Diffusion Limits for MCMC Paths

Alexandros Beskos

Statistical Science, UCL

Warwick, 19 March 2009

Joint work with:
Natesh Pillai, Gareth Roberts, Andrew Stuart

Outline

Outline

Metropolis-Hastings

- **Objective:** Sample distribution $\pi_n : \mathbb{R}^n \mapsto \mathbb{R}^+$.
- **Method:** Construct Markov chain reversible w.r.t. π_n and simulate up to stationarity.

1. Propose move $X \rightarrow Y$ according to user-specified

$$q_n(X, dY) = q_n(X, Y)dY$$

2. Accept Y with probability

$$a_n(X, Y) = 1 \wedge \frac{\pi_n(Y)q_n(Y, X)}{\pi_n(X)q_n(X, Y)}$$

otherwise stay at X.

3. Simulate $X^{(1)}, X^{(2)}, \dots$ up to equilibrium.

Local MCMC Algorithms

- Proposed move could be:

$$Y = X + \sqrt{h_n}\, Z, \quad Z \sim N(0, I_n)$$

giving Random-Walk Metropolis (RWM) algorithm.

- It could also be:

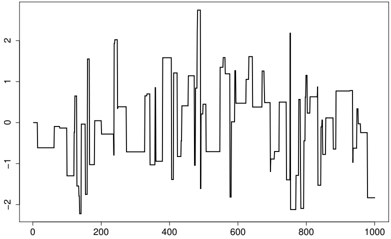
$$Y = X + \frac{h_n}{2} \nabla \log \pi_n(X) + \sqrt{h_n}\, Z$$

giving Metropolis-adjusted Langevin algorithm (MALA).

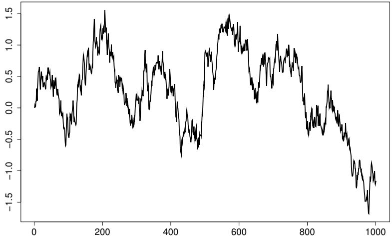
- Goldilocks Principle:
Step-size h_n should neither be "too small" or "too big".

Goldilocks Principle

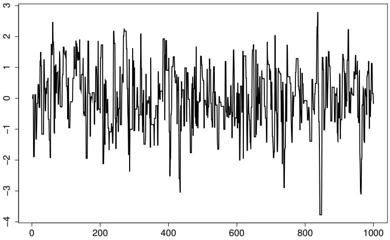
$h_1 = 100,\ a_1 = 0.085$



$h_1 = 0.01,\ a_1 = 0.975$



$h_1 = 4,\ a_1 = 0.509$:



Some Questions

- What is the "optimal" choice of h_n, a_n?
- What is the limiting behaviour of MCMC as $n \to \infty$?
- Adaptive schemes have tried to address h_n selection dynamically (e.g. Haario, Atchad, Roberts, Rosenthal, Andrieu, Moulines).
- Here we look at non-dynamic setting.

Outline

Results

- Consider iid target distribution

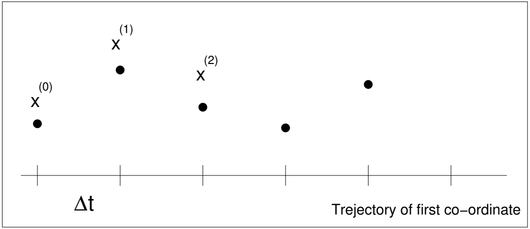
$$\pi_n(X) = \Pi_{i=1}^{n} f(x_i)$$

  and apply RWM, MALA for $h_n = l \times \Delta t$.

- Scale step-size as:

$$RWM: \quad \Delta t = n^{-1}, \qquad MALA: \quad \Delta t = n^{-1/3}$$

  and bring MCMC points close:



Results

- **Theorem** (Roberts et al., 97; Roberts & Rosenthal, 98)

  Continuous time process $x^{([t/\Delta t])}$ converges weakly to:

$$\frac{dx}{dt} = \frac{1}{2} s(l) \, (\log f)'(x) + \sqrt{s(l)} \, \frac{dW}{dt}$$

  for speed function:

$$s(l) = l^2 a(l)$$

  where $a(l)$ is limiting acceptance probability:

$$a(l) = \lim_n E\,[\, a_n(x, y) \,] \in (0, 1)$$

Remarks

- MCMC similar (for large n) to **Euler scheme** on diffusion.
- Speed function $s(l)$ is maximised for

$$RWM: \quad a(l) = 0.234$$
$$MALA: \quad a(l) = 0.574$$

- 'Mixing' cost of algorithms is $\mathcal{O}(\Delta t^{-1})$, so:

$$RWM: \quad n$$
$$MALA: \quad n^{1/3}$$

Remarks: Optimality

- Efficiency can be compared in terms of integrated autocorrelation time:

$$\tau_g(h_n) = 1 + \sum_{j=1}^{\infty} Corr(g(x^{(0)}), g(x^{(j)}))$$

so that:

$$Var(\frac{\sum_{j=1}^{J} g(x^{(j)})}{J}) \approx \frac{Var[g(x)]}{J} \times \tau_g$$

