Lecture 2: Markov Chains (I)

Readings Strongly recommended:

- Grimmett and Stirzaker (2001) 6.1, 6.4-6.6

Optional:

- Hayes (2013) for a lively history and gentle introduction to Markov chains.
- Norris (1997), for a canonical reference on Markov chains.
- Koralov and Sinai (2010) 5.1-5.5, pp.67-78 (more mathematical)

We will begin by discussing *Markov chains*. In Lectures 2 & 3 we will discuss *discrete-time* Markov chains, and Lecture 4 will introduce *continuous-time* Markov chains.

2.1 Setup and definitions

We consider a stochastic process $(X_t)_{t\in T}$ with a discrete (finite or countable) state space S, which depends on discrete time $T = \{0,1,2,\dots\}$. We sometimes refer to the process as X_t when it is clear that we mean the process, not the random variable. Since S is countable, we may index it with the positive or nonnegative integers, as in $S = \{1,2,3,\dots\}$.

Definition. The process $X_t = X_0, X_1, X_2, \dots$ is a *discrete-time Markov chain* if it satisfies the *Markov property*:

$$P(X_{n+1} = s|X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} = s|X_n = x_n) \tag{1}$$

or all $x_0, x_i, \dots, s \in S$ and for all $n \geq 0$.

Definition. The quantities $P(X_{n+1} = j|X_n = i)$ in (1) are called the *transition probabilities*. It is convenient to write them as

$$P_{ij}(n) = P(X_{n+1} = j|X_n = i)\,. \tag{2}$$

Definition. The *transition matrix*[1] at time n is the matrix $P(n) = (P_{ij}(n))$, i.e. the (i,j)th element of $P(n)$ is $P_{ij}(n)$.

Notice that the transition matrix has the following properties

(i) $P_{ij}(n) \geq 0 \quad \forall i,j$ (the entries are non-negative)

(ii) $\sum_j P_{ij}(n) = 1 \quad \forall i$ (the rows sum to 1)

Any matrix that satisfies (i), (ii) above is called a *stochastic matrix*. Hence, the transition matrix is a stochastic matrix.

[1] We call P a matrix even if $|S| = \infty$. If the state space is infinite, then we interpret products such as PQ to mean infinite matrix with entries given by the infinite sum

$$(PQ)_{ij} = \sum_k P_{ik} Q_{kj}\,.$$

Remark. Note that a "stochastic matrix" is *not* the same thing as a "random matrix"! Usually "random" can be substituted for "stochastic" but not here. A random matrix is a matrix whose entries are random variables. A stochastic matrix has deterministic entries.

***Exercise* 2.1.** Show that the transition probabilities satisfy properties (i), (ii) above.

***Exercise* 2.2.** Show that if X_t is a discrete-time Markov chain, then

$$P(X_n = s|X_0 = x_0, X_1 = x_1, \ldots, X_m = x_m) = P(X_n = s|X_m = x_m),$$

for any $0 \leq m < n$. That is, the probabilities at the current time, depend only on the most recent known state in the past, even if it's not exactly one step before.

Definition. The Markov chain X_t is *time-homogeneous* if $P(X_{n+1} = j|X_n = i) = P(X_1 = j|X_0 = i)$, i.e. the transition probabilities do not depend on time n. If this is the case, we write $P_{ij} = P(X_1 = j|X_0 = i)$ for the probability to go from i to j in one step, and $P = (P_{ij})$ for the transition matrix. Otherwise, it is *time-inhomogeneous*.

We will mainly consider time-homogeneous Markov chains in this course, though we will occasionally remark on how some results may be generalized to the time-inhomogeneous case.

Examples 2.1

1. *Weather model.* Let X_n be the state of the weather on day n in New York, which we assume is either *rainy* or *sunny*. We could use a Markov chain as a crude model for how the weather evolves day-by-day. The state space is $S = \{\text{rain}, \text{sun}\}$. One transition matrix might be

$$P = \begin{array}{c c} & \begin{array}{cc} \text{sun} & \text{rain} \end{array} \\ \begin{array}{c} \text{sun} \\ \text{rain} \end{array} & \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix} \end{array}$$

   This says that if it is sunny today, then the chance it will be sunny tomorrow is 0.8, whereas if it is rainy today, then the chance it will be sunny tomorrow is 0.4.

   Some questions you might be interested in include: if it is sunny today, what is the probability that it is sunny in two days? Or, what is the long-run fraction of sunny days in New York?

2. *Coin flipping.* Another two-state Markov chain is based on coin flips. Usually coin flips are used as the canonical example of independent Bernoulli trials. However, Diaconis et al. (2007) studied sequences of coin tosses empirically, and found that outcomes in a sequence of coin tosses are *not* independent. Rather, they are well-modelled by a Markov chain with the following transition probabilities:

$$P = \begin{array}{c c} & \begin{array}{cc} \text{heads} & \text{tails} \end{array} \\ \begin{array}{c} \text{heads} \\ \text{tails} \end{array} & \begin{pmatrix} 0.51 & 0.49 \\ 0.49 & 0.51 \end{pmatrix} \end{array}$$

   This shows that if you throw a Heads on your first toss, there is a very slightly higher chance of throwing heads on your second, and similarly for Tails.

3. *Random walk on the line.* Suppose we perform a walk on the integers, starting at some integer k. At each step we move to one unit right with probability p or one unit left with probability $1-p$. The position of the random walker is a Markov chain, which can be constructed explicitly as

$$X_n = \sum_{j=1}^{n} \xi_j, \qquad \xi_j = \begin{cases} +1 & \text{with probability } p \\ -1 & \text{with probability } 1-p \end{cases}, \qquad \xi_i \text{ i.i.d.}$$

The transition probabilities are

$$P_{i,i+1} = p, \quad P_{i,i-1} = 1-p, \quad P_{i,j} = 0 \quad (j \neq i \pm 1).$$

The state space is $S = \{\dots,-1,0,1,\dots\}$, which is countably infinite.

One canonical problem this models is a gambling game. A gambler starts with k\$, and at each game she[2] either wins 1\$ with probability p, or loses 1\$ with probability $1-p$. We might be interested in questions such as: what is her average earnings after n games? What is the probability that she wins 20\$, before she goes broke? On average, how long does it take for her to go broke? We'll show the phenomenon called the *Gambler's Ruin*, which says that even for a fair game with $p = 1/2$, the gambler will go broke with probability 1.

