6.262: Discrete Stochastic Processes 2/23/11

Lecture 7: Finite-state Markov Chains; the matrix approach

Outline:

- The transition matrix and its powers
- Convergence of $[P^n] > 0$
- Ergodic Markov chains
- Ergodic unichains
- Other finite-state Markov chains

Recall that the state X_n of a Markov chain at step n depends on the past only through the previous step, i.e.,

$$\Pr\{X_n = j | X_{n-1} = i, X_{n-2}, \ldots, X_0\} = P_{ij}$$

This implies that the joint probability of $X_0, X_1, \ldots, X_n$ can be expressed as a function of $\mathsf{p}_{X_0}(x_0)$ and of the transition probabilities, $\{P_{ij};\ 1 \leq i, j \leq \mathsf{M}\}$.

The transition probabilities are conveniently represented in terms of a transition matrix,

$$[P] = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{16} \\ P_{21} & P_{22} & \cdots & P_{26} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ P_{61} & P_{62} & \cdots & P_{66} \end{bmatrix}$$

If we condition only on the state at time 0, and define $P_{ij}^n = \Pr\{X_n = j \mid X_0 = i\}$, then, starting with $n = 2$, we have

$$\begin{aligned} P_{ij}^2 &= \sum_k \Pr\{X_2 = j | X_1 = k, X_0 = i\} \Pr\{X_1 = k | X_0 = i\} \\ &= \sum_k P_{ik} P_{kj} \end{aligned}$$

Note that $\sum_k P_{ik} P_{kj}$ is the i, j term of the product of the transition matrix $[P]$ with itself, which is $[P^2]$.

Thus the 2-step transition probabilities $\{P_{ij}^2;\ 1 \le i, j \le$ M$\}$ are simply the elements of $[P^2]$.

Iterating to find P_{ij}^n for successively larger n,

$$\begin{aligned} P_{ij}^n &= \sum_k \mathsf{Pr}\{X_n{=}j|X_{n-1}{=}k\}\,\mathsf{Pr}\{X_{n-1}{=}k|X_0=i\} \\ &= \sum_k P_{ik}^{n-1} P_{kj} \end{aligned}$$

Thus P_{ij}^n is the i,j element of $[P^n]$, i.e., the matrix $[P]$ to the n^{th} power.

Computational hack: To find high powers of $[P]$, calculate $[P^2],[P^4],[P^8]$, etc. and then multiply the required powers of 2.

Chapman-Kolmogorov eqns: Since $[P^{m+n}] = [P^m][P^n]$,

$$P_{ij}^{m+n} = \sum_k P_{ik}^m\, P_{kj}^n$$

Convergence of $[P^n] > 0$

An important question for Markov chains, and one that effects almost everything else, is whether the effect of the initial state dies out with time, i.e., whether $\lim_{n\to\infty} P_{ij}^n = \pi_j$ for all i and j, where π_j is a function only of j and not of i or n.

If this limit exists, we can multiply both sides by P_{jk} and sum over j, getting

$$\lim_{n\to\infty} \sum_j P_{ij}^n P_{jk} = \sum_j \pi_j P_{jk}$$

The left side is $\lim_{n\to\infty} P_{ik}^{n+1} = \pi_k$. Thus if this limit exists, the vector $\vec{\pi}$ must satisfy $\pi_k = \sum_j \pi_j P_{jk}$ for each k.

In matrix terms, does $\lim_{n\to\infty}[P^n]$ exist, and is each row is the same vector, $\vec{\pi}$? If so, then $\vec{\pi}$ must satisfy the matrix equation $\vec{\pi} = \vec{\pi}[P]$.

Def: A probability vector is a vector $\vec{\pi} = (\pi_1, \dots, \pi_{\mathsf{M}})$ for which each π_i is nonnegative and $\sum_i \pi_i = 1$. A probability vector $\vec{\pi}$ is called a steady-state vector for the transition matrix $[P]$ if $\vec{\pi} = \vec{\pi}[P]$.

One would now think that we have reduced the question of whether $\lim_{n\to\infty}[P^n]$ exists to the study of the steady-state equation $\vec{\pi} = \vec{\pi}[P]$.

Surprisingly, studying $\lim_{n\to\infty}[P^n]$ is relatively simple, whereas understanding the set of solutions to $\vec{\pi} = \vec{\pi}[P]$ is more complicated. We will find that $\vec{\pi} = \vec{\pi}[P]$ always has one (and often more) probability vector solutions, but this does't imply that $\lim_{n\to\infty}[P^n]$ exists.

