subsection 3.3 $p(\mu, \sigma, \gamma, \delta, \zeta) = p(\mu)p(\sigma)p(\gamma)p(\delta)p(\zeta)$ with $\mu \sim \text{Unif}(-10, 10), \sigma \sim \text{HalfCauchy}(0, s), \gamma \sim$
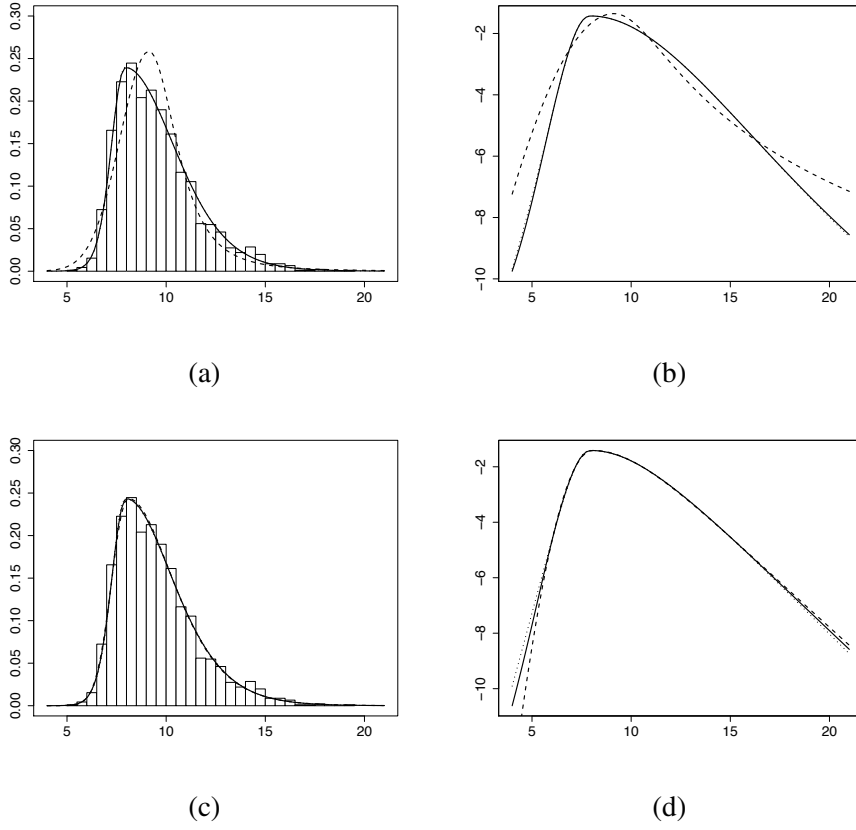
15

Figure 6: Aon data (histogram) with (a) Predictive densities and (b) Log-predictive densities: DTP t (continuous line); TPSH t (dashed line); TPSC t (dotted line). (c) Predictive densities and (d) Log-predictive densities: DTP SMN-BS (continuous line); TPSH SMN-BS (dashed line); TPSC SMN-BS (dotted line).

Unif$(-1, 1), \delta \sim p(\delta), \zeta \sim$ Unif$(-1, 1)$, with the prior shown in Figure 4 for δ, and $s = 1/5, 1, 5$. The results were not sensitive to the choice of s.

Figure 7 shows the posterior predictive densities for the treatment effect under different distributional assumptions for the random effects. Clearly, symmetric distributions put more predictive mass in the left tail $(-\infty, 0.05)$ than those with asymmetry. Therefore, the probability of a small or a negative effect is overestimated under symmetric random effects. The predictive distributions obtained for DTP, TPSC, and TPSH SAS models are fairly similar in this case. However, the Bayes factors, shown in Table 6, favour DTP and TPSH SAS models over the other competitors.