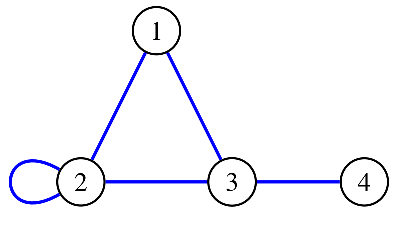
4. *Independent, identically distribute (i.i.d.) random variables.* A sequence of i.i.d. random variables is a Markov chain, albeit a somewhat trivial one. Suppose we have a discrete random variable X taking values in $S = \{1,2,\dots,k\}$ with probability $P(X = i) = p_i$. If we generate an i.i.d. sequence $X_0, X_1, \dots$ of random variables with this probability mass function, then it is a Markov chain with transition matrix

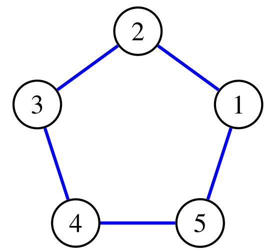
$$P = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & \cdots & k \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ k \end{array} & \begin{pmatrix} p_1 & p_2 & \cdots & p_k \\ p_1 & p_2 & \cdots & p_k \\ \vdots & \vdots & & \vdots \\ p_1 & p_2 & \cdots & p_k \end{pmatrix} \end{array}$$

5. *Random walk on a graph (undirected, unweighted).* Suppose we have a graph (a set of vertices and edge connecting them.) We can perform a random walk on the graph as follows: if we are at node i, choose an edge uniformly at random from the set of edges leading out of the node, and move along the edge to the node at the edge. Then repeat. If there are N nodes labelled by consecutive integers then this is a Markov chain on state space $S = \{1,2,\dots,N\}$.

   Here is are a couple of examples:

[2] In these notes, I will follow the convention of using a female pronoun for even-numbered lectures, and a male pronoun for odd-numbered lectures.



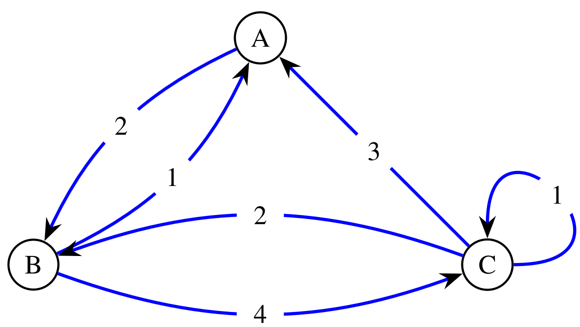


The corresponding transition matrices are:

$$P = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \qquad P = \begin{array}{c c} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix} \end{array}$$

6. *Random walk on a graph (weighted, directed).*

   Every Markov chain can be represented as a random walk on a weighted, directed graph. A weighted graph has a positive real number assigned to each edge, called the edge's "weight," and the random walker chooses an edge from the set of available edges, in proportion to each edge's weight. A directed graph assigns each edge a direction, and a walker can only move in that direction. Here is an example:



   The corresponding transition matrix is:

$$P = \begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{5} & 0 & \frac{4}{5} \\ \frac{3}{6} & \frac{2}{6} & \frac{1}{6} \end{pmatrix} \end{array}$$

   Such a directed graph forms the foundation for Google's Page Rank algorithm, which has revolutionized internet searches. Page Rank constructs a directed graph of the internet, where nodes are webpages and there is a directed edge from webpage A to webpage B if A contains a link to B. Page

Rank supposes an internet surfer clicks on links at random, and ranks pages according to the long-time average fraction of time that the surfer spends on each page.

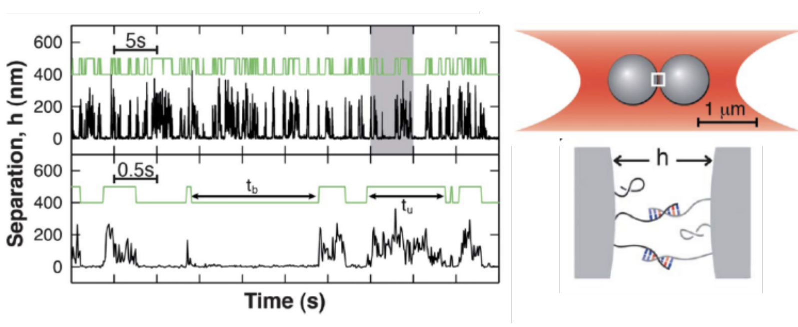
7. *Autoregressive model of order k (AR(k)).* Given constants $a_1,\dots,a_k \in \mathbb{R}$, let $Y_n = a_1Y_{n-1} + a_2Y_{n-2} + \dots + a_kY_{n-k} + W_n$, where W_n are i.i.d. random variables.

   This process is *not* a Markov chain, because it depends on the past k timesteps. However, we can form a Markov chain by defining $X_n = (Y_n, Y_{n-1},\dots,Y_{n-k+1})^T$. Then
   $$X_n = AX_{n-1} + \underline{W}_n,$$
   where $A = \begin{pmatrix} a_1 & a_2 & \cdots & & a_k \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \cdots & \cdots & \cdots & 1 & 0 \end{pmatrix}$, and $\underline{W}_n = (W_n, 0,\dots,0)^T$. Clear the vector-valued process X_n is a Markov chain.

8. *Card shuffling.* Shuffling a pack of cards can be modeled as a Markov chain. The state space S is the set of permutations of $\{1,2,\dots,52\}$. A shuffle takes one permutation $\sigma \in S$, and outputs another permutation $\sigma' \in S$ with some probability.

   A simple model is the top-to-random shuffle: at each step, take a card from the top of the deck, and put it back in at a random location. The transition matrix has elements
   $$P(X_1 = \sigma' | X_0 = \sigma) = \begin{cases} \frac{1}{52} & \text{if } \sigma' \text{ is obtained by taking an item in } \sigma \\ & \text{and moving it to the top,} \\ 0 & \text{otherwise.} \end{cases}$$
   One can also model more complicated shuffles, such as the riffle shuffle. While the state space is enormous ($|S| = 52!$) so you would not want to write down the whole transition matrix, one can still analyze these models, and study how many shuffles are needed to make the deck "close to random". For example, it takes seven riffle shuffles to get close to random, but it takes 11 or 12 to get so close that a gambler in a casino cannot exploit the deviations from randomness to win a typical game. See the online essay Austin (line) for an accessible introduction to these ideas, and Aldous and Diaconis (1986) for the mathematical proofs. (I first learned about the critical 7 riffle shuffles in the beautiful *Proofs from the Book*, by Aigner and Ziegler.)

9. *Markov chains to approximate continuous processes.* Sometimes a continuous stochastic process can be effectively modelled as a discrete one. Here is an example, from Rogers et al. (2013), where the authors measure the separation between two particles coated with velcro-like DNA strands:



The separation alternates between hovering near zero, and varying rapidly above zero. The authors found it effective to model the states of the particles as a two-state Markov chain, with states "bound" and "unbound," depending on whether the distance between the particles was small or large. They then constructed a theory for the transition rates between these states, based on their knowledge of the physics of the system.

Note that such a discrete approximation of a continuous process is rarely truely Markovian, but can nevertheless satisfy the Markov property approximately depending on the nature of the continuous process and the sets used in the discrete approximation.

10. *Language, and history of the Markov chain.* Markov chains were first invented by Andrei Markov to analyze the distribution of letters in Russian poetry (Hayes (2013)).[3] He meticulously constructed a list of the frequencies of vowel↔consonant pairs in the first 20,000 letters of Pushkin's poem-novel *Eugene Onegin*, and constructed a transition matrix from this data. His transition matrix was:

$$P = \begin{array}{r c} & \begin{array}{cc} \text{vowel} & \text{consonant} \end{array} \\ \begin{array}{r} \text{vowel} \\ \text{consonant} \end{array} & \begin{pmatrix} 0.175 & 0.825 \\ 0.526 & 0.474 \end{pmatrix} \end{array}$$

He showed that from this matrix one can calculate the average number of vowels and consonants in the whole poem. When he realized how powerful this idea was, he spent several years developing tools to analyze the properties of such random processes with memory.

