Introduction to Stochastic Processes
Math 310-2

Introduction

In Math 310-1, you studied random variables and sequences of random variables that were *independent*. Because they were independent, they satisfied properties like the central limit theorem and the law of large numbers.

But, many worldly phenomena are not independent. In this class we will study a basic type of non-independent process—Markov chains.

Markov Chain

DEF

A sequence of random variables (X_n) is a ***Markov chain*** (or Markov for short) if for n and any sequence of events (A_i) having positive probability,

$$\mathbb{P}(X_n \in A_n \mid X_{n-1} \in A_{n-1}) = \mathbb{P}(X_n \in A_n \mid X_{n-1} \in A_{n-1}, X_{n-2} \in A_{n-2}, \ldots).$$

In other words, (X_n) is *Markov* if knowing the previous value in the sequence gives you the same amount of information as knowing all previous values. Since we will be looking at conditional probabilities a lot, to keep things less cluttered, we will sometimes avoid writing the events. That is,

$$\mathbb{P}(X_n \in A_n \mid X_{n-1}, X_{n-2}, \ldots) \equiv \mathbb{P}(X_n \in A_n | X_{n-1} \in A_{n-1}, X_{n-2} \in A_{n-2}, \ldots)$$

where (A_i) is a sequence of positive probability events. Further, $\mathbb{P}(X_n)$, will mean the distribution of X_n.

Let's look at some sequences and decide whether or not they are Markov.

1. You're walking around Chicago. At each street intersection, you roll a die to decide which street to go down next. (X_n) is the sequence of streets you walk down.

   Markov! The next street you decide to go down depends only on your current position, not on how you got there.

2. You're trying to hack a computer network. Every minute you have a 50% chance to crack another password. If you've cracked $k/2$ passwords in the last k minutes, your heart-rate gets high, otherwise it is normal. (X_i) is your heart-rate in the ith minute.

   Markov, but not *stationary*. For minute n, if you are told that X_{n-1} =caught, you know with probability 1 that X_n will be caught. Alternatively, if you're told X_{n-1} =not caught, you know that $X_{n-2}, \ldots$ =not caught, and so you get know extra information knowing all previous states.

3. You're trying to hack a computer network. Every minute you have a 50% chance to crack another password. If you crack k passwords, you will be discovered. (X_i) is how many passwords you've cracked by the ith minute.

   Markov! If you've cracked n passwords in the ith minute, you still have a 50/50 chance of cracking your $(n+1)$th password in the $(i+1)$th minute, regardless of whether you cracked n or $n-1$ in the $(i-1)$th minute.

4. (X_i) is an independent and identically distributed (iid) sequence of Bernoulli $(1/3, 2/3)$ random variables.

   Markov! We trivially have $\mathbb{P}(X_i \mid X_{i-1}) = \mathbb{P}(X_i) = \mathbb{P}(X_i \mid X_{i-1}, X_{i-2}, \ldots)$ by independence.

5. You repeatedly take a test to become part of your favorite fan club. (X_i) is your average score on the first i tests.

   Markov, but not *stationary*. This one is subtle. Suppose that your score on the ith test is given by the independent sequence S_i. Then

   $$X_i = \frac{1}{n}\sum_{n=1}^{i} S_i.$$

   We then have $\mathbb{P}(X_i = a \mid X_{i-1} = b) = \mathbb{P}\left(\frac{S_i+(i-1)b}{i} = a\right)$. Since S_i is an independent sequence, knowing X_{i-2} doesn't change this probability.

Day 2

Notation

Let (X_i) have the Markov property. We call the codomain$(X_i) = X^{(i)}$ the *state space*, and we will assume $X^{(i)} = X^{(j)}$.

$\mathbb{P}(X_i = b \mid X_{i-1} = a)$ is called the *transition probability* from state a to state b at time i. If

$$\mathbb{P}(X_i = b \mid X_{i-1} = a) = \mathbb{P}(X_j = b \mid X_{j-1} = a) \qquad \text{for all } i \text{ and } j,$$

we call (X_i) *stationary*. We will only consider stationary processes.

For a stationary process (X_i),

$$P(a,b) \equiv \mathbb{P}(X_i = b \mid X_{i-1} = a) \qquad \text{and} \qquad P^n(a,b) \equiv \mathbb{P}(X_i = b \mid X_{i-n} = a)$$

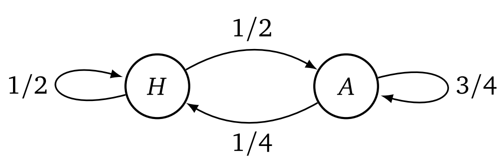
are called the *one-step transition function* and the *multi-step transition function*.

