Martingales, Markov chains, and Brownian motion

Mark Huber

2025-12-13

Contents

Preface

This book covers a rigorous one semester course in stochastic processes for students who have already taken a one semester course in probability that uses Calculus and Linear Algebra. It is intended to prepare students for advanced work in areas such as randomized algorithms, stochastic operations research, and finance where these ideas are essential to building effective models.

The introductory chapters begin with a fast review of probability, random variables, and expected value. The first stochastic process encountered is a simple random walk. Here the notions of expected value are reviewed, and conditional expectation introduced. First order logic is also covered in these chapters. The section ends with a rigorous discussion of random variables using measurable spaces, σ-algebras and measurable functions.

The next part builds on these ideas, formally defining expected value and proving useful theorems about when limits and expectation can be swapped.

After that, martingales are introduced. Formal definitions of this type of stochastic process, together with stopping times, are given. The two main theorems of this part are the Martingale Convergence Theorem and the Optional Sampling Theorem.

Markov chains follow. Again a formal definition is given, along with transition matrices, update functions, recurrence and transience, stationary measures, and stationary distributions. Harris chains are also defined. The main theorem is the Ergodic Theorems for finite state Markov chains, countable Markov chains, and continuous state space Markov chains. This theorem is proved rigorously using a coupling argument. The special case of branching processes is discussed together with generating functions.

Next stochastic processes with a continuous index are discussed. Brownian motion is defined and it is shown how to simulate from this process at a finite set of times. Continuous time Markov chains are built as well as Poisson point processes. Stationary and limiting distributions are extended to this setting.

This part also shows how differential equations can be modeled probabilistically. Uniform and square integrability are defined, and Wald's Equation is shown. The notion of stochastic integration is developed, together with Ito's Formula.

The final part shows how to construct Markov chains with a particular stationary distribution, using ideas such as reversibility, Metropolis-Hastings, and birth-death chains.

Chapter 1

Types of stochastic processes

Question of the Day

What is a stochastic process?

Summary

- **Random variables** are the primary mathematical object studied in probability, and represent variables about whose value some information is known.
- A collection of random variables that are indexed is called **stochastic process**.
- For numerical random variables, the **expected value** (also known as the **expectation**, **mean**, or **average** is a finite number.

In this text, you will learn what stochastic processes are, what are the most common types of stochastic processes, and what are the most common mathematical tools needed to understand and apply them. The basic object of study in probability is the *random variable*, so start there.

1.1 Random variables

A *random variable* is a variable where you only have partial information about the true value. For instance, the total number of people alive in the continental United States (so all states except Hawaii and Alaska) at this moment is a random variable. It is some specific integer value, but you do not have total information about it. This number could be represented by a random variable N_1.

Often (but not always) uppercase Roman letters are used to name random variables.

Another random variable would be the population of the United States including Alaska and Hawaii. Call this random variable N_2.

The lack of total information about these numbers can come about because it is simply too difficult to keep track (such as the continental US population example) or because the event has not happened yet! The winner of the next World Series is a random variable, for instance.

It is important to understand that just because we do not exactly know the value, does not mean that we are completely in the dark. For the US population example, one thing that I know with certainty is that

$$0 \leq N_1 \leq N_2.$$

But beyond that, it is possible to assign probabilities to the nonnegative integers that represent the state of knowledge about these values.

For instance, if two fair six sided dice are rolled, then the probability that the sum is 7 is $1/6$ whereas the probability the sum will be 2 is only $1/36$. Probability is the method used to codify partial information.

1.2 Stochastic Processes

The word *stochastic* just means random. It comes from the Greek word meaning to aim at a mark and was an archery term. (What could be more random than where an arrow strikes!)

For instance, a stock price at a particular time, the location of the last lightning strike in a field, the wind at a particular point, all of these are random.

Often there is not just one random variable, but many are being used at once. The stock prices over a year, the location of all lightning strikes in a forest, the wind over a county, all of these are stochastic processes. Random variables that are given an index are *stochastic processes.*

As another example, suppose that there are 27 wind turbines in a wind farm. On a particular day, turbine i generates X_i amount of power. Because the power output is unknown at the beginning of the day, each X_i is a random variable. The random variables $(X_1, X_2, \ldots, X_{27})$ that are indexed by $\{1, 2, \ldots, 27\}$ form a stochastic process.

1.3 The simplest process

The simplest nontrivial stochastic process consists of a sequence of random variables

$$X_1, X_2, X_3, \ldots$$

where each of the X_i have the same distribution, and any subset of the sequence are independent. Such a stochastic process forms an *independent, identically distributed* (iid) process. (Note that some authors put periods in iid so it becomes i.i.d., but here we follow the lead of laser, radar, and scuba and leave them out.)

For instance, suppose that for $B_1, B_2, \ldots$, $\mathbb{P}(B_i = 1) = 0.7$ and $\mathbb{P}(B_i = 0) = 0.3$. This type of random variable is said to have a *Bernoulli* distribution. Write $B_i \sim \text{Bern}(0.7)$. If for all n and $x_i \in \{0, 1\}$,

$$\mathbb{P}(B_1 = x_1, B_2 = x_2, \ldots, B_n = x_n) = \mathbb{P}(B_1 = x_1)\mathbb{P}(B_2 = x_2) \cdots \mathbb{P}(B_n = x_n),$$

this sequence will be iid. Taken together, the $\{B_i\}$ form a *Bernoulli process.*

However, if the sequence is $B_1, B_1, B_1, \ldots$, then the sequence is identically distributed, but they are not independent, since knowing the value of the first random variable in the sequence tells you everything about the random variables in the rest of the sequence.

1.4 Markov chains

A Markov chain is also a sequence of random variables $X_1, X_2, \ldots$, but unlike an iid sequence, they are not independent. In particular, the value of X_i is not independent of the rest of the random variables, but is allowed to depend on X_{i-1}.

These are used in several ways.

- They were introduced as a theoretical model of how letters appear in literature. For instance, in English the letter g is much more likely to be followed by an a than another g.

- For approximately sampling from high dimensional distribution. For instance, if you have ever shuffled a deck of cards, each time you shuffle can be modeled as a step in a Markov chain. As you shuffle, the chain moves towards its *limiting distribution*, which is uniform over the set of permutations of the cards. This feature appears in most Markov chains, and is the core of the *Ergodic Theorem*.

Suppose that $D_1, D_2, \ldots$ are an iid sequence of $\mathsf{Bern}(0.6)$ random variables. Now set $X_0 = 0$, and for all $i > 0$ let $X_i = X_{i-1} + D_i$. The sequence $X_0, X_1, X_2, \ldots$ forms a Markov chain. Knowing the value of X_i completely determines the distribution of X_{i+1}, and having the extra information about X_{i-1} does not affect the distribution of X_{i+1} in any way.

While the distribution of X_i is allowed to depend on X_{i-1}, it must not depend on $X_1, \ldots, X_{i-2}$. This makes Markov chains *forgetful* or *memoryless*.

1.5 Expected value

Some random variables have an *average* value, also known as the *expectation*, *expected value*, or *mean*. Suppose that many independent copies of a the random variable are generate. Then take the sample average of the copies by adding them up and dividing by the number of copies. This sample average will converge to the expected value of the original variable with probability 1 as the number of copies goes to infinity.

1.6 Martingales

A process where the average value at the next step of the process given all the previous values of the process is called a *martingale*. This can also be called a *fair game*, since on average the amount is not going up nor going down, but staying at the same level.

Note that Markov chains deal with the *distribution* of X_t given the past, while martingales deal with the *average value* of X_t given the past. Some Markov chains are also martingales and some are not. Similarly, there are stochastic processes that are martingales but not Markov chains.

Suppose a game is played where a player has a 50% chance of winning a dollar, and a 50% chance of losing a dollar, independent of what happened before. Then on average, the player will neither gain nor lose money at each step, which makes this a fair game.

Now suppose that the player wins a dollar with probability 80%, and only loses with probability 20%. Then the average gain for the player will be

$$(0.8)(1) + (0.2)(-1) = 0.6,$$

which is positive. So this is *not* a martingale, and favors the player.

1.7 Brownian motion

Up until now, our discussion has focused on sequences of random variables. Some calculations and models make more sense to use $\{X_t\}$, where t is any nonnegative number. These give rise to *continuous time stochastic processes*.

Probably the most important such process is *Brownian Motion*, which models many situations in the physical sciences and finance.

Brownian motion is both a Markov process and a martingale, and so the facts about these classes of processes can also be applied here.

1.8 What is in this text

To prove theorems rigorously in mathematics, it is necessary to have more precise definitions of the objects involved. That means that terms like

- probability function,
- random variable,
- distribution,
- expected value,
- and conditional expected value

need to all be defined carefully before theorems can be formulated. In this text, all of these terms will be defined in terms of first order logic.

Problems

1. Fill in the blank: In a __________, the distribution of X_n given $X_1, \ldots, X_{n-1}$ only depends on the value of X_{n-1}.

2. Fill in the blank: In a __________, the expected value of X_n given $X_1, \ldots, X_{n-1}$ equals X_{n-1}.

3. An alternate name for a martingale is what?

4. Give three alternate names for the expected value of a random variable.

5. Fill in the blank: A variable with partial information is a _________ variable.

6. Fill in the blank: In a __________, the expected value of X_n given $X_1, \ldots, X_{n-1}$ equals X_{n-1}.

7. Suppose B_1 and B_2 are independent Bernoulli random variable with parameter 0.6. This means that $\mathbb{P}(B_i = 1) = 0.6$ and $\mathbb{P}(B_i = 0) = 0.4$ for i either 1 or 2. What is $\mathbb{P}(B_1 = B_2 = 1)$?

8. Suppose $A_1 = A_2 = X$, where X is a Bernoulli random variable with parameter 0.7. True or false: the random variables A_1 and A_2 are independent.

9. Suppose you think that a company stock will go 10% with probability 70%, and will decline 5% with probability 30%. Is the change in the value of the stock a random variable?

10. A company employs exactly 5 people. Is the number of people employed by the company better modeled as a constant or a random variable?

Chapter 2

A simple stochastic process

Question of the Day

An American Roulette wheel has 38 spaces each of which can be modeled as equally likely to come up on a spin. Eighteen of these spaces are colored red, eighteen are colored black, and two are colored green.

Suppose a player starts with exactly one dollar. At each spin, the player bets one dollar and wins a dollar if the space is colored red, and loses the dollar otherwise. The game ends when the player has zero dollars left.

Let T denote the number of spins needed for the player's balance to reach zero. What is the expected value of T?

Summary

- Expected value is a linear operator, so for constants a and b,

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

- The **Strong Law of Large Numbers** (SLLN) says that if a random variable has an expected value, then the limit of the sample average of n iid draws from the same distribution as the random variable will be the expected value with probability 1.
- The conditional expectation $\mathbb{E}(X|Y)$ treats the random variable Y like a constant. The resulting average depends on Y, and so is a function of Y.
- Expected value can be calculated using an **expectation tree** whose branches reflect different events occurring. The branches are labeled in the middle with the probability that event occurs, and at the end with the conditional expectation given that event occurred. To calculate expectation, multiply the label of the branch times the value at the end of the branch, and sum the result.

2.1 Random variables

A *random variable* is a variable where you only have partial information about the true value. For instance, the total number of people alive in the continental United States (so all states except Hawaii and Alaska) at this moment is a random variable. It is some specific integer value, but you do not have total information about it. The unknown population could be represented by a random variable N.

In the roulette Question of the Day posed at the beginning of this chapter, the number of steps is unknown ahead of time. Therefore, this can be represented by a random variable T.

Of course, this random variable is a function of other random variables, namely, the spins of the wheel. Let D_i be the amount of money won by the player on the ith spin of the wheel. Because 18 out of 38 spaces are colored red, and 20 out of 38 spaces are not, the random variable D_i has distribution

$$\mathbb{P}(D_i = 1) = \frac{18}{38},\ \mathbb{P}(D_i = -1) = \frac{20}{38}.$$

The stream of random variables $D_1, D_2, D_3, \ldots$ from the spins have two properties.

First, they are *identical*, which means that all the D_i have the same probability distribution. The notation for this is for all positive integers i and j, $D_i \sim D_j$. The symbol $\sim$ here means that the left and right hand sides are random variables with the same distribution.

Second, they are *independent*. Recall the notation $[X|Y]$ means the distribution of the random variable X conditioned on the value of the random variable Y. The independence of the $D_1, D_2, D_3, \ldots$ means that for every integer i at least 2, $[D_i | D_{i-1}, D_{i-2}, \ldots, D_1] \sim D_i$.

With these two properties, call $D_1, D_2, \ldots$ an independent, identically distributed stream of random variables, or iid for short.

Anytime there exist random variables that are indexed by some set, the result is a *stochastic process*.

Definition 1
Random variables that have an index form a **stochastic process**.

Then because the player starts with \$1, and each D_i is the change in the amount of money the player has after the ith spin, the amount of money the player has after n spins will be

$$1 + D_1 + D_2 + \cdots + D_n.$$

The Question of the Day is asking about the smallest value of n that makes this sum 0. Mathematically, this can be written using the *infimum* function. The infimum of a set of positive integers is the smallest value in the set, unless the set is empty, in which case the infimum returns ∞. In this way, the first time that the sum reaches 0 can be written as:

$$T = \inf(\{t : 1 + D_1 + D_2 + \cdots + D_t = 0\}).$$

The Question of the Day asks about the expected value of the number of spins.

2.2 Expected value

The expected value of a random variable is the *average* of the result of many independent draws of the random variable.

Suppose that $X_1, X_2, \ldots$ are an iid stream of random variables where $X_i \sim X$ for all i. Then the Strong Law of Large Numbers gives that

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n}$$

exists and equals $\mathbb{E}[X]$ with probability 1 if $\mathbb{E}[X]$ exists and is finite. (Note that at this point $\mathbb{E}[X]$ has not been formally defined, a formal definition will come later.)

Theorem 1
The Strong Law of Large Numbers

For X such that $\mathbb{E}[X]$ exists and is finite, and $X_1, X_2 \ldots$ iid where $X_i \sim X$ for all i,

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mathbb{E}[X]\right) = 1.$$

When X is a random variable that only takes on a finite number of values, the expected value always exists and is finite. The value is found by summing the different values that the random variable takes on times the probability that it takes on those values.

For example, for X the roll of a fair six sided die, X is either 1, 2, 3, 4, 5, or 6. The probability for each of those values is $1/6$. Hence

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{6}1 + \frac{1}{6}2 + \cdots + \frac{1}{6}6 \\
&= \frac{1+2+3+4+5+6}{6} \\
&= \frac{21}{6} \\
&= \frac{7}{2} = 3.5.
\end{aligned}$$

So if a fair six sided die is rolled over and over, the limit of the sample average of the results will converge to 3.5 with probability 1.

2.3 Linearity of expected value

Recall that if

$$\lim_{n\to\infty} f_n = L, \ \lim_{n\to\infty} g_n = M,$$

then for any constants a and b,

$$\lim_{n\to\infty} af_n + bg_n = aL + bM.$$

Because the SLLN links limits and expected value, they must follow the same rules as limits. In particular, they are *linear* operators.

Theorem 2
Linearity of expected value

Suppose for random variables X and Y, $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ are finite. Then for any constants a and b,

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Note that it does not matter here if X and Y are independent, or dependent random variables!

Example 1
Suppose X_1 and X_2 are two fair throws of a six sided die, and $S = X_1 + X_2$. Then $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 3.5$, and

$$\mathbb{E}[S] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 7.$$

2.4 Expectation trees

One way to solve the Question of the Day is to use an *expectation tree*. This idea is a special case of a far reaching idea that will be called the *Fundamental Theorem of Probability* here because it allows the user to easily solve many problems in probability involving conditioning.

To start, consider the notation

$$\mathbb{E}[X|Y].$$

This represents the average value of X conditioned on the value of Y.

The simplest fact about these conditional means is

$$\mathbb{E}[X|X] = X.$$

That is, if you know the value of X, the mean of X will be that known value of X. (If you prefer rigorous definitions do not worry, they will come later in the text!)

If Y is independent of X, then $[X|Y] \sim X$ so

$$\mathbb{E}[X|Y] = \mathbb{E}[X].$$

Consider again the example where X_1 and X_2 are two fair throws of a six sided die, and $S = X_1 + X_2$. Then $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 3.5$, and

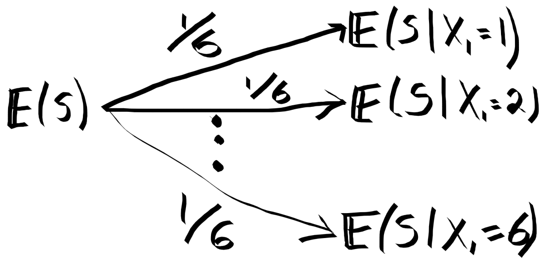
$$\mathbb{E}[S] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 7.$$

If the value of the first die roll was known, then

$$\mathbb{E}[S|X_1] = \mathbb{E}[X_1|X_1] + \mathbb{E}[X_2|X_1] = X_1 + 3.5.$$

An expectation tree says that if Y takes on a finite set of values, you can break apart the original conditioning into branching for each of the possible values in the conditioning.

In the example above, X_1 is either 1, 2, 3, 4, 5, or 6. So the $\mathbb{E}[S]$ can be found by considering all six of these possibilities, multiplied by their probability.



The equation is

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}(S|X_1 = 1)\mathbb{P}(X_1 = 1) + \cdots + \mathbb{E}(S|X_1 = 6)\mathbb{P}(X_1 = 6) \\ &= \frac{1}{6}\mathbb{E}(S|X_1 = 1) + \cdots + \frac{1}{6}\mathbb{E}(S|X_1 = 6) \\ &= \frac{1 + 3.5 + 2 + 3.5 + 3 + 3.5 + \cdots + 6 + 3.5}{6} \\ &= 3.5 + 3.5 \\ &= 7.\end{aligned}$$

This is the same value gotten through linearity!

Two useful facts about conditional expectation are as follows.

Fact 1
The expected value of a random value conditioned on itself is just the random variable. So

$$\mathbb{E}[X|X] = X.$$

Fact 2
If X and Y are independent, then

$$\mathbb{E}[X|Y] = \mathbb{E}[X]$$

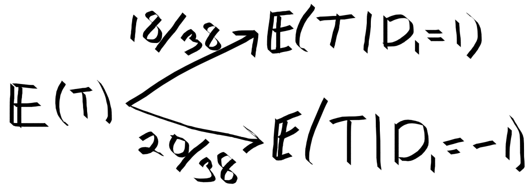
when $\mathbb{E}[X]$ exists.

2.4.0.1 Solving the Question of the Day

With expectation trees and conditioning it is possible to solve the Question of the Day! Consider the first spin of the roulette wheel. If the player loses the spin (and so their dollar) then $T = 1$. Otherwise, one spin has been spent, and the player has two dollars. To lose those two dollars, it is necessary to first lose the dollar just gained, and then the original dollar the player started with. This takes time $2\mathbb{E}[T]$. Combined with the first spin, this gives

$$\begin{aligned}\mathbb{E}[T|D_1 = -1] &= 1 \\ \mathbb{E}[T|D_1 = 1] &= 1 + 2\mathbb{E}[T].\end{aligned}$$

Now set up the expectation tree.



$$\begin{aligned}\mathbb{E}[T] &= \mathbb{E}[T|D_1=-1]\mathbb{P}(D_1=-1) + \mathbb{E}[T|D_1=1]\mathbb{P}(D_1=1)\\ &= 1(20/38) + (2\mathbb{E}[T]+1)(18/38)\\ &= 1 + (18/19)\mathbb{E}[T]\end{aligned}$$

Now solve for $\mathbb{E}(T)$:

$$\begin{aligned}(1/19)\mathbb{E}[T] &= 1\\ \mathbb{E}[T] &= \boxed{19}.\end{aligned}$$

Note that $\mathbb{E}(T)$ only had a solution because the chance of losing a dollar $20/38$ was higher than the chance of winning a dollar $18/38$. In this situation where the chance of winning was $20/38$ and the chance of losing $18/38$, solving the same equation would result in $\mathbb{E}[T] < 0$, which is not possible!

Because the solution is meaningless, the conclusion is that one of the assumptions fails. Really the only assumption made was that T had a finite expectation. Since that assumption fails and $T \geq 0$, it holds that $\mathbb{E}[T] = \infty$ in this case.

2.5 *Useful ideas*

To solve the Question of the Day, several techniques and ideas were used.

- Expectation of a random variable can be easier to understand than the variable itself.
- Expectation trees can be used to break a problem into pieces.
- Conditioning on one element of a problem can assist in solving it.

Problems

11. What is an indexed collection of random variables called?

12. For a collection $\{X_1, X_2, X_3\}$ what values does the index take on?

13. If $\mathbb{E}(X) = 3.2$, what is $\mathbb{E}[2X+3]$?

14. If $\mathbb{E}(Y) = -1.2$, what is $\mathbb{E}[4Y-2]$?

15. Suppose

$$\begin{aligned}\mathbb{E}(X \mid W=0) &= 4.2\\ \mathbb{E}(X \mid W=1) &= 3.2\\ \mathbb{P}(W=0) &= 0.6\\ \mathbb{P}(W=1) &= 0.4\end{aligned}$$

Find $\mathbb{E}(X)$.

16. Suppose

$$\begin{aligned}\mathbb{E}(Y|A=-1) &= -1.2\\ \mathbb{E}(Y|A=0) &= 0\\ \mathbb{E}(Y|A=1) &= 2.3\\ \mathbb{P}(A=-1) &= 0.2\\ \mathbb{P}(A=0) &= 0.2\\ \mathbb{P}(A=1) &= 0.6\end{aligned}$$

Find $\mathbb{E}(Y)$.

17.

Let

$$S = B_1 + \cdots + B_n$$

where the B_i are iid $\mathsf{Bern}(0.3)$. (This means $\mathbb{P}(B_i = 1) = 0.3,\ \mathbb{P}(B_i = 0) = 0.7$.

a. Find $\mathbb{E}(B_i)$.

b. Find $\mathbb{E}(S)$.

c. Find $\mathbb{E}(S|B_1)$.

18.

Let

$$N = G_1 + \cdots + G_k.$$

Suppose $\mathbb{E}(G_i) = 1.2$ for all i.

a. What is $\mathbb{E}(N)$?

b. What is $\mathbb{E}(N|G_1 = 3)$?

19.

In a *corner bet* in Roulette, a player is allowed to bet $1 on four numbers simulataneously. If any of these four numbers come up, the player wins $8 and their original bet is returned, otherwise they lose their original bet.

(a) Suppose a player is playing American Roulette with 38 slots on the wheel where the chance of winning a corner bet is $4/38$. They start with \$1 and play until their money is gone. If T is the number of plays this takes, and $\mathbb{E}[T]$ exists, find $\mathbb{E}[T]$.

(b) Repeat part (a) where the player is on a European Roulette table where the chance of winning a corner bet is $4/37$.

20.

A player is playing a game with a 10% chance of winning \$2, a 15% chance of winning \$1, and a 75% chance of losing \$1.

(a) Starting with \$1, let T be the number of plays until the player reaches no money. Given that $\mathbb{E}[T]$ exists, find its value.

(b) Repeat part (a), but now assume the player starts with \$5.

Chapter 3

Logic

Question of the Day

What is

$$(3 > 4) \vee (7 > 5)?$$

Summary

- A **logical statement** is either true or false.
- Logical AND is true if and only if every statement in the AND is true.
- Logical OR is true if and only if at least one statement in the OR is true.
- Logical NOT changes true to false and false to true.
- **Indicator functions** take as input a logical statement and has output 1 for a true statement and 0 for a false statement.
- **Probability functions** extend the indicator function by returning a number in $[0, 1]$ that gives the information the user has about the truth of the statement.

3.1 What is logic?

In using probability, often the goal is to understand the probability that an event is true or false. The mathematical objects that evaluate to true or false are called *logical statements*.

Definition 2
A **logical statement** is an expression that evaluates to be either true (T) or false (F).

3.2 Logical operators

The main *logical operators* are *logical AND*, *logical OR*, and *logical NOT*. The logical operators are capitalized to emphasize that they are not quite used in logic the same way they are used in natural language. First consider logical AND.

Definition 3
The **logical AND** of two logical statements p and q is written $p \wedge q$, and evaluates to

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

More generally, given a group of logical statements p_α indexed by α, the *logical AND* is true if and only if all the statements are true.

Definition 4
The **logical AND** of one or more statements $\{p_\alpha\}$ is written

$$\bigwedge_\alpha p_\alpha$$

and evaluates to true if and only if every indexed statement p_α is true.

When there are only a finite number of statements $(p_1, \ldots, p_n)$, the notation

$$p_1 \wedge p_2 \wedge \cdots \wedge p_n$$

is also used for the logical AND between the statements.

The *for all universal quantifier* notation can also be used for logical AND.

Definition 5
The **for all** (aka **for every**) existential quantifier is

$$(\forall \alpha)(p_\alpha) = \bigwedge_\alpha p_\alpha.$$

The logical OR is true if at least one of the statements is true. As before start with two statements.

Definition 6
The **logical OR** of two logical statements p and q is written $p \vee q$, and evaluates to

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

Definition 7
The **logical OR** of one or more statements $\{p_\alpha\}$ is written

$$\bigvee_\alpha p_\alpha$$

and evaluates to true if and only if at least one statement p_α is true.

When there are only a finite number of statements $(p_1, \ldots, p_n)$, the notation

$$p_1 \vee p_2 \vee \cdots \vee p_n$$

is also used for logical OR.

The *there exists existential quantifier* notation can also be used for logical OR.

Definition 8
The **there exists** (aka **there is**) existential quantifier is

$$(\exists \alpha)(p_\alpha) = \bigvee_\alpha p_\alpha.$$

Example 2
The Question of the Day can now be resolved:

$$(3 > 4) \vee (7 > 5) = \mathsf{F} \vee \mathsf{T} = \boxed{\mathsf{T}}.$$

The last commonly used logical operator is *logical NOT*, which flips true statements to false, and false statements to true.

Definition 9
If $p = \mathsf{T}$ then the **logical NOT** of p is $\neg p = \mathsf{F}$. If $p = \mathsf{F}$, then $\neg p = \mathsf{T}$.

Example 3
Try a modified version of the Question of the Day:

$$(3 > 4) \vee \neg(7 > 5) = \mathsf{F} \vee \neg\mathsf{T} = \mathsf{F} \vee \mathsf{F} = \boxed{\mathsf{F}}.$$

3.3 *Sets and their operators*

Informally, a *set* is a collection of elements. Every possible object under consideration is either in the set or not in the set. The notation $a \in S$ is read "a is an element of S", and $a \notin S$ is read "a is not an element of S". It cannot happen that a is both an element and not an element of S, but at least one of these things is true.

Definition 10
Say that S is a **set** if for all objects a, $(a \in S)$ is a logical statement. If $(a \in S) = \mathsf{T}$, say that a is an **element** of the set.

The slash through the element symbol means that the object is *not* an element of the set.

Notation 1
If $(a \in S) = \mathsf{F}$ for a set S, write $a \notin S$.

A common way of writing the elements of a set is through *list notation*. A list of objects is written inside curly braces $\{\}$. If an object appears in this list, it is an element of the set, otherwise it is not.

Example 4
Consider the set A written in list notation.

$$A = \{a, b, c\}.$$

This is equivalent to the following logical statement.

$$(\forall x)((x \in A) = (x = a) \lor (x = b) \lor (x = c)).$$

Notice that logical OR is commutative, and so order does not matter in list notation. The … symbol is often employed with integer sets to fill in a pattern.

Example 5
Consider the following sets.

$$\begin{aligned} \{4, \ldots, 10\} &= \{4, 5, 6, 7, 8, 9, 10\} \\ \{4, 5, \ldots\} &= \{i \text{ is an integer} : i \geq 4\} \\ \{\ldots, 0, 1, 2\} &= \{i \text{ is an integer} : i \leq 2\} \end{aligned}$$

A set A is a *subset* of a set S if every element in A is also an element of S.

Definition 11
Say that a set A is a **subset** of a set S if

$$(\forall a \in A)(a \in S).$$

The three set operators that correspond to logical AND, logical OR, and logical NOT are *intersection*, *union*, and *complement*.

Definition 12
For sets S_α, their **intersection** is the set of elements in all the sets:

$$(a \in \cap_\alpha S_\alpha) = \bigwedge_\alpha (a \in S_\alpha),$$

their **union** is the set of elements in at least one of the sets:

$$(a \in \cup_\alpha S_\alpha) = \bigvee_\alpha (a \in S_\alpha),$$

and the **complement** of set S is the set of elements not in the set:

$$(a \in S^C) = \neg(a \in S).$$

A special set is the *empty set*, which contains no elements at all.

Definition 13
The **empty set** (written $\emptyset$ or $\{\}$) is the set such that

$$(\forall a)(a \notin \emptyset).$$

In notation, the symbol for the empty set $\emptyset$ is often confused with the Greek letter ϕ. In fact, $\emptyset$ is a crossed out circle, indicating a set that does not even contain 0. It is also similar to (but not quite) a Danish letter Ø. Note that the main part of the symbol is a circle, whereas the curve in the Greek letter phi is wider than it is tall.

3.4 Indicator functions

The *indicator function* is a way of converting from logic to arithmetic.

Definition 14
The **indicator function** takes as input a logical statement and evaluates as:

$$\mathbb{I}(\mathsf{T}) = 1$$
$$\mathbb{I}(\mathsf{F}) = 0.$$

When using indicator functions, logical AND becomes multiplication.

Fact 3
For logical statements p and q,

$$\mathbb{I}(p \wedge q) = \mathbb{I}(p)\mathbb{I}(q).$$

For logical OR, it is a bit more complicated, but checking all the possibilities proves the following fact.

Fact 4
For logical statements p and q,

$$\mathbb{I}(p \vee q) = \mathbb{I}(p) + \mathbb{I}(q) - \mathbb{I}(p)\mathbb{I}(q).$$

Logical NOT is easy:

Fact 5
For a logical statement p,
$$\mathbb{I}(\neg p) = 1 - \mathbb{I}(p).$$

3.5 What exactly are probabilities?

If indicator functions indicate by 1 or 0 if the input is true or false, what do probabilities do? They *measure* our information about the truth of a statement. If the statement is true, the probability is 1 just like the indicator function. If the statement is false, the probability is 0 as well.

Things get interesting when the truth or falsehood of the statement is unknown. Saying something like $\mathbb{P}(X \geq 7) = 0.6$ means that the information that the statement $(X \geq 7)$ is at 0.6. It can be any number from 0 up to 1.

This means that probability is an extension of logic to include information.

Problems

21.

State whether each logical statement is true or false.

a. $(3 < 4) \wedge (7 = 7)$.

b. $(3 < 4) \wedge (7 = 8)$.

c. $(3 < 4) \vee (7 = 7)$.

d. $(3 < 4) \vee (7 = 8)$.

State whether each logical statement is true or false.

a. $(10 = 10) \wedge (7 < 8)$.

b. $(10 > 5) \wedge (7 < 8)$.

c. $(3 = 3) \vee (7 > 8)$.

d. $(3 > 4) \vee (7 > 8)$.

22.

Evaluate

a. $(3 > 4) \wedge (7 > 5)$.

b. $\neg(3 > 4) \wedge \neg(7 > 5)$.

23.

Evaluate

a. $(10 = 10) \wedge (-1 < 2)$.

b. $(10 = 10 \wedge \neg(-1 < 2)$.

24.

State whether each logical statement is true or false.

a. $(\forall x \in [3,4])(x < 5)$

b. $(\exists x \in [3,4])(x < 5)$

c. $(\forall x \in [3,4])(x < 3.5)$

d. $(\exists x \in [3,4])(x < 3.5)$

State whether each logical statement is true or false.

a. $(\exists x \in [-1,1])(x < -2)$

b. $(\exists x \in [-1,1])(x < 2)$

c. $(\forall x \in [-1,1])(x < 2)$

d. $(\exists x \in [-1,1])(x < 0.5)$

25.

State whether each statement about sets is true or false.

a. $3 \in \{1,2,3,4\}$.

b. $5 \in \{1,2,3,4\}$

c. $\{1,2\} \subseteq \{1,2,3,4\}$.

d. $\{1,5\} \subseteq \{1,2,3,4\}$.

26.

Evaluate

a. $\mathbb{I}((10 = 10) \wedge (-1 < 2))$.

b. $\mathbb{I}((10 = 10 \wedge \neg(-1 < 2))$.

27. If $\mathbb{I}(p) = \mathbb{I}(q) = 1$, what is $\mathbb{I}(p \wedge q)$?

28. If $\mathbb{I}(p) = \mathbb{I}(q) = 1$, what is $\mathbb{I}(p \vee q)$?

Chapter 4

Random variables

Question of the Day

How can random variables be modeled mathematically?

Summary

- *Events* are logical statements where a probability can be assigned.
- A collection of events are **mutually exclusive** aka **disjoint** if at most one of the statements can be true.
- A **σ-algebra of events** is a collection of logical statements that is closed under negation (logical NOT) and closed under countable logical OR.
- A function from a σ-algebra to $[0, 1]$ is a **probability function** aka **probability measure** if the probability of a true statement is 1 and for a countable collection of disjoint events, the probability of their logical or equals the sum of the probabilities of the events.
- A **σ-algebra of sets** is closed under complements and countable logical union.
- A **measurable space** is a state space together with a σ-algebra of subsets of the state space. The sets in the σ-algebra are called **measurable**.
- X is a **random variable** if there is a σ-algebra of sets such that for every set A in the algebra the value of $\mathbb{P}(X \in A)$ is defined.
- Given two measurable spaces A and B with σ-algebras $\mathcal{F}_A$ and $\mathcal{F}_B$, say that $f : A \to B$ is a **measurable function** if the inverse of any set in $\mathcal{F}_B$ is an element of $\mathcal{F}_A$.

4.1 Events

Logical statements are either true or false, there is no in between. So for instance, $(3 < 4) = \mathsf{T}$, and $(3 = 4) = \mathsf{F}$. In order to turn these true/false statements into numbers, the *indicator function* can be used.

$$\mathbb{I}(\mathsf{T}) = 1$$
$$\mathbb{I}(\mathsf{F}) = 0.$$

Note that indicator functions can be used to count the number of true statements in a collection of logical statements. For instance, given

$$\{(3 < 4), (3 = 4), (2 < 7)\}$$

it holds that

$$\mathbb{I}(3 < 4) + \mathbb{I}(3 = 4) + \mathbb{I}(2 < 7) = 1 + 0 + 1 = 2$$

and exactly two of these three statement were true.

This idea can be generalized.

Fact 6
For $\{p_\alpha\}$ a finite or countable set of logical statements, the number of p_α that are true equals the sum of the indicator functions of the events. That is,

$$\#(\{\alpha : p_\alpha = \mathsf{T}\}) = \sum_\alpha \mathbb{I}(p_\alpha).$$

Suppose now that N represents the number of people in a room. The number is unknown, and so the truth of a statement like $(N < 4)$ is unknown. *Probabilities* then assign a number between 0 and 1 (inclusive) that reflects the measure of the information about the truth of the statement. Usually the symbol $\mathbb{P}$ is used to indicate the probability of an logical statement. Like with the indicator function, the probability of a true statement is 1. A logical statement that can be assigned a probability is called an *event*.

Definition 15
Say that a logical statement is an **event** if $\mathbb{P}(p)$ is defined.

Probabilities cannot be assigned in any old way, they have to follow certain rules. They must fall into the interval $[0, 1]$, and the probability of a true event should be 1. That is, $\mathbb{P}(\mathsf{T}) = 1$.

Another rule is that when an event is broken into two events that cannot both happen at the same time, the probability that either event happens is the sum of the probabilities of the events. Two or more events such that at most one is true are called *mutually exclusive* or *disjoint*. Our indicator function fact can be used to make this idea precise.

Definition 16
A finite or countable set $\{p_\alpha\}$ of statements are **mutually exclusive** or **disjoint** if

$$\sum_\alpha \mathbb{I}(p_\alpha) \leq 1.$$

This allows us now to write a required property of probability functions, called *countable additivity*. For $\{p_\alpha\}$ a finite or countable set of disjoint events,

$$\mathbb{P}\left(\bigvee_\alpha p_\alpha\right) = \sum_\alpha \mathbb{P}(p_\alpha).$$

Example 6
Suppose that X is the roll of a fair six sided die, so $\mathbb{P}(i) = 1/6$ for $i \in \{1,2,3,4,5,6\}$. This could also have been written as

$$\mathbb{P}(X = i) = \frac{1}{6}\mathbb{I}(i \in \{1,2,3,4,5,6\}).$$

What is $\mathbb{P}(X \in \{2,4\})$?

Answer Since the event

$$(X \in \{2,4\}) = (X = 2) \vee (X = 4),$$

and $(X = 2)$ and $(X = 4)$ are disjoint,

$$\mathbb{P}(X \in \{2,4\}) = \mathbb{P}(X = 2) + \mathbb{P}(X = 4) = \frac{1}{6} + \frac{1}{6} = \boxed{0.3333\ldots},$$

Example 7
Suppose $(X = i)$ are all events for $i \in \{0,1,2,\ldots\}$, as is $X \in \{0,1,2,\ldots\}$. If $\mathbb{P}(X \in \{0,1,2,\ldots\}) = 1$, what is

$$\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \cdots ?$$

Answer Since the events $(X = i)$ for i a nonnegative integer are disjoint, countable additivity gives

$$\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \cdots = \mathbb{P}(X \in \{0,1,2,\ldots\}) = \boxed{1}.$$

To make countable additivity easier to use, it would help if whenever $p_1, p_2, \ldots$ are events, that their logical OR

$$\bigvee_{i=1}^{\infty} p_i$$

is also an event. Also, if the probability of event p is $\mathbb{P}(p)$, it should also be possible to assign a probability to the event that p does not happen, that is $\neg p$. In other words, $\neg p$ should also be an event. These two properties together make a collection of events a σ-algebra.

Definition 17
A nonempty collection $\mathcal{F}$ of logical statements is called a **σ-algebra** if the following two properties hold.

1. **Closed under negation** $(\forall p \in \mathcal{F})(\neg p \in \mathcal{F})$.
2. **Closed under countable logical OR** $(\forall p_1, p_2, \ldots \in \mathcal{F})(p_1 \vee p_2 \vee \cdots \in \mathcal{F})$.

The symbol $\mathcal{F}$ is often used for σ-algebras because the properties have the collection *closed* under complements and countable unions. The French word for closed is *fermé*, which starts with F.

With this definition, one can now prove facts about such collections.

Fact 7
For any σ-algebra of logical statements $\mathcal{F}$, it holds that $\mathsf{T} \in \mathcal{F}$ and $\mathsf{F} \in \mathcal{F}$.

Proof. Since all σ-algebras are nonempty, let $p \in \mathcal{F}$. Then $\neg p \in \mathcal{F}$ as well, which means that

$$p \vee \neg p \vee \neg p \vee \neg p \vee \cdots \in \mathcal{F}.$$

Since $p \vee \neg p \vee \neg p \vee \cdots = \mathsf{T}$, and $\neg \mathsf{T} = \mathsf{F}$, the proof is complete.

□

4.2 Distributions

Now a *probability function*, also known as a *distribution*, can be formally defined.

Definition 18
A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a **probability function** (aka **probability measure** aka **distribution**) if $\mathcal{F}$ is a σ-algebra, and

1. **Probability of true statements** $\mathbb{P}(\mathsf{T}) = 1$.
2. **Countable additivity** For all $p_1, p_2, \ldots \in \mathcal{F}$ disjoint,

$$\mathbb{P}(p_1 \vee p_2 \vee \cdots) = \mathbb{P}(p_1) + \mathbb{P}(p_2) + \cdots.$$

4.3 Random variables

Now consider a *real-valued random variable* X. Such a variable only takes on values in $\mathbb{R}$, furthermore, probabilities can be assigned to events of the form $(X \in A)$ for a collection of sets $A \subseteq \mathbb{R}$ that form a σ-algebra.

Definition 19
Say that X is a **real-valued random variable** if there exists a σ-algebra $\{\mathcal{F}_X\}$ of subsets of $\mathbb{R}$ such that the events $\{X \in A : A \in \mathcal{F}_X\}$ form a σ-algebra where $\mathbb{P}(X \in A)$ is defined for all $A \in \mathcal{F}_X$. Call the elements of $\mathcal{F}_X$ **measurable with respect to X**. Further, set

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A).$$

Call $\mathbb{P}_X(A)$ the **distribution of X**. Write $X \sim \mathbb{P}_X$ to indicate that this is the distribution. If $\mathbb{P}(X \in \Omega) = 1$, call Ω the **sample space** or **outcome space** of X.

Some notation for distributions is helpful.

Notation 2
Write $X \sim m$ to denote that m is the distribution of X. Write $X \sim Y$ if the random variables X and Y have the same distribution.

Example 8
For X the roll of a fair four sided die, the sample space is $\Omega = \{1, 2, 3, 4\}$. The set $\mathcal{F}_X$ is all subsets of Ω (the powerset of Ω), so

$$\mathcal{F}_X = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{1,4\},$$
$$\{2,3\}, \{2,4\}, \{3,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, \{1,2,3,4\}\}.$$

The distribution of X can be written as

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \frac{\#(A)}{4}.$$

This distribution is given a name, it is the *uniform distribution* over Ω. Write

$$X \sim \mathsf{Unif}(\{1, 2, 3, 4\}).$$

4.4 *σ-algebras of sets*

For a random variable X, events are all of the form $(X \in A)$ where A are sets that are measurable with respect to X. What can be said about these sets?

Because the events form a σ-algebra of logical statements, they are closed under negation. So if $(X \in A)$ is an event, so is $\neg(X \in A) = (X \in A^C)$.

Similarly, if $A_1, A_2, \ldots$ are sets, then

$$(X \in \cup_i A_i) = \vee_i (X \in A_i).$$

This leads to the following definitions of σ-algebra for sets.

Definition 20
A nonempty collection $\mathcal{F}$ of sets is called a **σ-algebra** if the following two properties hold.

1. **Closed under complements** $(\forall A \in \mathcal{F})(A^C \in \mathcal{F})$.

2. **Closed under countable unions** $(\forall A_1, A_2, \ldots \in \mathcal{F})(A_1 \cup A_2 \cup \cdots \in \mathcal{F})$.

A sample space together with a σ-algebra of sets is a *measurable space*.

Definition 21
Say that $(\Omega, \mathcal{F})$ is a **measurable space** if $\mathcal{F}$ is a collection of subsets of Ω that forms a σ-algebra. The elements of $\mathcal{F}$ are called **measurable sets**.

Example 9
Suppose that 1, 2, and 3 are the only mathematical objects being considered. Then

$$\mathcal{A} = \{\emptyset, \{1,2\}, \{3\}, \{1,2,3\}\},$$

and

$$\mathcal{B} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$$

are both σ-algebras of sets.

The collection $\mathcal{B}$ is another example of a powerset, the set of all subsets of $\{1,2,3\}$.

Fact 8
The powerset of a set forms a σ-algebra of sets.

Proof. Let $\mathcal{U}$ be a set of objects, and $\mathcal{P}(\mathcal{U})$ be the powerset. Let $A \in \mathcal{P}(\mathcal{U})$. Then $A^C \in \mathcal{P}(\mathcal{U})$ since $\mathcal{P}(\mathcal{U})$ contains every subset of $\mathcal{U}$.

Moreover, if $A_1, A_2, \ldots \in \mathcal{P}(\mathcal{U})$, then

$$(\forall i)(A_i \subseteq \mathcal{U}) \rightarrow \left(\bigcup_{i=1}^{\infty} A_i\right) \subseteq \mathcal{U},$$

hence

$$\left(\bigcup_{i=1}^{\infty} A_i\right) \in \mathcal{P}(\mathcal{U}),$$

so $\mathcal{P}(\mathcal{U})$ is a σ-algebra.

□

For countable sets, the powerset is often used as the standard σ-algebra for the set. However, under a commonly used set of axioms used for the real numbers called ZFC, it is not possible to assign probabilities consistently to all subsets of the real numbers. So instead, the *Borel sets* are usually used as the σ-algebra. In the Borel sets, intervals of the form $(-\infty, a] = \{x : x \leq a\}$ are all measurable, which is usually good enough for all needed applications. Formally, the Borel sets consists of any set that appears in every σ-algebra that contains all intervals of the form $(-\infty, a]$. Since the powerset contains all these intervals, the

Borel sets are nonempty.

4.5 Disjoint sets

Note that if two sets A and B do not overlap, that is, if $A \cap B = \emptyset$, then $(X \in A)$ and $(X \in B)$ cannot both happen, since

$$(X \in A) \wedge (X \in B) = (X \in AB) = (X \in \emptyset) = \mathsf{F}.$$

This motivates the following definition for sets.

Definition 22
Say that a collection of sets is **disjoint** or **mutually exclusive** if for any pair of sets A and B in the collection, $A \cap B = \emptyset$.

4.6 Functions of a random variable

The next thing to consider is what happens when a function is applied to a random variable? Is the new value still a random variable? For instance, if X is a random variable, and $W = X^2$, is W also a random variable. It depends on the σ-algebra associated with X and W.

Suppose that $f : A \to B$ where A and B are subsets of $\mathbb{R}$. Then for X a real-valued random variable, let $Y = f(X)$. Because X is a random variable, it has an associated σ-algebra of sets, call it $\mathcal{F}_A$.

For Y to be a random variable, there must be a σ-algebra associated with B, call it $\mathcal{F}_B$. So what is needed for Y to be a random variable is for the sets in $\mathcal{F}_B$ to be measurable. Let $C \in \mathcal{F}_B$. Then

$$(Y \in C) = (f(X) \in C).$$

Definition 23
For $f : A \to B$ and $C \subseteq B$, define the set $f^{-1}(C)$ to be all elements $a \in A$ such that $f(a) \in C$.

For example, if $f(x) = x^2$, then $f^{-1}([1,4]) = [-2,-1] \cup [1,2]$, since any value in $[-2,-1] \cup [1,2]$ squared gives a value in $[1,4]$, and no other real number squared gives a value in $[1,4]$.

Using this notation,

$$(Y \in C) = (f(X) \in C) = (X \in f^{-1}(C)).$$

Therefore, Y is a random variable if and only if whenever $C \in \mathcal{F}_B$, $f^{-1}(C) \in \mathcal{F}_A$. Such functions f are called *measurable*.

Definition 24
The function $f : A \to B$ is **measurable** with respect to $\mathcal{F}_A$ and $\mathcal{F}_B$ which are σ-algebras with respect to A and B respectively, if

$$(\forall C \in \mathcal{F}_B)(f^{-1}(C) \in \mathcal{F}_A).$$

Fact 9
If $f : A \to B$ is a measurable function and X is a random variable with $\mathbb{P}(X \in A) = 1$, then $f(X)$ is also a random variable.

Proof. This follows directly from the definitions.

□

Problems

29.

Suppose that W takes on values in $1, 2, 3, \ldots$, and

$$\mathbb{P}(W = i) = (3/4)(1/4)^i$$

for $i \in \{1, 2, \ldots\}$.

a. What is $\mathbb{P}(W = 2)$?

b. What is $\mathbb{P}(W \neq 2)$?

30. Suppose that X takes on values in $1, 2, 3, \ldots$, and

$$\mathbb{P}(X = i) = 2(1/3)^i$$

What is $\mathbb{P}(X \geq 2)$?

31. Suppose $\Omega = \{1, 2, 3, 4, 5\}$ and $\{1, 2\}$ are sets in a σ-algebra $\mathcal{F}$. Give three other sets that must also be in $\mathcal{F}$.

32. Suppose $\Omega = \{1, 2, 3, 4, \ldots\}$ and for all i, $\{i\}$ is a measurable set. Give at least five other measurable sets.

33.

Suppose $f(x) = 3x$.

a. What is $f^{-1}([0, 4])$?

b. What is $f^{-1}([-1, 4])$?

34.

Suppose $f(x) = x^2$.

a. What is $f^{-1}([0, 4])$?

b. What is $f^{-1}([-1, 4])$?

35. Consider the following four intervals.

$$A_1 = [-1, 1],\ A_2 = [0, 4],\ A_3 = [-3, -2],\ A_4 = [-4, 4].$$

Which pairs of these four sets are disjoint?

36. Consider the following three sets:

$$S_1 = \emptyset,\ S_2 = [-10, 10],\ S_3 = [5, 15]$$

Which pairs among these three sets are disjoint?

Chapter 5

Expected value

Question of the Day

How should the mean of a random variable be formally defined?

Summary

- The **mean of a simple random variable** is the sum of the outcomes of the variable times the probability that the random variable equals each of those outcomes.
- The **mean of a nonnegative random variable** X is the smallest number that is still greater than the mean of any finite random variable less than or equal to X.
- The **mean of a general random variable** X is the mean of the positive part $\max(X, 0)$ minus the mean of the negative part $\max(-X, 0)$ when both of these are finite.
- Random variables with a finite mean are called **integrable**. A random variable X is integrable if and only if $\mathbb{E}[|X|] < \infty$.

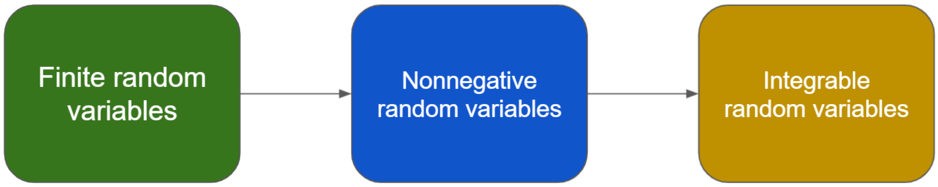
The *mean*, *expected value*, *average*, or *expectation* of a random variable all mean the same thing. The idea is that if $X_1, X_2, \ldots$ are iid draws with the same distribution of X, then the sample average

$$\lim_{n \to \infty} \frac{X_1 + X_2 + \cdots + X_n}{n}$$

will approach (with probability 1) some value that represents the average value of the random variable.

Sometimes this is true, and sometimes not. When it does happen, the random variable is called *integrable*, and the expectation is written as $\mathbb{E}[X]$.

Here is the plan of action. First the expectation will be defined for random variables that are finite, that only take on a finite set of values with probability 1. Second, these random variables will allow the extension of the definition to all nonnegative random variables. Finally, the definition will be extended to all integrable random variables.



5.1 Finite random variables

So how can this value be calculated? Consider first a random variable X where

$$\mathbb{P}(X = 1) = 0.2,\ \mathbb{P}(X = 3) = 0.3,\ \mathbb{P}(X = 4) = 0.5.$$

Then roughly 20% of the time, the $X_1, \ldots, X_n$ will equal 1, roughly 30% of the time, it will equal 3, and roughly 50% of the time, they will equal 4. So hopefully it makes sense that

$$\mathbb{E}[X] = 0.2(1) + 0.3(3) + 0.5(4) = 3.1.$$

A random variable that only takes on a finite number of values will be called *finite*.

Definition 25
Suppose that for a random variable X there is a set A that contains a finite number of elements, and

$$\mathbb{P}(X \in A) = 1.$$

Then call X **finite** or **simple**.

For any random variable that only takes on a finite number of values with probability 1, the mean of the variable is the sum of the values times the probability that X takes on those values.

Definition 26
Suppose $\mathbb{P}(X \in \{x_1, \ldots, x_n\}) = 1$. Then the **mean** (aka **expectation** aka **expected value** aka **average**) of the random variable is

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i).$$

Another way to write this sum is

$$\mathbb{E}[X] = \sum_{x} x\mathbb{P}(X = x),$$

since if you try to include an x value where X has probability 0 of hitting, that term of the sum just drops out.

With this simple definition, it is possible to prove important properties of expected value, such as linearity of expectations.

Fact 10
Suppose $\mathbb{P}(X \in \{x_1, \dots, x_n\}) = 1$ and $\mathbb{P}(Y \in \{y_1, \dots, y_m\}) = 1$. Then

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y],$$

and for any $c \in \mathbb{R}$, $\mathbb{E}[cX] = c\mathbb{E}[X]$.

Note that this holds true even if X and Y depend on each other. For instance,

$$\mathbb{E}[X + X^2] = \mathbb{E}[X] + \mathbb{E}[X^2].$$

Proof. There are at most $n \cdot m$ values that $X+Y$ can take on, so

$$\mathbb{E}[X+Y] = \sum_{s:(\exists i,j)(s=x_i+y_j)} s\mathbb{P}(X+Y=s)$$

There could be multiple ways that x_i and y_j add up to equal s. Since they are disjoint,

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_{s:(\exists i,j)(s=x_i+y_j)} \sum_i \sum_j s\mathbb{P}(X=x_i, Y=y_j)\mathbb{I}(x_i+x_j=s) \\
&= \sum_i \sum_j (x_i+y_j)\mathbb{P}(X=x_i, Y=y_j) \\
&= \sum_i x_i \sum_j \mathbb{P}(X=x_i, Y=y_j) + \sum_j y_j \sum_i \mathbb{P}(X=x_i, Y=y_j)
\end{aligned}$$

Note that $\sum_j \mathbb{P}(X=x_i, Y=y_j)$ covers all possible values for Y, and so is just $\mathbb{P}(X=x_i)$. Similarly, $\sum_i \mathbb{P}(X=x_i, Y=y_j) = \mathbb{P}(Y=y_j)$. So

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_i \sum_j (x_i+y_j)\mathbb{P}(X=x_i, Y=y_j) \\
&= \sum_i x_i\mathbb{P}(X=x_i) + \sum_j y_j\mathbb{P}(Y=y_j) \\
&= \mathbb{E}[X] + \mathbb{E}[Y],
\end{aligned}$$

and the proof of the first part is complete.

For the second part, let $c \in \mathbb{R}$ be nonzero. Then $\mathbb{P}(cX \in \{cx_1, \dots, cx_n\}) = 1$. So

$$\begin{aligned}
\mathbb{E}[cX] &= \sum_w w\mathbb{P}(cX=w) \\
&= c\sum_w (w/c)\mathbb{P}(X=w/c) \\
&= c\mathbb{E}[X].
\end{aligned}$$

If $c=0$, then $\mathbb{E}[cX] = \mathbb{E}[0] = 0 = 0\mathbb{E}[X]$, so either way, it holds!

□

5.2 Expectation of indicator functions are probabilities

One nice fact that follows directly from this definition is that the expected value of the indicator of an event equals the probability the event occurs.

Fact 11
For any random variable X with measurable set A,

$$\mathbb{E}[\mathbb{I}(X \in A)] = \mathbb{P}(X \in A).$$

Proof. Note that $\mathbb{I}(X \in A) \in \{0, 1\}$ so $\mathbb{I}(X \in A)$ is a finite random variable. So

$$\mathbb{E}[\mathbb{I}(X \in A)] = (1)\mathbb{P}(\mathbb{I}(X \in A) = 1) + (0)\mathbb{P}(\mathbb{I}(X \in A) = 0) = \mathbb{P}(X \in A).$$

□

Hence the idea of expected value extends the notion of probability! This is why most theorems in advanced probability focus on expected value, since that way they are as general as possible.

5.3 Nonnegative random variables

So that works for random variables that take on only a finite number of values with probability 1, but what about ones that have sample space $\{0, 1, 2, \ldots\}$. Or random variables like $\text{Unif}([0, 1])$, whose state space is uncountable?

Well, one property that it would be nice to have for random variables is if one random variable is smaller than another, then the mean of the first should be smaller than the mean of the second. This property is called *monotonicity*, and will help with our definition.

For a nonnegative random variable X, let W be any random variable where $\mathbb{P}(W \in \{w_1, \ldots, w_n\}) = 1$, such that $\mathbb{P}(W \leq X) = 1$ as well. Then $\mathbb{E}[X]$ should be at least $\mathbb{E}[W]$. So define $\mathbb{E}[X]$ to be the smallest number that is at least $\mathbb{E}[W]$ for *all* such finite W satisfying $\mathbb{P}(W \leq X) = 1$.

5.3.1 Example 1: Geometric random variables

Suppose $X \sim \text{Geo}(1/2)$, which means that

$$\mathbb{P}(X = i) = \left(\frac{1}{2}\right)^i \mathbb{I}(i \in \{1, 2, 3, \ldots\}).$$

Can a lower bound on $\mathbb{E}[X]$ be found?

Let $Y_m = \min(X, m)$, that is, Y_m is the smaller of either X, or the value m. Note that $Y_m \in \{1, 2, \ldots, m\}$, so for instance, $Y_4 \in \{1, 2, 3, 4\}$. A quick calculation shows:

$$\begin{aligned}
\mathbb{P}(Y_4 = 1) &= 1/2 \\
\mathbb{P}(Y_4 = 2) &= 1/4 \\
\mathbb{P}(Y_4 = 3) &= 1/8 \\
\mathbb{P}(Y_4 = 4) &= \mathbb{P}(X \geq 4) = 1 - 1/2 - 1/4 - 1/8 = 1/8.
\end{aligned}$$

Hence

$$\mathbb{E}[X] \geq \mathbb{E}[Y_4] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} = \frac{15}{8}.$$

In general, $\mathbb{E}[Y_m] = 2 - (1/2)^{m-1}$, which means that $\mathbb{E}[X] \geq 2$.

5.3.2 Example 2: Standard uniform

Now let $W \sim \mathsf{Unif}([0,1))$. A useful function to use here is the *floor* function, which rounds down to the nearest integer. So floor$(7.3) = 7$, floor$(3) = 3$, and floor$(-3.2) = -4$.

Because the floor function rounds down, for positive n and x, it holds that

$$\text{floor}(nx) \leq nx$$

so

$$\text{floor}(nx)/n \leq x.$$

Therefore, for our random variable W,

$$Z_n = \text{floor}(nW)/n \leq W.$$

Multiplying $W \in [0,1)$ by n puts it in $[0,n)$, taking the floor makes it in $\{0,1,2,\ldots,n-1\}$, and dividing by n makes

$$Z_n \in \left\{\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\right\}$$

Note that (for instance),

$$\mathbb{P}(Z_n = 3/n) = \mathbb{P}(nZ_n = 3) = \mathbb{P}(nW \in [3,4)) = \mathbb{P}(W \in [3/n, 4/n)) = 1/n.$$

Hence $Z_n \sim \mathsf{Unif}(\{0, 1/n, 2/n, \ldots, (n-1)/n\})$. A quick check shows that $(n-1)/n - Z_n$ has the same values and probabilities as Z_n, so

$$\mathbb{E}[((n-1)/n) - Z_n] = \mathbb{E}[Z_n].$$

Hence

$$((n-1)/n) = \mathbb{E}[Z_n + ((n-1)/n) - Z_n] = \mathbb{E}[Z_n] + \mathbb{E}[((n-1)/n) - Z_n] = \mathbb{E}[Z_n] + \mathbb{E}[Z_n],$$

so

$$\mathbb{E}[Z_n] = \frac{1}{2} \cdot \frac{n-1}{n}.$$

Therefore,

$$\mathbb{E}[W] \geq \frac{1}{2} \cdot \frac{n-1}{n}$$

for all $n \in \{1,2,3,\ldots\}$.

5.4 Formal definition of expectation for nonnegative random variables

So what should the expected value be? So far

$$\mathbb{E}[X] \geq \mathbb{E}[Y]$$

for any Y such that $\mathbb{P}(Y \leq X) = 1$. The idea is make $\mathbb{E}[X]$ the *smallest* number that is greater than all of these values. This idea goes by the name *supremum*, and for a set of real numbers S, is written $\sup(S)$. It equals the least upper bound on the set of numbers S.

For instance, $\sup([0,3)) = 3$, since 3 is larger than all the numbers in $[0,3)$, and it is the least such number.

Definition 27
Given a subset of real numbers S, the **supremum** of S is written as

$$\sup(S)$$

and is as follows

1. If $S = \emptyset$, then $\sup(S) = -\infty$.
2. If $S \neq \emptyset$ and there exists any number $w \in \mathbb{R}$ such that for all $s \in S$, $s \leq w$ then
$$\sup(S) = \min\{y : (\forall s \in S)(s \leq y)\}.$$
3. If $S \neq \emptyset$ and there does not exist a number $w \in \mathbb{R}$ such that $(\forall s \in S)(s \leq w)$, then $\sup(S) = \infty$.

In other words, if S is empty, then any number is an upper bound, and the least upper bound is $-\infty$. If S is nonempty and has an upper bound, then the supremum is the least upper bound. Finally, if S is nonempty and does not have an upper bound, then say the supremum is ∞.

With this idea, the expectation of a nonnegative random variable can now be defined!

Definition 28
Let W be a random variable such that $\mathbb{P}(W \geq 0) = 1$. Then

$$\mathbb{E}[W] = \sup_{Y \text{ finite}} (\{\mathbb{E}[Y] : \mathbb{P}(Y \leq W) = 1\}).$$

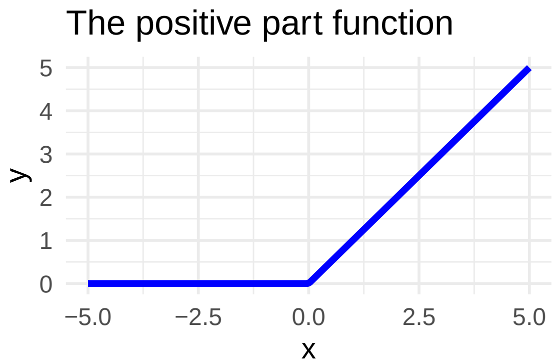
5.5 Integrable random variables

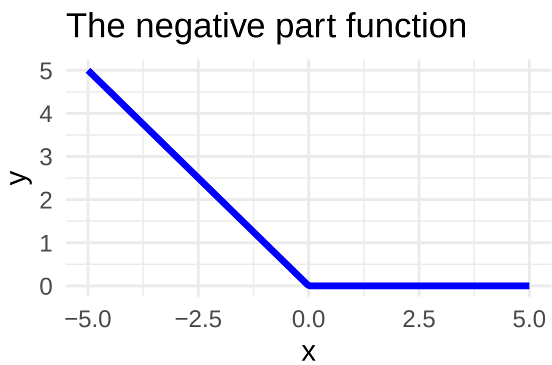
Because the supremum can be ∞, so can the expected value of a random variable. If the expected value is finite, the random variable is called *integrable*.

To extend this to all random variables, consider two functions, the *positive part* and the *negative part* function.

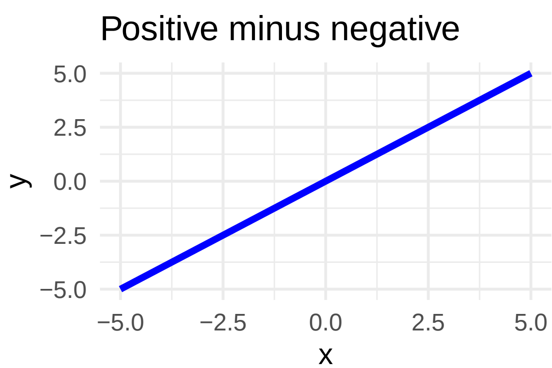
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

```
## generated.
```





If $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$ are the *positive part* and *negative* part of a real number, note that x^+ and x^- are always nonnegative, and that $x^+ - x^-$ is equal to just x.



This can be used to define the expected value of a

Definition 29
If $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ exist, then let

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

Note that $x^+ + X^- = |X|$. Hence $\mathbb{E}[|X|] = \mathbb{E}[X^+] + \mathbb{E}[X^-]$. The left hand side is finite if and only if both terms on the right hand side are finite. This gives the following fact.

Definition 30
A random variable is **integrable** if $\mathbb{E}[|X|] < \infty$.

Problems

37. Let X be a random variable such that

$$\mathbb{P}(X = 1) = 0.2,\ \mathbb{P}(X = 1.5) = 0.2,\ \mathbb{P}(X = 2) = 0.6.$$

Find $\mathbb{E}[X]$.

38. Suppose

$$\mathbb{P}(W = 1) = \mathbb{P}(W = 2) = \mathbb{P}(W = 3) = 0.1,$$

and $\mathbb{P}(W = 4) = 0.7$. Find $\mathbb{E}[W]$.

39. If X is a finite random variable, show that X^2 is also a finite random variable.

40. If X is a finite random variable and c is a real constant, show that $X + c$ is also a finite variable.

41.

Let $U \sim \text{Unif}([0,1])$ so U is a continuous random variable such that for all $0 \leq a \leq b \leq 1$,

$$\mathbb{P}(U \in [a,b]) = b - a.$$

Let

$$W = 0 \cdot \mathbb{I}(U < 1/3) + 0.3 \cdot \mathbb{I}(1/3 \leq U < 2/3) + 0.6 \cdot \mathbb{I}(U > 2/3).$$

a. Is W a finite random variable?

b. Does $W \leq U$ hold with probability 1?

c. What is $\mathbb{E}[W]$?

d. Give a lower bound for $\mathbb{E}[U]$ utilizing W.

42. Continuing the last problem, get a slightly better lower bound for $\mathbb{E}[U]$ using

$$R = 0 \cdot \mathbb{I}(U < 1/3) + (1/3) \cdot \mathbb{I}(1/3 \leq U < 2/3) + (2/3) \cdot \mathbb{I}(U > 2/3).$$

43. Prove that $\mathbb{E}[\mathbb{I}(X \in A)] = \mathbb{P}(X \in A)$ for any A where $\mathbb{P}(X \in A)$ is defined.

44. For A measurable with respect to X, find

$$\mathbb{E}[\mathbb{I}(X \in A)^2].$$

45. Let $X \sim \text{Exp}(1)$ be a standard exponential random variable, so

$$\mathbb{P}(X \in [a, b)) = \int_a^b \exp(-x)\, dx$$

for all $0 \leq a \leq b$.

Suppose the random variable Y is defined as

$$Y = 0 \cdot \mathbb{I}(X < 1) + 1 \cdot \mathbb{I}(1 \leq X < 2) + 2 \cdot \mathbb{I}(X \geq 2).$$

Use Y to give a lower bound on $\mathbb{E}[X]$.

46. Continuing the last problem, use

$$S = 0 \cdot \mathbb{I}(X < 1) + 1 \cdot \mathbb{I}(1 \leq X < 2) + 2 \cdot \mathbb{I}(2 \leq X < 3) + 3 \cdot \mathbb{I}(X \geq 3)$$

to give a better lower bound on $\mathbb{E}[X]$.

47. Some notation: for real numbers x and y,

$$x \vee y = \max(x, y).$$

Suppose that $\mathbb{E}[X \vee 0] = 7$ and $\mathbb{E}[-X \vee 0] = 12$. What is $\mathbb{E}[X]$?

48. If $\mathbb{E}[W \wedge 0] = 10$ and $\mathbb{E}[-W \wedge 0] = 5$, what is $\mathbb{E}[W]$?

49.

The *supremum* of a set of real numbers is the smallest number that is still an upper bound for the set. So for instance, the intervals $(1, 2)$ and $[1, 2]$ both have supremum 2, since that is the smallest number that is an upper bound for all elements of the set. By convention, if the set has no upper bound the supremum is ∞, and a set that is empty has supremum $-\infty$.

With that in mind, find the supremum of each of the following sets. You do not have to justify your answer.

a. $\{x : 3 < x < 10\}$.

b. $\{1, 2, 3, \ldots\}$.

c. $\emptyset$

50.

Find the supremum of the following sets.

a. $(-1, 1)$.

b. $[-1, 1]$.

c. $[1, \infty)$.

Chapter 6

Integrals

Question of the Day

Suppose $X \sim \text{Exp}(2)$. Find $\mathbb{E}[X]$.

Summary

- A **measure** gives the empty set a measure of 0, and the measure of the union of a countable collection of disjoint sets equals the sum of the measures of those sets.
- The **integral** of an indicator function with respect to a measure is the measure of the set where the indicator is true.
- Finite sums of constants times indicator functions are called **simple**.
- Integrals are linear operators over simple functions.
- For nonnegative functions f, the **integral** of the function is the smallest number that is at least as large as the integral of any simple function less than or equal to f,
- For general functions, the **integral** of the function is the integral of the positive part $\max(f, 0)$ minus the integral of the negative part $\max(-f, 0)$ when these two integrals are finite numbers.
- A function is **integrable** with respect to a number if the integral exists and is finite.
- When the Riemann integral of a function exists this has the same value of the integral of the function with respect to Lebesgue measure.
- When $\mathbb{P}(X \in A) = \int f \, d\mu$ for all measurable sets A, say that f is the **density** of X.
- For μ a measure and g a measurable function,

$$\mathbb{E}[g(X)] = \int g(x) f(x) \, d\mu(x).$$

While the previous chapter gave a formal method for defining the expected value, it really is not that great at allowing us to *calculate* the expected value. Instead, a connection needs to be made between integration and expectation.

6.1 Measures

Start by considering a *measure*. Earlier, a *probability measure* was defined over a σ-algebra of logical statements. In general, a *measure* is a nonnegative function that measures how big a mathematical object is.

Earlier, the mathematical objects were logical statements, now consider when the objects are sets. The domain of the function must be a σ-algebra of sets $\mathcal{F}$. Recall that such a σ-algebra is nonempty, and satisfies two properties:

1. If $A \in \mathcal{F}$, then $A^C \in \mathcal{F}$ so the collection is closed under the complement operation.
2. If $A_1, A_2, \ldots \in \mathcal{F}$, then $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$ as well. So the collection is closed under countable union.

Say that m is a *measure* if it takes as input members of a σ-algebra, and outputs a nonnegative number such that the measure of the empty set is 0, and the measure of the union of a disjoint sequence of sets equals the sum of the measures of the sets.

Definition 31
For $\mathcal{F}$ a σ-algebra of sets, say that $m : \mathcal{F} \to [0, \infty) \cup \{\infty\}$ is a **measure** if

1. $m(\emptyset) = 0$.
2. If $A_1, A_2, \ldots \in \mathcal{F}$ and for all $i \neq j$, $A_i \cap A_j = \emptyset$, then

$$m(A_1 \cup A_2 \cup \cdots) = \sum_{i=1}^{\infty} m(A_i).$$

Historically, this second property is seen in *Zeno's paradox*, which breaks the interval $[0, 1]$ into a sequence of intervals $[0, 1/2), [1/2, 3/4), [3/4, 7/8)$ and so on. The measure of the whole interval is 1, and the sum of the measures of the parts is $1/2 + 1/4 + 1/8 + \cdots = 1$ as well.

The two most commonly used measures are *counting measure* and *Lebesgue measure*.

6.2 Counting measure

In *counting measure* (usually denoted $\#(A)$ or $|A|$,) the measure of a set is the number of elements of the set. So $\#(\{1, 5, 7\}) = 3$. If there are not a finite number of points, then the counting measure is infinity. For instance,

$$\#(\{1, 2, 3, \ldots\}) = \#([0, 1]) = \infty.$$

6.3 Lebesgue measure

The *Lebesgue measure* of a set is length in one dimension, area in two dimensions, volume in three dimensions, and hypervolume in more than three dimensions. For instance

$$\begin{aligned}\text{Leb}([3.1, 7.2]) &= 7.2 - 3.1 = 4.1 \\ \text{Leb}([0,1] \times [4,7]) &= \text{Leb}(\{(x,y) : 0 \leq x \leq 1, 4 \leq y \leq 7\}) \\ &= (1-0)(7-4) \\ &= 3.\end{aligned}$$

6.4 Integration

Measures tell us the size of a set, while integrals represent a weighted size of the set, where element a is assigned weight $f(a)$. The first integrals that people are introduced to are usually the Riemann integral. The integrals considered here are more general, and allow the definition of integrals for a larger class of integrands.

To start, the integral of the function that is 1 over the set A and 0 elsewhere is just the measure of A. For instance,

$$\int_{x \in \mathbb{R}} \mathbb{I}(x \in [3.1, 7.2]) \, d\,\text{Leb}(x) = \int_{x \in [3.1, 7.2]} 1 \, d\,\text{Leb}(x) = 7.2 - 3.1 = 4.1.$$

This starts our definition of an integral.

Definition 32
The **integral of an indicator function** with respect to measure μ is the measure of the set where the indicator function is 1. That is

$$\int_{x \in \Omega} \mathbb{I}(x \in A) \, d\mu(x) = \mu(A).$$

As with the Riemann integral, the general measure integral should be a linear operator.

Definition 33
For $a_1, a_2, \ldots, a_n \in \mathbb{R}$ and $A_1, \ldots, A_n$ measurable sets,

$$\int_{x \in \Omega} \sum_{i=1}^{n} a_i \mathbb{I}(x \in A_i) \, d\mu(x) = \sum_{i=1}^{n} a_i \mu(A_i).$$

For nonnegative functions f which are not a simple linear combination of indicator functions, define the integral as with expectation. Make it the supremum over all simple linear combinations that are bounded above by f.

Definition 34
For f measurable,

$$\int_{x \in \Omega} f(x) \, d\mu(x) = \sup_{\text{measurable } g \leq f} \int_{x \in \Omega} g(x) \, d\mu(x).$$

If $|f|$ has finite integral over Ω, then f is an *integrable function*.

Definition 35
Let f be a measurable function. If

$$\int_{x\in\Omega} |f(x)|\,d\mu(x) < \infty,$$

then f is **integrable**.

Finally, if f is not nonnegative and integrable, break it up into the positive part and the negative part.

Definition 36
If f is an integrable function, then

$$\int_{x\in\Omega} f(x)\,d\mu(x) = \int_{x\in\Omega} \max(f,0)\,d\mu(x) - \int_{x\in\Omega} \max(-f,0)\,d\mu(x).$$

As with expected value, the integral of a function is not easy to calculate from the definition. However, two facts make these calculations much easier.

6.5 *Integrals with respect to counting measure*

Because counting measure assigns a value 1 to single points, integrals with respect to counting measure just become sums!

Fact 12
Integrals with respect to counting measure can be found as

$$\int_{x\in\Omega} f(x)\,d\#(x) = \sum_{x\in\Omega} f(x).$$

Example 10
Find

$$\int_{x\in\{1,2,3,4\}} x^2\,d\#(x).$$

Answer This is just

$$\begin{aligned}\int_{x\in\{1,2,3,4\}} x^2\,d\#(x) &= \sum_{x\in\{1,2,3,4\}} x^2 \\ &= 1^2+2^2+3^2+4^2 \\ &= \boxed{30}\end{aligned}$$

6.6 *Integrals with respect to Lebesgue measure.*

Fact 13
Suppose that the Riemann integral

$$I = \int_{A \subseteq \mathbb{R}^n} f(x_1, \ldots, x_n)\, d\mathbb{R}^n$$

exists. Then A is a measurable set, and the Lebesgue integral

$$\int_{A \subseteq \mathbb{R}^n} f(x_1, \ldots, x_n)\, d\mathbb{R}^n$$

is equal to I.

That means that all the wonderful Calculus techniques learned over the years still apply!

6.7 Densities

Definition 37
A random variable X has **density** f_X with respect to measure μ if for all measurable A,

$$\mathbb{P}(X \in A) = \int_{s \in A} f_X(s)\, d\mu(s).$$

If μ is counting measure, then if f_X is a density of X,

$$\mathbb{P}(X \in A) = \int_{s \in A} f_X(s)\, d\#(s) = \sum_{s \in A} f_X(s).$$

But from countable additivity,

$$\mathbb{P}(X \in A) = \sum_{s \in A} \mathbb{P}(X = s).$$

That shows the following fact.

Fact 14
If X is a discrete random variable, then it has density

$$f_X(s) = \mathbb{P}(X = s)$$

with respect to counting measure.

Continuous random variables, on the other hand, will have a density with respect to Lebesgue measure.

Fact 15
If X is a continuous random variable, then it has a density with respect to Lebesgue measure.

6.8 Actually calculating expected values

The main purpose of a density is to allow calculation of expected value through the use of an integral.

Fact 16
If f_X is a density of X with respect to measure μ and g is a measurable function, then

$$\mathbb{E}[g(X)] = \int_{x \in \Omega} g(s) f_X(s) \, d\mu(s)$$

when $g(X)$ is integrable.

So integrals are really a form of expected value! For this reason, finding $\mathbb{E}[g(X)]$ is often called *integration* (and of course when the value is finite, the random variable is *integrable*.) As seen, expected values can be difficult to calculate based on the definition. Fortunately, the Lebesgue integral is an extension of the Riemann integral, so all the techniques learned in a first Calculus course for finding integrals apply.

With that in mind, now the question of the day can be solved.

Example 11
Suppose that X is a random variable with an exponential distribution with rate 2. What is the mean of X?

