If the mathematics is not yet useful, it should at least be beautiful.

Albrecht Pietsch, Jena

Mathematics has advanced a bit since the time of Euclid.

Terence Tao, UCLA

Markov processes

Jan Vybíral

Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering
Czech Technical University

Contents

It is my pleasure to thank the students who provided a valuable feedback on these notes: Tobiáš Kohout, Martin Satranský.

1 Introduction: Gambler's ruin

The first part of this text follows quite closely [3]. We start the study of Markov processes by a motivating example, which will introduce the notation used, the questions studied and the methods used for the solution of these problems.

We consider the following game. Two players (for simplicity called "Player A" and "Player B") play repeatedly a certain game (like chess). Each time Player A wins, he gets one dollar from Player B and vice versa, if Player B wins, he gets one dollar from Player A. After that they repeat the game (for example, they start another chess game with a new board). The rounds played are independent of each other. In the beginning, they have a pot of S dollars. The probability, that Player A wins one round is p, $0 < p < 1$. The probability, that Player B wins, is $q = 1 - p$. If one or the other player has no dollars left (meaning that the other player has S dollars), the game stops and the player with no money left is *ruined.*

The last parameter, which we need to describe the initial state of the game, is the money Player A and Player B have in the beginning. We denote by X_0 the initial amount of money of Player A, Player B then starts with $S - X_0$ dollars. The game is then described by the input parameters (S, X_0, p, q), where $S \geq 0$ and $0 \leq X_0 \leq S$ are integers, $0 < p < 1$ and $q = 1 - p$.

To describe the whole game, it seems to be natural to denote by X_n the amount of dollars, which the Player A has after n rounds. Every time, the game (i.e., the series of the rounds) is played, X_n would have a different value, even if all the parameters (S, X_0, p, q) are the same. Therefore, X_n is a random variable, taking values in the range $\{0, 1, \ldots, S\}$. This notation allows us to reformulate the rules of the game.

- If $1 \leq k \leq S - 1$ and $n \geq 0$, then

$$\mathbb{P}(X_{n+1} = k + 1 | X_n = k) = p \quad \text{and} \quad \mathbb{P}(X_{n+1} = k + 1 | X_n = k) = q.$$

- The game stops after n rounds, if n is the smallest integer, for which $X_n = 0$ (A is ruined) or $X_n = S$ (B is ruined).

- Each series of rounds would have a different (random) length, which would lead to an unpleasant effect, that sometimes X_n would be defined and sometimes not. To avoid this confusion, we define $X_{n+1} = 0$ if $X_n = 0$ and $X_{n+1} = S$ if $X_n = S$, i.e.

$$\mathbb{P}(X_{n+1} = 0 | X_n = 0) = 1 \quad \text{and} \quad \mathbb{P}(X_{n+1} = S | X_n = S) = 1.$$

The questions, which we might be interested in, can be (very informally) divided into two groups, *local* and *global* ones. By a local question, we mean a question, which can be answered by knowing the distribution of X_n for one n (or a small number of n's). These include for example

- To find the distribution of X_n for one fixed n (i.e., what is the probability that Player A has k dollars after n rounds);

- To find the expected amount of dollars Player A has after n rounds, etc.

The global questions require the knowledge of a wide range of X_n's. They include the (arguably more interesting) questions

- What is the probability, that A gets ruined?
- What is the mean time before one of the players get ruined?
- If in the beginning Player A has more money then Player B, what is the probability that it stays so during the whole game, i.e., that $X_n \geq S - X_n$ for all $n \geq 0$?

Indeed, to answer if the game stopped after n rounds, we need to know more than just the value of X_n: If $X_n = 0$, we know that it stopped after $\ell \leq n$ rounds and if $X_n \neq 0$, we only know that it did not stop after the first n rounds.

Remark 1. Before we dive into the analysis of Gambler's ruin, let us make a remark on the independence of random variables. The random variables $X_0, X_1, X_2, \dots$ are in general *not* independent. Indeed, if we know, that $X_n = k$ for some $n \geq 0$ and $1 \leq k \leq S - 1$, then $X_{n+1} \in \{k-1, k+1\}$ and if $X_n = 0$ for some n then, necessarily, $X_{n+1} = 0$. Nevertheless, the *rounds* of the game are assumed to be independent. Therefore, we could define *independent* random variables

$$S_n = \begin{cases} +1 & \text{with probability } p \\ -1 & \text{with probability } q \end{cases}$$

to be the reward obtained by Player A in the n-th round. Then

$$X_n = X_0 + S_1 + \cdots + S_n \tag{1.1}$$

holds, but only until the game stops. Therefore, (1.1) is rather problematic - its validity is restricted to small n's and this restriction is essentially random.

1.1 Probability of ruin

We want to calculate the probability, that Player A gets ruined, if the initial parameters of the game (S, X_0, p, q) are fixed. We exclude trivial cases

- If $S = 0$ or $S = 1$, then one or both players start with zero dollars and the game does not even begin.
- If $X_0 = 0$ or $X_0 = S$, then A (or B) is ruined from the beginning - and the game again does not start at all.

We shall therefore assume that $S \geq 2$ and $1 \leq X_0 \leq S - 1$. Formally, we define the event of ruining Player A as

$$R_A = \bigcup_{n=0}^{\infty} \{X_n = 0\},$$

i.e., Player A gets ruined if (and only if) there is some $n \geq 0$ for which $X_n = 0$. We want to get $\mathbb{P}(R_A)$ as a function of (S, X_0, p, q). We start with the simplest cases, which are

$\underline{S = 2}$: Then we assume that $X_0 = 1$. In this case, the game always takes one round - if A wins the only round, B gets ruined. If A looses the only round, A gets ruined. Hence

$$R_A = \{X_1 = 0\}$$

and $\mathbb{P}(R_A) = q$.

$\underline{S = 3}$: We assume first that $X_0 = 1$. In this case, we can count all possible ways, how A can get ruined. A can loose the first game, i.e., $X_1 = 0$. Or it can win, loose, and loose again, i.e., $\{X_1 = 2, X_2 = 1, X_3 = 0\}$. The loop of winning and loosing one round after each other can be also longer and we get the decomposition

$$R_A = \bigcup_{n=0}^{\infty} \{X_1 = 2, X_2 = 1, \dots, X_{2n-1} = 2, X_{2n} = 1, X_{2n+1} = 0\} \tag{1.2}$$

and

$$\begin{aligned}\mathbb{P}(R_A) &= \sum_{n=0}^{\infty} \mathbb{P}(\{X_1 = 2, X_2 = 1, \dots, X_{2n-1} = 2, X_{2n} = 1, X_{2n+1} = 0\}) \\ &= \sum_{n=0}^{\infty} (pq)^n q = \frac{q}{1 - pq}.\end{aligned}$$

In the same way (☕), we would obtain

$$\mathbb{P}(R_A) = \frac{q^2}{1 - pq}$$

if $X_0 = 2$.

For $S \geq 4$ it seems rather infeasible (☕☕?) to enumerate all possible ways, how A can get ruined (i.e., to obtain a disjoint decomposition of R_A similar to (1.2)). Instead of that, we denote

$$f_S(k) = \mathbb{P}(R_A | X_0 = k), \quad 0 \leq k \leq S$$

and derive a system of $S + 1$ linear equations for $f_S(0), f_S(1), \dots, f_S(S)$. The first two equations are trivial: $f_S(0) = 1$ and $f_S(S) = 0$.

Lemma 1.1. *For all* $1 \leq k \leq S - 1$ *it holds*

$$\mathbb{P}(R_A | X_0 = k) = p \cdot \mathbb{P}(R_A | X_0 = k + 1) + q \cdot \mathbb{P}(R_A | X_0 = k - 1).$$

Proof. We use the independence and divide all possible ways how to ruin Player A by the first step from X_0 to X_1.

$$\begin{aligned}\mathbb{P}(R_A|X_0=k) &= \mathbb{P}(R_A, X_1=k+1|X_0=k) + \mathbb{P}(R_A, X_1=k-1|X_0=k) \\ &= \frac{\mathbb{P}(R_A, X_1=k+1, X_0=k)}{\mathbb{P}(X_0=k)} \cdot \frac{\mathbb{P}(X_1=k+1, X_0=k)}{\mathbb{P}(X_1=k+1, X_0=k)} \\ &\quad + \frac{\mathbb{P}(R_A, X_1=k-1, X_0=k)}{\mathbb{P}(X_0=k)} \cdot \frac{\mathbb{P}(X_1=k-1, X_0=k)}{\mathbb{P}(X_1=k-1, X_0=k)} \\ &= \mathbb{P}(R_A|X_1=k+1, X_0=k) \cdot \mathbb{P}(X_1=k+1|X_0=k) \\ &\quad + \mathbb{P}(R_A|X_1=k-1, X_0=k) \cdot \mathbb{P}(X_1=k-1|X_0=k) \\ &= p \cdot \mathbb{P}(R_A|X_1=k+1, X_0=k) + q \cdot \mathbb{P}(R_A|X_1=k-1, X_0=k) \\ &= p \cdot \mathbb{P}(R_A|X_0=k+1) + q \cdot \mathbb{P}(R_A|X_0=k-1).\end{aligned}$$

In the last step, we used the identity

$$\mathbb{P}(R_A|X_1=k+1, X_0=k) = \mathbb{P}(R_A|X_0=k+1),$$

which reflects the fact that the development depends only on the last state, and not on how this state was reached. We comment on this point also in Section 1.3.3. □

By Lemma 1.1, we have the following system of $S+1$ equations for $(f_S(k))_{k=0}^S$

$$f_S(0) = 1, \quad f_S(S) = 0 \tag{1.3}$$

$$f_S(k) = pf_S(k+1) + qf_S(k-1), \quad 1 \le k \le S-1. \tag{1.4}$$

We show that solving these equations leads to

$$f_S(k) = \frac{(q/p)^k - (q/p)^S}{1-(q/p)^S}, \quad 0 \le k \le S \quad \text{if} \quad p \ne q$$

and

$$f_S(k) = \frac{S-k}{S}, \quad 0 \le k \le S \quad \text{if} \quad p = q = 1/2.$$

We proceed similarly as in the solution of ordinary differential equations with boundary values. We first find two fundamental solutions of the system (1.4). Then we find a linear combination of these two solutions, which satisfies also the boundary conditions (1.3). The shortest way is to "expect" the solution in the form $f_S(k) = Ca^k$ - then (1.4) gives

$$Ca^k = f_S(k) = pf_S(k+1) + qf_S(k-1) = pCa^{k+1} + qCa^{k-1}$$

and, therefore, $a = pa^2 + q$, i.e., $0 = pa^2 - a + q = p(a-1)(a-q/p)$.

If $p \ne q$, then the two fundamental solutions are $f_S^1(k) = C_1$ and $f_S^2(k) = C_2 \cdot (q/p)^k$. Finally, we set the constants in

$$f_S(k) = C_1 + C_2\,(q/p)^k$$

to ensure that

$$1 = f_S(0) = C_1 + C_2 \quad \text{and} \quad 0 = f_S(S) = C_1 + C_2 (q/p)^S,$$

which gives

$$C_2 \cdot \left\{1 - (q/p)^S\right\} = 1, \quad C_2 = \frac{1}{1-(q/p)^S}, \quad C_1 = 1 - C_2, \quad C_1 = \frac{-(q/p)^S}{1-(q/p)^S}.$$

If $p = q = 1/2$, the general solution of (1.4) is $f_S(k) = C_1 + C_2\, k$ and the result follows.

1.2 Mean playing time

Even if the parameters of the game (S, X_0, p, q) are fixed, the game will stop after a different number of rounds. Actually, it can also happen that it never stops. Therefore we define the random variable

$$T := T_{0,S} = \inf\{n \geq 0 : X_n = 0 \text{ or } X_n = S\}.$$

Few remarks are in order:

- If the set $\{n \geq 0 : X_n = 0 \text{ or } X_n = S\}$ is empty, then (by the properties of infimum) we get $T = +\infty$.
- The range of T is therefore $\mathbb{N}_0 \cup \{+\infty\} = \{0, 1, 2, \ldots, +\infty\}$.
- The notation $T_{0,S}$ suggests, that it is the time needed to reach the set of states $\{0, S\}$, but $T_{0,S}$ depends (of course) also on X_0 and p.

Again, for S small, we can calculate the mean of $T = T_{0,S}$ directly. If $S = 2$ and $X_0 = 0$ or $X_0 = 2$, then $T_{0,2} = 0$ with probability 1 and also the mean time of each game is 0. If $X_0 = 1$, then $T_{0,2} = 1$ with probability 1 and its mean is also 1.

If $S = 3$ and $X_0 = 1$ (cases $X_0 = 0$ and $X_0 = 3$ are again trivial, the case $X_0 = 1$ is similar to $X_0 = 2$), then

$$\begin{aligned} \mathbb{P}(T_{0,3} = 2k) &= p^2 (pq)^{k-1}, \quad k \geq 1, \\ \mathbb{P}(T_{0,3} = 2k+1) &= q (pq)^k, \quad k \geq 0. \end{aligned}$$

And we can calculate directly (for $X_0 = 1$)

$$\begin{aligned} \mathbb{E}[T_{0,3}] &= \sum_{k=1}^{\infty} 2k \mathbb{P}(T_{0,3} = 2k) + \sum_{k=0}^{\infty} (2k+1) \mathbb{P}(T_{0,3} = 2k+1) \\ &= 2p^2 \sum_{k=1}^{\infty} k (pq)^{k-1} + q \sum_{k=0}^{\infty} (2k+1)(pq)^k \end{aligned}$$

Using that

$$\sum_{n=1}^{\infty} n x^{n-1} = \frac{1}{(1-x)^2}, \quad |x| < 1,$$

we get

$$
\begin{aligned}
\mathbb{E}[T_{0,3}] &= \frac{2p^2}{(1-pq)^2} + 2pq^2 \sum_{k=1}^{\infty} k(pq)^{k-1} + q \sum_{k=0}^{\infty} (pq)^k \\
&= \frac{2p^2}{(1-pq)^2} + \frac{2pq^2}{(1-pq)^2} + \frac{q}{1-pq} = \frac{2p^2 + 2pq^2 + q(1-pq)}{(1-pq)^2} \\
&= \frac{2p^2 + q + pq^2}{(1-pq)^2}.
\end{aligned}
$$

Similarly (or by interchanging p and q and X_0 and $S - X_0$), we get for $X_0 = 2$

$$\mathbb{E}[T_{0,3}] = \frac{2q^2 + p + qp^2}{(1-pq)^2}.$$

This approach seems to be again quite inconvenient (☕☕?) for $S \geq 4$.

