Stat 205B Lecture #4

Prof. Jim Pitman

January 30, 2003

Stationary Distributions

Motivation

We are working in a setting where (X_n) is a Markov chain over a countable state space S, with initial distribution $\mu(j)$ and transition distribution $P(j,k)$. The probability distribution $\mathbb{P}_\mu$ is defined so that $\mathbb{P}_\mu(X_0 = j) = \mu(j)$ and $\mathbb{P}_\mu(X_{n+1} = k \mid X_n = j) = P(j,k)$. Our goal is to understand when the distribution of X_n has a limit as $n \to \infty$, and how to characterize that limit.

Recall:

$$\mathbb{P}_\mu(X_n = k) = \sum_i \mu(i) P^n(i,k).$$

If we regard μ as a row vector, we can write this as $\mathbb{P}_\mu(X_n = k) = (\mu P^n)_k$. Suppose this converges (pointwise) to some π as $n \to \infty$:

$$(\mu P^n)_k \to \pi_k \quad \forall k \in S.$$

The nice case is when π is a probability distribution, but it doesn't have to be one. For example, $\pi_k \equiv 0$ is a possibility: this happens if the chain is transient (e.g., a biased random walk on the integers) or if the chain is null-recurrent (e.g. symmetric random walk on the integers; null recurrence discussed further below). Notice that if π is a probability distribution, then $\mu P^n \to \pi$ pointwise is equivalent to $\mu P^n \xrightarrow{d} \pi$, using the usual notion of convergence in distribution.

Suppose $\mu P^n \to \pi$. Then we can change the indexing and deduce:

$$\begin{aligned} \lim_{n\to\infty} \mu P^{n+1} &= \pi \\ \lim_{n\to\infty} \mu P^n P &= \pi \\ \pi P &= \pi \end{aligned}$$

For the last conclusion you must justify swapping the limit with a potentially infinite sum, but this is easily done. For now the point is just that distributions π such that $\pi P = \pi$ will inevitably arise in the theory of limit distributions of MCs.

Definition of stationary distribution

Definition 1. *Call a non-negative function* $\mu = (\mu(i) : i \in S)$ *a* measure *on* S*. We may also regard* μ *as a row vector. Call* μ *a* stationary distribution *for the transition matrix* P *if* $\mu P = \mu$*, that is:*

$$\sum_i \mu(i)P(i,j) = \mu(j) \quad \forall j \in S \tag{EQ}$$

A stationary distribution is also called an *equilibrium distribution* or *invariant distribution*. Note that this definition does not require μ to be a probability distribution. But if $\sum_i \mu(i) = 1$, then we call μ a *stationary probability distribution*.

If μ is a stationary probability distribution then, under $\mathbb{P}_\mu$, the process $(X_0, X_1, \ldots)$ is a *stationary process*:

$$(X_0, X_1, X_2, \ldots) \stackrel{d}{=} (X_1, X_2, \ldots)$$

or equivalently:

$$(X_0, X_1, \ldots, X_{n-1}) \stackrel{d}{=} (X_1, \ldots, X_n) \quad \text{for each } n = 1, 2, \ldots$$

That is, the joint distribution of variables in the MC is invariant under shifts.

Reversible chains

A special kind of stationary process is a *reversible chain*, where:

$$(X_0, X_1, \ldots, X_{n-1}) \stackrel{d}{=} (X_{n-1}, X_{n-2}, \ldots, X_0) \quad \text{for each} n = 1, 2, \ldots$$

It is easy to see that this condition holds for all n if and only if it holds for $n = 2$, that is, $(X_0, X_1) \stackrel{d}{=} (X_1, X_0)$, or equivalently

$$\mu(i)P(i,j) = \mu(j)P(j,i) \quad \forall i, j \in S \tag{REQ}$$

A measure μ that satisfies this equation is called a *reversible equilibrium distribution* for the transition matrix P.

Note that if μ solves (REQ), then μ solves (EQ). To see this, assume μ solves (REQ) and evaluate the lefthand side of (EQ) as follows:

$$\sum_i \mu(i)P(i,j) \;=\; \sum_i \mu(j)P(j,i) \;=\; \mu(j)\sum_i P(j,i) \;=\; \mu(j)$$

We get the righthand side of (EQ).

However, the converse does not hold. If $|S| = N$, then (EQ) is a system of N equations in N unknowns, while (REQ) is a highly overdetermined system of $\binom{N}{2}$ equations in N unknowns. So (REQ) may fail to have a solution even when (EQ) is solvable. On the other hand, if (REQ) has a solution, it is typically very easy to identify it. So if we are trying to find an equilibrium distribution for a given MC, it makes sense to look for a reversible equilibrium first.

Also, notice that both (EQ) and (REQ) are linear, so solutions are only determined up to constant multiples: if μ is a solution, then so is $c\mu$ for any constant c.

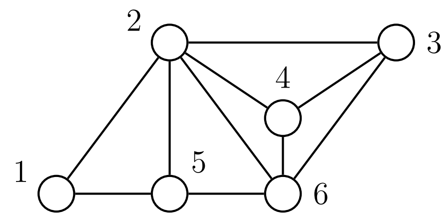

Figure 1: An undirected graph with 6 states.

