STAT 238 - Bayesian Statistics
Lecture Twenty Five

Spring 2026, UC Berkeley

Aditya Guntuboyina

01 April 2026

1 Markov Chains and Stationary Distributions

Let S be a state space. A sequence of random variables $\Theta^{(0)}, \Theta^{(1)}, \dots$ taking values in S is a Markov Chain on S provided the conditional distribution of $\Theta^{(t+1)}$ given $\Theta^{(t)} = \theta^{(t)}, \Theta^{(t-1)} = \theta^{(t-1)}, \dots, \Theta^{(0)} = \theta^{(0)}$ only depends on $\theta^{(t)}$. Further, if this conditional distribution is the same for all t, then the Markov chain is said to be time-homogeneous. We will only deal with time-homogeneous Markov Chains in this course.

A (time-homogeneous) Markov Chain is described via the conditional distribution of $\Theta^{(t+1)}$ given $\Theta^{(t)} = x$. If S is finite or countable, then these conditional distributions can be described by the probabilities:

$$P(x, y) := \mathbb{P}\{\Theta^{(t+1)} = y \mid \Theta^{(t)} = x\}.$$

Clearly $P(x, y) \geq 0$ and $\sum_{y \in S} P(x, y) = 1$ for every $x \in S$.

If S is an open subset of $\mathbb{R}^d$, then one usually describes the conditional distribution of $\Theta^{(t+1)}$ given $\Theta^{(t)} = x$ via the conditional density function $P(x, y)$:

$$\mathbb{P}\{\Theta^{(t+1)} \in A \mid \Theta^{(t)} = x\} = \int_A P(x, y) dy.$$

Here $P(x, y) \geq 0$ and $\int_S P(x, y) dy = 1$ for all $x \in S$.

1.1 Stationary Distribution

A probability distribution π on S is said to be the stationary or invariant distribution of the Markov chain $\Theta^{(0)}, \Theta^{(1)}, \dots$ if the following holds:

$$\Theta^{(t)} \sim \pi \implies \Theta^{(t+1)} \sim \pi.$$

If S is discrete, this can be written as:

$$\pi(y) = \sum_{x \in S} \pi(x) P(x, y) \quad \text{for every } y \in S. \tag{1}$$

In the continuous case, this becomes:

$$\pi(y) = \int_S \pi(x) P(x, y) dx \quad \text{for every } y \in S. \tag{2}$$

1.2 Ergodic Theorem

Suppose $\{\Theta^{(t)}\}$ is a time-homogeneous Markov chain that is **irreducible**. Suppose π is the stationary distribution of the Markov Chain (irreducibility guarantess uniqueness of the stationary distribution). Then

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} g(\Theta^{(t)}) = \int g(\theta) d\pi(\theta)$$

almost surely. For a reference to this result, see Roberts and Rosenthal [1, Theorem 4 and the remark after Corollary 6]. For a comprehensive book-length reference, see Meyn and Tweedie [2] (Chapter 17 of this book contains this result).

When the state space is finite, irreducibility means that for every pair of points x and y in the state space, there is a positive probability that the chain goes from x to y. For continuous state spaces, irreducibility is somewhat technical to define. If you are interested, see again the aforementioned references: Roberts and Rosenthal [1], Meyn and Tweedie [2].

1.3 Detailed Balance Condition

Note that (1) is equivalent to:

$$\sum_{x\in S} \pi(y)P(y,x) = \sum_{x\in S} \pi(x)P(x,y).$$

This is because $\sum_{x\in S} P(y,x) = 1$. Thus a **sufficient** condition for π being the stationary distribution for the Markov chain given by P is:

$$\pi(y)P(y,x) = \pi(x)P(x,y) \text{ for all } x, y \in S. \tag{3}$$

Note that one gets the same condition by also arguing from (2) (in this case, $P(x,y)$ and $P(y,x)$ appearing in (3) should be interpreted as conditional pdfs as opposed to pmfs).

Here is a simple example of a Markov Chain satisfying detailed balance.

Example 1.1 (Discrete Markov Chain)**.** *Suppose each $\Theta^{(t)}$ takes values in a finite state space $S = \{1,\ldots,N\}$ of cardinality N. Consider a graph with vertices S and degrees $d_1,\ldots,d_N$. Consider the random walk on the graph. The transition probabilities are then given by:*

$$P(i,j) = \frac{I\{j \text{ is a neighbour of } i\}}{d_i}$$

The detailed balance condition then becomes:

$$\pi(i)\frac{I\{j \text{ is a neighbour of } i\}}{d_i} = \pi(j)\frac{I\{i \text{ is a neighbour of } j\}}{d_j}$$

It is then clear that this Markov chain satisfies detailed balance with $\pi(i) \propto d_i$. So this $\pi(\cdot)$ is stationary. If the graph is connected, then this Markov chain is irreducible and this $\pi(\cdot)$ will be the unique stationary distribution.

Our next example of a Markov Chain with detailed balance is the Metropolis-Hastings chain.

1.4 Metropolis-Hastings

Suppose π is a probability on a state space S. If S is discrete, π should be interpreted as a pmf and if S is continuous, π should be interpreted as a pdf. Let $Q(x,y)$ be some transition probability. Again if S is discrete, interpret $Q(x,y)$ as a pmf over y for each fixed x. If S is continuous, interpret $Q(x,y)$ as a pdf over y for each fixed x.

