Natural (non-)informative priors for skew-symmetric distributions

Holger Dette[a], Christophe Ley[b], and Francisco J. Rubio[c]

[a] Ruhr-Universität Bochum, Germany.

[b] Ghent University, Belgium.

[c] London School of Hygiene & Tropical Medicine, UK.

Abstract

In this paper, we present an innovative method for constructing proper priors for the skewness (shape) parameter in the skew-symmetric family of distributions. The proposed method is based on assigning a prior distribution on the perturbation effect of the shape parameter, which is quantified in terms of the Total Variation distance. We discuss strategies to translate prior beliefs about the asymmetry of the data into an informative prior distribution of this class. We show via a Monte Carlo simulation study that our noninformative priors induce posterior distributions with good frequentist properties, similar to those of the Jeffreys prior. Our informative priors yield better results than their competitors from the literature. We also propose a scale- and location-invariant prior structure for models with unknown location and scale parameters and provide sufficient conditions for the propriety of the corresponding posterior distribution. Illustrative examples are presented using simulated and real data.

Keywords: Measure of skewness; Prior elicitation; Skew-symmetric distributions; Total variation distance.

1 Introduction

It is a well-known fact that several data sets cannot be modeled by means of symmetric distributions, and hence even less via the normal distribution, due to skewness inherent to the data. Such data are frequently encountered in domains such as biometry, finance, materials sciences or environmetrics, to cite but these. See for instance Ley (2015) for detailed explanations.

Given these needs, there exists a plethora of distinct proposals for skew distributions in the literature; for a recent and extensive overview of the state-of-the-art, we refer the reader to the

discussion paper Jones (2015). A popular class of such distributions are the *skew-symmetric densities* of the form

$$s_{f;G}(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) G\left(\lambda \omega \left(\frac{x-\mu}{\sigma}\right)\right), \quad x \in \mathbb{R}, \tag{1}$$

with f the symmetric density (to be skewed), G any symmetric, univariate, absolutely continuous cumulative distribution function (cdf), and ω an odd function (Azzalini and Capitanio, 2003; Wang et al., 2004). In (1), $\mu \in \mathbb{R}$ is a location, $\sigma \in \mathbb{R}_0^+$ a scale, $\lambda \in \mathbb{R}$ a skewness parameter, and we omit the dependence on ω in $s_{f;G}$ for the sake of notation. These distributions generalize the popular *skew-normal* distribution, corresponding to f and G respectively the density and cdf of the standard normal distribution and ω the identity function, which was introduced in the seminal paper Azzalini (1985). For a recent account on skew-symmetric distributions and, in particular, the skew-normal distribution, we refer the reader to the monograph Azzalini and Capitanio (2014). We focus on the study of skew-symmetric models of type (1) where ω is positive on $\mathbb{R}^+$ in order to be able to identify right- and left-skewness with the sign of λ and to obtain general results.

Bayesian inference within these families is a challenge. The prior elicitation for λ is complicated since this parameter not only controls the asymmetry of (1) but also the mode, spread, and tail behaviour. Numerous priors for λ have been proposed in the literature, *inter alia* by Liseo and Loperfido (2006), Cabras et al. (2012), Branco et al. (2013) and Rubio and Liseo (2014). These references focus on the construction of "noninformative priors" from different viewpoints. However there are several situations where we do have *a priori* information on how the data shall behave, and hence at least we know the sign of λ. For instance, when modeling BMI (body-mass index) data, we know the data will be right-skewed for biometric reasons, see *e.g.* Heinz et al. (2003). The same holds true for other biometric indicators and size measurements. Given the popularity of skew-symmetric distributions it is thus of paramount importance to construct informative priors for λ that reflect our *a priori* knowledge of the situation. To our knowledge, only Canale et al. (2016), who proposed the use of normal and skew-normal priors for λ in the skew-normal model, have studied informative priors. Their main motivation for using these kinds of priors is that they facilitate sampling from the corresponding posterior distribution. Their skew-normal prior requires the specification of 3 hyperparameters, which is a difficult task given the interdependence of the parameters. The prior elicitation proposed by Canale et al. (2016) also relies in the interpretation of λ as a parameter regulating solely the asymmetry, hence ignores its overall effect on the density.

In the present paper we tackle the problem of constructing priors for λ by interpreting it as a *perturbation parameter* turning the initial symmetric density f into a skew-symmetric density

of the form (1). Indeed f is modified by multiplication with a "skewing function" $2G(\lambda \cdot)$, which is also referred to as "modulation of symmetry" (see, e.g., Azzalini and Capitanio, 2014). This perturbation effect becomes obvious when we consider $\lambda = 0$: only then do we retrieve the initial (symmetric) density f, while any non-zero value of λ induces a perturbation. Viewing λ as perturbation parameter actually reflects its very nature as foreseen by Fernando de Helguero (1880–1908), the early pioneer of skew-symmetric distributions. Quoting him "*But it may happen, and indeed this must often take place, that other perturbation causes join in [...] The curve will be abnormal, asymmetrical*"[1].

With this interpretation of λ as perturbation parameter it is appealing to invoke its perturbation capacity as a principle on which to construct prior distributions. In Section 2 we shall therefore measure this effect of λ by calculating the Total Variation distance between f and its skew-symmetric counterpart (1). Rather than putting a prior on the parameter λ, whose values are difficult to interpret, we shall put a prior on this interpretable distance. We opt in Section 3 to assign Beta distributions on the range of values taken by this distance. This allows us, by varying the choice of the Beta hyperparameters, to build informative as well as noninformative priors, which moreover enjoy a clear interpretability. In Section 4, we first compare the performance of our priors to existing priors by means of a Monte Carlo simulation study, and then we illustrate their usefulness by analyzing a data set. Finally some proofs are provided in the Appendix. The present paper is complemented by an online Supplementary Material containing further details on the simulation study and a short application of our methodology to other distributions containing a shape parameter.

2   Measuring the perturbation within skew-symmetric families

There exist several distinct measures for the distance between two distributions. Those are called probability distances (or metrics, if the distance happens to be a true metric, see Gibbs and Su, 2002). Our choice in the present paper for the Total Variation metric has been driven by the fact that this distance allows precisely to measure mass relocation when passing from f to $s_{f;G}$ for a given value of the parameter λ. Moreover, contrary to other distances such as the Hellinger distance or Kullback-Leibler divergence, the Total Variation distance seems tailor-made for the problem at hand as it gives rise to simple expressions which is mostly not the case for other distances but is obviously crucial for our goal of building a prior for λ.

The Total Variation distance between two probability measures $\mu(\cdot)$ and $\nu(\cdot)$ on $\mathbb{R}$ is defined

---

[1]This is a passage from de Helguero (1909) translated to English in Azzalini and Regoli (2012).

as

$$d_{TV}(\mu, \nu) = \sup_{A \subset \mathbb{R}} |\mu(A) - \nu(A)|,$$

explaining why this distance represents the largest possible difference between the probability assigned to the same event by two such measures. One easily sees that $0 \leq d_{TV}(\mu, \nu) \leq 1$. If the probability measures admit Radon-Nikodym derivatives f_1 and f_2, supported on the interval $\mathbb{R}$, then the definition becomes

$$d_{TV}(f_1, f_2) = \frac{1}{2} \int_{\mathbb{R}} |f_1(x) - f_2(x)| dx.$$

Using this expression, the Total Variation distance between the baseline symmetric density f and its skew-symmetric counterpart $s_{f;G}$ from (1), for fixed $\lambda \in \mathbb{R}$, can be written as

$$d_{TV}(f, s_{f;G}|\lambda) = \frac{1}{2} \int_{\mathbb{R}} |2G(\lambda \omega(x)) - 1| f(x) dx.$$

The symmetry of G implies that $d_{TV}(f, s_{f;G}|\lambda) = d_{TV}(f, s_{f;G}| - \lambda)$, hence this distance is not a one-to-one function of the parameter λ. This suggests using as measure of perturbation the quantity

$$M_{TV}(\lambda) = \text{sign}(\lambda) d_{TV}(f, s_{f;G}|\lambda), \tag{2}$$

which enjoys some appealing properties. First, for f and G fixed, $M_{TV}(0) = 0$, which corresponds to the case $s_{f;G} = f$. Since $\lambda \mapsto M_{TV}(\lambda)$ is monotone increasing (see equation (3) below), the largest difference is obtained for $\lambda \to \pm\infty$, when $s_{f;G}$ converges to the positive/negative half-f. This largest difference equals $\pm 1/2$, hence $M_{TV}(\lambda) \in (-1/2, 1/2)$. Given that we only consider the case when f and $s_{f;G}$ have the same location and scale parameters, it follows that this measure is also invariant under affine transformations. By construction, we have that $M_{TV}(\lambda) = -M_{TV}(-\lambda)$. These properties resemble the desirable conditions P.1–P.3 discussed in Arnold and Groeneveld (1995) for a measure of skewness. However, the condition P.4 (convexity ordering) in Arnold and Groeneveld (1995) is not satisfied in general since we are measuring the perturbation effect of λ only with respect to the symmetric baseline density f. Thus, M_{TV} should not be interpreted as a measure of skewness, but well as a function that quantifies the overall perturbation effect of the parameter λ.

By using the symmetry properties of f and G, we can re-express (2) as

$$M_{TV}(\lambda) = \frac{1 - 2S_{f;G}(0; \lambda)}{2}, \tag{3}$$

where $S_{f;G}$ is the cdf associated with $s_{f;G}$. This expression reveals that, for a fixed choice of f and G, M_{TV} is simply a re-scaling of the difference between the mass cumulated on either side

4

of 0 by the distribution $S_{f;G}$ (since $1 - 2S_{f;G}(0;\lambda) = \{1 - S_{f;G}(0;\lambda)\} - S_{f;G}(0;\lambda)$). Therefore $M_{TV}(\lambda)$ measures the effect of the parameter λ in terms of the relocation of mass on either side of the symmetry center of f.

