

Bayesian Inference for Skew-Symmetric Distributions

 Fatemeh Ghaderinezhad ¹, Christophe Ley ^{1,*}  and Nicola Loperfido ²
¹ Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281, S9, Campus Sterre, 9000 Gent, Belgium; Fatemeh.Ghaderinezhad@UGent.be

² Dipartimento di Economia, Società e Politica, Università degli Studi di Urbino “Carlo Bo”, Via Saffi 42, 61029 Urbino (PU), Italy; nicola.loperfido@uniurb.it

* Correspondence: christophe.ley@ugent.be

Received: 27 February 2020; Accepted: 18 March 2020; Published: 25 March 2020



Abstract: Skew-symmetric distributions are a popular family of flexible distributions that conveniently model non-normal features such as skewness, kurtosis and multimodality. Unfortunately, their frequentist inference poses several difficulties, which may be adequately addressed by means of a Bayesian approach. This paper reviews the main prior distributions proposed for the parameters of skew-symmetric distributions, with special emphasis on the skew-normal and the skew-*t* distributions which are the most prominent skew-symmetric models. The paper focuses on the univariate case in the absence of covariates, but more general models are also discussed.

Keywords: Bayesian analysis; Jeffreys prior; matching prior; reference prior; symmetry

1. Introduction

The need to model skewed data led to the development of many skewed distributions which are obtained by adding to a symmetric distribution a parameter that controls skewness [1]. Arguably, the best known example is the skew-normal distribution introduced by Azzalini (1985) [2]. Its probability density function (pdf) is

$$sn(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\lambda \frac{x - \mu}{\sigma}\right), x \in \mathbb{R}, \quad (1)$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma \in \mathbb{R}^+$ the scale parameter, both inherited from the standard normal distribution with pdf denoted by ϕ and (cumulative distribution function) cdf Φ , and $\lambda \in \mathbb{R}$ is called the skewness parameter given that density (1) is asymmetric for $\lambda \neq 0$ and reduces to the standard normal pdf for $\lambda = 0$. Several extensions and generalizations followed, see [3], a very general and highly popular being the skew-symmetric distributions of Wang et al. (2004) [4] with pdf

$$s_{f;\Pi}(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} f\left(\frac{x - \mu}{\sigma}\right) \Pi\left(\frac{x - \mu}{\sigma}, \lambda\right), x \in \mathbb{R},$$

where f is the symmetric density to be skewed and $\Pi: \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$ is a so-called skewing function satisfying $\Pi(z, \lambda) + \Pi(-z, \lambda) = 1 \forall z, \lambda \in \mathbb{R}$ and $\Pi(z, 0) = 1/2 \forall z \in \mathbb{R}$. The most widely used subfamily of skew-symmetric distributions has densities of the form

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

where G is any symmetric, univariate, absolutely continuous cumulative distribution function. In (2), $\mu \in \mathbb{R}$ is a location, $\sigma \in \mathbb{R}^+$ a scale and $\lambda \in \mathbb{R}$ a skewness parameter. The function G might be replaced by a function $w(\cdot)$ satisfying $0 \leq w(-x) = 1 - w(x) \leq 1$, as done in [4,5]. Different skew

models might then be obtained by different choices for f and G , as for example the pdf and the cdf of the power exponential distribution [6], the Student t distribution [5] and the logistic distribution [7].

The skew-normal distribution admits several stochastic representations, of which a very simple is given as follows [8]: let U and V be independent, standard normal variables, then

$$Z_\lambda = \frac{\lambda}{\sqrt{1+\lambda^2}}|U| + \frac{1}{\sqrt{1+\lambda^2}}V \quad (3)$$

has the standard skew-normal distribution $SN(\lambda) = SN(0,1,\lambda)$. This representation models departures from normality, allows for simple generation of random numbers and the calculation of odd moments. Despite its numerous appealing probabilistic properties [3], the skew-normal distribution suffers from serious inferential problems, as remarked already by Azzalini (1985). In the first place, the method of moments might lead to complex estimates of the parameters. In the second place, when $\mu = 0$, $\sigma = 1$ and all observations are positive (negative) the likelihood function will be monotonically increasing (decreasing) in λ , thus making the maximum likelihood estimate (MLE) of λ (minus) infinite. In the third place, the MLE is quite unstable even in the presence of negative and positive observations. In the fourth place, skew-normal data offer little help in discriminating between different values of the shape parameter λ : very different values of λ might correspond to very similar skew-normal densities. In the fifth place, the sampling distribution of the MLE does not allow for analytically tractable standard errors and confidence intervals. In the sixth place, if all three parameters are assumed to be unknown, the profile likelihood function for λ has a stationary point at $\lambda = 0$, regardless of the observed sample, and the Fisher information matrix is singular when λ approaches zero.

The latter problem, also known as the singularity problem, is very serious and has motivated an active line of research both for multivariate and semiparametric generalizations of the skew-normal distribution. It has been discussed in various papers such as [9–15]. Different parameterizations have been proposed by, among others, [2,11,13,16]. While most authors were pointing at some special status for the normal distribution as symmetric base distribution, Ref. [12] showed that this singularity can occur for very general symmetric densities f and provided a full characterization, in a general multivariate context, of the singularity problem, showing that it is due to unfortunate mismatches between the symmetric density f and the skewing function $\Pi(\cdot, \cdot)$. It is to be noted that in the context of models (2) singularity occurs only when f is normal, irrespective of the choice of function G . However, even the proposed solutions of reparameterization do not remove all problems as remarked by Azzalini and Capitanio (1999) [9]: “*there are cases where the likelihood shape and the MLE are problematic. We are not referring here to difficulties with numerical maximization, but to the intrinsic properties of the likelihood function, not removable by change of parameterization. In case of this sort, the behaviour of the MLE appears quite unsatisfactory, and an alternative estimation method is called for*”.

These difficulties arise from the shape of the likelihood function, which can be modified by appropriate weighting functions [17]. The most natural and best known approach to weight the likelihood function is the Bayesian one, where the prior distribution plays the role of the weight function. The Bayesian approach might adequately address both point estimation and hypothesis testing of the skewness parameter λ . Most papers in this area focused on objective priors for the skew-normal distribution, whereas only a few of them dealt with its multivariate or semiparametric generalizations. Bayesian inference and prior elicitation of λ for skew-symmetric distributions are challenging since λ not only controls symmetry but also spread, modes and tail behaviour. This paper is not meant to be a comprehensive review of this very active research topic, but rather a handy source for interested readers to the Bayesian analysis of skew-symmetric distributions. The paper is structured as follows. Section 2 recalls some basic concepts regarding the default choices for the prior distribution. Sections 3 and 4 review the prior distributions for the parameters indexing the skew-normal and some other skew-symmetric models, respectively, while Section 5 reviews some of the literature on generalizations of the univariate skew-symmetric model. Final comments are provided in Section 6.

