Contents

5 Markov chains

5a Random walk on a regular graph

Assume that a weakly connected[1] finite directed graph has m vertices and is regular (that is, each vertex has k outgoing edges and k incoming edges, the same k for all vertices). In addition, we assume *aperiodicity:* there exists no $p \in \{2, 3, \dots\}$ such that the length of every cycle is divisible by p. A random walk started at a given vertex. Denote by S_n the position of the walk after n steps.

5a1 Proposition. For each vertex x of the graph,

$$\mathbb{P}\big(S_n = x\big) \to \frac{1}{m} \quad \text{as } n \to \infty\,.$$

This fact is a special case of a convergence theorem for Markov chains (see 5b3).[2]

Now assume that the random walk, started at a given vertex, stops on the first return to this vertex.

5a2 Proposition. The expected number of moves is equal to m (the number of vertices).

The proof uses Markov chains.[3] (Aperiodicity is not needed.)

Think, what happens if the graph consists of two large pieces connected by a thin neck.

Prop. 5a2 will be proved in Sect. 5c; Prop. 5a1 — in the end of this Sect. 5a.

First, some graph theory.

[1] That is, the corresponding undirected graph is connected.
[2] [D, Sect. 5.4, Example 4.5; Sect. 5.5(a)]; [KS, Sect. 5.3].
[3] [D, Sect. 5.4, (4.6) and Example 4.5].

We consider an aperiodic regular weakly connected finite directed graph. The graph has a set V of vertices and a set $E \subset V \times V$ of edges.[1] Weak connectedness:

$$\emptyset \subsetneq A \subsetneq V \quad \Longrightarrow \quad E \cap \big((A \times (V \setminus A)) \cup ((V \setminus A) \times A) \big) \ne \emptyset$$

for all $A \subset V$. Regularity:

$$\#\{y : (x,y) \in E\} = k = \#\{y : (y,x) \in E\}$$

for all $x \in V$.

5a3 Lemma. For every $A \subset V$ the number of incoming edges is equal to the number of outgoing edges; that is,

$$\#\big(E \cap (A \times (V \setminus A))\big) = \#\big(E \cap ((V \setminus A) \times A)\big) \, .$$

Proof. Denoting $B = V \setminus A$ we have

$$E \cap (A \times V) = E \cap (A \times B) \uplus E \cap (A \times A) \, ,$$
$$E \cap (V \times A) = E \cap (B \times A) \uplus E \cap (A \times A) \, ,$$

thus $\#\big(E \cap (A \times B)\big) = k \cdot (\#A) - \#\big(E \cap (A \times A)\big) = \#\big(E \cap (B \times A)\big)$. □

5a4 Corollary. Strong connectedness:

$$\emptyset \subsetneq A \subsetneq V \quad \Longrightarrow \quad E \cap \big(A \times (V \setminus A)\big) \ne \emptyset$$

for all $A \subset V$. (Closed sets: $\emptyset$ and V only.)

5a5 Corollary. For all $x, y \in V$ there exists a path (of *some* length) from x to y.

5a6 Lemma. There exists n such that for all $x, y \in V$, every $t \ge n$ is the length of some (at least one) path from x to y.

Proof. The set L_x of lengths of all loops from x to x is a semigroup, therefore $L_x - L_x$ is a group, $L_x - L_x = p_x \mathbb{Z}$ for some p_x. By 5a5, $L_x - L_x$ does not depend on x. Thus, $p_x = 1$ for all x. It means existence of N_x such that $N_x \in L_x$ and $N_x + 1 \in L_x$. We take $n_x = N_x^2$ and note that[2] $N_x^2 + kN_x + r = N_x(N_x+k)+r = N_x(N_x+k)-N_x r+(N_x+1)r = N_x(N_x+k-r)+(N_x+1)r \in L_x$. We take m such that a path of length $\le m$ exists from every x to every y; then $n = m + \max_x n_x$ fits. □

[1] May intersect the diagonal. Multiple edges are excluded, but all said can be easily generalized to graphs with multiple edges.

[2] Example: $\{10k + 11l\} \not\ni 78, 79, 89$.

Now we return to probability.

We want to show that the initial point x_0 is ultimately forgotten by the random walk (S_n).

