Discrete Space Markov Chains

Matthew Wascher

February 9, 2025

Doubly Stochastic Chains

A transition matrix p is called **doubly stochastic** if $\sum_x p(x, y) = 1$, that is, the **columns** sum to 1.

(Note that since p is by definition a stochastic matrix the rows must also sum to 1.)

Doubly Stochastic Chains

Example: Symmetric reflecting random walk

Consider the state space $\{0, 1, \ldots, L\}$ and let

$$X_t = \max(X_{t-1} - 1, 0) \text{ with probability } 1/2$$
$$X_t = \min(X_{t-1} + 1, L) \text{ with probability } 1/2$$

That is, X_t behaves like a symmetric random walk except that if it ever tries to move to the right at L or the left at 0 it instead remains where it is.

Doubly Stochastic Chains

Example: Symmetric reflecting random walk

Suppose $L = 4$. What is the transition matrix?

$$
\begin{array}{c|ccccc}
 & 0 & 1 & 2 & 3 & 4 \\
\hline
0 & 1/2 & 1/2 & 0 & 0 & 0 \\
1 & 1/2 & 0 & 1/2 & 0 & 0 \\
2 & 0 & 1/2 & 0 & 1/2 & 0 \\
3 & 0 & 0 & 1/2 & 0 & 1/2 \\
4 & 0 & 0 & 0 & 1/2 & 1/2
\end{array}
$$

Doubly Stochastic Chains

Theorem 1.24: If p is a doubly stochastic transition matrix for a Markov chain with N states, then $\pi(x) = 1/N$ for all x is a stationary distribution.

Let $\pi(x) = 1/N$ for all x

Proof: $\sum_x \pi(x) p(x,y) = \frac{1}{N} \sum p(x,y) = \frac{1}{N}(1) = \frac{1}{N}$

Thus $\pi(x) = \frac{1}{N}$ for all is a stationary distribution.

Doubly Stochastic Chains

Example: Tiny Board Game

Suppose you play a board game with spaces labeled $\{0, 1, 2, 3, 4, 5\}$. Each turn you roll a die that has 3 sides with 1, 2 sides with 2,1 side with 3, and move that number of spaces. On the board, 5 is adjacent to 0 so that if you are currently on space i and move j spaces, the result is that you end on space $i + j$ mod 5. What is the transition matrix?

$$\begin{array}{c|cccccc}
 & 0 & 1 & 2 & 3 & 4 & 5 \\
\hline
0 & 0 & 1/2 & 1/3 & 1/6 & 0 & 0 \\
1 & 0 & 0 & 1/2 & 1/3 & 1/6 & 0 \\
2 & 0 & 0 & 0 & 1/2 & 1/3 & 1/6 \\
3 & 1/6 & 0 & 0 & 0 & 1/2 & 1/3 \\
4 & 1/3 & 1/6 & 0 & 0 & 0 & 1/2 \\
5 & 1/2 & 1/3 & 1/6 & 0 & 0 & 0
\end{array}$$

Detailed Balance Condition

π is said to satisfy the detailed balance condition for a transition matrix p if for all x, y

$$\pi(x)p(x,y) = \pi(y)p(y,x)$$

Note that this is stronger than $\pi p = \pi$:

$$\sum_x \pi(x)\, p(x,y) = \sum_x \pi(y)\, p(y,x)$$

$$= \pi(y) \sum_x p(y,x)$$

$$\sum_x \pi(x) p(x,y) = \pi(y)$$

Many chains do not have stationary distributions that satisfy the detailed balance condition.

Detailed Balance Condition

Consider the following transition matrix:

	1	**2**	**3**
1	.5	.5	0
2	.3	.1	.6
3	.2	.4	.4

for any π, $\pi(1)\, p(1,3) = 0$

If detailed balance $\pi(3)\, p(3,1) = \pi(1)\, p(1,3) = 0$

But $p(3,1) > 0 \Rightarrow \pi(3) = 0$

But p is doubly stochastic so $\pi = (1/3, 1/3, 1/3)$ is stationary

Example: Ehrenfest Chain

Consider the Ehrenfest chain with 3 balls so the transition matrix is

	0	**1**	**2**	**3**
0	0	3/3	0	0
1	1/3	0	2/3	0
2	0	2/3	0	1/3
3	0	0	3/3	0

What is π? $\pi(0) = c$ $\quad \pi(1) = 3c$ $\quad \pi(2) = 3c$

$\pi(3) = c$

$c = 1/8$ $\quad \pi(0) = 1/8$ $\quad \pi(1) = 3/8$, $\quad \pi(2) = 3/8$ $\quad \pi(3) = 1/8$

$\pi(x)p(x,y) = \pi(y)p(y,x)$ then π is

stationary

Example: Ehrenfest Chain

Does the Ehrenfest chain satisfy the detailed balance condition?