2. Default Prior Choices in Bayesian Statistics

The Bayesian approach to quantify uncertainty in statistical inference can be broken down into three steps [18]. The first step consists in choosing a joint probability distribution for observable and unobservable quantities, consistently with the available knowledge about the underlying scientific problem and the data collection process. The second step is to condition on the observed data, which is carried out by means of several computational techniques. The third step is assessing the model's fit and interpreting the implications of the resulting posterior distribution. In this section we focus on the first step and review the most common default choices for the prior distribution. We use the following notation in the sequel: θ is our parameter of interest, $\pi(\theta)$ is the prior distribution, $\pi(\theta|t)$ is the posterior distribution given data information t , $p(t|\theta)$ is the data likelihood, $p(t, \theta) = p(t|\theta)\pi(\theta)$ is the joint distribution of t and θ , and $p(t)$ is the marginal distribution of t . With this notation in hand, we of course have that the posterior equals

$$\pi(\theta|t) = \frac{\pi(\theta)p(t|\theta)}{p(t)}.$$

2.1. Jeffreys Priors

In Bayesian analysis there are situations in which the available prior information is too vague to be formalized into a probability distribution, too controversial to be acceptable in scientific communities or too complicated to allow for a reliable statistical analysis. Hence the need for priors with minimal effect on the posterior distribution, so that “the chosen prior would let the data speak for themselves” [19]. Reference analysis aims at an “objective” Bayesian solution to statistical inference in the same way as conventional statistical methods, where solutions only depend on model assumptions and observed data.

One of the earliest non-informative (objective) priors is the uniform prior for the Binomial proportion [20,21]. Unfortunately, this prior suffers from its lack of invariance under one-to-one reparameterization. Jeffreys' prior is a non-informative prior which is invariant under one-to-one reparameterization and is proportional to the positive square root of the Fisher information associated with the parameter of interest. For regular models where asymptotic normality holds, the Jeffreys prior enjoys some optimality properties in the absence of nuisance parameters, but suffers from serious difficulties in the presence of nuisance parameters. As a first example, in the Neyman–Scott problem it leads to a strong inconsistency in Bayes estimation of the error variance [22]. As a second example, when estimating the product of two independent normal means, a circular symmetric prior was found to be superior to Jeffreys' prior [23]. As a third example, Jeffreys himself supported the use of another prior for location-scale models.

2.2. Reference Priors

Intuitively, a reference prior for some real-valued parameter θ is a prior of the form $\pi(\theta) = \pi(\theta|\mathcal{T}, \mathcal{P})$ which maximizes the missing information about θ within the class \mathcal{P} of prior distributions compatible with the available prior knowledge \mathcal{T} [19]. More formally, let D be a set of observations, generated by some random mechanism $p(D|\theta)$ that only depends on a real-valued parameter $\theta \in \Theta$. Furthermore, let $t = t(D) \in \mathcal{T}$ be any sufficient statistic (which may be the complete data set D). In Shannon's general information theory the amount of information $I^\theta[\mathcal{T}, \pi(\theta)]$, which may be expected to be provided by D or equivalently by $t(D)$ about θ , is

$$I^\theta[\mathcal{T}, \pi(\theta)] = \int_{\mathcal{T}} \int_{\Theta} p(t, \theta) \log \frac{p(t, \theta)}{p(t)\pi(\theta)} d\theta dt = \mathbb{E}_t \left[\int_{\Theta} \pi(\theta|t) \log \frac{\pi(\theta|t)}{\pi(\theta)} d\theta \right], \quad (4)$$

which is the expected Kullback–Leibler divergence of the prior from the posterior (here \mathbb{E}_t indicates that the expectation is taken on the t -part). The functional $I^\theta[\mathcal{T}, \pi(\theta)]$ is concave, non-negative and invariant under one-to-one transformations of θ . Lindley (1956) [24] and Bernardo (1979) [19] defined the reference prior as the prior maximizing (4). There are some situations where we need the

asymptotic maximization of the above expectation, since for a fixed n , its maximization might lead to a discrete prior with finitely many jumps, which is hardly compatible with the concept of diffuse prior [25]. Ref. [19] proved that, in the absence of any nuisance parameter, Jeffreys' prior yields the necessary maximization.

2.3. Matching Priors

Matching priors allow for posterior probability statements which have an interpretation as confidence statements in the sampling model. Matching priors aim at achieving a compromise between Bayesian and frequentist inference based on some order of approximation, thus providing default priors for routine use in Bayesian inference and possibly more palatable to frequentist statisticians. The concept of matching prior appears to have been proposed first by Lindley (1958) [26] and several matching priors have been proposed since, such as for example quantile matching priors, matching priors for distribution functions, highest probability density matching priors and matching priors associated with likelihood ratio statistics [27].

In this subsection we illustrate the approach to matching priors introduced by Welch and Peers in the seminal paper [28]. Suppose that Y_1, \dots, Y_n are i.i.d. random variables with pdf $f(Y|\theta)$, where θ is real-valued. In addition, assume all the regularity conditions which allow to expand the posterior around the MLE $\hat{\theta}_n$. Furthermore, for $0 < \alpha < 1$, let $\theta_{1-\alpha}^\pi(Y_1, \dots, Y_n) \equiv \theta_{1-\alpha}^\pi$ denote the $(1 - \alpha)$ -th asymptotic posterior quantile of θ based on the prior π , that is

$$P^\pi[\theta \leq \theta_{1-\alpha}^\pi | Y_1, \dots, Y_n] = 1 - \alpha + O_p(n^{-r})$$

for some $r > 0$. If $r = 1$, π is called a first-order matching prior and if $r = 3/2$ the prior π will be a second-order probability matching prior. For instance, the Jeffreys prior is a first-order probability matching prior in the absence of nuisance parameters. We illustrate this appealing property with an example from [29]. Suppose that Y_1, \dots, Y_n are i.i.d. with pdf $N(\theta, 1)$ and that $\pi(\theta) = 1$ with $-\infty < \theta < \infty$. Then the posterior density $\pi(\theta | Y_1, \dots, Y_n)$ stems from the $N(\bar{Y}_n, n^{-1})$. By considering $z_{1-\alpha}$ as the $100(1 - \alpha)\%$ quantile of the $N(0, 1)$ distribution, we have

$$P[\sqrt{n}(\theta - \bar{Y}_n) \leq z_{1-\alpha} | Y_1, \dots, Y_n] = 1 - \alpha = P[\sqrt{n}(\bar{Y}_n - \theta) \geq -z_{1-\alpha} | \theta].$$

