# Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution

Arnošt Komárek *,1,2, Emmanuel Lesaffre [2]

*Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, B–3000 Leuven, Belgium*

**Abstract**

Generalized linear mixed models are popular to regress a discrete response when there is clustering, e.g. in longitudinal studies or in hierarchical data structures. It is standard to assume that the random effects have a normal distribution. Recently, it has been examined whether wrongly assuming a normal distribution for the random effects is important for the estimation of the fixed effects parameters. While it has been shown that misspecifying the distribution of the random effects has a minor effect in the context of linear mixed models, the conclusion for generalized mixed models is less clear. Some studies report a minor impact, while others report that the assumption of normality really matters especially when the variance of the random effect is relatively high. Since it is unclear whether the normality assumption is truly satisfied in practice, it is important that generalized mixed models are available which relax the normality assumption. A replacement of the normal distribution with a mixture of Gaussian distributions specified on a grid whereby only the weights of the mixture components are estimated using a penalized approach ensuring a smooth distribution for the random effects is proposed. The parameters of the model are estimated in a Bayesian context using MCMC techniques. The usefulness of the approach is illustrated on two longitudinal studies using R-functions.

*Key words:* Clustered data, Logistic regression, Longitudinal study, Markov chain Monte Carlo, Poisson regression

* Current address: Dept. of Probability and Mathematical Statistics, Charles University, Sokolovská 83, CZ–186 75, Praha 8, Czech Republic.

*Email addresses:* komarek@karlin.mff.cuni.cz (Arnošt Komárek), emmanuel.lesaffre@med.kuleuven.be (Emmanuel Lesaffre).

[1] Supported by the postdoctoral grant PDM/06/242, Research Funds of K.U.Leuven.

[2] Supported by the Interuniversity Attraction Poles Program P6/03 – Belgian State

## 1 Introduction

The generalized linear mixed model (GLMM) is a popular tool to regress a discrete response when the measurements are clustered (e.g., multicenter clinical trials and longitudinal studies). Let $\boldsymbol{Y}_i = (Y_{i,1}, \ldots, Y_{i,n_i})'$ $(i = 1, \ldots, N)$ be the response vector in the $i$-th cluster or the longitudinal profile of the $i$-th unit and let $\boldsymbol{Y} = (\boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_N)'$ be the whole response vector. In the GLMM, it is assumed that the distribution of the response belongs to an exponential family and the effect of the covariates on the $(i,l)$-th response is modelled using a linear predictor $\eta_{i,l}$ as

$$\eta_{i,l} = h^{-1}\Big\{\mathrm{E}(Y_{i,l} \,|\, \boldsymbol{\beta},\, \boldsymbol{b}_i)\Big\} = \boldsymbol{\beta}'\boldsymbol{x}_{i,l} + \boldsymbol{b}'_i\boldsymbol{z}_{i,l} \qquad (i = 1, \ldots, N, l = 1, \ldots, n_i), \tag{1}$$

where $h^{-1}$ is a known link function, $\boldsymbol{x}_{i,l} \in \mathbb{R}^p$ is a vector of covariates and $\boldsymbol{z}_{i,l} \in \mathbb{R}^q$ represents a subset of covariates for which the effect may vary randomly across units. Further, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is the vector of regression coefficients (fixed effects) and $\boldsymbol{b}_1 = (b_{1,1}, \ldots, b_{1,q})'$, $\ldots$, $\boldsymbol{b}_N = (b_{N,1}, \ldots, b_{N,q})'$ are unit specific zero-mean vectors of random effects. Further, we will denote $\mathbb{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N)'$ the matrix of random effects. For a systematic exposition to GLMM, see, e.g., Molenberghs and Verbeke (2005). In this paper we focus on the most popular GLMMs (1) where the response is either binary or a count. When the response is multinomial response, a multivariate GLMM (see, e.g., Fahrmeir and Tutz, 2001) is necessary, in which case expression (1) must be adjusted accordingly.

It is conventionally assumed that the random effects are normally distributed, i.e., that $\boldsymbol{b}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}_q(\boldsymbol{0}, \mathbb{D})$. However, it has been indicated at several places in the literature (Neuhaus, Hauck, and Kalbfleisch, 1992; Kleinman and Ibrahim, 1998b; Heagerty and Kurland, 2001; Agresti, Caffo, and Ohman-Strickland, 2004; Litière et al., 2008) that, in contrast to the linear mixed model, misspecification of the random effects distribution in GLMM might influence inference on the fixed effects (e.g., the treatment effect) which are usually of primary interest but the situation is not clear. Several authors claim that misspecification of the random effects distribution has often a minor impact on the estimation of the fixed effects. On the other hand, it has been reported that the bias in estimating the fixed effects in a GLMM with a misspecified random effects distribution increases with increasing variability of the random effects and especially when the true random effect distribution is skewed or multimodal. Clearly, an incorrectly assumed random effects distribution influences inference on the random effects. Further, estimated values of the individual random effects, e.g., their empirical Bayes estimates, cannot be used to check the validity of the distributional assumption due to the shrinkage (Verbeke

– Federal Office for Scientific, Technical and Cultural Affairs.

and Lesaffre, 1997).

We remark that the random effects distributional assumption concerns can often be alleviated or outright eliminated by specifying a marginal model estimated using, e.g., generalized estimating equations (GEE) method (Liang and Zeger, 1986). Many modifications of this approach have appeared in the literature since then. For instance recently, Jara, García-Zattera, and Lesaffre (2007a) suggested a Bayesian semiparametric marginal model. However, the interpretation of the parameters of the marginal model is generally different from the interpretation of the parameters in the conditional (random effects) model. Lee and Nelder (2004) provide a discussion on which of the two models is to be preferred with a conclusion that conditional models are fundamental, from which also marginal predictions should be made. For this and other reasons, we will concentrate on conditional models.

For the above mentioned reasons, there is a need for GLMMs with a more flexible random effects distribution. One approach is to estimate the distribution using nonparametric maximum-likelihood method (Laird, 1978) leading to a discrete solution. This approach has been suggested in the context of GLMM by, e.g. Follmann and Lambert (1989), Butler and Louis (1992), Zackin, De Gruttola, and Laird (1996). An alternative approach leading to a continuous random effects distribution is offered by the heterogeneity model of Verbeke and Lesaffre (1996), Fieuws, Spiessens, and Draney (2004), Molenberghs and Verbeke (2005, Chapter 23) who specify the random effects distribution as a heteroscedastic normal mixture with estimated number of components, means and variances of the components. Normal mixtures have also been used in a model for binary data by Caffo, Ming-Wen, and Rohde (2007). Alternatively, Chen, Zhang, and Davidian (2002) used the semi-nonparametric approach of Gallant and Nychka (1987) to specify the distribution of random effects estimated with a Monte Carlo EM algorithm.

The GLMM can also be specified in a Bayesian context. Several authors replaced the normal prior for the random effects and used Bayesian nonparametric methods for this purpose (Bush and MacEachern, 1996, Walker and Mallick, 1997, Mukhopadhyay and Gelfand, 1997, Kleinman and Ibrahim, 1998b, Hanson, 2006, Jara, Hanson, and Lesaffre, 2007b).

Recently, the 'penalized Gaussian mixture (PGM)' approach has been suggested which delivers a flexible model for the random effects distribution. The approach is based on the idea of penalized smoothing promoted by Eilers and Marx (1996). Ghidey, Lesaffre, and Eilers (2004) used the PGM to model the random effects distribution in a linear mixed model. Komárek, Lesaffre, and Hilton (2005) used it as a flexible specification of the distribution of a logarithm of the baseline event time in a survival regression model. Further, Komárek and Lesaffre (2006); Bogaerts and Lesaffre (2007) used the bivari-

ate PGM in a survival models for paired censored data. Finally, Komárek and Lesaffre (2007) exploited the PGM to model all distributional parts in a random effects survival regression model. In summary, in all previous applications, the PGM was involved in models for continuous response. The objective of this paper is to show how the PGM can be used as a model for the random effects distribution in a (multivariate) GLMM leading to the PGM GLMM which is primarily intended to analyze a discrete response. Furthermore, our approach can be used, as above mentioned competitive approaches as a diagnostic tool to check parametric assumptions on the distribution of the random effects. For inference, we will use a simulation based Markov chain Monte Carlo (MCMC) technique. For practical usage of the proposed methodology we prepared a contributed package glmmAK in R (R Development Core Team, 2007).

The rest of the paper is organized as follows. In Section 2, we review the PGM and incorporate it as a model for the random effects distribution in the GLMM. Section 3 discusses inference for the proposed model and related MCMC. In Section 5, we apply our model to two longitudinal studies with binary and count data. Concluding remarks are given in Section 6.

## 2 Penalized Gaussian mixture in the GLMM

### *2.1 PGM distribution of random effects*

We assume that the random effects $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$ are i.i.d. distributed with a density $g(\boldsymbol{b})$ and

$$g(\boldsymbol{b}) = g(b_1, \ldots, b_q) = \frac{1}{\tau_1 \cdots \tau_q}\, g^*(b_1^*, \ldots, b_q^*), \tag{2}$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)'$ is a vector of unknown scale parameters and $\boldsymbol{b}^* = (b_1^*, \ldots, b_q^*)' = (\frac{b_1}{\tau_1}, \ldots, \frac{b_q}{\tau_q})'$ denotes a standardized random effect vector.

