On the Independence Jeffreys prior for skew–symmetric models with applications

Francisco Javier Rubio* and Brunero Liseo†

May 11, 2014

Abstract

We study the Jeffreys prior of the skewness parameter of a general class of scalar skew–symmetric models. It is shown that this prior is symmetric about 0, proper, and with tails $O(\lambda^{-3/2})$ under mild regularity conditions. We also calculate the independence Jeffreys prior for the case with unknown location and scale parameters. Sufficient conditions for the existence of the corresponding posterior distribution are investigated for the case when the sampling model belongs to the family of skew–symmetric scale mixtures of normal distributions. The usefulness of these results is illustrated using the skew–logistic model and two applications with real data.

Key Words: Binary regression; coverage; posterior existence; skewness; stress-strength model.

1 Introduction

The need for modelling data presenting departures from symmetry has fostered the development of distributions that can capture skewness. A popular method to produce this sort of distributions consists of adding a parameter that controls skewness to a symmetric distribution. In this line, Azzalini (1985) proposed a transformation to produce an asymmetric normal density, termed skew-normal, as follows

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

(1)

where $y \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, $\lambda \in \mathbb{R}$, $\phi$ is the standard normal probability density function (PDF), and $\Phi$ is the standard normal cumulative distribution function (CDF). The parameter $\lambda$

---

*Universidad de Granada, Departamento de Estadística, Avogadro 8, 18071 Granada, Spain. E-mail: francesco.rubio@edu.ugr.es
†Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK. E-mail: francisco.rubio@warwick.ac.uk

†MEMOTEF, Sapienza Università di Roma. E-mail: brunero.liseo@uniroma1.it
is often interpreted as a skewness parameter given that the density (1) is asymmetric for $\lambda \neq 0$, and it reduces to the normal PDF for $\lambda = 0$. Subsequently, Wang et al. (2004) showed that, in particular, this method can be extended to any continuous symmetric density $f$, with support on $\mathbb{R}$ and mode at 0, through the transformation

$$ss(y; \mu, \sigma, \lambda) = \frac{2}{\sigma} f \left( \frac{y - \mu}{\sigma} \right) \pi \left( \lambda \frac{y - \mu}{\sigma} \right),$$

(2)

where $\pi$, termed *skewing function*, is a function that satisfies $0 \leq \pi(y) \leq 1$, and $\pi(-y) = 1 - \pi(y)$. It follows, then, that any symmetric CDF can be used as a skewing function. Several choices for $f$ and $\pi$ have been explored in the literature, such as the power exponential distribution with power $\delta \in \mathbb{R}_+$ (Azzalini, 1986), the Student-$t$ distribution with $\nu \in \mathbb{R}_+$ degrees of freedom (Azzalini and Capitanio, 2003), the logistic distribution (Nadarajah, 2009), among others. Distributions obtained by means of this method are called skew–symmetric distributions. These distributions are widely used nowadays in several contexts such as binary regression (Bazán et al., 2010), meta–analysis (Guolo, 2012), data fitting (Branco et al., 2012), among many others.

It has been found that several skew–symmetric models present inferential issues. For instance, Azzalini (1985) showed that the Fisher information matrix of the parameters $(\mu, \sigma, \lambda)$ is singular at $\lambda = 0$ for the skew–normal sampling model. In addition, the maximum likelihood estimator of the parameter $\lambda$ can be $\infty$ with positive probability. The cases with infinite estimators are more commonly found in small and moderate samples. These inferential issues are present in other skew–symmetric models (Hallin and Ley, 2012). Some authors have proposed the use of the Bayesian approach in order to avoid these inferential problems (Liseo and Loperfido, 2006; Branco et al., 2012). In Bayesian practice it is often of interest to employ noninformative priors given that they typically produce posterior inference with appealing frequentist properties. However, due to the singularity of the Fisher information matrix at $\lambda = 0$ of some skew–symmetric models, the use of the Jeffreys–rule prior, which is defined as the square root of the determinant of the Fisher information matrix, has been avoided in this kind of models. In addition, the calculation of this sort of prior is typically cumbersome. Reference priors, which are another kind of noninformative priors, have been studied for the skew–normal and the skew Student–$t$ models in Liseo and Loperfido (2006) and Branco et al. (2012). An alternative noninformative prior is the independence Jeffreys prior. This prior is constructed as the product of the Jeffreys priors for each parameter, while treating the remaining parameters as fixed.

In this paper, we study the independence Jeffreys prior associated to the class of skew–symmetric distributions obtained by using a CDF as a skewing function in (2). In Section 2 we analyse the Jeffreys prior of the skewness parameter $\lambda$ in skew–symmetric models without location and scale parameters. We show that this prior is proper, symmetric about 0, and with
tails $O(\lambda^{-\frac{3}{2}})$ under rather mild regularity conditions. Using these results, we construct the independence Jeffreys prior for the general model with location and scale parameters. In Section 3 we obtain easy to check sufficient conditions for the propriety of the posterior distribution when the sampling model $f$ in (2) belongs to the family of scale mixtures of normal distributions. The case of samples containing censored observations is covered as well. In Section 4 we present the use of these results on the skew–logistic distribution. In Section 5 we illustrate the use of these models in the context of binary regression and stress–strength models. We conclude with some discussion and extensions of this work in Section 6.