Therefore, the one-sided credible interval $\bar{Y}_n + z_{1-\alpha}/\sqrt{n}$ for θ has exact frequentist coverage probability $(1 - \alpha)$. This exact matching does not always exist. However, if Y_1, \dots, Y_n are i.i.d. random variables then

$$\hat{\theta}_n | \theta \stackrel{a}{\sim} N\left(\theta, \mathcal{J}^{-1}/n\right)$$

where \mathcal{J} is the expected Fisher information and $\stackrel{a}{\sim}$ means asymptotically equivalent in distribution. Using the delta method we have

$$g(\hat{\theta}_n) | \theta \stackrel{a}{\sim} N\left[g(\theta), (g'(\theta))^2 \mathcal{J}^{-1}/n\right].$$

Therefore, if $g'(\theta) = \mathcal{J}^{1/2}(\theta)$, then $g(\theta) = \int_\theta \mathcal{J}^{1/2}(t) dt$ and $\sqrt{n}[g(\hat{\theta}_n) - g(\theta)] | \theta$ is asymptotically $N(0, 1)$. In the absence of nuisance parameters, a first-order matching prior for θ is a solution of the differential equation

$$\frac{d}{d\theta} \left[\pi(\theta) \mathcal{J}^{-1/2}(\theta) \right] = 0,$$

so that the Jeffreys prior is the unique first-order matching prior, but it does not always hold for the second-order matching probability [27].

To obtain the second-order matching prior we need an asymptotic expansion of the posterior distribution function up to $O(n^{-1})$ and the differential equation given by Mukerjee and Dey (1993) [30] and Mukerjee and Ghosh (1997) [31], that is

$$\frac{1}{3} \frac{d}{d\theta} \left[(\pi(\theta) \mathcal{J}^{-2}(\theta) g_3(\theta)) \right] + \frac{d^2}{d\theta^2} \left[(\pi(\theta) \mathcal{J}^{-1}(\theta)) \right] = 0 \quad (5)$$

where $g_3(\theta) = -\mathbb{E} \left[\frac{d^3 \log f(Y_1|\theta)}{d\theta^3} \middle| \theta \right]$. Jeffreys' prior is the unique second-order matching prior if it satisfies (5), as it happens for the location-scale families: for $\pi_J(\theta) = J^{1/2}(\theta)$, (5) converts to

$$\frac{1}{3} \frac{d}{d\theta} [J^{-3/2}(\theta)g_3(\theta)] + \frac{d^2}{d\theta^2} J^{-1/2}(\theta) = 0$$

which requires

$$\frac{1}{3} J^{-3/2}(\theta)g_3(\theta) + \frac{d}{d\theta} J^{-1/2}(\theta)$$

to be constant for all values of θ . We refer the reader to [29] for more details on cases where, in the absence of nuisance parameters, there is not a second-order probability matching prior, and where, in the presence of nuisance parameters, there are first- and second-order matching priors and where there is not a second-order matching prior.

3. Prior Choices for the Skew-Normal Distribution

This section reviews the prior distributions proposed for Bayesian inference on the parameters of the skew-normal distribution: the reference prior by Liseo and Loperfido (2006) [32], the matching prior by Cabras et al. (2012) [33] and the informative prior by Canale and Scarpa (2013) [34].

3.1. The Reference Prior

Liseo and Loperfido (2006) [32] first proposed a default prior for the shape parameter of the location-scale-free (standard) skew-normal model $sn(z; \lambda) = 2\phi(z)\Phi(\lambda z), z \in \mathbb{R}$. The associated Jeffreys prior is

$$\pi_J(\lambda) \propto J^{1/2}(\lambda) \text{ where } J(\lambda) = \int_{-\infty}^{\infty} 2z^2\phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz.$$

This prior is proper, symmetric about $\lambda = 0$, decreasing in $|\lambda|$ and its tails are of order $O(\lambda^{-3/2})$. This prior is therefore suitable for testing the hypothesis of symmetry, which might be formalized in the skew-normal framework as $H_0 : \lambda = 0$ versus $H_1 : \lambda \neq 0$. The same authors investigated the frequentist performances of this prior with simulated data, concluding that the Bayesian approach might be beneficial in easing some inferential difficulties of the frequentist approach for the standard skew-normal distribution.

Ref. [32] also considered a default Bayes analysis for the general scalar case (1), where λ is the parameter of interest and the location parameter μ and the scale parameter σ are the nuisance parameters. They are assumed to be independent of λ and to have a normal-inverse gamma distribution:

$$\mu|\sigma \sim N\left(\mu_0, \frac{\sigma^2}{\tau}\right) \text{ with } \mu_0 \in \mathbb{R}, \tau > 0, \sigma^{-2} \sim \text{Gamma}(\alpha, \beta)$$

where $\text{Gamma}(\alpha, \beta)$ is a Gamma distribution with parameters $\alpha, \beta > 0$. The default prior $\pi(\mu, \sigma) \propto \sigma^{-1}$ is a limiting case and is the conditional reference prior for (μ, σ) given λ . These assumptions allow for a closed-form expression of the marginal likelihood for λ . The proposed method has been successfully applied to the infamous "frontier" dataset (see <http://azzalini.stat.unipd.it/SN/frontier.dat>), where the maximum likelihood estimate of the skewness parameter λ is infinite.

Bayes and Branco (2007) [35] highlighted the advantages of the Bayesian approach and proposed two priors. They considered the stochastic representation (3) of the skew-normal distribution and, following the Bayes-Laplace rule, chose the uniform distribution on the interval $[-1, 1]$ as a prior for $\lambda/\sqrt{1-\lambda^2}$, thus leading to a $t(0, 0.5, 2)$ distribution as prior for λ , where $t(a, b, c)$ denotes the Student t distribution centered in $a \in \mathbb{R}$ with scale $b > 0$ and $c > 0$ degrees of freedom, which is a non-vague

and non-subjective prior. They further proposed the tractable approximation $t(0, \pi^2/4, 1/2)$ for the Jeffreys prior from [32]. They motivated it by the following approximation (see [36]):

$$\frac{1}{\pi} \frac{\phi(x)}{\sqrt{\Phi(x)(1-\Phi(x))}} \approx \frac{1}{\sqrt{2\pi}(\pi/2)} \exp\left(-\frac{2x^2}{\pi^2}\right).$$

