

Bayesian Estimation for the Centered Parameterization of the Skew-Normal Distribution

Estimación bayesiana para la parametrización centrada de la
distribución normal-asimétrica

PAULINO PÉREZ-RODRÍGUEZ^a, JOSÉ A. VILLASEÑOR^b, SERGIO PÉREZ^c,
JAVIER SUÁREZ^d

SOCIOECONOMÍA ESTADÍSTICA E INFORMÁTICA-ESTADÍSTICA, COLEGIO DE POSTGRADUADOS,
TEXCOCO, MÉXICO

Abstract

The skew-normal (SN) distribution is a generalization of the normal distribution, where a shape parameter is added to adopt skewed forms. The SN distribution has some of the properties of a univariate normal distribution, which makes it very attractive from a practical standpoint; however, it presents some inference problems. Specifically, the maximum likelihood estimator for the shape parameter tends to infinity with a positive probability. A new Bayesian approach is proposed in this paper which allows to draw inferences on the parameters of this distribution by using improper prior distributions in the “centered parametrization” for the location and scale parameter and a Beta-type for the shape parameter. Samples from posterior distributions are obtained by using the Metropolis-Hastings algorithm. A simulation study shows that the mode of the posterior distribution appears to be a good estimator in terms of bias and mean squared error. A comparative study with similar proposals for the SN estimation problem was undertaken. Simulation results provide evidence that the proposed method is easier to implement than previous ones. Some applications and comparisons are also included.

Key words: Metropolis-Hastings Algorithm, Point Estimation, Prior Distribution.

^aPhD. E-mail: perpdgo@gmail.com

^bPhD. E-mail: jvillasr@colpos.mx

^cPhD. E-mail: sergiop@colpos.mx

^dPhD. E-mail: sjavier@colpos.mx

Resumen

La distribución Normal Asimétrica (SN) es una generalización de la distribución normal, incluye un parámetro extra que le permite adoptar formas asimétricas. La distribución SN tiene algunas de las propiedades de la distribución normal univariada, lo que la hace muy atractiva desde el punto de vista práctico; sin embargo presenta algunos problemas de inferencia. Particularmente, el estimador de máxima verosimilitud para el parámetro de forma tiende a infinito con probabilidad positiva. Se propone una solución Bayesiana que permite hacer inferencia sobre los parámetros de esta distribución asignando distribuciones impropias en la “parametrización centrada” para el parámetro de localización y el de escala y una distribución tipo Beta para el parámetro de forma. Las muestras de las distribuciones posteriores se obtienen utilizando el algoritmo de Metropolis-Hastings. Un estudio de simulación muestra que la moda de la distribución posterior parece ser un buen estimador, en términos de sesgo y error cuadrado medio. Se presenta también un estudio de simulación donde se compara el procedimiento propuesto contra otros procedimientos. Los resultados de simulación proveen evidencia de que el método propuesto es más fácil de implementar que las metodologías previas. Se incluyen también algunas aplicaciones y comparaciones.

Palabras clave: algoritmo de Metropolis-Hastings, distribuciones a priori, estimación puntual.

1. Introduction

The skew-normal distribution is a three parameter class of distribution with location, scale and shape parameters, and it contains the normal distribution when the shape parameter equals zero. A continuous random variable Z is said to obey the skew-normal law with shape parameter $\lambda \in \mathbb{R}$ and it is denoted by $SN(\lambda)$ if its density function is:

$$f_Z(z; \lambda) = 2\phi(z) \Phi(\lambda z) I_{(-\infty, \infty)}(z), \quad (1)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and distribution functions of a standard normal variable.

If Y is a random variable defined by

$$Y = \xi + \omega Z,$$

with $\xi \in \mathbb{R}$, $\omega \in \mathbb{R}^+$, then Y is said to have a skew-normal distribution with location-scale (ξ, ω) parameters and shape parameter λ , and it is denoted by $Y \sim SN_D(\xi, \omega, \lambda)$. The subscript D indicates the use of the “direct parametrization” (Azzalini 1985). The density function of Y is:

$$f_Y(y; \xi, \omega, \lambda) = 2 \frac{1}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left[\lambda \left(\frac{y - \xi}{\omega}\right)\right] I_{(-\infty, \infty)}(y). \quad (2)$$

The mean and variance of density in (1) are given by:

$$E(Z) = \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}},$$

$$Var(Z) = 1 - \frac{2}{\pi} \frac{\lambda^2}{1+\lambda^2}.$$

The coefficient of skewness of Y is given by:

$$\gamma_1 = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{4-\pi}{2} \frac{(E(Z))^3}{\{1-(E(Z))^2\}^{3/2}} = \frac{4-\pi}{2} \frac{\left(\sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}}\right)^3}{\left(1 - \frac{2}{\pi} \frac{\lambda^2}{1+\lambda^2}\right)^{3/2}},$$

where κ_2, κ_3 are the second and third degree cumulants, respectively. The range of γ_1 is $(-c_1, c_1)$ where $c_1 = \sqrt{2}(4-\pi)/(\pi-2)^{3/2} \approx 0.99527$.

The maximum likelihood estimation of parameters is troublesome. Although the likelihood function can be calculated without much trouble, several problems arise on maximizing it. For example, if $\xi = 0$, $\omega = 1$, and all the observations are positive (or negative), then the likelihood function is monotone, and its maximum occur on the upper (lower) boundary of the parameter space, corresponding to a non finite estimate of λ . In the case of unknown values for the parameters ξ , ω , λ , the problem can be even more difficult because there is always an inflexion point at $\lambda = 0$ for the likelihood function, leading to the Hessian singular (Chiogna 1998, Azzalini & Genton 2008). The second problem is solved by using a “centered parametrization” of the density function. The first is an open problem and some proposals already exist: Sartori’s (2006) stand out, who uses a modification of the score function to estimate the λ parameter, combined with maximum likelihood estimators of ξ and ω .