2 Independence Jeffreys prior for univariate skew–symmetric models

Throughout we focus on the study of skew–symmetric models of the type

$$s(y; \mu, \sigma, \lambda) = \frac{2}{\sigma} f \left( \frac{y - \mu}{\sigma} \right) G \left( \lambda \frac{y - \mu}{\sigma} \right),$$

(3)

where $f$ is a continuous symmetric density function with support on $\mathbb{R}$, and $G$ is a CDF with continuous symmetric density $g$ with support on $\mathbb{R}$. This structure covers many cases of practical interest such as the skew–normal distribution (Azzalini, 1985), the skew–$t$ distribution (Azzalini and Capitanio, 2003), the skew–logistic distribution (Nadarajah, 2009), among many others.

Consider first the particular skew–symmetric model (3) without location and scale parameters, this is, assuming that $\mu = 0$ and $\sigma = 1$. Recall that the Fisher information of the parameter $\lambda$ is defined as

$$I(\lambda) = \int_{\mathbb{R}} \left[ \frac{\partial \log s(y; 0, 1, \lambda)}{\partial \lambda} \right]^2 s(y; 0, 1, \lambda) dy.$$

The Jeffreys prior of the parameter $\lambda$ is defined, up to a proportionality constant, as the square root of the Fisher information $I(\lambda)$, that is, $\pi(\lambda) \propto \sqrt{I(\lambda)}$. The following result characterises the cases where this prior is well–defined at $\lambda = 0$.

**Remark 1** The Fisher information of $\lambda$ associated to model (3), and consequently the Jeffreys prior of $\lambda$, is well–defined at $\lambda = 0$ if and only if the second moment of $f$ exists.

**Proof.** See appendix.

Particular cases of Remark I have already been reported in the literature. For instance, Branco et al. (2012) report the presence of a pole at $\lambda = 0$ in the Jeffreys prior of $\lambda$ for the skew Student–$t$ model with $\nu \leq 2$ degrees of freedom. Remark I shows that this feature is
present in many other skew–symmetric models, and that this sort of singularity is linked to the existence of the moments of the underlying symmetric density f.

Liseo and Loperfido (2006) and Branco et al. (2012) show that the Jeffreys priors of the parameter λ, for the cases where f and g are normal or Student-t distributions, are proper, decreasing in |λ|, and with tails $O(|\lambda|^{-\frac{3}{2}})$. Their proofs rely upon basic properties of these models, which suggests that there may be other models that lead to a Jeffreys prior of λ with the same properties under some reasonable regularity conditions. In order to establish this result, we introduce the following set of sufficient conditions.

**Condition S** Let f and g be continuous density functions with support on $\mathbb{R}$ that satisfy the following conditions:

(i) $f$ and $g$ are symmetric about 0.

(ii) $f$ is unimodal and there exists a finite constant $M$ such that $0 < f(x) < M$, for all $x \in \mathbb{R}$.

(iii) The Fisher information $I(\lambda) < \infty$, for all $\lambda \in \mathbb{R}$.

Conditions S.i–S.ii include models of practical interest, such as: the normal distribution, the Student-t distribution, the exponential power distribution, the logistic distribution, among many others. Condition S.iii is simply used to restrict ourselves to those cases where the Jeffreys prior of λ exists. Theorem 1 provides conditions for the finiteness of $I(0)$, however, for $\lambda \neq 0$ the finiteness of $I(\lambda)$ may require a case by case analysis (for a more detailed study of this point we refer the reader to Hallin and Ley, 2012). The following theorem shows that the results in Liseo and Loperfido (2006) and Branco et al. (2012) can be extended to the family of distributions that satisfies Condition S.

**Theorem 1** Let f and g be density functions that satisfy Condition S. Then, the Jeffreys prior of λ associated to model (3) with $(\mu, \sigma) = (0, 1)$ satisfies the following:

(i) The Jeffreys prior of λ is given by

$$\pi(\lambda) \propto \sqrt{\int_0^\infty x^2 f(x) \frac{g(\lambda x)^2}{G(\lambda x)[1 - G(\lambda x)]} dx}. \quad (4)$$

(ii) $\pi(\lambda)$ is symmetric about 0.

(iii) The tails of $\pi(\lambda)$ are of order $O(|\lambda|^{-\frac{3}{2}})$.

(iv) $\pi(\lambda)$ is integrable.

**Proof.** See appendix
Based on the tail behaviour, symmetry, and properness of the Jeffreys prior of $\lambda$ shown for
the skew–normal model in Liseo and Loperfido (2006), Bayes and Branco (2007) proposed an
approximation to this prior using a Student-$t$ distribution with $\nu = 1/2$ degrees of freedom and
an empirical choice for the scale parameter ($\pi/2$). Branco et al. (2012) also proposed a similar
approximation for the Jeffreys prior of $\lambda$ of the skew Student-$t$ model. Theorem 1 shows that
this approximation might be reasonable in other cases as well. However, the quality of this
approximation and the choice for the scale parameter seem to require a case by case analysis.

In Section 4 we show that this approximation is reasonable for a skew–logistic sampling model.

Condition S.iii can be relaxed to those cases where $I(\lambda) < \infty$ for all $\lambda \neq 0$, possibly
leading to an undefined Jeffreys prior at $\lambda = 0$ such as those models studied in Branco et al.
(2012). The results (ii)–(iii) in Theorem 1 are valid under these relaxed assumptions given that
they can be proved using essentially the same technique. The results in Branco et al. (2012)
also suggest that it is possible to obtain a proper Jeffreys prior $\pi(\lambda)$ for some sampling models
despite the singularity of the Fisher information at $\lambda = 0$. However, the use of priors containing
singularities might be less appealing to practitioners.

