Markov Chain Monte Carlo

Galin L. Jones

School of Statistics

University of Minnesota

Draft: January 15, 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. 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 2.2. 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 2.2, 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.*

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

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

Since 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}.

If $\lambda = \lambda P$, then λ is *invariant* for P.

$$\lambda(\mathrm{d}x)P(x, \mathrm{d}y) = \lambda(\mathrm{d}y)P(y, \mathrm{d}x). \tag{1}$$

Integrating both sides of equation 1 shows that it implies λ is invariant for P.

3 Metropolis-Hastings

Exercises

Appendix

References