Lemma 1.2. *For $1 \leq k \leq S-1$, it holds*

$$\mathbb{E}[T|X_0 = k] = 1 + p\,\mathbb{E}[T|X_0 = k+1] + q\,\mathbb{E}[T|X_0 = k-1].$$

Proof. We calculate

$$
\begin{aligned}
\mathbb{E}[T|X_0 = k] &= \sum_{\ell=0}^{\infty} \ell\,\mathbb{P}(T = \ell | X_0 = k) = \sum_{\ell=0}^{\infty} \ell\,\frac{\mathbb{P}(T = \ell, X_0 = k)}{\mathbb{P}(X_0 = k)} \\
&= \frac{1}{\mathbb{P}(X_0 = k)} \sum_{\ell=0}^{\infty} \ell \Big\{ \mathbb{P}(T = \ell, X_1 = k+1, X_0 = k) + \mathbb{P}(T = \ell, X_1 = k-1, X_0 = k) \Big\} \\
&= \frac{\mathbb{P}(X_1 = k+1, X_0 = k)}{\mathbb{P}(X_0 = k)} \sum_{\ell=0}^{\infty} \ell\,\frac{\mathbb{P}(T = \ell, X_1 = k+1, X_0 = k)}{\mathbb{P}(X_1 = k+1, X_0 = k)} \\
&\quad + \frac{\mathbb{P}(X_1 = k-1, X_0 = k)}{\mathbb{P}(X_0 = k)} \sum_{\ell=0}^{\infty} \ell\,\frac{\mathbb{P}(T = \ell, X_1 = k-1, X_0 = k)}{\mathbb{P}(X_1 = k-1, X_0 = k)} \\
&= \mathbb{P}(X_1 = k+1 \,|\, X_0 = k) \cdot \mathbb{E}[T = \ell | X_1 = k+1, X_0 = k] \\
&\quad + \mathbb{P}(X_1 = k-1 \,|\, X_0 = k) \cdot \mathbb{E}[T = \ell | X_1 = k-1, X_0 = k] \\
&= p \cdot \mathbb{E}[T = \ell | X_1 = k+1, X_0 = k] + q \cdot \mathbb{E}[T = \ell | X_1 = k-1, X_0 = k] \\
&= p \cdot (1 + \mathbb{E}[T = \ell | X_0 = k+1]) + q \cdot (1 + \mathbb{E}[T = \ell | X_0 = k-1]) \\
&= 1 + p\,\mathbb{E}[T|X_0 = k+1] + q\,\mathbb{E}[T|X_0 = k-1].
\end{aligned}
$$

□

We introduce again a vector of variables $(h(k))_{k=0}^{S}$

$$h(k) = \mathbb{E}\,[T|X_0 = k], \quad 0 \leq k \leq S.$$

By Lemma 1.2, we have the following equations

$$h(0) = \mathbb{E}[T|X_0 = 0] = 0, \quad h(S) = \mathbb{E}[T|X_0 = S] = 0, \tag{1.5}$$

$$h(k) = 1 + p\,h(k+1) + q\,h(k-1), \quad 1 \leq k \leq S-1. \tag{1.6}$$

The homogeneous equation corresponding to (1.6) has a solution $C_1 + C_2(q/p)^k$, one particular solution of (1.6) can be found of the form $k \to C\,k$, which leads to $C = 1/(q-p)$ (for $p \neq q$). Therefore, the general solution of (1.6) is

$$h(k) = C_1 + C_2(q/p)^k + \frac{k}{q-p}.$$

Using the boundary conditions (1.5) we get

$$h(k) = \mathbb{E}[T|X_0 = k] = \frac{1}{q-p}\left(k - S\,\frac{1-r^k}{1-r^S}\right), \quad r = q/p. \tag{1.7}$$

1.3 Additional material for Section 1

1.3.1 Measurability

1.3.2 Markov's identity

1.3.3 First step lemma - fractal approach

1.3.4 Alternative solutions

We use the relation $p + q = 1$ to reformulate (1.6) as

$$p[h(k+1) - h(k)] - q[h(k) - h(k-1)] = -1.$$

We substitute $d(k) = h(k+1) - h(k)$, $0 \leq k \leq S-1$ and obtain a system of equations

$$\sum_{k=0}^{S-1} d(k) = \sum_{k=0}^{S-1}[h(k+1) - h(k)] = h(S) - h(0) = 0$$

and

$$pd(k) - qd(k-1) = -1, \quad 1 \leq k \leq S-1.$$

We get, iteratively ($p \neq q, r = q/p$),

$$\begin{aligned}
d(1) &= -\frac{1}{p} + \frac{q}{p}d(0),\\
d(2) &= -\frac{1}{p} + \frac{q}{p}d(1) = -\frac{1}{p} - \frac{q}{p^2} + \frac{q^2}{p^2}d(0),\\
&\vdots\\
q(j) &= -\frac{1}{p}\left[1 + \frac{q}{p} + \frac{q^2}{p^2} + \cdots + \frac{q^{j-1}}{p^{j-1}}\right] + \left(\frac{q}{p}\right)^j d(0) = -\frac{1}{p}\frac{1-r^j}{1-r} + r^j\,d(0).
\end{aligned}$$

Finally, we find $d(0)$ from

$$\begin{aligned}
0 = \sum_{k=0}^{S-1} d(k) &= -\frac{1}{p(1-r)}\sum_{k=0}^{S-1}(1-r^k) + d(0)\sum_{k=0}^{S-1} r^k\\
&= -\frac{1}{p(1-q/p)}\left[S - \frac{1-r^S}{1-r}\right] + d(0)\frac{1-r^S}{1-r},
\end{aligned}$$

i.e.,

$$d(0) = h(1) - h(0) = h(1) = \frac{1}{p-q}\left[S\,\frac{1-r}{1-r^S} - 1\right],$$

which is (1.7) for $k = 1$.

2 Markov chains with discrete time

Definition 2.1. Let $X = (X_n)_{n=0}^{\infty}$ be a sequence of random variables with countable (i.e., finite or infinite countable) state space $\mathbb{S}$. Then X is called *Markov chain with discrete time* if the distribution of X_{n+1} depends only on X_n and, conditioned on X_n, does not depend on $X_0, X_1, \dots, X_{n-1}$. In other words, it satisfies the *Markov identity*

$$\mathbb{P}(X_{n+1} = j | X_n = i_n, \dots, X_1 = i_1, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i_n) \tag{2.1}$$

for all $n \in \mathbb{N}_0$ and all $i_0, i_1, \dots, i_n, j \in \mathbb{S}$.

Few remarks are in order

1. The word *chain* refers to the fact, that the state space is countable; *discrete time* denotes the fact, that the index set $n \in \mathbb{N}_0$ is discrete.

2. If in the Markov identity (2.1), the right-hand side is well-defined, i.e., if $\mathbb{P}(X_n = i_n, \dots, X_1 = i_1, X_0 = i_0) > 0$, then also $\mathbb{P}(X_n = i_n) > 0$ and also the right-hand side is well-defined. Therefore, we assume the validity of (2.1) only if $\mathbb{P}(X_n = i_n, \dots, X_1 = i_1, X_0 = i_0) > 0$ and (2.1) makes sense.

The distribution of a Markov chain is completely determined by the initial distribution of X_0 and by the rules of transition from X_n to X_{n+1}. To observe this, we calculate

$$\begin{aligned}
\mathbb{P}(X_n = i_n, \dots, X_1 = i_1, X_0 = i_0) &= \frac{\mathbb{P}(X_n = i_n, \dots, X_1 = i_1, X_0 = i_0)}{\mathbb{P}(X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0)} \cdot \\
&\cdot \frac{\mathbb{P}(X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0)}{\mathbb{P}(X_{n-2} = i_{n-2}, \dots, X_1 = i_1, X_0 = i_0)} \cdot \frac{\mathbb{P}(X_{n-2} = i_{n-2} \dots, X_1 = i_1, X_0 = i_0)}{\mathbb{P}(X_{n-3} = i_{n-3}, \dots, X_1 = i_1, X_0 = i_0)} \cdot \\
&\cdots \frac{\mathbb{P}(X_1 = i_1, X_0 = i_0)}{\mathbb{P}(X_0 = i_0)} \cdot \mathbb{P}(X_0 = i_0) \\
&= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \cdot \\
&\cdot \mathbb{P}(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \cdot \\
&\cdot \mathbb{P}(X_{n-2} = i_{n-2} | X_{n-3} = i_{n-3}, \dots, X_1 = i_1, X_0 = i_0) \cdots \mathbb{P}(X_1 = i_1 | X_0 = i_0) \cdot \mathbb{P}(X_0 = i_0) \\
&= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}) \cdot \mathbb{P}(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}) \cdot \\
&\cdot \mathbb{P}(X_{n-2} = i_{n-2} | X_{n-3} = i_{n-3}) \cdots \mathbb{P}(X_1 = i_1 | X_0 = i_0) \cdot \mathbb{P}(X_0 = i_0).
\end{aligned}$$

We therefore introduce notation for the *initial distribution* of X and for the transition rules. We assume that these rules do not depend on n, i.e., that the Markov chain is homogeneous and

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i)$$

for all $n \in \mathbb{N}_0$ and all $i, j \in \mathbb{S}$ (which we again interpret in the sense that if one of the sides is defined, the other must be defined as well and the quantities are the same).

Definition 2.2. Let $X = (X_n)_{n=0}^{\infty}$ be a homogeneous Markov chain with state space $\mathbb{S}$.

(i) Then the *transition matrix* of X is defined as

$$P = [P_{i,j}]_{i,j \in \mathbb{S}} = [\mathbb{P}(X_1 = j | X_0 = i)]_{i,j \in \mathbb{S}}.$$

(ii) The *initial distribution* of X is defined as

$$\nu_\ell = \mathbb{P}(X_0 = \ell), \quad \ell \in \mathbb{S}.$$

This allows us to rewrite the previous calculation as

$$\mathbb{P}(X_n = i_n, \dots, X_1 = i_1, X_0 = i_0) = \mathbb{P}_{i_{n-1}, i_n} \cdot \mathbb{P}_{i_{n-2}, i_{n-1}} \cdots \mathbb{P}_{i_0, i_1} \cdot \nu_{i_0}.$$

The most important property of the transition matrix is that all its row sums are equal to one. Indeed,

$$\begin{aligned}\sum_{j \in \mathbb{S}} P_{i,j} &= \sum_{j \in \mathbb{S}} \mathbb{P}(X_1 = j | X_0 = i) = \sum_{j \in \mathbb{S}} \frac{\mathbb{P}(X_1 = j, X_0 = i)}{\mathbb{P}(X_0 = i)} \\ &= \frac{1}{\mathbb{P}(X_0 = i)} \sum_{j \in \mathbb{S}} \mathbb{P}(X_1 = j, X_0 = i) = \frac{1}{\mathbb{P}(X_0 = i)} \cdot \mathbb{P}(X_0 = i) = 1.\end{aligned}$$

If $p_\ell^{(n)} = \mathbb{P}(X_n = \ell)$ is the distribution of X_n, then

$$\begin{aligned}p_k^{(n+1)} &= \mathbb{P}(X_{n+1} = k) = \sum_{j \in \mathbb{S}} \mathbb{P}(X_{n+1} = k, X_n = j) \\ &= \sum_{j \in \mathbb{S}} \mathbb{P}(X_{n+1} = k | X_n = j) \cdot \mathbb{P}(X_n = j) = \sum_{j \in \mathbb{S}} P_{j,k} p_j^{(n)} = [p^{(n)} P]_k,\end{aligned}$$

where we interpreted $p^{(n)}$ and $p^{(n+1)}$ as row vectors. This gives

$$p^{(n+1)} = p^{(n)} P.$$

We denote a state $k \in \mathbb{S}$ *absorbing* if $P_{k,k} = 1$. For gambler's ruin on $\{0, 1, \dots, S\}$, the transition matrix is given by

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ q & 0 & p & 0 & \dots & 0 \\ 0 & q & 0 & p & \dots & 0 \\ & \vdots & & & \vdots & \\ 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}.$$

In the next chapter, we shall meet the *random walk on* $\mathbb{Z}$, which is a sequence $X = (X_0, X_1, X_2, \dots)$ of random variables defined by

$$X_0 = 0, \quad X_{n+1} = \begin{cases} X_n + 1 & \text{with probability } 1/2, \\ X_n - 1 & \text{with probability } 1/2. \end{cases}$$

Its transition matrix is an infinite matrix (the state space is $\mathbb{S} = \mathbb{Z}$)

$$P = \begin{bmatrix} \dots & & \ddots & & & & \vdots & & \dots \\ \dots & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & \dots \\ \dots & & & \vdots & & & \ddots & & \dots \end{bmatrix}.$$

2.1 Iterations

We could define also higher-order transition matrices by

$$P^{(m)} = [P^{(m)}_{i,j}]_{i,j\in\mathbb{S}} = [\mathbb{P}(X_m = j|X_0 = i)]_{i,j\in\mathbb{S}}.$$

If, moreover, $p^{(n)}_j = \mathbb{P}(X_n = j)$ denotes the distribution of X_n, then we actually did prove (or we would prove in a nearly the same way as before

$$\begin{aligned} p^{(n+1)} &= p^{(n)} \cdot P = p^{(n-1)} \cdot P^2 = \dots, \\ p^{(n)} &= p^{(0)} \cdot P^n = \nu \cdot P^n, \\ p^{(n)} &= p^{(0)} \cdot P^{(n)}. \end{aligned}$$

As this holds for arbitrary $p^{(0)} = \nu$, it follows that $P^{(m)} = P^m$ for all $m \geq 1$. The notation $P^{(m)}$ is therefore not necessary, we can just take powers of P, denoted by P^m.

From what we just showed, one can easily deduce the *Chapman-Kolmogorov identity*, which states that

$$P^{(m+n)} = P^{(m)} \cdot P^{(n)} \quad \text{for all} \quad m, n \in \mathbb{N}_0.$$

As we shall encounter this identity on several other occasions as well, we give also a direct proof

$$\begin{aligned} P^{(m+n)}_{i,j} &= \mathbb{P}(X_{m+n} = j|X_0 = i) = \frac{\mathbb{P}(X_{m+n} = j, X_0 = i)}{P(X_0 = i)} \\ &= \frac{1}{P(X_0 = i)} \sum_{\ell\in\mathbb{S}} \mathbb{P}(X_{m+n} = j, X_n = \ell, X_0 = i) \\ &= \sum_{\ell\in\mathbb{S}} \frac{\mathbb{P}(X_{m+n} = j, X_n = \ell, X_0 = i)}{\mathbb{P}(X_n = \ell, X_0 = i)} \cdot \frac{\mathbb{P}(X_n = \ell, X_0 = i)}{P(X_0 = i)} \\ &= \sum_{\ell\in\mathbb{S}} \mathbb{P}(X_{m+n} = j|X_n = \ell, X_0 = i) \cdot \mathbb{P}(X_n = \ell|X_0 = i) \\ &= \sum_{\ell\in\mathbb{S}} \mathbb{P}(X_{m+n} = j|X_n = \ell) \cdot \mathbb{P}(X_n = \ell|X_0 = i) \\ &= \sum_{\ell\in\mathbb{S}} P^{(m)}_{\ell,j} P^{(n)}_{i,\ell} = [P^{(n)} \cdot P^{(m)}]_{i,j}. \end{aligned}$$

Similarly, we would prove $P^{(m+n)} = P^{(m)} \cdot P^{(n)}$. Notationally, we would complement this by setting $P^{(0)} = Id$ (the identity matrix).

2.2 Markov chains with two states

We can calculate P^n for chains with two states. This (rather straightforward) exercise has some nice lessons to learn. So, let us take $\mathbb{S} = \{0, 1\}$ and a transition matrix between these two states

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}.$$

This corresponds to a Markov chain

$$\begin{aligned} \mathbb{P}(X_{n+1} = 1 | X_n = 0) &= a, & \mathbb{P}(X_{n+1} = 0 | X_n = 0) &= 1 - a, \\ \mathbb{P}(X_{n+1} = 0 | X_n = 1) &= b, & \mathbb{P}(X_{n+1} = 1 | X_n = 1) &= 1 - b. \end{aligned}$$

For $a = b = 0$, we have $P = Id$ and $P^n = Id$. We exclude this case in what follows.

Lemma 2.3. *Let $n \in \mathbb{N}$. Then it holds*

$$P^n = \frac{1}{a+b} \begin{bmatrix} b + a(1-a-b)^n & a - a(1-a-b)^n \\ b - b(1-a-b)^n & a + b(1-a-b)^n \end{bmatrix}.$$

Proof. The proof could be done easily by mathematical induction. We would assume the formula for P^n to be true, and by $P^{n+1} = P \cdot P^n$, we would get it also for $n + 1$. It also holds for $n = 1$ and that would finish the proof.

Instead of that, we use the singular value decomposition. Using the methods of linear algebra, we see that P has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1 - a - b$, and corresponding eigenvectors $v_1 = [1, 1]^T$ and $v_2 = [-a, b]^T$. Therefore, we can write

$$P = \begin{bmatrix} 1 & -a \\ 1 & b \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \cdot \begin{bmatrix} \dfrac{b}{a+b} & \dfrac{a}{a+b} \\ -\dfrac{1}{a+b} & \dfrac{1}{a+b} \end{bmatrix} = M \times D \times M^{-1}.$$

This allows to write $P^n = (M \times D \times M^{-1})^n = M \times D^n \times M^{-1}$. Finally, we use that $D^n = \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix}$ and obtain the formula for P^n. □

The formula for P^n allows us to study the limiting behavior of X_n. We observe the following

- If $(a, b) = (0, 0)$, then $P^n = Id$ and $\lim_{n\to\infty} P^n = Id$.
- If $(a, b) = (1, 1)$, then $P = P^3 = \cdots = P^{2n+1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $P^2 = P^4 = \cdots = P^{2n} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\lim_{n\to\infty} P^n$ does not exist.

- If $-1 < 1-(a+b) < 1$, then $\lambda_2^n \to 0$ and

$$P^n \to \begin{bmatrix} 1 & -a \\ 1 & b \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ -\frac{1}{a+b} & \frac{1}{a+b} \end{bmatrix} = \frac{1}{a+b} \begin{bmatrix} b & a \\ b & a \end{bmatrix}.$$

- If $-1 < 1-(a+b) < 1$ (i.e., if $(a,b) \neq (0,0)$ and $(a,b) \neq (1,1)$), then for every initial distribution $\nu = [\gamma, 1-\gamma]$ we have after n steps the distribution of X_n equal to $\nu \cdot P^n$, which converges to

$$\nu \cdot P^n \to \nu \cdot \frac{1}{a+b} \begin{bmatrix} b & a \\ b & a \end{bmatrix} = \frac{1}{a+b} [\gamma b + (1-\gamma)b, \gamma a + (1-\gamma)a] = \left[\frac{b}{a+b}, \frac{a}{a+b}\right] =: \pi.$$

  We see, that no matter where the system starts, it always converges to the same (uniquely defined) distribution π, which satisfies $\pi = \pi \cdot P$. We call a distribution vector π *stationary* if (and only if) it satisfies $\pi = \pi \cdot P$. If the system is in this distribution after n steps, it remains that also in the step $n+1$ and, therefore, forever (i.e., for all $m \geq n$).

- If $(a,b) = (0,0)$, then every distribution ν is stationary, $\lim_{n\to\infty} \nu \cdot P^n = \nu$ exists, and it depends on the initial distribution (actually, it is equal to the initial distribution).

