Course:	Introduction to Stochastic Processes
Term:	Fall 2019
Instructor:	Gordan Žitković

Lecture 7
Markov chains

7.1 The Markov property

Simply put, a stochastic process has the **Markov property** if probabilities governing its future evolution depend only on its current position, and not on how it got there. Here is a more precise, mathematical, definition. It will be assumed throughout this course that any stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ takes values in a countable set S - the **state space**. Usually, S will be either finite, $\mathbb{N}_0$ (as in the case of branching processes) or $\mathbb{Z}$ (random walks). Sometimes, a more general, *but still countable*, state space S will be needed. A generic element of S will be denoted by i or j.

Definition 7.1.1. A stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ taking values in a countable state space S is called a **Markov chain** (or is said to have the **Markov property**) if

$$\begin{aligned}\mathbb{P}[X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0] = \\ = \mathbb{P}[X_{n+1} = j|X_n = i], \quad (7.1.1)\end{aligned}$$

for all $n \in \mathbb{N}_0$, all $i, j, i_0, i_1, \dots, i_{n-1} \in S$, whenever the two conditional probabilities are well-defined, i.e., when $\mathbb{P}[X_n = i, \dots, X_1 = i_1, X_0 = i_0] > 0$.

The Markov property is typically checked in the following way: one computes the left-hand side of (7.1.1) and shows that its value does not depend on $i_{n-1}, i_{n-2}, \dots, i_1, i_0$ (why is that enough?). The condition $\mathbb{P}[X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] > 0$ will be assumed (without explicit mention) every time we write a conditional expression like to one in (7.1.1).

All chains in this course will be *homogeneous*, i.e., the conditional probabilities $\mathbb{P}[X_{n+1} = j|X_n = i]$ will not depend on the current time $n \in \mathbb{N}_0$, i.e., $\mathbb{P}[X_{n+1} = j|X_n = i] = \mathbb{P}[X_{m+1} = j|X_m = i]$, for $m, n \in \mathbb{N}_0$.

Markov chains are (relatively) easy to work with because the Markov property allows us to compute all the probabilities, expectations, etc. we might be interested in by using only two ingredients.

1. **Initial probability** $\boldsymbol{a}^{(0)} = \{a_i^{(0)} : i \in S\}$, $a_i^{(0)} = \mathbb{P}[X_0 = i]$ - the initial probability distribution of the process, and
2. **Transition probabilities** $p_{ij} = \mathbb{P}[X_{n+1} = j | X_n = i]$ - the mechanism that the process uses to jump around.

Indeed, if one knows all $a_i^{(0)}$ and all p_{ij}, and wants to compute a joint distribution $\mathbb{P}[X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0]$, one needs to use the definition of conditional probability and the Markov property several times (the *multiplication theorem* from your elementary probability course) to get

$$\begin{aligned}
&\mathbb{P}[X_n = i_n, \dots, X_0 = i_0] \\
&\quad = \mathbb{P}[X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0] \mathbb{P}[X_{n-1} = i_{n-1}, \dots, X_0 = i_0] \\
&\quad = \mathbb{P}[X_n = i_n | X_{n-1} = i_{n-1}] \mathbb{P}[X_{n-1} = i_{n-1}, \dots, X_0 = i_0] \\
&\quad = p_{i_{n-1} i_n} \mathbb{P}[X_{n-1} = i_{n-1}, \dots, X_0 = i_0]
\end{aligned}$$

Repeating the same procedure, we get

$$\mathbb{P}[X_n = i_n, \dots, X_0 = i_0] = p_{i_{n-1} i_n} \times p_{i_{n-2} i_{n-1}} \times \dots \times p_{i_0 i_1} \times a_{i_0}^{(0)}.$$

When S is finite, there is no loss of generality in assuming that $S = \{1, 2, \dots, n\}$, and then we usually organize the entries of $a^{(0)}$ into a row vector

$$\boldsymbol{a}^{(0)} = (a_1^{(0)}, a_2^{(0)}, \dots, a_n^{(0)}),$$

and the transition probabilities p_{ij} into a square matrix $\mathbf{P}$, where

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{bmatrix}$$

In the general case (S possibly infinite), one can still use the vector and matrix notation as before, but it becomes quite clumsy in the general case. For example, if $S = \bar{Z}$, $\mathbf{P}$ is an infinite matrix

$$\mathbf{P} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \iddots \\ \dots & p_{-1\,-1} & p_{-1\,0} & p_{-1\,1} & \dots \\ \dots & p_{0\,-1} & p_{0\,0} & p_{0\,1} & \dots \\ \dots & p_{1\,-1} & p_{1\,0} & p_{1\,1} & \dots \\ \iddots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

7.2 Examples

Here are some examples of Markov chains - for each one we write down the transition matrix. The initial distribution is sometimes left unspecified because it does not really change anything.

1. Random walks Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple (possibly biased) random walk. Let us show that it indeed has the Markov property (7.1.1). Remember, first, that

$$X_n = \sum_{k=1}^{n} \delta_k \text{ where } \delta_k \text{ are } \textit{independent} \text{ (possibly biased) coin-tosses.}$$

For a choice of $i_0, \dots, i_n, j = i_{n+1}$ (such that $i_0 = 0$ and $i_{k+1} - i_k = \pm 1$) we have

$$\begin{aligned}
&\mathbb{P}[X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0] \\
=&\mathbb{P}[X_{n+1} - X_n = i_{n+1} - i_n | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0] \\
=&\mathbb{P}[\delta_{n+1} = i_{n+1} - i_n | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0] \\
=&\mathbb{P}[\delta_{n+1} = i_{n+1} - i_n],
\end{aligned}$$

where the last equality follows from the fact that the increment δ_{n+1} is independent of the previous increments, and, therefore, also of the values of $X_1, X_2, \dots, X_n$. The last line above does not depend on $i_{n-1}, \dots, i_1, i_0$, so X indeed has the Markov property.