3.2. The Matching Prior

Cabras et al. (2012) [33] proposed another approach towards Bayesian inference about the shape parameter of the skew-normal distribution. It is based on a pseudo-likelihood function and a matching prior distribution for the shape parameter when location and scale parameters are unknown. First, they derive the marginal likelihood

$$L_m(\lambda) = \int_0^\infty \int_{-\infty}^{+\infty} \frac{L(\lambda, \eta)}{\sigma} d\mu d\sigma, \quad (6)$$

where $L(\lambda, \eta) = \prod_{i=1}^n sn(y_i; \eta, \lambda)$ is the skew-normal likelihood function, $\eta = (\mu, \sigma)$ is the nuisance parameter and σ^{-1} is the right-invariant Haar measure on the location-scale group of transformations, whose action on the parameter space leaves λ unchanged. By considering the fact that the marginal likelihood (6) can be approximated by the modified profile likelihood $L_{mp}(\lambda)$ of [37] since $L_m(\lambda) = L_{mp}(\lambda)(1 + O(n^{-1}))$ (see [38]) and by invoking results about the use of pseudo-likelihood functions in Bayesian analysis, the matching prior $\pi(\lambda)$ is simply proportional to the square root of the inverse of the asymptotic variance of the MLE of λ . Based on Ventura et al. (2009) [39], the matching prior for λ corresponding to (6) is

$$\pi(\lambda) \propto I_{\lambda\lambda, \eta}(\lambda, \hat{\eta}_\lambda)^{1/2}, \quad (7)$$

where $\hat{\eta}_\lambda$ is the constrained MLE of η for a given λ and $I_{\lambda\lambda, \eta}(\lambda, \eta) = I_{\lambda\lambda}(\lambda, \eta) - I_{\lambda\eta}(\lambda, \eta)I_{\eta\eta}(\lambda, \eta)^{-1}I_{\eta\lambda}(\lambda, \eta)$ is the partial information with $I_{\lambda\lambda}(\lambda, \eta)$, $I_{\lambda\eta}(\lambda, \eta)$, $I_{\eta\eta}(\lambda, \eta)$ and $I_{\eta\lambda}(\lambda, \eta)$ blocks of the expected Fisher information

$$I(\lambda, \eta) = \begin{bmatrix} I_{\lambda\lambda}(\lambda, \eta) & I_{\lambda\eta}(\lambda, \eta) \\ I_{\eta\lambda}(\lambda, \eta) & I_{\eta\eta}(\lambda, \eta) \end{bmatrix}.$$

For the interested reader we provide the detailed quantities of this matrix:

$$I_{\lambda\lambda}(\lambda, \eta) = a_2, I_{\lambda\mu}(\lambda, \eta) = \frac{1}{\sigma} \left(\frac{b}{A^{3/2}} - \lambda a_1 \right), I_{\lambda\sigma}(\lambda, \eta) = -\frac{\lambda a_2}{\sigma}, I_{\mu\mu}(\lambda, \eta) = \frac{(1 + \lambda^2 a_0)}{\sigma^2},$$

$$I_{\mu\sigma}(\lambda, \eta) = \frac{1}{\sigma^2} \left[\frac{b\lambda(1 + 2\lambda^2)}{A^{3/2}} + \lambda^2 a_1 \right], I_{\sigma\sigma}(\lambda, \eta) = \frac{2 + \lambda^2 a_2}{\sigma^2},$$

where $b = \sqrt{\frac{2}{\pi}}$, $A = 1 + \lambda^2$, and $a_i = \mathbb{E} \left[Z^i \left(\frac{\phi(\lambda Z)}{\Phi(\lambda Z)} \right)^2 \right]$, $i = 0, 1, 2$, with Z following the standard skew-normal with parameter λ . However, the prior (7) might be data dependent because of the presence of $\hat{\eta}_\lambda$. A prior for λ which does not suffer from this problem is proportional to

$$\sqrt{\frac{a_2 A^2 [\pi(1 + a_0 \lambda^4) + \lambda^2(\pi(1 + a_0) - 4)] + 2\sqrt{2\pi} a_1 \lambda A^{3/2} - \pi a_1^2 \lambda^2 A^3 - 2}{\pi A^3 [2 + \lambda^2(2a_0 + a_2) + \lambda^4(a_0 a_2 - a_1^2)] - 2(\lambda + 2\lambda^3)^2 - 2\sqrt{(2\pi)} a_1 \lambda^3 \sqrt{A}(1 + 3\lambda^2 + 2\lambda^4)}}.$$

This prior is proper, symmetric at the origin and with tails of order $O(\lambda^{-3/2})$. It also compensates for the possible monotonicity of the modified profile likelihood (6) and possesses good frequentist properties.

3.3. The Informative Priors of Canale and Scarpa (2013)

Canale and Scarpa (2013) [34] discuss two informative priors for the skewness parameter of the skew-normal distribution. Their study is motivated by an interesting data set on marks referring to first-year undergraduate students for the program in Economics at the University of Padua. The skew-normal model is implemented on students' grades in the first mandatory class of Statistics. Making inference on the grades of the previous years shows that the distribution of Statistics grades is skewed to the right around a certain mean, which explains why they need informative priors for their endeavour.

The first prior is the normal density with hyperparameters reflecting prior beliefs about the expectation and variance of λ in order to center the prior on a particular guess for λ . The resulting posterior belongs to the family of unified skew-normal (SUN) distributions, introduced in [40]. The explicit expressions for the mean and the variance of the posterior are not very tractable but they allow for a simple interpretation. The second informative prior is itself a skew-normal, motivated by the distribution of grades of university examinations [41]. The skew-normal prior includes location and scale hyperparameters as well as a skewness hyperparameter reflecting the beliefs on the direction of skewness. The posterior distribution also belongs to the class of SUN distributions. The authors set the location hyperparameter of the skew-normal prior equal to zero in order to have a rough prior information only on the skewness side of the distribution of the data: considering negative or positive values for the skewness hyperparameter puts more prior mass on the positive or negative semi-axis. In both cases the resulting posteriors are intractable, but the SUN parametrization eases efficient sampling methods for posterior computation via Markov Chain Monte Carlo (MCMC). For both prior choices for λ , they have specified an independent normal inverse gamma distribution for the location and scale parameters. To perform the related Bayesian inference, the authors have presented an algorithm to simulate the full conditional distribution of the skewness parameter λ given the location and scale parameter. This algorithm uses a Gibbs sampler for the stochastic representation of the SUN model. To get the posterior, the authors introduced normal latent variables, say η_1, \dots, η_n . Conditionally on these latent variables, the generic i -th observation will be normally distributed with a specific mean and variance. This way of constructing the Gibbs sampler leads to conjugacy for the location and scale parameters. For the detailed computations we refer the reader to [34]. This sampling method is useful in MCMC methods to approximate the posterior distribution.

We also wish to mention that generally the MCMC method in Bayesian statistics bears a particular importance in model selection. Suppose we have a set of models reflecting competing hypotheses about the underlying data set, where each model is characterized by a specific vector of parameters of interest. From the Bayesian viewpoint, these models are compared pairwise through their Bayes factor which is the ratio of relative marginal likelihoods. Obviously, finding the marginal likelihood is often not feasible in particular analytically. We refer the reader to [42] and references therein for estimation methods of the marginal likelihood, specifically in general non-nested models.

