Institute of Mathematics and Computational Sciences

ISSN - 0103-2569

A Study of Skewed Heavy-tailed Distributions as Scale Mixtures

Ricardo S. Ehlers

No 387

ICMC TECHNICAL REPORT

São Carlos, Brazil
October 2012

A Study of Skewed Heavy-tailed Distributions as Scale Mixtures

Ricardo S. Ehlers*
University of São Paulo - Brazil

Abstract

In this paper, we study and compare different proposals of heavy-tailed (possibly skewed) distributions as robust alternatives to the normal model. The density functions are all represented as scale mixtures which enables efficient Bayesian estimation via Markov chain Monte Carlo (MCMC) methods. However, while the symmetric versions of these distributions are able to model heavy tails they of course fail to capture asymmetry for example when the data set contains extreme values in one of the tails. Therefore, distributions that accommodate skewness as well as fat tails are also taken into account.

Key words: Scale mixture of normals, scale mixture of uniforms, skewed distributions, MCMC.

1 Introduction

In this paper, we study and compare different proposals of heavy-tailed (possibly skewed) distributions as robust alternatives to the normal model. The density functions are all represented as scale mixtures as for example in Choy and Smith (1997) and Choy and Chan (2008). This enables efficient Bayesian estimation via Markov chain Monte Carlo (MCMC) methods. We begin by describing and reviewing a number of scale mixture representation proposals in the literature for symmetric distributions.

A rich class of continuous symmetric and unimodal distributions can be expressed as a scale mixture of normal distributions as defined below.

Definition 1.1. *A continuous random variable X is said to have a scale mixture of normal distributions with location parameter μ and scale parameter σ if,*

$$f(x) = \int_0^\infty (2\pi\kappa(\lambda)\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\kappa(\lambda)\sigma^2}(x-\mu)^2\right\} dH(\lambda).$$

*Corresponding author. Email: ehlers@icmc.usp.br

where $\kappa(\cdot)$ is a positive function, λ is a positive random variable and H is a distribution function such that $H(0) = 0$.

H is usually refered to as the mixing distribution function. In this case, the random variable X admits the stochastic representation

$$X = \mu + \kappa(\lambda)^{1/2} Y,$$

where $Y \sim N(0, \sigma^2)$ so that $X|\lambda \sim N(\mu, \kappa(\lambda)\sigma^2)$. Therefore, the unconditional mean and variance of X are given by,

$$\begin{aligned} E(X) &= E[E(X|\lambda)] = \mu \\ Var(X) &= E[Var(X|\lambda)] = \sigma^2 E(\kappa(\lambda)). \end{aligned}$$

The standardized version of the distribution of X is then obtained by setting

$$Z = \frac{X - \mu}{\sigma E^{1/2}(\kappa(\lambda))},$$

so that $E(Z) = 0$ and $Var(Z) = 1$ and the density function of Z is given by,

$$f(z) = \frac{1}{\sigma E^{1/2}(\kappa(\lambda))} \int_0^\infty (2\pi\kappa(\lambda)\sigma^2)^{-1/2} \exp\left\{-\frac{E(\kappa(\lambda))y^2}{2\kappa(\lambda)}\right\} dH(\lambda).$$

For example, the Student t distribution with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\nu > 0$ is obtained by setting $\kappa(\lambda) = 1/\lambda$ and assigning $\lambda \sim Gamma(\nu/2, \nu/2)$. In this case, we can use the following hierarchical form

$$\begin{aligned} X|\mu, \sigma, \lambda &\sim N(\mu, \sigma^2/\lambda) \\ \lambda|\nu &\sim Gamma(\nu/2, \nu/2). \end{aligned}$$

The unconditional mean and variance of X are given by,

$$\begin{aligned} E(X) &= \mu \\ Var(X) &= \sigma^2 E(\lambda^{-1}) = \sigma^2 \frac{\nu}{\nu - 2}, \; \nu > 2, \end{aligned}$$

since $\lambda^{-1} \sim IG(\nu/2, \nu/2)$ where $IG(\cdot, \cdot)$ denotes the Inverse Gamma distribution.

Many other distributions can be represented in this fashion. For example, if the mixing distribution is $Gamma(\nu/2, \delta/2)$ we have a Pearson Type VII distribution (Johnson et al. 1994) with shape parameters ν and δ (when $\nu = \delta$ we recover the Student t distribution). If the mixing distribution is $Beta(\nu, 1)$ we obtain the Slash distribution (when $\nu \to \infty$ we recover the normal distribution). Barndorf-Nielsen (1978) obtained the generalized hyperbolic distribution as a scale mixture of normals using the generalized inverse Gaussian as a mixing distribution. In the following sections we describe other heavy-tailed distributions and other forms of scale mixtures proposed in the literature.

2 Other heavy-tailed distributions and scale mixtures

As another heavy-tailed distribution we consider the generalized error distribution (GED) also known as exponential power distribution (Box and Tiao 1973). Here, we use the definition that appears for example in Gómez et al. (1998).

Definition 2.1. *A continuous random variable X has a generalized error distribution with parameters μ, σ and β if its density function is given by,*

$$f(x|\mu,\sigma,\beta) = \frac{1}{\sigma\Gamma(1+\frac{1}{2\beta})2^{1+1/2\beta}} \exp\left(-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right), \quad x \in \mathbb{R}.$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameters respectively and $\beta > 0$ determines the kurtosis of the distribution.

We refer to this distribution as $GED(\mu, \sigma, \beta)$. Important special cases of the GED are, the Laplace (or double exponential) distribution for $\beta = 1/2$ and the normal distribution for $\beta = 1$. The kurtosis of this GED distribution is given by $\Gamma(1/2\beta)\Gamma(5/2\beta)/\Gamma(3/2\beta)^2 - 3$ so that values $0 < \beta < 1$ lead to leptokurtic distributions while values $\beta > 1$ lead to tails thinner than in the normal distribution.

Choy and Smith (1997) obtained the GED distribution as a scale mixture of normals for $\beta \in (0.5, 1]$, a result latter extended by Walker and Gutiérrez-Peña (1999) who proposed a new representation that is valid for the entire range of β. The representation proposed is a scale mixture of uniform distributions (see also Fung and Seneta 2008).

