UC Berkeley
Department of Electrical Engineering and Computer Sciences

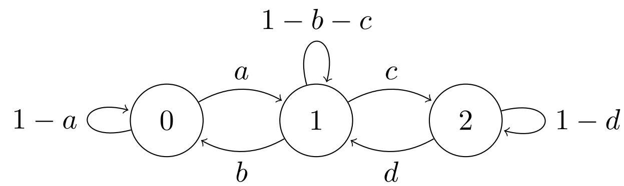
ELECTRICAL ENGINEERING 126: PROBABILITY AND RANDOM PROCESSES

Discussion 8
Fall 2017

---

1. **Markov Chain Big Theorem**

For this problem we will consider the following three-state chain and illustrate the ideas behind the Markov chain convergence theorem. Here, $a, b, c, d \in (0, 1)$.



(a) Let $T_0 = \min\{n \in \mathbb{Z}_+ : X_n = 0\}$ be the first passage time to state 0. Let $\mu_y := \mathbb{E}_0[\sum_{n=0}^{T_0-1} \mathbb{1}\{X_n = y\}]$ for $y = 0, 1, 2$ be the mean number of visits to state y, starting at 0 and ending right before we return to 0. Explain why $\mu = \mu P$.

(b) Therefore, if we define π to be μ after we normalize it so that the entries sum to 1, π is a stationary distribution. Why is π unique?

(c) Now deduce that $\pi_0 = 1/\mathbb{E}_0[T_0]$. In words, $\mathbb{E}_0[T_0]$ is the mean return time from state 0 to itself.

(d) Explain why the fraction of times $\sum_{m=1}^{n} \mathbb{1}\{X_m = 0\}$, where n is a positive integer, converges a.s. to π_0 as $n \to \infty$. (Hint: Define $T_0^{(1)} := T_0$ and for integers $k \geq 2$, define

$$T_0^{(k)} = \min\{n > T_0^{(k-1)} : X_n = 0\} - T_0^{(k-1)}$$

to be the additional time it takes to return to 0 for the kth time. Then $T_0^{(1)}, T_0^{(2)}, T_0^{(3)}, \ldots$ are i.i.d. and one can apply the SLLN.)

(e) Consider two copies of the above chain $(X_n, Y_n)_{n\in\mathbb{N}}$, where the chains move independently of each other, Y_0 is picked from the stationary distribution, and X_0 is started from any fixed state x. Explain why the two chains will meet after a finite time, and think about why this implies that the chain started from state x converges in distribution to the stationary distribution π.

Solution:

(a) Here, $\mu_0 = 1$ (since $X_0 = 0$) and $(\mu P)_0 = (1-a)\mu_0 + b\mu_1 = 1 - a + b\mu_1$. The expected number of visits to state 1, μ_1, is computed as follows. With probability a, $X_1 = a$. Conditioned on $X_1 = a$, the mean number of visits to state 1 before returning to state 0 is $1/b$, since every time we are at state 1 we have a probability b of transitioning to state 0, and so the number of times we stay at state 1 is geometric with parameter b. Plugging in, $(\mu P)_0 = 1 - a + b \cdot a(1/b) = 1$.

Now consider μ_y for $y = 1, 2$. μ_y is the mean number of visits to state y in the period $0, \dots, T_0 - 1$. Meanwhile, $(\mu P)_y = \sum_{x=0,1,2} \mu_x P_{x,y}$, and since μ_x is the mean number of visits to x in times $0, \dots, T_0 - 1$ and $P_{x,y}$ is the probability of transitioning to y, then $\mu_x P_{x,y}$ is the mean number of visits to y in times $1, \dots, T_0$. The insight here is that since we start at state 0 at time 0, and we end at state 0 at time T_0, the times $0, \dots, T_0 - 1$ and $1, \dots, T_0$ look the same, so the mean number of visits to y is the same for each period.

Thus, $\mu = \mu P$.