4. Prior Choices for Other Skew-Symmetric Distributions

There exists a wide literature on the Bayesian analysis of skew-symmetric distributions different from the skew-normal. Azzalini (1986) [6] and Naranjo et al. (2012) [43] provided a Bayesian analysis of a skewed exponential power distribution. This family includes the symmetric exponential distribution as well as the skew-normal distribution, and provides flexible distributions with lighter and heavier tails. Interestingly, this family of densities can fit each tail separately. Hossianzadeh and Zare (2016) [44] estimated the parameter of the discrete skewed Laplace distribution by an empirical Bayesian analysis and compared it with the maximum likelihood approach. In what follows, we will first consider the popular skew- t distribution and then focus on two general approaches for skew-symmetric densities.

4.1. Jeffreys' Prior for Skew- t Distributions

Skew- t distributions are the best-known alternatives to skew-normal ones, due to their flexibility: they can model any level of skewness and excess kurtosis. However, they pose some further inferential

problems, which we illustrate in the simpler case of the Student t distribution with known location and scale parameters. Ref. [45] discussed that the likelihood function approaches infinity when the degrees of freedom go to zero, and showed that the supremum of the likelihood function may be achieved when the degrees of freedom go to infinity. There have been several frequentist attempts to solve the inferential problems of the skew- t distribution with all parameters unknown. Sartori (2006) [46] used the modified score function, which requires the degrees of freedom to be fixed. Azzalini and Genton (2008) [10] proposed a deviance approach which is only partially satisfactory, since its implementation might not be straightforward. We illustrate the problem with the univariate skew- t distribution. The deviance approach replaces the boundary maximum likelihood estimate of (λ, v) by the smallest vector (λ_0, v_0) for which the null hypothesis $H_0 : (\lambda, v) = (\lambda_0, v_0)$ is not rejected. The deviance approach assumes that such smallest vector exists but neither theoretical results nor simulation studies support this assumption. For these reasons we cannot exclude the existence of samples admitting two vectors (λ_a, v_a) and (λ_b, v_b) satisfying $\lambda_a > \lambda_b$ and $v_a < v_b$ for which the hypotheses $H_0 : (\lambda, v) = (\lambda_a, v_a)$ and $H_0 : (\lambda, v) = (\lambda_b, v_b)$ are not rejected, while no vector (λ_0, v_0) exists satisfying $(\lambda_0, v_0) < (\lambda_a, v_a)$, $(\lambda_0, v_0) < (\lambda_b, v_b)$ and for which the null hypothesis $H_0 : (\lambda, v) = (\lambda_0, v_0)$ is not rejected. We reckon that this situation is likely to happen, given that when either of the parameters λ or v is large the shape of the skew- t density function remains almost unchanged if one parameter is substantially increased while the other is substantially decreased.

Given these shortcomings, Branco, Genton and Liseo (2012) [47] studied Bayesian analysis for various forms of skew- t distributions. Denoting by $\nu > 0$ the degrees of freedom parameter, they first considered skew- t densities of the form

$$2t(z|\nu)T(\lambda z|\nu), \quad z \in \mathbb{R},$$

where $t(\cdot|\nu)$ and $T(\cdot|\nu)$ are the pdf and the cdf of a Student t distribution with ν degrees of freedom. The corresponding Jeffreys prior for the skewness parameter λ when ν is known and finite is

$$\pi_J(\lambda|\nu) \propto \sqrt{\int_{-\infty}^{\infty} z^2 t(z|\nu) \frac{t^2(\lambda z|\nu)}{T(\lambda z|\nu)} dz}.$$

It is proper, symmetric about zero and with tails of order $O(\lambda^{-3/2})$. The same authors further investigated the case of the skew- t distribution of [5] with pdf

$$2t(z|\nu)T\left(\lambda z \sqrt{\frac{\nu+1}{\nu+z^2}} \middle| \nu+1\right), \quad z \in \mathbb{R}.$$

The corresponding Jeffreys prior for λ for known and finite ν is

$$\pi_J(\lambda|\nu) \propto \sqrt{\int_0^{\infty} \frac{z^2 t(z|\nu) t^2(\lambda z|\nu)}{(\nu+z^2) T\left(\lambda z \sqrt{\frac{\nu+1}{\nu+z^2}} \middle| \nu+1\right) \left[1 - T\left(\lambda z \sqrt{\frac{\nu+1}{\nu+z^2}} \middle| \nu+1\right)\right]} dz}.$$

Again, this prior is proper, symmetric about zero and the tails are of order $O(\lambda^{-3/2})$.

4.2. Jeffreys Prior for General Skew-Symmetric Models

Rubio and Liseo (2014) [48] investigated the Jeffreys prior for the skewness parameter of a general class of scalar skew-symmetric models. The Jeffreys prior cannot be used for some skew-symmetric models at $\lambda = 0$ because of the singularity of the Fisher information at this point; see the Introduction for details about this issue. They showed that under mild conditions, including knowledge of location and scale parameters, the Jeffreys prior of the skewness parameter λ in the skew-symmetric model is proper, symmetric about zero and tails are of order of $O(|\lambda|^{-3/2})$. They used these results to construct the independence Jeffreys prior for the model including the location and scale parameters: it is the product of the Jeffreys prior of each parameter, under the assumption that the remaining parameters are held fixed. The same authors also provided sufficient conditions for the existence of

the posterior distribution and briefly discussed the existence of a proper independence Jeffreys prior for the skew-logistic model described in [7] and gave a Student t approximation to that prior.

The approach in [48] might be sketched as follows. The Fisher information for the shape parameter in (2) is

$$I(\lambda) = \int_{-\infty}^{\infty} 2z^2 f(z) \frac{g^2(\lambda z)}{G(\lambda z)} dz = \int_0^{+\infty} 2z^2 f(z) [\pi h(\lambda z)]^2 dz$$

where

$$h(z) = \frac{1}{\pi} \frac{g(z)}{\sqrt{G(z) [1 - G(z)]}} \tag{8}$$

and therefore in this case the Jeffreys prior for λ is $\pi_J(\lambda) \propto \sqrt{I(\lambda)}$. The first step for approximating the function $h(z)$ in (8) is to see that the transformed random variable $Z = G^{-1}(X) \sim h(z)$, where the random variable $X \sim \text{Beta}(1/2, 1/2)$ and G is a cumulative distribution function of an absolutely continuous symmetric random variable, that is $G(-z) = 1 - G(z)$ for all $z \in \mathbb{R}$. The corresponding cumulative distribution function H is

$$H(z) = \frac{2}{\pi} \arcsin \left[\sqrt{G(z)} \right].$$

An approximation of h in terms of g might then be achieved by choosing the scale parameter σ such that

$$h(z) \approx \frac{1}{\sigma} g\left(\frac{z}{\sigma}\right).$$

The quality of this approximation depends on the thickness of the tails of g . The authors illustrate this point by considering the case of $g(z)$ having a Student t distribution with ν degrees of freedom and comparing the approximations using quantiles. Alternatively, σ might be chosen to minimize the Kullback-Leibler divergence between $h(z)$ and $g(z/\sigma)/\sigma$. Ref. [32] approximated the Jeffreys prior using the parameterization $\delta = \lambda/\sqrt{1 + \lambda^2}$. They also proposed to use the symmetric Beta prior $\text{Beta}(\tau, \tau)$ for $\beta = (\delta - 1)/2$, thus leading to the Student t prior for λ

$$\pi_J(\lambda) = \frac{\Gamma(2\tau)}{2^{2\tau-1} \Gamma(\tau) \Gamma(\tau) (1 + \lambda^2)^{\tau+1/2}}$$

which reduces to the Cauchy distribution for $\tau = 0.5$, see [32].