Model	DTP SAS	TPSH SAS	TPSC SAS	TPSC normal	Sym. SAS	normal
BF	1	1.27	0.30	0.05	0.02	5.2×10^{-5}

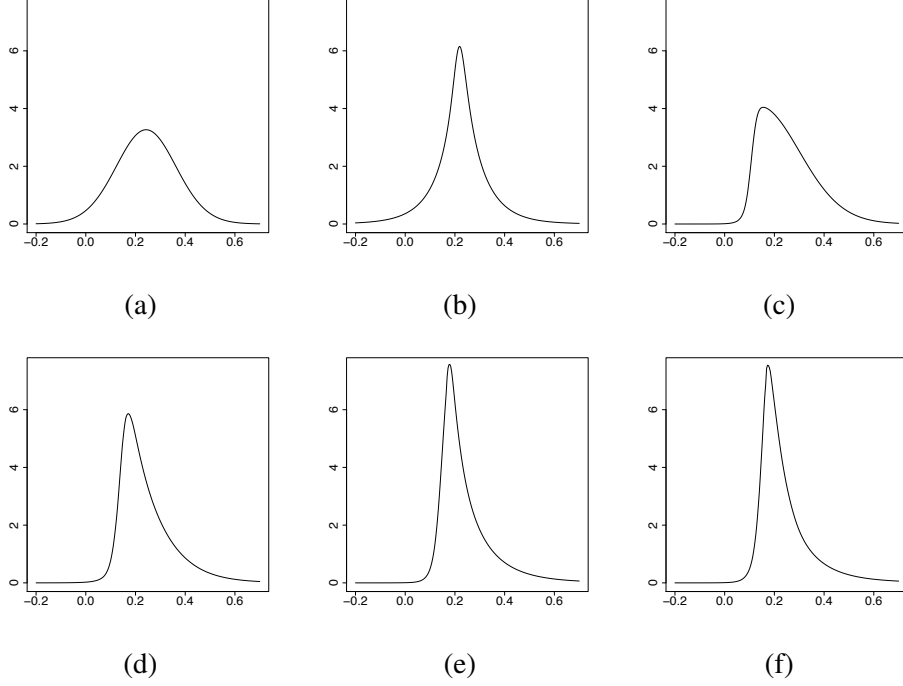Table 6: Fluoride data: Bayes factors of submodels vs. the DTP SAS model.

Figure 7: Predictive densities for the treatment effect: (a) Normal; (b) Symmetric SAS; (c) TPSC normal; (d) TPSC SAS; (e) TPSH SAS; (f) DTP SAS.

5 Concluding Remarks

We introduce a simple, intuitive and general class of transformations (DTP) that produces flexible unimodal and continuous distributions with parameters that separately control main-body skewness and tails on each side of the mode. Although some particular cases of DTP models have already appeared (Zhu and Galbraith, 2010, 2011), we formalise the idea and extend it to a wide range of symmetric "base" distributions $\mathcal{F}$. We also distinguish two subclasses of transformations and examine their interpretation as skewing mechanisms. A considerable advantage of the DTP class of transformations is the interpretability of its parameters (see Jones, 2014 for the importance of interpretability) which, in the Bayesian context, also facilitates prior elicitation. We propose a scale and location-invariant prior structure and derive conditions for posterior existence, also taking into account repeated and set observations.

As illustrated by the applications, DTP families provide a flexible way of modelling unimodal data (or latent effects with unimodal distributions) and we provide a Bayesian framework for inference with sensible prior assumptions. In addition, we can conduct formal model comparison through Bayes factors for selecting models within the following classes:

- subclasses of DTP models with the same underlying symmetric base distribution f: this is possible through the clearly separated roles of the parameters and the ensuing product prior structure with proper priors on γ and ζ.

- classes of DTP models with different underlying f: in nested cases this is easy, given the separate roles of the parameters and the ensuing product prior structure with proper priors on δ, and in non-nested cases the priors on different shape parameters δ are matched through a common prior

17

on the kurtosis measure κ.

DTP, TPSC and TPSH transformations can be used to construct robust models and, since they capture different kinds of asymmetry, selecting between these models provides more insight into the features of a data set. We have used Bayes factors for model choice, but other criteria, such as log-predictive scores, might be considered as well.

DTP families can be extended to the multivariate case in several ways. For TPSC models, Ferreira and Steel (2007) propose the use of affine transformations to produce a multivariate extension while Rubio and Steel (2013) propose to use copulas. In a similar fashion, the DTP (and consequently the TPSH) family can be used to construct multivariate distributions.

