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

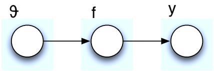
Maurizio Filippone

University of Glasgow

Royal Statistical Society meeting
October 13, 2010

Logistic Regression with Gaussian Process priors

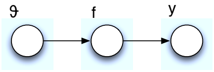
- Latent variables $\mathbf{f}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{f}|\mathbf{0}, K(\boldsymbol{\theta}))$
- Response $y_i|f_i \sim \mathrm{Bern}(y_i|\sigma(f_i))$



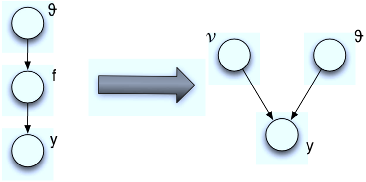
- Gibbs style samplers - also used in this paper in the Log-Gaussian Cox model
- Efficient sampling $\mathbf{f}$ and $\boldsymbol{\theta}$ is complex because of their coupling

Logistic Regression with Gaussian Process priors

- Latent variables $\mathbf{f}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{f}|\mathbf{0}, K(\boldsymbol{\theta}))$
- Response $y_i|f_i \sim \mathrm{Bern}(y_i|\sigma(f_i))$



- Gibbs style samplers - also used in this paper in the Log-Gaussian Cox model
- Efficient sampling $\mathbf{f}$ and $\boldsymbol{\theta}$ is complex because of their coupling
- Natural decoupling: whitening $\mathbf{f}$ (e.g., Murray and Adams '10)



GP Logistic Regression - Metric tensor

- The metric tensor:

$$G = \begin{pmatrix} G_{\mathbf{f}} & \mathbf{0} \\ \mathbf{0} & G_{\boldsymbol{\theta}} \end{pmatrix}$$

is block diagonal, so the geometric based argument in favor of the decoupling of $\mathbf{f}$ and $\boldsymbol{\theta}$, when sampled jointly, does not hold.

GP Logistic Regression - Metric tensor

- The metric tensor:

$$G = \begin{pmatrix} G_{\mathbf{f}} & \mathbf{0} \\ \mathbf{0} & G_{\boldsymbol{\theta}} \end{pmatrix}$$

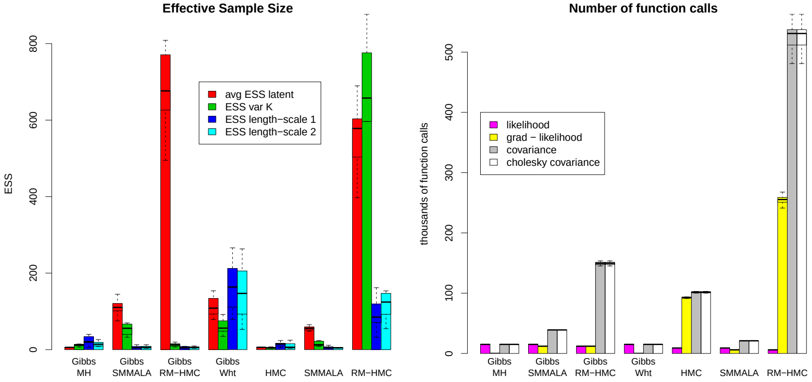
  is block diagonal, so the geometric based argument in favor of the decoupling of $\mathbf{f}$ and $\boldsymbol{\theta}$, when sampled jointly, does not hold.

- Note the complex form of $G_{\boldsymbol{\theta}}$

$$G_{\theta_j,\theta_i} = \frac{1}{2}\mathrm{Tr}\left(K^{-1}\frac{\partial K}{\partial \theta_i}K^{-1}\frac{\partial K}{\partial \theta_j}\right) - \frac{\partial^2 \log[p(\boldsymbol{\theta})]}{\partial \theta_i \partial \theta_j}$$

Results for 100 covariates - Bivariate example

Number of MCMC samples: 2000



Discussion

- Gibbs style samplers are very inefficient in sampling the length-scale parameters
- Gibbs with RM-HMC seems suboptimal in this problem, as it may well be for the Log-Gaussian Cox model presented by the Authors

Discussion

- Gibbs style samplers are very inefficient in sampling the length-scale parameters
- Gibbs with RM-HMC seems suboptimal in this problem, as it may well be for the Log-Gaussian Cox model presented by the Authors
- RM-HMC shows some potential in achieving a ESS comparable to the whitening; investigate guiding Hamiltonians capturing the essence of G_{θ} with less computational effort (G_{θ} and its derivatives are very expensive)
- start off from the whitened model and see whether manifold based samplers improve mixing