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

Lecture 10
Stationary and Limiting Distributions

Transitions between different states of a Markov chain describe *short-time* behavior of the chain. In most models used in physical and social sciences, systems change states many times per second. In a rare few, the time scale of the steps can be measured in hours or days. What is of interest, however, is the long-term behavior of the system, measured in thousands, millions, or even billions of steps. Here is an example: for a typical liquid stock traded on the New York Stock Excahenge, there is a trade every few seconds (or, even, milliseconds), and each trade changes the price (state) of the stock a little bit. What is of interest to an investor is, however, the distribution of the stock-price in 6 months, in a year or, in 30 years - just in time for retirement. A back-of-an-envelope calculation shows that there are, approximately, 50 million trades in 30 years. So, a grasp of very-long time behavior of a Markov chain is one of the most important achievments of probability in general, and stochastic-process theory in particular. We only scratch the surface in this lecture.

10.1 Stationarity and stationary distributions

Definition 10.1.1. A stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ is said to be **stationary** if the random vectors

$$(X_0, X_1, X_2, \dots, X_k) \text{ and } (X_m, X_{m+1}, X_{m+2}, \dots, X_{m+k})$$

have the same (joint) distribution for all $m, k \in \mathbb{N}_0$.

For stationary processes, all random variables $X_0, X_1, \dots$ have the same distribution (just take $k = 0$ in the definition). That condition is, however, only necessary. The pairs (X_0, X_1) and (X_m, X_{m+1}) should be equally distributed as random vectors, the same for triplets, etc. Intuitively, a stochastic process is stationary if, statistically speaking, it does not evolve. Its probabilistic behavior today is the same as its probabilistic behavior in a billion years. It is somethings useful to think about stationarity in the following way; if a system is let to evolve for a long time, it will reach an equilibrium state and fluctuate around it forever. We can expect that such a system will look similar a million years from now and a billion years from now. It might, however, not resemble its present state at all. Think about a glass of water in which we drop a tiny drop of ink. Immediately after that, the glass will be clear, with a tiny black speck. The ink starts to diffuse and the speck starts to grow immediately. It won't be long before the whole glass is of uniform black color - the ink has permeated every last "corner" of the glass. After that, nothing much happens. The ink will never spontaneously return to its initial state[1]. Ink is composed of many small particles which do not interact with each other too much. They do, however, get bombarded by the molecules of water, and this bombardment makes them behave like random walks[2] which simply bounce back once they hit the glass wall. Each ink particle will wander off in its own direction, and quite soon, they will be "everywhere". Eventually, the distribution of the ink in the glass becomes very close to uniform and no amount of further activity will change that - you just cannot get more "random" then the uniform distribution in a glass of water.

Let us get back to mathematics and give two simple examples; one of a process which is not stationary, and the other of a typical stationary process.

Example 10.1.2.

1. The simple random walk is not stationary. Indeed, X_0 is a constant, while X_1 takes two values with equal probabilities, so they cannot have the same distribution. Indeed, the distribution of X_n is more and more "spread-out" as time passes. Think of an ink drop in an infinite ocean. The dark, ink-saturated, region will get larger and larger, but it will never stabilize as there is always more ocean to invade.

2. For an example of a stationary process, take a regime switching chain $\{X_n\}_{n\in\mathbb{N}_0}$ with $p_{01} = p_{10} = 1$, and the initial distribution $\mathbb{P}[X_0 = 0] = \mathbb{P}[X_0 = 1] = \frac{1}{2}$. Then $X_n = X_0$ if n is even, and $X_n = 1 - X_0$ if n is odd. Moreover, X_0 and $1 - X_0$ have the same distribution (Bernoulli with $p = \frac{1}{2}$), and, so $X_0, X_1, \dots$ all have the same distribution. How about k-tuples? Why do $(X_0, X_1, \dots, X_k)$ and $(X_m, X_{m+1}, \dots, X_{m+k})$ have the same distribution? For $i_0, i_1, \dots, i_k \in \{0, 1\}$, by the Markov property, we have

$$\begin{aligned}\mathbb{P}[X_0 = i_0, X_1 = i_i, \dots, X_k = i_k] &= \mathbb{P}[X_0 = i_0] p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k} \\ &= \tfrac{1}{2} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k}.\end{aligned}$$

[1] It might, actually, but it will take an unimaginably long time.

[2] this phenomenon is called **diffusion**

In the same manner,

$$\begin{aligned}\mathbb{P}[X_m = i_0, X_1 = i_i, \dots, X_{m+k} = i_k] &= \mathbb{P}[X_m = i_0] p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k} \\ &= \tfrac{1}{2} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k},\end{aligned}$$

so the two distributions are identical.