The state space S of $\{X_n\}_{n\in\mathbb{N}_0}$ is the set $\bar{Z}$ of all integers, and the initial distribution $\boldsymbol{a}^{(0)}$ is very simple: we start at 0 with probability 1 (so that $a_0^{(0)} = 1$ and $a_i^{(0)} = 0$, for $i \neq 0$.). The transition probabilities are simple to write down

$$p_{ij} = \begin{cases} p, & j = i+1 \\ q, & j = i-1 \\ 0, & \text{otherwise.} \end{cases}$$

These can be written down in an infinite matrix,

$$\mathbf{P} = \begin{bmatrix}
\ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\
\dots & 0 & p & 0 & 0 & 0 & \dots \\
\dots & q & 0 & p & 0 & 0 & \dots \\
\dots & 0 & q & 0 & p & 0 & \dots \\
\dots & 0 & 0 & q & 0 & p & \dots \\
\dots & 0 & 0 & 0 & q & 0 & \dots \\
\dots & 0 & 0 & 0 & 0 & q & \dots \\
\iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}$$

but is typically not as useful as the matrix representation in the finite case.

2. Branching processes Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a branching process with the offspring distribution $\{p_n\}_{n\in\mathbb{N}_0}$. As you surely remember, it is constructed as follows[1] : $X_0 = 1$ and $X_{n+1} = \sum_{k=1}^{X_n} X_{n,k}$, where $\{X_{n,k}\}_{n\in\mathbb{N}_0, k\in\mathbb{N}}$ is a family

[1] We switch the notation from Z to X to make it uniform across all examples.

of independent random variables with distribution $\{p_n\}_{n\in\mathbb{N}_0}$. It is now not very difficult to show that $\{X_n\}_{n\in\mathbb{N}_0}$ is a Markov chain

$$
\begin{aligned}
&\mathbb{P}[X_{n+1}=j|X_n=i_n,X_{n-1}=i_{n-1},\dots,X_1=i_1,X_0=i_0]\\
=&\mathbb{P}[\sum_{k=1}^{X_n} X_{n,k}=j|X_n=i_n,X_{n-1}=i_{n-1},\dots,X_1=i_1,X_0=i_0]\\
=&\mathbb{P}[\sum_{k=1}^{i_n} X_{n,k}=j|X_n=i_n,X_{n-1}=i_{n-1},\dots,X_1=i_1,X_0=i_0]\\
=&\mathbb{P}[\sum_{k=1}^{i_n} X_{n,k}=j],
\end{aligned}
$$

where, just like in the random-walk case, the last equality follows from the fact that the random variables $X_{n,k}$, $k\in\mathbb{N}$ are independent of all $X_{m,k}$, $m<n$, $k\in\mathbb{N}$. In particular, they are independent of $X_n, X_{n-1},\dots,X_1,X_0$, which are obtained as combinations of $X_{m,k}$, $m<n$, $k\in\mathbb{N}$. The computation above also reveals the structure of the transition probabilities, p_{ij}, $i,j\in S=\mathbb{N}_0$:

$$p_{ij}=\mathbb{P}[\sum_{k=1}^{i} X_{n,k}=j].$$

There is little we can do to make the expression above more explicit, but we can use generating functions and write $P_i(s)=\sum_{j=0}^\infty p_{ij}s^j$ (remember that each row of the transition matrix is a probability distribution). Thus, $P_i(s)=(P(s))^i$ (why?), where $P(s)=\sum_{k=0}^\infty p_k s^k$ is the generating function of the branching probability. Since we always start from a single individual, the initial distribution is given by

$$a_i^{(0)}=\begin{cases}1, & i=1,\\ 0, & i\neq 1.\end{cases}$$

3. Gambler's ruin In Gambler's ruin, a gambler starts with \$$x$, where $0\leq x\leq a\in\mathbb{N}$ and in each play wins a dollar (with probability $p\in(0,1)$) and loses a dollar (with probability $q=1-p$). When the gambler reaches either 0 or a, the game stops. For mathematical convenience, it is usually a good idea to keep the chain defined, even after the modeled phenomenon stops. This is usually accomplished by **absorption**. We simply assume that the process "stays alive" but remains "frozen" instead of disappearing. In our case, once our gambler reaches either of the states 0 and a, he/she simply stays there forever.

Therefore, the transition probabilities are similar to those of a random walk, but differ from them at the boundaries 0 and a. The state space is finite

$S = \{0, 1, \dots, a\}$ and the matrix $\mathbf{P}$ is given by

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

The initial distribution is deterministic:

$$a_i^{(0)} = \begin{cases} 1, & i = x, \\ 0, & i \neq 1. \end{cases}$$

4. Regime Switching Consider a system with two different states; think about a simple weather forcast (rain/no rain), high/low water level in a reservoire, high/low volatility regime in a financial market, high/low level of economic growth, the political party in power, etc. Suppose that the states are called 1 and 2 and the probabilities p_{12} and p_{21} of switching states are given. The probabilities $p_{11} = 1 - p_{12}$ and $p_{22} = 1 - p_{21}$ correspond to the system staying in the same state. The transition matrix for this Markov with $S = \{1, 2\}$ is

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22}. \end{bmatrix}$$

When p_{12} and p_{21} are large (close to 1) the system nervously jumps between the two states. When they are small, there are long periods of stability (staying in the same state).

One of the assumptions behind regime-switching models is that the transitions (switches) can only happen in regular intervals (once a minute, once a day, once a year, etc.). This is a feature of all *discrete-time* Markov chains. One would need to use a *continuous-time* model to allow for the transitions between states at any point in time.

