IEOR 6711: Stochastic Models I

Fall 2013, Professor Whitt

Class Lecture Notes: Tuesday, October 29.

Markov Chains

The Contraction approach to $\pi = \pi P$

The limit for aperiodic irreducible finite-state DTMC's.

There is a nice simple limit for aperiodic irreducible finite-state Markov chains. For any initial probability vector $u \equiv (u_1, \ldots, u_m)$, the probability vector at time n is

$$P(X_n = j) = (uP^n)_j = \sum_{i=1}^{m} u_i P_{i,j}^n \ .$$

The key limiting result is

Theorem 0.1 *If P is the transition matrix of an aperiodic irreducible finite-state Markov chain with transition matrix P, then, for any initial probability vector u,*

$$uP^n \to \pi \quad as \quad n \to \infty \ ,$$

where the limiting probability vector π is the unique stationary probability vector, i.e., the unique solution to the fixed-point equation

$$\pi = \pi P \quad or \quad \pi_j = \sum_{i=1}^{m} \pi_i P_{i,j} \quad for\ all \quad j \ ,$$

where $\pi_j \geq 0$ for all j and $\sum_j \pi_j = 1$.

Note the conditions: Of course, irreducibility is essential. And aperiodicity is essential to get full convergence, as opposed to convergence of averages, or convergence through appropriate subsequences. The method of proof here is designed to apply to finite-state chains. The proof extends to infinite-state chains under the condition that there is some state j such that $P_{i,j} \geq \epsilon > 0$ for all states i, or $P_{i,j}^k \geq \epsilon > 0$ for some k. This is a strong extra condition saying that there is a state j such that there is a probability of at least $\epsilon > 0$ of going to j in one step (or in k steps, as a weaker version of the same condition), from any other state. With that extra condition, we not only get convergence, we get convergence quickly, geometrically fast. We actually provide a proof without this condition, but we do not get such quick convergence unless the condition holds.

The Contraction Proof.

One way to prove this result and others is to apply renewal theory. That is done in the Ross textbook. An alternative way to prove the theorem is to consider the transition matrix

P as an operator on the space of all probability vectors, here taken to be of dimension m, corresponding to there being m states. An operator on a space maps the space into itself. If u is a probability vector, then P maps u into the probability vector uP, corresponding to the probability vector starting with u and then taking one step according to P, i.e.,

$$(uP)_j = \sum_{i=1}^{m} u_i P_{i,j} \quad \text{for all} \quad j \ .$$

We want the underlying space to be a complete metric space and the operator to be a contraction map. Then we can apply the *Banach fixed-point theorem*, also called the *Banach-Picard fixed-point theorem* or the *contraction fixed-point theorem*; see pages 220-221 from the blue Rudin book, *Principles of Mathematical Analysis*, posted on line.

The proof can be done in two steps:

step 1.

In the first step, you prove that some power of P has all positive entries (using the assumption that P is an $m \times m$ transition matrix of an irreducible aperiodic Markov chain).

Example 0.1 We remark at the outset that the worst case has $P_{1,2} = P_{2,3} = \cdots = P_{m-1,m} = 1$, while $P_{m,1} > 0$, $P_{m,2} > 0$ and $P_{m,j} = 0$ for all $j \geq 3$. Note that, for this example, $P_{1,1}^k > 0$ for $k = m$, $k = 2m$, $k = 2m-1$, $k = 3m$, $k = 3m-1$, $k = 3m-2$, $k = 4m$, $k = 4m-1$, $k = 4m-2$, $k = 4m-3$, and so forth. In this example, we have $P_{1,1}^k > 0$ for $k = (m-2)m$, but $P_{1,1}^k = 0$ for $k = (m-2)m+1$, but then we have $P_{1,1}^k > 0$ for all $k \geq (m-1)m - (m-2) = m^2 - 2m + 2$.