The shape of the standardized density $g^*(\boldsymbol{b}^*)$ is modelled flexibly using a penalized approach motivated by the work of Eilers and Marx (1996) in the following way. For the $m$th margin ($m = 1, \ldots, q$), we choose a fine grid of a relatively high number $L_m = 2K_m + 1$ (equal to $31-41$) of *equidistant* knots $\boldsymbol{\mu}_m = (\mu_{m,-K_m}, \ldots, \mu_{m,K_m})'$ centered around zero, i.e., $\mu_{m,0} = 0$. That is, if we denote the distance between two consecutive knots in the $m$th margin as $\delta_m$, we can write $\mu_{m,j_m} = j_m \delta_m$ ($j_m = -K_m, \ldots, K_m$). Further, we choose for each

margin a basis standard deviation $\sigma_m$ and write the model for $g^*$ as:

$$g^*(b_1^*, \ldots, b_q^*) = \frac{1}{\sigma_1 \cdots \sigma_q} \sum_{j_1=-K_1}^{K_1} \cdots \sum_{j_q=-K_q}^{K_q} w_{j_1,\ldots,j_q} \, \varphi\left(\frac{b_1^* - \mu_{1,j_1}}{\sigma_1}\right) \cdots \varphi\left(\frac{b_q^* - \mu_{q,j_q}}{\sigma_q}\right), \tag{3}$$

where $\varphi$ is a density of the standard normal distribution $\mathcal{N}(0,\, 1)$ and $\boldsymbol{w} = \left\{w_{j_1,\ldots,j_q} :\; j_1 = -K_1, \ldots, K_1,\; \ldots,\; j_q = -K_q, \ldots, K_q\right\}$ is a set of $L = L_1 \times \cdots \times L_q$ mixture weights that have to be estimated. In the estimation procedure, a roughness penalty on $\boldsymbol{w}$ will be introduced to avoid identifiability problems and overfitting the data. Hence the name of the method: *penalized Gaussian mixture (PGM).*

Eilers and Marx (1996) proposed to use B-splines as basis functions to smooth unknown curves. However, when estimating a density, it seems natural to use parametric densities, like normals, instead. Moreover, due to the fact that an appropriately standardized B-spline approaches a normal density as its degree tends to infinity (Unser, Aldroubi, and Eden, 1992), the PGM can be viewed as a limiting case of the penalized smoothing with B-splines.

It is seen from equation (3) that the unknown function $g^*$ is expressed as a mixture of densities of the $q$-variate normal distributions with means taken from a multivariate grid of knots $\boldsymbol{M} = \left\{M_{j_1,\ldots,j_q} = (\mu_{1,j_1}, \ldots, \mu_{q,j_q})' :\; j_1 = -K_1, \ldots, K_1,\; \ldots,\; j_q = -K_q, \ldots, K_q\right\}$ and a common diagonal covariance matrix $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_q^2)$. Although all mixture components are uncorrelated, the off-diagonal elements of the covariance matrix of the distribution $g^*$ are not necessarily equal to zero. Indeed, only specific combinations of the weights $\boldsymbol{w}$ would lead to zero correlation, see expression (7).

To obtain a reasonable model for the unknown distribution, the knots should be put in the area where the true random effects density has a significant amount of probability mass. Since scale parameters $\boldsymbol{\tau}$ are included in model (3), the knots should cover an area where a zero-mean and unit-variances distribution has most of its probability mass. To this end, the choice of the boundary knots $\mu_{m,-K_m} \approx -5$, $\mu_{m,K_m} \approx 5$ $(m = 1, \ldots, q)$ usually suffices. Further, the distance $\delta_m$ between the two consecutive knots in the $m$th margin equal to 0.3 is small enough to approximate $g^*$ with satisfactory precision, as is illustrated in Komárek et al. (2005). Finally, the value $\sigma_m = (2/3)\delta_m$ is motivated by the correspondence to cubic B-splines, as explained in Komárek et al. (2005).

Satisfactory conditions for the PGM weights to ensure that (3) is a density are $w_{j_1,\ldots,j_q} > 0$ $(j_1 = -K_1, \ldots, K_1,\; \ldots,\; j_q = -K_q, \ldots, K_q)$ and $\sum_{j_1} \cdots \sum_{j_q} w_{j_1,\ldots,j_q} = 1$. To avoid constrained estimation, the transformed weights $\boldsymbol{a} = \left\{a_{j_1,\ldots,j_q} :\right.$

$j_1 = -K_1, \ldots, K_1, \ \ldots, \ j_q = -K_q, \ldots, K_q\Big\}$ given by

$$a_{j_1,\ldots,j_q} = \log\Big(\frac{w_{j_1,\ldots,j_q}}{w_{0,\ldots,0}}\Big), \qquad w_{j_1,\ldots,j_q}(\boldsymbol{a}) = \frac{\exp(a_{j_1,\ldots,j_q})}{\sum_{k_1}\cdots\sum_{k_q}\exp(a_{k_1,\ldots,k_q})} \tag{4}$$
$$(j_1 = -K_1, \ldots, K_1, \ \ldots, \ j_q = -K_q, \ldots, K_q)$$

are estimated instead of $\boldsymbol{w}$. For identifiability reasons $a_{0,\ldots,0} = 0$.

### *2.2 Mean effect of the random effects covariates and variance components of the random effects*

Unless we impose non-linear constraints on the PGM weights, the means of the margins of the PGM (3) are not necessarily equal to zero and their variances are not necessarily equal to one. This fact has to be taken into account when quantifying the mean effect of the covariates involved in $\boldsymbol{z}_{i,l}$ and the variance components of the random effects.

Let $\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_q^*)'$ be the mean of the PGM (3) and $\mathbb{D}^* = (d_{m,s}^*)_{m=1,\ldots,q,s=1,\ldots,q}$ its covariance matrix. Further, let for a fixed $m \in \{1, \ldots, q\}$ and a fixed $j_m \in \{-K_m, \ldots, K_m\}$, $w_{j_m}^{m+}$ be the weight obtained by summing up $w_{k_1,\ldots,k_{m-1},j_m,k_m,\ldots,k_q}$ over $k_1, \ldots, k_{m-1}, k_{m+1}, \ldots, k_q$. Similarly, let $w_{j_m,j_s}^{m,s+}$ $(m \neq s)$ be the weight obtained by performing the above summation but now fixing two subscripts. With a little algebra, we can derive that

$$\mathrm{E}(b_m^*) = \beta_m^* = \sum_{j_m=-K_m}^{K_m} w_{j_m}^{m+}\mu_{m,j_m}, \tag{5}$$

$$\mathrm{var}(b_m^*) = d_{m,m}^* = \sum_{j_m=-K_m}^{K_m} w_{j_m}^{m+}(\mu_{m,j_m} - \beta_m^*)^2 \ + \ \sigma_m^2, \tag{6}$$

$$\mathrm{cov}(b_m^*, b_s^*) = d_{m,s}^* = \sum_{j_m=-K_m}^{K_m}\sum_{j_s=-K_s}^{K_s} w_{j_m,j_s}^{m,s+}(\mu_{m,j_m} - \beta_m^*)\cdot(\mu_{s,j_s} - \beta_s^*) \tag{7}$$
$$(m = 1, \ldots, q, \ s = 1, \ldots, q, \ m \neq s).$$

In our specification of the model, we assume that the random effect covariate vector $\boldsymbol{z}_{i,l}$ is a subset of the fixed effect covariate vector $\boldsymbol{x}_{i,l}$. Without loss of generality, we will assume that the first $q$ components of $\boldsymbol{x}_{i,l}$ are equal to $\boldsymbol{z}_{i,l}$. We denote the mean effect of the covariates involved in $\boldsymbol{z}_{i,l}$ as $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$, which is equal to the sum of the corresponding fixed effect and the mean of the corresponding random effect, i.e.

$$\gamma_m = \beta_m + \mathrm{E}(b_m) = \beta_m + \tau_m\beta_m^* \qquad (m = 1, \ldots, q). \tag{8}$$

The variance components of the random effects, represented by their covariance matrix $\text{var}(\boldsymbol{b}) = \mathbb{D} = (d_{m,s})_{m=1,\ldots,q,s=1,\ldots,q}$ are equal to

$$\text{var}(b_m) = d_{m,m} = \tau_m^2 \, d^*_{m,m} \qquad (m = 1, \ldots, q), \tag{9}$$

$$\text{cov}(b_m, b_s) = d_{m,s} = \tau_m \tau_s \, d^*_{m,s} \qquad (m = 1, \ldots, q, \; s = 1, \ldots, q, \; m \neq s). \tag{10}$$

## 3 Estimation and inference

It is well known that even with normally distributed random effects, the likelihood of the GLMM (1) cannot be evaluated analytically, except in the special case of a linear mixed model with a normally distributed response. That is, even in relatively simple situations, the classical maximum-likelihood (ML) estimation of parameters of the GLMM must rely on approximations like Gaussian quadrature.

Alternatively, we can specify the model from a Bayesian perspective and use characteristics of the posterior distribution obtained using a simulation based Markov chain Monte Carlo (MCMC) methodology (see, e.g., Robert and Casella, 2004) for the inference. This approach will be followed here. However, we want to stress that the Bayesian specification of the model is used purely for computational feasibility. For this reason, prior distributions for all parameters, except for the mixture weights will be specified as vague. Nevertheless, the priors can easily be adopted if there is any prior information available, e.g., when using the GLMM for a meta-analysis. For the transformed mixture weights $\boldsymbol{a}$, the prior distribution based on the intrinsic Gaussian Markov random field (IGMRF, Rue and Held, 2005) will be used which is a Bayesian form of the roughness penalty (see Lang and Brezger, 2004).

### *3.1 Prior distributions*

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\tau}', \boldsymbol{a}', \boldsymbol{b}'_1, \ldots, \boldsymbol{b}'_N, \boldsymbol{r}'_1, \ldots, \boldsymbol{r}'_N, \boldsymbol{\lambda})'$ be the vector of all model parameters. Namely, it includes traditional parameters $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, $\boldsymbol{a}$, as well as latent quantities (values of random effects $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$ and component labels $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$ introduced below) , and smoothing hyperparameters $\boldsymbol{\lambda}$ defined below. Analogously to the survival models of Komárek and Lesaffre (2006, 2007), the prior distribution for $\boldsymbol{\theta}$ is specified hierarchically using the following decomposition:

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta}) \times p(\boldsymbol{\tau}) \times \prod_{i=1}^{N} \Big\{ \frac{\underbrace{p(\boldsymbol{b}_i \,|\, \boldsymbol{r}_i, \, \boldsymbol{\tau}) \times p(\boldsymbol{r}_i \,|\, \boldsymbol{a})}}{p(\boldsymbol{b}_i, \, \boldsymbol{r}_i \,|\, \boldsymbol{\tau}, \, \boldsymbol{a})} \Big\} \times p(\boldsymbol{a} \,|\, \boldsymbol{\lambda}) \times p(\boldsymbol{\lambda}). \tag{11}$$

For 'fixed effects' $\boldsymbol{\beta}$, we take a normal prior with possibly high variances. For inverse variance parameters $\boldsymbol{\tau}^{-2} = (\tau_1^{-2}, \ldots, \tau_q^{-2})'$, we take a product of independent gamma priors with possibly small values of the shape and rate parameters. That is,

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\beta}_0, \mathbb{S}_{\beta_0}) \tag{12}$$

$$\boldsymbol{\tau}^{-2} \sim \prod_{m=1}^{q} \text{Gamma}(\zeta_{m,1}, \zeta_{m,2}), \tag{13}$$

which are the choices made conventionally in the context of hierarchical regression models (see Gelman et al., 2004)