5. Deterministically monotone Markov chain A stochastic process $\{X_n\}_{n \in \mathbb{N}_0}$ with state space $S = \mathbb{N}_0$ such that $X_n = n$ for $n \in \mathbb{N}_0$ (no randomness here) is called Deterministically monotone Markov chain (DMMC). The transition matrix looks like this

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

It is a pretty boring chain; it main use is as a counterexample.

6. Not a Markov chain Consider a frog jumping from a lily pad to a lily pad in a small forest pond. Suppose that there are N lily pads so that the state space can be described as $S = \{1, 2, \dots, N\}$. The frog starts on lily pad 1 at time $n = 0$, and jumps around in the following fashion: at time 0 it chooses any lily pad except for the one it is currently sitting on (with equal probability) and then jumps to it. At time $n > 0$, it chooses any lily pad other than the one it is sitting on *and the one it visited immediately before* (with equal probability) and jumps to it. The position $\{X_n\}_{n\in\mathbb{N}_0}$ of the frog is not a Markov chain. Indeed, we have

$$\mathbb{P}[X_3 = 1 | X_2 = 2, X_1 = 3] = \frac{1}{N-2},$$

while

$$\mathbb{P}[X_3 = 1 | X_2 = 2, X_1 = 1] = 0.$$

A more dramatic version of this example would be the one where the frog remembers all the lily pads it had visited before, and only chooses among the remaining ones for the next jump.

7. Making a non-Markov chain into a Markov chain How can we turn the process of Example 6. into a Markov chain. Obviously, the problem is that the frog has to remember the number of the lily pad it came from in order to decide where to jump next. The way out is to make this information a part of the state. In other words, we need to change the state space. Instead of just $S = \{1, 2, \dots, N\}$, we set $S = \{(i_1, i_2) : i_1, j_2 \in \{1, 2, \dots N\}\}$. In words, the state of the process will now contain not only the number of the current lily pad (i.e., i_1) but also the number of the lily pad we came from (i.e., i_2). This way, the frog will be in the state (i_1, i_2) if it is currently on the lily pad number i_2, and it arrived here from i_1. There is a bit of freedom with the initial state, but we simply assume that we start from $(1, 1)$. Starting from the state (i_1, i_2), the frog can jump to any state of the form (i_2, i_3), $i_3 \neq i_1, i_2$ (with equal probabilities). Note that some states will never be visited (like (i, i) for $i \neq 1$), so we could have reduced the state space a little bit right from the start.

8. A more complicated example (*) Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk. The absolute-value process $Y_n = |X_n|$, $n \in \mathbb{N}_0$, is also a Markov chain. This processes is sometimes called the **reflected random walk**.

In order to establish the Markov property, we let $i_0, \dots, i_n, j = i_{n+1}$ be non-negative integers with $i_{k+1} - i_k = \pm 1$ for all $0 \le k \le n$ (the state space is $S = \mathbb{N}_0$). We need to show that the conditional probability

$$\mathbb{P}\left[|X_{n+1}| = j \,\middle|\, |X_n| = i_n, \dots, |X_0| = i_0\right] \tag{7.2.1}$$

does not depend on $i_{n-1}, \dots, i_0$. We start by splitting the event $A = \{|X_{n+1}| = j\}$ into two parts:

$$\mathbb{P}[|X_{n+1}| = j|B] = \mathbb{P}[A^+|B] + \mathbb{P}[A^-|B],$$

where $A^+ = \{X_{n+1} = j\}$, $A^- = \{X_{n+1} = -j\}$ and

$$B = \{|X_n| = i_n, |X_{n-1}| = i_{n-1}, \dots, |X_0| = i_0\}.$$

We assume that $i_n > 0$; the case $i_n = 0$ is similar, but easier so we leave it to the reader.

The event B is composed of all trajectories of the (original) random walk X whose absolute values are given by $i_0, \dots, i_n$. Depending on how many times $i_k = 0$ for some $k = 1, \dots, n$, there will be 2 or more such trajectories (draw a picture!) - let us denote the events corresponding to those single trajectories of X by $B_1, \dots, B_N$. In other words, each B_l, $l = 1, \dots, N$ looks like

$$B_l = \{X_0 = i'_0, X_1 = i'_1, \dots, X_n = i'_n\},$$

where $i'_k = i_k$ or $-i_k$ and $i'_{k+1} - i'_k = \pm 1$, for all k. These trajectories can be divided into two groups; those with $i'_n = i_n$ and those with $i'_n = -i_n$; let us label them so that $B_1, \dots, B_m$ correspond to the first group and $B_{m+1}, B_{m+2}, \dots, B_N$ to the second. Since $i_n > 0$, the two groups are disjoint and we can flip each trajectory in the first group to get one in the second and vice versa. Therefore, $N = 2m$ and

$$\tfrac{1}{2}\mathbb{P}[B] = \sum_{l=1}^{m} \mathbb{P}[B_l].$$

The conditional probability $\mathbb{P}[A^+|B_l]$ is either equal to $\frac{1}{2}$ or to 0, depending on whether $l \leq m$ or $l > m$ (why?). Therefore,

$$\begin{aligned}\mathbb{P}[A^+|B] &= \tfrac{1}{\mathbb{P}[B]}\mathbb{P}[A^+ \cap (B_1 \cup \dots \cup B_N)] = \tfrac{1}{\mathbb{P}[B]} \sum_{l=1}^{N} \mathbb{P}[A^+ \cap B_l] \\ &= \tfrac{1}{\mathbb{P}[B]} \sum_{l=1}^{N} \mathbb{P}[A^+|B_l]\mathbb{P}[B_l] = \frac{\frac{1}{2}\sum_{l=1}^{m} \mathbb{P}[B_l]}{\mathbb{P}[B]} = \tfrac{1}{4}.\end{aligned}$$

Similarly, $\mathbb{P}[A^-|B] = \frac{1}{4}$, which implies that

$$\mathbb{P}[A|B] = \mathbb{P}[A||X_n| = i_n],$$

and the Markov property follows. It may look like $\mathbb{P}[A|B]$ it is also independent of i_n but it is not; this probability is equal to 0 unless $|j| - |i_n| = \pm 1$).