Given another starting point $x'_0 \in V$, we introduce the probability space Ω' of paths (of length n) starting at x'_0, and random variables $S'_0, \dots, S'_n : \Omega' \to V$. We take the product

$$\tilde\Omega = \Omega \times \Omega'$$

and treat S_t, S'_t as maps $\tilde\Omega \to V$. We get two *independent* random walks, one starting at x_0, the other at x'_0. In addition, we let $\tilde S_t = (S_t, S'_t) : \tilde\Omega \to \tilde V = V \times V$.

The reflection helps again! The transformation $(x, y) \mapsto (y, x)$ of $\tilde V$ will be treated as reflection, and the diagonal of $\tilde V$ as the barrier. We define $M_n : \tilde\Omega \to \{0, 1\}$ by

$$M_n = \begin{cases} 0 & \text{if } S_0 \ne S'_0, S_1 \ne S'_1, \dots, S_n \ne S'_n, \\ 1 & \text{otherwise.} \end{cases}$$

5a7 Exercise. $\mathbb{E}\big(f(\tilde S_n)1\!\!1_{M_n=1}\big) = 0$ for every antisymmetric function $f : \tilde V \to \mathbb{R}$ ("antisymmetric" means $f(y, x) = -f(x, y)$).

Prove it.

Hint: similar to the proof of Lemma 4a2.

That is, the conditional distribution of $\tilde S_n$ given $M_n = 1$ is symmetric (if defined).

And again (recall 4a3), $\mathbb{E} f(\tilde S_n) = \mathbb{E}\big(f(\tilde S_n)1\!\!1_{M_n=0}\big)$.

5a8 Lemma. $|\mathbb{P}\big(S_n = x\big) - \mathbb{P}\big(S'_n = x\big)| \le \mathbb{P}\big(M_n = 0\big)$.

Proof. Take $f(a, b) = 1\!\!1_{\{x\}}(a) - 1\!\!1_{\{x\}}(b)$ in 5a7. ☐

The probability of the event $M_n = 0$ depends on n, x_0 and x'_0. We maximize it in x_0, x'_0:

$$\varepsilon_n = \max_{x_0, x'_0 \in V} \mathbb{P}\big(M_n = 0\big).$$

5a9 Lemma. $\varepsilon_n \to 0$ as $n \to \infty$.

The proof will be given later.

Let $p_n(x, y)$ denote the n-step transition probability from x to y. (Thus, $\mathbb{P}\big(S_t = y\big) = p_t(x_0, y)$ and $\mathbb{P}\big(S'_t = y\big) = p_t(x'_0, y)$.)

Clearly, $\sum_{y \in V} p_1(x, y) = 1$ for all $x \in V$; but regularity ensures also $\sum_{x \in V} p_1(x, y) = 1$ for all $y \in V$. By induction, $\sum_{y \in V} p_n(x, y) = 1$ for all $x \in V$, and $\sum_{x \in V} p_n(x, y) = 1$ for all $y \in V$.

Proof of Prop. 5a1. By Lemma 5a8, $|p_n(x_0,y) - p_n(x_0',y)| \le \varepsilon_n$. We average it in x_0'; taking into account that $\frac1m \sum_{x_0'\in V} p_n(x_0',y) = \frac1m$ we get $|p_n(x_0,y) - \frac1m| \le \varepsilon_n$; finally, $\varepsilon_n \to 0$ by Lemma 5a9. $\square$

Proof of Lemma 5a9. Lemma 5a6 gives us n such that $p_n(x,y) \ne 0$ for all x,y. Clearly, $p_n(x,y) \ge k^{-n}$. Thus,

$$\mathbb{P}\big(M_n = 1 \big) \ge \mathbb{P}\big(S_n = y, S_n' = y \big) \ge k^{-2n} \,,$$

no matter which y is used. We put $\theta = 1 - k^{-2n}$ and see that $\mathbb{P}\big(M_n = 0 \big) \le \theta$. But moreover, $\mathbb{P}\big(M_{t+n} = 0 \big| M_t = 0, S_t = a, S_t' = b \big) \le \theta$ for all a,b (provided that the condition is of non-zero probability). It follows that

$$\begin{gathered} \mathbb{P}\big(M_{t+n} = 0 \big| M_t = 0 \big) \le \theta \quad \text{for all } t \,; \\ \mathbb{P}\big(M_{t+n} = 0 \big) \le \theta \cdot \mathbb{P}\big(M_t = 0 \big) \quad \text{for all } t \,; \\ \mathbb{P}\big(M_{jn} = 0 \big) \le \theta^j \quad \text{for all } j \end{gathered}$$

and, of course, $\theta^j \to 0$ as $j \to \infty$. $\square$

Interestingly, $\varepsilon_n \to 0$ exponentially fast. However, the constant nk^{2n} can be quite large.