We now study the independence Jeffreys prior associated to the skew–symmetric model (3)
including location and scale parameters. In the next section we also show that this prior leads
to a proper posterior distribution under mild conditions.

**Theorem 2**

The independence Jeffreys prior of $(\mu, \sigma, \lambda)$ corresponding to a skew–symmetric model (3) that
satisfies Condition S is given by

$$
\pi_I(\mu, \sigma, \lambda) \propto \frac{1}{\sigma^2} \pi(\lambda),
$$

where $\pi(\lambda)$ is the function defined in (4).

**Proof.** See appendix

### 3 Existence of the posterior

In this section, we provide sufficient conditions for the existence of the posterior distribution
under the use of the priors studied in the previous section.

**Corollary 1** Let $y = (y_1, \ldots, y_n)$ be an i.i.d. sample from a skew–symmetric model (3) with
$(\mu, \sigma) = (0, 1)$ that satisfies Condition S. Then, the corresponding posterior distribution of this
parameter is proper.

**Proof.** The result follows by the properness of (4) under Condition S.
Liseo and Loperfido (2006) show that (4) is proper for the skew–normal sampling model (Azzalini, 1985); and Branco et al. (2012) show that this is also the case for the prior (4) associated to a skew–symmetric Student–t sampling model (Azzalini and Capitanio, 2003). In Section 4 we show that the prior (4) associated to a skew–logistic sampling model (Nadarajah, 2009) is also proper.

For the general model (3), with unknown location and scale parameters, the independence Jeffreys prior (5) is improper. Then, in order to conduct valid Bayesian inference it is necessary to check conditions for the existence of the corresponding posterior distribution. The following result provides sufficient conditions for the existence of the posterior distribution for the case when f is a scale mixture of normal distributions. The family of scale mixtures of normals contains important models such as the Normal distribution, the Student–t distribution with ν degrees of freedom, the exponential power distribution with power $1 ≤ δ ≤ 2$, the logistic distribution, the symmetric α-stable family of distributions, among others.

**Theorem 3** Let $y = (y_1, \ldots, y_n)$ be an i.i.d. sample from a skew–symmetric model (3) that satisfies Condition S. Suppose also that f is a scale mixture of normals. Then, the posterior distribution of $(\mu, \sigma, \lambda)$ associated to the independence Jeffreys prior (5) is proper if $n ≥ 2$ and all the observations are different.

**Proof.** See appendix

Since the skew–symmetric distributions of interest are continuous, it follows that the probability of obtaining repeated observations is zero. This implies that we can conduct valid Bayesian inference based on this prior whenever $n ≥ 2$ for almost any sample. In the Appendix we show that the proof of the propriety of the posterior distribution of $(\mu, \sigma, \lambda)$, under the assumptions in Theorem 3 can be reduced to proving the propriety of the posterior distribution in the symmetric case. This is, assuming that y is an i.i.d. sample from a scale mixture of normals f with location and scale parameters $(\mu, \sigma)$ and adopting the prior structure $\pi(\mu, \sigma) \propto \sigma^{-1}$. The propriety of the posterior distribution under the latter assumptions is studied in Fernández and Steel (1998), who also show that the presence of repeated observations in the sample may destroy the existence of the posterior distribution for some scale mixture of normal sampling models. They also present sufficient conditions for the propriety of the posterior distribution in cases when the sample contains repeated observations. We refer the reader to Fernández and Steel (1998) for further details on this.

Another scenario of interest is when the observations are recorded as sets of positive probability due to some kind of censoring mechanism. This is, when the collected sample consists of sets $S_1, \ldots, S_n$ with $P(y_i \in S_i) > 0$, $i = 1, \ldots, n$. This framework clearly covers all kinds of interval censoring. The following result shows that the independence Jeffreys prior (5) produces a proper posterior distribution in this case as well.
Theorem 4  Let \( S_1, \ldots, S_n \) be a sample of censored observations from a skew–symmetric model (3) that satisfies Condition S. Suppose also that \( f \) is a scale mixture of normals. Then, the posterior distribution of \( (\mu, \sigma, \lambda) \) associated to the Bayesian model (3)–(5) is proper if \( n \geq 2 \) and there exist two sets, say \( S_i, S_j \), such that

\[
\inf_{y_i \in S_i, y_j \in S_j} |y_i - y_j| > 0.
\]

Proof. See appendix

This result implies that the posterior distribution of \( (\mu, \sigma, \lambda) \) exists whenever the sample of set observations contains at least two observations that do not overlap.

4  Skew–Logistic model

Nadarajah (2009) showed that an interesting member of the skew–symmetric family (3) is the skew–logistic distribution, obtained by using the logistic PDF and CDF, \( f(t) = \frac{e^{-t}}{(1 + e^{-t})^2} \) and \( G(t) = \frac{1}{1 + e^{-t}} \). The skew–logistic density can be written in closed form, after some algebra, as follows

\[
sl(y; \mu, \sigma, \lambda) = \frac{1}{4} \text{sech}^2 \left( \frac{y - \mu}{2\sigma} \right) \left[ 1 + \tanh \left( \frac{\lambda y - \mu}{2\sigma} \right) \right],
\]

where \( \tanh(\cdot) \) and \( \text{sech}(\cdot) \) represent the hyperbolic tangent and the hyperbolic secant functions, respectively. For this sampling model, the Jeffreys prior (4) can be written as indicated below:

\[
\pi(\lambda) \propto \sqrt{\int_0^{\infty} x^2 \text{sech}^2 \left( \frac{x}{2} \right) \text{sech}^2 \left( \frac{\lambda x}{2} \right) dx}.
\]

It is easy to check that (7) satisfies Condition S and therefore it is proper, as a consequence of Theorem 1. The tail behaviour, symmetry, and properness shown in this result suggest the use of a Student-\( t \) approximation, such as the one proposed in Bayes and Branco (2007) for the skew-normal model. Empirically, we have found that \( \pi(\lambda) \) can be reasonably well approximated by a Student-\( t \) distribution with \( 1/2 \) degrees of freedom and scale parameter \( 4/3 \). Figure [1](#) illustrates the quality of this approximation.