A Simple Example

Your very conservative aunt decided to go to a Chicago comedy club. Each joke leaves her happy (H) or angry (A). If she is happy, the next joke has a 50/50 chance of leaving her happy or angry. If she is angry, the next joke has a 1/4th chance of making her happy and a 3/4ths chance of making her angry.

Day 3

We can model this situation with a directed graph:



DEF The ***transition matrix*** associated with a stationary Markov chain (X_i) having n states $\{1,2,\ldots,n\}$ is the matrix $T = [t_{ij}]$ where $t_{ij} = P(i,j)$.

For this example, letting happy be the first state and angry be the second state, we have

$$T = \begin{array}{c c} & \begin{array}{cc} H & A \end{array} \\ \begin{array}{c} H \\ A \end{array} & \left[\begin{array}{cc} \mathbb{P}(H \to H) & \mathbb{P}(H \to A) \\ \mathbb{P}(A \to H) & \mathbb{P}(A \to A) \end{array}\right] \end{array} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}.$$

Let's compute $P^2(A,A) = \mathbb{P}(X_i = A \mid X_{i-2} = A)$. There are only two ways for this to happen. Either the states went

$$A \to A \to A \qquad \text{or} \qquad A \to H \to A.$$

$\mathbb{P}(A \to A \to A) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}$ and $\mathbb{P}(A \to H \to A) = \frac{1}{4} \cdot \frac{1}{2} = \frac{2}{16}$, and so

$$P^2(A,A) = \mathbb{P}(A \to A \to A) + \mathbb{P}(A \to H \to A) = \frac{11}{16}.$$

The P^2 transition matrix now looks like $\begin{bmatrix} ? & ? \\ ? & \frac{11}{16} \end{bmatrix}$. Computing every entry we get $\begin{bmatrix} \frac{6}{16} & \frac{10}{16} \\ \frac{5}{16} & \frac{11}{16} \end{bmatrix}$.

Let's look at what happened from a matrix perspective: To compute $P^2(A,A)$, we need to compute the quantity $Q = \mathbb{P}(A \to A \to A) + \mathbb{P}(A \to H \to A)$, and by the Markov property,

$$\mathbb{P}(A \to A \to A) = \mathbb{P}(A \to A)\mathbb{P}(A \to A) \qquad \text{and} \qquad \mathbb{P}(A \to H \to A) = \mathbb{P}(A \to H)\mathbb{P}(H \to A).$$

If we squint, we see

$$Q = \begin{bmatrix} \mathbb{P}(H \to A) \\ \mathbb{P}(A \to A) \end{bmatrix} \cdot \begin{bmatrix} \mathbb{P}(A \to H) \\ \mathbb{P}(A \to A) \end{bmatrix} = \begin{bmatrix} \mathbb{P}(H \to A) & \mathbb{P}(A \to A) \end{bmatrix} \begin{bmatrix} \mathbb{P}(A \to H) \\ \mathbb{P}(A \to A) \end{bmatrix} = (\begin{bmatrix} 0 & 1 \end{bmatrix} T) \left(T \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 0 & 1 \end{bmatrix} T^2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

In other words, the $(2,2)$ entry of the transition matrix for $P^2(x,y)$ is the $(2,2)$ entry of T^2. In fact, the transition function $P^2(x,y)$ is given by the matrix T^2!

Using induction, we can show that the transition function $P^n(x,y)$ is given by the matrix T^n for any n, but first let's prove a lemma we've already used once.

LEM

If (X_i) is a Markov chain and (s_i) is a sequence of states of (X_i), then

$$\mathbb{P}(s_1 \to s_2 \to \cdots \to s_n) = \prod_{i=1}^{n-1} \mathbb{P}(s_i \to s_{i-1}).$$

We used this lemma already when we said $\mathbb{P}(A \to H \to A) = \mathbb{P}(A \to H)\mathbb{P}(H \to A)$, and it seems quite intuitive given the definition of Markov.

Proof. We will proceed by induction. Trivially, $\mathbb{P}(s_1 \to s_2) = \mathbb{P}(s_1 \to s_2)$. Suppose $\mathbb{P}(s_1 \rightsquigarrow s_n) = \prod \mathbb{P}(s_i \to s_{i+1}$, and consider $\mathbb{P}(s_1 \rightsquigarrow s_n \to s_{n+1})$. By definition

$$\mathbb{P}(s_n \to s_{n+1}) = \mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n)$$
$$= \mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n, X_{n-1} = s_{n-1}, \ldots)$$

with the last equality following from the Markov property. Now by the definition of conditional probability we have

$$\mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n, X_{n-1} = s_{n-1}, \ldots) = \frac{\mathbb{P}(X_{n+1} = s_{n+1}, X_n = s_n, X_{n-1} = s_{n-1}, \ldots)}{\mathbb{P}(X_n = s_n, X_{n-1} = s_{n-1}, \ldots)} = \frac{\mathbb{P}(s_1 \rightsquigarrow s_n \to s_{n+1})}{\mathbb{P}(s_1 \rightsquigarrow s_n)}.$$