Yes $\pi(x)p(x,y) = \pi(y)p(y,x)$ for all x,y

Reversibility

Consider a Markov chain with transition matrix $p(i,j)$ and stationary distribution π. What happens if we watch the process $X_m, 0 \leq m \leq n$ in reverse?

Theorem 1.25: Fix n and let $Y_m = X_{n-m}$ for $0 \leq m \leq n$. Then Y_m is a Markov chain with transition probability

$$\hat{p}(i,j) = P(Y_{m+1} = j | Y_m = i) = \frac{\pi(j)p(j,i)}{\pi(i)}$$

Reversibility

Proof of Theorem 1.25:

$$P(Y_{m+1} = i_{m+1} \mid Y_m = i_m, \dots, Y_0 = i_0) = \frac{\pi(i_{m+1})\, p(i_{m+1}, i_m)}{\pi(i_m)}$$

$$= \frac{P(Y_{m+1} = i_{m+1}, Y_m = i_m, \dots, Y_0 = i_0)}{P(Y_m = i_m, \dots, Y_0 = i_0)}$$

$$= \frac{P(X_{n-(m+1)} = i_{m+1}, \dots, X_n = i_0)}{P(X_{n-m} = i_m, \dots, X_n = i_0)} =$$

$$= \frac{\pi(i_{m+1})\, p(i_{m+1}, i_m)\, P(X_{n-m+1} = i_{m-1}, \dots X_n = i_0 \mid X_{n-m} = i_m)}{\pi(i_m)\, P(X_{n-m+1} = i_{m-1}, \dots X_n = i_0 \mid X_{n-m} = i_m)}$$

$$= \frac{\pi(i_{m+1})\, p(i_{m+1}, i_m)}{\pi(i_m)}$$

Reversibility

We can verify that $\hat{p}(i,j)$ defines a valid transition matrix:

$$\hat{p}(i,j) = \frac{\pi(j)\,p(j,i)}{\pi(i)}$$

$$\sum_j \frac{\pi(j)\,p(j,i)}{\pi(i)} = \frac{1}{\pi(i)} \sum_j \pi(j)\,p(j,i) = \frac{1}{\pi(i)}\left(\pi(i)\right) = 1$$

What happens if π satisfies the detailed balance condition?

Detailed balance says that $\pi(j)\,p(j,i) = \pi(i)\,p(i,j)$ for all i,j

$$\hat{p}(i,j) = \frac{\pi(j)\,p(j,i)}{\pi(i)} = \frac{\cancel{\pi(i)}\,p(i,j)}{\cancel{\pi(i)}} = p(i,j)$$

Reversibility

When π satisfies the detailed balance condiiton, we see that Y_m has the same distribution as X_n.

A Markov chain with this property (that the distribution is the same when the chain is reversed) is called **reversible.**

The detailed balance condition is a sufficient, but not necessary condition for a Markov chain to be reversible.

(Kolmogorov's Criterion)

Metropolis-Hasting Algorithm

The **Metropolis-Hasting algorithm** is an algorithm for sampling from a distribution $\pi(x)$.

The Metropolis-Hasting algorithm is important in Bayesian statistics as it is commonly used to sample from a posterior distribution in MCMC (Markov chain Monte Carlo) methods.

Metropolis-Hasting Algorithm

Begin with a Markov chain with transition matrix $q(x, y)$ that we call the proposal distribution. We wish to generate a sequence $X_1, \ldots, X_N$ in the following way. Given $X_n = x$, we propose the next move y according to $q(x, y)$. Next we define

$$r(x, y) = \min\left(\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right)$$

$r(x, y)$ is called the acceptance ratio. We flip a coin with probability of head $r(x, y)$

- If heads, set $X_{n+1} = y$ (accept the new move)
- If tails, set $X_{n+1} = x$ (reject the new move)

Metropolis-Hasting Algorithm

Generating the sequence $X_1, \ldots, X_N$ in this way results in a Markov chain with transition probabilities

$$p(x,y) = q(x,y)r(x,y)$$

We can verify that the distribution π satisfies the detailed balance condition for $p(x,y)$: Suppose WLOG that $\pi(y)q(y,x) > \pi(x)q(x,y)$

$$\pi(x)p(x,y) = \pi(x)q(x,y) \cdot 1$$

$$\pi(y)p(y,x) = \cancel{\pi(y)q(y,x)}\,\frac{\pi(x)q(x,y)}{\cancel{\pi(y)q(y,x)}} = \pi(x)q(x,y)$$

So $\pi(x)$ satisfies detailed balance

Metropolis-Hasting Algorithm

To generate a samples from $\pi(x)$

1. Run the chain defined by $p(x, y)$ for a long time until it reaches equilibrium to obtain a single sample
2. Repeat 1. by taking samples at widely separated times

The question of how long is a sufficiently "long time" is not always easy to answer.