For the general skew–logistic model with unknown location and scale parameters it follows that the posterior distribution of \( (\mu, \sigma, \lambda) \) using the independence Jeffreys prior (5) is proper, given that the logistic distribution can be represented as a scale mixture of normals (Stefanski, 1991), under the conditions in Theorem 3. Consequently, the results in Theorem 4 also hold for a skew–logistic sampling model.
Figure 1: (a) Jeffreys prior of $\lambda$ (continuous line) and Student-$t$ approximation (dashed line); (b) Absolute difference between the Jeffreys prior of $\lambda$ and the Student-$t$ approximation.

### 4.1 A simulation study

In this section we analyse the empirical coverage of the 95% posterior credible intervals, defined by the 2.5th and 97.5th percentiles, using the independence Jeffreys prior \textsuperscript{[5]}. We simulate $N = 1000$ data sets of size $n = 10, 30, 100, 1000$ from the skew-logistic distribution \textsuperscript{[6]} with parameters $\mu_0 = 0$, $\sigma_0 = 1$, and $\lambda_0 = 0.5, 1, 2, 5, 10$. The simulation step is implemented by numerical inversion of the corresponding CDF and the probability integral transform. For each of these samples, a posterior sample of size 1000 was obtained using the $t$-walk algorithm \cite{Christen_and_Fox_2010} with a burn-in period of 10,000 iterations and a thinning period of 50 iterations. The proportion of 95% posterior credible intervals that contain the true value of the parameters is reported in Tables 1–4. Overall, the frequentist coverage properties of this Bayesian model are good.

| Parameter | $\lambda_0 = 0.5$ | $\lambda_0 = 1$ | $\lambda_0 = 2$ | $\lambda_0 = 5$ | $\lambda_0 = 10$ |
|-----------|-------------------|-----------------|-----------------|-----------------|-----------------|
| $\mu$     | 0.993             | 0.970           | 0.934           | 0.906           | 0.914           |
| $\sigma$  | 0.933             | 0.965           | 0.978           | 0.947           | 0.936           |
| $\lambda$ | 0.999             | 0.985           | 0.965           | 0.950           | 0.953           |

Table 1: Coverage proportions: $n = 10$. The value of $\lambda_0$ is specified in each column.

| Parameter | $\lambda_0 = 0.5$ | $\lambda_0 = 1$ | $\lambda_0 = 2$ | $\lambda_0 = 5$ | $\lambda_0 = 10$ |
|-----------|-------------------|-----------------|-----------------|-----------------|-----------------|
| $\mu$     | 0.956             | 0.948           | 0.930           | 0.909           | 0.920           |
| $\sigma$  | 0.943             | 0.968           | 0.958           | 0.930           | 0.947           |
| $\lambda$ | 0.961             | 0.959           | 0.934           | 0.932           | 0.965           |

Table 2: Coverage proportions: $n = 30$. The value of $\lambda_0$ is specified in each column.
| Parameter | $\lambda_0 = 0.5$ | $\lambda_0 = 1$ | $\lambda_0 = 2$ | $\lambda_0 = 5$ | $\lambda_0 = 10$ |
|-----------|------------------|------------------|------------------|------------------|------------------|
| $\mu$     | 0.965            | 0.952            | 0.937            | 0.920            | 0.934            |
| $\sigma$  | 0.956            | 0.981            | 0.947            | 0.938            | 0.954            |
| $\lambda$ | 0.966            | 0.960            | 0.934            | 0.918            | 0.932            |

Table 3: Coverage proportions: $n = 100$. The value of $\lambda_0$ is specified in each column.

| Parameter | $\lambda_0 = 0.5$ | $\lambda_0 = 1$ | $\lambda_0 = 2$ | $\lambda_0 = 5$ | $\lambda_0 = 10$ |
|-----------|------------------|------------------|------------------|------------------|------------------|
| $\mu$     | 0.958            | 0.946            | 0.943            | 0.947            | 0.946            |
| $\sigma$  | 0.956            | 0.946            | 0.931            | 0.953            | 0.931            |
| $\lambda$ | 0.952            | 0.950            | 0.938            | 0.944            | 0.949            |

Table 4: Coverage proportions: $n = 1000$. The value of $\lambda_0$ is specified in each column.

## 5 Applications

In this section we present two applications of the Bayesian models studied in the previous sections for skew–symmetric distributions using real data. In the first application we consider the use of two skew–symmetric distributions in the context of binary regression to produce a more robust model. Using the Jeffreys prior (4), we propose a hierarchical proper prior based on a modification of the Jeffreys prior proposed in Chen et al. (2008) for binary regression models. We show that this prior is scale-invariant with respect to the covariates. The second application consists of a Bayesian model using the prior (5) in the context of stress–strength models with dependent observations.

