Brazilian Journal of Probability and Statistics (2007), 21, pp. 141–163.
©Associação Brasileira de Estatística

Bayesian inference for the skewness parameter of the scalar skew-normal distribution

Cristian Luis Bayes and **Márcia D'Elia Branco**
Universidade de São Paulo

Abstract: The skew-normal distribution includes the normal distribution as a special case. This family of distributions has a shape parameter that defines the direction of the asymmetry of the distribution, also called skewness parameter. The main focus of this paper will be showing that the Bayesian approach using MCMC methods is a good alternative to make inference under the skewness parameter. We present a brief discuss about prior choice and propose an approximation for the Jeffreys prior developed by Liseo and Loperfido (2006). We also provide an approximation for the bias correction factor proposed by Sartori (2006). A simulation study is presented to compare the bias, mean square error and interval estimates using the maximum likelihood and different Bayes estimators. Finally, hypotheses tests are considered using Bayes Factor and the likelihood ratio test.

Key words: skewness, uniform prior, Jeffreys prior, bias correction, maximum likelihood and Bayes factor.

1 Introduction

The skew-normal distribution introduced by Azzalini (1985) is a useful family that preserves some normal distribution properties and include unimodal skewness distributions. A recent overview including inference problems, can be seen in Dalla Valle (2004).

A random variable Z has a standard skew normal distribution if its probability density function (pdf) is given by

$$f_Z(z) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R} \quad \lambda \in \mathbb{R}, \tag{1.1}$$

where ϕ and Φ are the pdf and cumulated distribution function (cdf) of a standard normal, respectively, and λ is the skewness parameter. We consider the notation $Z \sim SN(\lambda)$. Note that, if $\lambda = 0$ then $f_Z(z) = \phi(z)$. If $\lambda < 0$ the distribution has a negative skewness and if $\lambda > 0$ the skewness is positive.

A more flexible variable can be built by adding location and scale parameters, such as $Y = \xi + \tau Z$. Then, Y has a pdf given by

$$f_Y(y) = 2\frac{1}{\tau}\phi\left(\frac{y-\xi}{\tau}\right)\Phi\left(\lambda\frac{y-\xi}{\tau}\right), \ \xi \in \mathbb{R}, \ \tau > 0. \tag{1.2}$$

We consider the notation $Y \sim SN(\xi, \tau^2, \lambda)$, called the model of three parameters. In this case, the average and the variance for Y are given by

$$E[Y] = \xi + \tau\delta\sqrt{\frac{2}{\pi}} \quad \text{and} \quad \text{var}[Y] = \tau^2[1 - \frac{2}{\pi}\delta^2],$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. Note that, δ can be used as an alternative parametrization with a similar interpretation of λ, whose characteristic is to be bounded ($\mid \delta \mid < 1$).

The skewness index for the standard skew-normal distribution, obtained by Azzalini (1985), is given by

$$\gamma = \sqrt{\frac{2}{\pi}}\left(\frac{4}{\pi} - 1\right)\left(\frac{\lambda}{\sqrt{1+\lambda^2}}\right)^3\left(1 - \frac{2}{\pi}\frac{\lambda^2}{1+\lambda^2}\right)^{-\frac{3}{2}}.$$

Thus, it can be seen that $\mid \gamma \mid < 0.9953$, which indicates the skew-normal (SN) distribution is adequate only to model small and moderate skewness.

Inference about the parameters of this distribution based on the maximum likelihood approach has some problems, such as

(i) The maximum likelihood(ML) estimator for the skewness parameter can be infinite;
(ii) The Fisher information matrix is singular when $\lambda = 0$;
(iii) Existence of local maximum.

To deal with the second problem Azzalini (1985) and Pewsey (2000) proposed a reparametrization, called centered parametrization. The centered parameters are the mean, the variance and the skewness index of the SN distribution. The new parametrization solves the singularity problem of the Fisher information matrix. This paper will focus on the first problem not solved by reparametrization.

Sartori (2006), Liseo (2004) and Liseo and Loperfido (2006) proposed different methods to deal with the bias problem of the ML estimator for the skewness parameter λ, using classical and Bayesian approaches.

Let $y_1, \ldots, y_n$ be a sample from $SN(\xi, \tau^2, \lambda)$. Then the likelihood function is given by

$$L(\xi, \tau^2, \lambda) = \prod_{i=1}^{n} \phi(\frac{y_i - \xi}{\tau})\Phi(\lambda\frac{y_i - \xi}{\tau}). \tag{1.3}$$

We can see from (1.3) that if $\xi = 0, \tau = 1$ and $y_i > 0, \forall i$ (or $y_i < 0, \forall i$) the likelihood function is an increasing (decreasing) monotonous function and the ML will be infinite (minus infinite). Liseo and Loperfido (2006) showed that the cases which the ML is infinite are completely characterized by the samples where all values have the same sign and the probability of this to occur is given by $[(0.5 - arctan\lambda/\pi)^n + (0.5 + arctan\lambda/\pi)^n]$.

Sartori (2006) proposed an estimator to correct bias, based on the Firth (1993) method and proved that this estimator is always finite. Sartori´s estimator $\tilde{\lambda}$ is the solution of the equation

$$U(\tilde{\lambda}) + M(\tilde{\lambda}) = 0,$$

where U is the score function (the first derivative of the log likelihood function) and

$$M(\lambda) = -\frac{\lambda}{2}\frac{a_4(\lambda)}{a_2(\lambda)}, \tag{1.4}$$

with $a_k = E[Z^k(\frac{\phi(\lambda Z)}{\Phi(\lambda Z)})^2], k = 2$ and 4. These expected values are calculated under the standard skew-normal distribution and they need to be numerically computed.

Liseo and Loperfido (2006) proposed Bayesian inference under reference prior, based on Berger and Bernardo (1992) method (BB reference prior). For the uniparametric model, the BB reference prior agrees with the Jeffreys prior, given by the square root of the Fisher information value. Then, the BB reference prior or Jeffreys prior for the standard skew-normal model is given by

$$f^J(\lambda) \propto \sqrt{\int 2x^2\phi(x)\frac{\phi^2(\lambda x)}{\Phi(\lambda x)}dx}. \tag{1.5}$$

Liseo and Loperfido (2006) obtained the following properties for the density (1.5):

(i) $f^J(\lambda)$ is symmetric around $\lambda = 0$ and decreasing in $|\lambda|$;
(ii) the tail of $f^J(\lambda)$ is order $O(\lambda^{-3/2})$;
(iii) the posterior maximum is finite;
(iv) the posterior mean is infinite for the same samples for which the ML is infinite.

Since, from Bayes' theorem, $\log f(\lambda \mid y) = \log h(\lambda) + \log \mathrm{L}(\lambda) + C$, where $h(\lambda)$ is a prior density, $f(\lambda \mid y)$ the posterior density and C do not depend on λ. Then, the derivative of the logarithm of the prior density can be seen as an alternative to the M-factor used by Sartori (2006) to correct the bias problem.