The second example above is quite instructive. We took a Markov chain and gave it an initial distribution with the property that X_0 and X_m have the same distribution for all $m \in \mathbb{N}_0$. Magically, the whole process became stationary. This is not a coincidence; we can play the same trick with any Markov chain, as long as the initial distribution with the above property can be found. Actually, such a distribution is so important that it even has a name:

Definition 10.1.3. A distribution $\pi = (\pi_i)_{i \in S}$ on the state space S of a Markov chain with transition matrix P is called a **stationary distribution** if

$$\mathbb{P}[X_1 = i] = \pi_i \text{ for all } i \in S, \text{ whenever } \mathbb{P}[X_0 = i] = \pi_i, \text{ for all } i \in S.$$

In words, π is called a stationary distribution if the distribution of X_1 is equal to that of X_0 when the distribution of X_0 is π. Here is a hands-on characterization:

Proposition 10.1.4. *A nonnegative vector $\pi = (\pi_i, i \in S)$ with $\sum_{i \in S} \pi_i = 1$ is a stationary distribution if and only if*

$$\pi P = \pi,$$

when π is interpreted as a row vector. In that case the Markov chain with initial distribution π and transition matrix P is stationary and the distribution of X_m is π for all $m \in \mathbb{N}_0$.

Proof. Suppose, first, that π is a stationary distribution, and let $\{X_n\}_{n \in \mathbb{N}_0}$ be a Markov chain with initial distribution $\boldsymbol{a}^{(0)} = \pi$ and transition matrix P. Then,

$$\boldsymbol{a}^{(1)} = \boldsymbol{a}^{(0)} P = \pi P.$$

By the assumption, the distribution $\boldsymbol{a}^{(1)}$ of X_1 is π. Therefore, $\pi = \pi P$.

Conversely, suppose that $\pi = \pi P$. Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a Markov chain with initial distribution π and transition matrix P. We need to show that $\{X_n\}_{n \in \mathbb{N}_0}$ is stationary. In order to do that, we first note that all random variables X_m, $m \in \mathbb{N}_0$, have the same distribution. Indeed, the distribution $\boldsymbol{a}^{(m)}$ of X_m is given by

$$\boldsymbol{a}^{(m)} = \boldsymbol{a}^{(0)} P^m = \pi P^m = (\pi P) P^{m-1} = \pi P^{m-1} = \cdots = \pi.$$

Next, we pick $m, k \in \mathbb{N}_0$ and a $k+1$-tuple $i_0, i_1, \dots, i_k$ of elements of S. By the Markov property, we have

$$\begin{aligned}\mathbb{P}[X_m = i_0, X_{m+1} = i_1, \dots, X_{m+k} = i_k] &= \mathbb{P}[X_m = i_0] p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k} \\ &= \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k}.\end{aligned}$$

This last expression does not depend on m, so we can conclude that $\{X_n\}_{n \in \mathbb{N}_0}$ is stationary. □

Example 10.1.5 (Social mobility)**.** A model of social mobility of families posits three different social classes (strata), namely "lower", "middle", and "upper". The transitions between these classes (states) for a given family are governed by the following transition matrix:

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

A stationary distribution $\pi = (\pi_1, \pi_2, \pi_3)$ would correspond to a steady-state distribution of the entire society into these three classed. To compute it, we start by writing down the following system of three linear equations:

$$\begin{array}{rclclcl} \pi_1 & = & {}^1\!/_2\, \pi_1 & + & {}^1\!/_3\, \pi_2 & & \\ \pi_2 & = & {}^1\!/_2\, \pi_1 & + & {}^1\!/_3\, \pi_2 & + & {}^1\!/_2\, \pi_3 \\ \pi_3 & = & & & {}^1\!/_3\, \pi_2 & + & {}^2\!/_3\, \pi_3 \end{array}$$

We solve the system and obtain that any triplet (π_1, π_2, π_3) with

$$\pi_1 = \tfrac{2}{3}\pi_2 \text{ and } \pi_3 = \pi_2 \tag{10.1.1}$$

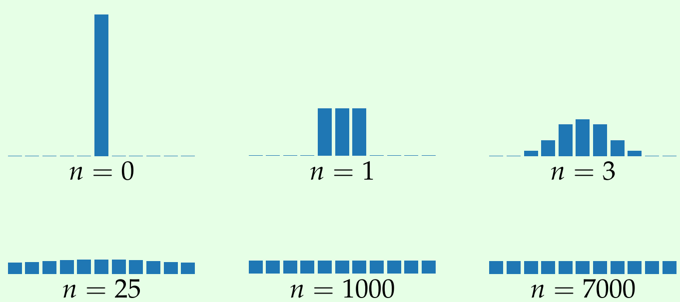
is a solution. Of course, π needs to be a probability distribution, so we also need to require that $\pi_1 + \pi_2 + \pi_3 = 1$. We plug in the expressions (10.1.1) for π_1 and π_3 in terms of π_2 into it to conclude that $\pi_2 = 3/8$. From there, the unique stationary distribution is given by