9. A function of the simple symmetric random walk which is not a Markov chain Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov chain on the state space S, and let $f : S \to T$ be a function. The stochastic process $Y_n = f(X_n)$ takes values in T; is it necessarily a Markov chain? We will see in this example that the answer is *no*. Before we present it, we note that we already encountered this situation in example 8. above. In it $\{X_n\}_{n\in\mathbb{N}_0}$ is the simple symmetric random walk, $f(x) = |x|$ and $T = \mathbb{N}_0$. In that, particular, case we have shown that $Y = f(X)$ has the Markov property. Let us keep the process X, but change the function f. First, let

$$R_n = X_n \ (\text{mod } 3) = \begin{cases} 0, & \text{if } X_n \text{ is divisible by 3,} \\ 1, & \text{if } X_n - 1 \text{ is divisible by 3,} \\ 2, & \text{if } X_n - 2 \text{ is divisible by 3,} \end{cases}$$

be the remainder obtained when X_n is divided by 3, and let $Y_n = (X_n - R_n)/3$ be the quotient, so that $Y_n \in \bar{Z}$ and $3Y_n \leq X_n < 3(Y_n + 1)$. Clearly, $Y_n = f(X_n)$, where $f(i) = \lfloor i/3 \rfloor$, where $\lfloor x \rfloor$ is the largest integer not exceeding x.

To show that Y is not a Markov chain, let us consider the the event $A = \{Y_2 = 0, Y_1 = 0\}$. The only way for this to happen is if $X_1 = 1$ and $X_2 = 2$ or $X_1 = 1$ and $X_2 = 0$, so that $A = \{X_1 = 1\}$. Also $Y_3 = 1$ if and only if $X_3 = 3$. Therefore

$$\mathbb{P}[Y_3 = 1 | Y_2 = 0, Y_1 = 0] = \mathbb{P}[X_3 = 3 | X_1 = 1] = 1/4.$$

On the other hand, $Y_2 = 0$ if and only if $X_2 = 0$ or $X_2 = 2$, so $\mathbb{P}[Y_2 = 0] = 3/4$. Finally, $Y_3 = 1$ and $Y_2 = 0$ if and only if $X_3 = 3$ and so $\mathbb{P}[Y_3 = 1, Y_2 = 0] = 1/8$. Therefore

$$\mathbb{P}[Y_3 = 1 | Y_2 = 0] = \mathbb{P}[Y_3 = 1, Y_2 = 0] / \mathbb{P}[Y_2 = 0] = \frac{1/8}{3/4} = \frac{1}{6}.$$

Therefore, Y is not a Markov chain. Try to modify this example so that one of the probabilities above is positive, but the other is zero.

10. A more realistic example. In a game of tennis, the scoring system is as follows: both players (let us call them Serena and Roger) start with the score of 0. Each time Serena wins a point (a.k.a. *rally*), her score moves a step up in the following hierarchy

$$0 \mapsto 15 \mapsto 30 \mapsto 40.$$

Once Serena reaches 40 and scores a point, three things can happen:

1. if Roger's score is 30 or less, Serena wins the game.
2. if Roger's score is 40, Serena's score moves up to "*advantage*", and

3. if Roger's score is "*advantage*", nothing happens to Serena's score, but Roger's score falls back to 40.

Finally, if Serena's score is "*advantage*" and she wins a point, she wins the game. The situation is entirely symmetric for Roger We suppose that the probability that Serena wins each point is $p \in (0, 1)$, independently of the current score. A situation like this is a typical example of a Markov chain in an applied setting. What are the states of the process? We obviously need to know both players' scores and we also need to know if one of the players has won the game. Therefore, a possible state space is the following:

$$S = \Big\{\text{0-0}, \text{0-15}, \text{0-30}, \text{0-40}, \text{15-0}, \text{15-15}, \text{15-30}, \text{15-40}, \text{30-0}, \text{30-15}, \text{30-30}, \text{30-40},$$

$$\text{40-0}, \text{40-15}, \text{40-30}, \text{40-40}, \text{40-A}, \text{A-40}, S_{win}, R_{win}\Big\},$$

where A stands for "*advantage*" and S_{win} (R_{win}) denotes the state where Serena (Roger) wins. It is not hard to assign probabilities to transitions between states. Once we reach either S_{win} or R_{win} the game stops. We can assume that the chain remains in that state forever, i.e., the state is absorbing. The initial distribution is quite simple - we aways start from the same state 0-0, so that $a^{(0)}_{0\text{-}0} = 1$ and $a^{(0)}_i = 0$ for all $i \in S \setminus \{\text{0-0}\}$.

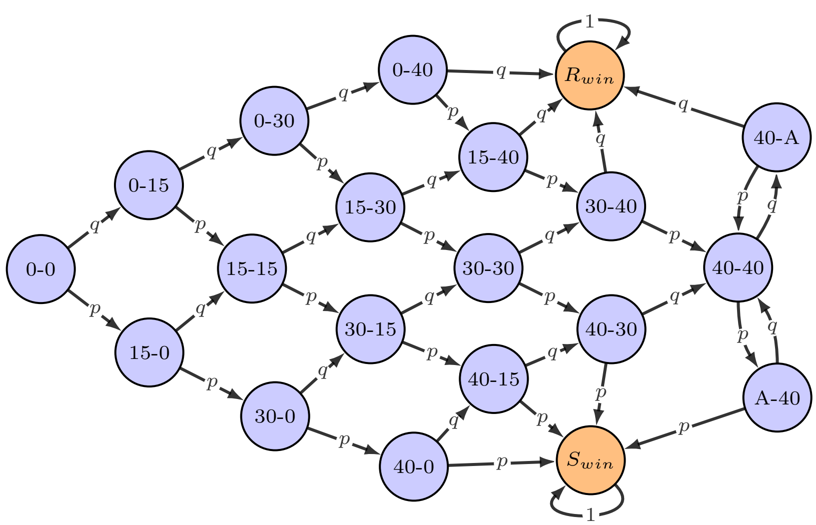

Figure 1. Markov chains with a finite number of states are usually represented by directed graphs (like the one in the figure above). The nodes are states, two states i, j are linked by a (directed) edge if the transition probability p_{ij} is non-zero, and the number p_{ij} is written above the link. If $p_{ij} = 0$, no edge is drawn.