In this paper we propose a good approximation for the bias correction M-factor and the Jeffreys prior. We believe it will make the computational aspects easier for both methods. In addition, we propose an alternative non informative Bayesian analysis using a uniform prior for a reparametrization of the skewness parameter.

In the next section we present the approximations for the Jeffreys prior and the bias correction M-factor, and also discuss a prior choice for the skewness parameter. In Section 3, we develop a Bayesian MCMC algorithm to calculate the posterior distribution using Jeffreys and uniform priors. Section 4 presents an analysis of a data set known in the literature as the frontier data set. In Section 5 we present a study of simulation and compare different point estimators for the skewness parameter. The classical and Bayesian approaches are also compared through interval estimation and hypotheses tests.

2 Prior specification and approximations

A stochastic representation for the skew normal random variable $Z \sim SN(\lambda)$ (Henze, 1986) is given by

$$Z = \sqrt{(1-\delta^2)}V + \delta U, \tag{2.1}$$

where V has a standard normal distribution and U a positive half normal distribution.

Following the principle of Bayes-Laplace, a natural choice for a non informative prior for δ is the uniform distribution. Since δ is bounded, there is a well defined proper uniform density. The uniform distribution will be our first prior proposed for δ. Noticing that, if $\delta \sim U_{[-1,1]}$ then using standard methods of variable transformation we can show that $\lambda \sim t(0, 1/2; 2)$; where $t(a, b; d)$ denotes a Student-t distribution centered in a with scale b and d degree of freedom. The last one is clearly a non vague distribution. Although it is non subjective, since it was not built using our prior knowledge about the parameter λ (see Figure 1). This remark reinforces the known fact that the Bayes-Laplace principle is not invariant by transformation. However, it is also known that Jeffreys prior has this invariance property (see Berger and Bernardo, 1992). The second prior we are considering is the Jeffreys prior given by Liseo and Loperfido (2006). As it is not a friendly expression, we propose an approximation for this Jeffreys prior that will help us to implement the Bayesian inference.

Observing the properties and the graph of the function given by (1.5) we can see the similarity with a Student-t density. In fact, the best t approximation is given by $t(0, \pi^2/4; 1/2)$. To show this we consider the following approximation

$$\frac{1}{\pi}\frac{\phi(z)}{\sqrt{\Phi(z)(1-\Phi(z))}} \approx \frac{1}{\sqrt{2\pi}(\pi/2)}\exp\{-\frac{z^2}{2(\pi^2/4)}\}\ , \tag{2.2}$$

given by Chaibub and Branco (2003). In the Appendix A we show how to obtain (2.2) using the Kulback-Liebler divergence measurement.

The Fisher information value can be written as

$$I(\lambda) = \int_{-\infty}^{\infty} 2z^2\phi(z)\frac{\phi^2(\lambda z)}{\Phi(\lambda z)}dz =$$

$$\int_0^{\infty} 2z^2\phi(z)\frac{\phi^2(\lambda z)}{\Phi(\lambda z)}dz + \int_0^{\infty} 2z^2\phi(z)\frac{\phi^2(\lambda z)}{[1-\Phi(\lambda z)]}dz = \int_0^{\infty} 2z^2\phi(z)\frac{\phi^2(\lambda z)}{\Phi(\lambda z)[1-\Phi(\lambda z)]}dz.$$

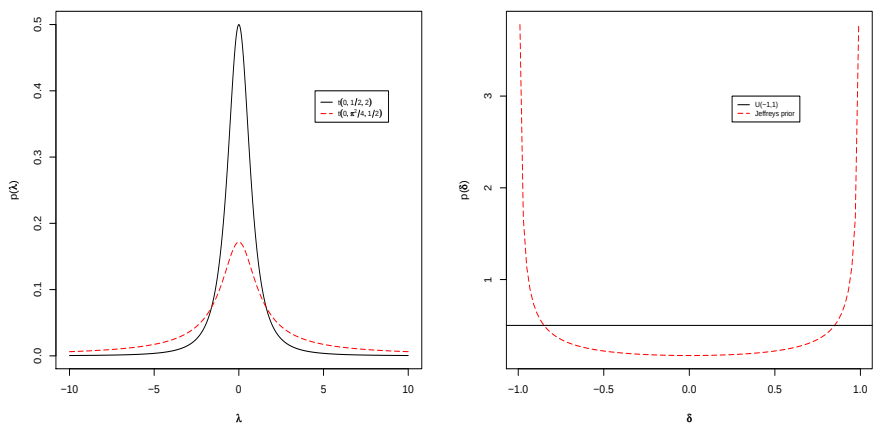
Now, using (2.2) we obtain

$$I(\lambda) \approx \frac{2}{\pi}\left(1 + \frac{2\lambda^2}{\pi^2/4}\right)^{-3/4}.$$

Then,

$$f^J(\lambda) \approx \sqrt{\frac{2}{\pi}}\left(1 + \frac{2\lambda^2}{\pi^2/4}\right)^{-3/2}.$$

The right side of the expression above is the pdf of the $t(0, \pi^2/4; 1/2)$. Then, this t-Student distribution can be use as an approximation for Jeffreys prior.



(a) (b)

Figure 1 *(a) Prior densities function for λ. (b)Prior densities function for δ*

In Figure 1 (a), we can see this prior density and compare with the first prior proposed. As expected the Jeffreys prior is more vague than the one based on the uniform. We also show in Figure 1 (b) the prior density under δ parametrization.

The M-factor proposed by Sartori (2006) can also be approximated using (2.2). This will be shown in Proposition 2.3.

The next proposition provides an approximation for $a_k = E\left[z^k \left(\frac{\phi(\lambda z)}{\Phi(\lambda z)}\right)^2\right]$, where the expected values are evaluated under the standard skew-normal distribution.

The proofs of all propositions are given in Appendix A.

Proposition 2.1 *When k is even, a good approximation for a_k is given by*

$$a_k \approx \frac{1}{2^{\frac{k}{2}-1}\pi} \frac{k!}{(k/2)!} \left(1 + \frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{k+1}{2}}. \tag{2.3}$$

Proposition 2.2 *When k is odd, a good approximation for a_k is given by*

$$a_k \approx -\frac{2^{\frac{k+2}{2}}}{\pi^{\frac{3}{2}}}\left(\frac{k-1}{2}\right)!\lambda\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{k+2}{2}}\left(1+\lambda^2\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-1}\right)^{-\frac{1}{2}}\times$$
$$\times\left\{\sum_{j=0}^{\frac{k-1}{2}} c_j\lambda^j\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{j}{2}}\right\}, \tag{2.4}$$

with $c_0 = 1$, $c_j = \frac{2j-1}{2j}c_{j-1}$.

