Using gaussian mixtures with unknown number of components for mixed model estimation

Laurence Watier[1], Sylvia Richardson[1] and Peter J Green[2]

[1] Institut National de la Santé et de la Recherche Médicale U170, 16 avenue Paul Vaillant Couturier 94807 Villejuif Cedex, France (watier@vjf.inserm.fr)

[2] Department of Mathematics, University of Bristol, University Walk, Bristol, BS81TW, United Kingdom

Abstract: Hierarchical mixed models are used to account for dependence between correlated data, in particular dependence created by a group structure within the sample. In such models, the correlation between observations is modelled by including, in the regression model, group-indexed parameters regarded as random variables, so called random effects. Gaussian distributions are commonly used for the random effects. However, this choice places a strong constraint on the shape of the random parameter distribution.In this presentation, we focus on misspecification in mixed model with random intercept, a commonly used model in epidemiology. We propose to model the prior distribution of the random intercept by gaussian mixtures with an unknown number of components in a Bayesian framework. This methodology has recently been developed by Richardson & Green (1997) to analyse heterogeneous data. Another use of gaussian mixtures with unknown number of components is that of density estimation.

Keywords: Mixed models; Bayesian estimation; gaussian mixtures; Misspecification.

1 Introduction

The influence of misspecification in random effects' distributions was studied by Neuhaus et al (1992) for logistic mixed model. These authors have shown cases of non consistency for fixed and random parameter estimation. The use of finite mixture for modelling random effects' distribution in linear mixed model has recently been proposed. Verbeke & Lesaffre (1996) used empirical Bayes estimation and defined the number of components of the mixture by a test. Magder & Zeger (1996) used Maximum Likelihood estimation and defined constraints on the variances to enforce smoothness of the distribution. Modelling the prior distribution of the random intercept by gaussian mixtures with an unknown number of components in a Bayesian framework is an appealing alternative. It requires to introduce an additional hierarchical level to the mixed model which comprises the unknown number of components and the mixture component parameters for the random intercept distribution.

2 Formulation of the model :

We use the notations :
- i group index, $i = 1$ to n
- j observation index in a group, $j = 1$ to J_i
- J_i size of group i
- Y_{ij} known outcome of observation j in group i
- U_{ij} known covariates for observation j in group i
- α_i random intercept
- β regression parameters (fixed effects)

The covariate subscripts for U and β are suppressed in order not to overburden the notation.

The complete model is defined by two submodels which are linked through their common parameters $\{\alpha_i\}$

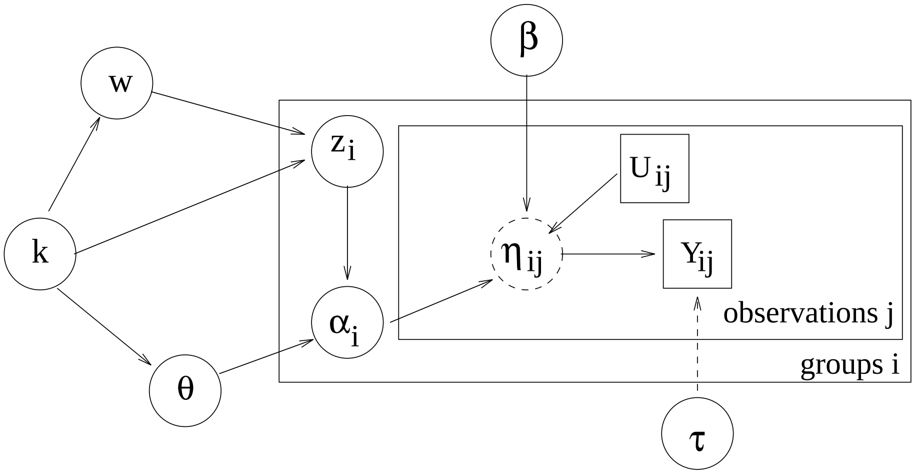
- Regression model $[Y_{ij}|U_{ij}, \alpha_i, \beta] = [Y_{ij}|\eta_{ij}]$ where $\eta_{ij} = \beta U_{ij}^T + \alpha_i$, with associated conditional independence assumptions of the $[Y_{ij}|\eta_{ij}]$ for each i and j,
- Mixture model for α_i :

$$\alpha_i \sim \sum_{p=1}^{k} w_p f(\cdot|\theta_p) \quad \text{independently for } i = 1, 2, \ldots, n$$

with $f(\cdot|\theta) \sim N(\mu_p, \sigma_p^2)$ and $\{\theta_p\}, \{w_p\}, k$ unknown parameters.