A different subclass of DTP transformations can be obtained by fixing $\sigma_1 = \sigma$ and $\sigma_2 = \dfrac{f(0; \delta_2)}{f(0; \delta_1)}\sigma$, leading to distributions with different shapes but equal mass cumulated on each side of the mode. This idea is proposed in Rubio (2013), who also composes this transformation with other skewing mechanisms to produce a different type of generalised skew-t distribution.

Rubio and Steel (2014) explore the use of Jeffreys priors in TPSC models. The use of Jeffreys priors for TPSH and DTP models is the object of further research.

Appendix A: Some density functions

(i) The symmetric Johnson-S_U distribution (Johnson, 1949):

$$f(x; \mu, \sigma, \delta) = \frac{\delta}{\sigma}\phi\left[\delta \operatorname{arcsinh}\left(\frac{x-\mu}{\sigma}\right)\right]\left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{1}{2}}.$$

(ii) The sinh-arcsinh distribution (Jones and Pewsey, 2009):

$$s_{JP}(x; \mu, \sigma, \delta) = \frac{\delta}{\sigma}\phi\left[\sinh\left(\delta \operatorname{arcsinh}\left(\frac{x-\mu}{\sigma}\right) - \varepsilon\right)\right]\frac{\cosh\left(\delta \operatorname{arcsinh}\left(\frac{x-\mu}{\sigma}\right) - \varepsilon\right)}{\sqrt{1 + \left(\frac{x-\mu}{\sigma}\right)^2}},$$

where $\varepsilon \in \mathbb{R}$ controls the asymmetry of the density and symmetry corresponds to $\varepsilon = 0$.

(iii) SMN-BS, a scale mixture of normals with Birnbaum-Saunders(δ, δ) mixing:

$$f(x; \mu, \sigma, \delta) = \frac{e^{\frac{1}{\delta^2}}\left(\sqrt{\delta x^2 + 1}K_0\left(\frac{\sqrt{\delta x^2 + 1}}{\delta^2}\right) + K_1\left(\frac{\sqrt{\delta x^2 + 1}}{\delta^2}\right)\right)}{2\pi\delta^{3/2}\sqrt{\delta x^2 + 1}}.$$

where $K_n(z)$ represents the modified Bessel function of the second kind.

(iv) The skew-t density from Jones and Faddy (2003):

$$s_{JF}(x; \mu, \sigma, a, b) = C_{a,b}^{-1}\left[1 + \frac{t}{\sqrt{a+b+t^2}}\right]^{a+1/2}\left[1 - \frac{t}{\sqrt{a+b+t^2}}\right]^{b+1/2},$$

where $a, b > 0$, $C_{a,b} = 2^{a+b-1}\operatorname{Beta}(a, b)\sqrt{a+b}$, and $t = \dfrac{x-\mu}{\sigma}$. The parameters (a, b) control the tails and skewness jointly. The density s_{JF} is asymmetric if and only if $a \neq b$, so that the density is skewed only when the tail behaviour differs in each direction.

18

(v) The skew-t density from Azzalini and Capitanio (2003):

$$s_{AC}(x; \mu, \sigma, \lambda, \delta) = 2f(x; \mu, \sigma, \delta)F\left(\lambda x \sqrt{\frac{\delta+1}{\delta+x^2}}; \mu, \sigma, \delta+1\right),$$

where $\lambda \in \mathbb{R}$ and f and F are, respectively, the density function and the distribution function of the Student-t.