The “centered parametrization” [Azzalini (1985), Azzalini & Capitanio (1999)] is obtained as follows given $Z \sim SN(\lambda)$, the random variable Y ,

$$Y = \mu + \sigma \left(\frac{Z - E(Z)}{\sqrt{Var(Z)}} \right),$$

is said to be a skew-normal variable with centered parameters μ, σ, γ_1 , $E(Y) = \mu$ and $Var(Y) = \sigma^2$. In this case the usual notation is $Y \sim SN_C(\mu, \sigma, \gamma_1)$. Here γ_1 is the skewness parameter of both Y and Z . One can easily change from one parametrization to another by using the identities in (3):

$$\begin{aligned}
\lambda &= \operatorname{sgn}(\gamma_1) \left(\frac{2}{4-\pi} \right)^{1/3} |\gamma_1|^{1/3} \left\{ \frac{2}{\pi} + |\gamma_1|^{2/3} \left(\frac{2}{4-\pi} \right)^{2/3} \left(\frac{2}{\pi} - 1 \right) \right\}^{-1/2}, \\
\xi &= \mu - \sigma \left(\frac{2}{4-\pi} \gamma_1 \right)^{1/3}, \\
\omega &= \sigma \left\{ 1 + \left(\frac{2}{4-\pi} \gamma_1 \right)^{2/3} \right\}^{1/2}.
\end{aligned} \tag{3}$$

Based on the Bayesian approach, a solution was proposed by Liseo & Loperfido (2006), who focused their inference on λ , considering ξ, ω as nuisance parameters in the direct parameterization. Wiper, Girón & Pewsey (2008) also studied the problem in the case of the half-normal distribution.

In this paper we propose a new Bayesian approach based on the “centered parametrization” to deal with the estimation of the shape parameter in the skew-normal family. The proposed methodology applies MCMC simulation techniques by using the Metropolis Hastings algorithm (Metropolis, Rosembluth, Teller & Teller 1953).

From a classical point of view, there are at least two reasons to consider the “centered parametrization”: *i*) it provides a more practical interpretation of the parameters, and *ii*) it solves the well known problems of the likelihood function under direct parametrization. Arellano-Valle & Azzalini (2008) stated that the standard likelihood based methods and also the Bayesian methods are problematic when they are applied to inference on the parameters in the direct parameterization near $\lambda = 0$. This is due to the fact that the direct parameterization is not numerically suitable for estimation.

The structure of this paper is the following: In section 2 a Bayesian method is proposed to obtain point estimates for the “centered parameters”, which can be back transformed using the equations in (3) to obtain point estimates of the direct parameters. Section 3 contains results from a simulation study concerning the proposed methodology. Applications are presented in section 4. Some conclusions are included in section 5.

2. Bayesian Estimation

Let $\mathbf{Y} = (Y_1, \dots, Y_n)'$ be a random sample from the skew-normal distribution with parameters μ , σ , and γ_1 , and suppose we wish to obtain Bayesian estimators for the parameters. Following Azzalini (1985), Azzalini & Capitanio (1999), and Pewsey (2000) we propose to use the “centered parametrization” to obtain Bayesian estimators for the parameters of interest. We propose the following prior distributions:

$$\begin{aligned}
 p(\mu) &\propto 1, \\
 p(\sigma) &\propto \frac{1}{\sigma}, \\
 p(\gamma_1) &\propto \left(\frac{\gamma_1 + c_1}{2c_1}\right)^{a-1} \left(1 - \frac{\gamma_1 + c_1}{2c_1}\right)^{b-1} I_{(-c_1, c_1)}(\gamma_1).
 \end{aligned}$$

Note that we have assigned the standard prior for location-scale parameters to μ and σ (Box & Tiao 1973). In the case of γ_1 , we know that it takes values on $(-c_1, c_1)$, so if $W \sim \text{Beta}(a, b)$, then the random variable $\gamma_1 = 2c_1W - c_1$ has a density the kernel of which is shown above. The prior density for γ_1 depends on two hyperparameters (a and b) and could lead to a rich varieties of shapes, just as in the case of the Beta distribution. In this paper, we set $a = b = 1$, which leads to a uniform distribution on $(-c_1, c_1)$, but other values can be selected, allowing the incorporation of prior information. The uniform prior is non-informative, but proper. An invariant and non informative prior for γ_1 could be derived using an approach similar to that employed by Liseo & Loperfido (2006); however, that approach is complicated from computational point of view because it involves numerical integration in the n -dimensional space. The priors proposed in this work allow the implementation of the well known Metropolis-Hastings algorithm in order to sample the posterior density. As the sample size increases, the effect of the prior becomes less relevant.

Then, under the assumption of independence, the joint prior density in the “centered parameterization” is:

$$p(\mu, \sigma, \gamma_1) \propto \frac{1}{\sigma} \left(\frac{\gamma_1 + c_1}{2c_1}\right)^{a-1} \left(1 - \frac{\gamma_1 + c_1}{2c_1}\right)^{b-1} I_{(-c_1, c_1)}(\gamma_1) I_{(0, \infty)}(\sigma).$$

Applying Bayes’ theorem, from (2) and (3) the posterior joint distribution of μ, σ, γ_1 is:

$$p(\mu, \sigma, \gamma_1 | \mathbf{y}) \propto \left\{ \prod_{i=1}^n f_{Y_i}(y_i; \mu, \sigma, \gamma_1) \right\} p(\mu, \sigma, \gamma_1).$$

The implied prior distribution can be obtained on the parameters in the original parameter space by the random variable transformation formulae, for which the inverse transformation turns out to be:

T^{-1}

$$\begin{aligned}\mu &= w_1(\xi, \omega, \lambda) = \xi + \sigma \left(\frac{2}{4 - \pi} \gamma_1 \right)^{1/3} \\ \sigma &= w_2(\xi, \omega, \lambda) = \omega \left[1 + \left(\frac{2}{4 - \pi} \gamma_1 \right)^{2/3} \right]^{-1/2} \\ \gamma_1 &= w_3(\xi, \omega, \lambda) = \frac{4 - \pi}{2} \frac{\left(\sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}} \right)^3}{\left(1 - \frac{2}{\pi} \frac{\lambda^2}{1 + \lambda^2} \right)^{3/2}},\end{aligned}$$

and its Jacobian is

$$\begin{aligned}|J| &= \left| \frac{12}{4 - \pi} \lambda \left[1 + \left(\frac{2}{4 - \pi} g(\lambda) \right)^{2/3} \right]^{-1/2} \left(1 - \lambda^2 (\lambda^2 + 1)^{-1} \right) \left(1 - \frac{2}{\pi} \frac{\lambda^2}{1 + \lambda^2} \right)^{-3/2} \right. \\ &\quad \left. \times (\lambda^2 + 1)^{-1} \left[\frac{\sqrt{2}}{\pi^{3/2}} (\lambda^2 + 1)^{-\frac{1}{2}} \lambda + \frac{1}{\pi} \left(\sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}} \right)^3 \left(1 - \frac{2}{\pi} \frac{\lambda^2}{1 + \lambda^2} \right)^{-1} \right] \right|\end{aligned}$$

and, therefore the implied joint prior density in the original parametrization is given by

$$p(\xi, \omega, \lambda) \propto p_{\mu, \sigma, \gamma_1}(w_1(\xi, \omega, \lambda), w_2(\xi, \omega, \lambda), w_3(\xi, \omega, \lambda)) |J|.$$

This implied original parameterized joint prior density is quite different from the one used by Liseo & Loperfido (2006).