The hierarchical formulation of the mixture model introduces *latent allocation variables* z_i indicating to which mixture component the random intercept α_i belongs.

The graphical structure of the model can be represented by the following **Directed Acyclic Graph** :



where $\theta_p = \{\mu_p, \sigma_p^2\}$.

The joint distribution is given by :

$$[\beta][\tau][k][\theta|k][w|k][z|w,k]\prod_i[\alpha_i|\theta,z]\prod_{ij}[Y_{ij}|\eta_{ij},\tau]$$

$\eta_{ij} = \beta U_{ij}^T + \alpha_i$

We use *weakly informative hyperpriors*, normal priors for μ_p, and gamma priors for σ_p^{-2}, which are based on a notional range of values of $\{\alpha_i\}$ (see Richardson & Green, 1997).
As k (the unknown number of mixture components) is altered, the estimation of the posterior distribution uses reversible jump MCMC with dimension-changing moves based on splitting/merging adjacent components while preserving their overall “combined shape” (Green, 1995). Moves for updating the fixed effects or the component parameters are performed either by Gibbs sampling or by using random walk Metropolis moves.

3 Simulation study

Misspecification of the random intercept distribution, for linear and logistic mixed model, and its consequences for fixed and random effects estimation are studied by simulations. These simulations will also allow an assessment of the performance of our proposed method in identifying the shape of the random intercept distribution.

3.1 Linear model

$$[Y_{ij}|\cdot] \sim N(\alpha_i + \beta U_{ij}^T, \tau^2) \quad i = 1,..,n; j = 1,..,J_i$$

equivalently $\quad Y_{ij} \sim \sum_{p=1}^{k} w_p N(\mu_p + \beta U_{ij}^T, \sigma_p^2 + \tau^2).$

In this case, the joint model is a gaussian mixtures with a particular structure on the components mean and variances. Parameters values for the simulations where inspired by paper of Magder & Zeger (1996). We have chosen a case where the value for the ratio $Var(\alpha_i)/\tau^2$ is equal to 0.5 rather than the value of 1.0 considered in Magder & Zeger because a previous analysis (Watier, Richardson & Green, 1998) has indicated that the performance of our method is closely linked to this ratio (results not shown).

The data sets consist of 180 clusters, with sizes varying from 1 to 6 (a total of 540 observations are obtained). Two fixed effects $\beta = (\beta_1, \beta_2) = (2, 5)$ are introduced. The covariate U_{ij1} linked with β_1 differs within group in contrast to U_{ij2} which it is constant within group. U_{ij1} are simulated as independent standard gaussian random variables, whereas the values of U_{ij2} are equal to zero for 90 clusters and to one for the others. The error term in the regression is an independant gaussian random vector with mean zero and variance 2.
Two different distributions, f_l, for the random intercept α_i were considered
$f_1 \,:\, \alpha_i \sim N(0, 2^2)$
$f_2 \,:\, \alpha_i \sim \{0.25N(14, \sqrt{10}^2) + 0.75\chi_4^2\}c$
The multiplicative constant term $c = 4/\sqrt{109}$ is chosen to ensure that f_2 has a variance equal to that of f_1.

A total of 20 simulations were done for each of the two cases. For each simulation, runs of 70 000 iterations of MCMC algorithm were obtained. As can be seen on Figure 1, there is a reasonable convergence of the posterior probability of k after a burn-in period. From these runs, parameters estimates (posteriors means, posterior standard deviations) where computed from the last 50 000 iterations. The results presented below are the average of posterior means and posterior standard deviation (SD) over the 20 simulations and the mean square error (MSE). For the sake of comparison, besides using our model with the mixture prior for α_i, we also analysed the data using a standard gaussian prior for α_i.