Ergodic Markov chains

Another way to express that $\lim_{n\to\infty}[P^n]$ **converges to a matrix of equal rows** $\vec{\pi}$ **is the statement that, for each column** j**,** $\lim_{n\to\infty} P_{ij}^n = \pi_j$ **for each** i**.**

The following theorem demonstrates this type of convergence, and some stronger results, for ergodic Markov chains.

Thm: Let an ergodic finite-state Markov chain have transition matrix $[P]$**. Then for each** j**,** $\max_i P_{ij}^n$ **is nonincreasing in** n**,** $\min_i P_{ij}^n$ **is nondecreasing in** n**, and**

$$\lim_{n\to\infty} \max_i P_{ij}^n \;=\; \lim_{n\to\infty} \min_i P_{ij}^n \;\doteq\; \pi_j > 0$$

with exponential convergence in n

The key to this theorem is the pair of statements that $\max_i P_{ij}^n$ **is nonincreasing in** n **and** $\min_i P_{ij}^n$ **is nondecreasing in** n**.**

It turns out, with an almost trivial proof, that these statements are true for all Markov chains, so we first establish this as a lemma.

Lemma 1: Let $[P]$ **be the transition matrix of an arbitrary finite-state Markov chain. Then for each** j**,** $\max_i P_{ij}^n$ **is nonincreasing in** n **and** $\min_i P_{ij}^n$ **is nondecreasing in** n**.**

Example 1: Consider the 2-state chain with $P_{12} = P_{21} = 1$**. Then** P_{12}^n **alternates between 1 and 0 for increasing** n **and** P_{22}^n **alternates between 0 and 1. The maximum of** P_{12}^n **and** P_{22}^n **is 1, which is nonincreasing, and the minimum is 0.**

<u>Lemma 1</u>: Let $[P]$ be the transition matrix of an arbitrary finite-state Markov chain. Then for each j, $\max_i P_{ij}^n$ is nonincreasing in n and $\min_i P_{ij}^n$ is non-decreasing in n.

<u>Example 2</u>: Consider the 2-state ergodic chain with $P_{12} = P_{21} = 3/4$. Then $P_{12}^n = \frac{3}{4}, \frac{3}{8}, \frac{9}{16}, \ldots$ for increasing n and $P_{22}^n = \frac{1}{4}, \frac{5}{8}, \frac{7}{16}, \ldots$.

Each sequence oscillates while approaching 1/2, but $\max(P_{12}^n, P_{22}^n) = \frac{3}{4}, \frac{5}{8}, \frac{9}{16}, \ldots$ which is decreasing toward 1/2. Similarly the minimum approaches 1/2 from below, $\min(P_{12}^n, P_{22}^n) = \frac{1}{4}, \frac{3}{8}, \frac{7}{16}, \ldots$

Lemma 1: **Let** $[P]$ **be the transition matrix of an arbitrary finite-state Markov chain. Then for each** j, $\max_i P_{ij}^n$ **is nonincreasing in** n **and** $\min_i P_{ij}^n$ **is nondecreasing in** n.

Proof: **For any states** i, j **and any step** n,

$$\begin{aligned} P_{ij}^{n+1} &= \sum_k P_{ik} P_{kj}^n \\ &\leq \sum_k P_{ik} \max_\ell P_{\ell j}^n \\ &= \max_\ell P_{\ell j}^n \end{aligned}$$

Since this holds for all states i, **it holds for the maximizing** i, **so** $\max_i P_{ij}^{n+1} \leq \max_\ell P_{\ell j}^n$. **Replacing maxima with minima and reversing inequalities,**

$$\min_i P_{ij}^{n+1} \geq \min_\ell P_{\ell j}^n.$$

Before completing the proof of the theorem, we specialize the theorem to the case where $[P] > 0$, i.e., where $P_{ij} > 0$ for all i, j.

<u>**Lemma 2**</u>**: Let $[P] > 0$ be the transition matrix of a finite-state Markov chain and let $\alpha = \min_{i,j} P_{ij}$. Then for all states j and all $n \geq 1$:**

$$
\begin{aligned}
\max_i P_{ij}^{n+1} - \min_i P_{ij}^{n+1} &\leq \left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n\right)(1-2\alpha). \\
\left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n\right) &\leq (1-2\alpha)^n. \\
\lim_{n\to\infty} \max_\ell P_{\ell j}^n &= \lim_{n\to\infty} \min_\ell P_{\ell j}^n \; > \; 0.
\end{aligned}
$$

Note that Lemma 1 implies that $\lim_{n\to\infty} \max_\ell P_{\ell j}^n$ must exist since this is the limit of a decreasing non-negative sequence. This lemma then shows the maxima and minima both have the same limit.

