Towards the Metropolis-Hastings Algorithm

1 A Brief Primer on Markov Chains

1.1 Definition, Notation and Transition Probabilities

A sequence of random variables $X_0, X_1, X_2, \ldots$, is a **Markov chain** if

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i). \tag{1}$$

Equation (1) is called the **Markov property**.

Imagine a processes transitioning through time. If it's a Markov chain, where you go next depends on where you **are** (and some random model to transition forward), but it does not depend on where you **were**.

This is not to say that X_{n+1} (for example), is independent of X_0. Clearly each value depends on the last so there is a building dependence. However, X_{n+1} is **conditionally independent** of X_0 (for example) once X_n is known. Actually, it is not the previous value that is important but rather the last known value. Equation (1) gives us, for example, that

$$P(X_5 = 4 | X_2 = 1, X_1 = 3, X_0 = 4) = P(X_5 = 4 | X_2 = 1).$$

To show this, one would write

$$P(X_5 = 4 | X_2 = 1, X_1 = 3, X_0 = 4) = \sum_j \sum_k P(X_5 = 4, X_4 = j, X_3 = k | X_2 = 1, X_1 = 3, X_0 = 4)$$

and proceed by rewritting the summand on the right-hand side as products of conditional probabilities. The sums are taken over the **state space** for the Markov chain $\{X_n\}_{n \geq 0}$.

The notation used so far here suggests that both time and space are discrete. Either or both of these could be continuous and we may switch as needed without going through all of the basics for these different cases.

Suppose that $\{X_n\}_{n \geq 0}$ is a Markov chain. If $P(X_{n+1} = j | X_n = i)$ is independent of n, we say that the Markov chain is **time homogeneous**.

For this course, all chains will be time homogeneous by default.

We will denote this **one-step transition probability** by p_{ij}. That is,

$$p_{ij} = P(X_{n+1} = j | X_n = i).$$

By our assumed time-homogenity, we also have that

$$p_{ij} = P(X_1 = j | X_0 = i).$$

We will denote/define the **n-step transition probability** by

$$p_{ij}^{(n)} = P(X_{n+m} = j | X_m = i) = P(X_n = j | X_0 = i).$$

Note that we must have

$$\sum_j p_{ij} = 1 \qquad \text{and} \qquad \sum_j p_{ij}^{(n)} = 1.$$

1.2 A Chapman-Kolmogorov Equation

The following **Chapman-Kolmogorov equation** will be useful.

For any $0 \leq m \leq n$,

$$p_{ij}^{(n)} = \sum_k p_{ik}^{(m)} \cdot p_{kj}^{(n-m)}.$$

Here, the sum is taken over the entire state space of the Markov chain and the zero-step transition probability is defined as

$$p_{ij}^{(0)} = \begin{cases} 1 & , \quad j = i \\ 0 & , \quad j \neq i. \end{cases}$$

(That makes perfect sense right? You are not going anywhere in zero time steps!)

Let's prove the C-K equation.

$$\begin{aligned} p_{ij}^{(n)} &= P(X_n = j | X_0 = i) = \textstyle\sum_k P(X_n = j, X_m = k, X_0 = i) \\ &= \textstyle\sum_k P(X_n = j | X_m = k, X_0 = i) \cdot P(X_m = k | X_0 = i) \\ &\stackrel{M.P.}{=} \textstyle\sum_k P(X_n = j | X_m = k) \cdot P(X_m = k | X_0 = i) \\ &= \textstyle\sum_k p_{kj}^{(n-m)} \cdot p_{ik}^{(m)} \quad \checkmark \end{aligned}$$

(The "$M.P$" denotes use of the Markov property.)

1.3 The Stationary Distribution of a Markov Chain

Let $\{X_n\}_{n\geq 0}$ be a Markov chain living on a state space $\mathcal{S}$ with transition probabilities p_{ij}.

Consider a probability mass function f over $\mathcal{S}$. That is, consider a function f with $f(i) \geq 0$ for all $i \in \mathcal{S}$ such that $\sum_{i\in\mathcal{S}} f(i) = 1$.