Definition 2.2. *A continuous random variable X is said to have a scale mixture of uniforms representation if its density function can be expressed as,*

$$f(x) = \int_0^\infty f_U(\mu - \kappa(u)\sigma, \mu + \kappa(u)\sigma)dH(u).$$

where $f_U(a,b)$ denotes the uniform density function on (a,b), μ and σ are the location and scale parameters, $\kappa(\cdot)$ is a positive function and u is a positive random variable.

Note that the uniform distributions in the mixture have the same mean μ but different supports controlled by the mixing variable u. The unconditional mean and variance of X are then obtained as,

$$\begin{aligned} E(X) &= \mu \\ Var(X) &= E[Var(X|u)] = \frac{\sigma^2}{3}E(\kappa^2(u)). \end{aligned}$$

Theorem 2.1. *The generalized error distribution can be expressed in the following hierarchical form,*

$$\begin{aligned} X|u &\sim U\left(\mu - 2^{(1-\beta)/2\beta}\sigma u^{1/2\beta}, \mu + 2^{(1-\beta)/2\beta}\sigma u^{1/2\beta}\right) \\ u &\sim Gamma\left(1 + \frac{1}{2\beta}, 2^{-\beta}\right). \end{aligned}$$

Proof. This follows from writing the joint density of X and u as,

$$f(x|u)f(u) \propto u^{-1/2\beta} I\left(u > \delta\right) u^{1/2\beta} \exp(-2^{-\beta}u)$$

where,

$$\delta = 2^{\beta-1}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}.$$

So, the marginal density of X is given by,

$$\begin{aligned} f(x) &\propto \int_0^\infty I\left(u > \delta\right) \exp(-2^{-\beta}u)du \\ &\propto \int_\delta^\infty \exp(-2^{-\beta}u)du \\ &\propto \exp\left(-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right) \end{aligned}$$

and we can conclude that $X \sim GED(\mu, \sigma, \beta)$. □

Under this hierarchical representation it is easy to obtain the unconditional mean and variance of X,

$$\begin{aligned} E(X) &= \mu \\ Var(X) &= E[Var(X|u)] = \frac{2^{(1-\beta)/\beta}\sigma^2}{3} E(u^{1/\beta}) \\ &= \frac{2^{(1-\beta)/\beta}\sigma^2}{3} \frac{(2^{-\beta})^{1+1/2\beta}}{\Gamma\left(1+\frac{1}{2\beta}\right)} \frac{\Gamma\left(1+\frac{3}{2\beta}\right)}{(2^{-\beta})^{1+3/2\beta}} \\ &= \frac{2^{1/\beta}\,\Gamma\left(\frac{3}{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}\sigma^2. \end{aligned}$$

Note that, since the normal distribution is a special case of the GED when $\beta = 1$ it also admits a scale mixture of uniforms representation,

$$\begin{aligned} X|u &\sim U\left(\mu - \sigma u^{1/2}, \mu + \sigma u^{1/2}\right) \\ u &\sim Gamma\left(\frac{3}{2}, \frac{1}{2}\right). \end{aligned}$$

Therefore, Definition 1.1 can be rewritten using the following hierarchical representation,

$$
\begin{aligned}
X|u,\lambda &\sim U\left(\mu - \kappa(\lambda)^{1/2}\sigma u^{1/2}, \mu + \kappa(\lambda)^{1/2}\sigma u^{1/2}\right) \\
u &\sim Gamma\left(\frac{3}{2}, \frac{1}{2}\right).
\end{aligned}
$$

so that, as noted in Choy and Chan (2008), in theory a density function represented as a scale mixture of normals also admits a scale mixture of uniforms representation.

McDonald and Newey (1988) introduced a flexible symmetric and unimodal distribution as another robust alternative to the normal distribution which they called the generalized t distribution. Its density function is given by,

$$
f(x) = \frac{p\Gamma\left(q + \frac{1}{p}\right)}{2q^{1/p}\sigma\Gamma\left(\frac{1}{p}\right)\Gamma(q)} \left(1 + \frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^{p}\right)^{-(q+1/p)}, \quad x \in \mathbb{R}
$$

where μ and σ are location and scale parameters respectively and $p > 0$ and $q > 0$ are two shape parameters. We refer to this distribution as $GT(\mu, \sigma, p, q)$. Larger values of p and q yield a density with thinner tails than the normal while smaller values are associated with thicker tailed densities. Also, it includes other well known symmetric unimodal distributions as special or limiting cases. For example, the t distribution with $\nu = 2q$ degrees of freedom and scale $\alpha = 2^{-1/2}\sigma$ is obtained when $p = 2$ and consequently the normal distribution with mean μ and variance $\sigma^2/2$ is obtained when $p = 2$ and $q \to \infty$. The GED in Definition 2.1 with location μ, scale $2^{-1/p}\sigma$ and $p = 2\beta$ is obtained when $q \to \infty$. For the purposes of this paper, the important feature of the generalized t distribution is that it can be represented as a scale mixture of a GED with a generalized Gamma as the mixing distribution, a result obtained by Arslan and Genç (2003).

Theorem 2.2. *The generalized t distribution with parameters μ, σ, p and q can be expressed in the following hierarchical form,*

$$
\begin{aligned}
X|s &\sim GED\left(\mu, \frac{2q^{2/p}\Gamma(3/p)\sigma^2}{s\Gamma(1/p)}, \frac{2}{p}\right) \\
s &\sim GG(q, 1, p/2),
\end{aligned}
$$

where $GG(\cdot,\cdot,\cdot)$ denotes the generalized Gamma distribution.

Proof. See Arslan and Genç (2003).

Recently, Choy and Chan (2008) proposed to represent the generalized t as a scale mixture of uniforms in order to overcome difficulties to handle the distribution in the above representation. This can be accomplished by expressing the distribution hierarchically as,

$$\begin{aligned} X|u,s &\sim U(\mu - \sigma q^{1/p} s^{-1/2} u^{1/p}, \mu + \sigma q^{1/p} s^{-1/2} u^{1/p}) \\ u &\sim Gamma\left(1 + \frac{1}{p}, 1\right) \\ s &\sim GG\left(q, 1, \frac{p}{2}\right). \end{aligned}$$

These authors compared various scale mixtures as sampling distributions for a real dataset. They adopted a Bayesian approach to estimate parameters via MCMC and performed model comparison in terms of Deviance Information Criterion (DIC, Spiegelhalter et al. 2002). However, while these distributions are able to model heavy tails they fail to capture asymmetry for example when the data set contains extreme values in one of the tails. Therefore, distributions that accommodate skewness as well as fat tails should be taken into account too. In the next section, we propose to consider a class of skewed distributions and show that they inherit the scale mixture representation of their symmetric counterparts.