Answer An exponential random variable with rate 2 has density

$$f_X(s) = 2\exp(-2s)\mathbb{I}(s \geq 0).$$

Hence the expected value is

$$\begin{aligned} \mathbb{E}[X] &= \int_{x \in \Omega} s \cdot 2\exp(-2s)\mathbb{I}(s \geq 0) \, ds \\ &= \int_{s \geq 0} 2s \exp(-2s) \, ds \end{aligned}$$

where the indicator function was taken into the limit.

WolframAlpha could be used to find the answer, but those wishing to get their hands into the engine would use integration by parts to slide over the derivative from one factor to another to get

$$\begin{aligned} \mathbb{E}[X] &= \int_{s \geq 0} 2s \exp(-2s) \, ds \\ &= \int_{s \geq 0} 2s[\exp(-2s)]'/(-2) \, ds \\ &= (2s)(\exp(-2s)/(-2))|_0^\infty - \int_{s \geq 0} [2s]' \exp(-2s)/(-2) \, ds \\ &= 0 + \int_{s \geq 0} \exp(-2s) \, ds \\ &= \exp(-2s)/(-2)|_0^\infty \\ &= 1/2 = \boxed{0.5000}. \end{aligned}$$

Problems

51. Suppose $\mathcal{F}$ is a σ-algebra for sets and for all $i \in \{1, 2, 3, 4, \ldots\}$,

$$m(i) = (1/3)^i.$$

What is $m(\{1, 2, 3, \ldots\})$?

52. Continuing the last problem, find

$$m(\{1, 3, 5, 7, \ldots\}).$$

53.

Suppose $\mu([1, 2)) = 1$, $\mu([2, 3)) = 2$, and $\mu([3, 4]) = 3$.

a. What is $\int_{[1,2)} 1 \, d\mu$?

b. What is $\int \mathbb{I}(x \in [1, 2)) \, d\mu(x)$?

54.

Suppose $\nu(\{1\}) = 1$, $\mu(\{2\}) = 2$, and $\nu(\{3\}) = 3$.

a. What is $\int_3 1 \, d\nu$?

b. What is $\int \mathbb{I}(x = 3) \, d\nu(x)$?

55. Suppose $f(x) \leq g(x)$ for all x and are measurable functions with respect to μ. If

$$\int f(x) \, d\mu(x) = 0.47,$$

what can be said about

$$\int g(x) \, d\mu(x)?$$

56. Suppose $f(x) \leq g(x) \leq h(x)$ are all measurable functions with respect to a measure ν. For

$$I_f = \int f(x) \, d\nu(x)$$
$$I_g = \int g(x) \, d\nu(x)$$
$$I_h = \int h(x) \, d\nu(x),$$

what can be said about the values I_f, I_g, I_h?

57. Suppose

$$f_1(x) \leq g(x)$$
$$f_2(x) \leq g(x)$$
$$f_3(x) \leq g(x)$$

are all measurable with respect to μ. Moreover,

$$\int f_1(x)\,d\mu(x) = 4.2, \int f_2(x)\,d\mu(x) = 8.3, \int f_3(x)\,d\mu(x) = -1.7,$$

What can be said about

$$\int g(x)\,d\mu(x)?$$

58. Suppose that $g(x), f_1(x), f_2(x), f_3(x), \ldots$ are measurable with respect to μ, and for every i, it holds that $g(x) \geq f_i(x)$ and

$$\int f_i(x)\,d\mu(x) = 10 - 1/i^2.$$

What can you say about

$$\int g(x)\,d\mu(x)?$$

59. Suppose that random variable W has density

$$f_W(w) = (1/2)\mathbb{I}(w \in [2,4]).$$

which respect to Lebesgue measure m.

What is $\mathbb{P}(W \geq 3)$?

60. Suppose random variable X has density $f_X(x) = 2\exp(-2x)\mathbb{I}(x \geq 0)$ with respect to Lebesgue measure. What is $\mathbb{P}(X \leq 1)$?

61. Continuing with X from the last problem, what is $\mathbb{E}[X^2]$?

62. Continuing with the last problem, what is $\mathbb{E}[X^3]$?

Chapter 7

Properties of Expectation

Question of the Day

For integrable X, does it always hold that $\mathbb{E}[|X|] \geq |\mathbb{E}[X]|$?

Summary

- There are random variables, such as the **Cauchy distribution** and **Zipf's Law** with parameter $\alpha \in (1, 2]$ where the expected value does not exist.

- Expectation is a **linear operator**, so for integrable random variables X and Y (possibly dependent) and constants a and b,
$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

- **Convex** functions always lie (at or) below their secant line.

- If f is convex and the means exist, **Jensen's inequality** says that
$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X)).$$

- One type of convergence of random variables is **convergence with probability 1**, where
$$\mathbb{P}\left(\lim_{t \to \infty} X_t = Y\right) = 1,$$

- Another type of convergence is **convergence in probability**. This means that for all $m \in \{1, 2, 3, \dots\}$,
$$\lim_{t \to \infty} \mathbb{P}(|X_t - Y| > 1/m) = 0.$$

- Convergence with probability 1 implies convergence in probability, but the converse does not hold.

At first glance it can be surprising that there can be random variables that are finite with probability 1, yet still do not have finite expectation and so are not integrable! Two canonical examples of this are the Cauchy distribution, and the Zipf (power) law distribution with $\alpha \leq 2$.

Definition 38
The standard Cauchy distribution has density

$$f(s) = \frac{2}{\tau} \cdot \frac{1}{1+s^2}$$

with respect to Lebesgue measure.

The Cauchy distribution is an example of a random variable that does *not* have an expected value, even though it is finite with probability 1!

Fact 17
The mean of a standard Cauchy does not exist.

Proof. For X a standard Cauchy, consider $\mathbb{E}(|X|)$.

$$\begin{aligned}
\mathbb{E}[|X|] &= \int \frac{2}{\tau} \cdot \frac{|s|}{1+s^2}\,ds \\
&= 2\int_{[0,\infty)} \frac{2}{\tau} \cdot \frac{s}{1+s^2}\,ds \\
&= 2\lim_{b\to\infty} \frac{1}{\tau} \ln(1+s^2)|_0^b \\
&= \infty.
\end{aligned}$$

Hence X is *not* integrable. □

For the Cauchy, the tail asymptotically goes down quadratically in the distance from the origin. For a *Zipf distribution* with parameter α, the tail goes down asymptotically as the distance raised to the α power.

Definition 39
Say that X has a **Zipf distribution with parameter α** if it has density

$$f(i) = \frac{C}{i^\alpha}\mathbb{I}(i \in \{1,2,3,\ldots\})$$

with respect to counting measure. Here C is the constant that makes this a probability density. Note that $\alpha > 1$ in order for C to exist.

Fact 18
For X a random variable with a Zipf distribution with parameter $\alpha \in (1,2]$, the mean of X does not exist.

Proof. The mean of X will be

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} i\frac{C}{i^\alpha} = \sum_{i=1}^{\infty} \frac{C}{i^{\alpha-1}}.$$

From the rules for this type of sum, this is only finite if $\alpha - 1 > 1$, so the mean only exists when $\alpha > 2$. □

7.1 Vector spaces and linear operators

Real valued integrable random variables together with the real numbers form an example of a *vector space*. Informally, this means that they have the following properties.

1. They can be added together: if X and Y are real valued integrable random variables, then $X + Y$ is a real valued integrable random variable.
2. They can be scaled: if X is a real valued integrable random variable, and $c \in \mathbb{R}$, then cX is a real valued integrable random variable.

Definition 40
For a vector space V with scalars S, we say that L is a **linear operator** if

$$(\forall x, y \in V)(\forall a, b \in S)(L[ax + by] = aL[x] + bL[y]).$$

Example

- Matrix multiplication is a linear operator over vectors in $\mathbb{R}^n$.

$$A(av + bw) = aAv + bAw$$

- Integration is a linear operator over functions with finite integral.

$$\int_A af(x) + bg(x)\, dx = a \int_A f(x)\, dx + b \int_A g(x)\, dx$$

- Differentiation is a linear operator over functions in C^1. Here C^1 means functions with a first derivative that is continuous almost everywhere.

$$[af + bg]' = af' + bg'.$$

- Limits are linear operators over sequences that have limits:

$$\lim_{n\to\infty} c_1 a_n + c_2 b_n = c_1 \lim_{n\to\infty} a_n + c_2 \lim_{n\to\infty} b_n.$$

Since the Strong Law of Large Numbers tells us that expected values can be written as limits, or they can be written as integrals, it should be no surprise that expected value is also a linear operator.

Fact 19
Expectation is a linear operator

For any integrable random variables X and Y, and real numbers a and b:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

- Important: this holds even when X and Y are not independent! For instance, X and X^2 are not independent. Still, $\mathbb{E}(X + X^2) = \mathbb{E}(X) + \mathbb{E}(X^2)$.
- Example: If $\mathbb{E}[X] = 4$ and $\mathbb{E}[Y] = 10$, what is $\mathbb{E}[2X - Y]$? Answer: $\mathbb{E}[2X - Y] = 2\mathbb{E}[X] - \mathbb{E}[Y] = 2(4) - 10 = -2$.

7.2 Domination and convexity

An important fact is that bigger random variables have bigger means.

Fact 20
Domination

If X and Y are integrable r.v. where $X \leq Y$ with probability 1, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Proof. First consider the case where $X \geq 0$.

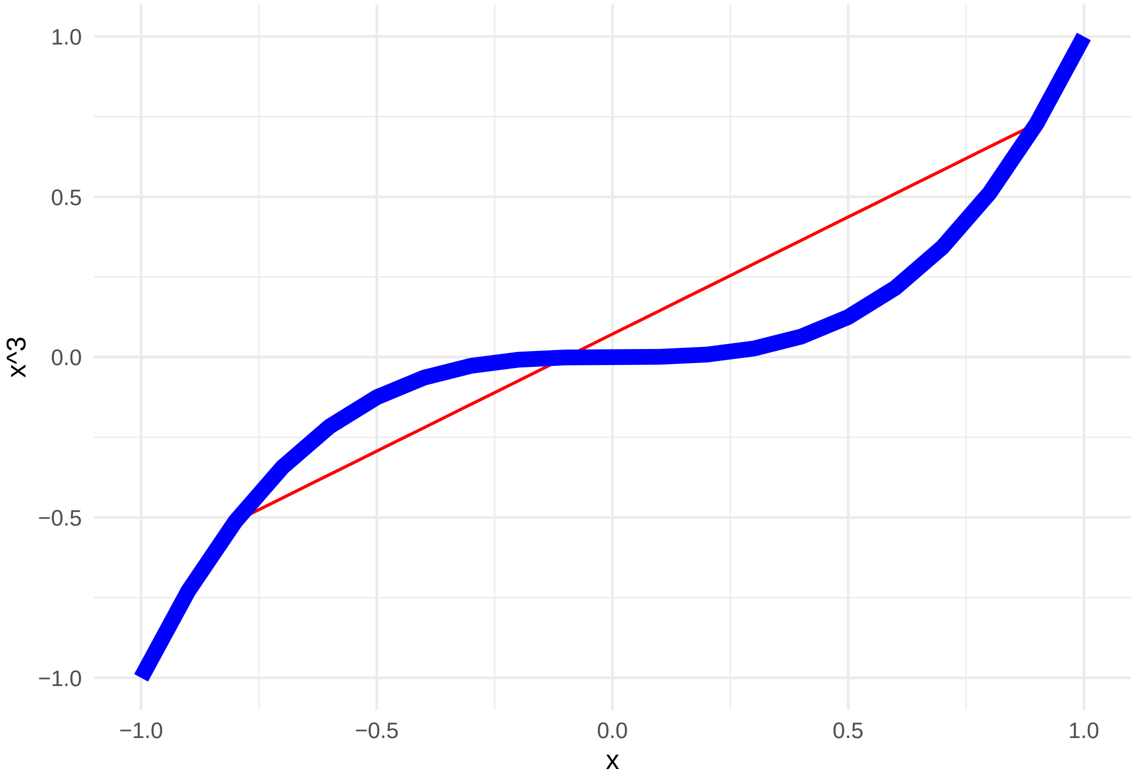
Let S_X be the set of simple functions dominated by X, and S_Y be the set of simple functions dominated by Y. Then $S_X \subseteq S_Y$, so

$$\sup_{S \in S_X} \mathbb{E}[S] \leq \sup_{S \in S_Y} \mathbb{E}[S].$$

The left hand side is just $\mathbb{E}[X]$, and the right hand side is just $\mathbb{E}[Y]$, so we are done with this case.

Now suppose $X \not\geq 0$. Then since $X \leq Y$, $Y - X \geq 0$, so the first case applies and $\mathbb{E}[Y - X] \geq \mathbb{E}[0] = 0$. By linearity $\mathbb{E}[Y - X] = \mathbb{E}[Y] - \mathbb{E}[X] \geq 0$, so we are done. □

Note that between any two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$, it is possible to draw a *secant line segment* between them.



In the picture above, sometimes the secant line connecting the two points on the graph is above the graph of the function, and sometimes it is below. If it is always above the graph of the function, no matter what two points are selected, the function is *convex*.

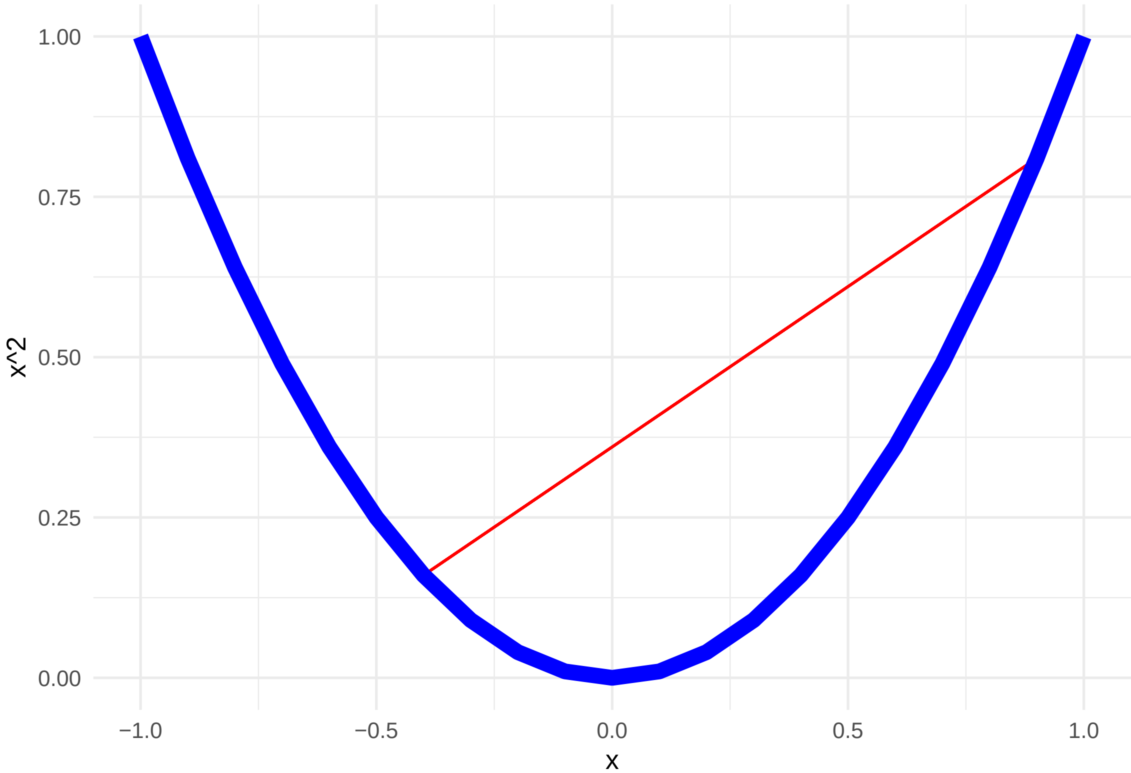
Definition 41
The **secant line segment** of a function f over $[x_1, x_2]$ consists of the points,

$$(\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda)f(x_2))$$

for $\lambda \in [0, 1]$.

Note that the point where $\lambda = 0$ is just $(x_1, f(x_1))$ and when $\lambda = 1$ is just $(x_2, f(x_2))$. When $\lambda = 1/3$, the point is $1/3$ of the way from the left endpoint of the line segment to the right endpoint.

Definition 42
A function is **convex** if the secant line connecting any two points on the graph lies on or above the graph. For example x^2, e^x and $|x|$ are all convex.



In Calculus, the following useful fact is shown which states that a function with continuous second derivative that is always nonnegative over $[a, b]$ is convex over $[a, b]$.

Fact 21
For $f \in C^2[a, b]$, if $f''(x) \geq 0$ for all $x \in [a, b]$, then f is convex over $[a, b]$.

Fact 22
Jensen's inequality

If $f(x)$ is a convex function and X is a random variable where $\mathbb{E}[X]$ and $\mathbb{E}[f(X)]$ exist, then $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$.

In particular, since $x \mapsto x^2$

$$\mathbb{E}[X^2] \geq \mathbb{E}[X]^2 \text{ and } \mathbb{E}\left[|X|\right] \geq |\mathbb{E}(X)|.$$

To see why Jensen's inequality is true, consider a random variable that just takes on two values. Suppose

$$\mathbb{P}(X = x_1) = p \text{ and } \mathbb{P}(X = x_2) = 1 - p.$$

Such a random variable is called *binary*. Then for a convex function X,

$$\mathbb{E}[f(X)] = pf(x_1) + (1-p)f(x_2) \geq f(px_1 + (1-p)x_2) = f(\mathbb{E}(X))$$

So the property of convexity is exactly what is needed to bring the function f out of the expectation.

7.3 Types of convergence for random variables

Earlier, *convergence with probability 1* was discussed:

$$(X_n \to X \text{ wp } 1) = \left(\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1\right)$$

7.4 Convergence in probability

A slightly weaker form of convergence that is still useful is called *convergence in probability*.

The idea is that instead of the limit converging to a fixed number, instead look at the probabilities that the limit is a fixed amount away from a target. These probabilities should go to 0.

Definition 43
Say that $X_1, X_2, \ldots$ **converges to X in probability** if

$$(\forall m \in \{1, 2, \ldots\}) \left(\lim_{t\to\infty} \mathbb{P}\left(|X_t - X| > 1/m\right) = 0\right).$$

Example 12
Suppose that $B_1, B_2, \ldots$ are an independent but *not* identically distributed sequence of Bernoulli random variables. Make $B_i \sim \text{Bern}(1/i)$, so as i gets larger, each B_i is more and more likely to be closer to 0.

Then choose any $m \in \{1, 2, \ldots\}$. Then

$$\lim_{i\to\infty} \mathbb{P}(|B_i - 0| > 1/m) = \lim_{i\to\infty} \mathbb{P}(B_i = 1)$$

since only if $B_i = 1$ will it be at least $1/m$ away from 0. But this probability goes to 0 as i goes to infinity.

Hence the sequence converges to 0 in probability.

So this sequence does converge to 0 in probability, but does it converge with probability 1. No!

Example 13
Continuing the above example,

$$\lim_{n\to\infty} B_i$$

does not exist with probability 1.

Because the B_i are either 0 or 1, the limit (if it existed) would have to be either 0 or 1.

It is definitely not 1, since it is easy to see that no matter how far out in the sequence you go, you would still see another 0 with probability 1.

It turns out it cannot be 0 either, since no matter how far out you go in the sequence you will still see another 1 with probability 1!

Let N be a positive integer. Then for any $M \geq N$,

$$\begin{aligned}\mathbb{P}(B_N \neq 1, B_{N+1} \neq 1, B_{N+2} \neq 1, \ldots) &\leq \mathbb{P}(B_N \neq 1, B_{N+1} \neq 1, B_{N+2} \neq 1, \ldots, B_M \neq 1) \\ &= \frac{N-1}{N}\frac{N}{N+1}\cdots\frac{M-1}{M} \\ &= \frac{N-1}{M}.\end{aligned}$$

since it is a telescoping sum.

This holds for all M, and so the upper bound on the probability can be made arbitrarily small, so the actual probability must be 0.

Hence no matter how far out you go in this sequence, you will find another 0 and then eventually another 1, and then eventually another 0, and so on, so you never converge!

So a sequence can converge in probability while not converging with probability 1. Is the reverse true? No!

Fact 23
If $X_t \to X$ with probability 1, then $X_t \to X$ in probability.

Because convergence with probability 1 implies convergence in probability, but not the other way round, say that convergence in probability is a *weaker* form of convergence, and convergence with probability 1 is *stronger*.

For theorems in probability, when one has convergence with probability 1 it is called a *strong law*, and when one only has convergence in probability, it is called a *weak law*.

For instance, let X be an integrable random variable and $X_1, X_2, \ldots$ an iid sequence of random variables with the same distribution as X. Then let

$$S_n = \frac{X_1 + \cdots + X_n}{n}.$$

Then the *weak law of large numbers* says that

$$S_n \to \mathbb{E}(X)$$

using convergence in probability. So for any $m \in \{1, 2, \ldots\}$, $\mathbb{P}(|S_n - \mathbb{E}(X)| > 1/m)$ converges to 0. It is the probabilities that are converging, not necessarily the random variables.

On the other hand, the *strong law of large number* says that

$$\mathbb{P}(S_n \to X) = 1,$$

which is a more powerful statement. This says that the actual numbers obtained from the sample average of a bunch of draws from the random variable will converge with probability 1 to the desired average of X.

Recall our earlier equation of when it is legal to swap limits and means. It turns out that it is possible to characterize exactly which weakly convergent sequences we can swap expected values and limits with a property called *uniform integrability* that will be discussed later.

Problems

63. For X a Zipf Law random variable, show that $\mathbb{E}[X^3]$ is only finite when $\alpha > 4$.

64. For X a Zipf Law random variable, show that $\mathbb{E}[X^4]$ is only finite when $\alpha > 5$.

65. Suppose $X_1, X_2, \ldots \to Y$ with probability 1. What can you say about the convergence in probability of the X_i to Y?

66. Suppose $W_1, W_2, \ldots \to W$ in probability. What can you say about the probability that W_i converges to Y?

67. Give an example of a strictly increasing sequence of random variables X_i that converge to X with probability 1.

68. Give an example of a sequence of random variables $X_1, X_2, X_3, \ldots$ that converges to X with probability 1, where $X_1, X_3, X_5, \ldots$ are all strictly smaller than X, while $X_2, X_4, X_6, \ldots$ are all strictly larger.

69. Let $t \geq 0$. Given that $\exp(tx)$ is a convex function for $t \geq 0$, if $\mathbb{E}[\exp(tX)] < \infty$, what can you say about this value and $\mathbb{E}[X]$?

70. Given that $\max(0, x)$ is a convex function, what can you say about $\mathbb{E}[\max(0, X)]$ and $\max(0, \mathbb{E}[X])$, given that both are finite?

Chapter 8

Swapping Limits and Expectation

Question of the Day

For $B_1, B_2, \ldots$ iid $\text{Bern}(1/2)$, find

$$\mathbb{E}\left(\sum_{i=1}^{\infty}(1/2)^i B_i\right).$$

Summary

- A finite set of random variables $X_1, \ldots, X_n$ is **independent** if the probability of a logical AND involving the X_i can be written as the product of the probability of the individual events.

- A sequence of random variables $X_1, X_2, X_3, \ldots$ is **independent** if any finite subset is independent.

- A sequence of random variables is **iid** (independent, identically distributed) if the sequence is independent and every one has the same distribution.

- The **extended reals** include the real numbers $\mathbb{R}$, ∞, and $-\infty$.

- You can swap limits and expectations of sequences of random variables if the random variables are monotonically increasing and bounded below. This is the **Monotonic Convergence Theorem**. (Sometimes called the **Monotone Convergence Theorem**.)

- You can swap limits and expectations when the absolute values of the random variables in the sequence has a common upper bound that is an integrable random variable. This is called the **Lebesgue Dominated Convergence Theorem** (aka the *Lebesgue Convergence Theorem* or *Dominated Convergence Theorem*.)

8.1 Limits of sequences

It is easier to work with limits with the extended reals, which include the regular real numbers denoted $\mathbb{R}$, together with the numbers $-\infty$ and ∞.

For instance,

$$\lim_{t\to\infty} t^2 = \infty$$
$$\lim_{t\to\infty} 1 - 1/t = 1$$
$$\lim_{t\to-\infty} t^3 = -\infty.$$

The limit of a real valued x_n equals L means that as you move further and further out in the sequence, the values of the $\{x_i\}$ get arbitrarily close to L. This can be written using logic notation as follows.

Definition 44
Let $x_1, x_2, \ldots$ be a sequence of real numbers. Then for $L \in \mathbb{R}$,

$$\left(\lim_{n\to\infty} x_n = L\right) = (\forall m \in \{1,2,\ldots\})(\exists N \in \{1,2,\ldots\})(\forall n \geq N)(|x_n - L| \leq 1/m),$$
$$\left(\lim_{n\to\infty} x_n = \infty\right) = (\forall m \in \{1,2,\ldots\})(\exists N \in \{1,2,\ldots\})(\forall n \geq N)(x_n > m),$$
$$\left(\lim_{n\to\infty} x_n = -\infty\right) = (\forall m \in \{1,2,\ldots\})(\exists N \in \{1,2,\ldots\})(\forall n \geq N)(x_n < -m).$$

8.2 Independence

Definition 45
A finite set of random variables $X_1, \ldots, X_n$ is **independent** if for all measurable $A_1, \ldots, A_n$ (with respect to the X_i),

$$\mathbb{P}\left(\bigvee_{i=1}^{n} X_i \in A_i\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i).$$

A sequence of random variables is independent if all finite subsets are.

Definition 46
A sequence of random variables $X_1, X_2, \ldots$ is **independent** if for all n, $X_1, \ldots, X_n$ are independent.

This allows us to define sequences that are *iid*, or *independent, identically distributed*.

Definition 47
A sequence of random variables $\{X_i\}$ is **iid**, or **independent, identically distributed** if the sequence is independent and all the random variables have the same distribution.

Example 14
Suppose that $X_1, X_2, \ldots$ are iid with distribution $\text{Bern}(1/2)$. Show that they do not converge to anything with probability 1.

8.2.0.0.1 Answer Because the X_i are discrete random variables, to converge there must be a point N such that $X_N = X_{N+1} = X_{N+2} = \cdots$. Let k be a positive integer. Then consider the chance that

$$X_N = X_{N+1} = X_{N+2} = \cdots = X_{N+k}.$$

Since the X_i are independent this is just $(1/2)^k$. If $X_N = X_{N+1} = \cdots$, then this event happens for all k. Therefore the probability of it happening is bounded above by $(1/2)^k$ for all k, but the only number in $[0, 1]$ where that is true is 0.

Hence the probability that the sequence converges is 0.

8.3 *Expected value and limits*

Now consider the following question. If X_t does converge to X as $t \to \infty$, then does

$$\mathbb{E}\left(\lim_{t\to\infty} X_t\right) = \lim_{t\to\infty} \mathbb{E}(X_t)?$$

In other words, can limits be brought inside (or outside) of expected value.

The answer is: sometimes!

Consider $U \sim \text{Unif}([0, 1])$ and $X_n = \mathbb{I}(U < 1/n)$ from earlier. Then

$$\mathbb{E}(X_n) = \mathbb{P}(U < 1/n) = 1/n,$$

and so

$$\lim_{n\to\infty} \mathbb{E}(X_n) = 0 = \mathbb{E}(0) = \mathbb{E}(X).$$

For another example, consider $W_n = n\mathbb{I}(U < 1/n)$. So if $U = 0.19248\ldots$,

$$\{W_n\} = \{1, 2, 3, 4, 5, 0, 0, 0, \ldots\},$$

so the W_n increase by 1 each time until $U \geq 1/n$, at which point they drop to 0 and stay there.

As with the $\{X_i\}$, it holds that $W_i \to 0$ with probability 1. However,

$$\mathbb{E}[W_n] = (1/n)(n) + (1 - 1/n)(0) = 1,$$

for all n. That is to say, their mean is always 1, and so

$$\lim_{n\to\infty} \mathbb{E}(W_n) = 1 \neq 0 = \mathbb{E}\left(\lim_{n\to\infty} W_n\right) = \mathbb{E}(0).$$

So extra conditions are required to ensure that limits and means can be swapped. The two most important such theorems are the Lebesgue dominated convergence theorem and the monotone convergence theorem.

8.4 The Lebesgue Dominated Convergence Theorem

One way of looking at what went wrong in the example was that the W_n are unbounded. Even though they go to 0 eventually, before that they had a positive probability of being n for any $n \in \{1, 2, \ldots\}$. The *Lebesgue dominated convergence theorem*, or often just the *dominated convergence theorem* prevents this type of problem by requiring that every random variable in the sequence be bounded above in magnitude by the *same* random variable. If this common random variable is integrable, then the limit of the mean of the sequence will equal the mean of the limit of the sequence.

Theorem 3
Lebesgue dominated convergence theorem (DCT)

Suppose $\lim_{t\to\infty} X_t = X$ with probability 1 and $|X_t| \leq Y$ for all t. Then if $\mathbb{E}[|Y|] < \infty$, then

$$\lim_{t\to\infty} \mathbb{E}[X_t] = \mathbb{E}\left[\lim_{t\to\infty} X_t\right].$$

8.5 The Monotone Convergence Theorem

An alternate explanation for what went wrong was that the W_t went from growing to shrinking back down to 0. If a nonnegative sequence is always increasing, then swapping limits and means will be okay.

Theorem 4
Monotone convergence theorem (MCT)

Suppose $0 \leq X_0 \leq X_1 \leq X_2 \leq \cdots$ with probability 1. Then $\lim_{t\to\infty} X_t$ is either a finite real number or ∞ with probability 1, and

$$\lim_{t\to\infty} \mathbb{E}[X_t] = \mathbb{E}\left[\lim_{t\to\infty} X_t\right].$$

This works even if the limit on both sides is infinity. In real analysis it is shown that any increasing sequence always converges either to a constant or to infinity.

Fact 24
Suppose $a_0 \leq a_1 \leq a_2 \leq \cdots$. Then $\lim_{i\to\infty} a_i$ is either a finite real number or ∞.

So $\lim_{t\to\infty} \mathbb{E}[X_t]$ always exists or is ∞.

The story so far is that there are two conditions that allow bringing limits in and out of a mean.

1. If the sequence is *dominated* by an integrable random variable (DCT.)

2. If the sequence is nonnegative and increasing (MCT.)

Example 15
Suppose $U \sim \text{Unif}([0,1])$ and $Y_t = U/t$. Show using the DCT that $\lim \mathbb{E}(Y_t) = \mathbb{E}(\lim Y_t)$.

Solution. Since $Y_t \in [0, 1/t]$, $Y_t \to 0$ with probability 1. Moreover, $|Y_t| \leq U$ for all $t \in \{1, 2, \ldots\}$, and U is integrable, so by the DCT,

$$\lim_{t\to\infty} \mathbb{E}(Y_t) = \mathbb{E}\left(\lim_{t\to\infty} Y_t\right).$$

Of course, in this case it was not necessary to use the DCT to show the limit:

$$\mathbb{E}(Y_t) = 1/(2t),$$

so

$$\lim \mathbb{E}(Y_t) = 0,$$

the same as $\mathbb{E}(\lim Y_t)) = \mathbb{E}(0) = 0$. Later on, more complicated examples where the DCT must be used will be given.

The MCT is commonly used to deal with series where the terms of the series are nonnegative random variables.

Example 16
Say $\{B_i\}$ is a sequence of independent r.v.'s with $B_i \sim \text{Bern}(1/2)$. Use the MCT to show that $\mathbb{E}\left(\sum_{i=1}^{\infty}(1/2)^i B_i\right) = 1/2$.

Solution. Let S_n be the partial sum of the first n terms:

$$S_n = (1/2)B_1 + (1/2)^2 B_2 + \cdots + (1/2)^n B_n.$$

Since the $B_i \geq 0$, $S_1, S_2, S_3, \ldots$ is a nonnegative, increasing sequence of random variables. Therefore the MCT applies!

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{\infty}(1/2)^i B_i\right) &= \mathbb{E}\left(\lim_{n\to\infty}\sum_{i=1}^{n}(1/2)^i B_i\right) \\
&= \mathbb{E}\left(\lim_{n\to\infty} S_n\right) \\
&= \lim_{n\to\infty} \mathbb{E}(S_n) && \text{by the MCT} \\
&= \lim_{n\to\infty} \mathbb{E}\left(\sum_{i=1}^{n}(1/2)^i B_i\right) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\left((1/2)^i B_i\right) && \text{by linearity} \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\left((1/2)^i B_i\right) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} (1/2)^{i+1} \\
&= \sum_{i=1}^{\infty} (1/2)^{i+1} \\
&= \frac{(1/2)^2}{1-1/2} && \text{by geometric series} \\
&= 1/2.
\end{aligned}$$

Problems

71. Suppose that $X_1, X_2, \ldots$ are independent and $\mathbb{P}(X_i \in A_i) = 1/i$. What is

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, X_3 \in A_3)?$$

72. Suppose that $\mathbb{P}(Y \in [0, a]) = a$ and $Y_1, Y_2, \ldots$ are iid with the same distribution as Y. Find

$$\mathbb{P}(Y_3 \in [0, 0.3], Y_6 \in [0, 0.5], Y_{13} \in [0, 0.7]).$$

73. Suppose $B_1, B_2, \ldots$ are independent and $B_i \sim \text{Bern}((1/3)^i)$. What is

$$\mathbb{E}\left[\sum_{i=1}^{\infty} B_i\right]?$$

74. Suppose that $X_1, X_2, \dots$ are random variables that are positive and $\mathbb{E}[X_i] = (1/4)^i$. What is

$$\mathbb{E}\left[\sum_{i=1}^{\infty} X_i\right].$$

75.

Suppose U is a standard uniform over $[0,1]$.

a. Find $\mathbb{E}[1/\sqrt{U}]$.

b. Find

$$\mathbb{E}\left[\lim_{n\to\infty} \frac{n\sin(U)}{1+n^2\sqrt{U}}\right].$$

76. Consider $Y \sim \text{Exp}(3)$, and $X_i = Y - 1/i$. What is

$$\mathbb{E}\left[\lim_{i\to\infty} X_i\right]?$$

77. Let T be a nonnegative random variable. The notation $\wedge$ can also be used for minimum, so $t \wedge T = \min(t, T)$. Note that $t \wedge T$ is an increasing sequence in t. When $\mathbb{P}(T < \infty) = 1$, it holds that

$$\lim_{t\to\infty} t \wedge T = T$$

since eventually t will be larger than the finite value of T with probability 1.
For T such that $\mathbb{P}(T < \infty) = 1$, what can be said about

$$\lim_{t\to\infty} \mathbb{E}[t \wedge T]?$$

78. Suppose $X \sim \text{Geo}(1/3)$. What is

$$\lim_{t\to\infty} \mathbb{E}[t \wedge X]?$$

Chapter 9

Proving Limit Theorems for Expectation

Question of the Day

Prove the monotone convergence theorem (MCT) and the dominated convergence theorem (DCT.)

Summary

- To show the MCT, break the proof into three cases

  1. All the random variables involved are integrable.

  2. At some point, there is an X_i that is not integrable. (This makes all later random variables in the sequence also not integrable.)

  3. All random variables in the sequence are integrable, but the limit is not.

- To show the DCT, first show **Fatou's Lemma**, which says that a limit infimum can be brought out of an expectation at the cost of making the result the same or larger.

- An application of the DCT is the **bounded convergence theorem** (BCT) which says if $|X| \leq M$ where M is a constant, then you can always bring

- An application of the MCT is as follows: If X is nonnegative, then

$$\lim_{m \to \infty} \mathbb{E}[X \wedge m] = \mathbb{E}[X].$$

  (Here $X \wedge m = \min(X, m)$ as usual.)

9.1 *When can you swap limits and mean?*

As a reminder, there are two commonly used sufficient (but not necessary) conditions to be able to bring limits in and out of expected value are as follows

1. **Monotonicity** For $X_t \geq 0$ and $X_0 \leq X_1 \leq X_2 \leq \cdots$ with probability 1. This is the monotone convergence theorem (MCT).

2. **Dominated** For $\mathbb{P}(\lim_{t\to\infty} X_t = X) = 1$ and $|X_t| \leq Y$ with $\mathbb{E}[|Y|] < \infty$. This is the dominated convergence theorem (DCT).

9.2 Proving the Monotone Convergence Theorem

Consider the case that $0 \leq X_1 \leq X_2 \leq \cdots$, and all the variables are integrable.

Fact 25
For $0 \leq X_1 \leq X_2 \leq \cdots$ and X integrable, where $X_n \to X$ with probability 1, it holds that

$$\lim_{n\to\infty} \mathbb{E}[X_i] = \mathbb{E}[X].$$

Proof. Since the X_i are increasing (with probability 1), for all i it must be that $X \geq X_i$ (with probability 1). Hence $\mathbb{E}[X_i] \leq \mathbb{E}[X]$ for all i, and so

$$\lim_{n\to\infty} \mathbb{E}[X_i] \leq \mathbb{E}[X].$$

What remains is to show the other direction of the inequality. To show this direction, fix $m \in \{1, 2, \ldots\}$. By the limit statement, with probability 1, there exists a T such that for all $t \geq T$,

$$|X_t - X| \leq 1/m.$$

Since the X_t are increasing towards X, it must be that

$$|X_t - X| = X - X_t \leq 1/m,$$

so

$$X \leq X_t + 1/m.$$

Therefore, for every $m \in \mathbb{Z}^+$,

$$\lim_{t\to\infty} \mathbb{E}[X] \leq \lim_{t\to\infty} \mathbb{E}[X_t + 1/m].$$

or more simply, for every $m \in \mathbb{Z}^+$,

$$\mathbb{E}[X] \leq \lim_{t\to\infty} \mathbb{E}[X_t] + 1/m$$

The only way that can happen is if

$$\lim_{n\to\infty} \mathbb{E}[X_i] \geq \mathbb{E}[X]$$

as desired. □

Now consider what happens if some of the X_i or X are not integrable.

Fact 26
For $0 \leq X_1 \leq X_2 \leq \cdots$ and $X = \lim_{n\to\infty} X_n$ where $X_n \to X$ with probability 1, and X_i not integrable, it holds that

$$\lim_{n\to\infty} \mathbb{E}[X_i] = \mathbb{E}[X].$$

Proof. For X_i not integrable, $\mathbb{E}[X_i] = \infty$. But $X \geq X_i$ as before, so also $\mathbb{E}[X] = \infty$ and

$$\lim_{n\to\infty} \mathbb{E}[X_i] = \mathbb{E}[X]$$

as before □

Fact 27
For $0 \leq X_1 \leq X_2 \leq \cdots$ integrable, $X = \lim_{n\to\infty} X_n$ not integrable, and $X_n \to X$ with probability 1, it holds that

$$\lim_{n\to\infty} \mathbb{E}[X_i] = \mathbb{E}[X].$$

Proof. The case where all the random variables are integrable is covered by the previous fact. Consider the case where all of the X_i are integrable, but X is not. Then $\mathbb{E}[X] = \infty$, and the goal becomes to show that the limit of the $\mathbb{E}[X_i]$ is infinity. That is, the goal is to show

$$(\forall M \in \mathbb{Z}^+)(\exists T \in \mathbb{Z}^+)(\forall t \geq T)(\mathbb{E}[X_t] \geq M)$$

By the definition of expected value,

$$\sup_{Y \leq X : Y \text{ simple}} \mathbb{E}[Y] = \infty.$$

Let W be a simple random variable such that $W \leq X$ and $\mathbb{E}[W] \geq M + 1$. Since $\lim_{t\to\infty} X_t = X$, there exists T such that for all $t \geq T$,

$$|X_t - X| \leq 1.$$

Since the X_t are increasing, this means

$$(|X_t - X| \leq 1) = (X - X_t \leq 1)$$

making $W - 1 \leq X - 1 \leq X_t$.

Since $M \leq \mathbb{E}[W - 1] \leq \mathbb{E}[X_t]$, that gives $\lim_{t\to\infty} \mathbb{E}[X_t] = \infty$, finishing the proof. □

9.3 *Proving the DCT*

The proof of the dominated convergence theorem will require a brief stop at a result called *Fatou's Lemma*. This Lemma states that the lim inf of a sequence of nonnegative random variables can always be brought inside expectation at the cost of decreasing the expectation.

What exactly is the lim inf? Consider the following sequence:

$$0, 1, 0, 1, 0, 1, 0, 1, \ldots.$$

This sequence does not have a limit. It will, however, have a lim inf.

9.3.1 The infimum

The *infimum* of a set of real numbers is the greatest lower bound on the set of numbers. So

$$\inf(\{1, 2, 3\}) = 1,$$

and

$$\inf([5, 10]) = 5.$$

So if the set has a minimum value, then the infimum is just the minimum value. The infimum also exists, however, for sets that do not have a minimum value. For instance,

$$\inf(\{x : 0 < x \leq 1\}) = \inf((0, 1]) = 0.$$

This is because 0 is a lower bound on $(0, 1]$, and it is also the greatest lower bound since no larger number is smaller than every number in $(0, 1]$. Formally, this can be defined as follows.

Definition 48
The **infimum** of a set of numbers A is the greatest lower bound on the numbers of A. There are three cases.

1. $A = \emptyset$. Then $\inf(A) = \infty$.
2. $(\exists b \in \mathbb{R})(A \subseteq [b, \infty))$. Then $\inf(A) = \max\{b : A \subseteq [b, \infty)\}$.
3. $(\forall b \in \mathbb{R})(\exists a \in A)(a < b)$ Then $\inf(A) = -\infty$.

Now how does bringing an infimum inside an expectation change things? It can only make things smaller.

Fact 28
Let S be a collection of nonnegative random variables. Then

$$\mathbb{E}[\inf(S)] \leq \inf\{\mathbb{E}[A] : A \in S\}.$$

Proof. Let $W \leq \inf(S)$ be simple. Then since W is a lower bound on the random variables in S, $W \leq A$ for all $A \in S$. Hence $\mathbb{E}[W] \leq \mathbb{E}[A]$ for all $A \in S$, which means (since infimum is the *greatest* lower bound), $\mathbb{E}[W] \leq \inf \mathbb{E}[S]$.

Therefore the supremum over all such $\mathbb{E}[W]$ is at most $\inf\{\mathbb{E}[A] : A \in S\}$, which is the same as saying

$$\mathbb{E}[\inf S] \leq \inf\{\mathbb{E}[A] : A \in S\}$$

□

9.3.2 The lim inf

Now, consider a sequence of values

$$x_1, x_2, \ldots$$

Then for every T, the set of values

$$\{x_T, x_{T+1}, x_{T+2}, \ldots\}$$

has an infimum. Moreover, because as T gets larger the set $\{x_t : t \geq T\}$ gets smaller, the infimum of these sets is an increasing sequence in its own right.

Real analysis can be used to show that any increasing sequence has a limit, therefore the limit of the infimum of $\{x_t : t \geq T\}$ always exists. This is called the lim inf.

Definition 49
The **infimum limit** (aka **limit inferior**) of a set of numbers $a_0, a_1, a_2, \ldots$ is

$$\liminf(a_0, a_1, \ldots) = \lim_{n\to\infty} \inf\{a_n, a_{n+1}, a_{n+2}, \ldots\}.$$

In the sequence from earlier

$$0, 1, 0, 1, 0, \ldots,$$

for all t,

$$\{x_t, x_{t+1}, \ldots\} = \{0, 1\},$$

and

$$\inf\{0, 1\} = 0$$

so the infimum sequence is just

$$0, 0, 0, \ldots$$

which has limit 0. So $\liminf 0, 1, 0, \ldots = 0$.

Fact 29
Fatou's Lemma

If $X_t \geq 0$ then

$$\liminf \mathbb{E}[X_t] \geq \mathbb{E}\left[\liminf X_t\right].$$

Proof. Let $Y_T = \inf_{t\geq T} X_t$. Then Y_T is a nonnegative increasing sequence, and so the MCT gives

$$\begin{aligned}\lim_{T\to\infty} \mathbb{E}[Y_T] &= \mathbb{E}\left[\lim_{T\to\infty} Y_T\right] \\ &= \mathbb{E}\left[\lim_{T\to\infty} \inf_{t\geq T} X_t\right] \\ &= \mathbb{E}\left[\liminf X_t\right].\end{aligned}$$

Also,

$$\begin{aligned}\mathbb{E}[Y_T] &= \mathbb{E}\left[\inf_{t\geq T} X_t\right] \\ &\leq \inf_{t\geq T} \mathbb{E}[X_t],\end{aligned}$$

so

$$\begin{aligned}\mathbb{E}\left[\liminf X_t\right] &= \lim_{T\to\infty} \mathbb{E}[Y_T] \\ &\leq \lim_{T\to\infty} \inf_{t\geq T} \mathbb{E}[X_t] \\ &= \liminf \mathbb{E}[X_t].\end{aligned}$$

□

9.3.3 The lim sup

The *limit supremum* is defined in a similar way to the limit infimum.

Definition 50
The **supremum limit** (aka **limit superior**) of a set of numbers $a_0, a_1, a_2, \ldots$ is

$$\limsup(a_0, a_1, \ldots) = \lim_{n\to\infty} \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\}.$$

For instance,

$$\limsup 0, 1, 0, 1, 0, \ldots = 1.$$

The limit inferior and limit superior are related in a simple way.

Fact 30
For a real valued sequence $a_1, a_2, \ldots$,

$$\limsup a_i = -\liminf(-a_i).$$

One way to characterize limits is that if the $\liminf$ and $\limsup$ of a sequence are equal, then that is the limit.

Fact 31
For a real valued sequence $a_0, a_1, \ldots$,

$$(\liminf a_i = \limsup a_i = L) = \left(\lim_{i\to\infty} a_i = L\right).$$

9.3.4 Proving the DCT

Another fact needed before proving the DCT is as follows.

Fact 32
If $A + B$ and A are integrable random variables, then so is B.

Proof. Note $|B| \leq |A + B| + |A|$. Taking the expectation of both sides gives

$$\mathbb{E}[|B|] \leq \mathbb{E}[|A + B|] + \mathbb{E}[|A|] < \infty,$$

finishing the proof. □

Now the DCT can be shown. Recall the DCT applies to a sequence $X_1, X_2, \ldots$ that converges to X with probability 1. If there exists an integrable Y such that for all t, $|X_t| \leq Y$, then X is also integrable, and

$$\lim_{t\to\infty} \mathbb{E}[X_t] = \mathbb{E}[X].$$

Proof. Since $|X_t| \leq Y$, it must be true that $X_t + Y \geq 0$. Applying Fatou's Lemma to $X_t + Y$ gives

$$\liminf \mathbb{E}[X_t + Y] \geq \mathbb{E}[\liminf(X_t + Y)] = \mathbb{E}[\lim(X_t + Y)] = \mathbb{E}[X + Y].$$

Note $|X+Y| \leq |X| + |Y| \leq |2Y|$, making $X+Y$ integrable. Since $X+Y$ and Y are integrable, so is X by our previous fact.

Since X is integrable, linearity gives

$$\liminf(\mathbb{E}[X_t] + \mathbb{E}[Y]) \geq \mathbb{E}[X] + \mathbb{E}[Y].$$

Since $\mathbb{E}[Y]$ is a constant, it can cancel on both sides to give $\liminf \mathbb{E}[X_t] \geq \mathbb{E}[X]$. (Note that this is just the result of Fatou's Lemma, but this holds even though the X_t could be negative.)

Repeat the whole argument with $-X_t + Y \geq 0$ to get

$$\liminf \mathbb{E}[-X_t] \geq -\mathbb{E}[X],$$

and multiplying by -1 gives

$$\limsup \mathbb{E}[X_t] \leq \mathbb{E}[X].$$

Putting these sides together gives

$$\limsup \mathbb{E}[X_t] \leq \mathbb{E}[X] \leq \liminf \mathbb{E}[X_t].$$

But

$$\liminf \mathbb{E}[X_t] \leq \limsup \mathbb{E}[X_t],$$

so everything must have the same value! □

9.3.5 Convergence in probability

The DCT above was stated for convergence with probability 1 but actually holds under the weaker condition of convergence in probability. This is more difficult to prove, however, so that proof will be deferred until later.

9.4 *Applications*

9.4.1 Bounded random variables

Often, the DCT is applied to bounded random variables.

Fact 33
Bounded convergence theorem

Suppose there exists M such that for all t, it holds that $|X_t| \leq M$. If $X_t \to X$ in probability then

$$\lim_{t\to\infty} \mathbb{E}[X_t] = \mathbb{E}\left[\lim_{t\to\infty} X_t\right] = \mathbb{E}[X].$$

Proof. This is just a special case of the DCT where the dominating random variable is a constant $Y = M$. □

9.4.2 Rounded random variables

In this section the minimum of two numbers a and b will often show up. Therefore, it will be nice to have some notation for this. Because minimum is similar to logical AND, the same symbol will be used:

$$\min(a, b) = a \wedge b.$$

In particular, with this notation

$$\mathbb{I}(p \wedge q) = \mathbb{I}(p) \wedge \mathbb{I}(q).$$

Note that for $X \geq 0$, the random variable $X \wedge m$ is bounded for any positive m. Basically any value of X that is above m is getting rounded down to m. As this bound weakens, the mean converges to the mean of the original random variable.

Fact 34
For X a nonnegative random variable,

$$\lim_{m \to \infty} \mathbb{E}[X \wedge m] = \mathbb{E}[X].$$

Proof. Note that $X \wedge m \leq X \wedge (m + 1)$, therefore, the MCT applies, giving the result. □

Problems

79.

Consider the sequence $0, 1, 0, 1, 0, 1, \ldots$.

a. What is the limit inferior of the sequence?

b. What is the limit superior of the sequence?

80.

Consider the sequence

$$1, -1/2, 1/4, -1/8, \ldots$$

a. What is the limit inferior of this sequence?

b. What is the limit superior of this sequence?

81.

Suppose that $X \sim \text{Geo}(1/2)$.

a. For $m \in \{1, 2, \ldots\}$, find

$$\mathbb{P}(X \wedge m = m)?$$

b. What is $\lim_{m \to \infty} \mathbb{E}[X \wedge m]$?

82.

Suppose that $Y \sim \text{Geo}(1/3)$.

a. For $m \in \{1, 2, \ldots\}$, find
$$\mathbb{E}[Y \wedge m].$$

b. What is $\lim_{m \to \infty} \mathbb{E}[Y \wedge m]$?

83. Suppose that you are given random variables $R_1, R_2, \ldots$ such that $\liminf R_i = X$ where $X \sim \text{Unif}([0, 4])$. Give a lower bound on
$$\liminf \mathbb{E}(R_i).$$

84. Suppose that $W_1, W_2, \ldots$ have a limit inferior that is $Y \sim \text{Geo}(0.2)$. Give a lower bound on
$$\liminf \mathbb{E}(W_i).$$

85. Suppose the S_i are random variables where $|S_i| \leq 10$ with probability 1, and $\lim S_i \sim A$ where $A \sim \text{Exp}(4)$. What can be said about $\lim \mathbb{E}(S_i)$?

86. Suppose that $C_1, C_2, \ldots$ are a sequence of random variables bounded in absolute value by 1. Furthermore, $\lim C_i = X$ where $C \sim \text{Unif}([-5, 5])$. What can be said about $\lim \mathbb{E}(C_i)$?

Chapter 10

Martingales

Question of the Day

Suppose I play a fair game where I either win or lose a dollar every play, independently, each with probability $1/2$. If I start with 15 dollars, what is the expected amount of money I have after 12 plays?

Summary

- The intuitive idea of the **conditional expectation** $\mathbb{E}[X|Y]$ is that this represents the average value of X given that Y is a known value rather than a random variable.

- Conditional expectation (like regular expectation) is a **linear operator**.

- X and Y indpendent implies that $\mathbb{E}(X|Y) = \mathbb{E}(X)$.

- X measurable with respect to Y implies $\mathbb{E}(X|Y) = X$.

- There is always a measurable function h such that $\mathbb{E}(X|Y) = h(Y)$.

- The **Fundamental Theorem of Probability** states that for any integrable X and Y, conditioning can be undone by taking the average over the conditioned variable. That is, $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$.

- A **martingale** is a fair game, where the expected value of the next value in the stream is just the previous value.

- Taking a fixed number of steps at a time in a martingale just results in a new martingale.

10.1 Intuitive notion of martingale

Martingales are the mathematical abstraction of the notion of a *fair game*, where on average players do not gain or lose money.

- If I win or lose a dollar each with probability $1/2$, I am playing a fair game.

- If I win a dollar with probability 30% and lose a dollar with probability 70%, it is an unfair game.
- If I win a dollar with probability 60% and lose a dollar with probability 40%, it is also an unfair game, even though the benefit is to me!
- Martingales are the amount of money you have at time t when playing a fair game.

In the Question of the Day, it is given that

$$\mathbb{P}(M_1 = M_0 + 1|M_0) = 1/2,\ \mathbb{P}(M_1 = M_0 - 1|M_0) = 1/2.$$

Consider the question, what is

$$\mathbb{E}[M_1 \mid M_0]?$$

In order to find the conditional expectation, just apply the regular rules for expectation, but treat M_0 as a constant rather than a random variable. There are two possible outcomes given that M_0 is a constant, $M_0 + 1$ and $M_0 - 1$. Both of these outcomes occur with probability $1/2$.

$$\begin{aligned}\mathbb{E}[M_1|M_0] &= (1/2)(M_0 + 1) + (1/2)(M_0 - 1)\\ &= (1/2)M_0 + 1/2 + (1/2)M_0 - 1/2\\ &= M_0\end{aligned}$$

So after one play of the game, the expected amount of money that the player has will be the same amount that they started with, M_0.

Now consider what happens after two plays:

$$\mathbb{E}[M_2|M_0, M_1] = (1/2)(M_1 + 1) + (1/2)(M_1 - 1) = M_1.$$

So this gives rise to the idea of a martingale. Before giving the formal definition, first some intuition.

A *martingale* $\{M_i\}$ consists of integrable random variables that form a fair game. That is to say:

1. $\mathbb{E}[|M_n|] < \infty$, and
2. for all $n > 0$, $\mathbb{E}[M_{n+1}|M_0, \ldots, M_n] = M_n$.

So is the Question of the Day game a martingale?

- Check condition one. To make this easier, let

$$D_0, D_1, D_2, \ldots \overset{\text{iid}}{\sim} \mathsf{Unif}(\{-1, 1\}),\ \text{and } M_n = 15 + \sum_{i=1}^{n} D_i.$$

- Recall that the absolute value function obeys the triangle inequality. That is, $|a+b| \leq |a| + |b|$. So

$$\begin{aligned}
\mathbb{E}[|M_n|] &= \mathbb{E}\left[\left|15 + \sum_{i=1}^{n} D_i\right|\right] \\
&\leq \mathbb{E}\left[|15| + \sum_{i=1}^{n} |D_i|\right] \\
&= \mathbb{E}\left[|15|\right] + \sum_{i=1}^{n} \mathbb{E}\left[|D_i|\right] \\
&= 15 + n < \infty.
\end{aligned}$$

- Note that it is okay for $\mathbb{E}[|M_n|]$ to increase in n, the condition is just that $\mathbb{E}[|M_n|]$ be finite, not that the limit as $n \to \infty$ be finite

- Now check condition 2:

$$\mathbb{E}[M_n | M_0, \ldots, M_{n-1}] = (1/2)(M_{n-1} + 1) + (1/2)(M_{n-1} - 1) = M_{n-1}.$$

10.2 *Properties of conditional expectation*

This second property of martingales involves conditional expectation. A formal definition will come later, but for now knowing some properties of conditional expectation will be helpful in calculations.

Fact 35
Properties of conditional expectation (assume all integrals exist and are finite):

1. Conditional expectations are linear (like regular expectation)

$$\mathbb{E}[aA + bB | C] = a\mathbb{E}[A|C] + b\mathbb{E}[B|C]$$

2. If X and Y are independent, then

$$\mathbb{E}[X|Y] = \mathbb{E}[X].$$

3. If X is a function of Y, so $X = f(Y)$, then

$$\mathbb{E}[X|Y] = f(Y).$$

4. $\mathbb{E}[X|Y]$ will always be some function of Y.

5. (Fundamental Theorem of Probability) For any X and Y

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$$

Comments

- Property 3 means that if the mean is conditioned on A, treat the random variable like a constant. For example: $\mathbb{E}[A^2|A] = A^2$.

- Property 4 means that if you condition on the value of Y, $\mathbb{E}[X \mid Y]$ can change. Another way to say that is that X depends on Y, or mathematically, $\mathbb{E}[X \mid Y] = f(Y)$ for some function f.

  For example, suppose

$$\begin{aligned} X_1, X_2 &\overset{\text{iid}}{\sim} \mathsf{Unif}(\{1,2,3,4,5,6\}) \\ S &= X_1 + X_2. \end{aligned}$$

  Then

$$\begin{aligned} \mathbb{E}[S|X_2] &= \mathbb{E}[X_1 + X_2|X_2] \\ &= \mathbb{E}[X_1|X_2] + \mathbb{E}[X_2|X_2] \end{aligned}$$

  Since X_1 and X_2 are independent, $\mathbb{E}[X_1 \mid X_2] = \mathbb{E}[X_1] = 3.5$. Also X_2 is a function of X_2, so $\mathbb{E}[X_2 \mid X_2] = X_2$. Hence

$$\mathbb{E}[S \mid X_2] = 3.5 + X_2.$$

  Note $3.5 + X_2$ is a function of X_2, the random variable that we are conditioning on.

  More generally, $\mathbb{E}[X|A_1, A_2, \ldots, A_n]$ will be a function of $A_1, A_2, \ldots, A_n$. Recall that functions might not use all of their inputs. For example, $f(x, y, z) = xy$ doesn't use z, yet is still a function of x, y, and z.

- Now Property 5 can be used to find $\mathbb{E}[S]$.

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S|X_2]] = \mathbb{E}[3.5 + X_2] = 3.5 + \mathbb{E}[X_2] = 7$$

10.3 Solving the Question of the Day

In order to solve the Question of the Day, the following fact is essential. Basically it says that if you have a fair game after one step, then you have a fair game after any fixed number of steps t.

Fact 36
For $M_0, M_1, \ldots$ a martingale and $t \geq 0$, $\mathbb{E}[M_t|M_0] = M_0$.

The basic martingale is fair after one step and the goal is to say that it is fair after t time steps. This is the ideal situation for doing a proof by induction.

Proof. Proof by induction!

For the base case where $t = 0$, $\mathbb{E}[M_0|M_0] = M_0$.

Induction hypothesis: Suppose $\mathbb{E}[M_{n-1}|M_0] = M_0$. Then

$$\begin{aligned} \mathbb{E}[M_n|M_0] &= \mathbb{E}[\mathbb{E}[M_n|M_{n-1}]|M_0] \\ &= \mathbb{E}[M_{n-1}|M_0] \end{aligned}$$

since the interior mean just took one step. But this last conditional mean is just M_0 by the induction hypothesis, which finishes the proof! □

Applying this fact to the Question of the Day gives

$$\mathbb{E}[M_{12}|M_0 = 15] = 15.$$

10.4 Stopping at a random number of steps

So if one step in the martingale is fair, then any fixed number of steps must also be fair.

A natural question to ask is what if a random number of steps is taken? If that random number does not depend on the values of the martingale, then the answer is again yes, the overall game is fair. But if the random number of steps is allowed to change based on the value of the martingale, the answer might change to no!

This is what makes working with martingales interesting. The *Optional Sampling Theorem* will be given later and provides a characterization of when the answer is still yes.

Problems

87. Suppose $\mathbb{E}(X|Y) = 3Y$ and $\mathbb{E}(W|Y) = -4Y$. What is $\mathbb{E}(2X + 5W|Y)$?

88. Suppose for $i \in \{1, 2, 3, ...\}$, $\mathbb{E}(R_i|A) = iA$. Find $\mathbb{E}(R_1 + R_2 + R_3|A)$.

89. Suppose $\mathbb{E}(W) = 4$, $\mathbb{E}(W|S) = 3/S$, and $\mathbb{E}(S|W) = 3W$. What is $\mathbb{E}(S)$?

90. Suppose $\mathbb{E}(S) = 8$, $\mathbb{E}(1/S) = 0.3$, $mean(R|S) = 2/S)$. What is $\mathbb{E}(R)$?

91.

Suppose that $\{M_i\}$ form a martingale.

a. What is $\mathbb{E}[M_5 \mid M_4]$?

b. What is $\mathbb{E}[M_5 \mid M_3, M_1]$?

c. What is $\mathbb{E}[M_5 \mid M_0]$?

92.

Let $\{W_i\}$ be a martingale. Find the following.

a. $\mathbb{E}[W_5 \mid W_4, W_3, W_2, W_1, W_0]$.

b. $\mathbb{E}[W_5 \mid W_0]$.

93. Let $\{R_i\}$ be a martingale with $R_0 = 10$. What is $\mathbb{E}[R_{15}]$?

94.

Suppose $S_0, S_1, S_2, ...$ is a martingale with $S_0 = -4$.

a. What is $\mathbb{E}[S_{10}]$?

b. What is $\mathbb{E}[S_{10} \mid S_5 = 7]$?

Chapter 11

Encoding Information

Question of the Day

Suppose that I have a radiation badge that turns black when I receive 100 or more millisieverts (mSv) of radiation. How can the information that this badge gives be represented?

Summary

- The information encoded in a particular random variable X can be represented by the **σ-algebra generated by X**, and is written $\sigma(X)$.
- If $\sigma(X) \subseteq \mathcal{F}$, say that X is measurable with respect to $\mathcal{F}$.
- A sequence $\mathcal{F}_1, \mathcal{F}_2, \ldots$ of sigma algebras is a **filtration** if

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots.$$

- For a sequence $X_0, X_1, \ldots$, the **natural filtration** (aka **adapted filtration**) is

$$\sigma(X_0), \sigma(X_0, X_1), \sigma(X_0, X_1, X_2), \ldots.$$

- Note that for $\mathcal{F}_t$, the natural filtration for $X_0, X_1, \ldots$ always has X_t being $\mathcal{F}_t$ measurable for all t.

Functions are mathematical objects that transform an input to an output. In this process, the function can keep some or all of the information in the input value, or it can destroy information.

- For $f(x) = 2x$, given $f(x)$ it is possible to figure out what x was. This function is *one-to-one.*
- For $g(x) = x^2$, given $g(x)$ it might not be possible to figure out what x was. For instance, if $g(x) = 4$, then x could be either -2 or 2. Information has been destroyed, specifically, the information about the sign of the random variable.

A measurable function f of a random variable X is also a random variable. In order to determine how much information is in the new random variable, a σ-algebra can be used.

11.1 The σ-algebra generated by a random variable

Consider the Question of the Day example.

- The total amount of radiation received is ω, a random variable.
- When measured in millisieverts, what the badge does is report $X = \mathbb{I}(\omega \geq 100)$.
- What information does X give me about ω?
- Given X, can determine if $(\omega \in [100, \infty))$ is true or false.
- Given X, can determine if $(\omega \in [0, 100))$ is true or false.
- Given X, cannot determine if $\omega \in [50, 150]$ is true or false. There is not enough information in X for that.

So how can the information content of X be encoded mathematically?

- Let $\mathcal{F}$ be those sets A such that given X, can determine if $\omega \in A$.
- Call $\mathcal{F}$ the *σ-algebra generated by* X, and write $\mathcal{F} = \sigma(X)$.

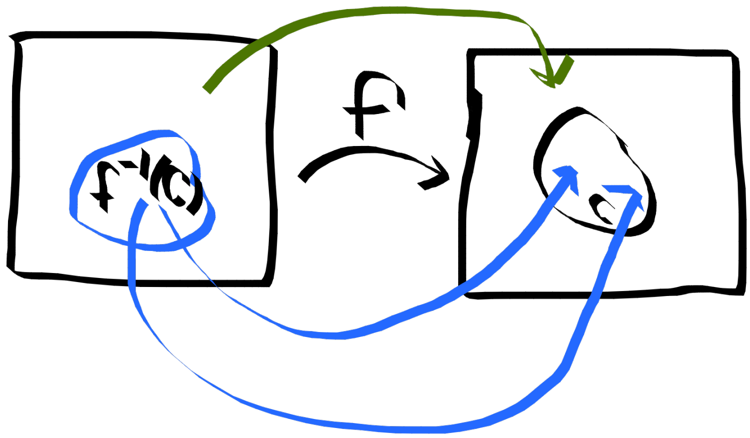
$$\sigma(X) = \mathcal{F} = \{[0, \infty), \emptyset, [100, \infty), [0, 100)\}.$$

- Then $\mathcal{F}$ will be a σ-algebra.

Formally, this idea uses the *inverse of a function applied to a set.* Recall from earlier that for a function $f : A \to B$, and $C \subseteq B$,

$$f^{-1}(C) = \{a \in A : f(a) \in C\},$$

that is, the inverse of a function applied to a set is all input values that map to the set.



Example 17
For $f(x) = 2x$,

$$f^{-1}([2,7]) = [1,3.5]$$
$$f^{-1}([-2,2]) = [-1,1]$$

Example 18
For $g(x) = x^2$,

$$g^{-1}([4,9]) = [-3,-2] \cup [2,3]$$
$$g^{-1}([-5,4]) = [-2,2]$$

For random variables that are functions, this allows us to define a σ-algebra associated with that set.

Recall also that f is *measurable* with respect to $\mathcal{F}_A$ a σ-algebra over A, and $\mathcal{F}_B$ a σ-algebra over B, if

$$(\forall B \in \mathcal{F}_B)(f^{-1}(B) \in \mathcal{F}_A).$$

Definition 51
Let ω have measurable sets $\mathcal{F}$, and $X = f(\omega)$ have measurable sets $\mathcal{F}_X$. Then

$$\sigma(X) = \{f^{-1}(A) : A \in \mathcal{F}_X\}.$$

A couple properties of the inverse function on sets will come in handy. First, the inverse of the complement of a set is exactly the complement of the inverse of the original set.

Fact 37
For a function $f : A \to B$, and $D \subseteq B$, $f^{-1}(D^C) = f^{-1}(D)^C$.

Proof. Here

$$\begin{aligned}(x \in f^{-1}(D^C)) &= (f(x) \in D^C) \\ &= \neg(f(x) \in D) \\ &= \neg(x \in f^{-1}(D)) \\ &= (x \in f^{-1}(D)^C),\end{aligned}$$

and we are done. □

Next, if you consider the inverse of a countable union of sets, that will be the same as the countable union of the inverses.

Fact 38
For a function $f : A \to B$, and $B_1, B_2, \ldots \subseteq B$,

$$f^{-1}(B_1 \cup B_2 \cup \cdots) = \cup_{i=1}^{\infty} f^{-1}(B_i).$$

Proof. Here

$$\begin{aligned}(x \in f^{-1}(\cup B_i)) &= (f(x) \in \cup B_i) \\ &= (\exists i \in \mathbb{Z}^+)(f(x) \in B_i) \\ &= (\exists i \in \mathbb{Z}^+)(x \in f^{-1}(B_i)) \\ &= \cup_{i=1}^{\infty}(x \in f^{-1}(B_i)).\end{aligned}$$

□

These two facts allow us to show that the sets $f^{-1}(A)$ actually form a σ-algebra.

Fact 39
For $X = f(\omega)$ a measurable function, $\sigma(X)$ is itself a σ-algebra.

Proof. Let f be measurable from $\mathcal{F}$ to $\mathcal{F}_X$. Then let $A \in \sigma(X)$, so $A = f^{-1}(D)$, where $D \in \mathcal{F}_X$. Then since $\mathcal{F}_X$ is a σ-algebra, $D^C \in \mathcal{F}_X$, and

$$f^{-1}(D^C) = f^{-1}(D)^C = A^C,$$

so $A^C \in \sigma(X)$ and $\sigma(X)$ is closed under complements. The fact that it is closed under countable unions is shown similarly from the previous fact. □

For example, consider $X = \mathbb{I}(\omega \leq 10)$. Then the sample space for ω is $\Omega = \mathbb{R}$. The σ-algebra for ω are the Borel sets. The state space for X is $\Omega_X = \{0, 1\}$, and the measurable sets for X are $\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$.

What values of ω map to $X = 1$? These are $X^{-1}(\{1\}) = [10, \infty)$. So that is one element of $\sigma(X)$. Since there are four sets measurable with respect to X, there are four inverse sets:

$$X^{-1}(\emptyset) = \emptyset$$
$$X^{-1}(\{0\}) = [0, 10)$$
$$X^{-1}(\{1\}) = [10, \infty)$$
$$X^{-1}(\{0, 1\}) = [0, \infty)$$

Therefore, in the radiation example,

$$\sigma(X) = \{\emptyset, [0, 10), [10, \infty), [0, \infty)\}.$$

This is in fact also a σ-algebra.

11.2 X measurable with respect to $\mathcal{F}$

Now consider what it means for a random variable X to be measurable with respect to a σ-algebra. Suppose that $X = X(\omega)$ is a measurable function of the random variable ω with measurable sets $\mathcal{F}$.

Also suppose that for any $A \in \mathcal{F}$, an oracle has told us that $\omega \in A$ is either true or false. Is that information enough to tell if $X \in B$ for any set B measurable with respect to X?

For instance, knowing if $\omega \in [0, 10)$ or $\omega \in [10, \infty)$ is enough information to tell us if $X \in \{0\}$ or $X \in \{1\}$. Mathematically, there is enough information to determine X if $\sigma(X) \subseteq \mathcal{F}$.

Definition 52
Say that X is **measurable with respect to $\mathcal{F}$** if $\sigma(X) \subseteq \mathcal{F}$.

Note that X is always measurable with respect to $\sigma(X)$, since $\sigma(X) \subseteq \sigma(X)$.

Fact 40
If X is measurable with respect to $\mathcal{F}$, then so is any measurable function of X.

Proof. Let f be a measurable function of X, and A be measurable with respect to ω. Then the inverse of $f \circ X$ over C is

$$[f \circ X]^{-1}(C) = f^{-1}(X^{-1}(C)).$$

Since X is measurable with respect to $\mathcal{F}$, $X^{-1}(C)$ is measurable with respect to X, which is a measurable set in the codomain of f, and so $f^{-1}(X^{-1}(C)$ is measurable in the domain of f. □

For instance, if X has $\sigma(X) \subseteq \mathcal{F}$, then $\sigma(X^2) \subseteq \mathcal{F}$. As seen at the beginning, X^2 holds even less information than X, so saying that $\mathcal{F}$ holds more info than X means that it also holds more info than X^2.

11.3 Adding information to a σ-algebra

The first badge gave very little information about the radiation level. That's why the σ-algebra $\sigma(X)$ only had four sets in it.

Suppose that now there is a second badge that tells me if the radiation absorbed is at least 200 mSv. This information could be held in a random variable $W = \mathbb{I}(\omega \geq 200)$. As before, ω is the true amount of radiation received.

How much information is contained in the values of both X and W? Knowing X and W, what possible intervals A could $(\omega \in A)$ be determined to be true or false?

All the sets from X are still measurable:

$$\sigma(X, W) \subseteq \{\emptyset, [0, 100), [100, \infty), [0, \infty)\},$$

but now W is giving extra information about ω:

$$\sigma(X, W) \subseteq \{\emptyset, [0, 200), [200, \infty), [0, \infty)\},$$

In addition, if $X = 1$ and $W = 0$ then

$$\omega \in [100, 200).$$

So this gives us some additional sets. Altogether, there will be 5 measurable sets, arising from

$$(X, W) \in \{(0, 0), (1, 0), (1, 1)\}$$

together with the empty set and the sample space.

So

$$\sigma(X, W) = \{\emptyset, [0, 100), [100, 200), [200, \infty), [0, \infty)\}.$$

Now our information has jumped to 5 sets!

In general, the more sets in $\omega(X)$, the more information that you have about the random variable X. In particular, for any two random variables X and W,

$$\sigma(X) \subseteq \sigma(X, W).$$

For a stochastic process, $X_1, X_2, \ldots$,

$$\sigma(X_1) \subseteq \sigma(X_1, X_2) \subseteq \sigma(X_1, X_2, X_3) \subseteq \cdots.$$

This type of increasing sequence of σ-algebras is called a *filtration*.

Definition 53
A sequence of σ-algebras $\mathcal{F}_1, \mathcal{F}_2, \ldots$ is a **filtration** if

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \cdots.$$

Definition 54
For random variables $X_1, X_2, \ldots$, the **natural** or **adapted filtration** is

$$\sigma(X_1), \sigma(X_1, X_2), \sigma(X_1, X_2, X_3), \ldots.$$

As noted earlier, adding a random variable can only increase the amount of information, or using set notation:

$$\sigma(X) \subseteq \sigma(X, Y).$$

So the adapted filtration is actually a filtration!

Problems

95. Consider random variable X in space $\Omega = \{1, 2, 3\}$ with

$$\sigma(X) = \{\emptyset, \{1, 2\}, \{3\}, \{1, 2, 3\}\}.$$

True or false: X is measurable with respect to

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}?$$

96. Suppose for random variable Y over space $\Omega = \{1, 2, 3, 4\}$, $\sigma(Y)$ contains $\{1\}$, $\{3\}$ and $\{4\}$. Suppose $\mathcal{F}$ does not contain $\{2\}$.

True or false: Y is measurable with respect to $\mathcal{F}$.

97. For $h(x) = x^2 + 1$, what is $h^{-1}([0, 4])$?

98. For $w(y) = 5y + 3$, find $w^{-1}([0, 4])$.

99. Let $\mathcal{P}(A)$ denote the *power set*, the set of all subsets of the set A. Then for $\mathcal{F}_i = \mathcal{P}(\{1, 2, \dots, i\})$, does $\mathcal{F}_i$ form a filtration?

100. Suppose $A_1, A_2, \dots$ is a sequence and $\mathcal{F}_i = \sigma(A_1, A_2, \dots, A_{2n})$. Does the $\mathcal{F}_i$ sequence form a filtration?

101. If $A \in \mathcal{F}_7$ but $A \notin \mathcal{F}_{10}$, is it possible for the $\mathcal{F}_i$ to form a filtration?

102. If $\{1, 2, 3\} \in \mathcal{F}_5$ but $\{1, 2, 3\} \notin \mathcal{F}_7$, can the $\mathcal{F}_i$ form a filtration?

103. Say $\mathcal{F}_t$ is the adapted filtration for $X_0, X_1, \dots$. Is X_7 necessarily measurable with respect to $\mathcal{F}_{10}$?

104. Say $\mathcal{F}_t$ is the adapted filtration for $Y_0, Y_1, \dots$. Is Y_{13} necessarily measurable with respect to $\mathcal{F}_{10}$?

Chapter 12

Conditional Expectation

Question of the Day

Suppose $Y \sim \text{Unif}([0, 100])$, and $X = \mathbb{I}(Y \geq 10)$. Calculate $\mathbb{E}[Y \mid X]$, and show that it obeys the formal definition of conditional expectation.

Summary

- The conditional expectation $\mathbb{E}(Y|\mathcal{F})$ is a random variable W that is measurable with respect to $\mathcal{F}$ and for any $A \in \mathcal{F}$, $\mathbb{E}(Y\mathbb{I}(W \in A)) = \mathbb{E}(W\mathbb{I}(W \in A))$.

- The conditional expectation $\mathbb{E}(Y|X)$ is a random variable W that is measurable with respect to X and for any set A measurable with respect to X, $\mathbb{E}(Y\mathbb{I}(X \in A)) = \mathbb{E}(W\mathbb{I}(X \in A))$.

- If X is measurable with respect to $\mathcal{F}$, then

$$\mathbb{E}(XY|\mathcal{F}) = X\mathbb{E}(Y|\mathcal{F}).$$

- Formally, M_n is a **martingale** with respect to filtration $\mathcal{F}_n$ if for all n three properties hold. First, M_n is measurable with respect to $\mathcal{F}_n$. Second, M_n is integrable. Third, $\mathbb{E}(M_{n+1}|\mathcal{F}) = M_n$.

Recall that information can be encoded as a σ-algebra. This encoding makes it possible to formally define conditional expectation.

Here is the idea. Given that $\mathcal{F}$ is a σ-algebra that encodes information,

$$\mathbb{E}[X \mid \mathcal{F}]$$

should be some value that depends on the information contained in $\mathcal{F}$.

For instance, if X is measurable with respect to $\mathcal{F}$, then knowing the information in $\mathcal{F}$ effectively makes X into a constant. Hence one thing that should be true about conditional expectation is that $\mathbb{E}[X \mid \mathcal{F}] = X$.

