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

Lecture 8
Classification of states

There will be a lot of definitions and some theory before we get to examples. You might want to peek ahead notions are being introduced; it will help your understanding.

8.1 The Communication Relation

Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov chain on the state space S. For a given set B of states, define the **hitting time** τ_B (or $\tau(B)$ if subscripts are impractical) of B as

$$\tau_B = \min\{n \in \mathbb{N}_0 \,:\, X_n \in B\}. \tag{8.1.1}$$

We know that τ_B is, in fact, a stopping time with respect to $\{X_n\}_{n\in\mathbb{N}_0}$. When B consists of only one element $B = \{i\}$, we simply write τ_i for $\tau_{\{i\}}$; τ_i is the first time the Markov chain $\{X_n\}_{n\in\mathbb{N}_0}$ "hits" the state i. As always, we allow τ_B to take the value $+\infty$; it means that no state in B is ever hit.

The hitting times are important both for applications, and for better understanding of the structure of Markov chains in general.

Example 8.1.1. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be the chain which models a game of tennis (from the previous lecture). The probability of winning for (say) Serena can be phrased in terms of hitting times:

$$\mathbb{P}[\text{Serena wins}] = \mathbb{P}[\tau_{i_S} < \tau_{i_R}],$$

where i_S = "Serena wins" and i_R ="Roger wins" (the two absorbing states of the chain). We will learn how to compute such probabilities in the subsequent lectures.

Having introduced the hitting times τ_B, let us give a few more definitions. It will be very convenient to consider the same Markov chain with different initial distributions. Most often, these distribution will correspond to starting from a fixed state (as opposed to choosing the initial state at random). We use the notation $\mathbb{P}_i[A]$ to mean $\mathbb{P}[A|X_0 = i]$ (for any event A), and $\mathbb{E}_i[A] = \mathbb{E}[A|X_0 = i]$ (for any random variable X). In practice, we use $\mathbb{P}_i$ and $\mathbb{E}_i$ to signify that we are starting the chain from the state i, i.e., $\mathbb{P}_i$ corresponds to a Markov chain whose transition matrix is the same as the one of $\{X_n\}_{n\in\mathbb{N}_0}$, but the initial distribution is given by $\mathbb{P}_i[X_0 = j] = 0$ if $j \neq i$ and $\mathbb{P}_i[X_0 = i] = 1$. Note also that $\mathbb{P}_i[X_1 = j] = p_{ij}$ and that $\mathbb{P}_i[X_n = j] = p_{ij}^{(n)}$, for any n.

Definition 8.1.2. A state $i \in S$ is said to **communicate** with the state $j \in S$, denoted by $i \to j$ if

$$\mathbb{P}_i[\tau_j < \infty] > 0.$$

Intuitively, i communicates with j if there is a non-zero chance that the Markov chain X will eventually visit j if it starts from i. Sometimes we also say that j is **a consequent of** i, that j **is accessible from** i, or that j **follows** i.

Example 8.1.3. In the tennis example, every state is accessible from $(0,0)$ (the fact that $p \in (0,1)$ is important here), but $(0,0)$ is not accessible from any other state. The consequents of $(0,0)$ are not only $(15,0)$ and $(0,15)$, but also $(30,15)$ or $(40,40)$. In fact, all states, except for $(0,0)$ itself are consequents of $(0,0)$.
The consequents of $(40,40)$ are $(40,40)$ itself, $(40, Adv)$, $(Adv, 40)$, "Serena wins" and "Roger wins".

Proposition 8.1.4. *$i \to j$ if and only if $p_{ij}^{(n)} > 0$ for some $n \in \mathbb{N}_0$.*

Proof. Leaving a rigorous mathematical proof aside, we note that the statement is intuitively easy to understand. If $i \to j$ then there must exist some time n such that $\mathbb{P}_i[\tau_j = n] > 0$. This, in turn, implies that it is possible to go from i to j in exactly n steps, where "possible" means "with positive probability". In our notation, that is exactly what $p_{ij}^{(n)} > 0$ means.

Conversely, if $p_{ij}^{(n)} > 0$ then

$$\mathbb{P}_i[\tau_j < \infty] \geq \mathbb{P}_i[\tau_j \leq n] \geq \mathbb{P}_i[X_n = j] = p_{ij}^{(n)} > 0. \qquad \square$$

Some immediate properties if the relation $i \to j$ are listed in the proposition below:

Proposition 8.1.5. *For all $i, j, k \in S$, we have*

1. *$i \to i$,*
2. *$i \to j, j \to k$ implies $i \to k$.*

Proof.

1. If we start from state $i \in S$ we are already there (note that 0 is allowed as a value for τ_B in (8.1.1)), i.e., $\tau_i = 0$ when $X_0 = i$.

2. Using Proposition 8.1.4, it will be enough to show that $p_{ik}^{(n)} > 0$ for some $n \in \mathbb{N}$. By the same Proposition, we know that $p_{ij}^{(n_1)} > 0$ and $p_{jk}^{(n_2)} > 0$ for some $n_1, n_2 \in \mathbb{N}_0$. By the Chapman-Kolmogorov relations, with $n = n_1 + n_2$, we have
$$p_{ik}^{(n)} = \sum_{l \in S} p_{il}^{(n_1)} p_{lk}^{(n_2)} \geq p_{ij}^{(n_1)} p_{jk}^{(n_2)} > 0. \qquad \square$$

Remark 8.1.6. The inequality $p_{ik}^{(n)} \geq p_{il}^{(n_1)} p_{lk}^{(n_2)}$ is valid for all $i, l, k \in S$, as long as $n_1 + n_2 = n$. It will come in handy later.