- Results for fixed effects

	Analysis with gaussian prior								
f_l	$\hat{\beta}_1$	SD	MSE	$\hat{\beta}_2$	SD	MSE	$\hat{\tau}$	SD	MSE
f_1	2.00	.073	.003	5.09	.328	.113	1.43	.053	.002
f_2	2.00	.072	.004	5.09	.325	.104	1.42	.053	.003

	Analysis with gaussian mixture prior								
f_l	$\hat{\beta}_1$	SD	MSE	$\hat{\beta}_2$	SD	MSE	$\hat{\tau}$	SD	MSE
f_1	1.99	.072	.003	5.10	.331	.114	1.43	.053	.002
f_2	2.01	.072	.003	4.95	**.248**	**.070**	1.42	.053	.003

For the two prior models, posterior means for fixed parameters are close to their true values. Posterior standard deviations and MSE are also similar between the two priors models, except for the parameter β_2 in the case of random intercept distribution simulated with f_2. In this case, the use of a gaussian mixture prior resulted in a 24% decrease of the posterior standard deviation for β_2 and a 33% decrease of the corresponding MSE. This remark is in accordance with Magder and Zeger. It is important to note that using a mixture when the random intercept is gaussian (case f_1), which is a substantial overparametrization, does not lead to a poorer performance.

- Results for the random intercept distribution

With the gaussian prior, posterior mean values for the parameters (μ, σ) are respectively equal to -0.17 and 1.97 for the distribution f_1, estimates are close to the original values and identical to the ones obtained with mixture prior when $k = 1$ (see Table below).
In the table below, average results obtained with the gaussian mixture prior are shown. Only components with probability greater than 10% are indicated.

	gaussian mixture prior				
f_l	k	$p(k\|y)$	$\hat{w}$	$\hat{\mu}$	$\hat{\sigma}$
f_1	1	.673	1	-0.17	1.97
	2	.193	.52	-1.23	1.78
			.48	1.01	1.81
f_2	2	.541	.68	1.40	0.81
			.32	5.11	1.38
	3	.269	.52	0.92	0.77
			.28	3.24	1.02
			.20	6.11	1.13
	4	.115	.40	0.39	0.71
			.28	2.32	0.84
			.19	4.31	0.94
			.13	6.97	0.97

In the case of f_1 one sees a high probability on $k = 1$. Posterior density estimate of the mixture is represented in Figure 1. As expected, gaussian mixtures with unknown number of components gave a good fit to the simulated mixing distribution f_2, using between 2 and 4 normal components.

3.2 Logistic model

$$logit\ p_{ij} = \alpha_i + \beta U_{ij}^T, i = 1,..,n, j = 1,..,J$$

The data sets consist of 100 clusters with size 10. Two fixed effects $\beta = (\beta_1, \beta_2) = (.5, 1)$ are introduced. The covariate U_{ij1} linked with β_1 differs within group in contrast to U_{ij2} which is constant within group. $U_{ijk}, \{k = 1, 2\}$ are independent standard gaussian random variables. The random intercept distribution is simulated from an asymmetric mixture :

$$\alpha_i \sim 0.50N(-2.0, (.5)^2) + 0.5N(2, 2^2)$$

As for linear model, a total of 20 simulations were done. For each simulation, runs of 300 000 iterations of MCMC algorithm were obtained because we found that in this case the convergence was slower. In fact in the 20 simulations we found that in about half the cases, the algorithm did not converge well. On Figure 1, we see a case where there is stability convergence of the posterior probability of k after a burn-in period. From these runs, parameters estimates (posteriors means, posterior standard deviations) where computed from the last 150 000 iterations. The results presented below are similar to those described for the linear model. For the sake of comparison, besides using the mixture prior for α_i, we also used a standard gaussian prior for α_i in the analysis.

- Results for fixed effects

Analysis	$\hat{\beta}_1$	SD	MSE	$\hat{\beta}_2$	SD	MSE
gaussian prior	0.53	.105	.016	1.08	.322	.105
gaussian mixture prior	0.53	.106	.016	1.09	**.288**	**.068**

For the two prior models, posterior means for fixed parameters are close to their true values. Posterior standard deviation and MSE are similar between the two priors models, except for the parameter β_2. Again we see that the use of a gaussian mixture prior resulted in a 11% decrease of the posterior standard deviation for β_2 and a 35% decrease of the corresponding MSE.