3 Introducing Skewness

There are a number of proposals in the literature to introduce skewness in symmetric unimodal distributions. The most common way to create a univariate skewed distribution is to introduce skewness into an originally symmetric distribution. In this approach, skewed distributions can be generated by for example, hidden truncation models (Azzalini 1985), inverse scale factors (Fernandez and Steel 1998), and order statistics (Jones 2004). All these methods have the advantage of preserving at least a subset of the properties of the original symmetric distribution, which are often well known.

In particular, Fernandez and Steel (1998) presented a general method for transforming any continuous unimodal and symmetric distribution into a skewed one by changing the scale at each side of the mode. They proposed the following class of skewed distributions indexed by a shape parameter $\gamma > 0$, which describes the degree of asymmetry,

$$s(\epsilon|\gamma) = \frac{2}{\gamma + 1/\gamma}\left\{f\left(\frac{\epsilon}{\gamma}\right) I_{[0,\infty)}(\epsilon) + f(\epsilon\gamma) I_{(-\infty,0)}(\epsilon)\right\}, \qquad (1)$$

where $f(\cdot)$ is a univariate density symmetric around zero and $I_C(\cdot)$ is an indicator function on C. Note that $\gamma = 1$ yields the symmetric distribution as $s(\epsilon|\gamma = 1) = f(\epsilon)$, and values of $\gamma > 1$ (< 1) indicate right (left) skewness.

Our preference for this skewing mechanism is mainly due to its simplicity and generality. Moments calculation is straightforward if the moments of the underlying symmetric distribution are available and it does not require calculation of cumulative distribution functions, which yields faster computations. Also, it entirely separates the effects of the skewness and tail parameters thus making prior independence between the two a plausible assumption, and hence facilitates the choice of their prior distributions.

A continuous random variable X with location and scale parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ can be represented as $X = \mu + \sigma\epsilon$ and its density function is then given by

$$s(x|\gamma) = \frac{2}{\sigma(\gamma + 1/\gamma)} \left\{ f\left(\frac{x-\mu}{\sigma\gamma}\right) I_{[0,\infty)}(x-\mu) + f\left(\frac{(x-\mu)\gamma}{\sigma}\right) I_{(-\infty,0)}(x-\mu) \right\}.$$

Thus, $s(x|\gamma)$ is the skewed version of the location-scale density $f(\cdot)$ preserving the mode μ.

Choosing $f(\cdot)$ to be the standard normal density we obtain the skewed normal distribution with parameters μ, σ and γ denoted $SN(\mu, \sigma, \gamma)$ and density function given by,

$$\begin{aligned} s(x|\gamma) &= \left(\frac{2}{\pi}\right)^{1/2} \frac{1}{\sigma(\gamma + 1/\gamma)} \\ &\quad \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \left(\frac{1}{\gamma^2} I_{[0,\infty)}(x-\mu) + \gamma^2 I_{(-\infty,0)}(x-\mu)\right)\right\} \end{aligned}$$

while choosing $f(\cdot)$ to be the standard Student t density we obtain the skewed Student distribution with parameters μ, σ, ν and γ denoted $SSTD(\mu, \sigma, \nu, \gamma)$ and density function given by,

$$\begin{aligned} s(x|\gamma) &= \frac{2\Gamma(\frac{\nu+1}{2})}{\sigma\Gamma(\frac{\nu}{2})(\gamma + 1/\gamma)(\pi\nu)^{1/2}} \\ &\quad \left[1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2 \left\{\frac{1}{\gamma^2} I_{[0,\infty)}(x-\mu) + \gamma^2 I_{[-\infty,0)}(x-\mu)\right\}\right]^{-\frac{\nu+1}{2}}. \end{aligned}$$

This can be represented as a scale mixture of skewed normals in the following hierarchical form,

$$\begin{aligned} X|\lambda &\sim SN(\mu, \sigma^2/\lambda, \gamma) \\ \lambda &\sim Gamma(\nu/2, \nu/2). \end{aligned}$$

We shall assume that all the parameters are a priori independent. Noninformative prior distributions for μ and $\tau = \sigma^{-1}$ are assigned as $p(\mu, \tau) \propto \tau^{-1}$. Following Fernandez and Steel (1998), we shall use a Gamma(a, b) prior distribution on $\phi = \gamma^2$ which is the ratio of probability masses above and below the mode, i.e. $\gamma^2 = Pr(\epsilon \geq 0)/Pr(\epsilon < 0)$. So, for observed data $\boldsymbol{x} = (x_1, \ldots, x_n)$ the complete conditional distributions in the skewed normal model are given by,

$$\tau^2|\boldsymbol{x}, \mu, \phi \sim Gamma\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 \phi^{-sign(x_i-\mu)}\right)$$

$$f(\phi|\boldsymbol{x}, \mu, \tau) \propto \phi^{a+n/2-1}(\phi+1)^{-n} \exp\left\{-\frac{\tau^2}{2}\sum_{i=1}^{n}(x_i-\mu)^2\phi^{-sign(x_i-\mu)} - b\phi\right\}$$

$$f(\mu|\boldsymbol{x},\tau,\phi)\propto\exp\left\{-\frac{\tau^2}{2}\sum_{i=1}^{n}(x_i-\mu)^2\left(\frac{1}{\phi}I_{[0,\infty)}(x_i-\mu)+\phi I_{(-\infty,0)}(x_i-\mu)\right)\right\}.$$

