International Journal of Approximate Reasoning 55 (2014) 1235–1251

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Bayesian robustness under a skew-normal class of prior distribution

Luciana Graziela de Godoi [a,*], Marcia D'Elia Branco [b]

[a] *Departamento de Estatística, Universidade Federal do Espírito Santo, Brazil*
[b] *Departamento de Estatística, Universidade de São Paulo, Brazil*

ARTICLE INFO

Article history:
Received 26 July 2013
Received in revised form 3 March 2014
Accepted 7 March 2014
Available online 13 March 2014

Keywords:
Bayesian robustness
Global sensitivity
Contamination class
Skew distributions

ABSTRACT

We develop a global sensitivity analysis to measure the robustness of the Bayesian estimators with respect to a class of prior distributions. This class arises when we consider multiplicative contamination of a base prior distribution. A similar structure was presented by van der Linde [12]. Some particular specifications for this multiplicative contamination class coincide with well known families of skewed distributions. In this paper, we explore the skew-normal multiplicative contamination class for the prior distribution of the location parameter of a normal model. Results of a Bayesian conjugation and expressions for some measures of distance between posterior means and posterior variance are obtained. We also elaborate on the behavior of the posterior means and of the posterior variances through a simulation study.

1. Introduction

Bayesian sensitivity analysis is concerned with the impact of different specifications for the prior distribution or the likelihood function on the posterior distribution. If, for instance, a specific posterior inference is not much affected by these choices, then we will say that this inference is robust. This approach is also known as robust Bayesian analysis.

In general, the robustness analysis supposes the likelihood $f(x \mid \theta)$ is fixed and consider a class Γ of prior distributions to deal with the uncertainty in specifying one prior distribution. Robustness of a given statistical procedure is measured by the size of the range of posterior measures obtained when the prior distribution varies over Γ. If this size is small, then we say that the inference is robust under the class of prior distributions Γ considered and we conclude that the posterior inference is not affected by a particular choice in Γ. This approach is called global robustness.

A possible class of priors Γ is obtained by a 'contamination' of an elicited base prior $f_0(\theta)$. In this context, a well-studied class of priors is the ϵ-contaminated class given by $\Gamma = \{f\colon\ f(\theta) = (1-\epsilon) f_0(\theta) + \epsilon g(\theta),\ \epsilon \in [0,1] \text{ and } g \in \mathcal{G}\}$ where θ is the parameter of interest and $\mathcal{G}$ is a class of contaminating density functions. This class Γ may be interpreted as an additive perturbation of the base density function $f_0(\theta)$ (see [15]). Clearly Γ depends on both the level ϵ of contamination and the contaminating density function $g(\cdot)$. Berger [6,7] discusses various other kinds of contamination.

Analogous to additive contamination, van der Linde [12] proposed multiplicative perturbation in both the likelihood and the prior distribution. Here we explore this idea with a multiplicative class of contaminated priors given by $\Gamma_M = \{f\colon\ f(\theta) = f_0(\theta) w(\theta) \text{ for some } w \in \mathcal{G}\}$, where $\mathcal{G}$ is a collection of non-negative functions such that $f_0(\theta)w(\theta)$ is a density function for each $w \in \mathcal{G}$. In addition, we assume that $\mathcal{G}$ contains the constant function $w(\theta) = 1$, $\forall\theta$, to ensure that Γ_M contains the base prior density $f_0(\theta)$.

* Corresponding author.

http://dx.doi.org/10.1016/j.ijar.2014.03.004

About the motivation behind the proposal to work with a multiplicative class of contamination, the main interest in working with this kind of contamination is the ability to change the symmetry of the distribution. For example, if $f_0(\theta)$ is symmetrical and unimodal then $w(\theta)$ induces an asymmetrization in the base density. Dey and Liu [8] discuss the use of asymmetric priors in the context of prior elicitation from expert opinion. When the quantile specified by the experts indicate a relationship like $q_{0.75} - q_{0.5} > q_{0.5} - q_{0.25}$ or $q_{0.75} - q_{0.5} > q_{0.5} - q_{0.25}$, where q_p is the p-quantile, an asymmetric prior is appropriate. However for easy of calculations a symmetric prior has been usually considered. A question of interest is to understand the impact on the posterior estimates of this assumption.

For particular specifications, the multiplicative contamination class coincides with families of skewed distributions known in the literature. For example, let Γ_M be given by the multiplicative contamination class, where $f_0(\boldsymbol{\theta})$ is given by $f_p(\boldsymbol{\theta}; \boldsymbol{\xi}_{\mathbf{V}}, \boldsymbol{\Omega}_{\mathbf{V}}, h^p) = |\boldsymbol{\Omega}_{\mathbf{V}}|^{-1/2} h^{(p)}(v(\boldsymbol{\theta}))$ denotes the density function of an elliptically contoured distribution with location $\boldsymbol{\xi}_{\mathbf{V}} \in \mathbb{R}^p$, positive definite $p \times p$ dispersion matrix $\boldsymbol{\Omega}_{\mathbf{V}}$, density generator $h^{(p)}$ and $v(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{\xi}_{\mathbf{V}})^T \boldsymbol{\Omega}_{\mathbf{V}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\xi}_{\mathbf{V}})$. For $w(\boldsymbol{\theta})$, we considered the function