Lemma 0.1 *For any states i and $j \neq i$, $P_{i,j}^k > 0$ for some k, with $1 \leq k \leq m-1$.*

Proof. We use the fact that the chain is irreducible. Let $S_{i,k}$ be the set of states reachable from state i in at most k steps, with $S_{i,0} \equiv \{i\}$. By the irreducibility, the sets $S_{i,k}$ have to be strictly increasing in k for every k until $S_{i,k} = \{1, 2, \ldots, m\}$, the full state space. Otherwise, the DTMC would not be irreducible. Since there are only $m-1$ other states, all these other states have to be reached in at most $m-1$ steps. If the increase is by more than a single state, then the number k will be strictly less than $m-1$. Hence, indeed, for any states i and $j \neq i$, $P_{i,j}^k > 0$ for some k, with $1 \leq k \leq m-1$. But, in general, the value k depends on the states i and j. ∎

As a corollary to the last conclusion, we deduce the following:

Corollary 0.1 *For any state i, there necessarily is a $k \leq m$ such that $P_{i,i}^k > 0$.*

Proof. To start, there must be some j such that $P_{i,j} > 0$. Then, by the reasoning above, $P_{j,i}^l > 0$ for some $l \leq m-1$. But then $P_{i,i}^{l+1} > P_{i,j} P_{j,i}^l > 0$. Finally, since $l \leq m-1$, necessarily $l + 1 \leq m$. ∎

Now we apply the aperiodicity to obtain a further result.

Lemma 0.2 *There exists a constant n_0 such that $P_{i,i}^k > 0$ for all $k \geq n_0$.*

Proof. We have shown above that $P_{i,i}^k > 0$ for some k with $1 \leq k \leq m$. Since the chain is aperiodic, there necessarily exist constants k_1 and k_2 such that the greatest common divisor of k_1 and k_2, $gcd(k_1, k_2)$, is 1, and $P_{i,i}^{k_1} > 0$ and $P_{i,i}^{k_2} > 0$. But then $P_{i,i}^k > 0$ for all $k \geq k_1 k_2$. This last step follows from the Euclidean algorithm for elementary diophantine linear equations; see Bezout's identity; i.e., we seek integers a and b such that

$$ak_1 + bk_2 = k_1k_2 + j$$

for each $j \geq 1$. It thus suffices to find integers a and b such that

$$ak_1 + bk_2 = j,$$

which is possible for all j because $gcd(k_1, k_2) = 1$. ∎

We can do better in this last step, but the reasoning is somewhat complicated. We have the following sharper result from Section 2.4 of Seneta, *Non-negative Matrices and Markov Chains*, second edition, Springer, 1981, in particular from Theorem 2.9 on page 58.

Theorem 0.2 *For all* $k \geq m^2 - 2m + 2$, $P_{i,j}^k > 0$ *for all* i *and* j.

In closing this part, we observe that the bound above on the number of steps required to get a completely positive power of P depends on the number m of states. Indeed, it essentially grows with the square of m. Note that $m^2 - 2m + 2 \leq m^2$ for all $m \geq 1$,

step 2.

In the second step, you assume that at least one column of P has all positive elements. (By step 1, that will necessarily occur after at most m^2 steps.)

We then want to show that P, regarded as an operator, is a *contraction map*, assuming that at least one column of P has all positive elements. That implies that there exists a unique fixed point and that there is convergence to that fixed point at a geometric rate: For any initial probability vector $u \equiv (u_1, \ldots, u_m)$,

$$d(uP^k, \pi) \leq c^k d(u, \pi) \quad \text{for all} \quad k \ ,$$

for the metric d. That yields a geometric rate of convergence. In Markov-chain theory we speak of *geometric ergodicity*. There is a literature on this topic, primarily focusing on DTMC's with infinitely many states.

Finding the appropriate norm on $\mathbb{R}^m$.

The space we will be looking at is a subset of $\mathbb{R}^m$, containing all probability measures and their differences. We will use a distance defined via a standard norm. That means that the distance is

$$d(x, y) \equiv ||x - y|| \ ,$$

where $|| \cdot ||$ is the norm. We will consider a standard norm on $\mathbb{R}^m$; the question is which one. Some norms do not work, but one does.

Consider the transition matrix:

$$P = \begin{pmatrix} 0.95 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.95 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.95 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.95 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.95 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.95 & 0.01 & 0.01 & 0.01 & 0.01 \end{pmatrix}$$

and the two probability vectors:

$$u = (1/3, 1/3, 1/3, 0, 0, 0)$$

and

$$v = (0, 0, 0, 1/3, 1/3, 1/3) \ .$$

Then consider the probability vectors uP and vP. We want to have

$$||uP - vP|| \leq c||u - v|| \ ,$$

where $0 \leq c < 1$, for some norm $|| \cdot ||$; then the desired distance is $d(u, v) \equiv ||u - v||$.