Proposition 2.3 *A good approximation for the Sartori M-factor is given by*

$$M(\lambda) = -\frac{\lambda}{2}\frac{a_4(\lambda)}{a_2(\lambda)} \approx -\frac{3\lambda}{2}\left[1+\frac{2\lambda^2}{(\pi^2/4)}\right]^{-1}.$$

In Figure 2 we can see the behavior of this function and the quality of this approximation.

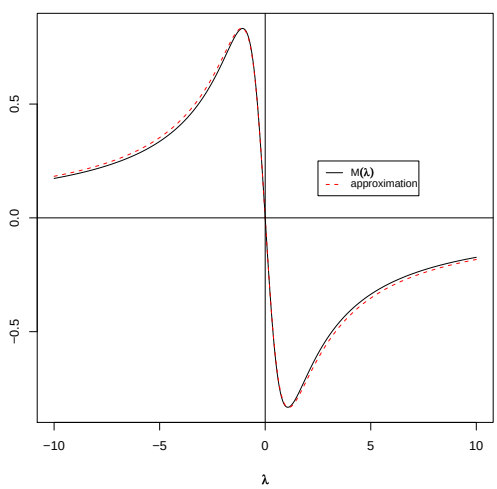


Figure 2 *Exact M-factor and approximation*

3 Bayesian inference

For the Bayesian inference we consider the stochastic representation given by (2.1), following the hierarchical representation

$$\begin{aligned} &Y_i|U_i,\lambda,\xi,\tau \sim N(\xi+\tau\delta U_i,\tau^2(1-\delta^2)),\\ &U_i \sim HN(0,1)\ ,\quad i=1,\ldots,n, \end{aligned} \tag{3.1}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Considering the usual Jeffreys prior for the location (ξ) and scale (τ) parameters and prior independence between the parameters, following the prior specification

$$p(\lambda,\xi,\tau) \propto \frac{1}{\tau}p(\lambda), \tag{3.2}$$

where $p(\lambda)$ is a $t(0,b;d)$ density. When $d=\frac{1}{2}$ and $b=\frac{\pi^2}{4}$ we have the Jeffreys prior, and when $d=2$ and $b=\frac{1}{2}$ we have the prior induced by the uniform distribution.

We consider the reparametrization $\beta=\tau\delta$ and $\eta=\tau\sqrt{1-\delta^2}$ to facilitate the computation. For this, we have the following hierarchical model

$$\begin{aligned} &Y_i|U_i,\xi,\eta,\beta \sim N(\xi+\beta U_i,\eta^2),\\ &U_i \sim HN(0,1),\quad i=1,\ldots,n. \end{aligned} \tag{3.3}$$

Considering the prior specifications given by (3.2) and using usual variable transformation methods, we induce the following prior distribution for the new parametrization,

$$f(\xi,\beta,\eta,w) \propto \frac{1}{\eta^2}\exp\left(-\frac{1}{2}\frac{\beta^2 w}{\eta^2 b}\right) w^{\frac{d+1}{2}-1}\exp\left(-\frac{d}{2}w\right). \tag{3.4}$$

From (3.3) and (3.4), we obtain the complete posterior conditional distributions given by

$$
\begin{aligned}
&w|\xi,\beta,\eta,U,Y \sim \text{Gamma}\left(\frac{d+1}{2},\frac{1}{2}\left(\frac{\beta^2}{b\eta^2}+d\right)\right)\\
&u_i|\xi,\beta,\eta,w,Y \sim N\left(\frac{(y_i-\mu)\beta}{\eta^2+\beta^2},\frac{\eta^2}{\eta^2+\beta^2}\right),\quad \forall i, i=1,...,n,\\
&\xi|\beta,\eta,w,U,Y \sim N\left(\frac{\sum_{i=1}^{n}(y_i-\beta u_i)}{n},\frac{\eta^2}{n}\right),\\
&\beta|\xi,\eta,w,U,Y \sim N\left(\frac{\sum_{i=1}^{n}(y_i-\xi)u_i}{\frac{w}{b}+\sum_{i=1}^{n}u_i^2},\frac{\eta^2}{\frac{w}{b}+\sum_{i=1}^{n}u_i^2}\right),\\
&\frac{1}{\eta^2}|\mu,\beta,w,U,Y \sim \text{Gamma}\left(\frac{n+1}{2},\frac{1}{2}\left(\frac{w\beta^2}{b}+\sum_{i=1}^{n}(y_i-\mu-\beta u_i)^2\right)\right).
\end{aligned}
\tag{3.5}
$$

The complete conditional distributions are used to implement a simple Gibbs sampling algorithm used in the next section and for the simulation study in Section 5.

4 Analysis of the frontier data set

In Azzalini´s webpage there is an interesting data set, called frontier data, which was used also by other authors. The frontier data consists of a simulated random sample of size 50 from a $SN(0,1,5)$. For this data set the ML estimator of λ is infinite although the histogram of the data (see Figure 3) shows that a finite value for λ fits this data set better. Using the bias correction method, Sartori (2006) obtains $\tilde{\lambda} = 9.14$. Considering our approximation for the M-factor we obtain $\tilde{\lambda} = 8.67$. They are pretty close to each other, specially considering that Sartori´s estimative is evaluated numerically.

The Bayesian point estimators given by posterior mean and median are obtained using the Gibbs sampling algorithm. For the posterior mode the *optim* routine of the R software was used.

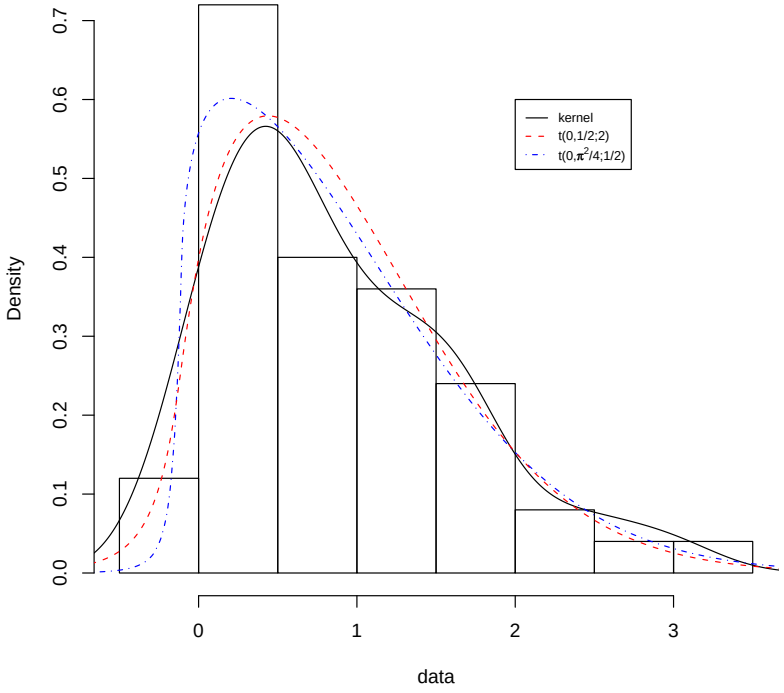


Figure 3 *Frontier data set with kernel estimate and posterior predictive distributions for two different priors*

Table 2 shows that the estimates for the parameters ξ and τ did not change much, but the estimates for the parameters λ are very sensitive to prior´s choice. For this data set the uniform prior provides better results, specially when considered the posterior median. Furthermore, under the Jeffreys´ prior the best estimate was given by the posterior mode. Moreover, the non existence of the posterior mean and the high value of the posterior median suggest a heavy tail posterior for λ, and an evident skewness.

