Bayesian inference for the multivariate skew-normal model: a Population Monte Carlo approach

Brunero Liseo and Antonio Parisi

Abstract Frequentist and likelihood based methods of inference encounter several difficulties with the multivariate skew-normal model. In spite of the popularity of this class of densities, there are no broadly satisfactory solutions for estimation and testing problems. In this paper we propose a general population Monte Carlo algorithm which exploits the stochastic representation of the skew-normal random variables in terms of latent structure to provide a full Bayesian analysis of the model. Our approach can be defined weakly informative since we use priors which approximate the actual reference prior for the shape parameter vector. We compare our results with the existing classical solutions and illustrate the practical implementation of the algorithm.

Key words: Bayes factor, Integrated likelihood, Monte Carlo, Objective Bayes inference, Reference prior, Skewness

1 Introduction

The Skew Normal (SN hereafter) model has been formally introduced in [2] and it was then generalized to the multivariate case by [5]. The main reason for the popularity of this class of distributions lies in the fact that it is able to capture, and explicitly model, mild departures from symmetry, without losing mathematical tractability; this can be particularly useful in real data applications.

Brunero Liseo
MEMOTEF - Sapienza Università di Roma

Antonio Parisi
Dipartimento SEFEMEQ, Università degli Studi di Roma "Tor Vergata"
Via Columbia 2 - 00133 Roma - Italy
e-mail: antonio.parisi@uniroma2.it Tel: +39 06 7259 5914 Fax: +39 06 20 40 219

The SN class is also very popular because it naturally arises in real data analysis when some special mechanism of data collection arises, i.e. hidden truncation or selective reporting: see [1] for a complete account of these aspects. A deeper analysis of the huge amount of literature, however, reveals that most of the existing results are related with the distributional theory of skew-normal and, more generally, skew-elliptical distributions. On the other hand, the theory of inference is still problematic even in the scalar cases [4]. These problems were anticipated since [2] and they are basically due to some anomalies of the likelihood function. These difficulties tend to be more challenging in the multivariate set-up where, in addition, "problematic" situations are not so easy to detect. Even ignoring these pathological cases, the likelihood surface arising from an i.i.d. sample of skew-normal random variables is often non regular and MLE computations tend to be unstable.

For all these reason, we propose to adopt a full Bayesian analysis of the multivariate SN model. In particular we propose to use objective priors, in order to correct the odd behavior of the likelihood function without introducing external information, and to exploit the latent structure of the SN model in order to tailor a specific version of a Monte Carlo algorithm and to produce valid posterior inferences, in terms of estimation and testing.

2 Augmented likelihood function and priors

In this paper we will consider the multivariate SN model. A random vector $\mathbf{X}$ is said to have a p-dimensional standard SN distribution, with correlation matrix Ω and shape parameter α when the density function is

$$f\left(\mathbf{x};\xi,\Omega,\alpha\right) = 2\varphi_p\left(\mathbf{x};\Omega\right)\cdot\Phi_1\left[\alpha'\mathbf{x}\right], \qquad \mathbf{x},\alpha\in\mathbb{R}^p. \tag{1}$$

It is easy to add location and scale parameters: let ξ a p-dimensional vector and let

$$\omega = \mathrm{diag}\left(\omega_\mathbf{1},\ldots,\omega_\mathbf{p}\right)$$

be the "vector" of the marginal scale parameters, that is $\Sigma = \omega\Omega\omega$ represents the scale matrix. Then $\mathbf{Y} = \xi + \omega\mathbf{X}$ has a p-dimensional SN distribution ($SN_p(\Sigma,\xi,\alpha)$, hereafter) with density

$$f(\mathbf{y};\xi,\Sigma,\alpha) = 2\varphi_p(\mathbf{x}-\xi;\Sigma)\Phi_1\left[\alpha'\omega^{-1}(\mathbf{x}-\xi)\right]$$

In this parameterization, each component of the shape parameter α can take any real value. There exists an alternative parameterization [3] defined in terms of $\delta\in[-1,1]^p$, given by

$$\delta = (1+\alpha'\Omega\alpha)^{-\frac{1}{2}}\Omega\alpha. \tag{2}$$

We adopt the same parameterization used in [6], which exploits the intrinsically latent structure of the SN density function so to produce an augmented likelihood function. In fact, the likelihood surface is not really manageable and the explicit introduction of latent variable allows us to turn back to a Gaussian context.