What if though, a different random variable W is involved? What should $\mathbb{E}[X|\sigma(W)]$ be? Well, it has all the information about W, so it should be some function of W. The following definition tells us what properties that $\mathbb{E}[X \mid \sigma(W)]$ (or really any $\mathbb{E}[X \mid \mathcal{F}]$ where $\mathcal{F}$ is an information encoding σ-algebra) should have.

Definition 55
The **conditional expectation of X with respect to $\mathcal{F}$** is a random variable $Y = \mathbb{E}[X|\mathcal{F}]$ that satisfies

1. Y is measurable with respect to $\mathcal{F}$.
2. For any $A \in \mathcal{F}$, $\mathbb{E}[Y\mathbb{I}(Y \in A)] = \mathbb{E}[X\mathbb{I}(Y \in A)]$.

Notation: $\mathbb{E}[X \mid W]$ should be read as $\mathbb{E}[X \mid \sigma(W)]$.

Definition 56
If $X(\omega)$ and $W(\omega)$ are two random variables, then the **conditional expectation of X given W** is

$$\mathbb{E}[X|W] = \mathbb{E}[X|\sigma(W)].$$

Some things to note about the definition of conditional expectation.

- Like the definition of expectation in general, this definition is nonconstructive. It does not tell us *how* to find conditional expectations, only what properties it has.
- To actually find $\mathbb{E}[X|Y]$, use the rules from earlier.
- This definition is solely for proving facts about conditional expectation.

12.1 *Pulling random variables out of conditional expectation*

Here is one more useful rule about conditional expectation.

Fact 41
If X is measurable with respect to $\mathcal{F}$, then

$$\mathbb{E}[XY|\mathcal{F}] = X\mathbb{E}[Y|\mathcal{F}].$$

Again, the intuition is that if the expectation is conditioned on the information in $\mathcal{F}$, and X is measurable with respect to $\mathcal{F}$, then in the context of the conditioning, X can be treated like a constant. This allows pulling X out of the expectation.

12.1.1 An example of conditional expectation

Now consider the following example where a conditional expectation is found, and then it is verified to obey the definition.

Let $Y \sim \text{Unif}([0, 100])$ and $X = \mathbb{I}(Y \geq 10)$. What is $\mathbb{E}[Y|X]$?

Before trying the formal definition, try to build this conditional expectation by considering what X tells us about the random variable Y. When $X = 1$, $Y \in [10, 100]$, and a useful fact about uniforms is that if $A \subseteq B$ and $Y \sim \text{Unif}(B)$, then $[Y|Y \in A] \sim \text{Unif}(A)$.

This means that $[Y|X = 1] \sim \text{Unif}([10, 100])$, which means $\mathbb{E}[Y|X = 1] = (10 + 100)/2 = 55$.

On the other hand, $[Y|X = 0] \sim \text{Unif}([0, 10))$, which means $\mathbb{E}(Y|X = 0) = (0 + 10)/2 = 5$.

So the conditional expectation is a function that when 0 is plugged in gives 5 and when 1 is plugged in gives 55. One such function is $h(s) = 5 + 50s$, which means

$$\mathbb{E}(Y|X) = 5 + 50X.$$

Now check that this satisfies the definition of conditional expectation. First, $5 + 50X$ is a function of X, so it is measurable with respect to X.

The second part of the definition is that $\mathbb{E}[(5 + 50X)\mathbb{I}(X \in A)$ should equal $\mathbb{E}[Y\mathbb{I}(X \in A)]$ for all sets A measurable with respect to X.

First, for what sets are X measurable? Since X is either 0 or 1, this is

$$\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}.$$

Now

$$\mathbb{I}(X \in \emptyset) = 0$$

always, so trivially

$$\mathbb{E}(Y\mathbb{I}(X \in \emptyset)) = 0 = \mathbb{E}((5 + 50X)\mathbb{I}(X \in \emptyset)).$$

Our next measurable set is $\{0\}$, and when $X = 0$ the expression $5 + 50X = 5$. Also recall that $\mathbb{E}(\mathbb{I}(p)) = \mathbb{P}(p)$. Hence

$$\mathbb{E}[(5 + 50X)\mathbb{I}(X = 0)] = \mathbb{E}[5\mathbb{I}(X = 0)] = 5\mathbb{P}(X = 0) = 5\mathbb{P}(Y < 10) = 5(10/100) = 0.5.$$

The density of $Y \sim \text{Unif}([0, 100])$ is $(1/100)\mathbb{I}(y \in [0, 100])$. Hence

$$\begin{aligned}
\mathbb{E}[Y\mathbb{I}(X = 0)] &= \mathbb{E}[Y\mathbb{I}(Y < 10)] \\
&= \int y\mathbb{I}(y < 10)(1/100)\mathbb{I}(y \in [0, 100])\, dy \\
&= \int_0^{10} y/100\, dy \\
&= (y^2/2)/100|_0^{10} \\
&= 0.5.
\end{aligned}$$

. Success!

Now for $\{1\}$. Then

$$\mathbb{E}[(5 + 50X)\mathbb{I}(X = 1)] = \mathbb{I}(55\mathbb{I}(X = 1)) = 55(90/100) = 49.5.$$

Also,

$$\begin{aligned}\mathbb{E}[Y\mathbb{I}(X=1)] &= \mathbb{E}[Y\mathbb{I}(Y \geq 10)] \\ &= \int y\mathbb{I}(y \geq 10)(1/100)\mathbb{I}(y \in [0,100])\, dy \\ &= \int_{10}^{100} y/100\, dy \\ &= (y^2/2)/100|_0^{10} \\ &= 50 - 0.5 = 49.5.\end{aligned}$$

.

Since $\mathbb{I}(X \in \{0,1\}) = \mathbb{I}(X=0) + \mathbb{I}(X=1)$ the previous work is enough to show that the conditional expectation definition holds for all sets measurable with respect to X.

Like the formal definition of expectation, the formal definition of conditional expectation is really only there for proving theorems and properties, it is not meant as a practical tool for calculating expectation.

12.2 Martingales

Now that filtrations and conditional expectation are formally defined, a martingale can be as well.

Definition 57
A stochastic process $M_0, M_1, \ldots$ is a **martingale** with respect to a filtration $\{\mathcal{F}_n\}$ if for all n:

1. M_n is measurable with respect to $\mathcal{F}_n$.
2. $\mathbb{E}[|M_n|] < \infty$
3. $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$.

Example 19
Let $D_1, D_2, D_3, \ldots \stackrel{\text{iid}}{\sim} \mathsf{Unif}(\{-1,1\})$, and $\mathcal{F}_n = \sigma(D_1, D_2, \ldots, D_n)$.

Then set $M_n = \sum_{i=1}^n D_i$. (Empty sums equal 0, so $M_0 = 0$.) Is M_n a martingale?

Yes! Here is why.

Let $n > 0$. Then $D_1, \ldots, D_n$ are measurable with respect to $\mathcal{F}_n = \sigma(D_1, \ldots, D_n)$, and M_n is a function of $D_1, \ldots, D_n$, so it is $\mathcal{F}_n$ measurable.

Next $M_n \in [-n, n]$, so $|M_n| \leq n$ and $\mathbb{E}[|M_n|] \leq n < \infty$.

Also,

$$\begin{aligned}\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[M_{n+1} \mid D_1, \ldots, D_n] \\ &= \mathbb{E}[D_{n+1} + D_1 + \cdots + D_n \mid D_1, \ldots, D_n] \\ &= D_1 + \cdots + D_n + \mathbb{E}[D_{n+1}] \\ &= M_n.\end{aligned}$$

Problems

105. Suppose $\mathbb{E}(X|R) = R^2$. Also, $\mathbb{P}(R^2 \in [0,1] = 0.3)$ and $\mathbb{E}(R^2\mathbb{I}(R^2 \in [0,1]) = 0.1)$. What is $\mathbb{E}(X\mathbb{I}(R^2 \in [0,1]))$?

106. Suppose that $\mathbb{E}(W|S) = 2S$ and the set $A = \{0,1,2\}$ is measurable with respect to S. Moreover, $mean(2S\mathbb{I}(S \in A)) = 45$. What is $\mathbb{E}(W\mathbb{I}(S \in A))$?

107. Suppose $\mathbb{E}[X \mid Y] = 4.2Y$. Write $\mathbb{E}[X]$ in terms of the mean of Y.

108. Suppose $\mathbb{E}[W \mid R] = 17 - R$. Write $\mathbb{E}[W]$ in terms of $\mathbb{E}[R]$.

109. Suppose $U \sim \mathsf{Unif}([0,1])$, and $W = \mathbb{I}(U \geq 0.2)$. Find

$$\mathbb{E}[U \mid W].$$

Prove that your answer is correct.

110. Suppose $U \sim \mathsf{Unif}([-1,1])$, and $T = \mathbb{I}(U \geq 0)$. Find $\mathbb{E}[U \mid T]$, and prove that your answer is correct.

111. Let $U_1, U_2, U_3, \ldots$ be iid uniform over $[-1,1]$. Set

$$M_i = \sum_{i=1}^{i} U_i.$$

Prove that $\{M_i\}$ is a martingale with respect to the adapted filtration.

112. Suppose that $B_1, B_2, \ldots$ are iid Bernoulli with mean 0.6. Show that

$$M_t = \left[\sum_{i=1}^{t} B_i\right] - 0.6t$$

is a martingale with respect to the adapted filtration.

Chapter 13

Stopping Times

Question of the Day

Suppose a player is playing a fair game where they either win or lose a dollar on each (independent) play with probability $1/2$. If they start with 3 dollars, and quit when they hit 0 or 10 dollars, what is the chance that the player walks away with 10 dollars?

Summary

- A **stopping time** T for a process with filtration $\mathcal{F}_n$ is a random variable such that for every positive integer n, the event that $T \leq n$ is $\mathcal{F}_n$ measurable.
- For a process M_t with stopping time T, call $M_{t \wedge T} = M_{\min(t,T)}$ the **stopped process**.
- If $\mathbb{P}(T < \infty) = 1$, then $\lim_{t \to \infty} M_{t \wedge T} = M_T$.

13.1 *What is a stopping time?*

13.1.0.1 Stopping times

In the Question of the Day, an example of the sequence of money the player has is

$$3, 2, 3, 4, 3, 2, 1, 0.$$

Of course if the player did not stop at 0 or 10, the sequence would continue after that.

$$3, 2, 3, 4, 3, 2, 1, 0, -1, -2, -1, 0, 1, 2, 1, 3, \ldots$$

Call the money that the player has after t plays of the game M_t. Then $M_0 = 3$ and $T = 7$ in the example sequence above. The idea is that T will be the smallest value such that M_T is either 0 or 10. This can be expressed in notation using an infinimum.

The set

$$\{t : M_t \in \{0, 10\}\}$$

in the example above is $\{7, 11, \ldots\}$ because it is known that $M_7 = 0$ and $M_{11} = 0$ as well. The infimum picks out the smallest value in this set.

$$T = \inf\{t : M_t \in \{0, 10\}\}.$$

This random variable T turns out to be an example of a *stopping time*. Intuitively, a stopping time is a condition under which a player could decide to stop playing a game.

Mathematically, this decision must be made using only information from the present and the past. That is, if the player has stopping at or before time n, this should be discoverable using only the information up to time n. Typically this is given through a filtration.

Definition 58
T is a **stopping time** with respect to a filtration $\{\mathcal{F}_n\}$ if for all n, the event $\{T \leq n\}$ is in $\mathcal{F}_n$.

For the example above, using the information in $(M_0, M_1, \ldots, M_n)$ the event $T \leq n$ is measurable with respect to $\sigma(M_0, M_1, \ldots, M_n)$ since

$$(T \leq n) = \bigvee_{i=1}^{n} (M_i \in \{0, 10\}).$$

13.2 Using stopping times with martingales

Recall that for a martingale $\{M_t\}$, for any fixed t it holds that

$$\mathbb{E}[M_t | M_0] = M_0,$$

which implies by taking the expected value again

$$\mathbb{E}[M_t] = \mathbb{E}[M_0].$$

A natural question is: does this hold for stopping times as well? For any stopping time T, is it true that $\mathbb{E}[M_T] = \mathbb{E}[M_0]$?

For some (but not all!) processes the answer is yes. For the process in the Question of the Day, to show that the answer is yes requires the *stopped process*.

Recall there was earlier an example of the Question of the Day process:

$$3, 2, 3, 4, 3, 2, 1, 0, -1, -2, -1, 0, 1, 2, 1, 3, \ldots$$

Here $T = 7$, because that was the first time this sequence hit either 0 or 10. The stopped process is the same as the original process, except when it hits 0 or 10, it just stops and repeats the same number:

$$3, 2, 3, 4, 3, 2, 1, 0, 0, 0, 0, 0, \ldots.$$

If the Question of the Day sequence had been

$$3, 4, 5, 6, 5, 6, 7, 6, 7, 8, 9, 10, 11, 10, 11, 12, \ldots$$

then the stopped process would be

$$3, 4, 5, 6, 5, 6, 7, 6, 7, 8, 9, 10, 10, 10, 10, 10 \ldots$$

Using the notation $a \wedge b = \min(a, b)$, the stopped process can be defined as follows.

Definition 59
If $\{M_t\}$ is a stochastic process with stopping time T, call

$$M'_t = M_{t \wedge T}$$

the **stopped process**.

For example, if $T = 3$, then $M_{0\wedge T} = M_0$, $M_{1\wedge T} = M_1$, $M_{2\wedge T} = M_2$, $M_{3\wedge T} = M_3$, $M_{4\wedge T} = M_T$, $M_{5\wedge T} = M_T$ and the stopped process gets stuck at the value M_T.

In the Question of the Day, M_T is either 0 or 10, because of the condition of the stopping time.

Because the value of the stopped process is constant after a while, it is also a martingale with respect to the same filtration.

Fact 42
If M_t is a martingale and T is a stopping time wrt to the same filtration, the stopped process $M'_t = M_{t\wedge T}$ is also a martingale.

Proof. For M'_t to be a martingale, we must show three things for all $t \geq 0$:

1. M'_t is $\mathcal{F}_t$ measurable,
2. $\mathbb{E}[|M'_t|] < \infty$, and
3. $\mathbb{E}[M'_{t+1}|\mathcal{F}_t] = M'_t$.

Consider these in order.

1. Let $t \geq 0$. Then $M'_t = M_{t\wedge T} = M_t \cdot \mathbb{I}(t < T) + M_T \cdot \mathbb{I}(T \leq t)$. All the information up to time t is enough to determine M_t, $\mathbb{I}(T \leq t)$, and $\mathbb{I}(T > t)$. When $T \leq t$, M_T is one of $M_0, \ldots, M_t$, and so it determined as well by $\mathcal{F}_t$. Hence M'_t is $\mathcal{F}_t$ measurable.

2. Next,
$$\begin{aligned}\mathbb{E}[|M'_t|] &= \mathbb{E}[|M_t|\mathbb{I}(t < T)|] + \mathbb{E}[|M_T|\mathbb{I}(T \leq t)] \\ &\leq \mathbb{E}[|M_t|] + \mathbb{E}[|M_0| + |M_1| + |M_2| + |M_3| + \cdots + |M_t|]\end{aligned}$$
since if $T \leq t$ then M_T has to equal one of $M_0, \ldots, M_t$. But
$$\mathbb{E}[|M_t|] + \mathbb{E}[|M_0| + |M_1| + |M_2| + |M_3| + \cdots + |M_t|] \leq \mathbb{E}[|M_t|] + \sum_{i=0}^{t} \mathbb{E}[|M_i|],$$
and all of these are finite since M_t is a martingale, so their sum is finite as well.

3. Now to show $\mathbb{E}[M'_{t+1}|\mathcal{F}_t] = M'_t$. Use that
$$M'_{t+1} = M_{(t+1)\wedge T} = M_{t+1}\mathbb{I}(T \geq t+1) + M_T\mathbb{I}(T \leq t).$$
$$\mathbb{E}[M'_{t+1}|\mathcal{F}_t] = \mathbb{E}[M_{t+1}\mathbb{I}(T \geq t+1) + M_T\mathbb{I}(T \leq t)|\mathcal{F}_t].$$

   Given the info up to time t, $\mathbb{I}(\{T \leq t\}$ is $\mathcal{F}_t$-measurable, which means $1 - \mathbb{I}(T \leq t) = \mathbb{I}(T \geq t+1)$ is also $\mathcal{F}_t$-measurable.

Also,

$$M_T \mathbb{I}(T \leq t) \in \{0, M_0, M_1, M_2, \dots, M_t\},$$

so is also $\mathcal{F}_t$-measurable.

The only piece which is not $\mathcal{F}_t$-measurable is M_{t+1}. So

$$\begin{aligned}
\mathbb{E}[M'_{t+1}|\mathcal{F}_t] &= \mathbb{I}(T \geq t+1)\mathbb{E}[M_{t+1}|\mathcal{F}_t] + M_T \mathbb{I}(T \leq t) \\
&= M_t \mathbb{I}(T \geq t+1) + M_T \mathbb{I}(T \leq t) \\
&= M_{t \wedge T},
\end{aligned}$$

which completes the proof.

□

Now the stopped process being a martingale gives

$$(\forall t)(\mathbb{E}[M'_t] = \mathbb{E}[M_0]).$$

So a limit can be taken:

$$\lim_{t\to\infty} \mathbb{E}[M_{t\wedge T}] = \mathbb{E}[M_0].$$

Because of the way that T was defined,

$$M_{t\wedge T} \in \{0, 1, 2, \dots, 10\},$$

and so $|M_{t\wedge T}| \leq 10$.

That means the Dominated Convergence Theorem (DCT) can be used to allow the limit to be brought inside the expectation.

$$\mathbb{E}[\lim_{t\to\infty} M_{t\wedge T}] = \mathbb{E}[M_0].$$

This brings us to the problem of what exactly is

$$\lim_{t\to\infty} M_{t\wedge T}$$

If $\mathbb{P}(T < \infty) = 1$, then sooner or later t will be greater than T, and $t \wedge T = T$. Otherwise, the limit could be anything.

So consider $\mathbb{P}(T < \infty)$. After any ten steps, there is a $(1/2)^{10}$ chance of winning 10 games in a row, therefore $\mathbb{P}(T > 10) \leq 1 - (1/2)^{10}$. The next ten steps then give (because they are independent of the first)

$$\mathbb{P}(T > 20) \leq [1 - (1/2)^{10}][1 - (1/2)^{10}] = [1 - (1/2)^{10}]^2.$$

Generalizing gives

$$\mathbb{P}(T > 10k) \leq [1 - (1/2)^{10}]^k,$$

so taking the limit as k goes to infinity gives

$$\mathbb{P}(T = \infty) \leq \lim_{k\to\infty} [1 - (1/2)^{10}]^k = 0.$$

Question of the Day

Now put these steps together followed by one last calculation.

1. $\mathbb{P}(T < \infty) = 1$ since for T to be infinite, every ten steps would have to avoid being ten wins, but ten wins happens with positive probability.

2. So $\lim_{t\to\infty} M_{t\wedge T} = M_T$.

3. $M_{t\wedge T}$ is a martingale, so for all t it holds that $\mathbb{E}[M_{t\wedge T}] = \mathbb{E}[M_0]$, and $\lim_{t\to\infty} \mathbb{E}[M_{t\wedge T}] = \mathbb{E}[M_0]$.

4. $\lim_{t\to\infty} \mathbb{E}[M_{t\wedge T}] = \mathbb{E}[\lim_{t\to\infty} \mathbb{E}[M_{t\wedge T}]]$ because $|M_{t\wedge T}| \leq 10$ so the DCT applies.

5. Hence
$$\mathbb{E}[M_T] = \mathbb{E}[M_0].$$

6. Note
$$\mathbb{E}[M_T] = \mathbb{P}(M_T = 10)(10) + (1 - \mathbb{P}(M_T = 10))(0) = 3,$$
so $\mathbb{P}(M_T = 10) = \boxed{0.3000}$.

13.3 *Does $\mathbb{E}[M_T]$ always equal $\mathbb{E}[M_0]$?*

It turns out not to be true that $\mathbb{E}[M_W] = \mathbb{E}[M_0]$ for all stopping times. Consider the Question of the Day, but set
$$W = \inf\{t : M_t = 4\}.$$

It turns out that $\mathbb{P}(W < \infty) = 1$ like before, although it is more difficult to show, so this will be shown later in the text. For now, just assume this is true. Then
$$\mathbb{E}[\lim_{t\to\infty} M_{t\wedge W}] = 4,$$
but $\mathbb{E}[M_0] = 3$, so they are not equal!

The moral of this example is that it is *not* always possible to say that $\mathbb{E}[M_W] = \mathbb{E}[M_0]$ for a stopping time W.

Problems

113. For a sequence $a_i = (-1)^i$, find
$$\{i \in \{0, 1, 2, 3, \ldots\} : a_i = 1\}.$$

114. For a sequence $b_i = i + (-1)^i$, find
$$\inf\{j : b_j = 10\}.$$

115. Suppose a player is playing a fair game where they either win or lose a dollar on each (independent) play with probability $1/2$. If they start with 4 dollars, and quit when they hit 0 or 16 dollars, what is the chance that the player walks away with 16 dollars?

116. Suppose a player is playing a fair game where they either win or lose two dollars on each (independent) play with probability $1/2$. If they start with 4 dollars, and quit when they hit 0 or 16 dollars, what is the chance that the player walks away with 16 dollars?

117.

Consider a stochastic process where each $X_t \in \{0, 1, 2\}$, with outcome

$$X_1, X_2, \ldots = 1, 1, 2, 1, 2, 2, 2, 2, 1, 0, 1, 2, 0, 0, 2, 1, \ldots$$

a. Let $T_1 = \inf\{t : X_t = 0\}$. What is T_1 for the outcome given above?

b. Let $T_2 = \inf\{t : X_t = 3\}$. What is T_2 for the outcome given above?

118. Continuing the last problem, suppose $T_3 = \inf\{t : X_t \in \{0, 3\}\}$. What is T_3 for the example sequence?

119. Suppose $M_0, M_1, M_2, \ldots = 3, 5, 2, 5, 3, 4, 1, 3, 4, -4, 5, \ldots$, and $T = 3$. What does the first eleven terms of the sequence $M_{t \wedge T}$ look like?

120. Suppose $M_0, M_1, M_2, \ldots = 3, 5, 2, 5, 3, 4, 1, 3, 4, -4, 5, \ldots$, and $W = \inf\{t : M_t \in \{1, 4\}\}$. What does the first eleven terms of the sequence $M_{t \wedge W}$ look like?

121. Consider two players playing a fair game where one player gives the other player $1 with probability $1/2$. Each play of the game is independent of the previous games. Let M_t be the amount of money owned by the first player after t steps in the game.

Then M_t forms a martingale with respect to the natural filtration. Let $T = \inf\{t : M_t \in \{0, 20\}\}$. If $\mathbb{P}(M_0 = 3) = 1$, prove that $\mathbb{P}(T < \infty) = 1$.

122.

Continuing the last problem, do the following.

a. Show that $\mathbb{E}[M_T] = 3$.

b. Find $\mathbb{P}(M_T = 20)$.

Chapter 14

Artificial Martingales

Question of the Day

Suppose that a player is playing an unfair game where they win a dollar with probability 45% and lose a dollar with probability 55%. If they start with 3 dollars, and quit when they hit 0 or 10 dollars, what is the chance that they walk away with 10 dollars?

Summary

- If M_t is not a martingale, often there is a value r such that $N_t = r^{M_t}$ is a martingale. Call this a **multiplicative martingale**.

- Similarly, it might be true that $W_t = M_t + \alpha t$ is a martingale. Call this an **additive martingale**.

- Collectively, call these **artificial martingales**. Once an artificial martingale is in place, it can be used to answer questions about the original process.

Last time the notion of a stopping time was introduced to make solving this type of problem easier. Recall that T is a stopping time with respect to filtration $\mathcal{F}_n$ if for every n, the event $T \leq n$ is measurable with respect to $\mathcal{F}_n$.

Note that stopping times exist for stochastic processes that are not martingales. In the Question of the Day, if M_t is the amount of money that the player has after playing t games, then

$$T = \inf\{t : M_t \in \{0, 10\}\}$$

is a stopping time with respect to the adapted filtration $\mathcal{F}_n = \sigma(M_0, \ldots, M_n)$.

Unfortunately, M_t is not a martingale, since

$$\mathbb{E}[M_{t+1} | M_0, \ldots M_t] = (0.45)(M_t + 1) + (0.55)(M_t - 1) = M_t - 0.1.$$

So it is not true that $\mathbb{E}[M_t] = \mathbb{E}[M_0]$ for $t > 0$.

14.1 Multiplicative martingales

To solve this issue, use M_t to create an *artificial martingale* that is a function of the original martingale.

First, let

$$D_1, D_2, D_3, \ldots \sim D$$

be iid, where

$$\mathbb{P}(D = 1) = 0.45,\ \mathbb{P}(D = -1) = 0.55.$$

Then write

$$M_t = M_0 + \sum_{i=1}^{t} D_i,$$

and use the filtration

$$\mathcal{F}_t = \sigma(M_0, D_1, D_2, \ldots, D_t).$$

Then each M_t is $\mathcal{F}_t$ measurable.

As noted before, M_t is not a martingale.

However, consider

$$N_t = (0.55/0.45)^{M_t}.$$

This is a martingale with respect to $\mathcal{F}_n$.

Check the three properties.

1. N_t is a function of M_t, so is $\mathcal{F}_n$ measurable
2. $|N_t| \leq (0.55/0.45)^t$, so N_t is a bounded random variable and so is integrable.
3. The conditional expectation of N_{t+1} given the information up to time t is

$$\begin{aligned}
\mathbb{E}[N_{t+1}|\mathcal{F}_t] &= \mathbb{E}[(0.55/0.45)^{M_0+\sum_{i=1}^t D_i} \cdot (0.55/0.45)^{D_{t+1}}|M_0, D_1, \ldots, D_t] \\
&= (0.55/0.45)^{M_0+\sum_{i=1}^t D_i}\mathbb{E}[(0.55/0.45)^{D_{t+1}}] \\
&= N_t[(0.45)(0.55/0.45)^1 + (0.55)(0.55/0.45)^{-1}] \\
&= N_t[0.55 + 0.45] \\
&= N_t
\end{aligned}$$

14.2 Solving the Question of the Day

With this new multiplicative martingale N_t, it is possible to solve the Question of the Day.

The first step is to show that $\mathbb{P}(T < \infty) = 1$. Note that in the first 10 games, there is a $(1/2)^{10}$ chance that they are all losses for the player, resulting in $T \leq 10$. This is true for the second 10 games, the third, and so on, giving for any positive integer k,

$$\mathbb{P}(T > 10k) \leq [1 - (1/2)^{10}]^k.$$

Taking the limit as k goes to infinity gives

$$\mathbb{P}(T = \infty) = \lim_{k\to\infty} [1 - (1/2)^{10}]^k = 0.$$

The second step is to use the stopped process

$$N_{t\wedge T}$$

which is a martingale since N_t is a martingale. Hence for all t,

$$\mathbb{E}[N_{t\wedge T}] = \mathbb{E}[N_0].$$

Since this is true for all t it is true in the limit.

$$\lim_{t\to\infty} \mathbb{E}[N_{t\wedge T}] = \mathbb{E}[N_0].$$

Since $|N_{t\wedge T}| \leq (0.55/0.45)^{10}$, the dominated convergence theorem can be used to bring the limit inside, giving

$$\begin{aligned}\mathbb{E}[N_0] &= \lim_{t\to\infty} \mathbb{E}[N_{t\wedge T}] \\ &= \mathbb{E}\left[\lim_{t\to\infty} \mathbb{E}[N_{t\wedge T}]\right] \\ &= \mathbb{E}[N_T]\end{aligned}$$

since $\mathbb{P}(T < \infty) = 1$.

That means

$$(0.55/0.45)^3 = (0.55/0.45)^{10}\mathbb{P}(N_T = 10) + (0.55/0.45)^0[1 - \mathbb{P}(N_T = 10)]$$

which gives

$$\mathbb{P}(N_T = 10) = \frac{(0.55/0.45)^3 - (0.55/0.45)^0}{(0.55/0.45)^{10} - (0.55/0.45)^0} = \boxed{0.1282\ldots}.$$

14.3 The Gambler's Ruin

Suppose instead of quitting at 0 or 10, the player quits at 0 or 20. Then the new probability of winning would be

$$\frac{(0.55/0.45)^3 - (0.55/0.45)^0}{(0.55/0.45)^{20} - (0.55/0.45)^0} = \boxed{0.01519\ldots}.$$

The only place the 20 appears is in the exponential of the denominator, so asymptotically the probability of winning more than 0 is going down exponentially in the target value.

Suppose the probability parameter of the problem was changed slightly, so that the player had only a 40% chance of winning and a 60% chance of losing. Then following the same path as before would give

$$\frac{(0.6/0.4)^3 - (0.6/0.4)^0}{(0.6/0.4)^{10} - (0.6/0.4)^0} = 0.04191\ldots.$$

Because of this exponential decay in the win probability, and the sensitivity of the result to changes in the probability of winning, this process is often called the *Gambler's Ruin*.

14.4 *Additive artificial martingales*

In the last section, an artificial martingale was created using multiplication to find the probabilities of the outcomes. By using addition instead, it is possible to find the expected time needed to reach the outcome.

Going back to the original Question of the Day, what is $\mathbb{E}[T]$?

To solve this problem, create an artificial additive martingale:

$$R_t = M_t + 0.1t = M_0 + \sum_{i=1}^{t} D_i + 0.1t.$$

Since R_t is a function of M_t, it is still $\mathcal{F}_t$ measurable.

Also, $|R_t| \leq t + 0.1t = 1.1t$, so R_t is integrable.

Finally,

$$\begin{aligned}
\mathbb{E}[R_{t+1}|\mathcal{F}_t] &= \mathbb{E}[M_{t+1} + 0.1(t+1)|\mathcal{F}_t] \\
&= (0.45)(M_t + 1) + (0.55)(M_t - 1) + 0.1t + 0.1 \\
&= M_t + 0.1t \\
&= R_t
\end{aligned}$$

so R_t is a martingale.

Hence for all t, $\mathbb{E}[R_{t\wedge T}] = \mathbb{E}[R_0]$, so

$$\begin{aligned}
\mathbb{E}[R_0] &= \lim_{t\to\infty} \mathbb{E}[R_{t\wedge T}] \\
&= \lim_{t\to\infty} \mathbb{E}[M_{t\wedge T}] + \lim_{t\to\infty} \mathbb{E}[0.1(t \wedge T)]
\end{aligned}$$

Note $|M_{t\wedge T}| \leq 10$, so the DCT can be used to bring the limit inside the expectation. The value of $t \wedge T$ is monotonically increasing in t, so the MCT can be used to bring that limit inside.

Since $\mathbb{E}[R_0] = \mathbb{E}[M_0] + 0.1(0) = 3$,

$$\begin{aligned}
3 &= \mathbb{E}\left[\lim_{t\to\infty} M_{t\wedge T}\right] + \mathbb{E}\left[\lim_{t\to\infty} 0.1(t \wedge T)\right] \\
&= \mathbb{E}[M_T] + 0.1\mathbb{E}[T],
\end{aligned}$$

so $\mathbb{E}[T] = (3 - \mathbb{E}[M_T])/0.1$.

The mean of M_T can be found from the probability that $M_T = 10$, so

$$\mathbb{E}[M_T|M_0 = 3] = 10\frac{(0.55/0.45)^3 - (0.55/0.45)^0}{(0.55/0.45)^{10} - (0.55/0.45)^0}.$$

This gives

$$\mathbb{E}[T] = \boxed{17.17\ldots}.$$

Problems

123. Suppose that I play an unfair game where I have a 60% chance of winning \$1 and a 40% of losing \$1. I start with \$10, and quit when I reach \$20 or \$0. What is the probability that when I quit I have \$0?

124. Suppose that I play an unfair game where I have a 65% chance of winning \$1 and a 35% of losing \$1. I start with \$10, and quit when I reach \$20 or \$0. What is the probability that when I quit I have \$0?

125. Suppose that I play an unfair game where I have a 60% chance of winning \$1 and a 40% of losing \$1. I start with \$10, and quit when I reach \$20 or \$0. What is the expected number of steps until I reach \$0 or \$20?

126. Suppose that I play an unfair game where I have a 55% chance of winning \$1 and a 45% of losing \$1. I start with \$10, and quit when I reach \$20 or \$0. What is the expected number of steps I need to take before quitting.

127. Suppose M_t is a process where $M_0 = 0$ and

$$\begin{aligned} \mathbb{P}(M_{t+1} = M_t + 1 | \mathcal{F}_t) &= 0.30 \\ \mathbb{P}(M_{t+1} = M_t | \mathcal{F}_t) &= 0.10 \\ \mathbb{P}(M_{t+1} = M_t - 1 | \mathcal{F}_t) &= 0.60 \end{aligned}$$

Find a value $r \neq 1$ such that $N_t = r^{M_t}$ is a martingale with respect to the natural filtration generated by M_t.

128. Suppose M_t is a process where $M_0 = 0$ and

$$\begin{aligned} \mathbb{P}(M_{t+1} = M_t + 1 | \mathcal{F}_t) &= 0.40 \\ \mathbb{P}(M_{t+1} = M_t | \mathcal{F}_t) &= 0.30 \\ \mathbb{P}(M_{t+1} = M_t - 1 | \mathcal{F}_t) &= 0.30 \end{aligned}$$

Find a value $r \neq 1$ such that $W_t = r^{M_t}$ is a martingale with respect to the natural filtration generated by M_t.

129. Suppose I play an unfair game where I have a 30% chance of winning \$1, a 20% chance of losing \$1, and a 50% chance of staying at my current dollar amount. If I start with \$10, and quit when I reach \$20 or \$0, what is the chance that I quit when I have \$0?

130. Suppose I play an unfair game where I have a 20% chance of winning \$1, a 15% chance of losing \$1, and a 65% chance of staying at my current dollar amount. If I start with \$10, and quit when I reach \$20 or \$0, what is the chance that I quit when I have \$0?

Chapter 15

Uniform Integrability

Question of the Day

Is there a necessary and sufficient condition for

$$\mathbb{E}\left[\lim_{t\to\infty} X_t\right] = \lim_{t\to\infty} \mathbb{E}[X_t]$$

when the $X_t \geq 0$?

Summary

- A stochastic process X_t is **uniformly integrable** if

$$\lim_{B\to\infty} \sup_t \mathbb{E}(|X_t|\mathbb{I}(|X_t| > B)) = 0.$$

- If $X_t \to X$ in probability and the X_t are uniformly integrable, then $\lim \mathbb{E}[X_t] = \mathbb{E}[X]$.

- For $X_t \to X$ in probability, the following are equivalent. (a) The X_t are uniformly integrable. (b) $\lim \mathbb{E}|X_n - X| = 0$. (c) $\lim \mathbb{E}|X_n| = \mathbb{E}|X|$.

- A sufficient condition for a stochastic process X_t to be uniformly integrable is that $|X_t| \leq Y$ where Y is an integrable random variable.

Two main theorems provide sufficient but not necessary conditions for swapping limits and expected value, the Dominated Convergence Theorem (DCT) and the Monotonic Convergence Theorem (MCT).

Is it possible to do better, and find a condition both necessary and sufficient? Today this question will be answered for nonnegative random variables.

Recall our counterexample for swapping limits and mean:

$$U \sim \textsf{Unif}([0,1]),\ X_n = n\mathbb{I}(U \leq 1/n).$$

The problem here is that there is a small chance of the X_n being extremely large. Preventing this from happening would allow the limit swap to work.

The condition that keeps the X_n from being too large is called *uniform integrability*. The main theorem of this section says the following.

Theorem 5
Suppose $X_n \to X$ in probability. Then the following are equivalent.

a. The $\{X_n\}$ are uniformly integrable.

b. $\lim_{n\to\infty} \mathbb{E}|X_n - X| = 0$.

c. $\lim_{n\to\infty} \mathbb{E}|X_n| = \mathbb{E}|X| < \infty$.

The condition $\mathbb{E}|X_n - X| \to 0$ is also called L^1 *convergence*.

In particular, $\mathbb{E}|X_n - X| = 0$ implies that $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

Fact 43
If $\mathbb{E}|X_n - X| = 0$ and $X_n \to X$ in probability, then

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}\left(\lim_{n\to\infty} X_n\right) = \mathbb{E}(X).$$

Proof. Note

$$\lim \mathbb{E}(X_n) = \mathbb{E}(X) \Leftrightarrow \lim \mathbb{E}(X_n - X) = 0 \Leftrightarrow \lim |\mathbb{E}(X_n - X)| = 0.$$

By Jensen's inequality,

$$|\mathbb{E}(X_n - X)| \leq \mathbb{E}|X_n - X|.$$

So $\lim \mathbb{E}|X_n - X| = 0 \to |\mathbb{E}(X_n - X)| = 0$ and the proof is complete. □

15.1 *Defining uniform integrability*

The path towards uniform integrability starts by understanding the mean of a single random variable first. Recall that a random variable X is integrable if $\mathbb{E}(|X|) < \infty$. The next fact says that a random variable is integrable if and only if its tails are small in some sense.

Fact 44
A random variable X is integrable if and only if

$$\lim_{B\to\infty} \mathbb{E}(|X|\mathbb{I}(|X| > B)) = 0.$$

This says that as less and less of the tail of $|X|$ is kept, the average of what is left of the tail must go towards 0. Another way to say this is for $\mathbb{E}[|X|] = \infty$, no matter how far out in the tail you go, still the average remains infinity!

Proof. ($\Rightarrow$) Assume $|X|$ is integrable. Then $\mathbb{P}(|X| < \infty) = 1$, so for any given $|X|$ value, with probability 1 eventually $B \to \infty$ will be greater than $|X|$.

So that means

$$\lim_{B\to\infty} |X|\mathbb{1}(|X| > B) = 0.$$

For all $B > 0$, $|X|\mathbb{1}(|X| > B) \leq |X|$ which is integrable, so we can use the DCT to say:

$$\lim_{B\to\infty} \mathbb{E}[|X|\mathbb{1}(|X| > B)] = \mathbb{E}\left[\lim_{B\to\infty} |X|\mathbb{1}(|X| > B)\right] = \mathbb{E}[0] = 0.$$

($\Leftarrow$) Now suppose $\lim_{B\to\infty} \mathbb{E}(|X|\mathbb{1}(|X| > B) = 0$. Then $|X| = |X|\mathbb{1}(|X| > B) + |X|\mathbb{1}(|X| \leq B)$. So

$$\mathbb{E}[|X|] = \mathbb{E}(|X|\mathbb{1}(|X| > B)) + \mathbb{E}(|X|\mathbb{1}(|X| \leq B)).$$

The assumption is that the first term on the RHS goes to 0 as $B \to \infty$, so there exists a value M such that

$$\mathbb{E}(|X|\mathbb{1}(|X| > M)) \leq 1.$$

Note $|X|\mathbb{1}(|X| \leq M) \leq M$, so

$$\mathbb{E}[|X|] \leq M + 1 < \infty,$$

which finishes the proof. □

Given a stochastic process $\{X_\alpha\}$ where for every index α, X_α is integrable, it holds that

$$\sup_\alpha \lim_{B\to\infty} \mathbb{E}[|X_\alpha|\mathbb{1}(|X_\alpha| > B)] = \sup_\alpha 0 = 0.$$

What if the limit and supremum are swapped, and the result still is 0? Then the resulting process is *uniformly integrable*.

Definition 60
A stochastic process $\{X_t\}$ is **uniformly integrable** if

$$\lim_{B\to\infty} \left(\sup_t \mathbb{E}(|X_t|\mathbb{1}(|X_t| > B))\right) = 0.$$

15.2 *Sufficient conditions for uniform integrability*

The direct definition of uniformly integrability can be difficult to apply.
Therefore, there are more restrictive conditions that imply uniform integrability.

Fact 45
Two conditions that imply $\{X_n\}$ is uniformly integrable are:

- Boundedness:
$$(\exists M)(\forall n)(|X_n| \leq M),$$
- Dominated by integrable random variable:
$$(\exists Y : \mathbb{E}[|Y|] < \infty)(\forall n)(\mathbb{P}(X_n \leq Y) = 1).$$

Proof. A bound is a special case of domination, so it suffices to prove only that condition.

Suppose that $|X_t| \leq Y$ for all t where Y is integrable. Then

$$|X_t|\mathbb{I}(X_t > B) \leq Y\mathbb{I}(Y > B)$$

for all t. Hence

$$\sup_t |X_t|\mathbb{I}(X_t > B) \leq Y\mathbb{I}(Y > B).$$

Since Y is integrable,

$$\lim_{B\to\infty} \mathbb{E}[Y\mathbb{I}(Y > B)] = 0$$

which then gives the same limit for the X_t. □

15.3 *A process that is not uniformly integrable*

Not all processes are uniformly integrable. Any process where you cannot bring a limit inside expectation will *not* be uniformly integrabile.

So consider this earlier example where you cannot bring the limit inside expectation. You can directly show that this process is not uniformly integrable.

- Let $U \sim \text{Unif}([0,1])$ and $Y_n = n\mathbb{I}(U < 1/n)$.
- So $Y_n \in \{0, n\}$.
- For $n > B$, $\mathbb{E}[Y_n\mathbb{I}(Y_n > B)] = 0(1 - 1/n) + n(1/n) = 1$.
- So $\mathbb{E}[Y_n] = 1$ for all n, so $\sup_n \mathbb{E}[Y_n] = 1 \neq 0$.
- Does not converge to 0, so Y_n cannot be uniformly integrable.
- Note $Y_n \to 0$ w.p. 1, and so also in probability.
- And $\lim_{n\to\infty} \mathbb{E}[Y_n] = 1$, but $\mathbb{E}[\lim_{n\to\infty} Y_n] = 0$.

15.4 *Proof of the uniform integrability theorem*

The theorem states that when $X_n \to X$ in probability the following three statements are equivalent.

a. The X_n are uniformly integrable.

b. $\lim \mathbb{E}|X_n - X| = 0$.

c. $\lim \mathbb{E}|X_n| = \mathbb{E}|X| < \infty$.

Proof. Start with $a \to b$. Suppose the X_n are uniformly integrable.

As noted earlier,

$$\lim \mathbb{E}(X_n) = \mathbb{E}(X) \Leftrightarrow \lim \mathbb{E}(X_n - X) = 0.$$

By Jensen's Inequality, the right hand side will be true if

$$\lim \mathbb{E}|X_n - X| = 0.$$

So this is the new goal!

Fix $m \in \{1, 2, 3, ...\}$. Then there exists B_1 such that for all $b \geq B_1$,

$$\sup_t \mathbb{E}[|X_t|\mathbb{I}(X_t > b)] < \frac{1}{3m}.$$

Let ϕ_M be the function that rounds x down to M if $x > M$, and up to $-M$ if $x < -M$. That is:

$$\phi_M(x) = M \cdot \mathbb{I}(x \geq M) + x \cdot \mathbb{I}(-M < x < M) - M \cdot \mathbb{I}(x \leq -M).$$

Then

$$X_n - X = (X_n - \phi_B(X_n)) + (\phi_B(X_n) - \phi_B(X)) + (\phi_B(X) - X).$$

Taking absolute values and means, and using the triangle inequality gives

$$\mathbb{E}|X_n - X| \leq \underbrace{\mathbb{E}|X_n - \phi_B(X_n)|}_{\text{term 1}} + \underbrace{\mathbb{E}|\phi_B(X_n) - \phi_B(X)|}_{\text{term 2}} + \underbrace{\mathbb{E}|\phi_B(X) - X|}_{\text{term 3}}.$$

The goal then becomes: show that each of these three terms is at most $1/(3m)$ for large enough n and B.

The second term is easiest: since $\phi_B(X_n) - \phi_B(X)$ is between $-2M$ and $2M$, the bounded convergence theorem gives that this converges to 0 as $n \to \infty$. Hence there is an N such that $\mathbb{E}|\phi_B(X_n) - \phi_B(X)| \leq 1/(3m)$ for all $n \geq N$.

Uniform integrability also deals easily with the first term. When $X_n \in [-B, B]$, $\mathbb{I}(|X_n| > B) = 0$ and $X_n - \phi_B(X_n) = 0$. When $|X_n| > B$, then $|X_n - \phi_B(X_n)| = |X_n| - B \leq |X_n|$. Hence

$$|X_n - \phi_B(X_n)| \leq |X_n|\mathbb{I}(|X_n| > B).$$

So uniform integrability gives that there exists B_2 such that for all $b \geq B_2$ and n, $\mathbb{E}|X_n - \phi_b(X_n)| \leq 1/(3m)$.

To understand the third term, first show that X must be integrable. Since $|X_n|$ and $|X| \geq 0$, Fatou's lemma says that

$$\liminf \mathbb{E}|X_n| \geq \mathbb{E}[\liminf |X_n|] = \mathbb{E}|X|.$$

So our new goal is to show $\liminf \mathbb{E}[|X_n|]$ is finite.

Since either $|X_n|$ is bigger than b or less than b, for all b,

$$|X_n| = |X_n|\mathbb{I}(|X_n| > b) + |X_n|\mathbb{I}(|X_n| \leq b) \leq |X_n|\mathbb{I}(|X_n| > b) + b,$$

For $b \geq B_1$,

$$\sup_n \mathbb{E}[|X_n|] \leq \left[\sup_n \mathbb{E}(|X_n|\mathbb{I}(X_n > b))\right] + b \leq 1/(3m) + b$$

which means

$$\liminf \mathbb{E}[|X_n|] \leq \sup_n \mathbb{E}[|X_n|] \leq 1 + B_1 < \infty,$$

making $\mathbb{E}[|X|] \leq \liminf \mathbb{E}[|X_n|]$ finite.

From integrability, for the third term there exists B_3 such that for $b \geq B_3$,

$$\mathbb{E}|\phi_b(X) - X| \leq \mathbb{E}[|X|\mathbb{I}(|X| > b) \leq 1/(3m)$$

Hence there for $B = \max\{B_1, B_2, B_3\}$, there exists an N such that for all $n \geq N$,

$$\mathbb{E}|X_n - X| \leq \frac{1}{3m} + \frac{1}{3m} + \frac{1}{3m} = \frac{1}{m}$$

and $\lim \mathbb{E}|X_n - X| = 0$. □

Now for the next equivalence.

Proof. Show that $b \to c$. Note that for any two real values x and y, $||x| - |y|| \leq |x - y|$. Combining this with Jensen's inequality gives

$$0 \leq |\mathbb{E}|X_n| - \mathbb{E}|X|| \leq \mathbb{E}||X_n| - |X|| \leq \mathbb{E}|X_n - X|.$$

So if the right hand side has limit 0, so does $|\mathbb{E}|X_n| - \mathbb{E}|X||$, which completes the result. □

The last equivalence can be shown as follows.

Proof. Show that $c \to a$. Assume statement c is true, so $\lim \mathbb{E}(|X_n|) = \mathbb{E}(|X|)$.

Create a continuous function that serves as a lower bound on $f(x) = x\mathbb{I}(x \leq M)$:

$$\psi_M(x) = x\mathbb{I}(x \in [0, M/2]) + (M - x)\mathbb{I}(x \in (M/2, M]))$$

Since $\psi_M(x) \leq x\mathbb{I}(x \leq M))$, it holds that

$$x\mathbb{I}(x > M) = x - x\mathbb{I}(x \leq M) \leq x - \psi_M(x).$$

That means

$$\begin{aligned}\mathbb{E}[|X_n|\mathbb{I}(|X_n| > M)] &\leq \mathbb{E}|X_n| - \mathbb{E}(\psi_M(|X_n|)) \\ &\leq \mathbb{E}|X_n| - \mathbb{E}|X| + \mathbb{E}|X| - \mathbb{E}(\psi_M(|X|)) + \mathbb{E}(\psi_M(|X|)) - \mathbb{E}(\psi_M(X_n))\end{aligned}$$

Because ψ_M is a continuous function differentiable almost everywhere with a derivative that is at most 1 in absolute value, for any two real values

$$\psi_M(|x|) - \psi_M(|y|) \leq ||x| - |y||.$$

That gives

$$\mathbb{E}(\psi_M(|X|)) - \mathbb{E}(\psi_M(|X_n|)) \leq |\mathbb{E}(|X| - |X_n|)|$$

and

$$\mathbb{E}[|X_n|\mathbb{I}(|X_n| > M)] \leq 2|\mathbb{E}|X_n| - \mathbb{E}|X|| + \mathbb{E}|X| - \mathbb{E}(\psi_M(|X|))$$

Let $\epsilon > 0$.
Because $\mathbb{E}|X_n| \to \mathbb{E}|X|$, there exists an N such that for all $n \geq N$, $2|\mathbb{E}|X_n| - \mathbb{E}|X|| \leq \epsilon/4$.

Because $|X|$ is integrable and $|X| - \psi_M(|X|) \leq |X|$, the dominated convergence theorem gives that there is an M_1 such that for all $m \geq M_1$,

$$\mathbb{E}|X| - \mathbb{E}(\psi_M(|X|)) \leq \epsilon/2.$$

Hence for $n \geq N$ and $m \geq M_1$,

$$\mathbb{E}[|X_n|\mathbb{I}(|X_n| > m)] \leq 2(\epsilon/2) + \epsilon/2 = \epsilon.$$

Since $X_1, \ldots, X_{N-1}$ is a finite set of integrable random variables, there exists and M_2 such that for all $m \geq M_2$ and $i \in \{1, \ldots, N-1\}$,

$$\mathbb{E}(|X_i|\mathbb{I}(|X_i| > m)) \leq \epsilon.$$

Combining these, for $m \geq \max(M_1, M_2)$,

$$\sup_n \mathbb{E}[|X_n|\mathbb{I}(|X_n| > m)] \leq \epsilon.$$

Since $\epsilon > 0$ was arbitrary, this gives

$$\lim_{M\to\infty} \sup_n \mathbb{E}(|X_n|\mathbb{I}(|X_n| > M)) = 0$$

and the X_n are uniformly integrable. □

Problems

131. Let $G \sim \text{Geo}(1/2)$. What is $\lim_{B\to\infty} \mathbb{E}(G\mathbb{I}(G > B))$?

132. Let $U \sim \text{Unif}([0, 10])$. What is $\lim_{B\to\infty} \mathbb{E}(U\mathbb{I}(U > B))$?

133. Let $U \sim \text{Unif}([0, 1])$, and $W_n = \sqrt{n} \cdot \mathbb{I}(U \leq 1/n)$. Show that the set of $\{W_n\}$ is uniformly integrable.

134. Let $U \sim \text{Unif}([0, 1])$, and $R_n = n^2 \cdot \mathbb{I}(U \leq 1/n)$. Show that the set of $\{R_n\}$ is not uniformly integrable.

135. Let $D_1, D_2, \ldots$ be iid $\text{Unif}(\{-1, 1\})$, and $X_t = \sum_{i=1}^t D_i$. Then X_t is a martingale, and $T = \inf\{t : X_t = 1\}$ is a stopping time with respect to the natural filtration.

If you are given the fact that $\mathbb{P}(T < \infty) = 1$, use this to show that $\{X_{t\wedge T}\}$ are not uniformly integrable.

136. Let $D_1, D_2, \ldots$ be iid $\text{Unif}(\{-1, 1\})$, and $X_t = \sum_{i=1}^t D_i$. Then X_t is a martingale, and $R = \inf\{t : X_t = 30\}$ is a stopping time with respect to the natural filtration.

If you are given the fact that $\mathbb{P}(R < \infty) = 1$, use this to show that the sequence $\{X_{t\wedge R}\}$ is not uniformly integrable.

137.

Let $D_1, D_2, \ldots$ be iid $\mathsf{Unif}(\{-1, 1\})$ and consider the martingale

$$M_n = \sum_{i=1}^{n} D_i.$$

For a fixed positive integer a, let

$$T = \inf\{n : M_n \in \{-a, a\}\}$$

a. Use the Martingale Convergence Theorem to show that $\mathbb{P}(T < \infty) = 1$.

b. Find $\mathbb{P}(M_T = a)$.

138.

Let $W_1, W_2, \ldots$ be iid $\mathsf{Unif}(\{-2, 2\})$ and consider the martingale

$$X_n = \sum_{i=1}^{n} W_i.$$

Then for a fixed positive integer a, let

$$T = \inf\{n : X_n \in \{-2a, 2a\}\}$$

a. Use the Martingale Convergence Theorem to show that $\mathbb{P}(T < \infty) = 1$.

b. Find $\mathbb{P}(X_T = 2a)$.

Chapter 16

The Martingale Convergence Theorem

Question of the Day

For a martingale $\{M_t\}$, when does $\lim_{t \to \infty} M_t$ exist?

Summary

- A sufficient condition for $\lim_{t \to \infty} M_t$ to exist is that the martingale is uniformly integrable. This fact is called the **Martingale Convergence Theorem**.

Last time a new characterization of integrable was introduced. A random variable X is integrable if and only if

$$\lim_{B \to \infty} \mathbb{E}[|X| \mathbb{1}(|X| > B)] = 0.$$

This idea gives us a way to characterize when a collection of random variables $\{X_\alpha\}$ are all integrable together, which is called uniform integrability:

$$\lim_{B \to \infty} \sup_\alpha \mathbb{E}[|X_\alpha| \mathbb{1}(|X_\alpha| > B)] = 0.$$

To why this condition might relate to martingales, consider an example of a martingale that does not converge to a value.

16.1 Simple symmetric random walk on the integers

Let $D_1, D_2, \ldots$ be iid D, where

$$\mathbb{P}(D = -1) = \mathbb{P}(D = 1) = 1/2.$$

Then let

$$M_t = \sum_{i=1}^{t} D_i.$$

This walk is called *simple symmetric random walk on the integers*. It is simple because the value of M_t can only change by one integer. It is symmetric because the chance of increasing by one equals the chance of decreasing by one. It is a random walk because it can be written:

$$M_{t+1} = M_t + D_{t+1},$$

that is, the next state M_{t+1} is the current state M_t plus a change D_{t+1} that is independent of what came before.

An example sequence in this chain would be

$$\{M_t\} = 0, 1, 2, 3, 2, 3, 4, 5, 4, 3, 2, 1, -1, 0, -1, 0, -1, -2, -1, 0, 1, \ldots .$$

Note that the value of the process M_t is always changing by 1, so it does not converge to any one particular value. Now consider the stopping time T which is the first time the process hits either 4 or -4:

$$T = \inf\{t : |M_t| = 4\}.$$

Then the stopped process looks like

$$\{M_{t \wedge T}\} = 0, 1, 2, 3, 2, 3, 4, 4, 4, \ldots .$$

This process does converge, to the value 4 or -4 and stays there thereafter.

So what is the difference between these two processes? They are both martingales, but M_t is *not* uniformly integrable, while $M_{t \wedge T}$ *is* uniformly integrable. In fact, all uniformly integrable martingales converge!

16.2 The Martingale Convergence Theorem

Theorem 6
Martingale Convergence Theorem

Let M_n be a uniformly integrable martingale. Then

$$M_\infty = \lim_{n \to \infty} M_n$$

exists with probability 1, and $\mathbb{E}[M_\infty | M_0] = M_0$.

This theorem has a lot of applications!

16.2.1 Applications of the Martingale Convergence Theorem

16.2.1.1 Showing a process is not uniformly integrable

Let $M_t = \sum_{i=1}^{t} D_i$ be simple symmetric random walk on the integers as before. As noted earlier, the process M_t cannot converge because at each time step it changes by 1. Hence it cannot be uniformly integrable.

Here the *contrapositive* is being used from logic. A logical statement $p \to q$ is true if and only if the contrapositive statement $\neg q \to \neg p$ is true. In this case, all uniformly integrable martingales converge, so if a martingale does not converge, it is not uniformly integrable.

16.2.1.2 Showing a stopping time is finite

For M_t, consider the stopping time

$$T = \inf\{t : M_t = -10 \text{ or } M_t = 5\}.$$

Then $|M_{t \wedge T}| \leq 10$, so the collection $\{M_{t \wedge T}\}$ is uniformly integrable. That means that

$$\lim_{t \to \infty} M_{t \wedge T}$$

exists with probability 1. But that can only be true if $\mathbb{P}(T < \infty) = 1$, because if $T = \infty$ then the process $M_{t \wedge T}$ always changes by 1 at each step.

16.3 Proof of the Martingale Convergence Theorem

In order to show the Martingale Convergence Theorem, the key ingredient is called an *upcrossing*. This is an interval $[a, b]$ such that at some time t_1 it holds that $M_{t_1} \leq a$, and for some time $t_2 > t_1$ it holds that $M_{t_2} \geq b$.

Notice that if the martingale represents a stock price, and you buy one share of the stock at time t_1 and sell the share at time t_2, then you must have made at least $t_2 - t_1$ in value. If the stock is a martingale, then if an infinite number of upcrossings happen, then you could make an infinite amount of money, even though the stock price represents a fair game!

Hence it is true (as will be shown) that the expected number of upcrossings in a martingale is finite with probability 1. On the other hand, if you have a sequence that has no limit, then it is either shooting off to infinity or minus infinity, or it is bouncing back and forth, which gives an infinite number of upcrossings.

Since a uniformly integrable martingale can do none of these things, it must have a limit with probability 1.

Now let's make these ideas more precise.

Definition 61
Let $a < b$ be rational numbers, and M_t a uniformly integrable martingale. Set

$$T_a = \inf\{t : M_t \leq a\}, \; T_b = \inf\{t > T_a : M_t \geq b\}.$$

When T_a and T_b are finite, call the interval $[T_a, T_b]$ an **upcrossing** of $[a, b]$ by the martingale.

The next fact shows that any particular upcrossing might never occur!

Fact 46
For $a < b$ on a uniformly integrable martingale, either $\mathbb{P}(T_a = \infty)$ or $\mathbb{P}(T_b = \infty)$ is positive.

Proof. Let $a \leq b$ for a uniformly integrable martingale, where $\mathbb{P}(T_a = \infty) = \mathbb{P}(T_b = \infty) = 0$. For such a martingale, the stopped process is also uniformly integrable and so limits can be brought inside expectations to give

$$M_0 = \lim_{t \to \infty} \mathbb{E}[M_{T_a \wedge t} | M_0] = \mathbb{E}[M_{T_a} | M_0].$$

Also

$$M_0 = \lim_{t\to\infty} \mathbb{E}[M_{T_a \wedge t} | M_0] = \mathbb{E}[M_{T_b} | M_0].$$

Since $M_{T_a} \le a$, it holds that $\mathbb{E}[M_{T_a} | M_0] \le a$ and similarly $\mathbb{E}[M_{T_b} | M_0] \ge b$. Hence

$$M_0 \le a \le b \le M_0$$

which implies that $a = b = M_0$.

Hence for all $a \le b$, this shows that for a uniformly integrable martingale with $\mathbb{P}(T_a < \infty) = \mathbb{P}(T_b < \infty) = 0$ it holds that $a = b$. The contrapositive of this statement is if $a < b$, then at least one of $\mathbb{P}(T_a = \infty)$ and $\mathbb{P}(T_b = \infty)$ must be strictly positive. □

Having a nonzero chance of never having another upcrossing after $T_a < T_b < \infty$ occurs means that the number of upcrossings is finite with probability 1.

Note that the set of pairs of rational numbers $a < b$ is countable, so with probability 1, a u.i. martingale M_t has a finite number of upcrossings for all rational numbers $a < b$.

On the other hand, for a sequence to *not* have a limit, there must be some interval with rational endpoints where an upcrossing occurs an infinite number of times.

Fact 47
Let $x_0, x_1, x_2, \ldots$ be a sequence of real numbers that does not converge to a real number or ∞ or $-\infty$. Then there exists rational numbers $a < b$ such that x_i upcrosses (a, b) infinitely often.

Proof. Recall

$$\limsup x_i = \lim_{n\to\infty} \sup_{i\ge n} x_i$$
$$\liminf x_i = \lim_{n\to\infty} \inf_{i\ge n} x_i$$

A useful real analysis fact is that $\lim x_i$ exists if and only if $\limsup x_i = \liminf x_i$.

Suppose $\liminf x_i < \limsup x_i$, then there must exist rational numbers a and b such that

$$\liminf x_i < a < b < \limsup x_i.$$

Since a and b are strictly inside $(\limsup x_i, \liminf x_i)$, there are an infinite number of x_i that are at least b, and an infinite number of x_i that are at most a. Hence x_i upcrosses (a, b) infinitely often. □

This combined with the upcrossing result for uniformly integrable martingales gives the Martingale Convergence Theorem.

16.4 Polya's Urn

Polya's Urn is a martingale that converges not to a discrete state (like in the earlier examples), but to a continuous random variable!

The idea is to start with one red and one blue marble in an urn (kind of a tall, rounded vase.) At each step of the process, pick a marble uniformly from the urn. Put that marble back into the urn, and add another of the same color.

For example, drawing a red marble at the first step has the player put the red marble back and add another red marble. This gives an urn with 2 red and 1 blue marbles. If R_n is the number of red marbles, and B_n is the number of blue marbles after n draws, let

$$M_n = \frac{R_n}{R_n + B_n} = \frac{R_n}{2+n}$$

be the fraction of red marbles after n draws.

Here is the weird, wild fact about this: M_n is a martingale with respect to the adapted filtration!

Since $|M_n| \leq 1$, it is bounded and integrable.

On draw $n+1$, M_n equals the chance of picking a red marble. Given that there are $n+2$ marbles at this time, the number of red marbles is $R_n = M_n(n+2)$. After one step, the number of red marbles is either $M_n(n+2)+1$ or $M_n(n+2)$. Hence the percentage of red marbles is

$$M_{n+1} \in \left\{ \frac{M_n(n+2)+1}{n+3}, \frac{M_n(n+2)}{n+3} \right\}.$$

The chance of picking these outcomes are M_n and $1-M_n$ respectively. Hence

$$\begin{aligned} \mathbb{E}[M_{n+1}|F_n] &= M_n \frac{M_n(n+2)+1}{n+3} + (1-M_n)\frac{M_n(n+2)}{n+3} \\ &= \frac{M_n}{n+3} + \frac{M_n(n+2)}{n+3} = \frac{M_n(n+3)}{n+3} = M_n \end{aligned}$$

Since $|M_n| \leq 1$, it is uniformly integrable. So do the fractional values in M_n converge to anything?

Yes! It turns out that R_n will always be uniform from 1 up to $n+1$.

Fact 48
In Polya's Urn with $R_0 = B_0 = 1$,

$$R_n \sim \mathsf{Unif}(\{1, 2, \ldots, n+1\}).$$

Proof. The proof is by induction. Since $R_0 = 1$ the base case holds. Suppose for some $n > 0$ that $R_n \sim \mathsf{Unif}(\{1, \ldots, n+1\})$.

Consider R_{n+1}. Let A_n be the indicator that the marble chosen at time $n+1$ is red.

$$\mathbb{P}(R_{n+1} = 1) = \mathbb{P}(R_n = 1, A_{n+1} = 0) = \frac{1}{n+1} \cdot \left(1 - \frac{1}{n+2}\right) = \frac{1}{n+1} \cdot \frac{n+1}{n+2} = \frac{1}{n+2}.$$

Similarly,

$$\mathbb{P}(R_{n+1} = n+1) = \mathbb{P}(R_n = n+1, A_{n+1} = 1) = \frac{1}{n+1} \cdot \frac{n+1}{n+2} = \frac{1}{n+2}.$$

and finally, for $i \in \{2, \ldots, n\}$

$$\begin{aligned}
\mathbb{P}(R_{n+1} = i) &= \mathbb{P}((R_n = i, A_{n+1} = 0) \vee (R_n = i-1, A_{n+1} = 1)) \\
&= \mathbb{P}((R_n = i, A_{n+1} = 0)) + (R_n = i-1, A_{n+1} = 1)) \\
&= \frac{1}{n+1} \cdot \frac{n+2-i}{n+2} + \frac{1}{n+1} \cdot \frac{i-1}{n+2)} \\
&= \frac{1}{n+1} \cdot \frac{n+1}{n+2} \\
&= \frac{1}{n+2}.
\end{aligned}$$

This completes the induction. □

So if R_n is always uniform over $\{1, 2, \ldots, n+1\}$, then for any $a \in [0, 1]$,

$$\mathbb{P}(M_n \leq a) = \mathbb{P}(R_n/(n+2) \leq a) = \lfloor (a(n+2) \rfloor /(n+1)) = (a(n+2) - \epsilon_n)/(n+1)$$

where each $\epsilon_n \in [0, 1]$. Taking the limit as $n \to \infty$ gives

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq a) = a,$$

which means the M_n are converging to the uniform distribution over $[0, 1]$.

Problems

139. If $T_{10} = 3$ and $T_{15} = 7$, call $[3, 7]$ an ____________ of $[10, 15]$.

140.

Suppose there is an upcrossing of $[5, 10]$ over the times $[17, 34]$.

a) What is T_5?

b) What is T_{10}?

141. Suppose that the $B_t \text{Unif}(\{-2, -1, 1, 2\})$ are iid. Show that

$$W_t = \sum_{i=0}^{t} B_i$$

is not uniformly integrable.

142. Suppose that M_t is a martingale such that $|M_{t+1} - M_t| \geq 0.3$ for all t. Show that M_t is not uniformly integrable.

143.

Suppose D is a random variable that is 3 with probability $1/4$, and -1 with probability $3/4$.

a) What is $\mathbb{E}[D]$?

b) Let $D_1, D_2, \ldots$ be iid with the same distribution as D. Then

$$M_n = \sum_{i=1}^{n} D_i$$

is a martingale. Prove that it is not uniformly integrable.

144. Suppose that W is 4 with probability $1/5$ and -1 with probability $4/5$. Let $W_1, \ldots, W_n$ be iid W. Then

$$\sum_{i=1}^{n} W_i$$

is a martingale. Show that this martingale is not uniformly integrable.

145. Suppose that $R \in [0, 1]$ with probability 1 and that $\mathbb{E}[R] = 0.3$. Then for $R_1, R_2, \ldots$ iid with the same distribution as R,

$$M_n = \sum_{i=1}^{n} (R_i - 0.3)/2^i$$

is a martingale. Show directly that no matter what the values of R_i are, that $\lim_{n\to\infty} M_n$ exists.

146. Suppose that $W \in [-1, 1]$ always and has mean 0. Then

$$S_n = \sum_{i=1}^{n} W_i/i^2.$$

is a martingale. Show directly that no matter what the values of the W_i are, that $\lim_{n\to\infty} S_n$ exists.

147.

Consider Polya's Urn, and suppose that it starts with 2 red and 1 blue marble.

a) After one step, what is the distribution of the number of blue marbles?

b) After two steps, what is the distribution of the number of blue marbles?

148. In Polya's Urn starting with 1 blue and 1 red marble, if N_t is the number of blue marbles after t steps, what does $\mathbb{P}(N_t/t \leq 0.4)$ converge to?

Chapter 17

The Optional Sampling Theorem

Question of the Day

A bet of x dollars on red in American Roulette returns $2x$ dollars with probability $18/38$, and 0 dollars with probability $20/38$. Is there a betting scheme that guarantees a player wins 1 dollar with probability 1?

Summary

The **Optional Sampling Theorem** says that for a martingale $\{M_t\}$ with stopping time T such that $M_{t \wedge T}$ is uniformly integrable, then $\mathbb{E}[M_T | M_0] = M_0$.

17.1 *A surprising answer*

The perhaps surprising answer to the Question of the Day is yes, there is! So why are casinos not worried about mathematicians using this method to win fortunes? Because there is a problem: you need to have an infinite amount of money available to accomplish this. Read on to see why.

17.1.1 Martingale betting scheme

This betting scheme has become known as the *martingale* betting scheme.

- Start by betting one dollar.
- If you win, quit, you've won one dollar!
- Otherwise, double the bet, and play again. Repeat until you win.

17.1.2 Analysis

Using this betting scheme can cause the player's worth to dip far into the negative numbers. But whenever the player wins, the *total* amount won will be exactly one dollar!

- Suppose the player wins on the third game. Then the total earning are

$$-1-2+4=1$$

.

- Suppose the player wins on the fifth game. Then the total earnings are

$$-1-2-4-8+16=1.$$

- Now let's generalize! Suppose the player wins on the ith game. Then the total earnings are

$$2^j - \sum_{j=1}^{i-1} 2^{j-1} = 2^j - \frac{2^j - 2^0}{2-1} = 1.$$

- So the player always wins 1 dollar total using this method whenever they win a game.

If M_t is the money after t steps of betting, and T is the first time you win a game,

$$M_T = M_0 + 1 \text{ so } \mathbb{E}[M_T|M_0] \neq M_0.$$

The Optional Sampling Theorem says that this type of behavior cannot happen when the stopped process is a uniformly integrable martingale.

Theorem 7
Optional Sampling Theorem

Suppose that $M_0, M_1, \ldots$ is a martingale and T is a stopping time with respect to $\{F_n\}$. If $M_{T\wedge t}$ is uniformly integrable, then

$$\mathbb{E}[M_T|M_0] = M_0.$$

Optional Time is another term for *stopping time* in the context of discrete time stochastic processes. So M_T is the value of the martingale at the optional time, that is, the martingale is being sampled at this time, hence the name of the theorem.

Proof. The uniform integrability of $M_{t\wedge T}$ gives

$$M_0 = \lim_{t\to\infty} \mathbb{E}[M_{T\wedge t}|M_0] = \mathbb{E}\left[\lim_{t\to\infty} M_{T\wedge t}|M_0\right] = \mathbb{E}[M_\infty|M_0].$$

where the last step comes from the Martingale Convergence Theorem.

If $T < \infty$, then $M_\infty = M_T$. If $T = \infty$, then $M_T = M_\infty$. Either way, $M_0 = \mathbb{E}[M_T|M_0]$, and we are done. □

17.2 *Supermartingales and submartingale*

So far the focus has been on martingales, fair games, but some games are biased upwards or downwards.

Definition 62
A stochastic process $M_0, M_1, \ldots$ is a **submartingale** with respect to a filtration F_n if for all n:

1. M_n is measurable with respect to F_n.
2. $\mathbb{E}[|M_n|] < \infty$
3. for all $n > 0$, $M_n \leq \mathbb{E}[M_{n+1}|F_n]$.

Definition 63
A stochastic process $M_0, M_1, \ldots$ is a **supermartingale** with respect to a filtration F_n if for all n:

1. M_n is measurable with respect to F_n.
2. $\mathbb{E}[|M_n|] < \infty$
3. for all $n > 0$, $M_n \geq \mathbb{E}[M_{n+1}|F_n]$.

- Note: if M_t is a submartingale, $-M_t$ is a supermartingale. If M_t is both a submartingale and a supermartingale, it is just a martingale.
- Our two main theorems (Martingale Convergence Theorem and Optional Sampling Theorem) hold for sub and supermartingales with the appropriate inequality sign.

Theorem 8
Martingale Convergence Theorem

Let $\{M_n\}$ be a uniformly integrable martingale, submartingale, or supermartingale. Then $M_\infty = \lim_{n\to\infty} M_n$ exists with probability 1, and

$$\mathbb{E}[M_\infty|M_0] \begin{cases} \geq M_0 & \text{for submartingales} \\ = M_0 & \text{for martingales} \\ \leq M_0 & \text{for supermartingales} \end{cases}$$

Theorem 9
Optional Sampling Theorem

Suppose that $M_0, M_1, \ldots$ is a sub or super or regular martingale and T is a stopping time with respect to $\{F_n\}$. If $M_{T\wedge t}$ is uniformly integrable, then

$$\mathbb{E}[M_T|M_0] \begin{cases} \geq M_0 & \text{for submartingales} \\ = M_0 & \text{for martingales} \\ \leq M_0 & \text{for supermartingales} \end{cases}$$

17.2.1 Applying the OST: Roulette

Is there a stopping time (betting scheme) that allows you to make money on Roulette, if the player must quit whenever their amount of money leaves a particular interval?

Start by letting M_t be the amount of money that the player has after t plays of Roulette, betting on red each time.

Next, our model should have the next bet amount allowed to depend on whatever happened previously. Let F_t be the information in the history of the t spins. Then let $f(F_t)$ be the amount of money bet on $(t+1)$st spin given what happened in the first t spins.

It follows that M_t is a supermartingale, since

$$\mathbb{E}[M_{t+1}|F_t] = M_t + f(F_n)(18/38) - (20/38)f(F_n) \leq M_t.$$

Now set up the stopping time. Call the interval of money where the player keeps playing $[a, b]$. Then

$$T = \inf\{t : M_t < a \text{ or } M_t > b\}.$$

Because it is bounded, $M_{t \wedge T}$ is a u.i. supermartingale. Then by the OST,

$$\mathbb{E}[M_T|M_0] \leq M_0.$$

In other words, no matter what betting scheme you use, there is no way for the player to make money *on average* as long as the player quits when the money left falls below a certain number or when it rises above a certain number.

17.2.2 An application: convergence of simple symmetric random walk

Can OST be used to show that simple symmetric random walk, which is $X_t = \sum_{i=1}^{t} D_i$ where $D_i \overset{\text{iid}}{\sim} \text{Unif}(\{-1, 1\})$, reaches 1 in finite time with probability 1?

To answer this, consider $T_{a,1} = \inf\{t : X_t \in \{1, a\}\}$, and $T_1 = \inf\{t : X_t = 1\}$. The reason to use $T_{a,1}$ is that $X_{t \wedge T_{a,1}}$ is a bounded martingale, and so is uniformly integrable. That means that the Martingale Convergence Theorem can be used to say that $\lim_{t \to \infty} X_{t \wedge T_{a,1}} = X_\infty$ exists with probability 1, which is equivalent to saying that $T_{a,1}$ is finite with probability 1.

Hence $X_\infty = X_{T_{a,1}}$, and

$$\mathbb{E}[X_{T_{a,1}}] = \mathbb{E}[X_0] = 0,$$

where

$$\mathbb{E}[X_{T_{a,1}}] = (1 - \mathbb{P}(X_{T_{a,1}} = a)) \cdot 1 + \mathbb{P}(X_{T_{a,1}} = a)a.$$

Solving gives

$$\mathbb{P}(X_{T_{a,1}} = a) = 1/(1 - a).$$

Now, if $T_1 = \infty$, then $X_{T_{a,1}} = a$ since it must have reached a before reaching 1. So

$$(\forall a \in \{-1, -2, \ldots\})(\mathbb{P}(T_1 = \infty) \leq \mathbb{P}(X_{T_{a,1}} = a) = 1/(1 - a)).$$

The only number in $[0, 1]$ less than $1/(1 - a)$ for all negative integers a is 0, so $\mathbb{P}(T_1 = \infty) = 0$.

Problems

149. Suppose that M_t is a martingale where $M_0 = 0$ and with stopping time T such that $M_{t \wedge T}$ is uniformly integrable. What can be said about $\mathbb{E}(M_T)$?

150. Suppose that N_t is a martingale where $N_0 = 0$ and with stopping time T_1 such that $N_{t \wedge T_1}$ is uniformly integrable. What can be said about $\mathbb{E}(N_{T_1})$?

151. If M_t is a martingale with stopping time T, and $M_{t \wedge T}$ is uniformly integrable, must it be true that $T < \infty$ with probability 1?

152. In Polya's Urn starting with one red and one blue ball, where M_t is the proportion of red balls, if $T = \inf t : M_t = 1/3$, is $\mathbb{P}(T = \infty) > 0$?

153. Suppose W_t is a martingale with stopping time S. Say $W_0 = 0$ and $|W_{t \wedge S}| \leq 40$. What is $\mathbb{E}[W_S]$?

154. Suppose R_t is a martingale with stopping time τ. Say $R_0 = 4$ and $|R_{t \wedge \tau}| \leq Y$, where $Y \sim \text{Geo}(1/2)$. What is $\mathbb{E}[R_\tau]$?

155. A gambler is trying to develop a betting scheme B_t based on a game that wins B_t with probability 0.4 and loses B_t with probability 0.6 at each play of a game. Is there any betting scheme where $1 \leq B_t \leq 1000$ at each step where the player quits at the first time T where $B_t > M_t$ or $M_t \geq 1000$ such that $\mathbb{E}[M_T | M_0] > M_0$?

156. A gambler is trying to develop a betting scheme B_t based on a game that wins B_t with probability 0.4 and loses B_t with probability 0.6 at each play of a game. Is there any betting scheme where $1 \leq B_t \leq 1000$ at each step where the player quits at the first time T where $M_t \geq 1000$ or $M_t < 0$ such that $\mathbb{E}[M_T | M_0] > M_0$?

