Discussion of the paper: "Sampling Schemes for Generalized Linear Dirichlet Process Random Effects Models" by M. Kyung, J. Gill, and G. Casella

M. Filippone[1], A. Mira[2], and M. Girolami[1]
1 - Department of Statistical Science, University College London, UK
2 - Department of Economics, University of Insubria, Italy

We congratulate the Authors for an interesting and thought provoking paper that compares different MCMC strategies for the class of generalized linear Dirichlet process random effects models. In our discussion we focus on the logistic regression model presented in Section 3.2 and would like to propose two alternative sampling schemes related to manifold methods that are quite general and could be employed every time a Metropolis-Hastings step is used within an MCMC simulation and the target distribution has closed form Fisher Information (FI). In particular we have implemented the same sampling scheme as presented by the Authors, but have sampled the parameters $\boldsymbol{\beta}$ using Simplified Manifold Metropolis Adjusted Langevin Algorithm (S-MMALA) and Riemann Manifold Hamiltonian Monte Carlo (RM-HMC), both presented in [1].

The starting point to apply these methods is the joint density of the observations and the parameters of interest, given by the likelihood of the model along with the prior distribution for $\boldsymbol{\beta}$. The gradient of the logarithm of this joint density and the FI (the expectation, taken with respect to the observed variables $\mathbf{y}$, of the negative Hessian of the log-likelihood) are then computed.

In order to keep the notation uncluttered, we define the logistic elements

$$l_i^+ = \text{logistic}(\mathbf{X}_i\boldsymbol{\beta} + (A\boldsymbol{\eta})_i) = \frac{1}{1 + \exp[-\mathbf{X}_i\boldsymbol{\beta} - (A\boldsymbol{\eta})_i]}$$

and similarly

$$l_i^- = 1 - l_i^+ = \frac{1}{1 + \exp[\mathbf{X}_i\boldsymbol{\beta} + (A\boldsymbol{\eta})_i]}.$$

The logarithm of the joint density of observations $\mathbf{y}$ and parameters $\boldsymbol{\beta}$ is:

$$\mathcal{L}_{\boldsymbol{\beta}} = \sum_{i=1}^{n} \left[y_i \log(l_i^+) + (1 - y_i) \log(l_i^-) \right] - \frac{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}}{2d^*\sigma^2}.$$

Using the property of the logistic function, for which $\nabla_{\boldsymbol{\beta}} l_i^+ = l_i^+ l_i^- \mathbf{X}_i$, the gradient with respect to $\boldsymbol{\beta}$ is easily evaluated as

$$\nabla_{\boldsymbol{\beta}} \mathcal{L}_{\boldsymbol{\beta}} = \sum_{i=1}^{n} \mathbf{X}_i \left(y_i - l_i^+ \right) - \frac{\boldsymbol{\beta}}{d^* \sigma^2}.$$

The Hessian of $\mathcal{L}_{\boldsymbol{\beta}}$ reads

$$\nabla_{\boldsymbol{\beta}} \nabla_{\boldsymbol{\beta}} \mathcal{L}_{\boldsymbol{\beta}} = -\sum_{i=1}^{n} \left(l_i^+ l_i^- \right) \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} - \frac{1}{d^* \sigma^2},$$

where there is no dependence from $\mathbf{y}$ anymore. As a consequence, the expectation with respect to $\mathbf{y}$ of the negative Hessian is equal to the negative Hessian itself, and therefore the FI, along with the negative Hessian of the prior, is

$$G = \sum_{i=1}^{n} \left(l_i^+ l_i^- \right) \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} + \frac{1}{d^* \sigma^2}$$

with derivatives with respect to the β_r, given by:

$$\frac{\partial G}{\partial \beta_r} = \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} \left(l_i^+ l_i^- \right) \left(l_i^- - l_i^+ \right) X_{ir}.$$

