Imperial College London

On Non-Linear MCMC

Ajay Jasra

ajay.jasra@imperial.ac.uk

Department of Mathematics

Imperial College London

Outline

- Introduction
- Non-Linear MCMC
- Some Algorithms
- The Theoretical Analysis
- Summary

Introduction

- The problem is to compute expectations

$$\pi(f) = \int_E f(x)\pi(x)dx$$

w.r.t a probability measure π, for many f.

- In many cases of interest it is not possible/sensible using analytic or deterministic numerical methods.

- This has a wide variety of real applications, from statistics to physics and much more.

- Basic idea: Sample an ergodic Markov chain of stationary distribution π. The estimate of $\pi(f)$ is

$$S_n^X(f) = \frac{1}{n+1}\sum_{i=0}^{n} f(X_i)$$

where $X_0, X_1, \ldots, X_n$ are samples from our Markov chain. Termed MCMC.

- For probability measures π
  - possessing many modes
  - and/or complex inter-dependence structure,

  MCMC can be slow to move around the space.

- As a result, many advanced MCMC algorithms:
  - Adaptive MCMC
  - Equi-energy sampler
  - Particle MCMC

  to name but a few.

- In this talk I present another alternative: Non-Linear MCMC.

Non-Linear MCMC

- Let $\mathscr{P}(E)$ be the class of probability measures on E.
- Standard MCMC has a kernel: $K : E \to \mathscr{P}(E)$.
- Non-Linear MCMC has a kernel: $K : E \times \mathscr{P}(E) \to \mathscr{P}(E)$.
- The kernel is to operate in some non-linear way on an input probability measure.
- The idea is to induce a kernel which mixes much faster than an ordinary MCMC kernel.

- A linear example:

$$K_\mu(x, dy) \quad = \quad (1-\epsilon)K(x, dy) + \epsilon\mu K(dy) \ ,$$

  where

  - K is a Markov kernel of invariant distribution π
  - $\epsilon \in (0, 1)$
  - $\mu K(dy) = \int \mu(dx)K(x, dy)$.

- Simulating from K_π allows regenerations from π, with K_π strongly uniformly ergodic.
- It is not possible to sample from K_π. The idea is then to approximate the kernel using simulated samples in the past.

- By this I mean to approximate, at time $n+1$ of the algorithm, the kernel by:

$$K_{S_n^X}(x,dy) \quad = \quad (1-\epsilon)K(x,dy) + \epsilon S_n^X K(dy) \ ,$$

- Such an algorithm, brings previously simulated samples back, with probability ϵ and then samples from K.
- Note that other approximation schemes are possible. In PMCMC, a new process is sampled (i.e. the particle filter) at every step and that yields a Markov chain.
- This leads us to our rather loose framework:
  - Identify a non-linear kernel, that admits π as an invariant distribution and can be expected to mix faster than an ordinary MCMC kernel
  - Construct a stochastic process that approximates the kernel, which can be simulated in practice.

Some Algorithms

- Consider the non-linear kernel:

$$\Pi_{\pi\times S_n^Y}((x_n,y_n),d(x_{n+1},y_{n+1})) =$$

$$\Big[(1-\epsilon)K(x_n,dx_{n+1})+\epsilon\Phi(S_n^Y)(dx_{n+1})\Big]P(y_n,dy_{n+1})$$

  where

  - P is a Markov kernel of invariant distribution η
  - $\Phi(\mu)(dx)=\mu(gK)(dx)/\mu(g)$
  - $g=d\pi/d\eta$
  - $(\pi\times\eta)\Pi_{\pi\times\eta}(d(x,y))=\pi\times\eta(d(x,y))$.

- η should be easier to sample than π but related to it.
- Φ will select a value of the chain $\{Y_n\}$ and try to help the $\{X_n\}$ process.

The Theoretical Analysis

- It is sought to prove a strong law of large numbers for the sample path:

$$S_n^X(f) \quad = \quad \frac{1}{n+1}\sum_{i=0}^{n} f(X_i).$$

- Introduce the sequence of probability distributions $\{S_n^\omega := 1/(n{+}1)\sum_{i=0}^n \omega(S_i^Y)\}_{n\geq 0}$ where $\omega(\mu)$ is the invariant probability distribution of K_μ.

- Adopt the decomposition

$$S_n^X(f) - \pi(f) \quad = \quad S_n^X(f) - S_n^\omega(f) + S_n^\omega(f) - \pi(f)\,.$$

- The analysis of the first term on the R.H.S relies upon a classical martingale argument.

- Under the assumptions in the paper the a solution to Poisson's equation exists

$$f(x) - \omega(\mu)(f) \quad = \quad \hat{f}_\mu(x) - K_\mu(\hat{f}_\mu)(x)\,.$$

- The first term on the R.H.S is

$$\begin{aligned}(n+1)[S_n^X - S_n^\omega](f) = M_{n+1} \\ + \sum_{m=0}^{n} [\hat{f}_{S_{m+1}^Y}(X_{m+1}) - \hat{f}_{S_m^Y}(X_{m+1})] + \hat{f}_{S_0^Y}(X_0) - \hat{f}_{S_{n+1}^Y}(X_{n+1})\ ,\end{aligned}$$

where

$$M_n = \sum_{m=0}^{n-1} [\hat{f}_{S_m^Y}(X_{m+1}) - K_{S_m^Y}(\hat{f}_{S_m^Y})(X_m)]\ ,$$

is such that $\{M_n, \mathcal{G}_n\}$ is a martingale.

- The Martingale can be controlled using Burkholder's inequality.

- The other expressions can be dealt with via continuity properties of the solution to the Poisson equation.
- The expression $S_n^{\omega}(f) - \pi(f)$ is more complex and appears to require a strong law of large numbers for $U-$statistics, in order to prove the result.
- The assumptions are based upon standard Foster-Lyapunov drift inequalities and are relatively standard.

Summary + Extensions

- Investigated a new approach to stochastic simulation: Non-Linear MCMC.
- The conditions required for convergence may be relaxed; e.g. using sub-geometric kernels.
- To design more elaborate methods to control the evolution of the empirical measure.
- To design ‘better’ algorithms.

Acknowledgements

- Joint work with Christophe Andrieu, Arnaud Doucet and Pierre Del Moral.
- During this work I was supported by the University of Cambridge, Universitié Nice, ISM Tokyo Japan and Imperial College London.
- Paper is available from: `http://stats.ma.ic.ac.uk/a/aj2/public_html/papers/NLREV1.PDF`.