Appendix B: Proofs

Proof of Theorem 1

The marginal likelihood of the data can be bounded from below as follows

$$
\begin{aligned}
m(\mathbf{x}) \quad &\propto \quad \int_\Delta \int_\Delta \int_\Gamma \int_{\mathbb{R}_+} \int_{\mathbb{R}} \left[\prod_{j=1}^n s(x_j; \mu, \sigma, \gamma, \delta_1, \delta_2)\right] p(\mu)p(\sigma)p(\gamma)p(\delta_1)p(\delta_2) \, d\mu d\sigma d\gamma d\delta_1 d\delta_2 \\
&\geq \quad \int_\Delta \int_\Delta \int_\Gamma \int_{\mathbb{R}_+} \int_{-\infty}^{x_{(1)}} \frac{f(0;\delta_1)^n}{\sigma^n H(\gamma)^n [f(0;\delta_1) + f(0;\delta_2)]^n} \left[\prod_{j=1}^n f\left(\frac{x_j - \mu}{\sigma a(\gamma)}; \delta_2\right)\right] \\
&\times \quad p(\mu)p(\sigma)p(\gamma)p(\delta_1)p(\delta_2) \, d\mu d\sigma d\gamma d\delta_1 d\delta_2 \quad\quad\quad (20)
\end{aligned}
$$

where $s(\cdot)$ is given by (8) in the paper, $H(\gamma) = \max\{a(\gamma), b(\gamma)\}$, and $x_{(1)}$ represents the smallest order statistic of $\mathbf{x}$. Therefore:

(i) follows by noting that the lower bound (20) does not depend upon δ_1.

(ii) follows by using the following inequality, provided $f(0;\delta) \leq U$ for some $U > 0$

$$\frac{f(0;\delta_1)^n}{[f(0;\delta_1) + f(0;\delta_2)]^n} \geq \frac{f(0;\delta_1)^n}{2^n U^n},$$

which leads to the necessary condition (11).

(iii) Given that $f(0;\delta)$ is continuous and monotonic, there exist $M > 0$ and a set $\Delta_2(M) \subset \Delta$ such that $f(0;\delta) < M$ for all $\delta \in \Delta_2(M)$. If we integrate δ_2 over Δ_2, we obtain the following lower bound, up to a proportionality constant, for $m(\mathbf{x})$

$$
\begin{aligned}
&\int_{\Delta_2} \int_\Delta \int_\Gamma \int_{\mathbb{R}_+} \int_{-\infty}^{x_{(1)}} \frac{f(0;\delta_1)^n}{\sigma^n H(\gamma)^n [f(0;\delta_1) + f(0;\delta_2)]^n} \left[\prod_{j=1}^n f\left(\frac{x_j - \mu}{\sigma a(\gamma)}; \delta_2\right)\right] \\
&\times \quad p(\mu)p(\sigma)p(\gamma)p(\delta_1)p(\delta_2) \, d\mu d\sigma d\gamma d\delta_1 d\delta_2 \\
&\geq \quad \int_{\Delta_2} \int_\Delta \int_\Gamma \int_{\mathbb{R}_+} \int_{-\infty}^{x_{(1)}} \frac{f(0;\delta_1)^n}{\sigma^n H(\gamma)^n [f(0;\delta_1) + M]^n} \left[\prod_{j=1}^n f\left(\frac{x_j - \mu}{\sigma a(\gamma)}; \delta_2\right)\right] \\
&\times \quad p(\mu)p(\sigma)p(\gamma)p(\delta_1)p(\delta_2) \, d\mu d\sigma d\gamma d\delta_1 d\delta_2.
\end{aligned}
$$

From the last expression we obtain the necessary condition (12).

19

Analogous results can be obtained for δ_2 by integrating μ over $(x_{(n)}, \infty)$, where $x_{(n)}$ represents the largest order statistic of $\mathbf{x}$.

Proof of Theorem 2

(i) In this parameterization, ε in (2) does not depend on σ. This fact will be used implicitly in a change of variable below. We obtain