The complete conditional density of ϕ is not of any standard form and we use a Metropolis-Hastings algorithm. The complete conditional density of μ will be rewritten in a form which allows to easily draw values. Ordering the observations $x_{(1)}\le x_{(2)}\le\cdots\le x_{(n)}$ and defining the sub-intervals $S_0=(-\infty,x_{(1)}]$, $S_h=(x_{(h)},x_{(h+1)}]$, $h=1,\ldots,n-1$ and $S_n=(x_{(n)},\infty)$ it follows that,

$$\begin{aligned}
f(\mu|\boldsymbol{x},\tau,\phi,\mu\in S_h) &\propto \exp\left\{-\frac{\tau^2}{2}\left[\sum_{i=1}^{h}\phi x_{(i)}^2+\sum_{i=h+1}^{n}x_{(i)}^2/\phi\right]\right\}\times\\
&\exp\left\{-\frac{\tau^2}{2}\left[-2\mu\left(\sum_{i=1}^{h}\phi x_{(i)}+\sum_{i=h+1}^{n}x_{(i)}/\phi\right)+\mu^2(h\phi+(n-h)/\phi)\right]\right\}\\
&\propto \exp\left\{-\frac{\tau^2}{2}\left[\sum_{i=1}^{h}\phi x_{(i)}^2+\sum_{i=h+1}^{n}x_{(i)}^2/\phi\right]\right\}\exp\left\{-\frac{\tau^2p_h}{2}(\mu^2-2\mu\mu_h)\right\}\\
&\propto \exp\left\{-\frac{\tau^2}{2}\left[\sum_{i=1}^{h}\phi x_{(i)}^2+\sum_{i=h+1}^{n}x_{(i)}^2/\phi-p_h\mu_h^2\right]\right\}\exp\left\{-\frac{\tau^2p_h}{2}(\mu-\mu_h)^2\right\}
\end{aligned}$$

where $p_h=h\phi+(n-h)/\phi$ and $p_h\mu_h=\sum_{i=1}^{h}\phi x_{(i)}+\sum_{i=h+1}^{n}x_{(i)}/\phi$. Then, the complete conditional density of μ can be written as,

$$f(\mu|\boldsymbol{x},\tau,\phi)\propto\sum_{h=0}^{n}\frac{1}{\sqrt{p_h}}\exp\left\{-\frac{\tau^2}{2}l_h\right\}f_N(\mu|\mu_h,1/\tau^2p_h)I_{S_h}(\mu),$$

where $l_h=\sum_{i=1}^{h}\phi x_{(i)}^2+\sum_{i=h+1}^{n}x_{(i)}^2/\phi-p_h\mu_h^2$ and $f_N(\cdot|m,c)$ denotes the density function of a normal distribution with mean m and variance c. So, in order to draw a value of μ we randomly choose a sub-interval S_h, $h=0,\ldots,n$ and sample a value from a $N(\mu_h,1/\tau^2p_h)$ distribution truncated to S_h.

In the skewed Student model written in the hierarchical form $X_i|\lambda_i\sim SN(\mu,(\tau^2\lambda_i)^{-1},\gamma)$ and $\lambda_i\sim Gamma(\nu/2,\nu/2)$, we need to specify a prior distribution for the degrees of freedom parameter ν. In order to avoid introducing strong prior information while not bounding ν away from zero we assign an exponential prior distribution with both mean and standard deviation equal to 10. Also, we assume that the mixing parameters λ_i are a priori independent. The complete conditional distributions are then obtained in a similar vein using this hierarchical form of the skewed Student. It follows immediately that,

$$\tau^2|\boldsymbol{x},\mu,\phi,\boldsymbol{\lambda}\sim Gamma\left(\frac{n}{2},\frac{1}{2}\sum_{i=1}^{n}\lambda_i(x_i-\mu)^2\phi^{-sign(x_i-\mu)}\right)$$

$$f(\phi|\boldsymbol{x},\mu,\tau,\boldsymbol{\lambda}) \propto \phi^{a+n/2-1}(\phi+1)^{-n}\exp\left\{-\frac{\tau^2}{2}\sum_{i=1}^{n}\lambda_i(x_i-\mu)^2\phi^{-sign(x_i-\mu)}-b\phi\right\}$$

$$f(\mu|\boldsymbol{x},\tau,\phi,\boldsymbol{\lambda}) \propto \sum_{h=0}^{n}\frac{1}{\sqrt{p_h}}\exp\left\{-\frac{\tau^2}{2}l_h\right\}f_N(\mu|\mu_h,1/\tau^2 p_h)I_{S_h}(\mu),$$

where

$$\begin{aligned}
p_h &= \sum_{i=1}^{h}\lambda_i\phi+\sum_{i=h+1}^{n}\lambda_i/\phi \\
p_h\mu_h &= \sum_{i=1}^{h}\phi\lambda_i x_{(i)}+\sum_{i=h+1}^{n}\lambda_i x_{(i)}/\phi \\
l_h &= \sum_{i=1}^{h}\phi\lambda_i x_{(i)}^2+\sum_{i=h+1}^{n}\lambda_i x_{(i)}^2/\phi-p_h\mu_h^2
\end{aligned}$$

$$\begin{aligned}
f(\nu|\boldsymbol{x},\mu,\tau,\phi,\boldsymbol{\lambda}) &\propto p(\boldsymbol{\lambda}|\nu)p(\nu) \\
&\propto \frac{(\nu/2)^{n\nu/2}}{\Gamma^n(\nu/2)}\exp\left\{-\nu\left[\beta+\frac{1}{2}\sum_{i=1}^{n}(\lambda_i-\log\lambda_i)\right]\right\}
\end{aligned}$$

$$\begin{aligned}
f(\boldsymbol{\lambda}|\boldsymbol{x},\mu,\tau,\phi) &\propto s(\boldsymbol{x}|\mu,\tau,\phi,\boldsymbol{\lambda})\ f(\boldsymbol{\lambda}) \\
&\propto \prod_{i=1}^{n}\lambda_i^{(\nu+1)/2-1}\exp\left\{-\lambda_i\left[\frac{\nu}{2}+\frac{\tau^2}{2}(x_i-\mu)^2\phi^{-sign(x_i-\mu)}\right]\right\}
\end{aligned}$$

We now have that the complete conditional densities of ϕ and ν are not of standard forms and Metropolis-Hastings schemes are used to sample their values.

Finally, we obtain the skewed GED with parameters μ, σ, β and γ, denoted $SGED(\mu,\sigma,\beta,\gamma)$ by choosing $f(\cdot)$ to be the standard symmetric GED. Its density function is given by,

$$\begin{aligned}
s(x|\gamma) &= \frac{2}{\sigma(\gamma+1/\gamma)\Gamma(1+\frac{1}{2\beta})2^{1+1/2\beta}} \\
&\quad \exp\left\{-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\left(\frac{1}{\gamma^{2\beta}}I_{[0,\infty)}(x-\mu)+\gamma^{2\beta}I_{(-\infty,0)}(x-\mu)\right)\right\} \qquad (2)
\end{aligned}$$

and the $SN(\mu,\sigma,\gamma)$ distribution is obtained as a special case when $\beta=1$. In the next theorem we show that this skewed version of the GED can also be represented as a scale mixture.