Example: Random walk on a graph. In a graph, each state i has a set of neighbors $\text{nbrs}(i)$. The random walk is defined as follows:

$$P(i,j) = \begin{cases} \frac{1}{|\text{nbrs}(i)|} & \text{if } j \in \text{nbrs}(i) \\ 0 & \text{otherwise} \end{cases}$$

Consider the graph in Fig. 1. Lets look for a reversible equilibrium, say with $\mu(1) = 2$. What must $\mu(2)$ be? Well, state 1 has 2 neighbors, so $P(1,2) = 1/2$. State 2 has 5 neighbors, so $P(2,1) = 1/5$. So, solving $\mu(1)P(1,2) = \mu(2)P(2,1)$, we get $\mu(2) = 5$.

Continuing through the graph in this way, we can determine $\mu(i)$ for all i. We find that $\mu(i) = |\text{nbrs}(i)|$ is the unique reversible equilibrium (up to constant multiples). This is true for general *connected* graphs. But if the graph has more than one connected component, we can set the scale constants for the various components independently, so the reversible equilibrium is not unique up to constant multiples.

We argued at the beginning of this lecture that the limit distribution must be a stationary distribution. It can be shown (see the book) that in a connected graph, $\mu(i) = |\text{nbrs}(i)|$ is not only the unique reversible equilibrium, but in fact the only stationary distribution of any kind. So X_n has a limi distribution, it must be a normalized version of μ:

$$\lim_{n\to\infty} \mathbb{P}(X_n = j) = \frac{\mu(j)}{\sum_i \mu(i)}$$

Example: Non-reversible equilibrium. Consider the chain in Fig. 2, where p and q are transition probabilities (hence $q = 1 - p$). It is easy to see that the uniform distribution is stationary for any p. However, this distribution is reversible only if $p = q$.

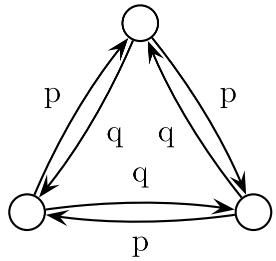

Figure 2: A 3-state Markov chain with transition probabilities p and q.

Aside: Periodic chains. State i has *period* d if:

$$d = \gcd\{n \,:\, P^n(i,i) > 0\}$$

That is, if $X_0 = i$, then the set of times when the chain may return to i are all divisible by d. With some number theory, one can prove that this set contains all sufficiently large multiples of d. If $d > 1$, then the chain is called *periodic*. If $d = 1$, then the chain may return to i at all sufficiently large times, and the chain is called *aperiodic*. It is also easy to see that if a chain is irreducible, then all the states have the same period. In the chain of Figure 2, if $p = 0$ or $p = 1$ then the chain is periodic (with period 3).

Existence and uniqueness of stationary distributions

Theorem 1. *If P governs an irreducible and recurrent Markov chain then there exists a stationary distribution μ which is unique up to constant multiples.*

This theorem allows two possibilities:

- the chain is *positive recurrent*: $\sum_i \mu(i) < \infty$. Then the unique stationary probability distribution is:
$$\pi(i) = \frac{\mu(i)}{\sum_j \mu(j)}$$
- the chain is *null recurrent*: $\sum_i \mu(i) = \infty$. Then the chain has no stationary probability distribution.

Note that if P is irreducible and the state space is finite, then obviously the chain is recurrent (it obviously must hit some state infinitely often with non-zero probability). So by the theorem above, the chain has a unique stationary probability distribution.

To prove existence, assume P governs an irreducible and recurrent MC, and construct a stationary distribution μ as follows. Fix a base state i. Let T_i be the first $n \geq 1$ such that $X_n = i$. Because the chain is recurrent, we know $\mathbb{P}_i(T_i < \infty) = 1$. Also, let N_{ji} be the number of times the chain hits j before T_i; that is:

$$N_{ji} = \sum_{n=0}^{\infty} \mathbb{1}(X_n = j,\, T_i > n)$$

Note that $N_{ii} = 1$. Finally, define:

$$\mu_i(j) = \mathbb{E}_i N_{ji}$$

so $\mu_i(j)$ is the expected number of js in an i-block. We claim that for each fixed i, the mapping $j \mapsto \mu_i(j)$ is a stationary distribution. Also, these stationary distributions are all the same up to constant multiples, regardless of i.

The first claim is that for any fixed i:

$$\sum_j \mu_i(j) P(j,k) = \mu_i(k) \quad \forall k \in S$$

There are two cases: $k = i$, and $k \neq i$. However, the proofs for both cases rely on the following observation:

$$\begin{aligned}
\mu_i(j)P(j,k) &= \mathbb{E}_i(\text{num } j \to k \text{ transitions in first } i\text{-block}) \\
&= \mathbb{E}_i\left(\sum_{n=0}^{\infty} \mathbb{1}(X_n = j,\, X_{n+1} = k,\, T_i > n)\right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j,\, X_{n+1} = k,\, T_i > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j,\, T_i > n)P(j,k) \quad \text{by Markov property}
\end{aligned}$$

Durrett derives a similar fact, and uses it to complete the proof for the two cases ($k = i$ and $k \neq i$). See the book for those cases (p. 304), and for the uniqueness part of the proof (p. 305).