$$
\begin{aligned}
p(\mathbf{x}) \quad &\propto \quad \int_\Delta \int_\Delta \int_0^\infty \int_{-\infty}^\infty \int_{\mathbb{R}_+^n} \frac{1}{[a(\gamma)+b(\gamma)]^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^n \frac{\lambda_j}{i_j(\gamma)^2}(x_j-\mu)^2\right] \\
&\times \quad p(\gamma\delta_1,\delta_2)\prod_{j=1}^n \left\{\varepsilon dP_{\lambda_j|\delta_1}I(x_j<\mu)+(1-\varepsilon)dP_{\lambda_j|\delta_2}I(x_j\geq\mu)\right\}d\mu d\sigma d\gamma d\delta_1 d\delta_2 \\
&\leq \quad \int_\Delta \int_\Delta \int_0^\infty \int_{-\infty}^\infty \int_{\mathbb{R}_+^n} \frac{1}{[a(\gamma)+b(\gamma)]^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2 h(\gamma)^2}\sum_{j=1}^n \lambda_j(x_j-\mu)^2\right] \\
&\times \quad p(\gamma)p(\delta_1,\delta_2)\prod_{j=1}^n \left\{\varepsilon dP_{\lambda_j|\delta_1}I(x_j<\mu)+(1-\varepsilon)dP_{\lambda_j|\delta_2}I(x_j\geq\mu)\right\}d\mu d\sigma d\gamma d\delta_1 d\delta_2,
\end{aligned}
$$

where $i_j(\gamma)=a(\gamma)I(x_j\geq\mu)+b(\gamma)I(x_j<\mu)$ and $h(\gamma)=\max\{a(\gamma),b(\gamma)\}$. Now, consider the change of variable $\theta=\sigma h(\gamma)$, then we get that this upper bound can be written as follows

$$
\begin{aligned}
&\int_\Delta \int_\Delta \int_0^\infty \int_{-\infty}^\infty \int_{\mathbb{R}_+^n} \frac{h(\gamma)^n}{[a(\gamma)+b(\gamma)]^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\theta^{n+1}} \exp\left[-\frac{1}{2\theta^2}\sum_{j=1}^n \lambda_j(x_j-\mu)^2\right] \\
&\times \quad p(\gamma)p(\delta_1,\delta_2)\prod_{j=1}^n \left\{\varepsilon dP_{\lambda_j|\delta_1}I(x_j<\mu)+(1-\varepsilon)dP_{\lambda_j|\delta_2}I(x_j\geq\mu)\right\}d\mu d\theta d\gamma d\delta_1 d\delta_2.
\end{aligned}
$$

By using that $0\leq\varepsilon\leq 1$, $\frac{1}{2}\leq\frac{h(\gamma)^n}{[a(\gamma)+b(\gamma)]^n}\leq 1$ it follows that the propriety of the posterior of $(\mu,\sigma,\gamma,\delta_1,\delta_2)$ under this prior structure is equivalent to the propriety of the posterior distribution of a TPSH sampling model with parameters $(\mu,\sigma,\delta_1,\delta_1)$ and prior structure $\pi(\mu,\sigma,\delta_1,\delta_1)\propto\sigma^{-1}p(\delta_1,\delta_2)$, where $p(\delta_1,\delta_2)$ is a proper prior. The rest of the proof thus focuses on the latter model, for which, by construction, we have

$$
\begin{aligned}
f(x_j;\mu,\sigma,\delta_1,\delta_2) \quad &= \quad \int_0^\infty \frac{2\lambda_j^{\frac{1}{2}}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\lambda_j}{2\sigma^2}(x_j-\mu)^2\right] \\
&\times \quad \left\{\varepsilon dP_{\lambda_j|\delta_1}I(x_j<\mu)+(1-\varepsilon)dP_{\lambda_j|\delta_2}I(x_j\geq\mu)\right\},
\end{aligned}
$$

with ε as in (2). Then, we can write the marginal of $\mathbf{x}$ as follows

$$
\begin{aligned}
p(\mathbf{x}) \quad &\propto \quad \int_\Delta \int_\Delta \int_0^\infty \int_{-\infty}^\infty \int_{\mathbb{R}_+^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^n \lambda_j(x_j-\mu)^2\right] p(\delta_1,\delta_2) \\
&\times \quad \prod_{j=1}^n \left\{\varepsilon dP_{\lambda_j|\delta_1}I(x_j<\mu)+(1-\varepsilon)dP_{\lambda_j|\delta_2}I(x_j\geq\mu)\right\}d\mu d\sigma d\delta_1 d\delta_2.
\end{aligned}
$$