Multiplying both sides by $\mathbb{P}(s_1 \rightsquigarrow s_n)$ gives us

$$\mathbb{P}(s_1 \rightsquigarrow s_n \to s_{n+1}) = \mathbb{P}(s_1 \rightsquigarrow s_n)\mathbb{P}(s_n \to s_{n+1}).$$

Since we assume that $\mathbb{P}(s_1 \rightsquigarrow s_n)$ could be written as a product, we've completed the proof. □

Day 4

We're almost ready to tackle our first theorem, but let's first prove two more easy facts.

LEM

If (X_i) is a Markov chain with finite state space S, then

$$\mathbb{P}(X_n = b) = \sum_{s \in S} \mathbb{P}(X_n = b \mid X_{n-1} = s)\mathbb{P}(X_{n-1} = s).$$

Proof. Since $\mathbb{P}(X_i \in S) = 1$, we have

$$\mathbb{P}(X_n = b) = \mathbb{P}(X_n = b \text{ and } X_{n-1} \in S) = \mathbb{P}\left(\bigcup_{s \in S} X_n = b \text{ and } X_{n-1} = s \right).$$

Since the events ($X_n = b$ and $X_{n-1} = s$) and ($X_n = b$ and $X_{n-1} = s'$) are disjoint if $s \neq s'$, we have

$$\mathbb{P}\left(\bigcup_{s\in S} X_n = b \text{ and } X_{n-1} = s\right) = \sum_{s\in S} \mathbb{P}(X_n = b \text{ and } X_{n-1} = s).$$

Noticing the summand is just $\mathbb{P}(X_n = b \mid X_{n-1} = s)\mathbb{P}(X_{n-1} = s)$ completes the proof. □

LEM If (X_i) is a Markov chain with finite state space S, then $\mathbb{P}(X_n = b \mid X_1 = a)$ is the sum of $\mathbb{P}(a \to s_2 \to s_3 \to \cdots \to s_{n-1} \to b)$ for every possible combination of $s_2, \ldots, s_{n-1} \in S$.

Proof. Like the previous proof, this follows from manipulations of the basic definitions.

First note, if $s_2, \ldots, s_{n-1}$ and $s_2', \ldots, s_{n-1}'$ are not the exact same sequence, then the events $a \to s_2 \to \cdots \to s_{n-1} \to b$ and $a \to s_2' \to \cdots \to s_{n-1}' \to b$ are disjoint. Thus

$$\mathbb{P}\left(\bigcup_{s_2,\ldots,s_{n-1}\in S} a \to s_2 \to \cdots \to s_{n-1} \to b\right) = \sum_{s_2,\ldots,s_{n-1}\in S} \mathbb{P}(a \to s_2 \to \cdots \to s_{n-1} \to b).$$

Now, by the definition of conditional probability,

$$\mathbb{P}(a \to s_2 \to \cdots \to s_{n-1} \to b) = \frac{\mathbb{P}(X_1 = a, X_2 = s_2, \ldots, X_{n-1} = s_{n-1}, X_n = b)}{\mathbb{P}(X_1 = a)},$$

and so

$$\mathbb{P}\left(\bigcup_{s_2,\ldots,s_{n-1}\in S} a \to s_2 \to \cdots \to s_{n-1} \to b\right) = \frac{\mathbb{P}(X_1 = a, X_2 \in S, \ldots, X_{n-1} \in S, X_n = b)}{\mathbb{P}(X_1 = a)} = \frac{\mathbb{P}(X_1 = a, X_n = b)}{\mathbb{P}(X_1 = a)},$$

with the last equality following because $X_i \in S$ always. The right side of this equation is just $\mathbb{P}(X_n = b \mid X_1 = a)$, and so combining all of our equations proves the result. □

Now we're ready to tackle our theorem linking matrix multiplication and multi-step transition functions.

THM For a Markov chain (X_i) having states $\{1, \ldots d\}$ and transition function $P(x, y)$, let T be the corresponding transition matrix. Then, the transition matrix corresponding to the multi-step transition function $P^n(x, y)$ is T^n.

