The Stationary Distribution of a Markov Chain

Alessandro Panconesi
DI, La Sapienza of Rome

May 15, 2005

In this note I present a concise proof of the existence and uniqueness of the limit distribution of an ergodic markov chain. This nice proof was described to me by David Gilat of Hebrew University during a hot summer afternoon in Perugia.

A finite $n \times n$ *transition probability matrix* $P := [p_{ij}]$ is a stochastic matrix where p_{ij} is the *transition probability* of going from state i to state j. We can think of a directed graph whose edges are weighted with transition probabilities. Note that since P is stochastic, the sum of the probabilities of the arcs outgoing from a vertex i sum up to 1, i.e. $\sum_j p_{ij} = 1$.

We are interested in keeping track of the random walk of a pebble. The pebble is initially placed on the graph according to an initial distribution X_0 and then it proceeds by traversing the edges ij's with their associated probabilities p_{ij}'s. X_0 is a probability distribution, X_0^i being the probability of placing the pebble on vertex i initially. After one step the position of the pebble is given by the probability distribution

$$X_1 = X_0 P$$

and in general after t steps we have

$$X_t = X_{t-1} P = X_0 P^t.$$

The infinite sequence $X_0, X_1, X_2, \ldots$ is called a *markov chain*. In what follows we shall use the term markov chain somewhat loosely, sometimes referring to the sequence proper, sometimes to the transition matrix P or the underlying graph. The meaning will be clear from the context.

We want to study the convergence of X_t. Under what conditions does it converge to a (unique) limit *for any starting distribution*? As we shall see there are two conditions that P should satisfy. Each of them is by itself necessary. It is remarkable that together they are also sufficient.

Let us start by asking what conditions might imply that the limit distribution is not unique or does not exist. Assume that the graph has a two sinks, i.e. two connected components where it is possible to enter but from which it is not possible to get out. In this case the limit, if it exists, is certainly not unique. Therefore *strong connectivity* of the underlying graph (the one defined by edges with strictly positive weight) is certainly a necessary condition. A strongly connected markov chain is also called *irreducible*.

Consider now a bipartite graph and assume that X_0 places the pebble with probability 1 on a specific vertex i. At even times the pebble can only be on the same side of the bipartition, while at odd times it is on the opposite side. Therefore X_t oscillates forever and does not converge to any limit. This motivates the following definitions and lemmata.

Definition 1 *The* period *of a state i is defined as*

$$p(i) := gcd\{n \geq 1 : p_{ii}^n > 0\}.$$

A state i is aperiodic *if its period is 1. A markov chain is aperiodic if every state is.*

Consider a markov chain in which a state i is at the intersection of two cycles, one of length 4 and the other of length 5. The reader can verify that for all $t \geq 12, p_{ii}^t > 0$. This is no coincidence.

Lemma 1 *If a markov chain is aperiodic then there exists N such that, for all i, if $t \geq N$ then $p_{ii}^n > 0$.*[1]

Strong connectivity and aperiodicity imply the following useful lemma.

Lemma 2 *If a markov chain is strongly connected and aperiodic then there exists N such that, for all i and j, if $t \geq N$ then $p_{ij}^t > 0$.*

Proof: Exercise. ■

We want to show that if a markov chain is strongly connected and aperiodic then there exist a unique limit distribution, and moreover this distribution is *stationary*. A distribution X is stationary if $X = XP$. Both results will be derived as corollary of the following theorem.

[1]For the proof see for instance the nice booklet *Finite Markov Chains and Algorithmic Apllications* by Olle Häggström. Recall that we consider here finite transition matrices.

Theorem 1 *Let P be the transition matrix of a markov chain that is aperiodic and strongly connected. Then, $\lim_{t\to\infty} P^t = P^\infty$ where*

$$P^\infty := \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_n \\ \pi_1 & \pi_2 & \dots & \pi_n \\ & \dots & \dots & \\ \pi_1 & \pi_2 & \dots & \pi_n \end{bmatrix}$$

Before the proof let us derive some consequences of the theorem. Let $\pi := (\pi_1, \pi_2, \dots, \pi_n)$.

Corollary 1 *π is a stationary distribution.*

Proof:

$$\begin{aligned} P^\infty P &= \left(\lim_{t\to\infty} P^t\right) P \\ &= \left(\lim_{t\to\infty} P^{t+1}\right) \\ &= P^\infty. \end{aligned}$$

The claim follows. ∎

Corollary 2 *For any initial distribution X_0, the sequence $X_t = X_{t-1}P = X_0P^t$ converges to π, i.e. $\lim_{t\to\infty} X_t = \pi$.*

Proof:

$$\begin{aligned} \lim_{t\to\infty} X_t &= \lim_{t\to\infty} \left(X_0\ P^t\right) \\ &= X_0 \left(\lim_{t\to\infty} P^t\right) \\ &= X_0 P^\infty \\ &= \sum_{i=1}^{n} X_0^i\, \pi \\ &= \left(\sum_{i=1}^{n} X_0^i\right) \pi \\ &= \pi. \end{aligned}$$

∎

Thus, if the chain is irreducible and aperiodic there is a unique limit, stationary distribution.

Proof of Theorem 1: The idea of the proof is to keep track of the smallest and largest values of a fixed column j of P^t, as t goes to infinity. Fix j, and let m_t and M_t be the smallest and largest values of the jth column, respectively. By lemma 2 we can assume without loss of generality that $p_{ij} > 0$. Let

$$\delta := \min_{ij} p_{ij}.$$

Since P is stochastic, from the assumption above we have

$$0 < \delta := \min_{ij} p_{ij} \leq \frac{1}{2}.$$

We ill prove that:

1. The sequence $\{m_t\}$ is non-decreasing;
2. The sequence $\{M_t\}$ is non-increasing;
3. $\Delta_t := M_t - m_t$ goes to zero (exponentially fast!).

Exercise 1 *Show that thee 3rd condition does not imply that m_t and M_t converege to the same limit, while the three conditions together do.*

We now prove that the first condition holds.

$$\begin{aligned} m_{t+1} &= \min_i p_{kj}^{t+1} \\ &= \min_i \sum_k p_{ik}\, p_{kj}^t \\ &\geq \min_i \sum_k p_{ik}\, m_t \\ &= \min_i \left(\sum_k p_{ik} \right) m_t \\ &= m_t. \end{aligned}$$

The second condition can be established similarly. Let us now establish the 3rd condition. Let ℓ be the row where M_t lies, i.e. $M_t = p_{\ell j}^t$. Then,

$$m_{t+1} = \min_i p_{kj}^{t+1}$$

$$
\begin{aligned}
&= \min_i \sum_k p_{ik}\; p^t_{kj} \\
&= \min_i p_{\ell j}\; M_t + \sum_{k \neq \ell} p_{ik} p^t_{kj} \\
&\geq \min_i p_{\ell j}\; M_t + \sum_{k \neq \ell} p_{ik} m_t \\
&= \min_i p_{\ell j}\; M_t + (1 - p_{\ell j})\; m_t \\
&\geq \min_i \delta\; M_t + (1-\delta)\; m_t \\
&= \delta\; M_t + (1-\delta)\; m_t.
\end{aligned}
$$

Similarly,

$$M_t \leq \delta m_t + (1-\delta) M_t.$$

Taking the two together we get,

$$
\begin{aligned}
\Delta_{t+1} &= M_{t+1} - m_{t+1} \\
&\leq (1-2\delta)(M_t - m_t) \\
&\leq (1-2\delta)\Delta_t \\
&\leq (1-2\delta)^t.
\end{aligned}
$$

■