Pseudo-Marginal Hamiltonian Monte Carlo with Efficient Importance Sampling

Kjartan Kloster Osmundsen[1]
Tore Selland Kleppe[1] Roman Liesenfeld[2]

[1]Department of Mathematics and Physics
University of Stavanger, Norway

[2]Institute of Econometrics and Statistics
University of Cologne, Germany

EcoSta 2018
City University of Hong Kong
20th June 2018

Background and motivation

- Simulate from target distributions with strong nonlinear dependencies
  - Joint posterior of latent variables and parameters in Bayesian hierarchical models
- Current methods include:
  - Variants of Gibbs sampling (nonlinear dependencies across the blocks)
  - Jointly updating latent variables and parameters (need to ensure that proposals are properly aligned)
  - Pseudo-marginal methods

Pseudo-marginal methods

- Target marginal posteriors of the parameters directly, by integrating out the latent variables
- Relies on ability to produce unbiased, low-variance Monte Carlo estimate of said posterior
  - Sequential Monte Carlo methods
- Our approach: Combining pseudo-marginal Hamiltonian Monte Carlo (Lindsten and Doucet, 2016) with Efficient Importance Sampling (Liesenfeld and Richard, 2003; Richard and Zhang, 2007)

Hamiltonian Monte Carlo (HMC)

- General purpose MCMC method
- Energy preserving dynamical system as the proposal mechanism
  - Approximated by numerical integrator which preserves the dynamics
- Produces close to iid samples
- The main sampling algorithm in Stan, the popular Bayesian modeling software

Pseudo-marginal HMC

- Directly targeting the marginal posterior $p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})$
- $p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x}$ is approximated numerically, using a set of random generated numbers $\mathbf{u}$
- An augmented target distribution corrects for Monte Carlo variation: $\bar{\pi}(\boldsymbol{\theta}, \mathbf{u}) \propto p(\boldsymbol{\theta})\hat{p}(\mathbf{y}|\boldsymbol{\theta}, \mathbf{u})p(\mathbf{u})$
- Regular HMC is applied to the augmented target
- The HMC integrator needs to evaluate $\nabla\bar{\pi}(\boldsymbol{\theta}, \mathbf{u})$, implemented using automatic differentiation software
- To ensure good performance, $\widehat{p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{u})}$ should be a smooth function of both $\mathbf{u}$ and $\boldsymbol{\theta}$
  - Typically not the case for sequential Monte Carlo methods

Efficient Importance Sampling (EIS)

- Our chosen algorithm for calculating $p\widehat{(\mathbf{y}|\boldsymbol{\theta}, \mathbf{u}})$
- EIS chooses importance densities that minimizes the Monte Carlo variance of importance sampling estimates
- A suitable density class $m(\mathbf{x}|\boldsymbol{a}, \theta)$ is chosen, where the EIS parameter $\boldsymbol{a}$ is chosen so MCMC variance is minimized
- The local minimization problems for $\boldsymbol{a}$ (one for each observation) are reduced to linear least squares problems, solved iteratively from a starting value $\boldsymbol{a}_0$
- $p\widehat{(\mathbf{y}|\boldsymbol{\theta}, \mathbf{u}}) = \frac{1}{n}\sum_{i=1}^{n} \frac{p(\mathbf{y}|\mathbf{x}^{(i)},\boldsymbol{\theta})p(\mathbf{x}^{(i)}|\boldsymbol{\theta})}{m(\mathbf{x}^{(i)}|\boldsymbol{a},\boldsymbol{\theta})}, \quad \mathbf{x}^{(i)} \sim m(\cdot|\boldsymbol{a}, \boldsymbol{\theta}, \mathbf{u})$

Simulation experiments

- State space models
- $y_t|x_t, \boldsymbol{\theta} \sim g_t(\cdot|x_t, \boldsymbol{\theta}),\ t = 1, \dots, T,$
  $x_t|x_{t-1}, \boldsymbol{\theta} \sim \mathcal{N}(\cdot|\mu_t(x_{t-1}, \boldsymbol{\theta}), \sigma_t^2(x_{t-1}, \boldsymbol{\theta})),\ t = 2, \dots, T,$
  $x_1|\boldsymbol{\theta} \sim \mathcal{N}(\cdot|\mu_1(\boldsymbol{\theta}), \sigma_1^2(\boldsymbol{\theta}))$
- Stan is used as a benchmark

One-parameter model

- $y_t \sim \exp(x_t/2) \cdot \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1), t \in (1,2,\ldots,T),$
  $x_t \sim \theta + \eta_t, \quad \eta_t \sim \mathcal{N}(0,1), t \in (1,2,\ldots,T)$
- Simulated observations

θ	CPU time (s)	Post. mean	Post. std.	ESS	ESS/s
HMC-EIS (0 reg)	16.4	0.026	0.063	631.8	38.4
HMC-EIS (1 reg)	74.2	0.026	0.063	876.5	11.8
Stan	2.1	0.026	0.063	319	151.2

Stochastic volatility model

- $y_t = \exp(x_t/2) \cdot \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1), t \in (1, 2, \ldots, T),$
  $x_t = \gamma + \delta x_{t-1} + v\eta_t, \quad \eta_t \sim \mathcal{N}(0,1), t \in (2, 3, \ldots, T),$
  $x_1 = \frac{\gamma}{1-\delta} + \frac{v}{\sqrt{1-\delta^2}}\eta_1, \quad \eta_1 \sim \mathcal{N}(0,1)$
- Dollar/Pound exchange rates

δ	CPU time (s)	Post. mean	Post. std.	ESS	ESS/s
HMC-EIS (2 reg)	245	0.976	0.01	469	1.92
Stan	10	0.976	0.01	284	28.6

Constant elasticity of variance diffusion model

- $y_t = x_t + \sigma_y \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1), t \in (1,2,\dots,T),$
  $x_t = x_{t-1} + \Delta(\alpha - \beta x_{t-1}) + \sigma_x x_{t-1}^{\gamma} \sqrt{\Delta} \eta_t,$
  $\eta_t \sim \mathcal{N}(0,1), t \in (2,3,\dots,T), \quad x_1 \sim \mathcal{N}(y_1, 0.01^2),$
- Short-term interest rates
- Stan is not converging (limited information in the observations, $\sigma_y = 0.0005$)
  - Compare our results to modified Cholesky Riemann manifold Hamiltonian Monte Carlo (MCRMHMC) and Particle Gibbs.

α	CPU time (s)	Post. mean	Post. std.	ESS	ESS/s
HMC-EIS (1 reg)	473	0.01	0.009	1000	2.11
MCRMHMC	16200	0.01	0.009	1000	0.06
Particle Gibbs	90	0.01	0.009	456	5.07

σ_x	Post. mean	Post. std.	ESS	ESS/s
HMC-EIS (1 reg)	0.41	0.06	945	1.73
MCRMHMC	0.41	0.06	579	0.04
Particle Gibbs	0.41	0.06	79	0.88

Conclusion

- We have combined HMC with EIS
- Produces stable, effective and accurate results.
- Competitive computational cost for models with advanced latent processes.

Thank you for your attention!