Pseudo-Marginal MCMC
(... for Bayesian Inference for Gaussian Processes)

Heiko

Gatsby Tea Talk

May 27, 2014

Outline

Hierarchical Latent Model

- y - Observations
- $f(y|u)$ - Link function
- $u|\theta$ - Latent Process
- θ - Hyperparameters

Bayesian Predictions

$$p(y^*|y) = \int du^* p(y^*|u^*)p(u^*|y)$$

Reminder: Metropolis Hastings Markov Chains

Accept $x_{\text{new}} \sim q(\cdot|x_{\text{old}})$ with probability

$$\min\left(\frac{\pi(x_{\text{new}})q(x_{\text{old}}|x_{\text{new}})}{\pi(x_{\text{old}})q(x_{\text{new}}|x_{\text{old}})}, 1\right)$$

Problems

Joint:

- $\pi(u, \theta|y)$ - super hard, [5, 3]

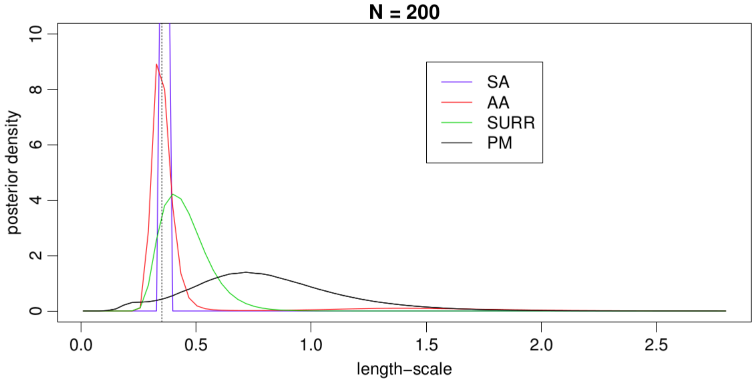
Metropolis-within-Gibbs:

- $\pi(u|\theta, y)$ - easy
- $\pi(\theta|u, y)$ - hard, [6]

Marginal:

- $\pi(\theta|y)$ - easy and hard, [4]

Problems with $\pi(\theta|u,\dots)$



Outline

Auxiliary Stochastic Process

Simulations:

$$u^{(1)}, \ldots, u^{(m)} := \mathbf{u} \sim q_{\mathsf{aux}}(\mathbf{u}|\theta)$$

Estimate:

$$\hat{p}_{\mathbf{u}}(y|\theta) := \frac{1}{m} \sum_{i} f(y|u^{(i)})$$

'Pseudo-Marginal' Metropolis Hastings

Define $\hat{\pi}_{\mathbf{u}}(\theta) := p(\theta)\hat{p}_{\mathbf{u}}(y|\theta)$

MH on $(\theta, \mathbf{u})$:

$$\underbrace{\frac{q_{\text{aux}}(\mathbf{u}^*|\theta^*) \times \hat{\pi}_{\mathbf{u}^*}(\theta^*)}{q_{\text{aux}}(\mathbf{u}|\theta) \times \hat{\pi}_{\mathbf{u}}(\theta)}}_{\text{Target}} \times \underbrace{\frac{q(\theta|\theta^*) \times q_{\text{aux}}(\mathbf{u}|\theta)}{q(\theta^*|\theta) \times q_{\text{aux}}(\mathbf{u}^*|\theta^*)}}_{\text{Proposal}}$$

Exact-Approximate Inference

Repeat:

- Simulate $\mathbf{u} \sim q_{\text{aux}}(\mathbf{u}|\theta)$
- Standard MH using $\hat{\pi}_{\mathbf{u}}(\theta)$ instead of $\pi(\theta)$
- Store: $\cdots \to (\theta_j, \hat{\pi}_{\mathbf{u}_j}(\theta_j)) \to \ldots$

Solve integral.

Magic?

Pro:

- Easy yet exact
- Mixing
- Intractable Likelihoods

Con:

- Hard
- Scaling?
- Sticky

Outline

Filippone and Girolami [4]

Gaussian Process Classification

- $p(\mathbf{y}|\mathbf{f}) = \prod_i \Phi(y_i f_i)$
- $p(\mathbf{f}|\theta) = \mathcal{N}(\mathbf{0}, \Sigma_\theta)$
- $\theta \sim p(\theta)$

Unbiased Estimator $\hat{p}(\mathbf{y}|\theta)$

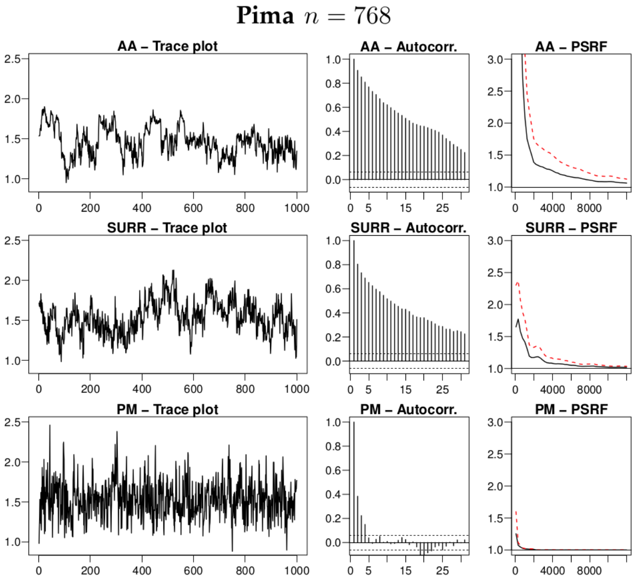
Approximate Inference:

$$q_{\mathrm{EP}}(\mathbf{f}|\theta) \approx p(\mathbf{f}|\mathbf{y},\theta) = \text{Sigmoid} \times \text{Gaussian}$$