Just for fun, here's an example (from Hayes (2013)) based on Markov's original work, to show how Markov chains can be used to generate realistic-looking text. In each of these excerpts, a Markov chain was constructed by considering the frequencies of strings of k letters from the English translation of the novel *Eugene Onegin* by Pushkin, for $k = 1,3,5,7$, and was then run from a randomly-generated initial condition. You can see that when $k = 3$, there are English-looking syllables, when $k = 5$ there are English-looking words, and when $k = 7$ the words themselves almost fit together coherently.

[3] Markov actually invented Markov chains to disprove a colleague's statement that the Law of Large Numbers can only hold for independent sequences of random variables, and he illustrated his new ideas on this vowel/consonant example.

First order

Theg sheso pa lyiklg ut. cout Scrpauscricre cobaives wingervet Ners, whe ilened te o wn taulie wom uld atimorerteansouroocono weveiknt hef ia ngry'sif farll t mmat and, tr iscond frnid riliofr th Gureckpeag

Third order

At oness, and no fall makestic to us, infessed Russion-bently our then a man thous always, and toops in he roguestill shoed to dispric! Is Olga's up. Italked fore declaimsel the Juan's conven night toget nothem,

Fifth order

Meanwhile with jealousy bench, and so it was his time. But she trick. Let message we visits at dared here bored my sweet, who sets no inclination, and Homer, so prose, weight, my goods and envy and kin.

Seventh order

My sorrow her breast, over the dumb torment of her veil, with our poor head is stooping. But now Aurora's crimson finger, your christening glow. Farewell. Evgeny loved one, honoured fate by calmly, not yet seeking?

2.2 Forward and backward Kolmogorov equations

Example 2.2 Consider the gambler from Example 2.1(3). Suppose she has a 0.4 chance of winning each game, and a 0.6 chance of losing, and if she goes broke, she stops playing (so she doesn't go into debt). The transition matrix describing how her money evolves is

$$P = \begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0.6 & 0 & 0.4 & 0 & \cdots \\ 0 & 0.6 & 0 & 0.4 & \cdots \\ 0 & 0 & 0.6 & 0 & \cdots \\ \vdots & & & & \ddots \end{pmatrix} \end{array}$$

In this lecture we'll answer questions such as: after playing n games, what is the probability she has gone broke? What is the average amount of money she can expect to have after n games, or the variance of this amount? If she plays for a very very long time (and has some probability of earning money each game if she goes broke), what is the probability distribution of her earnings? How long does it take her, on average, to go broke for the first time?

These questions can be answered using, respectively, the forward and backward Kolmogorov equations (this section, Section 2.2), the limiting distribution and stationary distribution (Section 2.3), and the mean first passage time (Section 2.4).

In general, suppose we have a Markov chain with transition probabilities $P(n)$, and let the probability distribution at time n be $\alpha^{(n)}$, as in

$$X_n \sim \alpha^{(n)} = (\alpha_1^{(n)}, \alpha_2^{(n)}, \ldots), \qquad \text{where} \qquad \alpha_i^{(n)} = P(X_n = i)\,. \tag{3}$$

Here $\alpha^{(n)}$ is a *row vector*, not a column vector – this is a convention for discrete probability distributions, which simplifies notation and makes it easier to generalize to continuous-state Markov processes. The initial state of the chain X_0 is also random variable with probability distribution $\alpha^{(0)}$. We'll consider how to calculate $\alpha^{(n)}$ from $\alpha^{(0)}$, and also how to calculate $\mathbb{E}f(X_n)$, the expectation of a function of X_n. We do this for a time-homogeneous Markov chain in Section 2.2.1, and extend these calculations to a time-inhomogeneous chain in Section 2.2.2.

2.2.1 Time-homogeneous Markov chain

Suppose we have a time-homogeneous Markov chain, so that $P(n) = P$, and initial condition $X_0 = i$. The probability distribution of X_1 is simply the ith row of P: $P(X_1 = j|X_0 = i) = P_{ij}$. At later times, we would like to know the n-step transition probabilities $P^{(n)}$, defined by

$$P_{ij}^{(n)} = P(X_n = j|X_0 = i). \tag{4}$$

For $n = 2$, we calculate

$$\begin{aligned}
P(X_2 = j|X_0 = i) &= \sum_k P(X_2 = j|X_1 = k, X_0 = i)P(X_1 = k|X_0 = i) && \text{Law of Total Probability} \\
&= \sum_k P(X_2 = j|X_1 = k)P(X_1 = k|X_0 = i) && \text{Markov Property} \\
&= \sum_k P_{kj}P_{ik} && \text{time-homogeneity} \\
&= (P^2)_{ij}
\end{aligned}$$

That is, the two-step transition matrix is $P^{(2)} = P^2$. This calculation illustrates a technique called *first-step analysis*, where one conditions on the first step of the Markov chain and uses the Law of Total Probability.

This calculation generalizes easily by induction:

Theorem. *Let $X_0, X_1, \ldots$ be a time-homogeneous Markov chain with transition probabilities P. The n-step transition probabilities are $P^{(n)} = P^n$, i.e.*

$$P(X_n = j|X_0 = i) = (P^n)_{ij}. \tag{5}$$

To make the notation cleaner we will write $P_{ij}^n = (P^n)_{ij}$. Note that P_{ij}^n does *not* equal $(P_{ij})^n$.

***Exercise* 2.3.** Prove this theorem, using a first-step analysis.

Example 2.3 Suppose the gambler in Example 2.2 starts with 2\$. Find the probability she has 0\$ after 3 games.

Solution. We start by calculating P^3. We could do this by hand, but the calculations below use a computer. To calculate P^3 on a computer we truncate the transition matrix at the rows and columns corresponding to 5\$, to obtain matrix $\tilde{P}$. Then P^3 and $\tilde{P}^3$ will differ only in rows 3 and higher, not the row we are interested in. The 3-step transition probabilities are

$$\tilde{P}^3 = \begin{array}{c c} & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.744 & 0 & 0.192 & 0 & 0.064 & 0 \\ 0.360 & 0.288 & 0 & 0.288 & 0 & 0.064 \\ \vdots & & & & & \vdots \end{pmatrix} \end{array}$$

Therefore after 3 games the gambler's money has probability distribution $(0.360, 0.288, 0, 0.288, 0, 0.064)$. That is, the probability she has 0\$ is 0.36, the probability she has 1\$ is 0.288, etc. $\bowtie$

From the n-step transition probabilities we can work out the probability distribution $\alpha^{(n)}$ of X_n. We use $\alpha_j^{(n)} = \sum_i P(X_n = j|X_0 = i)P(X_0 = i)$ to obtain

$$\alpha^{(n)} = \alpha^{(0)} P^n\,. \tag{6}$$

There is an evolution equation for $\alpha^{(n)}$ that will be useful in more general situations.

