Statistics and Probability Letters 96 (2015) 154–161

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On the propriety of the posterior of hierarchical linear mixed models with flexible random effects distributions

F.J. Rubio

University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK

ARTICLE INFO

Article history:
Received 4 July 2014
Received in revised form 22 September 2014
Accepted 23 September 2014
Available online 2 October 2014

Keywords:
Half-Cauchy prior
Improper priors
Skew-normal
Two-piece normal

ABSTRACT

The use of improper priors in the context of Bayesian hierarchical linear mixed models has been studied under the assumption of normality of the random effects. We study the propriety of the posterior under more flexible distributional assumptions and general improper prior structures.

1. Introduction

Hierarchical linear mixed models (LMM) are often used to account for parameter variation across groups of observations. The general formulation of this type of models is given by the model equation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of data, $\boldsymbol{\beta}$ is a $p \times 1$ vector of *fixed effects*, $\mathbf{u}$ is a $q \times 1$ vector of *random effects*, $\mathbf{X}$ and $\mathbf{Z}$ are known design matrices of dimension $n \times p$ and $n \times q$, respectively, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of residual errors. We assume that $\text{rank}(\mathbf{X}) = p$ and $n > p$ throughout. The random vectors $\boldsymbol{\varepsilon}$ and $\mathbf{u}$ are typically assumed to be normally distributed. Given that the normality assumption can be restrictive in practice, alternative distributional assumptions have been explored, such as finite mixtures of normal distributions (Zhang and Davidian, 2001), scale mixtures of skew-normal distributions (Lachos et al., 2010), and Bayesian nonparametric approaches (Dunson, 2010). However, in a Bayesian context with improper priors, only the model with normal assumptions has been studied. In this line, Hobert and Casella (1996) analysed the following hierarchical structure:

$$\begin{aligned}
\boldsymbol{\varepsilon}|\sigma_0 &\sim N_n(\mathbf{0}, \sigma_0^2 \mathbf{I}_n)\\
\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma_0 &\sim N_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma_0^2 \mathbf{I}_n),\\
\mathbf{u}|\sigma_1, \dots, \sigma_r &\sim N_q(\mathbf{0}, \mathbf{D}),
\end{aligned} \tag{2}$$

with an improper prior structure as follows:

$$\pi(\boldsymbol{\beta}, \sigma_0, \sigma_1, \dots, \sigma_r) \propto \prod_{i=0}^{r} \frac{1}{\sigma_i^{2a_i+1}}, \tag{3}$$

E-mail address: Francisco.Rubio@warwick.ac.uk.

http://dx.doi.org/10.1016/j.spl.2014.09.023
0167-7152/

where $\boldsymbol{u} = (\mathbf{u}_1^\top, \ldots, \mathbf{u}_r^\top)^\top$, $\mathbf{u}_i$ is $q_i \times 1$, $\sum_{i=1}^r q_i = q$, $\mathbf{D} = \oplus_{i=1}^r \sigma_i^2 \mathbf{I}_{q_i}$, $i = 1, \ldots, r$, and $N_d(\mathbf{m}, \mathbf{S})$ denotes a d-variate normal distribution with mean $\boldsymbol{m}$ and covariance matrix $\mathbf{S}$. The r subvectors of $\mathbf{u}$ correspond to the r different random factors in the experiment. These assumptions imply that the random effects are assumed to be independent. This independence assumption restricts the applicability of this sort of models, however, it covers some models of interest in meta-analysis (Hobert and Casella, 1996; Tan and Hobert, 2009). Hobert and Casella (1996) obtained necessary and sufficient conditions for the propriety of the corresponding posterior distribution. This kind of improper priors may be of interest in the lack of strong prior information, although their use often represents a controversial topic. In particular, the choice $a_0 = 0$ and $a_1 = \cdots = a_r = -1/2$ is referred to as the *standard diffuse prior* (Tan and Hobert, 2009). Sun et al. (2001) studied an extended improper prior structure which allows for assigning either proper or improper priors on the parameters of the distributions of the residual errors and the random effects as follows:

$$\pi(\boldsymbol{\beta}, \sigma_0, \sigma_1, \ldots, \sigma_r) \propto \prod_{i=0}^r \frac{1}{\sigma_i^{2a_i+1}} \exp(-b_i/\sigma_i^2). \tag{4}$$

This prior corresponds to assigning an inverse gamma distribution to the σ_i^2's when $a_i > 0$ and $b_i > 0$, and it contains the prior structure (2) for $b_0 = \cdots = b_r = 0$ (note that here we are presenting the priors on $(\sigma_0, \sigma_1, \ldots, \sigma_r)$ rather than the equivalent priors on $(\sigma_0^2, \sigma_1^2, \ldots, \sigma_r^2)$ presented in Hobert and Casella, 1996 and Sun et al., 2001).