Remember that the **greatest common denominator (GCD)** of a set A of natural numbers if the largest number $d \in \mathbb{N}$ such that d divides each $k \in A$, i.e., such that each $k \in A$ is of the form $k = ld$ for some $l \in \mathbb{N}$.

Definition 8.1.7. A **period** $d(i)$ of a state $i \in S$ is the greatest common denominator of the **return-time set** $R(i) = \{n \in \mathbb{N} : p_{ii}^{(n)} > 0\}$ of the state i. When $R(i) = \emptyset$, we set $d(i) = 1$. A state $i \in S$ is called **aperiodic** if $d(i) = 1$.

Example 8.1.8. Consider the Markov chain with three states and the transition matrix
$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$
The return set for each state $i \in \{1, 2, 3\}$ is given by
$$R(i) = \{3, 6, 9, 12, \dots\},$$
so $d(i) = 3$ for all $i \in \{1, 2, 3\}$. However, if we change the probabilities a bit:
$$\hat{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix},$$
the situation changes drastically:
$$\begin{aligned} R(1) &= \{3, 4, 5, 6, \dots\}, \\ R(2) &= \{2, 3, 4, 5, 6, \dots\}, \\ R(3) &= \{1, 2, 3, 4, 5, 6, \dots\}, \end{aligned}$$
so that $d(i) = 1$ for $i \in \{1, 2, 3\}$.

8.2 Classes

Definition 8.2.1. We say that the states i and j in S **intercommunicate**, denoted by $i \leftrightarrow j$ if $i \to j$ and $j \to i$. A set $B \subseteq S$ of states is called **irreducible** if $i \leftrightarrow j$ for all $i, j \in S$.

Unlike the relation of communication, the relation of intercommunication is symmetric. Moreover, we have the following three properties (the result follows directly from Proposition 8.1.4, so we omit it):

Proposition 8.2.2. *The relation $\leftrightarrow$ is an* equivalence relation *on S, i.e., for all $i, j, k \in S$, we have*

1. $i \leftrightarrow i$ (reflexivity) *,*
2. $i \leftrightarrow j$ *implies* $j \leftrightarrow i$ (symmetry)*, and*
3. $i \leftrightarrow j, j \leftrightarrow k$ *implies* $i \leftrightarrow k$ (transitivity)*.*

The fact that $\leftrightarrow$ is an equivalence relation allows us to split the state-space S into equivalence classes with respect to $\leftrightarrow$. In other words, we can write

$$S = S_1 \cup S_2 \cup S_3 \cup \dots,$$

where $S_1, S_2, \dots$ are mutually exclusive (disjoint) and all states in a particular S_n intercommunicate, while no two states from different equivalence classes S_n and S_m do. The sets $S_1, S_2, \dots$ are called **classes** of the chain $\{X_n\}_{n\in\mathbb{N}_0}$. Equivalently, one can say that classes are *maximal irreducible sets*, in the sense that they are irreducible and no class is a subset of a (strictly larger) irreducible set. A cookbook algorithm for class identification would involve the following steps:

1. Start from an arbitrary state (call it 1).
2. Identify *all* states j that intercommunicate with it (1, itself, always does).
3. That is your first class, call it C_1. If there are no elements left, then there is only one class $C_1 = S$. If there is an element in $S \setminus C_1$, repeat the procedure above starting from that element.

The notion of a class is especially useful in relation to another natural concept:

Definition 8.2.3. A set $B \subseteq S$ of states is said to be **closed** if $i \not\to j$ for all $i \in B$ and all $j \in S \setminus B$. A state $i \in S$ such that the set $\{i\}$ is closed is called **absorbing**.

Here is a nice characterization of closed sets:

Proposition 8.2.4. *A set B of states is closed if and only if $p_{ij} = 0$ for all $i \in B$ and all $j \in B^c = S \setminus B$.*

Proof. Suppose, first, that B is closed. Then for $i \in B$ and $j \in B^c$, we have $i \not\to j$, i.e., $p^{(n)}_{ij} = 0$ for all $n \in \mathbb{N}$. In particular, $p_{ij} = 0$.

Conversely, suppose that $p_{ij} = 0$ for all $i \in B$, $j \in B^c$. We need to show that $k \not\to l$ (i.e. $p^{(n)}_{kl} = 0$ for all $n \in \mathbb{N}$) for all $k \in B$, $l \in B^c$. Suppose, to the contrary, that there exist $k \in B$ and $l \in B^c$ such that $p^{(n)}_{kl} > 0$ for some $n \in \mathbb{N}$. That means that we can find a sequence of states

$$k = i_0, i_1, \dots, i_n = l \text{ such that } p_{i_{m-1} i_m} > 0 \text{ for all } m = 1, \dots, n.$$

The first state, $k = i_0$ is in B and the last one, $l = i_n$, is in B^c. Therefore there must exist an index m such that $i_{m-1} \in B$ but $i_m \in B^c$. We also know that $p_{i_m i_{m+1}} > 0$, which is in contradiction with out assumption that $p_{ij} = 0$ for all $i \in B$ and $j \in B^c$. □

Intuitively, a set of states is closed if it has the property that the chain $\{X_n\}_{n \in \mathbb{N}_0}$ stays in it forever, once it enters it. In general, if B is closed, it does not have to follow that $S \setminus B$ is closed. Also, a class does not have to be closed, and a closed set does not have to be a class. Here are some examples:

Example 8.2.5. Consider the *tennis* chain of the previous lecture and consider the following three sets of states:

1. $B = \{\text{“Serena wins”}\}$: closed and a class, but $S \setminus B$ is not closed
2. $B = S \setminus \{(0,0)\}$: closed, but not a class, and
3. $B = \{(0,0)\}$: class, but not closed.

