Lecture 9

David Aldous

16 September 2015

Ideas used in Lecture 8.

- Communicating classes, defined in terms of possible transitions.
- Definition of stationary distribution.
- Special structures making it easy to calculate the stationary distribution: doubly stochastic, detailed balance, RW on weighted undirected graph, success runs.

To repeat the first item, recall that the transition matrix $\mathbf{P}$ of a Markov chain can be represented as a weighted directed graph. In the previous lecture we first looked at some "structure theory" – some qualitative aspects of the chain's behavior do not depend on actual numerical transition probabilities but only on the graph of possible transitions.

- j is **accessible** from i if there is a (directed) path from i to j (or $i = j$).
- i and j **communicate** if each is accessible from the other.
- Because "communicate" is an equivalence relation, the state space **States** can be partitioned into **communicating classes** (CCs), say $C_1, C_2, \ldots$, such that i and j **communicate** if and only if they are in the same CC.
- A class C is **open** if it is possible to leave; that is if $p_{ij} > 0$ for some $i \in C$ and $j \notin C$. Otherwise it is **closed**.
- The graph is **strongly connected** if there is only one CC, that is if all states communicate. In the Markov chain context this property is called **irreducible**.

Here is another definition that depends only on the graph of possible transitions. Suppose we can partition the states into $k \geq 2$ subsets $D_0, D_1, \ldots, D_{k-1}$ such that every transition $i \to j$ takes the particle from its current subset to the next subset. That is

$$\text{if } i \in D_u \text{ and } p_{ij} > 0 \text{ then } j \in D_{u+1} \tag{1}$$

where $u + 1$ is taken modulo k.

If (1) holds for some k the chain is said to have **period** k. (More precisely, the period is the largest k for which (1) holds). If not, the chain is called **aperiodic**.

Some later theory will involve the assumption that a chain is irreducible and aperiodic. Given irreducible, to check the chain is aperiodic it is sufficient to know that $p_{ii} > 0$ for some i. The general necessary-and-sufficient condition – see [PK] section 4.3.2 – is that for some state i

$$\text{greatest common divisor of } \{t : p_{ii}^{(t)} > 0\} \text{ is } 1.$$

Stationary distributions

Recall the distribution $\mu(t)$ of X_t evolves as $\mu(t) = \mu(t-1)\mathbf{P}$ in vector-matrix notation. So suppose a probability distribution $\pi = (\pi_i, i \in \mathbf{States})$ satisfies

$$\pi = \pi\mathbf{P}; \quad \text{that is } \sum_i \pi_i p_{ij} = \pi_j \ \forall j. \tag{2}$$

If the chain has initial (time-0) distribution $\mu(0) = \pi$ then $\mu(t) = \pi$ for every time t. A distribution π satisfying (2) is called **stationary**.

This language is a bit confusing, when we imagine a Markov chain as a particle jumping between states. The particle continues to move even when we have a stationary distribution; **stationary** refers to the fact that the **probabilities** (of where the particle is at time t) do not change with time t.

If $\mu(t)$ and $\mu(\infty)$ are probability distributions on **States**, then convergence $\mu(t) \to \mu(\infty)$ as $t \to \infty$ means $\mu_i(t) \to \mu_i(\infty)$ as $t \to \infty$ for each $i \in$ **States**.

So if $\mu(t)$ is the distribution of $X(t)$ then $\mu(t) \to \mu(\infty)$ means

$$\mu_i(t) = \mathbb{P}(X(t) = i) \to \mu_i(\infty) \text{ for each } i \in \textbf{States}. \qquad (3)$$

Suppose that for a chain with transition matrix **P** we know (3) holds. Then (3) implies

$$\mu(t+1) = \mu(t)\mathbf{P} \to \mu(\infty)\mathbf{P}$$

which implies, because $\mu(t+1) \to \mu(\infty)$,

$$\mu(\infty) = \mu(\infty)\mathbf{P}$$

That is, the limit distribution of $X(t)$, if it exists, must be a stationary distribution, which we will now call π.

Mean occupation times. Consider a state i and time t.

$$\sum_{s=0}^{t-1} \mathbb{1}_{(X(s)=i)} = \text{ number of visits to } i \text{ before } t$$

$$\frac{1}{t} \sum_{s=0}^{t-1} \mathbb{1}_{(X(s)=i)} = \text{ proportion of time at } i \text{ before } t$$

$$\mathbb{E}[\tfrac{1}{t} \sum_{s=0}^{t-1} \mathbb{1}_{(X(s)=i)}] = \text{ mean proportion of time at } i \text{ before } t$$

$$= \frac{1}{t} \sum_{s=0}^{t-1} \mathbb{P}(X(s) = i).$$

From algebra/calculus, if $a(t) \to a(\infty)$ then $\frac{1}{t} \sum_{s=0}^{t-1} a(s) \to a(\infty)$.
Conclusion. For a Markov chain, if $\mu(t) \to \pi$ then

$$(\text{mean proportion of time at } i \text{ before } t) \to \pi_i.$$

We can extend this idea by imagining **costs** $c(i)$ (or gains). That is, suppose spending unit time in state i incurs a cost $c(i)$. Then, by summing over all states i.

$$\frac{1}{t}\mathbb{E}\left(\text{total cost during times } \{0, 1, \ldots, t-1\}\right) \to \sum_i c_i \pi_i .$$

In our earlier context of setting up first-step analysis of hitting times $\mathbb{E}_i T_A$, we can also consider the mean total cost up to time T_A – see [PK] sec. 3.4.2.

We now come to the central part of theory for Markov chains. Everyone says this theory in slightly different ways. See [PK] section 4.4; also a concise theory treatment with proofs is in [BZ] sections 5.3 - 5.4.

Assume irreducible; state space may be finite or countable infinite.

Recall $T_i = \min\{t \geq 0 : X_t = i\}$ and define also the **return time**

$$T_i^+ = \min\{t \geq 1 : X_t = i\}.$$

Fix a reference state b.

Theorem

Suppose irreducible.
(a) If state space is finite then $\mathbb{E}_b T_b^+ < \infty$.
(b) Suppose $\mathbb{E}_b T_b^+ < \infty$. *Define*

$$a(b,i) = \mathbb{E}_b \sum_{s=1}^{T_b^+} \mathbb{1}_{(X(s)=i)}$$

$=$ *mean number of visits to* i *before returning to* b. *So* $a(b,b) = 1$. *Then*

$$\pi_i = \frac{a(b,i)}{\mathbb{E}_b T_b^+}$$

is a stationary distribution, and is the **only** *stationary distribution.*

Discussion.

(a) There is a calculation which checks this π does satisfy $\pi = \pi \mathbf{P}$.

(b) Because π is the same for each choice of b we have another formula

$$\pi_i = \frac{1}{\mathbb{E}_i T_i^+} \text{ for each } i.$$

(c) For an irreducible chain, the properties

$\mathbb{E}_i T_i^+ < \infty$ for some i

$\mathbb{E}_i T_i^+ < \infty$ for all i

are equivalent. When these hold we call the chain **positive-recurrent**.

(d) The theorem implies that every finite-state irreducible chain is positive-recurrent. So every finite-state irreducible chain has a unique stationary distribution.