Example 1 For the skew-normal density we use the standard normal probability density function (pdf) ϕ and cdf Φ for f and G in (1), respectively, and $\omega(x) = x$, and obtain from Godoi et al. (2017) and (2) the representation

$$M_{TV}(\lambda) = \frac{\mathrm{ArcTan}(\lambda)}{\pi}, \tag{4}$$

for the perturbation measure M_{TV}. For the skew-Laplace density (obtained when f and G are the Laplace pdf and cdf, respectively, and $\omega(x) = x$) we have

$$M_{TV}(\lambda) = \frac{1}{2}\frac{\lambda}{1 + |\lambda|}.$$

Finally, let t_ν and T_ν denote the pdf and cdf of the Student t distribution with $\nu > 0$ degrees of freedom, respectively. The density of the skew-t distribution with ν degrees of freedom proposed by Azzalini and Capitanio (2003) is given by

$$\frac{2}{\sigma}t_\nu\left(\frac{x - \mu}{\sigma}\right)T_{\nu+1}\left(\lambda(x - \mu)\sqrt{\frac{\nu+1}{\nu\sigma^2+(x-\mu)^2}}\right), \quad x \in \mathbb{R}.$$

This distribution is a special case of the class of densities defined in (1). In the Appendix, we show that its perturbation measure M_{TV} is also given by (4) and therefore coincides with the corresponding measure for the skew-normal distribution (which is a special case of the skew-t when $\nu \to \infty$).

3 Proposed priors

The proposed perturbation measure $M_{TV}(\lambda)$ allows us to build informative as well as non-informative priors for the perturbation parameter λ in skew-symmetric models. Recall that M_{TV} varies in $\left(-\frac{1}{2}, \frac{1}{2}\right)$ and is an injective function of λ. Consequently any probability distribution on $\left(-\frac{1}{2}, \frac{1}{2}\right)$ as prior choice for M_{TV} induces a proper prior on λ. For these distributions we choose the very versatile beta distribution with density

$$\frac{1}{B(\alpha, \beta)}\left(u + \frac{1}{2}\right)^{\alpha-1}\left(\frac{1}{2} - u\right)^{\beta-1}, \quad u \in \left(-\frac{1}{2}, \frac{1}{2}\right),$$

where $B(\alpha, \beta)$ represents the beta function and $\alpha, \beta > 0$. We refer to this class of priors as the Beta Total Variation priors $BTV(\alpha, \beta)$ with hyperparameters $\alpha, \beta > 0$. Of course, any other distribution with support $\left(-\frac{1}{2}, \frac{1}{2}\right)$ can be employed instead of the beta distribution, however, this choice facilitates some aspects of our study thanks to its flexibility and interpretability.

Our way of proceeding leads to tractable and interpretable priors. In the previous section we have seen that the perturbation effect of λ basically consists of a mass relocation (3). This mass relocation affects the shape of the density differently in distinct skew-symmetric models (1). Figure 1 shows the density shapes obtained for several values of the percentage of total relocated mass $(M_{TV}(\lambda))$ in the skew-normal, skew-logistic, and skew-Laplace cases. Thus, the interpretation of the perturbation function $M_{TV}(\lambda)$ together with the aid of visualizing the shape of the density for different values of $M_{TV}(\lambda)$ can be used to translate prior believes about the shape of the density into a prior distribution as follows.

- **Informative priors**. If, *a priori*, we favour right/left asymmetry and hence need informative priors, we choose the hyperparameters α and β in such a way that the beta prior assigns mass to the appropriate range of values of M_{TV}.
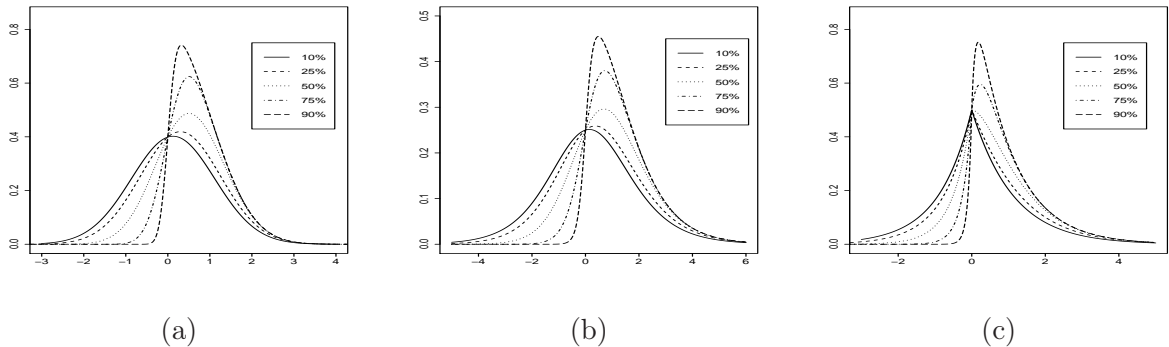

Figure 1: Shapes of the density for $\mu = 0$, $\sigma = 1$ and the percentage of total relocated mass equal to $10\%, 25\%, 50\%, 75\%, 90\%$: (a) skew-normal; (b) skew-logistic; and (c) skew-Laplace.

- **Non-informative priors**. For those cases where there is no reliable prior information about the asymmetry of the data, we explore the use of two types of noninformative priors, obtained for (i) $\alpha = \beta = 1$, the uniform distribution, which gives equal probability mass to any pair of subintervals of $[0, 1]$ of equal length, and (ii) $\alpha = \beta = 1/2$, corresponding to a U-shape beta density. The second choice is motivated as follows. By assigning a $Beta(\alpha, \beta)$ prior to an interpretable measure of perturbation, we implicitly associate a probability p with values that produce right-skewed distributions, and a probability $1 - p$ with values that produce left-skewed distributions. We can interpret this scenario as a Bernoulli trial with parameter p. A noninformative prior that has been widely studied for the parameter p of the Bernoulli distribution is the Jeffreys prior, which is precisely the $Beta(1/2, 1/2)$ prior. This is, the idea is to assign noninformative (or vaguely informative) priors to an interpretable function of the shape parameter λ. This strategy has been discussed in a

more general framework in Seaman III et al. (2012).

In the remainder of this section, we shall first describe and investigate the resulting $BTV(\alpha, \beta)$ priors for the location-scale-free densities $2f(x)G(\lambda\omega(x))$ (Section 3.1), and then discuss joint location-scale-skewness priors for the skew-symmetric models of interest (1) (Section 3.3). A simple remark on the invariance of these sorts of priors is presented below.

Remark 1 The $BTV(\alpha, \beta)$ priors are invariant under one-to-one transformations of λ. This implies that the BTV priors associated to a reparameterization $\alpha = h(\lambda)$, where $h : \mathbb{R} \to D \subset \mathbb{R}$ is a diffeomorphism, can be derived from the corresponding priors on λ using a change of variable.

3.1 Beta-TV priors

Putting a $Beta(\alpha, \beta)$ prior on $M_{TV}(\lambda)$ induces a prior on the parameter λ with pdf

$$\pi_{TV}(\lambda|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \left(M_{TV}(\lambda) + \frac{1}{2} \right)^{\alpha-1} \left(\frac{1}{2} - M_{TV}(\lambda) \right)^{\beta-1} \frac{d}{d\lambda} M_{TV}(\lambda). \tag{5}$$

In order to analyze the general priors $BTV(\alpha, \beta)$, we first investigate some properties of the simpler $BTV(1, 1)$ prior which reduces to $\pi_{TV}(\lambda|1, 1) = \frac{d}{d\lambda} M_{TV}(\lambda)$. Sufficient conditions for the well-definiteness of this prior are stated in the following result.

Lemma 1 *Consider the class of skew-symmetric densities of the type* (1). *If g is a bounded pdf and $\int_0^\infty \omega(x)f(x)dx < \infty$, the $BTV(1, 1)$ prior is well-defined for all λ and given by*

$$\pi_{TV}(\lambda|1, 1) = 2 \int_0^\infty \omega(x)f(x)g(\lambda\omega(x))dx. \tag{6}$$

In the following we provide some general properties of the prior (6), including a characterization of its tails in the important case $\omega(x) = x$.

Theorem 2 *Consider the class of skew-symmetric densities of the type* (1), *where g is a bounded pdf and $\int_0^\infty \omega(x)f(x)dx < \infty$. Then, the prior* (6) *has the following properties:*

(i) *$\pi_{TV}(\lambda|1, 1)$ is symmetric about $\lambda = 0$.*

(ii) *If g is unimodal, then $\pi_{TV}(\lambda|1, 1)$ is decreasing in $|\lambda|$.*

(iii) *For $\omega(x) = x$, and under the assumptions that f is unimodal, $f(0) = M < \infty$ and $\int_0^\infty xg(x)dx < \infty$, the tails of $\pi_{TV}(\lambda|1, 1)$ are of order $O(|\lambda|^{-2})$.*

Example 2 Using expression (6) with $\omega(x) = x$ we obtain $\pi_{TV}(\lambda|1, 1) = 1/\left(\pi \left(1 + \lambda^2 \right) \right)$ as $BTV(1, 1)$ prior for the skew-normal and skew-t distributions, and $\pi_{TV}(\lambda|1, 1) = 1/(2 \left(1 + |\lambda| \right)^2)$ for the skew-Laplace distribution.

Thanks to (5), any $BTV(\alpha, \beta)$ prior possesses a nice closed-form expression whenever the $BTV(1,1)$ prior does. The following result describes the tail behaviour of the density $\pi_{TV}(\lambda|\alpha, \beta)$ of the $BTV(\alpha, \beta)$ prior and is a consequence of Theorem 2 and the tail behaviour of Beta-transformations of symmetric distributions, see Section 4.5 of Jones (2004).