- If $(a,b) = (1,1)$, then $[1/2, 1/2]$ is the only stationary distribution. For every other initial distribution ν, the limit $\lim_{n\to\infty} \nu \cdot P^n$ does not exist.

2.3 Additional material

2.3.1 Stochastic matrices

Every transition matrix of a Markov chain on a given state space $\mathbb{S}$ satisfies two conditions

- $0 \leq P_{i,j} \leq 1$ for all $i, j \in \mathbb{S}$ and
- $\sum_{j \in \mathbb{S}} P_{i,j} = 1$ for all $i \in \mathbb{S}$.

On the other hand, these are the only general properties of transition matrices - this means, that every matrix satisfying these two properties can arise as a transition matrix of some Markov chain on $\mathbb{S}$. Matrices with these two properties are sometimes called *stochastic matrices*.

The set of stochastic matrices is easily seen to be convex, i.e., if P^0 and P^1 are two stochastic matrices of the same dimension, than $\lambda P^0 + (1-\lambda)P^1$ is also stochastic matrix for all $0 \leq \lambda \leq 1$. Therefore, it is a (closed) convex hull of its extreme points, which are matrices, for which each row contains one entry equal to one and the others are equal to zero.

2.3.2 Chapman-Kolmogorov identities

2.3.3 Infinite transition matrices

If $\mathbb{S}$ is infinite, say $\mathbb{S} = \mathbb{N}_0$ for simplicity, then

$$P = [P_{i,j}]_{i,j\in\mathbb{N}_0} = [\mathbb{P}(X_1 = j|X_0 = i)]_{i,j=0}^{\infty}$$

is formally an *infinite matrix*. It can be also easily mathematically formalized as an linear operator acting on ℓ_1.

Let us recall, that the space ℓ_1 is defined as the set of all infinite absolutely summable sequences, i.e.

$$\ell_1 = \left\{ x = (x_0, x_1, x_2, \dots) : \|x\|_1 = \sum_{j=0}^{\infty} |x_j| < \infty \right\}.$$

Using this notation, $\|\cdot\|_1$ is a norm and $(\ell_1, \|\cdot\|_1)$ is a Banach space. We observe that densities of random variables X with values in $\mathbb{N}_0$ can be identified with those elements of ℓ_1, which have unit norm and all entries non-negative. This means, that they lie in the unit sphere of ℓ_1 and in the cone of non-negative sequences, which is defined by

$$\mathcal{C} = \{x = (x_0, x_1, x_2, \dots) : x_j \geq 0 \text{ for all } j \in \mathbb{N}_0\}.$$

Let now $P = [P_{i,j}]_{i,j=0}^{\infty}$ be a double-indexed set of real numbers, which corresponds to some infinite transition matrix. Then $P_{i,j} \geq 0$ for all $i, j \in \mathbb{N}_0$ and $\sum_{j=0}^{\infty} P_{i,j} = 1$ for all $i \in \mathbb{N}_0$. We can now define/prove the following

- If $x \in \ell_1$ is arbitrary, we define

$$[P(x)]_\ell = [xP]_\ell = \sum_{k=0}^{\infty} x_k P_{k,l}.$$

- This series indeed converges as we see easily from

$$\left|\sum_{k=0}^{\infty} x_k P_{k,l}\right| \leq \sum_{k=0}^{\infty} |x_k| = \|x\|_1.$$

- P maps boundedly ℓ_1 into ℓ_1

$$\begin{aligned}\|P(x)\|_1 = \sum_{\ell=0}^{\infty} |[P(x)]_\ell| = \sum_{\ell=0}^{\infty} \left|\sum_{k=0}^{\infty} x_k P_{k,\ell}\right| \leq \sum_{\ell=0}^{\infty}\sum_{k=0}^{\infty} |x_k P_{k,\ell}| \\ = \sum_{k=0}^{\infty} |x_k| \sum_{\ell=0}^{\infty} |P_{k,\ell}| = \sum_{k=0}^{\infty} |x_k| = \|x\|_1.\end{aligned}$$

- If x is non-negative, the previous inequality becomes identity. Hence, P maps distributions into distributions.

3 Random walks

We now review the second example of a Markov chain with discrete time, the so-called *Random walks.* We start with a rather general definition.

Definition 3.1. Let $X_1, X_2, \dots$ be a sequence of independent random variables with values in $\mathbb{R}^d$, $d \geq 1$. Then

$$S_0 = 0 = (0, \dots, 0) \quad \text{and} \quad S_n := X_1 + \cdots + X_n \quad \text{for} \quad n \geq 1$$

is called a d-dimensional random walk with steps $X_1, X_2, \dots$.

We start with a (much) simpler case, namely when $d = 1$ and X_i's are (Rademacher) random variables with values $-1, +1$.

Definition 3.2. The simple random walk on $\mathbb{Z}$ is a random process with $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$, where $X_1, X_2, \dots$ are independent random variables with distribution $(0 < p < 1)$

$$\mathbb{P}(X_i = 1) = p \quad \text{and} \quad \mathbb{P}(X_i = -1) = q = 1 - p.$$

We call the simple random walk on $\mathbb{Z}$ *symmetric* if $p = q = 1/2$.

Some parameters of the simple random walk can be calculated directly

- $\mathbb{E}[X_n] = -1 \cdot q + 1 \cdot p = p - q,$
- $\text{Var}[X_n] = \mathbb{E}[X_n - \mathbb{E}X_n]^2 = \mathbb{E}[X_n]^2 - (\mathbb{E}X_n)^2 = 1 - (p-q)^2 = (p+q)^2 - (p-q)^2 = 4pq$
- $\mathbb{E}[S_n] = n\,\mathbb{E}[X_n] = n(p - q),$
- $\text{Var}[S_n] = n\,\text{Var}[X_1] = 4npq.$

3.1 Distribution of S_n

Using rather simple combinatorics, we can find the distribution of S_n. First, we observe that $|S_n| \leq n$ - after doing n steps of length one we can not get beyond n or below $-n$. Next observation is that the simple random walk preserves parity. Indeed, S_{2n} is always an even number and S_{2n+1} is odd.

To calculate $\mathbb{P}(S_{2n} = 2k)$ for $-n \leq k \leq n$, we count the steps of the random walk. If it does a steps to the right, b steps to the left and, after that, $S_{2n} = 2k$, then we have $a + b = 2n$ and $a - b = 2k$, which gives $a = n + k$ and $b = n - k$. In how many ways this can happen?

If $p = q = 1/2$, we can proceed by just counting all the possibilities. The random vector $(X_1, \dots, X_{2n}) \in \{-1, +1\}^{2n}$ has 2^{2n} possible values. If there should be $n+k$ ones and $n-k$ minus ones in this vector, we have $\binom{2n}{n+k}$ possibilities and

$$\mathbb{P}(S_{2n} = 2k) = 2^{-2n} \cdot \binom{2n}{n+k}.$$

If $p \neq q$, then the 2^{2n} possible outcomes of $(X_1, \ldots, X_{2n})$ do not have the same probability. We fix the positions of $n+k$ ones (the remaining indices will be occupied by -1) and obtain

$$\mathbb{P}(S_{2n} = 2k) = \binom{2n}{n+k} p^{n+k} q^{n-k}, \quad -n \le k \le n.$$

Similarly, we obtain

$$\mathbb{P}(S_{2n+1} = 2k+1) = \binom{2n+1}{n+k+1} p^{n+k+1} q^{n-k}, \quad -n \le k \le n.$$

With this, we essentially characterized the *local* properties of a random walk.

3.2 First return to the origin

We shall again study some of the *global* characteristics - namely (again) the first time when certain states are reached. This time we will be interested in the first time of return to the origin, i.e., of the first non-zero $n \in \mathbb{N}$ with $S_n = 0$. We define

$$T_0^r := \inf\{n \in \mathbb{N} : S_n = 0\}.$$

If there is no positive integer n with $X_n = 0$, then $T_0^r = \infty$. We want to calculate the distribution and the mean value (and possibly also variance and other parameters) of T_0^r. We put $g(n) = \mathbb{P}(T_0^r = n)$. Simple cases include $g(2n+1) = \mathbb{P}(T_0^r = 2n+1) = 0$ for every $n \in \mathbb{N}_0$, $g(2) = \mathbb{P}(T_0^r = 2) = 2pq, g(4) = \mathbb{P}(T_0^r = 4) = 2p^2q^2$.

We derive again a system of equations for $g(n)$. Let us denote

$$h(2n) = \mathbb{P}(S_{2n} = 0) = \binom{2n}{n} p^n q^n$$

as already calculated before. Naturally, $g(2n) \le h(2n)$.

Lemma 3.3. *For* $n \ge 1$, *the* convolution identity *holds*

$$h(n) = \sum_{k=0}^{n-2} g(n-k)h(k).$$

Together with $g(1) = 0, g(2) = 2pq, g(3) = 0$ and $g(4) = 2p^2q^2$, Lemma 3.3 applied to $n = 1, \ldots, N$ gives a system of equations for $(g(n))_{n=1}^N$. Also note that adding the terms with $k = n-1$ and $k = n$ means adding $g(1)h(n-1) = 0$ and $g(0)h(n) = 0$, i.e., we could rewrite the identity also as

$$h(n) = \sum_{k=0}^{n} g(n-k)h(k).$$

Proof. We decompose the event $S_n = 0$ by the last-but-one return to zero, i.e.,

$$\{S_n = 0\} = \bigcup_{k=0}^{n-2} \{S_k = 0, S_{k+1} \neq 0, \ldots, S_{n-1} \neq 0, S_n = 0\}.$$

This decomposition is disjoint and gives

$$\begin{aligned}
h(n) = \mathbb{P}(S_n = 0) &= \sum_{k=0}^{n-2} \mathbb{P}(S_k = 0, S_{k+1} \neq 0, \ldots, S_{n-1} \neq 0, S_n = 0) \\
&= \sum_{k=0}^{n-2} \mathbb{P}(S_{k+1} \neq 0, \ldots, S_{n-1} \neq 0, S_n = 0 | S_k = 0) \cdot \mathbb{P}(S_k = 0) \\
&= \sum_{k=0}^{n-2} \mathbb{P}(S'_1 \neq 0, \ldots, S'_{n-k-1} \neq 0, S'_{n-k} = 0 | S'_0 = 0) \cdot \mathbb{P}(S_k = 0) \\
&= \sum_{k=0}^{n-2} \mathbb{P}(T_0^r = n-k) \cdot \mathbb{P}(S_k = 0) = \sum_{k=0}^{n-2} h(k) g(n-k).
\end{aligned}$$

In this calculation, we defined $S'_m = S_{m+k}$, which is again a random walk under the condition that $S_k = 0$. □

We can solve the convolution identity using the method of the *generating function*. We define (whenever it converges, but $0 \leq g(n) \leq 1$ and $0 \leq h(n) \leq 1$ ensure that it holds at least for $|s| < 1$)

$$\begin{aligned}
G(s) &:= \sum_{n=0}^{\infty} s^n g(n), \\
H(s) &:= \sum_{k=0}^{\infty} s^k h(k).
\end{aligned}$$

Lemma 3.4. *These functions satisfy* $H(s) = (1 - 4pqs^2)^{-1/2}$ *and* $G(s)H(s) = H(s) - 1$ *for* $4pqs^2 < 1$.

Proof. We calculate

$$\begin{aligned}
H(s) &= \sum_{k=0}^{\infty} s^k \mathbb{P}(S_k = 0) = \sum_{k=0}^{\infty} s^{2k} \mathbb{P}(S_{2k} = 0) = \sum_{k=0}^{\infty} s^{2k} \binom{2k}{k} p^k q^k \\
&= \sum_{k=0}^{\infty} (pqs^2)^k \frac{(2k)(2k-1)\ldots 3 \cdot 2 \cdot 1}{[k!]^2} = \sum_{k=0}^{\infty} (pqs^2)^k 2^{2k} \frac{k(k-1/2)\ldots 3/2 \cdot 1 \cdot 1/2}{[k!]^2} \\
&= \sum_{k=0}^{\infty} (4pqs^2)^k \frac{(k-1/2) \cdot (k-3/2) \ldots 3/2 \cdot 1/2}{k!} \\
&= \sum_{k=0}^{\infty} (-4pqs^2)^k \frac{(1/2-k) \cdot (3/2-k) \ldots (-3/2) \cdot (-1/2)}{k!} = \frac{1}{\sqrt{1 - 4pqs^2}},
\end{aligned}$$

where we used the identity

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n = \sum_{n=0}^{\infty} \frac{\alpha \cdot (\alpha-1) \cdots (\alpha-n+1)}{n!} x^n \quad \text{for} \quad |x| < 1.$$

For the next step, we use the Cauchy-product of two power series and the convolution identity

$$\begin{aligned}
G(s)H(s) &= \left(\sum_{n=1}^{\infty} s^n g(n)\right)\left(\sum_{k=0}^{\infty} s^k h(k)\right) = \sum_{n=1}^{\infty}\sum_{k=0}^{\infty} s^{n+k} g(n)h(k) \\
&= \sum_{k=0}^{\infty}\sum_{n=2}^{\infty} s^{n+k} g(n)h(k) \underset{k \text{ fixed}, \ell=n+k}{=} \sum_{k=0}^{\infty}\sum_{\ell=k+2}^{\infty} s^{\ell} g(\ell-k)h(k) \\
&= \sum_{\ell=2}^{\infty} s^{\ell} \sum_{k=0}^{\ell-2} g(\ell-k)h(k) = \sum_{\ell=2}^{\infty} s^{\ell} h(\ell) = \sum_{\ell=0}^{\infty} s^{\ell} h(\ell) - 1 = H(s) - 1.
\end{aligned}$$

□

Next we decompose $G(s) = 1 - \dfrac{1}{H(s)} = 1 - \sqrt{1-4pqs^2}$, $4pqs^2 < 1$, into a power series and its coefficients will be identified as $g(n)$'s.

$$\begin{aligned}
G(s) = 1 - [1-4pqs^2]^{1/2} &= 1 - \sum_{k=0}^{\infty} \frac{1}{k!}(-4pqs^2)^k \cdot (1/2-0)\cdot(1/2-1)\cdot \dots \cdot (1/2-(k-1)) \\
&= -\sum_{k=1}^{\infty} \frac{1}{k!}(-4pqs^2)^k \cdot (1/2-0)\cdot(1/2-1)\cdot \dots \cdot (1/2-(k-1)) \\
&= -\sum_{k=1}^{\infty} \frac{1}{k!}(4pqs^2)^k \cdot (0-1/2)\cdot(1-1/2)\cdot \dots \cdot ((k-1)-1/2) \\
&= \frac{1}{2}\sum_{k=1}^{\infty} \frac{1}{k!} 2^{2k} p^k q^k s^{2k} \cdot (1-1/2)\cdot \dots \cdot ((k-1)-1/2) = \sum_{n=0}^{\infty} s^n g(n).
\end{aligned}$$

We see (and we actually knew that from the very beginning) that $g(2k+1) = 0$ for all $k \in \mathbb{N}$. And (using a similar idea only the other way round as before)

$$\begin{aligned}
g(2k) &= \frac{1}{2}\cdot\frac{(4pq)^k}{k!}(1-1/2)\cdot \dots \cdot ((k-1)-1/2) \\
&= (pq)^k \frac{2^{2k}}{k!} \cdot \frac{(1-1/2)\cdot 1\cdot(2-1/2)\cdot 2 \cdot \dots \cdot ((k-1)-1/2)\cdot(k-1)(k-1/2)\cdot k}{k!} \cdot \frac{1}{2(k-1/2)} \\
&= (pq)^k \frac{(2k)!}{[k!]^2} \cdot \frac{1}{2k-1} = \frac{1}{2k-1}\binom{2k}{k} p^k q^k = \frac{h(2k)}{2k-1}.
\end{aligned}$$

We can now calculate the probability of return to zero in finite time. Indeed, if $p \neq q$, we get $4pqs^2 < 1$ for $s = 1$ and

$$\mathbb{P}(T_0^r < \infty) = \sum_{k=0}^{\infty} \mathbb{P}(T_0^r = k) = \sum_{k=0}^{\infty} g(k) = G(1) = 1 - (1-4pq)^{1/2} = 1 - |p-q| = 2\min(p,q).$$

For $p = q = 1/2$, this follows from Abel's theorem for power series (details left to reader). To summarize, for $p = q = 1/2$, we have $\mathbb{P}(T_0^r < \infty) = 1$ and $\mathbb{P}(T_0^r = \infty) = 0$. For $p \neq q$, both these probabilities are strictly between 0 and 1.

The mean of the first return time to the origin is infinite as $\mathbb{P}(T_0^r = \infty) > 0$ and

$$\mathbb{E}[T_0^r] = \sum_{k=0}^{\infty} k \cdot \mathbb{P}(T_0^r = k) + (+\infty) \cdot \mathbb{P}(T_0^r = \infty) = +\infty.$$

If $p = q = 1/2$, then $\mathbb{P}(T_0^r = \infty) = 0$ and

$$\mathbb{E}[T_0^r] = \sum_{k=0}^{\infty} k \cdot \mathbb{P}(T_0^r = k) = \sum_{k=0}^{\infty} 2k \cdot g(2k) = \sum_{k=1}^{\infty} \frac{2k}{2k-1} \binom{2k}{k} \cdot 2^{-2k}.$$

This series diverges (which can be found - for example - by Stirling's formula $n! \approx \sqrt{2\pi n}(n/e)^n$). (☕) Is there a simpler way to deduce this? Consequently, $\mathbb{E}[T_0^r] = +\infty$ although $\mathbb{P}(T_0^r < +\infty) = 1$. This means that T_0^r is almost surely finite, but with an infinite mean value.