It is advantageous to introduce for each observation (sampled from the mixture) its latent component label and assume that the observation belongs to that mixture component. In our context, mixture observations are given by the values of random effects $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$. Let $\boldsymbol{r}_1 = (r_{1,1}, \ldots, r_{1,q})', \ldots, \boldsymbol{r}_N = (r_{N,1}, \ldots, r_{N,q})'$ be their component labels, i.e., discrete random vectors with values in $\{-K_1, \ldots, K_1\} \times \cdots \times \{-K_q, \ldots, K_q\}$. The PGM model, expressions (2) and (3), determines the factor $\prod_i \Big\{ p(\boldsymbol{b}_i \,|\, \boldsymbol{r}_i,\, \boldsymbol{\tau}) \times p(\boldsymbol{r}_i \,|\, \boldsymbol{a}) \Big\}$ in (11) as follows

$$\boldsymbol{b}_i \,\Big|\, \boldsymbol{r}_i,\, \boldsymbol{\tau} \sim \mathcal{N}_q \Big[ (\tau_1 \mu_{1,r_{i,1}}, \ldots, \tau_q \mu_{q,r_{i,q}})',\;\; \text{diag}\Big\{ (\tau_1\sigma_1)^2, \ldots, (\tau_q\sigma_q)^2 \Big\} \Big], \tag{14}$$

$$\text{P}\Big( \boldsymbol{r}_i = (j_1, \ldots, j_q)' \,\Big|\, \boldsymbol{a} \Big) = w_{j_1,\ldots,j_q}(\boldsymbol{a}) \tag{15}$$

$$(i = 1, \ldots, N, \quad j_1 = -K_1, \ldots, K_1, \ldots, j_q = -K_q, \ldots, K_q).$$

Note, that the use of component labels is a direct application of Bayesian data augmentation of Tanner and Wong (1987) since integration of the component label $\boldsymbol{r}_i$ out of the prior, i.e., $\int p(\boldsymbol{b}_i,\, \boldsymbol{r}_i \,|\, \boldsymbol{\tau},\, \boldsymbol{a})\, d\text{P}_{\boldsymbol{r}_i}$ leads to the mixture given by (2) and (3).

The term $p(\boldsymbol{a} \,|\, \boldsymbol{\lambda})$ in the prior (11) is specified as a combination of intrinsic Gaussian Markov random fields in each margin as

$$p(\boldsymbol{a} \,|\, \boldsymbol{\lambda}) \propto \exp\left[ -\left\{ \frac{\lambda_1}{2} \sum_{j_1} \cdots \sum_{j_q} \Big( \Delta_1^d a_{j_1,\ldots,j_q} \Big)^2 + \cdots + \frac{\lambda_q}{2} \sum_{j_1} \cdots \sum_{j_q} \Big( \Delta_q^d a_{j_1,\ldots,j_q} \Big)^2 \right\} \right], \tag{16}$$

where $\Delta_m^d$ is a difference operator of order $d$ in the $m$th margin (e.g., $\Delta_1^3 a_{j_1,j_2,\ldots,j_q} = a_{j_1,j_2,\ldots,j_q} - 3a_{j_1-1,j_2,\ldots,j_q} + 3a_{j_1-2,j_2,\ldots,j_q} - a_{j_1-3,j_2,\ldots,j_q}$), and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)'$ are smoothing hyperparameters. In fact, the prior density (16) is that of an (improper) multivariate normal distribution which can be seen from

$$p(\boldsymbol{a} \,|\, \boldsymbol{\lambda}) \propto \exp\Big\{ -\frac{1}{2} \boldsymbol{a}' \Big( \lambda_1\, \mathbb{P}'_{d,1} \mathbb{P}_{d,1} + \cdots + \lambda_q\, \mathbb{P}'_{d,q} \mathbb{P}_{d,q} \Big) \boldsymbol{a} \Big\}, \tag{17}$$

where $\mathbb{P}_{d,1}, \ldots, \mathbb{P}_{d,q}$ are corresponding difference operator matrices. The link between roughness penalties and the IGMRF prior is described in detail by Lang and Brezger (2004).

It is seen from (17) that the components of the vector $\boldsymbol{\lambda}$ determine directly the inverse variance of the IGMRF prior for $\boldsymbol{a}$. For this reason, we use a product of independent gamma priors, with possibly small values of the shape and rate parameters, for $\boldsymbol{\lambda}$, i.e.,

$$\boldsymbol{\lambda} \sim \prod_{m=1}^{q} \text{Gamma}(\xi_{m,1},\, \xi_{m,2}). \tag{18}$$

However, other commonly used priors for components of the inverse variance (see Gelman, 2006) are possible as well.

### *3.2 Posterior calculation*

The posterior distribution of the parameters of the PGM GLMM is given using the Bayes' formula as $p(\boldsymbol{\theta} \,|\, \boldsymbol{y}) \propto p(\boldsymbol{y} \,|\, \boldsymbol{\theta})\, p(\boldsymbol{\theta})$, where

$$p(\boldsymbol{y} \,|\, \boldsymbol{\theta}) = p(\boldsymbol{y} \,|\, \boldsymbol{\beta},\, \mathbb{B}) = \prod_{i=1}^{N} \prod_{l=1}^{n_i} p(y_{i,l} \,|\, \boldsymbol{\beta},\, \boldsymbol{b}_i) \tag{19}$$

is the likelihood determined by the distribution of the response. In the following, let

$$L_i(\boldsymbol{\beta},\, \boldsymbol{b}_i) = \prod_{l=1}^{n_i} p(y_{i,l} \,|\, \boldsymbol{\beta},\, \boldsymbol{b}_i), \qquad L(\boldsymbol{\beta},\, \mathbb{B}) = p(\boldsymbol{y} \,|\, \boldsymbol{\beta},\, \mathbb{B}) = \prod_{i=1}^{N} L_i(\boldsymbol{\beta},\, \boldsymbol{b}_i). \tag{20}$$

To sample from the posterior distribution, we use a hybrid version of the Gibbs sampler (Gelfand and Smith, 1990) with a block update of subsets of the parameter vector $\boldsymbol{\theta}$. Detail are given in the appendix. MCMC simulations and tools for posterior computation in the case of (a) a logit model with a binomial response, (b) a log-linear model with a Poisson response and (c) a cumulative logit model with a multinomial response together with uni- or bivariate random effects ($q \leq 2$) are implemented in the R contributed package glmmAK which is available from the *Comprehensive R Archive Network (CRAN).*

## 4 Simulation study

To validate our approach we conducted a small simulation study which mimics to a certain extent the Toenail infection data analyzed in Section 5.1. For $N =$

Table 1
Simulation study, 100 datasets: Bias, Std. Dev. and MSE are average bias, standard deviation and mean squared error, respectively of the estimates (posterior means). For each quantity, a better value when comparing the PGM and the Normal model is written in bold.

| | | **PGM GLMM** | | | **Normal GLMM** | | |
|---|---|---|---|---|---|---|---|
| | | Bias | Std. Dev. | (MSE) | Bias | Std. Dev. | (MSE) |
| Setting | | | | $\gamma(\mathsf{Intercept}) = -1.5$ | | | |
| Log-Norm. | $N$ =50 | **−1.6923** | 1.7830 | (**6.0428**) | −2.4604 | **1.7575** | (9.1423) |
| | 100 | **−1.4725** | 1.3628 | (**4.0255**) | −2.2325 | **1.1023** | (6.1991) |
| | 300 | **−0.6318** | 0.6268 | (**0.7921**) | −1.9571 | **0.6085** | (4.2006) |
| | 600 | **−0.3688** | **0.3899** | (**0.2881**) | −2.0557 | 0.4000 | (4.3857) |
| Mixture | $N$ =50 | **0.6629** | 0.9632 | (**1.3673**) | 1.0747 | **0.7897** | (1.7786) |
| | 100 | **0.5480** | 0.8062 | (**0.9504**) | 1.0710 | **0.5571** | (1.4575) |
| | 300 | **0.1896** | 0.7879 | (**0.6567**) | 1.0736 | **0.3096** | (1.2484) |
| | 600 | **0.5427** | 0.5062 | (**0.5508**) | 1.1668 | **0.2183** | (1.4091) |
| | | | | $\beta(\mathsf{Trt}) = -0.5$ | | | |
| Log-Norm. | $N$ =50 | **−0.0062** | 1.8533 | (3.4346) | −0.1227 | **1.8132** | (**3.3028**) |
| | 100 | **−0.1108** | **1.2633** | (**1.6083**) | −0.1686 | 1.3581 | (1.8729) |
| | 300 | **−0.0019** | **0.5431** | (**0.2949**) | −0.1462 | 0.6474 | (0.4406) |
| | 600 | **0.0294** | **0.3338** | (**0.1123**) | −0.0942 | 0.4456 | (0.2074) |
| Mixture | $N$ =50 | **0.1021** | **0.8528** | (**0.7377**) | 0.1201 | 0.9390 | (0.8961) |
| | 100 | **0.0734** | **0.6258** | (**0.3970**) | 0.1379 | 0.7416 | (0.5690) |
| | 300 | **0.0717** | **0.3277** | (**0.1125**) | 0.1009 | 0.4232 | (0.1892) |
| | 600 | **0.0424** | **0.2303** | (**0.0548**) | 0.0485 | 0.2800 | (0.0808) |
| | | | | $\beta(\mathsf{Time}) = -0.4$ | | | |
| Log-Norm. | $N$ =50 | −0.0096 | 0.1276 | (0.0164) | **−0.0035** | **0.1231** | (**0.0152**) |
| | 100 | **0.0154** | 0.0712 | (**0.0053**) | 0.0271 | **0.0696** | (0.0056) |
| | 300 | **0.0230** | 0.0590 | (0.0040) | 0.0353 | **0.0384** | (**0.0027**) |
| | 600 | 0.0523 | 0.0940 | (0.0116) | **0.0365** | **0.0249** | (**0.0020**) |
| Mixture | $N$ =50 | **−0.0395** | **0.0794** | (**0.0079**) | −0.0589 | 0.0820 | (0.0102) |
| | 100 | **−0.0246** | **0.0549** | (**0.0036**) | −0.0498 | 0.0580 | (0.0058) |
| | 300 | **0.0079** | 0.0454 | (**0.0021**) | −0.0350 | **0.0332** | (0.0023) |
| | 600 | **0.0216** | 0.0523 | (0.0032) | −0.0305 | **0.0221** | (**0.0014**) |
| | | | | $\mathsf{sd}(b) = 4.0$ | | | |
| Log-Norm. | $N$ =50 | **0.4142** | **1.8488** | (**3.5895**) | 1.2038 | 2.0887 | (5.8119) |
| | 100 | **0.1187** | 1.4294 | (2.0573) | 0.7975 | **1.1778** | (**2.0233**) |
| | 300 | −0.6806 | 0.8751 | (1.2290) | **0.6207** | **0.6514** | (**0.8095**) |
| | 600 | −1.4469 | 1.1063 | (3.3175) | **0.5344** | **0.3617** | (**0.4165**) |
| Mixture | $N$ =50 | **−0.6706** | 0.7699 | (**1.0424**) | −0.9635 | **0.6123** | (1.3032) |
| | 100 | **−0.5531** | 0.7886 | (**0.9278**) | −0.9432 | **0.4185** | (1.0649) |
| | 300 | **−0.4193** | 1.1568 | (1.5140) | −1.0016 | **0.2194** | (**1.0514**) |
| | 600 | **−0.9540** | 1.1114 | (2.1453) | −1.0964 | **0.1610** | (**1.2279**) |