Corollary 3 *Consider the skew-symmetric densities defined by (1) for $\omega(x) = x$, together with the assumptions of Theorem 2(iii). The right tail of $\pi_{TV}(\lambda|\alpha, \beta)$ is of order $O(|\lambda|^{-\beta-1})$, while its left tail is of order $O(|\lambda|^{-\alpha-1})$. Moreover, if $\alpha = \beta$, then $\pi_{TV}(\lambda|\alpha, \beta)$ is symmetric.*

In particular, for the $BTV(1/2, 1/2)$ prior we obtain the following expression:

$$\pi_{TV}(\lambda|1/2, 1/2) = \frac{1}{\pi\sqrt{\frac{1}{4} - M_{TV}^2(\lambda)}}\pi_{TV}(\lambda|1,1). \tag{7}$$

This prior is symmetric and, for skew-symmetric models with $\omega(x) = x$, its tails are of order $O(|\lambda|^{-3/2})$, which interestingly coincide with those of the Jeffreys prior (Rubio and Liseo, 2014). However, the prior $\pi_{TV}(\lambda|1/2, 1/2)$ and the Jeffreys prior are not identical. In fact, the Jeffreys prior has no closed-form expression, and moreover it can be ill-defined for certain combinations of f and G due to singularities in the Fisher information matrix in the neighborhood of $\lambda = 0$, see Hallin and Ley (2012).

3.2 Heuristic approximations to the $BTV(1,1)$ priors

In general the expression (6) is not available in closed-form. However, we can appeal to the characterization of the tail behaviour of these priors in Theorem 2 to come up with tractable approximations. For example, in the case when $\omega(x) = x$ and f and G are the logistic pdf and cdf, respectively, the $BTV(1,1)$ prior is not available in closed-form but can be reasonably well approximated with a Student-t distribution with 1 degree of freedom and scale parameter 0.92. Figure 2 shows the quality of this approximation. The quality of Student-t approximations for $BTV(1,1)$ priors associated to other skew-symmetric models seems to require a case by case analysis.

3.3 Location-scale-skewness models: partial information priors

Consider now the initial densities of interest (1), which contain unknown location and scale parameters. For this model we adopt the prior structure

$$\pi(\mu, \sigma, \lambda) = \frac{p(\lambda)}{\sigma}, \tag{8}$$

where $p(\lambda)$ is a proper prior on λ, here $BTV(\alpha, \beta)$. This prior structure can be justified as a sort of *partial information prior* (Sun and Berger, 1998) in the sense that we are using the
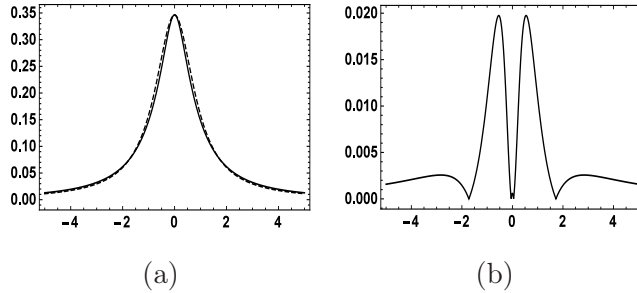
Figure 2: (a) Total variation prior of λ (continuous line) and Student-t approximation (dashed line); (b) Absolute difference between the Total variation prior of λ and the Student-t approximation.

reference prior for the location and scale parameters, $\pi(\mu, \sigma) \propto \sigma^{-1}$, while we allow for using a subjective prior on the perturbation parameter λ. Such structures can also be motivated as priors inspired by the form of the independence Jeffreys prior (Rubio and Steel, 2014, 2015). Theorem 4 below presents sufficient conditions for the propriety of the posterior distribution under the prior structure (8). We restrict our study to the cases when f belongs to scale mixtures of normals. This is a wide family of symmetric distributions which contains many models of practical interest such as the normal, logistic, Laplace, symmetric hyperbolic, Student-t, among many other distributions.

Theorem 4 *Let* $\mathbf{x} = (x_1, \ldots, x_n)$ *be an i.i.d. sample from a skew-symmetric model (1). Suppose that f is a scale mixture of normals. Then the posterior distribution of (μ, σ, λ) associated with the prior structure (8) is proper if $n \geq 2$ and if all the observations are different.*

This theorem, proved in the Appendix, guarantees that the priors proposed in the present paper for skew-symmetric densities lead almost surely to proper posterior distributions.

4 Finite sample properties and practical performance

4.1 Monte Carlo simulation study

Noninformative priors

In order to compare the performance of the priors proposed in Section 3 with that of the Jeffreys prior (Liseo and Loperfido, 2006, Rubio and Liseo, 2014), we have conducted a thorough simulation study, of which we only present certain results here, the others being provided in the Supplementary Material. We have generated $N = 1,000$ samples of size $n = 50$ from the skew-normal, skew-logistic and skew-Laplace distributions with location parameter $\mu = 0$, scale

parameter $\sigma = 1$, and perturbation parameter $\lambda = 0, 2.5, 5$. Results for the sample sizes $n = 100$ and $n = 200$ can be found in the Supplementary Material. For each of these samples, we simulate a posterior sample of size $1,000$ from (μ, σ, λ) using the $BTV(1,1)$, $BTV(1/2, 1/2)$ and Jeffreys priors. We employ a self-adaptive MCMC sampler (Christen and Fox, 2010) to obtain the posterior samples. For each posterior sample, we calculate the coverage proportions of the 95% credible intervals of each parameter (that is, the proportion of credible intervals that contain the true value of the parameter) as well as the 5%, 50% and 95% quantiles of the posterior medians and maximum *a posteriori* (MAP) estimators. In addition, we obtain the median of the Bayes factors (BFs) associated to the hypothesis $H_0 : \lambda = 0$. The Bayes factors are approximated using the Savage-Dickey density ratio.

The BTV priors for the skew-normal and skew-Laplace models enjoy nice closed-form expressions. For the skew-logistic model, we employ the Student-t approximation for the $BTV(1,1)$ prior described in Section 3.2, the $BTV(1/2, 1/2)$ prior following then immediately from (7). In order to implement the respective Jeffreys priors, we need to work with approximations. In the skew-normal and skew-logistic cases, we respectively use the Student-t approximation proposed in Bayes and Branco (2007) (1/2 degrees of freedom and scale $\pi/2$) and the Student-t approximation proposed in Rubio and Liseo (2014) (1/2 degrees of freedom and scale $4/3$). For the skew-Laplace model, we propose a new approximation to the Jeffreys prior:

$$\pi_J(\lambda) = \frac{1}{4s_0(1 + |\lambda/s_0|)^{3/2}},$$

where $s_0 = 0.77$.

The results are reported in Tables 1, 2 and 3. Overall, we observe that the $BTV(1/2, 1/2)$ and Jeffreys priors exhibit the best, and very similar, performance. This is an expected result as the Jeffreys prior is first-order or second-order probability matching under mild regularity conditions (Ghosh, 2011). However, we emphasize that the $BTV(1/2, 1/2)$ prior is more tractable than the Jeffreys prior and it is well-defined under less restrictive conditions. These conclusions are further supported by the simulation studies of the Supplementary Material.

Informative priors

We now explore the use of the proposed informative priors. We simulate $N = 1,000$ samples of size $n = 50$ from a skew-normal distribution with parameters $\mu = 0$, $\sigma = 1$ and $\lambda = 5$. We employ again a self-adaptive MCMC sampler to obtain the posterior samples. For each of these samples, we simulate a posterior sample of size $1,000$ from (μ, σ, λ) using the $BTV(3, 1/2)$ prior. This prior assigns 5% of the mass to values of $\lambda < 0$ while being vaguely informative about $\lambda > 0$. We also consider the skew-normal prior proposed in Canale et al. (2016) with

Prior	MAP			Median			Coverage	BF
	5%	50%	95%	5%	50%	95%		
$\lambda = 0$								
BTV(1/2,1/2)								
μ	-1.157	-0.021	1.169	-0.820	-0.005	0.922	0.990	–
σ	0.904	1.103	1.493	0.992	1.195	1.515	0.874	–
λ	-1.597	0.011	1.453	-1.769	0.004	1.484	0.990	1.715
Jeffreys								
μ	-1.170	-0.076	1.229	-0.871	-0.018	1.008	0.983	–
σ	0.923	1.115	1.541	0.999	1.216	1.528	0.858	–
λ	-1.854	0.015	1.663	-1.876	0.017	1.589	0.986	1.824
BTV(1,1)								
μ	-1.059	0.004	1.089	-0.647	-0.007	0.731	0.997	–
σ	0.897	1.081	1.344	0.974	1.163	1.412	0.892	–
λ	-0.712	0.003	0.552	-1.158	-0.011	0.938	0.996	1.245
$\lambda = 2.5$								
BTV(1/2,1/2)								
μ	-0.281	0.039	0.921	-0.224	0.189	0.821	0.899	–
σ	0.610	0.832	1.220	0.667	0.880	1.202	0.931	–
λ	-0.273	1.033	5.290	-0.103	1.414	7.759	0.869	0.949
Jeffreys								
μ	-0.283	0.036	0.994	-0.233	0.170	0.837	0.897	–
σ	0.614	0.847	1.220	0.674	0.891	1.213	0.936	–
λ	-0.307	1.342	5.964	-0.119	1.560	8.571	0.877	0.988
BTV(1,1)								
μ	-0.225	0.093	0.891	-0.163	0.308	0.815	0.862	–
σ	0.602	0.782	1.171	0.647	0.845	1.147	0.917	–
λ	-0.163	0.415	4.094	-0.076	1.032	5.345	0.843	0.797
$\lambda = 5$								
BTV(1/2,1/2)								
μ	-0.174	-0.004	0.341	-0.157	0.026	0.576	0.918	–
σ	0.594	0.958	1.197	0.662	0.960	1.193	0.926	–
λ	-15.792	3.132	30.601	0.557	4.759	31.230	0.891	0.140
Jeffreys								
μ	-0.180	-0.007	0.318	-0.153	0.019	0.552	0.919	–
σ	0.595	0.958	1.200	0.666	0.963	1.199	0.925	–
λ	-7.616	3.265	38.095	0.609	4.849	32.032	0.896	0.136
BTV(1,1)								
μ	-0.141	0.028	0.623	-0.114	0.072	0.642	0.895	–
σ	0.581	0.918	1.153	0.639	0.921	1.152	0.909	–
λ	-0.010	2.921	8.071	0.344	3.595	11.755	0.874	0.164

Table 1: Skew-normal data for noninformative priors: $\mu = 0, \sigma = 1, n = 50$.