How about the transition matrix? When the number of states is big ($\#S = 20$ in this case), transition matrices are useful in computer memory, but not so much on paper. Just for the fun of it, here is the transition matrix for our game-of-tennis chain, with the states ordered as in the definition of the set S above:

$$\mathbf{P} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & q & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & q & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & q & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & q & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & q & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0 & p & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0 & p & 0 & 0 & 0 \\
0 & q & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p & 0 & 0 \\
p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0 & 0 & 0 \\
p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0 & 0 \\
p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & p \\
0 & q & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p & 0 & 0 \\
p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0 & 0
\end{bmatrix}$$

Question 7.2.1. Does the structure of a game of tennis make is easier or harder for the better player to win? In other words, if you had to play against Roger Federer (I am rudely assuming that he is better than you), would you have a better chance of winning if you only played a point (rally), or if you actually played the whole game?

We will give a precise answer to this question in a little while. In the meantime, try to guess.

7.3 Chapman-Kolmogorov relations

The transition probabilities p_{ij}, $i, j \in S$ tell us how a Markov chain jumps from a state to a state in one step. How about several steps, i.e., how does one compute the probabilities like $\mathbb{P}[X_{k+n} = j | X_k = i]$, $n \in \mathbb{N}$? Since we are assuming that all of our chains are homogeneous (transition probabilities do not change with time), this probability does not depend on the time k, and we set

$$p_{ij}^{(n)} = \mathbb{P}[X_{k+n} = j | X_k = i] = \mathbb{P}[X_n = j | X_0 = i].$$

It is sometimes useful to have a more compact notation for this, last, conditional probability, so we write

$$\mathbb{P}_i[A] = \mathbb{P}[A | X_0 = i], \text{ for any event } A.$$

Therefore,

$$p_{ij}^{(n)} = \mathbb{P}_i[X_n = j].$$

For $n = 0$, we clearly have

$$p_{ij}^{(0)} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Once we have defined the **multi-step transition probabilities** $p^{(n)}_{ij}$, $i, j \in S$, $n \in \mathbb{N}_0$, we need to be able to compute them. This computation is central in various applications of Markov chains: they relate the small-time (one-step) behavior which is usually easy to observe and model to a long-time (multi-step) behavior which is really of interest. Before we state the main result in this direction, let us remember how matrices are multiplied. When A and B are $N \times N$ matrices, the product $C = AB$ is also an $N \times N$ matrix and its ij-entry C_{ij} is given as

$$C_{ij} = \sum_{k=1}^{N} A_{ik} B_{kj}.$$

There is nothing special about finiteness in the above definition. If A and B were infinite matrices $A = (A_{ij})_{i,j \in S}$, $B = (B_{ij})_{i,j \in S}$ for some countable set S, the same procedure could be used to define $C = AB$. Indeed, C will also be an "$S \times S$"-matrix and

$$C_{ij} = \sum_{k \in S} A_{ik} B_{kj},$$

as long as the (infinite) sum above converges absolutely. In the case of a typical transition matrix $\mathbf{P}$, convergence will not be a problem since $\mathbf{P}$ is a **stochastic matrix**, i.e., it has the following two properties (why?):

1. $p_{ij} \geq 0$, for all $i, j \in S$, and
2. $\sum_{j \in S} p_{ij} = 1$, for all $i \in S$ (in particular, $p_{ij} \in [0, 1]$, for all i, j).

When $\mathbf{P} = (p_{ij})_{i,j \in S}$ and $\mathbf{P}' = (p'_{ij})_{i,j \in S}$ are two $S \times S$-stochastic matrices, the series $\sum_{k \in S} p_{ik} p'_{kj}$ converges absolutely since $0 \leq p'_{kj} \leq 1$ for all $k, j \in S$ and so

$$\sum_{k \in S} |p_{ik} p'_{kj}| \leq \sum_{k \in S} p_{ik} \leq 1, \text{ for all } i, j \in S.$$

Moreover, a product C of two stochastic matrices A and B is always a stochastic matrix: the entries of C are clearly positive and

$$\sum_{j \in S} C_{ij} = \sum_{j \in S} \sum_{k \in S} A_{ik} B_{kj} = \sum_{k \in S} \sum_{j \in S} A_{ik} B_{kj} = \sum_{k \in S} A_{ik} \underbrace{\sum_{j \in S} B_{kj}}_{1} = \sum_{k \in S} A_{ik} = 1.$$

The proof of fact that we can switch the j-sum and the k-sum in the second step above is beyond the scope of these notes (and follows from a general result called Fubini-Tonelli theorem).