Chapter 18

Markov chains

Question of the Day

In a model for a queue, the queue length starts at 0. At each step, with probability 40% someone arrives to the queue, otherwise with probability 60% someone leaves the queue, unless the queue is empty in which case it stays empty. If the queue already has 4 people in it, an arrival does not join the queue. Given the queue is empty, what is the expected time needed for the next time the queue is empty?

Summary

- Informally, a **Markov chain** is a stochastic process that has the memoryless property:

$$[X_{t+1}|X_0, X_1, \ldots, X_t] \sim [X_{t+1}|X_t]$$

  so that the next value in the process only depends on the current value, and no other history.

- Formally, a stochastic process $\{X_t\}$ is a **Markov chain** with respect to a filtration $\mathcal{F}_t$ if for every t, X_t is $\mathcal{F}_t$ measurable, and for all A measurable with respect to X_{t+1}, $\mathbb{P}(X_{t+1} \in A|\mathcal{F}_t) = \mathbb{P}(X_{t+1} \in A|X_t)$.

- A Markov chain is **time homogeneous** if the distribution of the next state only depends on the current state and not on the time value t. So for all t, $[X_{t+1}|X_t] \sim [X_1|X_0]$.

- For a finite state Markov chain, **first step analysis** allows us to find the expected time needed to travel from one state to another.

In the Question of the Day, the model of a queue is an example of a *Markov chain*, a process $X_0, X_1, X_2, \ldots$ where the distribution of X_{t+1} only depends on X_t, and not on the earlier values of the chain.

Unfortunately, this particular Markov chain is not a martingale, submartingale, or supermartingale, since when the queue is empty the state value can only increase, and when the queue is full at 4 the state value can only decrease. Hence new techniques will be needed here.

The queue length at the next step only depends on the current queue length, not on past history. Mathematically, this means that the distribution of the state at the next step only depends on the current state, and not any state prior to that. This can be formalized using a filtration.

18.1 Markov chain definition

Definition 64
A stochastic process $X_0, X_1, X_2, \ldots$ is a **Markov chain** with respect to filtration F_t if for all t,

1. X_t is F_t measurable and
2. $[X_{t+1}|F_t] \sim [X_{t+1}|X_t]$.

Definition 65
A Markov chain $X_0, X_1, \ldots$ is **time-homogeneous** if for all t,

$$[X_{t+1}|X_t] \sim [X_1|X_0].$$

- Markov chains are often called *memoryless* processes because they can only remember the last state X_t, and forget $X_1, \ldots, X_{t-1}$ when generating X_{t+1}.
- For state space Ω, a Markov chain has the form:

$$X_{t+1} = f(X_t, U_{t+1}),$$

  where $f : \Omega \times [0, 1] \to \Omega$ is a deterministic function.
- Unless indicated otherwise, all Markov chains in this course will be time-homogeneous.

18.1.1 Example: the queue in the Question of the Day

- The sequence of queue lengths could be:
  - 0,1,0
  - 0,0
  - 0,1,2,3,4,4,4,3,2,3,2,1,0
- Let $T = \inf\{t \geq 1 : X_t = 0\}$.
- What is $\mathbb{E}[T]$?

18.1.2 How to represent Markov chain

There are multiple ways to represent Markov chains, perhaps the easiest to follow is graphically.

Definition 66
A **directed graph** is a mathematical object that consists of a set of nodes V (also called vertices) and edges (also called arcs) $E \subset V^2$. For edge $(v, w) \in E$, call w the **head** of the arc and v the **tail**.

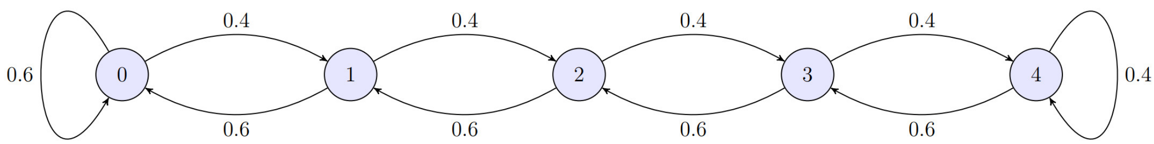
For a time-homogeneous Markov chain, nodes represent states of the chain, and an edge from v to w is marked with the probability of moving from state v to state w in one step.

Definition 67
For a Markov chain over state space Ω, the **transition graph** has node set Ω, edge set Ω^2, and every edge (x, y) receives label

$$p(x, y) = \mathbb{P}(X_{t+1} = y | X_t = x).$$

Typically arcs with label 0 are not drawn in the representation, although they still exist.

For the QotD MC, this transition graph can be drawn as follows.



18.1.3 Back to the Question of the Day

In the question of the day T was the number of steps needed to return to state 0 starting from state 0. It actually helps to have more random variables defined.

For $a \in \{0, 1, 2, 3, 4\}$, let

$$T_a = \inf\{t \geq 0 : X_t = 0 | X_0 = a\}.$$

For instance, T_3 is the number of steps needed to get to state 0 starting at state 3. And $\mathbb{E}[T_3]$ is the average number of steps needed to return to 0 starting at state 3.

To find $\mathbb{E}[T_3]$ (and $\mathbb{E}[T_1]$, $\mathbb{E}[T_2]$, etcetera), use **first step analysis**. This method considers what happens at the first step of the chain.

$$\begin{aligned}
\mathbb{E}[T_3] &= \mathbb{E}[\mathbb{E}[T_3 | X_1]] \\
&= \mathbb{E}[T_3 | X_1 = 2]\mathbb{P}(X_1 = 2) + \mathbb{E}[T_3 | X_1 = 4]\mathbb{P}(X_1 = 4) \\
&= (1 + \mathbb{E}[T_2])(0.6) + (1 + \mathbb{E}[T_4])(0.4).
\end{aligned}$$

This same technique can give an equation for each of the $\mathbb{E}[T_i]$ values.

$$\begin{aligned}
\mathbb{E}[T_0] &= 0 \\
\mathbb{E}[T_1] &= (1 + \mathbb{E}[T_0])(0.6) + (1 + \mathbb{E}[T_2])(0.4) \\
\mathbb{E}[T_2] &= (1 + \mathbb{E}[T_1])(0.6) + (1 + \mathbb{E}[T_3])(0.4) \\
\mathbb{E}[T_3] &= (1 + \mathbb{E}[T_2])(0.6) + (1 + \mathbb{E}[T_4])(0.4) \\
\mathbb{E}[T_4] &= (1 + \mathbb{E}[T_3])(0.6) + (1 + \mathbb{E}[T_4])(0.4).
\end{aligned}$$

It is more compact to write this as a linear system. Let $w_i = \mathbb{E}[T_i]$, then

$$\begin{aligned} w_0 &= 0 \\ w_1 &= 1 + 0.6w_0 + 0.4w_2 \\ w_2 &= 1 + 0.6w_1 + 0.4w_3 \\ w_3 &= 1 + 0.6w_2 + 0.4w_4 \\ w_4 &= 1 + 0.6w_3 + 0.4w_4, \end{aligned}$$

Or in matrix notation:

$$\begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -0.6 & 1 & -0.4 & 0 & 0 \\ 0 & -0.6 & 1 & -0.4 & 0 \\ 0 & 0 & -0.6 & 1 & -0.4 \\ 0 & 0 & 0 & -0.6 & 0.6 \end{pmatrix} \vec{w}.$$

These sorts of systems of equations can be solved using MATLAB, Mathematica, R, or many modern calculators.

Wolfram Alpha provides a web interface to Mathematia. To solve this problem in Wolfram Alpha, give the following command to www.wolframalpha.com:

```
inverse{{1,0,0,0,0},{-0.6,1,-0.4,0,0},{0,-0.6,1,-0.4,0},
        {0,0,-0.6,1,-0.4},{0,0,0,-0.6,0.6}}*
        {{0},{1},{1},{1},{1}}
```

The result to four sig figs is

$$\begin{pmatrix} 0 \\ 4.012 \\ 7.530 \\ 10.30 \\ 11.97 \end{pmatrix}$$

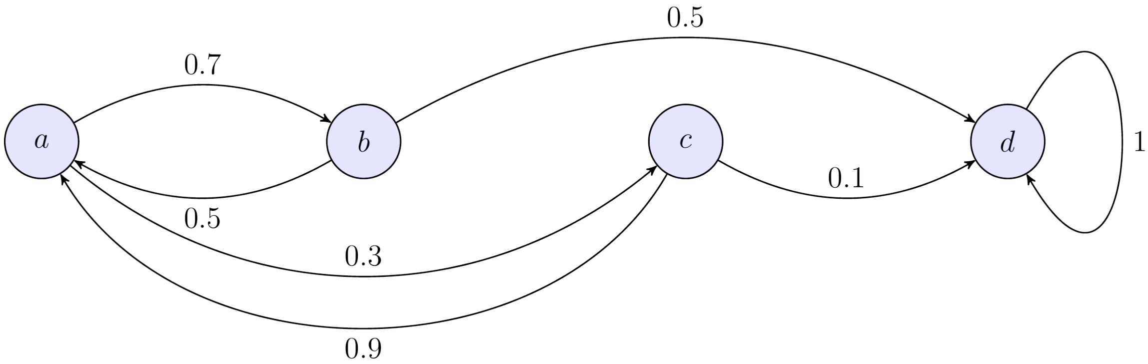
18.1.4 Finishing the QotD

So what is $\mathbb{E}[T]$ given the $\mathbb{E}[T_i]$ values? Again, use first step analysis! Consider where the chain moves from state $X_0 = 0$ at the first step, and then use the appropriate mean value.

$$\begin{aligned} \mathbb{E}[T] &= \mathbb{E}[\mathbb{E}[T|X_1]] \\ &= \mathbb{E}[T|X_1 = 1](0.4) + \mathbb{E}[T|X_1 = 0](0.6) \\ &= 0.4(1 + w_1) + 0.6(1) = 2.60494 = \boxed{2.60494\ldots}. \end{aligned}$$

18.2 An example Markov chain

Consider the following 4-state Markov chain.



- Starting at a, what is the expected number of steps needed to get to d?
- First step analysis to the rescue!
- Let w_i be expected number of steps to reach d starting from i.

$$
\begin{aligned}
w_a &= 0.7(1 + w_b) + 0.3(1 + w_c) \\
w_b &= 0.5(1 + w_a) + 0.5(1 + w_d) \\
w_c &= 0.9(1 + w_a) + 0.1(1 + w_d) \\
w_d &= 0
\end{aligned}
$$

This can be undertaken in Wolfram Alpha using

```
solve a = 0.7(1 + b) + 0.3(1 + c) and b = 0.5(1 + a)
+ 0.5(1 + d) and c = 0.9(1 + a) + 0.1(1 + d) and
d = 0
```

- Then $a = 100/19 = \boxed{5.263\ldots}$.

18.3 *Example of a stochastic process that is not a Markov chain*

Recall the Martingale betting scheme, where if we lose we double the bet, and if we win the bet returns to 1. So if the sequence of bets was

$$1, 1, 1, 2, 4, 8, 1, 1, 1, \ldots$$

then we know that we won the first two bets, lost the next three, then won the next three. (The 9th bet we wouldn't know if we won or lost until the next bet was revealed.) Suppose we started with $M_0 = 3$ dollars, then the $\{M_t\}$ sequence would be

$$3, 4, 5, 4, 2, -2, 6, 7, 8,$$

This is definitely not a Markov chain! For instance, if the sequence was

$$3, 2, 0, 4, 5,$$

then we won the last bet, so the next bet would be 1 and $M_5 \in \{4, 6\}$. However, if the sequence was

$$3, 4, 5, 6, 5,$$

then we lost the last bet, but won the one before, so the next bet would be 2 and $M_5 \in \{3, 7\}$. Hence the final state is not enough information to determine the distribution of the next state, so this is not a Markov chain.

Problems

157. If X_t is a Markov chain, what can you say about the values of

$$\mathbb{P}(X_2 = c | X_0 = a, X_1 = b)$$

and

$$\mathbb{P}(X_2 = c | X_0 = d, X_1 = b)$$

158. If Y_t is a Markov chain, what can you say about the values of

$$\mathbb{P}(Y_7 = 4 | Y_3 = 2, Y_6 = -1)$$

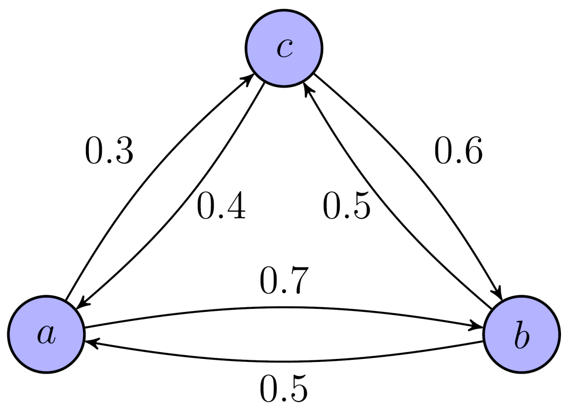
and

$$\mathbb{P}(Y_7 = 4 | Y_3 = 4, Y_6 = -1)$$

159. If Y_t is a time-homogeneous Markov chain, what can you say about the values

$$\mathbb{P}(Y_1 = d | Y_0 = a) \text{ and } \mathbb{P}(Y_7 = d | Y_6 = a)?$$

160. Consider the following Markov chain:



Let $T_{ab} = \inf\{t \geq 0 : X_t = b | X_0 = a\}$. Find $\mathbb{E}[T_{ab}]$.

161. For $\{X_t\}$ a time-homogeneous Markov chain, what can you say about $\mathbb{P}(X_5 = a | X_3 = b)$ and $\mathbb{P}(X_{12} = a | X_{10} = b)$?

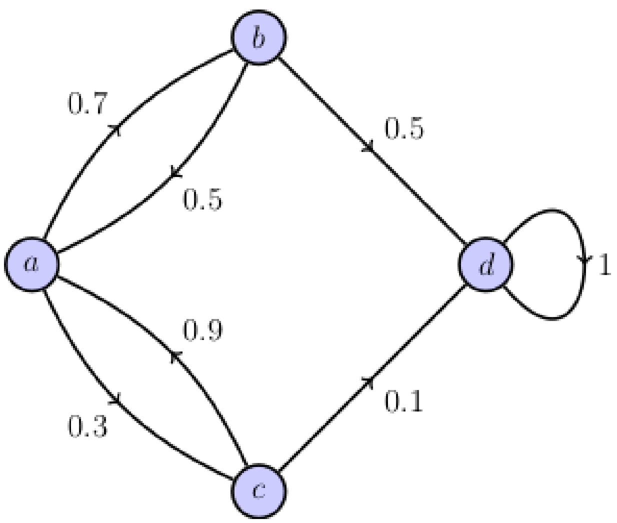
162. In a time-homogeneous Markov chain $\{Y_t\}$, what can be said about $\mathbb{P}(Y_{15} = 7 | Y_3 = 3)$ and $\mathbb{P}(Y_{20} = 7 | Y_8 = 3)$?

Chapter 19

Transition Matrices

Question of the Day

Consider the following Markov chain:



If the Markov chain is equally likely to start in any of the four states, what is the chance after 5 steps that the state is d?

Summary

- **Update functions** are deterministic functions that take the current state of a Markov chain and some random choices, and compute the next state of the Markov chain.
- A **transition matrix** of a finite state Markov chain with n states is an n by n matrix whose entry in the ith row and jth column is $\mathbb{P}(X_{t+1} = j | X_t = i)$.
- Consider a Markov chain $\{X_t\}$ with transition matrix A. Then $X_0, X_k, X_{2k}, \ldots$ is also a Markov chain

for any positive integer k, and has transition matrix A^5.

19.1 Representing Markov chains

There are multiple ways to represent Markov chains. One way is to use an *update function*, which takes the current state together with some source of randomness, and returns the new state. Typically the source of randomness is a standard uniform random variable, chosen for that step. Because the update function takes the current state and $U \sim \text{Unif}([0,1])$, its domain is $\Omega \times [0,1]$. Because it outputs the next state in the Markov chain, its codomain is Ω.

Definition 68
Say that $f : \Omega \times \Omega_R \to \Omega$ is an **update function** for a Markov chain $\{X_t\}$ if for all t and $R_1, R_2, \ldots$,

$$f(X_t, R_{t+1}) \sim [X_{t+1} | X_t]$$

With an update function and $U_1, U_2, \ldots$ iid $\text{Unif}([0,1])$, the state of the chain can be set to

$$X_{t+1} = f(X_t, U_{t+1})$$

For instance, consider the Markov chain that is a simple symmetric random walk on the integers. Then given the current state x, the next state will be $x+1$ with probability $1/2$ and $x-1$ with probability $1/2$. So

$$\mathbb{P}(X_{t+1} = X_t + 1 | X_t) = \mathbb{P}(X_{t+1} = X_t - 1 | X_t) = 1/2.$$

This Markov chain could be represented by an update function as follows.

$$f(x, u) = x + \mathbb{I}(u \leq 0.5) - \mathbb{I}(u > 0.5).$$

For $U_{t+1} \sim \text{Unif}([0,1])$, it is easy to verify that

$$\mathbb{P}(f(X_t, U_{t+1}) = X_t + 1 | X_t) = \mathbb{P}(f(X_t, U_{t+1}) = X_t - 1 | X_t) = 1/2.$$

Conversely, if a stochastic process is defined using an update function, then it turns out to also be a Markov chain.

Fact 49
Suppose that $R_1, R_2, \ldots$ is an iid sequence, $X_0 = x_0$, and for $t \geq 1$,

$$X_{t+1} = f(X_t, R_{t+1}).$$

Then X_t is a Markov chain with respect to the filtration $\mathcal{F}_t = \sigma(X_0, R_1, R_2, \ldots, R_t)$.

19.2 Transition graphs and matrices

When the state space is finite, there are alternate ways of representing the Markov chain. In the last chapter an edge labeled transition graph was used.

The Markov chain can also be represented by a square matrix with rows for every state and columns for every state and entries corresponding to the probability of moving from the row state to the column state.

Definition 69
For a finite state Markov chain with n possible states, let the n by n matrix whose entry for row i and column j is

$$p(i,j) = \mathbb{P}(X_{t+1} = j | X_t = i)$$

be the **transition matrix** of the Markov chain.

For instance, if the columns are rows are both labeled in order (a, b, c, d), the transition matrix of the transition graph in the Question of the Day becomes

$$\begin{pmatrix} 0 & 0.7 & 0.3 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0.9 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

19.3 Taking one step in the Markov chain

In the question of the day, the initial state was taken to be equally likely to be any of the four states $\{a, b, c, d\}$. That is,

$$X_0 \sim \text{Unif}(\{a, b, c, d\}).$$

This can be represented by a *probability vector*,

$$p_0 = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}.$$

In general, let the vector p_t be defined as $p_t(i) = \mathbb{P}(X_t = i)$ in the process.

Now consider the chance that $X_1 = d$. The event $X_1 = a$ can be partitioned into different possibilities based on the value of X_0:

$$\mathbb{P}(X_1 = a) = \mathbb{P}(X_1 = a, X_0 = a) + \mathbb{P}(X_1 = a, X_0 = b) + \mathbb{P}(X_1 = a, X_0 = c) + \mathbb{P}(X_1 = a, X_0 = d)$$

Note that for every i,

$$\mathbb{P}(X_1 = a, X_0 = i) = \mathbb{P}(X_0 = i)\mathbb{P}(X_1 = a | X_0 = i).$$

Using the notation $p_0(i) = \mathbb{P}(X_0 = i)$ and

$$p(i,j) = \mathbb{P}(X_1 = j | X_0 = i),$$

this gives

$$\mathbb{P}(X_1 = a, X_0 = i) = p_0(i)p(i,a).$$

With this notation

$$\begin{aligned} \mathbb{P}(X_1 = a) &= p_0(a)p(a,a) + p_0(b)p(b,a) + p_0(c)p(c,a) + p_0(d)p(d,a) \\ &= (0)(1/4) + (0.5)(1/4) + (0.9)(1/4) + (0)(1/4) \\ &= 0.35 \end{aligned}$$

This equation could also be written using matrix multiplication.

$$\mathbb{P}(X_1 = a) = \underbrace{(1/4 \quad 1/4 \quad 1/4 \quad 1/4)}_{p_0} \underbrace{\begin{pmatrix} 0 \\ 0.5 \\ 0.9 \\ 0 \end{pmatrix}}_{\text{first col of transition matrix}}.$$

Similar equations describe the probability that X_1 equals the rest of the states.

$$\begin{aligned}
\mathbb{P}(X_1 = b) &= p_0(a)p(a,b) + p_0(b)p(b,b) + p_0(c)p(c,b) + p_0(d)p(d,b) \\
\mathbb{P}(X_1 = c) &= p_0(a)p(a,c) + p_0(b)p(b,c) + p_0(c)p(c,c) + p_0(d)p(d,c) \\
\mathbb{P}(X_1 = d) &= p_0(a)p(a,d) + p_0(b)p(b,d) + p_0(c)p(c,d) + p_0(d)p(d,d)
\end{aligned}$$

Putting these four equations together gives the equation

$$p_1 = p_0 A,$$

where

$$A = \begin{pmatrix} p(a,a) & p(a,b) & p(a,c) & p(a,d) \\ p(b,a) & p(b,b) & p(b,c) & p(b,d) \\ p(c,a) & p(c,b) & p(c,c) & p(c,d) \\ p(d,a) & p(d,b) & p(d,c) & p(d,d) \end{pmatrix}$$

is just the transition matrix of the Markov chain.

In the Question of the Day, the transition from the zeroth step to the first step looks like

$$(1/4 \quad 1/4 \quad 1/4 \quad 1/4) \begin{pmatrix} 0 & 0.7 & 0.3 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0.9 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = (0.35 \quad 0.175 \quad 0.075 \quad 0.4)$$

19.3.1 Taking multiple steps

There was nothing special about X_0 to X_1 in last example. To move from prob. vector for X_5 to X_6, do something similar:

$$p_6 = p_5 A.$$

Now consider how to move from the probability vector to X_0 to that of X_2. First move to X_1, then go to X_2:

$$p_2 = p_1 A = (p_0 A)A = p_0(A \cdot A) = p_0 A^2.$$

Note that this calculation depended on the fact that while matrix multiplication is not commutative (you cannot change the order of multiplication), it is associative so you *can* move parenthesis around.

Fact 50
For a Markov chain X_t with transition matrix A, the process $X_0, X_k, X_{2k}, \ldots$ is also a Markov chain for any nonnegative integer k. Moreover, the transition matrix of this Markov chain is A^k.

19.3.2 Solving the Question of the Day

In the Question of the Day, the goal is to take five steps in Markov chain:

$$p_5 = p_4 A = p_3 A^2 = p_2 A^3 = p_1 A^4 = p_0 A^5.$$

Starting from a specified probability vector gives the following.

Fact 51
If $X_0 \sim p$ for a Markov chain with transition matrix A, then for $t \geq 0$,

$$X_t \sim pA^t.$$

19.3.2.1 Solving the Question of the Day in R

Matrix operations in R are accomplished with the `matrix` data type. Consider the following R code that sets up the matrix A row by row.

```
A <- matrix(c(  0, 0.7, 0.3,   0,
              0.5,   0,   0, 0.5,
              0.9,   0,   0, 0.1,
                0,   0,   0,   1),
            byrow = TRUE,
            nrow = 4)
A
```

```
##      [,1] [,2] [,3] [,4]
## [1,]  0.0  0.7  0.3  0.0
## [2,]  0.5  0.0  0.0  0.5
## [3,]  0.9  0.0  0.0  0.1
## [4,]  0.0  0.0  0.0  1.0
```

Now setup the first probability vector.

```
p_0 <- matrix(c(1 / 4, 1 / 4, 1 / 4, 1 / 4), nrow = 1)
p_0
```

```
##      [,1] [,2] [,3] [,4]
## [1,] 0.25 0.25 0.25 0.25
```

Matrix multiplication in R uses the `%*%` operator. So

```
p_0 %*% A %*% A %*% A %*% A %*% A
```

```
##         [,1]    [,2]    [,3]    [,4]
## [1,] 0.13454 0.06727 0.02883 0.76936
```

So the chance that the state is in d after five steps is about $0.7693\ldots$. Quite high, which is unsurprising because once the state enters d, it cannot leave.

19.3.2.2 Raising matrices to high powers in R

In the above calculation, the `%*%` operator was used five times to find A^5. This can be tedious as the exponent gets larger. To do matrix exponentiation in R, one way is to download a *package* (also known as a *library*), that adds capabilities to R. In this case, the library `expm` adds the operator `%^%` which calculates powers of a matrix.

A library needs to be installed once for each installation of R. So if `expm` is not already loaded into your installation of R, type `install.packages("expm")` into the console to accomplish the task. Again, this only needs to be done once for each installation of R.

The next command needs to be run each time you restart R. This `library` command loads the library into memory so that it can be used in your code.

```
library(expm)
```

Now play with the transition matrix! First consider A^{10}.

```
A %^% 10
```

```
##             [,1]       [,2]       [,3]      [,4]
## [1,] 0.09161328 0.00000000 0.00000000 0.9083867
## [2,] 0.00000000 0.05171718 0.02216450 0.9261183
## [3,] 0.00000000 0.09309092 0.03989611 0.8670130
## [4,] 0.00000000 0.00000000 0.00000000 1.0000000
```

Here the (i, j) entry is the probability that $X_{10} = j$ given that $X_0 = i$. So if you start in state a, the chance of being in state d after 10 steps is $0.9083\ldots$.

Higher powers!

```
A %^% 1000
```

```
##                [,1]          [,2]          [,3] [,4]
## [1,] 1.569802e-104  0.000000e+00  0.000000e+00    1
## [2,]  0.000000e+00 8.861783e-105 3.797907e-105    1
## [3,]  0.000000e+00 1.595121e-104 6.836232e-105    1
## [4,]  0.000000e+00  0.000000e+00  0.000000e+00    1
```

After 1000 steps, the chance of being in anything other than state d is smaller than one over the number of particles in the known universe (well the known universe as of 2023 anyway)!

That is because state d is absorbing, and the chance of avoiding this chain declines exponentially fast. That is, the chance the state stays away from d is declining exponentially in the number of steps taken.

19.4 *Another example of taking multiple steps in the Markov chain*

What if there is no absorbing state like d? Consider a 3-state Markov chain with transition matrix

$$A = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0.4 & 0.1 \end{pmatrix}$$

```
matrix(c(0.5, 0.5, 0, 0.3, 0, 0.7, 0.5, 0.4, 0.1), byrow = TRUE,
      nrow = 3) %^% 100
```

```
##           [,1]      [,2]      [,3]
## [1,] 0.4366197 0.3169014 0.2464789
## [2,] 0.4366197 0.3169014 0.2464789
## [3,] 0.4366197 0.3169014 0.2464789
```

Then after 100 steps

$$A^{100} \approx \begin{pmatrix} 0.4366197\ldots & 0.3169014\ldots & 0.2464789\ldots \\ 0.4366197\ldots & 0.3169014\ldots & 0.2464789\ldots \\ 0.4366197\ldots & 0.3169014\ldots & 0.2464789\ldots \end{pmatrix}.$$

This says that no matter where you start the process at time 0, after 100 steps the chances of being in state a, b, or c is the same (at least to 7 sig figs)!

This behavior is called *ergodicity* and is one of the major reasons why Markov chains are so useful.

Problems

163.

Suppose a Markov chain with states $\{a, b, c\}$ has transition matrix

$$A = \begin{pmatrix} 0.3 & 0 & 0.7 \\ 0 & 0.5 & 0.5 \\ 0.1 & 0.9 & 0 \end{pmatrix}$$

a. What is $\mathbb{P}(X_1 = c | X_0 = b)$?

b. What is $\mathbb{P}(X_7 = a | X_6 = a)$?

164.

Suppose a Markov chain with states $\{1, 2, 3\}$ has transition matrix

$$A = \begin{pmatrix} 0.7 & 0 & 0.3 \\ 0 & 0.7 & 0.3 \\ 0.7 & 0.3 & 0 \end{pmatrix}$$

a. What is $\mathbb{P}(X_1 = 3 | X_0 = 1)$?

b. What is $\mathbb{P}(X_7 = 2 | X_6 = 2)$?

165. Suppose that there is a sequence D_t that are iid draws from the uniform distribution over $\{-1, 1\}$. Prove that

$$X_t = \sum_{i=1}^{t} D_t$$

is a Markov chain formed from the filtration $\mathcal{F}_t = \sigma(D_1, \ldots, D_t)$.

166. Suppose that there is a sequence R_t that are iid draws from the uniform distribution over $\{-1, 2\}$. Prove that

$$X_t = \sum_{i=1}^{t} R_t$$

is a Markov chain formed from the filtration $\mathcal{F}_t = \sigma(R_1, \dots, R_t)$.

167.

Suppose that there is a sequence D_t that are iid draws from the uniform distribution over $\{-1, 1\}$ and

$$X_t = \sum_{i=1}^{t} D_t$$

a. Let $T = \inf\{t : X_t = 2\}$. Prove that $M_t = X_{t \wedge T}$ is a Markov chain with respect to the natural filtration.

b. Prove that $M_t = X_{t \wedge (T+1)}$ is *not* a Markov chain.

168.

To show that $\{X_t\}$ is a Markov chain over a discrete state space, the condition

$$\mathbb{P}(X_{t+1} \mid \mathcal{F}_n) = \mathbb{P}(X_{t+1} \mid X_t)$$

follows from

$$\mathbb{P}(X_{t+1} = x | F_t) = f(x, R_{t+1})$$

for some function f and random variable R_{t+1} independent of $\mathcal{F}_t$.

For example, let D_i be iid $\text{Unif}(\{-1, 1\})$, and $X_t = \sum_{i=1}^{t} D_i$. Then X_t is a Markov chain, since

$$\mathbb{P}(X_{t+1} = x | F_t) = (1/2)\mathbb{I}(X_t \in \{x+1, x-1\}) = \begin{cases} 1/2 & X_t = x+1 \\ 1/2 & X_t = x-1 \end{cases}.$$

for all t.

On the other hand, to show that $\{Y_t\}$ is *not* a Markov chain, it suffices to find 2 sequences $y_0, y_1, \dots, y_{t-1}$ and $y'_0, \dots, y'_{t-1}$ and elements y_t, y_{t+1} such that

$$\mathbb{P}(Y_{t+1} = y_{t+1} | Y_0 = y_0, \dots, Y_{t-1} = y_{t-1}, Y_t = y_t) \neq \mathbb{P}(Y_{t+1} = y_{t+1} | Y_0 = y'_0, \dots, Y_{t-1} = y'_{t-1}, Y_t = y_t).$$

For example, if $B \sim \text{Bern}(1/2)$ and $Y_0 \sim \text{Unif}\{1, 2, 3, 4, 5, 6\}$ are independent and $Y_t = Y_0 + tB$, then $\mathbb{P}(Y_3 = 5 | Y_0 = 2, Y_1 = 3, Y_2 = 4) = 1$, but $\mathbb{P}(Y_3 = 5 | Y_0 = 4, Y_1 = 4, Y_2 = 4) = 0$, so $\{Y_t\}$ is not a Markov chain.

a. Let $T_1 = \inf\{t : X_t = 4\}$. Is $X_{t \wedge T_1}$ a Markov chain? (Be sure to prove your answer.)

b. Let $T_2 = \inf\{t > T_1 : X_t = 4\}$ (so T_2 is the *second* time that the X_t process touches 4.) Is $X_{t \wedge T_2}$ a Markov chain? (Be sure to prove your answer.)

169. If $\mathbb{P}(X_{t+1} = i | X_t = i) = 1$, call the state i of the Markov chain *absorbing*.

Given transition matrix

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0.1 & 0.9 \end{pmatrix},$$

for state space $\{a, b, c, d\}$, which states are absorbing states?

170. Given transition matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.1 & 0.4 & 0.5 & 0 \\ 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

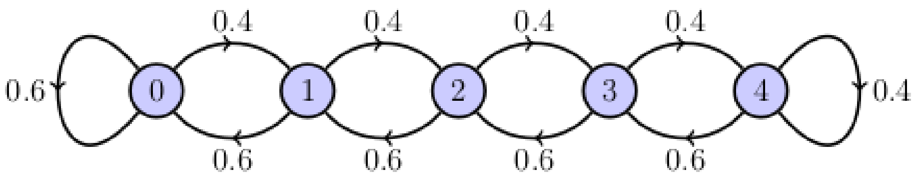
for state space $\{a, b, c, d\}$, which states are absorbing states?

Chapter 20

Limiting and Stationary Distributions

Question of the Day

For the queue model



the state represents the number of customers waiting in the queue. What is the long term average number of people waiting in the queue?

Summary

- A finite state Markov chain with transition matrix A has a **limiting distribution** with probability vector π if

$$\lim_{t\to\infty} p_0 A^T = \pi$$

regardless of the starting vector p_0.

- A Markov chain with transition matrix A has a **stationary distribution** with probability vector π if

$$\pi A = \pi.$$

- A limiting distribution for a Markov chain will always be a stationary distribution, but a stationary distribution might not be a limiting distribution.

20.1 Stochastic matrices

Recall that the *transition matrix* for a finite state Markov chain has the ith row and jth column entry equal to

$$p(i,j) = \mathbb{P}(X_{t+1} = j | X_t = i).$$

In the Question of the Day, the transition matrix is

$$A = \begin{pmatrix} 0.6 & 0.4 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0.6 & 0.4 \end{pmatrix}.$$

The sum of the entries in a row equals the probability that the state ends up in some state given that the state started in the row state. The probability the state ends up somewhere is 1, so the entries of each row have to add up to 1. A matrix with this property will be called *stochastic*.

Definition 70
A matrix is **stochastic** (or more specifically *row stochastic*) if the sum of the entries of each row is 1.

Recall that the (i,j) entry of A^t will be

$$\mathbb{P}(X_t = j | X_0 = i).$$

So this matrix will be stochastic as well! You can also derive this property straight from linear algebra properties of matrix multiplication, but the probabilistic argument is more fun.

Fact 52
Transition matrices and their powers are stochastic matrices.

20.1.1 Using R to find powers of matrices

In R, the matrix A can be stored in a variable by giving the `matrix` function the entries as a single long vector and then indicating that the entries should be read in by row. Giving the number of rows as 5 ensures that the result is a 5 by 5 matrix.

```
A <- matrix(c(0.6, 0.4, 0, 0, 0,
              0.6, 0, 0.4, 0, 0,
              0, 0.6, 0, 0.4, 0,
              0, 0, 0.6, 0, 0.4,
              0, 0, 0, 0.6, 0.4),
            byrow = TRUE,
            nrow = 5)
A
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  0.6  0.4  0.0  0.0  0.0
## [2,]  0.6  0.0  0.4  0.0  0.0
```

```
## [3,]  0.0  0.6  0.0  0.4  0.0
## [4,]  0.0  0.0  0.6  0.0  0.4
## [5,]  0.0  0.0  0.0  0.6  0.4
```

Now consider raising this matrix to the 100th power. The `expm` library allows us to do this in R.

```r
library(expm)
```

This gives the `%^%` operator for matrix powers.

```r
A %^% 100
```

```
##           [,1]      [,2]      [,3]      [,4]       [,5]
## [1,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
## [2,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
## [3,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
## [4,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
## [5,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
```

Note that the columns seem to be identical! What this means is that no matter where the state starts, the chance of ending up in state 0, 1, 2, 3, 4 will be the same! This behavior is not a coincidence.

20.2 Limiting distribution

Say that a Markov chain has a *limiting distribution* if after a large number of steps, the Markov chain forgets where it started at.

Definition 71
A probability distribution π is a **limiting distribution** for a Markov chain if for all $x_0 \in \Omega$, and $A \in F$,

$$\lim_{t \to \infty} \mathbb{P}(X_t \in A | X_0 = x_0) = \pi(A).$$

Note that if Ω is a discrete set $\{1, 2, 3, \ldots\}$, then this is the same as saying that for all $n, m \in \{1, 2, 3, \ldots\}$,

$$\lim_{t \to \infty} \mathbb{P}(X_t = n | X_0 = m) = \pi(\{i_n\}).$$

In terms of powers of matrices, this is equivalent to saying,

$$\lim_{t \to \infty} p_0 A^t = \pi,$$

no matter what the starting vector p_0 is!

20.3 Eigenvalues and eigenvectors

Recall that if for matrix A and nonzero vector v, if

$$vA = \lambda v$$

then call v an *eigenvector* and λ is an *eigenvalue*. (The German prefix *eigen* means *same* here.) To be precise, v is a *left* eigenvector because it is multiplying the matrix on the left. Note that left eigenvectors of A are right eigenvectors of A^T (read as A transpose).

Note that eigenvectors are not unique. That is because if v is an eigenvector, so is any nonzero constant times v.

20.3.1 Eigenvalues and eigenvectors in R

The `eigen` function in R returns eigenvalues together with right eigenvectors. To get left eigenvectors, use the `t` function to take the transpose of a matrix.

```
eigen(t(A))
```

```
## eigen() decomposition
## $values
## [1]  1.0000000  0.7926715 -0.7926715 -0.3027736  0.3027736
##
## $vectors
##            [,1]       [,2]       [,3]        [,4]        [,5]
## [1,] -0.7519041 -0.7572879  0.2575373  0.47227785 -0.64557266
## [2,] -0.5012694 -0.2431797 -0.5977748 -0.71059994  0.31980208
## [3,] -0.3341796  0.1835892  0.6180402  0.04373292  0.59176114
## [4,] -0.2227864  0.4046631 -0.4179883  0.45166467  0.08541468
## [5,] -0.1485243  0.4122153  0.1401855 -0.25707550 -0.35140525
```

The first column of the `vectors` part of `eigen` goes with the eigenvalue of 1. If this vector is normalized to sum to 1, the following vector is obtained:

```
e <- eigen(t(A))
e$vectors[,1] / sum(e$vectors[,1])
```

```
## [1] 0.38388626 0.25592417 0.17061611 0.11374408 0.07582938
```

That looks familiar! It is the same as the *limiting* distribution found earlier.

```
A %^% 100
```

```
##           [,1]      [,2]      [,3]      [,4]       [,5]
## [1,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
## [2,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
## [3,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
## [4,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
## [5,] 0.3838863 0.2559242 0.1706161 0.1137441 0.07582938
```

Again, this is not a coincidence!

20.4 *Stationary distributions*

Suppose that a chain is started in a distribution π. If after one step in the Markov chain, it still holds that the distribution of the state is π, then call the distribution *stationary*.

Definition 72
Distribution π is a **stationary distribution** for a Markov chain if $X_t \sim \pi$ implies that $X_{t+1} \sim \pi$. (When π is given in the form of a probability vector, say that π is a stationary probability vector.)

Note that the *state* of the chain is typically changing from step to step, it is only the *distribution* of the state that is unchanging.

For instance, suppose the current state of a Rubik's cube is uniform over all possible reachable states. Turn the right hand side of the cube clockwise 90 degrees. The resulting state is not independent of the previous state, is different from the previous state, but it is straightforward to show that the distribution of the new state will still be uniform over the reachable states.

20.4.1 Another example

Suppose

$$A = \begin{pmatrix} 0.1 & 0.4 & 0.5 \\ 0.6 & 0 & 0.4 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$$

Then

$$A^{100} = \begin{pmatrix} 0.313725 & 0.254902 & 0.431373 \\ 0.313725 & 0.254902 & 0.431373 \\ 0.313725 & 0.254902 & 0.431373 \end{pmatrix}$$

So $v = (0.313725 \quad 0.254902 \quad 0.431373)$ is a limiting distribution.
Also

$$vA = v$$

So this is a stationary distribution as well!

20.5 *Limiting distributions are always stationary distributions*

The relationship between limiting distributions and stationary distributions is one of the more important facets of Markov chain theory. One direction of the relationship is easy: limiting distributions are always stationary for a Markov chain.

Fact 53
Suppose for transition matrix A there is a probability vector π such that for all probability vectors v,

$$\lim_{t\to\infty} vA^t = \pi.$$

Then π is a stationary distribution probability vector.

Proof. Recall that you can bring linear operators in and out of limits:

$$\lim_{n\to\infty} (c_1 a_n + c_2 b_n) = c_1 \lim a_n + c_2 \lim b_n.$$

Hence

$$\begin{aligned} \pi A &= \left[\lim_{t\to\infty} vA^t\right] A \\ &= \lim_{t\to\infty} vA^t A \\ &= \lim_{t\to\infty} vA^{t+1} \\ &= \pi, \end{aligned}$$

making π a stationary probability vector. □

Sadly, the reverse is not true.

20.6 *An example where a stationary distribution is not a limiting distribution*

Consider a chain with states $\{a, b, c, d\}$ and transition matrix:

$$A = \begin{pmatrix} 0.4 & 0 & 0.6 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.6 & 0 & 0.4 \end{pmatrix}.$$

From state a, the chain can only travel to a or c. Similarly c can only travel to a or c.

The states b and d form a second collection of states that can only travel to each other.

Raising this matrix to a high power reveals that all rows are *not* the same! This means that there is no limiting distribution.

$$A^{1000} = \begin{pmatrix} 0.4545 & 0 & 0.5455 & 0 \\ 0 & 0.4615 & 0 & 0.5385 \\ 0.4545 & 0 & 0.5455 & 0 \\ 0 & 0.4615 & 0 & 0.5385 \end{pmatrix}.$$

Now it is the case that the chain restricted to states$\{a, c\}$ there is a limiting distribution. Similarly, if the chain only contained states $\{b, d\}$ there is a limiting distribution. But because it has these two collections of states that cannot reach one another (called *connected components* of nodes in graph theory), there is no overall limiting distribution.

This same fact that there are two connected components means that there are an infinite number of stationary distributions! How does that work?

Each of the two types of rows in the limit of the transition matrix corresponds to a stationary distribution. So $(0.4545, 0, 0.5455, 0)$ and $(0, 0.4615, 0, 0.5385)$ are both stationary distributions.

From these two stationary distributions, it is possible to build many more! For any $\lambda \in [0, 1]$,

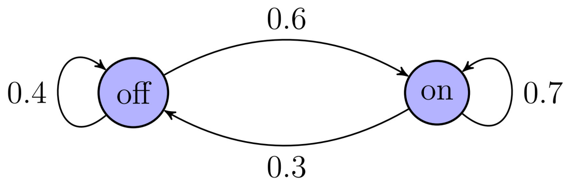
$$\pi_\lambda = \underbrace{\lambda \pi_1 + (1 - \lambda)\pi_2}_{\text{convex linear combination}}$$

is also stationary.

A Markov chain with a transition graph with two or more strongly connected components is *reducible*. Another way to describe the results of this section is that reducible Markov chains will not have a limiting distribution, but can have an infinite number of stationary distributions.

Problems

171. This Markov chain models a telecommunications circuit that is either on or off.



If the state at time 0 is off, what is the chance that in three steps the state is on?

172.

Continuing the last problem, consider the following questions.

a. Find the limiting distribution by raising the transition matrix to a high power.

b. Verify that the limiting distribution is a stationary distribution.

173.

Consider a Markov chain with transition matrix over state space $\{1, 2, 3, 4\}$:

$$A = \begin{pmatrix} 0 & 0.1 & 0.9 & 0 \\ 0 & 0.3 & 0.3 & 0.4 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}$$

a. Find the limiting distribution by raising the transition matrix to a high power.

b. Verify that the limiting distribution is a stationary distribution.

174.

Suppose a Markov chain just always jumps from one state to another, so

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

a. What is T^i where i is even?

b. What is T^i where i is odd?

c. Does this Markov chain have a limiting distribution?

175.

Consider a Markov chain with transition matrix over state space $\{a, b, c, d\}$:

$$T = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0 & 0 & 0.4 & 0.6 \end{pmatrix}$$

a. Find $\lim_{t \to \infty} T^t$ by raising T to a high power.

b. Does this chain have a limiting distribution?

176.

Consider a Markov chain with transition matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.3 & 0.7 \\ 0 & 0.5 & 0.5 \end{pmatrix}$$

a. Does it have a limiting distribution?

b. Does it have a stationary distribution?

177.

A poker player either wins \$1 or loses \$1 with equal probability. The player starts with \$2, and quits when reaching \$0 or \$5.

a. Write this down as a transition matrix where 0 and 5 are absorbing states.

b. Raise the transition matrix to a high power to discover the probability that the poker player ends up with \$5.

c. Was this what you expected from martingale theory?

178.

Consider the following transition matrix.

$$T = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}$$

a. What is(are) the absorbing state(s)?

b. What is the limiting distribution?

179.

Consider a Markov chain on state space $\Omega = \{a, b, c\}$ with transition matrix

$$\begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.8 & 0 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$

Raise the matrix to a high power.

a. Does this Markov chain have a limiting distribution?

b. Find a stationary distribution for the chain.

180. A matrix A is said to be in *block form* if the rows can be partitioned into classes such that $A(i, j) > 0$ if and only if i and j belong to the same class.

For instance,

$$A = \begin{pmatrix} 0.97 & 0.03 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0 & 0 & 0.63 & 0.37 \end{pmatrix}$$

is in block form with partition of rows $\{1, 2\}, \{3, 4\}$.

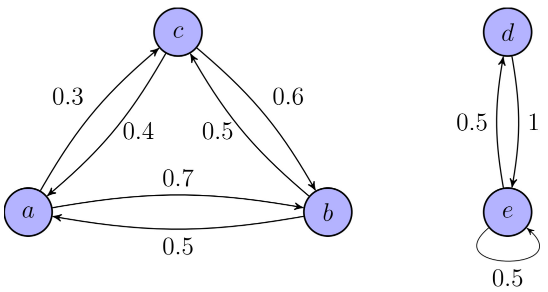
Consider

$$B = \begin{pmatrix} 0.3 & 0 & 0.7 & 0 \\ 0 & 0.9 & 0 & 0.1 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.9 & 0 & 0.1 \end{pmatrix}.$$

Can this matrix be put in block form with a nontrivial partition? If so, find the partition.

181.

Consider the following Markov chain:



(a) Write down the transition matrix for the chain. [Notice that the matrix you find is in block form.]

(b) What does the transition matrix look like after taking a lot of steps in the Markov chain?

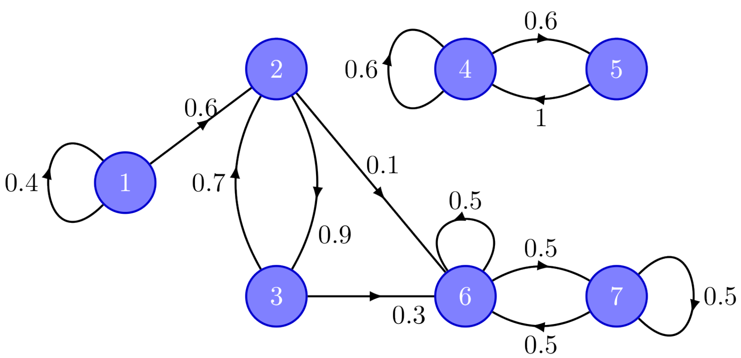
182. Give an example of a transition matrix for a four state Markov chain which has block form with row partition $\{1\}, \{2, 3, 4\}$.

Chapter 21

Recurrent and Transient states

Question of the Day

Consider the following Markov chain. For which states is the probability of returning to that state equal to 1?



A Markov chain consists of two types of states: recurrent and transient. A state is recurrent if the chain keeps returning to the state over and over again with probability 1. If a state is not recurrent, then it is transient.

21.1 Return times, recurrence, and transience

For any state x, the *return time* to x is the positive (strictly greater than 0) number of steps needed to return to state x starting from state x. Mathematically, this can be written as follows.

Definition 73
Let $R_x = \inf\{t > 0 : X_t = x | X_0 = x\}$ be the **return time** for state x.

For example:

- Sequence 01210 has return time 4
- Sequence 00 has return time 1

Note that the smallest a return time can be is 1!

Definition 74
A state x is **transient** if $\mathbb{P}(R_x < \infty) < 1$. A state x is **recurrent** if $\mathbb{P}(R_x < \infty) = 1$.

Example 20
Consider the transition graph in the Question of the Day. Starting at state 1, with probability 0.6 the state moves to 2 and then cannot get back to 1. Hence

$$\mathbb{P}(R_1 = \infty) \geq 0.6.$$

So state 1 is transient.

Now consider starting at state 4. Then either the state moves back to state 4 after one step, or it moves to state 5 and then a second step takes it back to state 4. Hence $R_4 \in \{1, 2\}$ and $\mathbb{P}(R_4 < \infty) = 1$.

21.2 Communication classes

Definition 75
States x and y of a Markov chain **communicate** if $\exists n, m \in \{0, 1, 2 \ldots\}$ such that $\mathbb{P}(X_n = y | X_0 = x) > 0$ and $\mathbb{P}(X_m = x | X_0 = y) > 0$. Write $x \leftrightarrow y$.

So $x \leftrightarrow y$ if there is a directed path (possibly of zero length) both from x to y and from y to x in the transition graph using positive probability edges.

Fact 54
Communication ($\leftrightarrow$) is an equivalence relation.

Definition 76
Say that $\equiv$ is an **equivalence relation** if it is reflexive, symmetric, and transitive.

1. Reflexive: $x \equiv x$.
2. Symmetric: $x \equiv y$ implies $y \equiv x$.
3. Transitive: $x \equiv y$ and $y \equiv z$ implies $x \equiv z$.

Definition 77
If $x \leftrightarrow y$, say that x and y are in the same **communication class**.

The next fact gives us a simpler way to establish if a state is recurrent or transient.

Fact 55
If state i is in communication class C, the state is recurrent if and only if no edges with positive probability leave C.

To simplify writing probabilities, let

$$p(i,j) = \mathbb{P}(X_{t+1} = j | X_t = i)$$

be the chance of moving from i to j in one step of the Markov chain.

Proof. Saying that a recurrent communication class has no outgoing edges is equivalent to the contrapositive statement that a communication class with an outgoing edge is transient.

Let C be a communication class with edge (i,j) where $i \in C$, $j \notin C$, $p(i,j) > 0$. Since i can reach j, but j isn't in the same communication class, so j cannot reach i. Hence $\mathbb{P}(R_i < \infty) \leq 1 - p(i,j) < 1$. So the communication class is not recurrent.

Now suppose that state i is in a communication class C with no outgoing edges. Let $n = \#C$.

Let j be any state in C reachable from i. Then since C is a communication class, there is an integer m_j such that $\mathbb{P}(X_{m_j} = i | X_0 = j) > 0$. Let $M = \max\{m_j\}$.

So no consider first taking a step in the chain. If we are back at i, then $R_i = 1$. Otherwise, after M steps there is a positive chance that we will have returned to i at least once. Let

$$\alpha = \min\{\mathbb{P}(i \in \{X_1, X_2, \ldots, X_M\} | X_0 = j)\}.$$

Then the probability we haven't returned to i in $1 + M$ steps is at most $1 - \alpha$. The probability we haven't returned to i in $1 + 2M$ steps is at most $(1-\alpha)^2$. And in general, the probability we haven't returned to i in $1 + kM$ steps is $(1-\alpha)^k$. Since that goes to 0 as $k \to \infty$, $\mathbb{P}(R_i < \infty) = 1$. □

Using this fact allows us to classify the states as follows:

1. First write down the communication classes.
2. Elements of classes with no outgoing edges are recurrent, the rest are transient.

In the QotD chain:

$$\text{Communication classes} = \{\{a\}, \{b,c\}, \{d,e\}, \{f,g\}\}$$

$$\text{transient} = \{a,b,c\}, \quad \text{recurrent} = \{d,e,f,g\}.$$

Fact 56
If any element of a class is recurrent, they all are.

Proof. This sollows directly from the previous fact: When x is recurrent, the class it is in has no outgoing edges, so every element in the class is in a class with no outgoing edges, and so is recurrent. □

This leads to the following definition. :::: {.defn data-latex=" "} If all states in a communication class are recurrent, call the class **recurrent**. Otherwise, the class is **transient**. ::::

In the Question of the Day chain, the communication classes can be sorted as follows.

Classes	Type
$\{1\}$	transient
$\{2,3\}$	transient
$\{4,5\}$	recurrent
$\{6,7\}$	recurrent

In general, a nonnegative random variable X can have $\mathbb{P}(X < \infty) = 1$, but still have $\mathbb{E}[X] = \infty$. Fortunately, when you are dealing with return times for a finite state Markov chain, you do not have to worry about this happening.

Fact 57
If state k is recurrent in a finite state Markov chain, then $\mathbb{E}[R_k] < \infty$.

Proof. Since k is recurrent, there are no outgoing edges from its communication class, which is of size at most $\#(\Omega)$. Hence after $\#(\Omega)$ steps, there is an $\alpha > 0$ chance of returning to k at least once.

Let $n = \#(\Omega)$. Then $R_k \in \{0, 1, 2, 3, \ldots\}$, so can use the tail sum formula:

$$\begin{aligned}
\mathbb{E}[R_k] &= \sum_{i=0}^{\infty} \mathbb{P}(R_k > i) \\
&= \mathbb{P}(R_k > 0) + \cdots + \mathbb{P}(R_k > n-1) + \\
&\quad \mathbb{P}(R_k > n) + \cdots + \mathbb{P}(R_k > 2n-1) + \\
&\quad \mathbb{P}(R_k > 2n) + \cdots + \mathbb{P}(R_k > 3n-1) + \cdots
\end{aligned}$$

Now $\mathbb{P}(R_k > n) \leq 1 - \alpha$, $\mathbb{P}(R_k > 2n) \leq (1-\alpha)^2$ and so on, which gives:

$$\begin{aligned}
\mathbb{E}[R_k] &\leq 1 + 1 + \cdots + 1 + \\
&\quad (1-\alpha) + (1-\alpha) + \cdots + (1-\alpha) + \\
&\quad (1-\alpha)^2 + (1-\alpha)^2 + \cdots + (1-\alpha)^2 + \cdots \\
&= n + n(1-\alpha) + n(1-\alpha)^2 + \cdots \\
&= \frac{n}{1-(1-\alpha)} = n\alpha^{-1} < \infty.
\end{aligned}$$

□

For QotD, can find these exactly:

$$\begin{aligned}
\mathbb{E}(R_d) &= \mathbb{E}(\mathbb{E}(R_d | X_1)) \\
&= \mathbb{E}(R_d | X_1 = d)\mathbb{P}(X_1 = d) + \mathbb{E}(R_d | X_1 = e)\mathbb{P}(X_1 = e) \\
&= (1)(0.7) + (1 + (1/0.6))(0.3) = \boxed{1.500},
\end{aligned}$$

where $1/0.6$ is the mean of a geometric random variable with parameter 0.6.

Fact 58
For a transient state y, for all starting states x, $\lim_{t\to\infty} \mathbb{P}(X_t = y|X_0 = x) = 0$.

Proof. Fix y, and call its transient communication class C. Note that once the state leaves C, it can never return. Since C is transient, it has at least one outgoing edge, call it (a, b).

When the state is in C, after $\#(\Omega)$ steps, there is a positive chance α of moving across the edge (a, b), never to return. Let k be a positive integer, then

$$\mathbb{P}(X_{k\#(\Omega)} \in C|X_0 \in C) \leq (1-\alpha)^k.$$

Therefore in the limit as the number of steps goes to infinity, the chance of staying in C goes to 0. □

Fact 59
Every finite state Markov chain has at least one recurrent state.

Proof. Fix $x \in \Omega$. Then

$$1 = \lim_{t\to\infty} \mathbb{P}(X_t \in \Omega|X_0 = x) = \sum_{y\in\Omega} \lim_{t\to\infty} \mathbb{P}(X_t = y|X_0 = x),$$

so there must be at least one $y \in \Omega$ with $\lim_{t\to\infty} \mathbb{P}(X_t = y|X_0 = x) > 0$. That state y is recurrent. □

Problems

183.

Suppose a Markov chain has communication classes $\{a, b\}, \{c\}, \{d, e\}$. States a, c, d are all recurrent.

a. Is state b recurrent or transient?

b. Is state e recurrent or transient?

184. Suppose a Markov chain with states $\{1, 2, 3, 4, 5, 6\}$ has communication classes $\{1, 2, 3\}$ and $\{4, 5, 6\}$. Furthermore, state 2 is transient. Classify each of the remaining states as either recurrent or transient.

185. Suppose a Markov chain over state space $\{1, 2, 3\}$ has transition matrix

$$A = \begin{pmatrix} 0.3 & 0.7 & 0 \\ 0.6 & 0.4 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Classify each state as either recurrent or transient.

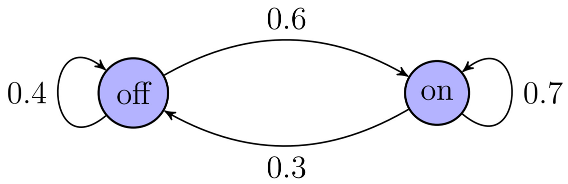
186. Suppose a Markov chain over state space $\{1, 2, 3\}$ has transition matrix

$$B = \begin{pmatrix} 0.3 & 0 & 0 & 0.7 \\ 0.6 & 0.3 & 0 & 0.1 \\ 0 & 0.4 & 0.6 & 0 \\ 0.6 & 0 & 0 & 0.4 \end{pmatrix}.$$

Classify each state as either recurrent or transient.

187.

This Markov chain models a telecommunications circuit that is either on or off.



a. What is the transition matrix for the chain?

b. Find the limiting distribution by raising the transition matrix to a high power.

c. Verify that the limiting distribution is a stationary distribution.

188. Suppose a Markov chain over state space $\{1, 2, 3, 4\}$ has transition matrix

$$B = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.6 & 0.4 \end{pmatrix}.$$

Classify each state as either recurrent or transient.

189.

Consider a Markov chain with transition matrix over state space $\{a, b, c, d\}$:

$$T = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0 & 0 & 0.4 & 0.6 \end{pmatrix}$$

a. Find $\lim_{t\to\infty} T^t$ by raising T to a high power.

b. Does this chain have a limiting distribution?

190. Consider the following transition matrix.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0 & 1 \end{pmatrix}.$$

If the matrix starts in state b, what is the chance that it reaches state a before it reaches state c?

191.

A poker player either wins \$1 or loses \$1 with equal probability. The player starts with \$2, and quits when reaching \$0 or \$5.

a. Write this down as a transition matrix where 0 and 5 are absorbing states.

b. Raise the transition matrix to a high power to discover the probability that the poker player ends up with \$5.

c. Was this what you expected from martingale theory?

192.

A poker player either wins \$1 with probability 0.6 or loses \$1 with probability 0.4. The player starts with \$2, and quits when reaching \$0 or \$5.

a. Write this down as a transition matrix where 0 and 5 are absorbing states.

b. Raise the transition matrix to a high power to discover the probability that the poker player ends up with \$5.

c. Was this what you expected from martingale theory?

193.

Consider a Markov chain on state space $\Omega = \{a, b, c\}$ with transition matrix

$$\begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.8 & 0 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$

Raise the matrix to a high power.

a. Does this Markov chain have a limiting distribution?

b. Find a stationary distribution for the chain.

194.

Consider a Markov chain on state space $\Omega = \{a, b, c\}$ with transition matrix

$$\begin{pmatrix} 0 & 0 & 1 \\ 0.1 & 0.4 & 0.5 \\ 0.2 & 0.8 & 0 \end{pmatrix}$$

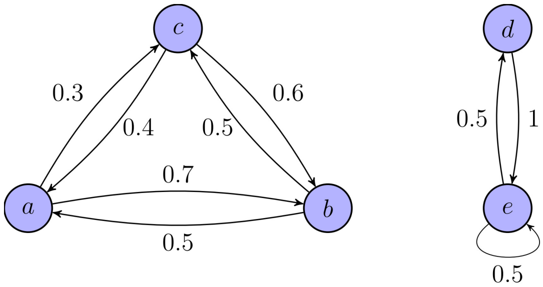
Raise the matrix to a high power.

a. Does this Markov chain have a limiting distribution?

b. Find a stationary distribution for the chain.

195.

Consider the following Markov chain:



(a) Write down the transition matrix for the chain. [Notice that the matrix you find is in block form.]

(b) What does the transition matrix look like after taking a lot of steps in the Markov chain?

196. Consider the following Markov chain transition matrix.

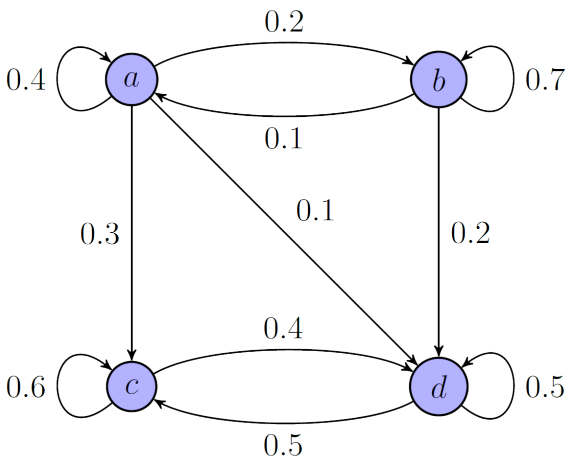
$$B = \begin{pmatrix} 0.3 & 0 & 0.7 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0.6 & 0 & 0.4 & 0 \\ 0 & 0.1 & 0 & 0.9 \end{pmatrix}.$$

This is in *block form* with rows partitioned into $\{1, 3\}$ and $\{2, 4\}$ providing two smaller submatrices, and everything else being 0.

Raise this transition matrix to a high power. Is it still in block form?

197.

Consider this Markov chain:



(a) What are the communication classes?

(b) Which communication classes are transient?

(c) What is the limiting distribution π?

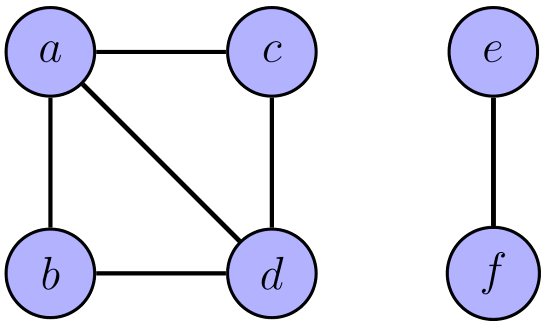
(d) What is $\pi(i)$ for the transient states i?

198. Consider the following transition matrix:

$$T = \begin{pmatrix} 0.3 & 0 & 0.7 \\ 0 & 0.5 & 0.5 \\ 0 & 0.3 & 0.7 \end{pmatrix}.$$

If π is the limiting distribution, what is $\pi(1)$?

199. In a *random walk on an undirected graph* you start at a node on the graph, and then take a step uniformly at random to move to a node that there is an edge to. For instance, consider the following graph.



The node a is connected by an edge to three other nodes, b, c, and d. Hence a has a $1/3$ chance of moving to b, a $1/3$ chance of moving to c, and a $1/3$ chance of moving to d.

Write down the transition matrix for this Markov chain.

200. Continuing the last problem, write down the communication classes for the Markov chain and state whether each communication class is recurrent or transient.

Chapter 22

Building Stationary Measures

Question of the Day

Do all finite state Markov chains have a stationary measure?

Summary

- Given a Markov chain $\mathcal{M}$ and measure μ over a finite or countable state space, set

$$\mathcal{M}(\mu)(z) = \sum_{x \in \Omega} \mu(x) p(x, z).$$

- Say that μ is a **stationary measure** if $\mathcal{M}(\mu) = \mu$.
- For a recurrent state x, let N_y be the random number of visits to state y in between visits to state x.
- Fact: $\mu(y) = \mathbb{E}_x[N_y]$ is a stationary measure for the Markov chain.
- If $\mathbb{E}[R_x] < \infty$ then $\pi(y) = \mathbb{E}_x[N_y]/\mathbb{E}[R_x]$ is a stationary distribution.

Recall that a distribution π is stationary if $X_t \sim \pi \to X_{t+1} \sim \pi$. The goal in this chapter is to show the following important fact.

> **Fact 60**
> All finite state Markov chains have at least one stationary distribution.

22.1 *The stationary measure*

A *probability distribution* or more simply *distribution* is a measure that has a finite value over the whole space. For instance, Lebesgue measure over $\mathbb{R}$ cannot be normalized to a probability distribution, since $\text{Leb}(\mathbb{R}) = \infty$. However, Lebesgue measure over $[0, 2]$ is 2, hence $\mu(A) = \text{Leb}(A \cap [0, 2])/2$ is a probability measure.

So far, Markov chains have been treated as operators that take a distribution and return another distribution, but they can also be viewed as an operator that takes a measure and returns another measure.

This works as follows. Suppose there is an amount of clay placed at each state of the chain. The total amount of clay used might be infinite (especially if the state space is countable infinite!) An atom at state a has a 20% chance of moving to b and an 80% chance of moving to state c. Then because there are so many atoms in the clay, approximately 20% of the clay ends up moving to b and 80% of the clay moves to c.

Hence if there was $\mu(a)$ weight of clay at state a, then $0.2\mu(a)$ ends up at state b and $0.8\mu(a)$ ends up at state c.

Now technically measures apply to sets, but often the goal is to understand the measure of a set that contains only a single element. In this case, it is customary to drop the curly braces around the set.

Notation 3
For a measure μ over state space Ω, let

$$\mu(i) = \mu(\{i\}).$$

Definition 78
Given a Markov chain with a finite or countably infinite state space, the **Markov chain operator on measures** $\mathcal{M}$ can be defined as

$$[\mathcal{M}(\mu)](i) = \sum_j \mu(j)\mathbb{P}(X_{t+1} = i | X_t = j).$$

Or with our p notation,

$$[\mathcal{M}(\mu)](i) = \sum_j \mu(j)p(j, i).$$

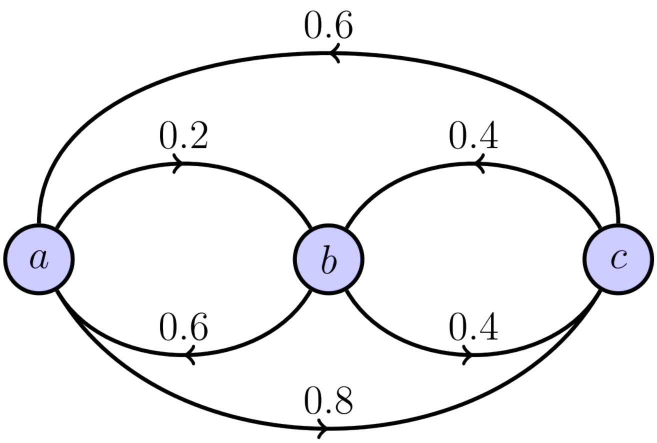
Definition 79
Say that μ is a **stationary measure** for a discrete Markov chain if

$$\mathcal{M}(\mu) = \mu.$$

Note: it is not possible to use matrix notation anymore because Ω might be countably infinite!

22.1.1 An example of measures and Markov chains

Consider the following Markov chain.



- Suppose each node a, b, c has one unit of clay on it.

$$\mu_0(a) = 1,\ \mu_0(b) = 1,\ \mu_0(c) = 1.$$

- Node a receives 0.6 clay from node b plus 0.6 clay from node c.
- Node b receives 0.2 clay from node a plus 0.4 clay from node c.
- Node c receives 0.4 clay from node b plus 0.8 clay from node a.
- New measure:

$$\mu_1(a) = 1.2,\ \mu_1(b) = 0.6,\ \mu_1(c) = 1.2.$$

22.2 *Counting visits to states on a cycle*

Now suppose that x is a state of the Markov chain. If the Markov chain starts at that state, it returns to x after R_x steps. If it never returns, then $R_x = \infty$.

Note that for $X_0 = x$ and $X_{R_x} = x$, the states $X_0, X_1, X_2, \ldots, X_{R_x}$ form a *directed cycle* in the graph,

For another state y, consider how many times the chain visits y while on this cycle that loops from x back to x.

Definition 80
Given $X_0 = x$, let N_y be the number of visits of the Markov chain to y in times $\{0, 1, 2, \ldots, R_x - 1\}$. That is,

$$N_y = \sum_{i=0}^{R_x - 1} \mathbb{1}(X_i = y | X_0 = x).$$

Having a random variable R_x in the limit of the summation is difficult to deal with. Fortunately it is possible to get rid of it in the limit by adding an indicator function in the summand.

$$N_y = \sum_{i=0}^{\infty} \mathbb{I}(X_i = y | X_0 = x)\mathbb{I}(i < R_x)$$
$$= \sum_{i=0}^{\infty} \mathbb{I}(X_i = y, i < R_x | X_0 = x).$$

Now here is the interesting fact. If you take the expected value of the number of visits to y, and make that the measure of $\{y\}$, that will be a stationary measure!

Fact 61
Fix x a recurrent state in the Markov chain. For all y, set

$$\mu(y) = \mathbb{E}_x(N_y).$$

Then μ is a stationary measure.

Proof. Start with the case $y \neq x$. Then to show stationarity, consider

$$\mathcal{M}(\mu)(z) = \sum_{y \in \Omega} \mu(y) p(y, z)$$
$$= \sum_{y \in \Omega} \mathbb{E}_{X_0 = x} \left[\sum_{i=0}^{R_x - 1} \mathbb{I}(X_i = y) \right] p(y, z).$$

The upper limit on the summation can be brought into the indicator function.

$$\mathcal{M}(\mu)(z) = \sum_{y \in \Omega} \mathbb{E}_{X_0 = x} \left[\sum_{i=0}^{\infty} \mathbb{I}(X_i = y, i < R_x) \right] p(y, z)$$

Since the terms in the infinite sum are nonnegative, the expected value can be brought inside by the Monotone Convergence Theorem. Also, $\mathbb{E}_q(\mathbb{I}(p)) = \mathbb{P}(p|q)$, so

$$\mathcal{M}(\mu)(z) = \sum_{y \in \Omega} \left[\sum_{i=0}^{\infty} \mathbb{P}(X_i = y, i < R_x | X_0 = x) \right] p(y, z)$$

Again the terms are nonnegative, which means the sums can be interchanged by Tonelli's Theorem.

$$\mathcal{M}(\mu)(z) = \sum_{i=0}^{\infty} \sum_{y \in \Omega} \mathbb{P}(X_i = y, i < R_x | X_0 = x) p(y, z)$$

Now $p(y, z) = \mathbb{P}(X_{i+1} = z | X_i = y)$, so

$$\mathcal{M}(\mu)(z) = \sum_{i=0}^{\infty} \sum_{y \in \Omega} \mathbb{P}(X_i = y, X_{i+1} = z, i < R_x | X_0 = x).$$

Note that the events $X_i = y$ are disjoint for all the different states y, and summing over all $y \in \Omega$ gives

$$\begin{aligned}\mathcal{M}(\mu)(z) &= \sum_{i=0}^{\infty} \mathbb{P}(X_i \in \Omega, X_{i+1} = z, i < R_x | X_0 = x) \\ &= \sum_{i=0}^{\infty} \mathbb{P}(X_{i+1} = z, i < R_x | X_0 = x) \\ &= \sum_{i=0}^{\infty} \mathbb{E}_{X_0=x}\left[\mathbb{I}(X_{i+1} = z, i < R_x | X_0 = x)\right]\end{aligned}$$

Again the mean can be brought outside by the MCT and the limits can be changed.

$$\begin{aligned}\mathcal{M}(\mu)(z) &= \mathbb{E}_{X_0=x}\left[\sum_{i=0}^{R_x-1} \mathbb{I}(X_{i+1} = z, i < R_x | X_0 = x)\right] \\ &= \mathbb{E}_{X_0=x}\left[\sum_{j=1}^{R_x} \mathbb{I}(X_j = z, i < R_x | X_0 = x)\right]\end{aligned}$$

Here it is important to note that the number of visits to any state in the time $0, 1, \ldots, R_x - 1$ is exactly the same as the number of visits to any state in the time $1, 2, \ldots, R_x$. That is because $X_0 = X_{R_x} = x$ in the cycle. Hence

$$\mathcal{M}(\mu)(z) = \mathbb{E}_{X_0=x}\left[\sum_{j=0}^{R_x-1} \mathbb{I}(X_j = z, i < R_x | X_0 = x)\right] = \mathbb{E}_{X_0=x}(z) = \mu(z)$$

and the measure is stationary. □

Note that any recurrent state x can be used to get a stationary measure. It turns out that two recurrent states in the same communication class will give the same stationary measure.

However, starting at states in two different communication classes will give you different stationary measures, because you will only be able to reach states (and have positive $\mathbb{E}(N_y)$ for states in your own recurrent communication class.

22.3 The stationary distribution

Now the tools are in place to show that a discrete state Markov chain with a recurrent class has a stationary distribution.

Fact 62
For a discrete state Markov chain with state x where $\mathbb{E}_{X_0=x}(R_x) < \infty$, there is always a stationary distribution.

Proof. Let x be any state in the recurrent communication class, and consider the stationary measure $\mu(y) = \mathbb{E}_x(N_y)$. By definition $\sum_{y \in \Omega} N_y = R_x$, so they have the same mean. Since Ω is finite, linearity of expectation gives:

$$\sum_{y \in \Omega} \mathbb{E}[N_y] = \mathbb{E}[R_x] < \infty,$$

where the last inequality must be true for finite state Markov chains. So finite state Markov chains always have at least one stationary distribution:

$$\pi(y) = \frac{\mathbb{E}[N_y]}{\mathbb{E}[R_x]}.$$

□

A couple notes.

- By definition, $N_x = 1$, so $\mathbb{E}[N_x] = 1$. So the stationary distribution associated with a recurrent state x is
$$\pi(x) = \frac{1}{\mathbb{E}[R_x]}.$$
- All finite state Markov chains have to have a recurrent class, so they must have a stationary distribution.
- It turns out, that if you pick $x \neq y$ in the same recurrent class, then they will lead to the same stationary distribution!

22.4 Example of a stationary measure

- Consider a Markov chain with the following transition matrix for states $\{a, b, c, d\}$:
$$\begin{pmatrix} 0.4 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$
From state a, the chain might move back to a or to state c. From b, the chain always moves to d, from c it might move to a or back to c, and from d it always moves back to state b.
  - Start with $x = b$.
  - The expected number of visits to d before returning to b is 1, because the sequence of states always goes b, d, b.
  - The expected number of visits to b before return is always 1. (In general $N_x = 1$ given $X_0 = x$.)
  - The expected number of visits to a or c are both 0.
  - So a stationary measure is $(0, 1, 0, 1)$
  - This can be normalized to a stationary distribution by noting $0 + 1 + 0 + 1 = 2$, so
  $$\frac{1}{2}(0, 1, 0, 1) = (0, 1/2, 0, 1/2)$$
  is a stationary dist.
- Now do $x = a$.
  - As before, $N_a = 1$.
  - With a 40% chance, the state moves back to a in the first step and $N_c = 0$.

  - With a 60% chance, the state first moves to c. Now there is a geometric number of steps until the move to a occurs. In this case $N_c \sim \text{Geo}(1/2)$.
  - So $\mathbb{E}[N_c] = (0.4)(0) + (0.6)(1/(1/2)) = 1.2$.
  - Final stationary measure:
$$(1, 0, 1.2, 0).$$
  - For stat. dist.:
$$\frac{1}{1 + 0 + 1.2 + 0}(1, 0, 1.2, 0) = (5/11, 0, 6/11, 0).$$

- Check answer. For μ_b:
$$\begin{pmatrix} 0 & 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0.4 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \end{pmatrix}$$

- Check answer: For μ_a:
$$\begin{pmatrix} 5/11 & 0 & 6/11 & 0 \end{pmatrix} \begin{pmatrix} 0.4 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 5/11 & 0 & 6/11 & 0 \end{pmatrix}$$

So it works! This tells us that stationary distributions exist when the expected time to return to a state is finite. But do limiting distributions? That is another story!

Problems

201. Give an example of a finite state aperiodic recurrent Markov chain (so exactly 1 recurrent communication class) with at least 4 states and stationary measure $\mu(i) = 1$ for all $i \in \Omega$.

202. Give an example of a Markov chain with two states $\{a, b\}$ such that $(1, 1/2)$ is a stationary measure.

203. Consider a Markov chain $\mathcal{M}$ on state space $\Omega = \{a, b, c\}$ with transition matrix
$$\begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.8 & 0 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$

Consider the measure
$$\mu(a) = 3,\ \mu(b) = 2,\ \mu(c) = 5.$$

What is $\mathcal{M}(\mu)$?

