Understanding MCMC, Lancaster 2003

Exercises

1. Let $\mathcal{X} = \{1, 2, 3\}$, and consider the Markov chain with transitions $P(1, \{2\}) = P(2, \{3\}) = P(3, \{1\}) = 3/4$, and $P(1, \{3\}) = P(2, \{1\}) = P(3, \{2\}) = 1/4$.
   (a) Prove that the uniform distribution on $\mathcal{X}$ is stationary for this chain.
   (b) Prove that the chain is not reversible with respect to its stationary distribution.

2. Let $\mathcal{X} = \mathbf{R}$, and consider the Markov chain defined as follows. Given X_n, we choose $X_{n+1} \sim N(X_n/2,\ 3/4)$. Let $\pi(\cdot) = N(0, 1)$. Prove that $\pi(\cdot)$ is stationary for this Markov chain, in two ways:
   (a) Given that X_n has standard normal density, compute directly the density of X_{n+1} and show it is the same.
   (b) Use the fact that we can (why?) write $X_{n+1} = X_n/2 + \sqrt{3/4}Z_{n+1}$, where $\{Z_n\}$ are i.i.d. standard normal.

3. For the multiplicative RWM algorithm with proposal
$$Q(x, \cdot) = xe^{N(0,\sigma^2)} ,$$
   show that
$$\alpha(x, y) = \min \left[1, \frac{x\pi(y)}{y\pi(x)}\right] .$$

4. Let $\mathcal{X} = \mathbf{R}$, and let $\pi(\cdot) = N(0, 1)$ be the standard normal distribution. Consider the Random-Walk Metropolis algorithm which uses the proposal kernel $Q(x, \cdot) = \text{Uniform}[x - 1,\ x + 1]$.
   (a) Describe in detail how this algorithm proceeds.
   (b) Prove that the resulting algorithm is ϕ-irreducible.
   (c) Prove that the resulting algorithm is aperiodic.
   (d) What can we conclude from this?

5. Let $\mathcal{X} = [0, 1] \times [0, 1]$, and let $\pi(d\mathbf{x}) = \pi(\mathbf{x})\, d\mathbf{x}$, where $d\mathbf{x}$ is two-dimensional Lebesgue measure, and where $\pi(\mathbf{x}) = 4x_1^2 x_2 + 2x_2^5$. Consider running the Gibbs sampler on this distribution.
   (a) Describe in detail how this algorithm proceeds.
   (b) Prove that the resulting algorithm is ϕ-irreducible.
   (c) Prove that the resulting algorithm is aperiodic.
   (d) Prove that the resulting algorithm is Harris recurrent.
   (d) What can we conclude from all of this?

6. Suppose Markov chain transitions $P(x, \cdot)$ on a state space $\mathcal{X}$ have a density with respect to some reference measure $\nu(\cdot)$: $P(x, dy) = p(x, y)\, \nu(dy)$. Let $C \subseteq \mathcal{X}$. Show that $P(x, \cdot) \geq \epsilon\rho(\cdot)$ for all $x \in C$, for some probability measure $\rho(\cdot)$ on $\mathcal{X}$, where $\epsilon = \int_{y\in\mathcal{X}} \left(\inf_{x\in C} p(x, y)\right)\nu(dy)$.

7. Let $\mathcal{X} = \mathbf{R}$, and consider again the Markov chain such that given X_n, we choose $X_{n+1} \sim N(X_n/2,\ 3/4)$. Recall that $\pi(\cdot) = N(0,1)$ is stationary for this Markov chain. Let $C = [-\sqrt{3}, \sqrt{3}]$, and let $V(x) = 1 + x^2$.
(a) Compute $\mathrm{E}[V(X_{n+1}) \,|\, X_n = x]$ explicitly.
(b) Use this to obtain a drift condition of the form $PV(x) \leq \lambda V(x) + b\mathbf{1}_C(x)$ for some $\lambda < 1$ and $b < \infty$.
(c) Establish a minorisation condition of the form $P(x, \cdot) \geq \epsilon\nu(\cdot)$ for all $x \in C$. [Hint: Use the previous exercise.]
(d) Put this all together, to obtain a quantitative bound on the time to stationarity of this Markov chain.