Prior	MAP			Median			Coverage	BF
	5%	50%	95%	5%	50%	95%		
$\lambda = 0$								
BTV(1/2,1/2)								
μ	-1.513	-0.006	1.624	-1.260	0.042	1.368	0.964	–
σ	0.871	1.080	1.350	0.922	1.148	1.432	0.905	–
λ	-0.771	0.005	0.753	-1.278	-0.027	1.099	0.972	2.049
Jeffreys								
μ	-1.524	0.017	1.580	-1.279	0.046	1.384	0.966	–
σ	0.871	1.089	1.372	0.927	1.153	1.433	0.901	–
λ	-0.862	-0.002	0.802	-1.327	-0.026	1.107	0.967	2.138
BTV(1,1)								
μ	-1.378	0.030	1.397	-1.163	0.031	1.187	0.980	–
σ	0.857	1.065	1.307	0.913	1.126	1.379	0.923	–
λ	-0.675	0.002	0.629	-1.009	-0.022	0.920	0.983	1.436
$\lambda = 2.5$								
BTV(1/2,1/2)								
μ	-0.459	0.080	0.810	-0.390	0.175	0.924	0.913	–
σ	0.617	0.871	1.252	0.650	0.909	1.266	0.923	–
λ	0.022	1.172	7.476	0.314	1.809	11.010	0.912	0.507
Jeffreys								
μ	-0.476	0.065	0.815	-0.378	0.162	0.935	0.912	–
σ	0.624	0.875	1.269	0.658	0.917	1.266	0.919	–
λ	0.073	1.321	7.835	0.316	1.894	9.999	0.905	0.518
BTV(1,1)								
μ	-0.357	0.167	0.859	-0.286	0.276	0.989	0.899	–
σ	0.606	0.835	1.204	0.641	0.875	1.196	0.899	–
λ	0.209	0.937	4.343	0.252	1.439	5.855	0.879	0.463
$\lambda = 5$								
BTV(1/2,1/2)								
μ	-0.316	0.011	0.524	-0.263	0.049	0.621	0.921	–
σ	0.614	0.938	1.230	0.656	0.955	1.236	0.919	–
λ	-7.980	3.013	33.228	0.995	4.600	33.555	0.915	0.111
Jeffreys								
μ	-0.307	0.010	0.486	-0.262	0.052	0.600	0.921	–
σ	0.618	0.938	1.230	0.662	0.958	1.242	0.919	–
λ	-10.231	3.015	33.053	1.099	4.535	33.336	0.906	0.115
BTV(1,1)								
μ	-0.238	0.060	0.632	-0.204	0.124	0.729	0.897	–
σ	0.594	0.894	1.181	0.640	0.913	1.185	0.902	–
λ	0.543	2.585	8.449	0.826	3.483	13.135	0.894	0.109

Table 2: Skew-logistic data for noninformative priors: $\mu = 0, \sigma = 1, n = 50$.

Prior	MAP			Median			Coverage	BF
	5%	50%	95%	5%	50%	95%		
$\lambda = 0$								
BTV(1/2,1/2)								
μ	-0.585	0.000	0.587	-0.524	-0.005	0.562	0.946	–
σ	0.803	1.032	1.293	0.835	1.067	1.332	0.960	–
λ	-0.369	0.001	0.385	-0.621	0.001	0.582	0.946	3.921
Jeffreys								
μ	-0.591	-0.007	0.629	-0.523	-0.009	0.575	0.948	–
σ	0.795	1.034	1.290	0.835	1.072	1.333	0.959	–
λ	-0.415	-0.001	0.388	-0.654	-0.000	0.574	0.948	3.759
BTV(1,1)								
μ	-0.549	-0.003	0.554	-0.509	-0.011	0.545	0.957	–
σ	0.798	1.031	1.286	0.824	1.065	1.318	0.958	–
λ	-0.347	-0.001	0.329	-0.546	-0.002	0.529	0.954	2.567
$\lambda = 2.5$								
BTV(1/2,1/2)								
μ	-0.273	0.014	0.381	-0.230	0.035	0.375	0.932	–
σ	0.661	0.939	1.267	0.706	0.972	1.299	0.936	–
λ	0.280	1.558	6.824	0.697	2.234	10.721	0.926	0.137
Jeffreys								
μ	-0.272	0.016	0.416	-0.239	0.038	0.385	0.930	–
σ	0.668	0.933	1.266	0.710	0.967	1.296	0.932	–
λ	0.270	1.493	6.402	0.684	2.194	11.039	0.922	0.136
BTV(1,1)								
μ	-0.229	0.041	0.421	-0.192	0.063	0.402	0.931	–
σ	0.656	0.917	1.230	0.694	0.949	1.254	0.933	–
λ	0.338	1.438	4.853	0.614	1.963	7.188	0.936	0.106
$\lambda = 5$								
BTV(1/2,1/2)								
μ	-0.187	0.000	0.244	-0.164	0.015	0.262	0.931	–
σ	0.667	0.942	1.234	0.697	0.967	1.263	0.932	–
λ	-3.128	3.112	30.605	1.445	4.790	34.839	0.922	0.054
Jeffreys								
μ	-0.182	0.002	0.257	-0.166	0.019	0.264	0.929	–
σ	0.668	0.942	1.246	0.701	0.971	1.267	0.933	–
λ	-6.021	3.107	30.692	1.424	4.824	34.424	0.926	0.053
BTV(1,1)								
μ	-0.157	0.021	0.291	-0.124	0.041	0.296	0.935	–
σ	0.658	0.914	1.211	0.688	0.944	1.237	0.933	–
λ	0.775	2.897	9.682	1.281	4.018	14.715	0.927	0.041

Table 3: Skew-Laplace data for noninformative priors: $\mu = 0, \sigma = 1, n = 50$.

hyperparameters $(\mu_0, \sigma_0, \lambda_0) = (0, 1, 6.5)$. This prior also assigns 5% of the mass to values of $\lambda < 0$ and is vaguely informative about $\lambda > 0$, however, it has lighter tails than the BTV prior. We calculate the coverage proportions of the 95% credible intervals of each parameter as well as the 5%, 50% and 95% quantiles of the posterior medians and MAP estimators. Results are reported in Table 4. We observe that the $BTV(3, 1/2)$ prior exhibits better frequentist properties than its competitor. This, together with the intuitive nature of our priors, underlines the strength of our new approach.

Prior	MAP			Median			Coverage
	5%	50%	95%	5%	50%	95%	
$\lambda = 5$							
BTV(3,1/2)							
μ	-0.185	-0.020	0.241	-0.161	0.001	0.309	0.935
σ	0.680	0.976	1.200	0.722	0.986	1.210	0.945
λ	-22.182	3.348	40.942	1.546	5.269	36.241	0.919
SN(0,2.5,6.5)							
μ	-0.054	0.078	0.247	-0.045	0.093	0.311	0.890
σ	0.692	0.902	1.087	0.703	0.904	1.084	0.908
λ	1.118	2.962	4.147	1.494	3.091	4.204	0.804

Table 4: Skew-normal data for informative priors: $\mu = 0, \sigma = 1, n = 50$.

4.2 Body Mass Index

The goal of this section is to investigate the Body Mass Index (BMI) of physically active adult men on basis of data provided in Heinz et al. (2003). We analyse three groups from this data set corresponding to individuals with ages in the ranges: $18 - 24$, $25 - 34$, and $35 - 44$, as these represent parts of the population with similar biological and sociological characteristics. The sample sizes are $n = 75, 94$ and 45, respectively. Biometric reasons entail that BMI data are typically asymmetric with a longer right tail, and we have consulted an expert on BMI data for our prior specification (Helena Carreira, LSHTM). Consequently, we fit a skew-normal distribution to these data sets together with the prior (8). In cases with small or moderate sample sizes, the profile likelihood of λ tends to be flat (see Azzalini and Capitanio, 2014). Thus, we expect informative priors to yield better results since the prior distribution has a relevant effect on the shape of the posterior distribution.