$$\pi = (\tfrac{2}{8}, \tfrac{3}{8}, \tfrac{3}{8}).$$

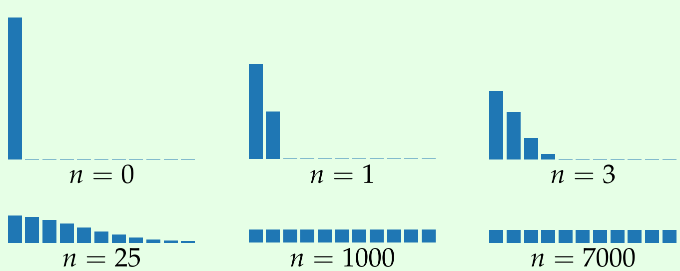
Example 10.1.6 (A model of diffusion in a glass)**.** Let us get back to the story about the glass of water and let us analyze a simplified model of that phenomenon. Our glass will be represented by the set $\{0, 1, 2, \dots, a\}$, where 0 and a are the positions adjacent to the walls of the glass. The ink particle performs a version simple random walk inside the glass - it moves to the left, to the right or stays put with equal probabilities, namely $\frac{1}{3}$. Once it reaches the state 0 further passage to the left is blocked by the wall, so it either takes a step to the right to position 1 (with probability $\frac{1}{3}$) or stays put (with probability $\frac{2}{3}$). The same thing happens at the other wall. All in all, we get a Markov chain with the following transition matrix

$$P = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & \dots & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \dots & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \dots & 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

Let us see what happens when we start the chain with a distribution concentrated at $a/2$; a graphical representation (histogram) of the distributions of X_0, X_1, X_3, X_{12}, X_{30} and X_{7000} when $a = 10$ represents the behavior of the system very well :



How about if we start from a different initial distribution? Here are the same plots when the initial distribution is concentrated at 0:



As you can see, the distribution changes rapidly at first, but then, once it has reached the "equilibrium" it changes remarkably little (compare X_{1000} and X_{7000}; when $X_0 = a/2$, even X_{25} is not very far). Also, the "equilibrium" distribution seems to be uniform and *does not depend on the initial distribution*; this is exactly what you would expect from a long-term distribution of ink in a glass.

Let us show that the uniform distribution $\pi = (1/(a+1), 1/(a+1), \dots, 1/(1+a))$ is indeed the stationary distribution. By Proposition 10.1.4, we need to show that it solves the system $\pi = P\pi$, which, expanded, looks like this

$$\begin{aligned}
\pi_0 &= \tfrac{2}{3}\pi_0 + \tfrac{1}{3}\pi_1 \\
\pi_1 &= \tfrac{1}{3}(\pi_0 + \pi_1 + \pi_2) \\
\pi_2 &= \tfrac{1}{3}(\pi_1 + \pi_2 + \pi_3) \\
&\vdots \\
\pi_{a-1} &= \tfrac{1}{3}(\pi_{a-2} + \pi_{a-1} + \pi_a) \\
\pi_a &= \tfrac{1}{3}\pi_{a-1} + \tfrac{2}{3}\pi_a
\end{aligned}$$

It is immediate that $\pi_0 = \pi_1 = \dots = \pi_a = \frac{1}{a+1}$ is a probability distribution that solves the system above. On the other hand, the first equation yields $\pi_1 = \pi_0$, the second one that $\pi_2 = 2\pi_1 - \pi_0 = \pi_0$, the third $\pi_3 = 2\pi_2 - \pi_1 = \pi_0$, etc. Therefore all π_i must be the same, and, since $\sum_{i=0}^{a} \pi_i = 1$, we conclude that the uniform distribution is the only stationary distribution.

Can there be more than one stationary distribution? Can there be none? Sure, here is an example:

Example 10.1.7.

1. For $P = I$, any distribution is stationary, so there are are infinitely many stationary distributions.

2. A simple example where no stationary distribution exists can be constructed on an infinite state space (but not on a finite space, as we will soon see). Take the Deterministically Monotone Markov chain. The transition "matrix" looks like the identity matrix, with the diagonal of ones shifted to the right. Therefore, the system of equations $\pi = \pi P$ reads

$$\pi_1 = \pi_2, \pi_2 = \pi_3, \dots, \pi_n = \pi_{n+1}, \dots,$$

and so, for π to be a stationary distribution, we must have $\pi_n = \pi_1$ for all $n \in \mathbb{N}$. Now, if $\pi_1 = 0$, π is not a distribution (it sums to 0, not 1). But if $\pi_1 > 0$, then the sum is $+\infty$, so π is not a distribution either. Intuitively, the chain never stabilizes, it just keeps moving to the right ad infinitum.