Forward Kolmogorov Equation. *(for a time-homogeneous, discrete-time Markov Chain)*

$$\alpha^{(n+1)} = \alpha^{(n)} P. \tag{7}$$

Proof.

$$\begin{aligned} \alpha_j^{(n+1)} &= \sum_i P(X_{n+1} = j|X_n = i)P(X_n = i) && \text{Law Of Total Probability} \\ &= \sum_i P_{ij}\alpha_i^{(n)} && \text{time-homogeneity and defn of } \alpha^{(n)}. \end{aligned}$$

$\square$

Now consider what happens if we ask for the expected value of some function of the state of the Markov chain, such as $\mathbb{E}X_n^2$, $\mathbb{E}X_n^3$, $\mathbb{E}|X_n|$, etc. Can we solve derive a similar formula and evolution equation for this quantity?

Let $f : S \to \mathbb{R}$ be a function defined on state space, and let $u^{(n)} = (u_i^{(n)})_{i\in S}$ be a vector with components

$$u_i^{(n)} = \mathbb{E}_i f(X_n) = \mathbb{E}[f(X_n)|X_0 = i]. \tag{8}$$

You should think of $u^{(n)}$ as a *column vector*; again this is a convention whose convenience will become more transparent later in the course. We can easily solve for $u^{(n)}$ using the n-step transition probabilities as

$$u_i^{(n)} = \sum_j f(j)|P(X_n = j|X_0 = i)P(X_0 = i) = \sum_j u_j^{(0)} P_{ij}^n = (P^n u^{(0)})_i\,, \tag{9}$$

since $u_j^{(0)} = f(j)$. Hence,

$$u^{(n)} = P^n u^{(0)} . \tag{10}$$

Also useful will be the equation for how $u^{(n)}$ evolves in time:

Backward Kolmogorov Equation. *(for a time-homogeneous, discrete-time Markov Chain)*

$$u^{(n+1)} = Pu^{(n)}, \qquad u_i^{(0)} = f(i) \quad \forall i \in S. \tag{11}$$

Proof. One proof uses the formula above. Another, more generalizable, uses a first-step analysis:

$$\begin{aligned}
u_i^{(n+1)} &= \sum_j f(j)P(X_{n+1}{=}j|X_0 = i) && \text{definition of expectation} \\
&= \sum_j \sum_k f(j)P(X_{n+1}{=}j|X_1{=}k, X_0{=}i)P(X_1{=}k|X_0{=}i) && \text{LoTP} \\
&= \sum_j \sum_k f(j)P(X_{n+1}{=}j|X_1{=}k)P(X_1{=}k|X_0{=}i) && \text{Markov property} \\
&= \sum_j \sum_k f(j)P^n_{kj}P_{ik} && \text{time-homogeneity} \\
&= \sum_k \sum_j f(j)P^n_{kj}P_{ik} && \text{switch order of summation} \\
&= \sum_k u^{(n)}(k)P_{ik} && \text{definition of } u^{(n)} \\
&= (Pu^{(n)})_i
\end{aligned}$$

We can switch the order of summation above, provided $\mathbb{E}_i|f(X_n)| < M < \infty$ for each i and each n, since then the double sum is absolutely convergent. □

Example 2.4 Returning to the gambler in Example 2.2, calculate the gambler's expected earnings after 3 games, assuming she starts with 2$.

Solution. We have that $u^{(0)} = (0, 1, 2, 3, \ldots)^T$, and calculate

$$u^{(3)} = P^3 u^{(0)} = \begin{pmatrix} 0 \\ 0.64 \\ 1.472 \\ \vdots \end{pmatrix},$$

so the gambler's expected earnings are $u_2^{(3)} = 1.472$. ⋈

Remark. What is so backward about the backward equation? It gets its name from the fact that it can be used to describe how conditional expectations propagate backwards in time. To see this, suppose that instead of (8), which propagates the expectation of a function into the future, given a fixed starting point, we choose a fixed time T and compute the expectation at that time, given an earlier, varying starting time. That is, for each $n \leq T$, define a column vector $v^{(n)}$ with components

$$v_i^{(n)} = \mathbb{E}[f(X_T)|X_n{=}i] . \tag{12}$$

Such a quantity is studied frequently in financial applications, where X_n may represent the price of a stock at time n, f the value of an option to sell, T the time at which you decide (in advance) to sell a stock, and quantities of the form (12) represent the expected payout, conditional on being in state i at time n. One is interested in solving for $v^{(0)}$ and in finding the element which maximizes the expected payoff. One can show that the vector $v^{(n)}$ evolves according to

$$v^{(n)} = P(n)v^{(n+1)}, \qquad v_i^{(T)} = f(i) \quad \forall i \in S. \tag{13}$$

That is, you find $v^{(n)}$ by evolving it *backwards* in time – you are given a final condition at time T, and you can solve for $v^{(n)}$ at all earlier times $n \leq T$. Note that the equation above holds even when the chain is time-inhomogeneous. This same statement is not true for (11) – for a general Markov chain there is no backward Kolmogorov equation that can be solved forward in time.

***Exercise* 2.4.** Derive (13) for a time-inhomogeneous Markov chain, using a first-step analysis. (Another derivation is in the next section.)

2.2.2 Time-Inhomogeneous Markov chains

The forward and backward Kolmogorov equations for a time-inhomogeneous Markov chain are derived from the Chapman-Kolmogorov equations, a relationship satisfied by all Markov chains and Markov processes more generally[4]:

Chapman-Kolmogorov Equation.

$$P(X_n = j|X_0 = i) = \sum_k P(X_n = j|X_m = k)P(X_m = k|X_0 = i). \tag{14}$$

Proof.

$$\begin{aligned}
P(X_n = j|X_0 = i) &= \sum_k P(X_n = j, X_m = k|X_0 = i) && \text{Law of Total Probability} \\
&= \sum_k P(X_n = j|X_m = k, X_0 = i)P(X_m = k|X_0 = i) && \because P(A\cap B|C) = P(A|B\cap C)P(B|C) \\
&= \sum_k P(X_n = j|X_m = k)P(X_m = k|X_0 = i). && \text{Markov property (see Ex.(2.2))}
\end{aligned}$$

□

To use this relationship, define a function $P(j,t|i,s)$ to be the transition probability to be in state j at time t, given the system started in state i at time s, i.e.

$$P(j,t|i,s) = P(X_t = j|X_s = i). \tag{15}$$

[4] Note that while all Markov processes satisfy a form of Chapman-Kolmogorov equations, from which many other equations can be derived, not all processes which satisfy Chapman-Kolmogorov equations, are Markov processes. See Grimmett and Stirzaker (2001), p.218 Example 14 for a counterexample.

The *forward Kolmogorov equation* comes from considering how $P(j,t|i,s)$ evolves in t, forward in time. The Chapman-Kolmogorov equations imply that

$$P(j,t+1|i,s) = \sum_k P(k,t|i,s)P(j,t+1|k,t). \tag{16}$$

Then, since $\alpha^{(t)}$ has components $\alpha_j^{(t)} = \sum_i P(j,t|i,0)\alpha_i^{(0)}$, we multiply (16) by $\alpha^{(0)}$ on the left (contracting it with index i) and let $s=0$ to obtain

Forward Kolmogorov Equation. *(general Markov chain)*

$$\alpha_j^{(t+1)} = \sum_k \alpha_k^{(t)} P(j,t+1|k,t) \qquad \Leftrightarrow \qquad \alpha^{(t+1)} = \alpha^{(t)}P(t)\,, \tag{17}$$

where $P(t)$ is the 1-step transition matrix at time t, i.e. $(P(t))_{kj} = P(j,t+1|k,t)$.

