1 Random walk on a graph

Assume that a connected finite oriented graph has m vertices, and each vertex has k outgoing edges and k incoming edges (the same k for all vertices). Denote the set of vertices by V and the set of edges by E; $E \subset V \times V$ (it may intersect the diagonal). (Multiple edges are thus excluded, but all said can be easily generalized to graphs with multiple edges.) It is assumed that $E \cap \big((A \times (V \setminus A)) \cup ((V \setminus A) \times A) \big) \ne \emptyset$ for every $A \subset V$ such that $A \ne \emptyset$ and $V \setminus A \ne \emptyset$ (weak connectedness). Also, $\#\{y \in V : (x,y) \in E\} = \#\{y \in V : (y,x) \in E\} = k$ for all $x \in V$. In addition, we assume *aperiodicity:* there exists no $p \in \{2,3,\dots\}$ such that every loop length is divisible by p. (A loop is a sequence of vertices $(y_0, y_1, \dots, y_t)$ such that $(y_0,y_1) \in E, \dots, (y_{t-1}, y_t) \in E$ and $y_t = y_0$; its length is t.)

A random walk started at a given vertex x_0. A path (of length n) of the random walk is a sequence $(s_0, \dots, s_n)$ of vertices such that the pairs (s_{k-1}, s_k) belong to E (for $k = 1, \dots, n$) and $s_0 = x_0$. There are k^n such paths; each has the probability k^{-n} (by definition). We have the probability space Ω of paths, and random variables $S_0, \dots, S_n : \Omega \to V$.

We start with some graph-theoretic (non-probabilistic) statements.

1.1 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. We have

$$\begin{aligned} E \cap (A \times V) &= E \cap (A \times (V \setminus A)) \uplus E \cap (A \times A) , \\ E \cap (V \times A) &= E \cap ((V \setminus A) \times A) \uplus E \cap (A \times A) \end{aligned}$$

and

$$\#\big(E \cap (A \times V)\big) = k \cdot \#A = \#\big(E \cap (V \times A)\big) .$$

□

1.2 Corollary. (Strong connectedness.)

$E \cap (A \times (V \setminus A)) \ne \emptyset$ for every $A \subset V$ such that $A \ne \emptyset$ and $V \setminus A \ne \emptyset$.

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

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

Proof. (Sketch.)

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. The period p_x does not depend on x. Thus, $p_x = 1$ for all x. It means existence of N such that $N \in L_x$ and $N+1 \in L_x$. We take $T = N^2$ and note that $N^2 + kN + r = N(N+k) + r = N(N+k-r) + (N+1)r \in L_x$. Generalization from $x = y$ to all x, y is easy. □

Now we return to probability. We want to show that the initial point x_0 is ultimately forgotten by the Markov chain.

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 $\tilde S_t = (S_t, S'_t) : \tilde\Omega \to \tilde V = V \times V$.

Recall the reflection principle, instrumental in 'Extremal values, etc.' It will help 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

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

1.5 Lemma. The conditional distribution of $\tilde S_n$ given $M_n = 1$ is symmetric.

That is, $\mathbb{E}\big(f(\tilde S_n) \,\big|\, M_n = 1\big) = 0$ for every antisymmetric function $f : \tilde V \to \mathbb{R}$ (antisymmetric means $f(y, x) = -f(x, y)$).

The proof is similar to the proof of Lemma 1 in 'Extremal values, etc.'

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

Prove it.

Hint: $f(a, b) = \mathbf{1}_{\{x\}}(a) - \mathbf{1}_{\{x\}}(b)$.

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

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

1.7 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 $\mathbb{P}\big(S'_t = y\big) = p_t(x'_0,y)$.)

1.8 Exercise. $\sum_{y\in V} p_1(x,y) = 1$ for all $x \in V$, and $\sum_{x\in V} p_1(x,y) = 1$ for all $y \in V$.

Prove it.

Hint: k times $1/k$...

1.9 Exercise. $\sum_{y\in V} p_n(x,y) = 1$ for all $x \in V$, and $\sum_{x\in V} p_n(x,y) = 1$ for all $y \in V$.

Prove it.

Hint: induction in n.

1.10 Theorem. For each vertex x of the graph,

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

Proof. By 1.6, $|p_n(x_0,y) - p_n(x'_0,y)| \le \varepsilon_n$. By 1.9, $\frac{1}{m}\sum_{x'_0\in V} p_n(x'_0,y) = \frac{1}{m}$. Thus, $|p_n(x_0,y) - \frac{1}{m}| \le \varepsilon_n$; finally, $\varepsilon_n \to 0$ by 1.7. □

Proof of Lemma 1.7. Lemma 1.4 gives us T such that $p_T(x,y) \ne 0$ for all x,y. Clearly, $p_T(x,y) \ge k^{-T}$. Thus,

$$\mathbb{P}\big(M_T = 1\big) \ge \mathbb{P}\big(S_T = y, S'_T = y\big) \ge k^{-2T}\,,$$

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

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

however, $\theta^j \to 0$ as $j \to \infty$. □

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

2 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: for all $x, y \in V$ there exists a path from x to y (recall 1.3).

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 1.4.)

2.1 Lemma. If the Markov chain is irreducible then

$$\lim_{n\to\infty} \mathbb{P}\big(S_1 \ne y, \dots, S_n \ne y\big) = 0$$

for each $y \in V$.