In this paper, we explore extensions of the hierarchical structure (2)–(4) by considering more flexible random effect distributions. In Section 2, we present an extension of this model for the case when the random effects are distributed according to a two-piece normal distribution. We show that the corresponding posterior is proper essentially under the conditions presented in Sun et al. (2001) for the model with normal assumptions (2)–(4). In Section 3, we characterise a rich class of parametric distributions, which are obtained by adding a shape parameter to the normal distribution, that also preserves the existence of the posterior distribution when they are used for modelling the random effects in (1). In Section 4, we introduce an alternative improper prior structure with proper heavy-tailed priors on the scale parameters $(\sigma_1, \ldots, \sigma_r)$. We conclude with a discussion on the implementation of this sort of models as well as possible extensions of this work.

2. Two-piece normal random effects

A random variable $U \in \mathbb{R}$ is said to be distributed according to a two-piece normal distribution, denoted $U \sim \text{TPN}(\mu, \sigma, \gamma)$, if its density function can be written as Arellano-Valle et al. (2005):

$$s(u|\mu, \sigma, \gamma) = \frac{2}{\sigma[a(\gamma) + b(\gamma)]} \left[\phi\left(\frac{u-\mu}{\sigma b(\gamma)}\right) I(u < \mu) + \phi\left(\frac{u-\mu}{\sigma a(\gamma)}\right) I(u \geq \mu)\right], \tag{5}$$

where ϕ denotes the standard normal density function, I denotes the indicator function, and $\{a(\cdot), b(\cdot)\}$ are positive differentiable functions. This density is continuous, unimodal, with mode at $\mu \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}_+$, and skewness parameter $\gamma \in \Gamma \subset \mathbb{R}$. It coincides with the normal density when $a(\gamma) = b(\gamma)$, and it is asymmetric for $a(\gamma) \neq b(\gamma)$ while keeping the normal tails in each direction. The most common choices for $a(\cdot)$ and $b(\cdot)$ correspond to the *inverse scale factors* parameterisation $\{a(\gamma), b(\gamma)\} = \{\gamma, 1/\gamma\}$, $\gamma \in \mathbb{R}_+$ (Fernández and Steel, 1998), and the ϵ*-skew* parameterisation $\{a(\gamma), b(\gamma)\} = \{1-\gamma, 1+\gamma\}$, $\gamma \in (-1, 1)$ (Mudholkar and Hutson, 2000).

Consider now the following extension of the hierarchical structure (2),

$$\begin{aligned}
&\boldsymbol{\varepsilon}|\sigma_0 \sim N_n(\mathbf{0}, \sigma_0^2 \mathbf{I}_n)\\
&\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma_0 \sim N_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma_0^2 \mathbf{I}_n),\\
&u_{ik}|\sigma_i, \gamma_i \overset{ind.}{\sim} \text{TPN}(0, \sigma_i, \gamma_i), \quad k = 1, \ldots, q_i,\ i = 1, \ldots, r,
\end{aligned} \tag{6}$$

with prior structure

$$\pi(\boldsymbol{\beta}, \sigma_0, \sigma_1, \ldots, \sigma_r, \gamma_1, \ldots, \gamma_r) \propto \frac{1}{\sigma_0^{2a_0+1}} \exp(-b_0/\sigma_0^2) \prod_{i=1}^r \frac{\pi(\gamma_i)}{\sigma_i^{2a_i+1}} \exp(-b_i/\sigma_i^2), \tag{7}$$

where $\pi(\gamma_i)$ are proper priors, and $\mathbf{u}_i = (u_{i1}, \ldots, u_{iq_i})^\top$. The following result presents necessary and sufficient conditions for the propriety of the corresponding posterior.

Theorem 1. *Consider the LMM (1) with hierarchical structure (6)–(7), and denote* $t = \text{rank}\{(\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top)\mathbf{Z}\}$. *Suppose that there exist constants* $m, M > 0$ *such that* $h(\gamma) = \min\{a(\gamma), b(\gamma)\} < m$ *and* $H(\gamma) = a(\gamma) + b(\gamma) > M$. *Consider also the following conditions:*

(a) *For* $i = 1, \ldots, r$ *either* $a_i < b_i = 0$ *or* $b_i > 0$,
(b1) $q_i + 2a_i > 0$,
(b2) $q_i + 2a_i > q - t$ *for all* $i = 1, \ldots, r$,
(c1) $n - p + 2\sum_{i=0}^r a_i > 0$,