There is a relationship between classes and the notion of closedness:

Proposition 8.2.6. *Every closed set B is a union of classes.*

Proof. Let $\hat{B}$ be the union of all classes C such that $C \cap B \neq \emptyset$. In other words, take all the elements of B and throw in all the states which intercommunicate with them. I claim that $\hat{B} = B$. Clearly, $B \subset \hat{B}$, so we need to show that $\hat{B} \subseteq B$. Suppose, to the contrary, that there exists $j \in \hat{B} \setminus B$. By construction, j intercommunicates with some $i \in B$. In particular $i \to j$. By closedness of B, we must have $j \in B$. This is a contradiction with the assumptions that $j \in \hat{B} \setminus B$. □

Example 8.2.7. A converse of Proposition 8.2.6 is not true. Just take the set $B = \{(0,0),(0,15)\}$ in the "tennis" example. It is a union of classes, but it is not closed.

8.3 Transience and recurrence

It is often important to know whether a Markov chain will ever return to its initial state, and if so, how often. The notions of transience and recurrence are used to address this questions.

Definition 8.3.1. The **(first) visit time** to state j, denoted by $T_j(1)$ is defined as

$$T_j(1) = \min\{n \in \mathbb{N} : X_n = j\}.$$

As usual $T_j(1) = +\infty$ if $X_n \neq j$ for all $n \in \mathbb{N}$.

Similarly second, third, etc., visit time are defined as follows:

$$\begin{aligned} T_j(2) &= \min\{n > T_j(1) : X_n = j\}, \\ T_j(3) &= \min\{n > T_j(2) : X_n = j\}, \text{ etc.,} \end{aligned}$$

with the understanding that if $T_j(n) = +\infty$, then also $T_j(m) = +\infty$ for all $m > n$.

Note that the definition of the random variable $T_j(1)$ differs from the definition of τ_j in that the minimum here is taken over the set $\mathbb{N}$ of natural numbers, while the set of non-negative integers $\mathbb{N}_0$ is used for τ_j. When $X_0 \neq j$, the hitting time τ_j and the first visit time $T_j(1)$ coincide. The important difference occurs only when $X_0 = j$. In that case $\tau_j = 0$ (we are already there), but it is always true that $T_j(1) \geq 1$. It can even happen that $\mathbb{P}_j[T_j(1) = \infty] = 1$.

Definition 8.3.2. A state $i \in S$ is said to be

1. **recurrent** if $\mathbb{P}_i[T_i(1) < \infty] = 1$,
2. **positive recurrent** if $\mathbb{E}_i[T_i(1)] < \infty$
3. **null recurrent** if it is recurrent, but not positive recurrent,
4. **transient** if it is not recurrent.

A state is recurrent if we are sure we will come back to it eventually (with probability 1). It is positive recurrent if it is recurrent and the time between two consecutive visits has finite expectation. Null recurrence means the we will return, but the waiting time may be very long. A state is transient if there is a positive chance (however small) that the chain will never return to it.

8.4 A criterion for recurrence

The definition of recurrence from above is conceptually simple, but it gives us no clue about how to actually go about deciding whether a particular state in a specific Markov chain is recurrent. A criterion stated entirely in terms of the transition matrix P would be nice. Before we give it, we need to introduce some notation. Given a state i, let f_i denote the probability that the chain will visit i again, if it starts there, i.e.,

$$f_i = \mathbb{P}_i[T_i(1) < \infty].$$

Clearly, i is recurrent if and only if $f_i = 1$.

The interesting thing is that every time our chain visits the state i, its future evolution is independent from the past (except for the name of the current state) and it behaves exactly like a new and independent chain started from i would. This is a special case of so-called **strong Markov property** which states that the usual Markov property also holds at stopping times (and not only fixed times n). We will not prove this property it these notes, but we will gladly use it to prove the following dichotomy:

Proposition 8.4.1. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov chain on a countable state space S, with the (deterministic) initial state $X_0 = i$. Then exactly one of the following two statements hold:*

1. *either the chain will return to i infinitely many times, or*
2. *the chain will return to i a finite number N_i of times, where N_i is geometrically distributed with parameter $1 - f_i$, where $f_i = \mathbb{P}_i[T_i(1) < \infty]$.*

In the first case, i is recurrent and, in the second, it is transient.

Proof. If $f_i = 1$, then X is guaranteed to return to i at least once. When that happens, however, the strong Markov property "deletes" the past, and the process "renews" itself. This puts us back in the original situation where we are looking at a chain which starts at i and is guaranteed to return there at least once. Continuing like that, we get a whole infinite sequence of stopping times

$$T_i(1) < T_i(2) < \dots$$

at which X finds itself at i.

If $f_i < 1$, a similar story can be told, but with a significant difference. Every time X returns to i, there is a probability $1 - f_i$ that it will never come back to i, and, this is independent of the past behavior. If we think of the return to i as a success, the number of successes before the first failure, i.e., the number of return visits to i, is nothing but a geometrically distributed random variable with parameter f_i. □

The following interesting fact follow almost directly from Proposition 8.4.1

Proposition 8.4.2. *Suppose that the state space S is finite. Then there exists at least one recurrent state.*

Proof. Suppose that all the states are transient. Starting the chain from some, arbitrary, state $i_0 \in S$, the total number of visits V_i to each state i is always finite. Indeed, if $i = i_0$ that is precisely the conclusion of Proposition 8.4.1. For a state $i \neq i_0$, the number of visits is either 0 - if we never even get to i, or $1 + V_i$ if we do. In either case, it is a finite number (not $+\infty$).