Separating the integral with respect to μ into $n+1$ integrals over the domains $(-\infty, x_{(1)})$, $[x_{(1)}, x_{(2)})$, ..., $[x_{(n)}, \infty)$, we have that

$$
\begin{aligned}
I_1 &= \int_\Delta \int_\Delta \int_0^\infty \int_{-\infty}^{x_{(1)}} \int_{\mathbb{R}_+^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^n \lambda_j (x_j - \mu)^2 \right] \\
&\times\; p(\delta_1, \delta_2)(1-\varepsilon)^n \prod_{j=1}^n dP_{\lambda_j|\delta_2} d\mu d\sigma d\delta_1 d\delta_2.
\end{aligned}
$$

By noting that $0 \le \varepsilon \le 1$, extending the integration domain on μ to the whole real line and integrating out δ_1 we obtain

$$
I_1 \le \int_\Delta \int_0^\infty \int_{-\infty}^\infty \int_{\mathbb{R}_+^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^n \lambda_j (x_j - \mu)^2 \right] p(\delta_2) \prod_{j=1}^n dP_{\lambda_j|\delta_2} d\mu d\sigma d\delta_2 < \infty.
$$

The finiteness of this integral is obtained using Theorem 1 from Fernández and Steel (1998b). Now, using similar arguments we have that

$$
\begin{aligned}
I_2 &= \int_\Delta \int_\Delta \int_0^\infty \int_{x_{(n)}}^\infty \int_{\mathbb{R}_+^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^n \lambda_j (x_j - \mu)^2 \right] \\
&\times\; p(\delta_1, \delta_2)\varepsilon^n \prod_{j=1}^n dP_{\lambda_j|\delta_1} d\mu d\sigma d\delta_1 d\delta_2 \\
&\le \int_\Delta \int_0^\infty \int_{-\infty}^\infty \int_{\mathbb{R}_+^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^n \lambda_j (x_j - \mu)^2 \right] \\
&\times\; p(\delta_1) \prod_{j=1}^n dP_{\lambda_j|\delta_1} d\mu d\sigma d\delta_1 < \infty.
\end{aligned}
$$

Finally, for an intermediate region we have

$$
\begin{aligned}
I_3 &= \int_\Delta \int_\Delta \int_0^\infty \int_{x_{(k)}}^{x_{(k+1)}} \int_{\mathbb{R}_+^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^n \lambda_j (x_{(j)} - \mu)^2 \right] p(\delta_1, \delta_2)\varepsilon^k (1-\varepsilon)^{n-k} \\
&\times\; \prod_{j=1}^k dP_{\lambda_j|\delta_1} \prod_{j=k+1}^n dP_{\lambda_j|\delta_2} d\mu d\sigma d\delta_1 d\delta_2 \\
&\le \int_\Delta \int_\Delta \int_0^\infty \int_{-\infty}^\infty \int_{\mathbb{R}_+^n} \frac{\prod_{j=1}^n \lambda_j^{\frac{1}{2}}}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^n \lambda_j (x_{(j)} - \mu)^2 \right] p(\delta_1, \delta_2) \\
&\times\; \prod_{j=1}^k dP_{\lambda_j|\delta_1} \prod_{j=k+1}^n dP_{\lambda_j|\delta_2} d\mu d\sigma d\delta_1 d\delta_2 < \infty.
\end{aligned}
$$

The finiteness follows again from Theorem 1 from Fernández and Steel (1998b). Combining the finiteness of I_1, I_2 and I_3 the result follows.

(ii) This follows using the previous proof together with Theorems 1, 2, and 3 from Fernández and Steel (1998b).

Proof of Corollary 1

From the proof of point (i) in Theorem 1 it follows that the propriety of the posterior distribution of $(\mu, \sigma, \gamma, \delta_1, \delta_2)$ is equivalent to proving the propriety of (μ, σ, δ), assuming that $S_1, \ldots, S_n$ is an i.i.d. sample of set observations from a scale mixture of normals $f(\cdot; \mu, \sigma, \delta)$ and adopting the prior $\pi(\mu, \sigma, \delta) \propto \sigma^{-1} p(\delta)$, where $p(\delta)$ is proper. The result then follows by combining this fact with Theorem 4 from Fernández and Steel (1998b).

References

Aas, K., and Haff, I. H. (2006), "The Generalized Hyperbolic Skew Student's $t-$distribution," *Journal of Financial Econometrics*, 4, 275–309.