If $p \neq q$, we can still try to compute $\mathbb{E}[T_0^r | T_0^r < +\infty]$, i.e., the mean of those trajectories which have finite length.

$$\begin{aligned}
\mathbb{E}[T_0^r | T_0^r < \infty] &= \sum_{n=0}^{\infty} \frac{n \cdot \mathbb{P}(T_0^r = n)}{\mathbb{P}(T_0^r < \infty)} = \frac{1}{\mathbb{P}(T_0^r < \infty)} \sum_{n=1} ng(n) \\
&= \frac{1}{2\min(p,q)} \cdot \sum_{n=1}^{\infty} ns^{n-1} g(n) \Big|_{s=1} = \frac{G'(1)}{2\min(p,q)} \\
&= \frac{1}{2\min(p,q)} \cdot \frac{4pqs}{\sqrt{1-4pqs^2}} \Big|_{s=1} = \frac{2\max(p,q)}{|p-q|}.
\end{aligned}$$

3.3 Sailor's parrot

We now consider the *simple random walk in* $\mathbb{Z}^d$, which is defined by its (independent) steps

$$\mathbb{P}(X_j = e_k) = \mathbb{P}(X_j = -e_k) = \frac{1}{2d}.$$

Here, $e_k, k = 1, \dots, d$ are the canonical vectors in $\mathbb{R}^d$, which have 1 at the kth coordinate and zero elsewhere. As usually, $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. We plan to study its *long-time behavior*, i.e., if and how many times it returns to zero, We define

- N: the number of appearances at zero

$$N := \#\{n \in \mathbb{N}_0 : S_n = 0\} \geq 1.$$

- τ: the first time of return to zero

$$\tau := \inf\{n \geq 1 : S_n = 0\}.$$

  Again, if $S_n \neq 0$ for all $n \geq 1$, we put $\tau = +\infty$.

Observe that $N = 1$ and $\tau = +\infty$ both describe the fact, that the random walk never returned to zero., i.e., $\{N = 1\} = \{\tau = +\infty\}$.

Definition 3.5. We call the random walk *recurrent*, if $\mathbb{P}(N = +\infty) = 1$, i.e., if it almost surely returns to 0 infinitely many times. We call the random walk *transient*, if $\mathbb{P}(N < +\infty) = 1$, i.e., if the random walk almost surely returns to zero finitely many times and then it stops to appear that forever.

In general, it is not clear, that nothing between $\mathbb{P}(N = +\infty) = 1$ (recurrent) and $\mathbb{P}(N = +\infty) = 0$ (transient) exists. It will follow later and requires a proof.

Lemma 3.6. *For $n \geq 1$ it holds that $\mathbb{P}(N = n) = \mathbb{P}(\tau = +\infty) \cdot \mathbb{P}(\tau < \infty)^{n-1}$.*

Proof. First, we show that

$$\mathbb{P}(N = n + 1) = \mathbb{P}(N = n) \cdot \mathbb{P}(\tau < +\infty). \tag{3.1}$$

We shall exploit that by the return to zero at time $k \geq 2$, the random walk splits into two parts. The first one is $(S_n)_{n=0}^{k}$ and has $S_0 = 0 = S_k$, the other one is the infinite part $(S_n)_{n=k}^{\infty} = (S_{k+j})_{j=0}^{\infty}$. Conditioned on the event $S_k = 0$, these two parts are independent. We denote the second part by S', i.e., $S'_j = S_{k+j}$. The symbol N' is defined similarly as N, but corresponding to S', i.e., $N' = \#\{j \geq 0 : S_{k+j} = 0\} = \#\{j \geq 0 : S'_j = 0\} = N - 1$.

If $N = n + 1 \geq 2$, then (at least) one return to zero appeared, and τ is finite. We therefore obtain

$$\begin{aligned}
\mathbb{P}(N = n + 1) &= \sum_{k=1}^{\infty} \mathbb{P}(N = n + 1, \tau = k) = \sum_{k=1}^{\infty} \mathbb{P}(N' = n, \tau = k)\\
&= \sum_{k=1}^{\infty} \mathbb{P}(N' = n) \cdot \mathbb{P}(\tau = k) = \mathbb{P}(N' = n) \sum_{k=1}^{\infty} \mathbb{P}(\tau = k)\\
&= \mathbb{P}(N = n) \cdot \mathbb{P}(\tau < +\infty)
\end{aligned}$$

as N and N' are equidistributed. This finishes the proof of (3.1).

The proof of the lemma is then finished by induction. Indeed, $n = 1$ is clear from $\mathbb{P}(N = 1) = \mathbb{P}(\tau = +\infty)$. Observe here, that $\mathbb{P}(\tau < \infty)$ is always positive - we can always make one step in some direction and then return back already in the second step. Also the induction step is easy:

$$\begin{aligned}
\mathbb{P}(N = n + 1) &= \mathbb{P}(N = n) \cdot \mathbb{P}(\tau < \infty) = \mathbb{P}(\tau = +\infty) \cdot \mathbb{P}(\tau < \infty)^{n-1} \cdot \mathbb{P}(\tau < \infty)\\
&= \mathbb{P}(\tau = +\infty) \cdot \mathbb{P}(\tau < \infty)^n.
\end{aligned}$$

□

As a corollary, we obtain that every random walk (in every dimension) is either recurrent, or transient.

Corollary 3.7. *Every random walk is either recurrent, or transient.*

Proof. If $\mathbb{P}(\tau = +\infty) = 0$, then (by the previous lemma) $\mathbb{P}(N = n) = 0$ for all $n \geq 1$ and

$$\mathbb{P}(N < +\infty) = \sum_{n=1}^{\infty} \mathbb{P}(N = n) = 0,$$

i.e., $P(N=+\infty)=1$ and the random walk is recurrent.

If, on the other hand, $\mathbb{P}(\tau=+\infty)>0$, then (using that $\mathbb{P}(\tau<+\infty)<1$)

$$\begin{aligned}\mathbb{P}(N<+\infty) &= \sum_{n=1}^{\infty}\mathbb{P}(N=n)=\sum_{n=1}^{\infty}\mathbb{P}(\tau=+\infty)\cdot\mathbb{P}(\tau<\infty)^{n-1}\\ &= \mathbb{P}(\tau=+\infty)\cdot\frac{1}{1-\mathbb{P}(\tau<+\infty)}=1,\end{aligned}$$

and the random walk is transient. □

There is a "simple" criterion, which can help us to decide in which dimension the simple random walk is recurrent and when it is transient.

Lemma 3.8. *The simple random walk in $\mathbb{Z}^d$ is transient if, and only if $\mathbb{E}[N]<\infty$ and it is recurrent if, and only if, $\mathbb{E}[N]=+\infty$.*

Roughly speaking, being transient and N having finite mean both tell, that we do not return to the origin too often. But it is not clear (but it follows from this lemma), that these two statements are actually equivalent.

Proof. By definition, we have

$$\mathbb{E}[N]=\sum_{n=1}^{\infty}n\cdot\mathbb{P}(N=n)+\infty\cdot\mathbb{P}(N=+\infty),$$

where we (as usually) define $0\cdot\infty=0$.

If $E[N]$ is finite, then necessarily $\mathbb{P}(N=+\infty)=0$ and the random walk is transient. On the other hand, if $\mathbb{P}(N=+\infty)=0$, then

$$\begin{aligned}\mathbb{E}[N] &= \sum_{n=1}^{\infty}n\cdot\mathbb{P}(N=n)=\sum_{n=1}^{\infty}n\cdot\mathbb{P}(\tau=+\infty)\cdot\mathbb{P}(\tau<\infty)^{n-1}\\ &= \mathbb{P}(\tau=+\infty)\sum_{n=1}^{\infty}n\cdot\mathbb{P}(\tau<\infty)^{n-1}=\frac{\mathbb{P}(\tau=+\infty)}{[1-\mathbb{P}(\tau<+\infty)]^2}=\frac{1}{\mathbb{P}(\tau=+\infty)}<+\infty.\end{aligned}$$

In this calculation, we used (actually two times) that if $\mathbb{P}(N=+\infty)=0$, then $\mathbb{P}(N<+\infty)=1$ and by the previous lemma also $\mathbb{P}(\tau=+\infty)>0$, i.e., $\mathbb{P}(\tau<+\infty)<1$.

We have shown that the random walk is transient if, and only if, $\mathbb{E}[N]<+\infty$. By simple logic, also the opposites of these two statements are equivalent, which together with the previous corollary finishes the proof. □

Lemma 3.9. *If the simple random walk in $\mathbb{Z}^d$ has steps $X_1,X_2,\dots$ and if we denote $\varphi(\alpha)=\mathbb{E}[e^{i\alpha\cdot X_1}]=\mathbb{E}[\cos(\alpha\cdot X_1)+i\sin(\alpha\cdot X_1)]$ for $\alpha\in\mathbb{R}^d$, then*

$$\mathbb{E}[N]=\lim_{t\to 1^-}\int_{[-\pi,\pi]^d}\frac{1}{1-t\varphi(\xi)}\cdot\frac{d\xi}{(2\pi)^d}.$$

Proof. For $d = 1$, we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{im\theta} d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\cos(m\theta) + i\sin(m\theta))d\theta = \begin{cases} 1, & \text{if } m = 0, \\ 0, & \text{if } m \in \mathbb{Z} \setminus \{0\}. \end{cases}$$

For general $d \geq 2$ we get

$$\frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} e^{im\cdot\xi} d\xi = \prod_{j=1}^{d} \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{im_j\xi_j} d\xi_j = \begin{cases} 1, & \text{if } m = (0,\dots,0), \\ 0, & \text{if } m \in \mathbb{Z}^d \setminus \{(0,\dots,0)\}. \end{cases}$$

We now plug in for m the value of the random walk after n steps, i.e., we put $m := S_n$ and obtain

$$\frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} e^{iS_n\cdot\xi} d\xi = \begin{cases} 1, & \text{if } S_n = (0,\dots,0), \\ 0, & \text{if } S_n \in \mathbb{Z}^d \setminus \{(0,\dots,0)\}. \end{cases}$$

Let χ be the indicator random variable of the set $S_n = 0$, i.e., $\chi = 1$ if $S_n = 0$ and $\chi = 0$ otherwise. Then we interchange the order of integration (we have a bounded function on a finite measure space)

$$\begin{aligned} \mathbb{P}(S_n = 0) = \mathbb{E}[\chi] = \mathbb{E}\left[\frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} e^{iS_n\cdot\xi} d\xi\right] &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} \mathbb{E}[e^{iS_n\cdot\xi}] d\xi \\ &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} \mathbb{E}[e^{iX_1\cdot\xi + iX_2\cdot\xi + \dots + iX_n\cdot\xi}] d\xi \\ &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} \mathbb{E}[e^{iX_1\cdot\xi} \cdot e^{iX_2\cdot\xi} \cdot \dots \cdot e^{iX_n\cdot\xi}] d\xi \\ &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} \mathbb{E}[e^{iX_1\cdot\xi}] \cdot \mathbb{E}[e^{iX_2\cdot\xi}] \cdot \dots \cdot \mathbb{E}[e^{iX_n\cdot\xi}] d\xi \\ &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} \varphi(\xi)^n d\xi. \end{aligned}$$

We fix $t \in [0,1)$, multiply this calculation by t^n and sum up (which means that we actually produce the generating function) and obtain

$$\begin{aligned} \sum_{n=0}^{\infty} t^n \mathbb{P}(S_n = 0) &= \sum_{n=0}^{\infty} \frac{t^n}{(2\pi)^d}\int_{[-\pi,\pi]^d} \varphi(\xi)^n d\xi \\ &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} \sum_{n=0}^{\infty} [t\varphi(\xi)]^n d\xi = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} \frac{1}{1 - t\varphi(\xi)} d\xi, \end{aligned}$$

where we used that $|t\varphi(\xi)| = |t\mathbb{E}[e^{i\xi\cdot X_1}]| \leq t\mathbb{E}|e^{i\xi\cdot X_1}| = t < 1$.

If now $t \to 1^-$, then the left-hand side converges to

$$\sum_{n=0}^{\infty} \mathbb{P}(S_n = 0) = \sum_{n=0}^{\infty} \mathbb{E}[\chi_{\{S_n = 0\}}] = \mathbb{E}\left(\sum_{n=0}^{\infty} \chi_{\{S_n = 0\}}\right) = \mathbb{E}[N].$$

□

Finally, we can show that the simple random walk is recurrent for some d's and transient for the others. The theorem is sometimes interpreted as the following saying:

A drunken man will always find his way home but his drunken parrot may get lost forever.

Theorem 3.10. *The simple random walk is recurrent for $d = 1$ and $d = 2$ and is transient for $d \geq 3$.*

Proof. For simple random walk in $\mathbb{Z}^d$, we have $\mathbb{P}(X_1 = \pm e_k) = \frac{1}{2d}$ and, therefore,

$$\begin{aligned}\varphi(\alpha) = \mathbb{E}[e^{i\alpha\cdot X}] &= \frac{e^{i\alpha_1} + e^{-i\alpha_1} + e^{i\alpha_2} + e^{-i\alpha_2} + \cdots + e^{i\alpha_d} + e^{-i\alpha_d}}{2d} \\ &= \frac{\cos(\alpha_1) + \cdots + \cos(\alpha_d)}{d}.\end{aligned}$$

We see, that the only point where $1 - t\varphi(\xi) \to 0$ for $\xi \in [-\pi, \pi]^d$ and $t \to 1^-$ is $\xi = 0$ (where $\varphi(\xi) = 1$).

We use some elementary analysis to show that

$$\lim_{t\to 1^-} \int_{[-\pi,\pi]^d} \frac{d\xi}{1 - t\varphi(\xi)}$$

is infinite if $d = 1, 2$ and finite for $d \geq 3$. We can reduce the domain of integration from $[-\pi, \pi]^d$ to (say) $[-\pi/2, \pi/2]^d$ as the function is bounded elsewhere.

First, we rewrite $1 - \cos(x) = 2\sin^2(x/2)$ and use the estimate $2x/\pi \leq \sin x \leq x$ for $x \in [0, \pi/2]$. Hence, $[x/\pi]^2 \leq [\sin(x/2)]^2 \leq [x/2]^2$ and $2x^2/\pi^2 \leq 1 - \cos(x) \leq x^2/2$. Finally,

$$1 - t\varphi(\xi) = 1 - t + t\cdot \frac{(1 - \cos(\xi_1)) + \cdots + (1 - \cos(\xi_d))}{d} \begin{cases} \geq 1 - t + \dfrac{t}{d}\cdot\dfrac{2\|\xi\|_2^2}{\pi^2}, \\ \\ \leq 1 - t + \dfrac{t}{2d}\cdot\|\xi\|_2^2. \end{cases}$$

If $d = 1, 2$, we therefore get

$$\frac{1}{1 - t\varphi(\xi)} \geq \frac{1}{1 - t + \frac{t}{2d}\cdot\|\xi\|_2^2} \geq \frac{1}{1 - t + \frac{1}{2d}\cdot\|\xi\|_2^2} \nearrow \frac{2d}{\|\xi\|_2^2}.$$

And for $d \geq 3$ we obtain

$$\frac{1}{1 - t\varphi(\xi)} \leq \frac{1}{1 - t + \frac{t}{d}\cdot\frac{2\|\xi\|_2^2}{\pi^2}} \leq \frac{\pi^2 d}{2t\|\xi\|_2^2}.$$

Finally, we use that $\int \|\xi\|_2^{-2} d\xi$ does not converge around zero for $d = 1, 2$ and converges at the origin for $d \geq 3$.

□

3.4 Additional material

3.5 Details to the convolution identity

3.5.1 Elementary proof of $g(2k) = h(2k)/(2k - 1)$

4 First step analysis

We wish to elaborate on the method of the analysis of the first step of a Markov chain and make it a powerful general approach to solve a number of typical problems.

4.1 Probability of hitting a subset

Let $X = (X_n)_{n=0}^{\infty}$ be a Markov chain with the state space $\mathbb{S}$. Let $A \subset \mathbb{S}$. We denote by T_A the first time, when X reaches A, i.e.,

$$T_A = \inf\{n \in \mathbb{N}_0 : X_n \in A\}.$$

If $X_0 = k$ and $k \in A$, then $T_A = 0$. If, on the other hand, the chain never reaches A, i.e., if the set $\{n \in \mathbb{N}_0 : X_n \in A\}$ is empty, then we define $T_A = +\infty$.