Markov "people" tend to use the symbol "π" for distributions involving Markov chains. So, let π be a function with $\pi(i) \geq 0$ for all $i \in \mathcal{S}$ such that $\sum_{i\in\mathcal{S}} \pi(i) = 1$.

We will use the notation π_i for $\pi(i)$.

Definition: The probability distribution π is said to be a **stationary distribution** for the chain if

$$X_0 \sim \pi \qquad \Rightarrow \qquad X_1 \sim \pi \qquad \Rightarrow \qquad X_2 \sim \pi \qquad \Rightarrow \qquad \cdots$$

That is, if you start this chain according to a draw from π and iterate forward using the transition probabilities p_{ij}, it will maintain this distribution at all fixed time points.

Note that

$$P(X_1 = j) = \sum_{i \in S} P(X_1 = j, X_0 = i) = \sum_{i \in S} P(X_2 = i | X_0 = j) \cdot P(X_0 = i).$$

So, the stationary distribution must satisfy the **stattionary equation**

$$\boxed{\pi_j = \sum_i \pi_i p_{ij}.}$$

Since

$$X_0 \sim \pi \qquad \Rightarrow \qquad X_1 \sim \pi \qquad \Rightarrow \qquad X_2 \sim \pi,$$

we have by transitivity that

$$X_0 \sim \pi \qquad \Rightarrow \qquad X_2 \sim \pi.$$

This means that a stationary distribution π must also satisfy

$$\pi_j = \sum_i \pi_i p_{ij}^{(2)}.$$

Similarly, we can get

$$\boxed{\pi_j = \sum_i \pi_i p_{ij}^{(n)}.}$$

for any fixed $n \geq 1$.

For a given Markov chain, such a stationary distribution may or may not exist and, if it exists, it may or may not be unique. Since this is not a course on Markov chains, I'd rather not go into conditions for existence and uniqueness. We are not going to be studying a given Markov chain but instead will be constructing our own. We will always do so in a way where there is a unique stationary distribution.

1.4 A Limiting Distribution

Given a Markov chain with n-step transition probabilities $p_{ij}^{(n)}$, suppose that the following limit exists and is independent of i.

$$\lim_{n \to \infty} p_{ij}^{(n)}.$$

In this case, we will give the limit a name.

$$\gamma_j := \lim_{n \to \infty} p_{ij}^{(n)}.$$

Then $\{\gamma_j : j \in \mathcal{S}\}$ is a probability distribution on $\mathcal{S}$ since the values are non-negative and

$$\sum_{j \in \mathcal{S}} \gamma_j = \sum_{j \in \mathcal{S}} \lim_{n \to \infty} p_{ij}^{(n)} = \lim_{n \to \infty} \sum_{j \in \mathcal{S}} p_{ij}^{(n)} = \lim_{n \to \infty} 1 = 1.$$

This distribution is also stationary with respect to the transition probabilies p_{ij}! To see this, note that

$$\sum_j \gamma_j p_{ji} = \sum_j \left(\lim_{n \to \infty} p_{kj}^{(n)} \right) p_{ji} = \lim_{n \to \infty} \sum_j p_{kj}^{(n)} \cdot p_{ji} \stackrel{C.K}{=} \lim_{n \to \infty} p_{ki}^{(n+1)} = \gamma_i$$

(Here, the "$C.K.$" denotes use of the Chapman-Kolmogorov equation.)

This is the stationary equation, boxed on the previous page, with γ in place of π and different indices.

1.5 Who Cares?

Suppose we are interested in finding the stationary distribution π for a given Markov chain. If we can solve the stationary equation

$$\pi_j = \sum_i \pi_i p_{ij}$$

subject to the constraint that $\sum \pi_i = 1$, we absolutely should!

If we can't, we should take advantage of the fact that (in "nice Markov world") the limiting distribution exists and is stationary and (in "nice Markov world") the stationary distribution is unique. That is, we should go for the limiting (= stationary) distribution via simulation of the process "for a long time".

Defining "a long time" is difficult to do in a general setting (and even in most specific settings!). This will depend on the process of interest and our method of simulation.