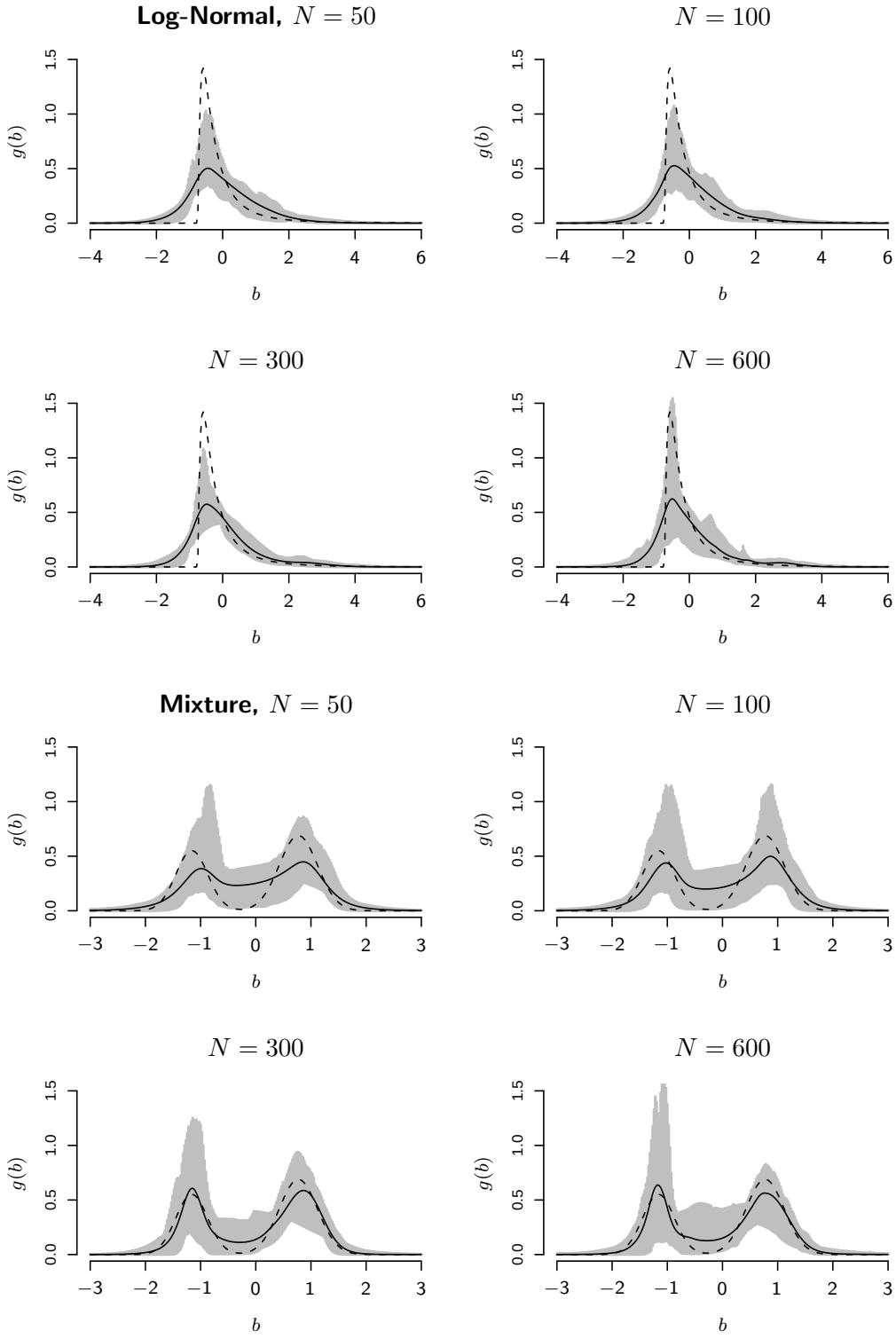


Fig. 1. Simulation study, 100 datasets. Results for the estimated standardized density of the random intercept in the PGM GLMM. Dashed line: true density, solid line: simulation based point-wise mean, gray region: simulation based 95% confidence interval.

50, 100, 300, 600, binary longitudinal responses $y_{i,j}$ ($i = 1, \ldots, N$, $j = 1, \ldots, 7$) were generated according to the following random intercept logit model

$$\text{logit}\Big\{\text{P}(Y_{i,l} = 1 \,|\, \boldsymbol{\beta},\, b_i)\Big\} = \beta_1 + \beta_2 \mathsf{Trt}_i + \beta_3 \mathsf{Time}_{i,j} + b_i, \tag{21}$$

where $\beta_1 = \gamma(\mathsf{Intercept}) = -1.5$, $\beta_2 = \beta(\mathsf{Trt}) = -0.5$, $\beta_3 = \beta(\mathsf{Time}) = -0.4$. Further, the binary covariate $\mathsf{Trt}_i$ was equal to 1 for $N/2$ subjects and equal to 0 for $N/2$ subjects. The continuous covariate $\mathsf{Time}_{i,j}$ was independently generated from a normal distribution $\mathcal{N}(\mu_{T,j},\, \sigma^2_{T,j})$ with $\boldsymbol{\mu}_T = (\mu_{T,1}, \ldots, \mu_{T,7})'$ $= (0,\ 1,\ 2,\ 3,\ 6,\ 9,\ 12)'$ and $\boldsymbol{\sigma}_T = (\sigma_{T,1}, \ldots, \sigma_{T,7})' = (0,\ 0.1,\ 0.2,\ 0.3,\ 0.5,$ $0.6,\ 0.8)'$. The random intercepts $b_i$ were obtained as $b_i = \tau\, b_i^*$ with $\tau =$ $\text{sd}(b_i) = 4.0$. Standardized random intercepts $b_i^*$ were generated from a zero-mean shifted and unit-variance scaled (a) log-normal distribution, (b) normal mixture $0.4\mathcal{N}(-2,\, 0.5^2) + 0.6\mathcal{N}(1.33,\, 0.6^2)$. Note that the log-normal density in (a) is a typical representative of a skewed distribution whose support moreover does not cover the whole real line. On the other hand, the mixture in (b) is asymmetric and bimodal (see also Figure 1). For each setting (combination of the sample size $N$ and a random intercept density), we generated 100 datasets. For comparison purposes, each simulated dataset was analyzed by the PGM GLMM and also by the GLMM under the assumption of normally distributed random intercepts (Normal GLMM). The same prior distributions were used as in subsection 5.1. Posterior means based on the MCMC sample of length 10 000 with 1:5 thinning and a burn-in of 10 000 iterations are considered as point estimates for each model.

Summaries of the results for important model parameters are given in Table 1. Results for the estimation of the random intercept density are shown in Figure 1. With respect to bias in the estimation of parameters reported in Table 1, the PGM approach performed in 28 scenarios out of 32 better than the Normal model. Moreover, especially in the estimation of the intercept, the decrease in bias provided by the PGM model is quite considerable as compared to the Normal model. As could be expected, the more flexible PGM approach leads to estimates which are often more variable than estimates provided by the Normal model. However, the increase in variability is usually low, still leading to a lower value of the MSE in 22 out of 32 scenarios when comparing the PGM and Normal models. Furthermore, as seen in Figure 1, the random effects density is reasonably reproduced as well.

## 5 Applications

Practical use of the proposed method will be illustrated on two real examples. For the sake of comparison, we fitted all models also under two different assumptions concerning the distribution of random effects. Firstly, we compare

our PGM GLMM to the most classical (parametric) model with normally distributed random effects (Normal GLMM). Secondly, we fitted also non-parametric Bayesian models in which random effects are assumed to follow a Dirichlet process (DP GLMM, see, Mukhopadhyay and Gelfand, 1997). With respect to the flexibility concerning the distribution of random effects, the DP GLMM could be considered as a natural competitor to our PGM GLMM. The computation for the PGM and Normal GLMMs has been conducted using the R package glmmAK and the scripts for the analyzes are available in the documentation to the package. Estimation of the DP GLMM's has been performed using the R package DPpackage (Jara, 2007).

In all models in this section relatively flat $\mathcal{N}(0,\, 100^2)$ priors were chosen for the components of the fixed effects vector $\boldsymbol{\beta}$. For PGM GLMMs, the following choices for the parameters defining the PGM were taken. Common to all margins $m$ ($m = 1, \ldots, q$): $K_m = 15$ ($L_m = 31$), $\mu_{m,-K_m} = -4.5$, $\mu_{m,K_m} = 4.5$, $\delta_m = 0.3$, $\sigma_m = 0.2$. Further, for all margins $m$, we used a vague Gamma(1, 0.005) prior for both $\tau_m^{-2}$ and $\lambda_m$. In the Normal GLMMs, the random effects followed a normal distribution $\mathcal{N}_q(\mathbf{0},\, \mathbb{D})$. In subsection 5.1, $q = 1$, $\mathbb{D} = d_{1,1}$ and a vague Gamma(1, 0.005) prior was taken for $d_{1,1}^{-1}$. In subsection 5.2, $q = 2$, and a Wishart prior for $\mathbb{D}^{-1}$ with two degrees of freedom and an inverse scale matrix equal to diag(0.005, 0.005) was considered. In the DP GLMMs, the random effects shifted by a corresponding fixed effect were assumed to follow a Dirichlet process $\mathcal{DP}(\nu_0\, G_0)$ with the Gamma(1, 0.005) prior on the precision parameter $\nu_0$. The DP base distribution $G_0$ was assumed to be normal $\mathcal{N}_q(\boldsymbol{\gamma}_0,\, \mathbb{D}_0)$. Flat $\mathcal{N}(0,\, 100^2)$ hyperpriors were assumed for the components of the base mean $\boldsymbol{\gamma}_0$. In subsection 5.1, $q = 1$, $\mathbb{D}_0 = d_{0,1,1}$ and an inverse-Gamma(1.5, 0.5) prior was taken for $d_{0,1,1}$. In subsection 5.2, $q = 2$, and an inverse-Wishart prior with four degrees of freedom and a scale matrix equal to diag(1, 1) was considered for $\mathbb{D}_0$.