First, we discuss the probability that the set A is ever reached if we start in $\ell \in \mathbb{S}$, i.e., the (conditional) probability that $T_A < +\infty$ if $X_0 = \ell$. We denote this quantity by $t(\ell) := \mathbb{P}(T_A < +\infty | X_0 = \ell)$. If $\ell \in A$, then $T_A = 0$ and $t(\ell) = \mathbb{P}(T_A < +\infty | X_0 = \ell) = 1$. If $\ell \notin A$, then we calculate

$$\begin{aligned}
t(\ell) = \mathbb{P}(T_A < +\infty | X_0 = \ell) &= \sum_{m \in \mathbb{S}} \mathbb{P}(T_A < +\infty, X_1 = m | X_0 = \ell) \\
&= \sum_{m \in \mathbb{S}} \frac{\mathbb{P}(T_A < +\infty, X_1 = m, X_0 = \ell)}{\mathbb{P}(X_1 = m, X_0 = \ell)} \cdot \frac{\mathbb{P}(X_1 = m, X_0 = \ell)}{\mathbb{P}(X_0 = \ell)} \\
&= \sum_{m \in \mathbb{S}} \mathbb{P}(T_A < +\infty | X_1 = m, X_0 = \ell) \cdot P_{\ell,m} = \sum_{m \in \mathbb{S}} P_{\ell,m} t(m) \\
&= \sum_{m \in A} P_{\ell,m} + \sum_{m \notin A} P_{\ell,m} t(m),
\end{aligned}$$

which gives a system of linear equations for $(t(\ell))_{\ell \in \mathbb{S}}$ or, better said, for $(t(\ell))_{\ell \notin A}$.

In a similar way, we can investigate, which is the first state from A, which gets reached by X. We therefore define

$$g_\ell(k) := \mathbb{P}(T_A < +\infty, X_{T_A} = \ell | X_0 = k), \quad \ell \in A.$$

This is the probability, that A is reached and - when it is reached - X enters A through ℓ. Again, if $k \in A$, the question is simple and we get

$$g_\ell(k) = \begin{cases} 1 & \text{if } k = \ell \in A, \\ 0 & \text{if } k \in A, k \neq \ell. \end{cases} \tag{4.1}$$

If $k \in \mathbb{S} \setminus A$, then we need at least one step to reach A. Therefore

$$\begin{aligned}
g_\ell(k) = \mathbb{P}(T_A < +\infty, X_{T_A} = \ell | X_0 = k) &= \sum_{m \in \mathbb{S}} \mathbb{P}(T_A < +\infty, X_{T_A} = \ell, X_1 = m | X_0 = k) \\
&= \sum_{m \in \mathbb{S}} \mathbb{P}(T_A < +\infty, X_{T_A} = \ell | X_1 = m, X_0 = k) \mathbb{P}(X_1 = m | X_0 = k) \qquad (4.2) \\
&= \sum_{m \in \mathbb{S}} P_{k,m} g_\ell(m) \quad \text{for} \quad \ell \in A, k \notin A.
\end{aligned}$$

Hence, for every $\ell \in A$, the vector $g_\ell = (g_\ell(k))_{k\in\mathbb{S}}$ satisfies the previous two sets of equations (4.1) and (4.2). The solution of this system depends of course on the transition matrix P.

Furthermore, we get (for all $k \in \mathbb{S}$)

$$\begin{aligned} 1 &= \mathbb{P}(T_A = +\infty | X_0 = k) + \sum_{\ell\in A} \mathbb{P}(T_A < +\infty, X_{T_A} = \ell | X_0 = k) \\ &= \mathbb{P}(T_A = +\infty | X_0 = k) + \sum_{\ell\in A} g_\ell(k). \end{aligned}$$

4.2 Mean time of absorption

Next, we shall discuss, how long it takes until we reach $A \subset \mathbb{S}$. We denote

$$h(k) = h_A(k) = \mathbb{E}[T_A | X_0 = k].$$

If $k \in A$, then $T_A = 0$ and $h_A(k) = 0$. If $k \notin A$, then we have to make at least one step before reaching A. Therefore,

$$\begin{aligned} h_A(k) = \mathbb{E}[T_A|X_0 = k] &= \sum_{\ell\in\mathbb{S}} \mathbb{P}(X_1 = \ell|X_0 = k) \cdot [1 + \mathbb{E}[T_A|X_0 = \ell]] \\ &= \sum_{\ell\in\mathbb{S}} \mathbb{P}(X_1 = \ell|X_0 = k) + \sum_{\ell\in\mathbb{S}} \mathbb{P}(X_1 = \ell|X_0 = k) \cdot \mathbb{E}[T_A|X_0 = \ell] \\ &= 1 + \sum_{\ell\in\mathbb{S}} P_{k,\ell} h_A(\ell). \end{aligned}$$

This gives the system of equations

$$\begin{aligned} h_A(k) &= 1 + \sum_{\ell\in\mathbb{S}} P_{k,\ell} h_A(\ell), \quad k \in \mathbb{S} \setminus A, \\ h_A(\ell) &= 0, \quad \ell \in A. \end{aligned}$$

4.3 Time of the first return

Again, if $j \in \mathbb{S}$, we define the first non-zero time, when j is reached

$$T_j^r = \inf\{n \geq 1 : X_n = j\}.$$

This is again a random variable with values in $\{1, 2, \dots\} \cup \{+\infty\}$. If $X_0 \neq j$, then $T_j^r = T_{\{j\}}$ in the sense of the previous section. Now we denote (and we keep(!) this notation for later use as well)

$$\mu_j(i) := \mathbb{E}[T_j^r | X_0 = i], \quad i, j \in \mathbb{S}.$$

By the first-step analysis, we get again

$$\begin{aligned} \mu_j(i) &:= \mathbb{E}[T_j^r|X_0 = i] = 1 \times \mathbb{P}(X_1 = j|X_0 = i) + \sum_{\ell\in\mathbb{S},\ell\neq j} \mathbb{P}(X_1 = \ell|X_0 = i)[1 + \mathbb{E}[T_j^r|X_0 = \ell]] \\ &= P_{i,j} + \sum_{\ell\in\mathbb{S},\ell\neq j} P_{i,\ell}[1 + \mathbb{E}[T_j^r|X_0 = \ell]] = 1 + \sum_{\ell\in\mathbb{S},\ell\neq j} P_{i,\ell}\mu_j(\ell). \end{aligned}$$

4.4 The number of returns

The first-step analysis is a robust method, which can be (sometimes with some modifications) used to answer many different questions about Markov chains. In this part, we discuss the number of returns of a Markov chain to some (fixed) state $i \in \mathbb{S}$. Warning: the results of this part will be very important in the subsequent classification of states and in the study of limiting behavior of Markov chains.

For $i \in \mathbb{S}$, we define

$$R_i := \#\{n \geq 1 : X_n = i\},$$

the number of returns of the Markov chain $X = (X_n)_{n=0}^{\infty}$ to the state i. Again, this is a random variable with values in $\{0, 1, \dots\} \cup \{+\infty\}$. We are interested in the expected value of this random variable, given that the initial state is given. This means, that we would like to study

$$\mathbb{E}[R_j | X_0 = i] \quad \text{for} \quad i, j \in \mathbb{S}.$$

Before we come to that, we need some tools. First, we define

$$\begin{aligned} p_{i,j} &:= \mathbb{P}(T_j^r < +\infty | X_0 = i) \\ &= \mathbb{P}(X_n = j \text{ for some } n \geq 1 | X_0 = i) \end{aligned}$$

the probability of the event, that the process ever reaches j if it started in i. The calculation of $p_{i,j}$ is rather involved. For that sake, we decompose the event $\{\exists n \geq 1 : \ X_n = j\}$ into disjoint subset by the first arrival to j

$$\begin{aligned} p_{i,j} &= \mathbb{P}(X_n = j \text{ for some } n \geq 1 | X_0 = i) \\ &= \sum_{n=1}^{\infty} \underbrace{\mathbb{P}(X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j, X_0 = i | X_0 = i)}_{=: f_{i,j}^{(n)}} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}. \end{aligned} \tag{4.3}$$

Note, that it is not really the "first-step analysis" but more something like "the analysis of the first return". Few facts about $f_{i,j}^{(n)}$ are easy to observe, namely

- $f_{i,j}^{(n)} = \mathbb{P}(T_j^r = n | X_0 = i)$, i.e., $f_{i,j}^{(n)}$ is actually the distribution of the random variable T_j^r conditioned to $X_0 = i$,
- $f_{i,j}^{(0)} = 0$ for $i \neq j$ and $f_{i,i}^{(0)} = 1$,
- $f_{i,j}^{(1)} = P_{i,j} = \mathbb{P}(X_1 = j | X_0 = i)$.

To calculate $f_{i,j}^{(n)}$, one could build the following system of equations

$$\begin{aligned}
[P^n]_{i,j} &:= \mathbb{P}(X_n = j | X_0 = i) = \frac{\mathbb{P}(X_n = j, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{k=1}^{n} \frac{\mathbb{P}(X_n = j, X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&\qquad \cdot \frac{\mathbb{P}(X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j, X_0 = i)}{\mathbb{P}(X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j, X_0 = i)} \\
&= \sum_{k=1}^{n} \mathbb{P}(X_n = j | X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j, X_0 = i) \\
&\qquad \cdot \mathbb{P}(X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j | X_0 = i) \\
&= \sum_{k=1}^{n} \mathbb{P}(X_n = j | X_k = j) \cdot f_{i,j}^{(k)} = \sum_{k=1}^{n} f_{i,j}^{(k)} [P^{n-k}]_{j,j}.
\end{aligned}$$

At least in theory, we could now calculate $f_{i,j}^{(n)}$ from this convolution identity and the initial conditions and then we find $p_{i,j}$ by (4.3).

Let us assume for the moment, that we succeeded to find $p_{i,j}$'s. How do they help us to find the distribution of R_i's? Observe that $R_j = m$ conditioned on $X_0 = i$ if the process starts in i, then it arrives to j, next it returns to j (and that $m-1$ times) and finally it leaves j and never comes back. Hence, for $m \geq 1$,

$$\mathbb{P}(R_j = m | X_0 = i) = p_{i,j} \cdot p_{j,j}^{m-1} \cdot (1 - p_{j,j}).$$

Here, we used that every arrival to j splits the process into two parts, which are independent conditioned on the fact, that we just reached j. If $m = 0$, we obtain

$$\mathbb{P}(R_j = 0 | X_0 = i) = 1 - p_{i,j}.$$

We observe that

$$\begin{aligned}
\mathbb{P}(R_j < +\infty | X_0 = i) &= \sum_{m=0}^{\infty} \mathbb{P}(R_j = m | X_0 = i) = (1 - p_{i,j}) + \sum_{m=1}^{\infty} p_{i,j} \cdot p_{j,j}^{m-1} \cdot (1 - p_{j,j}) \\
&= \begin{cases} \text{if } p_{j,j} = 1 : 1 - p_{i,j}, \\ \text{if } p_{j,j} < 1 : (1 - p_{i,j}) + p_{i,j} \cdot (1 - p_{j,j}) \sum_{m=1}^{\infty} p_{j,j}^{m-1} \\ \qquad\qquad = (1 - p_{i,j}) + \frac{(1 - p_{j,j}) \cdot p_{i,j}}{1 - p_{j,j}} = 1. \end{cases}
\end{aligned}$$

Taking complements, we can reformulate this as

$$\mathbb{P}(R_j = +\infty | X_0 = i) = \begin{cases} p_{i,j} & \text{if } p_{j,j} = 1, \\ 0 & \text{if } p_{j,j} < 1. \end{cases}$$

And specially, for $i = j$, we get

$$\mathbb{P}(R_i < +\infty | X_0 = i) = \begin{cases} 0 & \text{if } p_{i,i} = 1, \\ 1 & \text{if } p_{i,i} < 1. \end{cases}$$

and

$$\mathbb{P}(R_i = +\infty | X_0 = i) = \begin{cases} 1 & \text{if } p_{i,i} = 1, \\ 0 & \text{if } p_{i,i} < 1. \end{cases}$$

Finally, we may calculate $\mathbb{E}[R_j | X_0 = i]$:

If $p_{j,j} = 1$, we get $\mathbb{P}(R_j = +\infty | X_0 = i) = p_{i,j}$. Therefore, if $p_{j,j} = 1$ and $p_{i,j} > 0$, then $\mathbb{E}[R_j | X_0 = i] = +\infty$. And if $p_{j,j} = 1$ and $p_{i,j} = 0$, we obtain

$$\mathbb{E}[R_j | X_0 = i] = \sum_{m=0}^{\infty} m \cdot \underbrace{\mathbb{P}(R_j = m | X_0 = i)}_{=0} + (+\infty) \cdot \underbrace{\mathbb{P}(R_j = +\infty | X_0 = i)}_{=0} = 0.$$

And if $p_{j,j} < 1$, we get $\mathbb{P}(R_j = +\infty | X_0 = i) = 0$ and

$$\begin{aligned} \mathbb{E}[R_j | X_0 = i] &= \sum_{m=0}^{\infty} m \cdot \mathbb{P}(R_j = m | X_0 = i) + (+\infty) \cdot \underbrace{\mathbb{P}(R_j = +\infty | X_0 = i)}_{=0} \\ &= \sum_{m=0}^{\infty} m \cdot p_{i,j} \cdot p_{j,j}^{m-1} \cdot (1 - p_{j,j}) = p_{i,j} \cdot (1 - p_{j,j}) \cdot \underbrace{\sum_{m=1}^{\infty} m \cdot p_{j,j}^{m-1}}_{= \frac{1}{(1-p_{j,j})^2}} \\ &= \frac{p_{i,j}}{1 - p_{j,j}}. \end{aligned}$$

5 Classification of states

We want to distinguish states according to their properties with respect to the Markov chain. The main difference will be between states, which get visited repeatedly and states, which are visited only rarely, i.e., finitely many times. But before we come to that, we exclude the cases, when some parts of the state space are essentially disjoint - there is no possibility of changing from one group of states to the other group. One can think about two not connected mazes. If a rat start in the first one, it will always stay there - and the same is true for the other one.

Recall, that we already defined and used the following two notions.

(i) The state $i \in \mathbb{S}$ is *absorbing* if $\mathbb{P}(X_1 = i | X_0 = i) = 1$.

(ii) The distribution $\pi = [\pi_j]_{j \in \mathbb{S}}$ is called *stationary* if $\pi = \pi \cdot P$.

5.1 Communicating states

We say that the state $j \in \mathbb{S}$ is accessible from $i \in \mathbb{S}$ if it is possible to travel from i to j in a finite time with positive probability. In other words, j is accessible from i, if there exists $n \geq 0$ such that

$$[P^n]_{i,j} = \mathbb{P}(X_n = j | X_0 = i) > 0.$$

We denote this by $i \to j$. Note that (by choosing $n = 0$) we have $i \to i$ for all $i \in \mathbb{S}$.

If $i \to j$ and $j \to i$, then we say that the states i and j communicate, and denote this by $i \leftrightarrow j$.

The binary relation $\leftrightarrow$ satisfies

(i) Reflexivity: For all $i \in \mathbb{S}$, we always have $i \leftrightarrow i$ (by choosing $n = 0$ above).

(ii) Symmetry: If $i \leftrightarrow j$, then also $j \leftrightarrow i$.

(iii) Transitivity: If $i \leftrightarrow j$ and $j \leftrightarrow k$, then also $i \leftrightarrow k$.

For the last property, note that if $\mathbb{P}(X_n = j | X_0 = i) = a > 0$ and $\mathbb{P}(X_m = k | X_0 = j) = b > 0$, then

$$\begin{aligned}
\mathbb{P}(X_{m+n} = k | X_0 = i) \geq \mathbb{P}(X_{m+n} = k, X_n = j | X_0 = i) &= \frac{\mathbb{P}(X_{m+n} = k, X_n = j, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \frac{\mathbb{P}(X_{m+n} = k, X_n = j, X_0 = i)}{\mathbb{P}(X_n = j, X_0 = i)} \cdot \frac{\mathbb{P}(X_n = j, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \mathbb{P}(X_{m+n} = k | X_n = j, X_0 = i) \cdot \mathbb{P}(X_n = j | X_0 = i) = b \cdot a > 0.
\end{aligned}$$

Relation satisfying these three properties is called *equivalence* and it allows to split the set into *equivalence classes* $A_1, A_2, \ldots$, which form a disjoint decomposition of $\mathbb{S}$, i.e.,

(i) $\mathbb{S} = A_1 \cup A_2 \cup \ldots$;

(ii) $A_p \cap A_q = \emptyset$ if $p \neq q$;

(iii) If $i, j \in A_p$ for any fixed p, then $i \leftrightarrow j$;

(iv) if $i \in A_p$, $j \in A_q$ and $p \neq q$, then $i \not\leftrightarrow j$.

The states communicate with each other if, and only if, they belong to the same class A_p. If there is only one class of communicating states, we call the Markov chain *irreducible*. Otherwise, it is called *reducible*.

