MCMC In High Dimensions

Andrew Stuart

Mathematics Institute and Centre for Scientific Computing,
University of Warwick

SIAM, Oxford, 5th January 2007

Collaboration with:
Alexandros Beskos

Funded by EPSRC

Outline

Outline

Markov Chain Monte Carlo

- **Objective:** Sample distribution $\pi_n : \mathbb{R}^n \mapsto \mathbb{R}^+$.
- **Method:** Construct Markov chain $\{x^{(k)}\}$ with π_n invariant.
- **Ergodicity:** the Markov chain samples π_n after mixing time is reached and

$$\frac{1}{K}\sum_{k=1}^{J} f(x^{(k)}) \to \int_{\mathbb{R}^n} f(x)\pi_n(dx) \quad \text{as} \quad J \to \infty.$$

- **Question:** How do methods behave as $n \to \infty$?

Metropolis-Hastings

- **Propose** move $x \rightarrow y$ according to user-specified

$$q_n(x, dy) = q_n(x, y)dy$$

- **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.
- **New Markov chain** has π_n as invariant.

Local Metropolis-Hastings Algorithms

- Random Walk Metropolis (**RWM**):

$$y = x + \sigma_n Z, \quad Z \sim \mathcal{N}(0, I_n)$$

- Metropolis-Adjusted Langevin Algorithm (**MALA**)
  The Langevin SDE

$$dX_t = \frac{1}{2} \nabla \log \pi_n(X_t) dt + dW_t$$

  has invariant distribution π_n. Suggests the proposal:

$$y = x + \frac{\sigma_n^2}{2} \nabla \log \pi_n(x) + \sigma_n Z, \quad Z \sim \mathcal{N}(0, I_n)$$

- **Question:** What is the appropriate σ_n for large n?
- **Courant restriction** is computational PDE analogy

The Context

- Existing work concerns product targets:

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

- Roberts and coworkers (1997–2001) have shown:

$$\text{RWM:} \quad \sigma_n^2 = \mathcal{O}(n^{-1})$$
$$\text{MALA:} \quad \sigma_n^2 = \mathcal{O}(n^{-1/3})$$

in the sense that, in stationarity, for these scalings:

$$\lim_{n\to\infty} \mathbb{E}[a_n(x,y)] \in (0,1)$$

and, for larger time-steps,

$$\lim_{n\to\infty} \mathbb{E}[a_n(x,y)] = 0.$$

- Mixing time $M(n)$ ="number of steps to reach stationarity",

RWM: $M(n) = \mathcal{O}(n)$, MALA: $M(n) = \mathcal{O}(n^{1/3})$

The Context

Our work:

- We investigate non-product targets using a new approach, extending existing results and, in the process, simplfying the proofs.
- Furthermore, we exploit ideas from numerical analysis to construct new schemes which, in important applications, give $\sigma_n^2 = \mathcal{O}(1)$.
- We demonstrate the relevance of our results for infinite dimensional sampling applications.

Outline

The Family of Targets

- We consider changes of measure from product measures:

$$\tilde{\pi}_n(x) = \prod_{j=1}^{n} \frac{1}{\lambda_j} f\left(\frac{x_j}{\lambda_j}\right)$$

- λ_j is standard deviation of x_j.
- Our target π_n is defined as:

$$\frac{d\pi_n}{d\tilde{\pi}_n}(x) = \exp\Big(-G_n(x)\Big)$$

for $G_n : \mathbb{R}^n \mapsto \mathbb{R}$.

- Motivated by applications, we assume that

$$\lambda_j = j^{-\kappa}, \quad j = 1, 2, \ldots n$$

for integer $\kappa \geq 0$.

Infinite-Dimensional Motivation

- We find that if $\tilde{\pi}_n$ is the target then:

$$\begin{aligned} \text{RWM:} \quad & \sigma_n^2 = \mathcal{O}(\lambda_n^2 n^{-1}) \\ \text{MALA:} \quad & \sigma_n^2 = \mathcal{O}(\lambda_n^2 n^{-1/3}) \end{aligned}$$

- We anticipate similar MALA, RWM behaviour for π_n, $\tilde{\pi}_n$, in the presence of absolute continuity in the limit $n = \infty$:

$$\frac{d\pi_\infty}{d\tilde{\pi}_\infty}(x) = \exp\Big(-G_\infty(x)\Big)$$

- Such systems appear in many applications: conditioned diffusions, conditioned Gaussian random fields.
- In such applications π_n is then a discretization of π_∞.