As posterior summary statistics, we report posterior means, standard deviations, MC errors and 95% highest posterior density (HPD) intervals. For a specific regression parameter $\beta$, we report $\mathrm{P} = 2\min\Big\{\mathrm{P}(\beta < 0 \,|\, \boldsymbol{y}),\ \mathrm{P}(\beta > 0 \,|\, \boldsymbol{y})\Big\}$ which will be called here the P-value and and which can be viewed as the counterpart of a classical two-sided P-value (see Held, 2004). The convergence and behavior of the Markov chain was assessed by a critical examination of the trace and autocorrelation plots, and using the methods of Geweke (1992) and Raftery and Lewis (1992).

### *5.1 Binary data: Toenail infection*

A longitudinal clinical trial in dermatology was set up to compare the efficacy of two oral treatments for toenail infection (De Backer et al., 1998). In this

Table 2
Toenail infection. Posterior summary statistics for model parameters. For each parameter we report posterior mean (posterior standard deviation; MC error) on the first row, 95% HPD interval and the P-value (for effect of covariates) on the second row.

| | **PGM GLMM** | **Normal GLMM** | **DP GLMM** |
|---|---|---|---|
| $\gamma_1$ (Intercept) | $-1.694$ (0.698; 0.041) | $-1.636$ (0.442; 0.005) | $-2.702$ (1.210; 0.011) |
| | ($-3.133$, $-0.499$) | ($-2.535$, $-0.803$) | ($-5.013$, $-0.873$) |
| $\beta_2$ (Time) | $-0.388$ (0.046; 0.001) | $-0.395$ (0.045; 0.000) | $-0.388$ (0.046; 0.000) |
| | ($-0.477$, $-0.300$) P<0.001 | ($-0.484$, $-0.308$) P<0.001 | ($-0.481$, $-0.302$) P<0.001 |
| $\beta_3$ (Trt) | 0.398 (0.433; 0.009) | $-0.153$ (0.590; 0.008) | 0.334 (0.444; 0.003) |
| | ($-0.444$, 1.247) P=0.356 | ($-1.341$, 0.965) P=0.799 | ($-0.578$, 1.170) P=0.437 |
| $\beta_4$ (Trt:Time) | $-0.129$ (0.071; 0.001) | $-0.139$ (0.069; 0.000) | $-0.128$ (0.071; 0.000) |
| | ($-0.267$, 0.011) P=0.069 | ($-0.278$, $-0.008$) P=0.040 | ($-0.267$, 0.011) P=0.068 |
| $\sqrt{d_{1,1}} = \mathsf{sd}(b)$ | 3.586 (0.651; 0.039) | 4.054 (0.388; 0.003) | |
| | (2.590, 4.926) | (3.330, 4.835) | |

paper, we will analyze a dichotomized version of the degree of onycholysis which expresses the degree of separation of the nail plate from the nail-bed (0 = absent or mild; 1 = moderate or severe). The response was evaluated at seven visits (approximately on weeks 0, 4, 8, 12, 24, 36 and 48). In total 937 and 971 measurements ($n = 1\,908$) were obtained on 146 and 148 patients ($N = 294$) in the control group (itraconazole 200 mg/day) and in the treatment group (terbinafine 250 mg/day), respectively. The effect of the treatment

on the dichotomized onycholysis has already been analyzed by Lesaffre and Spiessens (2001) with a logistic random effects model assuming normally distributed random intercepts. The methodology of our paper allows us, among other things, to evaluate whether the assumption of normality of the random effects was reasonable.

Let $Y_{i,l}$ represent the dichotomized onycholysis of the $i$-th subject at the $l$-th visit. We will model it using the following random intercept logit model:

$$\text{logit}\Big\{\text{P}(Y_{i,l}=1\,|\,\boldsymbol{\beta},\,b_i)\Big\}=\beta_1+\beta_2\,\mathsf{Trt}_i+\beta_3\,\mathsf{Time}_{i,l}+\beta_4\,\mathsf{Time}_{i,l}\cdot\mathsf{Trt}_i+b_i,\;\;(22)$$

where Trt denotes the binary treatment indicator and Time the visit time in months.

For posterior calculations, we generated a sample of length 25 000 obtained using an MCMC simulation with 1:100 thinning after a burn-in of 25 000 iterations which took, on an AMD Opteron 244 processor (1.8 GHz) with 2 GB RAM running Unix OS, 308 minutes for the PGM model, 268 minutes for the Normal model and 2 865 minutes for the DP model. Note that the implementation of the DP model computes also two different measures of model fit which requires one more reading of the dataset at each of the MCMC scans. Hence the computational times of the PGM and DP models cannot be

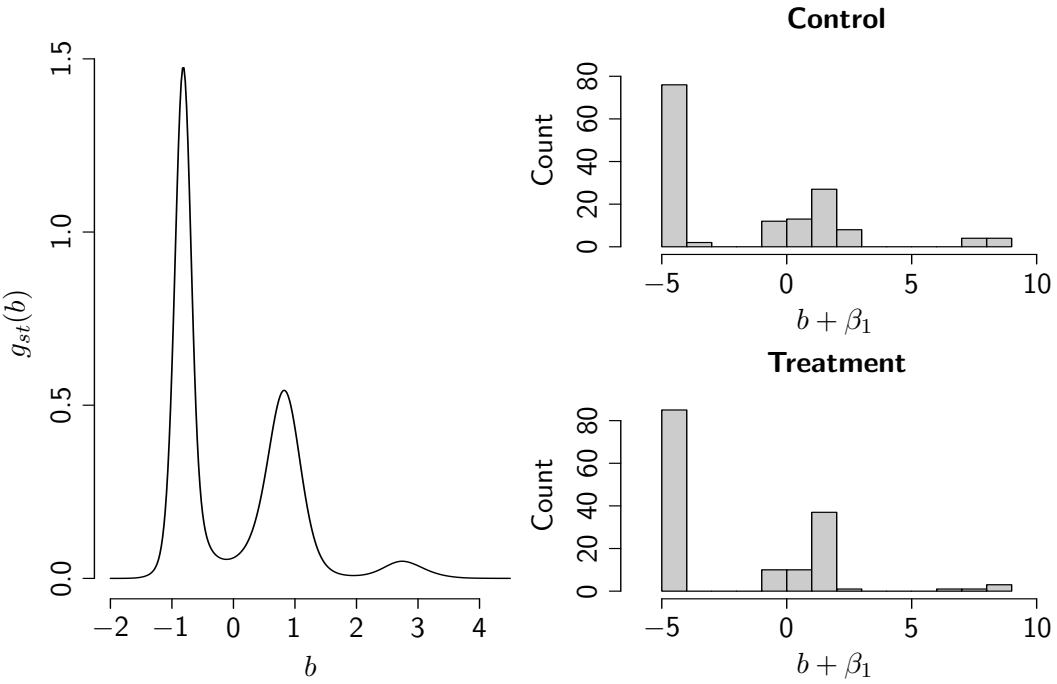


Fig. 2. Toenail infection, PGM model. Posterior point-wise mean of the random intercept density (standardized to have zero-mean and unit-variance) in the left panel and posterior medians for individual random intercepts shifted by $\beta_1$ in the right panel.

directly compared. The reason for thinning of the sample was a rather high autocorrelation for parameters Intercept and sd($b$) (about 0.6 for lag=10 and 0.3 for lag=50) in the PGM model.

The left panel of Figure 2 shows an estimate of a standardized version of the random intercept density in the PGM model. It is clear that the random intercept distribution is quite distinct from normality, suggesting that the patients could be divided into two or three groups according to their resistance against the infection and hence that an important covariate has been omitted from the model. Further, this covariate does not seem to interact with the treatment as the estimated distributions of the random intercept (histograms of posterior medians of the individual random effects) in both treatment groups are practically the same, see the right panel of Figure 2.

The effect of assuming incorrectly a normal distribution for the random intercept is shown in Table 2 which reports posterior summary statistics for all models. As was expected, due to proper randomization, there is no significant difference between the two treatment groups at baseline which is expressed by the value of $\beta_3$. The estimates of treatment over time (coefficient $\beta_4$) do not differ greatly between the three methods, however the PGM and DP models lead to a non-significant result whereas the Normal model suggests a slightly significant improvement of the treatment. Overally, the effect of covariates is estimated to be practically the same by both PGM and DP models. That the results of the Normal model are somewhat different from those of the PGM and DP models, especially with respect to the treatment effect $\beta_3$, are in agreement with the previous findings regarding the effect of misspecification of the random effect distribution in the GLMM (see Section 1). Indeed, the fitted random intercept density is multimodal here and the variability of the random intercept is quite high, certainly on the logit scale, see Table 2.