8. Let $\mathcal{X} = [0, \infty)$, and let $\pi(dx) = e^{-x}dx$ be the standard exponential distribution. Consider the Random-Walk Metropolis algorithm which uses the proposal kernel $Q(x, \cdot) = \mathrm{Uniform}[x - \delta,\ x + \delta]$ for some $\delta > 0$.
(a) Compute the rejection probability $\mathbf{P}[X_{n+1} = X_n \,|\, X_n = x]$ for $x \in \mathcal{X}$.
(b) What value of δ do you think will lead to the most efficient algorithm? Why?

9. Let π denote the discrete uniform density on the following subset of $S = \{0, 1\}^6$. Let

$$\mathcal{X}(1) = \{(a_1, a_2, \ldots, a_6);\ \sum_{i=1}^{5} |a_{i+1} - a_i| = 0 \text{ or } 1\}$$

and let

$$\mathcal{X}(2) = \{(a_1, a_2, \ldots, a_6);\ \sum_{i=1}^{5} |a_{i+1} - a_i| = 4 \text{ or } 5\},$$

so that π is the uniform distribution on $\mathcal{X} = \mathcal{X}(1) \cup \mathcal{X}(2)$. We consider the Gibbs sampling algorithm which updates in turn each of the 6 components. Write down explicitly the elements of $\mathcal{X}$.

By considering how the Gibbs sampler changes $\sum_{i=1}^{5} |a_{i+1} - a_i|$, show that the Gibbs sampler is reducible in this example.

Suppose we decided to try and 'diagnose' convergence by monitoring a_1 from independent runs of the Gibbs sampler started at a collection of different starting points. Would we be able to 'detect' non-convergence? Why?

Methods which empirically monitor Markov chain output until approximate stationarity is observed are called *convergence diagnostics*. What conclusions can you draw about the use of one-dimensional convergence diagnostics from this simple example?

10. Suppose we consider the independence sampler with $q(x, y) = q(y)$ and suppose that

$$\frac{q(y)}{\pi(y)} \geq \beta > 0, \quad \forall y \in \mathcal{X} \tag{1}$$

then show that the transition density of the sampler (for y not equal to x is given by

$$p(x, y) = \left(q(y) \wedge \frac{q(x)\pi(y)}{\pi(x)} \right) \geq \beta\pi(y) \ .$$

Hence show that

$$\|P^n(x, \cdot) - \pi\| \leq 2(1 - \beta)^n \ .$$

11. Consider the following random walk Metropolis sampler on the geometric distribution:

$$\pi(i) = (1 - a)a^i, \quad i = 0, 1, 23, \ldots$$

fo some constant $0 < a < 1$. ¿From state x we propose a move to $x + 1$ or $x - 1$ with equal probability, 1/2.

Verify that for $x \geq 1$, the downward move (ie to $x - 1$) is always accepted, whereas upward moves are accepted with probability a. Now consider the Lyapunov drift function, $V(x) = e^{\beta x}$. Show that for $x \geq 1$,

$$PV(x) = \mathbf{E}\left(V(X_1)|X_0 = x\right) = \frac{1}{2}\left(ae^{\beta(x+1)} + (1 - a)e^{\beta x} + e^{\beta(x-1)}\right) \ .$$

Show that the right hand side can be written as $\lambda V(x)$ where

$$\lambda = 1 - \frac{(1 - e^{\beta})(a - e^{-\beta})}{2} \ .$$

Hence by a suitable choice of β, show that the algorithm is geometrically ergodic.

12. Consider the bivariate normal distribution, π, with unit variances and correlation ρ. If $(X, Y) \sim \pi$ show that the conditional densities are given by

$$(X|Y) \sim N(\rho Y, (1 - \rho^2))$$

and

$$(Y|X) \sim N(\rho X, (1 - \rho^2)) \ .$$

Hence show that if $\{X_n\}$ is the X sequence of a Gibbs sampler under this parameterisation, then

$$X_{n+1} \sim N(\rho^2 X_n, 1 - \rho^4)$$

and that

$$X_n \sim N(\rho^{2n} X_0, 1 - \rho^{4n}) \ .$$