$$\frac{F_q(\boldsymbol{\xi}_{\mathbf{U}} + \boldsymbol{\Delta}^T \boldsymbol{\Omega}_{\mathbf{V}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\xi}_{\mathbf{V}}); \mathbf{0}, \boldsymbol{\Omega}_{\mathbf{U}} - \boldsymbol{\Delta}^T \boldsymbol{\Omega}_{\mathbf{V}}{}^{-1} \boldsymbol{\Delta}, h^{(q)}_{v(\boldsymbol{\theta})})}{F_q(\boldsymbol{\xi}_{\mathbf{U}}; \mathbf{0}, \boldsymbol{\Omega}_{\mathbf{U}}, h^{(q)})},$$

where Δ is a $q \times p$ real matrix controlling shape. The function $F_r(\mathbf{x}; \mathbf{0}, \Sigma, h^{(r)})$ denotes the r-dimensional centered elliptical cumulative distribution with $r \times r$ dispersion matrix Σ and density generator $h^{(r)}$, and $h^{(q)}_{v(\boldsymbol{\theta})}(u) = h^{(p+q)}\{u + v(\boldsymbol{\theta})\}/h^{(p)}(v(\boldsymbol{\theta}))$. Then, the prior distribution of $\boldsymbol{\theta}$ is a unified skew elliptical distribution (*SUE*), denoted here by $\theta \sim SUE_{p,q}(\boldsymbol{\xi}_{\mathbf{V}}, \boldsymbol{\Omega}_{\mathbf{V}}, \boldsymbol{\Delta}, h^{(p+q)}, \boldsymbol{\xi}_{\mathbf{U}}, \boldsymbol{\Omega}_{\mathbf{U}})$. The *SUE* distribution has been introduced in [1] and it is a particular case of distributions generated through selection mechanisms. More details about such selection mechanisms can be found in [2].

A particular distribution of *SUE* is obtained when the function $h^{(p)}$ is the p-variate normal generator function given by $(2\pi)^{-p/2}e^{-u/2}$, for $u > 0$. In this situation, we obtain the unified skew-normal (*SUN*) distribution (see [1]) and denote it by $SUN_{p,q}(\boldsymbol{\xi}_{\mathbf{V}}, \boldsymbol{\Omega}_{\mathbf{V}}, \boldsymbol{\Delta}, \boldsymbol{\tau}_{\mathbf{U}}, \boldsymbol{\Omega}_{\mathbf{U}})$. The univariate skew-normal distribution $SUN_{1,1}(\mu, \sigma^2, \lambda\sigma^2, 0, \sigma^2(1+\lambda^2))$ is the distribution whose density function is given by $f(\theta) = 2\phi(\theta; \mu, \sigma^2)\Phi(\lambda\theta, \lambda\mu, \sigma^2)$, where $\phi(\cdot; \mu, \sigma^2)$ and $\Phi(\cdot; \mu, \sigma^2)$ corresponds to the density and the cumulative distribution functions of a normal distribution with location parameter μ and scale parameter σ^2, respectively. The most common notation about this distribution in the literature is $SN(\mu, \sigma^2, \lambda)$ where the λ represents the shape parameter controlling skewness. The standard skew-normal distribution is obtained when $\mu = 0$ and $\sigma^2 = 1$. In this situation, $f(\theta)$ corresponds to a density function of a standard skew-normal distribution with the shape parameter λ, as proposed by Azzalini [4]. We denote this distribution by $SN(\lambda)$.

Mukhopadhyay and Vidakovic [14] consider the skewed prior distributions to study the performance of linear Bayes rules in estimating a normal mean. They showed that linear Bayes rules have reasonably well in comparison to exact Bayes rules when the prior distribution is given by $SN(\lambda)$. To justify the use of skew prior distributions, they study the situation where the parameter space is truncated, for example $\Theta = [\theta_0, \infty)$, but the truncation point θ_0 is unknown. According to the authors, 'one way to incorporate this prior information about the parameter is elicitation of a prior with lighter tails on the truncated side' and they suggest to consider a family of skewed prior distributions and to study the results in the robust Bayesian point of view.

In the context of contamination class of prior, if we consider $\Phi(.)$ fixed, $\phi(\theta; \mu, \sigma^2)$ as the base prior distribution and the parameter λ varying over the $\mathbb{R}$, then λ reflects the degree of contamination of the base prior, indicating less contamination when it is near zero. The base prior distribution is obtained when $\lambda = 0$. The robustness study in the multiplicative contamination class is developed considering how the changes in λ can affect the posterior distribution of θ or posterior measures of this distribution.

Through hierarchical Bayes approach, Liseo and Loperfido [13] consider a skew prior distribution in the multivariate case, but they do not explore robust Bayesian procedures.

In this work, we discuss the robustness of posterior mean and posterior variance under the skew-normal class of prior for the location parameter of a normal distribution $N(\theta, \tau^2)$, in two situations: when τ^2 is known and unknown. In Section 2 we review some properties of a general skew-normal distribution and present new results of Bayesian conjugation. In Section 3 we obtain expressions for some robustness measures. In Section 4 we perform a simulation study to analyze the measures obtained and discuss a possible conflict between the information provided by the likelihood and by the prior distribution. We present our conclusions in Section 5.

The symbols $\phi(.)$ and $\Phi(.)$ correspond respectively to the density and the cumulative distribution functions of the standard normal distribution.

2. The general multivariate skew-normal distribution and results of Bayesian conjugation

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with location parameter θ and scale parameter τ^2.

Usually, the prior distribution specified for θ is the normal, that is conjugated to the statistical model considered. The idea here is to propose a class of prior distributions that contains the normal distribution, but allows the inclusion of the