### *5.2 Count data: Epileptic seizures*

Thall and Vail (1990) report the data from a longitudinal study of seizures in epileptic patients. In total, $N = 59$ patients were randomized to receive either the anti-epileptic drug progabide (Trt=1) or placebo (Trt=0), as an adjuvant to standard chemotherapy. Patients underwent four successive post-randomization clinic visits ($n = 236$). For the $i$th patient, the response variable $Y_{i,l}$ denotes the number of seizures during the 2-weeks period before the $l$th visit. GLMMs to these data were fitted using an approximate method of penalized quasi-likelihood (PQL) under the assumption of normality of random effects by Breslow and Clayton (1993). A semi-parametric Bayesian approach was taken by Kleinman and Ibrahim (1998a) to analyze these data. Booth et al. (2003) analyzed the data assuming a model where the counts are con-

Table 3
Epileptic seizures. Posterior summary statistics for model parameters. For each parameter we report posterior mean (posterior standard deviation; MC error) on the first row, 95% HPD interval and the P-value (for effect of covariates) on the second row.

| | **PGM GLMM** | **Normal GLMM** |
|---|---|---|
| $\gamma_1$ (Intercept) | $-1.377$ (1.238; 0.025) | $-1.419$ (1.256; 0.007) |
| | ($-3.800$, 1.087) | ($-3.904$, 1.034) |
| $\gamma_2$ (Visit) | $-0.275$ (0.167; 0.008) | $-0.273$ (0.156; 0.001) |
| | ($-0.602$, 0.057) P=0.101 | ($-0.574$, 0.036) P=0.083 |
| $\beta_3$ (Base) | 0.866 (0.137; 0.003) | 0.885 (0.137; 0.001) |
| | (0.604, 1.143) P<0.001 | (0.616, 1.152) P<0.001 |
| $\beta_4$ (Trt) | $-0.995$ (0.423; 0.009) | $-0.944$ (0.418; 0.003) |
| | ($-1.849$, $-0.184$) P=0.021 | ($-1.765$, $-0.115$) P=0.024 |
| $\beta_5$ (Base:Trt) | 0.371 (0.213; 0.004) | 0.346 (0.213; 0.001) |
| | ($-0.045$, 0.795) P=0.082 | ($-0.069$, 0.764) P=0.107 |
| $\beta_6$ (Age) | 0.487 (0.361; 0.006) | 0.492 (0.370; 0.002) |
| | ($-0.202$, 1.223) P=0.174 | ($-0.245$, 1.207) P=0.182 |
| sd($b_1$) | 0.543 (0.075; 0.004) | 0.530 (0.065; 0.000) |
| | (0.408, 0.696) | (0.411, 0.661) |
| sd($b_2$) | 0.711 (0.186; 0.007) | 0.613 (0.205; 0.002) |
| | (0.367, 1.092) | (0.096, 0.972) |
| corr($b_1, b_2$) | 0.061 (0.220; 0.026) | 0.042 (0.314; 0.002) |
| | ($-0.437$, 0.486) | ($-0.571$, 0.714) |

| | **DP GLMM** |
|---|---|
| $\gamma_1$ (Intercept) | $-1.427$ (1.252; 0.056) |
| | ($-3.871$, 1.080) |
| $\gamma_2$ (Visit) | $-0.269$ (0.157; 0.001) |
| | ($-0.568$, 0.045) P=0.087 |
| $\beta_3$ (Base) | 0.885 (0.137; 0.002) |
| | (0.619, 1.158) P<0.001 |
| $\beta_4$ (Trt) | $-0.930$ (0.438; 0.003) |
| | ($-1.783$, $-0.070$) P=0.032 |
| $\beta_5$ (Base:Trt) | 0.340 (0.226; 0.002) |
| | ($-0.094$, 0.787) P=0.133 |
| $\beta_6$ (Age) | 0.493 (0.367; 0.016) |
| | ($-0.215$, 1.232) P=0.175 |

ditionally independent negative binomial variables. We will specify the linear predictor of the GLMM in the same way as in Breslow and Clayton's Model IV, i.e.,

$$\log\Big\{\mathrm{E}(Y_{i,l} \,|\, \boldsymbol{\beta},\, \boldsymbol{b}_i)\Big\} = \\ \beta_1 + \beta_2 \mathsf{Visit}_{i,l} + \beta_3 \mathsf{Base}_i + \beta_4 \mathsf{Trt}_i + \beta_5 \mathsf{Base}_i \cdot \mathsf{Trt}_i + \beta_6 \mathsf{Age}_i + b_{i,1} + b_{i,2} \mathsf{Visit}_{i,l}, \tag{23}$$

where Visit is the centered visit time in weeks divided by 10 ($-0.3$, $-0.1$, 0.1, 0.3), Base is the logarithm of $\frac{1}{4}$ the 8-week pre-randomization seizure count and Age is the logarithm of age in years.

Posterior calculations are based on a sample of length 25 000 generated using an MCMC simulation with 1:200 thinning after a burn-in of 25 000 iterations.

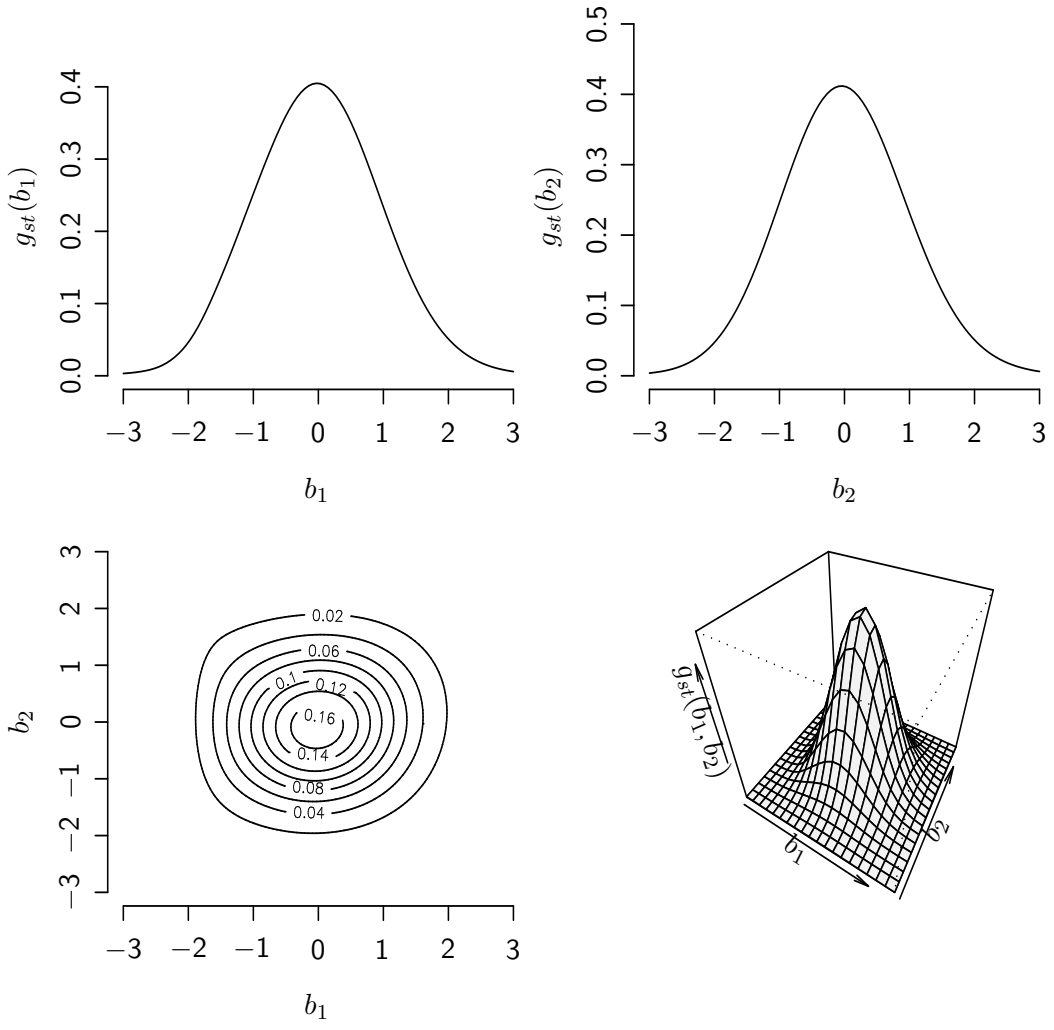


Fig. 3. Epileptic seizures, PGM model. Posterior point-wise mean of the random effect density (standardized to have zero-means and unit-variances). Marginal densities in the upper panel, joint density in the lower panel.

This took, on the same processor as in subsection 5.1, 594 minutes for the PGM model, 46 minutes for the Normal model and 740 minutes for the DP model. For the parameters reported in Table 3, a high autocorrelation was found only for the correlation coefficient between the random effects in the PGM model. It valued 0.97 for lag=10 and 0.92 for lag=50, hence a rather drastic thinning of the MCMC sample was needed. The lag=10 autocorrelation values lies below 0.08 for the parameters Intercept, Base, Trt, Base:Trt, Age, sd($b_2$) and between 0.12 and 0.17 for the parameters Visit, sd($b_1$).

The estimated density of the random effects (standardized to have zero-mean and unit-variances) is shown in Figure 3. It is seen that for this example, the normality assumption for the random effects holds. Consequently, see Table 3, the posterior summary for both fixed effects and variance components of the random effects are similar in all three models, with the Normal model showing in several cases somewhat narrower HPD intervals. Based on posterior medians of the individual random effects, one can identify patients with extreme values of seizure counts and/or changes in seizure counts over the time, even after the adjustment for covariates, see Figure 4. Not surprisingly, due to the normality of random effects, this exercise led to practically the same figure as that resulted from the PQL analysis of Breslow and Clayton (1993).

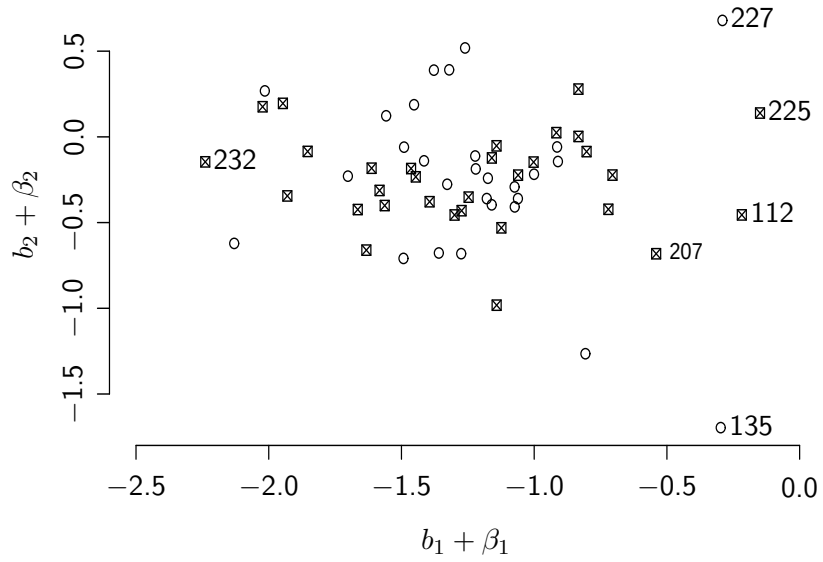


Fig. 4. Epileptic seizures, PGM model. Posterior medians for individual random effects shifted by corresponding fixed effect (circles for placebo group, crossed squares for treatment group). Extraordinary patients are identified by the ID numbers from Table 2 of Thall and Vail (1990).