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.

Azzalini, A. (1986), "Further Results on a Class of Distributions Which Includes the Normal Ones," *Statistica*, 46, 199–208.

Azzalini, A., and Capitanio, A. (2003), "Distributions Generated by Perturbation of Symmetry With Emphasis on a Multivariate Skew-t Distribution," *Journal of the Royal Statistical Society* B, 65, 367–389.

Barndorff-Nielsen, O. E. (1977), "Exponentially Decreasing Distributions for the Logarithm of Particle Size," *Proceedings of the Royal Society of London* A, 353, 401-419.

Berlaint, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications,* Wiley, New York.

Christen, J. A., and Fox, C. (2010), "A General Purpose Sampling Algorithm for Continuous Distributions (The t-walk)," *Bayesian Analysis*, 5, 1–20.

Critchley, F., and Jones, M. C. (2008), "Asymmetry and Gradient Asymmetry Functions: Density-Based Skewness and Kurtosis," *Scandinavian Journal of Statistics*, 35, 415-437.

Doss, H., and Hobert, J. P. (2010), "Estimation of Bayes Factors in a Class of Hierarchical Random Effects Models Using Geometrically Ergodic MCMC Algorithm," *Journal of Computational and Graphical Statistics*, 19, 295–312.

Dunson, D. B. (2010), "Nonparametric Bayes Applications to Biostatistics," In *Bayesian Nonparametrics* (Hjort, N. L., Holmes, C. .C, Müller, P. Walker, S. G. Eds.), pp. 223–273. Cambridge University Press, Cambridge.

Fernández, C., Osiewalski, J., and Steel, M. F. J. (1995), "Modeling and Inference With v-Spherical Distributions," *Journal of the American Statistical Association*, 90, 1331-1340.

Fernández, C., and Steel, M. F. J. (1998a), "On Bayesian Modeling of Fat Tails and Skewness," *Journal of the American Statistical Association*, 93, 359–371.

Fernández, C. and Steel, M. F. J. (1998b). On the dangers of modelling through continuous distributions: A Bayesian perspective, in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. eds., *Bayesian Statistics 6*, Oxford University Press (with discussion), pp. 213–238.

Fernández, C., and Steel, M. F. J. (2000), "Bayesian Regression Analysis With Scale Mixtures of Normals," *Econometric Theory*, 16, 80–101.

Ferreira, J. T. A. S., and Steel, M. F. J. (2006), "A Constructive Representation of Univariate Skewed Distributions," *Journal of the American Statistical Association*, 101, 823–829.

Ferreira, J. T. A. S., and Steel, M. F. J. (2007), "A New Class of Skewed Multivariate Distributions With Applications to Regression Analysis," *Statistica Sinica*, 17, 505–529.

Finley, A. O., and Banerjee, S. (2013), "spBayes: Univariate and Multivariate Spatial-temporal Modeling," R package version 0.3-8, http://CRAN.R-project.org/package=spBayes.

Fischer, M., and Klein, I. (2004), "Kurtosis Modelling by Means of the $J-$Transformation," *Allgemeines Statistisches Archiv*, 88, 35–50.

Goerg, G. M. (2011), "Lambert W Random Variables - A New Generalized Family of Skewed Distributions with Applications to Risk Estimation," *The Annals of Applied Statistics*, 5, 2197–2230.

Groeneveld, R. A., and Meeden, G. (1984), "Measuring Skewness and Kurtosis," *The Statistician*, 33, 391-399.

Hansen, B. E. (1994), "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705–730.

Haynes, M. A., MacGilllivray, H. L., and Mergersen, K. L. (1997), "Robustness of Ranking and Selection Rules Using Generalized g and k Distributions," *Journal of Statistical Planning and Inference*, 65, 45–66.

Johnson, N. L. (1949), "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, 36, 149–176.

Jones, M. C. (2014), "On Families of Distributions With Shape Parameters," *International Statistical Review*, in press (with discussion).

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.

Jones, M. C., and Faddy, M. J. (2003), "A Skew Extension of the t-Distribution, With Applications," *Journal of Royal Statistical Society* Series B, 65, 159–174.