204. Consider a Markov chain $\mathcal{M}$ on state space $\Omega = \{a, b, c\}$ with transition matrix
$$\begin{pmatrix} 0.1 & 0.6 & 0.3 \\ 0.2 & 0.7 & 0.1 \\ 0 & 1 & 0 \end{pmatrix}$$

Consider the measure

$$\mu(a) = 1,\ \mu(b) = 5,\ \mu(c) = 10.$$

What is $\mathcal{M}(\mu)$?

205. Let M_t be a Markov chain on state space x, y, z. Suppose that $X_0 = x$, and $X_{R_x} = x$ where $R_x = \inf\{t > 0 : X_t = x\}$. During the trip from x back to x, y is visited an average of 3.2 times, and z is visited an average of 2.7 times.

Give a stationary measure for the Markov chain.

206.

Consider a four state Markov chain $\{1, 2, 3, 4\}$ where starting from state 4, state 1 is visited on average 0 times, state 2 is visited 1.8 times, and state 3 is visited 2.2 times.

a. Give a stationary measure for the Markov chain.

b. What is $\mathbb{E}[R_4]$?

207.

Suppose a Markov chain has a single communication class $\{1, 2, 3, 4\}$. If you start at state 1, the expected number of visits to the other states before returning to 1 are

$$\mathbb{E}[N_2] = 4.3$$
$$\mathbb{E}[N_3] = 1.2$$
$$\mathbb{E}[N_4] = 0.6$$

a. What is $\mathbb{E}[R_1]$?

b. Give at least one stationary measure for this chain.

208.

Suppose a Markov chain has a single communication class $\{a, b, c\}$. If you start at state c, the expected number of visits to other states before returning to c are

$$\mathbb{E}[N_a] = 7.2$$
$$\mathbb{E}[N_b] = 10.1$$

a. Give a stationary measure for this chain.

b. Normalize your stationary measure to give a stationary distribution for this chain.

Chapter 23

Stationary Distributions

Question of the Day

What are the stationary distributions of a Markov chain?

Summary

- A state is **positive recurrent** if $\mathbb{E}[R_x] < \infty$.
- All recurrent states in a finite state Markov chain are positive recurrent.
- For $\mathcal{C}$ a recurrent communication class there is exactly one stationary distribution π such that $\pi(\mathcal{C}) = 1$, and it has $\pi(x) = 1/\mathbb{E}[R_x]$ for all states x in the class.
- Simple symmetric random walk on the integers is an example of a countable state space Markov chain with one recurrent communication class that is not positive recurrent.

A stationary probability distribution π has the property that

$$\mathbb{P}(X_1 \in A | X_0 \sim \pi) = \pi(A),$$

or in other words

$$X_0 \sim \pi \to X_1 \sim \pi.$$

If you start in the stationary distribution, then after one step of the Markov chain you are still in the stationary distribution.

23.1 *Postive recurrence*

Recall the following facts about stationary measures.

- μ is a stationary measure if

$$\mu(z) = \sum_{y \in \Omega} \mu(y) \mathbb{P}(X_{t+1} = z | X_t = y).$$

- Let N_y be # of visits to y in between visits to x. Then $\mu(y) = \mathbb{E}[N_y]$ is a stationary measure.
- Also, $\sum_{y\in\Omega} \mathbb{E}[N_y] = \sum_{y\in\Omega} \mu(y) = \mathbb{E}[R_x]$ Note that we can bring the expectation inside the sum when Ω is finite by linearity of expectation, and when Ω is a countable set by the Monotonic Convergence Theorem.
- If $\mathbb{E}[R_x] < \infty$, then $\mu(z)/\mathbb{E}[R_x]$ is a stationary distribution.

It helps to have a term for when $\mathbb{E}[R_x] < \infty$.

Definition 81
A recurrent state x is **positive recurrent** if

$$\mathbb{E}[R_x] < \infty.$$

Earlier it was shown that $\mathbb{E}[R_x] < \infty$ always for recurrent states in finite state Markov chain. Given our new definition, that idea can be expressed as follows.

Fact 63
All recurrent states in a finite state Markov chain are positive recurrent.

23.2 *Uniqueness of stationary distributions concentrating on a positive recurrent communication class.*

Recall that for a given communication class, all states are either recurrent or transient. Classes with recurrent states are called recurrent.

So the next thing is to show that each recurrent communication class only has a single stationary distribution associated with it. The first step is to show that if $\pi(x) > 0$ for some x in a recurrent communication class, then $\pi(y) > 0$ for all y in that class.

Fact 64
Let $C \subseteq \Omega$ be a recurrent communication class and π be a stationary distribution. If there exists $x \in C$ with $\pi(x) > 0$, then for all $y \in C$ it holds that $\pi(y) > 0$.

Proof. Suppose $x \in C$ has $\pi(x) > 0$. Let $y \in C$. Then there exists n such that $\mathbb{P}(X_n = y|X_0 = x) > 0$. Recall if $X_0 \sim \pi$, then $X_n \sim \pi$. Put together,

$$\begin{aligned}
\pi(y) &= \mathbb{P}(X_n = y) \\
&\geq \mathbb{P}(X_n = y, X_0 = x) \\
&= \mathbb{P}(X_0 = x)\mathbb{P}(X_n = y|X_0 = x) \\
&= \pi(x)\mathbb{P}(X_n = y|X_0 = x) > 0.
\end{aligned}$$

□

Fact 65
Let C be a recurrent communication class. There is a unique stationary distribution such that $\pi(x) > 0$ for all $x \in C$ and $\pi(y) = 0$ for all $y \notin C$.

First worry about stationary measures.

Fact 66
Let C be a recurrent communication class. Then the set of stationary measures ν such that $\nu(C^C) = 0$ is unique up to constant multiples.

As usual, use the notation
$$p(x, y) = \mathbb{P}(X_1 = y | X_0 = x).$$

Proof. Let ν be a stationary measure with $\nu(C^C) = 0$. For any two states c and z in C, by definition of stationary measure,

$$\nu(z) = \sum_y \nu(y)p(y, z) = \nu(c)p(c, z) + \sum_{y \neq c} \nu(y)p(y, z).$$

We could write this same equation using different dummy variables as

$$\nu(y) = \nu(c)p(c, y) + \sum_{x \neq c} \nu(x)p(x, y).$$

Now use this identity recursively for $\nu(y)$ to get:

$$\begin{aligned} \nu(z) &= \nu(c)p(c, z) + \sum_{y \neq c} \left[\nu(c)p(c, y) + \sum_{x \neq c} \nu(x)p(x, y) \right] p(y, z) \\ &= \nu(c)p(c, z) + \sum_{y \neq c} \nu(c)p(c, y)p(y, z) + \sum_{y \neq c} \sum_{x \neq c} \nu(x)p(x, y)p(y, z) \end{aligned}$$

Use $\mathbb{P}_c$ to denote the probability distribution of the Markov chain given that $X_0 = c$ with probability 1, and use $\mathbb{P}_\nu$ to denote the probability distribution of the Markov chain given that $X_0 \sim \nu$. Then

$$\nu(z) = \nu(c) \left[\mathbb{P}_c(X_1 = z) + \mathbb{P}_c(X_1 \neq c, X_2 = z) \right] + \mathbb{P}_\nu(X_0 \neq c, X_1 \neq c, X_2 = z)$$

Repeat this recursion n times to get:

$$\nu(z) = \nu(c) \sum_{m=1}^{n} \mathbb{P}_x(X_k \neq c \, \forall k \in \{1, \ldots, m\}, X_m = z) + \sum_{x_0 \neq c} \nu(x_0)\mathbb{P}(X_0 = x_0, X_1 \neq c, X_2 \neq c, \ldots, X_{n-1} \neq c, X_n = z).$$

Formally use an induction to show this fact.

Now take the limit as n goes to infinity. Since the probability of an event is the expected value of the indicator function of the event, and indicator functions are bounded by 1 in absolute value, the limit can be brought inside the probability. In the last term, no matter what x_0 is, it bounded above by the probability that the first $n-1$ steps never hit c, and this goes to 0 as n goes to infinity since c and x_0 have to be in the same communication class for $\nu(x_0) > 0$.

The sum in the first term converges by the monotone convergence theorem to

$$\sum_{m=1}^{\infty} \mathbb{P}_c(X_k \neq c \, \forall k \in \{1, \ldots, m\}, X_m = z) = \mathbb{E}[N_z].$$

Hence

$$\nu(z) = \nu(c)\mathbb{E}[N_z].$$

For fixed $c \in C$, $\nu(c)$ is a fixed constant independent of z. So what this says is that ν must be a multiple of the stationary distribution we already know about, $\mathbb{E}[N_z]$, which is 0 for any z outside of the communication class. □

Fact 67
Let C be a recurrent communication class. Then the set of stationary distributions $\{\pi : \pi(C^C) = 0\}$ contains exactly one distribution:

$$\pi(x) = \frac{1}{\mathbb{E}[R_x]}.$$

Proof. From earlier if two stationary measures π and π' exist, then $\pi = c\pi'$ for a constant c. But $\sum_{x \in C} \pi(x) = \sum_{x \in C} c\pi'(x) = 1$, so $c = 1$ and $\pi = \pi'$.

Let x be any element of C. Then $\pi(y) = \mathbb{E}[N_y]/\mathbb{E}[R_x]$ is the unique stationary distribution, and $N_x = 1$ since x is always visited exactly once in $\{0, 1, \ldots, R_x - 1\}$. So $\pi(x) = 1/\mathbb{E}[R_x]$. □

23.3 The intuition behind the travel time/stationary distribution relationship

Consider a state x in the Markov chain that on average takes 5 steps to return to x. Then as the number of steps grows to infinity, about $1/5$ of the steps will be landing in x. That is, the limiting distribution for x will be $1/5$, or one over the average number of steps to return to x starting from x.

23.4 Example of a countably infinite chain with no stationary distribution

Consider the state space $\Omega = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$. Then have the Markov chain either add or subtract 1 from the current state with equal probability. This is called *simple symmetric random walk* on the integers.

Fact 68
Simple symmetric random walk on $\mathbb{Z}$ is a Markov chain with no stationary distribution.

Proof. Suppose $\mu(i) \geq 0$ exists such that

$$\mu(i) = (1/2)\mu(i-1) + (1/2)\mu(i+1),$$

so that

$$\mu(i+1) = 2\mu(i) - \mu(i-1)$$

or

$$\mu(i+2) = 2\mu(i+1) - \mu(i).$$

This is a recurrence relation. If $\mu(0) = \mu(1)$, then this equation gives $\mu(2) = 2\mu(0) - \mu(0) = \mu(0)$.

Then

$$\mu(3) = 2\mu(2) - \mu(1) = 2\mu(0) - \mu(0) = \mu(0).$$

An easy induction proof gives $\mu(i) = \mu(0)$ for all i. When $\mu(0) > 0$ then $\sum_i \mu(i) \neq 1$, and when $\mu(0) = 0$, $\sum_i \mu(i) = 0$. Either way, μ is not a probability distribution!

Suppose $\mu(1) \neq \mu(0)$. Then

$$\mu(2) = 2\mu(1) - \mu(0) = \mu(1) + (\mu(1) - \mu(0)).$$

Similarly

$$\mu(3) = 2\mu(2) - \mu(1) = \mu(2) + (\mu(2) - \mu(1)) = \mu(2) + (\mu(1) - \mu(0)),$$

and in general

$$\mu(i+1) = \mu(i) + (\mu(1) - \mu(0)).$$

Again an easy induction gives $\mu(i+1) = \mu(0) + i(\mu(1) - \mu(0))$. when $\mu(1) - \mu(0) > 0$ then when i is large enough $\mu(i) > 1$, and when $\mu(1) - \mu(0) < 0$ then for i large enough $\mu(i) < 0$. Either way, μ is not a probability distribution! □

This has two immediate consequences for simply symmetric random walk.

- Recall that any limiting distribution is also a stationary distribution. Therefore there is no limiting distribution for this chain, since there is no stationary distribution.
- If $\mathbb{E}[R_0] < \infty$, then it would be possible to build a stationary distribution. But there is no stationary distribution. Hence $\mathbb{E}[R_0] = \infty$, and the chain cannot be positive recurrent.

Problems

209. A finite state Markov chain has one communication class, $\{a, b, c\}$. How many stationary distributions π for the chain have $\pi(\{a, b, c\}) = 1$?

210.

Suppose that a finite state Markov chain has two recurrent communication classes, $\{a, b\}$ and $\{c, d, e\}$.

a) How many stationary distributions π for the chain are there with $\pi(\{a, b\}) = 1$?

b) How many stationary distributions π for the chain are there with $\pi(\{c, d, e\}) = 1$?

211.

State if state x is recurrent (but not positive recurrent), positive recurrent, or transient.

a. $\mathbb{E}[R_x] = 4.2$.

b. $\mathbb{E}[R_x] = \infty, \mathbb{P}(R_x < \infty) = 1$.

c. $\mathbb{P}(R_x = \infty) = 0.3$.

212.

Suppose that $\{a, b, c\}, \{d, e\}$ are communication classes where starting at state b, N_y is the random number of visits to state y before returning to state b. Classify state b as recurrent (but not positive recurrent), positive recurrent, or transient based on the following information.

a. $\mathbb{E}[N_a] = 4.2,\ \mathbb{E}[N_c] = 1.7$.

b. $\mathbb{P}(N_a = 0) = 0.8, \mathbb{P}(N_c = 0) = 0.7$ and $p(b, b) = 0$.

213. Suppose that a Markov chain has recurrent communication class $\{a, b, c\}$ with stationary distribution $(0.3, 0.2, 0.5)$. What is $\mathbb{E}[R_c]$?

214.

Suppose that a Markov chain has recurrent communication class $\{a, b, c, d\}$ and transient communication class $\{e, f\}$. The stationary distribution is $(0.25, 0.25, 0.4, 0.1, 0, 0)$.

a. What is $\mathbb{E}[R_b]$?

b. What is $\mathbb{E}[R_c]$?

Chapter 24

The Ergodic Theorem for Finite State Markov chains

Question of the Day

Finite state Markov chains always have a stationary distribution π. When is π also a limiting distribution?

Summary

- A chain has **period** k if there is a partition $P_0, \dots, P_{k-1}$ of states such that $(\forall i \in \{0, \dots, k-1\})(\forall x \in P_i)(\mathbb{P}(X_{t+1} \in P_{i+1-k\mathbb{1}(i=k-1)}) = 1)$.
- A chain with period 1 is **aperiodic**.
- A chain is **irreducible** if it consists of exactly one recurrent communication class.
- A chain is **ergodic** if it is aperiodic and irreducible.
- The **Ergodic Theorem for finite state Markov chains** says that for a finite state Markov chain the following holds.
    1) There is at least one stationary distribution π.
    2) π is unique if and only if there is exactly one recurrent communication class.
    3) If C is a recurrent communication class, then for $x \in C$ making $\pi(x) = 1/\mathbb{E}[R_x]$ and for $x \notin C$ setting $\pi(x) = 0$ is a stationary distribution.
    4) If there is one recurrent aperiodic communication class, then the stationary distribution is also the limiting distribution.
- The **total variation distance** between two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ is $\text{dist}_{\text{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{A \in F} |\mathbb{P}_1(A) - \mathbb{P}_2(A)|$.

So far the following facts about Markov chains have been established.

- Every finite state Markov chain has at least one stationary distribution.
- For applications such as shuffling cards and Markov chain Monte Carlo (MCMC), it is important to know when the stationary distribution to equal the limiting distribution.
- To answer that question, it helps to think about what can go wrong? In other words, what can prevent the limiting distribution from being the stationary distribution.
- Fortunately, there are only two problems that can come up.
  - First, there could be more than one stationary distribution, in which case the limiting distribution cannot be both!
  - Second, the Markov chain could be *periodic*, moving in a pattern between 2 or more sets of states.
- It turns out, these are the only two problems!
- To solve the first problem: make sure that there is only one recurrent communication class.
- To solve the second problem: make sure the chain is aperiodic.

24.1 Periodicity

Saying that a Markov chain has periodicity k can be thought of as having a state space which has been sliced up like a pizza.

Whenever the state of the Markov chain falls into a particular slice, it has to move clockwise to somewhere in the next slice.

Formally, the division of the state space into slices is called a *partition* of the state space. A partition consists of a set of subsets of the original set where every element of the original set is an element of exactly one element of the partition. There are several ways to describe this, one way is using indicator functions.

Definition 82
Say that $P_0, P_1, \ldots, P_{k-1}$ is a **partition** of a set A, if for all $i \in \{0, 1, \ldots, k-1\}$,

$$\sum_{a \in A} \mathbb{I}(a \in P_i) = 1.$$

Definition 83
For a recurrent communication class C, let k be the largest integer such that there is a partition C into $P_0, P_1, \ldots, P_{k-1}$ that satisfies

$$(\forall i \in \{0, 1, \ldots, k-1\})(\forall x \in P_i)(\mathbb{P}(X_{t+1} \in P_{i+1-k\mathbb{I}(i=k-1)} | X_t = x) = 1).$$

Call k the **period** of C. Say that the chain has **periodicity** k.

24.1.1 Example of a period 2 chain

Suppose that a Markov chain has state space $\{a, b, c, d\}$, and transition matrix

$$A = \begin{pmatrix} 0 & 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.6 & 0 & 0.4 \\ 0.1 & 0 & 0.9 & 0 \end{pmatrix}.$$

- In the example above $P_0 = \{a, c\}$ and $P_1 = \{b, d\}$.
- From any state in P_0 you land in P_1 with probability 1.
- From any state in P_1 you land in P_0 with probability 1.
- The period is 2.

Definition 84
A recurrent communication class C is **aperiodic** if it has period 1.

24.2 The Ergodic Theorem for finite state Markov chains

Theorem 10
Ergodic Theorem for finite state Markov chains

For finite state Markov chains

1. There is at least one stationary distribution.
2. The stationary distribution π is unique if and only if there is exactly one recurrent communication class.
3. If C is a recurrent communication class, then setting for all $x \in C$,

$$\pi(x) = \frac{1}{\mathbb{E}[R_x]}.$$

   and for all $y \notin C$, $\pi(y) = 0$, π is a stationary distribution.

4. For one recurrent aperiodic communication class,

$$(\forall x \in \Omega)(\forall A \in F)\left(\lim_{t\to\infty} \mathbb{P}(X_t \in A | X_0 = x) = \pi(A)\right)$$

Definition 85
A chain is **irreducible** if it consists of exactly one recurrent communication class. If that class is also aperiodic, then the chain is **ergodic**.

The short version of the ergodic theorem is:

> The limiting distribution equals the unique stationary distribution in a finite state Markov chain if it is irreducible and aperiodic.

The slightly longer version of ergodic theorem:

In a finite state Markov chain, the limiting distribution exists and equals the unique stationary distribution iff the chain has one recurrent communication class which is aperiodic.

24.3 *The total variation distance*

Another way to describe the limiting distribution is by using a distance between probability measures. The *total variation distance* is often used for this purpose.

Definition 86
The **total variation distance** between two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ with measurable events in $\mathcal{F}$ is

$$\text{dist}_{\text{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{a \in F} |\mathbb{P}_1(a) - \mathbb{P}_2(a)|.$$

Definition 87
A **distance** (aka a **metric**) d is a function that takes pairs of states and returns a real number that satisfies for all states x, y and z:

1. $d(x, y) \geq 0$ (nonnegativity)
2. $d(x, y) = 0 \Leftrightarrow x = y$ (identity)
3. $d(x, y) = d(y, x)$ (symmetry)
4. $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality)

It is straightforward to show that d_{TV} is a distance.

Recall: The probability distribution associated with r.v. X is

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A).$$

So for instance, the total variation distance from X to π is:

$$d_{\text{TV}}(X, \pi) = \sup_A |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$$

where $Y \sim \pi$. (Note $\mathbb{P}(Y \in A)$ is often written as $\pi(A)$ for convenience.)

With this notation, the ergodic theorem is:

Finite state Markov chains that are irreducible and aperiodic with stationary distribution π have

$$(\forall x \in \Omega) \left(\lim_{t \to \infty} d_{\text{TV}}([X_t | X_0 = x], \pi) = 0 \right).$$

24.4 *Aperiodicity*

It turns out to be easy to force a Markov chain to be aperiodic. As long as there is a state in a recurrent communication class with a positive chance of staying in the same state after one step, the class is aperiodic.

Fact 69
Suppose x is in a recurrent communication class and $\mathbb{P}(X_1 = x|X_0 = x) > 0$. Then the class is aperiodic.

Proof. Let C_1, C_2 be any partition of the state space where $x \in C_1$. Then $\mathbb{P}(X_2 \in C_2|X_1 = x) \leq 1 - \mathbb{P}(X_2 = x|X_1 = x) < 1$. Hence the chain is aperiodic. □

24.4.1 Using holding to give aperiodicity

Definition 88
A state x where $\mathbb{P}(X_1 = x|X_0 = x) > 0$ is said to have a **holding** probability.

Suppose the goal is to give states a holding probability without changing the stationary distribution π. To accomplish this, at each step, flip a fair coin. If heads, the chain stays where it is, otherwise the state changes normally according to the Markov chain.

In this new chain (sometimes called a *lazy chain*) every state has a holding probability. So in particular all the recurrent states have a holding probability.

From a linear algebra point of view, if the old transition matrix was A, the new transition matrix with holding is $(1/2)A + (1/2)I$. Note

$$\pi A = \pi \Rightarrow \pi[(1/2)A + (1/2)I] = (1/2)\pi + (1/2)\pi = \pi$$

so the lazy chain has the same stationary distributions as the original chain.

24.5 *Eigenvalues and aperiodicity*

The period can be related to the eigenvalues of the transition matrix.

- Period k chains cycle through a partition $P_0, P_1, P_2, \ldots, P_{k-1}$.
- After k steps, back in P_1.
- So raising transition matrix to k power gives a chain that starts in P_0 and returns to P_0.
- If A^k is a Markov chain on P_0 with eigenvalue 1, then A must have had an eigenvalue that is the kth root of 1.
- For example, with period 3,

  Eigenvalues $1, -1/2 \pm (\sqrt{3}/2)i$, cube roots of unity

24.6 *Periodicity through greatest common divisors*

Another way to view periodicity is through the lens of greatest common divisors. This is a more number theory way of looking at periodicity. First we need the notion of when an integer divides another integer.

A positive integer k divides a positive integer n if there exists an integer ℓ such that $k\ell = n$. Write $k|n$.

For example: 2 divides 6, 14 divides 14, and 1 divides every positive integer.

Definition 89
Let A be a set of positive integers. Then the **greatest common divisor** (or gcd) of A is

$$\text{gcd}(A) = \max\{k \in \mathbb{Z}^+ | (\forall a \in A)(k|a)\}.$$

For example, $\text{gcd}(\{3,4\}) = 1$, $\text{gcd}(\{2,4,6,\ldots\}) = 2$. Now we can find out the periodicity as a gcd.

Let a be any state in a recurrent communication class C. Let $M = \{m : \mathbb{P}(X_m = a | X_0 = a)\}$. Then the period of C equals $\text{gcd}(M)$.

Let a be a state in recurrent communication class C. Suppose that the period of C is k. Then there exists a partition $(P_0, P_1, \ldots, P_{k-1})$ of C such that

$$(\forall i \in \{0,1,2,\ldots,k-1\})(\mathbb{P}(X_1 \in P_{i+1} | X_0 \in P_i) = 1).$$

where $P_k = P_0$. In fact for $d \in \{0,1,\ldots\}$ and $i \in \{0,1,\ldots,k-1\}$ set $P_{kd+i} = P_i$. Then it is straightforward to show

$$(\forall i \in \{0,1,2,\ldots\})(\mathbb{P}(X_1 \in P_{i+1} | X_0 \in P_i) = 1).$$

Now let m be a positive integer such that $\mathbb{P}(X_m = a | X_0 = a) > 0$. Then since $X_0 = a$, $a \in P_0$ in the partition. Then it must be true that $a \in P_m$, so $P_m = P_0$. Hence $k|m$, and k is a divisor of every element of M.

This is true for any partition of the form

$$(\forall i \in \{0,1,2,\ldots,k-1\})(\mathbb{P}(X_1 \in P_{i+1} | X_0 \in P_i) = 1).$$

But the period is the greatest such k where a partition exists, and so it is the greatest common divisor of the integers in M.

Problems

215. Suppose a Markov chain has state space $\{a,b,c\}$, where $\{a,b\}$ is a recurrent communication class and $\{c\}$ is a transient communication class. Does the chain have a limiting distribution?

216. Suppose a Markov chain has state space $\{a,b,c,d,e\}$ where $\{a,b\}$ and $\{c,d\}$ are recurrent communication classes and $\{e\}$ is a transient communication class. Does the chain have a limiting distribution?

217.

Suppose $X \sim \text{Unif}(\{1,2,3\})$ and $Y \sim \text{Unif}(\{1,2,3,4,5\})$.

a. By adding in all elements with $\mathbb{P}(X = a) > \mathbb{P}(Y = a)$, find the set $A \subseteq \{1,2,3,4,5\}$ such that $\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)$ is as large as possible?

b. What is the set $B \subseteq \{1,2,3,4,5\}$ such that $\mathbb{P}(Y \in B) - \mathbb{P}(X \in B)$ is as large as possible?

c. What is the relationship between A and B?

d. What is the total variation distance between X and Y?

218. A useful fact about total variation distance for random variables with densities with respect to the same measure μ is

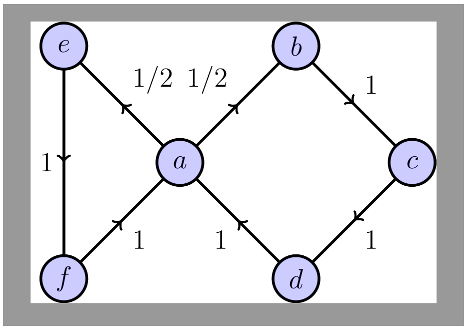
$$\text{dist}_{\text{TV}}(X, Y) = \int_{x: f_X(x) \geq f_Y(x)} f_X(x) - f_Y(y)\, d\mu.$$

As always, if μ is Lebesgue measure (for continuous random variables) this integral is the same as the Riemann integral, and for μ equal to counting measure (for discrete random variables) this integral becomes a sum.

Use this fact to find the total variation distance between $U_1 \sim \text{Unif}(\{1, 2, 3, 4\})$ with density $f_{U_1}(u) = (1/4)\mathbb{I}(u \in \{1, 2, 3, 4\})$ and $U_2 \sim \text{Unif}(\{2, 3, 4, 5, 6\})$ with density $f_{U_2}(u) = (1/5)\mathbb{I}(u \in \{2, 3, 4, 5, 6\})$.

219.

Consider the Markov chain with transition graph:



If the chain has period k, then it holds that $\{t : \mathbb{P}(X_t = a | X_0 = a) > 0\} \subseteq \{k, 2k, 3k, 4k, \ldots\}$.

(a) Find enough elements in the set $\{t : \mathbb{P}(X_t = a | X_0 = a) > 0\}$ to show that the chain is aperiodic.

(b) Find the limiting distribution of the chain.

220.

Consider a Markov chain over state space $\{a, b, c\}$ with transition matrix

$$A = \begin{pmatrix} 0 & 0.2 & 0.8 \\ 0.4 & 0 & 0.6 \\ 0 & 1 & 0 \end{pmatrix}$$

a. Find $\mathbb{P}(X_2 = a | X_0 = a)$.

b. Find $\mathbb{P}(X_3 = a | X_0 = a)$.

c. Do the previous two parts give enough information to show that the Markov chain is aperiodic?

221. Suppose that transition matrix A has eigenvalues 1 and -1, but there are no other eigenvalues that when raised to a nonnegative integer power equals 1. What is the period of A?

222. Suppose a Markov chain with transition matrix B has one recurrent comm class with period 3. How many comm classes are there in the Markov chain with transition matrix B^3?

Chapter 25

Travel Times

Question of the Day

Consider simple symmetric random walk with partially reflecting boundaries on $\{0, 1, \dots, n\}$. So

$$p(i, i+1) = (1/2)\mathbb{1}(i < n)$$
$$p(i, i-1) = (1/2)\mathbb{1}(i > 0)$$
$$p(0, 0) = 1/2$$
$$p(n, n) = 1/2$$

What is the expected number of steps needed to return to 0 starting from 0?

Summary

- One way to find expected travel times from i to j by adding an artificial state k, making $p(j, k) = p(k, i) = 1$, and using $\mathbb{E}[R_k] = 2 + \mathbb{E}[T_{i,j}]$.
- If the uniform distribution is stationary for a finite state Markov chain with all states in one recurrent communication class, then the expected return time for every state is the number of states in the space.

Consider a four state Markov chain with transition matrix

$$\begin{pmatrix} 0 & 0.4 & 0 & 0.6 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

What is the expected time to return to state a starting at state a?

- Note a and b are connected since $p(a, b)$ and $p(b, a)$ are both positive.

- Also, c and d are connected since $p(c, d)$ and $p(d, c)$ are both positive.
- Finally, a and d are connected since $p(a, d)$ and $p(d, a)$ are both positive.
- Hence $\{a, b, c, d\}$ form one communication class, and since it is the only comm class in the finite state chain, it must be recurrent.
- So there is a unique stationary distribution. To find it, solve

$$\begin{pmatrix} a & b & c & d \end{pmatrix} \begin{pmatrix} 0 & 0.4 & 0 & 0.6 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix} = \begin{pmatrix} a & b & c & d \end{pmatrix}$$

Using R

```
A <- matrix(c(0, 0.4, 0, 0.6,
              1, 0, 0, 0,
              0, 0, 0.1, 0.9,
              0.5, 0, 0.5, 0),
            byrow = TRUE,
            nrow = 4)
eigen(t(A))
```

```
## eigen() decomposition
## $values
## [1]  1.0000000 -0.9658911  0.4658911 -0.4000000
##
## $vectors
##            [,1]       [,2]       [,3]       [,4]
## [1,] -0.5731205 -0.6134566 -0.5340868  0.4526787
## [2,] -0.2292482  0.2540479 -0.4585508 -0.4526787
## [3,] -0.3820804 -0.3175599  0.5731885 -0.5432145
## [4,] -0.6877446  0.6769685  0.4194491  0.5432145
```

- The first eigenvalue is 1, so the first eigenvector in column 1, normalized to add up to 1, is

```
A <- matrix(c(0, 0.4, 0, 0.6,
              1, 0, 0, 0,
              0, 0, 0.1, 0.9,
              0.5, 0, 0.5, 0),
            byrow = TRUE,
            nrow = 4)
v <- eigen(t(A))$vectors[,1]
pi <- v / sum(v)
pi
```

```
## [1] 0.3061224 0.1224490 0.2040816 0.3673469
```

So $\pi(a) \approx 0.306122$.

By the Ergodic Theorem

$$\mathbb{E}[R_a] = \frac{1}{\pi(a)} = \frac{1}{0.3061224\ldots}.$$

```
1 / pi[1]
```

```
## [1] 3.266667
```

So the average return time to a is about $\boxed{3.266}$.

25.1 Travel times for the example

What if the goal is not to get the average return time, but instead to get the average travel time between states. For instance, what is the expected time to travel from a to c?

The idea is to alter the Markov chain somewhat. Start with the original transition matrix.

$$\begin{pmatrix} 0 & 0.4 & 0 & 0.6 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

Next, add an artificial node e. This node will help us keep track of travel from a to c. Change node c so that it always moves to e. Once at node e, always move back to node a. The new transition matrix then looks as follows:

$$\begin{pmatrix} 0 & 0.4 & 0 & 0.6 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Consider starting the new chain in state e and consider the path needed to return to e. The first move has to be to state a, then you have to travel from a to c, and then the final move will be state c back to state e. Using our earlier notation T_{ij} for the number of steps needed to move from state i to state j,

$$R_e = 2 + T_{ac}.$$

The entire state space is still connected, so the Ergodic Theorem holds, which means the eigenvalues can be used to find the stationary distribution of this new chain.

```
library(expm)
```

```
B <- matrix(c(0, 0.4, 0, 0.6, 0,
               1, 0, 0, 0, 0,
               0, 0, 0, 0, 1,
               0.5, 0, 0.5, 0, 0,
               1, 0, 0, 0, 0),
            byrow = TRUE,
            nrow = 5)
B
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  0.0  0.4  0.0  0.6    0
## [2,]  1.0  0.0  0.0  0.0    0
## [3,]  0.0  0.0  0.0  0.0    1
## [4,]  0.5  0.0  0.5  0.0    0
## [5,]  1.0  0.0  0.0  0.0    0
```

```r
eigen(t(B))
```

```
## eigen() decomposition
## $values
## [1]  1.000000e+00+0.0000000i -1.000000e+00+0.0000000i  6.938894e-18+0.5477226i
## [4]  6.938894e-18-0.5477226i  0.000000e+00+0.0000000i
##
## $vectors
##                [,1]          [,2]                         [,3]
## [1,] 0.7669650+0i -0.7669650+0i -1.266788e-16-3.761774e-01i
## [2,] 0.3067860+0i  0.3067860+0i -2.747211e-01+1.992483e-17i
## [3,] 0.2300895+0i -0.2300895+0i  1.029425e-16+3.761774e-01i
## [4,] 0.4601790+0i  0.4601790+0i -4.120817e-01+4.013026e-17i
## [5,] 0.2300895+0i  0.2300895+0i  6.868028e-01+0.000000e+00i
##                             [,4]             [,5]
## [1,] -1.266788e-16+3.761774e-01i  9.220692e-17+0i
## [2,] -2.747211e-01-1.992483e-17i  7.071068e-01+0i
## [3,]  1.029425e-16-3.761774e-01i -8.107333e-17+0i
## [4,] -4.120817e-01-4.013026e-17i  4.710277e-16+0i
## [5,]  6.868028e-01+0.000000e+00i -7.071068e-01+0i
```

This chain has eigenvalues of 1 and -1. These are the two square roots of 1, which means that this new chain has period 2. So even though there is not a limiting distribution, the stationary distribution still exists, and $\pi(i) = 1/\mathbb{E}[R_i]$.

```r
v <- eigen(t(B))$vectors[, 1]
pi <- v / sum(v)
pi
```

```
## [1] 0.3846154+0i 0.1538462+0i 0.1153846+0i 0.2307692+0i 0.1153846+0i
```

So the final answer is $1/\pi(e) - 2$. R uses brackets `[` and `]` to access entries, so

```r
1 / pi[5] - 2
```

```
## [1] 6.666667+0i
```

gives us our answer of $\boxed{6.666\ldots}$.

25.2 *First step analysis*

Recall that we already had a method for finding $\mathbb{E}[T_{a,c}]$, first step analysis! This works as follows.

- Start by letting $w_i = \mathbb{E}[T_{i,c}]$.

- Then

$$\begin{aligned} w_a &= 1 + 0.4w_b + 0.6w_d \\ w_b &= 1 + w_a \\ w_c &= 0 \\ w_d &= 1 + 0.5w_a + 0.5w_c \end{aligned}$$

- This has solution $w_a = 20/3$, so $6.666\ldots$ steps are needed on average.

It is worth noting: you should get the same answer using either method!

25.3 *Solving the Question of the Day*

Simple symmetric random walk with partially reflecting boundaries:

In order to use the Ergodic Theorem, we need to have a stationary distribution. The detailed balance equations, $\pi A = \pi$, are:

$$\mathbb{P}(X_{t+1} = j) = \sum_i \mathbb{P}(X_t = i)\mathbb{P}(X_{t+1} = j | X_t = i).$$

$$\begin{aligned} \pi(0) &= (1/2)\pi(0) + (1/2)\pi(1) \\ \pi(1) &= (1/2)\pi(0) + (1/2)\pi(2) \\ \vdots &= \vdots \\ \pi(i) &= (1/2)\pi(i-1) + (1/2)\pi(i+1) \\ \vdots &= \vdots \\ \pi(n-1) &= (1/2)\pi(n-2) + (1/2)\pi(n) \\ \pi(n) &= (1/2)\pi(n-1) + (1/2)\pi(n). \end{aligned}$$

Solving these equations one by one gives:

$$\begin{aligned} \pi(0) &= \pi(1) \\ \pi(1) &= \pi(2) \\ \pi(2) &= \pi(3) \\ \vdots &\vdots . \end{aligned}$$

So that means that the $\pi(i)$ values are all equal! Since the probability vector must sum to 1,

$$\pi(i) = \frac{1}{n+1}.$$

So the time to return to 0 is $\boxed{n+1}$ for all the states!

The time to return to 5 starting from 5 is also $n+1$, or the time to return to 100 starting from 100, also $n+1$. Wild!

Problems

223.

Consider a two state Markov chain with transition matrix.

$$A = \begin{pmatrix} 0.4 & 0.6 \\ 0.5 & 0.5 \end{pmatrix}.$$

a. What is the expected travel time from b to a?

b. What is the expected return time to a?

c. What is the stationary distribution of a?

224.

Consider a two state Markov chain with transition matrix.

$$A = \begin{pmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{pmatrix}.$$

a. What is the expected travel time from b to a?

b. What is the expected return time to a?

c. What is the stationary distribution of a?

225. Consider the following four state Markov chain transition matrix.

$$A = \begin{pmatrix} 0.1 & 0.3 & 0.4 & 0.2 \\ 0 & 0.5 & 0 & 0.5 \\ 0.3 & 0.1 & 0.2 & 0.4 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

What is the expected time needed to travel from state a to state d? Calculate this through the artificial node method.

226. Repeat the calculation of the expected time needed to travel from state a to state d using first step analysis.

227.

Consider simple symmetric random walk with partially reflecting boundaries on states $\{0, 1, \dots, 10\}$.

a. What is the expected time needed to return to state 0 starting from state 0?

b. What is the expected time needed to return to state 7 starting from state 7?

228. A Markov chain with 100 states and a recurrent communication class that is the whole state space has the uniform distribution as its unique stationary distribution. For state a, what is the expected time needed to return to a?

Chapter 26

Coupling

Question of the Day

Why does the stationary distribution equal the limiting distribution for irreducible and aperiodic Markov chains?

Summary

- A **coupling** between two random variables X and Y refers to the joint distribution (X, Y) of X and Y.
- The total variation distance between X and Y is bounded above by $\mathbb{P}(X \neq Y)$ in any coupling between the two variables.
- With exactly one recurrent aperiodic communication class, the chance that two Markov chains have not coupled goes to zero as the number of steps goes to infinity.
- This means that under these conditions, if $Y_0 \sim \pi$ for π the unique stationary distribution, then no matter what $X_0 = x$ is, the total variation distance between X_t and Y_t will go to 0. That is, π will be the limiting distribution.

So why does the limiting distribution equal the stationary distribution from a probability perspective?

Think about starting a Markov chain Y_0 in the stationary distribution π. No matter how many steps in the chain you take, Y_t will still be in the stationary distribution.

Now picture starting another copy of the Markov chain X_0 at a particular state x. As the chains run according to an update function, perhaps they meet at the stopping time T.

$$T = \inf\{t : X_t = Y_t\}.$$

Once they meet, have the token tracking the X_t process stick to the token tracking the Y_t process. that is, for $t > T$, $X_t = Y_t$. The term for linking two train cars together so that they move as one is called *coupling*, and that is the term that will be used here to describe the (X_t, Y_t) process.

Since $Y_t \sim \pi$ always, X_t inherits that stationary distribution from Y_t. Which means that the more likely it is that $X_t = Y_t$, the closer X_t is to the stationary distribution.

Before defining a coupling for Markov chains, start with a coupling for just two random variables.

Definition 90
A **coupling** of random variables X and Y is just another name for the bivariate distribution of (X, Y).

It turns out that the total variation distance between the two random variables can be upper bounded by the probability that the two random variables take on different values.

Fact 70
The Coupling Lemma

For any two random variables X and Y with coupling (X, Y),

$$\text{dist}_{\text{TV}}(X, Y) \leq \mathbb{P}(X \neq Y).$$

Proof. Recall that $\text{dist}_{\text{TV}}(X, Y) = \sup_A \mathbb{P}(X \in A) - \mathbb{P}(Y \in A)$. Let A be any measurable set. Then

$$\begin{aligned}
|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| &= |\mathbb{P}(X \in A, X = Y) + \mathbb{P}(X \in A, X \neq Y) - \\
&\quad (\mathbb{P}(Y \in A, X = Y) + \mathbb{P}(Y \in A, X \neq Y))| \\
&= |\mathbb{P}(Y \in A, X = Y) + \mathbb{P}(X \in A, X \neq Y) - \\
&\quad \mathbb{P}(Y \in A, X = Y) - \mathbb{P}(Y \in A, X \neq Y)| \\
&= |\mathbb{P}(X \in A, X \neq Y) - \mathbb{P}(Y \in A, X \neq Y)|
\end{aligned}$$

Note that for real values a and b with $0 \leq a \leq c$ and $0 \leq b \leq c$, it holds that $|a - b| \leq |c|$. Both $\mathbb{P}(X \in A, X \neq Y)$ and $\mathbb{P}(Y \in A, X \neq Y)$ are bounded above by $\mathbb{P}(X \neq Y)$. Hence

$$|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \leq \mathbb{P}(X \neq Y).$$

The right hand side is independent of A, and so the supremum of the left hand side over all A is at most $\mathbb{P}(X \neq Y)$. $\square$

Example 21
Suppose $X \sim \textsf{Unif}([0,2])$ and $Y \sim \textsf{Unif}([1,5])$.

a. Upper bound $\text{dist}_{\textsf{TV}}(X,Y)$ using the coupling where X and Y are chosen independently.

b. Let $Y \sim \textsf{Unif}([1,5])$ and $U \sim \textsf{Unif}([0,2])$ be independent. Let

$$X = U\mathbb{I}(Y \geq 2 \vee U \leq 1) + Y\mathbb{I}(Y < 2 \wedge U > 1)$$

Then it is straightforward to verify that $X \sim \textsf{Unif}([0,2])$. Use this coupling to upper bound $\text{dist}_{\textsf{TV}}(X,Y)$.

c. Let $Y \sim \textsf{Unif}([1,5])$, $B \sim \textsf{Bern}(1/3)$, $U \sim \textsf{Unif}([0,1])$ be independent. Then set

$$X = Y\mathbb{I}(Y \leq 2) + [U+B]\mathbb{I}(Y > 2).$$

Again it can be shown that $X \sim \textsf{Unif}([0,2])$. Use this coupling to upper bound $\text{dist}_{\textsf{TV}}(X,Y)$.

Solution

a. Because X and Y are independent and continuous, the probability that they are equal is 0, and the probability that they are unequal is 1. Hence $\boxed{\text{dist}_{\textsf{TV}}(X,Y) \leq 1}$. Note that this is a trivial bound since the most the total variation distance can ever be is 1!

b. Here $X = Y$ if $Y < 2, U > 1$, which happens with probability

$$\mathbb{P}(Y < 2, U > 1) = \mathbb{P}(Y < 2)\mathbb{P}(U > 1) = (1/4)(1/2) = 1/8.$$

Hence $\mathbb{P}(X \neq Y) = 7/8 = \boxed{0.8750}$, which is the upper bound given by the coupling lemma for $\text{dist}_{\textsf{TV}}(X,Y)$.

c. Here $X = Y$ when $Y < 2$, which happens with probability $(2-1)/(4-2) = 0.2500$. Hence the upper bound on $\text{dist}_{\textsf{TV}}(X,Y) = \boxed{0.7500}$.

Note that for $A = [2,5]$, $\mathbb{P}(Y \in A) - \mathbb{P}(X \in A) = 3/4 - 0 = 0.7500$, so that must be the total variation distance between X and Y.

26.1 *Coupling two stochastic processes*

Rather than just coupling a single pair of variables, it is also possible to couple two entire stochastic processes.

Definition 91
A **coupling** of two stochastic processes X_t and Y_t is a sequence $\{(X_t, Y_t)\}$ such that $\{X_t\}$ and $\{Y_t\}$ viewed by themselves have the correct distribution for their process.

26.1.1 Example: coupling simply symmetric random walks

Here is an example of how to couple simple symmetric random walks over the integers.

- Suppose $D_1, D_2, D_3, \ldots \overset{\text{iid}}{\sim} \textsf{Unif}(\{-1, 1\})$.
- $X_0 = 0$, $Y_0 = 5$, $X_{t+1} = X_t + D_{t+1}$, $Y_{t+1} = Y_t + D_{t+1}$.
- Then X_t and Y_t both move according to the same Markov chain.
- Now suppose $W_0 = 0$, $W_{t+1} = W_t - D_{t+1}$.
- Since $-D_i \sim D_i$, W_t is also a simple symmetric random walk, but it moves differently than X_t or Y_t.

Now consider a stochastic process $Y_0, Y_1, Y_2, \ldots$ where $Y_0 \sim \pi$. Then because the process is started in the stationary distribution all of the $Y_i \sim \pi$.

26.2 *Using coupling to show the ergodic theorem*

Suppose $Y_0 \sim \pi$ where π is stationary, so $Y_t \sim \pi$ for all t. Further, suppose $\mathbb{P}(X_t \neq Y_t) \to 0$. From the coupling lemma this means

$$\text{dist}_{\textsf{TV}}(X_t, Y_t) = \text{dist}_{\textsf{TV}}(X_t, \pi) \to 0,$$

that is, the distribution of X_t converges to the stationary distribution!

To see why sooner or later the X_t and Y_t processes run into each other in an irreducible, aperiodic finite state Markov chain, the following fact will be useful.

Fact 71
Let $\{X_t\}$ be an irreducible, aperiodic finite state Markov chain over Ω. For any two states x and y in Ω, there is a time t such that $\mathbb{P}(X_t = x | X_0 = x) > 0$ and $\mathbb{P}(X_t = x | X_0 = y) > 0$.

A fact from number theory (presented below without proof) is needed to prove the previous fact.

Fact 72
Let $A \subseteq \mathbb{Z}^+$ with $\gcd(A) = 1$ satisfying $(\forall a_1, a_2 \in A)(a_1 + a_2 \in A)$. Then

$$(\exists n)(\forall N \geq n)(N \in A).$$

Since y and x communicate, let c be the length of a path from y to x. Let A be the set of times t' such that $\mathbb{P}(X_{t'} = x | X_0 = x) > 0$. If $t_1, t_2 \in A$, then $\mathbb{P}(X_{t_1+t_2} = x | X_0 = x) \geq \mathbb{P}(X_{t_1} = x | X_0 = x) \cdot \mathbb{P}(X_{t_2} = x | X_0 = x) > 0$, so $t_1 + t_2 \in A$.

Then since the chain is aperiodic, $\gcd(A) = 1$, and so satisfies the condition of the number theory fact. Hence there is a time t' such that $t = t' + c$ is also in the set, which gives the desired time.

Now to prove the ergodic theorem.

Proof. Since the chain is finite there exists a recurrent state z (and class). The cycle trick proof shows that there is a stationary measure, and since $\mathbb{E}[R_z] < \infty$, it can be normalized to give a stationary distribution.

If there are at least two recurrent communication class, then the stationary measure from the cycle trick gives two stationary measures, each of which can be converted to a stationary distribution, and which give probability 1 to disjoint states. Hence the stationary distribution is not unique.

If there is one recurrent communication class.

Let x be a recurrent state. The stationary measure cycle trick normalized to a stationary distribution gives π with $\pi(x) = 1/\mathbb{E}[R_x]$. For any y that is transient, this cannot be reached from x and so the same stationary distribution has $\pi(y) = 0$. By the previous lemma this stationary distribution is unique, so this holds for any recurrent x and transient y.

Fix $x \in \Omega$ and let $X_0 = x$. Let $Y_0 \sim \pi$, where π is stationary. Then $Y_t \sim \pi$ for all t.

Advance the X_t and Y_t chains independently if $X_t \neq Y_t$, otherwise just advance X_t to X_{t+1} and set $Y_{t+1} = X_{t+1}$. With this coupling both $\{X_t\}$ and $\{Y_t\}$ are following the transition probabilities for the Markov chain. Hence

$$\text{dist}_{\text{TV}}(X_t, \pi) = \text{dist}_{\text{TV}}(X_t, Y_t) \leq \mathbb{P}(X_t \neq Y_t).$$

Now, for every pair of states $x, y \in \Omega$, there is a time $t_{x,y}$ such that there is a positive chance of moving to a state z from both states x and y in that many steps. So let

$$\alpha_{x,y} = \min\{\mathbb{P}(X_{t_{x,y}} = z | X_t = x), \mathbb{P}(X_{t_{x,y}} = z | X_t = y)\} > 0.$$

So after $t_{x,y}$ steps, there is at most a $1 - \alpha_{x,y}$ chance that $X_{t_{x,y}} \neq Y_{t_{x,y}}$. Let $\alpha = \min_{x,y} \alpha_{x,y}$ and $t = \max_{x,y} t_{x,y}$.

Then for any $k \in \{1, 2, 3, \ldots\}$, $\mathbb{P}(X_{kt} \neq Y_{kt}) \leq (1 - \alpha)^k \to 0$. □

Problems

229. Suppose $A \sim \text{Unif}(\{1, \ldots, 90\})$ and $B \sim \text{Unif}(\{10, 11, \ldots, 110\})$ have a coupling where there is a 0.4 chance the states are equal. What can be said about the total variation distance?

230. Suppose $A \sim \text{Unif}(\{1, \ldots, 90\})$ and $B \sim \text{Unif}(\{10, 11, \ldots, 110\})$ have a coupling where there is a $81/101$ chance the states are equal. What can be said about the total variation distance?

231. Suppose $X \sim \text{Exp}(1/2)$ and $Y \sim \text{Unif}([0, 1])$ have a 0.1 chance of being equal with a certain coupling. What can be said about the total variation distance?

232. Suppose $A \sim \text{Unif}(\{1, \ldots, 90\})$ and $B \sim \text{Unif}(\{10, 11, \ldots, 110\})$ have a coupling where there is a $9/10$ chance the states are equal. What can be said about the total variation distance?

233.

Consider $X \sim \text{Unif}([0, 2])$ and $Y \sim \text{Unif}([1, 3])$. Note that $\mathbb{P}(X \in [1, 2])$ and $\mathbb{P}(Y \in [1, 2])$ both are $1/2$. Also recall that for W a uniform random variable over set B, for $A \subseteq B$, $[W | W \in A] \sim \text{Unif}(A)$. This gives a way to couple uniform random variables.

Let $B \sim \text{Bern}(1/2)$ and $U \sim \text{Unif}([0, 1])$

a. Let $X = 2U$ and $Y = 1 + 2U$. What is $\mathbb{P}(X \neq Y)$?

b. Let $X = B(1 + U) + (1 - B)(U)$ and $Y = B(1 + U) + (1 - B)(2 + U)$. What is $\mathbb{P}(X \neq Y)$?

234. Suppose $A \sim \text{Unif}([0, 10])$ and $B \sim \text{Unif}([6, 16])$. Find a coupling (perhaps by drawing extra random variables) such that $\mathbb{P}(A \neq B) = 0.6$.

235.

Suppose that $X_0 = x_0$ and $Y_0 \sim \pi$ are two copies of a Markov chain which has π as the stationary distribution. Suppose $\mathbb{P}(Y_0 \neq X_0) \leq 10 \exp(-t/5)$.

a. What can be said about $\text{dist}_{\text{TV}}(X_t, Y_t)$ when $t = 20$?

b. What can be said about $\text{dist}_{\text{TV}}(X_t, Y_t)$ when $t = 30$?

236.

Suppose that $X_0 = x_0$ and $Y_0 \sim \pi$ are two copies of a Markov chain which has π as the stationary distribution. Suppose $\mathbb{P}(Y_0 \neq X_0) \leq 100 \exp(-t/10)$.

a. What can be said about $\text{dist}_{\text{TV}}(X_t, Y_t)$ when $t = 60$?

b. What can be said about $\text{dist}_{\text{TV}}(X_t, Y_t)$ when $t = 70$?

c. What can be said about $\text{dist}_{\text{TV}}(X_t, Y_t)$ when $t = 80$?

Chapter 27

Coupling for mixing times

Question of the day

Simple symmetric random walk with partially reflecting boundaries on the integers $\{0, 1, \ldots, n-1\}$ has update function

$$f(x, u) = [x + \mathbb{1}(u < 0.5) - \mathbb{1}(u \geq 0.5)] \wedge (n-1) \vee 0.$$

Bound the mixing time of this Markov chain in terms of n.

Summary

- The **mixing time** τ_ϵ of a Markov chain is the minimum number of steps needed to move from any starting state to within ϵ total variation distance of the limiting distribution.
- Using the coupling lemma, any particular coupling gives an upper bound on τ_ϵ.

Recall that for irreducible, aperiodic Markov chains, the limiting distribution equals the unique stationary distribution. Roughly speaking, the mixing time of a Markov chain is how many steps must be taken before the chain forgets the state that it is currently in. Or in other words, it is how many steps must be taken before the distribution of the state is close to stationarity.

Definition 92
For a Markov chain with unique stationary distribution π, let

$$\tau_{x,\epsilon} = \inf\{t : \text{dist}_{\text{TV}}(X_t | X_0 = x, \pi) \leq \epsilon\}.$$

Call

$$\tau_\epsilon = \sup_{x \in \Omega} \tau_{x,\epsilon}$$

the **mixing time** of the Markov chain.

Recall the coupling lemma says that $\text{dist}_{\text{TV}}(X_t | X_0 = x, \pi) \leq \mathbb{P}(X_t \neq Y_t | X_0 = x, Y_0 \sim \pi)$.

One way to build a coupling for two Markov chains is to use an update function to run both chains using the same source of randomness.

Let $R \in \Omega_R$ be a source of randomness. Recall that $f : \Omega \times \Omega_R \to \Omega$ is an *update function* for a time-homogeneous Markov chain over Ω if f is a computable function that satisfies $f(x, R) \sim [X_1 | X_0 = x]$.

In other words, an update function takes the current state plus some random choices, and returns the next state in the Markov chain. For the simple symmetric random walk on $\{0, \ldots, n-1\}$,

a simple update function is

$$f(x, U) = x + D \cdot \mathbb{I}(x + D \in \{0, \ldots, n-1\})$$

where $D \sim \text{Unif}\{-1, 1\}$. The indicator function makes the partially reflecting boundaries: if we try to move outside of the state space, then the indicator function forces us to stay at the current state instead.

Now we can couple (X_t, Y_t) together by saying for all t,

$$X_{t+1} = f(X_t, D_{t+1}) \text{ and } Y_{t+1} = f(Y_t, D_{t+1}).$$

In other words, use the same random choice to move both X_t and Y_t.

For example, say $n = 4$ so $\Omega = \{0, 1, 2, 3\}$.

t D_t (X_t, Y_t) :—: :—-: :———: 0 $(0, 2)$ 1 +1 $(1, 3)$ 2 +1 $(2, 3)$ 3 -1 $(1, 2)$ 4 +1 $(2, 3)$ 5 +1 $(3, 3)$

so coupling occurred at $t = 5$.

Notice that for this chain, $X_t \leq Y_t$ implies $X_{t+1} \leq Y_{t+1}$. So if $X_t = n-1$, then $Y_t = n-1$, and $T_C \leq t$ where

$$T_C = \inf\{t : X_t = Y_t\}$$

we have for the stopping time

$$T = \inf\{t : X_t = n-1\},$$

that $T_C \leq T$.

To bound $\mathbb{E}[T]$ (and hence $\mathbb{E}[T_C]$), we consider a third process whose update function is

$$W_{t+1} = W_t + D_{t+1}\mathbb{I}(W_t + D_{t+1} \geq 0),$$

so it is like the X_t process but can keep going up forever.

Still, $W_{t \wedge T} = X_{t \wedge T}$, so studying the W_t process can help us understand the X_t process. Consider a *potential function* ϕ that will be constructed so that

$$\mathbb{E}[\phi(W_{t+1}) | \phi(W_t)] = \phi(W_t) + 1.$$

First a starting point. Set $\phi(0) = 0$. Next, consider that

$$\mathbb{E}[\phi(W_1) | W_0 = 0] = (1/2)\phi(0) + (1/2)\phi(1)$$

so to make this 1 larger than $\phi(0)$, set $\phi(1) = 2$.

Next

$$\begin{aligned}\mathbb{E}[\phi(W_1)|W_0 = 1] &= (1/2)\phi(0) + (1/2)\phi(2)\\ &= \phi(0) - 2 + (1/2)\phi(2),\end{aligned}$$

so to make this 1 more than $\phi(1)$ set $\phi(2) = 6$.

Continuing in this way, one gets the recursion

$$\phi(i+1) = 2\phi(i) + 2 - \phi(i-1).$$

and the first few entries are:

i	0	1	2	3	4	5
$\phi(i)$	0	2	6	12	20	30

Fact 73
The solution to the recursion

$$\phi(i+1) = 2\phi(i) + 2 - \phi(i-1),$$

with $\phi(0) = 0$ is

$$\phi(i) = i(i+1).$$

Proof. Usually proofs of this sort are by induction, and this is no exception!

The base case holds because $\phi(0) = (0)(1) = 0$ and $\phi(1) = (1)(2) = 2$.

Now suppose it holds for $\phi(i)$ and $\phi(i-1)$ and consider $\phi(i+1)$. Then

$$\begin{aligned}\phi(i+1) &= 2\phi(i) + 2 - \phi(i-1)\\ &= 2(i)(i+1) + 2 - (i-1)(i)\\ &= 2i^2 + 2i + 2 - i^2 + i\\ &= i^2 + 3i + 2\\ &= (i+1)(i+2),\end{aligned}$$

completing the induction. □

Then it follows directly that

$$M_t = \phi(W_t) - t$$

is a martingale, so $M_{t\wedge T}$ is also a martingale. Hence for all t

$$\mathbb{E}[M_{t\wedge T}|M_0 = 0] = 0,$$

and so

$$\mathbb{E}[\phi(W_{t\wedge T}) - (t \wedge T)] = 0,$$

and

$$\mathbb{E}[\phi(W_{t\wedge T)}] = \mathbb{E}[(t \wedge T)]$$

By the Ergodic Theorem, $\mathbb{P}(T < \infty) = 1$, so

$$\lim_{t\to\infty} \phi(W_{t\wedge T}) = \phi(W_T) = n(n+1).$$

Similarly,

$$\lim_{t\to\infty} t \wedge T = T.$$

By the bounded convergence theorem,

$$\lim_{t\to\infty} \mathbb{E}[\phi(W_{t\wedge T)}] = \mathbb{E}[\lim_{t\to\infty} \phi(W_{t\wedge T})] = n(n+1),$$

and by the monotone convergence theorem

$$\lim_{t\to\infty} \mathbb{E}[t \wedge T] = \mathbb{E}[\lim_{t\to\infty}(t \wedge T)] = \mathbb{E}[T].$$

Hence

$$\mathbb{E}[T] = n(n+1).$$

So that tells us the average number of steps needed to couple. Can we use that to bound the mixing time? Recall Markov's inequality:

For an integrable random variable X,

$$\mathbb{P}(|X| \geq a) \leq \mathbb{E}[X]/a.$$

So $\mathbb{P}(T > 2n(n+1) \leq n(n+1)/[2n(n+1)] = 1/2$. In other words, after every $2n(n+1)$ steps, we have at least a $1/2$ chance of coupling. Hence

$$\mathbb{P}(T_C > k \cdot 2n(n+1) \leq \mathbb{P}(T_C \leq \lfloor k \rfloor 2n(n+1)) \leq (1/2)^{\lfloor k \rfloor - 1} \leq (1/2)^{k-2}.$$

for all $k > 0$. Then to make this at most ϵ,

$$k = \ln(1/\epsilon)/\ln(2) + 2,$$

so the mixing time is

$$\boxed{\tau_\epsilon \leq 2n(n+1)[2 + \ln(1/\epsilon)/\ln(2)].}$$

Problems

237. Suppose two copies of a Markov chain $\{X_t\}$ and $\{Y_t\}$ are coupled together so that each follows the same transition matrix.

Further, say that $X_0 = x_0$ where $x_0 \in \Omega$, $Y_0 \sim \pi$, and $\mathbb{P}(X_{100} = Y_{100} | X_0 = x_0) = 0.99$. Bound the total variation distance between X_{100} and π.

238. Suppose two copies of a Markov chain $\{X_t\}$ and $\{Y_t\}$ are coupled together so that each follows the same transition matrix.

Further, say that $X_0 = x_0$ where $x_0 \in \Omega$, $Y_0 \sim \pi$, and $\mathbb{P}(X_{50} = Y_{50} | X_0 = x_0) = 0.9$. Bound the total variation distance between X_{50} and π.

239. Suppose that for all states x and y,

$$\mathbb{P}(X_{100} = Y_{100} | X_0 = x, Y_0 = y) = 0.1,$$

and that if $X_t = Y_t$, then $X_{t'} = Y_{t'}$ for all $t' \geq t$. Create an upper bound for

$$\mathbb{P}(X_{200} = Y_{200})$$

240. Suppose that for all states x and y,

$$\mathbb{P}(X_{100} = Y_{100} | X_0 = x, Y_0 = y) = 0.1$$

and that if $X_t = Y_t$, then $X_{t'} = Y_{t'}$ for all $t' \geq t$.

Create an upper bound for

$$\mathbb{P}(X_{400} = Y_{400})$$

241. Suppose (X_t, Y_t) is a coupling for two copies of a Markov chain such that

$$\mathbb{P}(X_t \neq Y_t) \leq 100 \exp(-t/10).$$

Give an upper bound on $\tau_{0.05}$.

242. Suppose (X_t, Y_t) is a coupling for two copies of a Markov chain such that

$$\mathbb{P}(X_t \neq Y_t) \leq 50 \exp(-t/7).$$

Give an upper bound on $\tau_{0.05}$.

243.

Suppose that

$$(\forall x_0 \in \Omega)(\text{dist}_{\text{TV}}([X_t | X_0 = x_0], \pi) \leq 1000 \exp(-t/100)).$$

a. Give an upper bound on $\tau_{0.01}$.

b. Give an upper bound on $\tau_{0.000001}$.

244.

Suppose that

$$(\forall x_0 \in \Omega)(\text{dist}_{\text{TV}}([X_t | X_0 = x_0], \pi) \leq 50 \exp(-t/16)).$$

a. Give an upper bound on $\tau_{0.01}$.

b. Give an upper bound on $\tau_{0.000001}$.

Chapter 28

Countable state space Markov chains

Question

When does the stationary distribution of a countable state space Markov chain equal the limiting distribution?

Summary

- A countable state space Markov chain has limiting distribution equal to the stationary distribution if it consists of one communication class which is positive recurrent and aperiodic.

- Simple symmetric random walk on the integers with partially reflecting boundary at 0 is recurrent but not positive recurrent.

Intuitively, a set is countable if its elements can be put into a sequence, that is $A = \{a_1, a_2, a_3, \ldots\}$. Formally, this can be written in terms of a function from $\{1, 2, 3, \ldots\}$ onto the set. Recall that a function $f : S \to T$ is onto if for every $t \in T$ there exists $s \in S$ such that $f(s) = t$.

Definition 93
A set A is **countable** if there exists a function $f : \{1, 2, \ldots\} \to A$ that is onto.

With this definition, finite sets are also countable. It helps to have a term to distinguish sets that are countable but not finite.

Definition 94
A set is **countably infinite** if it is countable but not finite.

Not all sets are countable.

Definition 95
A set is **uncountable** if it is not countable.

Famously, the real numbers, or even a finite length interval such as $[0, 1]$ are uncountable.

How does the Ergodic Theorem change for countably infinite state spaces? For some results the proofs given for finite state spaces go through nicely in the countably infinite case.

- The cycle trick still works to make a stationary measure for a recurrent state.
- For a recurrent communication class C (so there is $x \in C$ with $\mathbb{P}(R_x < \infty) = 1$), every stationary measure that puts positive measure on the states in C and 0 everywhere else has a unique stationary measure up to nonzero constants.
- For a positive recurrent communication class C (so there is $x \in C$ with $\mathbb{E}[R_x] < \infty$), there is a unique stationary distribution with $\pi(C) = 1$ defined as $\pi(y) = 1/\mathbb{E}[R_y]$ for all $y \in C$.
- The Coupling Lemma can be used to show that if C is not only positive recurrent, but also aperiodic, then

$$(\forall x_0 \in C)\left(\lim_{t \to \infty} \text{dist}_{\text{TV}}([X_t | X_0 = x_0], \pi) = 0\right).$$

Putting these facts together gives the Ergodic Theorem for countable state spaces.

Theorem 11
Consider a Markov chain with countable state space Ω.

1) If there is a recurrent communication class C, there is a unique stationary distribution π with $\pi(C) = 1$ if and only if C is positive recurrent. Moreover, $\pi(x) = 1/\mathbb{E}[R_x]$ for all $x \in C$.

2) Let C be a positive recurrent communication class with stationary distribution π. Then C is aperiodic if and only if

$$(\forall x \in C)(\lim_{t \to \infty} \text{dist}_{\text{TV}}([X_t | X_0 = x], \pi) = 0).$$

Note that if there is only one communication class and it is positive recurrent, this implies that π is the limiting distribution.

28.1 The balance equations

The equations for a stationary measure applied to the stationary distribution are referred to as the *balance equations.*

Definition 96
A distribution π on countable state space Ω is **stationary** if

$$(\forall i \in \Omega)(\pi(i) = \sum_j \pi(j)\mathbb{P}(X_1 = i | X_0 = j)).$$

Call these the **balance equations**.

28.1.1 Example: Simple symmetric random walk on the integers with partially reflecting boundary at $\{0\}$

Suppose the state space is $\{0, 1, 2, \ldots\}$ and

$$p(0,0) = 1/2$$
$$p(0,1) = 1/2$$

and for all $i \in \{1, 2, \ldots\}$,

$$p(i, i-1) = 1/2$$
$$p(i, i+1) = 1/2$$

This chain is recurrent, but not positive recurrent! Let X_t be the state of the chain at time t.

To see why this chain is recurrent, use a martingale approach. If the chain moves from 0 back to 0, then $R_0 = 1$. If the chain first moves to 1, then the question is what is the probability that $T = \inf\{t : X_t = 0 | X_1 = 1\}$ is finite?

Let $D_1, D_2, \ldots$ be iid $\mathsf{Unif}(\{-1, 1\})$. Then set

$$M_t = m_0 + \sum_{i=1}^{t} D_t$$

Note for $t < T$, $M_t = X_t$. Now let

$$T_i = \inf\{t : X_t \in \{0, i\} | X_1 = 1\}$$

where $i \in \{2, 3, 4, \ldots\}$. Then note that

$$(T = \infty)$$

if and only if

$$X_{T_2} = 2, X_{T_3} = 3, X_{T_4} = 4, \ldots.$$

That is, the only way that $T = \infty$ if the X_t process reaches every integer at least 2 before it reaches 0. That means

$$\mathbb{P}(T = \infty) \leq \mathbb{P}(X_{T_i} = i)$$

for all i.

Now the stopped process $X_{t \wedge T_i}$ is bounded by i, and so uniformly integrable. This means the Optional Sampling Theorem applies, and

$$\mathbb{E}[X_{T_i}] = \mathbb{E}[X_1] = 1,$$

Using $\mathbb{E}[X_{T_i}] = \mathbb{P}(X_{T_i} = i)i + \mathbb{P}(X_{T_i} = 0)(0)$ gives

$$\mathbb{P}(X_{T_i} = i) = 1/i$$

for all $i \in \{2, 3, \ldots\}$.

At this point $\mathbb{P}(T = \infty) \leq 1/i$ for all integer i at least 2, so must be 0. That is, the chain is recurrent!

However, it is not positive recurrent. To see why, suppose π is a positive measure that satisfies the balance equations. Then

$$\pi(0) = (1/2)\pi(0) + (1/2)\pi(1)$$
$$\pi(1) = (1/2)\pi(0) + (1/2)\pi(2)$$
$$\pi(2) = (1/2)\pi(1) + (1/2)\pi(3)$$

and so on. Solving these one at a time gives

$$\pi(0) = \pi(1)$$
$$\pi(2) = \pi(0)$$
$$\pi(3) = \pi(0)$$

and so on. If $\pi(0) = 0$, then the result is just the trivial measure. If $\pi(0) > 0$, then the sum of the $\pi(i)$ is infinite, making this *not* a stationary distribution.

So there is no stationary distribution, and the chain is not positive recurrent.

28.1.2 Example: biased random walk

Now consider the Markov chain with the same state space but biased to move left.

$$p(0,0) = 2/3$$
$$p(0,1) = 1/3$$

and for all $i \in \{1, 2, ...\}$,

$$p(i, i-1) = 2/3$$
$$p(i, i+1) = 1/3$$

Then again writing down the balance equations gives

$$\pi(0) = (2/3)\pi(0) + (2/3)\pi(1)$$
$$\pi(1) = (1/3)\pi(0) + (1/3)\pi(2)$$
$$\pi(2) = (1/3)\pi(1) + (1/3)\pi(3)$$

and so on. Solving these one at a time gives

$$\pi(0) = 2\pi(1)$$
$$\pi(1) = 2\pi(2)$$
$$\pi(2) = 2\pi(3)$$

and so on.

Using that $\pi(0) + \pi(1) + \cdots = 1$ gives

$$\pi(0)[1 + 1/2 + 1/4 + \cdots] = 1,$$

which gives $\pi(0) = 1/2$ and

$$\pi(i) = (1/2)^i$$

for all other i.

This gives that the communication class is positive recurrent!

Problems

245.

State if the following are true or false. (You do not need to justify your answer.)

a. In finite state Markov chains, recurrent communication classes are also positive recurrent.

b. In countable state space Markov chains, recurrent communication classes are also positive recurrent.

c. Positive recurrent communication classes are always aperiodic.

d. In a countable state space Markov chain with one recurrent communication class, there is always a stationary measure.

246.

State if the following are true or false.

a. There is a version of the Ergodic Theorem for countable state space Markov chain.

b. Finite sets are also countable.

c. In a countable state space Markov chain with one recurrent communication class, there is always a stationary distribution.

d. Positive recurrence can be shown by showing recurrence and finding a stationary distribution.

247. Give an example of a Markov chain with a countably infinite state space that has one recurrent communication class of period 2.

248. Give a countably infinite Markov chain with period 3.

249.

Consider a Markov chain on the nonnegative integers that has $\pi(i) = (1/2)^{i+1}$ as a stationary distribution.

a. Is this enough information to find $\mathbb{E}[R_0]$? If so, what is it?

b. What if we add the extra condition that $(\forall i, j)(\exists t)(\mathbb{P}(X_t = j | X_0 = i) > 0)$, and $\mathbb{P}((\exists t)(X_t = 0 | X_0 = 0)) = 1$? Now do we have enough information to find $\mathbb{E}[R_0]$, and if so, what is it?

250.

Consider a Markov chain on the positive integers where $\pi(i) = (1/2)^i$ is a stationary distribution.

a) Does this mean the chain has one communication class?

b) If the answer to a) is yes, this answer: must the chain be periodic? If the answer to a) is no, give an example of a chain that has this stationary distribution but more than one communication class.

251. Suppose a Markov chain on $\Omega = \{2, 4, 6, \ldots\}$ has stationary distribution

$$\pi(i) = \frac{C}{i^2},$$

where C is a constant.

Suppose

$$(\forall i \in \Omega)(\mathbb{P}(X_1 = i | X_0 = 2) > 0),$$

and

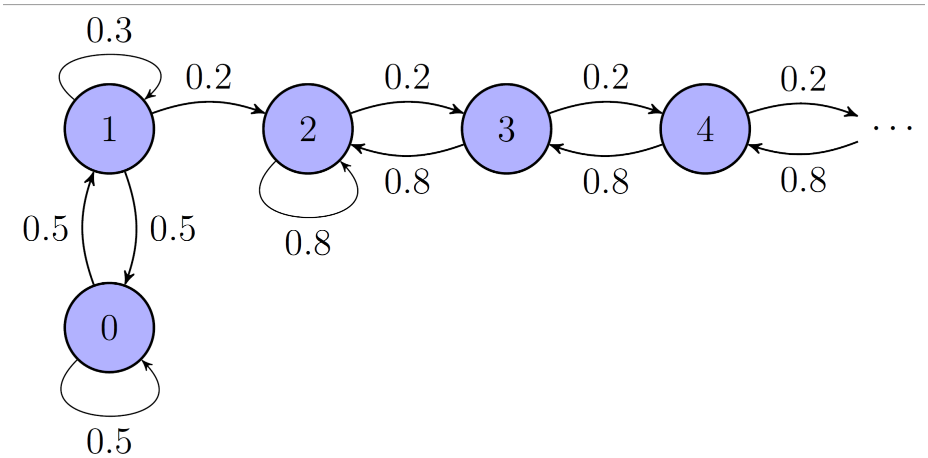
$$(\forall i \in \Omega)(\mathbb{P}(X_1 = 2 | X_0 = i) > 0),$$

Prove that this chain has one positive recurrent communication class.

252. Suppose that for state space $\Omega = \{1, 2, 3, \ldots\}$, $p(1,1) = p(1,2) = 1/2$ and for all $i \geq 2$, $\mathbb{P}(i,1) = p(i, i+1) = 1/2$. Prove that this chain has exactly one communication class.

253.

Consider the following Markov chain:



a. What are the communication classes of the chain?

b. Write down the balance equations for the chain.

c. Does a stationary distribution for this chain exist? If so, what is it?

d. For each communication class, state if it is recurrent or transient.

254. Suppose that the state space of the chain is $\{0, 1, 2, \ldots\}$ and that

$$p(0,0) = 0$$
$$p(0,1) = 1$$
$$p(i,i-1) = 0.3$$
$$p(i,i+1) = 0.7$$

for $i \geq 1$.

Write down the balance equations and solve to find the stationary distribution.

Chapter 29

General State Space Markov chains

Question of the day

When is the stationary distribution also limiting for a continuous state space?

Summary

- A **Harris chain** is a Markov chain with two properties. First, there exists a set A such that from any state x in the chain, there is a number of steps t where $\mathbb{P}(X_t \in A | X_0 = x) > 0$. Second, there exists $\epsilon > 0$ and probability distribution ρ such that for all $x \in A$ there is a probability distribution ν_x such that $[X_1 | X_0 = x] = \epsilon\rho + (1 - \epsilon)\nu_x$.
- A Harris chain is **recurrent** if for all $x \in A$, if $X_0 = x$, then the probability that the chain returns to A in one or more steps is 1.
- A recurrent Harris chain is **aperiodic** if for all $x \in A$, if $X_0 = x$, then there exists n such that for all $n' \geq n$, the probability that the chain returns to A in n' steps is positive.
- **Ergodic Theorem for Harris chains** (aka the **Fundamental Theorem of Markov chains**: Suppose $\mathcal{M}$ is an aperiodic recurrent Harris chain with stationary distribution π and for all states x, the probability the chain reaches A is 1. Then π is also the limiting distribution.

In practice, Markov chains are used to model a wide variety of situations. In particular, for most statistics applications the state space is some subset of n dimension real space, that is, $\mathbb{R}^n$. This is an example of a *continuous* space. The rules for when a stationary distribution is also limiting will be different for such a space than for a countable state space.

There are multiple ways to approach this problem. One way is to try to make the coupling argument from earlier work. That is, can a chain over a continuous state space be built so that it is possible to couple two copies of the Markov chain so that they end up meeting with probability 1? This particular type of Markov chain will be called a *Harris chain*.

29.1 *Harris chain*

A Harris chain is a Markov chain with two properties.

1. There is a special set A such that from any starting state, there is a positive chance of moving to A within a finite number of steps.
2. If the state is in A, then there is an ϵ chance that the chain can "forget" the exact location of the state within A in deciding the next move.

This idea leads to the following definition.

Definition 97
A Markov chain $\{X_t\}$ over state space Ω is a **Harris chain** if there exists a measurable set $A \subseteq \Omega$, $\epsilon > 0$, and a probability measure ρ where

1. For $T_A = \inf\{t \geq 0 : X_t \in A\}$,
$$(\forall z \in \Omega)(\mathbb{P}(T_A < \infty | X_0 = z) > 0).$$
2. For all $x \in A$ there is a distribution ν_x such that
$$[X_1 | X_0 = x] = \epsilon\rho + (1 - \epsilon)\nu_x.$$

The Harris chain definition encompasses two parts:

1. The first condition is the continuous equivalent of the requirement that the chain should consist of only one communication class.
2. The second condition will allow the chain to couple effectively. It says that for $x \in A$, the distribution of $[X_1|X_0 = x]$ has an ϵ chance of being ρ, which does not depend on the value of x! With probability $1 - \epsilon$, the distribution of $X_1|X_0 = x$ is ν_x, which does depend on x.

So it turns out that a countable state space where every state communicates with every other state is always a Harris chain.