## 6 Concluding remarks

It has been indicated in the literature that misspecification of the random effects distribution in GLMM can influence the estimation of quantities of the primary interest, like the fixed effects. To avoid such misspecification, we have suggested to model the distribution of the random effects in a flexible way using the penalized Gaussian mixture.

It seems that the price for flexibility is the computational burden as the models need to be fitted using a simulation based MCMC methodology. Moreover, as we have seen in our examples a rather drastic thinning of the MCMC sample is necessary to decrease the autocorrelation in the chains which on its turn increases the required computational time. However, the amount of thinning that we have used is not exceptional in the applications where IGMRF priors are involved (compare to, e.g., Knorr-Held and Rue, 2002). Furthermore, for both our applications, the required computational time could have been shortened in two ways. Firstly, as indicated by rather low MC errors in Tables 2 and 3, the chains of length 25 000 used there provide estimates having (unnecessarily) high precision implying that shorter chains could have been used for the inference. Secondly, a burn-in period of the MCMC might have been shortened as well, especially in models with univariate random effects. We have computed the quantities presented in Tables 2 and 3 also using the chains with a shorter burn-in of 5 000 iterations. In the toenail infection example, the relative distance between the estimated posterior means obtained using the chain with a burn-in of 5 000 and 25 000, respectively, remained between 0.3% and 3.8%. In the epileptic seizures example which involved bivariate random effects, shorter burn-in led to a relative change of the posterior means of 0.6% to 19.1% for all parameters of Table 3 except $\mathsf{corr}(b_1, b_2)$ where the posterior mean changed from 0.061 to $-0.172$ and 95% HPD interval changed from $(-0.437, 0.486)$ to $(-0.384, 0.131)$. Moreover, many methods that relax the normality assumption on the random effects require a Monte Carlo component in the estimation procedure which leads to somewhat longer computational time.

A possible drawback of the current implementation of our method is that model selection can only be based on quantities like "P-values" from Tables 2 and 3. Note that these can also easily be computed in a simultaneous manner for a subset of the parameter vector as indicated in Besag et al. (1995, p. 30). Model selection however, could have been based on measures of the model complexity and fit as suggested by Spiegelhalter et al. (2002) but this option must further be explored.

One could also pose a question whether all the effort is worthwhile and whether our model does not overemphasize model fit or makes too ambitious attempts

to prevent misspecification. With respect to the model fit, we controls only one component of the GLMM – the random effects distribution. All other components, especially the link function and the linear form of the predictor, are fixed in advance. Moreover, the random effects distribution is not allowed to adapt too closely the data because of the penalty term included in the estimation procedure. Finally, although the number of parameters of the random effects distribution is relatively high, it is always limited and does not increase with the sample size as is the case, e.g., when using fully nonparametric approaches. Therefore we believe that the PGM is a parsimonious way to express an unknown distribution.

To conclude, we argue that our approach is worthwhile since it can offer the user more certainty about the effect of the distributional assumptions of the random effects on the estimation of the fixed effects, as exemplified here in the analysis of the toenail infection. Additionally, the results of our more complex model can serve as a scientifically sound justification of the assumptions in simpler models. For example, based on the results from our PGM model, one can justify a GLMM with normally distributed random effects for the analysis of the epileptic data in subsection 5.2.

**Acknowledgments**

We want to thank Alejandro Jara for his help with Bayesian non-parametric models. We thank the Associate Editor and two anonymous referees for their valuable comments that led to an improvement of the paper. The authors thank Novartis, Belgium, for permission to use their dermatological data for statistical research.

## A Markov chain Monte Carlo sampling

In this appendix, we will discuss the shape of the full conditional distributions for blocks of parameters that are jointly updated in one iteration of the MCMC sampler. We also give hints on how to sample from the full conditional distributions if their shape is not standard.

### *A.1 Updating the fixed effects $\boldsymbol{\beta}$*

The full conditional distribution for the fixed effects $\boldsymbol{\beta}$ depends only on the response vector $\boldsymbol{y}$ and the values of random effects $\mathbb{B}$:

$$p(\boldsymbol{\beta} \,|\, \cdots) \propto p(\boldsymbol{\beta}) \times L(\boldsymbol{\beta}, \mathbb{B}) \propto \exp\Big\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbb{S}_{\beta_0}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\Big\} \times L(\boldsymbol{\beta}, \mathbb{B}). \tag{A.1}$$

Let $\boldsymbol{\beta}^{(t)}$ and $\mathbb{B}^{(t)}$ denote the current values of $\boldsymbol{\beta}$ and $\mathbb{B}$, respectively. To sample from (A.1) we use a Metropolis-Hastings algorithm with a multivariate normal proposal with the following mean and covariance matrix:

$$\begin{aligned} \boldsymbol{m}(\boldsymbol{\beta}^{(t)}) &= \mathbb{V}(\boldsymbol{\beta}^{(t)}) \left\{\mathbb{S}_{\beta_0}^{-1}\boldsymbol{\beta}_0 \;+\; \mathcal{I}\Big(\boldsymbol{\beta}^{(t)}\Big)\boldsymbol{\beta}^{(t)} \;+\; \mathcal{U}\Big(\boldsymbol{\beta}^{(t)}\Big)\right\} \\ \mathbb{V}(\boldsymbol{\beta}^{(t)}) &= \left\{\mathbb{S}_{\beta_0}^{-1} \;+\; \mathcal{I}\Big(\boldsymbol{\beta}^{(t)}\Big)\right\}^{-1}, \end{aligned} \tag{A.2}$$

where

$$\mathcal{U}\Big(\boldsymbol{\beta}^{(t)}\Big) = \frac{\partial L(\boldsymbol{\beta}, \mathbb{B})}{\partial \boldsymbol{\beta}}\Big(\boldsymbol{\beta}^{(t)}, \mathbb{B}^{(t)}\Big), \qquad \mathcal{I}\Big(\boldsymbol{\beta}^{(t)}\Big) = -\mathrm{E}\left\{\frac{\partial^2 L(\boldsymbol{\beta}, \mathbb{B})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} \,\middle|\, \boldsymbol{\beta}^{(t)}, \mathbb{B}^{(t)}\right\}. \tag{A.3}$$

This procedure corresponds to conjugacy of $p(\boldsymbol{\beta})$ with the second-order normal approximation of $L(\boldsymbol{\beta}, \mathbb{B})$ constructed around its mode located using one Fisher-scoring step. Written in this form, the procedure can directly be applied to an arbitrary regression model where the likelihood $L(\boldsymbol{\beta}, \mathbb{B})$ is not conjugate with a normal prior $p(\boldsymbol{\beta})$. We remark, that for univariate GLMM's, our proposal is the same as a weighted least squares proposal of Gamerman (1997).

### *A.2 Updating the random effects $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$*

For a fixed unit $i$, the full conditional distribution of its random effect $\boldsymbol{b}_i$ is given by

$$p(\boldsymbol{b}_i \,|\, \cdots) \propto p(\boldsymbol{b}_i \,|\, \boldsymbol{r}_i, \boldsymbol{\tau}) \times L_i(\boldsymbol{b}_i, \boldsymbol{\beta}) \propto \exp\Big\{-\sum_{m=1}^{q} \frac{(b_{i,m} - \tau_m \mu_{m,r_{i,m}})^2}{2(\tau_m \sigma_m)^2}\Big\} \times L_i(\boldsymbol{b}_i, \boldsymbol{\beta}). \tag{A.4}$$

In (A.4), the likelihood of the GLMM is combined with a normal "prior" in the same way as in (A.4). To update a vector of unit specific random effects $\boldsymbol{b}_i$, we apply the same procedure as in subsection A.1.

### *A.3 Updating the inverse variance parameters* $\boldsymbol{\tau}^{-2}$

The full conditional distribution of the inverse variance parameters $\boldsymbol{\tau}^{-2}$ is the following product

$$\begin{aligned} p(\boldsymbol{\tau^{-2}} \mid \cdots) &\propto p(\boldsymbol{\tau}^{-2}) \times \prod_{i=1}^{N} p(\boldsymbol{b}_i \mid \boldsymbol{r}_i, \boldsymbol{\tau}) \\ &\propto \prod_{m=1}^{q} \left\{ \left(\tau_m^{-2}\right)^{\zeta^*_{m,1}-1} \exp\left(\zeta^*_{m,3}\sqrt{\tau_m^{-2}} \; - \; \zeta^*_{m,2}\tau_m^{-2}\right) \right\}, \end{aligned} \tag{A.5}$$

where

$$\zeta^*_{m,1} = \zeta_{m,1} + \frac{N}{2}, \quad \zeta^*_{m,2} = \zeta_{m,2} + \frac{\sum_{i=1}^{N} b_{i,m}^2}{2\sigma_m^2}, \quad \zeta^*_{m,3} = \frac{\sum_{i=1}^{N} \mu_{m,r_{i,m}}\, b_{i,m}}{\sigma_m^2} \quad (m = 1, \ldots, q). \tag{A.6}$$

Sampling from each component of the product (A.5) is somewhat complicated by the fact that it is generally not log-concave. However, it can be shown that it is unimodal which allows us to use a simpler version of the slice sampler (Neal, 2003).