In Figure 3, we have the data histogram with a kernel estimator and the posterior predictive distributions, using different priors. We can see that the predictive distribution based in the uniform prior is closer to the kernel estimate. Finally, both predictive distributions are much better than the half normal distribution ($\lambda = \infty$).

Table 2 *Bayesian and ML estimatives for the frontier data set.*

		Jeffreys			Uniform		
Param	ML	Mode	Mean	Median	Mode	Mean	Median
λ	∞	6.85	∞	31.27	3.97	7.61	5.26
ξ	-0.11	-0.04	-0.10	-0.11	0.02	0.06	-0.02
τ	1.51	1.35	1.61	1.58	1.47	1.30	1.27

5 Simulation study

In our simulation study we generated 10000 samples with sample sizes $n = 30, 50, 100$ and 200 of a $SN(\lambda)$, for different values of λ. For each one, the following estimators were considered: ML, posterior mode, posterior mean and posterior median, under the two prior specification. We also obtained the classical asymptotic 95 % confidence interval and the Bayesian HPD credibility interval, with probability 0.95.

Table 3 and Table 4 present the empirical mean bias and mean square error, respectively. The last column of both tables shows the percentage of samples that were used to obtain the ML and Jeffreys posterior mean, since samples with infinite estimates for these estimators were not considered. From these tables we can see that the posterior mean under uniform prior presents lower values of the bias and of the mean square error in general.

It is convenient to emphasize that the posterior mode under Jeffreys prior also showed a good behaviour as a point estimator. Sartori (2006) noted that in full exponencial model, the estimator proposed by Firth (1993) coincides with the posterior mode under Jeffreys prior. In the present simulation study, we also calculated the Sartori corrected estimator, $\tilde{\lambda}$, and observed that it is very close to the Jeffreys posterior mode.

Table 5 and Table 6 present the mean lower limit(ll) and the mean upper limit(ul) of the interval estimates. The columns denoted by $0 \in I$ present the percentage of samples in which zero is included in the interval, and the columns denoted by $\lambda \in I$ the percentage of samples in which the real value of λ is included in the interval.

For the asymptotic intervals (Table 5) we observe an unusual fact, the percentage of samples that contains the zero increases for big values of λ. This fact is explained by the large sizes of the intervals. This unpleasant situation does not occur with the Bayesian HPD intervals. In Table 6, as expected, the percentage of samples that contains zero decreases when λ increases. Therefore, we observe that the uniform prior produces shorter intervals than Jeffreys prior without damaging its credibility.

Table 3 *Bias of the Bayesian and ML estimates for λ.*

			Prior I : Jeffreys			Prior II : uniform			
n	λ	ML	Mode	Mean	Median	Mode	Mean	Median	%
30	0,5	0,05	0,01	0,04	0,03	-0,04	-0,01	-0,02	100,00
	1	0,14	0,03	0,12	0,10	-0,09	0,00	-0,03	99,97
	2	0,57	0,07	0,52	0,60	-0,31	0,08	-0,07	99,23
	3	1,16	-0,01	1,05	1,85	-0,75	0,15	-0,24	96,18
	5	2,38	-0,59	2,21	5,24	-2,07	0,00	-1,01	86,26
	8	3,43	-2,32	3,20	9,59	-4,59	-1,09	-2,92	69,95
	10	3,62	-3,77	3,77	11,59	-6,43	-2,22	-4,48	61,98
50	0,5	0,03	0,00	0,02	0,01	-0,02	-0,01	-0,01	100,00
	1	0,07	0,01	0,06	0,04	-0,05	-0,01	-0,02	100,00
	2	0,27	0,05	0,23	0,19	-0,16	0,02	-0,04	99,95
	3	0,87	0,06	0,67	0,56	-0,39	0,06	-0,11	99,73
	5	1,87	0,00	1,77	3,28	-1,21	0,29	-0,37	95,93
	8	4,45	-0,83	3,48	8,55	-3,18	0,13	-1,48	86,53
	10	4,91	-1,80	4,25	11,69	-4,77	-0,40	-2,60	79,86
100	0,5	0,01	0,00	0,01	0,01	-0,01	0,00	-0,01	100,00
	1	0,03	0,01	0,03	0,02	-0,02	0,00	-0,01	100,00
	2	0,10	0,02	0,09	0,06	-0,08	0,00	-0,03	100,00
	3	0,25	0,04	0,22	0,16	-0,15	0,02	-0,04	100,00
	5	0,87	0,13	0,79	0,72	-0,47	0,12	-0,09	99,86
	8	2,43	0,16	2,22	3,41	-1,45	0,40	-0,35	98,26
	10	3,96	-0,01	3,53	6,75	-2,41	0,62	-0,71	95,33
200	0,5	0,01	0,00	0,00	0,00	-0,01	0,00	0,00	100,00
	1	0,02	0,00	0,01	0,01	-0,01	0,00	0,00	100,00
	2	0,05	0,00	0,04	0,03	-0,04	-0,01	-0,02	100,00
	3	0,10	0,02	0,09	0,07	-0,07	0,00	-0,02	100,00
	5	0,30	0,04	0,26	0,19	-0,21	0,01	-0,06	100,00
	8	0,87	0,11	0,76	0,62	-0,59	0,07	-0,15	99,95
	10	1,67	0,23	1,46	1,47	-0,96	0,22	-0,20	99,79

Finally, we consider here the hypothesis test problem. Two pairs of hypotheses are considered $H_0 : \lambda = 0$ vs $H_1 : \lambda \neq 0$, to test the skewness of the data set, and $H_0 : \lambda = \lambda^*$ vs $H_1 : \lambda \neq \lambda^*$, where λ^* is the true parameter value.

From the classical perspective, the statistic considered for the test is the likelihood ratio in favour of H_0, given by

$$LR = 2\left[\sum_{i=1}^{n} log\Phi(\widehat{\lambda} z_i) - \sum_{i=1}^{n} log\Phi\left(\lambda^* z_i\right)\right], \tag{5.1}$$

where $\hat{\lambda}$ is the maximum likelihood estimator for λ.