Proof. We will proceed by induction. By definition, T gives the transition probabilities for $P(x, y)$. Suppose that T^{n-1} gives the transition probabilities for $P^{n-1}(x, y)$. Since (X_i) is a Markov processes,

$$\mathbb{P}(s_1 \rightsquigarrow s_{n-1} \to s_n) = \mathbb{P}(s_1 \rightsquigarrow s_{n-1})\mathbb{P}(s_{n-1} \to s_n),$$

where $s_1, \ldots, s_n$ are states. Further, from our previous lemmas,

$$\mathbb{P}(X_n = s_n \mid X_1 = s_1) = \sum_{s\in S} \mathbb{P}(X_n = s_n \mid X_{n-1} = s_{n-1})\mathbb{P}(X_{n-1} = s_{n-1} \mid X_1 = s_1)$$

$$= \sum_{s\in S} P(s, s_n)P^{n-1}(s_1, s) = \sum_{s\in S} P^{n-1}(s_1, s)P(s, s_n).$$

However, the right hand side is the same as the dot product of the s_n column of T with the s_1 row of T^{n-1}, which is the definition of the (s_1, s_n) entry of $TT^{n-1} = T^n$. □

Now we know how to analyze transition probabilities for any finite-state Markov chain using matrices!

Let's see how our probability questions become linear algebra questions. Suppose we'd like to know the distribution after one joke supposing that your aunt started happy. That is, we'd like to know $f(x) = P(H, x)$. This is just the first row of T, and so the distribution is given by $\begin{bmatrix} 1 & 0 \end{bmatrix} T$. The distribution after 4 steps supposing your aunt started off angry is $\begin{bmatrix} 0 & 1 \end{bmatrix} T^7$. And $P^7(A, H)$ is $\begin{bmatrix} 0 & 1 \end{bmatrix} T^7 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

But we can do more. Suppose we wanted to know the distribution after one joke supposing that there was a 50/50 chance your aunt came into the club happy or angry. That's just $\begin{bmatrix} 1/2 & 1/2 \end{bmatrix} T$. In general, if $\vec{d}$ is a *row vector* specifying an initial distribution, then $\vec{d}T^n$ gives the resulting distribution after n steps.

Day 5

Consider now the following questions: After many many jokes, what is the probability that your aunt leaves the club happy? What's the probability she leaves angry? Does the state she comes into the club with affect these probabilities? How does the number of jokes told affect these probabilities?

We can answer all of these questions with linear algebra! Since we're going to be computing large powers of T, let's diagonalize it.

$$T = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1/4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix} = SDS^{-1}.$$

Since $(1/4)^n \to 0$ as $n \to \infty$, we can easily compute

$$T^\infty \equiv \lim_{n\to\infty} T^n = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{bmatrix}.$$

From this we can see $\begin{bmatrix} 1 & 0 \end{bmatrix} T^\infty = \begin{bmatrix} 0 & 1 \end{bmatrix} T^\infty$ and so her chances of being happy or angry upon leaving the club after many jokes doesn't depend on her initial state. In either case, she leaves happy with a $1/3$ chance and angry with a $2/3$ chance. We can also conclude that the number of jokes told doesn't really matter as long as the number is large. Indeed

$$T^n - T^m = S(D^n - D^m)S^{-1} = S \begin{bmatrix} \frac{1}{4^n} - \frac{1}{4^m} & 0 \\ 0 & 0 \end{bmatrix} S^{-1}.$$

As long as n and m are large, $\frac{1}{4^n} - \frac{1}{4^m} \approx 0$, and so the difference is approximately the zero matrix.

Day 6

Hitting Times

Consider a Markov chain for a gambler. The state space would be \$$d \geq 0$. If the gambler has any money, he can bet it and transition to a state with more money or a state with less money. But, if the gambler ever has \$0, he cannot bet any further.

The *Gambler's Ruin* problem asks, how long does it take to end up in state \$0?

Hitting Time (DEF)

Let (X_i) be a Markov chain. A ***hitting time*** for an event B is the random variable

$$T_B = \min\{i \geq 0 : X_i \in B\}.$$

We will not compute hitting times right now, but hitting times give us another way to think about Markov chains. For example,

$$P^n(x,y) = \sum_{m\leq n} P(T_{\{y\}} = m \mid X_1 = x) P^{n-m}(y,y).$$

(Why is this formula intuitive?)

Different Types of States

Consider the gambler's ruin problem. If $a, b > 0$, there's a non-zero chance of starting at state a and ending at state b. (Is the correct way to mathematical quantify this statement $P(a,b) > 0$?) However, $P^n(0,b) = 0$ unless $b = 0$. That is, you cannot transition out of state \$0.

Absorbing (DEF)

A state a of a Markov chain is called ***absorbing*** if $P(a,b) = 0$ whenever $b \neq a$.

In plain language, a state is absorbing if once you enter it, you can never leave.