Theorems

Note: For MALA we use proposal as if target was $\tilde{\pi}_n$.

Theorem (MALA): Assume

$$\exists\, M > 0 : \textit{for all } n,\ |G_n| \leq M$$

and conditions on f. Then the average acceptance probability of MALA in stationarity satisfies:

$$\liminf_{n\to\infty} \mathbb{E}[a_n(x,y)] > 0, \quad \textit{if } \sigma_n^2 \leq \mathcal{O}(\lambda_n^2 n^{-1/3}),$$

$$\lim_{n\to\infty} \mathbb{E}\,[a_n(x,y)] = 0, \qquad \textit{if } \sigma_n^2 > \mathcal{O}(\lambda_n^2 n^{-1/3}).$$

Theorem (RWM): Similar; replace $\mathcal{O}(\lambda_n^2 n^{-1/3}) \to \mathcal{O}(\lambda_n^2 n^{-1})$.

Sketch of Proof

Average acceptance probability:

$$\alpha_n = \mathbb{E} a_n = \mathbb{E}\left(1 \wedge e^{R_n}\right)$$

- $\sigma_n^2 \leq \mathcal{O}(\lambda_n^2 n^{-1/3})$: we have $\sup_n \mathbb{E}|R_n| < \infty$ and, for any $\gamma > 0$,

$$\alpha_n \geq e^{-\gamma}\mathbb{P}(|R_n| \leq \gamma) \geq e^{-\gamma}\left(1 - \frac{\mathbb{E}|R_n|}{\gamma}\right) > 0.$$

- $\sigma_n^2 > \mathcal{O}(\lambda_n^2 n^{-1/3})$: we get $\mathbb{E}\, R_n = -2c_n \downarrow -\infty$ and

$$\alpha_n \leq e^{-c_n} + \mathbb{P}\Big(R_n \geq -c_n)\Big) \leq e^{-c_n} + \frac{\mathbb{E}\,|\,R_n - \mathbb{E}\,R_n\,|}{c_n}.$$

Special Case: MALA + Gaussian Reference Measure

- **Target:** $\pi_n(x) = \exp\Big(-G_n(x)\Big)\ \tilde{\pi}_n(x)$ with

$$\tilde{\pi}_n = \prod_{j=1}^{n} N(0, \lambda_j^2)$$

- **Proposal:** Implicit MALA ($\theta = 0$ in previous)

$$y = x - (1-\theta)\,\frac{\sigma_n^2}{2}\,L_n\,x - \theta\,\frac{\sigma_n^2}{2}\,L_n\,y + \sigma_n\,Z, \quad Z \sim \mathcal{N}(0, I_n)$$

$\theta \in [0,1]$, L_n diagonal $n \times n$ with j-th diagonal element λ_j^{-2}.

Theorem: If $\theta \neq 1/2$ then $\sigma_n^2 = \mathcal{O}(\lambda_n^2 n^{-1/3})$.
If $\theta = 1/2$ then $\sigma_n^2 = \mathcal{O}(1)$.

Outline

Conditioned Diffusions

- **Sample** $X(t) \in L^2([0,1],\mathbb{R})$:

$$\frac{dX}{dt} = f(X) + \frac{dW}{dt}$$

- **Given**

$$X(0) = X^- \quad \& \quad X(1) = X^+$$

  - Target measure: π_∞
  - Reference measure: Brownian Bridge ($f \equiv 0$): $\tilde{\pi}_\infty$
  - From the Girsanov theorem:

$$\frac{d\pi_\infty}{d\tilde{\pi}_\infty}(x) = \exp\Big(-G_\infty(x)\Big)$$

for $G_\infty : L^2([0,1],\mathbb{R}) \mapsto \mathbb{R}$.