Table 4 *Mean square error of the Bayesian and ML estimates fo λ.*

			Prior I : Jeffreys			Prior II : uniform			
n	λ	ML	Mode	Mean	Median	Mode	Mean	Median	%
30	0,5	0,32	0,28	0,31	0,30	0,25	0,28	0,27	100,00
	1	0,57	0,43	0,55	0,72	0,35	0,44	0,40	99,97
	2	2,19	0,92	2,42	3,83	0,61	1,10	0,80	99,23
	3	4,61	1,27	4,12	8,05	0,98	1,77	1,13	96,18
	5	11,27	1,80	9,62	14,64	2,18	2,58	1,73	86,26
	8	26,02	3,06	22,43	20,87	4,65	3,21	3,30	69,95
	10	18,17	4,37	23,27	23,20	6,48	3,87	4,76	61,98
50	0,5	0,22	0,21	0,22	0,22	0,20	0,20	0,20	100,00
	1	0,34	0,30	0,33	0,32	0,28	0,30	0,29	100,00
	2	0,96	0,66	0,92	1,37	0,52	0,70	0,61	99,95
	3	16,80	1,14	5,78	3,93	0,81	1,32	1,01	99,73
	5	7,02	2,06	7,97	14,27	1,56	2,91	1,81	95,93
	8	79,11	2,97	17,74	24,43	3,38	4,25	2,75	86,53
	10	40,59	3,60	25,00	28,90	4,92	4,66	3,57	79,86
100	0,5	0,15	0,15	0,15	0,15	0,14	0,15	0,15	100,00
	1	0,21	0,20	0,21	0,20	0,19	0,20	0,19	100,00
	2	0,45	0,40	0,45	0,43	0,38	0,40	0,39	100,00
	3	0,96	0,73	0,92	0,84	0,63	0,72	0,68	100,00
	5	3,92	1,73	3,68	6,69	1,25	1,97	1,57	99,86
	8	9,95	3,36	9,42	19,38	2,25	4,39	2,88	98,26
	10	35,64	4,07	18,75	28,35	3,12	5,79	3,58	95,33
200	0,5	0,10	0,10	0,10	0,10	0,10	0,10	0,10	100,00
	1	0,15	0,14	0,14	0,14	0,14	0,14	0,14	100,00
	2	0,28	0,27	0,28	0,27	0,26	0,27	0,26	100,00
	3	0,49	0,45	0,49	0,47	0,43	0,45	0,44	100,00
	5	1,14	0,98	1,12	1,07	0,89	0,97	0,94	100,00
	8	3,10	2,30	2,97	4,37	1,81	2,43	2,11	99,95
	10	11,05	3,51	7,61	13,52	2,46	4,03	3,12	99,79

When $\lambda^* = 0$, we denote this statistic by LR_0 and it is given by

$$LR_0 = 2\left[\sum_{i=1}^{n} log\Phi(\widehat{\lambda} z_i) + nlog2\right]. \tag{5.2}$$

From the Bayesian perspective we can use the regular Bayes Factor without problem, since the prior is proper. The Bayes Factor in favour of H_0 is given by

$$BF = \frac{\prod_{i=1}^{n} \Phi(\lambda^* z_i)}{\int_0^\infty \Phi_n(0|I_n + \frac{b}{w}\boldsymbol{z}\boldsymbol{z}^T)\text{Gamma}(w|\frac{d}{2}, \frac{d}{2})dw}, \tag{5.3}$$

Table 5 *The percentage of intervals which contains zero or the real value of λ, for the asymptotic 95 % confidence interval.*

n	λ	ll	ul	$0 \in I$	$\lambda \in I$	n	λ	ll	ul	$0 \in I$	$\lambda \in I$
	0,5	0,00	1,10	50,00	95,84		0,5	0,23	0,80	3,46	94,93
	1	0,33	1,96	3,86	97,08		1	0,64	1,43	0,00	95,79
	2	0,50	4,65	2,49	96,78		2	1,32	2,88	0,00	96,03
30	3	-0,11	8,43	9,89	96,48	100	3	1,88	4,62	0,01	96,68
	5	-3,38	18,13	34,87	95,87		5	2,51	9,22	0,42	96,75
	8	-11,92	34,77	67,32	94,58		8	1,98	18,88	3,84	96,58
	10	-12,42	39,66	80,81	92,94		10	-3,06	30,99	8,19	95,75
	0,5	0,11	0,94	25,67	95,64		0,5	0,30	0,71	0,00	95,24
	1	0,49	1,65	0,08	95,96		1	0,74	1,29	0,00	95,10
	2	1,02	3,51	0,06	97,35		2	1,52	2,58	0,00	95,78
50	3	-0,89	8,63	0,98	96,74	200	3	2,21	4,00	0,00	95,78
	5	0,20	13,55	8,89	96,59		5	3,40	7,20	0,00	96,06
	8	-24,00	48,91	29,87	95,44		8	4,76	12,98	0,04	96,73
	10	-13,79	43,60	45,45	95,74		10	4,94	18,40	0,38	96,66

where $\text{Gamma}(w|a,b)$ denotes a gamma density of probability with parameters a and b, and $\Phi_n(0 \mid \Sigma)$ is a distribution function of a n-dimensional normal with zero mean and covariance matrix Σ, evaluated at zero vector.

Again, when $\lambda^* = 0$, the Bayes factor is simplified and given by

$$BF_0 = \frac{1}{2^n \int\limits_0^\infty \Phi_n(0|I_n + \frac{b}{w}\boldsymbol{z}\boldsymbol{z}^T)\text{Gamma}(w|\frac{d}{2},\frac{d}{2})dw}. \tag{5.4}$$

For details on how to obtain this expression see the Appendix B.

Tables 7, 8 and 9 present the percentage of times that the hypotheses H_0 was accepted, for each value of sample n and skewness λ. For hypotheses test purpose we present only samples generated with $\lambda \leq 1$, because for values bigger than 1 the percentage of acceptance H_0 is almost 100 % (or 0 %).

We emphasize the difference between the Bayesian and classical criteria to decide in favour of H_0. We follow here what is usually done under these two different perspectives. Under the classical approach we fix the significancy level in 5 % and define a critical region. Under the Bayesian perspective, we consider evidence in favour of H_0 when $BF > 2$, and consider evidence in favour of H_1 when $BF < 0.5$.