The example with many stationary distributions can be constructed on any state space, but the other one, where no stationary distribution exists, had to use an infinite one. Was that necessary? Yes. Before we show this fact, let us analyze the relation between stationary distributions and the properties of recurrence and transience. Here is our first result:

Proposition 10.1.8. *Suppose that the state space S of a Markov chain is finite and let $S = C_1 \cup C_2 \cup \dots \cup C_m \cup T$ be its canonical decomposition into recurrent classes $C_1, \dots, C_m$ and the set of transient states T. Then the following two statements are equivalent:*

1. *π is a stationary distribution, and*
2. *$\pi_{C_k} = \pi_{C_k} P_k$, $k = 1, \dots, m$, and $\pi_T = (0, 0, \dots, 0)$,*

where

$$P = \begin{bmatrix} P_1 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & 0 \\ 0 & 0 & P_m & 0 \\ & R & & Q \end{bmatrix},$$

is the canonical form of the transition matrix, $\pi_{C_k} = (\pi_i, i \in C_k)$, $k = 1, 2, \dots, m$ and $\pi_T = (\pi_i, i \in T)$.

Proof. We write the equation $\pi = \pi P$ coordinatewise as $\pi_j = \sum_{i \in S} \pi_i p_{ij}$ and, by distinguishing the cases $i \in C_k$, $k \in \{1, 2, \dots, m\}$, and $i \in T$, we get the following sytem of matrix equations (alternatively, just write the system $\pi = \pi P$ in the block-matrix form according to the cannonical decomposition above):

$$\pi_{C_k} = \pi_{C_K} P_{C_k} + \pi_T R, \ k = 1, \dots, m, \text{ and } \pi_T = \pi_T Q.$$

The last equality can be read as follows: π_T is in a row null-space of $I - Q$. We know, however, that $I - Q$ admits an inverse, and so it is a regular square matrix. Its row null-space (as well as its column null-space) must be trivial, and, consequently, $\pi_T = 0$.

Having established that $\pi_T = 0$, we can de-couple the system of equations above and write it as

$$\pi_{C_k} = \pi_{C_K} P_k, \ k = 1, \dots, m, \text{ and } \pi_T = (0, 0, \dots, 0),$$

which is exactly what we needed to prove.

The other implication - the proof of which consists of a verification of the fact that each distribution from *(2)* above is indeed a stationary distribution - is left to the reader. □

The moral of the story of Proposition 10.1.8 is the following: in order to compute the stationary distribution(s), classify the states and find the canonical decomposition of the state space. Then, set $\pi_i = 0$ for any transient state i. What remains are recurrent classes, and you can analyize each one separately. Note, however, that π_{C_k} does not need to be a real distribution on C_k, since $\sum_{i \in C_k} (\pi_{C_k})_i$ does not need to equal 1. However, unless $\pi_{C_k} = (0, 0, \dots, 0)$, we can always multiply all its elements by a constant to make the sum equal to 1.

10.2 Stationary distributions for finite irreducible chains and Kac's lemma

We now know what the general structure of the set of all stationary distributions is, but we still have no clue as to whether they actually exist. Indeed, our results so far had the following form: "If π is a stationary distribution, then …". Luckily, these results also allow us to focus our search on single recurrent classes, or, more comfortably, chains consisting of a single recurrent class. They are important enough to get a name:

Definition 10.2.1. A Markov chain is said to be **irreducible** if it has only one class.

We also assume from now on that the state space is finite. That rules out, in particular, our non-existence example (Example 10.1.7) which required an infinite chain. We will see that that is not a coincidence, and that the situation is much cleaner in the finite setting. In fact, our next result (known as Kac's lemma, but we state it here as a theorem) gives a very nice answer to the question of existence and uniqueness, with an unexpected benefit:

Theorem 10.2.2 (Kac). *Let $\{X_n\}_{n \in \mathbb{N}_0}$ be an irreducible Markov chain with a finite state space. Then*

1. *there exists a unique stationary distribution $\pi = (\pi_j)_{j \in S}$,*
2. *moreover, it is given by the following formula:*

$$\pi_j = \frac{\nu_{ij}}{m_i},\ j \in S,$$

where i is an arbitrary but fixed state, and

$$\nu_{ij} = \mathbb{E}_i \left[\sum_{n=0}^{T_i(1)-1} \mathbf{1}_{\{X_n=j\}} \right] \text{ and } m_i = \mathbb{E}_i[T_i(1)] < \infty$$

are the expected number of visits to state j in between two consecutive visits to state i, and the expected return time to i, respectively.

