Vincent's CS70 Discussion 13B Notes

April 22, 2016

1 More on Markov chains

Before deriving a few interesting properties of Markov chains, let's state a few definitions that will come in handy for this worksheet.

Definition 1. *A Markov chain is irreducible if it is possible to get from any state to any other state.*

If we view our Markov chain as a directed graph (that is, we simply ignore the probabilities associated with each arrow in the diagram of our Markov chain and ignore those arrows with transition probability zero), then we can see that a Markov chain is irreducible precisely when its corresponding digraph is strongly connected.

We remark that the connections between Markov chains and graphs do not end here. In fact, an alternative, completely equivalent definition of a Markov chain would be a directed graph $D = (V, E)$ along with functions $\pi_0 : V \to [0, 1]$, $P : E \to [0, 1]$, and sequence of random variables $\{X_n\}_{n=0}^{\infty}$. The precise details are left as an exercise to the reader (Vincent's note: Ha! I've always wanted to say that).

Definition 2. *The period of a state $i \in \mathscr{X}$ of a Markov chain is defined to be*

$$gcd\{n > 0 : \mathbb{P}(X_n = i | X_0 = i) > 0\}.$$

*State i is **aperiodic** if the number above is defined and equals 1, and a Markov chain is aperiodic if all of its states are aperiodic.*

Intuitively the state i has period $k > 1$ if, given that we start at state i at time 0, it is only possible to return to i at times that are multiples of k.

2 Linear algebra connections!

2.1 The distribution of X_n

Given a Markov chain $(\mathscr{X}, \pi_0, P, \{X_n\}_{n=0}^{\infty})$, suppose that we want to compute the probability of being in state $j \in \mathscr{X}$ at time n. Note that we don't care about the path we take to get there, so long as we reach our target state at our target time. By the law of total probability and the definition of the matrix P,

$$\begin{aligned}\mathbb{P}(X_n = j) &= \sum_{i \in \mathscr{X}} \mathbb{P}(X_n = j, X_{n-1} = i) \\ &= \sum_{i \in \mathscr{X}} \mathbb{P}(X_n = j | X_{n-1} = i)\mathbb{P}(X_{n-1} = i) \\ &= \sum_{i \in \mathscr{X}} \mathbb{P}(X_{n-1} = i)\mathbb{P}(X_n = j | X_{n-1} = i) \\ &= \sum_{i \in \mathscr{X}} \mathbb{P}(X_{n-1} = i)P(i, j)\end{aligned}$$

If we encode the values of $\mathbb{P}(X_{n-1} = i)$ as a row vector π_{n-1}, where $\pi_{n-1}(i) = \mathbb{P}(X_{n-1} = i)$ for $i \in \mathscr{X}$ then we have that the last equation above is precisely the definition of matrix multiplication for computing the jth entry (we end up with a row-vector so only need to specify one coordinate) of the matrix product $\pi_{n-1}P$! In other words, $\pi_n = \pi_{n-1}P$.

We have thus found an iterative method for computing $\mathbb{P}(X_n = i)$ for any $n \in \mathbb{N},\ i \in \mathscr{X}$:

$$\pi_n = \pi_{n-1}P = (\pi_{n-2}P)P = \pi_{n-2}P^2 = \ldots = \pi_0 P^n,$$

so $P(X_n = i)$ is the ith entry of $\pi_0 P^n$. This result should be at least mildly exciting for those of you who have taken or are taking Math 54 — we've reduced the problem of calculating probabilities to purely mechanical matrix computations!

2.2 Eigenv... I mean... invariant distributions

Definition 3. *We say that a distribution π (represented as a row-vector) is invariant under the matrix P if*

$$\pi P = \pi$$

It is necessary to remark that we can phrase this in terms of linear algebra: an invariant distribution π is precisely an eigenvector of P corresponding to eigenvalue 1. A nice, short exercise for those of you that know a bit of linear algebra is as follows: prove that if a distribution π is an eigenvector of a right stochastic matrix P, then it must correspond to eigenvalue 1 (Hint: $\pi P = \lambda \pi$ must also be a distribution).

Let's now return from our excursions into other branches of math and state a theorem. The proof of the following result was covered in lecture, but it's a good exercise to try to prove it yourself.

Theorem 1. *Let π_n denote the distribution of X_n in row-vector form. We have that $\pi_n = \pi_0$ for all $n \in \mathbb{N}$ if and only if π_0 is invariant under P.*

3 Conditional Expectation

Let's now step back to a topic that was covered in lecture some time ago, but (if I remember correctly) we have yet to talk about in section.

Definition 4. *Given two random variables X and Y, the conditional expectation of Y given $X = x$ is defined to be*

$$E[Y|X = x] = \sum_y y\mathbb{P}(Y = y|X = x).$$

Thus, we see that $E[Y|X]$ is a function of X, and it itself a random variable. We now derive a result that's quite similar to the law of total probability.

Theorem 2. *For two random variables X, Y, $E[Y] = E[E[Y|X]]$*

Proof. Noting that $E[Y|X]$ is itself a random variable, it's valid to take its expectation. Thus, we have that

$$\begin{aligned} E[E[Y|X]] &= \sum_x E[Y|X = x]\mathbb{P}(X = x) \\ &= \sum_x \sum_y y\mathbb{P}(Y = y|X = x)\mathbb{P}(X = x) \\ &= \sum_x \sum_y y\mathbb{P}(Y = y, X = x) \\ &= \sum_y y \sum_x \mathbb{P}(Y = y, X = x) \\ &= \sum_y y\mathbb{P}(Y = y) \\ &= E[Y] \end{aligned}$$

□