For $p(\lambda)$ we use the following priors: (i) the $BTV(1, 1)$ prior (6), (ii) the $BTV(1/2, 1/2)$ prior (7), (iii) the Jeffreys prior, (iv) the informative $BTV(15.6, 4.8)$ prior, (v) the informative skew-normal prior of Canale et al. (2016) with hyperparameters $(\mu_0, \sigma_0, \lambda_0) = (0.5, 1.1, 3.5)$ (which are selected in order to resemble the $BTV(15.6, 4.8)$ prior), and (vi) the matching prior

of Cabras et al. (2012). The informative prior (iv) is selected by choosing the 5% and 95% quantiles of the beta distribution on $M_{TV}(\lambda)$ to be 0.1 and 0.4, according to the expert opinion. This means favouring between 20% and 80% of relocation of mass to the right-hand side of the symmetry point of the baseline distribution. For each of these models, we simulate, using an adaptive MCMC sampler, a posterior sample of size $N = 10,000$ from (μ, σ, λ) (with a burn-in period of $10,000$ iterations and a thinning period of 100 iterations). Table 5 shows a summary of the posterior simulations, maximum likelihood estimator of the parameters and the 95% quantile bootstrap-confidence intervals, and the Bayes factors associated to the hypothesis $H_0 : \lambda = 0$ (obtained using the Savage-Dickey density ratio for priors (i)–(v), and a Laplace approximation for prior (vi)). The posterior inference for μ and σ is similar throughout the different Bayesian models. However, we observe significant differences with respect to λ. The credible intervals obtained with the Jeffreys, $BTV(1,1)$ and $BTV(1/2,1/2)$ priors contain negative values of λ. The reasons are that (i) these priors assign half of the mass on negative values on λ, a feature that is further increased by their heavy tails and the flatness of the profile likelihood of λ, and (ii) the second critical point at $\lambda = 0$ (see Figure 3). The Matching prior induces credible intervals that do not contain the value $\lambda = 0$ due to its bimodality, which pushes the credible intervals away from zero, even in cases where the distribution of the data is close to symmetry (group 35–44). These four noninformative priors include high values of λ due to the combination of their heavy tails and the flatness of the likelihood surface in the direction of λ. On the other hand, the informative priors are centred around values which are coherent with the expert prior knowledge, in particular they only contain positive values of λ. This shows the effectiveness of working with informative priors, and the attractiveness of our new intuitive approach.
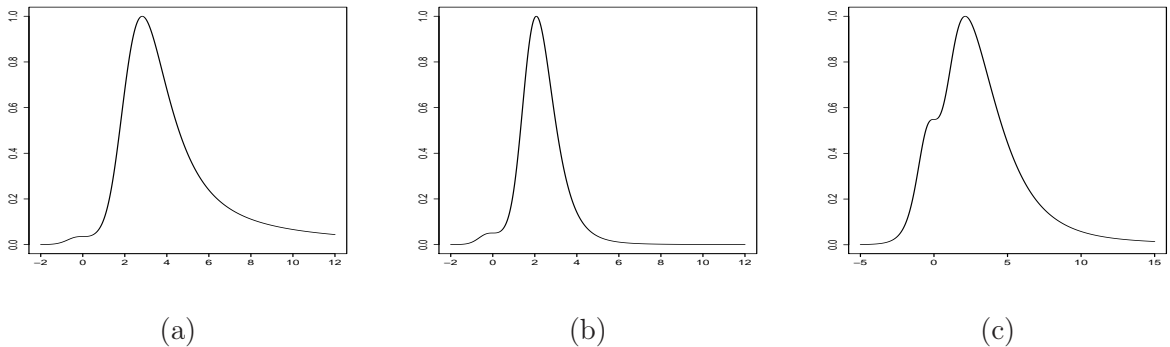

Figure 3:   (a) Profile likelihood of λ (group 18–24); (b) Profile likelihood of λ (group 25–34); (c) Profile likelihood of λ (group 35–44).

Prior	μ	σ	λ	BF
Age group: 18–24				
Jeffreys	21.02 (19.67,23.96)	3.88 (2.52,4.92)	2.57 (-1.11,5.97)	0.31
BTV(1/2,1/2)	21.05 (19.71,24.04)	3.84 (2.55,4.95)	2.48 (-0.85,6.09)	0.27
BTV(1,1)	21.17 (19.82,24.60)	3.75 (2.50,4.87)	2.24 (-0.78,5.69)	0.28
BTV(15.6,4.8)	21.33 (20.20,22.61)	3.61 (2.69,4.56)	1.94 (0.50,3.73)	0.16
SN	21.33 (20.39,22.54)	3.62 (2.79,4.47)	1.94 (0.59,3.23)	0.15
Matching	20.88 (19.73,21.96)	3.99 (2.98,5.04)	2.90 (0.84,7.17)	0.23
MLE	20.94 (19.73,21.87)	3.93 (2.94,4.97)	2.83 (1.47,∞)	–
Age group: 25–34				
Jeffreys	21.73 (20.20,25.02)	4.16 (2.89,5.23)	1.80 (-0.43,3.70)	0.45
BTV(1/2,1/2)	21.73 (20.16,25.00)	4.14 (2.92,5.29)	1.79 (-0.63,3.51)	0.44
BTV(1,1)	21.85 (20.34,25.37)	4.05 (2.86,5.17)	1.64 (-0.62,3.38)	0.36
BTV(15.6,4.8)	21.83 (20.67,23.25)	4.06 (3.19,5.13)	1.64 (0.51,2.98)	0.21
SN	21.80 (20.74,23.18)	4.10 (3.17,5.04)	1.70 (0.63,2.92)	0.23
Matching	21.47 (20.26,22.81)	4.36 (3.31,5.38)	2.11 (0.78,4.13)	0.39
MLE	21.53 (20.53,24.32)	4.29 (2.70,5.37)	2.07 (0.02,3.87)	–
Age group: 35–44				
Jeffreys	24.30 (21.53,28.33)	3.34 (2.25,5.02)	0.74 (-2.07,4.87)	1.54
BTV(1/2,1/2)	24.29 (21.42,28.15)	3.31 (2.29,5.04)	0.75 (-1.79,5.10)	1.38
BTV(1,1)	24.72 (21.71,28.21)	3.20 (2.24,4.77)	0.49 (-1.96,3.87)	1.07
BTV(15.6,4.8)	23.53 (21.94,25.30)	3.48 (2.42,4.69)	1.32 (0.05,3.16)	0.93
SN	23.48 (22.01,25.17)	3.50 (2.50,4.67)	1.36 (0.23,2.88)	0.78
Matching	22.98 (21.22,28.79)	3.96 (2.73,5.61)	2.05 (-2.33,7.00)	2.63
MLE	22.95 (21.58,26.18)	3.84 (2.26,5.10)	2.12 (0,∞)	–

Table 5: BMI data: posterior median, 95% posterior credible intervals, and Bayes factors associated to $\mathcal{H}_0 : \lambda = 0$.

5 Discussion

The construction of meaningful priors, either informative or noninformative, is of central importance in Bayesian inference. Prior elicitation is particularly challenging when the model parameters control several features. Such is the case of the skewness, or rather perturbation, parameter in skew-symmetric distributions. This parameter controls the mode, asymmetry, tail

behaviour, and spread of the pdf. We proposed a new method for constructing priors for this parameter based on its overall effect on the shape of the density. For this purpose, we studied the perturbation effect of the shape parameter through the Total Variation distance. We showed that the priors induced by the Total Variation distance are very intuitive and hence user-friendly, have very good frequentist properties and enjoy tractable expressions, especially compared to the popular Jeffreys prior which moreover can have singularities.

The constructive strategy proposed in this paper can be extended to shape parameters in other distributions. In the Supplementary Material, we provide a brief study on the construction of priors using the Total Variation distance for log-skew-symmetric distributions and two-piece distributions. It is shown that the priors for the entire family of two-piece distributions have closed-form expressions, which are linked to a family of priors proposed in Rubio and Steel (2014). Applying this new strategy of prior construction to various other families with shape parameters represents a promising research direction. Simpson et al. (2017) recently proposed a new strategy to construct prior distributions based on the Kullback-Leibler divergence. The reasoning behind those priors is to penalise model complexity by constructing a prior that favours the baseline nested model, which differs from the approach presented here. The study of Penalised Complexity priors for skew-symmetric models represents a possible research direction. More generally, the use of other distances between the distributions, instead of the Total Variation, is another interesting research direction, although this has to be done with some care as different distances may have very different properties. For instance, a possible competitor is the Wasserstein distance, which is defined for two distributions F_1 and F_2 on $\mathbb{R}$, with finite first moment, by (Vallender, 1974)

$$d_{\mathcal{W}}(F_1, F_2) = \int_{\mathbb{R}} |F_1(x) - F_2(x)| dx.$$

Unfortunately, this distance is not invariant under monotone transformations, a property used in Section 2 to construct a measure of perturbation. Other distances, such as the Energy distance, Kolmogorov distance, and the Hellinger distance, do not seem to lead to closed-form or interpretable expressions, in contrast to our choice.

The priors proposed in this paper can be extended to the family of multivariate skew-symmetric distributions with density function:

$$s_n(\mathbf{x}; \boldsymbol{\lambda}) = 2f_n(\mathbf{x})G(\boldsymbol{\lambda} \cdot \mathbf{x}),$$

where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$, f_n is a symmetric pdf with support on $\mathbb{R}^n$, G is a symmetric cdf with support on $\mathbb{R}$, and $\boldsymbol{\lambda} \cdot \mathbf{x} = \sum_{j=1}^n \lambda_j x_j$. In order to construct a joint

prior for the skewness parameter $\boldsymbol{\lambda}$, we can decompose it into conditional priors:

$$\pi(\boldsymbol{\lambda}) = \pi(\lambda_1|\lambda_2, \dots, \lambda_n)\pi(\lambda_2|\lambda_3, \dots, \lambda_n)\dots\pi(\lambda_n),$$

and sequentially apply the strategy proposed in Section 3 to construct the univariate conditional priors. A more detailed study of this prior will be considered for future research.

Appendix: Proofs

Proof of the representation (4) for skew-t distributions

Let $X_{\lambda,\nu}$ be a random variable following a skew-t distribution. By using the stochastic representation of the skew-t distribution as a scale mixture of skew-normal distributions (Azzalini and Capitanio, 2003) it follows that

$$S_{t_\nu, T_{\nu+1}}(0; \lambda, \nu) \;=\; \mathbb{P}\left(X_{\lambda,\nu} \le 0\right) = \mathbb{P}\left(V_\nu^{-1/2}Z_\lambda \le 0\right) = \mathbb{P}\left(Z_\lambda \le 0\right), \tag{9}$$

where $V_\nu \sim \chi_\nu^2/\nu$, and Z_λ is a skew-normal random variable with location 0, unit scale, and skewness parameter λ. The result follows from this relationship together with equation (3).