Therefore the sum $\sum_{i \in S} V_i$ is also finite - a contradiction with the fact that there are infinitely many time instances $n \in \mathbb{N}_0$, and that the chain must be in some state in each one of them. □

Remark 8.4.3. If S is not finite, it is not true that recurrent states must exist. Just remember the "Deterministically-Monotone Markov Chain" example, or the random walk with $p \neq \frac{1}{2}$. All states are transitive there.

Proposition 8.4.1 above is the engine behind the following important criterion:

Theorem 8.4.4. *A state $i \in S$ is recurrent if and only if*

$$\sum_{n \in \mathbb{N}} p_{ii}^{(n)} = \infty.$$

Proof. Let $V_i^{(n)}$ denote the number of visits to that state i during the first n steps, with the initial visit at time 0 not counted. More precisely,

$$V_i^{(n)} = \sum_{k=1}^{n} \mathbf{1}_{\{X_k = i\}}.$$

Taking the expectation yields

$$\mathbb{E}_i[N_n^i] = \sum_{k=1}^{n} \mathbb{E}_i[\mathbf{1}_{\{X_n = i\}}] = \sum_{k=1}^{n} \mathbb{P}_i[X_n = i] = \sum_{k=1}^{n} p_{ii}^{(k)}.$$

Let $V_i = \lim_{n\to\infty} V_i^{(n)}$ be the total number of visits to the state i. Using the intuitively acceptable (but not rigorously proven) fact that $\mathbb{E}_i[V_i] = \lim_n \mathbb{E}_i[V_i^{(n)}]$, we conclude that

$$\mathbb{E}_i[V_i] = \lim_n \mathbb{E}_i[V_i^{(n)}] = \lim_n \sum_{k=1}^{n} p_{ii}^{(k)} = \sum_{k=1}^{\infty} p_{ii}^{(k)}.$$

By Proposition 8.4.1 above, if i is transient, i.e., if $f_i < 1$, we have

$$\mathbb{E}_i[V_i] = \tfrac{f_i}{1-f_i} < \infty, \text{ and so } \sum_{k=1}^{\infty} p_{ii}^{(k)} = \mathbb{E}_i[V_i] < \infty.$$

On the other hand, if i is recurrent, we necessarily have $V_i = \infty$ (by the same Proposition). Hence,

$$\sum_{k=1}^{\infty} p_{ii}^{(k)} = \mathbb{E}_i[V_i] = \infty.$$

□

Remark 8.4.5. Let N be a random variable taking values in $\mathbb{N}_0 \cup \{+\infty\}$. If $\mathbb{E}[N] < \infty$, then, clearly $\mathbb{P}[N = \infty] = 0$ so that N only takes values in $\mathbb{N}_0$. On the other hand, it is not true that $\mathbb{P}[N = \infty] = 0$ implies that $\mathbb{E}[N] < \infty$. It suffices to take a random variable with the following distribution

$$\mathbb{P}[N = n] = c/n^2 \text{ for } n \in \mathbb{N},$$

where the constant c is chosen so that $\sum_n c/n^2 = 1$ (in fact, we can compute that $c = 6/\pi^2$ explicitly in this case). The expected value of N is given by

$$\mathbb{E}[N] = \sum_{n=1}^{\infty} n\mathbb{P}[N = n] = c \sum_{n=1}^{\infty} \tfrac{1}{n} = +\infty.$$

The message is that, in general, you cannot detect whether something happened infinitely many times or not based only on its expectation. Such a detection, however, becomes possible in the special case when $N = V_i$ denotes the total number of returns to the state i of a Markov chain. This is exactly the content of proof of Theorem 8.4.4 above: each time the chain leaves i, it comes back to it (or does not) with the same probability, independently of the past. This gives us extra information about the random variable N (namely that it is either infinite with probability 1 or geometrically distributed) and allows us to test its finiteness by using the expected value only.

Here is an application of our recurrence criterion - a beautiful and unexpected result of George Pólya from 1921.

Example 8.4.6 (Optional)**.** In addition to the simple symmetric random walk on the line ($d = 1$) we studied before, one can consider random walks whose values are in the plane ($d = 2$), the space ($d = 3$), etc.

These are usually defined as follows: the random walk in d dimensions is the Markov chain with the state space $S = \mathbb{Z}^d$; starting from the state $(x_1, \dots, x_d)$, it picks one of its $2d$ neighbors $(x_1 + 1, \dots, x_d)$, $(x_1 - 1, \dots, x_d)$, $(x_1, x_2 + 1, \dots, x_d)$, $(x_1, x_2 - 1, \dots, x_d)$, $\dots$, $(x_1, \dots, x_d + 1)$, $(x_1, \dots, x_d - 1)$ randomly and uniformly and moves there.

In this example, we are interested in recurrence properties of the d-dimensional random walk and and their dependence on d. We already know that the simple symmetric random walk on $\mathbb{Z}$ is recurrent (i.e., every $i \in \mathbb{Z}$ is a recurrent state). The easiest way to proceed when $d \geq 2$ is to use Theorem 8.4.4, i.e., to estimate the values $p_{ii}^{(n)}$, for $n \in \mathbb{N}$. By symmetry, we can focus on the origin, i.e., it is enough to compute, for each $n \in \mathbb{N}$,

$$p^{(n)} = p_{00}^{(n)} = \mathbb{P}_0[X_n = (0, 0, \dots, 0)].$$

For that, we should count all "trajectories" from $(0, \dots, 0)$ that return to $(0, \dots, 0)$ in n steps. First of all, it is clear that n needs to be even, i.e., $n = 2m$, for some $m \in \mathbb{N}$. It helps if we think of any trajectory as a sequence of "increments" $\xi_1, \dots, \xi_n$, where each ξ_i takes its value in the set $\{1, -1, 2, -2, \dots, d, -d\}$. In words, $\xi_i = +k$ if the k-th coordinate increases by 1 on the i-th step, and $\xi_i = -k$, if the k-th coordinate decreases[a]. This way, the problem becomes combinatorial:

In how many ways can we put one element of the set $\{1, -1, 2, -2, \dots, d, -d\}$ into each of $n = 2m$ boxes so that the number of boxes with k in them equals to the number of boxes with $-k$ in them?