To obtain the marginal posterior distributions of interest $p(\mu | \mathbf{y})$, $p(\sigma | \mathbf{y})$, $p(\gamma_1 | \mathbf{y})$ it is necessary to use numerical based integration as the Markov Chain Monte Carlo techniques. In the case of the Gibbs sampler, it is necessary to have the complete conditionals $p(\mu | \sigma, \gamma_1, \mathbf{y})$, $p(\sigma | \mu, \gamma_1, \mathbf{y})$, $p(\gamma_1 | \mu, \sigma, \mathbf{y})$ to implement it; however, their closed forms are not available, therefore we propose to use the Metropolis-Hastings algorithm (e.g. Metropolis et al. 1953, Chib & Greenberg 1995).

To apply the Metropolis-Hastings algorithm, the candidate generating distribution has to be selected first. One has to be very careful in this step since an inadequate selection of such a distribution can cause the Metropolis-Hastings algorithm to have a poor performance due to a high rejection rate.

Pewsey (2000) obtained large sample theory results for the moment's estimators from the "centered parametrization". For $Y \sim SN_C(\mu, \sigma, \gamma_1)$, let

$$\begin{aligned} \beta_2 &= 3 + \tau^4(\pi - 3), \\ \beta_3 &= 10\gamma_1 + \tau^5(3\pi^2 - 40\pi + 96)/4, \\ \beta_4 &= 15 \{1 + \tau^4(2\pi - 6)\} - \tau^6(9\pi^2 - 80\pi + 160)/2, \end{aligned}$$

where $\tau = (2/(4 - \pi)\gamma_1)^{1/3}$. By using the delta method, Pewsey (2000) obtained:

$$\begin{aligned} \text{Var}(\tilde{\mu}) &= \sigma^2/n, \\ \text{Var}(\tilde{\sigma}) &= \sigma^2(\beta_2 - 1)/4n + O(n^{-3/2}), \\ \text{Var}(\tilde{\gamma}_1) &= \{9 - 6\beta_2 - 3\gamma_1\beta_3 + \beta_4 + \gamma_1^2(35 + 9\beta_2)/4\} /n + O(n^{-3/2}), \\ \text{Cov}(\tilde{\mu}, \tilde{\sigma}) &= \sigma^2\gamma_1/2n + O(n^{-3/2}), \\ \text{Cov}(\tilde{\mu}, \tilde{\gamma}_1) &= \sigma(\beta_2 - 3 - 3\gamma_1^2/2)/n + O(n^{-3/2}), \\ \text{Cov}(\tilde{\sigma}, \tilde{\gamma}_1) &= \sigma \{2\beta_3 - \gamma_1(5 + 3\beta_2)\} /4n + O(n^{-3/2}), \end{aligned}$$

where $\tilde{\mu}, \tilde{\sigma}, \tilde{\gamma}_1$ are the moment estimators of μ, σ and γ_1 . As a consequence of the definition of the moment estimators, Slutsky's Theorem and the Central Limit Theorem, the joint distribution of the estimators is asymptotically trivariate normal. We can use these results to select the multivariate normal distribution as the candidate generator in the Metropolis-Hastings algorithm.

To start the Metropolis-Hastings algorithm, the trivariate normal distribution $N_3(\tilde{\theta}, \tilde{\Sigma})$ is used as a proposal density, where:

$$\tilde{\Sigma} = \begin{pmatrix} \widetilde{\text{Var}(\tilde{\mu})} & \widetilde{\text{Cov}(\tilde{\mu}, \tilde{\sigma})} & \widetilde{\text{Cov}(\tilde{\mu}, \tilde{\gamma}_1)} \\ \widetilde{\text{Cov}(\tilde{\mu}, \tilde{\sigma})} & \widetilde{\text{Var}(\tilde{\sigma})} & \widetilde{\text{Cov}(\tilde{\sigma}, \tilde{\gamma}_1)} \\ \widetilde{\text{Cov}(\tilde{\mu}, \tilde{\gamma}_1)} & \widetilde{\text{Cov}(\tilde{\sigma}, \tilde{\gamma}_1)} & \widetilde{\text{Var}(\tilde{\gamma}_1)} \end{pmatrix}, \tag{4}$$

and $\tilde{\theta} = (\tilde{\mu}, \tilde{\sigma}, \tilde{\gamma}_1)$ are the moment's estimators of (μ, σ, γ_1) in the "centered parametrization". The variances and covariances in (4) are obtained by plugging in the parameter estimates into the variance and covariance formulae given by Pewsey (2000). It is important to note that $\tilde{\Sigma}$ in (4) can be adjusted in order to mimic the posterior distribution (Carlin & Louis 2000). In this paper, routines were written in the R software (R Core Team 2015) to estimate the posterior distributions of μ, σ, γ_1 . These routines are designed to update the covariance matrix after having obtained s Markov Chain Monte Carlo samples from the parameters of interest. In this work we set $s = 1,000$. The variance covariance matrix is estimated as:

$$\tilde{\tilde{\Sigma}} = \frac{1}{s} \sum_{j=1}^s (\theta_j - \bar{\theta}) (\theta_j - \bar{\theta})',$$

where j indexes the Markov Chain Monte Carlo samples, $\bar{\theta} = \frac{1}{n} \sum_{j=1}^n \theta_j$ and θ_j is the j -th Markov Chain Monte Carlo sample. Once the variance covariance-matrix

is adjusted, the $N_3(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}})$ is used as the proposal distribution. Due to the fact that σ is positive and γ_1 is restricted to the interval $(-c_1, c_1)$, inadmissible values can be avoided simply by rejecting samples that do not meet these conditions. In order to verify the convergence of the Metropolis-Hastings algorithm, the Gelman & Rubin (1992) convergence test is used.