4.3. Distance-Based Priors

As already mentioned, the shape parameter λ does not only impact the skewness in skew-symmetric models, but also the mean, the variance, the modes and the kurtosis. Dette et al. (2018) [49] dealt with this issue by assigning a prior distribution on the perturbation effect of the skewness parameter, quantified by the Total Variation distance between the symmetric density f and its skew-symmetric counterpart $2f(x)G[\lambda w(x)]$, where w is an odd function, rather than on the skewness parameter itself. The rationale behind this choice is that such a distance is more easily interpretable than the parameter λ , and hence informative as well as non-informative priors can more readily be found for the effect of λ than for λ itself. The Total Variation distance between two probability measures $\mu(\cdot)$ and $\nu(\cdot)$ on \mathbb{R} is

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

that is, the maximum difference between the probabilities assigned to the same event by the two measures. It is bounded between zero and one, $0 \leq d_{TV}(\mu, \nu) \leq 1$. The Total Variation distance between f and $2f(x)G[\lambda w(x)]$ is given by

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

The symmetry of G implies that $d_{TV}(f, G|\lambda)$ is not a one-to-one function of the parameter λ : $d_{TV}(f, G|\lambda) = d_{TV}(f, G|-\lambda)$. It is therefore convenient to use $M_{TV}(\lambda) = \text{sign}(\lambda)d_{TV}(f, G|\lambda)$ as a measure of perturbation, due to its appealing properties: $M_{TV}(0) = 0$, the largest/smallest value of $M_{TV}(\lambda)$ is ± 0.5 (attained when $\lambda \rightarrow \pm\infty$) so that $M_{TV}(\lambda) \in (-1/2, 1/2)$, and $M_{TV}(\lambda)$ is invariant under affine transformations. Moreover, $M_{TV}(\lambda) = 0.5 \left[1 - 2S_{f;G}(0; \lambda) \right]$, where $S_{f;G}$ is the cdf associated with $s_{f;G}$, which means that $M_{TV}(\lambda)$ is a re-scaling of the difference between the mass cumulated on either side of the origin for a fixed choice of f and G by the distribution $S_{f;G}$. In summary, $M_{TV}(\lambda)$ quantifies the impact of the parameter λ on the relocation of the probability mass on either side of the symmetry center of f .

The proposed measure M_{TV} allows to build both informative and non-informative priors for the perturbation parameter λ in skew-symmetric models. Since $M_{TV} \in (-1/2, 1/2)$ is an injective function of λ any prior option for λ induces a proper prior. Ref. [49] proposed, for $M_{TV}(\lambda)$, Beta priors with support on the interval $(-1/2, 1/2)$ and with density

$$\frac{1}{\text{Beta}(\alpha, \beta)} \left(u + \frac{1}{2}\right)^{\alpha-1} \left(\frac{1}{2} - u\right)^{\beta-1},$$

where $\text{Beta}(\alpha, \beta)$ is a beta function with hyperparameters $\alpha, \beta > 0$. This induces on λ the proper prior with pdf

$$\pi(\lambda|\alpha, \beta) = \frac{1}{\text{Beta}(\alpha, \beta)} \left(M_{TV}(\lambda) + \frac{1}{2}\right)^{\alpha-1} \left(\frac{1}{2} - M_{TV}(\lambda)\right)^{\beta-1} \frac{d}{d\lambda} M_{TV}(\lambda).$$

Priors of this type are called Beta Total Variation priors and are denoted by $BTV(\alpha, \beta)$; they are flexible, interpretable and lead to tractable posterior distributions. The behaviour of the prior $BTV(\alpha, \beta)$ is well illustrated by the special case $BTV(1, 1)$, that is a uniform prior giving equal probability mass to any pair of subintervals of equal length belonging to the support. If g is a bounded pdf and $\int_0^1 w(x)f(x)dx < \infty$, then $BTV(1, 1)$ is well-defined for all λ and is given by

$$\pi_{TV}(\lambda|1, 1) = 2 \int_0^\infty w(x)f(x)g[(\lambda w(x))] dx. \tag{9}$$

Since (9) does not have a closed-form, the authors proposed to approximate it by a Cauchy distribution centered at the origin and with scale parameter equal to 0.92. A Monte Carlo study showed that the proposed non-informative prior induces a posterior distribution with good frequentist properties and similar to those of the Jeffreys prior.

4.4. Prior Choices in the Presence of Kurtosis Parameters

Rubio and Steel (2015) [50] have proposed a general strategy for constructing weakly informative priors for kurtosis parameters by assigning a uniform prior to a bounded measure of kurtosis applied to the symmetric baseline density $f(\cdot|\delta)$ in which δ is the tail parameter and is a one-to-one function of the kurtosis. This methodology, used in [51], induces a proper prior on δ that can be interpreted as weakly non-informative prior, in that it assigns a flat prior on a function that incorporates the influence of the parameter δ on the shape of the density. This prior can be coupled with the Jeffreys prior for the skewness parameter in order to produce a joint prior for (δ, λ) in skew-symmetric models by using $p(\lambda, \delta) = p(\lambda|\delta)p(\delta)$ where

$$p(\lambda|\delta) \propto \sqrt{\int_0^\infty \frac{x^2 f(x|\delta) g(\lambda x)^2}{G(\lambda x) [1 - G(\lambda x)]} dx}.$$

For each value of δ the tails are of order $O(|\lambda|^{-3/2})$. A simulation study showed that this prior produces a posterior density with good frequentist properties.

5. Overview on Related Topics

So far this paper has focussed on the univariate case without covariates. This section briefly reviews some of the literature on more general settings related to skew-symmetric distributions.

Ref. [52] proposed a general population Monte Carlo algorithm in order to conduct a full Bayesian analysis of the multivariate skew-normal distribution, also in the presence of constrained parameters. Since the prior distribution approximates the actual reference prior for the shape parameter vector, this approach can be considered as a weakly informative prior. In addition, a generalization to the matrix variate regression model with skew-normal error is also provided.

Ref. [53] carried out a Bayesian analysis of a p -variate skew- t distribution by providing a new parameterization, considering a set of non-informative priors and a sampler designed to obtain the posterior model based on the parameters. The methodology can be extended to multivariate regression models with skewed errors and also stochastic frontier models.