Jones, M. C., and Pewsey A. (2009), "Sinh-arcsinh Distributions," *Biometrika*, 96, 761–780.

Klein, I., and Fischer, M. (2006), "Power Kurtosis Transformations: Definition, Properties and Ordering," *Allgemeines Statistisches Archiv*, 90, 395–401.

Ley, C., and Paindaveine, D. (2010), "Multivariate Skewing Mechanisms: A Unified Perspective Based On the Transformation Approach," *Statistics & Probability Letters*, 80, 1685–1694.

Marinho, V. C. C., Higgins, J. P. T., Logan, S., and Sheiham, A. (2003), "Fluoride Toothpastes for Preventing Dental Caries in Children and Adolescents (Cochrane Review)," *The Cochrane Library*, (Issue 4 edn). Wiley: Chichester.

McCulloch, M. E., and Neuhaus, J. M. (2011), "Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter," *Statistical Science*, 26, 388–402.

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.

Polson, N., and Scott, J. G. (2012), "On the Half-Cauchy Prior for a Global Scale Parameter," *Bayesian Analysis*, 7, 887–902.

Ramirez-Cobo, P., Lillo, R. E., Wilson, S., and Wiper, M. P. (2010), "Bayesian Inference for Double Pareto Lognormal Queues," *The Annals of Applied Statistics*, 4, 1533–1557.

Reed, W., and Jorgensen, M. (2004), "The double Pareto-lognormal Distribution – A New Parametric Model for Size Distributions," *Communications in Statistics, Theory & Methods*, 33, 1733-1753.

Roberts, G. O., and Rosenthal, J. S. (2009), "Examples of Adaptive MCMC," *Journal of Computational and Graphical Statistics*, 18, 349–367.

Rosco, J. F., Jones, M. C., and Pewsey, A. (2011), "Skew t Distributions Via the Sinh-arcsinh Transformation," *TEST* 20: 630–652.

Rubio, F. J. (2013), *Modelling of Kurtosis and Skewness: Bayesian Inference and Distribution Theory*, PhD Thesis, University of Warwick, UK.

Rubio, F. J., and Steel, M. F. J. (2013), "Bayesian Inference for P$(X < Y)$ Using Asymmetric Dependent Distributions," *Bayesian Analyisis*, 8, 43–62.

Rubio, F. J., and Steel, M. F. J. (2014), "Inference in Two-Piece Location-Scale models With Jeffreys Priors (with discussion)," *Bayesian Analyisis*, 9, 1–22.

Thompson, S. G., and Lee, K. J. (2008), "Flexible Parametric Models for Random–Effects Distributions," *Statistics in Medicine*, 27, 418–434.

Tukey, J. M. (1977), *Exploratory Data Analysis*, Addison-Wesley, Reading, M. A.

van Zwet, W. R. (1964), *Convex Transformations of Random Variables*, Mathematisch Centrum, Amsterdam.

Venturini, S., Dominici, F., and Parmigiani, G. (2008), "Gamma Shape Mixtures for Heavy-tailed Distributions," *Annals of Applied Statistics*, 2, 756–776.

Wang, J., Boyer, J., and Genton M. C. (2004), "A Skew Symmetric Representation of Multivariate Distributions," *Statistica Sinica*, 14, 1259–1270.

Zhang, D., and Davidian, M. (2001). "Linear Mixed Models with Flexible Distributions of Random Effects for Longitudinal Data," *Biometrics* 57, 795–802.

Zhu, D., and Galbraith, J. W. (2010), "A Generalized Asymmetric Student-t Distribution With Application to Financial Econometrics," *Journal of Econometrics*, 157, 297–305.

Zhu, D., and Galbraith, J. W. (2011), "Modeling and Forecasting Expected Shortfall With the Generalized Asymmetric Student-t and Asymmetric Exponential Power Distributions," *Journal of Empirical Finance*, 18, 765–778.

Zhu, D., and Zinde-Walsh, V. (2009), "Properties and Estimation of Asymmetric Exponential Power Distribution," *Journal of Econometrics*, 148, 86-99.