(b) Uniqueness is harder to justify, but in fact Part (d) below implies that $n^{-1} \sum_{m=1}^{n} \mathbb{1}\{X_m = y\} \to 1/\mathbb{E}_y[T_y]$ for all states y as $n \to \infty$, so by tkaing expectations of both sides, we obtain $n^{-1} \sum_{m=1}^{n} \mathbb{P}(X_m = y) \to 1/\mathbb{E}_y[T_y]$. In particular, if we start the chain from the stationary distribution, then $\mathbb{P}(X_m = y) = \pi(y)$ so $\pi(y) = 1/\mathbb{E}_y[T_y]$, in particular, π is unique.

(c) Note that $\mu_0 + \mu_1 + \mu_2 = \mathbb{E}_0[T_0]$ and $\pi_0 = \mu_0/(\mu_0 + \mu_1 + \mu_2) = 1/\mathbb{E}_0[T_0]$.

(d) Observe that $\sum_{m=1}^{T_0^{(1)} + \cdots + T_0^{(k)}} \mathbb{1}\{X_m = 0\} = k$. Thus,

$$\frac{1}{T_0^{(1)} + \cdots + T_0^{(k)}} \sum_{m=1}^{T_0^{(1)} + \cdots + T_0^{(k)}} \mathbb{1}\{X_m = 0\} = \frac{k}{T_0^{(1)} + \cdots + T_0^{(k)}} \to \frac{1}{\mathbb{E}_0[T_0]}$$

a.s., as $k \to \infty$, by the SLLN. Also, since $T_0^{(1)} + \cdots + T_0^{(k)} \to \infty$ as $k \to \infty$, then we also have $n^{-1} \sum_{m=1}^{n} \mathbb{1}\{X_m = 0\} \to 1/\mathbb{E}_0[T_0]$ a.s., as $n \to \infty$. Finally, we use $1/\mathbb{E}_0[T_0] = \pi_0$ from the arguments in the previous parts.

(e) The original chain is aperiodic, which is the condition that we need in order for the product chain $(X_n, Y_n)_{n \in \mathbb{N}}$ to be irreducible (you can convince yourself that if the original chain is periodic, then the product chain is not irreducible). Then, the vector $\tilde{\pi}(x, y) := \pi(x)\pi(y)$ is stationary for the product chain, because the two chains are independent. In particular, $\tilde{\pi}(x, x) = \pi(x)\pi(x) > 0$, so $\mathbb{E}_{(x,x)}[T_{(x,x)}] = 1/\tilde{\pi}(x, x) < \infty$ for any state x, which means that the two chains will meet each other at the state x in finite time.

What is the big deal? In fact $\mathbb{P}(X_n \neq Y_n) \le \mathbb{P}(T > n)$ for any positive integer n. This is because at time T we can glue the chains together and force them to transition together for the rest of time, so then the event $\{X_n \neq Y_n\}$ exactly becomes the event $\{T > n\}$, i.e., at time n the two chains have not met yet. Now since we have argued that T is finite, $\mathbb{P}(T > n) \to 0$ as $n \to \infty$, so $\mathbb{P}(X_n \neq Y_n) \to 0$ as $n \to \infty$. However, recall that $(X_n)_{n \in \mathbb{N}}$ is the chain started at x and $(Y_n)_{n \in \mathbb{N}}$ is the stationary chain, so we have argued that the chain started at x is approaching stationarity!

2. **Random Walk on the Cube**

Consider the symmetric random walk on the vertices of the 3-dimensional unit cube where two vertices are connected by an edge if and only if the line connecting them is an edge of the cube. In other words, this is the random walk on the graph with 8 nodes each written as a string of 3 bits, so that the vertex set is $\{0,1\}^3$, and where two vertices are connected by an edge if and only if their corresponding bit strings differ in exactly one location.

This random walk is modified so that the nodes 000 and 111 are made absorbing.

(a) What are the communicating classes of the resulting Markov chain? For each class, determine its period, and whether it is transient or recurrent.

(b) For each transient state, what is the probability that the modified random walk started at that state gets absorbed in the state 000?