We emphasize that not all scale mixtures of skew-normal distributions can be written as a skew-symmetric distribution of the type (1). The skew-t is one of few cases reported in the literature which admit both representations. However, equation (9) reveals a more general property of scale mixtures of skew-normal distributions since $\mathbb{P}\left(V^{-1/2}Z_\lambda \le 0\right) = \mathbb{P}\left(Z_\lambda \le 0\right)$ for any positive random variable $V^{-1/2}$.

Proof of Theorem 2

(i) The symmetry property is immediate from expression (6).

(ii) It is easily seen that $\pi_{TV}(\lambda) = \int_0^\infty 2\omega(u)f(u)g(\lambda\omega(u))du$. For $u > 0$ and $|\lambda_1| > |\lambda_2| > 0$, it follows that $\omega(u)f(u)g(\lambda_1\omega(u)) < \omega(u)f(u)g(\lambda_2\omega(u))$ thanks to the unimodality and symmetry of g. Thus

$$\int_0^\infty \omega(u)f(u)g(\lambda_1\omega(u))du \le \int_0^\infty \omega(u)f(u)g(\lambda_2\omega(u))du$$

and hence the prior is decreasing in $|\lambda|$.

(iii) By using the change of variable $u = \lambda x$ and the maximality of f at 0, it follows that for $\lambda > 0$

$$\int_0^\infty xf(x)g(\lambda x)dx \le M \int_0^\infty xg(\lambda x)dx = \frac{M}{\lambda^2}\int_0^\infty ug(u)du.$$

18

The same results hold true for $\lambda < 0$ by the symmetry property shown in (i). Now, let $|\lambda| \geq L > 0$. Then, the unimodality and symmetry of f yield $f\left(\dfrac{x}{\lambda}\right) \geq f\left(\dfrac{x}{L}\right)$ for $x > 0$. By using the change of variable $u = \lambda x$ we find that for $\lambda > 0$

$$
\begin{aligned}
\int_0^\infty x f(x) g(\lambda x) dx &= \frac{1}{\lambda^2} \int_0^\infty u f\left(\frac{u}{\lambda}\right) g(u) du \\
&\geq \frac{1}{\lambda^2} \int_0^\infty u f\left(\frac{u}{L}\right) g(u) du.
\end{aligned}
$$

Analogously for $\lambda < 0$. The result follows by combining the previous inequalities.

Proof of Theorem 4

Recall that a posterior distribution is proper whenever the marginal distribution $P(x_1, \ldots, x_n) < \infty$ (Fernández and Steel, 1999). Now note that $s_{f;G}(x; \mu, \sigma, \lambda) \leq \dfrac{2}{\sigma} f\left(\dfrac{x - \mu}{\sigma}\right)$, which entails that

$$
\begin{aligned}
P(x_1, \ldots, x_n) &= \int_{\mathbb{R}} \int_{\mathbb{R}_+} \int_{\mathbb{R}} \left[\prod_{j=1}^n s(x_j; \mu, \sigma, \lambda)\right] \frac{p(\lambda)}{\sigma} d\mu d\sigma d\lambda \\
&\leq \int_{\mathbb{R}_+} \int_{\mathbb{R}} \left[\prod_{j=1}^n \frac{2}{\sigma} f\left(\frac{x_j - \mu}{\sigma}\right)\right] \frac{1}{\sigma} d\mu d\sigma \int_{\mathbb{R}} p(\lambda) d\lambda.
\end{aligned}
$$

Given that $p(\lambda)$ is proper, it follows that the posterior distribution of (μ, σ, λ) exists whenever the posterior distribution of (μ, σ) exists for a scale mixture of normals sampling model and the prior $\pi(\mu, \sigma) \propto \sigma^{-1}$. The propriety of the latter, for $n \geq 2$ and when all the observations are different, follows by Theorem 1 of Fernández and Steel (1999).

Acknowledgements

Supplementary Material: Extensions and Monte Carlo simulation study

6   Extension to other distributions

We here briefly show how our method applies to other types of flexible distributions where a (skewness) parameter has a perturbation effect on the original distribution.

Log-skew-symmetric distributions

The proposed priors from our main paper have the same interpretation if they are used for the perturbation parameter in log-skew-symmetric distributions. Recall that a positive random variable Y is said to be distributed according to a log-skew-symmetric distribution if it is distributed according to (10) below. This sort of distributions have been used for modelling environmental, medical, biological, and financial data (see Marchenko and Genton, 2010 and the references therein). The pdf of Y is given by

$$s_l(y; \lambda) = \frac{2}{y} f(\log y) G(\lambda \log y), \quad y > 0. \tag{10}$$

It follows that the TV distance between (10) and the corresponding baseline log-symmetric density $f_l(y) = \frac{1}{y} f(\log y)$ satisfies

$$d_{TV}(s_l, f_l | \lambda) \;\; = \;\; \frac{1}{2} \int_0^\infty |s_l(y; \lambda) - f_l(y)| \, dy = \frac{1}{2} \int_{-\infty}^\infty |s(x; \lambda) - f(x)| dx = d_{TV}(s, f | \lambda).$$

Consequently, the priors proposed in Section 3 of the main paper for the skew-symmetric family coincide with those obtained for the log-skew-symmetric family. It is also clear that one could use any other increasing diffeomorphism from $\mathbb{R}_+$ to $\mathbb{R}$ instead of the logarithmic transformation.

Two-piece distributions

Consider the family of two-piece distributions with the following parameterisation (see Rubio and Steel, 2014 for a general overview):

$$s_{tp}(x; \gamma) = f\left(\frac{x}{1 - \gamma} \right) I(x < 0) + f\left(\frac{x}{1 + \gamma} \right) I(x \geq 0), \quad x \in \mathbb{R}, \tag{11}$$

where $\gamma \in (-1, 1)$, and f is a unimodal symmetric pdf with mode at 0. The parameter γ controls the mass cumulated on either side of the mode ($x = 0$) while preserving the tail behaviour of f. Density (11) is asymmetric for $\gamma \neq 0$ and it reduces to f for $\gamma = 0$. The TV distance between s_{tp} and the baseline pdf f is given by:

$$d_{TV}(s_{tp}, f | \gamma) \;\; = \;\; \frac{1}{2} \int_{-\infty}^0 \left| f\left(\frac{x}{1 - \gamma} \right) - f(x) \right| dx + \frac{1}{2} \int_0^\infty \left| f\left(\frac{x}{1 + \gamma} \right) - f(x) \right| dx = \frac{|\gamma|}{2}.$$

If we define the measure of perturbation $M_{TV}(\gamma) = \gamma/2$, this coincides, up to a proportionality constant, with the AG measure of skewness proposed in Arnold and Groeneveld (1995) (see Rubio and Steel, 2014). Consequently, if we assume that $M_{TV}(\gamma) = \gamma/2 \sim Beta(\alpha, \beta)$ (using the notation in the main paper for the Beta distribution), we obtain the AG-Beta priors proposed in Rubio and Steel (2014) for this family of distributions.

7 Simulation study

This section is a complement to the Monte Carlo simulation study in Section 4.1 of the main paper. It is a performance comparison between noninformative priors built according to our new method and the Jeffreys prior.

We simulate $N = 1,000$ samples of sizes $n = 100$ and $n = 200$ from the skew-normal, skew-logistic and skew-Laplace distributions, in each case with location parameter $\mu = 0$, scale parameter $\sigma = 1$, and skewness parameter $\lambda = 0, 2.5, 5$. For each of these samples, we simulate a posterior sample of size $1,000$ from (μ, σ, λ) using the $BTV(1,1)$, $BTV(1/2, 1/2)$, and Jeffreys priors. We employ a self-adaptive MCMC sampler to obtain the posterior samples. For each posterior sample, we calculate the coverage proportions of the 95% credible intervals of each parameter (that is, the proportion of credible intervals that contain the true value of the parameter) as well as the 5%, 50% and 95% quantiles of the posterior medians and maximum *a posteriori* (MAP) estimators. In addition, we obtain the median of the Bayes factors (BFs) associated to the hypothesis $H_0 : \lambda = 0$. The Bayes factors are approximated using the Savage-Dickey density ratio. Results are reported in Tables 6–11. Overall, we observe that the Jeffreys and $BTV(1/2, 1/2)$ priors exhibit the best, and very similar, performance.

Prior	MAP			Median			Coverage	BF
	5%	50%	95%	5%	50%	95%		
$\lambda = 0$								
BTV(1/2,1/2)								
μ	-1.031	0.018	0.993	-0.797	0.009	0.775	0.983	–
σ	0.943	1.074	1.389	1.000	1.150	1.387	0.861	–
λ	-1.455	0.001	1.481	-1.244	-0.018	1.315	0.984	2.002
Jeffreys								
μ	-1.039	0.028	1.020	-0.864	0.011	0.800	0.982	–
σ	0.946	1.084	1.432	1.008	1.159	1.400	0.839	–
λ	-1.537	-0.021	1.482	-1.392	-0.026	1.401	0.982	2.204
BTV(1,1)								
μ	-0.997	-0.007	0.960	-0.708	0.015	0.683	0.992	–
σ	0.933	1.067	1.355	0.985	1.131	1.346	0.877	–
λ	-1.186	0.008	1.279	-0.992	-0.010	1.075	0.992	1.404
$\lambda = 2.5$								
BTV(1/2,1/2)								
μ	-0.188	0.014	0.646	-0.170	0.058	0.635	0.880	–
σ	0.650	0.955	1.177	0.704	0.947	1.175	0.898	–
λ	-0.029	2.111	4.318	0.212	2.139	5.127	0.879	0.386
Jeffreys								
μ	-0.202	0.009	0.478	-0.174	0.052	0.638	0.890	–
σ	0.651	0.959	1.180	0.708	0.952	1.177	0.919	–
λ	0.003	2.138	4.308	0.274	2.191	5.028	0.885	0.369
BTV(1,1)								
μ	-0.176	0.031	0.739	-0.148	0.088	0.667	0.870	–
σ	0.645	0.931	1.155	0.691	0.922	1.150	0.897	–
λ	-0.069	1.955	3.867	0.151	1.950	4.504	0.859	0.338
$\lambda = 5$								
BTV(1/2,1/2)								
μ	-0.142	-0.006	0.152	-0.120	0.004	0.185	0.922	–
σ	0.806	0.983	1.137	0.814	0.988	1.147	0.946	–
λ	1.724	4.245	10.966	2.289	4.995	14.257	0.931	0.004
Jeffreys								
μ	-0.134	-0.003	0.150	-0.123	0.004	0.184	0.925	–
σ	0.810	0.986	1.147	0.817	0.991	1.147	0.944	–
λ	1.893	4.336	10.033	2.409	5.044	15.041	0.927	0.004
BTV(1,1)								
μ	-0.117	0.011	0.177	-0.103	0.019	0.231	0.922	–
σ	0.788	0.975	1.133	0.780	0.975	1.131	0.935	–
λ	1.767	4.044	8.088	1.945	4.582	10.286	0.932	0.003

Table 6: Skew-normal data for noninformative priors: $\mu = 0, \sigma = 1, n = 100$.