### *A.4 Updating the component labels* $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$

Updating of the component labels $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$ is straightforward since for fixed $i$, the full conditional distribution is discrete with

$$\begin{aligned} \mathrm{P}(\boldsymbol{r}_i = (j_1, \ldots, j_q)' \mid \cdots) &\propto \mathrm{P}(\boldsymbol{r}_i = (j_1, \ldots, j_q)' \mid \boldsymbol{a}) \times p(\boldsymbol{b}_i \mid \boldsymbol{r}_i, \boldsymbol{\tau}) \\ &\propto w_{j_1 \ldots, j_q}(\boldsymbol{a}) \exp\left\{ - \sum_{m=1}^{q} \frac{(b_{i,m} - \tau_m \mu_{m,j_m})^2}{2(\tau_m \sigma_m)^2} \right\} \\ &\quad (j_1 = -K_1, \ldots, K_1, \; \ldots, \; j_q = -K_q, \ldots, K_q). \end{aligned} \tag{A.7}$$

### *A.5 Updating the transformed PGM weights* $\boldsymbol{a}$

Let $N_{j_1,\ldots,j_q}$ be the number of random effects currently allocated in the $(j_1, \ldots, j_q)$th mixture component, i.e.,

$$N_{j_1,\ldots,j_q} = \sum_{i=1}^{N} I[\boldsymbol{r}_i = (j_1, \ldots, j_q)'] \qquad (j_1 = -K_1, \ldots, K_1, \; \ldots, \; j_q = -K_q, \ldots, K_q). \tag{A.8}$$

The full conditional distribution of the $(j_1, \ldots, j_q)$th element of $\boldsymbol{a}$ is given by

$$\begin{aligned} p(a_{j_1,\ldots,j_q} \mid \cdots) &\propto p(\boldsymbol{a} \mid \boldsymbol{\lambda}) \times \prod_{i=1}^{N} p(\boldsymbol{r}_i \mid \boldsymbol{a}) \\ &\propto \exp\left[ -\frac{\left\{ a_{j_1,\ldots,j_q} - \mathrm{E}(a_{j_1,\ldots,j_q} \mid \boldsymbol{a}_{-(j_1,\ldots,j_q)}, \boldsymbol{\lambda}) \right\}^2}{2 \operatorname{var}(a_{j_1,\ldots,j_q} \mid \boldsymbol{a}_{-(j_1,\ldots,j_q)}, \boldsymbol{\lambda})} \right] \\ &\quad \times \frac{\exp(N_{j_1,\ldots,j_q} a_{j_1,\ldots,j_q})}{\left\{ \sum_{k_1=-K_1}^{K_1} \cdots \sum_{k_q=-K_q}^{K_q} \exp(a_{k_1,\ldots,k_q}) \right\}^N}, \end{aligned} \tag{A.9}$$

where $\boldsymbol{a}_{-(j_1,\ldots,j_q)}$ denotes the vector $\boldsymbol{a}$ with omitted $(j_1, \ldots, j_q)$th element, $\mathrm{E}(a_{j_1,\ldots,j_q} \mid \boldsymbol{a}_{-(j_1,\ldots,j_q)}, \boldsymbol{\lambda})$, and $\operatorname{var}(a_{j_1,\ldots,j_q} \mid \boldsymbol{a}_{-(j_1,\ldots,j_q)}, \boldsymbol{\lambda})$ are the conditional moments resulting from the IGMRF prior (17), see Komárek and Lesaffre (2006, 2007) for more details. To sample from (A.9) we use the slice sampling method of Neal (2003) which use is simplified by the fact that the distribution (A.9) is log-concave.

### *A.6 Updating the smoothing hyperparameters* $\boldsymbol{\lambda}$

Updating the smoothing hyperparameters $\boldsymbol{\lambda}$ is easy as their full conditional distribution is a product of independent gamma distributions:

$$\begin{aligned} p(\boldsymbol{\lambda} \mid \cdots) &\propto p(\boldsymbol{\lambda}) \times p(\boldsymbol{a} \mid \boldsymbol{\lambda}) \\ \boldsymbol{\lambda} \mid \cdots &\sim \prod_{m=1}^{q} \mathrm{Gamma}\left( \xi_{m,1} + \frac{L_m - d + 1}{2}, \; \xi_{m,2} + \frac{\boldsymbol{a}' \mathbb{P}'_{d,1} \mathbb{P}_{d,1} \boldsymbol{a}}{2} \right) \end{aligned} \tag{A.10}$$

## References


Agresti, A., Caffo, B., Ohman-Strickland, P., 2004. Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. Computational Statistics and Data Analysis 47, 639–653.

Besag, J., Green, P., Higdon, D., Mengersen, K., 1995. Bayesian computation and stochastic systems (with Discussion). Statistical Science 10, 3–66.

Bogaerts, K., Lesaffre, E., 2007. Estimating local and global measures of association for bivariate interval censored data with a smooth estimate of the density. Submitted.

Booth, J., Casella, G., Friedl, H., Hobert, J., 2003. Negative binomial loglinear mixed models. Statistical Modelling 3, 179–191.

Breslow, N. E., Clayton, D. G., 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88, 9–25.

Bush, C. A., MacEachern, S. N., 1996. A semiparametric Bayesian model for randomised block designs. Biometrika 83, 275–285.

Butler, S. M., Louis, T., 1992. Random effects models with nonparametric priors. Statistics in Medicine 11, 1981–2000.

Caffo, B., Ming-Wen, A., Rohde, C., 2007. Flexible random intercept models for binary outcomes using mixtures of normals. Computational Statistics and Data Analysis 51, 5220–5235.

Chen, J., Zhang, D., Davidian, M., 2002. A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. Biostatistics 3, 347–360.

De Backer, M., De Vroey, C., Lesaffre, E., Scheys, I., De Keyser, P., 1998. Twelve weeks of continuous onychomycosis caused by dermatophytes: A double blind comparative trial of terbafine 250 mg/day versus itraconazole 200 mg/day. Journal of the American Academy of Dermatology 38, S57–S63.

Eilers, P. H. C., Marx, B. D., 1996. Flexible smoothing with B-splines and penalties (with Discussion). Statistical Science 11, 89–121.

Fahrmeir, L., Tutz, G., 2001. Multivariate Statistical Modelling Based on Generalized Linear Models, Second Edition. Springer-Verlag, New York.

Fieuws, S., Spiessens, B., Draney, K., 2004. Mixture models. In: De Boeck, P., Wilson, M. (Eds.), Explanatory item response models: A generalized linear and nonlinear approach. Springer-Verlag, New York, Ch. 11, pp. 317–340.

Follmann, D. A., Lambert, D., 1989. Generalizing logistic regression by nonparametric mixing. Journal of the American Statistical Association 84, 295–300.

Gallant, A. R., Nychka, D. W., 1987. Semi-nonparametric maximum likelihood estimation. Econometrica 55, 363–390.

Gamerman, D., 1997. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing 7, 57–68.

Gelfand, A. E., Smith, A. F. M., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85, 398–409.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 515–533.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian Data Analysis, Second Edition. Chapman & Hall/CRC, Boca Raton.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments (with Discussion). In: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M. (Eds.), Bayesian Statistics. Vol. 4. Oxford University Press, Oxford, pp. 169–193.

Ghidey, W., Lesaffre, E., Eilers, P., 2004. Smooth random effects distribution in a linear mixed model. Biometrics 60, 945–953.

Hanson, T., 2006. Inference for mixtures of finite Polya tree models. Journal of the American Statistical Association 101, 1548–1565.

Heagerty, P. J., Kurland, B. F., 2001. Misspecified maximum likelihood esti-

mates and generalised linear mixed models. Biometrika 88, 973–985.
Held, L., 2004. Simultaneous posterior probability statements from Monte Carlo output. Journal of Computational and Graphical Statistics 13, 20–35.
Jara, A., 2007. Applied Bayesian non- and semi-parametric inference using DPpackage. Submitted.
Jara, A., García-Zattera, M. J., Lesaffre, E., 2007a. A Dirichlet process mixture model for the analysis of correlated binary responses. Computational Statistics and Data Analysis 51, 5402–5415.
Jara, A., Hanson, T., Lesaffre, E., 2007b. Robustifying generalized linear mixed models using mixtures of multivariate Polya trees. Submitted.
Kleinman, K. P., Ibrahim, J. G., 1998a. A semi-parametric Bayesian approach to generalized linear mixed models. Statistics in Medicine 17, 2579–2596.
Kleinman, K. P., Ibrahim, J. G., 1998b. A semiparametric Bayesian approach to the random effects model. Biometrics 54, 921–938.
Knorr-Held, L., Rue, H., 2002. On block updating in Markov random fields models for disease mapping. Scandinavian Journal of Statistics 29, 597–614.
Komárek, A., Lesaffre, E., 2006. Bayesian semi-parametric accelerated failure time model for paired doubly-interval-censored data. Statistical Modelling 6, 3–22.
Komárek, A., Lesaffre, E., 2007. Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. To appear in Journal of the American Statistical Association.
Komárek, A., Lesaffre, E., Hilton, J. F., 2005. Accelerated failure time model for arbitrarily censored data with smoothed error distribution. Journal of Computational and Graphical Statistics 14, 726–745.
Laird, N., 1978. Nonparametric maximum likelihood estimation of a mixing distribution. Journal of the American Statistical Association 73, 805–811.
Lang, S., Brezger, A., 2004. Bayesian P-splines. Journal of Computational and Graphical Statistics 13, 183–212.
Lee, Y., Nelder, J. A., 2004. Conditional and marginal models: Another view (with Discussion). Statistical Science 19, 219–238.
Lesaffre, E., Spiessens, B., 2001. On the effect of the number of quadrature points in a logistic random-effects model: An example. Applied Statistics 50, 325–335.
Liang, K. Y., Zeger, S. L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.
Litière, S., Alonso, A., Molenberghs, G., 2008. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. To appear in Statistics in Medicine.
Molenberghs, G., Verbeke, G., 2005. Models for Discrete Longitudinal Data. Springer Science+Business Media, New York.
Mukhopadhyay, S., Gelfand, A. E., 1997. Dirichlet process mixed generalized linear models. Journal of the American Statistical Association 92, 633–639.

Neal, R. M., 2003. Slice sampling (with Discussion). The Annals of Statistics 31, 705–767.

Neuhaus, J. M., Hauck, W. W., Kalbfleisch, J. D., 1992. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. Biometrika 79, 755–762.

R Development Core Team, 2007. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL `http://www.R-project.org`

Raftery, A. E., Lewis, S. M., 1992. One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. Statistical Science 7, 493–497.

Robert, C. P., Casella, G., 2004. Monte Carlo Statistical Methods, Second Edition. Springer-Verlag, New York.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC, Boca Raton.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A., 2002. Bayesian measures of model complexity and fit (with Discussion). Journal of the Royal Statistical Society, Series B 64, 583–639.

Tanner, M. A., Wong, W. H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–550.

Thall, P. F., Vail, S. C., 1990. Some covariance models for longitudinal count data with overdispersion. Biometrics 46, 657–671.

Unser, M., Aldroubi, A., Eden, M., 1992. On the asymptotic convergence of B-spline wavelets to Gabor functions. IEEE Transactions on Information Theory 38, 864–872.

Verbeke, G., Lesaffre, E., 1996. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association 91, 217–221.

Verbeke, G., Lesaffre, E., 1997. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. Computational Statistics and Data Analysis 23, 541–556.

Walker, S. G., Mallick, B. K., 1997. Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. Journal of the Royal Statistical Society, Series B 59, 845–860.

Zackin, R., De Gruttola, V., Laird, N., 1996. Nonparametric mixed-effects models for repeated binary data arising in serial dilution assays: An application to estimating viral burden in AIDS. Journal of the American Statistical Association 91, 52–61.