For a Markov chain, define

$$\rho_{xy} = \mathbb{P}(T_{\{y\}} < \infty \mid X_1 = x).$$

That is, ρ_{xy} is the probability you end up at state y at some time in the future having started at state x.

DEF

Recurrent & Transient

For a Markov chain (X_i), a state a is called ***recurrent*** if $\rho_{aa} = 1$, otherwise it is called ***transient*** .

What are the transient and recurrent states for the gambler's ruin problem? Well, the state \$0 is absorbing, and if the odds are in the house's favor, $\rho_{x0} = 1$. That is, you will always eventually end up at \$0. Thus, the state \$0 is recurrent. However, if you start at state \$$d > 0$, there is a positive probability you end up at state \$0, and so $\rho_{dd} < 1$. This means that every state \$$d > 0$ is transient.

Now we get to our first big theorem.

THEOREM

Let (X_i) be a Markov chain, let $N(x)$ be the number of times state x is visited, and let $G(y,x) = \mathbb{E}(N(x) \mid X_0 = y)$. Then,

(i) if x is recurrent, $\mathbb{P}(N(x) = \infty \mid X_1 = x) = 1$ and $G(x,x) = \infty$, and

(i) if x is transient, $\mathbb{P}(N(x) < \infty \mid X_1 = y) = 1$ and

$$G(y,x) = \frac{\rho_{yx}}{1 - \rho_{xx}}.$$

In plain words, the theorem says that recurrent states are visited infinitely often and transient states are visited finitely often. Also, we can precisely compute the expectation of the number of times a transient state is visited in terms of ρ.

Day 7

Proof. (i). Suppose x is a recurrent state. Let $T^{(n)}_{\{x\}} = \min\{i > n : X_i = x\}$ be the first return time to $\{x\}$ after time n. Since X_i is a stationary process,

$$\mathbb{P}(T_{\{x\}} < \infty \mid X_1 = x) = \mathbb{P}(T^{(n)}_{\{x\}} < \infty \mid X_n = x).$$

Since x is recurrent, $\mathbb{P}(T_{\{x\}} < \infty \mid X_1 = x) = 1$, and so $\mathbb{P}(T^{(n)}_{\{x\}} < \infty \mid X_n = x) = 1$. Thus if we are ever in state x we must again be in state x at some point in the future, and so $N(x) = \infty$. It immediately follows that $G(x,x) = \mathbb{E}(\infty) = \infty$.

(ii). Suppose x is transient. Let p_n be the probability that, starting at y, we hit x at least n times. Computing with the Markov property, we see

$$p_n = \rho_{yx}\rho_{xx}^{n-1} = \mathbb{P}(N(x) \geq n \mid X_1 = y).$$

We'd like to compute $\mathbb{P}(N(x) < \infty \mid X_1 = y)$. Applying an easy lower bound, we see

$$\begin{aligned}
\mathbb{P}(N(x) < \infty \mid X_1 = y) &= \sum_{i<\infty} \mathbb{P}(N(x) = i \mid X_1 = y) \\
&\geq \sum_{i \leq n} \mathbb{P}(N(x) = i \mid X_1 = y) \\
&= 1 - \mathbb{P}(N(x) \geq n \mid X_1 = y) = 1 - \rho_{yx}\rho_{xx}^{n-1},
\end{aligned}$$

and so $\mathbb{P}(N(x) < \infty \mid X_1 = y) = 1$.

Now, let $d(n) = \mathbb{P}(N(x) = n \mid X_1 = y)$ be the distribution of $N(x)$. We can explicitly write d as

$$\begin{aligned}
d(n) &= \mathbb{P}(N(x) \geq n \mid X_1 = y) - \mathbb{P}(N(x) \geq n+1 \mid X_1 = y) \\
&= \rho_{yx}\rho_{xx}^{n-1} - \rho_{yx}\rho_{xx}^{n} = \rho_{yx}\rho_{xx}^{n-1}(1 - \rho_{xx}),
\end{aligned}$$

and so

$$G(y,x) = \mathbb{E}(N(x) \mid X_1 = y) = \sum n d(n) = \sum n \rho_{yx}\rho_{xx}^{n-1}(1 - \rho_{xx}) = \frac{\rho_{yx}}{1 - \rho_{xx}}.$$

□

In light of the previous theorem, we can think of recurrent and transient states as synonymous with being visited a finite number of times.

To ease notation, since we will often talk about $\mathbb{P}(E \mid X_0 = x)$ for events E and states x, we will use the shorthand notation $\mathbb{P}_x(E)$.

Recurrent & Transient Chains

DEF A Markov chain is called a ***recurrent chain*** if every state is recurrent and it is called a ***transient chain*** if every state is transient.