Fact 74
Any irreducible finite state or countable state space chain with one communication class is a Harris chain.

Proof. Let A be any state x in the chain. Let $B = \{y : \mathbb{P}(X_1 = y|X_0 = x) > 0\}$, $\rho(C) = \mathbb{P}(X_1 \in C|X_0 = x)$, and $\epsilon = 1$.

Since all states communicate with x, $\mathbb{P}(T_A < \infty|X_0 = z) > 0$ for all states z. Also, for $C \subseteq B$,

$$\mathbb{P}(X_1 \in C|X_0 = x) = (1)\rho(C),$$

so the choice of ϵ and ρ works as well. □

29.1.1 Example: Random walk on $\mathbb{R}$

Another example of a Harris chain is the following random walk on the real line.

- Let $R_1, R_2, \ldots$ be iid $\mathsf{Unif}([-1, 1])$.
- Then create a Markov chain by letting $X_0 = 0$ and for all $t \in \{0, 1, 2, \ldots\}$, set $X_{t+1} = X_t + R_{t+1}$.
- It does not work to set $A = \{0\}$, since chain returns to exactly 0 with probability 0!
- Instead, give A positive Lebesgue measure. For instance, let $A = [0, 1]$. Then it turns out that from any state x, there is a finite length path that has positive probability of landing the state in A.
- If $X_t = z > 0$, then consider the event that

$$R_t, R_{t+1}, \ldots, R_{t+\lfloor z \rfloor} \in [-1, -(z-1)/\lfloor z \rfloor].$$

Then

$$X_{t+\lfloor z \rfloor} = X_t + R_{t+1} + \cdots + R_{t+\lfloor z \rfloor} \in [z - \lfloor z \rfloor, z - (z-1)] \in [0, 1].$$

This means that

$$\begin{aligned} \mathbb{P}(T_A < \infty) &\geq \mathbb{P}(R_t, R_{t+1}, \ldots, R_{t+\lfloor z \rfloor} \in [-1, -(z-1)/\lfloor z \rfloor]) \\ &= \mathbb{P}(R_t \in [-1, -(z-1)/\lfloor z \rfloor])^{\lfloor z \rfloor} \\ &> 0. \end{aligned}$$

- Notice from any point in A, there is at least a $1/2$ chance that the next point is also in $[0, 1]$. So make $B = [0, 1]$, and $\epsilon = 1/2$. Let $\rho(B) = \mathsf{Unif}(B)$. (Usually this is a good choice for ρ.)
- Let $C \subseteq B$. What is $\mathbb{P}(X_1 \in C | X_0 = x)$, where $x \in A$?
- If $x = 0$, $X_1 \sim \mathsf{Unif}([-1, 1])$.
- If $x = 1$, $X_1 \sim \mathsf{Unif}([0, 2]$.
- For any $x \in [0, 1]$, there is a $1/2$ chance that $X_1 \in [0, 1]$, so a $1/2$ chance that X_1 is uniform over $[0, 1]$.
- So $\mathbb{P}(X_1 \in C | X_0 = x) = (1/2)\rho(C)$ for all $x \in [0, 1]$.
- That makes this Markov chain a Harris chain.

29.2 *Why this definition?*

This definition of Harris chain will lead to a limiting distribution. The proof is once again by coupling.

When $X_t \in A$, there is an ϵ chance that the next state comes from ρ, and is independent of the current state! To make coupling happen, the chain can be simulated in the following way.

First, some notation.

$$\begin{aligned} \tau_x(D) &= \mathbb{P}(X_1 \in D | X_0 = x) \\ \psi_x(D) &= (\tau_x(D) - \epsilon\rho(D))/(1 - \epsilon) \end{aligned}$$

If $x \in A$, then either the chain forgets which part of A it was in with probability ϵ, and with probability $1 - \epsilon$ it remembers. Here ψ_x is the distribution of the next state if it does not forget where it is, and ρ is the distribution if it does forget.

Then the distribution of the next state is a *mixture* of these two distributions. The distribution τ_x is a convex linear combination of ψ_x and ρ, with weight ϵ for ρ and $1 - \epsilon$ for ψ_x.

Given this setup, here is how to advance the original Harris chain to the next state written as an algorithm.

One coupled step in a Harris chain 1) If $X_t \notin A$ 2) Draw $X_{t+1} \leftarrow \tau_{X_t}$ 3) Else 4) Draw $B \leftarrow \text{Bern}(\epsilon)$ 5) If $B = 1$ 6) Draw $X_{t+1} \leftarrow \rho$ 7) Else 8) Draw $X_{t+1} \leftarrow \psi_{X_t}$

Note that for any measurable set D,

$$\tau_x(D) = \epsilon\rho(D) + \psi_x(D),$$

so this way of updating the chain value is valid.

29.3 Coupling X_t

Simulating the chain in this way allows a coupling argument to be used. Suppose X_t and Y_t are individual copies of the Markov chain that need to be coupled together. The key observation is that if $X_t \in A$ and $Y_t \in A$, and $B = 1$, then at the next time step X_{t+1} and Y_{t+1} can both be chosen according to ρ.

When this happens $X_{t+1} = Y_{t+1}$ and coupling has occurred!

If is does not happen the chains both move independently according to τ_{X_t} and τ_{Y_t}. The chain then takes another step and so on. Eventually, they will come together with probability 1 as long as the Harris chain keeps returning the states to the set A. This will happen as long as the chain is recurrent and aperiodic. The definitions are similar to the finite and countable cases.

Definition 98
Let $R = \inf\{n > 0 : X_n \in A\}$. A Harris chain is **recurrent** if for all $x \in A$, $\mathbb{P}(R < \infty | X_0 = x) = 1$. A Harris chain that is not recurrent is **transient**.

Definition 99
A recurrent Harris chain is **aperiodic** if for all $x \in \Omega$, there exists n such that for all $n' \geq n$,

$$\mathbb{P}(X_{n'} \in A | X_0 = x) > 0.$$

As with finite state Markov chains, the easiest way to get aperiodicity is for every state in A to have a positive probability of holding.

Fact 75
Suppose for all $a \in A$ (from the Harris chain definition), $\mathbb{P}(X_1 \in A | X_0 = a) > 0$. Then X_t is aperiodic.

Proof. Fix $x \in \Omega$. Then for some n, $\mathbb{P}(X_n \in A | X_0 = x) > 0$. Then no matter where $X_n \in A$ is, there is a positive chance of landing in A at the next step. An induction yields

$$\mathbb{P}(X_{n'} \in A | X_0 = x) > 0$$

for all $n' \geq n$. □

The bad news is that even with aperiodicity and recurrence, we do not get everything in the ergodic theorem with Harris chains that we got in the finite or countable state space case. However, we do get the most important thing, which is that if we have a stationary distribution, then it will also be the limiting distribution.

Theorem 12
Ergodic Theorem for Harris chains

Let X_n be an aperiodic recurrent Harris chain with stationary distribution π. If $\mathbb{P}(R < \infty | X_0 = x) = 1$ for all x, then as $t \to \infty$,

$$d_{\text{TV}}([X_t | X_0 = x], \pi) \to 0.$$

29.4 *A tale of three Ergodic Theorems*

Recall that all states in a communication class communicate, and that the return time to a state x is $R_x = \inf t > 0 : X_t = x | X_0 = x$. If $\mathbb{P}(R_x < \infty) = 1$ for x in communication class $\mathcal{C}$, then $\mathcal{C}$ is recurrent. If $\mathbb{E}[R_x] < \infty$ for x in communication class $\mathcal{C}$, then $\mathcal{C}$ is positive recurrent. Both recurrence and positive recurrence are class properties, meaning that either every state in the class has the property or none of them do.

A recurrent communication class is aperiodic if there exists a state x in the class and a time t such that for all $t' \geq t$, $\mathbb{P}(X_{t'} = x | X_0 = x) > 0$.

The three Ergodic Theorems can be viewed as follows.

- In the Ergodic Theorem for finite Markov chains, the following holds.
  - If there is one recurrent communication class, a unique stationary distribution π exists.
  - If the one recurrent communication class is aperiodic then π is also the limiting distribution.
- In the Ergodic Theorem for countable Markov chains
  - If there is one recurrent communication class, a unique stationary measure μ exists.
  - For one recurrent communication class there is a stationary distribution if and only if the class is positive recurrent.
  - Need existence of π (equivalent to positive recurrence.)
  - If the one positive recurrent communication class is aperiodic then π is also the limiting distribution.
- In Ergodic Theorem for Harris chains
  - Consider a recurrent, aperiodic Harris chain.
  - Suppose there is a stationary distribution π.

– Then the limit of the distribtion of the state will be π starting from any state that reaches the small set A in finite time with probability 1.

Because the Ergodic theorem for Harris chains is the most general of the three statements, this is also called the **Fundamental Theorem of Markov chains**.

29.4.1 Example of finding π

Consider random walk on $[0, \infty]$ with partially reflecting boundaries. Let $X_0 = 1$, $D_1, D_2, \ldots$ be iid $\textsf{Unif}([-1, 1])$, and

$$X_{t+1} = X_t + D_{t+1}(\mathbb{I}(X_t + D_{t+1} > 0)).$$

The goal is to show that $\pi \sim \textsf{Unif}([0, 10])$ is stationary for this Markov chain.

Start by assuming that $X_0 \sim \textsf{Unif}([0, 10])$. Then the probability that the next state is at most a can be calculated as follows.

First consider if $a \in [1, 9]$. Then

$$\begin{aligned}
\mathbb{P}(X_0 + D_1 \le a) &= \mathbb{E}(\mathbb{I}(X_0 + D_1 \le a)) \\
&= \mathbb{E}(\mathbb{E}(\mathbb{I}(X_0 + D_1 \le a | D_1))) \\
&= \mathbb{E}(\mathbb{E}(\mathbb{I}(X_0 \le a - D_1 | D_1))) \\
&= \mathbb{E}(\mathbb{P}(X_0 \le a - D_1 | D_1)) \\
&= \mathbb{E}\left(\frac{a - D_1}{10 - 0}\right) \\
&= \frac{a}{10}
\end{aligned}$$

since $\mathbb{E}[D_1] = 0$.

That is the cdf for a uniform over $[0, 10]$! It can be shown that $\mathbb{P}(X_1 \le a) = a/10$ for $a \in [0, 1)$ and $a \in [9, 10)$ with similar calculations, but there are a lot more cases to consider.

Problems

255.

State whether or not the following are true. You do not have to justify your answer.

a. All connected countable state space chains are Harris.

b. All Harris chains have a countable state space.

c. Periodicity is not necessary for a Harris chain to have a limiting distribution.

d. Harris chains always return to the set A in the definition with probability 1.

256.

State whether or not the following are true.

a. Harris chains cannot have an uncountable state space.

b. The number of steps to return to the small set A always has finite expected value.

c. Some Harris chains have stationary distribution equal to the limiting distribution.

d. There is a version of the Ergodic Theorem for Harris chains.

257. Show that if $U_1, \ldots, U_{10}$ are iid $\mathsf{Unif}([0,1])$, then $\mathbb{P}(U_1 + \cdots + U_{10} \in [6, 6.1]) > 0$.

258. Suppose that $T_1, T_2, \ldots, T_5$ are iid $\mathsf{Exp}(1/2)$. Prove that

$$\mathbb{P}(T_1 + \cdots + T_5 \in [10, 10.5]) > 0.$$

259. Let $a \leq a' \leq b' \leq b$. Show that if $X \sim \mathsf{Unif}([a,b])$, then $[X | X \in [a', b']] \sim \mathsf{Unif}([a', b'])$.

260. Let $T_1, T_2, \ldots$ be iid $\mathsf{Exp}(1/2)$. Set $X_0 = x$ where $x > 10$. For $t > 0$, let

$$X_{t+1} = X_t + (T_{t+1} - 1)\mathbb{I}(X_t + T_{t+1} - 1 > 0).$$

Show that for $x \in [0, \infty)$, there exists s such that $\mathbb{P}(X_s \in [0,1] | X_0 = x) > 0$. (This is the set A that makes this a Harris chain.)

Chapter 30

Countably infinite Harris Chains

Question of the Day

Consider a chain is over state space $\{0, 1, 2, \ldots\}$ where $p(0, 0) = 2/3$ and for $i \geq 1$, $p(i, i-1) = 2/3$ and $p(i, i+1) = 1/3$. What is $\mathbb{E}[R_0]$?

Summary

- To verify that a countably infinite chain has a limiting distribution, first show the chain is one recurrent class, then find a stationary distribution π using the balance equations, then finally use $\pi(i) = 1/\mathbb{E}[R_i]$ to find $\mathbb{E}[R_i]$.

- The Ergodic Theorem for Harris chains can be proved using coupling in a similar fashion to that of the Ergodic Theorem for finite state Markov chains.

In the question of the day, any two states x and y can reach each other in at most $|x - y|$ steps. So the state space consists of one communication class. But is that class transient or recurrent?

30.1 The Question of the Day chain is recurrent

Suppose $X_0 = 0$. If $X_1 = 0$ it holds that $R_0 = 1 < \infty$. If $X_1 = 1$, then let $T_{0,a} = \inf\{t \geq 1 : X_t \in \{0, a\}\}$. Then $X_{t \wedge T_{0,a}}$ is bounded by a, hence uniformly integrable. It is easy to verify that $X_{t \wedge T}$ is a supermartingale. Hence the Optional Sampling Theorem applies to say,

$$\mathbb{E}[X_T] \leq 1,$$

and the Martingale Convergence Theorem gives $\mathbb{P}(T < \infty) = 1$.

Hence

$$\mathbb{P}(X_T = 0)(0) + \mathbb{P}(X_T = a)(a) \leq 1,$$

and $\mathbb{P}(X_T = a) \leq 1/a$.

For $R_0 = \infty$, it must hold that $X_{T_{0,a}} = a$ for all a. Hence $\mathbb{P}(R_0 = \infty \leq 1/a$ for all a. The only way that can be true is if $\mathbb{P}(R_0 = \infty)$.

30.2 Is the chain aperiodic?

The fact that $p(0, 0) = 2/3 > 0$ is enough to make the chain aperiodic.

30.3 Is the chain positive recurrent?

Once the communication class has been shown to be recurrent, is it positive recurrent?

To show this, it is necessary to find a solution to the balance equations. For the Question of the Day, these are

$$\begin{aligned} \pi(0) &= (2/3)\pi(1) + (2/3)\pi(0) \\ \pi(1) &= (2/3)\pi(2) + (1/3)\pi(0) \\ \pi(2) &= (2/3)\pi(3) + (1/3)\pi(1) \\ \vdots &= \vdots . \end{aligned}$$

Solving the first equation gives $\pi(1) = (1/2)\pi(0)$, which plugged into the second equation gives $\pi(2) = (1/2)\pi(1)$, and so on. Therefore

$$\pi(0) = (1/2)^{i+1}$$

is a stationary distribution.

Therefore, the Ergodic Theorem applies, which gives that $\mathbb{E}[R_0] = 1/(1/2) = 2$.

30.4 Proving the Ergodic Theorem for Harris chains

Recall that the distribution of X_1 given $X_0 = x$ is a mixture of two distributions, the first of which depends on x and the second which does that. That is,

$$[X_t | X_0 = x] \sim \epsilon\rho + (1 - \epsilon)\nu_x.$$

For measurable D and known ρ, ν_x can be found by solving the above expression to get

$$\nu_x(D) = (\mathbb{P}(X_1 \in D | X_0 = x) - \epsilon\rho(D))/(1 - \epsilon).$$

Now build a Harris chain X_t as follows. Given X_t, let W_{t+1} be a draw from ν_{X_t}. Let R_t be an independent draw from ρ, and B_t an independent Bernoulli random variable with mean (⊠).

For each t let Z_{t+1} be an independent draws from ν_{X_t}. Let W_{t+1} be a draw from the same distribution as X_{t+1} given $X_t = x$.

Then X_{t+1} can be set as follows:

$$X_{t+1} = W_{t+1}\mathbb{I}(X_t \notin A) + \mathbb{I}(X_{t+1} \in A)[R_{t+1}B_{t+1} + Z_{t+1}(1 - B_{t+1})]$$

In words, this says that if the current state is not in A, move forward one step normally. But if the current state is in A, first flip a coin with probability ϵ of heads.

If the coin is heads, then draw the next state according to ρ. Note that if this happens, the exact value of $X_t \in A$ is unimportant, no matter where in A you start, at the next step the chain moves to Y_t. If the coin is tails, draw the next state from the leftover distribution ν_x.

With this coupling of the chain, the proof that the stationary distribution is limiting for positive recurrent aperiodic Harris chains is as follows.

Proof. Fix x. Let $X_0 = x$ and $Y_0 \sim \pi$. Then the probability that X_t and Y_t hits A in finite time is 1. Since the chain is recurrent, Y_t will hit A infinitely often with probability 1 after that. Because the chain is aperiodic, after the first time X_t hits A, there exists n such that for $n' \geq n$, $\mathbb{P}(X'_n \in A) > 0$.

Taken together, that means there is a time t such that $\mathbb{P}(Y_t \in A, X_t \in A) = c > 0$. If $B_{t+1} = 1$ (which happens with probability ϵ), then using the coupling given above ensures that $X_{t+1} = Y_{t+1}$.

Suppose $B_{t+1} = 0$. Ater t more steps, the two chains have another chance to couple. After m such chances, the probability that $X_t \neq Y_t$ is at most $(1-c)^m$. Since this goes to 0 as m goes to infinity, as t goes to infinity the chance that X_t and Y_t have not met goes to 0. The Coupling Lemma then gives that the total variation distance between X_t and Y_t goes to 0. Since Y_t always has the stationary distribution, the proof is complete. □

Problems

261. Suppose that $p(a,a) = 0.3$ and the communication class containing a is recurrent. What can you say about the period of the class containing a?

262.

Consider a Markov chain where $\{a, b, c, d\}$ form a recurrent communication class.

Now consider a new chain that with probability 0.5 takes a step in the first chain, and with probability 0.5 stays at the current state.

a) Is $\{a, b, c, d\}$ still a recurrent communication class in the new chain?

b) If yes, what is the period of the class?

263. Suppose that $\{1, 2, 3, \ldots\}$ is a recurrent communication class in a Markov chain with stationary distribution π where $\pi(4) = 0.3$. What can you say about $\mathbb{E}[R_4]$?

264. Suppose that $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$ is a recurrent communication class in a Markov chain with stationary distribution π where $\pi(0) = 0.5$. What can you say about $\mathbb{E}[R_0]$?

265.

Suppose that the state space is $\{0, 1, 2, \ldots\}$ with (for $i \geq 1$) $p(0,0) = p(i, i-1) = 0.9$ and $p(i, i+1) = 0.1$.

a. Find the unique solution to the balance equations.

b. Find $\mathbb{E}[R_1]$ exactly.

266.

Suppose that the state space is $\{0, 1, 2, \ldots\}$ with (for $i \geq 1$) $p(0, 0) = p(i, i-1) = 0.6$ and $p(i, i+1) = 0.4$.

a. Find the unique solution to the balance equations.

b. Find $\mathbb{E}[R_1]$.

Chapter 31

The branching process

Question of the Day

How can the growth of a population be modeled as a stochastic process?

Summary

- If a population consists of independent individuals that at each generation each individual has a random number of children drawn from the same distribution, then the size of the population for all times forms a **branching process**.

- A branching process goes **extinct** when the population reaches 0. The probability a branching process goes extinct is called the **extinction probability**.

- For a random variable X, the function

$$\text{gf}_X(a) = \mathbb{E}[a^X]$$

is called the **generating function** of the random variable X. By convention $0^0 = 1$ so $\text{gf}_X(0) = \mathbb{P}(X = 0)$.

31.1 Making a fission bomb

The Manhattan project in the United States was the effort to build a nuclear weapon during World War II. The goal was to construct a *fission bomb*.

31.1.1 How fission bombs work

- Need a collection of atoms whose nuclei are prone to splitting in half.

- Uranium-235 (aka U-235) has 92 protons and 143 neutrons in its nucleus.

- When an extra neutron is added, the result is an extremely unstable isotope called U-236.

- U-236 quickly splits into Barium-141 and Krypton-92 and three more neutrons.
- Those neutrons might go on to split more U-235 nuclei.
- When this occurs with more and more nuclei splitting, this is called a *chain reaction.*
- If chain reaction dies out, the bomb fizzles out.
- Otherwise, BOOM!

31.2 Form of a branching process

There are a number of situations similar to this fission process. A burning tree might light one or more other trees on fire in a forest. In a population of bacteria, an individual bacterium might split into two or die, leaving no descendants. In an epidemic, an infected individual might infect a random number of people before their infection is defeated.

One model for these situations is a *branching process.* In this type of process, there is a starting population of independent individuals.

At each generation, the following happens.

1. Each individual in that generation independently has a random number of children, drawn from the same distribution
2. The generation dies out, leaving only the children.

Here *children* means the descendants of the current generation. For instance, in the fission bomb example a child is a neutron. In an epidemic model, if person A gives person B a disease, then B is the child of A. So the point is it doesn't have to be an actual child to be called that in the branching process.

- Each neutron has 0 or 3 children.
- Each of those goes on to have a random number of children.
- Suppose there are six neutrons at stage t ($X_t = 6$).
- Then the number of neutrons at stage $t + 1$ is

  $$X_{t+1} = \sum_{i=1}^{6} D_i,$$

  where D_i are iid and in $\{0, 3\}$.

The following is the formal definition of a branching process.

A **branching process** is a special kind of Markov chain with state space $\{0, 1, 2, \ldots\}$ where for $t, i \geq 0$, the random variables $Y_{t,i}$ are iid and

$$X_{t+1} = \sum_{i=1}^{X_t} Y_{t+1,i}.$$

- Let $\mu = \mathbb{E}[Y_{t,i}]$ denote the average # of children an individual has.

- Then
$$\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}[X_t|X_{t-1}]] = \mathbb{E}[X_{t-1}\mu] = \mu\mathbb{E}[X_{t-1}].$$

- This forms the basis of an induction proof that
$$\mathbb{E}[X_n] = \mu^n \mathbb{E}[X_0].$$

For a branching process where each individual has on average μ children,
$$\mathbb{E}[X_n] = \mu^n \mathbb{E}[X_0].$$

Note that if $\mu < 1$, μ^n is rapidly going towards zero!

If $\mu < 1$, then as $n \to \infty$, $\mathbb{E}[X_n] \to 0$, and $\mathbb{P}(X_n > 0) \to 0$.

Proof. Since $X_n \geq 0$, Markov's inequality applies: $\mathbb{P}(X_n \geq 1) \leq \mathbb{E}[X_n]/1$. Since $\mu < 1$, as $n \to \infty$,
$$\mu^n \to 0 \Rightarrow \mathbb{E}[X_n] \to 0 \Rightarrow \mathbb{P}(X_n \geq 1) \to 0.$$
□

A branching process goes **extinct** if there is some n for which $X_n = 0$.

When $\mu < 1$, the processes goes extinct with probability 1. What about when $\mu = 1$? $\mu > 1$? Notation:
$$\begin{aligned} a_n(k) &= \mathbb{P}(X_n = 0|X_0 = k) \\ a(k) &= \lim_{n\to\infty} a_n(k) \\ p(i) &= \mathbb{P}(X_1 = i|X_0 = 1). \end{aligned}$$

- So $a(k)$ is the probability that you go extinct starting with k people.
- Recall that individuals reproduce independently.
- So for k people to go extinct, each individuals line must fail.
- Hence $a(k) = a(1)^k$.
- If $p(i) = 0$, then never go extinct (everyone has at least one child).
- From now on, assume $p(0) > 0$.

The probability the stochastic process goes extinct is the **extinction probability**.

Let $a = a(1)$ be the extinction probability.

Let's do a little first step analysis!
$$\begin{aligned} a &= \mathbb{P}(\exists n : X_n = 0|X_0 = 1) \\ &= \sum_i \mathbb{P}(\exists n : X_n = 0|X_1 = i)\mathbb{P}(X_1 = i) \\ &= \sum_i a(i)p(i) = \sum_i a^i p(i) = \mathbb{E}[a^{Y_{1,1}}]. \end{aligned}$$

For $a \neq 0$,

$$\mathrm{gf}_X(a) = \mathbb{E}[a^X]$$

is the **generating function** of the random variable X. For $a = 0$, $\mathrm{gf}_X(0) = \mathbb{P}(X = 0)$.

31.2.1 Notes

- $\mathrm{gf}_X(0) = \mathbb{P}(X = 0)$ makes generating function continuous where it exists.
- Recall the moment generating function is $\mathrm{mgf}_X(t) = \mathbb{E}[e^{tX}]$.
- So generating function is mgf with $a = e^t$.

Notation: let $\phi(a) = \mathbb{E}[a^{Y_{1,1}}]$.

For ϕ the generating function of the number of children in a branching process, the following holds.

1. $\phi(0) = p(0)$.
2. $\phi'(1) = \mu$.
3. $\phi(1) = 1$.
4. $\phi''(a) \geq 0$ for all $a \in [0, 1]$.

Proof. Recall $0^a = 0$ for any $a \neq 0$, and $0^0 = 1$. So immediately

$$\mathrm{gf}_X(0) = 1 \cdot \mathbb{P}(X = 0),$$

and using branching processes notation, this gives

$$\phi(0) = p(0).$$

For the second fact, recall that it is legal to differentiate power series term by term inside the series' radius of convergence. Since $\phi(1) = \mathbb{E}[1] = 1$, the radius of convergence is at least 1. Hence for all $a < 1$,

$$\begin{aligned}\phi'(a) &= (d/da)\left[p(0) + \sum_{i=1}^{\infty} a^i p(i)\right] \\ &= \sum_{i=1}^{\infty} i a^{i-1} p(i) = \mathbb{E}[X \cdot a^{\max(X-1,0)}]\end{aligned}$$

For $a < 1$ this is bounded above by $\mathbb{E}[X] = \mu$, so the dominated convergence theorem applies and taking the limit as $a \to 1^-$ gives $\phi'(1) = \mu$.

The third fact is just

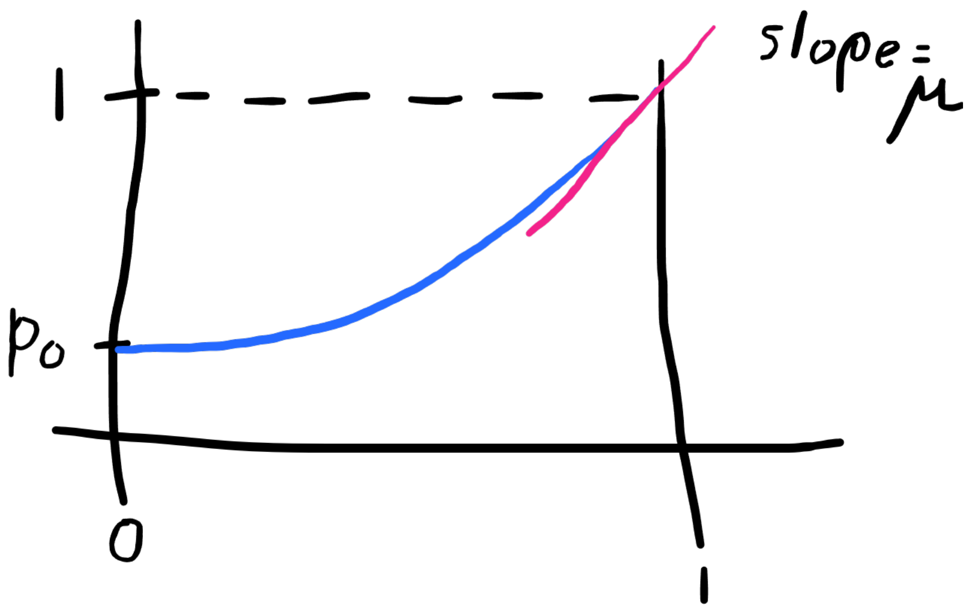
$$\phi(1) = \mathbb{E}[1^X] = 1,$$

and the last fact differentiates term by term again to get:

$$\phi''(a) = \sum_{i=2}^{\infty} i(i-1)a^{i-2}p(i) \geq 0.$$

□

So the picture of the ϕ function looks like:



When $\mu > 1$, $a = \phi(a)$ at $a = 1$, and once for $a \in (0, 1)$. When $\mu = 1$, $a = \phi(a)$ only at $a = 1$.

What does this mean for the bomb?

- There needs to be a greater than $1/3$ chance of a neutron causing a fission event to have a positive probability of causing a chain reaction.
- If the average number of hits by neutrons leaving a fission event is greater than $1/3$, every neutron that fires has a positive chance of never going extinct. By firing multiple neutrons into the mass of fissile material, an explosion is practically guaranteed!

Problems

267. Suppose $X_1, X_2, \ldots$ are iid uniform over $\{0, 1, 2\}$, and $\mathbb{P}(W = 0) = 0.2,\ \mathbb{P}(W = 2) = 0.8$. What is the generating function of

$$S = \sum_{i=1}^{W} X_i?$$

268. Suppose $B_1, B_2, \ldots$ are iid Bernoulli with mean $1/2$. If R is uniform over $\{0, 1, 2\}$, what is the generating function of

$$S = \sum_{i=1}^{R} B_i?$$

269. Suppose a branching process has a number of children that is uniform over $\{0, 1, 2, 3\}$. What is the extinction probability of this branching process?

270. Suppose a branching process has a number of children that is uniform over $\{0, 1, 2\}$. What is the extinction probability of this branching process?

271. Suppose in a branching process where each individual has Y children

$$\mathbb{P}(Y = 0) = 0.1,\ \mathbb{P}(Y = 1) = 0.6,\ \mathbb{P}(Y = 3) = 0.3.$$

What is the extinction probability of this branching process?

272. Suppose in a branching process where each individual has Y children

$$\mathbb{P}(Y = 0) = 0.35,\ \mathbb{P}(Y = 1) = 0.25,\ \mathbb{P}(Y = 3) = 0.4.$$

What is the extinction probability of this branching process?

Chapter 32

Generating Functions

Question of the day

Consider a branching process where the number of children is drawn from Y where

$$\mathbb{P}(Y = 0) = 1/6,\ \mathbb{P}(Y = 1) = 1/3,\ \mathbb{P}(Y = 2) = 1/2.$$

If the branching process starts with a single individual, what is the chance that the process goes extinct?

Summary

- Suppose W is a nonnegative integer random variable, $X_1, X_2, \ldots$ are iid with the same distribution as X. Then for
$$S = \sum_{i=1}^{W} X_i$$
it holds that $gf_S(a) = g_W(g_X(a))$.
- For a branching process with Y having the distribution of the number of children, if $\mathbb{P}(Y = 0) = 0$ then the extinction probability is 0, if $\mathbb{P}(Y < 1) > 0$ and $\mathbb{E}[Y] \leq 1$ then the extinction probability is 1, and if $\mathbb{P}(Y < 1) > 0$ and $\mathbb{E}[Y] > 1$, then the extinction probability is the unique solution to $a = \text{gf}_Y(a)$ for $a \in (0, 1)$.

32.1 *Generating functions for random sums of random variables*

Recall that the generating function of a random variable X is

$$\text{gf}_X(a) = \mathbb{E}[a^X].$$

If X and Y are independent, then so are a^X and a^Y, so

$$\text{gf}_{X+Y}(a) = \mathbb{E}[a^{X+Y}] = \mathbb{E}[a^X a^Y] = \mathbb{E}[a^X]\mathbb{E}[a^Y] = \text{gf}_X(a)\,\text{gf}_Y(b).$$

32.1.1 Example: Summing two fair six sided dice

- Let $X, Y \sim \textsf{Unif}(\{1,2,3,4,5,6\})$ be independent.
- What is $\text{gf}_{X+Y}(a)$?

$$\begin{aligned}\text{gf}_{X+Y}(a) &= \text{gf}_X(a)\,\text{gf}_Y(a)\\ &= \left[\frac{1}{6}a^1 + \cdots + \frac{1}{6}a^6\right]^2\\ &= \frac{1}{36}(a^2 + 2a^3 + 3a^4 + 4a^5 + 5a^6 + 6a^7 + 5a^8 + 4a^9 + 3a^{10} + 2a^{11} + a^{12}),\end{aligned}$$

so $\mathbb{P}(X + Y = 5) = 4/36 = 0.1111\ldots$.

32.2 *Adding a fixed number of random variables*

If there were 20 independent fair six sided dice being added together, then the generating function would be multiplied by itself 20 times! This idea can be generalized.

Fact 76
If $X_1, X_2, \ldots, X_n$ are iid X, then

$$\text{gf}_{X_1+\cdots+X_n}(a) = \text{gf}_X(a)^n.$$

32.3 *Adding a random number of random variables*

Now consider a trickier problem, adding a random number of random variables!

- Let $X_1, X_2, X_3 \overset{\text{iid}}{\sim} X \sim \textsf{Unif}(\{1,\ldots,6\})$.
- Let $N \sim \textsf{Unif}(\{1,2,3\})$.
- What is the generating function of $S = \sum_{i=1}^N X_i$?
- The key to these types of problems is the use of the Fundamental Theorem of Probability. Generating functions are just a type of expected value, and so conditioning on a random variable's value is allowed as long as the expected value is taken again.

$$\begin{aligned}\text{gf}_S(a) &= \mathbb{E}\left[a^{\sum_{i=1}^N X_i}\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[a^{\sum_{i=1}^N X_i} | N\right]\right]\\ &= \mathbb{E}\left[\text{gf}_X(a)^N\right]\\ &= \text{gf}_N(\text{gf}_X(a)).\end{aligned}$$

- Since

$$\text{gf}_N(a) = (1/3)[a + a^2 + a^3],$$

Using `expand (1/3)*(p(x) + p(x)^2 + p(x)^3) where p(x) = (1 / 6)*(x + x^2 + x^3 + x^4 + x^5 + x^6)` in Wolfram Alpha gives

$$\begin{aligned}\text{gf}_S(x) = & x^{18}/648 + x^{17}/216 + x^{16}/108 + (5x^{15})/324 + (5x^{14})/216 + (7x^{13})/216+ \\ & (31x^{12})/648 + (13x^{11})/216 + (5x^{10})/72 + (49x^9)/648 + (17x^8)/216 + (17x^7)/216+ \\ & (19x^6)/162 + (11x^5)/108 + (19x^4)/216 + (49x^3)/648 + (7x^2)/108 + x/18\end{aligned}$$

So, for instance,

$$\mathbb{P}\left(\sum_{i=1}^{N} X_i = 12\right) = \frac{31}{648} = 0.04783\ldots.$$

There was nothing special about X or N in the above, so a similar argument shows the following.

Fact 77
Suppose $X_1, X_2, \ldots$ are iid with distribution the same as X. Then for any nonnegative integer valued N,

$$\text{gf}_{X_1+\cdots+X_N}(a) = \text{gf}_N(\text{gf}_X(a)).$$

Using *composition* notation, this says that

$$\text{gf}_{X_1+\cdots+X_N} = \text{gf}_N \circ \text{gf}_X .$$

32.4 *Generating functions for branching processes*

Given the population at time t is X_t, then the population at time $t+1$ is a random sum of random variables, so the rule for generating functions applies.

Remember that

$$[X_n | X_{n-1}] = Y_1 + \cdots + Y_{X_{n-1}},$$

where the Y_i are iid draws from the same distribution as Y.

So

$$\begin{aligned}\text{gf}_{X_n} &= \text{gf}_{X_{n-1}} \circ \text{gf}_Y \\ &= (\text{gf}_{X_{n-2}} \circ \text{gf}_Y) \circ \text{gf}_Y \\ &= \vdots \\ &= \text{gf}_Y \circ \text{gf}_Y \circ \cdots \circ \text{gf}_Y .\end{aligned}$$

In words: the generating function of the population of a branching process after n generations is the n-fold composition of the generating function of Y with itself, where Y has the same distribution as the number of children an individual has.

The generating function of X satisfies

$$\text{gf}_X(0) = \mathbb{P}(X = 0).$$

So $\mathbb{P}(X_n = 0) = [\text{gf}_Y \circ \text{gf}_Y \circ \cdots \circ \text{gf}_Y](0)$.

Fact 78
Consider a branching process where the number of children has the same distribution as Y. The probability that a branching process started with 1 individual is extinct after n generations is the n-fold composition of gf_Y with itself evaluated at 0.

32.4.1 Example: Understanding extinction.

Consider a branching process where each individual has a number of children Y with

$$\mathbb{P}(Y = 0) = 1/2, \mathbb{P}(Y = 1) = 1/3, \mathbb{P}(Y = 2) = 1/6.$$

Then the random variable determining the distribution of the number of childred has generating function

$$\text{gf}_Y(a) = 1/2 + (1/3)a + (1/6)a^2.$$

To find the chance of extinction at the first generation, note that

$$\text{gf}_{X_1}(0) = \text{gf}_Y(0) = 1/2.$$

That is, there is a $1/2$ chance that $X_1 = 0$.

Now consider the chance of extinction after two generations:

$$\begin{aligned}\text{gf}_{X_2}(0) &= \text{gf}_Y(\text{gf}_Y(0)) = \text{gf}_Y(1/2) = (1/2) + (1/3)(1/2) + (1/6)(1/2)^2 \\ &= (12 + 4 + 1)/24 = 17/24.\end{aligned}$$

This leads to the following table:

n	0	1	2	3	4	5
$\mathbb{P}(X_n = 0)$	0	0.5	0.7083	0.8197	0.8852	0.9256

The average number of children is $\mu = (1/3)(1) + (1/6)(2) < 1$, which means the probability of going extinct goes to $a = 1$ as $n \to \infty$.

32.4.2 Another example

Suppose the probability vector for the number of children was

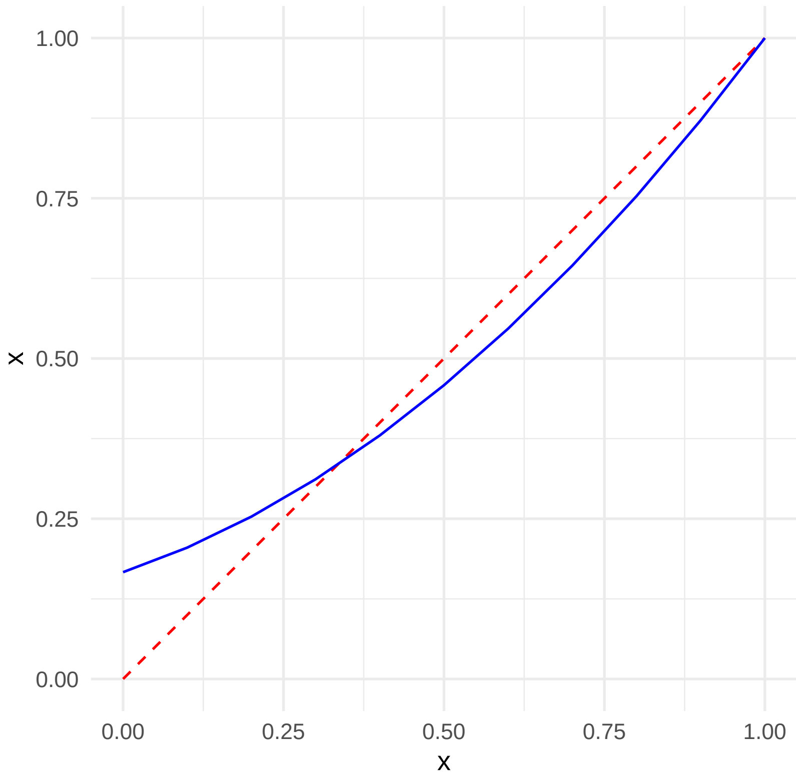
$$(1/6, 1/3, 1/2)$$

for 0, 1, or 2 children respectively.

Then $\text{gf}_Y(a) = (1/6) + (1/3)a + (1/2)a^2$. Using the same procedure as before:

n	0	1	2	3	4	5
$\mathbb{P}(X_n = 0)$	0	0.1666	0.2361	0.2732	0.2950	0.3085

This does not appear to be reaching 1 as $n \to \infty$. To see why this is true, look at the graph of $\phi(a) = \text{gf}_Y(a)$ and the identity function.



When the blue line is above the red dashed line, that means $\phi(a) > a$, and so the extinction probability at generation $n + 1$ will be larger than at n.

But when the blue line is below the red dashed line, it holds that $\phi(a) < a$. As long as the process starts out below point where $a = \text{gf}_Y(a)$, will converge to that point where $\phi(a) = a$. This is called a *fixed point*.

Note also that 1 is also always a fixed point since $\phi(1) = 1$. (Another way to say this is that if the probability of extinction is 1 at a generation, it stays at 1 forever.) The folowing fact summarizes these observations.

Fact 79
For a branching process with:

condition	extinction probability
$p_0 = 0$	$a = 0$
$p_0 \in [0,1),\ \mu \leq 1$	$a = 1$
$p_0 \in [0,1),\ \mu > 1$	unique solution to $a = \text{gf}_Y(a)$ in $(0,1)$

32.4.3 Convergence in the example

For general Y, it might not be easy to solve $a = \phi(a)$. So instead, just use convergence of the composition of ϕ. That is, start by plugging 0 into ϕ, then plug that answer into ϕ, and so on until convergence occurs.

Of course when Y has a polynomial of low degree as a generating function, it is possible to solve for a exactly.

$$\begin{aligned}
a &= (1/6) + (1/3)a + (1/2)a^2 \\
0 &= (1/6) - (2/3)a + (1/2)a^2 \\
0 &= 1 - 4a + 3a^2 \\
0 &= (a-1)(3a-1) \text{ (since } (a-1) \text{ always a factor)} \\
a &= 1/3.
\end{aligned}$$

32.5 *History: What happened with the German bomb?*

Werner Heisenberg (of uncertainty principle fame) was in charge of the German project to build a nuclear weapon. During his time leading the lab, he made an enormous mistake in the calculation of the amount of uranium needed to build a bomb.

Heisenberg told German high command needed several tons of U-235 to make it work. Now, with a proper branching process analysis it can be found that about 60 kg are needed. Later, Heisenberg told the British that he lied on purpose so that the Nazis would not get the weapon.

Problems

273. Suppose

$$\mathbb{P}(W = 0) = 0.5,\ \mathbb{P}(W = 2) = 0.5$$

What is the generating function of W?

274. Suppose

$$\mathbb{P}(Y = 0) = 0.3,\ \mathbb{P}(Y = 1) = 0.3, \mathbb{P}(Y = 2) = 0.4.$$

What is the generating function of W?

275. Suppose

$$\mathbb{P}(W = 0) = 0.5,\ \mathbb{P}(W = 2) = 0.5$$

is the distribution of the number of children for an individual in a branching process. Starting with 1 person, what is the probability that the population is extinct at the 3rd generation?

276. Suppose

$$\mathbb{P}(Y = 0) = 0.3,\ \mathbb{P}(Y = 1) = 0.3, \mathbb{P}(Y = 2) = 0.4.$$

is the distribution of the number of children in a branching process. Find the probability that the population is extinct after two generations.

277. Suppose $\mathbb{P}(R = 0) = 0.3$ and $\mathbb{P}(R = 2) = 0.7$ is the distribution for the number of children in a branching process. Find the probability of extinction starting with 1 person in the population.

278. Suppose $\mathbb{P}(S = 0) = 0.3$, $\mathbb{P}(S = 1) = 0.5$ and $\mathbb{P}(S = 3) = 0.2$ is the distribution for the number of children in a branching process. Find the probability of extinction starting with 1 person in the population. (If you do not have a means for solving cubic equations, recall that $x = 1$) is always a solution to $x = \phi(x)$!)

279.

Suppose that in a branching process, each individual has a number of children that is Poisson distributed with parameter λ. Such a distribution has generating function equal to $\exp(-\lambda(1 - x))$.

a) If $\lambda = 1$, find the probability that the population is extinct at the 2nd generation.

b) If $\lambda = 0.6$, find the probability that the population is extinct at the 2nd generation.

280. Suppose that in a branching process, each individual has a number of children that is Poisson distributed with parameter λ. Such a distribution has generating function equal to $\exp(-\lambda(1 - x))$.

Find the extinction probability when $\lambda = 2$ using WolframAlpha.

Chapter 33

Brownian Motion

Question of the Day

If B_t is standard Brownian motion, then what is $\mathbb{P}(B_2 > 1)$?

Summary

- **Brownian motion** started as a physical phenomenon noted by Robert Brown. Today a **standard Brownian motion** (aka a **standard Wiener process**) refers to a stochastic process with the following four properties:
  1. **Centered** $B_0 = 0$
  2. **Independent increments** For all $a < b \leq c < d$, the increments $B_d - B_c$ and $B_b - B_a$ are independent random variables.
  3. **Normal increments** For all $a < b$,

  $$B_b - B_a \sim \mathsf{N}(0, b - a).$$

  4. **Continuity** The process B_t is continuous with probability 1.

33.1 *History of Brownian motion*

There are two things that are called *Brownian motion*

1. The motion of small particles without any apparent driving force.

2. A mathematical stochastic process that is the limit of a scaled random walk. This is also called a *Wiener process* in honor of Norbert Wiener.

33.1.1 Physical Brownian motion

In 1595, the microscope was invented. A little more than a century later, in 1696, Gray noticed that small grains suspended in fluids were moving. While small, grains are still living matter, and Gray assumed that it was the *animus* of the grains that produced the movement.

In 1827, Robert Brown saw pollen grains moving around. This was much more significant, since pollen on its own will not produce a new plant. The idea at the time was that only organic material moved on its own. But Brown explored further and saw movement from inorganic material!

By the late 19th century, physicists were tackling a basic question about the universe: is matter continuous or discontinuous? Many thought that Newton had settled question on continuous side with his work in optics, but Brownian motion remained a mystery.

In 1905, Einstein had his "Miracle Year" where he produced four amazing papers that revolutionized different pieces of physics

1. Photoelectric Effect, where he treated light as discrete rather than continuous. He received the Nobel Prize for this work in 1921.

2. Brownian motion displacement prediction, which treated matter as small and discrete.

3. Special Relativity.

4. Mass-Energy Equivalence ($E = mc^2$).

In 1908, Perrin actually was able to measure the displacement, and received the Nobel Prize for his work in 1926.

33.1.2 Mathematical Brownian motion

Meanwhile, on the mathematics side, Bachelier proposed a new type of stochastic process for modeling stock prices in 1900. Unlike the physics debate, there was no debate here: the stock market moves according to many small discrete trades. However, this continuous model was proposed to try and replicate the success of differential equations as a modeling tool in the stochastic realm.

Consider a time variable t and a response variable y. A small change in t is represented by the *differential* dt. Given a change in time, y changes as well. Write

$$dy = y_{t+dt} - y_t.$$

Adding up these changes over an interval can give the overall change in the y_t process. *Integral* notation is used to reflect this.

$$y_T - y_0 = \int_{t=0}^{T} dy.$$

If y is changing as t changes in a direct fashion, then the Fundamental Theorem of Calculus

$$dy = y'(t)\, dx$$

can be used to write these integrals as

$$y_T - y_0 = \int_{t=0}^{T} y'(t)\, dt.$$

integrals can be used to understand how changes in t result in changes in y. For instance, if $y_t = t$, then summing up the changes in t results in the changes of y. This can be written

$$y_T = \int_0^T dt.$$

Things become more interesting when the function B_t is not deterministic, but instead is a random variable. Again

$$dB_t = B_{t+dt} - B_t.$$

Mathematical Brownian motion makes the choice that this differential of Brownian motion will have mean 0 and variance dt. The reason is that variance of the sum of independent variables inside, so this makes

$$\mathbb{V}(B_T - B_0) = \mathbb{V} \int_{t=0}^{T} dB_t = \int_{t=0}^{T} \mathbb{V}(dB_t) = \int_{t=0}^{T} dt = T.$$

What random variable has mean 0 and variance dt? One choice is

$$dB_t \sim \textsf{Unif}(\{-\sqrt{dt}, \sqrt{dt}\}).$$

Another choice that could be made is

$$dB_t \sim \textsf{N}(0, dt).$$

Note that here the second parameter of the normal distribution is the variance, *not* the standard deviation.

Because $B_T - B_0 = \int_{t=0}^{T} dB_t$ is an infinite sum of differential elements, the Central Limit Theorem indicates that is does not really matter what dB_t starts as. No matter how it starts, the CLT gives that the result of summing these independent identically distributions dB_t will be normally distributed.

Note that B_0 is not defined by this model, instead, only the increment $B_T - B_0$ is defined. For convenience, set $B_0 = 0$ in *standard Brownian motion*.

Next consider two intervals $[a, b]$ and $[c, d]$ that do not overlap. The integrals that form these increments use Brownian differentials that are different, and so independent. That is,

$$B_b - B_a = \int_{t=a}^{b} dB_t,\ B_d - B_c = \int_{t=c}^{d} dB_t.$$

Next, because $\mathbb{E}[dB_t] = 0$ and $\mathbb{V}(dB_t) = dt$, together with the fact that the mean of a sum (integral) and variance of a sum are the sum of means and sum of variances respectively, the Central Limit Theorem implies that

$$\begin{aligned} B_b - B_a &\sim \textsf{N}\left(\int_{t=a}^{b} \mathbb{E}[dB_t], \int_{t=a}^{b} \mathbb{V}(dB_t)\right) \\ &= \textsf{N}\left(\int_{t=a}^{b} 0\ dt, \int_{t=a}^{b} dt\right) \\ &= \textsf{N}(0, b - a) \end{aligned}$$

Finally, viewing dB_t as uniform over $\{-\sqrt{dt}, \sqrt{dt}\}$ means that dt small makes $\sqrt{dt}$ and so dB_t small as well.

For a function where dx implies dy small, the function is called *continuous*. So dt small makes dB_t small means that Brownian motion should be continuous as well. Of course, with random variables, the best that can be said is that it will be continuous with probability 1.

33.2 The definition of Brownian motion

None of the above is a *proof* of the properties of Brownian motion. Instead, the above arguments should be viewed as a motivation for why standard Brownian motion is defined the way that it is.

Definition 100
Say that a stochastic process B_t is a **standard Brownian motion** or **standard Wiener process** if it satisfies the following.

1. **Centered** $B_0 = 0$.
2. **Independent increments** For all $a < b \leq c < d$, $B_d - B_c$ and $B_b - B_a$ are independent random variables.
3. **Normal increments** For all $a < b$,
$$B_b - B_a \sim \mathsf{N}(0, b - a).$$
4. **Continuity** The process B_t is continuous with probability 1.

33.3 Answering the Question of the Day

In the Question of the Day, the question regards $\mathbb{P}(B_2 > 1)$. Since $B_0 = 0$,

$$\mathbb{P}(B_2 > 1) = \mathbb{P}(B_2 - B_0 > 1) = \mathbb{P}(X > 1),$$

where $X \sim \mathsf{N}(0, 2)$. This makes $X/\sqrt{2} \sim \mathsf{N}(0, 1)$, and so

$$\mathbb{P}(B_2 > 1) = 1 - \mathbb{P}(X \leq 1) = 1 - \mathbb{P}(Z \leq 1/\sqrt{2}),$$

where $Z \sim \mathsf{N}(0, 1)$ is a normal random variable.

The probability can be found in R as

```
1 - pnorm(1 / sqrt(2))
```

```
## [1] 0.2397501
```

So the answer is $\boxed{0.2397\ldots}$.

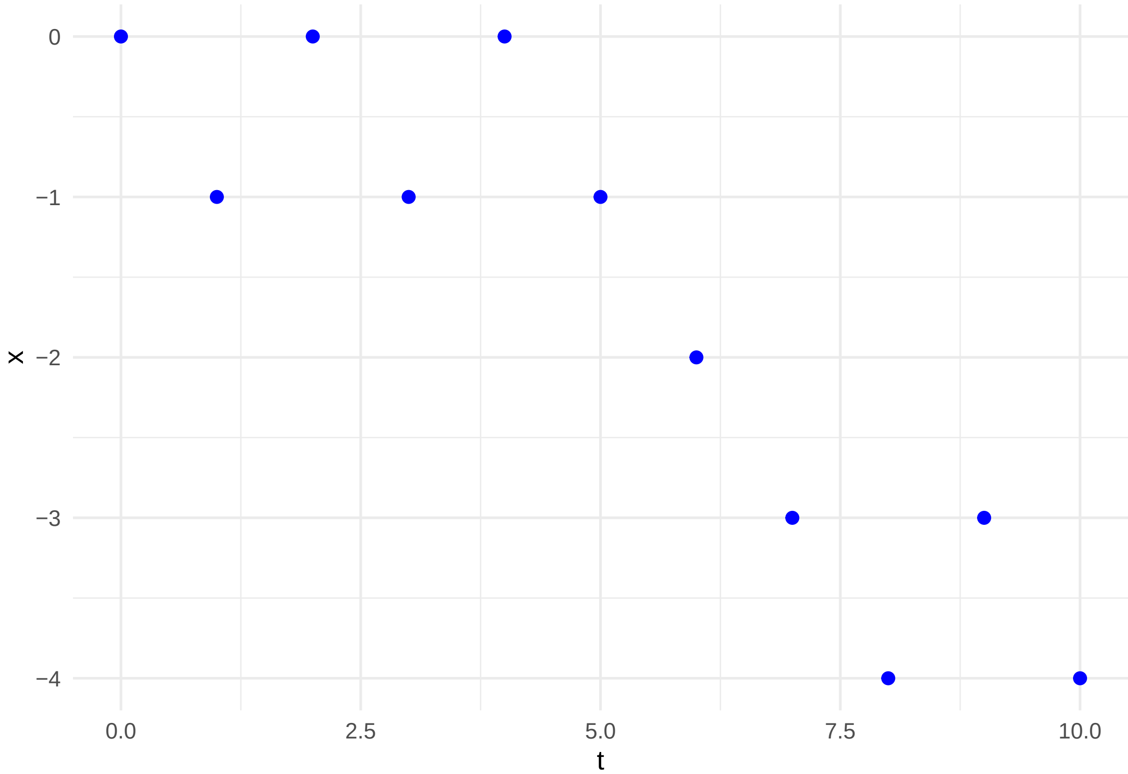
33.4 Another view of Brownian motion: the limit of simple symmtric random walk

An alternate method of viewing Brownian motion is to consider it the limit of simple symmetric random walk on the integers that is scaled properly.

Suppose a Markov chain has transition probabilities:

$$Z_{t+1} = \begin{cases} Z_t + 1 & \text{with probability } 1/2 \\ Z_t - 1 & \text{with probability } 1/2 \end{cases}$$

A realization of this process might look something like this:



Another description begins with an iid sequence of uniformly drawn random variables over $\{-1, 1\}$.

$$D_1, D_2, \ldots \overset{\text{iid}}{\sim} \text{Unif}(\{-1, 1\})$$
$$X_t = \sum_{i=1}^{t} D_i.$$

With this description it is possible to calculate the mean and variance of X_t.

$$\mathbb{E}[X_t] = \mathbb{E}\left(\sum D_i\right) = \mathbb{E}[D_1] + \cdots + \mathbb{E}[D_t] = t\mathbb{E}[D_1] = 0$$
$$\mathbb{V}[X_t] = \mathbb{V}\left(\sum D_i\right) = \mathbb{V}[D_1] + \cdots + \mathbb{V}[D_t] = t\mathbb{V}[D_1].$$

Now $\mathbb{V}[D_1] = \mathbb{E}[D_1^2] - \mathbb{E}[D_1]^2 = 1$, and standard deviation is square root of variance, so

$$\text{SD}(X_t) = \sqrt{t}\,\text{SD}(D_1) = \sqrt{t}.$$

Each time step differs from the others by 1. What if they differ by a smaller time step h instead? Then

$$Y_t = \sum_{i=1}^{t/h} D_i.$$

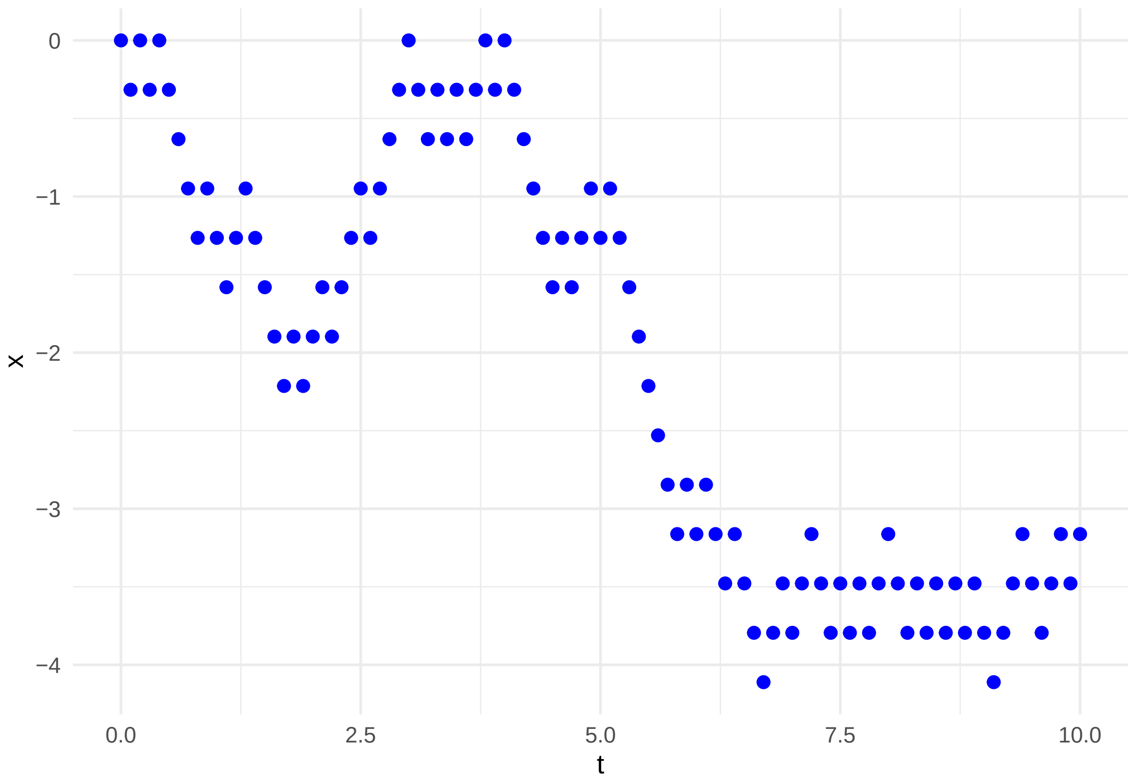
In the sum, there are t/h random variables each with standard deviation 1, so

$$\mathrm{SD}(Y_t) = \sqrt{t/h} D_i.$$

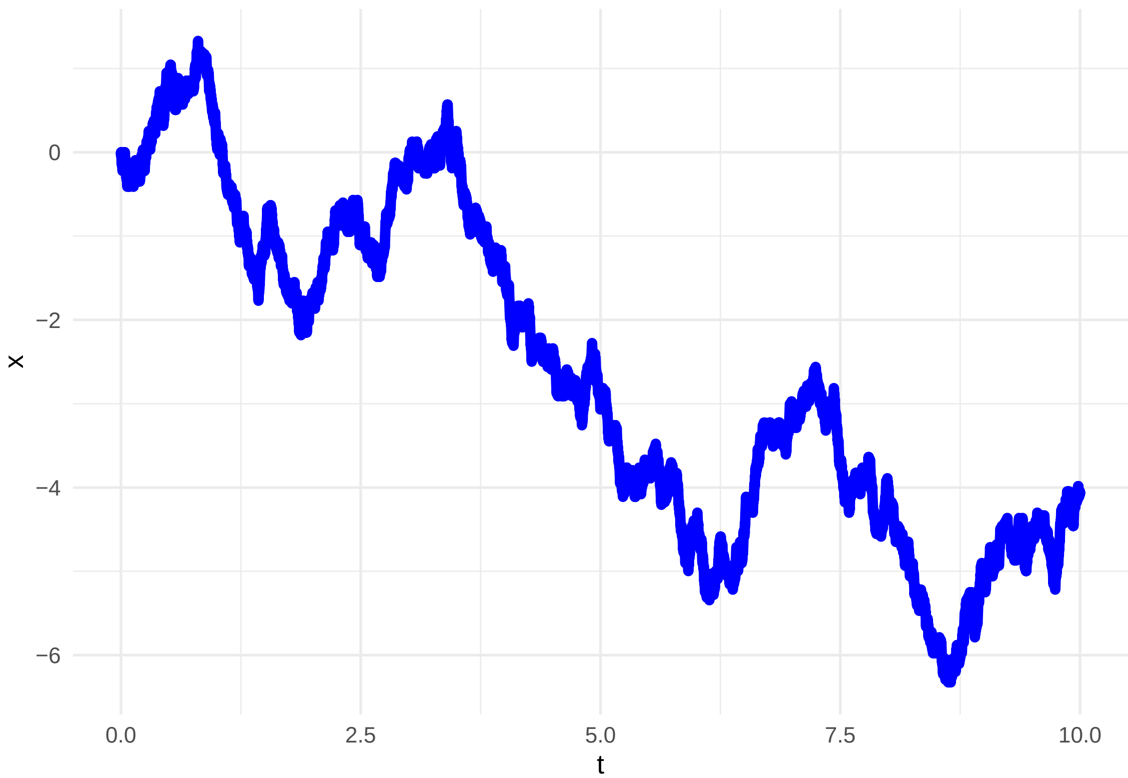
To keep the overall standard deviation at $\sqrt{t}$, instead of adding D_i, add $\sqrt{h} D_i$.

$$W_t = \sum_{i=1}^{t/h} \sqrt{h} D_i.$$

For $h = 0.1$, a realization might look like:



For $h = 0.001$, and connecting the points by lines gives



The idea of Brownian motion is to find a process that acts as the scaled limit of simple symmetric random walk as $h \to 0$. What would such a process look like?

- It starts at 0.
- For $a < b \le c < d$, the random variable $W_d - W_c$ is the sum of D_i variables that are different from the D_i variables used to find $W_b - W_a$. Hence they should be independent of each other.
- The sum of random variables tends towards the normal distribution by the Central Limit Theorem. For $a < b$, $W_b - W_a$ should be normal with mean 0 and standard deviation that is the square root of the change in time. That is, $W_b - W_a \sim \mathsf{N}(0, b - a)$.
- As $h \to 0$, $\sqrt{h} \to 0$ as well, so the process should be continuous.

These properties define the mathematical notion of Brownian motion.

33.5 *Existence of Brownian motion*

Note that we have not actually proved that Brownian motion exists. The definition just says that if a stochastic process has these four properties, then call it Brownian motion. A proof that Brownian motion exists is beyond the scope of this course.

With this definition, can prove facts about Brownian motion consistent with its use in models. The first is that the process is *stationary*.

Definition 101
A stochastic process is **stationary** (more precisely, has *stationary increments* if $\forall a < b$ and $s > 0$,

$$X_b - X_a \sim X_{b+s} - X_{a+s}$$

Of course, a process being stationary over time has nothing to do with the notion of a stationary distribution of a Markov chain.

Fact 80
Standard Brownian motion is stationary.

Proof. Let $a < b$ and $s > 0$. Then $X_b - X_a \sim \mathsf{N}(0, b-a)$ and $X_{b+s} - X_{a+s} \sim \mathsf{N}(0, (b+s)-(a+s))$, which is the same distribution. □

Definition 102
A **Lévy Process** is a stochastic process with independent and stationary increments.

Although Brownian motion is continuous, it is not differentiable!

Fact 81
With probability 1, Brownian motion is not differentiable anywhere.

Proof. A formal proof is beyond this course, but here's the intuition. Again thinking of dB_t as uniform over $\{-\sqrt{dt}, \sqrt{dt}\}$. The absolute value of the derivative would be

$$\left|\frac{dB_t}{dt}\right| = \frac{\sqrt{dt}}{dt} = \frac{1}{\sqrt{dt}}.$$

But dt is infinitesimally small, so $1/\sqrt{dt}$ is infinitely large! In other words, it does not exist. □

Problems

281.

Let B_t be a standard Brownian motion.

a) What is the mean of $B_8 - B_4$?

b) What is the variance of $B_8 - B_4$?

282. Suppose B_t is standard Brownian motion. Find the mena and variance of $B_{10} - B_4$.

283. Let B_t be a standard Brownian motion. What is the distribution of $B_4 - B_{1.5}$?

284. For W_t a standard Brownian motion, what is the distribution of $B_{10.1} - B_{8.2}$?

285. Let B_t be a standard Brownian motion. What is the probability that B_t is continuous?

286. For W_t a standard Brownian motion, find $\mathbb{P}\left(\lim_{t\to 1.1} W_t = W_{1.1}\right)$.

287. Suppose B_t is standard Brownian motion. Find $\mathbb{P}(B_5 > 2)$.

288. Suppose B_t is standard Brownian motion. Find $\mathbb{P}(B_3 \in [-1, 1])$.

Chapter 34

Simulating Brownian motion

Question of the Day

- Suppose that $B_0 = 0$ and $B_2 = -2.099779$. A draw from a standard normal is $Z = 0.349643$. Use this to simulate from $[B_1|B_0, B_2]$.

Summary

- A stochastic process has the **Markov property** if for all $s < t$,

$$[X_t|\mathcal{F}_s] \sim [X_t|X_s].$$

- Brownian motion has the Markov property.
- Extension of Brownian motion can be done using

$$B_b - B_a \sim \mathsf{N}(0, b-a).$$

- Interpolation of Brownian motion can be done using the following. suppose $t_2 > t_1$ and for $\lambda \in [0,1]$, $t = \lambda t_1 + (1-\lambda)t_2$. Then

$$[B_t - B_{t_1}|B_{t_1}, B_{t_2}] \sim \mathsf{N}(\lambda B_{t_1} + (1-\lambda)B_{t_2}, \lambda(1-\lambda)(t_2 - t_1))$$

34.1 *The Markov property*

The defining property of a Markov chain is that knowing the whole history of the process is not necessary to finding the distribution of the random variable at the next time step. Instead, just knowing the previous value of the process is enough.

This idea can be extended to times which are continuous over $[0, \infty)$.

Definition 103
A stochastic process $\{X_t\}_{t\geq 0}$ has the **Markov property** if for all $s \leq t$ and measurable A,

$$\mathbb{P}(X_t \in A|F_s) = \mathbb{P}(X_t \in A|X_s).$$

An important example of a process with the Markov property is Brownian motion.

Fact 82
Brownian motion has the Markov property.

Having the Markov property makes it much easier to simulate Brownian Motion. In particular, generating Brownian motion for a finite set of times $\{t_1, \ldots, t_n\}$ can be accomplished relatively easily.

Consider how to draw

$$B_{t_1}, B_{t_2}, \ldots, B_{t_n}.$$

For $t \in \{0, 2, 4, 6\}$, B_0, $B_2 - B_0$, $B_4 - B_2$, $B_6 - B_4$ are independent. Moreover, the increments are normally distributed.

$$B_0 = 0,\ B_2 - B_0 \sim \mathsf{N}(0, 2),\ B_4 - B_2 \sim \mathsf{N}(0, 2),\ B_6 - B_4 \sim \mathsf{N}(0, 2).$$

34.2 Using normals to simulate Brownian motion

Consider using draws from standard normals to simulate Brownian motion. For example, suppose that three standard normal draws are taken. The results are:

$$Z_1 = -1.484,\ Z_2 = 1.456,\ Z_3 = -0.09262.$$

Recall for $c \in \mathbb{R}$, $\mathbb{V}(cX) = c^2\mathbb{V}(X)$. So $Z_i \sim \mathsf{N}(0, 1) \Rightarrow \sqrt{2}Z \sim \mathsf{N}(0, 2)$.

$$\begin{aligned}
B_0 &= 0\\
B_2 &= B_0 + (B_2 - B_0) = 0 + \sqrt{2}(-1.484) = -2.099\\
B_4 &= B_2 + (B_4 - B_2) = -2.099 + \sqrt{2}(1.456) = -0.03991\\
B_6 &= B_4 + (B_6 - B_4) = -0.03991 + \sqrt{2}(-0.09262) = -0.1706.
\end{aligned}$$

This gives us the value of B_t for four values of t. This is sometimes called a *skeleton* of Brownian motion. Of course, there are still uncountably many values of B_t that have not been defined!

34.3 Interpolating Brownian motion

At this point B_0 and B_2 are set. Suppose the next time the user needs is $t = 1$. How can B_1 be generated?

In particular, given $B_0 = 0$ and $B_2 = -2.099$, what is the distribution of B_1. By the normal increment property of standard Brownian motion:

$$[B_1 - B_0|(B_2 - B_1) + (B_1 - B_0) = b] \sim [Z_1|Z_1 + Z_2 = b],$$

where $Z_1, Z_2 \overset{\text{iid}}{\sim} \mathsf{N}(0, 1)$.

So how does the distribution of two random variables given their sum behave? The following fact (a partial proof will be given later) will be useful here.

Fact 83
For X and Y continuous with densities f_X and f_y with respect to Lebesgue measure,

$$f_{X|X+Y=s}(x) \propto f_X(x) f_Y(s-x).$$

34.4 Adding normal random variables

This fact can be applied to normal random variables to interpolate random variables.

- Suppose $t \in [t_1, t_2]$, so $t = \lambda t_1 + (1-\lambda)t_2$.
- Suppose $X \sim \mathsf{N}(0, \lambda(t_2 - t_1))$
- $Y \sim \mathsf{N}(0, (1-\lambda)(t_2 - t_1))$
- (Then $X + Y \sim \mathsf{N}(0, t_2 - t_1) \sim B_{t_2} - B_{t_1}$)

$$\begin{aligned} f_{X|X+Y=s}(x) &\propto \exp\left(-\frac{x^2}{2\lambda(t_2 - t_1)}\right) \exp\left(-\frac{(s-x)^2}{2(1-\lambda)(t_2 - t_1)}\right) \\ &= \exp\left(-\frac{(1-\lambda)x^2 + \lambda(s-x)^2}{2\lambda(1-\lambda)(t_2 - t_1)}\right) \\ &= \exp(\left(-\frac{x^2 - 2sx\lambda - \lambda s^2 + (\lambda s)^2 - (\lambda s)^2}{2\lambda(1-\lambda)(t_2 - t_1)}\right) \\ &\propto \exp\left(-\frac{(x - s\lambda)^2}{2\lambda(1-\lambda)(t_2 - t_1)}\right) \end{aligned}$$

which means
$$[X|X+Y=s] \sim \mathsf{N}(\lambda s, \lambda(1-\lambda)(t_2 - t_1)).$$

- Recall $X = B_t - B_{t_1}$ and $Y = B_{t_2} - B_t$, so
$$[B_t - B_{t_1} | B_{t_1}, B_{t_2}] \sim \mathsf{N}(\lambda s, \lambda(1-\lambda)(t_2 - t_1)).$$
- This gives the following fact!

Fact 84
Suppose $t_1 < t_2$. For $\lambda \in [0,1]$ and $t = (1-\lambda)t_1 + \lambda t_2$:

$$[B_t | B_{t_1}, B_{t_2}] \sim \mathsf{N}(B_{t_1} + \lambda(B_{t_2} - B_{t_1}), \lambda(1-\lambda)(t_2 - t_1)).$$

Here is the intuition behind this result. Pretend that a Hooke's law spring is connecting B_t to B_{t-1}. Given B_{t_1}, B_{t_2}, one spring attaches B_t to B_{t_1}, and another to B_{t_2}. Each is pulling B_t towards it, and so the overall give in the spring is lessened.

34.5 Solving the Question of the Day

In the Question of the Day the draw from $Z \sim \mathsf{N}(0,1)$ was

$$Z = 0.349643.$$

The goal is to use this to find a value for B_1 given that $B_0 = 0$ and $B_2 = -2.099779$. First find the percentage of the way from 2 to 0 that time 1 is.

$$\lambda = (1 - 0)/(2 - 0) = 1/2.$$

This means that

$$B_1 - B_0 \sim \mathsf{N}(1/2(0) + (1/2)(-2.099779), (1/2)(1/2)(2)).$$

Using the value for Z, this gives

$$B_1 = -2.099779/2 + (1/2)^{1/2} Z = -0.8026545.$$

34.6 *The method for interpolation*

Given the Brownian Motion at times $\{t_1, \ldots, t_n\}$, $Z \sim \mathsf{N}(0, 1)$, and a new time t, find B_t as follows:

1. When $t > t_n$,
$$B_{t_n} + (t - t_n)^{1/2} Z.$$

2. When $t = (1 - \lambda)t_i + \lambda t_{i+1}$ for some $\lambda \in [0, 1]$
$$(1 - \lambda)B_{t_i} + \lambda B_{t_{i+1}} + [\lambda(1 - \lambda)(t_{i+1} - t_i)]^{1/2} Z.$$

3. When $t < t_1$:
$$B_{t_1} + (t_1 - t)^{1/2} Z.$$

   Note: Brownian motion looks the same when run forward in time or backwards in time. Called a *reversible* process.

34.7 *Proof of the additive density result*

For simplicity, assume that the partial derivatives of all densities are continuous. For any continuous random variables X and S with joint pdf $f_{X,S}(x, s)$:

$$f_{X|S=s}(x) = \frac{f_{X,S}(x, s)}{f_S(s)} \propto f_{X,S}(x, s).$$

So the key is finding the joint density of X and S when $S = X + Y$. The joint pdf is perhaps easier:

$$\begin{aligned}
\mathbb{P}(X \leq a, S \leq b) &= \mathbb{P}(X \leq a, X + Y \leq b) \\
&= \int_{(x,y):x \leq a \wedge x+y \leq b} f_{X,Y}(x, y)\, d\mathbb{R}^2 \\
&= \int_{-\infty}^{a} \int_{-\infty}^{b-a} f_{X,Y}(x, y)\, dx\, dy
\end{aligned}$$

We can make this an iterated integral by either Tonelli (since nonnegative) or Fubini (since the absolute integral is at most 1.)

With the assumption that the partial derivatives of the integrands are continuous everywhere, the partial derivatives with respect to b can be placed inside the integral to give

$$\frac{\partial}{\partial b}\mathbb{P}(X \leq a, S \leq b) = \int_{-\infty}^{a} f_{X,Y}(x, b-a)\, dx,$$

and then

$$\frac{\partial^2}{\partial a \partial b}\mathbb{P}(X \leq a, S \leq b) = f_{X,Y}(a, b-a)\, dx.$$

Hence $f_{X,S}(a,b) = f_{X,Y}(a, b-a)$, and we are done.

Problems

289. Suppose $B_4 = -2.4$ for B_t standard Brownian motion. What is the distribution of B_5?

290. Suppose $B_1 = 1.2$ and $B_3 = -2$. What is the distribution of B_6 given this information?

291. Suppose $B_1 = 1.3$ and $B_4 = -2.4$ where B_t is standard Brownian motion. What is the distribution of B_2?

292. Suppose $B_1 = 1.3$ and $B_4 = -2.4$ where B_t is standard Brownian motion. What is the distribution of B_3 given this information?

293. Suppose $B_4 = -2.4$ for B_t standard Brownian motion. What is the distribution of B_1?

294. Suppose $B_1 = 1.2$ and $B_3 = -2$. What is the distribution of B_2 given this information?

Chapter 35

Continuous Time Markov chains

Question of the Day

Suppose X_t has infinitesimal generator

$$A = \begin{pmatrix} -2.2 & 1 & 1.2 \\ 0.3 & -1.4 & 1.1 \\ 4.6 & 2.7 & -7.3 \end{pmatrix}$$

and states $\{a, b, c\}$. What is

$$\mathbb{P}(X_{3.2} = c | X_0 = a)?$$

Summary

- In a **continuous time Markov chain** or **jump chain**, after an infinitesimally small amount of time, the probability of moving from state i to j is the rate $\lambda_{i,j}$ times the infinitesimally small length of time.

- The **infinitesimal generator** is a matrix where for $i \neq j$, $A(i,j)$ is the rate that the chain jumps from i to j. $A(i,i) = -\sum_{j \neq i} A(i,j)$ so each row sums to 0.

- A distribution with probability vector π is **stationary** if $X_t \sim \pi$ implies $X_{t'} \sim \pi$ for all $t' \geq t$.

- Finite state CTMCs have an Ergodic Theorem similar to other finite state Markov chains.

- If πA is the zero vector, then π is stationary. (This indicates the rate of probability flow into each state equals the flow out of it.)

- In general, $p_t = p_0 \exp(tA)$.

- The time between changes in the state are exponential random variables whose rate depends on the current state.

