MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018
Lecture 21

MARKOV CHAINS

Contents

1 INTRODUCTION

Recall a model we considered earlier: random walk. We have $X_n \stackrel{d}{=} \mathrm{Ber}(p)$, i.i.d. Then $S_n = \sum_{1 \le j \le n} X_j$ was defined to be a simple random walk. One of its key property is that the distribution of S_{n+1} conditioned on the state $S_n = x$ at n is independent from the past history, namely $S_m, m \le n-1$. To see this formally note that

$$\begin{aligned}
&\mathbb{P}(S_{n+1} = y | S_n = x, S_{n-1} = z_1, \ldots, S_1 = z_{n-1}) \\
&\qquad = \frac{\mathbb{P}(X_{n+1} = y - x, S_n = x, S_{n-1} = z_1, \ldots, S_1 = z_{n-1})}{\mathbb{P}(S_n = x, S_{n-1} = z_1, \ldots, S_1 = z_{n-1})} \\
&\qquad = \frac{\mathbb{P}(X_{n+1} = y - x)\mathbb{P}(S_n = x, S_{n-1} = z_1, \ldots, S_1 = z_{n-1})}{\mathbb{P}(S_n = x, S_{n-1} = z_1, \ldots, S_1 = z_{n-1})} \\
&\qquad = \mathbb{P}(X_{n+1} = y - x),
\end{aligned}$$

where the second equality follows from the independence assumption for the sequence $X_n, n \ge 1$. A similar derivation gives $\mathbb{P}(S_{n+1} = y | S_n = x) = \mathbb{P}(X_{n+1} = y - x)$ and we get the required equality: $\mathbb{P}(S_{n+1} = y | S_n = x, S_{n-1} = z_1, \ldots, S_1 = z_{n-1}) = \mathbb{P}(S_{n+1} = y | S_n = x)$.

Definition 1. A discrete time stochastic process $(X_n, n \geq 1)$ is defined to be a Markov chain if it takes values in some countable set $\mathcal{X}$, and for every $x_1, x_2, \ldots, x_n \in \mathcal{X}$ it satisfies the property

$$\begin{aligned}\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1) \\ = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1})\end{aligned}$$

The elements of $\mathcal{X}$ are called *states*. We say that the Markov chain is in state $s \in \mathcal{X}$ at time n if $X_n = s$. Mostly for now we will consider the case when $\mathcal{X}$ is finite. In this case we call $(X_n, n \geq 1)$ a *finite state Markov chain* and, without the loss of generality, we will assume that $\mathcal{X} = \{1, 2, \ldots, N\}$.

Let us establish some properties of Markov chains.

Proposition 1. Given a Markov chain $X_n, n \geq 1$.

1. For every collection of states $s, x_1, x_2, \ldots, x_{n-1}$ and every m

$$\begin{aligned}\mathbb{P}(X_{n+m} = s | X_{n-1} = x_{n-1}, \ldots, X_1 = x_1) \\ = \mathbb{P}(X_{n+m} = s | X_{n-1} = x_{n-1}).\end{aligned}$$

2. For every collection of states $x_1, x_2, \ldots, x_n$ and $k = 1, 2 \ldots, n$

$$\begin{aligned}&\mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1 | X_k = x_k) \\ &= \mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_{k+1} = x_{k+1} | X_k = x_k) \times \\ &\times \mathbb{P}(X_{k-1} = x_{k-1}, \ldots X_1 = x_1 | X_k = x_k).\end{aligned}$$

Proof. Exercise. □

2 EXAMPLES

We already have an example of a Markov chain - random walk.

Consider now the following example (Exercise 2, Section 6.1 [1]). Suppose we roll a die repeatedly and X_n is the number of 6-s we have seen so far. Then X_n is a Markov chain and $\mathbb{P}(X_n = x + 1 | X_{n-1} = x) = 1/6, \mathbb{P}(X_n = x | X_{n-1} = x) = 5/6$ and $\mathbb{P}(X_n = y | X_{n-1} = x) = 0$ for all $y \neq x, x + 1$. Note, that we can think of X_n as a random walk, where the transition to the right occurs with probability $1/6$ and the transition to the same state with the probability $5/6$.