5.2 Recurrent and transient states

Definition 5.1. The state $i \in \mathbb{S}$ is called *recurrent* if

$$p_{i,i} = \mathbb{P}(T_i^r < +\infty | X_0 = i) = 1.$$

By the results of the first step analysis, this is equivalent to (recall that R_i is the number of returns to i)

(i) $\mathbb{P}(\exists n \in \mathbb{N} : X_n = i | X_0 = i) = 1$,

(ii) $\mathbb{E}[R_i | X_0 = i] = +\infty$,

(iii) $\mathbb{P}(R_i = +\infty | X_0 = i) = 1$.

Theorem 5.2. *State $i \in \mathbb{S}$ is recurrent if, and only if,* $\sum_{n=1}^{\infty} [P^n]_{i,i} = +\infty$.

Proof. Let $i, j \in \mathbb{S}$. Then we calculate

$$\begin{aligned}\mathbb{E}[R_j | X_0 = i] &= \mathbb{E}\left[\sum_{n=1}^{\infty} \chi_{\{X_n = j\}} | X_0 = i\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[\chi_{\{X_n = j\}} | X_0 = i\right] \\ &= \sum_{n=1}^{\infty} \mathbb{P}(X_n = j | X_0 = i) = \sum_{n=1}^{\infty} [P^n]_{i,j}.\end{aligned}$$

Finally, we apply this calculation to $i = j$. □

Theorem 5.3. *If $i \in \mathbb{S}$ is a recurrent state, then every state $j \in \mathbb{S}$, which communicates with i, is also recurrent.*

Proof. We assume that $i \neq j$. We know that i is recurrent and that i and j communicate. Therefore, there exist $a, b \geq 1$ such that

$$[P^a]_{i,j} > 0 \quad \text{and} \quad [P^b]_{j,i} > 0.$$

For $n \geq a + b$, we get

$$\begin{aligned}\mathbb{P}(X_n = j | X_0 = j) &= \sum_{\ell, m \in \mathbb{S}} \mathbb{P}(X_n = j | X_{n-a} = \ell) \cdot \mathbb{P}(X_{n-a} = \ell | X_b = m) \cdot \mathbb{P}(X_b = m | X_0 = j) \\ &\geq \mathbb{P}(X_n = j | X_{n-a} = i) \cdot \mathbb{P}(X_{n-a} = i | X_b = i) \cdot \mathbb{P}(X_b = i | X_0 = j) \\ &= [P^a]_{i,j} \cdot [P^{n-a-b}]_{i,i} \cdot [P^b]_{j,i}\end{aligned}$$

and

$$\sum_{n=a+b}^{\infty}[P^n]_{j,j} \geq [P^a]_{i,j} \cdot [P^b]_{j,i} \cdot \sum_{n=a+b}^{\infty}[P^{n-a-b}]_{i,i}$$
$$= [P^a]_{i,j} \cdot [P^b]_{j,i} \cdot \sum_{n=0}^{\infty}[P^n]_{i,i} = +\infty.$$

By the previous theorem, this finishes the proof. □

The opposite of *recurrent* is *transient.*

Definition 5.4. We call the state $i \in \mathbb{S}$ transient, if (and only if) it is not recurrent. By the previous, the following statements are equivalent

- $i \in \mathbb{S}$ is transient,
- $\mathbb{P}(R_i = +\infty | X_0 = i) < 1$,
- $\mathbb{P}(R_i = +\infty | X_0 = i) = 0$,
- $\mathbb{P}(R_i < +\infty | X_0 = i) > 0$,
- $\mathbb{P}(R_i < +\infty | X_0 = i) = 1$,
- $p_{i,i} = \mathbb{P}(T_i^r < +\infty | X_0 = i) = \mathbb{P}(\exists n \geq 1 : X_n = i | X_0 = i) < 1$,
- $\mathbb{P}(T_i^r = +\infty | X_0 = i) > 0$,
- $\mathbb{E}[R_i | X_0 = i] < +\infty$,
- $\sum_{n=1}^{\infty}[P^n]_{i,i} < +\infty$.

Clearly, if $i \in \mathbb{S}$ is transient and j communicates with i, then j is also transient. Therefore, the communication classes can be divided into two groups - of classes, which contain only recurrent states (recurrent classes) and classes, which contain only transient states (transient classes). If $\mathbb{S}$ is finite, then at least one class is recurrent.

Theorem 5.5. *Let $X = (X_n)_{n=0}^{\infty}$ be a Markov chain with finite state space $\mathbb{S}$. Then at least one state is recurrent.*

Proof. If $j \in \mathbb{S}$ is transient, then $p_{j,j} < 1$ and we get for all $i \in \mathbb{S}$

$$\mathbb{E}[R_j | X_0 = i] = \sum_{n=0}^{\infty} n \cdot \mathbb{P}(R_j = n | X_0 = i) = p_{i,j} \cdot (1 - p_{j,j}) \cdot \sum_{n=1}^{\infty} n \cdot (p_{j,j})^{n-1}$$
$$= \frac{p_{i,j}}{1 - p_{j,j}} < +\infty.$$

Therefore, also

$$\sum_{n=0}^{\infty}[P^n]_{i,j} = \mathbb{E}[R_j | X_0 = i] < +\infty$$

and, consequently, $\lim_{n\to\infty}[P^n]_{i,j} = 0$.[1]

If all $j \in \mathbb{S}$ would be transient, then we would get

$$0 = \sum_{j\in\mathbb{S}} \lim_{n\to\infty}[P^n]_{i,j} = \lim_{n\to\infty}\sum_{j\in\mathbb{S}}[P^n]_{i,j} = \lim_{n\to\infty} 1 = 1.$$

□

5.3 Positive recurrent and null recurrent states

If the state $i \in \mathbb{S}$ is recurrent, then $p_{i,i} = \mathbb{P}(T_i^r < +\infty | X_0 = i) = 1$ and T_i^r is almost surely a finite random variable. We therefore expect, that it returns from i to i over and over again. Nevertheless, these returns might happen after a very long time.

Definition 5.6. Let $i \in \mathbb{S}$ be a recurrent state. Then

- $i \in \mathbb{S}$ is positive recurrent, if $\mu_i(i) = \mathbb{E}[T_i^r | X_0 = i] < +\infty$ and
- $i \in \mathbb{S}$ is null recurrent, if $\mu_i(i) = \mathbb{E}[T_i^r | X_0 = i] = +\infty$.

Being positive recurrent or null recurrent is again a property of the whole communicating class.

Theorem 5.7. *Let $i \neq j \in \mathbb{S}$ be two communicating states. If i is positive recurrent, then also j is positive recurrent.*

The communicating classes therefore split into three groups - classes of transient states, classes of positive recurrent states, and classes of null recurrent states.

Proof. Step 1.
Let $i \in \mathbb{S}$ be positive recurrent, i.e, $\mu_i(i) = \mathbb{E}[T_i^r | X_0 = i] < +\infty$. Let also $n_0 \geq 1$ denote the smallest positive integer with $[P^{n_0}]_{i,j} > 0$. Then we show that also $\mu_i(j) = \mathbb{E}[T_i^r | X_0 = j] < +\infty$. We calculate

$$\begin{aligned}
+\infty > \mu_i(i) = \mathbb{E}[T_i^r | X_0 = i] &= \sum_{n=1}^{\infty} n \cdot \mathbb{P}(T_i^r = n | X_0 = i) = \sum_{n=1}^{\infty} n \cdot \frac{\mathbb{P}(T_i^r = n, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&\geq \sum_{n=1}^{\infty} n \cdot \frac{\mathbb{P}(T_i^r = n, X_{n_0} = j, X_{n_0-1} \neq i, \dots, X_1 \neq i, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&\qquad \cdot \frac{\mathbb{P}(X_{n_0} = j, X_{n_0-1} \neq i, \dots, X_1 \neq i, X_0 = i)}{\mathbb{P}(X_{n_0} = j, X_{n_0-1} \neq i, \dots, X_1 \neq i, X_0 = i)} \\
&= \mathbb{E}[T_i^r | X_{n_0} = j, X_{n_0-1} \neq i, \dots, X_1 \neq i, X_0 = i] \cdot \mathbb{P}(X_{n_0} = j, X_{n_0-1} \neq i, \dots, X_1 \neq i | X_0 = i) \\
&= (n_0 + \mathbb{E}[T_i^r | X_0 = j]) \cdot \underbrace{\mathbb{P}(X_{n_0} = j, X_{n_0-1} \neq i, \dots, X_1 \neq i | X_0 = i)}_{>0}.
\end{aligned}$$

Hence, $\mathbb{E}[T_i^r | X_0 = j] < +\infty$.

Step 2.
Let $X_0 = i$ and let $\{Y_m : m \geq 1\}$ be independent copies of T_i^r (conditioned onto $X_0 = i$). We can obtain Y_m's by setting

[1] Up to now, this holds also with $\mathbb{S}$ infinite.

- $Y_1 = \inf\{\ell \geq 1 : X_\ell = i\} = T_i^r$,
- $Y_2 = \inf\{\ell \geq 1 : X_{Y_1+\ell} = i\}$ - the time of the second return to i,
- $Y_m = \inf\{\ell \geq 1 : X_{Y_1+\cdots+Y_{m-1}+\ell} = i\}$ - the time of the mth return to i, $m \geq 2$.
- By the return to i, the chain X splits again into two parts, which are independent (under the condition of the current state i). This means that Y_m's are indeed independent.
- $\mathbb{E}[Y_\ell] = \mathbb{E}[T_i^r | X_0 = i] < +\infty$.

We define

$$\begin{aligned} p &= \mathbb{P}(X \text{ visits } j \text{ before it first returns to } i | X_0 = i) \\ &\geq \mathbb{P}(X_1 \neq i, \ldots, X_{n_0-1} \neq i, X_{n_0} = j | X_0 = i) > 0. \end{aligned}$$

Anytime, when X returns to i, it has probability $p > 0$ that it first gets to j before it returns to i. If N is the (random variable of the) number of returns to i before it first gets to j, then N has geometric distribution with parameter p, i.e.,

$$\mathbb{P}(N = k) = p(1-p)^k, \quad k \geq 0.$$

For $X_0 = i$ we have $T_j^r \leq Y_1 + \cdots + Y_{N+1}$ and $\mathbb{E}[T_j^r | X_0 = i] \leq \mathbb{E}(N+1) \cdot \mathbb{E}(Y) < +\infty$.

Step 3.
Finally, we combine both the previous steps

$$\mathbb{E}[T_j^r | X_0 = j] \leq \mathbb{E}[T_i^r | X_0 = j] + \mathbb{E}[T_j^r | X_0 = i] < +\infty$$

(the length of a path from j to j is at most that long as its length from j to i and then from i to j).

□

5.4 Periodic and aperiodic states

Definition 5.8. For $i \in \mathbb{S}$ we define the period of i as the greatest common divisor of the set $\{n \geq 1 : [P^n]_{i,i} > 0\}$. If the period of i is one, we call i aperiodic. The chain is called aperiodic if all its states are aperiodic.

If $P_{i,i} > 0$, then i is surely aperiodic. We do not define the period of i if $[P^n]_{i,i} = 0$ for all $n \geq 1$ (alternatively, we could define the period to be zero).

Theorem 5.9. *Every two communicating states have the same period.*

Proof. We assume that $i \neq j$, $i \leftrightarrow j$ and we denote by $d(i)$ and $d(j)$ the period of i and j, respectively. The path from i to i is possible with positive probability, therefore $d(i) \geq 1$ and, similarly, $d(j) \geq 1$. By definition, there are $k, \ell \geq 1$ such that

$$\alpha = [P^k]_{i,j} > 0 \quad \text{and} \quad \beta = [P^\ell]_{j,i} > 0.$$

Then $[P^{k+l}]_{i,i} > 0$ and, therefore, $d(i)|(k+l)$.

Let now $m \geq 1$ be arbitrary with $[P^m]_{j,j} > 0$. Then also $[P^{m+k+\ell}]_{i,i} > 0$ (with positive probability, we can go from i to j in k steps, from j to j in m steps, and from j to i in ℓ steps). Therefore, $d(i)$ divides also $k+\ell+m$ and, therefore, also m. We observe, that $d(i)$ is a common divisor of $\{m \geq 1 : [P^m]_{j,j} > 0\}$. But $d(j)$ was the largest integer with this property and, therefore, $d(i) \leq d(j)$. In the same way, we would prove $d(j) \leq d(i)$, which finishes the proof. □

6 Limiting behavior

This section composes the heart of the subject of Markov processes. For many researchers, this is actually the reason, why we study Markov processes at all. Let us briefly describe the main aim of this section. In the optimal scenario, the stationary distribution π of a Markov chain $X = (X_n)_{n=0}^{\infty}$ would exist and be uniquely defined. And, furthermore, if we start the Markov chain X in an arbitrary initial distribution $p^{(0)}$, the distribution of X_n would converge to π in a long-term horizon. And, if we could wish for even more, it would converge as fast as possible [2].

Unfortunately, this line can fail in several ways. For example, if the transition matrix P is equal to identity, i.e., $P = I$, then every distribution π is a stationary distribution as we always have $\pi = \pi I$. Hence, the uniqueness of stationary distribution can get lost. We encountered even more examples, which contradict the previous scenario, when we studied Markov processes with only two states.

6.1 Stationary distribution

Let us recall, that a stationary distribution of a Markov chain $X = (X_n)_{n=0}^{\infty}$ with a state space $\mathbb{S}$ is every (row) vector $\pi = [\pi_j]_{j\in\mathbb{S}}$, which satisfies $\pi = \pi \cdot P$, where P is the transition matrix of X. It goes without saying, that we assume the entries of π to be non-negative and summing to one.

We are interested in two crucial questions:

1. Does a stationary distribution always exist?
2. And (if yes) is it unique?

Unfortunately, the answer to both previous questions is negative in general. [(☕) Show that the random walk on $\mathbb{Z}$ does not have a stationary distribution for all $0 < p < 1$, even for $p = 1/2$!] And we have to pose some additional assumptions to get a positive answer.

Before we come to that, let us consider two simple concepts.

6.1.1 Random walk on a graph

Similarly to the random walk on $\mathbb{Z}$, we define a random walk on a graph. For that sake, we recall few basic notions from graph theory. Let $V = \{1, \dots, N\}$ be a finite set (the vertices of a graph) and let $E \subset \{\{i, j\} : 1 \le i < j \le N\}$ be its edges. By $\deg(i)$ we denote the degree of the vertex $i \in V$, i.e., the number of its neighbors

$$\deg(i) = \#\{j : \{i, j\} \in E\}.$$

We consider the following Markov chain. If $X_n = i$, then X_{n+1} is equal to one of the neighbors of i, with each of them having the same probability. This means, that we select uniformly and independently on previous choices one of the edges going from i and follow it to the next vertex. Formally,

$$X_{n+1} = j \begin{cases} \text{with probability} & \frac{1}{\deg(i)} & \text{if } j \text{ is a neighbor of } i, \\ \text{with probability} & 0 & \text{if } j \text{ is not a neighbor of } i \end{cases}$$

or

$$P_{i,j} = P(X_{n+1} = j | X_n = i) = \begin{cases} \frac{1}{\deg(i)} & \text{if } j \text{ is a neighbor of } i, \\ 0 & \text{if } j \text{ is not a neighbor of } i. \end{cases}$$

It is quite easy to check (see also the next section) that the vector

$$\pi = [\pi_i]_{i\in\mathbb{S}} = \left[\frac{\deg(i)}{2\cdot \#E}\right]_{i\in\mathbb{S}}$$

is a stationary distribution. Indeed,

$$\begin{aligned}[\pi\cdot P]_j &= \sum_{i\in\mathbb{S}} \pi_i P_{i,j} = \sum_{\substack{i\in\mathbb{S}\\ \{i,j\}\in E}} \frac{\deg(i)}{2\cdot\#E}\cdot\frac{1}{\deg(i)} = \frac{1}{2\cdot\#E}\sum_{\substack{i\in\mathbb{S}\\ \{i,j\}\in E}} 1 \\ &= \frac{\deg(j)}{2\cdot\#E} = \pi_j.\end{aligned}$$

Finally, we verify that π is indeed a distribution. Its entries are non-negative and

$$\sum_{j\in\mathbb{S}} \pi_j = \frac{1}{2\cdot\#E}\sum_{j\in\mathbb{S}} \deg(j) = 1.$$

[(☕) How is it with the uniqueness of this stationary distribution?]

6.1.2 Detailed balance equations

Definition 6.1. We say that the distribution $\nu = [\nu_j]_{j\in\mathbb{S}}$ satisfies the detailed balance equations, if for every $i, j \in \mathbb{S}$ it holds that

$$\nu_i P_{i,j} = \nu_j P_{j,i}.$$

Theorem 6.2. *If ν satisfies the detailed balance equations, then ν is stationary.*

Proof. The proof is straightforward. For every $j \in \mathbb{S}$ we have

$$(\pi P)_j = \sum_{i\in\mathbb{S}} \pi_i P_{i,j} = \sum_{i\in\mathbb{S}} \pi_j P_{j,i} = \pi_j \sum_{i\in\mathbb{S}} P_{j,i} = \pi_j,$$

i.e., $\pi = \pi\cdot P$. □

Remark 2. Previous theorem is only one implication (every distribution with detailed balance equations is stationary) and it can not be reversed. (☕) Find a Markov chain X, which has a stationary distribution, which does not satisfy the detailed balance equations.