- If $A_1, A_2, \ldots, A_n$ are independent exponential random variables where $\mathbb{E}[A_i] = 1/\lambda_i$, then $\min_i A_i$ is also an exponential random variable with rate equal to $\lambda_1 + \cdots + \lambda_n$. The probability that $A_j = \min_i A_i$

is $\lambda_j/(\lambda_1 + \cdots + \lambda_n)$.

Suppose that X_t is a *continuous time Markov chain*. This means that there is a rate $\lambda_{i,j}$ associated with each pair of states $i \neq j$ where (intuitively)

$$\mathbb{P}(X_{t+dt} = j | X_t = i) = \lambda_{i,j}\, dt.$$

So after a tiny amount of time, the chance of moving from i to j equals the rate at which the chain moves from state i to j times the tiny amount of time.

If this probability is summed up for all $j \neq i$, it gives the total rate at which probability is leaking out of state i.

$$\mathbb{P}(X_{t+dt} \neq i | X_t = i) = \sum_{j \neq i} \lambda_{i,j}\, dt.$$

Because this probability is leaving state i, make the value of $\lambda(i,i)$ equal to the negative of this sum. If the $\lambda(i,j)$ values are collected into a matrix, this is called the *infinitesimal generator* for the Markov chain.

Definition 104
The **infinitesimal generator** of a continuous time Markov chain with n states is the n by n matrix A, where for $j \neq i$, $A(i,j)$ is the rate at which the chain moves from state i to state j, and

$$A(i,i) = -\sum_{j \neq i} A(i,j).$$

Before defining a continuous time Markov chain, first define exactly what is meant by the differential. First recall the *little-o notation*.

Definition 105
A function $f(x)$ is in the set $o(x)$ if

$$(\forall n \in \{1,2,\ldots\})(\exists m \in \{1,2,\ldots\})(|x| \leq 1/m \rightarrow |f(x)/x| \leq 1/n).$$

Write $f(x) = o(x)$ as a shorthand for $f(x) \in o(x)$.

It is usually easier to use limits to prove that one function is little-o of x.

Fact 85
If

$$\lim_{x \to 0} \frac{f(x)}{x} = 0$$

for a function f, then $f(x) = o(x)$.

Definition 106
Write $\mathbb{P}(X \in dx) = \mathbb{P}(X \in [x, x+dx]) = \lambda\, dx$ to mean $\mathbb{P}(X \in [x, x+h]) = \lambda h + o(h)$. Similarly, write

$$\mathbb{P}(X_{t+dt} = j | X_t = i) = \lambda\, dt$$

to mean

$$\mathbb{P}(X_{t+h} = j | X_t = i) = \lambda h + o(h).$$

Definition 107
Say that $\{X_t\}_{t \in [0,\infty)}$ is a **continuous time Markov chain** (CTMC for short) if it has an infinitesimal generator A where for $j \neq i$,

$$\mathbb{P}(X_{t+dt} = j | X_t = i) = A(i,j)\, dt.$$

Ignoring the $o(h)$ term is called the *first order approximation*.

Example 22
In the Question of the Day, what is approximately $\mathbb{P}(X_{0.01} = c | X_0 = a)$? (Use the first order approximation.)

Answer Because the infinitesimal generator has $A(1,3) = 1.2$, this would be

$$\mathbb{P}(X_{0.01} = c | X_0 = a) \approx 1.2(0.01) = 0.012.$$

35.1 The Ergodic theorem

The Ergodic theorem is very similar to the regular one for finite state Markov chains. Because jumps between states happen at any time, there is no need for aperiodicity! However, it can no longer be said that the stationary distribution is the multiplicative inverse of the expected return time.

Definition 108
Say that distribution π is a **stationary distribution** for a CTMC if $X_t \sim \pi$ implies that $X_s \sim \pi$ for all $s \geq t$.

Definition 109
Say that distribution π is a **limiting distribution** for a CTMC if

$$(\forall x \in \Omega)(\lim_{t \to \infty} (\text{dist}_{\text{TV}}([X_t | X_0 = x], \pi) = 0).$$

It is unnecessary to change the definitions of transience or recurrence for continuous time, since they are defined in terms of the return times R_x. As in the discrete time case, when $\mathbb{P}(R_x < \infty) = 1$, call the state *recurrent*. Otherwise it is *transient*.

Theorem 13
Ergodic Theorem for finite state continuous time Markov chains

For finite state Markov chains, the following holds.

1. There is at least one stationary distribution.
2. The stationary distribution π is unique if and only if there is exactly one recurrent communication class.
3. If the stationary distribution π is unique, then

$$(\forall x)\left(\lim_{t\to\infty} \text{dist}_{\text{TV}}([X_t|x_0 = x], \pi) = 0\right).$$

35.2 *Matrix Differential Equations*

Another way to view the infinitesimal generator is as a system of differential equations.

Let $p_t(i) = \mathbb{P}(X_t = i)$. Then

$$\begin{aligned} p_t'(i) &= \frac{\mathbb{P}(X_{t+dt} = i) - \mathbb{P}(X_t = i)}{dt} \\ &= \frac{\sum_j \mathbb{P}(X_{t+dt} = i|X_t = j)\mathbb{P}(X_t = j) - \mathbb{P}(X_t = i)}{dt} \\ &= \frac{\sum_{j\neq i} p_t(j)A(j,i)\ dt + p_t(i)(1 + A(i,i)\ dt) - p_t(i)}{dt} \\ &= \sum_j p_t(j)A(j,i). \end{aligned}$$

In other words,

$$p_t' = p_t A$$

A formal proof would be more careful with the meaning of the rates in terms of $o(h)$ functions. But the proof is essentially what is given above.

Fact 86
For a CTMC $\{X_t\}$ with infinitesimal generator A and $p_t(i) = \mathbb{P}(X_t = i)$,

$$p_t' = p_t A.$$

This statement is similar to the statement for discrete time Markov chains:

$$p_{t+1} - p_t = p_t(T - I),$$

where I is the identity matrix.

For one dimensional functions, the solution to the differential equation $y' = ky$ is $y(x) = y(0)exp(kx)$. So it probably should not be a surprise that the solution to the system of differential equations is similar.

Definition 110
For an n by n matrix M, define $\exp(M)$ to be

$$\exp(M) = I + M + \frac{M^2}{2!} + \frac{M^3}{3!} + \cdots$$

when this expression exists.

Fact 87
For an infinitesimal generator A,

$$p_t = p_0 \exp(tA).$$

The proof consists of two parts: verifying that all infinitesimal generators have $\exp(tA)$ defined for all $t \geq 0$, and that this is actually the solution of the differential equations. The first requires knowledge of advanced linear algebra and will be omitted here. The second also requires advanced linear algebra and analysis to show that the Taylor series for matrix exponentials can be differentiated term by term.

Now, if $p(t)A$ is the zero vector, then the probabilities flowing into each state exactly match the probabilities flowing out. Since the derivative in this case p' is also the zero vector, the probability distribution is unchanging, which makes it stationary.

Fact 88
If p_t' is the zero vector for probability function p_t, then p_t is stationary.

In other words, the stationary distribution for a chain with one communication class corresponds to the normalized left eigenvector associated with eigenvalue 0.

35.3 Exponential time between jumps

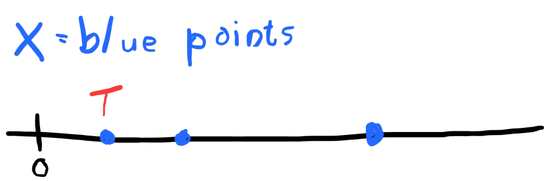
So given that the rate at which the chain moves from i to j is $A(i, j)$, what is the distribution of time between the moves?

Suppose X is a set of points in $[0, \infty)$ such that

$$\mathbb{P}(X \cap [x, x + dx] \neq \emptyset) = \lambda\, dx.$$

This can be used to understand the rate problem by considering the distribution of

$$T = \min(X).$$



In other words, T is the point in X closest to 0. Now, consider the problem of finding $\mathbb{P}(T > a)$ for some $a > 0$. Note that the event that $T > a$ is equivalent to saying that $X \cap [0, a) = \emptyset$.

Now if the interval $[0, a)$ was broken up into intervals $[0, h), [h, 2h), \ldots, [a-h, h)$ then the intervals are all the same length so the probability that they contain a point should all be the same. Moreover, the intervals are disjoint, so the probability that they contain a point of X should be independent.

At this point is in possible to state precisely what $\mathbb{P}(X \cap [x, x+dx] \neq \emptyset) = \lambda\, dx$ means.

Definition 111
Say that
$$\mathbb{P}(X \cap dt \neq \emptyset) = \lambda\, dt$$
if $\mathbb{P}(X \cap [t, t+h) \neq \emptyset) = \lambda h + o(h)$.

For $h = a/n$, the intervals $[0, h), [h, 2h), \ldots, [a-h, h)$ can be numbered $1, 2, \ldots, n$. Given the above definition, the chance that there is no point in $[0, a)$ is the chance that there is no point in any of the n length h intervals. This is

So
$$\mathbb{P}(X \cap [0, a) = \emptyset) = \prod_{i=1}^{n} (1 - \lambda h + o(h)).$$

Products are tough to deal with. Fortunately, the exponential function eats products for breakfast. Also, from the Taylor series expansion for the exponential function

$$1 - \lambda h + o(h) = \exp(-\lambda h + o(h))$$

from the Taylor series expansion for the exponential function.

Fact 89
For λ and h positive,
$$1 - \lambda h + o(h) = \exp(-\lambda h + o(h)).$$

The proof is mainly just dealing with technical details and relying on the second degree Taylor polynomial expansion of the exponential function.

Proof. To prove this, let $f(h)$ be any function that is $o(h)$. The Taylor series for the exponential function is

$$\exp(-x) = 1 - x + x^2/2! - x^3/3! + \cdots.$$

This alternates for any $x > 0$, so for any fixed $h > 0$ small enough that $f(h) < \lambda h$,

$$\exp(-\lambda h + f(h)) \in [1 - \lambda h + f(h), 1 - \lambda h + (\lambda h - f(h))^2/2!]$$

and
$$\exp(-\lambda h + (\lambda h + f(h))^2/2!) \geq 1 - \lambda h + (\lambda h - f(h))^2/2!.$$

Because the exponential function is continuous and increasing there is a value $x \in [-\lambda h + f(h), -\lambda h + (\lambda h - f(h))^2/2!]$ such that $\exp(x) = \exp(-\lambda h + f(h))$.

Let $g(h) = x - \lambda h + f(h)$. Since the width of the interval is $(\lambda h + f(h))^2/2!$, $g(h)$ is also $o(h)$ and $\exp(-\lambda h + g(h)) = 1 - \lambda h + f(h)$, which completes the proof. □

Using this fact

$$\begin{aligned}\mathbb{P}(X \cap [0,a] = \emptyset) &= \prod_{i=1}^{n} \exp(-\lambda h + o(h)) \\ &= \exp\left(-\lambda h n + n o(h)\right) \\ &= \exp(-\lambda a + o(1)).\end{aligned}$$

So as $h \to 0$, the $o(1)$ term disappears leaving $\mathbb{P}(T > a) = \exp(-\lambda a)$. By the way, this expression $\mathbb{P}(T > a) = 1 - \mathbb{P}(T \leq a)$ is called the *survival function* of the random variable T.

This is the survival function for an exponential random variable of rate λ. This proves the following fact.

Fact 90
For X a collection of points such that $\mathbb{P}(X \cap [x, x+dx] \neq \emptyset) = \lambda\, dx$, $\min(X) \sim \text{Exp}(\lambda)$.

So another way to think about a continuous time Markov chain is that it wants to jump from the current state x to another state y by setting an alarm clock that has an exponential distribution of rate $\lambda_{x,y}$ for all states x and y.

The state x has set such an alarm clock for every possible state it has a positive rate of visiting. Whichever alarm clock goes off first, that is the state that it jumps to. A helpful fact about independent exponential random variables will be useful here.

Fact 91
Let $A_1, \ldots, A_n$ be independent (but not necessary identical) exponentially distributed random variables. Say $A_i \sim \text{Exp}(\lambda_i)$. Then

$$\min(A_1, \ldots, A_n) \sim \text{Exp}(\lambda_1 + \cdots + \lambda_n).$$

Also,

$$\mathbb{P}(\min(A_1, \ldots, A_n) = A_i) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}.$$

This can be shown using the cdf of the exponentials.

In other words, the minimum of independent exponential random variables is another exponential whose rate is the sum of the exponentials being added. Moreover, the probability that a particular exponential is first is directly proportional to its rate.

Problems

295. Suppose that a continuous time Markov chain has infinitesimal generator

$$\begin{pmatrix} -3.4 & 3.4 \\ 1.2 & -1.2 \end{pmatrix}.$$

Approximate $\mathbb{P}(X_{0.0007} = 2 | X_0 = 1)$ using the first order approximation.

296. Suppose a CTMC has infinitesimal generator

$$B = \begin{pmatrix} -10.3 & 4.1 & 6.2 \\ 5.0 & -9.3 & 4.3 \\ 6.2 & 3.7 & -9.9 \end{pmatrix}$$

and states $\{a, b, c\}$.

Approximate $\mathbb{P}(X_{0.0002} = c | X_0 = a)$ using the first order approximation.

297. Suppose that a continuous time Markov chain has infinitesimal generator

$$\begin{pmatrix} -3.4 & 3.4 \\ 1.2 & -1.2 \end{pmatrix}.$$

What is $\mathbb{P}(X_{0.7} = 2 | X_0 = 1)$?

298. Suppose a CTMC has infinitesimal generator

$$B = \begin{pmatrix} -10.3 & 4.1 & 6.2 \\ 5.0 & -9.3 & 4.3 \\ 6.2 & 3.7 & -9.9 \end{pmatrix}$$

and states $\{a, b, c\}$. Find $\mathbb{P}(X_{0.2} = c | X_0 = a)$.

299. Suppose that a continuous time Markov chain has infinitesimal generator

$$\begin{pmatrix} -3.4 & 3.4 \\ 1.2 & -1.2 \end{pmatrix}.$$

Find the eigenvector associated with eigenvalue 0 and normalize to get the unique stationary distribution

300. Suppose a CTMC has infinitesimal generator

$$B = \begin{pmatrix} -10.3 & 4.1 & 6.2 \\ 5.0 & -9.3 & 4.3 \\ 6.2 & 3.7 & -9.9 \end{pmatrix}$$

and states $\{a, b, c\}$.

Find the left eigenvector with eigenvalue 0 and normalize to find the unique stationary distribution.

301.

Suppose a CTMC has infinitesimal generator

$$B = \begin{pmatrix} -10.3 & 4.1 & 6.2 \\ 5.0 & -9.3 & 4.3 \\ 6.2 & 3.7 & -9.9 \end{pmatrix}$$

and states $\{a, b, c\}$.

a. If the current state is a, what is the chance that the next jump is to b instead of c?

b. At what rate is the chain jumping away from state a?

302.

Consider a CTMC with infinitesimal generator

$$C = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

a. If the current state is b, what is the chance that the next jump is to a instead of c?

b. At what rate is the chain jumping away from state b?

Chapter 36

Poisson Point Processes

Question of the Day

The times of earthquakes are modeled using a Poisson process with rate 0.1 per year. What is the chance that there are exactly two earthquakes in one year?

Summary

- A **Poisson point process over $[0, \infty)$ of rate λ** X is a collection of points such that $\mathbb{P}(X \cap dt \neq \emptyset) = \lambda\, dt$
- For such a set, $\min(X) \sim \mathsf{Exp}(\lambda)$.
- The number of points in $[a, b]$ will have a Poisson distribution with parameter $(b - a)\lambda$.
- The **Poisson process** is defined as $N_t = \#(X \cap [0, t])$.

Last time a set of points $X \subset [0, \infty)$ was discussed where the probability that X contained a point in an interval was approximately the length of the interval times the rate of the process.

That is,

$$\mathbb{P}(X \cap [t, t + h] \neq \emptyset) = \lambda h + o(h).$$

For small intervals, it is unlikely that X has two points in the interval. (From the above, this is $O(h^2)$.) So $\mathbb{I}(X \cap [t, t+h]) \neq \emptyset$ is a Bernoulli with mean close to λh for small h. For a larger interval $[a, b]$ the expected number of points of X in the interval is

$$\int_a^b \mathbb{E}[\mathbb{I}(X \cap dt) \neq \emptyset] = \int_a^b \lambda\, dt = \lambda(b - a).$$

This motivates the following definition.

Definition 112
A random set of points $X \subset [0, \infty)$ is a **Poisson point process of rate λ over $[0, \infty)$** if it satisfies the following.

1. For any $0 \le a < b$,
$$\mathbb{E}(\#\{X \cap [a, b]\}) = \lambda(b - a).$$

2. For any $0 \le a < b \le c < d$, the two random variables
$$\#\{X \cap [a, b]\} \text{ and } \#\{X \cap [c, d]\}$$
are independent.

This is used as a model for a variety of problems.

- Arrivals to a service center (or any queue, really.)
- Times of radioactive decay.
- Defects in a steel bar.
- Typos in a document.
- Defunct pixels in a monitor, row by row.

Anywhere there are a bunch of possibilities for an error, with the probability very low, a Poisson point process can be a good moel.

The language of these processes derives from the queue application.

Definition 113
For a Poisson point process, the values of the points are sometimes called the **arrival times**. The distance between adjacent arrival times are called *interarrival times*.

If $0 < T_1 < T_2 < T_3 < \cdots$ are the arrival times, then recall from last time that the event $T_1 > a$ is equivalent to $\#(X \cap [0, a]) = \emptyset$.

Recall that the distribution of $\min(X)$ is exponentially distributed with rate λ. This extends to the interarrival times.

Fact 92
The random variables $T_i - T_{i-1}$ and $T_1 - 0$ are all independent and have exponential distribution $\text{Exp}(\lambda)$.

Example 23
Suppose typos in a text are modeled as occuring according to a Poisson point process of rate 0.01 typos / word. On average, how many words separate the third and fourth typos?

Answer The number of words separating the third and fourth typo is an exponential of rate 0.01. The expected value of this is one over the rate, so $\boxed{100}$.

So the arrival times themselves are a sum of iid exponential random variables.

$$T_n = (T_1 - 0) + (T_2 - T_1) + \cdots + (T_n - T_{n-1}).$$

The sum of iid exponentials gives a *gamma distribution*, also known as an *Erlang distribution*.

Fact 93
For a Poisson point process of rate λ, the distribution of T_n is

$$T_n \sim \mathsf{Gamma}(n, \lambda)$$

with density

$$f_{T_n}(s) = \lambda^n s^{n-1} \exp(-\lambda s)/(n-1)!$$

The Gamma distribution with integer first parameter is also called the Erlang distribution.

Example 24
For a Poisson point process of rate 4.1 over $[0, \infty)$, what is the chance that the third arrival lies in $[1, 2]$.

Answer To answer this, integrate the density of a gamma distribution with parameters 3 and 4.1 over the interval $[1, 2]$:

$$\int_1^2 4.1^3 s^2 \exp(-4.1s)/2!\, ds = \boxed{0.2120\ldots}.$$

36.1 *Counting points in an interval*

So how many points are there in an interval $[a, b]$? That is, what is the distribution of

$$\#(X \cap [0, t])?$$

This turns out to be where the *Poisson* point process gets its name. The average number of points in $[0, t]$ is λt.

Fact 94
For X a Poisson point process of rate λ over $[0, \infty)$,

$$\#(X \cap [0, t]) \sim \mathsf{Pois}(\lambda t)$$

with density

$$f_{\#(X \cap [0,t])}(i) = \frac{(\lambda t)^i \exp(-\lambda t)}{i!}.$$

Example 25
Suppose that X is a Poisson point process of rate 1.2 over $[0, \infty)$. What is the chance that there are two points of X in the interval $[10, 12]$?

Answer Because the location of the interval does not matter, only the length, this is the same as

$$\mathbb{P}(\#(X \cap [0, 2]) = 2).$$

Applying the Poisson density gives

$$(1.2 \cdot 2)^2 \exp(-1.2 \cdot 2)/2! = \boxed{0.2612 \ldots}.$$

Example 26
Suppose that cracks in a sidewalk are modeled as a PPP with rate 2 per meter. What is the probability that there are no cracks in the first meter?

Answer The rate (2 per meter) over the region of measure 1 meter gives a Poisson distributed number of points with parameter found by multiplying the rate times the Lebesgue measure of the region:

$$\frac{2}{\text{meter}} \cdot 1 \text{ meter} = 2.$$

Then the probability a Poisson with parameter 2 has value 0 is

$$\mathbb{P}(\#(P \cap [0, 1]) = 0) = \exp(-2) \approx \boxed{0.1353}.$$

Alternately, one could evaluate this as

$$\mathbb{P}(T_1 > t) = \exp(-\lambda t)$$

and then plug in $\lambda t = 2$ as before.

36.2 The Poisson process

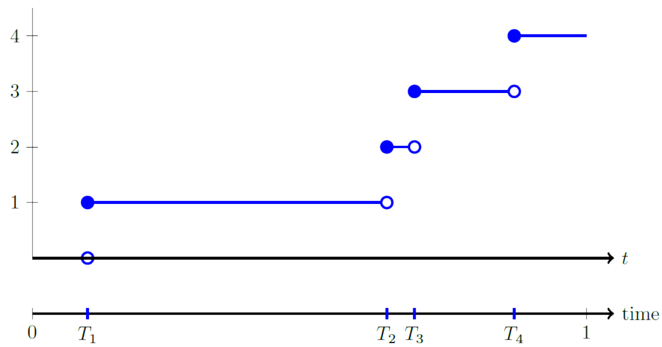
Suppose that goal is not to have the points themselves, but for a given time, the goal is to under the number of points that fall into the interval $[0, t]$.

Definition 114
For X a Poisson point process with rate λ over $[0, \infty)$, call

$$N_t = \#(X \cap [0, t])$$

the *Poisson process* of rate lambda.

Note that for $t \in [0, T_1)$, $N_t = 0$, then for $t \in [T_1, T_2)$, $N_t = 1$, and so on.



Because the interarrival times are exponentially distributed, this process can also be viewed as a countable state space continuous time Markov chain on state space $\{0, 1, 2, \ldots\}$ where $\lambda(i, i+1) = \lambda$ and $\lambda(i, j) = 0$ for all $j \neq i + 1$.

All of the states are transient from this point of view.

Fact 95
A Poisson process of rate λ is a Lévy process where for $a < b$,

$$N_b - N_a \sim \text{Pois}(\lambda(b - a)).$$

The Poisson process N_t has several nice properties similar to Brownian motion.

Fact 96
A Poisson process N_t of rate λ satisfies the following.

1. $N_0 = 0$.
2. **Poisson increments** For $a < b$, $N_b - N_a \sim \text{Pois}(\lambda(b - a))$.
3. **Independent increments** For $a < b \leq c < d$, $N_b - N_a$ and $N_d - N_c$ are independent.

Because the Poisson process hass stationary and independent increments, it is another example of a *Lévy process*.

Unlike Brownian motion, a Poisson point process has jumps of size 1, so is definitely not continuous.

36.3 *Conditioning on the number of points in an interval*

For a Poisson point process of constant rate, consider an interval $[a, b]$. Then for any small subinterval of equal length, the probability each contains a point is the time.

So if the number of points is fixed, then they are equally likely to lie in small regions of equal length. That makes the distribution of the points uniform and independent. This is reflecting in the following fact.

Fact 97
Given $N_t = n$, the points $X_1, \ldots, X_n$ are iid Unif(A).

Note that $X_1, \dots, X_n$ are the *unsorted* point values. The $T_1 < T_2 < \dots < T_n$ are *not* uniform and *not* independent because they have to be in order!

Example 27
Suppose a slab of marble has defects that are modeled as a Poisson point process over $[0, 1]$. Say there are 3 defects. What is the chance that all the defects lie in $[1/2, 1]$?

Answer Each of the three defects can be though of as uniformly distributed over $[0, 1]$. The chance that the uniform position of a defect falls in $[1/2, 1]$ is $(1 - 1/2)/(1 - 0) = 1/2$. The three points are independent, so the chance that all three points fall into $[1/2, 1]$) is $(1/2)^3 = 1/8 = \boxed{0.1250}$.

36.4 Understanding the gamma and Poisson densities

This section gives intuition (not proofs!) for the density of gamma and Poisson random variables.

First, some preliminaries. Recall that the chance that there is a point of X in a differential interval dt is

$$\lambda\ dt.$$

This makes the chance that there is no point in the differential interval is

$$1 - \lambda\ dt = \exp(-\lambda dt).$$

Intuitively, the chance that there are no points in an interval $[0, t]$ is the product of $1 - \lambda\ da$ for a from 0 to t. Switching to the exponential function, this is

$$\exp\left(\int_0^t -\lambda\ da\right) = \exp(-\lambda t).$$

36.5 Density of a gamma

It was claimed that the density of T_n, the nth arrival time, is that of a gamma with parameters n and λ. Consider $\mathbb{P}(T_n \in dt)$. For the nth arrival time to fall into a small differential interval around t, three things have to happen.

1. There has to be a point in dt. This happens with probability $\lambda\ dt$.

2. There have to be $n - 1$ points in the interval $[0, t)$. Picking a particular $t_1 < t_2 < \cdots < t_{n-1}$ for these points gives a probability of $(\lambda\ dt_1)(\lambda\ dt_2) \cdots (\lambda\ dt_{n-1})$. Then this should be integrated over all $t_1 < t_2 < \cdots < t_{n-1}$. But this is rough! It is easier to integrate over all $x_1 \in [0, t], x_2 \in [0, t], \dots, x_{n-1} \in [0, t]$. The problem here is the x_i are not sorted, so there are $(n - 1)!$ different x_i that lead to the same t_i. This is compensated for by dividing by $(n - 1)!$. Therefore, the chance of this second part happening is
$$(\lambda t)^{n-1}/(n - 1)!.$$

3. There have to be no other points in the interval. This is just $\exp(-\lambda t)$ as before.

Putting these three factors together give the density of a gamma times dt as

$$[\lambda\ dt]\left[\frac{(\lambda t)^{n-1}}{(n-1)!}\right][\exp(-\lambda t)] = \frac{\lambda^n t^{n-1}\exp(-\lambda t)}{(n-1)!}\ dt$$

as given in the fact earlier.

36.6 Density of a Poisson

Now consider how to get n points in $[0, t]$. Again this can be broken down into two parts.

1. What is the chance that for a set of arrival times $t_1 < t_2 < \cdots < t_n$, that there is a point near each of $dt_1,\ dt_2\ \ldots, dt_n$. This is $(\lambda\ dt_1)(\lambda\ dt_2)\cdots(\lambda\ dt_n)$ Integrating this over all $0 < t_1 < t_2 < \cdots < t_n < t$ is tough, so again use the trick where the integration is over point vaues $(x_1, \ldots, x_n) \in [0, t]^n$ instead with no ordering. This gives $(\lambda t)^n$, but again ignores that $n!$ different x_i give a particualr set of t_i. Compensate by dividing by $n!$ to get a factor for this part of $(\lambda t)^n/n!$.

2. The rest of the interval should be empty. As always, this happens with probability $\exp(-\lambda t)$.

Combining these factors gives the Poisson density of

$$\frac{(\lambda t)^n \exp(-\lambda t)}{n!}.$$

Problems

303. Arrivals to a queue are modeled as a Poisson point process with constant rate. If on average their are 3 arrivals in the first two hours, what is the rate of the PPP?

304. A Poisson point process on a rectangle of area 3.7 square inches has 2.4 points. What is the (constant) rate of the process?

305. Arrivals to a queue are modeled as a Poisson point process of rate 2.3/hour. What is the chance that the third arrival occurs more than 10 minutes after the second arrival?

306. Arrivals to a math tutoring center are modeled as a Poisson point process of rate 11.2 per hour. What is the chance that the 10th student to arrive is more than five minutes past the arrival of the 9th student?

307. Arrivals to a queue are modeled as a Poisson point process of rate 2.3/hour. What is the chance that there is exactly 1 arrival in the first five minutes?

308. Arrivals to a math tutoring center are modeled as a Poisson point process of rate 11.2 per hour. What is the chance that the number of arrivals in the first hour is exactly 10?

309.

Suppose that the arrivals of airport shuttles at a particular stop follow a Poisson process of rate $1/[10 \text{ min}]$.

a. On average, how many shuttles will arrive in an hour?

b. What is the chance that there are no shuttles in the first 20 minutes?

c. What is the chance that the second shuttle arrives somewhere in $[15, 25]$ minutes?

310.

Suppose that defects in a strip of metal is modeled as a Poisson point process of rate 0.27 per meter.

a. On average, how many defects are in five meters of metal?

b. What is the chance that there are no defects in a meter of metal?

c. What is the chance that the second defect falls between meter 3 and meter 4 of the metal?

311. A webpage receives hits at rate 4 per minute. Suppose that five hits are received in the first five minutes. What is the chance that exactly three of them arrived in the first two minutes?

312. A bookstore receives customers at rate 3.4 per hour as a Poisson point process. If they receive 6 customers in the first hour, what is the chance that exactly 2 of these customers arrive in the first half hour?

Chapter 37

Ergodic Theorem for Continuous Time Markov chains

Question of the Day

Consider the Markov chain with infinitesimal generator

$$A = \begin{pmatrix} -1.6 & 0.4 & 1.2 \\ 1 & -5.3 & 4.3 \\ 0.9 & 2.2 & -3.1 \end{pmatrix}$$

Find the limiting distribution of this chain.

Summary

- For a continuous time Markov chain, the **underlying discrete time Markov chain** just keeps track of the states visited rather than the times at which the state jumps.
- A **continuous time Harris chain** is a Markov chain whose underlying discrete time chain is Harris.
- As with earlier chain, a continuous time Harris chain is **irreducible** if the underlying discrete chain is irreducible.
- For an irreducible continuous time Harris chain with a stationary distribution π, if all states return to themselves in finite time with probability 1, then the limiting distribution is unique and equals the stationary distribution.
- The **balance equations** for a continuous time Markov chain require that the rate at which probability is leaving a state equals the rate at which probability flows into the state.

In the Harris chain ergodic theorem there were two key ingredients: aperiodicity, and irreducibility. Because the time until the state is left is random (exponential in fact), aperiodicity is not necessary in continuous time. Instead, *only* irreducibility is necessary to have the limiting distribution equal to a unique stationary distribution when the chain is positive recurrent.

For continuous time, recall that π is stationary if

$$X_t \sim \pi \Rightarrow (\forall s > t)(X_s \sim \pi).$$

That is, even though the state changes randomly, the *distribution* of the state remains unchanged as time moves forward in the Markov chain.

Because the distribution evolves continuously in time, it is possible to describe this using derivatives.

Fact 98
A continuous time Markov chain with distribution p_t is stationary if $p_t' = 0$.

Recall that for one dimensional functions $[\exp(kt)]' = k\exp(kt)$ when the derivative is with respect to t. The derivative rule for $\exp(tA)$ is the same as in one dimension The constant A comes down to give

$$[\exp(tA)]' = A\exp(tA)$$

Since $p_t = p_0 \exp(tA)$, $p_t' = p_0 A \exp(tA)$. The matrix $\exp(tA)$ holds the probabilities of landing in states at time t, and so if the chain is irreducible eventually all the entries of $\exp(tA)$ will be strictly greater than 0. Hence the distribution p_0 is stationary iff $p_0 A = 0$.

Fact 99
For an infinitesimal generator A, any stationary distribution π has $\pi A = 0$.

Example 28
Question Suppose the infinitesimal generator is

$$A = \begin{pmatrix} -1.6 & 0.4 & 1.2 \\ 1 & -5.3 & 4.3 \\ 0.9 & 2.2 & -3.1 \end{pmatrix}$$

Find the stationary distribution by finding the left eigenvector associated with the eigenvalue 0.

Answer The eigenvalues can be found using Wolfram Alpha. Use

```
eigenvalues {{-1.6,0.4,1.2},{1,-5.3,4.3},{0.9,2.2,-3.1}}
```

to give

$$\lambda_1 \approx -7.4,\ \lambda_2 \approx -2.5,\ \lambda_3 = 0.$$

To find the unique solution (up to a constant factor) for $\pi A = 0$, use

```
solve -1.6a + 1b + 0.9c = 0 and 0.4a -5.3b + 2.2c = 0 and 1.2a + 4.3b - 3.1c = 0 an
```

to give

$$\pi = (697, 388, 808)/1893 \approx (0.3681, 0.2049, 0.4268).$$

A natural question to ask is whether or not this this unique stationary solution also a limiting distribution?

37.1 The underlying discrete time chain

To help answer this question, consider the *underlying discrete time chain* of the continuous time chain. This underlying chain consists of the different states visited during the continuous time run, but ignoring the time stamps. So if the chain jumped from state a to b at time 3.4 and then from state b to a at time 4.1 and then from a to c at time 7.9 then the underlying discrete time chain is a process with $X_0 = a$, $X_1 = b$, $X_2 = a$, and $X_3 = c$.

Definition 115
Let τ_i be the time of the ith jump in the continuous time Markov chain. So

$$\tau_i = \inf\{t : \#(\{X_{t'} : t' \in [0,t]\} = i)\}.$$

Then call $X_0, X_{\tau_1}, X_{\tau_2}, \ldots$ the **discrete time underlying chain** for the original chain.

For a countable state space Markov chain, it is straightforward to calculate the probabilities associated with the underlying discrete time Markov chain.

Fact 100
For a countable state space Markov chain where $\lambda(a,a) \neq 0$ for all states a, the probability of moving from state a to b in the underlying discrete time Markov chain is

$$p(a,b) = \frac{\lambda(a,b)}{-\lambda(a,a)}.$$

Note that since $-\lambda(a,a) = \sum_{b \neq a} \lambda(a,b)$, this is the proper normalizing constant to turn the off diagonal entries of the infinitesimal generator into a probability distribution.

Example 29
In the Question of the Day, find

$$p(a,b) = \mathbb{P}(X_{\tau_1} = b | X_0 = a).$$

This will be

$$\frac{0.4}{1.6} = \boxed{0.2500}.$$

Definition 116
A **continuous time Harris chain** is a Markov chain whose underlying discrete chain is Harris.

Definition 117
A continuous time Harris chain is **irreducible** if the underlying discrete chain is irreducible.

Theorem 14
Ergodic Theorem for continuous time Harris chains

Let X_n be a continuous time Harris chain with stationary distribution π. If $\mathbb{P}(R < \infty | X_0 = x) = 1$ for all x, then as $t \to \infty$,

$$d_{\text{TV}}([X_t | X_0 = x], \pi) \to 0.$$

As with discrete time chains, stationary distributions for continuous time Markov chains must satisfy the balance equations, which are just a way of saying that πA is the vector of all zeros.

Definition 118
For a countable state space, the stationary distribution satisfies the **balance equations**:

$$(\forall i) \left(\pi(i) \sum_{j \neq i} \lambda(i,j) = \sum_{j \neq i} \pi(j) \lambda(j,i) \right).$$

37.2 The limiting behavior of matrix exponentials

If A is the infinitesimal generator for a chain, what is happening to $\exp(tA)$ as t goes to infinity?

Recall that $\exp(tA)$ is defined as:

$$exp(tA) = I + tA + \frac{t^2 A^2}{2!} + \frac{t^3 A^3}{3!} + \cdots.$$

A nice fact from linear algebra is that every infitesimal generator is similar to its Jordan Normal Form (also known as the Jordan Canonical Form.) That is,

$$A = PJP^{-1},$$

where eigenvalues of A are on diagonal.

For infinitesimal generators, all of the eigenvalues must be real, otherwise the probabilities would be complex! This is part of a general fact from linear algebra that *self-adjoint* operators always have real eigenvalues.

When the eigenvectors of A span n dimensional space, it turns out that J is a diagonal matrix D. For example:

$$\begin{pmatrix} -1.6 & 0.4 & 1.2 \\ 1.0 & -5.3 & 4.3 \\ 0.9 & 2.2 & -3.1 \end{pmatrix} = P \begin{pmatrix} -7.463 & 0 & 0 \\ 0 & -2.536 & 0 \\ 0 & 0 & 0 \end{pmatrix} P^{-1}$$

where

$$P = \begin{pmatrix} -0.03247 & 0.7745 & 0.5773 \\ -0.8897 & -0.4346 & 0.5773 \\ 0.4552 & -0.4594 & 0.5773 \end{pmatrix}.$$

Now consider A^3:

$$\begin{aligned} A^3 &= (PDP^{-1})(PDP^{-1})(PDP^{-1}) \\ &= PD(P^{-1}P)D(P^{-1}P)DP^{-1} \\ &= PD^3P^{-1}. \end{aligned}$$

To cube a diagonal matrix, just cube each of the entries.

Going back to the exponential:

$$e^{tA} = I + tPDP^{-1} + t^2PD^2P^{-1}/2! + \cdots,$$

and since this is all linear transformations, if the limit exists:

$$e^{tA} = P[I + tD + t^2D^2/2! + \cdots]P^{-1} = Pe^{tD}P^{-1}.$$

This gives a procedure for finding the limit as $t \to \infty$. First find the Jordan Normal Form PDP^{-1}. Then the limit will be

$$P \lim_{t\to\infty} \exp(tD)P^{-1},$$

where $\exp(tD)$ is just $\exp(td(i,i))$ where $d(i,i)$ is the ith diagonal entry of D.

Example 30

Suppose

$$A = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}$$

The following query to Wolfram Alpha (wolframalpha.com)

```
jordan form of {{-1,1},{2,-2}}
```

gives

$$S = \begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix}, \; D = \begin{pmatrix} -3 & 0 \\ 0 & 0 \end{pmatrix}, \; S^{-1} = \begin{pmatrix} -1/3 & 1/3 \\ 2/3 & 1/3 \end{pmatrix}.$$

Note that D shows one eigenvalue 0, and the other one is negative. So

$$\begin{aligned} \exp(tA) &= Se^{tD}S^{-1} \\ &= \begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} e^{-3t} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1/3 & 1/3 \\ 2/3 & 1/3 \end{pmatrix} \\ &= \begin{pmatrix} (2/3) + (1/3)e^{-3t} & (1/3) - (1/3)e^{-3t} \\ (2/3) - (2/3)e^{-3t} & (1/3) + (2/3)e^{-3t} \end{pmatrix} \\ &= \begin{pmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{pmatrix} + e^{-3t} \begin{pmatrix} 1/3 & -1/3 \\ -2/3 & 2/3 \end{pmatrix}. \end{aligned}$$

The rows $(2/3, 1/3)$ of the first part are the same, as $t \to \infty$ the second part goes to 0 because $d(2,2) < 0$ and $\exp(td(2,2))$ will go quickly to 0 as $t \to \infty$.

Hence $(2/3, 1/3)$ is the limiting distribution.

Note that $(2/3, 1/3)$ is also stationary. A natural question to ask is whether or not this unique stationary solution also a limiting distribution?

37.3 The underlying discrete chain

To help answer this question, consider the *underlying discrete chain* of the continuous time chain. For a continuous time Markov chain, the value of X_t is completely determined by the states visited and the times that the chain jumps from state to state. The underlying discrete chain just keeps track of the states visited and not the time at which the jumps occured.

So if a CTMC starts in state a, then jumps to state b at time 4.2, then back to a at time 7.3, then to c at time 7.9, all the underlying discrete chain sees is $abac$, the states visited.

So if there is a CTMC that starts at time 0 at state a, and τ is the first time that the state jumps, what is the chance it moves to state b? Recall that the rate at which it moves to b is $\lambda(a, b)$, and the rate at which it moves away from state a is $-\lambda(a, a)$. The probability of moving to state b will be proportional to $\lambda(a, b)$.

Definition 119
For a continuous time Markov chain where for all states a, $\lambda(a, a) \neq 0$, the **underlying discrete time chain** is a discrete time Markov chain where for all states a and b,

$$p(a, b) = \frac{\lambda(a, b)}{-\lambda(a, a)}.$$

Definition 120
A **continuous time Harris chain** is a Markov chain whose underlying discrete chain is Harris.

Definition 121
A continuous time Harris chain is **irreducible** if the underlying discrete chain is irreducible.

Theorem 15
Ergodic Theorem for continuous time Harris chains

Let X_n be a continuous time Harris chain with stationary distribution π. If $\mathbb{P}(R < \infty | X_0 = x) = 1$ for all x, then as $t \to \infty$,

$$d_{\text{TV}}([X_t | X_0 = x], \pi) \to 0.$$

37.4 Solving the question of the day

Here the infinitesimal generator is

$$A = \begin{pmatrix} -1.6 & 0.4 & 1.2 \\ 1 & -5.3 & 4.3 \\ 0.9 & 2.2 & -3.1 \end{pmatrix}$$

Now find the eigenvalues and eigenvectors of the matrix in R. First set up the matrix.

```
A <- matrix(c(-1.6, 0.4, 1.2,
              1, -5.3, 4.3,
```

```
            0.9, 2.2, -3.1),
          byrow = TRUE,
          nrow = 3)
```

Next find the eigenvalues and left eigenvectors:

```
eigen(t(A))
```

```
## eigen() decomposition
## $values
## [1] -7.463737 -2.536263  0.000000
##
## $vectors
##             [,1]       [,2]      [,3]
## [1,]  0.01438817  0.8064146 0.6138606
## [2,] -0.71419107 -0.2924301 0.3417187
## [3,]  0.69980290 -0.5139845 0.7116203
```

The vector associated with the zero eigenvalue can then be normalized to give the distribution.

```
pi_unnormalized <- eigen(t(A))[["vectors"]][,3]
pi <- pi_unnormalized / sum(pi_unnormalized)
pi
```

```
## [1] 0.3681986 0.2049657 0.4268357
```

Check to see if it is stationary:

```
pi %*% A
```

```
##              [,1]          [,2] [,3]
## [1,] 1.665335e-16 -2.220446e-16    0
```

To machine precision, $\pi A = 0$. Since the underlying discrete chain is irreducible, this stationary distribution must also be the limiting distribution.

Double check by exponenting the matrix at a high time value. The `expm` function in the `expm` library does this.

```
expm::expm(40 * A)
```

```
##           [,1]      [,2]      [,3]
## [1,] 0.3681986 0.2049657 0.4268357
## [2,] 0.3681986 0.2049657 0.4268357
## [3,] 0.3681986 0.2049657 0.4268357
```

It works!

Problems

313. The stationary distribution π is a left eigenvector for the intfitesimal generator with what eigenvalue?

314. Verify that $\pi = (0.3, 0.2, 0.5)$ is the probability vector of a stationary distribution for the infinitesimal generator

$$\begin{pmatrix} -1 & 1 & 0 \\ 0.5 & -2 & 1.5 \\ 0.4 & 0.2 & -0.6 \end{pmatrix}$$

315. For a continuous time Markov chain with states $\{a, b, c\}$ and infinitesimal generator

$$\begin{pmatrix} -1 & 1 & 0 \\ 0.5 & -2 & 1.5 \\ 0.4 & 0.2 & -0.6 \end{pmatrix},$$

what is $p(b, c)$ in the discrete time underlying Markov chain?

316. Continuing the above problem, write out the entire transition matrix for the discrete time underlying Markov chain.

317. Consider a Markov chain on $\{0, 1, 2, ...\}$ which $\lambda(i, i+1) = 2$ for $i \geq 0$ and $\lambda(i, i-1) = 3$ for $i \geq 1$. Show that this is an irreducible continuous time Harris chain.

318. Consider a Markov chain on $\{a, b, c\}$ with $\lambda(a, b) = \lambda(b, c) = \lambda(c, a) = 1.2$. Show that this chain is irreducible.

319. Consider the continuous time Markov chain with infinitesimal generator

$$A = \begin{pmatrix} -2.3 & 1.1 & 1.2 & 0 \\ 3.2 & -5.0 & 1.0 & 0.8 \\ 0.6 & 0.6 & -3.2 & 2.0 \\ 1.1 & 0.5 & 2.1 & -3.7 \end{pmatrix}$$

Find the Jordan Normal Form of this generator.

320. Continuing the last problem, what are the eigenvalues of the infinitesimal generator?

321. For the continuous time Markov chain with infinitesimal generator

$$A = \begin{pmatrix} -2.3 & 1.1 & 1.2 & 0 \\ 3.2 & -5.0 & 1.0 & 0.8 \\ 0.6 & 0.6 & -3.2 & 2.0 \\ 1.1 & 0.5 & 2.1 & -3.7 \end{pmatrix},$$

find the limiting distribution by calculating $\exp(tA)$ for t large.

322. For the Markov chain with infinitesimal generator

$$A = \begin{pmatrix} -10 & 5 & 5 \\ 3 & -7 & 4 \\ 2 & 4 & -6 \end{pmatrix}$$

find the limiting distribution by finding $\exp(tA)$ for a large value of t.

Chapter 38

Replacing differential equations with CTMCs

Question of the Day

Create a continuous time Markov chain model over $\{1, 2, 3, \ldots\}$ of the differential equation

$$y' = ky.$$

Summary

- A **differential equation** contains one or more derivatives in its statement. A **first order differential equation** only contains first derivatives.
- A first order differential equation where one variable should be discrete can be turned into a continuous time Markov chain model.

One of the simplest *differential equation* models is exponential population growth. For a population y and constant k, this model is

$$y' = ky$$

or in terms of time t

$$\frac{dy}{dt} = ky.$$

> **Definition 122**
> A **differential equation** is any equation that involves one or more derivatives. If only the first derivative is used, it is a **first order** differential equation.

The equation $y' = ky$ is an example of a first order differential equation.

38.1 Modeling differential equations as CTMC

In the population growth model, it should be that $y \in \{0, 1, 2, \ldots\}$, but because the differential equation $y' = ky$ has a solution $y(t) = y_0 \exp(kt)$ which varies continuously in t, the actual output will have fractional values of y.

One way to solve this problem is to force the state space to be $\{0, 1, \ldots\}$, and then treat the derivative with respect to t as a rate at which the model jumps from one state to another.

For instance, if $k = 4.3$ and $y = 2$, then the population jumps at rate $(4.3)(2) = 8.6$ to state $y = 3$.

The CTMC model for exponential growth with $k > 0$ has state space $\{0, 1, 2, 3, \ldots\}$ and for all states i in the state space:

$$\lambda(i, i+1) = ki.$$

What if $k < 0$? Then the population $y(t) = y_0 \exp(kt)$ is decreasing as t increases. This can be modeled by making the rate from i to $i - 1$ proportional to $-ki$. This way the rate is always positive.

38.2 One variable models

Consider a differential equation of the form

$$y'(t) = f(y) - g(y)$$

where f and g are nonnegative functions and it is desired to have $y \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$

Then make

$$\lambda(i, i+1) = f(i)$$
$$\lambda(i, i-1) = g(i)$$

in the continuous time Markov chain model.

38.3 Multivariate models

Many differential equation models involve more than one variable.

38.3.1 Classic predator-prey model

The Lotka-Volterra model is a classical differential equation model that has two interacting values: the number of predators and the number of prey in a system.

$$x = \#\text{ of prey}$$
$$y = \#\text{ of predators}$$
$$a, b, c, d = \text{constants of model}$$

Given these values, the numbers of prey and predators evolve in the model as follows.

$$\frac{dx}{dt} = \underbrace{ax}_{\text{births}} - \underbrace{bxy}_{\text{deaths}}, \quad \frac{dy}{dt} = \underbrace{cxy}_{\text{births}} - \underbrace{dy}_{\text{deaths}}.$$

Figure 38.1: Wikipedia. Retrieved 27 August, 2013 from http://en.wikipedia.org/wiki/Lotka

The solution to this model exhibits periodic behavior.

Any solution to these equations will yield a noninteger numbers of prey and predators. It also disallows the possibility that the predator or prey go extinct.

Now use these rates of change to build a CTMC with similar behavior. The state of the chain at time t will be a random process (X_t, Y_t) where both X_t and Y_t are nonnegative integers.

There are four events to consider: prey birth, prey death, predator birth, and predator death. Therefore in the CTMC model, there will be four edges leaving any particular state.

From current state (x, y):

new state	rate
$(x+1, y)$	ax
$(x-1, y)$	bxy
$(x, y+1)$	cxy
$(x, y-1)$	dy

Unlike the deterministic Lotka-Volterra model which continues indefinitely, the CTMC model has a steady state. This is the absorbin state $(0, 0)$ because the rate at which this state is left is 0.

The chain will take a *very* long time to reach this absorbing state if x_0 and y_0 are reasonably sized.

These type of systems which travel for a long time before reaching the final state are called *quasistable* or *quasistationary*.

38.4 Conservation

In some models, it is important to conserve quantities. For instance, in modeling the spread of a disease the population is often taking to be constant with individual people moving between various disease states. By setting up the rates properly in a CTMC, it is possible to maintain this constant.

38.4.1 Example: the SEIR model of disease spread

There are various models of how a disease can spread through a population. One such model assigns each member of the population to one of four stages: Susceptible, Exposed, Infected, and Recovered. The only thing that matters is the four-tuple that gives the number of people in each stage.

What is conserved here is the total number of people:

$$S_t + E_t + I_t + R_t = N.$$

This model assumes the population is constant: so if someone dies from other causes (for instance, they are hit by a car and killed), it is immediately replaced by new baby (which is susceptible to the disease).

One can argue that such a model is unrealistic, but the purpose of models like this is not to predict exactly how the disease will spread. Rather this type of model gives an *explanatory* view of why certain features of disease spread happen.

Now consider the model in more detail. First write down the differential equation model.

$$\begin{aligned}
\frac{dS}{dt} &= \underbrace{b(E+I+R)}_{\text{births}} - \underbrace{\beta SI}_{\text{exposure}} \\
\frac{dE}{dt} &= \underbrace{\beta SI}_{\text{exposure}} - \underbrace{\sigma E}_{\text{infection}} - \underbrace{bE}_{\text{deaths}} \\
\frac{dI}{dt} &= \underbrace{\sigma E}_{\text{infection}} - \underbrace{bI}_{\text{deaths}} - \underbrace{\gamma I}_{\text{recovery}} \\
\frac{dR}{dt} &= \underbrace{\gamma I}_{\text{recovery}} - \underbrace{bR}_{\text{deaths}} .
\end{aligned}$$

Note that terms in this model come in pairs: dS/dt has $-\beta SI$ term, while dE/dt has $+\beta SI$ term. This is to conserve the total number of people by having a person move from type S to type E.

This pairwise behavior can be enforced in a CTMC using the moves made by the chain. Let (s, e, i, r) be the current state (all nonnegative integers.) Then the rate at which the state jumps is given as follows.

new state	rate
$(s-1, e+1, i, r)$	βsi
$(s+1, e, i, r-1)$	br
$(s+1, e, i-1, r)$	bi
$(s+1, e-1, i, r)$	be
$(s, e-1, i+1, r)$	σe
$(s, e, i-1, r+1)$	γi

Note that a new state such as $(s+1, e-1, i, r)$ has the same total population as the original state (s, e, i, r). This jump just represents a susceptible individual becoming exposed.

Unlike the differential equation model, this model can reach a steady state where no one has the disease. Therefore it is possible to ask questions in this model such as what is the average time needed for the infection to die out in the population.

38.4.2 Example: Population genetics and the Hardy-Weinberg model

Individuals in a population carry *genes* that are passed down from parent to child. Different variants of a gene can arise due to imperfect copying of a gene during this process.

In animals with a *diploid chromosome*, every individual has two copies of a gene, the order of which does not matter. So if a particular gene comes in variations A and a, then an individual might be AA, Aa, or aa.

When two parents have a child, the first parent uniformly at random contributes one of their genes, and the other parent contributes uniform at random one of their genes.

So two parents that are both AA will have a child that is AA. But if one parent is AA and the other is Aa, then the child could be AA or Aa. And if both parents are Aa, then the child could be AA, Aa, or aa.

It is of interest to determine in a population if a gene actually makes the child more or less likely to survive in the population. If the gene variant does not affect survival, the population is expected to be in *Hardy-Weinburg equilibrium*.

Definition 123
A population with random mating and no selection, no migration is in **Hardy-Weinberg Equilibrium**.

A gene that does not affect mating or survival can be modeled using the following differential equations.

$$\begin{aligned}
\frac{dy_{AA}}{dt} &= b[y_{AA}^2 + (1/2)y_{AA}y_{Aa} + (1/2)y_{Aa}y_{AA} + (1/4)y_{Aa}y_{Aa}]/N - dy_{AA} \\
\frac{dy_{Aa}}{dt} &= b[y_{AA}y_{Aa} + (1/2)y_{Aa}^2 + y_{Aa}y_{aa}]/N - dy_{Aa} \\
\frac{dy_{aa}}{dt} &= b[(1/4)y_{Aa}^2 + (1/2)y_{Aa}y_{aa} + y_{aa}^2]/N - dy_{aa}.
\end{aligned}$$

Adding these equations together gives:

$$\frac{dN}{dt} = (b-d)N.$$

When the birth rate b does not equal the death rate d, the total population is not conserved, but it will always be true that

$$N = y_{AA} + y_{Aa} + y_{aa}.$$

To model this behavior as a CTMC, for current state $(n_{AA}, n_{Aa}, n_{aa}, n)$, where $n = n_{AA} + n_{Aa} + n_{aa}$, any change in the state should preserve this final equation. If the change in state is $(1, 0, 0, 1)$, then the jump moves from state $(n_{AA}, n_{Aa}, n_{aa}, n)$ to state $(n_{AA} + 1, n_{Aa}, n_{aa}, n + 1)$.

change in state	rate
$(1, 0, 0, 1)$	$b[n_{AA}^2 + (1/2)n_{AA}n_{Aa} + (1/2)n_{Aa}n_{AA} + (1/4)n_{Aa}^2]/n$
$(0, 1, 0, 1)$	$b[n_{AA}n_{Aa} + (1/2)n_{Aa}^2 + n_{Aa}n_{aa}]/n$
$(0, 0, 1, 1)$	$b[(1/4)n_{Aa}^2 + (1/2)n_{Aa}n_{aa} + n_{aa}^2]/n$
$(-1, 0, 0, -1)$	dn_{AA}
$(0, -1, 0, -1)$	dn_{Aa}

change in state	rate
$(0, 0, -1, -1)$	dn_{aa}

Note that these rates are always positive. Since the differential equation has negative terms, as earlier this is represented in the CTMC model by using positive rates for state changes that are negative. For instance, the move $(-1, 0, 0, -1)$ has a positive rate in the CTMC model, making the coefficient of the derivative negative in the DE model.

Problems

323. For the CTMC model of exponential growth $y' = ky$, if $\lambda(2, 3) = 10$, what is k?

324. For the CTMC model of exponential growth $y' = ky$, if $\lambda(2, 1) = 10$, what is k?

325.

For the state space $\{0, 1, 2, ...\}$, and CTMC where $\lambda(i, i + 1) = 3i$ for all i, answer the following.

a) What is the expected time spent in state 4 before jumping?

b) Will the state of the chain converge?

326.

For the state space $\{0, 1, 2, ...\}$, and CTMC where $\lambda(i, i - 1) = 3i$ for all $i \geq 1$, answer the following.

a) What is the expected time spent in state 4 before jumping?

b) Will the state of the chain converge?

327. Suppose that from state (i, j) there are positive rate edges to $(i + 1, j - 1)$ and $(i - 1, j + 1)$. What simple linear function of the i and j is being conserved here?

328. Suppose that from state (i, j) there are positive rate edges to $(i + 1, j - 2)$ and $(i - 1, j + 2)$. For what value of k is $i + kj$ being conserved in this chain?

::::

Chapter 39

Square Integrability

Question of the Day

Are stochastic processes with a common upper bound on the second moment always uniformly integrable?

Summary

- A process $\{X_\alpha\}$ is **dominated** when there exists Y with $\mathbb{E}[|Y|] < \infty$ such that $|X_\alpha| \leq Y$ for all α.
- A process is **square integrable** when there exists M such that $\mathbb{E}[X_\alpha^2] \leq M$ for all α.
- A process that is either dominated or square integrable is also uniformly integrable.

39.1 *The main result*

Recall the following two facts about stochastic processes.

1. $X_n \to X$ in probability means

$$(\forall \epsilon > 0)\left(\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0\right).$$

2. When $X_n \to X$ in probability $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$ if and only if the $\{X_n\}$ are uniformly integrable, which means

$$\lim_{B\to\infty}\left(\sup_n \mathbb{E}(|X_n|\mathbb{I}(|X_n| > B))\right) = 0.$$

Earlier, it was noted that if a process $\{X_\alpha\}$ was *dominated* by an integrable random variable, that is, if $(\forall \alpha)(|X_\alpha| \leq Y)$ where $\mathbb{E}[Y] < \infty$, then the $\{X_\alpha\}$ are uniformly integrable.

Now consider if the *second moment* of the $\{X_\alpha\}$ is bounded.

Definition 124
A stochastic process $\{X_\alpha\}$ is **square integrable** if there exists an M such that $\mathbb{E}[X_\alpha^2] \leq M$ for all α.

Square integrability is often much easier to show than uniform integrability. The good news is that the former implies the latter.

Theorem 16
Square integrable processes are uniformly integrable.

39.2 *Applications*

39.2.1 Simple symmetric random walk on the integers

Consider the Markov chain on the integers that moves left with probability $2/3$ (so from state i to $i-1$) and right with probability $1/3$ (so from state i to $i+1$).

Let $X_0 = x_0 > 0$ and $T = \inf\{t : X_t = 0\}$. To apply the Optional Sampling Theorem, it is necessary to learn if $X_{t\wedge T}$ is uniformly integrable.

Consider $X_0 = x_0 > 0$ and $X_{t+1} = X_t + D_{t+1}$ where the D_i are iid and $\mathbb{P}(D_i = -1) = 2/3$ and $\mathbb{P}(D_i = 1) = 1/3$ with $T = \inf\{t : X_t = 0\}$ Then $\{X_{t\wedge T}\}$ is square integrable

The proof is by induction. In fact, it is possible to show that

$$\mathbb{E}[X_{t\wedge T}^2] \leq \max\{5, x_0^2\}.$$

Base case: when $t = 0$, $X_0 = x_0$ so $\mathbb{E}[X_{t\wedge T}^2] = x_0^2$.

Induction hypothesis: $\mathbb{E}[X_{t\wedge T}^2] \leq \max\{5, x_0^2\}$, consider $\mathbb{E}[X_{(t+1)\wedge T}^2]$. Use our standard trick:

$$\mathbb{E}[X_{(t+1)\wedge T}^2] = \mathbb{E}[\mathbb{E}[X_{(t+1)\wedge T}^2 | X_{t\wedge T}]].$$

Look at the inside expectation first. Things will be different if $T \leq t$ or $T > t$, so break this up using indicator functions:

$$\begin{aligned} X_{(t+1)\wedge T} &= X_{(t+1)\wedge T}\mathbb{I}(T \leq t) + X_{(t+1)\wedge T}\mathbb{I}(T > t) \\ &= X_{(t+1)\wedge T}\mathbb{I}(T > t) \end{aligned}$$

because if $T \leq t$ then $X_{t\wedge T} = X_{(t+1)\wedge T} = 0$. Now

$$X_{(t+1)\wedge T}\mathbb{I}(T > t) = [X_{t\wedge T} + D_{t+1}]\mathbb{I}(T > t).$$

Plug this into out expectation to get:

$$\begin{aligned} \mathbb{E}[X_{(t+1)\wedge T}^2 | X_{t\wedge T}] &= \mathbb{E}[(X_{t\wedge T} + D_{t+1})^2]\mathbb{I}(T > t)^2 | X_{t\wedge T}] \\ &= \mathbb{E}[(X_{t\wedge T}^2 + 2X_{t\wedge T}D_{t+1} + D_{t+1}^2)\mathbb{I}(T > t) | X_{t\wedge T}]/ \end{aligned}$$

This last equation used the fact that the square of an indicator function is just the same indicator function since its value is 0 or 1.

To simplify the right hand side, note:

- For all t, D_{t+1}^2 is either $(-1)^2$ or 1^2, both of which equal 1.
- For all t, D_{t+1} is independent of $X_{t \wedge T}$.
- The function $\mathbb{I}(T > t) = \mathbb{I}(X_{t \wedge T} > 0)$, so it acts as a constant in the conditional expectation.
- $\mathbb{E}[D_{t+1}] = (1/3)(1) + (2/3)(-1) = -1/3$.

Taken together, these give us

$$\mathbb{E}[X_{(t+1)\wedge T}^2 | X_{t \wedge T}] = \mathbb{I}(T > t)\mathbb{E}[X_{t\wedge T}^2 + 2X_{t \wedge T}(-1/3) + 1].$$

Now if $X_{t \wedge T}$ is at most 2, that says:

$$\mathbb{E}[X_{(t+1)\wedge T}^2 | X_{t \wedge T}] = \mathbb{I}(T > t)[4 + 1] \leq 5$$

On the other hand, if $X_{t \wedge T}$ is at least 3, then

$$\begin{aligned}\mathbb{E}[X_{(t+1)\wedge T}^2 | X_{t \wedge T}] &\leq \mathbb{I}(T > t)[X_{t\wedge T} + 2(3)(-1/3) + 1] \\ &\leq X_{t \wedge T} \leq \max\{5, x_0^2\}\end{aligned}$$

which completes the induction.

Therefore X_t is square integrable, and so is uniformly integrable as well.

39.3 Proof square integrability implies uniform integrability

Now consider why any square integrable process must be uniformly integrable.

Proof. Suppose there exists M with $\mathbb{E}[X_\alpha^2] \leq M$ for all α. The goal is to show $\sup_\alpha \mathbb{E}[|X_\alpha|\mathbb{I}(|X_\alpha| > B)$ goes to 0 as $B \to \infty$.

The key observation in the proof is that

$$|X_\alpha|\mathbb{I}(|X_\alpha| > B) \leq |X_\alpha|^2\mathbb{I}(|X_\alpha| > B)/B.$$

Taking the mean of both sides:

$$\begin{aligned}\mathbb{E}[|X_\alpha|\mathbb{I}(|X_\alpha| > B)] &\leq \mathbb{E}[|X_\alpha|^2\mathbb{I}(|X_\alpha| > B)/B] \\ &\leq (1/B)\mathbb{E}[X_\alpha^2] \leq M/B.\end{aligned}$$

This holds for any α, so $\sup_\alpha \mathbb{E}[|X_\alpha|\mathbb{I}(|X_\alpha| > B)] \leq M/B$. This goes to 0 as B goes to infinity. □

Returning to the $X_t = X_0 + \sum_{i=1}^t D_t$ from earlier, consider letting

$$M_t = X_t + (1/3)t.$$

Then M_t is a function of X_t and so measurable with respect to $\sigma(D_1, D_2, \ldots, D_t)$. Moreover, $|M_t| \leq (4/3)t$, so each M_t is integrable.

Finally,

$$\begin{aligned}
\mathbb{E}[M_{t+1}|X_0,\ldots,X_t] &= \mathbb{E}[X_{t+1} + (1/3)(t+1)|X_0,\ldots,X_t] \\
&= \mathbb{E}[X_t + D_{t+1} + (1/3)t + (1/3)|X_0,\ldots,X_t] \\
&= X_t + (1/3)t + \mathbb{E}[D_{t+1} + 1/3] \\
&= M_t.
\end{aligned}$$

Hence M_t is a martingale with respect to $\mathcal{F}_t = \sigma(D_1,\ldots,D_t)$. As before, let $T = \inf\{t : X_t = 0\}$.

Now check if $M_{t\wedge T}$ is uniformly integrable by checking for square integrability.

$$\begin{aligned}
\mathbb{E}[M_{t\wedge T}^2] &= \mathbb{E}[\mathbb{E}[M_{t\wedge T}^2|M_{(t-1)\wedge T}]] \\
&= \mathbb{E}[\mathbb{E}[(M_{(t-1)\wedge T} + (D_t + 1/3)\mathbb{I}(T > t-1))^2|M_{(t-1)\wedge T}] \\
&= \mathbb{E}[\mathbb{E}[M_{(t-1)\wedge T}^2 + 2M_{(t-1)\wedge T}(D_t + 1/3)\mathbb{I}(T > t-1) + \\
&(D_t + 1/3)^2\mathbb{I}(T > t-1)|M_{(t-1)\wedge T}]].
\end{aligned}$$

As before D_t and $\mathbb{I}(T > t-1)$ are independent, and $\mathbb{E}[D_t + 1/3] = 0$ so the middle term goes away leaving

$$\mathbb{E}[M_{t\wedge T}^2] = \mathbb{E}[M_{(t-1)\wedge T}^2 + \mathbb{I}(T > t-1)\mathbb{E}[(D_t + 1/3)^2]].$$

Since $\mathbb{E}(D_t + 1/3)^2 = (4/3)^2(1/3) + (-2/3)^2(2/3) = 8/9$,

$$\begin{aligned}
\mathbb{E}[M_{t\wedge T}^2] &= \mathbb{E}[M_{(t-1)\wedge T}^2 + \mathbb{I}(T > t-1)(8/9)] \\
&= \mathbb{E}[M_{(t-1)\wedge T}^2] + (8/9)\mathbb{P}(T > t-1).
\end{aligned}$$

An induction proof then gives:

$$\mathbb{E}[M_{t\wedge T}^2] = x_0^2 + \mathbb{P}(T > 0) + \mathbb{P}(T > 1) + \cdots + \mathbb{P}(T > t-1).$$

The probabilities are all nonnegative, so

$$\mathbb{E}[M_{t\wedge T}^2] \leq x_0^2 + \mathbb{P}(T > 0) + \mathbb{P}(T > 1) + \cdots = x_0^2 + \mathbb{E}[T].$$

The last line comes from the tail sum formula for expected value.

Problems

329. Suppose $\mathbb{E}[X_i^2] \leq 5$ for all $i \in \{0,1,2,\ldots\}$. Are the $\{X_i\}$ uniformly integrable?

330. Suppose $\mathbb{E}[Z_i] \leq 4 + \cos(i\tau/7)$ for all $i \in \{0,1,2,\ldots\}$. Are the $\{Z_i\}$ uniformly integrable?

331. Suppose $\mathbb{E}[Y_i^2] = i$ for all $i \in \{0,1,2,\ldots\}$. Are the $\{Y_i\}$ uniformly integrable?

332. Suppose $\mathbb{E}[Z_i] \in [2i, 3i]$ for all $i \in \{0,1,2,\ldots\}$. Are the $\{Z_i\}$ square integrable?

333. Suppose $\mathbb{E}[A^2] \leq 10$. Show that $\mathbb{E}[|A|\mathbb{I}(|A| > 4)] \leq 10/4 = 2.5$.

334. Suppose $\mathbb{E}[X^2] \leq 11.3$. What is the best upper bound you can give for $\mathbb{E}[|X|\mathbb{I}(|X| > 10)]$?

Chapter 40

Walds equation

Question of the Day

Suppose that I roll a fair six sided die until the sum of the numbers is at least 100. Give an interval bounding the expected number of rolls needed.

Summary

- Wald's Equation: Consider $X_1, X_2, \ldots$ iid X and a stopping time T such that X is integrable and either $X_i \geq 0$ with probability 1 or $\mathbb{E}[T] < \infty$. Then the mean of the sum of the X_i up to a stopping time equals the product of the mean of the random variables times the mean of the stopping time.

Abraham Wald was a statistician and mathematician who contributed greatly to the problem of taking data up until a criterion is reached. He was born 1902 in Austria, and then fled to the U.S. in 1938 because he was Jewish.

He is best known for his work on sequential experiments. These are used when random number of experiments are conducted. For instance, during a vaccine trial, participants could be split into two groups, one which receives the vaccine and another that does not. The latter is called the *control group*. Then they are recorded whenever a participant becomes ill with the disease. The trial ends when the number of patients in the control group with the disease is much higher than the number of patients in the vaccine group. This was the type of testing done with the COVID-19 vaccine.

A natural question in statistics is how many patients must be used and what "much higher" is exactly in order to give the trial statistical significance. Wald's Equation (a.k.a Wald's Equality a.k.a. Wald's Identity) came out of Wald's attempt to answer questions like these. It provides a shortcut for many problems that could use the Optional Sampling Theorem instead for their solution.

40.1 Random sums of random variables

Recall that for any iid sequence $X_1, X_2, \ldots$ of integrable random variables and any fixed positive integer n,

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = n\mathbb{E}[X_i].$$

This follows from the linearity of expectation.

But does this necessarily hold for *random* sums of the random variables? For instance, if T is a random variable, does

$$\mathbb{E}\left(\sum_{i=1}^{T} X_i\right) = \mathbb{E}[T]\mathbb{E}[X_i]?$$

When T is a stopping time with respect to the X_i, Wald's Equation gives two very simple conditions under which this equation does hold.

Theorem 17
Wald's Equation

Let $X_1, X_2, \ldots$ be iid as X, and T be a stopping time with respect to the natural filtration on $X_1, X_2, \ldots$. Suppose that either

1. $\mathbb{P}(X \geq 0) = 1$ or
2. $\mathbb{E}[T] < \infty$.

Then it holds that

$$\mathbb{E}\left(\sum_{n=1}^{T} X_n\right) = \mathbb{E}[T]\mathbb{E}[X_i].$$

40.2 Solving the Question of the Day

In the Question of the Day, the rolls are of a fair six sided die. So say that $X_1, X_2, \ldots$ are iid with

$$X_i \sim \mathsf{Unif}(\{1, 2, 3, 4, 5, 6\}).$$

The dice are rolled until their sum is at least 100, so our stopping time (with respect to the adapted filtration) becomes:

$$T = \inf\{t : X_1 + \cdots + X_t \geq 100\}.$$

Since $X_i \geq 1$ and $X_i \leq 6$ for all i,

$$\text{ceil}\left(\frac{100}{6}\right) \leq T \leq 100,$$

and so the mean is in $[17, 100]$.

Since the mean is finite, Wald's Lemma applies, and

$$\mathbb{E}\left(\sum_{i=1}^{T} X_i\right) = \mathbb{E}[T]\mathbb{E}[X_i].$$

For a fair, six sided die,

$$\mathbb{E}[X_i] = \frac{1+6}{2} = 3.5.$$

Note that because the last die roll that makes the sum at least 100 is at most six, it holds that

$$\mathbb{E}\left(\sum_{i=1}^{T} X_i\right) \in \{100, 101, 102, 103, 104, 105\}.$$

Given as an inequality

$$100 \leq \mathbb{E}\left(\sum_{i=1}^{T} X_i\right) \leq 105.$$

Using Wald's equation then gives

$$100 \leq \mathbb{E}[T]\mathbb{E}[X_i] \leq 105,$$

or

$$\frac{100}{3.5} \leq \mathbb{E}[T] \leq \frac{105}{3.5}.$$

Simplifying gives

$$\boxed{28.57 \leq \mathbb{E}[T] \leq 30}.$$

40.3 *Proof of Wald's Equation*

Wald's Equation holds when the $X_i \geq 0$ with probability 1 or when the mean of the stopping time is finite. To prove this, start with the case that the X_i are nonnegative.

Proof. Assume $\mathbb{P}(X_i \geq 0) = 1$.

Begin by writing the sum over i from 1 to T as an infinite sum by putting the indicator function in the summand.

$$\mathbb{E}\left(\sum_{i=1}^{T} X_i\right) = \mathbb{E}\left(\sum_{i=1}^{\infty} X_i \mathbb{1}(i \leq T)\right) = \sum_{i=1}^{\infty} \mathbb{E}[X_i \mathbb{1}(i \leq T)],$$

where the last step uses the Monotone Convergence Theorem and the fact that the summand is nonnegative.

To break apart $\mathbb{E}[X_i \mathbb{1}(i \leq T)]$, note that X_i and $\mathbb{1}(i \leq T)$ are actually independent. This is because knowing $X_1, \dots, X_{i-1}$ are sufficient to determine if $T \leq i-1$, and $(i \leq T) = \neg(T \leq i-1)$. So $(T \geq i) \in \sigma(X_1, \dots, X_{i-1})$, and these are independent of X_i. Hence

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{T} X_i\right) &= \sum_{i=1}^{\infty} \mathbb{E}[X_i]\mathbb{E}[\mathbb{1}(i \leq T)] \\
&= \sum_{i=1}^{\infty} \mathbb{E}[X_i]\mathbb{P}(i \leq T) \\
&= \mathbb{E}[X] \sum_{i=1}^{\infty} \mathbb{P}(T \geq i)
\end{aligned}$$

since $\mathbb{E}[X_i] = \mathbb{E}[X]$ for all i.

Finally,

$$\sum_{i=1}^{\infty} \mathbb{P}(T \geq i) = \mathbb{E}[T],$$

as this is just the tail sum formula for the expected value of nonnegative integer random variables, finishing the proof. ☐

Now consider the case when T is an integrable random variable.

Proof. Suppose T is an integrable stopping time for $X_1, X_2, \dots$ with respect to the natural filtration, and the X_i are iid X where

$$\mu = \mathbb{E}(X_i) < \infty.$$

Create a martingale out of the sums of X_i using

$$M_n = \sum_{i=1}^{n} (X_i - \mu).$$

To see that the $\{M_n\}$ form a martingale with respect to

$$\mathcal{F}_n = \sigma(X_1, \dots, X_n),$$

first note that the M_n are a computable function of $(X_1, \dots, X_n)$. Next,

$$\mathbb{E}[|M_n|] \leq \mathbb{E}[|X_1 - \mu + X_2 - \mu + \cdots + X_n - \mu|] \leq \mathbb{E}[|X_1| + \cdots + |X_n| + n|\mu|].$$

Since X_i has finite expectation, $\mathbb{E}[|X_i|]$ is also finite, and $\mathbb{E}[|M_n|] \leq n(\mathbb{E}(|X_i|) + \mu) < \infty$.

Finally, consider

$$\begin{aligned}
\mathbb{E}[M_{n+1}|F_n] &= \mathbb{E}[(X_1 - \mu) + \cdots + (X_{n+1} - \mu)|X_1, \dots, X_n] \\
&= (X_1 - \mu) + \cdots + (X_n - \mu) + \mathbb{E}[X_{n+1} - \mu] \\
&= (X_1 - \mu) + \cdots + (X_n - \mu) \\
&= M_n.
\end{aligned}$$

So the $\{M_n\}$ do form a martingale.

Next show that $M_{t\wedge T}$ is a uniformly integrable martingale. The approach is to show that the $M_{t\wedge T}$ are dominated by an integrable random variable that is the sum of the absolute values of the summands in the martingale. Let

$$Y = \sum_{i=1}^{T} |X_i - \mu| \geq |M_{t\wedge T}|$$

by the triangle inequality.

Since each $|X_i - \mu|$ is nonnegative, can use the nonnegative version of Wald's Equation that was just shown to get

$$\mathbb{E}[Y] = \mathbb{E}[|X_i - \mu|]\mathbb{E}[T] < \infty.$$

Therefore $|M_{t\wedge T}| \leq Y$, where Y is an an integrable random variable.

Because the $M_{t\wedge T}$ stochastic process is uniformly integrable, the OST can be used to say

$$\mathbb{E}[M_T | M_0] = M_0,$$

or in this case

$$\mathbb{E}\left[\sum_{i=1}^{T}(X_i - \mu)\right] = \mathbb{E}\left[\sum_{i=1}^{T} X_i - \mu T\right] = 0.$$

Adding $\mathbb{E}(\mu T) = \mu\mathbb{E}[T]$ to both sides completes the proof. □

40.4 Example: American Roulette

An American Roulette wheel contains 18 red numbers, 18 black numbers, a green 0, and a green 00. Assuming the wheel is fair, a bet on red wins with probability $18/38$.

With this mind, if I start playing with \$20 and bet \$1 every spin, what is the expected number of plays made before my money is gone?