- Because of diffusion limit we can factorise:

$$(\tau_g(h_n))^{-1} \approx s(l)\, c_g \times \Delta t$$

Outline

Generalizing

- We considered targets:

$$\frac{d\pi}{d\pi_0}(X) \propto \exp(-\Phi_n(X))$$

where

$$\pi_0 = \prod_{i=1}^{n} \frac{1}{\lambda_i} f\left(\frac{x_i}{\lambda_i}\right); \quad \lambda_i = i^{-\kappa}$$

- RWM: $h_n = l \times n^{-2\kappa-1}$, MALA: $h_n = l \times n^{-2\kappa-1/3}$
- We use **mean square jump**:

$$M(n) = E\,|x^{(j+1)} - x^{(j)}|^2$$

Results

- Consider the norm $\|X\|_s = \left(\sum_{i=1}^{n} i^{2s} x_i^2\right)$
- Assume that Φ_n is such that:

$$\Phi_n(X) \geq M$$
$$|\Phi_n(Y) - \Phi_n(X)| \leq L(\|X\|_s, \|Y\|_s)\|X - Y\|_{s'}$$
$$|\Phi_n(X)| \leq C(1 + \|X\|_{s''})$$

for L continuous, and $s, s^{'}, s^{''} < \kappa - 1/2$.

- Then, we can factorise out $s(l)$:

$$MALA: \quad M(n) = s(l) \times n^{-2\kappa - 1/3}$$
$$RWM: \quad M(n) = s(l) \times n^{-2\kappa - 1}$$

More Results

- Target is on Hilbert space:

$$\frac{d\pi}{d\pi_0}(X) = \exp\{-\Phi(X)\}$$

  with $\pi_0 \sim N(0, C)$.

- $C\, e_i = \lambda_i^2 e_i, \quad \lambda_i = i^{-\kappa}$
- MCMC trajectory on discretised space converges to **SPDE**:

$$\frac{dX}{dt} = \frac{1}{2}\, s(l)\left(-X - C\nabla\Phi(X)\right) + \sqrt{s(l)\, C}\,\frac{dW}{dt}$$

- Pillai & Stuart, 09.

Outline

Conditioned Diffusions

- **Sample** $X = X(u)$:

$$\frac{dX}{du} = g(X) + \gamma \frac{dW}{du}$$

- **Conditioned on:**
  - Diffusion bridge: $X(0)$ and $X(1)$ (Molecular Dynamics)
  - Interpolate data: $X(i\Delta)_{i=0}^{N}$ (Econometrics)
  - Noisy observation of $X(u)$ (Signal Filtering)

Data Assimilation

- Navier-Stokes inverse problem for fluid dynamics.
- **Sample** $X_0 \in L^2(\Omega, R^2)$ initial condition for PDE:

$$\frac{\partial X}{\partial t} + X \cdot \nabla X = \nu \Delta X + g, \quad X(0) = X_0 \ .$$

- **Conditioned** on (Lagrangian) observations:

$$\frac{dz_j}{dt} = X(z_j, t), \quad z_j(0) = z_{j,0} \ ,$$
$$y_{j,k} = z_j(t_k) + N(0, \Sigma_{j,k}) \ .$$

- (Assuming Gaussian prior on X_0.)

Outline

Summary

- We studied complexity of Local MCMC algorithms in high (sometimes infinite!) dimensions.
- Obtained interpretation of RWM and MALA as Euler approximations of a diffusion.
- Diffusion limit allows for uniform optimization.
- Results can be extended to non-trivial targets.
- Non-product targets can give SPDE limits.

Some Further Directions

- Proving SPDE led to development of new framework for proving convergence.
- New framework avoids enormous complexities with using generators.
- In the case of RWM and MALA it avoids many of the conditions required in previous works.
- We are currently applying new framework to Hybrid Monte-Carlo algorithm, with evidence of hypoelliptic diffusion limit.

References

- Bédard, M. (07)
  Weak Convergence of Metropolis Algorithms for Non-iid Target Distributions, Ann. Appl. Probab. 17, 1222-44.
- Beskos, A., Pillai, N., Roberts, G., Stuart, A. (09)
  Optimal Tuning of Hybrid Monte Carlo, in progress.
- Beskos, A., Roberts, G., Stuart, A. (09)
  Optimal scalings for local Metropolis-Hastings chains on non-product targets in high dimensions, to appear in Ann. Appl. Probab..
- Pillai, N., Stuart, A. (09)
  SPDE Limits of Rescaled Random Walk Metropolis Algorithm in Hilbert Spaces, in progress.
- Roberts, G., Gelman, A., Gilks, W. (97)
  Weak convergence and optimal scaling of random walk Metropolis algorithms, Ann. Appl. Probab., 7, 110-120.
- Roberts, G., Rosenthal, J. (98)
  Optimal scaling of discrete approximations to Langevin diffusions, JRSSB, 60, 255-268.