Theorem 3.1. *The skewed generalized error distribution can be expressed in the following hierarchical form,*

$$\begin{aligned} X|u,\gamma &\sim SU\left(\mu - 2^{(1-\beta)/2\beta}\sigma u^{1/2\beta}, \mu + 2^{(1-\beta)/2\beta}\sigma u^{1/2\beta}, \gamma\right) \\ u &\sim Gamma\left(1+\frac{1}{2\beta}, 2^{-\beta}\right), \end{aligned}$$

where $SU(a,b,\gamma)$ denotes the skewed version of the Uniform distribution.

Proof. This result follows from,

$$s(x|u,\gamma) \propto u^{-1/2\beta}[I(u>\delta_1)I_{[0,\infty)}(x-\mu) + I(u>\delta_2)I_{(-\infty,0)}(x-\mu)]$$

where,

$$\delta_1 = 2^{\beta-1}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\frac{1}{\gamma^{2\beta}} \quad \text{and} \quad \delta_2 = 2^{\beta-1}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\gamma^{2\beta}.$$

So, integrating with respect to u the above density times the density function of u we obtain,

$$\begin{aligned} s(x|\gamma) &\propto \int_{\delta_1}^{\infty}\exp(-2^{-\beta}u)du I_{[0,\infty)}(x-\mu) + \int_{\delta_2}^{\infty}\exp(-2^{-\beta}u)du I_{(-\infty,0)}(x-\mu) \\ &\propto \exp\left\{-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\left(\frac{1}{\gamma^{2\beta}}\ I_{[0,\infty)}(x-\mu) + \gamma^{2\beta}\ I_{(-\infty,0)}(x-\mu)\right)\right\} \end{aligned}$$

and $X \sim SGED(\mu,\sigma,\beta,\gamma)$. □

This result allows us to propose a skewed version of the generalized t distribution thus increasing its flexibility while enabling efficient Bayesian estimation via MCMC methods by representing it as a scale mixture. Applying the same skewing mechanism and choosing $f(\cdot)$ to be the standard generalized Student t density we obtain the skewed generalized Student distribution with parameters μ, σ, p, q and γ denoted $SGT(\mu,\sigma,p,q,\gamma)$ with density function given by,

$$\begin{aligned} s(x|\gamma) &= \frac{p\Gamma\left(q+\frac{1}{p}\right)}{\sigma(\gamma+1/\gamma)q^{1/p}\Gamma\left(\frac{1}{p}\right)\Gamma(q)} \\ &\quad \left[1+\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^{p}\left(\frac{1}{\gamma^{p}}I_{[0,\infty)}(x-\mu)+\gamma^{p}I_{(-\infty,0)}(x-\mu)\right)\right]^{-(q+1/p)}. \end{aligned} \quad (3)$$

It is not difficult to see that setting $p=2$ we recover the skewed t distribution with $2q$ degrees of freedom and scale $2^{-1/2}\sigma$. The skewed normal distribution is then obtained when $q \longrightarrow \infty$. In what follows we propose alternative representations for this skewed distribution.

Theorem 3.2. *The skewed generalized t distribution can be expressed in the following hierarchical form,*

$$\begin{aligned} X|s,\gamma &\sim SGED\left(\mu, 2^{-1/p}s^{-1/2}q^{1/p}\sigma, \frac{p}{2}, \gamma\right) \\ s &\sim GG\left(p/2, 1, q\right). \end{aligned}$$

Proof. Taking density (2) with scale parameter $2^{-1/p}s^{-1/2}q^{1/p}\sigma$ and $\beta = p/2$ and the density of a $GG\left(p/2, 1, q\right)$ it follows that,

$$\begin{aligned} s(x|s,\gamma)f(s) &\propto s^{1/2}\exp\left\{-s^{p/2}\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p\left(\frac{1}{\gamma^p}I_{[0,\infty)}(x-\mu)+\gamma^p I_{(-\infty,0)}(x-\mu)\right)\right\} \\ & \quad s^{pq/2-1}\exp(-s^{p/2}) \\ &\propto s^{(pq-2)/2} \\ & \quad \exp\left\{-s^{p/2}\left[1+\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p\left(\frac{1}{\gamma^p}I_{[0,\infty)}(x-\mu)+\gamma^p I_{(-\infty,0)}(x-\mu)\right)\right]\right\}. \end{aligned}$$

Now setting $y = s^{p/2}$ it is easy to see that,

$$\begin{aligned} s(x|\gamma) &\propto \int_0^\infty y^{(q+1/p)-1} \\ & \quad \exp\left\{-y\left[1+\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p\left(\frac{1}{\gamma^p}I_{[0,\infty)}(x-\mu)+\gamma^p I_{(-\infty,0)}(x-\mu)\right)\right]\right\}dy \\ &\propto \left[1+\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p\left(\frac{1}{\gamma^p}I_{[0,\infty)}(x-\mu)+\gamma^p I_{(-\infty,0)}(x-\mu)\right)\right]^{-(q+1/p)} \end{aligned}$$

and the normalizing constant is the one given in (3). □

Lemma 3.1. *The skewed generalized t distribution can be expressed in the following hierarchical form,*

$$\begin{aligned} X|u,s,\gamma &\sim SU\left(\mu - q^{1/p}s^{-1/2}u^{1/p}\sigma, \mu + q^{1/p}s^{-1/2}u^{1/p}\sigma, \gamma\right) \\ u &\sim Gamma\left(1+\frac{1}{p}, 1\right) \\ s &\sim GG\left(p/2, 1, q\right), \end{aligned}$$

where $SU(a, b, \gamma)$ denotes the skewed version of the Uniform distribution.