3. Simulation Study

In this section, we present results from a simulation study concerning the proposed Bayesian procedure described in the previous section. Samples of size $n = 20, 50$ and 100 were simulated from $SN_C(\mu, \sigma, \gamma_1)$ for different values of μ, σ and γ_1 . A comparison with the method of moments (Pewsey 2000) and modified maximum likelihood estimators (Sartori 2006) in terms of the bias and the mean squared error is also included. For this comparison, the estimates in the “direct parameterization” were transformed to the “centered parameterization”.

Next, the algorithm used to estimate the bias and mean squared error using the Bayesian approach is briefly described.

1. Set $i = 1$.
2. Set (μ, σ, γ_1) .
3. Generate a random sample of size n from $SN_C(\mu, \sigma, \gamma_1)$.
4. Estimate (μ, σ, γ_1) using the method of moments.
5. Obtain the initial variance-covariance estimator using the Pewsey’s (2000) result.
 - a) Generate a trivariate normal sample using the results of steps 4) and 5).
 - b) Reject the samples that do not meet the conditions: $\sigma > 0$ and $\gamma_1 \in (-c_1, c_1)$, or keep the samples that meet the conditions to select one with the Metropolis algorithm.
 - c) Repeat steps a) and b), $B = 30,000$ times; update the variance-covariance matrix of the proposed distribution after 1,000 iterations.
 - d) Obtain the marginal posterior densities of (μ, σ, γ_1) ; discard the first 25,000 iterations (burn-in).
 - e) Obtain the mode of the marginal posterior density of (μ, σ, γ_1) using the results in step d), say $(\hat{\mu}_i, \hat{\sigma}_i, \hat{\gamma}_{1,i})$.
6. Set $i = i + 1$.
7. Repeat steps 3 to 6, 5,000 times.
8. Obtain the mean of $\hat{\mu}_1, \dots, \hat{\mu}_{5,000}, \hat{\sigma}_1, \dots, \hat{\sigma}_{5,000}, \hat{\gamma}_{1,1}, \dots, \hat{\gamma}_{1,5000}$.

9. Compute the bias and mean squared error of $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\gamma}_1$.

The algorithm to estimate the bias and mean squared error for the estimators obtained using the method of moments is as follows:

1. Set $i = 1$.
2. Set (μ, σ, γ_1) .
3. Generate a random sample of size n from $SN_C(\mu, \sigma, \gamma_1)$.
4. Obtain the moment estimators of (μ, σ, γ_1) using Pewsey's (2000) results, say $(\hat{\mu}_i, \hat{\sigma}_i, \hat{\gamma}_{1,i})$.
5. Set $i = i + 1$.
6. Repeat steps 3 to 6, 5,000 times.
7. Obtain the mean of $\tilde{\mu}_1, \dots, \tilde{\mu}_{5,000}$, $\tilde{\sigma}_1, \dots, \tilde{\sigma}_{5,000}$, $\tilde{\gamma}_{1,1}, \dots, \tilde{\gamma}_{1,5000}$.
8. Compute the bias and mean squared error of $\tilde{\mu}$, $\tilde{\sigma}$ and $\tilde{\gamma}_1$.

The algorithm to estimate the bias and mean squared error for the estimators obtained using the modified maximum likelihood (Sartori 2006) is analogous to that used in the moment's estimator case. However, the estimates in the "direct parameterization" were transformed to the "centered parameterization" before computing the bias and the mean squared error.

Tables 1-3 present the bias and mean squared error obtained when using the Bayesian approach, the method of moments, and the modified maximum likelihood method respectively. In all the cases considered the bias and the mean squared error decrease as the sample size increases for the three methods. The bias for the location and scale parameters are small in general. Sartori (2006) pointed out that the bias for the shape parameter in the skew-normal distribution has a lower asymptotic bias than the maximum likelihood estimator.

TABLE 1: Bias and mean squared error for Bayesian point estimators. The estimates for μ , σ , and γ_1 were obtained from 5,000 simulated samples of size n from $SN_C(\mu, \sigma, \gamma_1)$. The mode of the marginal posterior densities was used as a point estimate of the true parameters.