Table 6 *The percentage of intervals which contains zero or the real value of λ, for the 95 % HPD interval.*

		Prior I : Jeffreys				Prior II : Uniform			
n	λ	ll	ul	$0 \in I$	$\lambda \in I$	ll	ul	$0 \in I$	$\lambda \in I$
	0,5	0,02	1,11	49,00	95,60	-0,01	1,04	54,00	96,50
	1,0	0,39	2,05	4,00	96,00	0,32	1,82	6,00	96,30
	2,0	0,86	21,60	0,00	97,10	0,69	3,76	0,00	94,10
30	3,0	1,09	257,46	0,00	95,90	0,86	6,70	0,00	90,60
	5,0	1,26	664,20	0,00	96,80	0,99	11,41	0,00	90,10
	8,0	1,26	1209,19	0,00	94,40	1,01	16,84	0,00	83,40
	10,0	1,17	1804,99	0,00	96,30	0,99	20,16	0,00	84,50
	0,5	0,11	0,93	25,80	94,80	0,10	0,89	27,40	95,00
	1,0	0,52	1,67	0,20	95,50	0,47	1,58	0,30	95,30
	2,0	1,10	3,48	0,00	95,40	0,97	3,15	0,00	93,90
50	3,0	1,54	21,04	0,00	96,50	1,32	5,20	0,00	93,50
	5,0	2,06	264,33	0,00	96,10	1,68	10,41	0,00	91,20
	8,0	2,31	936,63	0,00	96,80	1,84	18,07	0,00	88,60
	10,0	2,33	1492,75	0,00	95,60	1,87	22,26	0,00	87,60
	0,5	0,23	0,80	3,60	95,10	0,22	0,78	4,00	95,00
	1,0	0,65	1,43	0,00	95,10	0,62	1,39	0,00	94,30
	2,0	1,35	2,88	0,00	95,40	1,28	2,76	0,00	94,20
100	3,0	1,94	4,57	0,00	93,80	1,81	4,30	0,00	92,80
	5,0	2,93	66,26	0,00	96,20	2,61	8,16	0,00	93,70
	8,0	3,82	243,03	0,00	96,40	3,25	14,90	0,00	91,70
	10,0	4,24	737,98	0,00	95,20	3,52	21,32	0,00	89,60
	0,5	0,30	0,70	0,00	94,60	0,30	0,69	0,00	94,30
	1,0	0,75	1,29	0,00	94,00	0,73	1,27	0,00	93,50
	2,0	1,53	2,58	0,00	97,00	1,50	2,52	0,00	96,80
200	3,0	2,24	4,02	0,00	95,40	2,17	3,90	0,00	95,10
	5,0	3,47	7,14	0,00	95,20	3,30	6,82	0,00	94,20
	8,0	5,06	13,00	0,00	95,60	4,64	11,98	0,00	95,10
	10,0	5,98	40,07	0,00	96,60	5,36	15,98	0,00	94,50

Table 7 shows a good behaviour of the likelihood ratio test. Tables 8 and 9, show the effect of the prior choice on the Bayes factor. The decision under uniform prior leads us to reject H_0 offen.

Table 7 *Percentage of times of the H_0 was accepted using the ratio likelihood test.*

n	λ	$H_0 : \lambda = 0$	$H_0 : \lambda = \lambda^*$	n	λ	$H_0 : \lambda = 0$	$H_0 : \lambda = \lambda^*$
	0	94,19	94,19		0	94,96	94,96
	0,1	92,13	94,41		0,1	86,61	94,76
	0,25	80,42	94,80		0,25	51,07	95,02
30	0,3	73,80	94,54	100	0,3	35,71	94,68
	0,5	44,88	93,84		0,5	3,10	94,57
	0,75	16,24	94,18		0,75	0,00	94,95
	1	3,95	93,89		1	0,00	94,88
	0	94,50	94,50		0	94,76	94,76
	0,1	91,45	94,82		0,1	79,59	94,83
	0,25	72,10	94,28		0,25	21,04	94,70
50	0,3	61,47	94,45	200	0,3	8,83	94,86
	0,5	23,14	94,91		0,5	0,02	94,97
	0,75	3,39	94,58		0,75	0,00	94,64
	1	0,23	94,72		1	0,00	94,49

6 Final comments

In this paper we provided a good approximation for the Jeffreys prior, the Fisher information and also for the terms $a_k = E[Z^k(\frac{\phi(\lambda Z)}{\Phi(\lambda Z)})^2]$ which we believe will be helpful for new researches in this area. The Bayesian approach using the uniform prior for δ parametrization is a solution for the challenge proposed by Azzalini with the frontier data set.

Therefore, the simulation study confirms the advantage of using this prior distribution. We recommend the uniform posterior mean as an alternative to classical point estimator, since the former was shown to have lower bias and mean square error. The Bayesian inference under uniform prior also showed better behaviour for interval estimation. However, the Bayes Factor under the uniform prior is very conservative, it has a tendency to reject H_0 more than other methods, even when H_0 is true. We believe it will be interesting for a future research to work with other methods for model choice as, for example, the DIC criterion, given by Spiegelhalter et al. (2002).

Table 8 *Percentage of time of the $H_0 : \lambda = 0$ is accepted (or rejected) using the Bayes Factor.*

		Prior I : Jeffreys		Prior II : Uniform	
n	λ	$BF_0 > 2$	$BF_0 < 0.5$	$BF_0 > 2$	$BF_0 < 0.5$
30	0,00	92,33	1,52	74,30	3,93
	0,10	89,49	2,76	68,38	5,78
	0,25	76,26	8,06	50,40	15,90
	0,30	69,12	11,90	42,38	21,52
	0,50	38,61	33,74	16,42	49,73
	0,75	11,71	67,74	2,85	80,67
	1,00	2,39	89,30	0,29	95,26
50	0,00	94,17	1,22	80,75	3,74
	0,10	90,89	2,45	74,62	5,76
	0,25	70,65	12,08	46,59	21,83
	0,30	60,61	18,32	36,59	30,97
	0,50	23,51	53,27	8,25	68,75
	0,75	2,64	89,68	0,48	95,54
	1,00	0,15	98,84	0,00	99,69
100	0,00	96,43	0,74	87,71	2,26
	0,10	90,61	3,07	76,66	6,68
	0,25	56,55	23,04	34,45	36,59
	0,30	41,10	35,90	21,36	51,74
	0,50	4,84	85,76	1,17	92,98
	0,75	0,06	99,70	0,00	99,93
	1,00	0,00	100,00	0,00	100,00
200	0,00	97,77	0,47	91,50	1,52
	0,10	87,28	4,64	73,26	9,80
	0,25	30,83	49,01	14,96	64,17
	0,30	14,44	69,02	5,56	81,94
	0,50	0,05	99,56	0,00	99,93
	0,75	0,00	100,00	0,00	100,00
	1,00	0,00	100,00	0,00	100,00

A Approximations

Approximation (2.2):

By observing the plot of the density function $\frac{1}{\pi}\,[\Phi(x)]^{-1/2}[1-\Phi(x)]^{-1/2}\phi(x)$ we noticed that this density has a shape similar to a normal density centered in zero. A possible way to estimate the value of the variance σ^2 that gives the best approximation, is to minimize the symmetric version of the Kullback-

Liebler divergence between $f_1(x) = \frac{1}{\pi}\,[\Phi(x)]^{1/2-1}[1-\Phi(x)]^{1/2-1}\phi(x)$ and $f_2(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{x^2}{2\,\sigma^2}\right\}$.