Ref. [54] investigated the time series of electricity spot prices, which exhibit heavy tails and skewness. The authors conducted Bayesian inference on the multivariate skew- t distribution by putting a normal prior on the multi-dimensional skewness parameter.

Ref. [51] proposed a general non-informative structure for regression models with skew-symmetric errors, showed that under some mild conditions the resulting posterior distribution is proper and extended the results to the cases where the response variables are censored. The authors also investigated accelerated failure time models, which are relevant in survival analysis. Different prior distributions have been implemented on the skewness parameter of the skew-normal model including a Jeffreys prior, a matching prior, an informative prior and a uniform, noninformative prior on the parameter $\delta = \lambda/\sqrt{1 + \lambda^2}$, leading to the proper prior

$$\pi(\lambda) \propto \frac{1}{(1 + \lambda^2)^{3/2}}.$$

Ref. [55] used finite mixtures of skewed distributions to model flow cytometry data, in order to describe their skewness, kurtosis and heterogeneity. The authors developed Bayesian inference based on data augmentation and MCMC sampling using the aforementioned model. Data augmentation in this case is based on stochastic representation of the skew-normal distribution in terms of a random-effects model with truncated normal random effects. Finite mixtures of skew-normals provide a Gibbs sampling scheme that can be drawn from standard densities only. The same MCMC scheme is extended to mixtures of skew- t distributions by considering the skew- t distribution as a scale mixture of skew-normals.

Ref. [56] proposed a new class of distributions by introducing a skewness parameter in multivariate elliptically symmetric densities. This class of densities contains many standard families such as skew- t and skew-normal distributions. They condition on some unobserved variables commonly used in regression modelling and model stock market returns, security options or risky financial assets subject to shocks. Within the Bayesian realm, they show inter alia that there exist posterior distributions and moments for regression coefficients derived under improper priors.

Linear mixed models (LMM) are commonly used to analyze repeated measure data since they allow for flexible modelling of within-subject correlations. Mostly LMM for continuous responses assume that the random effects and the within-subjects errors are normally distributed, which can be unrealistic. Ref. [57] considered the less restrictive assumption of skew-normality and Bayesian inference based on prior distributions very similar to non-informative ones. They illustrated the proposed approach with the Framingham cholesterol data, obtained from a well-known long-term study aimed at investigating the relationship between various risk factors and diseases and to characterize the natural history of chronic circulatory diseases.

6. Discussion

In this paper we have provided an overview on the various proposals of Bayesian inference within skew-symmetric models. We hope that the reader will consider it as a helpful tool and source

of information on this research domain. We refer the interested reader to the simulation study and real data analysis of the recent paper [49] for a performance comparison between several of the above described prior proposals. Digging further into performance comparisons is a promising research task in order to get a more complete picture on which prior to ideally use in which situation when dealing with skew-symmetric distributions.

A referee remarked that in the general case the posterior distribution is multimodal and it is therefore necessary to impose some conditions ensuring unimodality. Log-concavity implies unimodality and it is preserved under convolution, marginalization, affine transformations and conditioning. For example, the assumption that the joint distribution of the parameter and the observations is log-concave implies that the posterior distribution is log-concave, too. We illustrate this point with a simple example. Assume that we sampled just one observation from a standard skew-normal distribution and that our prior distribution on the shape parameter is standard normal: $f(z|\lambda) = 2\phi(z)\Phi(\lambda z)$ and $\pi(\lambda) = \phi(\lambda)$. The joint distribution of the observation and the parameter is then $f(z, \lambda) = 2\phi(z)\phi(\lambda)\Phi(\lambda z)$. A little calculation shows that $f(z, \lambda)$ is log-concave. Hence, without further calculation, we know that $f(\lambda|z)$ is log-concave and hence unimodal, too. MAP estimates are then uniquely defined and can be easily derived by noticing that the posterior distribution is skew-normal: $\pi(\lambda|z) = 2\phi(\lambda)\Phi(\lambda z)$. Ref. [58] provides a thorough review of the literature on log-concavity, both in the univariate and in the multivariate case. Would the assumption of log-concavity be too restrictive, one could resort to other multivariate generalizations of unimodality, as for example block-unimodality, which already appeared in the Bayesian literature ([59]).

Author Contributions: Conceptualization, F.G. and C.L.; methodology, F.G.; writing—original draft preparation, F.G.; writing—review and editing, C.L. and N.L.; supervision, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by a BOF Starting Grant of Ghent University.

Acknowledgments: We thank two anonymous reviewers for their constructive comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ley, C. Skew distributions. In *Statistical Theory and Methods, Encyclopedia of Environmetrics*, 2nd ed.; El-Shaarawi, A., Piegorisch, W., Eds.; Wiley: New York, NY, USA, 2012; pp. 1944–1949.
- Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
- Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*; IMS Monograph Series; Cambridge University Press: Cambridge, UK, 2014.
- Wang, J.; Boyer, J.; Genton, M. A skew-symmetric representation of multivariate distribution. *Stat. Sin.* **2004**, *14*, 1259–1270.
- Azzalini, A.; Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *J. R. Stat. Soc. Ser. B* **2003**, *65*, 367–389.
- Azzalini, A. Further results on a class of distributions which includes the normal ones. *Statistica* **1986**, *46*, 199–208.
- Nadarajah, S. The skew logistic distribution. *Statistica* **2009**, *93*, 197–203.
- Henze, N. A probabilistic representation of the skew-normal distribution. *Scand. J. Stat.* **1986**, *13*, 271–275.
- Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew-normal distribution. *J. R. Stat. Soc. Ser. B* **1999**, *61*, 579–602.
- Azzalini, A.; Genton, M. Robust likelihood methods based on the skew-*t* and related distributions. *Int. Stat. Rev.* **2008**, *76*, 106–129.
- Chiogna, M. A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Stat. Methods Appl.* **2005**, *14*, 331–341.
- Hallin, M.; Ley, C. Skew-symmetric distributions and Fisher information—A tale of two densities. *Bernoulli* **2012**, *18*, 747–763.
- Hallin, M.; Ley, C. Skew-symmetric distributions and Fisher information: The double sin of the skew-normal. *Bernoulli* **2014**, *20*, 1432–1453.