$\mu = 0.7969, \sigma = 0.6041, \gamma_1 = 0.9851$						
n	$E(\hat{\mu} - \mu)$	$E(\hat{\sigma} - \sigma)$	$E(\hat{\gamma}_1 - \gamma_1)$	$E(\hat{\mu} - \mu)^2$	$E(\hat{\sigma} - \sigma)^2$	$E(\hat{\gamma}_1 - \gamma_1)^2$
20	-0.0051	-0.0437	-0.3026	0.0187	0.0124	0.1734
50	-0.0029	-0.0330	-0.0864	0.0076	0.0053	0.0201
100	-0.0000	-0.0206	-0.0279	0.0036	0.0025	0.0025
$\mu = 0.7939, \sigma = 0.6080, \gamma_1 = 0.9556$						
n	$E(\hat{\mu} - \mu)$	$E(\hat{\sigma} - \sigma)$	$E(\hat{\gamma}_1 - \gamma_1)$	$E(\hat{\mu} - \mu)^2$	$E(\hat{\sigma} - \sigma)^2$	$E(\hat{\gamma}_1 - \gamma_1)^2$
20	-0.0026	-0.0394	-0.3060	0.0196	0.0121	0.1815
50	-0.0032	-0.0297	-0.0989	0.0074	0.0052	0.0282
100	-0.0021	-0.0189	-0.0363	0.0039	0.0026	0.0051
$\mu = 0.7824, \sigma = 0.6228, \gamma_1 = 0.8510$						
n	$E(\hat{\mu} - \mu)$	$E(\hat{\sigma} - \sigma)$	$E(\hat{\gamma}_1 - \gamma_1)$	$E(\hat{\mu} - \mu)^2$	$E(\hat{\sigma} - \sigma)^2$	$E(\hat{\gamma}_1 - \gamma_1)^2$
20	-0.0014	-0.0315	-0.3021	0.0204	0.0125	0.2032
50	-0.0038	-0.0202	-0.1158	0.0079	0.0050	0.0514
100	-0.0019	-0.0141	-0.0551	0.0043	0.0027	0.0187
$\mu = 0.7569, \sigma = 0.6535, \gamma_1 = 0.6670$						
n	$E(\hat{\mu} - \mu)$	$E(\hat{\sigma} - \sigma)$	$E(\hat{\gamma}_1 - \gamma_1)$	$E(\hat{\mu} - \mu)^2$	$E(\hat{\sigma} - \sigma)^2$	$E(\hat{\gamma}_1 - \gamma_1)^2$
20	-0.0024	-0.0232	-0.2635	0.0234	0.0128	0.2117
50	-0.0025	-0.0125	-0.1093	0.0092	0.0052	0.0797
100	-0.0027	-0.0072	-0.0467	0.0046	0.0026	0.0352
$\mu = 0.5642, \sigma = 0.8256, \gamma_1 = 0.1370$						
n	$E(\hat{\mu} - \mu)$	$E(\hat{\sigma} - \sigma)$	$E(\hat{\gamma}_1 - \gamma_1)$	$E(\hat{\mu} - \mu)^2$	$E(\hat{\sigma} - \sigma)^2$	$E(\hat{\gamma}_1 - \gamma_1)^2$
20	-0.0042	-0.0107	-0.0569	0.0378	0.0198	0.1968
50	-0.0002	0.0007	-0.0234	0.0148	0.0076	0.1187
100	-0.0024	0.0015	-0.0099	0.0071	0.0038	0.0686
$\mu = 0.3568, \sigma = 0.9342, \gamma_1 = 0.0239$						
n	$E(\hat{\mu} - \mu)$	$E(\hat{\sigma} - \sigma)$	$E(\hat{\gamma}_1 - \gamma_1)$	$E(\hat{\mu} - \mu)^2$	$E(\hat{\sigma} - \sigma)^2$	$E(\hat{\gamma}_1 - \gamma_1)^2$
20	0.0061	-0.0151	0.0020	0.0519	0.0278	0.1935
50	-0.0007	0.0017	-0.0049	0.0214	0.0109	0.1242
100	-0.0005	0.0059	-0.0070	0.0108	0.0057	0.0708

TABLE 2: Bias and mean squared error for modified maximum likelihood estimators (Sartori, 2006). The estimates for μ , σ , and γ_1 were obtained from 5,000 simulated samples of size n from $SN_C(\mu, \sigma, \gamma_1)$.

$\mu = 0.7969, \sigma = 0.6041, \gamma_1 = 0.9851$						
n	$E(\check{\mu} - \mu)$	$E(\check{\sigma} - \sigma)$	$E(\check{\gamma}_1 - \gamma_1)$	$E(\check{\mu} - \mu)^2$	$E(\check{\sigma} - \sigma)^2$	$E(\check{\gamma}_1 - \gamma_1)^2$
20	-0.0003	-0.0585	-0.2880	0.0176	0.0130	0.1588
50	-0.0014	-0.0286	-0.0544	0.0069	0.0048	0.0114
100	-0.0007	-0.0129	-0.0159	0.0035	0.0023	0.0015
$\mu = 0.7939, \sigma = 0.6080, \gamma_1 = 0.9556$						
n	$E(\check{\mu} - \mu)$	$E(\check{\sigma} - \sigma)$	$E(\check{\gamma}_1 - \gamma_1)$	$E(\check{\mu} - \mu)^2$	$E(\check{\sigma} - \sigma)^2$	$E(\check{\gamma}_1 - \gamma_1)^2$
20	-0.0001	-0.0552	-0.2792	0.0176	0.0126	0.1535
50	-0.0016	-0.0255	-0.0650	0.0072	0.0049	0.0193
100	-0.0013	-0.0117	-0.0226	0.0034	0.0023	0.0039
$\mu = 0.7824, \sigma = 0.6228, \gamma_1 = 0.8510$						
n	$E(\check{\mu} - \mu)$	$E(\check{\sigma} - \sigma)$	$E(\check{\gamma}_1 - \gamma_1)$	$E(\check{\mu} - \mu)^2$	$E(\check{\sigma} - \sigma)^2$	$E(\check{\gamma}_1 - \gamma_1)^2$
20	0.0008	-0.0476	-0.2719	0.0186	0.0127	0.1856
50	-0.0016	-0.0215	-0.0843	0.0075	0.0049	0.0445
100	-0.0005	-0.0089	-0.0356	0.0039	0.0025	0.0159
$\mu = 0.7569, \sigma = 0.6535, \gamma_1 = 0.6670$						
n	$E(\check{\mu} - \mu)$	$E(\check{\sigma} - \sigma)$	$E(\check{\gamma}_1 - \gamma_1)$	$E(\check{\mu} - \mu)^2$	$E(\check{\sigma} - \sigma)^2$	$E(\check{\gamma}_1 - \gamma_1)^2$
20	-0.0029	-0.0438	-0.2261	0.0215	0.0130	0.1998
50	-0.0039	-0.0168	-0.0831	0.0085	0.0053	0.0825
100	-0.0009	-0.0073	-0.0375	0.0042	0.0025	0.0360
$\mu = 0.5642, \sigma = 0.8256, \gamma_1 = 0.1370$						
n	$E(\check{\mu} - \mu)$	$E(\check{\sigma} - \sigma)$	$E(\check{\gamma}_1 - \gamma_1)$	$E(\check{\mu} - \mu)^2$	$E(\check{\sigma} - \sigma)^2$	$E(\check{\gamma}_1 - \gamma_1)^2$
20	-0.0025	-0.0434	-0.0494	0.0338	0.0185	0.2183
50	-0.0002	-0.0131	-0.0106	0.0136	0.0073	0.1208
100	-0.0009	-0.0067	-0.0053	0.0069	0.0035	0.0615
$\mu = 0.3568, \sigma = 0.9342, \gamma_1 = 0.0239$						
n	$E(\check{\mu} - \mu)$	$E(\check{\sigma} - \sigma)$	$E(\check{\gamma}_1 - \gamma_1)$	$E(\check{\mu} - \mu)^2$	$E(\check{\sigma} - \sigma)^2$	$E(\check{\gamma}_1 - \gamma_1)^2$
20	0.0019	-0.0466	0.0032	0.0512	0.0266	0.2189
50	-0.0009	-0.0132	0.0008	0.0196	0.0102	0.1298
100	0.0000	-0.0064	-0.0003	0.0100	0.0051	0.0630

TABLE 3: Bias and mean squared error for moment estimates (Pewsey, 2000). The estimates for μ , σ , and γ_1 were obtained from 5,000 simulated samples of size n from $SN_C(\mu, \sigma, \gamma_1)$.