To get the answer, we start by fixing a possible count $(i_1, \dots, i_d)$, satisfying $i_1 + \dots + i_d = m$ of the number of times each of the values in $\{1, 2, \dots, d\}$ occurs. These values have to be placed in m of the $2m$ slots and their negatives (possibly in a different order) in the remaining m slots. So, first, we choose the "positive" slots (in $\binom{2m}{m}$ ways), and then distribute i_1 "ones", i_2 "twos", etc., in those slots; this can be done in[b]

$$\binom{m}{i_1 i_2 \dots i_d}$$

ways. This is also the number of ways we can distribute the negative "ones", "twos", etc., in the remaining slots. All in all, for fixed $i_1, i_2, \dots, i_d$, all of this can be done in

$$\binom{2m}{m} \binom{m}{i_1 i_2 \dots i_d}^2$$

ways. Remembering that each path has the probability $(2d)^{-2m}$, and summing over all $i_1, \dots, i_d$ with $i_1 + \dots + i_d = m$, we get

$$p^{(2m)} = \frac{1}{(2d)^{2m}} \binom{2m}{m} \sum_{i_1+\dots+i_d=m} \binom{m}{i_1 i_2 \dots i_d}^2.$$

For $d = 1$, the situation is somewhat simpler:

$$p^{(2m)} = \frac{1}{4^m} \binom{2m}{m}$$

It is still too complicated sum over all $m \in \mathbb{N}$, but we can simplify it further by using Stirling's formula

$$n! \sim \sqrt{2\pi n} \left(\tfrac{n}{e}\right)^n,$$

where $a_n \sim b_n$ means $\lim_{n\to\infty} a_n/b_n = 1$. Indeed, from there,

$$\binom{2m}{m} \sim \frac{4^m}{\sqrt{\pi m}}, \tag{8.4.1}$$

and so

$$p^{(2m)} \sim \frac{1}{\sqrt{m\pi}}.$$

It is easy to apply our criterion:

$$\sum_{m=1}^{\infty} p^{(2m)} = \infty,$$

and we recover our previous conclusion that the simple symmetric random walk is recurrent.

Moving on to the case $d = 2$, we notice that the sum of the multinomial coefficients no longer equals 1; in fact it is given by[c]

$$\sum_{i=0}^{m} \binom{m}{i}^2 = \binom{2m}{m}, \tag{8.4.2}$$

and, so,

$$p^{(2m)} = \frac{1}{16^m} \left(\frac{4^m}{\sqrt{\pi m}} \right)^2 \sim \frac{1}{\pi m},$$

and

$$\sum_{m=1}^{\infty} p^{(2m)} = +\infty,$$

which implies that the two-dimensional random walk is also recurrent.

How about $d \geq 3$? Things are even more complicated now. The multinomial sum does not admit a nice closed-form expression as in (8.4.2), so we need to do some estimates; these are a bit tedious so we skip them, but report the punchline, which is that

$$p^{(2m)} \sim C\left(\frac{3}{m}\right)^{3/2},$$

for some constant C. This is where it gets interesting: this series converges:

$$\sum_{m=1}^{\infty} p^{(2m)} < \infty,$$

and, so, the random walk is transient for $d = 3$. This is enough to conclude that the random walk is transient for all $d \geq 3$, too (why?). In a nutshell, the random walk is recurrent for $d = 1, 2$, but transient for $d \geq 3$. In the words of Shizuo Kakutani

> *A drunk man will find his way home, but a drunk bird may get lost forever.*

---

[a] For $d = 2$ we could have used the values "up", "down", "left" and "'right", for $1, -1, 2$ or -2, respectively. In dimension 3, we could have added "forward" and "backward", but we run out of words for directions for larger d.

[b] $\binom{m}{i_1 \dots i_d}$ is called the *multinomial coefficient*. It counts the number of ways we can color m objects into one of d colors such that there are i_1 objects of color 1, i_2 of color 2, etc. It is a generalization of the binomial coefficient and is given by

$$\binom{m}{i_1 i_2 \dots i_d} = \frac{m!}{i_1! i_2! \dots i_d!}.$$

[c] Why is this identity true? Can you give a counting argument?

8.5 Class properties

Certain properties of states are shared between all elements in a class. Knowing which properties have this feature is useful for a simple reason - if you can check them for a single class member, you know automatically that all the other elements of the class share it.

Definition 8.5.1. A property is called a **class property** it holds for all states in its class, whenever it holds for any one particular state in the that class.

Put differently, a property is a class property if and only if either all states in a class have it or none does.