Note that, unlike states, there can be Markov chains that are neither recurrent nor transient.

Closed

DEF Let (X_i) be a Markov chain with state space $\mathcal{S}$. A subset $\mathcal{A} \subseteq \mathcal{S}$ is called ***closed*** if $P(a,b) = 0$ whenever $a \in \mathcal{C}$ and $b \in \mathcal{A}^C$, where $\mathcal{A}^C$ is the complement of $\mathcal{C}$.

In other words, a set $\mathcal{A}$ is closed if $\mathcal{A}$ as a whole acts like an absorbing state.

LEM Let $\mathcal{A}$ be a closed set for a Markov chain and suppose $a \in \mathcal{A}$ and $b \notin \mathcal{A}$. Then $\rho_{ab} = 0$.

Proof. Suppose $\rho_{ab} > 0$. Then there exists some sequence of states $a = x_0 \to x_1 \to \cdots \to x_{n-1} \to x_n = b$ such that $\mathbb{P}(x_0 \to x_1 \to \cdots \to x_{n-1} \to x_n) > 0$. For each i, $x_i \in \mathcal{A}$ or $x_i \notin \mathcal{A}$. Let k be the smallest index such that $x_k \notin \mathcal{A}$. Since $x_0 \in \mathcal{A}$, $k \geq 1$.

Now, since $\mathcal{A}$ is closed and $x_{k-1} \in \mathcal{A}$ and $x_k \notin \mathcal{A}$, we have that $P(x_{k-1}, x_k) = 0$. But this means, $\mathbb{P}(x_0 \to x_1 \to \cdots \to x_{n-1} \to x_n) = \prod_{i\geq 1}^{n} \mathbb{P}(x_{i-1} \to x_i) = 0$, which is a contradiction. □

Leads To

DEF For a Markov chain, we say a state a ***leads to*** a state b if $\rho_{ab} > 0$, and we notate it $a \underset{\text{leads to}}{\rightsquigarrow} b$.

Irreducible

DEF Let (X_i) be a Markov chain with state space $\mathcal{S}$. A subset $\mathcal{A} \subseteq \mathcal{S}$ is called ***irreducible*** if $\rho_{ab} > 0$ for any $a, b \in \mathcal{A}$.

Eventually we will learn how to decompose a Markov chain into closed and irreducible components.

THM Let (X_i) be a Markov chain with state space $\mathcal{S}$. For $x, y \in \mathcal{S}$, suppose that x is recurrent and $x \underset{\text{leads to}}{\rightsquigarrow} y$. Then y is recurrent and $\rho_{xy} = \rho_{yx} = 1$.

Proof. We will first prove the result for the one-step process. Fix $x, y \in \mathcal{S}$ such that x is recurrent and $\mathbb{P}(x \to y) > 0$. Since x is recurrent, $\mathbb{P}_x(T_x = \infty) = 0$, so

$$0 = \mathbb{P}_x(T_x = \infty) \geq \mathbb{P}(x \to y)(1 - \rho_{yx}).$$

But, we assumed $\mathbb{P}(x \to y) > 0$, so we conclude $\rho_{yx} = 1$ (because $1 - \rho_{yx} = 0$).

Let $N^j(a) = \#\{X_i = a : i \leq j\}$. Now, since x is recurrent, for any $\varepsilon > 0$ and any k, we may find an n such that $\mathbb{P}_x(N^n(x) \geq k) > 1 - \varepsilon$. Let $C_w^n(a) = \{\text{chains of length } n \text{ with } w \text{ occurrences of } a\}$. Now, we have

$$\mathbb{P}_x(N^n(x) \geq k) = \sum_{w=k}^{n} \sum_{C \in C_w^n(x)} \mathbb{P}_x(C) > 1 - \varepsilon,$$

and so in particular

$$\sum_{w=0}^{k} \sum_{C \in C_w^n(x)} \mathbb{P}_x(C) \leq \varepsilon.$$

Consider a chain $C \in C_w^n(x) \cap C_0^n(y)$. Since the chain is Markov, we know $\mathbb{P}(\cdots \to x \to s \to \ldots) = \mathbb{P}(\cdots \to x)\mathbb{P}(x \to s)\mathbb{P}(s \to \ldots)$. Since there are no occurrences of y, $\mathbb{P}(x \to s) = 1 - P(x, y)$. Since there are w occurences of x in C, $\mathbb{P}_x(C) \leq (1 - P(x, y))^k$.