Table 9 *Percentage of times of the $H_0 : \lambda = \lambda^*$ is accepted (or rejected) using the Bayes Factor.*

		Prior I : Jeffreys		Prior II : Uniform	
n	λ	$BF > 2$	$BF < 0.5$	$BF > 2$	$BF < 0.5$
30	0,00	92,33	1,52	74,30	3,93
	0,10	92,03	1,73	73,29	4,24
	0,25	91,77	1,97	75,60	4,49
	0,30	92,62	1,61	76,54	3,72
	0,50	92,25	1,47	79,93	3,36
	0,75	92,20	1,74	84,10	2,89
	1,00	91,84	1,84	87,77	2,30
50	0,00	94,17	1,22	80,75	3,74
	0,10	94,40	1,23	81,28	3,26
	0,25	94,40	0,99	82,14	3,13
	0,30	94,54	1,17	82,69	2,94
	0,50	94,74	1,07	85,50	2,55
	0,75	94,23	1,21	88,58	2,21
	1,00	94,25	1,22	90,40	1,75
100	0,00	96,43	0,74	87,71	2,26
	0,10	96,26	0,84	88,28	2,35
	0,25	96,11	0,70	88,51	2,39
	0,30	96,28	0,90	88,41	2,21
	0,50	96,08	0,82	90,02	2,10
	0,75	96,54	0,81	92,69	1,60
	1,00	96,53	0,75	94,20	1,16
200	0,00	97,77	0,47	91,50	1,52
	0,10	97,65	0,44	92,11	1,47
	0,25	97,56	0,52	92,21	1,48
	0,30	97,70	0,55	92,49	1,48
	0,50	97,72	0,35	93,90	1,21
	0,75	97,54	0,52	94,83	1,00
	1,00	97,12	0,50	95,63	0,87

The symmetric version of the Kullback-Liebler divergence between two probability densities is defined by

$$J[f_1 : f_2] = K[f_1 : f_2] + K[f_2 : f_1],$$

where $K[f_1 : f_2] = \int f_1(x) \log\{f_1(x)/f_2(x)\}\, dx$ is the Kullback-Liebler's divergence measure between f_1 and f_2. Differently from K, J is a symmetric measure because $J[f_1 : f_2] = J[f_2 : f_1]$.

Using Monte Carlo integration we calculated $J[f_1 : f_2]$ for 5000 different values of σ ranging from 1 to 2. Beginning with $\sigma = 1$ our algorithm calculated the respective $J_1[f_1 : f_2]$ and update the next value considering $\sigma = \sigma + 0.0002$, this procedure was repeated 5000 times until it reached the value $\sigma = 2$. Thus, we were able to numerically estimate the value of σ that minimizes $J[f_1 : f_2]$ and we found $\sigma = 1.54$.

A simpler alternative to estimate σ is to consider the value of σ that satisfies

$$\frac{1}{\pi}\,[\Phi(x)]^{-1/2}[1-\Phi(x)]^{-1/2}\,\phi(x) \;=\; \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{x^2}{2\,\sigma^2}\right\},$$

for an adequate value of x. By setting $x = 0$ we found that $\sigma = \pi/2 \approx 1.57$. Note that, this estimate is very close to the previous estimate. In this paper we adopted the value $\sigma = \pi/2$. From Figure 4 we can appreciate the precision of this approximation. The continuous line is the normal density and the density $\frac{1}{\sqrt{2\pi}(\pi/2)}\exp\left\{-\frac{x^2}{2(\pi^2/4)}\right\}$ is the dashed line.

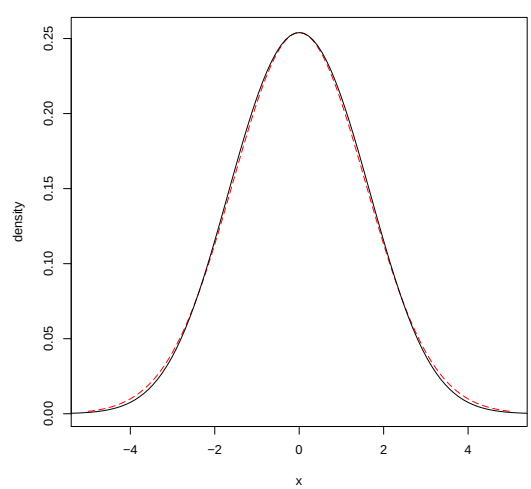


Figure 4 *Approximation (2.2)*

Proof. Proposition 2.1

Note that

$$\begin{aligned}
a_k &= \int_{-\infty}^{\infty} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz \\
&= \int_{0}^{\infty} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz + \int_{-\infty}^{0} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz \\
&= \int_{0}^{\infty} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz + \int_{0}^{\infty} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{1-\Phi(\lambda z)} dz \\
&= \int_{0}^{\infty} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)\left[1-\Phi(\lambda z)\right]} dz.
\end{aligned}$$

Then, using (2.2) it follows that

$$\begin{aligned}
a_k &\approx \int_{0}^{\infty} 2z^k \phi(z) \frac{\pi}{2} \frac{1}{(\pi^2/4)} \exp\left(-\frac{\lambda^2 z^2}{(\pi^2/4)}\right) dz \\
&= \frac{\pi}{(\pi^2/4)} \left(1 + \frac{2\lambda^2}{(\pi^2/4)}\right)^{-1/2} \int_{0}^{\infty} z^k \phi\left(z|0, \left(1 + \frac{2\lambda^2}{(\pi^2/4)}\right)^{-1}\right) dz \\
&= \frac{1}{2^{\frac{k}{2}-1}\pi} \frac{k!}{(k/2)!} \left(1 + \frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{k+1}{2}},
\end{aligned}$$

where $\phi(z \mid a, b)$ denotes the pdf of a $N(a, b)$.