Proposition 8.5.2. *Transience and recurrence are class properties.*

Proof. Suppose that the state i is recurrent, and that j is in its class, i.e., that $i \leftrightarrow j$. Then, there exist natural numbers m and k such that $p_{ij}^{(m)} > 0$ and $p_{ji}^{(k)} > 0$. By the Chapman-Kolmogorov relations, for each $n \in \mathbb{N}$, we have

$$p_{jj}^{(n+m+k)} = \sum_{l_1 \in S} \sum_{l_2 \in S} p_{jl_1}^{(k)} p_{l_1 l_2}^{(n)} p_{l_2 m}^{(m)} \geq p_{ji}^{(k)} p_{ii}^{(n)} p_{ij}^{(m)}.$$

In other words, there exists a positive constant c (take $c = p_{ji}^{(k)} p_{ij}^{(m)}$), independent of n, such that

$$p_{jj}^{(n+m+k)} \geq c p_{ii}^{(n)}.$$

The recurrence of i implies that $\sum_{n=1}^{\infty} p_{ii}^{(n)} = +\infty$, and so

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} \geq \sum_{n=m+k+1}^{\infty} p_{jj}^{(n)} = \sum_{n=1}^{\infty} p_{jj}^{(n+m+k)} \geq c \sum_{n=1}^{\infty} p_{ii}^{(n)} = +\infty,$$

and so, j is recurrent. Thus, recurrence is a class property, and since transience is just the opposite of recurrence, it is clear that transience is also a class property, too. □

Proposition 8.5.3. *Period is a class property, i.e., all elements of a class have the same period.*

Proof. Let $d = d(i)$ be the period of the state i, and let $j \leftrightarrow i$. Then, there exist natural numbers m and k such that $p_{ij}^{(m)} > 0$ and $p_{ji}^{(k)} > 0$. By Chapman-Kolmogorov,

$$p_{ii}^{(m+k)} \geq p_{ij}^{(m)} p_{ji}^{(k)} > 0,$$

and so $m + k \in R(i)$. Similarly, for any $n \in R(j)$,

$$p_{ii}^{(m+k+n)} \geq p_{ij}^{(m)} p_{jj}^{(n)} p_{ji}^{(k)} > 0,$$

so $m + k + n \in R(i)$. By the definition of the period, we see now that $d(i)$ divides both $m + k$ and $m + k + n$, and, so, it divides n. This works for each $n \in R(j)$, so $d(i)$ is a common divisor of all elements of $R(j)$; this, in turn, implies that $d(i) \leq d(j)$. The same argument with roles of i and j switched shows that $d(j) \leq d(i)$. Therefore, $d(i) = d(j)$. □

8.6 The canonical decomposition

Now that we know that transience and recurrence are class properties, we can introduce the notion of the **canonical decomposition** of a Markov chain. Let $S_1, S_2, \dots$ be the collection of all classes; some of them contain recurrent states and some transient ones. Proposition 8.5.2 tells us that if there is one recurrent state in a class, than all states in the class must be recurrent. Thus, it makes sense to call the whole class **recurrent**. Similarly, the classes which are not recurrent consists entirely of transient states, so we call them **transient**. There are at most countably many states, so the number of all classes is also at most countable. In particular, there are only countably (or finitely) many recurrent classes, and we usually denote them by $C_1, C_2, \dots$. Transient classes are denoted by $T_1, T_2, \dots$. There is no special rule for the choice of indices $1, 2, 3, \dots$ for particular classes. The only point is that they can be enumerated because there are at most countably many of them.

The distinction between different transient classes is usually not very important, so we pack all transient states together in a set $T = T_1 \cup T_2 \cup \dots$.

Definition 8.6.1. Let S be the state space of a Markov chain $\{X_n\}_{n\in\mathbb{N}_0}$. Let $C_1, C_2, \dots$ be its recurrent classes, and $T_1, T_2, \dots$ the transient ones, and let $T = T_1 \cup T_2 \cup \dots$ be their union. The decomposition

$$S = T \cup C_1 \cup C_2 \cup C_3 \cup \dots,$$

is called the **canonical decomposition** of the (state space of the) Markov chain $\{X_n\}_{n\in\mathbb{N}_0}$.

The reason that recurrent classes are important is simple - they can be interpreted as Markov chains themselves. In order for such an interpretation to be possible, we need to make sure that the Markov chain stays in a recurrent class if it starts there. In other words, we have the following important proposition:

Proposition 8.6.2. *Recurrent classes are closed.*

Proof. Suppose that C is a recurrent class which is not closed. Then, there exist states $i \in C$ and $j \in C^c$ such that $i \to j$. On the other hand, since $j \notin C$ and C is a class, we cannot have $j \to i$. Started at i, the chain will reach j with positive probability, and, since $j \not\to i$, never return. That implies that the number of visits to i will be finite, with positive probability. That is in contradiction with the fact that i is recurrent and the statement of Proposition 8.4.1. □

The fact we just proved implies the following nice dichotomy, valid for every finite-state-space chain:

Proposition 8.6.3. *A class of a Markov chain on a finite state space is recurrent if and only if it is closed.*

Proof. We know that recurrent classes are closed. In order to show the converse, we need to prove that transient classes are not closed. Suppose, to the contrary, the there exists a finite state-space Markov chain with a closed transient class T. Since T is closed, we can see it as a state space of the restricted Markov chain. This, new, Markov chain has a finite number of states so there exists a recurrent state. This is a contradiction with the assumption that T consists only of transient states. □

Remark 8.6.4. Finiteness is necessary. For a random walk on $\mathbb{Z}$, all states intercommunicate. In particular, there is only one class $\mathbb{Z}$ itself and it it trivially closed. If $p \neq \frac{1}{2}$, however, all states are transient, and, so, $\mathbb{Z}$ is a closed and transient class.