Simulations from the corresponding posterior distributions are obtained using a Metropolis-Hastings algorithm with a burn-in period of 50,000 iterations and a thinning period of 100 iterations. Model comparison is conducted using AIC, BIC and Bayes factors. The Bayes factors are calculated using an importance sampling technique. R codes for these examples are available upon request.

### 5.1 Binary regression

Binary and binomial observations are common in contexts such as biology, medicine, quality control, among others (see e.g. Collet, 1999 for a good survey of this). Generalised linear models are a useful tool for modelling this sort of observations given that the probability of observing $y$ successes (failures) of a binomial random variable $Y \in \{0, 1, \ldots, n\}$ can be related to a certain set of covariates $\mathbf{x} = (1, x_1, \ldots, x_k)^T$, through the model

$$
\mathbb{P}(Y = y|\mathbf{x}, \beta, \theta) = \binom{n}{y} F(\mathbf{x}^\top \beta; \theta)^y (1 - F(\mathbf{x}^\top \beta; \theta))^{n-y},
$$

(8)
\[ \beta = (\beta_0, ..., \beta_k)^\top \] is a vector of regression coefficients, \( F(\cdot; \theta) \) is a univariate distribution function with shape parameter \( \theta \in \Theta \), and \( F^{-1} \) is called the link function. The most common links correspond to the cases where \( F \) is the logistic distribution (logit) or the standard normal distribution (probit), which are often referred to as the canonical links. It has been found that these links do not always provide a good fit (see e.g. Aranda-Ordaz, 1981), and also that link misspecification can affect the inference about the parameters (Czado and Santner, 1992).

An approach for constructing more robust links consists of using a flexible distribution as a link function. In this line we can find a number of references such as Copenhaver and Mielke (1977), Chen et al. (1999), Basu and Mukhopadhyay (2000), Kim et al. (2008), Bazán et al. (2010), and Wang and Dey (2010). In this application, we propose a hierarchical prior based on the Jeffreys prior (4) for the generalised linear model (8) with a skew-symmetric link as in (3), as described below.

Let \( y_i \sim \text{Binomial} \{ n_i, S(x_i^\top \beta; \lambda) \} \), \( i = 1, \ldots, m \), where \( x_i = (1, x_{i1}, ..., x_{ik})^\top \) is a vector of covariates, \( \beta = (\beta_0, ..., \beta_k)^\top \) is a vector of regression coefficients, and \( S^{-1} \) is the link function defined by the skew-symmetric distribution (3) with \( (\mu, \sigma) = (0, 1) \) and unknown skewness parameter \( \lambda \in \mathbb{R} \). Then, the likelihood function of the parameters \( (\beta, \lambda) \) is given by

\[
\mathcal{L}(\beta, \lambda|X, y) \propto \prod_{i=1}^{m} S(x_i^\top \beta; \lambda)^{y_i} [1 - S(x_i^\top \beta; \lambda)]^{n_i-y_i}. \tag{9}
\]

where \( X = (x_1, x_2, ..., x_m)^\top \) is the design matrix.

We propose a hierarchical prior structure for the parameters of model (9), based on a modification of the Jeffreys prior presented in Chen et al. (2008). We adopt the hierarchical prior structure

\[
\pi(\beta, \lambda) = \pi(\beta|\lambda)\pi(\lambda), \tag{10}
\]

where

\[
\pi(\beta|\lambda) \propto \det \left[ X^\top W(\beta, \lambda) X \right]^{\frac{1}{2}}, \tag{11}
\]

\( \det \left[ X^\top W(\beta, \lambda) X \right] \) denotes the determinant of the matrix \( X^\top W(\beta, \lambda) X \),

\[
W(\beta, \lambda) = \text{diag}(w_1(\beta, \lambda), ..., w_m(\beta, \lambda)),
\]

\[
w_i(\beta, \lambda) = \frac{n_i S(x_i^\top \beta; \lambda)^2}{S(x_i^\top \beta; \lambda)\{1 - S(x_i^\top \beta; \lambda)\}},
\]

and \( \pi(\lambda) \) is given by (4). Since the prior (11) is proper for any \( \lambda \) fixed (Chen et al., 2008), it follows that the hierarchical structure (10) is also proper if the conditions in Theorem 1 are
satisfied. For example, under the use of the skew–normal or the skew–logistic link. In addition, for the skew–normal and the skew–logistic links we can employ the Student-\(t\) approximations of \(\pi(\lambda)\) described in Section 4 in order to facilitate the implementation of prior (10). The proposed prior also presents the following invariance property.

**Remark 2** Define

\[
C_0(X) = \int_{\mathbb{R}} \int_{\mathbb{R}^{k+1}} \det [X^TW(\beta, \lambda)X]^{\frac{1}{2}} \pi(\lambda) d\beta d\lambda,
\]

\[
C(X) = \int_{\mathbb{R}} \int_{\mathbb{R}^{k+1}} L(\beta, \lambda|X, y) \det [X^TW(\beta, \lambda)X]^{\frac{1}{2}} \pi(\lambda) d\beta d\lambda,
\]

and \(V = \text{diag}(1, v_1, \ldots, v_k)\) with \(v_1 > 0, \ldots, v_k > 0\). Then, \(C_0(X) = C_0(XV)\) and \(C(X) = C(XV)\).