Now let X_n be the largest of the six possible outcomes observed up to time n. Then X_n is again a Markov chain. What are its transition probabilities?

For our next example consider the following model of an inventory process. The inventory can hold finish goods up to capacity $C \in \mathbb{N}$. Every month n there is some current inventory level I_n and a certain fixed amount of product $x \in \mathbb{N}$ is produced, as long as limit is not reached, namely $I_n + x \le C$. If $I_n + x > C$, than just enough $C - I_n$ is produced to reach the capacity. Every month there is a random demand $D_n, n \ge 1$, which we assume is i.i.d. If the current inventory level is at least as large as the demand, then the full demand is satisfied. Otherwise as much of the demand is satisfied as possible, bringing the inventory level down to zero.

Let I_n be the inventory level in month n. Then I_n is a Markov chain. Note

$$I_{n+1} = \min((I_n - D_n)^+ + x, C).$$

Specifically, the probability distribution of I_{n+1} given $I_n = i$, is independent from the values $I_m, m \le n-1$. I_n is a Markov chain taking values in $0, 1, \ldots, C$.

3 HOMOGENEOUS FINITE STATE MARKOV CHAINS

We say that the Markov chain X_n is homogeneous if $\mathbb{P}(X_{n+1} = y | X_n = x) = \mathbb{P}(X_2 = y | X_1 = x)$ for all n. Observe that all of our examples are homogeneous Markov chains. For a homogenous Markov chain X_n we can specify transition probabilities $\mathbb{P}(X_{n+1} = y | X_n = x)$ by a sequence of values $p_{x,y} = \mathbb{P}(X_{n+1} = y | X_n = x)$. For the case of finite state Markov chain, say the state space is $\{1, 2, \ldots, N\}$. Then the transition probabilities are $p_{i,j}, 1 \le i, j \le N$. We call $P = (p_{i,j})$ the transition matrix of X_n. The transition matrix P has the following obvious property $\sum_j p_{i,j} = 1$ for all i. Any non-negative matrix with such property is called stochastic matrix, for obvious reason.

Observe that

$$\begin{aligned}
\mathbb{P}(X_{n+2} = j | X_n = i) \\
&= \sum_{1 \le k \le N} \mathbb{P}(X_{n+2} = j | X_{n+1} = k, X_n = i) \mathbb{P}(X_{n+1} = k | X_n = i) \\
&= \sum_{1 \le k \le N} \mathbb{P}(X_{n+2} = j | X_{n+1} = k) \mathbb{P}(X_{n+1} = k | X_n = i) \\
&= \sum_{1 \le k \le N} p_{k,j} p_{i,k}.
\end{aligned}$$

This means that the matrix P^2 gives the two-step transition probabilities of the underlying Markov chain. Namely, the (i, j)-th entry of P^2, which we denote by $p_{i,j}^{(2)}$ is precisely $\mathbb{P}(X_{n+2} = j | X_n = i)$. This observation is not hard to extend to the general case: for every $r \ge 1$, P^r is the transition matrix of r-steps of the Markov chain. One of our goals is understanding the long-term dynamics of P^r as $r \to \infty$. We will see that for a broad class of Markov chains the following property holds: the limit $\lim_{r \to \infty} p_{i,j}^{(r)}$ exists and depends on j only. Namely, the starting state i is irrelevant, as far as the limit is concerned. This property is called *mixing* and is a very important property of Markov chains.

Let e_j denote the j-th N-dimensional column vector. Namely e_j has j-th coordinate equal to one, and all the other coordinates equal to zero. We also let e denote the N-dimensional column vector consisting of ones. Suppose $X_0 = i$, for some state $i \in \{1, \ldots, N\}$. Then the probability vector of X_n can be written as $e_i^T P^n$ in vector form. Suppose at time zero, the state of the chain is random and is given by some probability vector μ. Namely $\mathbb{P}(X_0 = i) = \mu_i, i = 1, 2, \ldots, N$. Then the probability vector of X_n is precisely $\mu^T P^n$ in vector form.