Prior	MAP			Median			Coverage	BF
	5%	50%	95%	5%	50%	95%		
$\lambda = 0$								
BTV(1/2,1/2)								
μ	-0.886	-0.035	0.869	-0.682	-0.009	0.692	0.994	–
σ	0.960	1.054	1.330	1.008	1.108	1.296	0.846	–
λ	-1.286	0.028	1.295	-1.028	-0.004	1.037	0.993	2.306
Jeffreys								
μ	-0.890	0.032	0.890	-0.718	0.006	0.720	0.991	–
σ	0.961	1.058	1.351	1.009	1.115	1.319	0.844	–
λ	-1.335	-0.002	1.297	-1.082	-0.019	1.052	0.991	2.557
BTV(1,1)								
μ	-0.857	0.002	0.860	-0.602	-0.007	0.652	0.995	–
σ	0.952	1.048	1.285	0.999	1.098	1.275	0.862	–
λ	-1.207	-0.001	1.180	-0.992	-0.004	0.851	0.995	1.570
$\lambda = 2.5$								
BTV(1/2,1/2)								
μ	-0.124	0.008	0.211	-0.119	0.021	0.317	0.905	–
σ	0.699	0.981	1.123	0.760	0.980	1.124	0.909	–
λ	1.158	2.283	3.671	0.924	2.362	3.814	0.907	0.012
Jeffreys								
μ	-0.127	0.006	0.202	-0.119	0.019	0.321	0.914	–
σ	0.709	0.984	1.120	0.772	0.982	1.125	0.911	–
λ	1.191	2.319	3.645	0.922	2.369	3.886	0.916	0.006
BTV(1,1)								
μ	-0.123	0.015	0.225	-0.110	0.031	0.401	0.898	–
σ	0.687	0.976	1.123	0.743	0.972	1.117	0.902	–
λ	0.787	2.231	3.541	0.722	2.294	3.666	0.897	0.014
$\lambda = 5$								
BTV(1/2,1/2)								
μ	-0.090	0.000	0.097	-0.082	0.003	0.098	0.938	–
σ	0.874	0.992	1.108	0.874	0.995	1.107	0.935	–
λ	2.948	4.603	7.795	3.129	4.939	8.662	0.923	5×10^{-11}
Jeffreys								
μ	-0.087	-0.001	0.095	-0.081	0.004	0.098	0.940	–
σ	0.873	0.989	1.099	0.874	0.995	1.108	0.937	–
λ	2.922	4.615	7.907	3.134	4.947	8.943	0.926	4×10^{-11}
BTV(1,1)								
μ	-0.080	0.004	0.098	-0.076	0.008	0.106	0.945	–
σ	0.867	0.987	1.096	0.869	0.990	1.101	0.944	–
λ	2.853	4.457	7.475	3.048	4.748	8.275	0.929	5×10^{-11}

Table 7: Skew-normal data for noninformative priors: $\mu = 0, \sigma = 1, n = 200$.

Prior	MAP			Median			Coverage	BF
	5%	50%	95%	5%	50%	95%		
$\lambda = 0$								
BTV(1/2,1/2)								
μ	-1.139	0.048	1.191	-0.994	0.055	1.046	0.957	–
σ	0.906	1.060	1.244	0.936	1.099	1.282	0.911	–
λ	-0.682	-0.014	0.641	-0.842	-0.030	0.820	0.952	2.794
Jeffreys								
μ	-1.190	0.047	1.227	-1.044	0.059	1.097	0.952	–
σ	0.913	1.070	1.254	0.944	1.106	1.295	0.910	–
λ	-0.783	-0.014	0.719	-0.899	-0.044	0.865	0.946	2.913
BTV(1,1)								
μ	-1.030	0.029	1.082	-0.932	0.051	0.965	0.965	–
σ	0.901	1.054	1.230	0.938	1.092	1.270	0.916	–
λ	-0.614	-0.016	0.608	-0.766	-0.035	0.739	0.961	1.901
$\lambda = 2.5$								
BTV(1/2,1/2)								
μ	-0.312	0.037	0.546	-0.278	0.090	0.602	0.922	–
σ	0.706	0.947	1.196	0.734	0.950	1.203	0.923	–
λ	0.540	1.929	4.478	0.766	2.141	5.002	0.904	0.122
Jeffreys								
μ	-0.305	0.031	0.535	-0.277	0.089	0.604	0.925	–
σ	0.704	0.949	1.203	0.737	0.955	1.204	0.921	–
λ	0.565	1.951	4.205	0.742	2.137	5.028	0.911	0.115
BTV(1,1)								
μ	-0.264	0.074	0.642	-0.237	0.134	0.676	0.918	–
σ	0.685	0.917	1.174	0.718	0.926	1.180	0.913	–
λ	0.496	1.725	3.919	0.675	1.917	4.491	0.894	0.099
$\lambda = 5$								
BTV(1/2,1/2)								
μ	-0.222	-0.003	0.260	-0.198	0.018	0.294	0.931	–
σ	0.780	0.973	1.168	0.792	0.980	1.175	0.942	–
λ	1.649	4.077	10.223	2.174	4.733	13.117	0.916	0.006
Jeffreys								
μ	-0.210	0.001	0.264	-0.200	0.015	0.287	0.933	–
σ	0.788	0.975	1.168	0.796	0.981	1.172	0.943	–
λ	1.507	4.109	9.910	2.178	4.766	13.401	0.907	0.006
BTV(1,1)								
μ	-0.182	0.020	0.303	-0.170	0.040	0.330	0.930	–
σ	0.761	0.961	1.156	0.774	0.967	1.155	0.932	–
λ	1.601	3.792	7.933	1.964	4.338	10.000	0.909	0.005

Table 8: Skew-logistic data for noninformative priors: $\mu = 0, \sigma = 1, n = 100$.

Prior	MAP			Median			Coverage	BF
	5%	50%	95%	5%	50%	95%		
$\lambda = 0$								
BTV(1/2,1/2)								
μ	-0.907	-0.022	0.904	-0.824	-0.010	0.807	0.938	–
σ	0.931	1.037	1.168	0.950	1.059	1.187	0.913	–
λ	-0.529	-0.002	0.540	-0.594	-0.002	0.632	0.938	3.838
Jeffreys								
μ	-0.913	-0.012	0.902	-0.826	-0.007	0.820	0.936	–
σ	0.932	1.038	1.165	0.949	1.061	1.191	0.915	–
λ	-0.531	0.011	0.566	-0.606	0.002	0.662	0.931	4.186
BTV(1,1)								
μ	-0.879	-0.013	0.837	-0.780	-0.012	0.755	0.945	–
σ	0.926	1.035	1.154	0.946	1.058	1.181	0.921	–
λ	-0.481	0.007	0.542	-0.531	0.003	0.618	0.942	2.574
$\lambda = 2.5$								
BTV(1/2,1/2)								
μ	-0.225	0.011	0.363	-0.207	0.037	0.382	0.930	–
σ	0.787	0.975	1.142	0.804	0.975	1.142	0.930	–
λ	1.069	2.233	3.760	1.187	2.330	3.999	0.931	0.001
Jeffreys								
μ	-0.228	0.014	0.334	-0.211	0.034	0.392	0.924	–
σ	0.790	0.974	1.143	0.804	0.977	1.145	0.926	–
λ	1.042	2.253	3.745	1.188	2.349	3.936	0.927	0.001
BTV(1,1)								
μ	-0.207	0.029	0.370	-0.194	0.052	0.427	0.918	–
σ	0.777	0.966	1.128	0.790	0.968	1.128	0.918	–
λ	0.954	2.174	3.617	1.088	2.236	3.800	0.911	0.002
$\lambda = 5$								
BTV(1/2,1/2)								
μ	-0.141	0.003	0.157	-0.132	0.009	0.172	0.934	–
σ	0.859	0.989	1.125	0.859	0.992	1.128	0.940	–
λ	2.773	4.640	7.651	3.013	4.942	8.602	0.942	9×10^{-9}
Jeffreys								
μ	-0.141	0.004	0.169	-0.138	0.010	0.174	0.940	–
σ	0.854	0.987	1.126	0.859	0.992	1.132	0.933	–
λ	2.811	4.615	7.605	3.023	4.957	8.638	0.945	8×10^{-9}
BTV(1,1)								
μ	-0.131	0.012	0.179	-0.123	0.018	0.184	0.936	–
σ	0.851	0.983	1.114	0.851	0.985	1.123	0.932	–
λ	2.704	4.429	7.301	2.883	4.754	8.134	0.939	8×10^{-9}

Table 9: Skew-logistic data for noninformative priors: $\mu = 0, \sigma = 1, n = 200$.