We do not assume that $\pi(\cdot)$ and $Q(\cdot,\cdot)$ are related in any way. For example, there is no reason for Q to satisfy detailed balance with respect to π. In other words, for x and y, it may very well happen that

$$\pi(x)Q(x,y) \neq \pi(y)Q(y,x).$$

The Metropolis-Hastings uses $Q(\cdot,\cdot)$ and $\pi(\cdot)$ to construct a new Markov chain $\{\Theta^{(t)}\}$ which moves as follows. Given $\Theta^{(t)} = x$, first use $Q(x,\cdot)$ to generate y, then take $\Theta^{(t+1)}$ to be:

$$\Theta^{(t+1)} = \begin{cases} y, & \text{with probability } \min\left(1, \frac{\pi(y)\,Q(y,x)}{\pi(x)\,Q(x,y)}\right), \\ x, & \text{with probability } 1 - \min\left(1, \frac{\pi(y)\,Q(y,x)}{\pi(x)\,Q(x,y)}\right). \end{cases}$$

The key is to realize that this new Markov chain satisfies detailed balance with respect to π. To see this, we need to verify:

$$\pi(x)P(x,y) = \pi(y)P(y,x) \quad \text{for all } x,y \in S. \tag{4}$$

If $y = x$, there is nothing to prove. So let us assume that $y \neq x$. In this case, it is easy to see that

$$P(x,y) = Q(x,y)\min\left(1, \frac{\pi(y)\,Q(y,x)}{\pi(x)\,Q(x,y)}\right).$$

Consequently

$$\pi(x)P(x,y) = \pi(x)Q(x,y)\min\left(1, \frac{\pi(y)\,Q(y,x)}{\pi(x)\,Q(x,y)}\right) = \min(\pi(x)\,Q(x,y), \pi(y)Q(y,x)).$$

Similarly $\pi(y)P(y,x) = \min(\pi(x)\,Q(x,y), \pi(y)Q(y,x))$, which proves (4).

The Metropolis-Hastings Algorithm can be interpreted as a device for converting any Markov Chain into one which satisfies detailed balance with respect to π. It simply adds an acceptance step to the Markov Chain given by $Q(\cdot,\cdot)$ with acceptance probability:

$$\min\left(1, \frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right).$$

A simple special case of the Metropolis-Hastings arises when $Q(\cdot,\cdot)$ is symmetric in the sense that $Q(x,y) = Q(y,x)$. In this case, the acceptance probability is simply $\min(1, \pi(y)/\pi(x))$. For example, when Q is given by a random walk Markov chain: $y = x + z$ where z has a symmetric distribution around 0 such as $z \sim N(0,\Sigma)$ or $z \sim \text{uniform}(\prod_{j=1}^{d}[-a_j, a_j])$. In this special case, the Metropolis-Hastings chain is simply referred to as the Metropolis Chain.

1.5 The Gibbs Sampler

Our third example of a Markov Chain satisfying detailed balance with respect to π is the Gibbs sampler. Suppose the state space S consists of bivariate pairs (x_1, x_2). Suppose $\pi(\cdot)$ is such that its conditionals $\pi_{1|2}(\cdot \mid x_2)$ and $\pi_{2|1}(\cdot \mid x_1)$ are easy to sample from. Here $\pi_{1|2}(\cdot \mid x_2)$ is the conditional distribution of x_1 given x_2 when $(x_1, x_2) \sim \pi$. Similarly $\pi_{2|1}(\cdot \mid x_1)$ is the conditional distribution of x_2 given x_1 when $(x_1, x_2) \sim \pi$.

The Gibbs sampler employs the following Markov Chain. Given $\Theta^{(t)} = (x_1, x_2)$, first sample one of the indices 1 and 2 with equal probability 0.5. If we get 1, then use the update:

$$(x_1, x_2) \mapsto (x_1', x_2) \quad \text{where } x_1' \sim \pi_{1|2}(\cdot \mid x_2).$$

If we get 2, then use the update:

$$(x_1, x_2) \mapsto (x_1, x_2') \quad \text{where } x_2' \sim \pi_{2|1}(\cdot \mid x_1).$$

It turns out that this chain satisfies detailed balance with respect to π. For this, we need to prove that

$$\pi(x_1, x_2)P((x_1, x_2), (x_1', x_2')) = \pi(x_1', x_2')P((x_1', x_2'), (x_1, x_2)).$$

The transition $(x_1, x_2) \to (x_1', x_2')$ will only be possible if one of $x_1 = x_1'$ or $x_2 = x_2'$ hold true. Let $x_1 = x_1'$ so we need to verify:

$$\pi(x_1, x_2)P((x_1, x_2), (x_1, x_2')) = \pi(x_1, x_2')P((x_1, x_2'), (x_1, x_2)). \tag{5}$$

From the description of the Gibbs chain, it is clear that

$$P((x_1, x_2), (x_1, x_2')) = \frac{1}{2}\pi_{2|1}(x_2' \mid x_1) = \frac{1}{2}\frac{\pi(x_1, x_2')}{\pi_1(x_1)}$$

where $\pi_1(\cdot)$ is the marginal distribution of X_1 when $(X_1, X_2) \sim \pi$. As a result

$$\pi(x_1, x_2)P((x_1, x_2), (x_1, x_2')) = \frac{1}{2}\frac{\pi(x_1, x_2)\pi(x_1, x_2')}{\pi_1(x_1)}.$$

Clearly the right hand side above is symmetric in x_2 and x_2' which proves (5). One can similarly verify detailed balance when $x_2 = x_2'$.

References

[1] Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, **1**, 20–71.

[2] Meyn, S. P. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge University Press.