4 STATIONARY DISTRIBUTION

Consider the following simple Markov chain on states $1, 2$: $p_{1,1} = p_{1,2} = 1/2, p_{2,1} = 1, p_{2,2} = 0$. Suppose we start at random at time zero with the following probability distribution μ: $\mu_1 = \mathbb{P}(X_0 = 1) = 2/3, \mu_2 = \mathbb{P}(X_0 = 2) = 1/3$. What is the probability distribution of X_1? We have $\mathbb{P}(X_1 = 1) = (1/2)\mathbb{P}(X_0 = 1) + \mathbb{P}(X_0 = 2) = (1/2)(2/3) + (1/3) = 2/3$. From this we find $\mathbb{P}(X_1 = 2) = 1 - \mathbb{P}(X_1 = 1) = 1/3$. We see that the probability distribution of X_0 and X_1 are identical. The same applies to every n.

Definition 2. A probability vector $\pi = (\pi_i), 1 \le i \le N$ is defined to be a **stationary distribution** if $\mathbb{P}(X_n = i) = \pi_i$ for all times $n \ge 1$ and states $i = 1, \ldots, N$, conditioned on $\mathbb{P}(X_0 = i) = \pi_i, 1 \le i \le N$. In this case we also say that the Markov chain X_n is in **steady-state**.

Repeating the derivation above for the case of general Markov chains, it is not hard to see that the vector π is stationary iff it satisfies the following properties: $\pi_i \ge 0, \sum_i \pi_i = 1$ and

$$\pi_i = \sum_{1 \le k \le N} p_{k,i} \pi_k, \ \ \forall i.$$

In vector form this can be written as

$$\pi^T = \pi^T P, \tag{1}$$

where w^T denotes the (row) transpose of a column vector w.

One of the fundamental properties of finite state Markov chains is that a stationary distribution always exists.

Theorem 1. Given a finite state Markov chain with transition matrix P, there exists at least one stationary distribution π. Namely the system of equation (1) has at least one solution satisfying $\pi \ge 0, \sum_i \pi_i = 1$.

Proof. There are many proofs of this fundamental results. One possibility is to use Brower's Fixed Point Theorem. Later on we will give another probabilistic proof which provides important intuition about the meaning of π_i. For now let us give a quick proof, but one that relies on linear programming (LP). If you are not familiar with linear programming theory, you can simply skip the proof.

Consider the following LP problem in variables $\pi_1, \ldots, \pi_N$.

$$\max \sum_{1 \le i \le N} \pi_i$$

Subject to:

$$P^T \pi - \pi = 0,$$

$$\pi \ge 0.$$

Note that a stationary vector π exists iff this LP has an unbounded optimal solution. Indeed, if π is a stationary vector, then it clearly is a feasible solution to this LP. Note that $\alpha\pi$ is also a solution for every $\alpha > 0$. Since $\alpha \sum_{1\le i\le N} \pi_i = \alpha$, then we can obtain a feasible solution as large as we want. On the other hand, suppose this LP has an unbounded objective value. In particular, there exists a solution x satisfying $\sum_i x_i > 0$. Taking $\pi_i = x_i / \sum_i x_i$ we obtain a stationary distribution.

Now using LP duality theory, this LP has an unbounded solution iff the dual solution is infeasible. The dual solution is

$$\min \sum_{1\le i\le N} 0y_i$$

$$\text{Subject to:}$$

$$Py - y \ge e.$$

Let us show that indeed this dual LP problem is infeasible. Suppose the contrary is true. Namely, there exists y satisfying $Py - y \ge e$. Take any such y and find $k^*, 1 \le k^* \le N$ such that $y_{k^*} = \max_i y_i$. Observe that $\sum_i p_{k^*,i} y_i \le \sum_i p_{k^*,i} y_{k^*} = y_{k^*} < 1 + y_{k^*}$, since the rows of P sum to one. Thus the constraint $Py - y \ge e$ is violated in the k^*-th row. We conclude that the dual problem is indeed infeasible. Thus the primal LP problem is unbounded and the stationary distribution exists. □

As we mentioned above, stationary distribution π is not necessarily unique, though in many special cases it is. The uniqueness can be verified by checking whether the following system has a unique solution. $\pi^T = \pi^T P, \sum_j \pi_j = 1, \pi_j \ge 0$.