We are now ready to compute

$$\mathbb{P}_x(N(y)=0) \leq \mathbb{P}_x(N^n(y)=0) = \sum_{w=0}^{n} \sum_{C\in C_w^n(x)\cap C_0^n(y)} \mathbb{P}_x(C)$$

$$= \sum_{w=k+1}^{n} \sum_{C\in C_w^n(x)\cap C_0^n(y)} \mathbb{P}_x(C) + \sum_{w=0}^{k} \sum_{C\in C_w^n(x)\cap C_0^n(y)} \mathbb{P}_x(C)$$

$$\leq (1-P(x,y))^k + \varepsilon.$$

Since $P(x,y)>0$ and this holds for all ε and k, we conclude that, we conclude that $\mathbb{P}_x(N(y)\geq 1)=1$ and so $\rho_{xy}=1$.

Since

$$\rho_{yy} \geq \rho_{yx}\rho_{xy} = 1,$$

y is recurrent.

We have proven that if y is one step from x and $\rho_{xy}>0$ and x is recurrent, then y is recurrent.

Now, suppose $x \underset{\text{leads to}}{\rightsquigarrow} y$ and x is recurrent. This means there exists some chain $x=s_0\to s_1\to\cdots\to s_{n-1}\to s_n=y$ such that $\mathbb{P}(s_0\to\ldots\to s_n)>0$. Suppose one of the s_i is not recurrent, and let k be the minimum index of such an s_i. Since $s_0=x$ is recurrent, $k\geq 1$. We now have s_{k-1} is recurrent and $\mathbb{P}(s_{k-1}\to s_k)>0$, so s_k must be recurrent, a contradition which completes the proof. $\square$

COR Suppose $\mathcal{S}$ is the state space of a Markov chain and $\mathcal{B}\subseteq\mathcal{S}$ is a closed and irreducible set. Then, if $\mathcal{B}$ contains a recurrent state, $\rho_{xy}=1$ for all $x,y\in\mathcal{B}$.

THM If $\mathcal{B}\subseteq\mathcal{S}$ is a finite, closed, and irreducible subset of the state space $\mathcal{S}$ for a Markov chain, then every state in $\mathcal{B}$ is recurrent.

Proof. Fix $x\in\mathcal{B}$. By the pidgeon hole principle, there must be some state $y\in\mathcal{B}$ so that $\mathbb{E}(N(y)\mid X_0=x)=\infty$. By our first theorem, this means that y is recurrent. Since $\mathcal{B}$ is irreducible, $\rho_{yt}>0$ for any $t\in\mathcal{B}$, and so by our second theorem, $\rho_{yt}=1$. $\square$

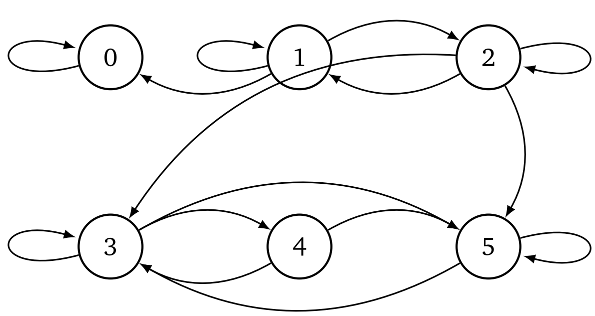
We've stated a lot of theorems. Let's look at a simple example.

Suppose the transition matrix for a Markov chain is

$$\begin{array}{c} \\ 0\\1\\2\\3\\4\\5 \end{array}\begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0\\ 1/4 & 1/2 & 1/4 & 0 & 0 & 0\\ 0 & 1/5 & 2/5 & 1/5 & 0 & 1/5\\ 0 & 0 & 0 & 1/6 & 1/3 & 1/2\\ 0 & 0 & 0 & 1/2 & 0 & 1/2\\ 0 & 0 & 0 & 1/4 & 0 & 3/4 \end{bmatrix} \end{array}.$$

Can we identify the closed, irreducible components? Can we identify the recurrent states?

Well, the transition graph for this chain looks like



Subsets of the graph with only incoming edges are closed. Subsets of the graph where there is a path going through every element are irreducible. We therefore see the closed, irreducible subsets of the state space are $C_0 = \{0\}$ and $C_1 = \{3,4,5\}$. Since these are finite, they must consist only of recurrent points. We can also see that states 1 and 2 are not recurrent since there is a positive probability of transitioning into C_0 or C_1.

THM

Let $\mathcal{S}_R \subseteq \mathcal{S}$ be the set of all recurrent states for a Markov chain with state space $\mathcal{S}$. Then $\mathcal{S}_R = \bigcup C_i$ where $\{C_i\}$ are disjoint, closed, and irreducible.

Proof. Suppose C is a closed set of recurrent points. Then $\underset{\text{leads to}}{\rightsquigarrow}$ is an equivalence relation on C.