Proof. We take T and ε such that for every $x \in V$ there exists a path from x to y of length $\le T$ and of probability $\ge \varepsilon$. Then (assuming $\mathbb{P}\big(S_1 \ne y, \dots, S_n \ne y\big) \ne 0$ for all n),

$$\mathbb{P}\big(S_{t+1} \ne y, \dots, S_{t+T} \ne y \,\big|\, S_t = x\big) \le 1 - \varepsilon\,.$$

Thus

$$\mathbb{P}\big(S_{t+1} \ne y, \dots, S_{t+T} \ne y \,\big|\, S_1 \ne y, \dots, S_t \ne y\big) =$$
$$= \sum_{x\in V} \mathbb{P}\big(S_{t+1} \ne y, \dots, S_{t+T} \ne y \,\big|\, S_t = x\big) \mathbb{P}\big(S_t = x \,\big|\, S_1 \ne y, \dots, S_t \ne y\big) \le$$
$$\le (1 - \varepsilon)$$

and
$$\mathbb{P}\big(S_1 \ne y, S_2 \ne y, \dots, S_{jT} \ne y\big) \le (1-\varepsilon)^j \quad \text{for } j = 1, 2, \dots$$
□

2.2 Exercise. Let y belong to each nonempty closed set. (In other words: for every x there exists a path from x to y.) Prove that
$$\lim_{n\to\infty} \mathbb{P}\big(S_1 \ne y, \dots, S_n \ne y\big) = 0\,.$$

2.3 Exercise. Let $A \subset V$ intersect every nonempty closed set. (In other words: for every x there exists a path from x to A.) Prove that
$$\lim_{n\to\infty} \mathbb{P}\big(S_1 \notin A, \dots, S_n \notin A\big) = 0\,.$$

2.4 Lemma. If the Markov chain is irreducible and aperiodic, then there exists T such that for all $x, y \in V$, every $t \ge T$ is the length of some (at least one) path from x to y. That is, $p_t(x, y) > 0$.

The proof is similar to that of 1.4. (Only the graph matters.)

We may consider two independent copies of the Markov chain:
$$V^2 = V \times V\,,$$
$$p^{(2)}\big((x_1, x_2), (y_1, y_2)\big) = p(x_1, y_1)p(x_2, y_2)\,.$$

2.5 Exercise. (a) If the Markov chain (V, p) is irreducible and aperiodic, then the Markov chain $(V^2, p^{(2)})$ is irreducible and aperiodic;

(b) it may happen that (V, p) is irreducible but $(V^2, p^{(2)})$ is not.

Prove (a) and find a counterexample for (b).

Assume that the Markov chain is irreducible and aperiodic (from now on, till Theorem 2.10).

2.6 Lemma. There exist $\varepsilon_n \to 0$ such that
$$|p_n(x_1, y) - p_n(x_2, y)| \le \varepsilon_n$$
for all $x_1, x_2, y \in V$ and $n = 0, 1, 2, \dots$

Proof. (Sketch.) We use the reflection-type argument similarly to 1.5, 1.6, 1.7. □

2.7 Exercise. For all probability measures μ on V,

$$\Big| \sum_{x_1} \mu(x_1) p_n(x_1, y) - p_n(x_2, y) \Big| \le \varepsilon_n$$

for all $x_2, y \in V$ and $n = 0, 1, \dots$

Prove it.

Hint: $| \sum_{x_1} \mu(x_1)(p_n(x_1, y) - p_n(x_2, y))|$.

Of course, $\mu(x)$ means $\mu(\{x\})$. Substituting for μ the distribution of S_t we get

$$|\mathbb{P}\big(S_{t+n} = y\big) - p_n(x_2, y)| \le \varepsilon_n$$

for all $x_2, y \in V$ and $t, n \in \{0, 1, \dots\}$.

2.8 Exercise. For all probability measures μ, ν on V,

$$\Big| \sum_{x_1} \mu(x_1) p_n(x_1, y) - \sum_{x_2} \nu(x_2) p_n(x_2, y) \Big| \le \varepsilon_n$$

for all $y \in V$ and $n = 0, 1, \dots$

Prove it.

2.9 Corollary.

$$|\mathbb{P}\big(S_t = y\big) - \mathbb{P}\big(S_u = y\big)| \le \varepsilon_n$$

for all $y \in V$ and $t, u \in \{n, n+1, \dots\}$.

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

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

exists for each $x \in V$.

Proof. By 2.9, $\big(\mathbb{P}\big(S_n = x\big)\big)_n$ is a Cauchy sequence. □

We still assume that the Markov chain is irreducible and aperiodic (from now on, till Theorem 2.15).

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

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

2.12 Exercise. The numbers

$$\mu(x) = \lim_{n\to\infty} \mathbb{P}\big(S_n = x\big)$$

are a stationary probability measure.

Prove it.

Hint: $\mathbb{P}\big(S_{n+1} = y\big) = \sum_x \mathbb{P}\big(S_n = x\big)p(x,y)$.

2.13 Exercise. $\mu(x) > 0$ for every x.

Prove it.

Hint: otherwise there exist x, y such that $\mu(x) > 0$, $\mu(y) = 0$ and $p(x,y) > 0$.

2.14 Exercise. The measure μ defined in 2.12 is the only stationary probability measure.

Prove it.

Hint: apply 2.8 to stationary μ, ν.

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

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

for every probability measure ν on V.

2.16 Exercise. Prove Theorem 2.15.

If a Markov chain (V,p) is irreducible but periodic, with the (least) period d, then the limit

$$\mu(x) = \lim_n \mathbb{P}\big(S_{nd} = x\big)$$

exists for each $x \in V$. The numbers $\mu(x)$ are a probability measure satisfying

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

That is, μ is stationary for the Markov chain (V, p_d). 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 (V,p).

Here is another property related to the graph only.

2.17 Definition. 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.*

2.18 Exercise. If x is transient then

$$\mathbb{P}\big(S_n = x\big) \to 0 \quad \text{as } n \to \infty\,.$$

Prove it.

Hint: apply 2.3 to the set A of all y such that there is no path from y to x.

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...