**Proof.** The result is a consequence of Theorem 5 from Chen et al. (2008).

As discussed by Chen et al. (2008), this result implies that \(C_0(X)\) and \(C(X)\) are scale-invariant with respect to the covariates, which is a desirable property in Bayesian modelling, particularly for conducting variable selection.

In order to illustrate the use of the proposed Bayesian model, we analyse the popular data set reported in Bliss (1935). The aim of this experiment was to model the response of confused flour beetles to gaseous carbon disulphide. Aranda-Ordaz (1981) mentioned that this data set presents asymmetric departures from the logistic model, hence the use of a skew–symmetric link seems appropriate. Figure 2 shows the predictive dose-response curves associated to 4 links: the skew–logistic link with the prior (10), the skew–normal link with the prior (10), the logit link together with the prior \(\pi(\beta|\lambda = 0)\) which corresponds to the Jeffreys prior described in Chen et al. (2008), and the probit link together with the prior \(\pi(\beta|\lambda = 0)\) which again corresponds to the Jeffreys prior described in Chen et al. (2008). The Bayes factors of the different links against the skew–logistic link, AIC, and BIC values, shown in Table 6, favour the use of an asymmetric model and slightly favour the skew–logistic link over the other competitors. Table 5 shows the predicted observations with these models obtained by multiplying the number of subjects \(n_j\) by the predicted probability at the corresponding dose level. This table suggests a better fit of the asymmetric models.
5.2 Stress-strength models

Let \((X, Y)\) be a pair of absolutely continuous random variables such that \(Z = X - Y\) has skew-symmetric distribution \((3)\). It follows that

\[
\theta = \mathbb{P}(X < Y) = \mathbb{P}(X - Y < 0) = S(0; \mu, \sigma, \lambda). \tag{12}
\]
The parameter $\theta$ is called the stress-strength coefficient and it has been applied in several contexts (see Rubio and Steel, 2013). Note that, unlike Rubio and Steel (2013) who model the joint distribution of $(X, Y)$, here we are making distributional assumptions on the difference $Z$. In a Bayesian context, if we have a sample of paired observations $(x_i, y_i), i = 1, \ldots, n,$ from $(X, Y)$, then we can obtain a sample from the posterior distribution of $\theta$ by first obtaining a sample from the posterior distribution of $(\mu, \sigma, \lambda)$, obtained in turn by using the sample of differences \(z_i = x_i - y_i\), and then by plugging these values into (12). It is worth pointing out that this approach can only be applied when the sample is complete and it does not contain censored observations. For a more general approach that covers these cases see Rubio and Steel (2013).

We consider the data set presented in Venkatraman and Begg (1996) which contains 72 lesion scores obtained using both a clinical scheme without a dermoscope ($X$ Test), and a dermoscopic scoring scheme ($Y$ Test). Their main interest was assessing the information provided by the use of the dermoscope. Here, we analyse the subset of $n = 51$ non-diseased patients. The sample skewness of the differences \(z_i = x_i - y_i\) is 0.57, and Figure 3a shows the histogram of these differences. These features suggest the need for using an asymmetric model. For this purpose, we compare the performance of the skew–normal distribution (1) and the skew-logistic distribution (6) together with the independence Jeffreys prior (5). Since the sample of differences does not contain repeated observations, it follows that the posterior distribution of $(\mu, \sigma, \lambda)$ is proper and consequently the posterior of $\theta$ is well-defined for both sampling models. Figure 3b shows the posterior distributions of $\theta$. We can observe that the inference on $\theta$ under both distributional assumptions is very similar. A 95\% posterior credible interval for $\theta$ for the skew–logistic model is (0.54, 0.76), while the corresponding interval for the skew–normal model is (0.52, 0.74). These intervals do not contain the value $\theta = 0.5$, therefore this approach leads to similar results as those obtained in Rubio and Steel (2013). The Bayes factors of the models of interest against the skew–logistic model, AIC, and BIC values, shown in Table 7, slightly favour the skew–normal model.

| Model          | AIC    | BIC    | Bayes factor |
|----------------|--------|--------|--------------|
| skew-logistic  | 176.17 | 181.96 | 1            |
| skew-normal    | 175.01 | 180.80 | 1.5          |

Table 7: Melanoma data: Model comparison.
Figure 3: Melanoma data: (a) Histogram of the difference of the observations; (b) Posterior distribution of $\theta$, skew-logistic model (continuous line), skew–normal model (dashed line).

6 Discussion

We have studied the Jeffreys prior of the skewness parameter of a general class of scalar skew–symmetric models as well as the independence Jeffreys prior for the same class of models with unknown location and scale parameters. We have shown that this sort of priors has appealing properties such as symmetry, properness, and identifiable tail behaviour that allow in many cases a tractable approximation that facilitate their implementation. We have also presented easy to check conditions for the existence of the posterior distribution for a general subclass of skew–symmetric sampling models. Given that the prior on the skewness parameter has heavy tails, $O(|\lambda|^{-3/2})$, it is expected to obtain good frequentist properties of the corresponding Bayesian models since heavy–tailed priors are usually employed as “vague priors”. This feature was illustrated using a simulation study in Section 4.1.

