

Discussion of

Center-Adjusted Inference for a Nonparametric
Bayesian Random Effect Distribution

by YISHENG LI, PETER MÜLLER and XIHONG LIN

O'Bayes - 09

sampling model

$$y_{ij} | x_{ij}, z_{ij} \sim EF \quad \text{with} \quad \left\{ \begin{array}{l} \text{mean } \mu_{ij}^{\mathbf{b}} \\ \text{variance } v_{ij}^{\mathbf{b}} = \phi v(\mu_{ij}^{\mathbf{b}}), \phi \text{ known} \\ \text{fixed effects } \boldsymbol{\beta} \\ \text{random effects } \mathbf{b}_i \\ \text{link function } g(\mu_{ij}^{\mathbf{b}}) = \eta_{ij}^{\mathbf{b}} \end{array} \right.$$

$$\eta_{ij}^{\mathbf{b}} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i$$

hierarchical priors for random effects

- conventional

fixed effects $\boldsymbol{\beta}$

$$\mathbf{b}_i \Big| G \stackrel{i.i.d.}{\sim} G$$

$$G \sim DP(M, G_0)$$

$$G_0 = N(\mathbf{0}, \mathbf{D})$$

$\mathbf{D} \sim$ proper (IW or USP)

- adjusted (Li-Müller-Lin)

fixed effects $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}^F, \boldsymbol{\beta}^R)$

$\boldsymbol{\beta}^R$ paired with $\mathbf{b}_i \Rightarrow \boldsymbol{\beta}_{pair} \equiv \boldsymbol{\beta}^R + \boldsymbol{\mu}_G$

$$(\boldsymbol{\beta}^R) + \mathbf{b}_i \Big| G \stackrel{i.i.d.}{\sim} G$$

$$G \sim DP(M, G_0)$$

$G_0 = N(\boldsymbol{\beta}_b, \mathbf{D})$, $\boldsymbol{\beta}_b$ *unknown vector*

of mean parameters

$\mathbf{D} \sim$ proper (IW or USP)

$(\boldsymbol{\beta}, \boldsymbol{\beta}_b) \sim$ flat (or diffuse Normal)

From the adjusted model for inference on the fixed effects paired with random effects and the variance components of the random effects, the main general results are:

- unbiased estimates (w.r.t. the conventional model);
- better credible interval length (*CLI*);
- good coverage probabilities for the fixed effects;

moreover

- USP on ***D*** works better than IW (or IG) proper prior leading to small variance components for random effects.

In general, adding one more hierarchical level in the prior modellization, you make the GLMM more flexible: I expect better results in estimation with the same sample sizes.

Three questions

- In the prediction of the outcome y are there any differences between the conventional and the adjusted model?
(I think that there could be more overfitting problems using the adjusted model with small samples).
- assuming USP prior for \mathbf{D} instead of IW (or IG), you obtain a smaller bias and better coverage probabilities: is this a natural consequence of the elicitation on IW's hyperparameters which is avoided using USP?
- how to assign the hyperparameters of IW (or IG)? Does it need a sensitivity analysis?

Other questions

About simulations

1. Normal outcome

$$10 \text{ obs} \times 50 \text{ subjects} = 500 \text{ observations}$$

2. Binary outcome

$$10 \text{ obs} \times 100 \text{ subjects} = 1000 \text{ observations}$$

How could be the comparison between the two models (conventional and adjusted) when one or more β are =0, and the sample size is smaller?

About the application

$$1 \div 65 \text{ obs} \times 286 \text{ subjects} = \text{large sample}$$

You use the IW for \mathbf{D} , instead of the USP after claiming with the simulation studies that the USP should be preferable to the IW. Why?

Two suggestions for future developments

- It could be interesting to consider as prior for G the Pólya-trees processes which have the advantage w.r.t. the DP in that they have continuous probability distributions as support instead of the discrete distributions of the DP.
- A further generalisation could be to consider the results in the paper of Regazzini E., Lijoi A. and Prunster I. (2003) *Distributional results for means of normalized random measures with independent increments*. Ann. Stat., 31, 2, 560-585.

One suggestion from the authors to me

- In the frequentist approach for GLMM, inference based on estimating equations is preferable to a fully parametric specification. For inference about a parameter of interest pseudo-likelihoods derived from estimating equations are available and they are motivated by asymptotic considerations. I am working with *Laura Ventura*, *Stefano Cabras* and other colleagues in “hybrid” methods to obtain simplifications in the Bayesian approach.

It could be interesting to explore comparisons with your results.

Thanks!

Discussion of

Bayesian Variable Selection and Grouping

by FEI LIU and SOUNAK CHAKRABORTY

O'Bayes - 09

The authors consider the normal linear regression model

$$Y = \mu \mathbf{1} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} , \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

with the aim of incorporating several regularisation methods into one Bayesian hierarchical model.

Prior distribution for $\boldsymbol{\beta}$

$$\boldsymbol{\beta} | \boldsymbol{\Lambda}, \sigma^2 \sim N_p(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda}^{-1}) \quad \text{where} \quad \left\{ \begin{array}{l} \boldsymbol{\Lambda} = [\lambda_{ij}] \text{ with } \lambda_{ii} = r + \sum_{j=1}^p \lambda_{ij} \\ r \geq 0, \lambda_{ij} > 0 \text{ hyperparameters} \\ \pi(\lambda_{ii} | a) = \frac{a^2}{2} \exp\left(-\frac{a^2}{2\lambda_{ii}}\right) 1(\lambda_{ii} > 0) \\ \pi(\lambda_{ij} | b) = \frac{b^2}{2} \exp\left(-\frac{b^2}{2\lambda_{ij}}\right) 1(\lambda_{ij} > 0), i \neq j \\ a, b \text{ hyperparameters} \end{array} \right.$$

- The authors show how to obtain the different regularisation methods discussed in the talk with particular choices of the hyperparameter a , b and r ;

moreover

- they provide bounds for the prior on $\beta|\sigma^2$ (integrating out λ_{ij}) which are essential in implementing *MCMC* computation.

Two questions

1. The sensitivity analysis w.r.t. the choice of the hyperparameters a , b and r seems to be quite robust when $a/b \leq 0.5$, but the authors say that the “*specification of the prior parameters a , b and r is crucial*”: what is the statistical meaning of $a/b \leq 0.5$?

2. In the proposed *MCMC* simulation to obtain the sampling draws $\{\mu^{(h)}, \sigma^{2(h)}, \boldsymbol{\beta}^{(h)} \boldsymbol{\Lambda}^{(h)}\}$, $h = 1, \dots, N$, from the posterior distribution $\pi(\mu, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\Lambda} | data)$, at each step the authors perform a K-means clustering algorithm to threshold small values in $\boldsymbol{\beta}^{(h)}$ to zero: does this modification of states in the chain alter transition probabilities?

Are you sure these simulations are draws from the actual posterior?