The *backward Kolmogorov equation* comes from considering how $P(j,t|i,s)$ evolves in s, backward in time. The Chapman-Kolmogorov equations imply that

$$P(j,t|i,s) = \sum_k P(j,t|k,s+1)P(k,s+1|i,s). \tag{18}$$

Now, let $f: S \to \mathbb{R}$, let $T > 0$ be a fixed time, and let $u_i^{(s)} = \mathbb{E}[f(X_T)|X_s = i]$ (recall (8),(12)). Notice that $u_i^{(s)} = \sum_k f(k)P(k,T|i,s)$, so multiplying (18) by the column vector $(f(1), f(2), \dots)^T$ on the right (contracting it with index j) and evaluating at $t = T$ gives

Backward Kolmogorov Equation. *(general Markov chain)*

$$u_i^{(s)} = \sum_k P(k,s+1|i,s)u_k^{(s+1)} \qquad \Leftrightarrow \qquad u^{(s)} = P(s)u^{(s+1)}\,, \tag{19}$$

where $P(s)$ is the 1-step transition matrix at time s as above.

Remark. Equations (16), (18) for the evolution of the transition probabilities forward and backward in time, are considered in some references to be the forward and backward Kolmogorov equations.

***Exercise* 2.5.** Show that the product $\alpha^{(t)}u^{(t)}$ is constant, and equal to $\mathbb{E}f(X_T)$.

Solution. We have

$$\alpha^{(t)}e^{(t)} = \sum_k \alpha_k^{(t)}u_k^{(t)} = \sum_k P(X(t)=k)\mathbb{E}(f(X_T)|X_t = k) = \mathbb{E}f(X_T).$$

$\bowtie$

2.3 Long-time behaviour and stationary distribution

Suppose we take a Markov chain and let it run for a long time. What happens? Clearly the random variable X_n does not generally converge to anything, because it is continually jumping around, but its probability distribution might converge. Let's look at some examples.

Example 2.5 Consider the two-state weather model in Example 2.1(1). Suppose it is raining today. What is the probability distribution for the weather in the future? Let's calculate $\alpha^{(n)}$ from (6), using $\alpha^{(0)} = (0,1)$:

n	P(sun)	P(rain)
0	0	1
1	0.4000	0.6000
2	0.5600	0.4400
3	0.6240	0.3760
4	0.6496	0.3504
5	0.6598	0.3402
6	0.6639	0.3361
7	0.6656	0.3344
8	0.6662	0.3338
9	0.6665	0.3335
10	0.6666	0.3334
11	0.6666	0.3334
12	0.6667	0.3333
13	0.6667	0.3333
14	0.6667	0.3333

The probability distribution seems to converge. After 12 days, the distribution doesn't change, to 4 digits. You can check that the distribution it converges to does not depend on the initial condition. For example, if we start with a sunny day, $\alpha^{(0)} = (1,0)$, then $\alpha^{(10)} = (0.6666, 0.3334)$, $\alpha^{(11)} = (0.6666, 0.3334)$, etc.

***Exercise* 2.6.** Work out the n-step transition probabilities for any initial condition analytically, and show they converge to $(2/3, 1/3)$. (Hint: diagonalize the transition matrix.)

Does the probability always converge? Not necessarily. The following two examples illustrate situations where it doesn't converge.

Example 2.6 Consider a Markov chain on state space $\{0,1\}$ with transition matrix

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

Suppose the random walker starts at state 0. Its distribution at time n is:

n	P(0)	P(1)
0	1	0
1	0	1
2	1	0
3	0	1
4	1	0
⋮	⋮	⋮

Clearly the distribution doesn't converge. Yet, if we start with initial distribution $\alpha^{(0)} = (0.5, 0.5)$, we obtain

n	P(0)	P(1)
0	0.5	0.5
1	0.5	0.5
2	0.5	0.5
$\vdots$	$\vdots$	$\vdots$

The distribution never changes!

Example 2.7 A *simple symmetric random walk* is a random walk on the integers with probability $p = 1/2$ to go left or right (see Example 2.1(3)). From the Binomial distribution, one can show that the transition probabilities at time n are

$$P^n_{ij} = \begin{cases} \binom{n}{k}\left(\frac{1}{2}\right)^n & j = i+2k-n, \quad k = 0,1,\ldots,n \\ 0 & \text{o.w.} \end{cases}$$

As $n \to \infty$, $P^n_{ij} \to 0$. So P^n converges – but not to a probability distribution, rather to the function that is identically zero. Here, the problem is that mass eventually escapes to infinity.

2.3.1 Limiting and stationary distributions

When does the distribution of a Markov chain converge? And when it does, what does it converge to, and how can we find the limiting distribution? The answers are given by understanding the *limiting distribution* and *stationary distribution*. We consider only time-homogeneous Markov chains.

Definition. Consider a time-homogeneous Markov chain with transition matrix P. A row vector λ is a *limiting distribution* if $\lambda_i \geq 0$, $\sum_j \lambda_j = 1$ (so that λ is a probability distribution), and if, for every i,

$$\lim_{n\to\infty} (P^n)_{ij} = \lambda_j \qquad \forall j \in S.$$

In other words,

$$P^n \to \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \ldots \\ \lambda_1 & \lambda_2 & \lambda_3 & \ldots \\ \lambda_1 & \lambda_2 & \lambda_3 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \qquad \text{as } n \to \infty.$$

As we saw in examples 2.6, 2.7, a limiting distribution doesn't have to exist. If it does exist, it is unique, by definition.

***Exercise* 2.7.** Show that, if $|S| < \infty$, then λ is a limiting distribution if and only if $\lim_{n\to\infty} \alpha^{(0)} P^n = \lambda$ for any initial probability distribution $\alpha^{(0)}$.

If $|S| = \infty$, then we could have that $\lim_{n\to\infty} P^n$ exists, but is not a probability distribution. For example, for the simple symmetric random walk in Example 2.7, $\lim_{n\to\infty} P^n = (\ldots,0,0,0,\ldots)$, however the zero vector is not a probability distribution.[5]

[5] The problem is we can't interchange a limit and sum in general: although $\sum_j P^n_{ij} = 1$, hence the limit as $n \to \infty$ is also 1, we have $\lim_{n\to\infty} \sum_j P^n_{ij} \neq \sum_j \lim_{n\to\infty} P^n_{ij}$.