Remark 10.2.3. Even though it is not exceedingly hard, the proof of this proposition is a bit technical, so we omit it. It is important, however, to understand what the result states:

1. The stationary distribution π exists and is unique in any irreducible finite Markov chain. Moreover, $\pi_i > 0$ for all $i \in S$.
2. The expected number of visits to the state j in between two consecutive visits to the state i can be related to a stationary distribution by $\nu_{ij} = m_i \pi_j$. By uniqueness, the quotient $\frac{\nu_{ij}}{m_i}$ does not depend on the state i.
3. When we set $j = i$, ν_{ii} counts the number of visits to i between two consecutive visits to i, which is always equal to 1 (the first visit is counted and the last one is not). Therefore, $\nu_{ii} = 1$, and so $\pi_i = \frac{1}{m_i}$ and $\nu_{ij} = \pi_j / \pi_i$.

Theorem 10.2.2 and Remark 10.2.3 are typically used in the following way: one first computes the unique stationary distribution π by solving the equation $\pi = \pi P$ and then uses it to determine m_i or the ν_{ij}'s. Here is a simple example:

Example 10.2.4. Suppose that traffic statistics on a given road are as follows: three out of every four trucks are followed by a car, but only one out of every five cars is followed by a truck. A truck passes by you. How many cars do you expect to see before another truck passes?

The type of the vehicle passing by you (car or truck) can be modeled by a Markov chain with two states and the transition matrix:

$$P = \begin{pmatrix} 4/5 & 1/5 \\ 3/4 & 1/4 \end{pmatrix},$$

with the first row (column) corresponding to the state "car". We are interested in the number of "visits" to the state "car" ($j = 1$) between two visits to the state "truck"($i = 2$), which we denoted by ν_{21} in Theorem 10.2.2. According to the same theorem, it is a good idea to find the (unique) stationary distribution first. The equations are

$$\pi_1 = \tfrac{4}{5}\pi_1 + \tfrac{3}{4}\pi_2$$
$$\pi_2 = \tfrac{1}{5}\pi_1 + \tfrac{1}{4}\pi_2,$$

which, with the additional requirement $\pi_1 + \pi_2 = 1$, give

$$\pi_1 = \tfrac{15}{19}, \pi_2 = \tfrac{4}{19}.$$

Since $v_{ij} = m_i \pi_j$, we have $v_{21} = m_2 \pi_1 = \frac{15}{19} m_2$. We also know that $m_2 = 1/\pi_2 = \frac{19}{4}$, and, so, we expect to see $\frac{15}{4} = 3.75$ cars between two consecutive trucks.

In this particular case, we could have answered the question by computing the expected return time $m_2 = \mathbb{E}_2[T_2(1)]$ to the state 2. As above, $m_2 = 1/\pi_2 = \frac{19}{4}$. We have to be careful, because the state 2 itself is counted exactly once in this expectation, so $m_2 = \frac{19}{4}$ does not only count all the cars between two trucks, but also one of the trucks. Therefore, the number of cars only is $\frac{19}{4} - 1 = \frac{15}{4}$, which is exactly the number we obtained above.

The family of Markov chains called **random walks on graphs** provides for many interesting and unexpected applications of Theorem 10.2.2. We start by remembering what a (simple) graph is.

Definition 10.2.5. A **simple graph** is defined by a finite set V (whose elements are called **vertices**) and a set E if unordered pairs of distinct vertices (whose elements are called **edges**).

Intuitively, a simple graph is a finite collection of points, some of which are connected by lines. We do not allow loops (edges from a vertex to itself) or multiple edged between vertices. For two vertices i and j write $i \sim j$ if there is an edge between i and j, and say that i and j are **neighbors**. The number of neighbors of the vertex i called the **degree** if i, and is denoted by $d(i)$.

Given a simple graph G, let the **random walk on** G be the Markov chain defined as follows:

1. The set of states S is the set of vertices V of the graph G.
2. To move from a state (vertex) to the next one, chose one of its neighbors, uniformly at random, and jump to it. More precisely,

$$p_{ij} = \begin{cases} 0 & i \not\sim j \\ 1/d(i) & i \sim j. \end{cases}$$

As long as the underlying graph is connected (you can move from any state to any other state by traveling on edges only), the corresponding random walk is irreducible and each state is recurrent. The interesting thing is that there is a very simple expression for the (unique) stationary distribution. Indeed, if we write our usual system of equations $\pi = \pi P$ that defines π in this case, we obtain

$$\pi_j = \sum_i \pi_i p_{ij} = \sum_{i:i\sim j} \pi_i \tfrac{1}{d(i)}.$$

If we plug in $\pi_i = d(i)$ into the right-hand side, we get $d(j)$, because there are $d(j)$ terms, each of which equals 1. This matches the left-hand side, so we have a solution to $\pi = \pi P$. The only thing missing is that these πs do not add up to 1. That is easily fixed by dividing by their sum, and we obtain the following nice expression for the stationary distribution:

$$\pi_i = \frac{d(i)}{\sum_{j\in V} d(j)}.$$

Example 10.2.6. A interesting example of a random walk on a graph can be constructed on a chessboard. We pick a piece, say knight, and make it choose, uniformly, from the set of all legal moves. This corresponds to a graph whose vertices are the squares the board, with the two vertices connected by an edge if and only if it is legal for a knight to go from one of them to the other in a single move.