One of the unpleasant properties of the priors studied in this paper is that they are well–defined at $\lambda = 0$ only when the transformed symmetric density $f$ has a finite second moment. This can be considered as a limitation for the use of these Bayesian models. However, things must be considered in perspective. In some cases where our prior cannot be defined at zero, maximum likelihood estimation fails also, and we do not know any other broadly satisfactory alternative method. Inspired by the structure of the independence Jeffreys prior (5) we can construct a more general benchmark prior for skew–symmetric models as follows

$$
\pi(\mu, \sigma, \lambda) \propto \sigma^{-1} p(\lambda),
$$

where $p(\lambda)$ is any proper prior. Using this prior structure, the corresponding posterior distribution is proper for any skewing function $G$ if $f$ in (5) is a scale mixture of normals, the sample size $n \geq 2$, and all the observations are different. The proof of this result is similar to that of Theorem 3. The study of appropriate choices for $p(\lambda)$ is a matter of further research.
Extensions to this work include the study of the independence Jeffreys prior of multivariate skew–symmetric models as well as the propriety of the corresponding posterior distributions. Also, the technique employed in the proof of the propriety of the posterior distribution in Theorem 3 can be extended, with some care, to the use of the prior structures (5) and (13) in the context of linear regression models with skew–symmetric residual errors using the results in Fernández and Steel (2000). This opens the door to a variety of applications.
Appendix

Proof of Remark

First, after some algebra we get

\[ \frac{\partial \log[s(x; 0, 1, \lambda)]}{\partial \lambda} = \frac{xg(\lambda x)}{G(\lambda x)}. \]

Using this calculation it follows that the Fisher information of \( \lambda \) can be written as follows

\[ I(\lambda) = \int_{-\infty}^{\infty} \left[ \frac{xg(\lambda x)}{G(\lambda x)} \right]^2 2f(x)G(\lambda x) dx = 2 \int_{-\infty}^{\infty} x^2f(x)\frac{g(\lambda x)^2}{G(\lambda x)} dx. \]

Therefore \( I(0) = 4g(0)^2 \int_{\mathbb{R}} x^2f(x)dx \), which establishes the relationship of \( I(0) \) and the finiteness of the second moment of \( f \).

Proof of Theorem

(i) Splitting the integration range in the expression of \( I(\lambda) \), using the symmetry of \( f \) and \( g \), and the equality \( G(-x) = 1 - G(x) \) we obtain

\[
I(\lambda) = 2 \int_{-\infty}^{0} x^2 f(x) \frac{g(\lambda x)^2}{G(\lambda x)} dx + 2 \int_{0}^{\infty} x^2 f(x) \frac{g(\lambda x)^2}{G(\lambda x)} dx \\
= 2 \int_{0}^{\infty} x^2 f(x) \frac{g(\lambda x)^2}{G(\lambda x)[1 - G(\lambda x)]} dx.
\]

The result follows by taking the square root of the latter expression.

(ii) The symmetry of \( \pi(\lambda) \) follows from that of \( I(\lambda) \), which in turns is a consequence of the symmetry of \( f \) and \( g \), and the equality \( G(-x) = 1 - G(x) \).

(iii) Using the fact that \( f \) is upper bounded we can obtain the following upper bound for the Fisher information of \( \lambda \)

\[
I(\lambda) = 2 \int_{0}^{\infty} x^2 f(x) \frac{g(\lambda x)^2}{G(\lambda x)[1 - G(\lambda x)]} dx \leq 2M \int_{0}^{\infty} x^2 \frac{g(\lambda x)^2}{G(\lambda x)[1 - G(\lambda x)]} dx.
\]

Now, consider the change of variable \( u = \lambda x \), with \( \lambda > 0 \), then we can rewrite this upper bound as follows

\[
I(\lambda) \leq \frac{2M}{\lambda^3} \int_{0}^{\infty} u^2 \frac{g(u)^2}{G(u)[1 - G(u)]} du,
\]

16
After the same change of variable \( u = \lambda x \) we can rewrite the Fisher information as

\[
I(\lambda) = \frac{2}{\lambda^3} \int_0^\infty u^2 f \left( \frac{u}{\lambda} \right) \frac{g^2(u)}{G(u)[1 - G(u)]} du.
\]

Note that for \( \lambda \geq L > 0 \), \( f \left( \frac{u}{\lambda} \right) \geq f \left( \frac{u}{L} \right) \), for all \( u > 0 \). Then

\[
I(\lambda) \geq \frac{2}{\lambda^3} \int_0^\infty u^2 f \left( \frac{u}{L} \right) \frac{g^2(u)}{G(u)[1 - G(u)]} du, \quad \text{(16)}
\]

for \( \lambda > L > 0 \). Therefore, by combining (15) and (16) it follows that \( I(\lambda) \) has tails of order \( O(|\lambda|^{-3}) \) which implies that \( \pi(\lambda) \) has tails \( O(|\lambda|^{-3/2}) \).

(iv) Let \( a > 0 \), using that \( \pi(0) < \infty \) (as a consequence of Condition S.iii), and that \( \pi(\lambda) \) is finite on \([0, \infty)\) by assumption, it follows that

\[
\int_0^a \pi(\lambda) d\lambda < \infty. \quad \text{(17)}
\]

Now, using (15) we have that

\[
\int_a^\infty \pi(\lambda) d\lambda \leq C \int_a^\infty \frac{1}{\lambda^2} d\lambda < \infty, \quad \text{(18)}
\]

where \( C \) is a positive constant. Combining (17) and (18), and using that \( \pi(\lambda) \) is symmetric about 0, the properness of \( \pi(\lambda) \) follows.