Prior	MAP			Median			Coverage	BF
	5%	50%	95%	5%	50%	95%		
$\lambda = 0$								
BTV(1/2,1/2)								
μ	-0.342	-0.007	0.347	-0.341	0.006	0.356	0.936	–
σ	0.850	1.021	1.197	0.868	1.038	1.213	0.950	–
λ	-0.240	0.001	0.226	-0.335	0.001	0.304	0.934	6.549
Jeffreys								
μ	-0.343	-0.007	0.350	-0.347	0.006	0.349	0.939	–
σ	0.851	1.021	1.199	0.869	1.038	1.210	0.953	–
λ	-0.244	0.001	0.221	-0.330	0.004	0.303	0.936	6.333
BTV(1,1)								
μ	-0.354	-0.003	0.328	-0.338	0.003	0.333	0.944	–
σ	0.853	1.021	1.200	0.868	1.035	1.212	0.952	–
λ	-0.243	0.001	0.214	-0.319	0.001	0.292	0.934	4.277
$\lambda = 2.5$								
BTV(1/2,1/2)								
μ	-0.173	0.002	0.237	-0.157	0.014	0.240	0.945	–
σ	0.781	0.969	1.195	0.799	0.983	1.216	0.952	–
λ	0.834	2.039	4.492	1.068	2.377	5.605	0.944	0.011
Jeffreys								
μ	-0.173	0.001	0.249	-0.155	0.012	0.238	0.938	–
σ	0.771	0.971	1.203	0.788	0.985	1.217	0.953	–
λ	0.841	2.065	4.552	1.060	2.386	5.598	0.940	0.010
BTV(1,1)								
μ	-0.163	0.012	0.257	-0.145	0.026	0.251	0.938	–
σ	0.772	0.958	1.192	0.788	0.974	1.200	0.944	–
λ	0.778	1.951	4.391	0.991	2.241	5.094	0.946	0.008
$\lambda = 5$								
BTV(1/2,1/2)								
μ	-0.120	0.001	0.147	-0.113	0.008	0.157	0.938	–
σ	0.786	0.974	1.187	0.801	0.983	1.194	0.947	–
λ	1.943	4.172	9.410	2.427	4.877	12.860	0.936	0.002
Jeffreys								
μ	-0.128	-0.000	0.153	-0.117	0.009	0.159	0.943	–
σ	0.787	0.973	1.182	0.798	0.985	1.194	0.945	–
λ	1.950	4.128	10.446	2.381	4.868	12.906	0.936	0.003
BTV(1,1)								
μ	-0.109	0.009	0.164	-0.099	0.017	0.177	0.947	–
σ	0.772	0.962	1.159	0.787	0.974	1.182	0.947	–
λ	1.852	3.848	8.546	2.224	4.479	10.951	0.942	0.002

Table 10: Skew-Laplace data for noninformative priors: $\mu = 0, \sigma = 1, n = 100$.

Prior	MAP			Median			Coverage	BF
	5%	50%	95%	5%	50%	95%		
$\lambda = 0$								
BTV(1/2,1/2)								
μ	-0.204	-0.002	0.225	-0.209	-0.002	0.223	0.941	–
σ	0.889	1.008	1.136	0.898	1.013	1.142	0.942	–
λ	-0.147	-0.000	0.149	-0.183	0.000	0.185	0.947	10.585
Jeffreys								
μ	-0.206	-0.001	0.224	-0.211	-0.000	0.214	0.939	–
σ	0.884	1.007	1.133	0.895	1.014	1.141	0.939	–
λ	-0.148	0.001	0.154	-0.180	0.001	0.190	0.945	10.440
BTV(1,1)								
μ	-0.208	-0.002	0.211	-0.206	0.001	0.210	0.941	–
σ	0.887	1.007	1.132	0.898	1.014	1.140	0.942	–
λ	-0.138	0.000	0.148	-0.180	-0.000	0.189	0.945	6.852
$\lambda = 2.5$								
BTV(1/2,1/2)								
μ	-0.126	-0.004	0.155	-0.119	0.003	0.155	0.945	–
σ	0.840	0.987	1.147	0.845	0.994	1.159	0.936	–
λ	1.307	2.311	3.849	1.429	2.469	4.217	0.941	7×10^{-6}
Jeffreys								
μ	-0.127	-0.001	0.157	-0.118	0.004	0.161	0.940	–
σ	0.836	0.982	1.148	0.851	0.992	1.154	0.939	–
λ	1.289	2.291	3.830	1.403	2.460	4.176	0.944	6×10^{-6}
BTV(1,1)								
μ	-0.119	0.004	0.159	-0.110	0.009	0.160	0.945	–
σ	0.828	0.978	1.140	0.843	0.988	1.145	0.938	–
λ	1.263	2.210	3.738	1.382	2.385	4.045	0.945	5×10^{-6}
$\lambda = 5$								
BTV(1/2,1/2)								
μ	-0.084	-0.001	0.097	-0.080	0.002	0.096	0.951	–
σ	0.849	0.984	1.137	0.855	0.991	1.142	0.939	–
λ	2.817	4.596	7.956	3.036	4.900	8.621	0.952	7×10^{-8}
Jeffreys								
μ	-0.085	-0.002	0.095	-0.079	0.002	0.097	0.951	–
σ	0.847	0.987	1.142	0.853	0.990	1.142	0.940	–
λ	2.784	4.567	7.767	3.033	4.879	8.693	0.945	1×10^{-7}
BTV(1,1)								
μ	-0.080	0.004	0.099	-0.074	0.006	0.102	0.951	–
σ	0.840	0.978	1.128	0.849	0.986	1.135	0.932	–
λ	2.615	4.427	7.547	2.859	4.744	8.132	0.943	7×10^{-8}

Table 11: Skew-Laplace data for noninformative priors: $\mu = 0, \sigma = 1, n = 200$.

References

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

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

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

A. Azzalini and A. Capitanio. *The Skew-Normal and Related Families*. Cambridge: IMS Monographs, Cambridge University Press, 2014.

A. Azzalini and G. Regoli. The work of Fernando de Helguero on non-normality arising from selection. *Chilean Journal of Statistics*, 3:113–128, 2012.

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

M.D. Branco, M.G. Genton, and B. Liseo. Objective Bayesian analysis of skew-t distributions. *Scandinavian Journal of Statistics*, 40:63–85, 2013.

S. Cabras, W. Racugno, M.E. Castellanos, and L. Ventura. A matching prior for the shape parameter of the skew-normal distribution. *Scandinavian Journal of Statistics*, 39:236–247, 2012.

Antonio Canale, Euloge Clovis Kenne Pagui, and Bruno Scarpa. Bayesian modeling of university first-year students' grades after placement test. *Journal of Applied Statistics*, 43(16):3015–3029, 2016.

J.A. Christen and C. Fox. A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis*, 5:263–281, 2010.

F. de Helguero. Sulla rappresentazione analitica delle curve statistiche. *Giornale degli Economisti*, 38:241–265, 1909.

C. Fernández and M.F.J. Steel. On the dangers of modelling through continuous distributions: a Bayesian perspective. *Bayesian Statistics*, 6:213–238, 1999.

M. Ghosh. Objective priors: An introduction for frequentists. *Statistical Science*, 26:187–202, 2011.

A.L. Gibbs and F.E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70:419–435, 2002.

L.G. Godoi, M.D. Branco, and F. Ruggeri. Concentration function for the skew-normal and skew-t distributions, with application in robust Bayesian analysis. *Brazilian Journal of Probability and Statistics; in press*, 31:373–393, 2017.

M. Hallin and C. Ley. Skew-symmetric distributions and Fisher information – a tale of two densities. *Bernoulli*, 18:747–763, 2012.

G. Heinz, L.J. Peterson, R.W. Johnson, and C.J. Kerk. Exploring relationships in body dimensions. *Journal of Statistics Education*, 11, 2003.

M.C. Jones. Families of distributions arising from distributions of order statistics. *Test*, 13:1–43, 2004.

M.C. Jones. On families of distributions with shape parameters (with discussion). *International Statistical Review*, 83:175–192, 2015.

C. Ley. Flexible modelling in statistics: past, present and future. *Journal de la Société Française de Statistique*, 156:76–96, 2015.

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

Y.V. Marchenko and M.G. Genton. Multivariate log-skew-elliptical distributions with applications to precipitation data. *Environmetrics*, 21:318–340, 2010.

F.J. Rubio and B. Liseo. On the independence Jeffreys prior for skew-symmetric models. *Statistics & Probability Letters*, 85:91–97, 2014.

F.J. Rubio and M.F.J. Steel. Inference in two-piece location-scale models with Jeffreys priors. *Bayesian Analysis*, 9:1–22, 2014.

F.J. Rubio and M.F.J. Steel. Bayesian modelling of skewness and kurtosis with two-piece scale and shape distributions. *Electronic Journal of Statistics*, 9:1884–1912, 2015.

J.W. Seaman III, J.W. Seaman Jr, and J.D. Stamey. Hidden dangers of specifying noninformative priors. *The American Statistician*, 66:77–84, 2012.

D.P. Simpson, T.G. Martins, A. Riebler, G.A. Fuglstad, H. Rue, and S.H. Sørbye. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32:1–28, 2017.

D. Sun and J.O. Berger. Reference priors with partial information. *Biometrika*, 85:55–71, 1998.

S. Vallender. Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability and Its Applications*, 18:784–786, 1974.

J. Wang, J. Boyer, and M.G. Genton. A skew-symmetric representation of multivariate distri-

butions. *Statistica Sinica*, 14:1259–1270, 2004.