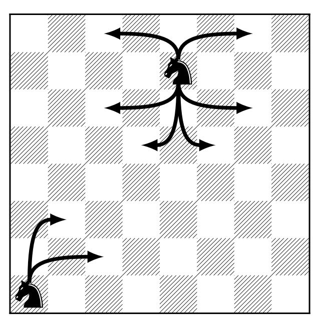


Figure 1. Examples of legal moves for a knight on a chessboard.

What makes all of this possible is the fact that the rules governing the knight's moves are symmetric. If it can jump from i to j, then it can also jump from j to i. We would not be able to construct a random walk on a graph based on the moves of a pawn, for example.

Once we have built the chain, we can check that it is irreducible (do it!), and compute its stationary distribution by computing degrees of all vertices. They are given by the number of different legal moves from each of the 64 squares, as in the following picture

2	3	4	4	4	4	3	2
3	4	6	6	6	6	4	3
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
3	4	6	6	6	6	4	3
2	3	4	4	4	4	3	2

Figure 2. The number in each square is the number of different squares a knight can jump to from it.

The value assigned to each square by stationary distribution is then simply the number on that square, divided by the sum of all the numbers on the board, which happens to be 336.
Theorem 10.2.2 can now be used to answer the following, classical question:

> *A knight starts from the lower left corner of a chessboard, and moves around by selecting one of its legal moves at random at each step, and taking it. What is expected number of moves it will take before it returns to the lower left corner?*

The question is asking for the value of $m_i = \mathbb{E}_i[T_i(1)]$ for $i =$ "lower left corner". By Theorem 10.2.2 this equals to $1/\pi_i = 336/2 = 168$. If the knight started from one of the central squares, this time would be 4 times shorter (42). It is also easy to compute the expected number of visits to another state, between two visits to the lower left corner. It is imply the degree of that state divided by 2 (the degree of the lower left corner).

10.3 Long-run averages

One of the most important properties of stationary distributions is that they describe the long-term behaviour of a Markov chain. Before we explain how, let us recall the classical Law of Large Numbers (LLN) for independent variables:

Theorem 10.3.1 (Law of Large Numbers for iid random variables)**.** *Let $\{Y_n\}_{n\in\mathbb{N}_0}$ be a sequence of independent and identically distributed random variables, such that $\mathbb{E}[|Y_0|] < \infty$. Then*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} Y_k = \mathbb{E}[Y_0],$$

(in an appropriate, mathematically precise, sense).

A special case goes by the name Borel's Law of Large Numbers and applies to the case where each Y_n is a Bernoulli random variable, i.e., takes only the values 0 and 1 (like tossing a biased coin repeatedly). In this case, we have $\sum_{k=0}^{n-1} Y_k = N_n$, where N_n is the number of "successes" (time instances k when $Y_k = 1$) up to time n. The quotient $\frac{1}{n}N_n$ is then the proportion of "successes" among the first n experiments, and Borel's law states that it converges towards the "theoretical frequency", i.e., the probability $p = \mathbb{P}[Y_0 = 1]$ of a single "success". Put differently, p is the long-run proportion of the times k when $Y_k = 1$.

If we try to do the same with a Markov chain, we run into two problems. First, the random variables $X_0, X_1, \dots$ are neither independent nor identically distributed. Second, X_k takes its values in the state space S which does not necessarily consist of numbers, so the expression $X_0 + X_1$ or $\mathbb{E}[X_0]$ does not make sense for every Markov chain. To deal with the second problem, we pick a numerical "reward" function $f : S \to \mathbb{R}$ and form sums of the form $f(X_0) + f(X_1) + \dots + f(X_{n-1})$. Independence is much more subtle, but the Markov property and irreducibility of the chain can be used as a replacement:

Theorem 10.3.2 (Ergodic theorem for Markov Chains)**.** *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a finite and irreducible Markov chain. For any function $f : S \to \mathbb{R}$ we have*

$$\lim_n \tfrac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \mathbb{E}_\pi[f(X_0)] := \sum_{j\in S} f(j)\pi_j,$$

where π is the (unique) stationary distribution of X.

An important special case - corresponding conceptually to Borel's law - is when the function f equals 0 except for one state, where it equals 1. In more compact notation, we pick a state i_0 and define the function f by the following formula

$$f(i) = \begin{cases} 1 & i = i_0 \\ 0 & i \neq i_0. \end{cases}$$

If we apply Proposition 10.3.2 with that particular f, we immediately get the following nice result:

Proposition 10.3.3. *Given a finite and irreducible Markov chain, let N_n^i denote the number of visits to the state i in the first n steps. Then,*

$$\pi_i = \lim_n \tfrac{1}{n} N_n^i,$$

where π is the chain's (unique) stationary distribution.