Together with the canonical decomposition, we introduce the **canonical form** of the transition matrix P. The idea is to order the states in S with the canonical decomposition in mind. We start from all the states in C_1, followed by all the states in C_2, etc. Finally, we include all the states in T. The resulting matrix looks like this

$$P = \begin{bmatrix} P_1 & 0 & 0 & \dots & 0 \\ 0 & P_2 & 0 & \dots & 0 \\ 0 & 0 & P_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Q_1 & Q_2 & Q_3 & \dots & \dots \end{bmatrix},$$

where the entries should be interpreted as matrices: P_1 is the transition matrix within the first class, i.e., $P_1 = (p_{ij}, i \in C_1, j \in C_1)$, etc. Q_k contains the transition probabilities from the transient states to the states in the (recurrent) class C_k. Note that Proposition 8.6.2 implies that each P_k is a stochastic matrix, or, equivalently, that all the entries in the row of P_k outside of P_k are zeros.

We finish the discussion of canonical decomposition with an important result and one of its consequences.

8.7 Examples

Random walks: $p \in (0,1)$.

- **Communication and classes**. Clearly, it is possible to go from any state i to either $i+1$ or $i-1$ in one step, so $i \to i+1$ and $i \to i-1$ for all $i \in S$. By transitivity of communication, we have $i \to i+1 \to i+2 \to \cdots \to i+k$. Similarly, $i \to i-k$ for any $k \in \mathbb{N}$. Therefore,

$i \to j$ for all $i, j \in S$, and so, $i \leftrightarrow j$ for all $i, j \in S$, and the whole S is one big class.

- **Closed sets**. The only closed set is S itself.
- **Transience and recurrence** We studied transience and recurrence in the lectures about random walks (we just did not call them that). The situation highly depends on the probability p of making an up-step. If $p > \frac{1}{2}$, there is a positive probability that the first step will be "up", so that $X_1 = 1$. Then, we know that there is a positive probability that the walk will never hit 0 again. Therefore, there is a positive probability of never returning to 0, which means that the state 0 is transient. A similar argument can be made for any state i and any probability $p \neq \frac{1}{2}$. What happens when $p = \frac{1}{2}$? In order to come back to 0, the walk needs to return there from its position at time $n = 1$. If it went up, the we have to wait for the walk to hit 0 starting from 1. We have shown that this *will* happen sooner or later, but that the expected time it takes is infinite. The same argument works if $X_1 = -1$. All in all, 0 (and all other states) are null-recurrent (recurrent, but not positive recurrent).
- **Periodicity**. Starting from any state $i \in S$, we can return to it after $2, 4, 6, \dots$ steps. Therefore, the return set $R(i)$ is always given by $R(i) = \{2, 4, 6, \dots\}$ and so $d(i) = 2$ for all $i \in S$.

Gambler's ruin: $p \in (0, 1)$.

- **Communication and classes**. The winning state a and the losing state 0 are clearly absorbing, and form one-element classes. The other $a - 1$ states intercommunicate among each other, so they form a class of their own. This class is not closed (you can - and will - exit it and get absorbed sooner or later).
- **Transience and recurrence**. The absorbing states 0 and a are (trivially) positive recurrent. All the other states are transient: starting from any state $i \in \{1, 2, \dots, a-1\}$, there is a positive probability (equal to p^{a-i}) of winning every one of the next $a - i$ games and, thus, getting absorbed in a before returning to i.
- **Periodicity**. The absorbing states have period 1 since $R(0) = R(a) = \mathbb{N}$. The other states have period 2 (just like in the case of a random walk).

Deterministically monotone Markov chain

- **Communication and classes**. A state i communicates with the state j if and only if $j \geq i$. Therefore $i \leftrightarrow j$ if and only if $i = j$, and so, each $i \in S$ is in a class by itself.

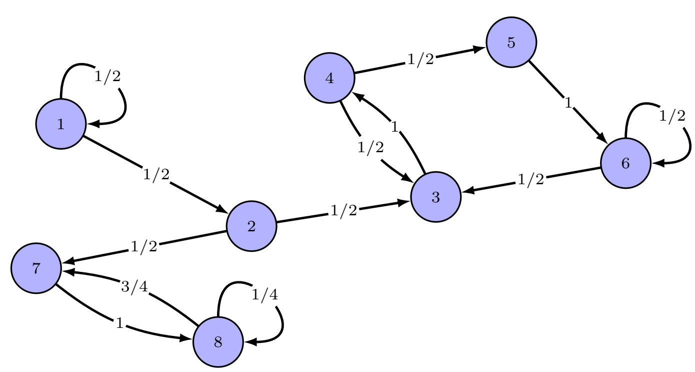
- **Closed sets**. The closed sets are precisely the sets of the form $B = i, i+1, i+2, \dots$, for $i \in \mathbb{N}$.
- **Transience and recurrence** All states are transient.
- **Periodicity**. The return set $R(i)$ is empty for each $i \in S$, so $d(i) = 1$, for all $i \in S$.

A game of tennis

- **Communication and classes**. All the states except for those in $E = \{(40, Adv), (40, 40), (Adv, 40)$, Serena wins, Roger wins$\}$ intercommunicate only with themselves, so each $i \in S \setminus E$ is in a class by iteself. The winning states *Serena wins* and *Roger wins* are absorbing, and, so, also form classes with one element. Finally, the three states in $\{(40, Adv), (40, 40), (Adv, 40)\}$ intercommunicate with each other, so they form the last class.
- **Periodicity**. The states i in $S \setminus E$ have have the property that $p_{ii}^{(n)} = 0$ for all $n \in \mathbb{N}$, so $d(i) = 1$. The winning states are absorbing so $d(i) = 1$ for $i \in \{$Serena wins, Roger wins$\}$. Finally, the return set for the remaining three states is $\{2, 4, 6, \dots\}$ so their period is 2.