What happens if we start a chain in a limiting distribution λ for a Markov chain with a finite state space? Let's calculate the distribution $\alpha^{(1)}$ at the next step of the chain, starting from $\alpha^{(0)} = \lambda$. We have

$$\alpha_j^{(1)} = (\lambda P)_j = \sum_k (\lim_{n\to\infty} P_{ik}^n) P_{kj} = \lim_{n\to\infty} \sum_k P_{ik}^n P_{kj} = \lim_{n\to\infty} P_{ij}^{n+1} = \lambda_j.$$

We can interchange the sum and limit, because it's a finite sum. Therefore if we start the chain in the limiting distribution, its distribution remains there forever. This motivates the following definition:

Definition. Given a Markov chain with transition matrix P, a *stationary distribution* is a probability distribution π which satisfies

$$\pi = \pi P \qquad \Longleftrightarrow \qquad \pi_j = \sum_i \pi_i P_{ij} \quad \forall\ j. \tag{20}$$

A stationary distribution may also be called an *invariant measure*, *invariant distribution*, *steady-state probability*, *equilibrium probability* or *equilibrium distribution* (the latter two are from physics).

The stationary distribution is stationary in the following sense: if we start the chain in the stationary distribution, $X_0 \sim \pi$, the distribution does not change: $X_1 \sim \pi$, $X_2 \sim \pi$, etc.

In applications we want to know the limiting distribution. We saw above that when $|S| < \infty$, a limiting distribution is a stationary distribution, but the converse is not always true. Indeed, in Example 2.6, you can calculate that a stationary distribution is $\pi = (0.5, 0.5)$, but this is not a limiting distribution.

Yes, the stationary distribution is easier to calculate than a limiting distribution: we can find it by solving a linear system of equations. In fact, you can see that π is a left eigenvector of P corresponding to eigenvalue 1. Therefore we will restrict our focus to the stationary distribution. Some questions we might ask about π include:

(i) Does it exist?
(ii) Is it unique?
(iii) When is it a limiting distribution, i.e. when does an arbitrary distribution converge to it?

This is the subject of a rich body of work on the limiting behaviour of Markov chains. In general there are two broad approaches to answering such questions. One approach is probabilistic, using tools such as recurrence times and coupling properties. Another approach uses linear algebra to study the transition matrix, for Markov chains with a finite state space. We will survey some results using this second approach, which should be more familiar to students from a range of applied backgrounds.

***Exercise* 2.8.** (a) Solve for the stationary distribution for the gambler in the example at the beginning of Section 2.2. (b) Solve again, but this time assume that if the gambler loses all her money, there is some small probability 0.1 per game that she finds 1$ on the ground and can play again (if she doesn't find money on the ground, she doesn't play that game).

***Exercise* 2.9.** Calculate the eigenvalues and the eigenvectors of the transition matrices in Examples 2.5, 2.6.

2.3.2 Stationary distributions and linear algebra

Let's focus on Markov chains with a finite state space, $|S| = N < \infty$, and ask what linear algebra tells us about the stationary and limiting distributions associated with the transition matrix P.

We know that P has an eigenvalue $\lambda = 1$, since the rows of P sum to 1 so we have

$$P\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix} = \begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix},$$

and therefore $(1,1,\ldots,1)^T$ is a right eigenvector. To ensure that the corresponding left eigenvector π is a stationary distribution, we need to know that its entries are all nonnegative.

Let's put the issue of the nonnegativity of π on hold for a moment, and ask if P has a limiting distribution. We do this by calculating P^n. Suppose that P can be diagonalized, as

$$P = B^{-1}\Lambda B, \qquad \text{where} \quad \Lambda = \begin{pmatrix}\lambda_1 & 0 & 0 & \cdots & 0\\ 0 & \lambda_2 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & 0\\ 0 & \cdots & 0 & 0 & \lambda_N\end{pmatrix}. \tag{21}$$

The rows of B are left eigenvectors of P, the columns of B^{-1} are right eigenvectors, and the eigenvalues are $\lambda_1, \lambda_2, \ldots, \lambda_N$; let's order them so that $\lambda_1 = 1$. Therefore

$$P^n = B^{-1}\Lambda^n B, \qquad \text{where} \quad \Lambda = \begin{pmatrix}\lambda_1^n & 0 & 0 & \cdots & 0\\ 0 & \lambda_2^n & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & 0\\ 0 & \cdots & 0 & 0 & \lambda_N^n\end{pmatrix}. \tag{22}$$

What happens as $n \to \infty$? For the first eigenvalue we have $\lambda_1^n = 1$. Any eigenvalue such that $|\lambda_i| < 1$ will converge to zero, $\lambda_i^n \to 0$ as $n \to \infty$. Therefore, if $|\lambda_i| < 1$ for $i \geq 2$, we have, using our knowledge of the right and left eigenvectors corresponding to λ_1,

$$\lim_{n\to\infty} P^n = B^{-1}\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & 0\end{pmatrix} B \quad = \quad \begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}\begin{pmatrix}\pi_1 & \pi_2 & \cdots & \pi_N\end{pmatrix} = \begin{pmatrix}\pi_1 & \pi_2 & \cdots & \pi_N\\ \pi_1 & \pi_2 & \cdots & \pi_N\\ \vdots & \vdots & \vdots & \vdots\\ \pi_1 & \pi_2 & \cdots & \pi_N\end{pmatrix}.$$

We argue that π must be a probability distribution: Since P has nonnegative entries, so does P^n, so π must be nonnegative. We must also have that $\sum_j \pi_j = 1$, since $\sum_j P_{ij}^n = 1$ so applying the limit and interchanging sum and limit (for a finite sum) gives $\sum_j \lim_{n\to\infty} P_{ij}^n = 1$.

We have just shown that if P is diagonalizable, such that all eigenvalues except λ_1 have $|\lambda_i| < 1$, then the left eigenvector π corresponding to λ_1 is a limiting distribution, and therefore it is also a stationary distribution. (If P is not diagonalizable, then we can do a similar calculation using the Jordan canonical form of the matrix, and obtain the same conclusion given similar conditions on the generalized eigenvalues.) It remains to ask: when do all eigenvalues (or generalized eigenvalues) satisfying $|\lambda_i| < 1$ for $i \neq 1$?

We can show quite easily that all eigenvalues must have norm less than or equal to 1.

Lemma. *The spectral radius of a stochastic matrix P is 1, i.e. $\rho(P) = \max_\lambda |\lambda| = 1$, where the maximum is over all eigenvalues of P.*

Proof. Let η be a left eigenvector with eigenvalue λ. Then $\lambda \eta_i = \sum_{j=1}^N \eta_j P_{ji}$,

$$|\lambda| \sum_{i=1}^N |\eta_i| = \sum_{i=1}^N \Big|\sum_{j=1}^N \eta_i P_{ji}\Big| \leq \sum_{i,j=1}^N |\eta_j| P_{ji} = \sum_{j=1}^N |\eta_j|.$$

Therefore $|\lambda| \leq 1$. $\square$

Whew. This is good news – it shows that no eigenvalue of P has complex norm greater than 1 – but it still doesn't rule out the possibility that there are other eigenvalues with complex norm equal to 1. To handle this possibility we turn to a powerful theorem from linear algebra, the Perron-Frobenius Theorem.

Definition. A matrix A is *positive* if it has all positive entries: $A_{ij} > 0$ for all i, j.

Remark. A is positive is *not* the same as A being positive-definite!