Solution:

(a) The communicating classes are $\{000\}$ with period 1; $\{111\}$ with period 1; $\{001, 010, 011, 100, 101, 110\}$ with period 2. $\{000\}$ and $\{111\}$ are recurrent, while the other communicating class is transient.

(b) The probability of absorption is the same for states 001, 010 and 100, so denote this probability by p. Similarly, the probability of absorption is the same for 011, 101, and 110, so denote this probability by q. Then,

$$\begin{aligned} p &= \frac{1}{3} + \frac{2}{3} \cdot q, \\ q &= \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot p, \end{aligned}$$

and so $p = 3/5$, $q = 2/5$.

3. **Hidden Markov Models**

A hidden Markov model (HMM) is a Markov chain $\{X_n\}_{n=0}^{\infty}$ in which the states are considered "hidden" or "latent". In other words, we do not directly observe $\{X_n\}_{n=0}^{\infty}$. Instead, we observe $\{Y_n\}_{n=0}^{\infty}$, where $Q(x, y)$ is the probability that state x will emit observation y. π_0 is the initial distribution for the Markov chain, and P is the transition matrix.

(a) What is $\mathbb{P}(X_0 = x_0, Y_0 = y_0, \ldots, X_n = x_n, Y_n = y_n)$, where n is a positive integer, $x_0, \ldots, x_n$ are hidden states, and $y_0, \ldots, y_n$ are observations?

(b) What is $\mathbb{P}(X_0 = x_0 \mid Y_0 = y_0)$?

(c) We observe $(y_0, \ldots, y_n)$ and we would like to find the most likely sequence of hidden states $(x_0, \ldots, x_n)$ which gave rise to the observations. Let

$$U(x_m, m) = \max_{x_{m+1}, \ldots, x_n \in \mathcal{X}} \mathbb{P}(X_m = x_m, X_{m+1:n} = x_{m+1:n}, Y_{0:n} = y_{0:n})$$

denote the largest probability for a sequence of hidden states beginning at state x_m at time $m \in \mathbb{N}$, along with the observations $(y_0, \ldots, y_n)$. Develop a recursion for $U(x_m, m)$ in terms of $U(x_{m+1}, m+1)$, $x_{m+1} \in \mathcal{X}$.

Solution:

(a) The probability is

$$\pi_0(x_0)Q(x_0, y_0)\prod_{i=1}^{n} P(x_{i-1}, x_i)Q(x_i, y_i).$$

(b) This is a simple application of Bayes rule.

$$\mathbb{P}(X_0 = x_0 \mid Y_0 = y_0) = \frac{\mathbb{P}(X_0 = x_0, Y_0 = y_0)}{\mathbb{P}(Y_0 = y_0)} = \frac{\pi_0(x_0)Q(x_0, y_0)}{\sum_{x\in\mathcal{X}} \pi_0(x)Q(x, y_0)}.$$

(c) The probability of transitioning to x_{m+1} is $P(x_m, x_{m+1})$. The probability of emission is $Q(x_{m+1}, y_{m+1})$. Once we are in state x_{m+1}, the most likely sequence of hidden states for the observations $(y_0, \ldots, y_n)$, beginning at x_{m+1} at time $m + 1$, is $U(x_{m+1}, m + 1)$. Hence,

$$U(x_m, m) = \max_{x_{m+1}\in\mathcal{X}} P(x_m, x_{m+1})Q(x_{m+1}, y_{m+1})U(x_{m+1}, m + 1). \quad (1)$$

To avoid numerical issues, we often work with the logarithms of the above quantities instead.

Note also that the recursion should be solved backwards for efficiency. If we simply try to solve for $U(x_0, m)$, we would have to evaluate all $|\mathcal{X}|^{n+1}$ possible paths, which is computationally prohibitive. Instead, if we solve the equations backwards using (1), then each step requires taking the maximum over $|\mathcal{X}|$ possibilities, so the algorithm will terminate in at most $O(n\mathcal{X})$ steps.