$\mu = 0.7969, \sigma = 0.6041, \gamma_1 = 0.9851$						
n	$E(\tilde{\mu} - \mu)$	$E(\tilde{\sigma} - \sigma)$	$E(\tilde{\gamma}_1 - \gamma_1)$	$E(\tilde{\mu} - \mu)^2$	$E(\tilde{\sigma} - \sigma)^2$	$E(\tilde{\gamma}_1 - \gamma_1)^2$
20	0.0009	-0.0258	-0.3292	0.0182	0.0135	0.2115
50	-0.0015	-0.0105	-0.1872	0.0073	0.0053	0.0804
100	-0.0007	-0.0048	-0.1280	0.0038	0.0026	0.0404
$\mu = 0.7939, \sigma = 0.6080, \gamma_1 = 0.9556$						
n	$E(\tilde{\mu} - \mu)$	$E(\tilde{\sigma} - \sigma)$	$E(\tilde{\gamma}_1 - \gamma_1)$	$E(\tilde{\mu} - \mu)^2$	$E(\tilde{\sigma} - \sigma)^2$	$E(\tilde{\gamma}_1 - \gamma_1)^2$
20	0.0010	-0.0235	-0.3057	0.0192	0.0135	0.1987
50	-0.0007	-0.0107	-0.1822	0.0076	0.0054	0.0828
100	-0.0007	-0.0050	-0.1149	0.0036	0.0027	0.0401
$\mu = 0.7824, \sigma = 0.6228, \gamma_1 = 0.8510$						
n	$E(\tilde{\mu} - \mu)$	$E(\tilde{\sigma} - \sigma)$	$E(\tilde{\gamma}_1 - \gamma_1)$	$E(\tilde{\mu} - \mu)^2$	$E(\tilde{\sigma} - \sigma)^2$	$E(\tilde{\gamma}_1 - \gamma_1)^2$
20	0.0016	-0.0230	-0.2667	0.0188	0.0135	0.2024
50	-0.0000	-0.0112	-0.1474	0.0077	0.0051	0.0850
100	0.0002	-0.0051	-0.0913	0.0040	0.0026	0.0468
$\mu = 0.7569, \sigma = 0.6535, \gamma_1 = 0.6670$						
n	$E(\tilde{\mu} - \mu)$	$E(\tilde{\sigma} - \sigma)$	$E(\tilde{\gamma}_1 - \gamma_1)$	$E(\tilde{\mu} - \mu)^2$	$E(\tilde{\sigma} - \sigma)^2$	$E(\tilde{\gamma}_1 - \gamma_1)^2$
20	-0.0012	-0.0270	-0.2112	0.0215	0.0136	0.2068
50	-0.0021	-0.0124	-0.1025	0.0086	0.0054	0.0973
100	-0.0006	-0.0065	-0.0525	0.0044	0.0027	0.0560
$\mu = 0.5642, \sigma = 0.8256, \gamma_1 = 0.1370$						
n	$E(\tilde{\mu} - \mu)$	$E(\tilde{\sigma} - \sigma)$	$E(\tilde{\gamma}_1 - \gamma_1)$	$E(\tilde{\mu} - \mu)^2$	$E(\tilde{\sigma} - \sigma)^2$	$E(\tilde{\gamma}_1 - \gamma_1)^2$
20	-0.0026	-0.0345	-0.0406	0.0337	0.0180	0.2040
50	-0.0007	-0.0130	-0.0063	0.0136	0.0071	0.1130
100	-0.0012	-0.0069	-0.0063	0.0070	0.0035	0.0580
$\mu = 0.3568, \sigma = 0.9342, \gamma_1 = 0.0239$						
n	$E(\tilde{\mu} - \mu)$	$E(\tilde{\sigma} - \sigma)$	$E(\tilde{\gamma}_1 - \gamma_1)$	$E(\tilde{\mu} - \mu)^2$	$E(\tilde{\sigma} - \sigma)^2$	$E(\tilde{\gamma}_1 - \gamma_1)^2$
20	-0.0039	-0.0366	-0.0077	0.0493	0.0255	0.2025
50	-0.0000	-0.0154	0.0085	0.0199	0.0102	0.1041
100	0.0020	-0.0059	0.0001	0.0098	0.0052	0.0565

4. Some Examples

In this section, we present three examples of the proposed Bayesian method. For each example, we estimate the marginal posterior densities of μ, σ , and γ_1 using three parallel Metropolis sampling chains; each are run for 50,000 iterations and a burn-in period of 25,000 iterations. In all the cases the proposal acceptance rate was at least 30%. Convergence was checked by inspecting trace plots and applying the Gelman & Rubin (1992) test.

4.1. Example 1: The Frontier Data

The data in this example corresponds to $n = 50$ random numbers generated from $SN_D(0, 1, 5)$ or equivalently $SN_C(0.7824, 0.6228, 0.8510)$. The data are available within the R's `sn` package (Azzalini 2016). The maximum of the likelihood function occurs on the upper boundary of the parameter space, corresponding to an infinite estimate of λ . So, one could expect to obtain an estimate near 0.99527 for γ_1 .

Figure 1 shows histograms of simulated draws from the posterior distribution for μ, σ , and γ_1 . Table 4 shows the posterior summaries for μ, σ , and γ_1 . If the posterior mode of the marginal posterior distribution is used for estimation purposes, then $\hat{\mu} = 0.8873$, $\hat{\sigma} = 0.7323$, $\hat{\gamma}_1 = 0.9682$. Note that those point estimates are similar to those obtained with the method proposed by Sartori (2006), that is $\check{\mu} = 0.8811(0.08602)$, $\check{\sigma} = 0.7295(0.0777)$, $\check{\gamma}_1 = 0.9481(0.0570)$, where the values in parentheses are the standard errors. Table 5 shows the result for the Gelman & Rubin's test. We used three parallel Metropolis sampling chains with different initial values. Approximated convergence is diagnosed when the upper C.I. limit is close to 1.

