Lecture 21. Convergence of Markov Chain Monte Carlo.

Note Title 6/8/2006

Claim: MCMC will converge exponentially fast.

This lecture will not give a rigorous mathematical proof. But will sketch a proof which can be made rigorous by further work.

This claim will be sketched for MCMC which obey detailed balance. Similar results can be obtained for MCMC which don't obey detailed balance.

Results depend on the eigenvalues of $K(x|y)$.

Subclaim if $K(x|y)$ obeys detailed balance, then $Q(x,y) = P(y)^{1/2} K(x|y) P(x)^{-1/2}$ is a symmetric matrix.

This can be checked as follows.

If $Q(x,y) = Q(y,x)$, then

$$P(y)^{1/2} K(x|y) P(x)^{-1/2} = P(x)^{1/2} K(y|x) P(y)^{-1/2}$$

$$\Rightarrow K(x|y) P(y) = K(y|x) P(x) \quad \text{detailed balance.}$$

(Converse follows directly).

Treat $Q(x,y)$ as a symmetric matrix and exploit standard linear algebra results.

(page 2)

μ index of the eigenvalues/eigenvectors

Linear Algebra

(1) $Q(x,y)$ has real eigenvalues λ^μ and eigenvectors $e^\mu(x)$

s.t. $\sum_y Q(x,y)\, e^\mu(y) = \lambda^\mu e^\mu(x)$

(2) These eigenvectors are orthogonal

$$\sum_x e^\mu(x) e^\nu(x) = 0, \text{ if } \mu \neq \nu$$
$$= 1, \text{ if } \mu = \nu.$$

(3) We can express: (spectral theorem)

$$Q(x,y) = \sum_\mu \lambda^\mu e^\mu(x) e^\mu(y)$$

(4) which implies that

$$Q^M(x,y) = \sum_\mu \{\lambda^\mu\}^M e^\mu(x) e^\mu(y)$$

$Q^M(x,y)$ is the matrix obtained by taking the product of $Q(\cdot,\cdot)$ with itself M times.

Page 3.

Hence, we can write

$$K^M(y|x) = \sum_\mu (\lambda_\mu)^M P(x)^{-1/2} P(y)^{1/2} e^\mu(x) e^\mu(y).$$

Re-express as

$$K^M(y|x) = \sum_\mu (\lambda_\mu)^M u^\mu(x) v^\mu(y)$$

with $u^\mu(x) = e^\mu(x) P(x)^{-1/2}$
$v^\mu(x) = e^\mu(x) P(x)^{1/2}$

It can be checked that

$$\sum_x v^\mu(x) K(y|x) = \lambda^\mu v^\mu(y)$$

$$\sum_y u^\mu(y) K(y|x) = \lambda^\mu u^\mu(x)$$

u & v are left and right eigenvectors of $K(y|x)$ respectively.

They come in pairs $u^\mu(x)$ & $v^\mu(x)$ with same eigenvalue λ^μ (easy to check)

(If $K(y|x)$ is symmetric, then $u^\mu(x) = v^\mu(x)$ and we are back to standard linear algebra)

Also $\sum_x v^\mu(x) u^\nu(x) = 1$, if $\mu = \nu$
0, if $\mu \neq \nu$

This implies that any function (vector) can be expanded as:

$$f(x) = \sum_{\mu} \{ \sum_{y} f(y) u^{\mu}(y) \} v^{\mu}(x)$$

$$f(x) = \sum_{\mu} \{ \sum_{y} f(y) v^{\mu}(y) \} u^{\mu}(x)$$

Now, the first eigenvalue λ^1, and its left and right eigenvectors $u^1(x)$ & $v^1(x)$ can be computed to be

$$v^1(x) = P(x) \leftarrow \text{target distribution.}, \quad u^1(x) = 1, \quad \lambda^1 = 1.$$

(This follows directly from the detailed balance equations).

The other eigenvalues, $\lambda^i : i = 2, \ldots, N$ obeys $|\lambda^i| < 1$ (except for pathological cases).

This follows from the condition that $\sum_{y} K(y|x) = 1$, for all x.

Page 5.

Now suppose we start the MCMC with distribution $P_0(x)$ (any distribution).

We can express

$$P_0(x) = \sum_\mu \left(\sum_y P_0(y) u^\mu(y)\right) v^\mu(x)$$

$$P_0(x) = \underbrace{P(x)}_{\text{target distribution}} + \sum_{\mu=2}^{N} \left(\sum_y P_0(y) u^\mu(y)\right) v^\mu(x)$$

(Because $u^1(y) = 1$, hence $\sum_y P_0(y) u^1(y) = 1$, normalization
and $v^1(x) = P(x)$).

Now $\sum_y K^m(x|y) P_0(y) = \sum_{\mu=1}^{N} \alpha^\mu \{\lambda^\mu\}^m v^\mu(x)$

$$= P(x) + \sum_{\mu=2}^{N} \alpha^\mu \{\lambda^\mu\}^m v^\mu(x)$$

($\lambda^1 = 1$, so $\{\lambda^1\}^m = 1$)

decays exponentially fast, because $|\lambda^\mu| < 1$ for $\mu = 2$ to N.

Hence $K^m(y|x) P_0(x) \to P(x)$ exponentially fast independent of the starting condition $P_0(x)$.

So MCMC converges exponentially fast.

Practical Problem

It is nice to know that MCMC convergences exponentially fast with fall off $|\lambda_2|^M$, where λ_2 is the second largest (modulus) eigenvalue of

$$Q(x,y) = p(y)^{1/2} K(x|y)\, p(x)^{-1/2}.$$

Problem: It is impossible to calculate λ_2, except for very simple transition kernels (ones you would never want to use).

Some very clever people have put bounds on the magnitude of $|\lambda_2|$. But these bounds are not tight, which means convergence in practice is a lot faster than the theory says.

So you just have to run MCMC on a computer and see how fast it converges (heuristics like the autocorrelation can help).