Although this parameterization is more suitable in terms of practical implementation and computation, we prefer the original parameterization for the elicitation of priors. Our primary goal here is to propose a general method of inference for the parameters of the multivariate SN distribution. For these reasons we have tried to be as much "objective" as possible in choosing the prior for the parameter vector. However, it is not easy to derive a formal Jeffreys or reference prior for the parameters of a multivariate skew-normal distribution. In this paper we have assumed a priori, as usual, $\xi \perp\!\!\!\perp (\delta, \Sigma)$. Also we assumed a flat prior for the "location" parameter ξ and a conjugate Normal Inverse Wishart prior for the "scale" parameter Σ;

$$\pi(\xi) \propto 1 \qquad \text{and} \qquad \Sigma \sim IW_p(m, \Lambda).$$

Obviously, one can always consider the limiting case ($m \to 0, \Lambda \to 0$) to get the classical Jeffreys prior

$$\pi(\xi, \mathbf{G}) \propto \frac{1}{|\, \Sigma \,|^{\frac{p+1}{2}}}.$$

The choice of a good objective prior for δ is more delicate. in the univariate SN model, [7] have shown that the Jeffreys' prior for the shape parameter α is proper; its use, in a sense, automatically and pragmatically solves the problem of potentially non vanishing likelihood function, which can happen with the SN model (see [3]).

In the multivariate case, an objective analysis can be done only for one component of the shape vector. For practical purposes, a prior can be chosen in the following way: in the scalar case the approximate Jeffreys' prior for $\beta = (1+\delta)/2$, with $\delta = \alpha/\sqrt{1+\alpha^2}$ is a Beta$(0.25, 0.25)$ prior; in analogy with that, one can use, in the multivariate case, the prior

$$\pi^{IND}(\delta) \propto \prod_{j=1}^{p} \left(1 - \delta_j^2\right)^{-\frac{3}{4}}, \tag{3}$$

that is, we assume that the components of the skewness vector are, a priori, independent and identically distributed. Although independence can be considered a strong assumption, it is hard to conceive any non subjective form of dependence. Notice also that any prior distribution on δ should be considered only for those values which satisfy the positive definiteness of $\mathbf{G}$. In this perspective we will consider the parameter constraint as generated from the prior rather than from the likelihood. In the rest of the paper, we will then consider, as the prior for δ,

$$\tilde{\pi}^{IND}(\delta \mid \Omega) = \frac{1}{A(\Omega)} \prod_{j=1}^{p} \left(1 - \delta_j^2\right)^{-\frac{3}{4}}, \tag{4}$$

where $A(\Omega)$ is the integral of (3) over the parameter values such that $\det(\Omega) > 0$. In a simulation framework, the function $A(\Omega)$ should be evaluated for any proposed vector of parameters, and this translates in a very high computational burden. Hence, we propose an approximation of the function $A(\Omega)$ which represents a very efficient solution to this problem.

3 Simulations

A population Monte Carlo sampler (PMC, hereafter) is employed to obtain a sample from the joint posterior distribution of θ. We preferred not to adopt a MCMC strategy in order to avoid the problems that usually arise when simulating a Markov chain to explore the posterior distribution of the parameters of a model with latent variables. In particular, the coexistence of different particles, and the competition between them, allows to better consider the issue of multimodality of the posterior density. In similar situations the Gibbs sampler tends to be attracted by one of the modes and hardly escapes from a neighborhood of it. Besides, from a model selection perspective, the estimation of the normalising constant can be performed as a simple by-product of any PMC (and MC) sampler. Finally, the use of MC algorithms allows the simoultaeous draw of all the particles: this fact dramatically improves the efficiency of the algorithm compared with generic MCMC approaches.

We consider the frequentist properties of our Bayesian procedure: even though the median of the sampling distributions of the MLE is quite precise, this kind of estimator always shows a non-negligible probability of obtaining infinite values for the estimates. The Bayesian approach, via the use of a weakly informative prior, eliminates this problem and provides estimates with good frequentist coverage properties.

References

1. Arnold, B.C., Beaver, R.J.: Skewed multivariate models related to hidden truncation and/or selective reporting. Test **11**(1), 7–54 (2002). With discussion and a rejoinder by the authors
2. Azzalini, A.: A class of distributions which includes the normal ones. Scand. J. Statist. **12**, 171–178 (1985)
3. Azzalini, A., Capitanio, A.: Statistical applications of the multivariate skew-normal distributions. J. R. Statist. Soc. B **61**, 579–602 (1999)
4. Azzalini, A., Capitanio, A.: Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t* distribution. J. R. Statist. Soc. B **65**, 367–389 (2003)
5. Azzalini, A., Dalla Valle, A.: The multivariate skew-normal distribution. Biometrika **83**, 715–726 (1996)
6. Fruhwirth-Schnatter, S., Pyne, S.: Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. Biostatistics (to appear) (2010)
7. Liseo, B., Loperfido, N.: A note on the reference prior for the scalar skew normal distribution. J. Statist. Plann. Inference **136**(2), 373–389 (2006)