TABLE 4: Posterior summary for the frontier data.

	Quantile 2.5%	Median	Mean	Mode	Sd	Quantile 97.5%
μ	0.7040	0.8840	0.8875	0.8873	0.1026	1.1011
σ	0.5976	0.7301	0.7369	0.7323	0.0812	0.9150
γ_1	0.4238	0.8811	0.8355	0.9682	0.1539	0.9910

TABLE 5: Results of the convergence test (Gelman & Rubin 1992).

Potential Scale Reduction Factors		
	Point est.	Upper C.I.
μ	1.04	1.06
σ	1.04	1.06
γ_1	1.04	1.10
Multivariate psrf		
1.03		

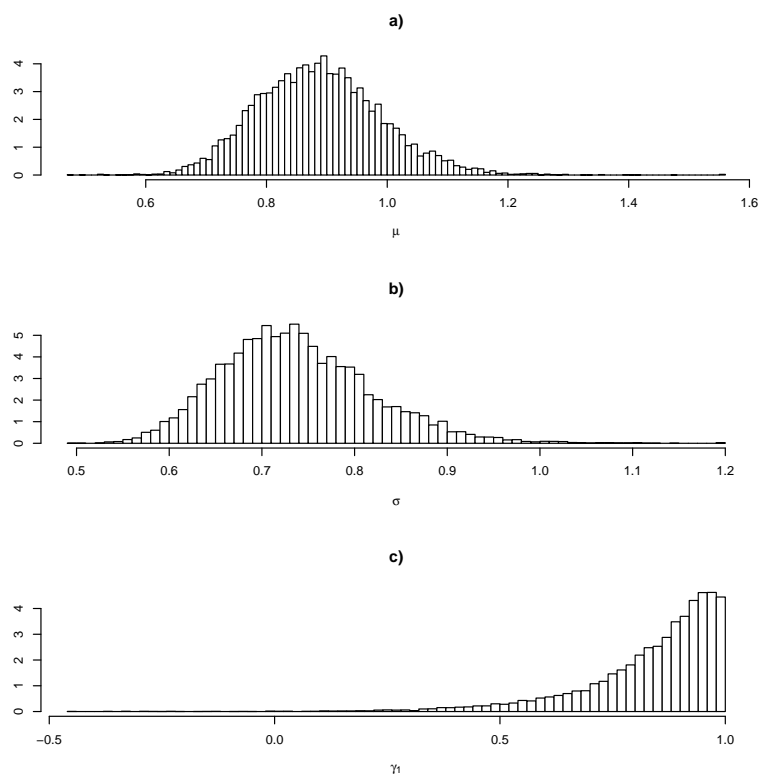


FIGURE 1: Histogram of simulated draws from the posterior distributions for a) μ , b) σ and c) γ_1 .

4.2. Example 2: Sartori's Data

The data in Table 6 are 20 random numbers from $SN_D(0, 1, 10)$ or equivalently $SN_C(0.7939, 0.6080, 0.9556)$ published in Sartori's article (2006). The usual maximum likelihood estimator for λ and that obtained by using Sartori's method are both finite.

TABLE 6: Sartori's data.

0.195	0.847	0.726	-0.139	1.788	0.570	2.069	0.452	0.868	1.199
0.894	0.887	1.258	0.918	0.496	0.183	0.119	1.207	0.446	2.579

Figure 2 shows the histograms of simulated draws from the posterior distribution for μ, σ and γ_1 . Table 7 shows the posterior summaries for the mean, standard deviation, and the coefficient of skewness. When the posterior mode of the marginal posterior distribution is used as point estimate, then $\hat{\mu} = 0.8627$, $\hat{\sigma} = 0.6382$, $\hat{\gamma}_1 = 0.6179$. Note that those point estimates are similar to those obtained with the method proposed by Sartori (2006): that is $\check{\mu} = 0.8812(0.2376)$, $\check{\sigma} = 0.6338(0.1171)$, $\check{\gamma}_1 = 0.6289(0.3884)$.

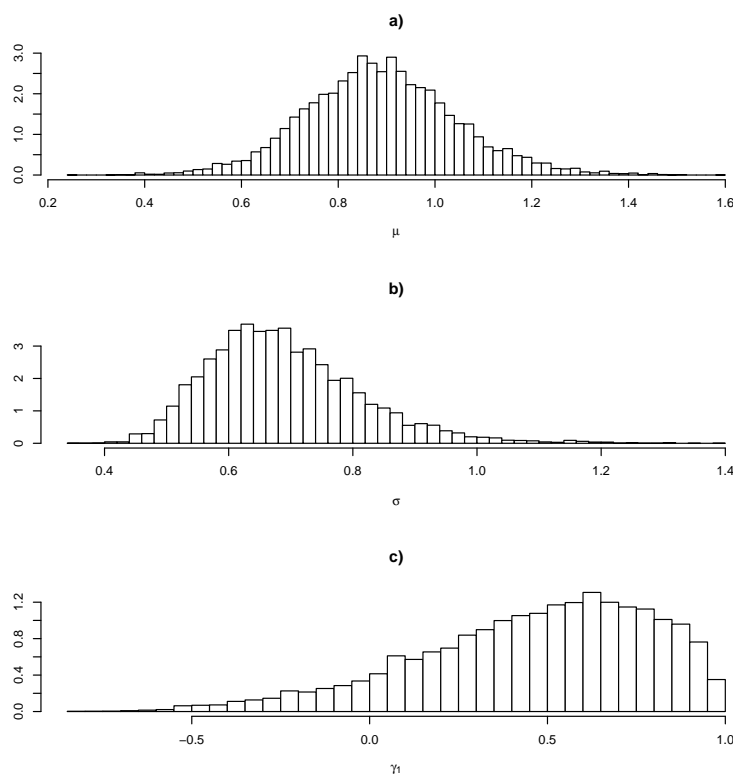


FIGURE 2: Histogram of simulated draws from the posterior distributions for a) μ , b) σ and c) γ_1 .

TABLE 7: Posterior summary Sartori's data.