It is clear that since for any two states x, y, either $\rho_{xy} > 0$ or $\rho_{xy} = 0$, $\underset{\text{leads to}}{\rightsquigarrow}$ is a relation on the set of states. We will now show it satisfies the properties of an equivalence relation.

(reflexive) If $x \in C$, then $x \underset{\text{leads to}}{\rightsquigarrow} x$ because x is recurrent.

(symmetric) If $x, y \in C$ and $x \underset{\text{leads to}}{\rightsquigarrow} y$, then $y \underset{\text{leads to}}{\rightsquigarrow} x$ by our second theorem.

(transitive) If $x, y, z \in C$ and $x \underset{\text{leads to}}{\rightsquigarrow} y$ and $y \underset{\text{leads to}}{\rightsquigarrow} z$, then $\rho_{xz} \geq \rho_{xy}\rho_{yz} > 0$, so $x \underset{\text{leads to}}{\rightsquigarrow} z$.

Now, notice that $\mathcal{S}_R$ is closed. Suppose $y \notin \mathcal{S}_R$. Then,

$$1 = \rho_{xx} \leq 1 - \rho_{xy},$$

and so $\rho_{xy} = 0$ which is equivalent to $\mathcal{S}_R$ beging closed. Now, let $\{C_i\}$ be the partition coming from the equivalence relation $\underset{\text{leads to}}{\rightsquigarrow}$. □

DEF

For a set of states A, let $\rho_{xA} = \mathbb{P}_x(T_A < \infty)$ be the probability of transitioning from the state x to any state in A in finite time.

If A is a closed set of states, we can think of ρ_{xA} as the probability that starting at x, you will be absorbed by the set A. In light of the previous theorem, we now have a strategy for studying Markov chains.

(i) Find the closed, irreducible subsets of the state space $\{C_i\}$.

(ii) For a transient state x, compute ρ_{xC_i}.

(iii) Study irreducible Markov chains.

For a moment, let's consider (ii). Suppose $\{C_i\}$ is the collection of closed, irreducible subsets of the state space. Then, if $x \in C_i$,

$$\rho_{xC_i} = 1 \qquad \text{and} \qquad \rho_{xC_j} = 0 \text{ for } i \neq j.$$

If x is transient, things are a little harder. But, we do know if x transitions to C_i, it does so in one step or more than one step. Thus,

$$\rho_{xC_i} = \sum_{y \in C_i} P(x,y) + \sum_{y \notin C_i} P(x,y)\rho_{yC_i} = \sum_{y \in C_i} P(x,y) + \sum_{y \in \mathcal{S}_T} P(x,y)\rho_{yC_i}, \tag{1}$$

where $\mathcal{S}_T$ is the set of all transient points.

If we squint at equation (1), we see that taking the collection of $\{\rho_{zC_i}\}$ as unknowns, equation (1) gives us a system of linear equations for $\{\rho_{zC_i}\}$. Thus, finding ρ_{xC_i} should just be a little linear algebra (if the system actually has a solution).

Suppose now that there are only a finite number of transient states $\{t_1, \ldots, t_n\}$. Then, we can rewrite the system corresponding to (1) as a matrix equation of the form

$$\vec{\rho}_{C_i} = P_T \vec{\rho}_{C_i} + \vec{p}_{C_i}, \tag{2}$$

where

$$\vec{\rho}_{C_i} = \begin{bmatrix} \rho_{t_1C_i} \\ \rho_{t_2C_i} \\ \vdots \end{bmatrix} \qquad \vec{p}_{C_i} = \begin{bmatrix} P(t_1, C_i) \\ P(t_2, C_i) \\ \vdots \end{bmatrix}$$

and P_T is the matrix with j,k entry $P(t_j, t_k)$.

Equation (2) has a solution if and only if $(I - P_T)$ is invertible. As it turns out, it always is.

THM If a Markov chain has a finite number of transient states and P_T is the transition matrix between transient states, then $(I - P_T)$ is invertible.

Proof. Showing that $(I - P_T)$ is invertible is equivalent to showing that 1 is not an eigenvalue of P_T.

Suppose that 1 were an eigenvalue for P_T and $\vec{v}$ a corresponding eigenvector. Then

$$\lim_{n\to\infty} P_T^n \vec{v} = \vec{v}.$$

However, the i,j entry of P_T^n is the transition probability $P^n(t_i, t_j)$. Since t_j is transient, it is only hit a finite number of times and so

$$\lim_{n\to\infty} P^n(i,j) = 0.$$

Thus we have that

$$\lim_{n\to\infty} P_T^n \vec{v} = 0\vec{v} = \vec{0},$$

which contradicts $\vec{v}$ being an eigenvector. $\square$