We can also use Theorem 1.23 to compute expected values of functions of $X \sim \pi(x)$, since this theorem tells us

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) \to \sum_{x} f(x)\pi(x)$$

Metropolis-Hasting Algorithm

Consider the geometric distribution with success probability θ, so

$\pi(x) = 0$ if $x \notin \mathbb{Z}^+$

$$\pi(x) = \theta(1-\theta)^{x-1}, x = 1, 2, \ldots$$

$q(x,y) = 1/2$ if $|x-y| = 1$ $\qquad X_0 = 1$

0 else

$$r(x,y) = \frac{\theta(1-\theta)^{y-1}\,(1/2)}{\theta(1-\theta)^{x-1}\,(1/2)} \quad x,y \in \mathbb{Z}^+,\ |x-y| = 1$$

$= 0$ else

Connection to Bayesian Statistics

Why is the Metropolis-Hastings Algorithm useful in Bayesian statistics?

Consider the problem of inference for a parameter θ after observing data

$$Y_1, \ldots, Y_n \sim \text{ iid } f(y|\theta)$$

Bayesian statistics begins with a prior distribution on θ, $h(\theta)$, and expresses belief about the parameter θ after observing the data by the posterior

$$g(\theta | Y_1, \ldots, Y_n)$$

Connection to Bayesian Statistics

Let $f(\mathbf{Y}|\theta) = \prod_{i=1}^{n} f(Y_i|\theta)$. Using Bayes Theorem, the posterior can be expressed as

$$P(B|A) = \frac{P(A|B)\,P(B)}{\sum_i P(A|B_i)\,P(B_i)}$$

$$g(\theta|Y_1, \ldots, Y_n) = \frac{f(\mathbf{Y}|\theta)h(\theta)}{\sum_\theta f(\mathbf{Y}|\theta)h(\theta)}$$

However, in many cases the denominator $\sum_\theta f(\mathbf{Y}|\theta)h(\theta)$ is difficult to compute. In particular, if θ is a high dimensional quantity or if θ takes on continuous values in which case we replace the sum by an integral and

$$\int_\theta f(\mathbf{Y}|\theta)h(\theta)$$

may not have a closed form solution.

Connection to Bayesian Statistics

To see why Metropolis-Hastings is useful here, set

$$\pi(\theta) = \frac{f(\mathbf{Y}|\theta)h(\theta)}{\sum_\theta f(\mathbf{Y}|\theta)h(\theta)}$$

Now for any $q(\theta, \theta')$ we might choose we can observe

$$r(\theta, \theta') = \min\left(\frac{\pi(\overset{\theta'}{\cancel{y}})q(\overset{\theta',}{\cancel{y}}, \overset{\theta}{\cancel{x}})}{\pi(\underset{\theta}{\cancel{x}})q(\underset{\theta,}{\cancel{x}}, \underset{\theta'}{\cancel{y}})}, 1\right) = \min\left(\frac{f(\mathbf{Y}|\theta')h(\theta')q(\theta', \theta)}{f(\mathbf{Y}|\theta)h(\theta)q(\theta, \theta')}, 1\right)$$

and in particular we can compute this quantity and run the algorithm without needing to compute $\cancel{\sum_\theta f(\mathbf{Y}|\theta)h(\theta)}$.

$\cancel{\sum_{\theta'} f(Y|\theta')h(\theta')}$

Connection to Bayesian Statistics

Example: Suppose k-sided die is rolled 10 times. k is unknown, but a priori we believe $P(k=4) = P(k=6) = P(k=8) = 1/3$. Of the 10 rolls, 4 are 1s. What is the posterior distribution of k?

$$P(k \mid X) = \frac{P(X \mid k)\, P(k)}{\sum_{k \in \{4,6,8\}} P(X \mid k)\, P(k)}$$

$$= \frac{\binom{10}{4} (.25)^4 (.75)^6 (1/2)}{\sum \ldots}$$

$$= \frac{\binom{10}{4} (1/6)^4 (5/6)^6 (1/2)}{\sum \ldots}$$

$$= \frac{\binom{10}{4} (1/8)^4 (7/8)^6\, 1/2}{\sum \ldots}$$

Connection to Bayesian Statistics

How can we use the Metropolis-Hastings algorithm to sample from the posterior of k?

$$q(x,y) = \begin{cases} 1/3 & y = 4, 6, 8 \\ 0 & \text{else} \end{cases}$$

$$r(x,y) = \frac{f(X|y)\,h(y)\,q(y,x)}{f(X|x)\,h(x)\,q(x,y)} = \frac{\binom{10}{4}(1/y)^4(1-1/y)^6(1/3)(1/3)}{\binom{10}{4}(1/x)^4(1-1/x)^6(1/3)(1/3)}$$

$$\frac{(1/y)^4(1-1/y)^6}{(1/x)^4(1-1/x)^6}$$