	Quantile 2.5%	Median	Mean	Mode	Sd	Quantile 97.5%
μ	0.5895	0.8812	0.8850	0.8802	0.1548	1.1980
σ	0.5001	0.6763	0.6929	0.6746	0.1245	0.9810
γ_1	-0.3391	0.4996	0.4451	0.5873	0.3365	0.9329

4.3. Example 3: Half-Normal Case

Perhaps, one of the hardest cases to estimate the parameters of the $SN_D(0, 1, \lambda)$ is when λ is large. Because it is known that as λ tends to infinity, and the $SN_D(0, 1, \lambda)$ tends to the half-normal density, we test the proposed estimation procedure for the half-normal case and compared it with the results of Wiper et al. (2008). In this case, we expect to obtain an estimate near 0.9936 for γ_1 . In Table 8, we present the simulation results for the proposed estimation method when a sample comes from a half-normal random variable with parameters $\xi = 0, \eta = 1; HN(0, 1)$. This is approximately a $SN_C(0.7979, 0.6028, 0.99527)$.

The point estimate for γ_1 is very close to the expected value, 0.99527, when $n = 50, 100$; this result is a good one because if one graphed together the halfnor-

TABLE 8: Performance of the proposed estimation method when a sample comes from a $HN(0, 1)$. The mean square error (mse) and bias are obtained when the mode of the marginal posterior densities is used as a point estimate of $\mu = 0.7979, \sigma = 0.6028, \gamma_1 = 0.99527$. Results are based on 5,000 samples of size n .

n	$E(\hat{\mu} - \mu)$	$E(\hat{\sigma} - \sigma)$	$E(\hat{\gamma}_1 - \gamma_1)$	$E(\hat{\mu} - \mu)^2$	$E(\hat{\sigma} - \sigma)^2$	$E(\hat{\gamma}_1 - \gamma_1)^2$
20	0.0004	-0.0436	-0.3053	0.0176	0.0117	0.1695
50	-0.0055	-0.0381	-0.0849	0.0075	0.0055	0.0213
100	-0.0013	-0.0238	-0.0221	0.0035	0.0027	0.001

mal and the $SN_C(0.7979, 0.6028, 0.99527)$ densities, then we would not be able to distinguish between the two densities. Results show that the bias and mean squared error of the $\hat{\gamma}_1$ tends to zero as n increases, which provides evidence that the proposed estimator is consistent (Table 8). Although the following is not a fair comparison because in the Wiper's case only two parameters are estimated while in the SN case three parameters are estimated, at least we have an idea how well the proposed estimator is working in this case. For the location parameter, the proposed estimator seems to be better since its mean squared error is smaller than the one proposed by Wiper et al.'s (2008). For the scale parameter, the Wiper et al.'s (2008) estimator seems to be better than the one proposed in this paper (Table 9).

TABLE 9: Estimated bias and mean square error of the Bayesian posterior mean estimators when a sample comes from $HN(0, 1)$. Taken from Wiper's et al. (2008) results.

n	$E(\hat{\xi} - \xi)$	$E(\hat{\eta} - \eta)$	$E(\hat{\xi} - \xi)^2$	$E(\hat{\eta} - \eta)^2$
20	-0.0029	0.0298	0.0303	0.0036
50	-0.0005	0.0114	0.0109	0.0006
100	-0.0002	0.0055	0.0052	0.0002

5. Concluding Remarks

The simulation results in Section 3 provide evidence on the advantages of using the Metropolis Hastings algorithm to estimate the parameters of the skew-normal family.

The results from a simulation study show that the bias and the mean squared error decreases as the sample size increases. Also, it seems that the mode appears to be a precise synthetic index of the posterior distribution.

It turns out that the estimates provided by the new methodology for the shape parameter are finite in comparison with the estimates obtained by using the direct parameterization and the maximum likelihood method, which can lead to convergence problems.

Since we are using the Bayesian approach, it is possible to obtain HPD for the parameters of interest, similarly to Wiper et al. (2008).

[Received: January 2016 — Accepted: November 2016]

References

- Arellano-Valle, R. B. & Azzalini, A. (2008), ‘The centred parametrization for the multivariate skew-normal distribution’, *Journal of Multivariate Analysis* **99**, 1362–1382.
- Azzalini, A. (1985), ‘A class of distributions which includes the normal ones’, *Scandinavian Journal of Statistics* **12**, 171–178.
- Azzalini, A. (2016), *The R package sn: The Skew-Normal and Skew-t distributions (version 1.4-0)*, Università di Padova, Italia.
*<http://azzalini.stat.unipd.it/SN>
- Azzalini, A. & Capitanio, A. (1999), ‘Statistical applications of the multivariate skew normal distribution’, *Journal of the Royal Statistical Society* **61**, 579–602.
- Azzalini, A. & Genton, M. G. (2008), ‘Robust likelihood methods based on the skew- t and related distributions’, *International Statistical Review* **76**, 106–129.
- Box, G. & Tiao, G. (1973), *Bayesian inference in statistical analysis*, Addison-Wesley series in behavioral science: quantitative methods, Addison-Wesley Pub. Co.
- Carlin, B. P. & Louis, T. A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, second edn, Chapman - Hall/CRC, New York.
- Chib, S. & Greenberg, E. (1995), ‘Understanding the Metropolis-Hastings Algorithm’, *The American Statistician* **49**(4), 327–335.
- Chiogna, M. (1998), ‘Some results on the scalar skew-normal distribution’, *Journal of the Italian Statistical Society* **7**, 1–13.
- Gelman, A. & Rubin, D. B. (1992), ‘Inference from iterative simulation using multiple sequences’, *Statistical Science* **7**, 457–511.
- Liseo, B. & Loperfido, N. (2006), ‘A note on reference priors for the scalar skew-normal distribution’, *Journal of Statistical Planning and Inference* **136**, 373–389.
- Metropolis, N., Rosenbluth, A. W., Teller, M. & Teller, E. (1953), ‘Equations of state calculations by fast computing machines’, *Journal of Chemical Physics* **21**, 1087–1092.
- Pewsey, A. (2000), ‘Problems of inference for Azzalini’s skew-normal distribution’, *Journal of Applied Statistics* **27**, 859–870.

R Core Team (2015), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

*<http://www.R-project.org/>

Sartori, N. (2006), 'Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions', *Journal of Statistical Planning and Inference* **136**, 4259–4275.

Wiper, M., Girón, F. J. & Pewsey, A. (2008), 'Objective Bayesian inference for the half-normal and half-*t* distributions', *Communications in Statistics-Theory and Methods* **37**(20), 3165–3185.