Look at the two vectors we get when we apply P to u and v:

$$uP = (0.95, 0.01, 0.01, 0.01, 0.01, 0.01)$$

and

$$vP = (0.01, 0.95, 0.01, 0.01, 0.01, 0.01) \ .$$

There are three natural norms to consider on $\mathbb{R}^m$: the l_∞, l_2 and l_1 norms:

$$\begin{aligned} ||x||_\infty &\equiv \max\{|x_i|\} \ , \\ ||x||_2 &\equiv \sqrt{\left(\sum_{i=1}^{m} |x_i|^2\right)} \ , \\ ||x||_1 &\equiv \sum_{i=1}^{m} |x_i| \ . \end{aligned} \tag{1}$$

It is not difficult to see that

$$\begin{aligned} ||u - v||_\infty &= 1/3 \ , \\ ||u - v||_2 &= \sqrt{6/9} = \sqrt{2/3} < 1 \ , \\ ||u - v||_1 &= 2 \end{aligned} \tag{2}$$

and It is not difficult to see that

$$\begin{aligned} ||uP - vP||_\infty &= 0.94 \ , \\ ||uP - vP||_2 &= \sqrt{2 \times (0.94)^2} = 1.329 \ , \\ ||uP - vP||_1 &= 2 \times 0.94 = 1.88 \ . \end{aligned} \tag{3}$$

By this example, we prove that the norms $||x||_\infty$ and $||x||_2$ do *not* work. However, it turns out that the norm $||x||_1$ does work.

Theorem 0.3 *Let P be a $m \times m$ Markov-chain transition matrix associated with an irreducible Markov chain. Assume that $P_{i,1} \geq \epsilon > 0$ for all i, $1 \leq i \leq m$. Then*

$$||uP - vP||_1 \leq (1 - \epsilon)||u - v||_1 \ .$$

Proof. Note that

$$||uP - vP||_1 = \sum_{j=1}^{m} |\sum_{i=1}^{m} u_i P_{i,j} - \sum_{i=1}^{m} v_i P_{i,j}| \ .$$

Now write

$$P_{i,1} \equiv \epsilon + Q_{i,1} \quad \text{and} \quad P_{i,j} \equiv Q_{i,j} \quad \text{for} \quad j \neq i \ , \qquad \text{for all} \quad i \ .$$

Then Q is a nonnegative $m \times m$ matrix with row sums $1 - \epsilon$. Now observe that

$$\begin{aligned} \sum_{j=1}^{m} |\sum_{i=1}^{m} u_i P_{i,j} - \sum_{i=1}^{m} v_i P_{i,j}| &= |\sum_{i=1}^{m} u_i(\epsilon + Q_{i,1}) - \sum_{i=1}^{m} v_i(\epsilon + Q_{i,1})| + \sum_{j=2}^{m} |\sum_{i=1}^{m} u_i Q_{i,j} - \sum_{i=1}^{m} v_i Q_{i,j}| \\ &= |\sum_{i=1}^{m} u_i Q_{i,1} - \sum_{i=1}^{m} v_i Q_{i,1}| + \sum_{j=2}^{m} |\sum_{i=1}^{m} u_i Q_{i,j} - \sum_{i=1}^{m} v_i Q_{i,j}| \\ &= \sum_{j=1}^{m} |\sum_{i=1}^{m} u_i Q_{i,j} - \sum_{i=1}^{m} v_i Q_{i,j}| \\ &\leq \sum_{j=1}^{m} \sum_{i=1}^{m} |u_i - v_i| Q_{i,j} = \sum_{i=1}^{m} \sum_{j=1}^{m} |u_i - v_i| Q_{i,j} = (1 - \epsilon) ||u - v||_1 \ . \end{aligned} \quad (4)$$

with the equality in the second line holding because the epsilon terms can be dropped out, using

$$\sum_{i=1}^{m} (u_i - v_i) = 0 \ ,$$

because u and v are probability vectors, summing to 1. ∎

We remark that we would obtain a stronger contraction property if we applied the above reasoning to all m columns. We would strengthen the assumption in the theorem to: Assume that, for each column j, $P_{i,j} \geq \epsilon_j \geq 0$ for all i, $1 \leq i \leq m$, with $\epsilon_j > 0$ for at least one j. We would then use the reasoning above to obtain the stronger conclusion:

$$||uP - vP||_1 \leq c||u - v||_1 \ ,$$

where

$$c = (1 - \sum_{j=1}^{m} \epsilon_j) < 1.$$

It is not difficult to see that we must have $\sum_{j=1}^{m} \epsilon_j < 1$.