Importance Sampling:

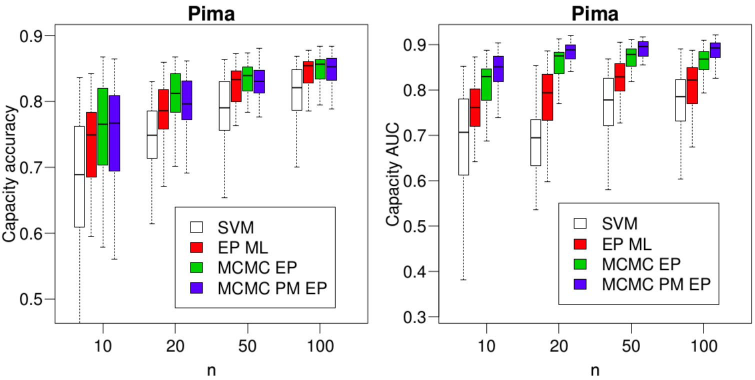
$$\hat{p}_{\mathbf{u}}(y|\theta) = \frac{1}{m}\sum_i p(\mathbf{y}|\mathbf{f}^{(i)})\frac{p(\mathbf{f}^{(i)}|\theta)}{q_{\mathrm{EP}}(\mathbf{f}^{(i)}|\theta)}$$

Mixing

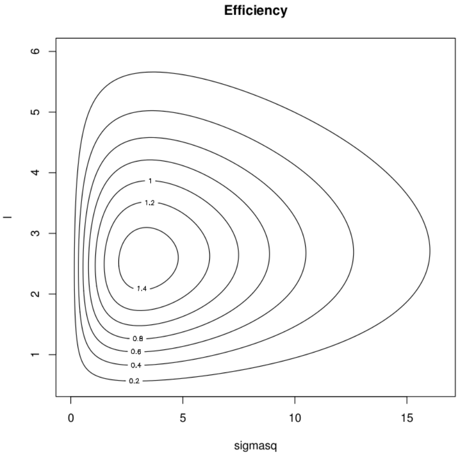


Pima $n = 768$
AA – Trace plot
AA – Autocorr.
AA – PSRF
SURR – Trace plot
SURR – Autocorr.
SURR – PSRF
PM – Trace plot
PM – Autocorr.
PM – PSRF

Bayesian θ



Scaling? ?]



Optimal at $\hat{\mu} \approx 2.56$ and $\hat{\sigma}^2 \approx 3.28$ at which point $\alpha \approx 7.00\%$.
$\hat{\mu}(\sigma)$ insensitive to σ (and vice versa).

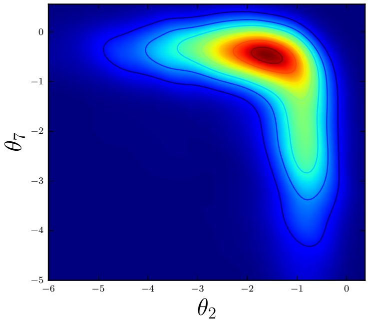
Outline

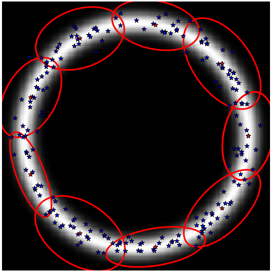
Summary

- Replace $\pi(\theta)$ with unbiased estimator
- Exact-Approximate Inference
- Pseudo-Marginal: Better mixing
- Intractable Likelihoods
- Proposal?

Proposal?

- Random Walk: $q(\theta^*|\theta) = \mathcal{N}(\theta^*|\theta, \Sigma)$
- $\nabla\pi(\theta), \nabla^2\pi(\theta)$ unavailable
- Non-Linear? Kernels! [7]





References

[1] Christophe Andrieu and Gareth O. Roberts. The pseudo-marginal approach for efficient monte carlo computations. The Annals of Statistics, pages 697–725, 2009. URL http://www.jstor.org/stable/30243645.

[2] Mark A. Beaumont. Estimation of population growth or decline in genetically monitored populations. Genetics, 164(3):1139–1160, 2003. URL http://www.genetics.org/content/164/3/1139.short.

[3] M. Filippone, A. F. Marquand, C. R. V. Blain, S. C. R. Williams, J. Mourao-Miranda, and M. Girolami. PROBABILISTIC PREDICTION OF NEUROLOGICAL DISORDERS WITH a STATISTICAL ASSESSMENT OF NEUROIMAGING DATA MODALITIES. The annals of applied statistics, 6(4): 1883–1905, December 2012. ISSN 1932-6157. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3918662/. PMID: 24523851 PMCID: PMC3918662.

[4] Maurizio Filippone and Mark Girolami. Pseudo-marginal bayesian inference for gaussian processes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014. URL http://eprints.gla.ac.uk/93194/.

[5] Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(2):123–214, 2011. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2010.00765.x/full.

[6] Iain Murray and Ryan Prescott Adams. Slice sampling covariance hyperparameters of latent gaussian models. In NIPS, pages 1732–1740, 2010. URL https://papers.nips.cc/paper/4114-slice-sampling-covariance-hyperparameters-of-latent-gaussian-models.pdf.

[7] Dino Sejdinovic, Heiko Strathmann, Maria Lomeli Garcia, Christophe Andrieu, and