5 Markov chains

5a Random walk on a regular graph

Assume that a weakly connected\footnote{That is, the corresponding undirected graph is connected.} 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 $p \in \{2, 3, \ldots \}$ 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,

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

This fact is a special case of a convergence theorem for Markov chains (see 5b3).\footnote{[D, Sect. 5.4, Example 4.5; Sect. 5.5(a)]; [KS, Sect. 5.3].}

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.\footnote{[D, Sect. 5.4, (4.6) and Example 4.5].} (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.
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.\footnote{May intersect the diagonal. Multiple edges are excluded, but all said can be easily generalized to graphs with multiple edges.} Weak connectedness:

$$\emptyset \subset A \subset V \implies E \cap ((A \times (V \setminus A)) \cup ((V \setminus A) \times A)) \neq \emptyset$$

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,

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

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

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

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

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

**5a4 Corollary.** Strong connectedness:

$$\emptyset \subset A \subset V \implies E \cap (A \times (V \setminus A)) \neq \emptyset$$

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 $t \geq n$ 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 \ref{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 $2N_x^2 + kN_x + r = N_x(N_x + k) + r = N_x(N_x + k) - N_x r + (N_x + 1) r = N_x(N_x + k - r) + (N_x + 1) r \in L_x$. We take $m$ such that a path of length $\leq m$ exists from every $x$ to every $y$; then $m = m + \max_x n_x$ fits. \qed

$\{10k + 11l\} \not\in 78, 79, 89.$
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 \( \Omega' \) of paths (of length \( n \)) starting at \( x'_0 \), and random variables \( S'_0, \ldots, S'_n : \Omega' \to V \). 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

\[
M_n = \begin{cases} 
0 & \text{if } S_0 \neq S'_0, S_1 \neq S'_1, \ldots, S_n \neq S'_n, \\
1 & \text{otherwise.}
\end{cases}
\]

5a7 Exercise. \( \mathbb{E} (f(\tilde{S}_n) \mathbb{I}_{M_n=1}) = 0 \) 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), \( \mathbb{E} f(\tilde{S}_n) = \mathbb{E} (f(\tilde{S}_n) \mathbb{I}_{M_n=0}) \).

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

Proof. Take \( f(a, b) = \mathbb{I}_{\{a\}}(a) - \mathbb{I}_{\{a\}}(b) \) in 5a7. \( \square \)

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

\[
\varepsilon_n = \max_{x_0,x'_0 \in V} \mathbb{P} \left( M_n = 0 \right).
\]

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, \( \mathbb{P} (S_t = y) = p_t(x_0, y) \) and \( \mathbb{P} (S'_t = y) = p_t(x'_0, y) \).)

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, \(|p_n(x_0, y) - p_n(x_0', y)| \leq \varepsilon_n\). We average it in \(x_0'\); taking into account that \(\frac{1}{m} \sum_{x_0' \in V} p_n(x_0', y) = \frac{1}{m}\) we get \(|p_n(x_0, y) - \frac{1}{m}| \leq \varepsilon_n\); finally, \(\varepsilon_n \to 0\) by Lemma 5a9.

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

\[
\mathbb{P}(M_n = 1) \geq \mathbb{P}(S_n = y, S'_n = y) \geq k^{-2n},
\]

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

\[
\begin{align*}
\mathbb{P}(M_{t+n} = 0 | M_t = 0) & \leq \theta \quad \text{for all } t; \\
\mathbb{P}(M_{t+n} = 0) & \leq \theta \cdot \mathbb{P}(M_t = 0) \quad \text{for all } t; \\
\mathbb{P}(M_{jn} = 0) & \leq \theta^j \quad \text{for all } j
\end{align*}
\]

and, of course, \(\theta^j \to 0\) as \(j \to \infty\).

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

\[
(p(x, y))_{x, y \in V}
\]

satisfying

\[
p(x, y) \geq 0; \quad \forall x \sum_y p(x, y) = 1.
\]

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

\[
E = \{(x, y) \in V^2 : p(x, y) \neq 0\}
\]

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

\[
p(s_0, \ldots, s_n) = p(s_0, s_1) \cdots p(s_{n-1}, s_n);
\]

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 \(E \cap (A \times (V \setminus A)) = \emptyset\).  


A Markov chain is \textit{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 \textit{aperiodic}, if there exists no $p \in \{2, 3, \ldots\}$ 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.

\textbf{5b1 Theorem.} If a Markov chain is irreducible and aperiodic then the limit

$$\lim_n P(S_n = x)$$

exists for each $x \in V$.

\textbf{5b2 Definition.} A probability measure $\mu$ on $V$ is \textit{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).

\textbf{5b3 Theorem.} If a Markov chain is irreducible and aperiodic then it has one and only one stationary probability measure $\mu$, and

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

for every probability measure $\nu$ on $V$.

If a Markov chain $(V, p)$ is irreducible but periodic, with the (least) period $d$, then $V = V_1 \cup \cdots \cup V_d$ and $p_i(x, y) \neq 0$ 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 $\mu(x) = \lim_n P(S_{nd} = x)$ (assuming $x_0 \in V_1$), and the measure

$$\nu(x) = \lim_n \frac{1}{d} \left( P(S_{nd} = x) + P(S_{nd+1} = x) + \cdots + P(S_{nd+d-1} = x) \right)$$

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

Here is another property related to the graph only.

\textbf{5b4 Definition.} \footnote{Only for \textit{finite} Markov chains.} A state $x \in V$ is \textit{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 \textit{recurrent}. 

If $x$ is transient then $\mathbb{P}(S_n = x) \to 0$ 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 $T = \inf \{ n > 0 : S_n = x_0 \}$.

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

**Proof.**

\[
\exists n \forall t \quad \mathbb{P}(T \leq t + n \mid S_0, \ldots, S_t) > 0 \text{ a.s.}; \\
\exists n \exists \varepsilon \forall t \quad \mathbb{P}(T \leq t + n \mid S_0, \ldots, S_t) \geq \varepsilon \text{ a.s.}; \\
\mathbb{P}(T > t + n \mid S_0, \ldots, S_t) \leq (1 - \varepsilon) \mathbb{I}_{T > t} \text{ a.s.}; \\
\forall j \quad \mathbb{P}(T > jn) \leq (1 - \varepsilon)^j.
\]

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

“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. $\mathbb{E} \sum_{n=1}^{T} \delta_{S_n} = \left( \mathbb{E} \sum_{n=0}^{T-1} \delta_{S_n} \right) p$.

**Proof.**

\[
\mathbb{E} \left( \delta_{S_{n+1}} - \delta_{S_n} p \mid S_0, \ldots, S_n \right) = 0; \\
\mathbb{E} \left( (\delta_{S_{n+1}} - \delta_{S_n} p) \mathbb{I}_{T > n} \right) = 0;
\]

\(^1\)A priori, taking on values in $\{1, 2, \ldots\} \cup \{\infty\}$. 

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

$$
\mathbb{E} \sum_{n=0}^{\infty} (\delta_{s_{n+1}} - \delta_{s_n}p) \mathbb{I}_{T>n} = 0; \\
\mathbb{E} \sum_{n=0}^{T-1} (\delta_{s_{n+1}} - \delta_{s_n}p) = 0; \\
\mathbb{E} \sum_{n=0}^{T-1} \delta_{s_{n+1}} = \left( \mathbb{E} \sum_{n=0}^{T-1} \delta_{s_n} \right) p.
$$

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, $\mathbb{E} T = m$. 

$\square$