Markov Chain Monte Carlo

Galin L. Jones

School of Statistics

University of Minnesota

Draft: January 28, 2023

Contents

1 Introduction

Later.

2 Markov Chains

Let $(\mathsf{X}, \mathcal{B})$ be a measurable space. A sequence of X-valued random variables $\{X_1, X_2, X_3, \ldots\}$ is a Markov chain if for all g

$$E\left[g(X_{n+1}, X_{n+2}, \ldots) \mid X_n, X_{n-1}, \ldots, X_1\right] = E\left[g(X_{n+1}, X_{n+2}, \ldots) \mid X_n\right].$$

Then P is a *Markov kernel* if $P : \mathsf{X} \times \mathcal{B} \to \mathbb{R}$ satisfying (i) for each fixed $x \in \mathsf{X}$, $P(x, \cdot)$ is a probability measure and (ii) for each fixed $B \in \mathcal{B}$, $P(\cdot, B)$ is a measurable function.

When X is a discrete set a Markov kernel can be represented as a square matrix whose entries are nonnegative and whose rows sum to 1.

Example 2.1. Suppose

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{pmatrix}$$

Then P is a Markov matrix on two states $\{0, 1\}$, say. The first row for example, is interpreted as the probability of moving in one step from state 0 to state 0 is 1/2 which is the same as the probability of moving in one step from state 0 to start 1.

Example 2.2. Let $\mathsf{X} = \mathbb{Z}$ and let $0 < \theta < 1$. If $x \geq 1$, then a Markov kernel is defined by the matrix P with elements

$$P(x, x+1) = P(-x, -x-1) = \theta, \qquad P(x, 0) = P(-x, 0) = 1 - \theta,$$

and $P(0, 1) = P(1, 0) = 1/2$.

Most often X will be uncountable and $\mathcal{B}$ will be countably generated. If X is topological, then $\mathcal{B}$ will be the Borel σ-algebra generated by X.

Example 2.3. Let $\mathsf{X} = (0, 1)$ and consider the Markov chain that evolves as follows. Draw $U \sim$ Uniform$(0, 1)$. If $u \leq 0.5$, $X_{n+1} \sim$ Uniform$(0, X_n)$, but if $u > 0.5$, $X_{n+1} \sim$ Uniform$(X_n, 1)$. Then if $X_n = x$ and $B \in \mathcal{B}$

$$P(x, B) = \int_B \left[\frac{1}{2}\frac{1}{x} I_y((0, x)) + \frac{1}{2}\frac{1}{1-x} I_y(x, 1)\right] \mathrm{d}y.$$

In Example 2.3, the integrand in the Markov kernel is a conditional density on X. This is a setting that will be encountered repeatedly throughout. If there is a conditional density $k(y \mid x)$ such that the Markov kernel satisfies for $B \in \mathcal{B}$

$$P(x, B) = \int_B k(y \mid x) \mathrm{d}y$$

then say k is a *Markov transition density.*

Example 2.4. Suppose $f(x, y)$ is a joint density with support $\mathbb{R}^2$ and conditional densities $f_{X|Y}(x \mid y)$ and $f_{Y|X}(y \mid x)$. Then

$$k(x', y' \mid x, y) = f_{X|Y}(x' \mid y) f_{Y|X}(y' \mid x')$$

is a Markov transition density. The Markov chain evolves from $(X_k = x, Y_k = y)$ to (X_{k+1}, Y_{k+1}) by drawing $X_{k+1} \sim F_{X|Y}(\cdot \mid y)$ followed by $Y_{k+1} \sim F_{Y|X}(\cdot \mid X_{k+1})$. This is a special case of the so-called two-variable Gibbs sampler.

Suppose λ is a positive measure on $(\mathsf{X}, \mathcal{B})$, define

$$\lambda P(B) = \int_{\mathsf{X}} \lambda(\mathrm{d}x) P(x, B). \tag{1}$$

When λ is a probability measure, the encouraged interpretation is that $X_{n+1} \mid X_n \sim P(X_n, \cdot)$ and $X_n \sim \lambda$, the product $\lambda(\mathrm{d}x)P(x, \cdot)$ is the joint distribution of (X_n, X_{n+1}) and λP is the marginal distribution of X_{n+1}.

Since Markov kernels act to the left on measures (1),

$$P^2(x, B) = \int_{\mathsf{X}} P(x, \mathrm{d}x_k) P(x_k, B).$$

Continuing in this fashion obtain for every $n \geq 2$

$$P^n(x, B) \int_{\mathsf{X}} P(x, \mathrm{d}x_k) P(x_k, \mathrm{d}x_{k+1}) \cdots P(x_{k+n-2}, B).$$

More generally, the so-called Chapman-Kolmogorov equations hold for $n \geq m \geq 0$

$$P^n(x, B) \int_{\mathsf{X}} P^m(x, \mathrm{d}y) P^{n-m}(y, B).$$

If $\lambda = \lambda P$, then λ is *invariant* for P. Notice that if λ is invariant for P and $X_n \sim \lambda$, then $X_{n+1} \sim \lambda$. That is, the marginal distribution does not depend upon n in which case the Markov chain is *stationary*.

Example 2.5.

Example 2.6. Recall the Markov chain defined in Example (2.4)

One common way of establishing invariance of MCMC Markov chains is to verify a *detailed balance condition*; see Exercise 3.1. Detailed balance holds if

$$\lambda(\mathrm{d}x)P(x, \mathrm{d}y) = \lambda(\mathrm{d}y)P(y, \mathrm{d}x). \tag{2}$$

When λ is a probability measure, one interpretation is that the joint distribution of (X_k, X_{k+1}) is the same as the distribution of (X_{k+1}, X_k) so that this is also often called the *reversibility condition*. Another name often encountered is that P is λ-*symmetric*.

2.1 Stability

Most MCMC applications are constructed so that a specific probability distribution F is invariant. However, in applications where MCMC is required it is typically difficult to simulate from the invariant distribution. The most that can be hoped for is that the simulation will eventually produce a representative sample from F. This long-run behavior is in no way guaranteed.

Example 2.7. Suppose F lives on $\{1, 2\}$ with $F(1) = 1 - F(2) = 1/4$ and

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Since the Markov chain moves deterministically between the two states, it will over represent state 1 and underrepresent state 2 no matter how many iterations there are.

3 Metropolis-Hastings

Exercises

Exercise 3.1. Prove that if Equation 2 holds, then λ is invariant for P.

Exercise 3.2. Whta is the invariant distribution of the Markov chain defined in Example 2.6?

Appendix

References