Proposition 7.3.1. *Let $\mathbf{P}^n$ be the n-th (matrix) power of the transition matrix $\mathbf{P}$. Then $p^{(n)}_{ij} = (\mathbf{P}^n)_{ij}$, for $i, j \in S$.*

Proof. We proceed by induction. For $n = 1$ the statement follows directly from the definition of the matrix $\mathbf{P}$. Supposing that $p_{ij}^{(n)} = (\mathbf{P}^n)_{ij}$ for all i, j, we have

$$\begin{aligned} p_{ij}^{(n+1)} &= \mathbb{P}[X_{n+1} = j | X_0 = i] \\ &= \sum_{k \in S} \mathbb{P}[X_1 = k | X_0 = i] \mathbb{P}[X_{n+1} = j | X_0 = i, X_1 = k] \\ &= \sum_{k \in S} \mathbb{P}[X_1 = k | X_0 = i] \mathbb{P}[X_{n+1} = j | X_1 = k] \\ &= \sum_{k \in S} \mathbb{P}[X_1 = k | X_0 = i] \mathbb{P}[X_n = j | X_0 = k] \\ &= \sum_{k \in S} p_{ik} p_{kj}^{(n)}. \end{aligned}$$

where the second equality follows from the law of total probability, the third one from the Markov property, and the fourth one from homogeneity. The last sum above is nothing but the expression for the matrix product of $\mathbf{P}$ and $\mathbf{P}^n$, and so we have proven the induction step. □

Using Proposition 7.3.1, we can write a simple expression for the distribution of the random variable X_n, for $n \in \mathbb{N}_0$. Remember that the initial distribution (the distribution of X_0) is denoted by $\boldsymbol{a}^{(0)} = (a_i^{(0)})_{i \in S}$. Analogously, we define the vector $\boldsymbol{a}^{(n)} = (a_i^{(n)})_{i \in S}$ by

$$a_i^{(n)} = \mathbb{P}[X_n = i],\ i \in S.$$

Using the law of total probability, we have

$$a_i^{(n)} = \mathbb{P}[X_n = i] = \sum_{k \in S} \mathbb{P}[X_0 = k] \mathbb{P}[X_n = i | X_0 = k] = \sum_{k \in S} a_k^{(0)} p_{ki}^{(n)}.$$

We usually interpret $\boldsymbol{a}^{(0)}$ as a (row) vector, so the above relationship can be expressed using vector-matrix multiplication

$$\boldsymbol{a}^{(n)} = \boldsymbol{a}^{(0)} \mathbf{P}^n.$$

The following theorem shows a simple, yet fundamental, relationship between different multi-step transition probabilities $p_{ij}^{(n)}$.

Theorem 7.3.2 (Chapman-Kolmogorov relations)**.** *For $n, m \in \mathbb{N}_0$ and $i, j \in S$ we have*

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}.$$

Proof. The statement follows directly from the matrix equality

$$\mathbf{P}^{m+n} = \mathbf{P}^m \mathbf{P}^n. \qquad \square$$

It is usually difficult to compute $\mathbf{P}^n$ for a general transition matrix $\mathbf{P}$ and a large n. We will see later that it will be easier to find the limiting values $\lim_{n\to\infty} p_{ij}^{(n)}$. In the mean-time, here is a simple example where all of the above can be done by hand

Example 7.3.3. In the setting of a Regime Switching chain (Example 4.), let us write a for p_{12} and b for p_{21} to simplify the notation, so that the transition matrix looks like this:

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

The characteristic equation $\det(\lambda I - \mathbf{P}) = 0$ of the matrix $\mathbf{P}$ is

$$\begin{aligned} 0 = \det(\lambda I - \mathbf{P}) &= \begin{vmatrix} \lambda - 1 + a & -a \\ -b & \lambda - 1 + b \end{vmatrix} \\ &= ((\lambda-1)+a)((\lambda-1)+b) - ab = (\lambda-1)(\lambda-(1-a-b)). \end{aligned}$$

The eigenvalues are, therefore, $\lambda_1 = 1$ and $\lambda_2 = 1-a-b$. The eigenvectors are $v_1 = \binom{1}{1}$ and $v_2 = \binom{a}{-b}$, so that with

$$V = \begin{bmatrix} 1 & a \\ 1 & -b \end{bmatrix} \text{ and } D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & (1-a-b) \end{bmatrix}$$

we have

$$\mathbf{P}V = VD, \text{ i.e., } \mathbf{P} = VDV^{-1}.$$

This representation is very useful for taking (matrix) powers:

$$\begin{aligned} \mathbf{P}^n &= (VDV^{-1})(VDV^{-1})\dots(VDV^{-1}) = VD^nV^{-1} \\ &= V \begin{bmatrix} 1 & 0 \\ 0 & (1-a-b)^n \end{bmatrix} V^{-1} \end{aligned}$$

Assuming $a+b>0$ (i.e., $\mathbf{P} \neq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$), we have

$$V^{-1} = \tfrac{1}{a+b} \begin{bmatrix} b & a \\ 1 & -1 \end{bmatrix},$$

and so

$$\begin{aligned}
\mathbf{P}^n &= VD^nV^{-1} = \begin{bmatrix} 1 & a \\ 1 & -b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1-a-b)^n \end{bmatrix} \tfrac{1}{a+b} \begin{bmatrix} b & a \\ 1 & -1 \end{bmatrix} \\
&= \frac{1}{a+b} \begin{bmatrix} b & a \\ b & a \end{bmatrix} + \frac{(1-a-b)^n}{a+b} \begin{bmatrix} a & -a \\ b & -b \end{bmatrix} \\
&= \begin{bmatrix} \frac{b}{a+b} + (1-a-b)^n \frac{a}{a+b} & \frac{a}{a+b} - (1-a-b)^n \frac{a}{a+b} \\ \frac{b}{a+b} + (1-a-b)^n \frac{b}{a+b} & \frac{a}{a+b} - (1-a-b)^n \frac{b}{a+b} \end{bmatrix}
\end{aligned}$$

The expression for $\mathbf{P}^n$ above tells us a lot about the structure of the multi-step probabilities $p_{ij}^{(n)}$ for large n. Note that the second matrix on the right-hand side above comes multiplied by $(1-a-b)^n$ which tends to 0 as $n \to \infty$, unless we are in the uninteresting situation $a = b = 0$ or $(a = b = 1)$. Therefore,

$$\mathbf{P}^n \sim \frac{1}{a+b} \begin{bmatrix} a & b \\ a & b \end{bmatrix} \text{ for large } n.$$

The fact that the rows of the right-hand side above are equal points to the fact that, for large n, $p_{ij}^{(n)}$ does not depend (much) on the initial state i. In other words, this Markov chain forgets its initial condition after a long period of time. This is a rule more than an exception, and we will study such phenomena in the following lectures.