Theorem. (Perron-Frobenius Theorem.) *Let M be a positive $k \times k$ matrix, with $k < \infty$. Then the following statements hold:*

(i) There is a positive real number λ_1 which is an eigenvalue of M. All other eigenvalues λ of M satisfy $|\lambda| < \lambda_1$.

(ii) The eigenspace of eigenvectors associated with λ_1 is one-dimensional.

(iii) There exists a positive right eigenvector v and a positive left eigenvector w associated with λ_1. Furthermore,

$$\lim_{n\to\infty} \frac{1}{\lambda_1^n} M^n = vw^T,$$

where the eigenvectors are normalized so that $w^T v = 1$.

(iv) M has no other eigenvectors with nonnegative entries.

For a proof, see an advanced linear algebra textbook, such as Lax (1997), Chapter 16, or Horn and Johnson (2012), Theorem 8.2.8. There is also a brief description of the proof in Strang (1988), Section 5.3.

We can combine the Perron-Frobenius Theorem with the Lemma above, to obtain a statement about Markov chains.

Theorem. *Let $X_0, X_1, X_2, \ldots$ be a time-homogeneous Markov chain with a finite state space. Suppose the transition matrix P is positive. Then there exists a unique stationary probability distribution π such that π is positive. Furthermore, π is also a limiting distribution.*

Proof. From the Lemma above, the spectral radius of P is 1. The Perron-Frobenius theorem tells us there is a one-dimensional eigenspace associated with the eigenvalue $\lambda_1 = 1$, and the corresponding left eigenvector π is positive. Therefore, π is the unique stationary distribution.All other eigenvalues have complex norm less than 1. Furthermore, since the corresponding right eigenvector is $v = (1, \ldots, 1)^T$, we obtain that $\lim_{n\to\infty} P_{ij}^n = \pi_j$, which is the definition of a limiting distribution. $\square$

One can extend these results to a Markov chain such that P^s is positive, for some integer $s > 0$. This means that there is a time s such that, no matter where you start, there is a non-zero probability of being at any other state.

Theorem. *Let $X_0, X_1, X_2, \ldots$ be a time-homogeneous Markov chain with a finite state space and transition matrix P. Suppose there exists some integer $s > 0$ such that P^s is positive. Then there exists a unique stationary probability distribution π such that π is positive. Furthermore, π is also a limiting distribution.*

Proof. This theorem results from a combinations of theorems and definitions from Horn and Johnson (2012), outlined here. Since P and P^s are positive, P is irreducible and has only one nonzero eigenvalue of maximum complex norm (Definitions 6.2.22, 8.5.0, Theorem 8.5.2). Theorem 8.4.4 (another version of the Perron-Frobenius Theorem) shows that since P is irreducible and nonnegative, there is a one-dimensional eigenspace associated with the eigenvalue $\lambda_1 = 1$, and the corresponding left eigenvector π is positive. Theorem 8.5.1 shows that $\lim_{n\to\infty} P^n_{ij} = \pi_j$, so π is a limiting distribution. (See also Koralov and Sinai (2010) p.72, for a more direct proof that analyzes the evolution of probability and shows it converges exponentially quickly to the stationary distribution.) $\square$

This theorem can be weakened slightly by allowing for Markov chains with some kind of periodicity. We need to consider a chain which can move between any two states (i, j), but not necessarily at a time s that is the same for all pairs.

***Exercise* 2.10.** Calculate the stationary distribution for the gambler in Example 2.2, assuming that if she goes broke, there is a 10% chance per game that she finds 1$ and can play again.

Definition. A stochastic matrix is *irreducible* if, for every pair (i, j) there exists an $s > 0$ such that $(P^s)_{ij} > 0$.

There are limit theorems for irreducible chains, with slightly weaker conditions. Irreducible chains also have a unique stationary distribution (Horn and Johnson (2012), Theorem 8.4.4). However, it is not true that an arbitrary distribution converges to it; rather, we have that the average distribution converges (Horn and Johnson (2012), Theorem 8.6.1):

$$\alpha^{(0)}\bar{P}^{(n)} \to \pi \quad \text{as } n \to \infty, \qquad \text{where } \bar{P}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} P^k .$$

We need to form the average, because there may be a built-in periodicity, as in Example 2.6, where $P^{2n} = I$, and $P^{2n+1} = P$, so $\alpha^{(n)}$ oscillates between two distributions, instead of converging to a fixed limit.

***Exercise* 2.11.** Show that the transition matrix in Example 2.6 is irreducible, however there is no s such that P^s is positive.

Remark. So far we have analyzed only Markov chains with a finite state space. There are also results for Markov chains with infinite state spaces, which are often proved using probabilistic approaches. To state the results, one is interested in the mean recurrence time μ_i of each state i, defined by

$$\mu_i = \mathbb{E}(T_i | X_0 = i) \tag{23}$$

where $T_i = \min\{n \geq 1 : X_n = i\}$ is the first return time to i. We could have that μ_i is finite or infinite. If a Markov chain is irreducible, and $\mu_i < \infty$ for all states i, then there is a unique stationary distribution π given by $\pi_i = \mu_i^{-1}$. (Grimmett and Stirzaker (2001), Section 6.4 Theorem 3, or Norris (1997), Theorem 1.7.7.) Furthermore, if the chain is irreducible, has stationary distribution π, and is *aperiodic* (there exists an s such that $P_{ii}^s > 0$ for some state i), then π is a limiting distribution (Grimmett and Stirzaker (2001), Section 6.4 Theorem 17, or Norris (1997), Theorem 1.8.3).

2.4 Mean first passage time

Sometimes we want to ask how long it takes a Markov chain to do something: how long until the weather turns sunny again, how long does it take a gambler to go broke, etc? Answering these questions requires asking about the probability distributions of random times, which depend on the realization of a Markov chain. We can't handle question about any kind of random time, but we can handle questions about the time it takes to hit a given subset of the state space, using tools from linear algebra.

Definition. The *first-passage time* of a set $A \subset S$ is defined by

$$T_A = \min\{n \geq 0 : X_n \in A\}.$$

A common quantity of interest is the average time it takes to hit set A.

Definition. The *mean first passage time* (mfpt) to set A starting at state i is

$$\tau_i = \mathbb{E}(T_A | X_0 = i). \tag{24}$$

Let's compute τ_i using a first-step analysis assuming (for now) that $P(T_A < \infty | X_0 = i) = 1$ for all $i \in S$. We consider only time-homogeneous Markov chains.