Proof of lemma 2: We tighten the proof of lemma 1 slightly to make use of the positive elements. For a given n and j, let ℓ_{min} be a state that minimizes P_{ij}^n over i. Then

$$
\begin{aligned}
P_{ij}^{n+1} &= \sum_k P_{ik} P_{kj}^n \\
&\leq \sum_{k \neq \ell_{\text{min}}} P_{ik} \max_\ell P_{\ell j}^n \; + \; P_{i\ell_{\text{min}}} \min_\ell P_{\ell j}^n \\
&= (1 - P_{i\ell_{\text{min}}}) \max_\ell P_{\ell j}^n + P_{i\ell_{\text{min}}} \min_\ell P_{\ell j}^n \\
&= \max_\ell P_{\ell j}^n - P_{i\ell_{\text{min}}} \left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \right) \\
&\leq \max_\ell P_{\ell j}^n - \alpha \left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \right) \\
\max_i P_{ij}^{n+1} &\leq \max_\ell P_{\ell j}^n - \alpha \left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \right)
\end{aligned}
$$

We have shown that

$$\max_i P_{ij}^{n+1} \le \max_\ell P_{\ell j}^n - \alpha \left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \right).$$

Interchanging max with min and $\le$ with $\ge$, we get

$$\min_i P_{ij}^{n+1} \ge \min_\ell P_{\ell j}^n + \alpha \left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \right).$$

Subtracting these equations,

$$\max_i P_{ij}^{n+1} - \min_i P_{ij}^{n+1} \le \left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \right)(1 - 2\alpha).$$

Since $\min_\ell P_{\ell j} \ge \alpha$ **and** $\max_\ell P_{\ell j} \le 1 - \alpha$**,**

$$\max_\ell P_{\ell j} - \min_\ell P_{\ell j} \le 1 - 2\alpha$$

Iterating on n**,**

$$\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \le (1 - 2\alpha)^n$$

Finally, we can get back to arbitrary finite-state ergodic chains with transition matrix $[P]$.

We have shown that $[P^h]$ is positive for $h = (\mathsf{M}-1)^2+1$, so we can apply Lemma 2 to $[P^h]$, with $\alpha = \min_{ij} P_{ij}^h$.

We don't much care about the value of α, but only that it is positive. Then

$$\lim_{m\to\infty} \max_{\ell} P_{\ell j}^{hm} = \min_{\ell} P_{\ell j}^{hm} = \pi_\ell > 0$$

To show that the limit applies for all n rather than than just multiples of h, we use Lemma 1, showing that $\max_\ell P_{\ell j}^n$ is non-increasing in n, so it must have the same limit as $\max_\ell P_{\ell j}^{hm}$. The same argument applies for the minima. QED

Ergodic unichains

We have now seen that for ergodic chains, $\lim_{n\to\infty} P_{ij}^n = \pi_j$ for all i where $\vec{\pi}$ is a probability vector. The resulting vector $\vec{\pi}$ is also a steady-state vector and is the unique probability vector solution to $\vec{\pi}[P] = \vec{\pi}$ (see Thm 3.3.1).

It is fairly easy to extend this result to a more general class called ergodic unichains. These are chains containing a single ergodic class along with an arbitrary set of transient states.

If a state is in a singleton transient class, then there is a fixed probability, say α, of leaving the class at each step, and the probability of remaining in the class for more than n steps is $(1-\alpha)^n$.

Th probability of remaining in an arbitrary set of transient states also decays to 0 exponentially with n. Essentially each transient state has at least one path to a recurrent state, and one of those paths must be taken eventually.

For an ergodic unichain, the ergodic class is eventually entered, and then steady state in that class is reached.

For every state j then,

$$\lim_{n\to\infty} \max_i P_{ij}^n = \lim_{n\to\infty} \min_i P_{ij}^n = \pi_j$$

The difference here is that $\pi_j = 0$ for each transient state and $\pi_j > 0$ for each recurrent state.

Other finite-state Markov chains

First consider a Markov chain with several ergodic classes, $\mathcal{C}_1, \ldots, \mathcal{C}_m$. The classes don't communicate and should be considered separately.

If one insists on analyzing the entire chain, $[P]$ will have m independent steady state vectors, one nonzero on each class. $[P^n]$ will then converge, but the rows will not all be the same.

There will be m sets of rows, one for each class, and the row for class k will be nonzero only for the elements of that class.

Next consider a periodic recurrent chain of period d. This can be separated into d subclasses with a cyclic rotation between them.

If we look at $[P^d]$, we see that each subclass becomes an ergodic class, say $\mathcal{C}_1, \ldots, \mathcal{C}_d$. Thus $\lim_{n\to\infty}[P^{nd}]$ exists.

A steady state is reached within each subclass, but the chain rotates from one subclass to another.

6.262 Discrete Stochastic Processes
Spring 2011