7.4 Problems

Problem 7.4.1. In a *Gambler's ruin* problem with the state space $S = \{0, 1, 2, 3, 4\}$ and the probability $p \in (0, 1)$ of winning in a single game, the number of zeros in the transition matrix is

(a) 17 (b) 15 (c) 14 (d) 5 (e) depends on p

Problem 7.4.2. In a *Gambler's ruin* problem with the state space $S = \{0, 1, 2, 3, 4\}$ and the probability $p = \frac{1}{2}$ of winning in a single game, the 4-step transition probability $p_{24}^{(4)} = \mathbb{P}[X_{n+4} = 4 | X_n = 2]$ is

(a) 2^{-4} (b) 2^{-3} (c) 3×2^{-4} (d) 2^{-2} (e) none of the above

Problem 7.4.3. In a *Gambler's ruin* problem with the state space $S = \{0, 1, 2, 3, 4\}$ and the probability $p = 1/3$ of winning in a single game, the 4-step transition probability

$$p_{22}^{(4)} = \mathbb{P}[X_{n+4} = 2 | X_n = 2]$$

is

(a) $16/81$ (b) $32/81$ (c) $1/4$ (d) $1/2$ (e) none of the above

Problem 7.4.4. Let X be a Markov chain on N states, with the $N \times N$ transition matrix P. We construct a new Markov chain Y from X as follows: at each point in time, we toss a biased coin (probability of *heads* $p \in (0,1)$). If it shows *heads* we follow a single transition of the Markov chain X. If it shows *tails*, we remain in the same state. If I denotes the identity $N \times N$-matrix, find an expression for transition matrix of Y in terms of p, I and P. **Note:** Y is called the *lazy chain* associated to X.

Problem 7.4.5. Glass A contains 12 oz of milk, and glass B 12 oz of water. The following procedure is then performed twice: first, half of the content of the glass A is transfered into class B. Then, the contents of glass B are thoroughly mixed, and a third of its entire content transfered back to A. Finally, the contents of the glass A are thoroughly mixed. What is the final amount of milk in glass A? What does this have to do with Markov chains?

Problem 7.4.6. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk, and let $\{Y_n\}_{n\in\mathbb{N}_0}$ be a random process whose value at time $n \in \mathbb{N}_0$ is equal to the amount of time (number of steps, including possibly n) the process $\{X_n\}_{n\in\mathbb{N}_0}$ has spent above 0, i.e., in the set $\{1, 2, \dots\}$. (For example, if a trajectory of $\{X_n\}_{n\in\mathbb{N}_0}$ takes the values $0, 1, 2, 1, 0, -1, 0, 1, \dots$, then the values of the correspondin trajectory of $\{Y_n\}_{n\in\mathbb{N}_0}$ are $0, 1, 2, 3, 3, 3, 3, 4, \dots$).
Then

(a) $\{Y_n\}_{n\in\mathbb{N}_0}$ is a Markov process

(b) Y_n is a function of X_n for each $n \in \mathbb{N}$.

(c) X_n is a function of Y_n for each $n \in \mathbb{N}$.

(d) Y_n is a stopping time for each $n \in \mathbb{N}_0$.

(e) None of the above

Problem 7.4.7. 100 red balls, and the blue container has 100 blue balls. In each step a container is selected; red with probability 1/2 and blue with probability 1/2. Then, a ball is selected from it - all balls in the container are equally likely to be selected - and placed in the other container. If the selected container is empty, no ball is transfered.

Once there are 100 blue balls in the red container and 100 red balls in the blue container, the game stops.

We decide to model the situation as a Markov chain.

1. What is the state space S we can use? How large is it?
2. What is the initial distribution?
3. What are the transition probabilities between states? Don't write the matrix, it is way too large; just write a general expression for p_{ij}, $i, j \in S$.

Problem 7.4.8. Let $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of independent coin tosses (i.e., random variables with values T or H with equal probabilities). Let $X_0 = 0$, and, for $n \in \mathbb{N}$, let X_n be the number of times two consecutive ξs take the same value in the first $n+1$ tosses. For example, if the outcome of the coin tosses is TTHHTTTH ..., we have $X_0 = 0$, $X_1 = 1$, $X_2 = 1$, $X_3 = 2$, $X_4 = 2$, $X_5 = 3$, $X_6 = 4$, $X_7 = 4, \dots$

Is $\{X_n\}_{n\in\mathbb{N}_0}$ a Markov chain? Explain using the definition. If it is, describe its state space, the transition probabilities and the initial distribution. If it is not, show exactly how the Markov property is violated.

Problem 7.4.9. A country has $m+1$ cities ($m \in \mathbb{N}$), one of which is the capital. There is a direct railway connection between each city and the capital, but there are no tracks between any two "non-capital" cities. A traveler starts in the capital and takes a train to a randomly chosen non-capital city (all cities are equally likely to be chosen), spends a night there and returns the next morning and immediately boards the train to the next city according to the same rule, spends the night there, ..., etc. We assume that her choice of the city is independent of the cities visited in the past. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be the number of visited non-capital cities up to (and including) day n, so that $X_0 = 1$, but X_1 could be either 1 or 2, etc.