Unveil Product Structure

- Karhunen-Loève representation of $x \in L^2$ from Gaussian measure $\mathcal{N}(0, \mathcal{C})$ is

$$x(t) = \sum_{j=1}^{\infty} x_j e_j(t).$$

- Here $x_j \sim \mathcal{N}(0, \lambda_j^2)$ and $\mathcal{C}$ has evalues/evectors $(\lambda_j, e_j(t))$.
- For Brownian bridge

$$\lambda_j^2 = (\pi j)^{-2}, e_j(t) = \sin(j\pi t).$$

- Using the isometry between L^2 and ℓ^2 this (random) Fourier series shows

$$\mathcal{N}(0, \mathcal{C}) \leftrightarrow \prod_{j=1}^{\infty} \mathcal{N}(0, \lambda_j^2).$$

Conditioned Diffusions

- **Infinite-Dimensional** diffusion-bridge target π_∞:

$$\pi_\infty(x) = \exp\Big(-G_\infty(x)\Big)\,\tilde{\pi}_\infty(x), \quad \tilde{\pi}_\infty = \prod_{i=1}^{\infty} \mathcal{N}(0, \lambda_i^2)$$

  with Brownian bridge eigenvalues $\lambda_i^2 = \pi^{-2} i^{-2}$.
- **Spectral Method** π_n:
  Use Fourier expansion truncation:

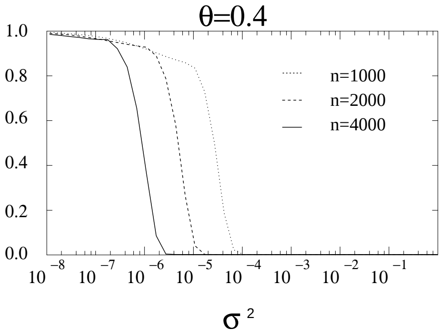
$$x = \sum_{j=1}^{\infty} x_j e_j \approx \sum_{j=1}^{n} x_j e_j$$

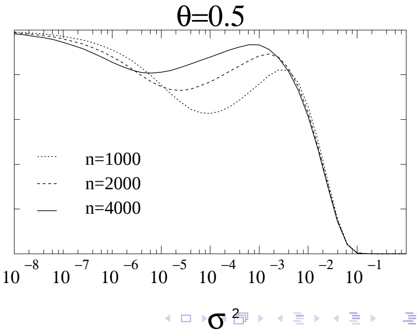
- Theory suggests implicit MALA with $\theta = 1/2$ giving $\sigma_n^2 = \mathcal{O}(1)$.

Example - MALA Results

- We applied implicit MALA to sample a non-Gaussian bridge.
- We verified $\sigma_n^2 = \mathcal{O}(\lambda_n^2 n^{-1/3}) = \mathcal{O}(n^{-7/3})$ for $\theta \neq 1/2$.
- And at $\theta = 1/2$, $\sigma_n^2 = \mathcal{O}(1)$.

Average Acceptance Probability in Stationarity





Outline

What We Have Shown

1. We have found scaling of step σ_n^2 in Metropolis-Hastings proposals for non-product targets in high dimensions.
2. We have thus extended existing results in literature in a manner which makes them much more applicable.
3. When the reference measure is Gaussian, an implicit scheme gives MALA with scaling $\sigma_n^2 = \mathcal{O}(1)$.
4. Changes of measure from Gaussian law appears in many applications.

What Remains Open

1. Relax conditions on $\{G_n\}$ for theorems:
   - $|G_\infty(x)|_\beta \le M|x|_\gamma \quad \forall x.$
   - $|G_\infty(x) - G_\infty(y)|_\beta \le M|x-y|_\gamma \quad \forall x, y.$
2. Does the step scaling $O(n^{-\rho})$ imply mixing $O(n^\rho)$?
   - For product measure MCMC method has an SDE limit for any fixed component x_j (Roberts et al);
   - this facilitates proof of mixing time;
   - we conjecture existence of a limiting SPDE for entire vector x in non-product case.

References

- G.O. Roberts, A. Gelman and W.R. Gilks. ”Weak convergence and optimal scaling of random walk Metropolis algorithms”. Ann. Appl. Prob. **7**(1997), 110–120.
- G.O. Roberts and J.S. Rosenthal. ”Optimal scaling of discrete approximations to Langevin diffusions”. J. Roy. Stat. Soc. **60B**(1998), 255–268.
- A. Beskos, G.O. Roberts, A.M. Stuart and J. Voss. ”An MCMC Method for diffusion bridges.” See:

  http://www.maths.warwick.ac.uk/~stuart/sample.html

- A. Beskos and A.M. Stuart. ”Scalings for local Metropolis-Hastings chains on non-product targets.” In preparation.