Proof. The (skewed) conditional density of X given u and s can be written as,

$$s(x|u,s,\gamma) \propto s^{1/2}u^{-1/p}[I(u > \delta_1)I_{[0,\infty)}(x-\mu) + I(u > \delta_2)I_{(-\infty,0)}(x-\mu)]$$

where,

$$\delta_1 = \frac{s^{p/2}}{q}\left|\frac{x-\mu}{\sigma}\right|^p \frac{1}{\gamma^p} \quad \text{and} \quad \delta_2 = \frac{s^{p/2}}{q}\left|\frac{x-\mu}{\sigma}\right|^p \gamma^p.$$

So, the marginal density of X is given by,

$$s(x|\gamma) \propto \int_0^\infty s^{1/2} \quad \left[\int_{\delta_1}^\infty \exp(-u)du I_{[0,\infty)}(x-\mu) + \int_{\delta_2}^\infty \exp(-u)du I_{(-\infty,0)}(x-\mu)\right] f_{GG}(s|p/2,1,q)ds,$$

where $f_{GG}(\cdot)$ denotes the density function of the Generalized Gamma distribution. Now, solving the inner integrals and multiplying by $s^{1/2}$ we obtain (up to a constant) the density function of a $SGED(\mu, 2^{-1/p}s^{-1/2}q^{1/p}\sigma, p/2, \gamma)$. From Theorem 3.2 it then follows that $X \sim SGT(\mu, \sigma, p, q, \gamma)$ and this completes the proof. □

The next step is to obtain the complete conditional distributions of the SGED and SGT using their scale mixture of uniforms representations. In the SGED model written in the hierarchical form of Theorem 3.1 we specify the prior distribution for the kurtosis parameter β as an inverse Gamma with hyperparameters a and b and assume that the mixing parameters u_i are a priori independent. The complete conditional densities are obtained below.

$$\begin{aligned} f(\boldsymbol{u}|\boldsymbol{x}, \mu, \tau, \beta, \phi) &\propto s(\boldsymbol{x}|\boldsymbol{u}, \mu, \tau, \beta, \phi)p(\boldsymbol{u}) \\ &\propto \prod_{i=1}^n \exp(-2^{-\beta}u_i) I\left(u_i > 2^{\beta-1}|x_i-\mu|^{2\beta}\tau^{2\beta}\phi^{-\beta sign(x_i-\mu)}\right). \end{aligned}$$

So,

$$u_i|\boldsymbol{u}_{-i}, x_i, \mu, \tau, \beta, \phi \sim Gamma(1, 2^{-\beta}), \quad i = 1, \dots, n$$

truncated to $u_i > 2^{\beta-1}|x_i-\mu|^{2\beta}\tau^{2\beta}\phi^{-\beta sign(x_i-\mu)}$.

$$\begin{aligned} f(\tau^2|\boldsymbol{x}, \mu, \beta, \phi, \boldsymbol{u}) &\propto (\tau^2)^{n/2-1}\prod_{i=1}^n I(0<\tau^2<2^{(1-\beta)/\beta}u_i^{1/\beta}\phi^{sign(x_i-\mu)}(x_i-\mu)^{-2}) \\ &\propto (\tau^2)^{n/2-1} I(0<\tau^2<\delta) \end{aligned}$$

where $\delta = \min\{2^{(1-\beta)/\beta}u_i^{1/\beta}\phi^{sign(x_i-\mu)}(x_i-\mu)^{-2}\}$.

$$\begin{aligned}
f(\beta|\boldsymbol{x},\mu,\tau,\phi,\boldsymbol{u}) &\propto 2^{-n(1-\beta)/2\beta}\prod_{i=1}^{n}u_i^{-1/2\beta}\beta^{-(a+1)}\exp(-b/\beta) \\
&\prod_{i=1}^{n}I\left(2^{(1-\beta)/2\beta}u_i^{1/2\beta} > |x_i-\mu|\,\tau\phi^{-sign(x_i-\mu)/2}\right) \\
&\propto \exp\left\{-\frac{1}{2\beta}\left(-n\log(2)+\sum_{i=1}^{n}\log(u_i)\right)\right\}\beta^{-(a+1)}\exp(-b/\beta) \\
&\prod_{i=1}^{n}I\left(0<\beta<\frac{\log(2u_i)}{2\log(2^{1/2}\,|x_i-\mu|\,\tau\phi^{-sign(x_i-\mu)/2})}\right) \\
&\propto \beta^{-(a+1)}\exp\left\{b+-\frac{1}{\beta}\left(-\frac{n}{2}\log(2)+\frac{1}{2}\sum_{i=1}^{n}\log(u_i)\right)\right\}I(0<\beta<\delta),
\end{aligned}$$

where $\delta=\min\{\log(2u_i)/2\log(2^{1/2}\,|x_i-\mu|\,\tau\phi^{-sign(x_i-\mu)/2})\}$. It then follows that,

$$\beta|\boldsymbol{x},\mu,\tau,\phi,\boldsymbol{u}\sim IG\left(a,b+\frac{1}{2}\left[-n\log(2)+\sum_{i=1}^{n}\log(u_i)\right]\right)$$

truncated to $0<\beta<\delta$ where $IG(a,b)$ denotes the Inverse Gamma distribution with parameters a and b and mean $b/(a-1)$.

$$\begin{aligned}
f(\phi|\boldsymbol{x},\mu,\tau,\beta,\boldsymbol{u}) &\propto \phi^{a+n/2-1}(\phi+1)^{-n}\exp(-b\phi) \\
&\prod_{i=1}^{n}I\left(\phi^{-sign(x_i-\mu)}<2^{(1-\beta)/\beta}u_i^{1/\beta}\tau^2(x_i-\mu)^{-2}\right)
\end{aligned}$$

$$\begin{aligned}
f(\mu|\boldsymbol{x},\sigma,\beta,\gamma,\boldsymbol{u}) &\propto \prod_{i=1}^{n}I\left(-2^{(1-\beta)/2\beta}u_i^{1/2\beta}<\frac{x_i-\mu}{\sigma}\gamma^{-sign(x_i-\mu)}<2^{(1-\beta)/2\beta}u_i^{1/2\beta}\right) \\
&\propto \prod_{i=1}^{n}[I(x_i-\delta_i\gamma<\mu<x_i+\delta_i\gamma)I(x_i\geq\mu)+ \\
&\quad I(x_i-\delta_i/\gamma<\mu<x_i+\delta_i/\gamma)I(x_i<\mu)]
\end{aligned}$$