Put another way,

> *In a finite irreducible Markov chain, the component π_i of the stationary distribution π can be interpreted as the portion (percentage) of time the chain spends in the state i, over a long run.*

10.4 Limiting distributions

Example 10.1.6 shows vividly how the distribution of ink quickly reaches the uniform equilibrium state. It is no coincidence that this "limiting" distribution happens to be a stationary distribution. Before we make this claim more precise, let us define rigorously what we mean by a limiting distribution:

Definition 10.4.1. A distribution $\pi = (\pi_i, i \in S)$ on the state space S of a Markov chain with transition matrix P is called a **limiting distribution** if

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j,$$

for all $i, j \in S$.

Remark 10.4.2.

1. Note that for π to be a limiting distribution, all the limits in Definition 10.4.1 must exist. Once they do (and S is finite), π is automatically a probability distribution: $\pi_j \geq 0$ (as a limit of non-negative numbers) and

$$\sum_{j\in S} \pi_j = \sum_{j\in S} \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \sum_{j\in S} p_{ij}^{(n)} = \lim_{n\to\infty} 1 = 1.$$

   In a Deterministically Monotone Markov Chain we have $p_{ij}^{(n)} = 0$ for $n > j - i$. Therefore $\pi_j := \lim_{n\to\infty} p_{ij}^{(n)} = 0$ for each i, but these πs do not define a probability distribution because they do not add up to 1.

2. Note that the independence on the initial state i is built into the definition of the limiting distribution: the sequence $\{p_{ij}^{(n)}\}_{n\in\mathbb{N}}$ must tend to the same limit π_j for all $i \in S$.

3. Since limits are unique, there can be at most one limiting distribution in a given chain.

The connection with stationary distributions is spelled out in the following propositions:

Proposition 10.4.3. *Suppose that a Markov chain with transition matrix P admits a limiting distribution $\pi = (\pi_i, i \in S)$. Then π is a stationary distribution.*

Proof. To show that π is a stationary distribution, we need to verify that it satisfies $\pi = \pi P$, i.e., that

$$\pi_j = \sum_{i\in S} \pi_i p_{ij}.$$

We use the Chapman-Kolmogorov equation $p_{ij}^{(n+1)} = \sum_{k\in S} p_{ik}^{(n)} p_{kj}$ and start from the observation that $\pi_j = \lim_{n\to\infty} p_{ij}^{(n+1)}$ to get exactly what we need:

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n+1)} = \lim_{n\to\infty} \sum_{k\in S} p_{ik}^{(n)} p_{kj} = \sum_{k\in S} (\lim_{n\to\infty} p_{ik}^{(n)}) p_{kj} = \sum_{k\in S} \pi_k p_{kj}. \qquad \square$$

Example 10.4.4. Limiting distributions don't need to exist, even when there are stationary ones. Here are two examples:

1. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be an extreme regime-switching chain which alternates deterministically between states 0 and 1. Its transition matrix is given by

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and it is easy to see that

$$P^{2n-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for } n \in \mathbb{N}.$$

That means that the values of each $p_{ij}^{(n)}$ oscillate between 0 and 1, and, therefore, cannot converge to anything when $n \to \infty$. In other words, there is no limiting distribution. On the other hand, a stationary distribution $\pi = (\frac{1}{2}, \frac{1}{2})$ clearly exists.

2. In the previous example the limiting distribution did not exists because the limits of the sequences $p_{ij}^{(n)}$ did not exist. A more subtle reason for the non-existence of limiting distributions can be dependence on the initial conditions: the limits $\lim_n p_{ij}^{(n)}$ may exist for all i, j, but their values can depend on i (which it outlawed in Definition 10.4.1). The simplest example is a Markov chain with two states $i = 1, 2$ which does not move at all, i.e., where $p_{11} = p_{22} = 1$. It follows that for each n, we have $p_{ij}^{(n)} = 1$ if $i = j$ and $p_{ij}^{(n)} = 0$ if $i \neq j$. Therefore, the limits $\lim_{n\to\infty} p_{ij}^{(n)}$ exist for each pair i, j, but their values depend on the initial state i:

$$\lim_{n\to\infty} p_{12}^{(n)} = 0 \text{ and } \lim_{n\to\infty} p_{22}^{(n)} = 1.$$

The first part of the example above shows that no limiting distribution needs to exist even in the simplest of irreducible finite chains. Luckily, it also identifies the problem: the chain is periodic and it looks very differently on even vs. odd time points. Clearly, when the chain exhibits this kind of a periodic behavior, it will always be possible to separate even from odd points in time, and, therefore, no equilibrium can be achieved. For this reason, we need to assume, additionally, that the chain is aperiodic. The beauty of the following result (which we give without a proof) is that nothing else is needed.