In S-MMALA we set an integration step size of $\varepsilon = 1$, whereas for RM-HMC we set a maximum number of leapfrog steps to 10 with integration step $\varepsilon = 0.75$. The autocorrelation for the components of $\boldsymbol{\beta}$ are reported in Fig. 1, where it can be seen that the autocorrelation is lower than the one obtained by the Authors and reported in the right panel of Fig. 2 in the paper. The use of a Metropolis-Hastings step within the Gibbs sampler avoids the use of any auxiliary variable in sampling $\boldsymbol{\beta}$, that in the expression of the likelihood are effectively integrated out. As pointed out by the Authors in section 3.3.2, one of the challenges in setting up efficient MCMC methods using Metropolis-Hastings steps, however, is how to tune the parameters of the proposal. Here, by using the idea presented in [1], the proposal is automatically tuned based on the geometry of the underlying statistical model, whereby only the integration step ε and/or the number of leapfrog steps need to be tuned. Such a proposal mechanism allows to efficiently deal with multivariate correlated posterior distributions as confirmed by the results.

References

[1] M. Girolami and B. Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

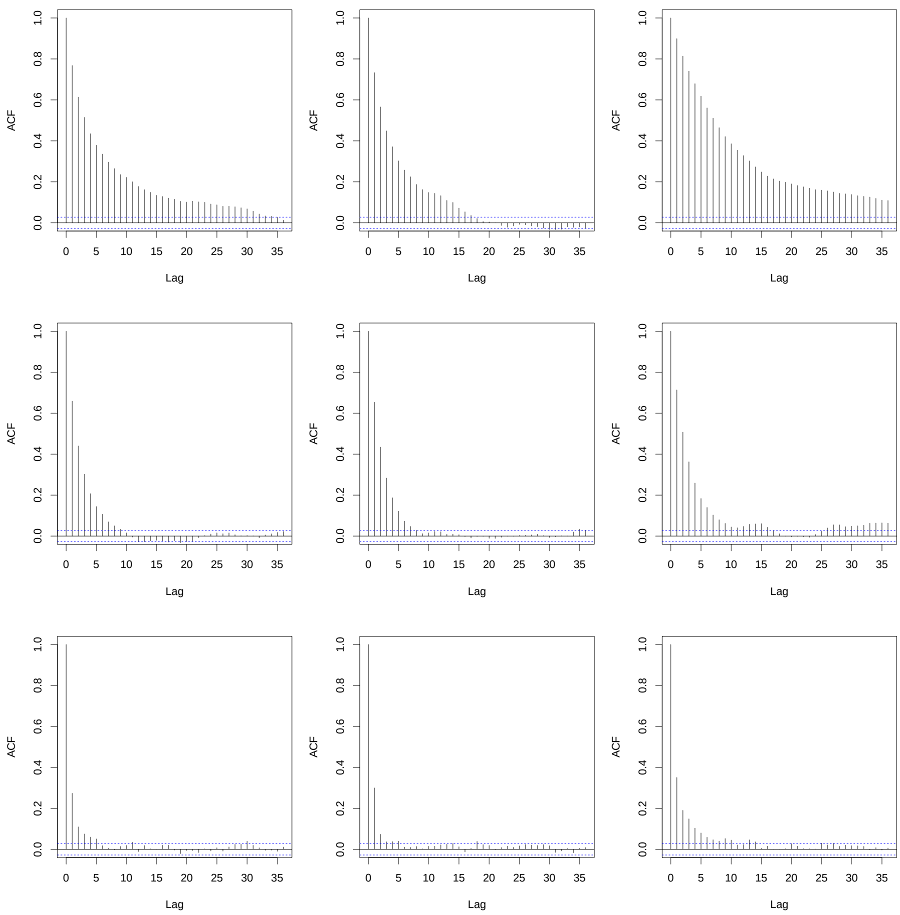

Figure 1: Autocorrelation of the samples for $\boldsymbol{\beta}$ obtained by the KS mixture representation by the Authors (first row or right panel of Fig. 2 in the paper), S-MMALA (second row), and RM-HMC (third row)