For $i \in A$, we have $T_A = 0$ so $\tau_i = 0$. Consider $i \notin A$. Then

$$\begin{aligned}
\tau_i &= \sum_{t=1}^{\infty} t P(T_A{=}t | X_0{=}i) \\
&= \sum_{t=1}^{\infty} \sum_{j \in S} t P(T_A{=}t | X_1{=}j) P(X_1{=}j | X_0{=}i) && \text{LOTP \& Markov property}
\end{aligned}$$

Because the chain is time-homogeneous, we expect that $P(T_A{=}t | X_1{=}j) = P(T_A{=}t-1 | X_0{=}j)$. To show this explicitly, write

$$\begin{aligned}
P(T_A{=}t | X_1{=}j) &= P(X_2 \in A^c, \ldots, X_{t-1} \in A^c, X_t \in A | X_1 = j) && \text{by definition} \\
&= P(X_1 \in A^c, \ldots, X_{t-2} \in A^c, X_{t-1} \in C | X_0 = j) && \text{by time-homogeneity} \\
&= P(T_A{=}t-1 | X_0{=}j).
\end{aligned}$$

Therefore, substituting into the above and changing the index $t \to t+1$, we have

$$\begin{aligned}
\tau_i &= \sum_{t=0}^{\infty} \sum_{j \in S} (t+1) P(T_A{=}t | X_0{=}j) P_{ij} \\
&= \sum_{j \in S} \sum_{t=0}^{\infty} t P(T_A{=}t | X_0{=}j) P_{ij} + \sum_{j \in S} \sum_{t=0}^{\infty} P(T_A{=}t | X_0{=}j) P_{ij} \\
&= \sum_{j \in S} \tau_j P_{ij} \quad + \ 1.
\end{aligned}$$

The second term is 1, because $\sum_{t=0}^{\infty} P(T_A{=}t|X_0{=}j) = 1$, since this sum is the probability that T_A takes any value and we assumed $P(T_A < \infty) = 1$, and then we sum $\sum_{j\in S} P_{ij} = 1$. We can interchange the order of summation in the second step because all the terms we are adding up are nonnegative. We obtain:

Theorem. *Let $\tau = (\tau_i)_{i\in S}$ be a vector of mean first passage times from each state $i \in S$. Then τ solves the system of linear equations:*

$$\begin{cases} \tau_i = 0 & i \in A \\ \tau_i = 1 + \sum_j P_{ij}\tau_j & i \notin A. \end{cases} \tag{25}$$

In fact, τ is the minimal nonnegative solution to these equations, meaning that any other nonnegative solution y to (28) *has $y_i \geq \tau_i$ for all i.*

Remark. The condition about τ being the minimal nonnegative solution is required to handle Markov chains with an infinite state space. See Norris (1997), Theorem 1.3.5.

Remark. The calculations above actually don't depend on the assumption $P(T_A < \infty|X_0 = i) = 1$ – one can carry through the same calculations, starting from the identity $\tau_i = \sum_{t=1}^{\infty} tP(T_A{=}t|X_0{=}i) + \infty P(T_a = \infty|X_0{=}i)$, and obtain the same result (28) even without this assumption. If $\tau_i = \infty$ (which can happen even when $P(T_A < \infty|X_0{=}i){=}1$), then there won't exist a nonnegative solution to (28).

Equation (28) gives a way to find the mean first passage time by solving a linear system of equations. We can write (28) as

$$(P' - I)\tau' = -1\,, \tag{26}$$

where P' is P with the rows and columns corresponding to elements in A removed, and τ' is τ with the elements in A removed. This form of the equations will make it easier to make the connection to continuous-time Markov chains and processes later in the course, and is convenient to use on a computer.

Note that we can't find the mfpt from state i in isolation; we have to solve for the mfpt from all states i simultaneously. For systems that are not too large, this means we can use built-in linear algebra solvers to calculate mfpts. If the problem has some extra structure, we can sometimes even find analytical solutions.

Example 2.8 Consider the gambler from the beginning of Section 2.2. Let $A = \{0\}$, the event that she has 0\$. Show that the average time it takes her to go broke, starting from k\$, is $\tau_k = \frac{k}{1-2p}$.

Solution. Solved by verifying that (28) is satisfied. Note that if the gambler were to stop when she wins some amount M\$, then the mfpt to the set $B = \{0, M\}$ can be found by solving the inhomogeneous recurrence relation in (28) to be[6]

$$\tau_k^M = \frac{k}{1-2p} - \frac{M}{1-2p} \cdot \frac{\left(\frac{1-p}{p}\right)^k - 1}{\left(\frac{1-p}{p}\right)^M - 1} \quad (p \neq 1/2), \qquad \tau_k^M = Mk - k^2 \quad (p = 1/2).$$

As $M \to \infty$, $\tau_k^M \to \tau_k$. ⋈

***Exercise* 2.12.** Suppose you perform a random walk on the integers where at each step you jump left or right with equal probability, and let X_n be your position at time n. Calculate the mean first passage time τ_0 to leave the interval $(-6, 6)$, starting at $X_0 = 0$.

[6] See `https://web.mit.edu/neboat/Public/6.042/randomwalks.pdf`

We can calculate the probability that T_A is finite, using a similar calculation.

Definition. The *hitting probability* of set A starting at state i is

$$h_i = P(T_A < \infty | X_0 = i). \tag{27}$$

Theorem. *The vector of hitting probabilities $h = (h_i)_{i \in S}$ is the minimal nonnegative solution to the system of linear equations*

$$\begin{cases} h_i = 1 & i \in A \\ h_i = \sum_j P_{ij} h_j & i \notin A. \end{cases} \tag{28}$$

***Exercise* 2.13.** Prove the above theorem, using a first-step analysis. (See Norris (1997), Theorem 1.3.2.)

Example 2.9 Consider the gambler from the beginning of Section 2.2. Calculate the probability she eventually goes broke, starting from k\$.

Solution. (See Norris (1997), Example 1.3.3.) We must solve the recurrence relation

$$h_0 = 1, \qquad h_i = p h_{i+1} + (1-p) h_{i-1}, \quad i = 1, 2, \ldots.$$

If $p \neq 1/2$ the general solution is

$$h_i = C + D \left(\frac{1-p}{p} \right)^i$$

for some constants C, D. For $p < 1/2$, the condition $h_i \leq 1$ requires $D = 0$ so $h_i = 1$ for all i – the gambler will always end up broke.

If $p = 1/2$, the recurrence relation has a general solution

$$h_i = C + Di,$$

and again $D = 0$ so $h_i = 1$ for all i. Therefore, even if the gambler is playing a fair game, no matter how much money she starts with, she is certain to end up broke. This phenomenon is called the *Gambler's Ruin*.

$\bowtie$

References

Aldous, D. and Diaconis, P. (1986). Shuffling cards and stopping times. *American Mathematical Monthly*, 93:333–348.

Austin, D. (online). How many times do I have to shuffle this deck? `http://www.ams.org/samplings/feature-column/fcarc-shuffle`.

Diaconis, P., Holmes, S., and Montgomery, R. (2007). Dynamical Bias in the Coin Toss. *SIAM Rev.*, 49(2):211–235.

Dobrow, R. P. (2016). *Introduction to Stochastic Processes with R*. Wiley.

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford University Press.

Hayes, B. (2013). First links in the Markov chain. *American Scientist*, 101.

Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis*. Cambridge university press.

Koralov, L. B. and Sinai, Y. G. (2010). *Theory of Probability and Random Processes*. Springer.

Lax, P. (1997). *Linear Algebra*. John Wiley & Sons.

Norris, J. R. (1997). *Markov Chains*. Cambridge University Press.

Rogers, W. B., Sinno, T., and Crocker, J. C. (2013). Kinetics and non-exponential binding of DNA-coated colloids. *Soft Matter*, 9(28):6412–6417.

Strang, G. (1988). *Linear Algebra and its Applications*. Brooks/Cole, 3rd edition.