5b Finite Markov chains

A *Markov chain* (discrete in space and time, and homogeneous in time) is described by a *transition probability matrix*

$$\big(p(x,y) \big)_{x,y\in V}$$

satisfying

$$p(x,y) \ge 0 \,; \quad \forall x \;\; \sum_y p(x,y) = 1 \,.$$

The set V is assumed to be finite. We turn V into a graph putting

$$E = \{ (x,y) \in V^2 : p(x,y) \ne 0 \}$$

and define the probability of a path $(s_0,\dots,s_n)$ as the product of n probabilities

$$p(s_0,\dots,s_n) = p(s_0,s_1)\dots p(s_{n-1},s_n) \,;$$

as before, s_0 must be equal to a given initial point $x_0 \in V$. Here are some definitions that depend on the graph only.

A set $A \subset V$ is *closed* if $E \cap \big(A \times (V \setminus A) \big) = \emptyset$.

A Markov chain is *irreducible* if $\emptyset$ and V are the only closed sets. In other words: the graph is strongly connected. Equivalently: for all $x, y \in V$ there exists a path from x to y (recall 5a5).

An irreducible Markov chain is *aperiodic,* if there exists no $p \in \{2, 3, \dots\}$ such that every loop length is divisible by p. (This property does not depend on the initial point; recall the proof of 5a6.)

Here are some results stated here without proofs.

5b1 Theorem. If a Markov chain is irreducible and aperiodic then the limit

$$\lim_n \mathbb{P}\big(S_n = x\big)$$

exists for each $x \in V$.

5b2 Definition. A probability measure μ on V is *stationary,* if

$$\mu(y) = \sum_{x \in V} \mu(x) p(x, y) \quad \text{for all } y \in V \, .$$

Irreducibility implies that $\mu(x) > 0$ for all x (since the set $\{x : \mu(x) > 0\}$ is closed).

5b3 Theorem. If a Markov chain is irreducible and aperiodic then it has one and only one stationary probability measure μ, and

$$\forall y \quad \sum_{x \in V} \nu(x) p_n(x, y) \to \mu(y) \quad \text{as } n \to \infty$$

for every probability measure ν on V.

If a Markov chain (V, p) is irreducible but periodic, with the (least) period d, then $V = V_1 \uplus \dots \uplus V_d$ and $p_1(x, y) \ne 0$ only when $x \in V_i, y \in V_{i+1}$ for some i (here $n + 1 = 1$, of course). The Markov chain (V_1, p_d) is irreducible and aperiodic, its stationary probability measure is $\mu(x) = \lim_n \mathbb{P}\big(S_{nd} = x\big)$ (assuming $x_0 \in V_1$), and the measure

$$\nu(x) = \lim_n \frac{1}{d} \Big(\mathbb{P}\big(S_{nd} = x\big) + \mathbb{P}\big(S_{nd+1} = x\big) + \dots + \mathbb{P}\big(S_{nd+d-1} = x\big) \Big)$$

is stationary for the original Markov chain (V, p).

Here is another property related to the graph only.

5b4 Definition. [1] A state $x \in V$ is *transient,* if there exists $y \in V$ such that a path from x to y exists, but a path from y to x does not exist. Otherwise, x is called *recurrent.*

[1] Only for *finite* Markov chains.

If x is transient then $\mathbb{P}\big(S_n = x\big) \to 0$ as $n \to \infty$.

Recurrent states x, y are called equivalent, if there exists a path from x to y, and a path from y to x. (Well, the latter follows from the former.) Equivalence classes are irreducible closed sets...

5c Return time

Similarly to Sect. 5a we consider a regular (weakly) connected finite directed (but maybe periodic) graph (V, E), and the random walk (S_n) on it, starting at a given $x_0 \in V$.