Let us verify, that the stationary distribution of a random walk on a finite graph could have been found by detailed balance equations. Indeed, if $i \neq j$ are not neighbors, then $P_{i,j} = P_{j,i} = 0$ and the equations are satisfied. If i and j are connected by an edge, then we require that

$$\pi_i P_{i,j} = \frac{\pi_i}{\deg(i)} = \frac{\pi_j}{\deg(j)} = \pi_j P_{j,i}.$$

We observe, that we can fulfill these equations by choosing $\pi_i = c\cdot\deg(i)$ and $c = 2\#E$ to ensure that the entries of π sum up to one.

Theorem 6.3. *Let $X = (X_n)_{n=0}^{\infty}$ be irreducible Markov chain.*

a) If X has a stationary distribution π, then necessarily π is given by

$$\pi_i = \frac{1}{\mu_i(i)}, \quad i \in \mathbb{S}, \tag{6.1}$$

where $\mu_i = \mathbb{E}[T_i^r | X_0 = i]$ is the mean time of return from i to i. This also means, that π is uniquely determined and all states are positive recurrent.

b) On the other hand, if a Markov chain is irreducible and positive recurrent, then the distribution π defined by (6.1) is the only stationary distribution.

Proof. The proof is quite long and we split it into several parts.
Step 1. We start with part a) and show that X is recurrent. Let π be a stationary distribution of an irreducible Markov chain X. For contradiction, let us suppose that X is transient. Then $\lim_{n\to\infty} [P^n]_{i,j} = 0$ for all $i, j \in \mathbb{S}$[2]. Hence

$$\begin{aligned}\pi_j = [\pi \cdot P]_j = [\pi \cdot P^n]_j = \lim_{n\to\infty} [\pi \cdot P^n]_j = \lim_{n\to\infty} \sum_{i\in\mathbb{S}} \pi_i [P^n]_{i,j} \\ = \sum_{i\in\mathbb{S}} \pi_i \lim_{n\to\infty} [P^n]_{i,j} = 0\end{aligned}$$

for all $j \in \mathbb{S}$. Note, that here we used the Dominated convergence theorem (a.k.a. Lebesgue limit theorem), with $\pi = [\pi_i]_{i\in\mathbb{S}}$ being the integrable majorant. This is a contradiction with π being a distribution and, therefore, X is recurrent.

Step 2. We show that $\pi_i > 0$ for every $i \in \mathbb{S}$. Indeed, if for some $i \in \mathbb{S}$ it holds that $\pi_i = 0$, then we get

$$0 = \pi_i = \sum_{k\in\mathbb{S}} \pi_k [P^n]_{i,k} \geq \pi_j [P^n]_{i,j}$$

for all $j \in \mathbb{S}$ and all $n \geq 1$. Now it is enough to choose some $j \in \mathbb{S}$ with $\pi_j > 0$ and some $n \geq 1$ with $[P^n]_{i,j} > 0$ (which is always possible, as i and j communicate).

Step 3. Next, we show that (6.1) holds if the assumptions of a) are satisfied. Let X be irreducible, with stationary distribution π. Then X is recurrent and we start it with the initial distribution equal to π, i.e., $p^{(0)} = \pi$. Then (knowing that T_i^r is finite almost surely)

$$\begin{aligned}\mu_i = \mathbb{E}[T_i^r | X_0 = i] = \sum_{n=1}^{\infty} n \cdot \mathbb{P}(T_i^r = n | X_0 = i) = \sum_{n=1}^{\infty} \mathbb{P}(T_i^r \geq n | X_0 = i) \\ = \sum_{n=1}^{\infty} \frac{\mathbb{P}(T_i^r \geq n, X_0 = i)}{\mathbb{P}(X_0 = i)}.\end{aligned}$$

For $n \geq 1$, we get $\mathbb{P}(T_i^r \geq 1, X_0 = i) = \mathbb{P}(X_0 = i) = p_i^{(0)} = \pi_i$.
For $n \geq 2$ we use the identity $\mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A \cap B^c)$ for the sets $A = \{X_m \neq i, m =$

[2] See Theorem 5.5

$1, \dots, n-1\}$ and $B = \{X_0 = i\}$. We obtain

$$\begin{aligned}\mathbb{P}(T_i^r \geq n, X_0 = i) &= \mathbb{P}(X_0 = i, X_1 \neq i, \dots, X_{n-1} \neq i) \\ &= \mathbb{P}(X_1 \neq i, \dots, X_{n-1} \neq i) - \mathbb{P}(X_0 \neq i, X_1 \neq i, \dots, X_{n-1} \neq i) \\ &= \mathbb{P}(X_0 \neq i, \dots, X_{n-2} \neq i) - \mathbb{P}(X_0 \neq i, X_1 \neq i, \dots, X_{n-1} \neq i) \\ &= \alpha_{n-2} - \alpha_{n-1},\end{aligned}$$

where

$$\alpha_n = \mathbb{P}(X_0 \neq i, \dots, X_n \neq i) =: \mathbb{P}(A_n).$$

Observe, that $A_0 \supset A_1 \supset A_2 \supset \dots$ and that

$$\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=0}^{\infty} A_n\right) = \mathbb{P}(X_m \neq i \text{ for all } m \geq 0) = 0$$

as X is recurrent (as well as all its states).[3]

We therefore have

$$\mu_i = \sum_{n\geq 1} \frac{\mathbb{P}(T_i^r \geq n, X_0 = i)}{\pi_i} = \sum_{n\geq 2} \frac{\alpha_{n-2} - \alpha_{n-1}}{\pi_i} + \frac{\pi_i}{\pi_i}.$$

We multiply this identity by π_i and obtain

$$\mu_i \cdot \pi_i = \pi_i + \sum_{n=2}^{\infty} (\alpha_{n-2} - \alpha_{n-1}) = \pi_i + \alpha_0 - \lim_{n\to\infty} \alpha_n = \mathbb{P}(X_0 = i) + \mathbb{P}(X_0 \neq i) = 1.$$

This finishes the proof of a).

Step 4. We now prove the part b).

First, we assume that X is irreducible and positive recurrent. Then $0 < \mu_i < +\infty$ for all states $i \in \mathbb{S}$ and we can therefore define

$$\pi_i := \frac{1}{\mu_i}, \quad i \in \mathbb{S}.$$

With this choice, $\pi_i > 0$ for all $i \in \mathbb{S}$. To show that π is a stationary distribution, we have to discuss two facts:

b1) $\sum_{j\in\mathbb{S}} \pi_j = 1$,

b2) $\pi = \pi \cdot P$.

Let $N_j(n)$ denote the number of occurrences in $j \in \mathbb{S}$ in the first n steps of the Markov chain, i.e,

$$N_j(n) := \#\{1 \leq \ell \leq n : X_\ell = j\} = \sum_{\ell=1}^{n} \chi_{\{X_\ell = j\}}.$$

[3] details?

Then

$$\mathbb{E}[N_j(n)|X_0 = i] = \sum_{\ell=1}^{n} \mathbb{P}(X_\ell = j|X_0 = i) = \sum_{\ell=1}^{n} [P^\ell]_{i,j}$$

and

$$\sum_{j\in\mathbb{S}} \frac{1}{n} \cdot \mathbb{E}[N_j(n)|X_0 = i] = \frac{1}{n} \sum_{j\in\mathbb{S}} \sum_{\ell=1}^{n} [P^\ell]_{i,j} = \frac{1}{n} \sum_{\ell=1}^{n} \underbrace{\sum_{j\in\mathbb{S}} [P^n]_{i,j}}_{=1} = 1.$$

We show that

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}[N_j(n)|X_0 = i] = \frac{1}{\mu_j} \quad \text{for all } i, j \in \mathbb{S}. \tag{6.2}$$

Then we get[4]

$$\begin{aligned} \sum_{j\in\mathbb{S}} \pi_j &= \sum_{j\in\mathbb{S}} \frac{1}{\mu_j} = \sum_{j\in\mathbb{S}} \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[N_j(n)|X_0 = i] \\ &= \lim_{n\to\infty} \sum_{j\in\mathbb{S}} \frac{1}{n} \mathbb{E}[N_j(n)|X_0 = i] = \lim_{n\to\infty} 1 = 1. \end{aligned}$$

Step 5. Now we prove (6.2).
We denote by T_m the time of the m-th visit to j, i.e.,

$$T_m = \min\{n \geq 1 : N_j(n) = m\}.$$

Observe that $T_{N_j(n)} \leq n$. Also, the random variables $T_2 - T_1, T_3 - T_2, \ldots, T_m - T_{m-1}$ are independent and equidistributed (and are independent copies of T_j^r conditioned on $X_0 = j$.) We get, by the law of large numbers, that

$$\frac{T_m}{m} = \frac{(T_m - T_{m-1}) + (T_{m-1} - T_{m-2}) + \cdots + (T_2 - T_1) + T_1}{m} \to \mu_j = \mathbb{E}[T_j^r|X_0 = j].$$

Due to

$$T_{N_j(n)} \leq n \leq T_{N_j(n)+1}$$

we get

$$\frac{T_{N_j(n)}}{N_j(n)} \leq \frac{n}{N_j(n)} \leq \frac{T_{N_j(n)+1}}{N_j(n)} \cdot \frac{N_j(n)+1}{N_j(n)+1}.$$

As $N_j(n) \to +\infty$ as $n \to +\infty$ (recall that X is positive recurrent), then we get

$$\frac{n}{N_j(n)} \to \mu_j \quad \text{and} \quad \frac{N_j(n)}{n} \to \frac{1}{\mu_j}.$$

almost surely. Taking the mean value, (6.2) follows.

[4] We assume that $\mathbb{S}$ is finite for this step and leave out the details for infinite $\mathbb{S}$.

Step 6. Finally, we show that $\pi_j = 1/\mu_j$ is stationary, i.e, we prove the condition b2).[5] First, we calculate

$$\begin{aligned}\sum_{j\in\mathbb{S}} \frac{1}{n}\cdot \mathbb{E}[N_j(n)|X_0=i]P_{j,k} &= \sum_{j\in\mathbb{S}} \frac{1}{n}\left(\sum_{m=1}^{n}[P^m]_{i,j}\right)P_{j,k}\\ &= \sum_{m=1}^{n}\frac{1}{n}\sum_{j\in\mathbb{S}}[P^n]_{i,j}\cdot P_{j,k} = \sum_{m=1}^{n}[P^{m+1}]_{j,k}\\ &= \frac{1}{n}\sum_{m=2}^{n+1}[P^m]_{i,k} = \frac{1}{n}\left(\sum_{m=1}^{n+1}[P^m]_{i,k} - P_{i,k}\right)\\ &= \frac{1}{n}\left\{\mathbb{E}[N_k(n+1)|X_0=i] - P_{i,k}\right\}.\end{aligned}$$

If we now pass to the limit $n\to\infty$ (and use that $\mathbb{S}$ is finite), then we obtain

$$\sum_{j\in\mathbb{S}}\frac{1}{\mu_j}\cdot P_{j,k} = \frac{1}{\mu_k}.$$

□

6.2 Limiting distribution

Definition 6.4. We say that X has a limiting distribution if the limit

$$\lim_{n\to\infty}\mathbb{P}(X_n=j|X_0=i) = \lim_{n\to\infty}[P^n]_{i,j}$$

exists for all $i,j\in\mathbb{S}$ and these limits form a distribution on $\mathbb{S}$, i.e.,

$$\forall i\in\mathbb{S}: \sum_{j\in\mathbb{S}}\lim_{n\to\infty}\mathbb{P}(X_n=j|X_0=i)=1.$$

Further, we will assume that $\mathbb{S}$ is finite.

Theorem 6.5. *Let $\mathbb{S}$ be finite. If for some $i\in\mathbb{S}$ the limits $\nu_j := \lim_{n\to\infty}[P^n]_{i,j}$ exist for all $j\in\mathbb{S}$, then $\nu=(\nu_j)_{j\in\mathbb{S}}$ is a stationary distribution.*

Proof. We have

$$\sum_{j\in\mathbb{S}}\nu_j = \sum_{j\in\mathbb{S}}\lim_{n\to\infty}[P^n]_{i,j} = \lim_{n\to\infty}\sum_{j\in\mathbb{S}}[P^n]_{i,j}=1$$

and

$$\begin{aligned}\nu_j &= \lim_{n\to\infty}[P^{n+1}]_{i,j} = \lim_{n\to\infty}\sum_{k\in\mathbb{S}}[P^n]_{i,k}P_{k,j} = \sum_{k\in\mathbb{S}}\lim_{n\to\infty}[P^n]_{i,k}P_{k,j}\\ &= \sum_{k\in\mathbb{S}}\nu_k P_{k,j} = [\nu\cdot P]_j.\end{aligned}$$

□

[5] Again, only for $\mathbb{S}$ finite.

The following is one of the most important theorems about Markov chains - it studies the relation between limiting and stationary distribution. Let us recall that by Theorem 6.3 we know that irreducible and positive recurrent Markov chain has an uniquely determined stationary distribution.

Theorem 6.6. *Let $X = (X_n)_{n=0}^{\infty}$ be irreducible and aperiodic Markov chain and let us assume that it has a stationary distribution π. Let $\mathbb{P}(X_0 = j) = \lambda_j$ be an arbitrary distribution. Then*

$$\mathbb{P}(X_n = j) \to \pi_j$$

if $n \to \infty$ and this holds for every $j \in \mathbb{S}$. In particular, we have that $[P^n]_{i,j} \to \pi_j$ for every $i \in \mathbb{S}$.

Proof. We use the *coupling method.*

- We consider Markov chain $Y = (Y_n)_{n=0}^{\infty}$ with the initial distribution $\mathbb{P}(Y_0 = j) = \pi_j$ and the transition matrix P, i.e., with the same transition matrix as X.
- We take $b \in \mathbb{S}$ fixed and put $T := \inf\{n \geq 1 : X_n = Y_n = b\}$. This means that T is a random variable, which denotes the first time, when X and Y meet in b.
- We show that $\mathbb{P}(T < +\infty) = 1$. The process $W_n := (X_n, Y_n)$ is a Markov chain on the state space $\mathbb{S} \times \mathbb{S}$ with transition matrix

  $$\tilde{P}_{(i,k),(j,\ell)} = P_{i,j} \cdot P_{k,\ell}$$

  and the initial distribution $\mu_{i,k} = \lambda_i \pi_k$ (we assume that X and Y are independent on each other).
- We shall use the statement[6] that if Z is an irreducible aperiodic Markov chain on a state space $\mathbb{S}$, then for every $u, v \in \mathbb{S}$ there exists $n_0 \geq 1$ large enough such that $[P^n]_{u,v} > 0$ for every $n \geq n_0$.
- Therefore, for fixed $i, k, j, \ell \in \mathbb{S}$ we have that

  $$[\tilde{P}^n]_{(i,k),(j,\ell)} = [P^n]_{i,j} \cdot [P^n]_{k,\ell} > 0$$

  for n large enough. And, consequently, $W = (W_n)_{n=0}^{\infty}$ with the transition matrix $\tilde{P}$ is also aperiodic. It has a stationary distribution $\tilde{\pi}_{(i,k)} = \pi_i \cdot \pi_k$. This means that W is positive recurrent. Hence,[7] T is the time of the first passage of W through the state $(b, b) \in \mathbb{S}$ and $\mathbb{P}(T < +\infty) = 1$.
- Let us define

  $$Z_n = \begin{cases} X_n : & n < T, \\ Y_n : & n \geq T. \end{cases}$$

  Essentially, Z starts as X and in the moment, when it meets Y in b, it changes from X to Y.

[6]... which we (unfortunately) did not prove ...

[7]We use that irreducible positive recurrent Markov chain visits a fixed state with probability 1 in a finite time.

- As Y was started in the stationary distribution π, we have $\mathbb{P}(Y_n = j) = \mathbb{P}(Y_0 = j) = \pi_j$ for every $j \in \mathbb{S}$ and every $n \geq 0$. Furthermore, X and Z have the same initial distribution λ and the same transition matrix P.

- We estimate

$$\begin{aligned} |\mathbb{P}(X_n = j) - \pi_j| &= |\mathbb{P}(Z_n = j) - \mathbb{P}(Y_n = j)| \\ &= |\mathbb{P}(Z_n = j, n < T) + \mathbb{P}(Z_n = j, n \geq T) \\ &\qquad - \mathbb{P}(Y_n = j, n < T) - \mathbb{P}(Y_n = j, n \geq T)| \\ &= |\mathbb{P}(X_n = j, n < T) + \mathbb{P}(Y_n = j, n \geq T) \\ &\qquad - \mathbb{P}(Y_n = j, n < T) - \mathbb{P}(Y_n = j, n \geq T)| \\ &= |\mathbb{P}(X_n = j, n < T) - \mathbb{P}(Y_n = j, n < T)| \leq 2\mathbb{P}(n < T) \to 0 \end{aligned}$$

  as n tends to infinity.

□

7 Additional material and applications

In this section, we describe some additional material, which is closely connected to the theory of Markov processes, including some applications of Markov processes.