where $\delta_i=2^{(1-\beta)/2\beta}u_i^{1/2\beta}\sigma$. The complete conditional distribution of μ is obtained in a similar vein as with the SN and SSTD models, i.e. by ordering the observations $x_{(1)}\leq x_{(2)}\leq\cdots\leq x_{(n)}$ and defining the sub-intervals $S_0=(-\infty,x_{(1)}]$, $S_h=(x_{(h)},x_{(h+1)}]$, $h=1,\ldots,n-1$ and $S_n=(x_{(n)},\infty)$. It then follows that,

$$\begin{aligned}
f(\mu|\boldsymbol{x},\sigma,\beta,\gamma,\boldsymbol{u},\mu\in S_h) &\propto \prod_{i=1}^{h}I\left(x_{(i)}-\frac{\delta_i}{\gamma}<\mu<x_{(i)}+\frac{\delta_i}{\gamma}\right) \\
&\prod_{i=h+1}^{n}I\Big(x_{(i)}-\delta_i\gamma<\mu<x_{(i)}+\delta_i\gamma\Big)
\end{aligned}$$

from which we obtain that,

$$
\begin{aligned}
f(\mu|\boldsymbol{x},\sigma,\beta,\gamma,\boldsymbol{u},\mu\in S_0) &\propto \prod_{i=1}^{n} I\Big(x_{(i)}-\delta_i\gamma<\mu<x_{(i)}+\delta_i\gamma\Big)\\
&\propto I\Big(\max\{\min(x_{(i)}-\delta_i\gamma),x_{(1)}-\delta_1\gamma\}<\mu<x_{(1)}\Big)\\
f(\mu|\boldsymbol{x},\sigma,\beta,\gamma,\boldsymbol{u},\mu\in S_n) &\propto \prod_{i=1}^{n} I\Big(x_{(i)}-\delta_i/\gamma<\mu<x_{(i)}+\delta_i/\gamma\Big)\\
&\propto I\Big(x_{(n)}<\mu<\min\{\max(x_{(i)}+\delta_i/\gamma),x_{(n)}+\delta_n/\gamma\}\Big)\\
f(\mu|\boldsymbol{x},\sigma,\beta,\gamma,\boldsymbol{u},\mu\in S_h) &\propto I(a<\mu<b)
\end{aligned}
$$

where $a=\max\{\min(x_{(i)}-\delta_i\gamma),x_{(h)}\}$ and $b=\min\{\max(x_{(i)}+\delta_i/\gamma),x_{(h+1)}\}$ for $h=1,\dots,n-1$. So, in order to draw a value from the complete conditional distribution of μ we randomly choose a sub-interval S_h, $h=0,\dots,n$ and then sample a value from a uniform distribution defined in one of the above intervals.

Finally, using the hierarchical form given in Lemma 3.1 for the SGT distribution we specify prior distributions for p and q as Inverse Gamma and assume that the mixing parameters u_i and s_i are a priori independent.

$$
\begin{aligned}
f(\boldsymbol{u}|\boldsymbol{x},\mu,\tau,p,q,\boldsymbol{s},\phi) &\propto s(\boldsymbol{x}|\mu,\tau,p,q,\boldsymbol{u},\boldsymbol{s},\phi)\ p(\boldsymbol{u})\\
&\propto \prod_{i=1}^{n}\Big[\exp(-u_i)I\left(u_i>|x_i-\mu|^p\tau^p\phi^{-(p/2)sign(x_i-\mu)}s_i^{p/2}/q\right)\Big]
\end{aligned}
$$

$$
\begin{aligned}
f(\boldsymbol{s}|\boldsymbol{x},\mu,\tau,p,q,\boldsymbol{u},\phi) &\propto s(\boldsymbol{x}|\mu,\tau,p,q,\boldsymbol{u},\boldsymbol{s},\phi)\ p(\boldsymbol{s})\\
&\propto \prod_{i=1}^{n} s_i^{(pq+1)/2-1}\exp(-s_i^{p/2})\\
&\quad\ I\left(\left|\frac{x_i-\mu}{\sigma}\right|\gamma^{-sign(x_i-\mu)}<q^{1/p}s_i^{-1/2}u_i^{1/p}\right)
\end{aligned}
$$

$$
\begin{aligned}
f(\tau^2|\boldsymbol{x},\mu,p,q,\boldsymbol{u},\boldsymbol{s},\phi) &\propto (\tau^2)^{n/2-1}\prod_{i=1}^{n} I\left(0<\tau^2<\frac{q^{2/p}s_i^{-1}u_i^{2/p}\phi^{sign(x_i-\mu)}}{(x_i-\mu)^2}\right)\\
&\propto (\tau^2)^{n/2-1}I\left(0<\tau^2<\min(\delta_i)\right)
\end{aligned}
$$

where $\delta_i=q^{2/p}s_i^{-1}u_i^{2/p}\phi^{sign(x_i-\mu)}/(x_i-\mu)^2$.

$$
\begin{aligned}
f(\phi|\boldsymbol{x},\mu,\tau,p,q,\boldsymbol{u},\boldsymbol{s}) &\propto \phi^{a+n/2-1}(\phi+1)^{-n}\exp(-b\phi)\\
&\quad\ \prod_{i=1}^{n} I\left(\phi^{-sign(x_i-\mu)}<\frac{q^{2/p}s_i^{-1}u_i^{2/p}\tau^2}{(x_i-\mu)^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
f(q|\boldsymbol{x},\mu,\tau,p,\boldsymbol{u},\boldsymbol{s},\phi) &\propto q^{-(c+n/p+1)}\exp(-d/q) \\
&\quad \prod_{i=1}^{n} I\left(q > \left[|x_i-\mu|\tau\phi^{-sign(x_i-\mu)/2}s_i^{1/2}\right]^p u_i^{-1}\right) \\
&\propto q^{-(c+n/p+1)}\exp(-d/q)\ I(q > \max(\delta_i))
\end{aligned}
$$

where $\delta_i = \left[\tau|x_i-\mu|\phi^{-sign(x_i-\mu)/2}s_i^{1/2}\right]^p u_i^{-1}$ and c and d are the hyperparameters in the Inverse Gamma prior distribution of q. The complete conditional distribution of q is then Inverse Gamma with parameters $c+n/p$ and d truncated to $q > \max(\delta_i)$.