Proof. Proposition 2.2

Note that

$$\begin{aligned}
a_k &= \int_{-\infty}^{\infty} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz \\
&= \int_{0}^{\infty} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz + \int_{-\infty}^{0} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz \\
&= \int_{0}^{\infty} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz - \int_{0}^{\infty} 2z^k \phi(z) \frac{\phi^2(\lambda z)}{1-\Phi(\lambda z)} dz
\end{aligned}$$

$$= \int_0^\infty 2z^k\phi(z)\phi^2(\lambda z)\left(\frac{1}{\Phi(\lambda z)} - \frac{1}{1-\Phi(\lambda z)}\right)dz$$

$$= \int_0^\infty 2z^k\phi(z)\frac{\phi^2(\lambda z)}{\Phi(\lambda z)\left[1-\Phi(\lambda z)\right]}dz - \int_0^\infty 4z^k\phi(z)\frac{\phi^2(\lambda z)}{\Phi(\lambda z)\left[1-\Phi(\lambda z)\right]}\Phi(\lambda z)dz.$$

Then, using (2.2) it follows that

$$a_k \approx \int_0^\infty 2z^k\phi(z)\frac{\pi}{2}\frac{1}{(\pi^2/4)}\exp\left(-\frac{\lambda^2 z^2}{(\pi^2/4)}\right)dz - \int_0^\infty 4z^k\phi(z)\frac{\pi}{2}\frac{1}{(\pi^2/4)}\exp\left(-\frac{\lambda^2 z^2}{(\pi^2/4)}\right)\Phi(\lambda z)dz$$

$$= \frac{4}{\pi}\int_0^\infty z^k\phi\left[\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{1/2} z\right]dz - \frac{8}{\pi}\int_0^\infty z^k\phi\left[\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{1/2} z\right]\Phi(\lambda z)dz.$$

From the transformation $w = \left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{\frac{1}{2}} z$ it follows that

$$a_k \approx \frac{4}{\pi}\left(\frac{2}{\pi}\right)^{\frac{3}{2}}\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{k+1}{2}}\int_0^\infty w^k\phi(w)dw$$

$$- \frac{8}{\pi}\left(\frac{2}{\pi}\right)^{\frac{3}{2}}\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{k+1}{2}}\int_0^\infty w\phi(w)\Phi\left(\lambda\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{1}{2}} w\right)dw.$$

Finally, using Owen (1980, p.394, 399), we obtain

$$a_k \approx \frac{2^{\frac{k+2}{2}}}{\pi^{\frac{3}{2}}}\left(\frac{k-1}{2}\right)!\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{k+1}{2}} - \frac{2^{\frac{k+2}{2}}}{\pi^{\frac{3}{2}}}\left(\frac{k-1}{2}\right)!\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{k+1}{2}}$$

$$\times\left[1+\lambda\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{1}{2}}\left(1+\lambda^2\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-1}\right)^{-\frac{1}{2}}\left\{\sum_{j=0}^{\frac{k-1}{2}} c_j\lambda^j\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{j}{2}}\right\}\right]$$

$$= -\frac{2^{\frac{k+2}{2}}}{\pi^{\frac{3}{2}}}\left(\frac{k-1}{2}\right)!\lambda\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{k+2}{2}}\left(1+\lambda^2\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-1}\right)^{-\frac{1}{2}}$$

$$\times\left\{\sum_{j=0}^{\frac{k-1}{2}} c_j\lambda^j\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{j}{2}}\right\}.$$

In particular a_1 is approximated by

$$a_1 \approx -\left(\frac{2}{\pi}\right)^{\frac{3}{2}}\lambda\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-\frac{3}{2}}\left(1+\lambda^2\left(1+\frac{2\lambda^2}{(\pi^2/4)}\right)^{-1}\right)^{-\frac{1}{2}}.$$

Proof. Proposition 2.3
It is straight forward from Proposition 2.1 .

B The Bayes factor

Lemma B.1 *If* $V \sim N_k(\eta, \Sigma)$*, then* $E\left[\Phi_m(a + AV|\gamma, \Gamma)\right] = \Phi_m(a|\gamma - A\eta, \Gamma + A\Sigma A^T)$*.*

Proof. See Gupta et al. (2004).

Proposition B.1 *If* $\lambda \sim t(0, b; d)$*, then the Bayes factor is given by*

$$BF = \frac{\prod_{i=1}^{n} \Phi(\lambda^* z_i)}{\int_0^\infty \Phi_n(0|I_n + \frac{b}{w}\boldsymbol{z}\boldsymbol{z}^T)\, Gamma(w|\frac{d}{2}, \frac{d}{2})dw}. \tag{B.1}$$

Proof. The Bayes Factor in favour of H_0 is given by

$$BF = \frac{\prod_{i=1}^{n} \Phi(\lambda^* z_i)}{\int_{-\infty}^{\infty} t(\lambda \mid 0, b, d) \prod_{i=1}^{n} \Phi(\lambda z_i) d\lambda}.$$

Then, using the hierarchical form of the Student-t distribution, we rewrite the above expression as

$$BF = \frac{\prod_{i=1}^{n} \Phi(\lambda^* z_i)}{\int_{-\infty}^{\infty} \int_0^\infty \phi(\lambda \mid 0, \frac{b}{w}) \prod_{i=1}^{n} \Phi(\lambda z_i) \mathrm{Gamma}(w|\frac{d}{2}, \frac{d}{2}) dw d\lambda}.$$

Changing the order integration, we have

$$BF = \frac{\prod_{i=1}^{n} \Phi(\lambda^* z_i)}{\int_0^\infty E_S\left[\Phi_n(\lambda \boldsymbol{z})\right] \mathrm{Gamma}(w \mid \frac{d}{2}, \frac{d}{2}) dw},$$

where $S \sim N(0, \frac{b}{w})$.

Finally, using Lemma B.1, we obtain

$$BF = \frac{\prod_{i=1}^{n} \Phi(\lambda^* z_i)}{\int_0^\infty \Phi_n(0|I_n + \frac{b}{w}\boldsymbol{z}\boldsymbol{z}^T)\text{Gamma}(w|\frac{d}{2}, \frac{d}{2})dw}.$$

Acknowledgements

The authors are thankful to a referee and the editor for valuable suggestions. The first author was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) during his master program under supervision of the second author.

(Received November, 2005. Accepted November, 2006.)

References

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

Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors. *Bayesian Statistics 4.* Oxford University Press, 35-60.

Chaibub Neto, E. and Branco, M. D.(2003). Bayesian reference analysis for binomial calibration problem. *Technical Report, IME-USP,* 2003-12.

Dalla Valle, A. (2004). The skew-normal distribution. *In Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Genton, M.G., Ed., Chapman and Hall/CRC, Boca Raton, FL, pp. 3−24.

Firth, D. (1993). Bias redution of maximum likelihood estimates. *Biometrika*, **80**, 27−38.

Gupta, A. K., Gonzalez-Farias, G. and Dominguez-Molina, J.A (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis*, **89**, 181-190.

Liseo, B. (2004). Skew-elliptical distributions in Bayesian inference. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Genton, M. G., Ed., Chapman and Hall / CRC, Boca Raton, FL, pp. 153-171.

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.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, **27**, 859−870.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates: skew normal and skew-t distributions. *Journal of Statisitical Planning and Inference*, **136**, 4259-4275.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures od model complexity and fit. *Journal Royal Statistical Society B*, **64**, 583-639.

Cristian Luis Bayes and **Marcia D'Elia Branco**
Departamento de Estatística
Instituto de Matemática e Estatística
Universidade de São Paulo
Rua do Matão,1010
Cidade Universitária.
05508-090 São Paulo/SP, Brazil
E-mails: cbayesr@ime.usp.br and mbranco@ime.usp.br