Example.[6.6 from [2]] An absent-minded professor has two umbrellas, used when commuting from home to work and back. If it rains and umbrella is available, the professor takes it. If umbrella is not available, the professor gets wet. If it does not rain the professor does not take the umbrella. It rains on a given commute with probability p, independently for all days. What is the steady-state probability that the professor will get wet on a given day?

We model the process as a Markov chain with states $j = 0, 1, 2$. The state j means the location where the professor is currently in has j umbrellas. Then the corresponding transition probabilities are $p_{0,2} = 1, p_{2,1} = p, p_{1,2} = p, p_{1,1} = 1 - p, p_{2,0} = 1 - p$. The corresponding equations for $\pi_j, j = 0, 1, 2$ are then $\pi_0 = \pi_2(1-p), \pi_1 = (1-p)\pi_1 + p\pi_2, \pi_2 = \pi_0 + p\pi_1$. From the second equation $\pi_1 = \pi_2$. Combining with the first equation and with the fact $\pi_0 + \pi_1 + \pi_2 = 1$, we obtain $\pi_1 = \pi_2 = \frac{1}{3-p}, \pi_0 = \frac{1-p}{3-p}$. The steady-state probability that the professor gets wet is the probability of being in state zero times probability that it rains on this day. Namely it is $\mathbb{P}(\text{wet}) = \frac{(1-p)p}{3-p}$.

5 CLASSIFICATION OF STATES. RECURRENT AND TRANSIENT STATES

Given a finite state homogeneous Markov chain with transition matrix P, construct a directed graph as follows: the nodes are $i = 1, 2, \ldots, N$. Put edges (i, j) for every pair of states such that $p_{i,j} > 0$. Given two states i, j suppose there is a directed path from i to j. We say that i communicates with j and write $i \to j$. By allowing paths of lengths zero, we obtain that i communicates with itself ($i \to i$), although it is possible that starting from $X_0 = i$, after time $n = 0$ the chain never returns to i. What is the probabilistic interpretation of this? It means there is a positive probability of getting to state j starting from i. Formally $\sum_n p_{i,j}^{(n)} > 0$. Suppose, there is a path from i to j, but not from j to i. This means that if the chain starting from i, got to j, then it will never return to i again. Since, there is a positive chance of going from i to j, intuitively, this will happen with probability one. Thus with probability one we will never return to i. We would like to formalize this intuition.

Definition 3. A state i is called transient if there exists a state j such that $i \to j$, but $j \nrightarrow i$. Otherwise i is called recurrent.

We write $i \leftrightarrow j$ if states i and j communicate with each other. Observe that if $i \leftrightarrow j$ then $j \leftrightarrow i$. Also, if $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$. Finally, observe that if i is recurrent then it must be the case that $i \leftrightarrow i$. Indeed, consider any state j (possibly i itself) such that $p_{i,j} > 0$. If there is a path from j to i, then there is a path from i to i as well and the assertion is established. Otherwise, we find that $i \to j$, but $j \not\to i$, and therefore i is not recurrent. We conclude that $\leftrightarrow$ is an equivalency relationship on the set of recurrent states, and we can partition all the recurrent states into equivalency classes $R_1, R_2, \ldots, R_r$. The entire states space $\{1, 2, \ldots, N\}$ then can be partitioned as $T \cup R_1 \cup \cdots \cup R_r$, where T is the (possibly empty) set of transient states.

References

[1] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, Oxford University Press, 3rd edition, 2001.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to probability,* Athena Scientific, 2002.