14. Ley, C.; Paindaveine, D. On the singularity of multivariate skew-symmetric models. *J. Multivar. Anal.* **2010**, *101*, 1434–1444.
15. Pewsey, A. Problems of inference for Azzalini's skew-normal distribution. *J. Appl. Stat.* **2000**, *27*, 859–870.
16. Arellano-Valle, R.B.; Azzalini, A. The centred parametrization for the multivariate skew-normal distribution. *J. Multivar. Anal.* **2008**, *99*, 1362–1382; Erratum in **2009**, *100*, 816.
17. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **1993**, *80*, 27–38.
18. Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B. *Bayesian Data Analysis*, 3rd ed.; Chapman & Hall/CRC: London, UK, 2013.
19. Bernardo, J.M. Reference posterior distributions for Bayesian inference (with discussion). *J. R. Stat. Soc. Ser. B* **1979**, *41*, 113–147.
20. Bayes, T. An essay towards solving a problem in the doctrine of chances. *Philos. Trans. R. Soc. Ser. A* **1763**, *53*, 370–418.
21. Laplace, P.S. *Théorie Analytique des Probabilités*; Courcier: Paris, France, 1812.
22. Bernardo, J.M.; Berger, J.O. On the development of reference priors (with discussion). *Bayesian Stat.* **1992**, *4*, 35–60.
23. Efron, B. Why isn't everyone a Bayesian? (with discussion). *Am. Stat.* **1986**, *40*, 1–11.
24. Lindley, D.V. On the measure of the information provided by an experiment. *Ann. Math. Stat.* **1956**, *27*, 986–1005.
25. Berger, J.O.; Bernardo, J.M. Estimating a product of means. Bayesian analysis with reference priors. *J. Am. Stat. Assoc.* **1989**, *84*, 200–207.
26. Lindley, D.V. Fiducial distributions and Bayes' theorem. *J. R. Stat. Soc. Ser. B* **1958**, *20*, 102–107.
27. Datta, G.S.; Mukerjee, R. *Probability Matching Priors: Higher Order Asymptotics*; Springer: New York, NY, USA, 2004.
28. Welch, B.; Peers, H.W. On formulae for confidence points based on integrals of weighted likelihoods. *J. R. Stat. Soc. Ser. B* **1963**, *25*, 318–329.
29. Ghosh, M. Objective priors: an introduction for frequentists. *J. Stat. Sci.* **2011**, *26*, 187–202.
30. Mukerjee, R.; Dey, D.K. Frequentist validity of posterior quantiles in the presence of a nuisance parameter: Higher order asymptotics. *Biometrika* **1993**, *80*, 499–505.
31. Mukerjee, R.; Ghosh, M. Second-order probability matching priors. *Biometrika* **1997**, *84*, 970–975.
32. Liseo, B.; Loperfido, N. A note on reference priors for the scalar skew-normal distribution. *J. Stat. Plan. Inference* **2006**, *136*, 373–389.
33. Cabras, S.; Racugno, W.; Castellanos, M.E.; Ventura, L. A matching prior for the shape parameter of the skew-normal distribution. *Scand. J. Stat.* **2012**, *39*, 236–247.
34. Canale, A.; Scarpa, B. Informative Bayesian inference for the skew-normal distribution. *arXiv* **2013**, arXiv:1305.3080.
35. Bayes, C.; Branco, E. Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *J. Stat. Plan. Inference* **2007**, *21*, 141–163.
36. Chaibub Neto, E.; Branco, M.D. *Bayesian Reference Analysis for Binomial Calibration Problem*; Technical Report; IME-USP: São Paulo, Brazil, 2003; p. 12.
37. Barndorff-Nielsen, O.E. On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **1983**, *70*, 343–365.
38. Pace, L.; Salvan, A.; Ventura, L. Likelihood based discrimination between separate regression models. *J. Stat. Plan. Inference* **2006**, *136*, 3539–3553.
39. Ventura, L.; Cabras, S.; Racugno, W. Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. *J. Am. Stat. Assoc.* **2009**, *104*, 768–774.
40. Arellano-Valle, R.B.; Azzalini, A. On the unification of families of skew-normal distributions. *Scand. J. Stat.* **2006**, *33*, 561–574.
41. Canale, A.; Pagui, E.C.K., Scarpa, B. Bayesian modelling of university first-year students' grades after placement test. *J. Appl. Stat.* **2016**, *43*, 3015–3029.
42. Chib, S.; Jeliazkov, I. Marginal likelihood from the Metropolis-Hastings output. *J. Am. Stat. Assoc.* **2001**, *96*, 270–291.

43. Naranjo, L.; Perez, C.J.; Martin, J. Bayesian analysis of a skewed exponential power distribution. In Proceedings of the Computational Statistics 2012, Limassol, Cyprus, 27–31 August 2012; Colubi, A., Ed.; Physica-Verlag (Springer) Publishing: Heidelberg, Germany, 2012; pp. 641–652.
44. Hossianzadeh, A.; Zare, K. Bayesian analysis of discrete skewed Laplace distribution. *J. Mod. Appl. Stat. Methods* **2016**, *15*, 696–702.
45. Fernández, C.; Steel, M.F.J. Multivariate Student t regression models: pitfalls and inference. *Biometrika* **1999**, *86*, 153–167.
46. Sartori, N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *J. Stat. Plan. Inference* **2008**, *136*, 4259–4275.
47. Branco, E.; Genton, M.; Liseo, B. Objective Bayesian analysis of skew- t distributions. *Scand. J. Stat.* **2012**, *40*, 63–85.
48. Rubio, F.J.; Liseo, B. On the independence Jeffreys prior for skew-symmetric models. *Stat. Probab. Lett.* **2014**, *85*, 91–97.
49. Dette, H.; Ley, C.; Rubio, F.J. Natural (non-)informative priors for skew-symmetric distributions. *Scand. J. Stat.* **2018**, *45*, 405–420.
50. Rubio, F.J.; Steel, J.M. Bayesian modelling of skewness and kurtosis with two-piece scale and shape distributions. *Electron. J. Stat.* **2015**, *9*, 1884–1912.
51. Rubio, F.J.; Genton, M. Bayesian linear regression with skew-symmetric error distributions with applications to survival analysis. *Stat. Med.* **2016**, *35*, 2441–2454.
52. Liseo, B.; Parisi, A. Bayesian inference for the multivariate skew-normal model: A population Monte Carlo approach. *Comput. Stat. Data Anal.* **2013**, *63*, 125–138.
53. Parisi, A.; Liseo, B. Objective Bayesian analysis for the multivariate skew- t model. *Stat. Methods Appl.* **2017**, *27*, 277–295.
54. Panagiotelisa, A.; Smith, M. Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions. *Int. J. Forecast.* **2008**, *24*, 710–727.
55. Frühwirth-Schnatter, S.; Pyne, S. Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew- t distributions. *Biostatistics* **2010**, *11*, 317–336.
56. Sahu, S.K.; Dey, D.K.; Branco, M.D. A new class of multivariate skew distributions with applications to Bayesian regression models. *Can. J. Stat.* **2003**, *31*, 129–150.
57. Arellano-Valle, R.B.; Bolfarine, H.; Lachos, V.H. Bayesian inference for skew-normal linear mixed models. *J. Appl. Stat.* **2007**, *34*, 663–682.
58. Samworth, R.J. Recent progress in log-concave density estimation. *Stat. Sci.* **2018**, *33*, 493–509.
59. Liseo, B.; Petrella, L.; Salinetti, G. Block unimodality for multivariate Bayesian robustness. *J. Ital. Stat. Soc.* **1993**, *2*, 55–71.