1. Explain why $\{X_n\}_{n\in\mathbb{N}_0}$ is a Markov chain on the appropriate state space $\mathcal{S}$ and the find the transition probabilities of $\{X_n\}_{n\in\mathbb{N}_0}$, i.e., write an expression for
$$\mathbb{P}[X_{n+1} = j | X_n = i], \text{ for } i, j \in \mathcal{S}.$$
2. Let τ_m be the first time the traveler has visited all m non-capital cities, i.e.
$$\tau_m = \min\{n \in \mathbb{N}_0 \,:\, X_n = m\}.$$
What is the distribution of τ_m, for $m = 1$ and $m = 2$.
3. (*) Compute $\mathbb{E}[\tau_m]$ for general $m \in \mathbb{N}$. What is the asymptotic behavior of $\mathbb{E}[\tau_m]$ as $m \to \infty$? More precisely, find a simple function $f(m)$ of m (like m^2 or $\log(m)$) such that $\mathbb{E}[\tau_m] \sim f(m)$, i.e., $\lim_{m\to\infty} \frac{\mathbb{E}[\tau_m]}{f(m)} = 1$.

Problem 7.4.10. A math professor has 4 umbrellas. He keeps some of them at home and some in the office. Every morning, when he leaves home, he checks the weather and takes an umbrella with him if it rains. In case all the umbrellas are in the office, he gets wet. The same procedure is repeated in the afternoon when he leaves the office to go home. The professor lives in a tropical region, so the chance of rain in the afternoon is higher than in the morning; it is $1/5$ in the afternoon and $1/20$ in the morning. Whether it rains of not is independent of whether it rained the last time he checked.

On day 0, there are 2 umbrellas at home, and 2 in the office. Describe a Markov chain that can be used to model this situation; make sure to specify the state space, the transition probabilities and the initial distribution.

Problem 7.4.11. An airline reservation system has two computers. A computer in operation may break down on any given day with probability $p \in (0,1)$, independently of the other computer. There is a single repair facility which takes two days to restore a computer to normal. The facilities are such that only one computer at a time can be dealt with. Form a Markov chain that models the situation and sketch its transition graph.

(*Hint:* Make sure to keep track of the number of machines in operation as well as the status of the machine - if there is one - at the repair facility.)

Problem 7.4.12. A monkey is sitting in front of a typewritter in an effort to re-write the complete works of William Shakespeare. She has two states of mind "inspired" and "in writer's block". If the monkey is inspired, she types the letter `a` with probability $1/2$ and the letter `b` with probability $1/2$. If the monkey is in writer's block, she will type `b`. After the monkey types a letter, her state of mind changes independently of her current state of mind, but depending on the letter typed as follows:

- into "inspired" with probability $1/3$ and "in writer's block" with probability $2/3$, if the letter typed was `a` and,
- into "inspired" with probability $2/3$ and "in writer's block" with probability $1/3$, if the letter typed was `b`.

1. What is the probability that the monkey types `abbabaabab` in the first 10 strokes, if she starts in the inspired state?
2. Another monkey is sitting next to her, trying to do the same (re-write Shakespeare). He has no states of mind, and types `a` or `b` with equal probability each time, independently of what he typed before. A piece of paper with `abbabaabab` on it is found, but we don't know who produced it. Which monkey is more likely to have done it, the she-monkey or the he-monkey?

Problem 7.4.13. A "deck" of cards starts with 2 red and 2 black cards. A "move" consists of the following:

(i) pick a random card from the deck (if the deck is empty, do nothing),

(ii) if the card is black <u>and</u> the card drawn on the previous move was also black, return it back to the deck,

(iii) otherwise, throw the card away (this, in particular, applies to any card drawn on the first move, since there is no "previous" move at that time).

1. Model the situation using a Markov chain: find an appropriate state space, and sketch the transition graph with transition probabilities. **Note:** If you manage to do it with 10 states (or fewer), you get extra credit of 10 points.

2. What is the probability that the deck will be empty after exactly 4 moves? What is the probability that the deck will be empty eventually? (*Hint:* Both of these should be easy. If they are not, there is probably something wrong with your answer to 1. above.)

Problem 7.4.14. Let $\{Y_n\}_{n\in\mathbb{N}_0}$ be a sequence of die-rolls, i.e., a sequence of independent random variables with distribution

$$Y_n \sim \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{array}.$$

Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a stochastic process defined by

$$X_n = \max(Y_0, Y_1, \dots, Y_n),\ n \in \mathbb{N}_0.$$

In words, X_n is the maximal value rolled so far. Explain why $\{X_n\}_{n\in\mathbb{N}_0}$ is a Markov chain, and find its transition matrix and the initial distribution.

Problem 7.4.15. We throw a fair 6-sided die repeatedly and define the process Y as follows: $Y_0 = 0$ and Y_n is the number of different values (among the possible ones $\{1,2,3,4,5,6\}$) observed in the first n throws:

(a) Y is <u>not</u> a homogeneous Markov chain

(b) The sequence $\{Y_n\}_{n\in\mathbb{N}_0}$ is iid (independent, identically distributed),

(c) The sequence $Y_{n+1} - Y_n$ is iid,

(d) $\{Y_n\}_{n\in\mathbb{N}_0}$ has an absorbing state

(e) none of the above.

Problem 7.4.16. A car-insurance company classifies drivers in three categories: *bad*, *neutral* and *good*. The reclassification is done in January of each year and the probabilities for transitions between different categories is given by

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/5 & 2/5 & 2/5 \\ 1/5 & 1/5 & 3/5 \end{bmatrix},$$

where the first row/column corresponds to the *bad* category, the second to *neutral* and the third to *good*. The company started in January 1990 with 1400 drivers in each category. Estimate the number of drivers in each category in 2090. Assume that the total number of drivers does not change in time and use software for your computations.