Theorem 10.4.5. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a finite-state, irreducible and aperiodic Markov chain. Then the limiting distribution exists.*

Example 10.4.6. In Example **??** we considered a two-state "Regime-switching" Markov chain with the following transition matrix:

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix},$$

where $0 < a, b < 1$. Using diagonalization, we produced the following expression for P^n:

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

Since $|1-a-b| < 1$, we have $(1-a-b)^n \to 0$ as $n \to \infty$, and, so

$$\lim_n P^n = \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{pmatrix}$$

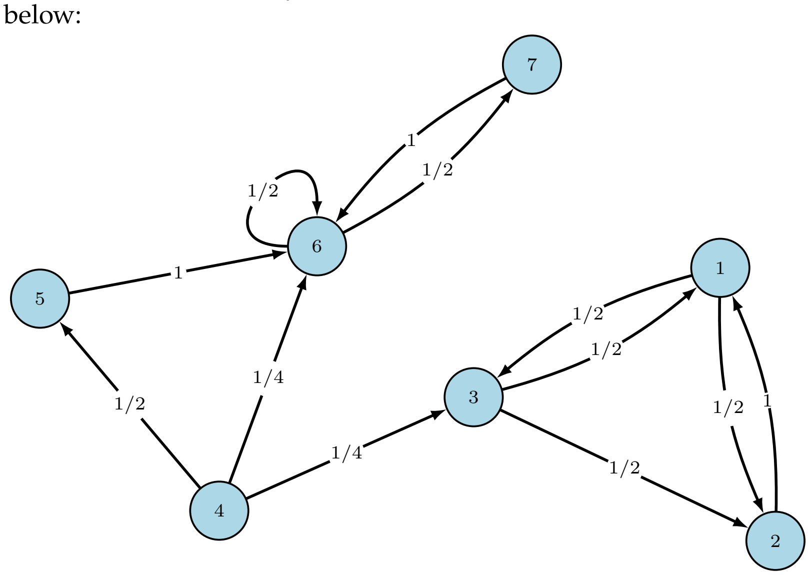
It follows that $(\frac{b}{a+b}, \frac{a}{a+b})$ is a limiting distribution. Of course, we could have concluded that directly from Theorem 10.4.5, since our chain is clearly finite, irreducible and aperiodic.

10.5 Problems

Problem 10.5.1. A spider is walking from a vertex to a vertex of a cube, along its edges. Each time it reaches a vertex, it chooses one of the three edges that meet there with probability $1/3$, independently of its previous choices. Assuming that it takes 1 min for the spider to cross an edge, what is the expected amount of time it to return to its initial vertex? How about if the cube is replaced by a tetrahedron?

Problem 10.5.2. Redo Example 10.2.6 for other chess pieces (except the pawn whose movements are not symmetric).

Problem 10.5.3. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov chain whose transition graph is given below:

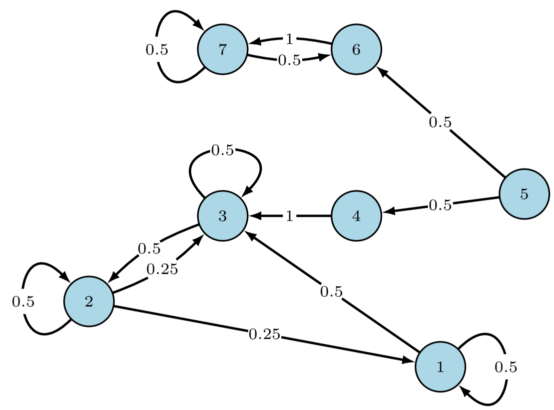


1. Find all stationary distributions.

2. Is there a limiting distribution?

3. If the chain starts from the state 5, what is the asymptotic proportion of time it will spend at state 7? How about the state 5?

Problem 10.5.4. Consider the Markov chain below:



Find all stationary distributions. For each $i \in \{1, 2, \ldots, 7\}$, compute the long-run proportion of time this chain will spend in the state 1.

Problem 10.5.5. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov chain with the transition matrix

1. Find all stationary distributions.

2. The chain starts from the state $i = 1$. What is the expected number of steps before it returns to 1?

3. How many times, on average, does the chain visit 2 between two consecutive visits to 1?

4. Each time the chain visits the state 1, \$1 is added to an account, \$2 for the state 2, and nothing in the state 3. Estimate the amount of money on the account after 10000 transitions? You may assume that the ergodic theorem (Theorem 10.4.5) provides an adequate approximation.

Problem 10.5.6. 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*.

1. 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.

2. A yearly premium charged to a driver depends on his/her category; it is \$1000 for *bad* drivers, \$500 for *neutral* drivers and \$200 for *good* drivers. A typical driver has a total of 56 years of driving experience. Estimate the total amount he/she will pay in insurance premiums in those 56 years. You may assume that 56 years is long enough for the ergodic theorem (Theorem 10.4.5) to provide an accurate approximation.