**Proof of Theorem 2**

The diagonal entries of the Fisher information matrix of \((\mu, \sigma, \lambda)\) are given by

\[
\begin{align*}
I_{\mu,\mu} &= \frac{2}{\sigma^2} \int_{-\infty}^{\infty} \left[ f'(t) \left( \frac{f'(t)}{f(t)} + \lambda \frac{g(\lambda t)}{G(\lambda t)} \right)^2 \right] f(t)G(\lambda t) dt, \\
I_{\sigma,\sigma} &= \frac{2}{\sigma^2} \int_{-\infty}^{\infty} \left[ 1 + f'(t) \left( \frac{f'(t)}{f(t)} + \lambda \frac{g(\lambda t)}{G(\lambda t)} \right) \right]^2 f(t)G(\lambda t) dt, \\
I_{\lambda,\lambda} &= \int_{-\infty}^{\infty} t^2 f(t) \frac{g(\lambda t)^2}{G^2(\lambda t)} dt.
\end{align*}
\]

Note that \( I_{\mu,\mu} \) does not depend on \( \mu \); \( I_{\sigma,\sigma} \) depends on \( \sigma \) through the factor \( \sigma^{-2} \); and \( I_{\lambda,\lambda} \) coincides with \( I(\lambda) \), defined in the proof of Theorem 1. The result follows from these observations and by the definition of the independence Jeffreys prior.
Proof of Theorem 3

First of all, recall that a posterior distribution is proper whenever the marginal distribution $p(y_1, \ldots, y_n) < \infty$ (Fernández and Steel, 1998). Now, note that

$$s(y; \mu, \sigma, \lambda) \leq \frac{2}{\sigma} f \left( \frac{y - \mu}{\sigma} \right).$$ \hspace{1cm} (19)

Then, if follows that

$$p(y_1, \ldots, y_n) = \int_{R^+} \int_{R^+} \prod_{j=1}^{n} \left[ s(y_j; \mu, \sigma, \lambda) \right] \frac{\pi(\lambda)}{\sigma} d\mu d\sigma d\lambda$$

$$\leq \int_{R^+} \int_{R^+} \left[ \prod_{j=1}^{n} \frac{2}{\sigma} f \left( \frac{y_j - \mu}{\sigma} \right) \right] \frac{1}{\sigma} d\mu d\sigma \int_{R^+} \pi(\lambda) d\lambda$$

By Theorem 1, we have that $\pi(\lambda)$ is proper. Then, it follows that the posterior distribution of $(\mu, \sigma, \lambda)$ exists whenever the posterior distribution of $(\mu, \sigma)$ exists for a scale mixture of normals sampling model and the prior $\pi(\mu, \sigma) \propto \sigma^{-1}$. The properness of the latter, for $n \geq 2$, follows by Theorem 1 from Fernández and Steel (1998).

Proof of Theorem 4

The result follows again by using inequality (19) and Theorem 4 from Fernández and Steel (1998).

References

Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika* 68: 357–363.

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

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica* 46: 199–208.

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

Bayes, C. L. and Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics* 21: 141–163.
Basu, S. and Mukhopadhyay, S. (2000). Bayesian analysis of binary regression using symmetric and asymmetric links. *Sankhya: The Indian Journal of Statistics*, Series B 62: 372–387.

Bazán, J. L., Bolfarine, H. and Branco, M. (2010). A framework for skew-probit links in binary regression. *Communications in Statistics – Theory and Methods* 39: 678–697.

Bliss, C. I. (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology* 22: 134-167.

Branco, M. D., Genton, M. G. and Liseo, B. (2012). Objective Bayesian Analysis of Skew-t Distributions. *Scandinavian Journal of Statistics* 40: 63-85.

Chen, M. H., Dey, D. K. and Shao, Q. M. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association* 94: 1172–1186.

Chen, M. H., Ibrahim, J. G. and Kim, S. (2008). Properties and implementation of Jeffreys’s priors in binomial regression models. *Journal of the American Statistical Association* 103: 1659–1664.

Christen, J. A. and Fox, C. (2010). A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis* 5: 263–282.

Collett, D. (1999). *Modelling Binary Data*. Chapman & Hall/CRC, Boca Raton Florida.

Copenhaver, T. W. and Mielke, P. W. (1977). Quantit analysis: a quantal assay refinement. *Biometrics* 33: 175–186.

Czado, C. and Santner, T. J. (1992). The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference* 33: 213–231.

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

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

Guolo, A. (2012). Flexibly modeling the baseline risk in meta-analysis. *Statistics in Medicine*, in press.

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

Kim, S., Chen, M. H. and Dey, D. K. (2008). Flexible generalized $t$–link models for binary response data. *Biometrika* 95: 93–106.
Liseo, B. and Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* 136: 373–389.

Nadarajah, S. (2009). The skew logistic distribution. *Advances in Statistical Analysis* 93: 197-203.

Rubio, F. J. and Steel, M. F. J. (2013). Bayesian Inference for $P(X < Y)$ using Asymmetric Dependent Distributions. *Bayesian Analysis* 8: 43-62.

Stefanski, L. A. (1991). A normal scale mixture representation of the logistic distribution. *Statistics & Probability Letters* 11: 69–70.

Venkatraman, E. S. and Begg, C. B. (1996). A distribution–free procedure for comparing operating characteristic curves from a paired experiment. *Biometrika* 83: 835–848.

Wang, J., Boyer, J. and Genton M. C. (2004). A skew symmetric representation of multivariate distributions. *Statistica Sinica* 14: 1259–1270.

Wang, X. and Dey, D. K. (2010). Generalized extreme value regression for binary response data: an application to B2B electronic payments system adoption. *Annals of Applied Statistics* 4: 2000–2023.