$$
\begin{aligned}
f(p|\boldsymbol{x},\mu,\tau,q,\boldsymbol{u},\boldsymbol{s},\phi) &\propto q^{-n/p}\prod_{i=1}^{n} u_i^{-1/p}\ p^{-(a+1)}\exp(-b/p) \\
&\quad \prod_{i=1}^{n} I\left(\left[\tau|x_i-\mu|\phi^{-sign(x_i-\mu)/2}s_i^{1/2}\right]^p < qu_i\right) \\
&\propto p^{-(a+1)}\exp\left\{-\frac{1}{p}\left(b+n\log(q)+\sum_{i=1}^{n}\log(u_i)\right)\right\} \\
&\quad \prod_{i=1}^{n} I\left(0 < p < \frac{\log(qu_i)}{\log(\tau|x_i-\mu|\phi^{-sign(x_i-\mu)/2}s_i^{1/2})}\right).
\end{aligned}
$$

The complete conditional distribution of p is then Inverse Gamma with parameters a and $b + n\log(q) + \sum_{i=1}^{n}\log(u_i)$ truncated to $0 < p < \min(\delta_i)$ where $\delta_i = \log(qu_i)/\log(\tau|x_i-\mu|\phi^{-sign(x_i-\mu)/2}s_i^{1/2})$.

$$
\begin{aligned}
f(\mu|\boldsymbol{x},\tau,p,q,\boldsymbol{u},\boldsymbol{s},\phi) &\propto \prod_{i=1}^{n} I\left(-q^{1/p}s_i^{-1/2}u_i^{1/p} < \frac{x_i-\mu}{\sigma}\gamma^{-sign(x_i-\mu)} < q^{1/p}s_i^{-1/2}u_i^{1/p}\right) \\
&\propto \prod_{i=1}^{n}[I(x_i-\delta_i\gamma < \mu < x_i+\delta_i\gamma)I(x_i \geq \mu)+ \\
&\quad I(x_i-\delta_i/\gamma < \mu < x_i+\delta_i/\gamma)I(x_i < \mu)]
\end{aligned}
$$

where $\delta_i = q^{1/p}s_i^{-1/2}u_i^{1/p}\sigma$. So, using the same trick of ordering the observations it is not difficult to see that,

$$\begin{aligned}
f(\mu|\boldsymbol{x}, \sigma, p, q, \gamma, \boldsymbol{u}, \mu \in S_0) &\propto \prod_{i=1}^{n} I\Big(x_{(i)} - \delta_i\gamma < \mu < x_{(i)} + \delta_i\gamma\Big) \\
&\propto I\Big(\max\{\min(x_{(i)} - \delta_i\gamma), x_{(1)} - \delta_1\gamma\} < \mu < x_{(1)}\Big) \\
f(\mu|\boldsymbol{x}, \sigma, p, q, \gamma, \boldsymbol{u}, \mu \in S_n) &\propto \prod_{i=1}^{n} I\Big(x_{(i)} - \delta_i/\gamma < \mu < x_{(i)} + \delta_i/\gamma\Big) \\
&\propto I\Big(x_{(n)} < \mu < \min\{\max(x_{(i)} + \delta_i/\gamma), x_{(n)} + \delta_n/\gamma\}\Big) \\
f(\mu|\boldsymbol{x}, \sigma, p, q, \gamma, \boldsymbol{u}, \mu \in S_h) &\propto I(a < \mu < b),\ h = 1, \ldots, n-1,
\end{aligned}$$

where $a = \max\{\min(x_{(i)} - \delta_i\gamma), x_{(h)}\}$ and $b = \min\{\max(x_{(i)} + \delta_i/\gamma), x_{(h+1)}\}$.

4 Discussion

In this paper, we considered flexible families of distributions which can accommodate both heavy tails and skewness. For a particular skewing mechanism applied to heavy tailed distributions we showed how to obtain scale mixture representations.

From a Bayesian viewpoint, our proposed scale mixture representations of skewed distributions have the advantage of simplifying the complete conditional distributions thus making Bayesian computations more efficient. We notice that most of the complete conditional distributions are either uniform, Gamma or truncated Gamma, thus being easy to sample from. Some complete conditionals are nonetheless still non-standard which is the case for the tail parameter β in the (S)GED and the skewness parameter ϕ in all skewed distributions.

A natural extension of the results in this paper would be to consider skewed multivariate distributions deriving their scale mixture representations and the associated complete conditional distributions. There are different proposals in the literature to construct multivariate skewed distributions. Arslan and Genç (2009) have already derived a scale mixture representation of their In particular, Bauwens and Laurent (2005) proposed to generalize the method described in Fernandez and Steel (1998) to the multivariate case, i.e. to construct a multivariate skew distribution from a symmetric one. This is object of current and future research by the author.

Acknowledgements

This work was supported by FAPESP - Brazil, under grant number 2011/22317-0.

References

Arslan, O. and A. I. Genç (2003). Robust location and scale estimation based on the univariate generalized t (GT) distribution. *Communications in Statistics: Theory and Methods 32*, 1505–1525.

Arslan, O. and A. I. Genç (2009). The skew generalized t (sgt) distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation. *Statistics 43*(5), 481–498.

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

Barndorf-Nielsen, O. E. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics 5*, 151–157.

Bauwens, L. and S. Laurent (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics 23*, 346–354.

Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Publishing Reading, MA.

Choy, S. T. B. and J. S. K. Chan (2008). Scale mixtures distributions in statistical modelling. *Australia and New Zealand Journal of Statistics 50*(2), 135–146.

Choy, S. T. B. and A. F. M. Smith (1997). On robust analysis of a normal location parameter. *Journal of the Royal Statistical Society, Series B 59*, 463–474.

Fernandez, C. and M. Steel (1998). On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association 93*, 359–371.

Fung, T. and E. Seneta (2008). A characterisation of scale mixtures of the uniform distribution. *Statistics and Probability Letters 78*, 2883–2888.

Gómez, E., M. A. Gómez-Villegas, and J. M. Marín (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics - Theory and Methods 27*(3), 589–600.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions* (2nd ed.), Volume 1. John Wiley, New York.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics (with discussion). *Test 13*, 1–43.

McDonald, J. B. and W. K. Newey (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Econometric Theory 4*, 428–457.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B 64*, 1–34.

Walker, S. G. and E. Gutiérrez-Peña (1999). Robustifying Bayesian procedures. In *Bayesian Statistics 6*, pp. 685–710. New York: Oxford University Press.