Discussion of the paper 'Riemann manifold Langevin and Hamiltonian Monte Carlo methods' by M. Girolami and B. Calderhead

Vassilios Stathopoulos†
University of Glasgow, Glasgow, UK.
Maurizio Filippone
University of Glasgow, Glasgow, UK.

We consider a univariate binomial probit model where we use X, W, and $Y \in \{0, 1\}$ to denote the observed covariates, the latent variables, and binomial responces respectively. The latent variables W are modelled as:

$$W = X\beta + \varepsilon$$

with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. The binomial variable is $Y(W) = 0$ if $W < 0$ and $Y(W) = 1$ if $W > 0$. The model in this form is known to be non-identifiable as the likelihood is constant along straight lines out of the origin of the β, σ plane and therefore is only informative about their ratio (Nobile, 1998, 2000; McCulloch et al., 2000; Imai and van Dyk, 2005). This poses significant challenges to the MCMC methods discussed in this paper since the Fisher information (FI) matrix is not positive definite. The problem can be resolved by considering an informative prior Nobile (1998) and by adding the negative of its Hessian to the FI as suggested by the Authors in Section 4.2. The resulting posterior, however, is strongly skewed and, as we discuss here, this can lead to very poor mixing of the chains.

For the experiments presented here, we generated a synthetic dataset for the binomial model as described in Nobile (1998) and used the priors $p(\beta) = \mathcal{N}(0, 100)$ and $p(1/\sigma^2) = \mathcal{G}(3/2, 1/6)$ which ensure weak identifiability. Furthermore, we re-parameterise σ, such that $\psi = \log(\sigma^2)$, and sample ψ. The log-likelihood is given by:

$$\mathcal{L} = \sum_i y_i \log\left[\Phi\left(\frac{\beta x_i}{\sigma}\right)\right] + \sum_i (1 - y_i) \log\left[\Phi\left(-\frac{\beta x_i}{\sigma}\right)\right]$$

where Φ is the cumulative function of $\mathcal{N}(0, 1)$. The gradient of the log likelihood and the FI follow as

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_i a_i y_i + b_i \qquad \frac{\partial \mathcal{L}}{\partial \psi} = \sum_i c_i y_i + d_i$$

$$\mathcal{I}_i = \begin{bmatrix} b_i^2 & b_i d_i \\ b_i d_i & d_i^2 \end{bmatrix} + \Phi\left(\frac{\beta x_i}{\sigma}\right) \begin{bmatrix} a_i^2 + 2a_i b_i & (a_i c_i + a_i d_i + b_i c_i) \\ (a_i c_i + a_i d_i + b_i c_i) & c_i^2 + 2c_i d_i \end{bmatrix}$$

with:

$$a_i = \frac{x_i}{\sigma}(\xi(\rho) + \xi(-\rho)) \qquad b_i = -\frac{x_i}{\sigma}\xi(-\rho)$$

†Correspondence address: Sir Alwyn Williams Building, Lilybank Gardens, Glasgow, G12 8QQ, Scotland

Table 1. ESS of the proposed algorithms and MH with Gaussian proposal. ESS is calculated from 3,000 posterior samples after a burn-in period of 1,000 samples. The low ESS for RMHMC is due to the small step-size required to achieve accurate integration of the Hamiltonian system.

	Metropolis. Hastings	*MALA*	*MMALA*	*Simplified MMALA*	*HMC*	*RMHMC*
minimum ESS	244 ± 24.5	36 ± 7.6	133 ± 42.6	100 ± 38.3	143 ± 18.3	26 ± 17

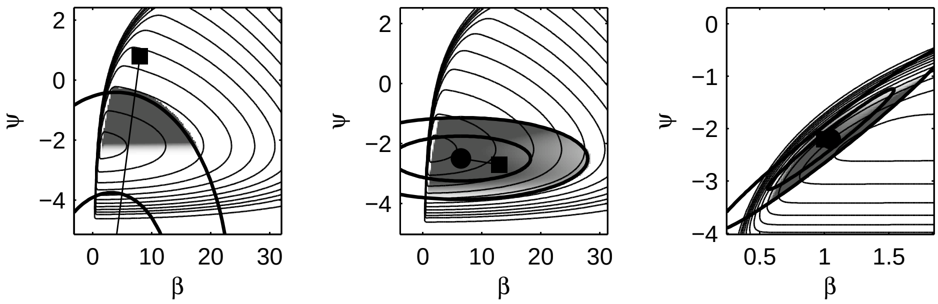

Figure 1. Illustration of the point adaptive proposal mechanism of simplified MMALA. A rectangle denotes the current state while the mean of the proposal is denoted by a filled circle. The 90% and 50% of the Gaussian proposal are presented by the ellipses. The shaded area is the acceptance rate for the underlying regions with dark areas denoting high acceptance rate.

$$c_i = -\frac{\beta x_i}{2\sigma}(\xi(\rho) + \xi(-\rho)) \qquad d_i = \frac{\beta x_i}{2\sigma}\xi(-\rho)$$

$$\rho = \frac{\beta x_i}{\sigma} \qquad \xi(\rho) = \frac{\mathcal{N}(\rho)}{\Phi(\rho)}$$

In Table 1 we compare the proposed MCMC algorithms with a component wise-adaptive Metropolis Hastings (MH) algorithm in terms of ESS. Figure 1 also illustrates the problems associated with the skew posterior distribution and the position dependent proposal mechanisms of MMALA and simplified MMALA. From the right column of Figure 1, we see that in "steep" regions of the posterior the proposal distribution adapts to the curvature forcing the algorithm to make small steps. On the other hand, in smoother regions the proposal allows for larger steps which can sometimes overshoot. This behaviour also leads to very low acceptance rates for large regions where the log joint likelihood is higher than the current state. This is illustrated in the left and middle columns of Figure 1 and is due to the acceptance ratio for non-symetric proposal mechanisms.

References

Imai, K. and D. A. van Dyk (2005, February). A bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics 124*(2), 311–334.

McCulloch, R. E., N. G. Polson, and P. E. Rossi (2000, November). A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics 99*(1), 173–193.

Nobile, A. (1998). A hybrid markov chain for the bayesian analysis of the multinomial probit model. *Statistics and Computing 8*(3), 229–242.

Nobile, A. (2000, December). Comment: Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics 99*(2), 335–345.