We introduce the "return time" random variable[1] $T = \inf\{n > 0 : S_n = x_0\}$.

5c1 Lemma. $T < \infty$ almost surely, and moreover, $\mathbb{E}\, T < \infty$.

Proof.

$$\begin{gathered}
\exists n\ \forall t\ \ \mathbb{P}\big(T \le t+n \,\big|\, S_0, \dots, S_t\big) > 0 \text{ a.s.}\,; \\
\exists n\ \exists \varepsilon\ \forall t\ \ \mathbb{P}\big(T \le t+n \,\big|\, S_0, \dots, S_t\big) \ge \varepsilon \text{ a.s.}\,; \\
\mathbb{P}\big(T > t+n \,\big|\, S_0, \dots, S_t\big) \le (1-\varepsilon) 1\!\!1_{T>t} \text{ a.s.}\,; \\
\mathbb{P}\big(T > t+n\big) \le (1-\varepsilon)\mathbb{P}\big(T > t\big)\,; \\
\forall j\ \ \mathbb{P}\big(T > jn\big) \le (1-\varepsilon)^j\,.
\end{gathered}$$

□

Treating the (one step) transition function $p(\cdot,\cdot)$ as a matrix and measures on V as row vectors we write $\mu p = \nu$ rather than $\nu(\{y\}) = \sum_x \mu(\{x\}) p(x,y) = \int p(\cdot, y)\, \mathrm{d}\mu$, and in particular, $\delta_x p$ rather than $\sum_y p(x,y)\delta_y$. Thus, distributions of S_n are: $\mathrm{Distr}(S_0) = \delta_{x_0}$, $\mathrm{Distr}(S_1) = \delta_{x_0} p$, and so on. We also use expectations of random vectors (in the m-dimensional linear space of signed measures on V): $\mathrm{Distr}(S_n) = \mathbb{E}\, \delta_{S_n}$ (and in general, $\mathrm{Distr}(X) = \mathbb{E}\, \delta_X$).

"The cycle trick": $\sum_{n=0}^{T-1} \delta_{S_n} = \sum_{n=1}^{T} \delta_{S_n}$ a.s. (just because $S_0 = x_0 = S_T$ a.s).

5c2 Lemma. $\mathbb{E} \sum_{n=1}^{T} \delta_{S_n} = \big(\mathbb{E} \sum_{n=0}^{T-1} \delta_{S_n}\big) p$.

Proof.

$$\begin{gathered}
\mathbb{E}\big(\delta_{S_{n+1}} - \delta_{S_n} p \,\big|\, S_0, \dots, S_n\big) = 0\,; \\
\mathbb{E}\big((\delta_{S_{n+1}} - \delta_{S_n} p) 1\!\!1_{T>n}\big) = 0\,;
\end{gathered}$$

[1] A priori, taking on values in $\{1, 2, \dots\} \cup \{\infty\}$.

taking into account that $\sum_{n=0}^\infty \mathbb{P}\big(T>n\big) = \mathbb{E}\, T < \infty$ (and vectors $\delta_{S_{n+1}} - \delta_{S_n} p$ are a bounded set) we have

$$\mathbb{E} \sum_{n=0}^\infty (\delta_{S_{n+1}} - \delta_{S_n} p) 1\!\!1_{T>n} = 0\,;$$

$$\mathbb{E} \sum_{n=0}^{T-1} (\delta_{S_{n+1}} - \delta_{S_n} p) = 0\,;$$

$$\mathbb{E} \sum_{n=0}^{T-1} \delta_{S_{n+1}} = \Big(\mathbb{E} \sum_{n=0}^{T-1} \delta_{S_n} \Big) p\,.$$

□

Proof of Prop. 5a2. The measure $\mathbb{E} \sum_{n=0}^{T-1} \delta_{S_n}$ is invariant, therefore, proportional to the uniform (or the counting) measure. The measure at x_0 is equal to 1 ($n=0$ only...); thus the measure of the whole V must be m. On the other hand, it is $\mathbb{E} \sum_{n=0}^{T-1} 1 = \mathbb{E}\, T$; thus, $\mathbb{E}\, T = m$. □