- Results for the random intercept distribution

In the table below, average results obtained with gaussian mixture prior are shown for $k = 2$.

gaussian mixture prior				
k	$p(k\|y)$	$\hat{w}$	$\hat{\mu}$	$\hat{\sigma}$
2	.291	.61	-1.77	1.38
		.39	4.33	1.87

Over the 20 simulations, the gaussian mixture prior did not recover well the underlying true random effect distribution. However, in the cases were the algorithm converge, the 2 components were reasonably well estimated (see the Figure 1). A previous analysis (Watier, Richardson & Green, 1998) has indicated that the performance for the logistic model is conditioned, notably, by the cluster size. Indeed we found that the underlying random effect distribution is well recovered for cluster size equal to 50 (results not shown). Posterior density estimate of the random intercept in a case of good convergence can be appreciated in Figure 1.

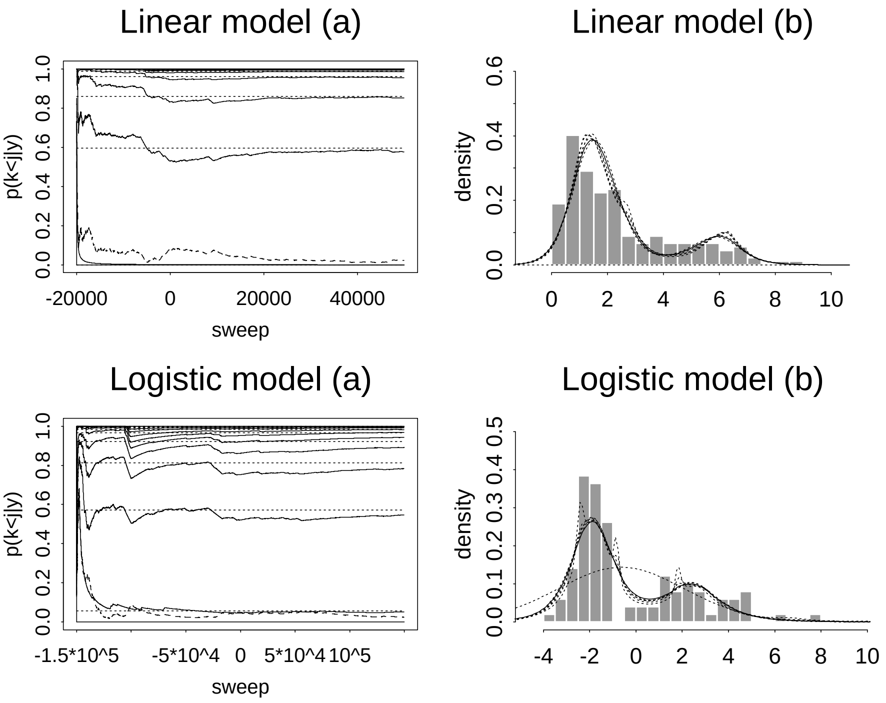


Figure 1 : For one simulation (a) Cumulative occupancy fractions (b) Comparison of simulated random intercept distribution (Histogramm) and posterior density estimate of the mixture.

4 Conclusion

For linear and logistic mixed models, our simulations did not shown an important effect of misspecification on fixed effect associated with covariates differing within cluster. This is not true when the covariates are constant within cluster, for which the gaussian mixture prior improves the results with a decrease in posterior standard deviation as well as the MSE. If the interest is in the shape of the between groups variability, analyses with standard gaussian priors are not, for the linear model, appropriate and mixture priors are a viable alternative. For the logistic model, convergence for gaussian mixture prior necessitates long runs. To recover the true random intercept distribution large number of cluster size is needed.

References

Green P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711-732.

Magder L. S. and Zeger S. L. (1996). A Smooth Nonparametric Estimate of a Mixing Distribution Using Mixtures of Gaussians *Journal of the American Statistical Association* **91**, 1141-1151.

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

Richardson S. and Green P. J. (1997). On Bayesian Analysis of Mixtures with Unknown Number of Components. *Journal of the Royal Statistical Society B* **59**, 731-792.

Verbeke G. and Lesaffre E. (1996). A Linear Mixed-Effects Model With Heterogeneity in the Random-Effects Population. *Journal of the American Statistical Association* **91**, 217-221.

Watier L., Richardson S. and Green P. J. (1998). Modelling random effect distribution in mixed models using gaussian mixtures. XIX^{th} International Biometrics Conference IBC98, Cape Town, December 1998.