7.1 Algebraic methods

Markov chain X is fully described by its transition matrix P and the initial distribution, i.e., the distribution of X_0, which we denote $p^{(0)}$. If the state space is finite, say $\#\mathbb{S} = n$, then $P \in [0,1]^{n\times n} \subset \mathbb{R}^{n\times n}$ and it seems quite intuitive, that we could use the methods of linear algebra to study P and, indirectly, also of X.

First, let us observe that the vector $e = [1, \dots, 1]$ is the right eigenvector of P with the eigenvalue 1. Indeed, the identity

$$Pe^T = P \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = e^T$$

follows from the fact, that the row sum of P is one for every row. (☕) Show, that - therefore - there exists also a left eigenvector of P corresponding to the eigenvalue 1. If we denote it $\pi \neq 0$, this means that $\pi = \pi \cdot P$. If all the entries of π would be non-negative, then we can re-normalize π by taking

$$\frac{1}{\sum_{j=1}^{n} \pi_j} \cdot (\pi_1, \dots, \pi_n)$$

and this is a stationary distribution.

The non-negativity of the entries of an eigenvector is the subject of the *Perron-Frobenius theorem*. Its proof is slightly simpler if we assume that $A_{i,j} > 0$ instead of $A_{i,j} \geq 0$.

Theorem 7.1. *Let $A \in \mathbb{R}^{n\times n}$ be a matrix with $A_{i,j} > 0$ for all $1 \leq i, j \leq n$. Denote $\varrho = \max_j |\lambda_j|$ its spectral radius. Then $\varrho > 0$ and it holds*

1. *ϱ is an eigenvalue of A;*
2. *ϱ has algebraic multiplicity 1, i.e., $\det(\lambda I - A)$ has simple root in $\lambda = \varrho$;*
3. *There exists an eigenvector ν corresponding to ϱ, which has all its entries positive;*
4. *If $\lambda \neq \varrho$ is an eigenvalue of A, then $|\lambda| < \varrho$;*
5. *If u is an eigenvector of A with positive entries, then u is a multiple of ν.*

Proof. We shall present only the main ideas of the proof of some parts of the theorem.

- Let $A_{i,j} > 0$. Denote $S := \{x \in \mathbb{R}^n : \|x\|_2 = 1, x_j \geq 0 \text{ for all } 1 \leq j \leq n\}$. This means that S is the part of the unit sphere in $\mathbb{R}^n$, which consists of vectors with non-negative entries. Or, in another way, it is the intersection of the unit sphere with the cone of vectors with non-negative entries. Note, that it is compact.

- For $x \in S$, Ax has (strictly) positive entries.
- Define
$$L(x) = \min\left\{\frac{(Ax)_i}{x_i}, \text{ over } i\text{'s with } x_i \neq 0\right\}.$$
Note that $L(x)$ is the largest $\alpha > 0$ with $\alpha x_i \leq (Ax)_i$ for all $1 \leq i \leq n$.
- L is continuous on S and S is compact. Therefore, the maximal value of L on S exists and we denote it by $L(\nu) = \alpha$. We show that α is an eigenvalue, ν is the corresponding eigenvector and all its entries are (strictly) positive.
- Observe, that $L(\nu) = \alpha$ means, that $\alpha\nu_i \leq (A\nu)_i$, hence $A\nu - \alpha\nu \geq 0$ (in all coordinates). Let us assume, that $A\nu \neq \alpha\nu$. Then $A(A\nu - \alpha\nu) > 0$ and we can find $\varepsilon > 0$ small enough so that $A(A\nu - \alpha\nu) > \varepsilon A\nu$.

  Therefore, we have
$$\left[A\left(\frac{A\nu}{\|A\nu\|_2}\right)\right]_i = \frac{1}{\|A\nu\|_2}(A(A\nu))_i > \frac{\alpha+\varepsilon}{\|A\nu\|_2}(A\nu)_i = (\alpha+\varepsilon)\left(\frac{A\nu}{\|A\nu\|_2}\right)_i.$$
  This means that $L\left(\frac{A\nu}{\|A\nu\|_2}\right)_i \geq \alpha + \varepsilon$, which is a contradiction with maximality of the value of L in ν.
- As $\nu \geq 0$ (recall that $\nu \in S$), then $A\nu > 0$ and also $\nu > 0$, i.e., all the coordinates of ν are strictly positive.
- Next we show that $\varrho = \alpha$. Let $\mu \in \mathbb{C}$ be an eigenvalue of A and let $y \in \mathbb{C}^n$ be a corresponding eigenvector with $\|y\|_2 = 1$. Then
$$(\mu y)_i = (Ay)_i = \sum_{j=1}^n A_{i,j} y_j$$
  and it follows that
$$|\mu| \cdot |y_i| \leq \sum_{j=1}^n A_{i,j} \cdot |y_j|.$$
  This means that the vector z with $z_i = |y_i|$ satisfies $|\mu| \cdot z_i \leq (Az)_i$ and that $L(z) \geq |\mu|$. On the other hand, $L(z) \leq \alpha$ and we obtain $|\mu| \leq \alpha$. As α is an eigenvalue itself, we get $\varrho = \alpha$.

This gives 1) and 3) in the statement of the theorem - and we leave out the rest. □

Let us note that the theorem holds (with minor modifications) also for matrices A, for which $(A^k)_{i,j} > 0$ for all $1 \leq i, j \leq n$.

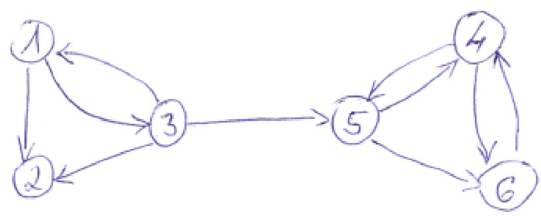
7.2 In this section, we present Google

The aim of this section is to give one prominent application of Markov processes, as it appeared in [1]. This fundamental paper starts with a historical statement:

> *In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext.*

Let us describe (very roughly) the main setting and ideas. We represent the World Wide Web by oriented graph, where the vertices correspond to web pages and the edges represent the links from a one page to another page.

Here comes a picture of a small WWW with 6 web pages:



The algorithm PageRank is based on the assumption, that important web pages link to important web pages. We consider a *random surfer*, who starts from a random web page and then continues randomly to further web pages according to the existing links. He/she chooses randomly one of the links on the current webpage and follows this randomly chosen link. In our example, the transition matrix would look like follows

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ ? & ? & ? & ? & ? & ? \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Obviously, this idea has a number of problems. For example, it is not clear how to incorporate the second web page, which does not have any links to other web pages. There are (at least) two ideas how to proceed. We could take the second line of P as (0 1 0 0 0 0), which would make it an absorbing state. Essentially, the random surfer would get stuck on this page. Or we could define the second row of P as (1/6 1/6 1/6 1/6 1/6 1/6), which would mean that the random surfer restarts and selects a random page of the model WWW.

The importance of a web page will be proportional to the time which the random surfer spends on this page in a long run. We shall therefore find a stationary distribution π, i.e., we solve the equation $\pi = \pi \cdot P$. Again, this idea might run into troubles. In our example, there are no links from the block $\{4,5,6\}$ to $\{1,2,3\}$. In some sense, the block $\{4,5,6\}$ is absorbing and the stationary distribution would be supported on $\{4,5,6\}$, making the webpages 1, 2, and 3 unimportant. To eliminate this (and other) troubles, we assume that the random surfer has a non-zero probability $p = 1-\alpha > 0$ of restart, i.e., we put

$$P' = \alpha P + \frac{1-\alpha}{n} e\, e^T,$$

where $e = (1, 1, \dots, 1)^T$ is the vector full of ones (and ee^T is a rank-1 matrix full of ones). In our example, with $\alpha = 0.9$, we would obtain

$$P' = \begin{pmatrix} 1/60 & 28/60 & 28/60 & 1/60 & 1/60 & 1/60 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 19/60 & 19/60 & 1/60 & 1/60 & 19/60 & 1/60 \\ 1/60 & 1/60 & 1/60 & 1/60 & 28/60 & 28/60 \\ 1/60 & 1/60 & 1/60 & 28/60 & 1/60 & 28/60 \\ 1/60 & 1/60 & 1/60 & 55/60 & 1/60 & 1/60 \end{pmatrix}.$$

The importantce of a certan web page will then be given by the size of the corresponding entry of the stationary distribution π' with $\pi' = \pi' \cdot P'$. Under mild assumptions (satisfied for P'), π' is unique and has all entries positive. The solution of the equation $\pi' = \pi' \cdot P'$ is computationaly infeasible (as P' is an $n \times n$ matrix with n of the order 10^{10}). But we can approximate π' iteratively. We start with an arbitrary distribution ν and calculate $\nu \cdot P'$, $\nu \cdot (P')^2$, etc. The iterations are still computationally difficult but already feasible (P' is a sum of a simple matrix and a sparse matrix).

Finally, let us note that the speed of the convergence depends on the second eigenvalue λ_2 of P' (with the simple largest eigenvalue of P' being $\lambda_1 = 1$). Indeed, the eigenvalues of $(P')^r$ are $\lambda_1^r = 1$, λ_2^r, etc. and the smaller is $|\lambda_2|$, the smaller is also $|\lambda_2|^r$.

7.3 Markov Chain Monte Carlo (MCMC)

The next application of Markov processes is the so-called *Markov Chain Monte Carlo* method. Before we come to that, we first recall the Monte Carlo method, which essentially refers to a calculation of deterministic quantities by a randomized algorithm. As a motivating example, we consider a (measurable) set $\Omega \subset [0,1]^d$. We would like to calculate $|\Omega|$, the Lebesgue measure of Ω. We assume that

- The measure $|\Omega| = \displaystyle\int_{[0,1]^d} \chi_\Omega(x)dx$ is difficult to calculate directly,
- but for a fixed point $x \in [0,1]^d$, it is easy to decide if $x \in \Omega$ or $x \notin \Omega$.

We approximate $|\Omega| = \displaystyle\int_\Omega 1dx = \int_{[0,1]^d} \chi_\Omega(x)dx$ by $\displaystyle\frac{1}{n}\sum_{i=1}^n \chi_\Omega(x_i)$, where $x_i \in [0,1]^d$ are chosen independently a uniformly distributed at random.

If we define the random variable X by

$$X = \chi_\Omega(x) = \begin{cases} 1, & \text{if } x \in \Omega, \\ 0, & \text{if } x \notin \Omega, \end{cases}$$

where $x \in [0,1]^d$ is chosen uniformly distributed in $[0,1]^d$, then we essentially replaced X by $\displaystyle\frac{1}{n}\sum_{j=1}^n X_j$, where $X_1, \dots, X_n$ are independent copies of X.

The averge error of this algorithm is easily calculated

$$\mathbb{E}\left|\mathbb{E}X-\frac{1}{n}\sum_{j=1}^{n}X_j\right|^2=\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}(\mathbb{E}X-X_j)\right|^2=\frac{1}{n^2}\mathbb{E}\sum_{j,k=1}^{n}(\mathbb{E}X-X_j)(\mathbb{E}X-X_k)$$
$$=\frac{1}{n^2}\sum_{j=1}^{n}\mathbb{E}(\mathbb{E}X-X_j)^2=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}.$$

The average error of the algorithm (in the L_2-sense) is therefore $\sigma/\sqrt{n}$, where $\sigma^2=\text{var}(X)$ and n is the number of repetitions used.

Let us give some examples of the Monte Carlo method

1. Calculation of π: Let $\Omega\subset[0,1]^2$, where $\Omega=\{(x_1,x_2)\in[0,1]^2:x_1^2+x_2^2\le 1\}$. We choose $(x_1,x_2)\in[0,1]^2$ randomly and uniformly distributed and define
$$X=\begin{cases}1 & \text{if } x\in\Omega,\\ 0 & \text{if } x\notin\Omega.\end{cases}$$
Then $\mathbb{E}X=|\Omega|=\pi/4$ can be approximated by $\frac{1}{n}\sum_{j=1}^{n}X_j$ up to a precision of $\asymp\frac{1}{\sqrt{n}}$.

2. Similarly, we can handle the *Buffon's needle problem.* In this experiment, we let a needle of an unit length fall down on a sheet of paper with parallel lines of unit mutual distance. We are interested in the probability, that the needle crosses one of the lines. If we denote by $x\in[0,1/2]$ the distance of the middle point of the needle from the closest line and by $\alpha\in[0,\pi/2]$ the angle between the needle and the line, then x and α are chosen uniformly and the set of parameters, where the needle intersects the line is
$$B=\{(x,\alpha)\in[0,1/2]\times[0,\pi/4]:0\le x\le(\sin\alpha)/2\}$$
and its measure is $\mathbb{P}(B)=\dfrac{4}{\pi}\displaystyle\int_0^{\pi/2}(\sin\alpha)/2d\alpha=2/\pi.$

3. Properties of random polygons (Sylvester, 1864): For a convex set $K\subset\mathbb{R}^2$, we choose $x_1,\dots,x_4\in K$ independent and uniformly distributed over K. What is the probability, that $\text{conv}(x_1,x_2,x_3,x_4)$ is a triangle?

   To answer the question it is enough to calculate the mean value of the area of $\text{conv}(x_1,x_2,x_3)$. This leads to an integral of a function of six variables. For some K's the exact value can be comupted analytically and is known, but for a general K it can be approximated by the Monte Carlo method.

4. If we want to calculate the integral
$$\int_{\mathbb{R}^d}f(x)\psi(x)dx,$$
where $\psi(x)$ is a density on $\mathbb{R}^d$, we can evaluate $\dfrac{1}{n}\displaystyle\sum_{j=1}^{n}f(x_j)$, where x_j's a generated randomly according to ψ. Then $X=f(x)$ and $\mathbb{E}X=\int f(x)\psi(x)dx$.

To apply the Monte Carlo method, it is necessary to generate random samples according to the given distribution. We produce a Markov chain, which will have the given distribution as its stationary distribution, we initialize this chain in an arbitrary way and then we take the value of X_n for some large n. Hence, for a given (large but finite) set $\mathbb{S}$ and a given distribution π on $\mathbb{S}$, we look for a Markov chain X on state space $\mathbb{S}$ such that

- π is the stationary distribution of X,
- the distribution of X_n converges quickly to π if $n \to \infty$,
- and which might possibly satisfy also other conditions, i.e., that the step from X_n to X_{n+1} can be performed quickly.

We describe the so-called *Metropolis algorithm*. If π and $\mathbb{S}$ are given and if we already have some Markov chain with (symmetric) transition matrix $\Psi = (\Psi_{x,y})_{x,y\in\mathbb{S}}$. We modify this Markov chain in such a way that its stationary distribution becomes π. We accept the transition from x to y with probability $a(x,y)$ and we refuse this transition (i.e., we stay at x) with probability $1 - a(x,y)$. We therefore define

$$P_{x,y} = \begin{cases} \Psi_{x,y} a(x,y), & \text{if } y \neq x, \\ 1 - \sum_{z:z\neq x} \Psi_{x,z} a(x,z), & \text{if } y = x. \end{cases}$$

We choose $a(x,y)$ to ensure that P and π satisfy the detailed balance equations. Consequently, we will know that π is the stationary distribution of P. This means that we require

$$\pi(x)P_{x,y} = \pi(y)P_{y,x} \quad x \neq y$$

and, equivalently

$$\pi(x)\Psi_{x,y}a(x,y) = \pi(y)\Psi_{y,x}a(y,x), \quad x \neq y.$$

Due to the symmetry of Ψ, this can be further simplified to

$$\pi(x)a(x,y) = \pi(y)a(y,x), \quad x \neq y.$$

We want to choose $0 \leq a(x,y) \leq 1$ and $0 \leq a(y,x) \leq 1$ as large as possible (small $a(x,y)$ reduces the speed of X and also its speed of convergence to π). Therefore, we define

$$\pi(x)a(x,y) = \pi(y)a(y,x) = \min(\pi(x), \pi(y)), \quad x \neq y,$$

i.e.,

$$a(x,y) = \frac{\min(\pi(x), \pi(y))}{\pi(x)} = \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$$

and

$$a(y,x) = \frac{\min(\pi(x), \pi(y))}{\pi(y)} = \min\left(1, \frac{\pi(x)}{\pi(y)}\right).$$

Altogther, we define

$$P_{x,y} = \begin{cases} \Psi_{x,y} \min\left(1, \frac{\pi(y)}{\pi(x)}\right), & \text{if } y \neq x, \\ 1 - \sum_{z:z\neq x} \Psi_{x,z} \min\left(1, \frac{\pi(y)}{\pi(z)}\right), & \text{if } y = x. \end{cases}$$

7.4 Ising model?

7.5 Strong Markov property

References

[1] S. Brin, and L. Page, *The anatomy of a large-scale hypertextual web search engine*, Computer networks and ISDN systems 30, no. 1-7 (1998): 107-117.

[2] D. A. Levin, Y. Peres, and E. Wilmer, Markov chains and mixing times, American Mathematical Society, Providence, RI, 2009

[3] Nicolas Privault, Understanding Markov chains, Springer Undergrad. Math. Ser., Springer Singapore, Singapore, 2013