8.8 Problems

Problem 8.8.1. Consider a Markov Chain whose transition graph is given below



1. Identify the classes.
2. Find transient and recurrent states.
3. Find periods of all states.

4. Compute $f_{13}^{(n)}$, for all $n \in \mathbb{N}$, where $f_{ij}^{(n)} = \mathbb{P}_i[T_j(1) = n]$.

5. Using software, we can get that, approximately,

$$P^{20} = \begin{bmatrix} 0 & 0 & 0.15 & 0.14 & 0.07 & 0.14 & 0.21 & 0.29 \\ 0 & 0 & 0.13 & 0.15 & 0.07 & 0.15 & 0.21 & 0.29 \\ 0 & 0 & 0.3 & 0.27 & 0.15 & 0.28 & 0 & 0 \\ 0 & 0 & 0.27 & 0.3 & 0.13 & 0.29 & 0 & 0 \\ 0 & 0 & 0.29 & 0.28 & 0.15 & 0.28 & 0 & 0 \\ 0 & 0 & 0.28 & 0.29 & 0.14 & 0.29 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.43 & 0.57 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.43 & 0.57 \end{bmatrix},$$

   where P is the transition matrix of the chain. Compute the probability $\mathbb{P}[X_{20} = 3]$, if the initial distribution (the distribution of X_0) is given by $\mathbb{P}[X_0 = 1] = 1/2$ and $\mathbb{P}[X_0 = 3] = 1/2$.

Problem 8.8.2. Suppose that all classes of a Markov chain are recurrent, and let i, j be two states such that $i \to j$. For each of the 4 statements before, either explain why it is true, or give an example of a Markov chain in which it fails.

(a) for each state k, either $i \to k$ or $j \to k$

(b) $j \to i$

(c) $p_{ji} > 0$ or $p_{ij} > 0$

(d) $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$

Problem 8.8.3.

1. Identify the recurrent states in the setting of (the solution of) Problem **??** of Lecture 7,

2. What are the periods of states in (the solution of) Problem **??** of Lecture 7? Which states are recurrent and which are transient?

Problem 8.8.4. A fair 6-sided die is rolled repeatedly, and for $n \in \mathbb{N}$, the outcome of the n-th roll is denoted by Y_n (it is assumed that $\{Y_n\}_{n\in\mathbb{N}}$ are independent of each other). For $n \in \mathbb{N}_0$, let X_n be the remainder (taken in the set $\{0, 1, 2, 3, 4\}$) left after the sum $\sum_{k=1}^{n} Y_k$ is divided by 5, i.e. $X_0 = 0$, and

$$X_n = \sum_{k=1}^{n} Y_k \pmod 5, \text{ for } n \in \mathbb{N},$$

making $\{X_n\}_{n\in\mathbb{N}_0}$ a Markov chain on the state space $\{0, 1, 2, 3, 4\}$ (no need to prove this fact).

Write down the transition matrix of the chain, classify the states, separate recurrent from transient ones, and compute the period of each state.

Problem 8.8.5. Let $\{Z_n\}_{n\in\mathbb{N}_0}$ be a Branching process (with state space $S = \{0,1,2,3,4,\dots\} = \mathbb{N}_0$) with the offspring probability given by $p_0 = 1/2$, $p_2 = 1/2$. Classify the states (find classes), and describe all closed sets.

Problem 8.8.6. Let C be a class in a Markov chain. For each of the following statements either explain why it is true, or give an example showing that it is false.

(a) C is closed,

(b) C^c is closed,

(c) at least one state in C is recurrent,

(d) for all states $i, j \in C$, $p_{ij} > 0$,

Problem 8.8.7. Consider a Markov chain whose state space has n elements ($n \in \mathbb{N}$). For each of the following statements either explain why it is true, or give an example showing that it is false.

(a) all classes are closed

(b) at least one state is transient,

(c) not more than half of all states are transient,

(d) there are at most n classes,

Problem 8.8.8. Let C_1 and C_2 be two (different) classes. For each of the following statements either explain why it is true, or give an example showing that it is false.

(a) $i \to j$ or $j \to i$, for all $i \in C_1$, and $j \in C_2$,

(b) $C_1 \cup C_2$ is not a class,

(c) if $i \to j$ for some $i \in C_1$ and $j \in C_2$, then $k \not\to l$ for all $k \in C_2$ and $l \in C_1$,

(d) if $i \to j$ for some $i \in C_1$ and $j \in C_2$, then $k \to l$ for some $k \in C_2$ and $l \in C_1$,

Problem 8.8.9. Let i be a recurrent state with period 5, and let j be another state. For each of the following statements either explain why it is true, or give an example showing that it is false.

(a) if $j \to i$, then j is recurrent,

(b) if $j \to i$, then j has period 5,

(c) if $i \to j$, then j has period 5,

(d) if $j \not\to i$ then j is transient,

Problem 8.8.10. Let i and j be two states such that i is transient and $i \leftrightarrow j$. For each of the following statements either explain why it is true, or give an example showing that it is false.

(a) $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \sum_{n=1}^{\infty} p_{ii}^{(n)}$,

(b) if $i \to k$, then k is transient,

(c) if $k \to i$, then k is transient,

(d) period of i must be 1,

Problem 8.8.11. Suppose there exists $n \in \mathbb{N}$ such that $P^n = I$, where I is the identity matrix and P is the transition matrix of a finite-state-space Markov chain. For each of the following statements either explain why it is true, or give an example showing that it is false.

(a) $P = I$,

(b) all states belong to the same class,

(c) all states are recurrent

(d) the period of each state is n,