- First create a random variable X which is the amount of money won on a single spin. Set $\mathbb{P}(X = 1) = 18/38$ and $\mathbb{P}(X = -1) = 20/38$.
- So $\mu = \mathbb{E}[X] = (18/38) + (-1)(20/38) = -2/38 = -1/19$.
- Let $T = \inf\{t : X_1 + \cdots + X_t = -20\}$.
- Problem: X is not nonnegative, so to apply Wald, it is necessary to determine if $\mathbb{E}[T] < \infty$.
- One way to tackle this is to let $T_a = \inf\{t : X_1 + \cdots + X_t \in \{-20, a\}\}$.
- Then for $S_t = X_1 + \cdots + X_t$, $S_{t\wedge T_a}$ is a finite state Markov chain with two recurrent communication classes: $\{a\}$ and $\{-20\}$. This makes $\mathbb{E}[T_a] < \infty$.
- Hence
$$\mathbb{E}\left[\sum_{i=1}^{T_a} X_i\right] = \mathbb{E}[T_a](-1/19).$$
- Since $-20 \leq \mathbb{E}[\sum_{i=1}^{T_a} X_i]$,
$$\mathbb{E}[T_a] \leq 20(19).$$
- Since T_a are nonnegative, increasing, and converge to T,
$$\mathbb{E}[T] \leq 20(19).$$
- That means we can apply Wald to T! This gives
$$\mathbb{E}[\sum_{i=1}^{T} X_i] = \mathbb{E}[T](-2/38) \Rightarrow \mathbb{E}[T] = 20(38/2) = \boxed{380}.$$

To summarize this approach:

1. Find stopping times that converge to our stopping time of interest but which are bounded.

2. Use Wald to upper bound the value of the converging stopping times.

3. Use them to show that the limit of the stopping times is finite.

4. Use Wald with the original stopping time to get the final answer.

Problems

335. Suppose $\mathbb{P}(X = 0) = 0.3$ and $\mathbb{P}(X = 2) = 0.7$. Say $X_1, X_2, \ldots$ with the same distribution as X. What is $\mathbb{E}[T]$ for
$$T = \inf\{t : X_1 + X_2 + \cdots + X_t = 10\}?$$

336. Suppose $\mathbb{P}(X = 0) = 0.9$ and $\mathbb{P}(X = 5) = 0.1$. Say $X_1, X_2, \ldots$ with the same distribution as X. What is $\mathbb{E}[T]$ for
$$T = \inf\{t : X_1 + X_2 + \cdots + X_t = 10\}?$$

337. Suppose $X \in [0, 1]$, and $\mathbb{E}[X] = 0.7$. Say $X_1, X_2, \ldots$ with the same distribution as X. Give the tightest bound you can on $\mathbb{E}[T]$ for
$$T = \inf\{t : X_1 + X_2 + \cdots + X_t \geq 10\}?$$

338. Suppose $X \in [0, 0.5]$, and $\mathbb{E}[X] = 0.2$. Say $X_1, X_2, \ldots$ with the same distribution as X. Give the tightest bound you can on $\mathbb{E}[T]$ for
$$T = \inf\{t : X_1 + X_2 + \cdots + X_t \geq 8.2\}?$$

339. Let $B_1, B_2, \ldots$ be iid Bernoulli random variables with mean p. Given that a geometric random variable can be defined as
$$G = \inf\{t : B_1 + \cdots B_t = 1\},$$
use Wald's Equality to find $\mathbb{E}[G]$.

340. Let $B_1, B_2, \ldots$ be iid Bernoulli random variables with mean p. Given that a negative binomial random variable with parameters r and p can be defined as
$$N = \inf\{t : B_1 + \cdots B_t = r\},$$
use Wald's Equality to find $\mathbb{E}[N]$.

341. Suppose $X = -1$ with probability 0.6 and $X = 1$ with probability 0.4. For $X_1, X_2, \ldots$ iid X, find the expected value of T, where
$$T = \inf\{t : X_1 + \cdots + X_t = -10\}.$$

342.

A player betting \$1 on American Roulette on 17 wins \$35 dollars with probability $1/38$ and otherwise loses the dollar.

a. What is the probability that if the player starts with $35 and plays until they either lose their money or their number comes up, that the player wins before they lose their $35?

b. If the player starts with $35 and plays until they have nothing, what is the expected number of spins of the wheel the player takes?

Chapter 41

Stochastic Integration

Question of the Day

Suppose a person models a a stock price as $10+B_t$, where B_t is standard Brownian motion. If the Brownian motion falls below -10, then the stock price is 0. Hence the price is

$$S_t = (10 + B_t)\mathbb{I}(B_t \geq -10).$$

The person holds five shares of this stock if $S_t \in [0, 20]$, ten shares if $S_t \geq 20$, and zero shares if $S_t = 0$. What is the distribution of the random variable for the profit made from this portfolio?

Summary

- For a stock at price S_t, and if the number of shares is constant at y, then $y(S_T - S_0)$ is the amount of money made by owning the stock from

- For a stock at price S_t, and Y_t the number of shares owned at time t where Y_t is measurable with respect to $\{S_r\}_{r \leq t}$, then

  $$\int_{t=0}^{T} Y_t \, dS_t$$

  represents the profit made over the interval $[0, T]$

Integral in differential calculus can be thought of as the *accumulation* of something over time.

- Area under a curve accumulates as you move along the horizontal axis.

$$\int_{x=a}^{b} f(x) \, dx.$$

- Water filling a swimming pool accumulates at a (possibly changing rate) over time.

$$\int_{t=0}^{T} r(t)\,dt.$$

- Volume accumulates at rate equal to the surface area of a solid object as you pass through that object.

$$\int_{z=0}^{z_{\max}} A(z)\,dz.$$

Note that in all these examples the differential dx, dt, dz are nonnegative. The integrands $f(x)$ and $r(t)$ are allowed to be positive, zero, or negative.

41.1 Financial models

Today the problem at hand will involve the money to be made by investing in a financial asset. For instance, if I buy five shares of stock at $10 per share, this costs me $50. If the stock goes up to $12 per share at time $T = 2$ days, then my five shares of stock are now worth a total of $5(12) = 60$ dollars. My profit if I sell the stock (assuming no transaction fees) is $60 - 50 = 10$ dollars.

In general, if S_t represents the (possibly random) evolution of the price of the stock, then

$$\int_{t=0}^{T} 5\,dS_t = 5(S_T - S_0)$$

is the amount of profit made by holding the five shares of stock from time 0 up to time T. (This can be generalized to $t \in [a, b]$.)

Unlike in the area, water, and volume examples above, the dS_t here is random and can be positive or negative!

> In mathematical finance, it is common to use T for a fixed time, whereas earlier it usually represented a stopping time. Be careful to examine the problem and see if T is fixed or not!

Example 31
Suppose a stock price evolves deterministically, and is 300 up until time 10 when a shock drops the price to 100.

What is

$$\int_{t=0}^{20} 3\,dS_t$$

Answer At time $t = 0$ the 3 shares cost $3S_0 = 900$, and at time $t = 20$ the 3 shares cost $3S_{20} = 300$. Therefore the integral evaluates to $300 - 900 = \boxed{-600}$.

Example 32
Suppose a stock price evolves deterministically, and is

$$S_t = 100 + 3t.$$

What is

$$\int_{t=0}^{20} 3 \, dS_t$$

Answer At time $t = 0$ the 3 shares cost $3S_0 = 300$, and at time $t = 20$ the 3 shares cost $3S_{20} = 480$. Therefore the integral evaluates to $480 - 300 = \boxed{180}$.

41.2 *Varying shares*

The next step is to allow the ownership of a number of shares Y_t that depends on the values of S_r for all $r \in [0, t]$. In other words, based on the history of stock prices, the investor is allowed to decide how many shares of the stock to own.

In this case, the amount of profit taken by owning the stock from time 0 to time T is written as the *stochastic integral*

$$\int_{t=0}^{T} Y_t \, dS_t.$$

Example 33
Suppose that a stock $S_t = 100 + 3t$, and at time t I own $Y_t = S_t$ shares of stock. What is

$$\int_0^{20} Y_t \, dS_t = \int_0^{20} S_t \, dS_t$$

intuitively?

Answer Think how much profit is made from time t up to time $t + dt$. How much profit $S_t \, dS_t$ is made from the differential change at time t?

Since S_t is deterministic,

$$dS_t = 3 \, dt$$

at time t, so

$$S_t \, dS_t = (100 + 3t)3 \, dt,$$

and

$$\int_0^{20} S_t \, dS_t = \int_0^{20} (100 + 3t)3 \, dt = 300t + (9/2)t^2|_0^{20} = \boxed{7800}.$$

41.3 Nondifferential, random stock prices

Generalizing the last example, when S_t is differentiable:

$$\int_0^T Y_t \, dS_t = \int_0^T Y_t S_t' \, dt.$$

But as seen earlier, standard Brownian motion is *not* differentiable, so if S_t is a function of Brownian Motion, it is not possible to use this approach.

In addition, if the stock price is random, then the profit made will be random as well.

Fact 101
If S_t is a random stochastic process, and Y_t measurable with respect to the information in S_r for $r \leq t$, then

$$\int_{t=0}^T Y_t \, dS_t$$

is a random variable.

Example 34
Suppose $S_t = B_t$ and I own one share of stock. What is the profit made over two time units?

Answer This is

$$\int_{t=0}^T 1 \, dB_t = B_T - B_0 \sim \mathsf{N}(0, T).$$

So in this case the stochastic integral is normally distributed.

41.4 Defining the stochastic integral

The stochastic integral that will be used here is the *Ito integral.* In the Ito integral the number of shares owned must be based on the current or past value of the stock price. There exist other integrals that look a tiny amount of differential time into the future. There are physics applications for this type of integral, but it is not suited for financial models!

Start by defining the Ito integral in the case where the number of shares held only changes at fixed times. Such a share strategy is called a *simple strategy.*

Definition 125
Say that Y_t is a **simple strategy** if it can be written in the form

$$Y_t = y_1 \mathbb{I}(t \in [0, t_1)) + y_2 \mathbb{I}(t \in [t_1, t_2)) + \cdots + y_k \mathbb{I}([t_{k-1}, t_k)),$$

that is, Y_t changes value a finite number of times.

For a simple strategy, the Ito integral can be defined as follows.

Definition 126
For $Y_t = y_1\mathbb{1}(t \in [0, t_1)) + y_2\mathbb{1}(t \in [t_1, t_2)) + \cdots + y_k\mathbb{1}([t_{k-1}, T])$, the Ito integral is defined as

$$\int_0^T Y_t\, dB_t = y_1(B_{t_1} - B_0) + y_2(B_{t_2} - B_{t_1}) + \cdots + y_k\mathbb{1}(B_T - B_{t_{k-1}}).$$

Given this definition, it is possible to prove several important properties of the Ito integral.

Fact 102
The Ito integral has the following three properties.

1. Linearity:
$$\int_0^T (aX_t + bY_t)\, dB_t = a\int_0^T X_t\, dB_t + b\int_0^T Y_t\, dB_t.$$
2. Martingale: $Z_T = \int_0^T Y_t\, dB_t$ is a martingale.
3. Second moment:
$$\mathbb{E}\left(\int_0^T Y_t\, dB_t\right)^2 = \int_0^T \mathbb{E}[Y_t^2]\, dt.$$

41.5 *What if Y_t is not simple?*

What happens if the share strategy Y_t is not simple? Things do become a lot more complicated in this situation. The approach to use is similar to that used to construct the Riemann integral.

Namely, approximate Y_t over intervals of small width with an average value. Then take the limit as width approaches 0. This approach works well for strategies that are right-continuous with left limits. This is called a *cadlag* process as an acronym from the French phrase *continue à droite, limite à gauche*.

Definition 127
A stochastic process Y_t is **cadlag** if it is right-continuous and also has limits from the left.

An example of such a process is the Poisson process seen earlier.

Problems

343. What is
$$\int_{t=0}^{10} 4\, dS_t$$
where $S_t = 50 + 60\mathbb{1}(t \geq 4)$.

344. What is
$$\int_{t=0}^{10} -3\, dS_t$$
where $S_t = 100 + 50\mathbb{1}(t \geq 8)$.

345. What is the distribution of

$$W = \int_{t=0}^{20} [3\mathbb{1}(t \in [0,10)) + 5\mathbb{1}(t \in [10,20])]\, dB_t.$$

346. What is the distribution of

$$R = \int_{t=0}^{20} [-6\mathbb{1}(t \in [0,8)) + 6\mathbb{1}(t \in [8,20])]\, dB_t?$$

347. What is

$$\mathbb{E}\left(\int_{t=0}^{30} (t + B_t)\, dB_t\right)^2.$$

348. Recall for a normal random variable $X \sim \mathsf{N}(\mu, \sigma^2)$,

$$\mathbb{E}[\exp(\alpha X)] = \mu\alpha + \alpha^2\sigma^2/2.$$

What is

$$\mathbb{E}\left(\int_{t=0}^{20} \exp(B_t)\, dB_t\right)^2?$$

Chapter 42

Ito's Formula

Question of the Day

For

$$S_t = 100\exp(3t + 4B_t),$$

write $\int_{t=0}^{T} Y_t \, dS_t$ as an integral whose differential is dB_t.

Summary

- Ito's Formula allows us to write dS_t in terms of dt and dB_t.

The power of the Fundamental Theorem of Calculus is that it allows us to change the differential of the integral. For instance, if $y = x^2$, then $dy = 2x \, dx$, which means that

$$\int_{x=0}^{4} 2x \, dx = \int_{y=0}^{16} dy = 16 - 0 = 16.$$

In the same way, it would be helpful to chance stochastic integrals with respect to a differential stock price dS_t to one with a differential dt and/or dB_t.

The Ito integral

$$\int_{t=0}^{T} Y_t \, dB_t$$

tells us how much profit is made when investing in a stock whose price evolves as B_t while owning Y_t shares at time t.

However, standard Brownian motion is not actually a great model for a stock price! Usually, the stochastic process driving stock prices (or interest rates, or other financial values of interest) are functions of Brownian motion. That is,

$$S_t = f(t, B_t)$$

where $f(t, b)$ is a function that takes as input two real numbers, and returns another real number.

42.1 Exponential Brownian motion

An example of this is the *exponential Brownian motion* in the Question of the Day:

$$S_t = 100 \exp(3t + 4B_t).$$

More generally, one might have

$$S_t = s_0 \exp(\mu t + \sigma B_t)$$

where the constant μ is called the *drift* and the constant $sigma$ is called the *spread.*

42.2 Working out dS_t

For a function f, what is

$$dS_t = S_{t+dt} - S_t = f(t + dt, B_{t+dt}) - f(t, B_t)?$$

The change in time is affecting both the first and the second coordinate. To understand how this affects the differential of S_t it helps to recall how Taylor series expansions work when there is more than one variable input to the function.

Before trying two input functions, let's try one input functions. Suppose that all the derivatives of f that maps x to y are continuous. This

$$f(x + h) - f(x) = f'(x)h + (1/2)f''(x)h^2 + (1/3!)f'''(x)h^3 + \cdots$$

is the Taylor series expansion of f. For $x \in [a, b]$, this can also be written as

$$f(x + h) - f(x) = f'(x)h + (1/2)f''(x)h^2 + O(h^3).$$

Now suppose f maps (t, b) to a third value s. Then the two dimension Taylor series expansion involves partial derivatives. These act like regular derivatives where only one variable is treated as a variable and the rest act like constants. So for $\partial f(t, b)/\partial t$ to mean that t acts like a variable and t acts like a constant in the derivative. Then for f with continuous second partial derivatives:

$$f(t+h, b+g) - f(t, b) = \frac{\partial f}{\partial t}h + \frac{\partial f}{\partial b}g + \frac{1}{2}\left[\frac{\partial^2 f}{\partial t^2}h^2 + 2\frac{\partial^2 f}{\partial t \partial b}hg + \frac{\partial^2 f}{\partial b}g^2\right] + O(h^3 + h^2 g + hg^2 + g^3).$$

When B_t is plugged in for b, then the square of the change in B_t is about the change in t. That is, $(dB_t)^2 = dt$. So this gives us

$$\begin{aligned} dS_t &= f(t + h, B_t \pm h^{1/2}) - f(t, B_t) \\ &= \frac{\partial f}{\partial t}h + \frac{\partial f}{\partial b}(\pm h^{1/2}) + \frac{1}{2}\left[\frac{\partial^2 f}{\partial t^2}h^2 + 2\frac{\partial^2 f}{\partial t \partial b}(\pm h^{3/2}) + \frac{\partial^2 f}{\partial b}h\right] + O(h^3 + h^{2.5} + h^2 + h^{1.5}). \end{aligned}$$

Putting in dt for h gives Ito's Lemma.

Theorem 18
Ito's Lemma

Let $f(t, b)$ be a function with continuous second partial derivatives. Then

$$df(t, B_t) = \left[\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial b^2}\right] dt + \frac{\partial f}{\partial b}\, dB_t.$$

42.3 *Using Ito's Lemma on the Question of the Day*

For the Question of the Day,

$$S_t = f(t, B_t),$$

where $f(t, b) = 100 \exp(3t + 4b)$.

The partial derivatives are then:

$$\frac{\partial f}{\partial t} = 300 \exp(3t + 4b), \; \frac{\partial f}{\partial b} = 400 \exp(3t + 4b), \; \frac{\partial^2 f}{\partial b^2} = 1600 \exp(3t + 4b).$$

All partial derivatives of $f(t, b)$ are continuous, so Ito's Lemma applies and

$$\begin{aligned} dS_t &= df(t, B_t) \\ &= \left[300 \exp(3t + 4B_t) + \frac{1600}{2} \exp(3t + 4B_t)\right] dt + 400 \exp(3t + 4B_t)\, dB_t \\ &= f(t, B_t)[1100\, dt + 400\, dB_t]. \end{aligned}$$

Hence

$$\int_{t=0}^{T} Y_t\, dS_t = \boxed{\int_{t=0}^{T} 1100 Y_t S_t\, dt + \int_{t=0}^{T} 400 Y_t S_t\, dB_t}.$$

42.4 *Analytic solutions of stochastic integrals*

Many Riemann integrals have analytical solutions, but very few Ito integrals do. Consider:

$$\int_0^T x \exp(x)\, dx = e^T(T - 1) + 1$$

because for Riemann integrals, $dy = f'(x)\, dx$ is relatively simple.

However, because Ito's Lemma contains three terms, most Ito integrals cannot be solved analytically. There are two exceptions.

First, the basic integration of the random differential by itself is solvable:

$$\int_{t=0}^{T} 1\, dB_t = B_T - B_0 = B_T.$$

Second, consider the slightly less trivial:

$$\int_{t=0}^{T} B_t \, dB_t.$$

To solve this, set

$$f(t,b) = (1/2)[b^2 - t].$$

Then the partial derivatives are

$$\frac{\partial f}{\partial t} = -1/2, \ \frac{\partial f}{\partial b} = b, \ \frac{\partial^2 f}{\partial b^2} = 1.$$

So Ito's Lemma gives:

$$df(t, B_t) = [-1/2 + 1/2] \ dt + B_t dB_t = B_t dB_t.$$

That is,

$$\begin{aligned} I_T &= \int_{t=0}^{T} B_t \, dB_t \\ &= \int_{t=0}^{T} df(t, B_T) \\ &= f(T, B_T) - f(0, B_0) \\ &= \boxed{(1/2)[B_T^2 - T]}. \end{aligned}$$

42.5 *Checking properties of the solution*

Recall that I_T should be a martinagle. In fact, I_T is measurable against the natural filtration on standard Brownian motion since it is a function of the standard Brownian motion. Also

$$\mathbb{E}[|(1/2)(B_T^2 - T)|] \leq \mathbb{E}[(1/2)B_T^2 + T] < \infty,$$

so it is integrable.

Finally, for $T' < T$,

$$\begin{aligned} \mathbb{E}[I_T | F_{T'}] &= \mathbb{E}[(1/2)[B_T^2 - T] | B_{T'}] \\ &= (1/2)\mathbb{E}[(B_{T'} + (B_T - B_{T'}))^2 - T | B_{T'}] \\ &= (1/2)\mathbb{E}[B_{T'}^2 + 2B_{T'}(B_T - B_{T'}) + (B_T - B_{T'})^2 - T | B_{T'}] \\ &= (1/2)[B_{T'}^2 + 0 + (T - T') - T] \\ &= (1/2)[B_{T'}^2 - T'] \\ &= I_{T'}. \end{aligned}$$

That makes I_T a martingale!

Does the variance result hold? Consider

$$\begin{aligned} \mathbb{E}[I_T^2] &= \mathbb{E}[[(1/2)(B_T^2 - T)]^2] \\ &= (1/4)\mathbb{E}[B_T^4 - 2B_T^2 T + T^2] \\ &= (1/4)[3T^2 - 2T^2 + T^2] = T^2/2. \end{aligned}$$

On the other hand

$$\int_{t=0}^{T} \mathbb{E}[(B_t)^2]\,dt = \int_{t=0}^{T} t\,dt$$
$$= T^2/2$$

so the variance property does hold as well!

Problems

349.

Say $f(t,x) = 4\exp(3t - x)$.

a) Find $\partial f/\partial t$.

b) Find $\partial f/\partial x$.

c) Find $\partial^2 f/\partial x^2$.

350.

Suppose that $g(t,b) = tb^2$.

a) Find $\partial g/\partial t$.

b) Find $\partial g/\partial b$.

c) Find $\partial^2 g/\partial b^2$.

351. Say $f(t,x) = 4\exp(3t - x)$.

Find $df(t, B_t)$ in terms of dt and dB_t.

352. Suppose that $g(t,b) = tb^2$. Find $dg(t, B_t)$ in terms of dt and dB_t.

353. For $R_t = B_t^2$, write

$$\int_{t=0}^{T} Y_t\,dR_t$$

as a stochastic integral with respect to dB_t using Ito's Lemma.

354. For $Q_t = (t + B_t)^3$, write

$$\int_{t=0}^{T} Y_t\,dQ_t$$

as a stochastic integral with respect to dB_t using Ito's Lemma.

Chapter 43

Reversibility

Question of the Day

Is there a simpler condition to show a distribution is stationary?

Summary

- A Markov chain can be described using a **transition density** $p(x, y)$ such that $\mathbb{P}(X_{t+1} \in A | X_t = x) = \int_{y \in A} p(x, y) \, dy$.

- Let $\pi(y)$ be a density. Say that a Markov chain with transition density q is **reversible** if for all states x and y,

$$\pi(y)p(x, y) = \pi(y)p(y, x).$$

  These are also called the **balance equations**.

Generall, recall that a stochastic process is a *Markov chain* if

$$\mathbb{P}(X_{t+1} \in A | X_0, X_1, \dots, X_t) = \mathbb{P}(X_{t+1} \in A | X_t)$$

for all measurable sets A. This is very general, but not so helpful in practice.

One way to describe how a particular Markov chain operates is through an *update function*, which says that $X_{t+1} = \phi(X_t, R_t)$, where R_t is a source of randomness.

For today's topic, it helps to have a different way of describing Markov chains, the *transition density*. This generalizes the notion of *transition probabilities* to Markov chains with continuous state spaces.

Definition 128
A Markov chain $\{X_t\}$ has **transition density** $p(x, y)$ if for all states x and measureable sets A,

$$\mathbb{P}(X_{t+1} \in A | X_t = x) = \int_{y \in A} p(x, y) \, dy.$$

Still another way to think about Markov chains is that they let probability flow around the individual states. At each step of the chain, all the probability currently stored at state y leaves it.

The flow into state y at a step can be written in terms of the transition probabilities and the density $r(x)$ which is the current density of probability at state x.

$$\text{flow into } y = \int_{x \in \Omega} r(x)p(x,y)\,dx.$$

The *balance equations* state that probabilities are in balance if the flow of probability into a state equals the flow of probability out of a state.

Definition 129
The **balance equations** for transition density q and stationary density π are

$$(\forall y \in \Omega)(\pi(y) = \int_{x \in \Omega} \pi(x)p(x,y)\,dx$$

Fact 103
A Markov chain where the detailed balance equations are satisfied has stationary distribution given by density π.

Proof. Let A be a measurable set. Suppose X_t has density p. Then

$$\begin{aligned}
\mathbb{P}(X_{t+1} \in A) &= \mathbb{E}[\mathbb{I}(X_{t+1} \in A)] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{I}(X_{t+1} \in A)|X_t]] \\
&= \int_{x \in \Omega} \mathbb{E}[\mathbb{I}(X_{t+1} \in A)|X_t = x]\pi(x)\,dx \\
&= \int_{x \in \Omega} \mathbb{P}(X_{t+1} \in A|X_t = x)\pi(x)\,dx \\
&= \int_{x \in \Omega} \int_{y \in A} \pi(x)p(x,y)\,dy\,dx \\
&= \int_{y \in A} \int_{x \in \Omega} \pi(x)p(x,y)\,dx\,dy \text{ by Tonelli} \\
&= \int_{y \in A} \pi(y)\,dy \text{ by detailed balance} \\
&= \pi(A)
\end{aligned}$$

thereby showing the π is stationary. □

Often the first three equations are omitted. They are included here to emphasize that the next statement is a direct consequence of the Fundamental Theorem of Probability.

A stronger notion than this is the idea of *reversibility* or *detailed balance*. The balance equations state that the flow of probability out of a state matches the flow of probability into the state. The detailed balance equations say that the flow of probability from state x to y is exactly matched by the flow of probability from state y to x.

Definition 130
Say that a Markov chain with transition density q satisfies the **detailed balance** equations with respect to density π if for all states x and y

$$\pi(x)p(x,y) = \pi(y)p(y,x).$$

Flow from x to y equalling flow from y to x is a stronger statement than

Fact 104
If a Markov chain satisfies the detailed balance equations, then they also satisfy the balance equations.

Proof. Suppose the detailed balance equations are satisfied. Let y be any state of the chain. Then

$$\begin{aligned}\int_{x\in\Omega} \pi(x)p(x,y)\,dx &= \int_{x\in\Omega} \pi(y)p(y,x)\,dx \\ &= \pi(y)\mathbb{P}(X_{t+1}\in\Omega|X_t = x) \\ &= \pi(y),\end{aligned}$$

and so the chain does satisfy the balance equations as well. □

43.1 *Building a reversible chain for a specific distribution*

Because there are many fewer equations, it is often easier to make a chain satisfy detailed balance rather than the overall balance equations.

Consider the following example. Suppose that the goal is to make a chain on three states with stationary density

$$\pi(a) = 0.5,\ \pi(b) = 0.3,\ \pi(c) = 0.2.$$

To be reversible, $\pi(a)p(a,b) = \pi(b)p(b,a)$. Hence

$$\frac{p(a,b)}{p(b,a)} = \frac{\pi(b)}{\pi(a)} = \frac{3}{5}.$$

Similarly,

$$\frac{p(a,c)}{p(c,a)} = \frac{\pi(c)}{\pi(a)} = \frac{2}{5}$$

and

$$\frac{p(b,c)}{p(c,b)} = \frac{\pi(c)}{\pi(b)} = \frac{2}{3}$$

A first try might make

$$p(a,b) = 3/5, p(b,a) = 1, p(a,c) = 2/5, p(c,a) = 1, p(b,c) = 2/3, p(c,b) = 1$$

Unfortunately, these are not probabilities since (for instance) $p(b,a)+p(b,c) = 5/3 > 1$. However, dividing all the probabilities by 2, and filling in the holding probabilities does the job.

$$p(a,b) = 3/10, p(b,a) = 1/2, p(a,c) = 1/5, p(c,a) = 1/2, p(b,c) = 1/3, p(c,b) = 1/2, p(a,a) = 1/2, p(b,b) = 1/6, p(c,$$

Notice that the holding probabilities do not affect reversibility, so this satisfies the target distribution!

43.1.1 Alternate chains

There was of course many different ways to get this stationary distribution. If the first try is divided by 3 instead of two, the transition density obtained is

$$p(a,b) = 1/5, p(b,a) = 1/3, p(a,c) = 2/15, p(c,a) = 1/3, p(b,c) = 2/9, p(c,b) = 1/3, p(a,a) = 2/3, p(b,b) = 4/9, p(c,c$$

Using any constant at least 2 gives a different stationary distribution reversible with respect to the same density!

43.2 *Symmetric Markov chains*

Fact 105
Suppose $p(x,y) = p(y,x)$. Then the uniform distribution is stationary if the measure of Ω is finite.

Proof. Let π be the uniform distribution. Then for all x and y in the state space, $\pi(x) = \pi(y)$, and $\pi(x)p(x,y) = \pi(y)p(y,x)$, so π is reversible, and hence stationary. $\square$

43.2.1 Uniform permutations

- Consider the Markov chain on permutations where two elements of the permutation are chosen uniformly at random and swapped (transposed).
- So if permutation is 43215 and elements 2 and 4 are swapped, new permutation is 41235.
- Consider two permutations x and y of $\{1, 2, \ldots, n\}$ connected by a transposition.
- Then $p(x,y) = p(y,x) = 1/n^2$.
- Hence the uniform distribution is stationary.
- Any sorting algorithm shows this chain is irreducible, so it is also the unique stationary distribution.
- Aperiodic since if you pick the same element twice to swap, you hold position.
- So also the limiting distribution.

43.3 *Why "reversibility"?*

- Consider two Markov chains on $\{1, 2, 3, 4\}$.
- Chain 1: $p(1,2) = 1,\ p(2,3) = 1,\ p(3,4) = 1,\ p(4,1) = 1$.

- Chain 2: $p(1,2) = p(2,1) = 1/2$, $p(2,3) = p(3,2) = 1/2$, $p(3,4) = p(4,3) = 1/2$, $p(4,1) = p(1,4) = 1/2$.

- A run of chain 1: $1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, \ldots$.

- If you reverse it, you immediately can tell its been reversed: $4, 3, 2, 1, 4, 3, 2, 1, \ldots$.

- A run of chain 2: $1, 4, 3, 4, 3, 4, 1, 2, 3, 4, 3, 4, 2, \ldots$.

- Looks the same run forward or backwards!

Fact 106
Suppose a Markov chain is reversible with respect to π and $X_0 \sim \pi$. Then

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t) = \mathbb{P}(X_0 = x_t, X_1 = x_{t-1}, \ldots, X_t = x_0).$$

That is, the distribution of $(X_0, X_1, \ldots, X_t)$ given $X_0 \sim \pi$ has the same distribution of $(X_t, X_{t-1}, \ldots, X_0)$ given $X_t \sim \pi$.

Proof idea For simplicity, suppose that the state space is discrete. Then

$$\begin{aligned}
\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t) & \\
&= \pi(x_0)p(x_0, x_1)p(x_1, x_2) \cdots p(x_{t-1}, x_t) \\
&= \pi(x_1)p(x_1, x_0)p(x_1, x_2) \cdots p(x_{t-1}, x_t) \\
&= \pi(x_2)p(x_1, x_0)p(x_2, x_1)p(x_2, x_3) \cdots p(x_{t-1}, x_t) \\
&= \quad \vdots \\
&= \pi(x_t)p(x_t, x_{t-1}) \cdots p(x_1, x_0) \\
&= \mathbb{P}(X_0 = x_t, X_1 = x_{t-1}, \ldots, X_t = x_0).
\end{aligned}$$

With many more integral signs, this idea can be extended to continuous state spaces.

43.4 Methodical approaches

There are more methodical ways to construct reversible Markov chains whose stationary distribution matches a target. In the next chapter the most commonly used of these methods, Metropolis-Hastings, will be discussed.

Problems

355. What is the stationary distribution of a Markov chain whose transition matrix is symmetric?

356. What is the stationary distribution of a Markov chain if it is reversible with respect to π?

357. Suppose a Markov chain has two states a and b, and that $p(a, b) = 0.6$ while $p(b, a) = 0.4$. Show that this chain is reversible with respect to the probability vector $(0.4, 0.6)$.

358. Suppose a Markov chain has two states a and b, and that $p(a, b) = 0.3$ while $p(b, a) = 0.6$. Find a distribution π that the chain is reversible with respect to.

359. Create a Markov chain whose limiting distribution is uniform over $\{a, b, c, d, e\}$.

360. Create a Markov chain over $\{a, b, c\}$ with uniform limiting distribution where $p(a, b) = 0.3$ and $p(c, a) = 0$.

Chapter 44

Metropolis-Hastings

Question of the Day

Using simple symmetric random walk on the integers with a partially reflecting boundary at 0, construct a Markov chain whose stationary distribution has unnormalized density w where

$$w(i) = \frac{1}{i^3}$$

Summary

- The **Metropolis-Hastings** method takes an existing random walk over a space, and uses it to build a Markov chain whose stationary distribution is an arbitrary distribution over the space.
- A user must be somewhat careful, as even if the original random walk was connected, the new chain might not be.
- If it is connected and the distribution is anything other than uniform, the chain will automatically also be aperiodic.

44.1 Normalizing constants

The target density in the Question of the Day is

$$w(i) = \frac{1}{i^3}$$

for $i \in \{1, 2, \ldots\}$. This is an example of an *unnormalized density* since the terms do not add up to 1.

Call $Z = \sum_{j=1}^{\infty} w(j)$ the **normalizing constant** of the distribution. It turns out (as in this example) finding Z exactly can be difficult.

In fact, the calculation of Z for high dimensional problems often falls into the category of $\#P$-complete problems.

44.2 Reversibility and normalizing constants

The value of the reversibility property of Markov chains is that it can be described entirely in terms of w. There is no need to know Z to build a reversible chain.

If the target density is $\pi(x) = w(x)/Z$, then reversibility for a Markov chain with transitions $p(x, y)$ requires that

$$\pi(x)p(x, y) = \pi(y)p(y, x)$$

for all states x and y. This condition will be true if and only if

$$w(x)p(x, y) = w(y)p(y, x),$$

so there is no need to know Z to verify reversibility.

44.3 The Metropolis-Hastings approach

This idea goes back to a 1953 paper with 5 authors: Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller. (The two Rosenbluths and two Tellers were husband-wife pairs.) So occasionally you see this method referred to as MR2T^2.

Nicholas Metropolis was both the project leader and first in alphabetical order, so his name was first on paper. From a historical perspective it is unclear whether he actually helped develop the algorithm.

What is known is that he coined the term for this type of algorithm: Monte Carlo methods.

Since the original authors were physicists, they stuck to a special type of distribution called a Boltzmann distributions. Originally the proposal densities where all *symmetric*, so $q(x, y) = q(y, x)$ for all x and y.

Later, a statistician named Hastings found that the method could be extended to allow choices of q that were not symmetric.

For this reason, today the method is usually called the *Metropolis-Hastings method* or *MH chain*.

44.4 The two steps of an MH chain

1. Propose moving from $X_t = x$ to y using transitions given by $q(\cdot, \cdot)$.

2. With probability
$$\min\left\{\frac{w(y)q(y, x)}{w(x)q(x, y)}, 1\right\}$$
accept the move and set $X_{t+1} = y$. Otherwise, retain $X_{t+1} = x$.

44.5 Using Metropolis-Hastings on the Question of the Day

In the Question of the Day, the original chain is a simple symmetric random walk on the integers with a partially reflecting boundary at 0. This can be represented using the update function:

$$\phi(i, u) = i + \mathbb{I}(u > 1/2) - \mathbb{I}(u < 1/2, i > 0).$$

Another way to represent it is through the transition probabilities.

$$q(0, 0) = 1/2, \forall i \geq 0, q(i, i + 1) = q(i + 1, i) = 1/2.$$

In words, there is a fifty-fifty chance of adding or subtracting 1 to the current state, but if that move would make the chain go negative, stay at the current position.

The goal is to build a chain with stationary distribution with unnormalized density $p(i) = 1/i^3$ for {i ∈{1, 2, 3, ...} }.

The Metropolis Hastings protocol constructs such a chain from the original random walk as follows.

Given current state i:

1) Choose m uniformly from $\{-1, 1\}$.
2) If $m = 1$ propose moving the state to $i + 1$.
   a) Accept this move with probability
   $$\frac{1/(i+1)^3}{1/i^3} = i^3/(i+1)^3.$$
   b) Else reject and stay at the current position.
3) If $m = -1$, then if $i > 1$ move to state $i - 1$, else stay where you are.

44.6 *Recurrence and transience*

An interesting fact about the Question of the Day example is that the original simply symmetric random walk with partially reflecting boundary does *not* have a stationary distribution at all! However, the Metropolis-Hastings chain does.

The point is that Metropolis-Hastings can be used on any proposal chain, regardless of whether that chain has a stationary distribution. However, for the limiting distribution of the new chain to equal the target stationary distribution, it must be the case that the original chain was connected.

44.7 *The main result*

The main result is that Metropolis-Hastings constructs a reversible Markov chain with a stationary distribution equal to the

Fact 107
Suppose $q(x, y) > 0 \Rightarrow q(y, x) > 0$ for all states x and y. Then the Metropolis-Hastings chain constructed from this chain is reversible with respect to $\pi(i) = w(i)/Z$.

Proof. Let x and y be any two states in Ω with $w(x), w(y), q(x, y)$ and $q(y, x)$ all positive. Without loss of generality $w(x)q(x, y) \leq w(y)q(y, x)$. Then

$$p(x, y) = q(x, y) \min\left\{\frac{w(y)q(y, x)}{w(x)q(x, y)}, 1\right\} = q(x, y)$$

and

$$p(y, x) = q(y, x) \min\left\{\frac{w(x)q(x, y)}{w(y)q(y, x)}, 1\right\} = \frac{w(x)q(x, y)}{w(y)}.$$

Hence
$$\pi(y)p(y,x) = \frac{w(y)}{Z}\frac{w(x)q(x,y)}{w(y)} = \frac{w(x)}{Z}q(x,y) = \pi(x)p(x,y).$$

□

44.8 Continuous Time Markov chains and reversibility

Metropolis-Hastings needs the weird $\min\{\cdot, 1\}$ formulation to make sure that the probabilities created lie in $[0, 1]$. With continuous time Markov chains, this is not necessary, rates can be any positive value.

The basic idea of reversibility is that the flow of the measure from state x to state y should equal the flow from y to x. In discrete time, this requires that

$$\pi(x)p(x,y) = \pi(y)p(y,x).$$

In continuous time, this is
$$\pi(x)\lambda(x,y) = \pi(y)\lambda(y,x).$$

> **Definition 131**
> A probability distribution π on a discrete state space Ω satisfies the **detailed balance equations** (also called **reversibility** for a continuous time Markov chain if for all $a, b \in \Omega$:
>
> $$\pi(b)\lambda(b,a) = \pi(a)\lambda(a,b).$$

To get a distribution
$$\pi(x) = \frac{w(x)}{Z},$$
just make $\lambda(a,b) = 1/w(a)$ and $\lambda(b,a) = 1/w(b)$ to ensure that this happens.

44.9 Continuous state spaces

Reversibility (and Metropolis-Hastings) also work with continuous state spaces. Recall X has density f if $\mathbb{P}(X \in dx) = f(x)\,dx$.

> **Definition 132**
> A probability distribution π on a continuous state space Ω satisfies the **detailed balance equations** (also called **reversibility** for a discrete time Markov chain if for all $a, b \in \Omega$:
>
> $$\pi(dy)\mathbb{P}(X_{t+1} \in dx, X_t = y) = \pi(dx)\mathbb{P}(X_{t+1} \in dy, X_t = x).$$

When the probabilities are given in terms of densities, then both sides have a $dx\,dy$ factor that cancels out.

Problems

361.

Suppose a chain has $p(a,b) = p(b,a) = 0.4$. The goal is to create a chain with stationary measure $w(a) = 4$ and $w(b) = 5$ using Metropolis-Hastings.

a) What is the chance of accepting a move from a to b?

b) What is the chance of accepting a move from b to a?

362.

Suppose a chain has $p(a,b) = p(b,a) = 0.8$. The goal is to create a chain with stationary measure $w(a) = 10$ and $w(b) = 4$ using Metropolis-Hastings.

a) What is the chance of accepting a move from a to b?

b) What is the chance of accepting a move from b to a?

363.

Suppose a chain has $p(a,b) = 0.6$ and $p(b,a) = 0.5$. The goal is to create a chain with $w(a) = 4$ and $w(b) = 5$ using Metropolis-Hastings.

a) What is the chance of accepting a move from a to b?

b) What is the chance of accepting a move from b to a?

364.

Suppose a chain has $p(i,j) = 0.3$ and $p(j,i) = 0.5$. The goal is to create a chain with stationary measure $w(i) = 2.3$ and $w(j) = 1.7$ using Metropolis-Hastings.

a) What is the chance of accepting a move from i to j?

b) What is the chance of accepting a move from j to i?

Chapter 45

Birth Death Chains

Question of the Day

Are there chains over $\{0, 1, 2, \ldots\}$ for which if π is a stationary distribution, then the chain is reversible with respect to π?

Summary

- A Markov chain is a **birth death** chain if the state only changes by at most value 1 at each step.
- If π is stationary for a birth death chain, then the chain is reversible.
- For a continuous time birth death chain, let λ_i be the rate of movement from i to $i+1$, and μ_i be the rate of movement from i to $i-1$. Then the chain has a stationary distribution if and only if

$$\sum_{i=1}^{\infty} \frac{\lambda_0 \cdots \lambda_i}{\mu_1 \cdots \mu_{i+1}} < \infty.$$

If a Markov chain is reversible with respect to π, then π is a stationary distribution of the chain. But the reverse is not always true. That is to say, reversibility implies stationarity, but not the other way around.

Still, there is a class of chain called a *birth death chain* where reversibility and stationarity are the same thing!

Definition 133
A Markov chain on $\{0, 1, 2, \ldots\}$ is a **birth death chain** if the change in the state after one time step or jump is at most 1.

As before, let $p(x, y) = \mathbb{P}(X_{t+1} = y | X_t = x)$ for discrete time Markov chains. Then the birth death requirement is that

$$(p(x, y) > 0) \to |x - y| \leq 1.$$

Similarly, for $r(x, y)$ the rate of transition from x to y for continuous time Markov chains, the birth death requirement is that

$$(r(x, y) > 0) \rightarrow |x - y| \leq 1.$$

Fact 108
For birth death chains, if π is stationary, then the chain is reversible with respect to π.

Proof. Assume π is stationary. For a continuous time Markov chain, the balance equations are

$$\begin{aligned}
\pi(0)r(0,1) &= \pi(1)r(1,0) \\
\pi(1)[r(1,0) + r(1,2)] &= \pi(0)r(0,1) + \pi(2)r(2,1) \\
\pi(2)[r(2,1) + r(2,3)] &= \pi(1)r(1,2) + \pi(3)r(3,2) \\
\vdots &= \vdots
\end{aligned}$$

□

The first balance equation is just the first detailed balance equation between 0 and 1. To get the remaining detailed balance equations between states i and $i+1$, just add the first $i+1$ equations. For $j < i$, there will be a $\pi(j)r(j, j+1)$ and $\pi(j+1)r(j+1, j)$ on both the left and the right hand sides. The only terms that do not cancel will be

$$r(i)r(i, i+1) = r(i+1)r(3, 2),$$

giving reversibility.

From a graph perspective, the balance equations enforce the rule that for any given cut, probability flow from one side of the cut to the other is balanced by probability flow from the other to the original side. In detailed balance (reversibility), flow from an edge from i to $i+1$ must be balanced by flow from the edge from $i+1$ to i. But this is exactly the same as balance using the cut $\{0, 1, \ldots, i\})$ to $\{i+1, i+2, \ldots\}$. So really reversibility is guaranteed by the topology of the transition graph.

45.1 Births and deaths

One way to view these chains is as a population that changes from i to $i+1$ (called a *birth*) or from $i+1$ down to i (called a *death*). So $r(i, i+1)$ is the *birth rate* at state i and $r(i+1, i)$ is the *death rate* at state $i+1$. So simplify the notation, often

$$\lambda_i = r(i, i+1),\ \mu_i = r(i, i-1).$$

Note the Greek letter lambda starts with the Roman letter ell, which also starts life, and the Greek letter mu starts with the Roman letter em, which stands for the Latin *mortis* meaning death. (This is where English words such as mortal and mortuary come from.)

45.2 The Poisson process

Recall that if P is a Poisson point process of rate λ over $[0, \infty)$, then

$$N_t = \#(P \cap [0, t])$$

is a *Poisson process* of rate λ. This Poisson process can only ever increase by 1 as time advances, so it is a birth death chain where the rate of births is λ and the rate of deaths is 0.

Because the state of this chain never stops increasing, there is no stationary distribution.

45.3 Exponential growth chain

A simple model of population growth in organisms that reproduce via cloning (such as infected individuals in a pandemic) is the *exponential growth birth death process. In this process, for all i,

$$\lambda_i = \lambda i, \ \mu_i = \mu i$$

for constants λ and μ.

When $\lambda > \mu$ there is no stationary distribution, but when $\mu > \lambda$, the greater rate of deaths forces the chain to converge to 0 with probability 1.

45.4 Stationary distributons of birth death processes

More generally, a birth death process will have a stationary distribution if and only if there is a point where the rate of deaths is greater than the rate of births in a particular way.

Fact 109
A birth death process has a stationary distribution if and only if

$$\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_i}{\mu_1 \mu_2 \cdots \mu_{i+1}} < \infty.$$

Proof. Earlier, the equivalence of the balance and detailed balance equations was shown. Reversibility using the birth notation is

$$(\forall i)(\pi(i)\lambda_i = \pi(i+1)\mu_{i+1}),$$

or equivalently,

$$(\forall i)\left(\pi(i+1) = \pi(i)\frac{\lambda_i}{\mu_{i+1}}\right).$$

A simple induction uses these equations to show that for all $i \geq 1$,

$$\pi(i) = \pi(0)\frac{\lambda_0 \lambda_1 \cdots \lambda_i}{\mu_1 \mu_2 \cdots \mu_{i+1}}.$$

This means that $\sum_{i=0}^{\infty} \pi(i)$ equals

$$\pi(0)\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_i}{\mu_1 \mu_2 \cdots \mu_{i+1}} < \infty,$$

and so the distribution can be normalized as a probability distribution if and only if the sum is finite. □

45.5 The $M/M/1$ queue

A model of a queue called $M/M/1$ in queueing theory, is a birth death chain that operates as follows

- There are constants λ and μ such that $\lambda_i = \lambda$ and $\mu_i = \mu$ for all $i \geq 0$.
- The constant λ is called the *arrival rate*, and measures the rate at which new customers arrive to the queue. The constant μ is called the *service rate*, and measures the rate at which customers are served and leave the queue.

With this setup, a natural question to ask is what is the stationary distribution for this model? Since the birth and death rates are all the same,

$$\frac{\lambda_0 \cdots \lambda_i}{\mu_1 \cdots \mu_{i+1}} = \left(\frac{\lambda}{\mu}\right)^i.$$

If $\lambda < \mu$, the sum of the above quantity is a geometric series that converges to yield

$$\pi(i) = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^i.$$

45.6 The Yule process

The *Yule process* (named after the British statistician Udny Yule) came about from a process that grows faster than exponentially because of *preferential attachment*. Consider an Empire who population is naturally growing exponentially fast, but as it grows it makes it more likely that people outside the Empire will wish to join, resulting in faster than exponential growth.

Specifically, the model has a constant λ, and for all $i \geq 1$,

$$\lambda_i = \lambda i^2,\ \mu_i = 0.$$

Recall that an exponential of rate λi^2 has a mean of $1/[\lambda i^2]$, and so the average time to go from 1 to 2 is $1/\lambda$, from 2 to 3 is $1/(4\lambda)$, from 3 to 4 is $1/(9\lambda)$, and so on.

Then the time needed to get to infinity is

$$\frac{1}{1^2\lambda} + \frac{1}{2^2\lambda} + \frac{1}{3^2\lambda} + \cdots = \frac{\tau^2}{24\lambda}.$$

That is to say, the population (on average) becomes larger than any finite number after only $1.65/\lambda$ time!

When the state moves past any finite number in finite time, the model is said to have a *population explosion*! Naturally, such models typically ignore effects (such as resource limiting) that come into play as a population expands.

Problems

365. Consider an $M/M/1$ queue with $\lambda = 2$ and $\mu = 3$. What is the stationary distribution for the chain?

366. Consider an $M/M/1$ queue with $\lambda = 5$ and $\mu = 8$. What is the stationary distribution for the chain?

367. Prove that

$$\sum_{i=1}^{\infty} \frac{1}{i^{1.5}} < \infty.$$

368. Prove that if $\mu_i = 0$ for all i and $\lambda_i = 2i^{1.5}$, that the population explodes.

Problem Solutions

Chapter 1

1. Fill in the blank: In a __________, the distribution of X_n given $X_1, \dots, X_{n-1}$ only depends on the value of X_{n-1}.

Solution. This is a $\boxed{\text{Markov chain.}}$

3. An alternate name for a martingale is what?

Solution. A martingale is also called a $\boxed{\text{fair game.}}$

5. Fill in the blank: A variable with partial information is a _________ variable.

Solution. This would be a $\boxed{\text{random}}$ variable.

7. Suppose B_1 and B_2 are independent Bernoulli random variable with parameter 0.6. This means that $\mathbb{P}(B_i = 1) = 0.6$ and $\mathbb{P}(B_i = 0) = 0.4$ for i either 1 or 2. What is $\mathbb{P}(B_1 = B_2 = 1)$?

Solution. Because they are independent:

$$\mathbb{P}(B_1 = B_2 = 1) = \mathbb{P}(B_1 = 1)\mathbb{P}(B_2 = 1) = (0.6)(0.6) = \boxed{0.3600}.$$

9. Suppose you think that a company stock will go 10% with probability 70%, and will decline 5% with probability 30%. Is the change in the value of the stock a random variable?

Solution. $\boxed{\text{Yes}}$, because you only have partial information about the result.

Chapter 2

11. What is an indexed collection of random variables called?

Solution. This is a $\boxed{\text{stochastic process}}$.

12. If $\mathbb{E}(X) = 3.2$, what is $\mathbb{E}[2X+3]$?

Solution. Using linearity of expectation, this is

$$\mathbb{E}[2X+3] = 2\mathbb{E}[X] + 3 = 2(3.2) + 3 = \boxed{9.400}.$$

14. Suppose

$$\begin{aligned} \mathbb{E}(X \mid W = 0) &= 4.2 \\ \mathbb{E}(X \mid W = 1) &= 3.2 \\ \mathbb{P}(W = 0) &= 0.6 \\ \mathbb{P}(W = 1) &= 0.4 \end{aligned}$$

Find $\mathbb{E}(X)$.

Solution. This can be done with an expectation tree:

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}[X \mid W = 0]\mathbb{P}(W = 0) + \mathbb{E}[X \mid W = 1]\mathbb{P}(W = 1) \\ &= (4.2)(0.6) + (3.2)(0.4) \\ &= \boxed{3.800}. \end{aligned}$$

16.

Let

$$S = B_1 + \cdots + B_n$$

where the B_i are iid $\text{Bern}(0.3)$. (This means $\mathbb{P}(B_i = 1) = 0.3,\ \mathbb{P}(B_i = 0) = 0.7$.

a. Find $\mathbb{E}(B_i)$.

b. Find $\mathbb{E}(S)$.

c. Find $\mathbb{E}(S|B_1)$.

Solution. a. This is $(0.3)(1) + (0.7)(0) = \boxed{0.3000}$.

b. Adding this together n times gives

$$\mathbb{E}[S] = \mathbb{E}[B_1 + \cdots + B_n] = \mathbb{E}[B_1] + \cdots + \mathbb{E}[B_n] = \boxed{0.3n}.$$

c. This would be

$$\begin{aligned} \mathbb{E}[S \mid B_1] &= \mathbb{E}[B_1 + \cdots + B_n \mid B_1] \\ &= \mathbb{E}[B_1 \mid B_1] + \mathbb{E}[B_2 \mid B_1] + \cdots + \mathbb{E}[B_n \mid B_1] \\ &= B_1 + \mathbb{E}[B_2] + \cdots + \mathbb{E}[B_n] \\ &= \boxed{B_1 + 0.3(n-1)}. \end{aligned}$$

18.

In a *corner bet* in Roulette, a player is allowed to bet \$1 on four numbers simulataneously. If any of these four numbers come up, the player wins \$8 and their original bet is returned, otherwise they lose their original bet.

(a) Suppose a player is playing American Roulette with 38 slots on the wheel where the chance of winning a corner bet is $4/38$. They start with \$1 and play until their money is gone. If T is the number of plays this takes, and $\mathbb{E}[T]$ exists, find $\mathbb{E}[T]$.

(b) Repeat part (a) where the player is on a European Roulette table where the chance of winning a corner bet is $4/37$.

Solution. Let $D_1 = 8$ if the player wins their bet, and $D_1 = -1$ if the player loses their bet. Then the Fundamental Theorem of Probability (an expectation tree) gives:

$$\begin{aligned}\mathbb{E}[T] &= \mathbb{E}[\mathbb{E}[T|D_1]] \\ &= \mathbb{E}[T|D_1 = 8]\mathbb{P}(D_1 = 8) + \mathbb{E}[T|D_1 = -1]\mathbb{P}(D_1 = -1).\end{aligned}$$

If $D_1 = -1$ the player loses immediately, making $T = 1$. If $D_1 = 8$, then for the player to lose all their money, they must lose $8 + 1 = 9$ individual dollars, one after the other. Including the first spin gives

$$\mathbb{E}[T|D_1 = 8] = 1 + \mathbb{E}[T_1 + \cdots + T_9] = 1 + 9\mathbb{E}[T].$$

So

$$\mathbb{E}[T] = \mathbb{P}(D_1 = -1) + \mathbb{P}(D_1 = 8)(1 + 9\mathbb{E}[T]),$$

and since the two probabilities add to 1,

$$\mathbb{E}[T] = 1/(1 - 9\mathbb{P}(D_1 = 8)).$$

(a) Plugging in $\mathbb{P}(D_1 = 8) = 4/38$ gives $\mathbb{E}[T] = \boxed{19}$.

(b) Plugging in $\mathbb{P}(D_1 = 8) = 4/37$ gives $\mathbb{E}[T] = \boxed{37}$.

Chapter 3

19.

State whether each logical statement is true or false.

a. $(3 < 4) \wedge (7 = 7)$.

b. $(3 < 4) \wedge (7 = 8)$.

c. $(3 < 4) \vee (7 = 7)$.

d. $(3 < 4) \vee (7 = 8)$.

a. $\boxed{\text{T}}$ since both statements are true.

b. $\boxed{\text{F}}$ since not all statements are false.

c. $\boxed{\mathsf{T}}$ since at least one statement is true.

d. $\boxed{\mathsf{T}}$ since at least one statement is true.

21.

Evaluate

a. $(3 > 4) \wedge (7 > 5)$.

b. $\neg(3 > 4) \wedge \neg(7 > 5)$.

a. Here

$$\begin{aligned}(3 > 4) \wedge (7 > 5) &= \mathsf{F} \wedge \mathsf{T} \\ &= \boxed{\mathsf{F}}\end{aligned}$$

b. Here

$$\begin{aligned}\neg(3 > 4) \wedge \neg(7 > 5) &= \neg\mathsf{F} \wedge \neg\mathsf{T} \\ &= \mathsf{T} \wedge \mathsf{F} \\ &= \boxed{\mathsf{F}}\end{aligned}$$

23.

State whether each logical statement is true or false.

a. $(\forall x \in [3, 4])(x < 5)$

b. $(\exists x \in [3, 4])(x < 5)$

c. $(\forall x \in [3, 4])(x < 3.5)$

d. $(\exists x \in [3, 4])(x < 3.5)$

Solution. a. $\boxed{\mathsf{T}}$ since every number in the closed interval from 3 up to 4 inclusive is less than 5.

b. $\boxed{\mathsf{T}}$ since the number 3 in the closed interval from 3 up to 4 inclusive is less than 5.

c. $\boxed{\mathsf{F}}$ since the number 4 is in the closed interval from 3 up to 4 but is not less than 3.5.

d. $\boxed{\mathsf{T}}$ since the number 3 is in the closed interval from 3 up to 4 and is less than 3.5.

25.

State whether each statement about sets is true or false.

a. $3 \in \{1, 2, 3, 4\}$.

b. $5 \in \{1, 2, 3, 4\}$

c. $\{1,2\} \subseteq \{1,2,3,4\}$.

d. $\{1,5\} \subseteq \{1,2,3,4\}$.

Solution. a. $\boxed{\text{T}}$

b. $\boxed{\text{F}}$

c. $\boxed{\text{T}}$

d. $\boxed{\text{F}}$

27. If $\mathbb{I}(p) = \mathbb{I}(q) = 1$, what is $\mathbb{I}(p \wedge q)$?

Solution. From the rule,

$$\mathbb{I}(p \wedge q) = \mathbb{I}(p)\mathbb{I}(q) = (1)(1) = \boxed{1}.$$

Chapter 4

29.

Suppose that W takes on values in $1, 2, 3, \ldots$, and

$$\mathbb{P}(W = i) = (3/4)(1/4)^i$$

for $i \in \{1, 2, \ldots\}$.

a. What is $\mathbb{P}(W = 2)$?

b. What is $\mathbb{P}(W \neq 2)$?

Solution. a. Using the formula given, this is

$$\mathbb{P}(W = 2) = (3/4)(1/4)^2 = 3/64 = \boxed{0.04688}.$$

b. Using negation, this is

$$\mathbb{P}(W \neq 2) = 1 - \mathbb{P}(W = 2) = 1 - 3/64 = \boxed{0.9531}.$$

31. Suppose $\Omega = \{1,2,3,4,5\}$ and $\{1,2\}$ are sets in a σ-algebra $\mathcal{F}$. Give three other sets that must also be in $\mathcal{F}$.

All σ-algebras over Ω contain both the empty set and Ω itself, so that's two more sets. Finally, σ-algebras are closed under complements. So three sets also in $\mathcal{F}$ are

$$\boxed{\{\{3,4,5\}, \emptyset, \{1,2,3,4,5\}\}}.$$

33.

Suppose $f(x) = 3x$.

a. What is $f^{-1}([0,4])$?

b. What is $f^{-1}([-1,4])$?

Solution. a. This would be $\boxed{[0,4/3]}$.

b. This would be $\boxed{[-1/3,4/3]}$.

35. Consider the following four intervals.

$$A_1 = [-1,1],\ A_2 = [0,4],\ A_3 = [-3,-2],\ A_4 = [-4,4].$$

Which pairs of these four sets are disjoint?

Solution. The set A_4 intersects everything, so cannot be part of a disjoint pair. A_2 intersects A_1. That leaves

$$\boxed{\{A_1, A_3\}, \{A_2, A_3\}}.$$

Chapter 5

37. Let X be a random variable such that

$$\mathbb{P}(X = 1) = 0.2,\ \mathbb{P}(X = 1.5) = 0.2,\ \mathbb{P}(X = 2) = 0.6.$$

Find $\mathbb{E}[X]$.

Solution. Since X is simple, this is

$$\mathbb{E}[X] = (0.2)(1) + (0.2)(1.5) + (0.6)(2) = \boxed{1.700}.$$

39. If X is a finite random variable, show that X^2 is also a finite random variable.

Solution. X finite means that there exists $\{x_1,\ldots,x_n\}$ such that $\mathbb{P}(X \in \{x_1,\ldots,x_n\}) = 1$. So

$$\mathbb{P}\left(X^2 \in \cup_{i=1}^n x_i^2\right) = 1.$$

Hence X^2 is also finite.

41.

Let $U \sim \text{Unif}([0,1])$ so U is a continuous random variable such that for all $0 \leq a \leq b \leq 1$,

$$\mathbb{P}(U \in [a,b]) = b - a.$$

Let

$$W = 0 \cdot \mathbb{I}(U < 1/3) + 0.3 \cdot \mathbb{I}(1/3 \leq U < 2/3) + 0.6 \cdot \mathbb{I}(U > 2/3).$$

a. Is W a finite random variable?

b. Does $W \leq U$ hold with probability 1?

c. What is $\mathbb{E}[W]$?

d. Give a lower bound for $\mathbb{E}[U]$ utilizing W.

Solution. a. $\boxed{\text{Yes.}}$, since $\mathbb{P}(W \in \{0, 0.3, 0.6\}) = 1$.

b. $\boxed{\text{Yes.}}$. Since $U \geq 0$, $W = 0 \leq U$. If $W = 0.3$ then $U > 1/3 \geq 0.3 = W$. Finally, if $W = 0.6$ then $U \geq 2/3 > 0.6 = W$. So in all cases, $W \leq U$.

c. Since W is finite, this is

$$\begin{aligned}\mathbb{E}[W] &= 0 \cdot \mathbb{P}(U < 1/3) + 0.3\mathbb{P}(U \in [1/3, 2/3]) + 0.6\mathbb{P}(2/3 < U \leq 1) \\ &= (0.3)(2/3 - 1/3) + 0.6(1 - 2/3) \\ &= \boxed{0.3000}\end{aligned}$$

d. Since $W \leq U$, $\mathbb{E}[W]$ is also a lower bound on $\mathbb{E}[U]$. Hence $\boxed{0.3000}$ is a lower bound on $\mathbb{E}[U]$.

43. Prove that $\mathbb{E}[\mathbb{I}(X \in A)] = \mathbb{P}(X \in A)$ for any A where $\mathbb{P}(X \in A)$ is defined.

Solution. If $\mathbb{P}(X \in A)$ is defined, then $\mathbb{I}(X \in A) \in \{0, 1\}$, and so is a simple random variable. Therefore,

$$\mathbb{E}[\mathbb{I}(X \in A)] = 0 \cdot \mathbb{P}(X \notin A) + 1 \cdot \mathbb{P}(X \in A) = \mathbb{P}(X \in A)$$

as desired. □

45. Let $X \sim \text{Exp}(1)$ be a standard exponential random variable, so

$$\mathbb{P}(X \in [a, b)) = \int_a^b \exp(-x)\, dx$$

for all $0 \leq a \leq b$.

Suppose the random variable Y is defined as

$$Y = 0 \cdot \mathbb{I}(X < 1) + 1 \cdot \mathbb{I}(1 \leq X < 2) + 2 \cdot \mathbb{I}(X \geq 2).$$

Use Y to give a lower bound on $\mathbb{E}[X]$.

Solution. Since $Y \leq X$, $\mathbb{E}[Y] \leq \mathbb{E}[X]$. And Y is simple, so

$$\mathbb{E}[Y] = 0 \cdot \mathbb{P}(X < 1) + \mathbb{P}(1 \leq X < 2) + 2\mathbb{P}(X \geq 2).$$

The probabilities can be found using integrals:

$$\mathbb{P}(1 \leq X < 2) = \int_1^2 \exp(-x)\,dx = \exp(-1) - \exp(-2),$$
$$\mathbb{P}(X \geq 2) = \int_2^\infty \exp(-x)\,dx = \exp(-2),$$

so

$$\mathbb{E}[Y] = \exp(-2) + \exp(-1) \approx \boxed{0.5032}.$$

47. Some notation: for real numbers x and y,

$$x \vee y = \max(x, y).$$

Suppose that $\mathbb{E}[X \vee 0] = 7$ and $\mathbb{E}[-X \vee 0] = 12$. What is $\mathbb{E}[X]$?

Solution. Since both the positive part of X and the negative part are finite, the mean of X is the positive part minus the negative part, or $7 - 12 = \boxed{-5}$.

49.

The *supremum* of a set of real numbers is the smallest number that is still an upper bound for the set. So for instance, the intervals $(1, 2)$ and $[1, 2]$ both have supremum 2, since that is the smallest number that is an upper bound for all elements of the set. By convention, if the set has no upper bound the supremum is ∞, and a set that is empty has supremum $-\infty$.

With that in mind, find the supremum of each of the following sets. You do not have to justify your answer.

a. $\{x : 3 < x < 10\}$.

b. $\{1, 2, 3, \ldots\}$.

c. $\emptyset$

Solution. The answers are

a. $\boxed{10}$. This is since 10 is an upper bound because $x < 10$ is part of the set, and there is no smaller upper bound.

b. $\boxed{\infty}$. Here the numbers in the set do not have an upper bound.

c. $\boxed{-\infty}$. Every number is an upper bound for the empty set, and so the supremum is infinitely small.

Chapter 6

51. Suppose $\mathcal{F}$ is a σ-algebra for sets and for all $i \in \{1, 2, 3, 4, \ldots\}$,

$$m(i) = (1/3)^i.$$

What is $m(\{1, 2, 3, \ldots\})$?

Solution. Since $\{1\}, \{2\}, \ldots$ are disjoint sets,

$$\begin{aligned} m(\{1,2,3,\ldots\}) &= m(\{1\}) + m(\{2\}) + \cdots \\ &= \sum_{i=1}^{\infty} \left(\frac{1}{3}\right)^i \\ &= \frac{1/3}{1-1/3} \\ &= 1/2 = \boxed{0.5000}. \end{aligned}$$

53.

Suppose $\mu([1,2)) = 1$, $\mu([2,3)) = 2$, and $\mu([3,4]) = 3$.

a. What is $\int_{[1,2)} 1 \, d\mu$?

b. What is $\int \mathbb{I}(x \in [1,2)) \, d\mu(x)$?

Solution. a. This is just the μ measure of $[1,2)$, which is $\boxed{1}$.

b. This is also just the μ measure of set inside the indicator, which is $\mu([1,2)) = \boxed{1}$.

55. Suppose $f(x) \leq g(x)$ for all x and are measurable functions with respect to μ. If

$$\int f(x) \, d\mu(x) = 0.47,$$

what can be said about

$$\int g(x) \, d\mu(x)?$$

Solution. Because $g(x)$ dominates $f(x)$, its integral has to be larger. That is

$$\boxed{\int g(x) \, d\mu(x) \geq 0.47}.$$

That's all that can be said with this information.

57. Suppose

$$\begin{aligned} f_1(x) &\leq g(x) \\ f_2(x) &\leq g(x) \\ f_3(x) &\leq g(x) \end{aligned}$$

are all measurable with respect to μ. Moreover,

$$\int f_1(x) \, d\mu(x) = 4.2, \int f_2(x) \, d\mu(x) = 8.3, \int f_3(x) \, d\mu(x) = -1.7,$$

What can be said about

$$\int g(x) \, d\mu(x)?$$

Solution. Because $g(x)$ dominates $f_1(x), f_2(x), f_3(x)$, its integral has to be larger than each of the three integrals. That is

$$\boxed{\int g(x)\,d\mu(x) \geq 8.3}.$$

That's all that can be said with this information.

59. Suppose that random variable W has density

$$f_W(w) = (1/2)\mathbb{I}(w \in [2,4]).$$

which respect to Lebesgue measure m.

What is $\mathbb{P}(W \geq 3)$?

Solution. Setting up the integral, this is

$$\mathbb{P}(W \geq 3) = \int_{w\in[3,\infty)} (2/4)\mathbb{I}(w \in [2,4])\,dm(w).$$

Note that the only time $\mathbb{I}(w \in [2,4]) > 0$ when $w \geq 3$ is when $w \in [3,4]$. So this integral becomes

$$\begin{aligned}
\mathbb{P}(W \geq 3) &= (1/2)\int_{w\in[3,4]} 1\,dm(w) \\
&= (1/2)m([3,4]) \\
&= \boxed{0.5000}.
\end{aligned}$$

61. Continuing with X from the last problem, what is $\mathbb{E}[X^2]$?

Solution. From the fact above, the square that is applied to the random variable X inside the mean is applied to the variable of integration inside the integral. So

$$\begin{aligned}
\mathbb{E}[X^2] &= \int s^2 2\exp(-2s)\mathbb{I}(s \geq 0)\,ds \\
&= \int_{s\geq 0} 2s^2\exp(-2s)\,ds \\
&= \boxed{0.5000}.
\end{aligned}$$

Chapter 7

63. For X a Zipf Law random variable, show that $\mathbb{E}[X^3]$ is only finite when $\alpha > 4$.

Solution. Here

$$\begin{aligned}
\mathbb{E}[X^3] &= \sum_{i=1}^{\infty} i^3 \frac{C}{i^\alpha} \\
&= \sum_{i=1}^{\infty} \frac{C}{i^{\alpha-3}},
\end{aligned}$$

which only converges when the exponent of i in the denominator is greater than 1. So $\alpha - 3 > 1$, so $\boxed{\alpha > 4}$.

65. Suppose $X_1, X_2, \ldots \to Y$ with probability 1. What can you say about the convergence in probability of the X_i to Y?

Solution. Since convergence with probability 1 implies convergence in probability, it holds that X_i converges to Y in probability.

67. Give an example of a strictly increasing sequence of random variables X_i that converge to X with probability 1.

Solution. Just set $X_i = X - 1/i$. Then the X_i are strictly increasing, and converge to X with probability 1.

69. Let $t \geq 0$. Given that $\exp(tx)$ is a convex function for $t \geq 0$, if $\mathbb{E}[\exp(tX)] < \infty$, what can you say about this value and $\mathbb{E}[X]$?

Solution. By Jensen's inequality, the convex function can be pulled out of the expectation at the cost of making the value smaller. Hence the result is

$$\exp(t\mathbb{E}[X]) \leq \mathbb{E}[\exp(tX)].$$

Chapter 8

71. Suppose that $X_1, X_2, \ldots$ are independent and $\mathbb{P}(X_i \in A_i) = 1/i$. What is

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, X_3 \in A_3)?$$

Solution. Because the X_i are independent,

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, X_3 \in A_3) = \prod_{i=1}^{3} \mathbb{P}(X_i \in A_i) = (1/1)(1/2)(1/3) = \boxed{0.1666\ldots}.$$

73. Suppose $B_1, B_2, \ldots$ are independent and $B_i \sim \textsf{Bern}((1/3)^i)$. What is

$$\mathbb{E}\left[\sum_{i=1}^{\infty} B_i\right]?$$

Solution. Because Bernoulli random variables are nonnegative, the partial sums

$$S_n = \sum_{i=1}^{n} B_i$$

are increasing, so by the monotone convergence theorem

$$\begin{aligned}\mathbb{E}\left[\lim_{n\to\infty} S_n\right] &= \lim_{n\to\infty} \mathbb{E}[S_n] \\ &= \lim_{n\to\infty} \mathbb{E}[B_1] + \mathbb{E}[B_2] + \cdots + \mathbb{E}[B_n] \\ &= \lim_{n\to\infty} (1/3) + (1/9) + \cdots + (1/3)^n \\ &= \sum_{i=1}^{\infty} (1/3)^i \\ &= \frac{1/3}{1-1/3} \\ &= \boxed{0.5000}.\end{aligned}$$

75.

Suppose U is a standard uniform over $[0,1]$.

a. Find $\mathbb{E}[1/\sqrt{U}]$.

b. Find

$$\mathbb{E}\left[\lim_{n\to\infty} \frac{n\sin(U)}{1+n^2\sqrt{U}}\right].$$

Solution. a. Using the law of the unconscious statistician,

$$\begin{aligned}\mathbb{E}[1/\sqrt{U}] &= \int (1/\sqrt{u})\mathbb{I}(u \in [0,1])\,du \\ &= \int_0^1 u^{-1/2}\,du \\ &= u^{1/2}/(1/2)|_0^1 \\ &= \boxed{2}.\end{aligned}$$

b. Note that for $U \in [0,1]$ and $n \geq 1$, $\sin(U) \leq 1$, so

$$\frac{n\sin(U)}{1+n^2/\sqrt{U}} \leq \frac{n}{n^2\sqrt{U}} = \frac{1}{n\sqrt{U}} \leq \frac{1}{\sqrt{U}}.$$

The right hand side is integrable (in part a you found it was 2), so the Dominated Convergence Theorem says that

$$\begin{aligned}\mathbb{E}\left[\lim_{n\to\infty} n\sin(U)/(1+n^2\sqrt{U}\right] &= \lim_{n\to\infty} \mathbb{E}\left[\frac{n\sin(U)}{1+n^2\sqrt{U}}\right] \\ &\leq \lim_{n\to\infty} \mathbb{E}\left[\frac{n}{1+n^2\sqrt{U}}\right] \\ &\leq \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n\sqrt{U}}\right] \\ &= \lim_{n\to\infty} 2/n \\ &= \boxed{0}.\end{aligned}$$

77. Let T be a nonnegative random variable. The notation $\wedge$ can also be used for minimum, so $t \wedge T = \min(t, T)$. Note that $t \wedge T$ is an increasing sequence in t. When $\mathbb{P}(T < \infty) = 1$, it holds that

$$\lim_{t \to \infty} t \wedge T = T$$

since eventually t will be larger than the finite value of T with probability 1.
For T such that $\mathbb{P}(T < \infty) = 1$, what can be said about

$$\lim_{t \to \infty} \mathbb{E}[t \wedge T]?$$

Solution. Since $t \wedge T$ is increasing in T, the MCT says that

$$\lim_{t \to \infty} \mathbb{E}[t \wedge T] = \mathbb{E}[\lim_{t \to \infty} t \wedge T] = \mathbb{E}[T].$$

Chapter 9

79.

Consider the sequence $0, 1, 0, 1, 0, 1, \ldots$.

a. What is the limit inferior of the sequence?

b. What is the limit superior of the sequence?

Solution. a. If you skip the first n entries in the sequence, the smallest value is 0 and there will always be another 0. So

$$\liminf\{0, 1, 0, 1, \ldots\} = \boxed{0}.$$

b. If you skip the first n entries in the sequence, the largest value is 1 and there will always be another 1. So

$$\limsup\{0, 1, 0, 1, \ldots\} = \boxed{1}.$$

81.

Suppose that $X \sim \text{Geo}(1/2)$.

a. For $m \in \{1, 2, \ldots\}$, find

$$\mathbb{P}(X \wedge m = m)?$$

b. What is $\lim_{m \to \infty} \mathbb{E}[X \wedge m]$?

Solution. a. $X \wedge m$ rounds down to m if X is larger than m. So

$$\mathbb{P}(X \wedge m = m) = \mathbb{P}(X \geq m) = \boxed{(1/2)^{m-1}}.$$

b. From our last theorem, this is $\mathbb{E}[X] = \boxed{2}$.

83. Suppose that you are given random variables $R_1, R_2, \ldots$ such that $\liminf R_i = X$ where $X \sim$ Unif$([0,4])$. Give a lower bound on

$$\liminf \mathbb{E}(R_i).$$

Solution. By Fatou's Lemma,

$$\liminf \mathbb{E}(R_i) \geq \mathbb{E}(\liminf R_i) = \mathbb{E}[X] = \boxed{2}.$$

85. Suppose the S_i are random variables where $|S_i| \leq 10$ with probability 1, and $\lim S_i \sim A$ where $A \sim$ Exp(4). What can be said about $\lim \mathbb{E}(S_i)$?

Solution. By the Bounded Convergence Theorem, the limit can be brought inside the expectation to say

$$\lim \mathbb{E}(S_i) = \mathbb{E}(\lim S_i) = \mathbb{E}(A) = 1/4 = \boxed{0.2500}.$$

Chapter 10

87. Suppose $\mathbb{E}(X|Y) = 3Y$ and $\mathbb{E}(W|Y) = -4Y$. What is $\mathbb{E}(2X + 5W|Y)$?

Solution. By linearity of conditional expectation, this is

$$2\mathbb{E}(X|Y) + 5\mathbb{E}(W|Y) = (2)(3Y) + (5)(-4Y) = 6Y - 20Y = \boxed{-14Y}.$$

89. Suppose $\mathbb{E}(W) = 4$, $\mathbb{E}(W|S) = 3/S$, and $\mathbb{E}(S|W) = 3W$. What is $\mathbb{E}(S)$?

Solution. By the Fundamental Theorem of Probability,

$$\mathbb{E}(S) = \mathbb{E}(\mathbb{E}(S|W)) = \mathbb{E}(3W) = 3\mathbb{E}(W) = \boxed{12}.$$

91.

Suppose that $\{M_i\}$ form a martingale.

a. What is $\mathbb{E}[M_5 \mid M_4]$?

b. What is $\mathbb{E}[M_5 \mid M_3, M_1]$?

c. What is $\mathbb{E}[M_5 \mid M_0]$?

Solution. With a martingale, the mean of a future state will be the latest state that we have information for. So

a. $\boxed{M_4}$.

b. $\boxed{M_3}$.

c. $\boxed{M_0}$.

93. Let $\{R_i\}$ be a martingale with $R_0 = 10$. What is $\mathbb{E}[R_{15}]$?

Solution. This is
$$\mathbb{E}[R_{15}] = \mathbb{E}[\mathbb{E}[R_{15} \mid R_0]] = \mathbb{E}[R_0] = \boxed{10}.$$

Chapter 11

95. Consider random variable X in space $\Omega = \{1, 2, 3\}$ with
$$\sigma(X) = \{\emptyset, \{1, 2\}, \{3\}, \{1, 2, 3\}\}.$$

True or false: X is measurable with respect to
$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}?$$

Solution. This is true, since $\sigma(X) \subseteq \mathcal{F}$.

97. For $h(x) = x^2 + 1$, what is $h^{-1}([0, 4])$?

Solution. Since $x^2 + 1 \geq 0 + 1 = 1$, only the $h^{-1}([1, 4])$ contributes. To get $x^2 + 1 \leq 4$ requires $x^2 \leq 3$ which happens when $x \in \boxed{[-\sqrt{3}, \sqrt{3}]}$.

99. Let $\mathcal{P}(A)$ denote the *power set*, the set of all subsets of the set A. Then for $\mathcal{F}_i = \mathcal{P}(\{1, 2, \dots, i\})$, does $\mathcal{F}_i$ form a filtration?

Solution. $\boxed{\text{Yes}}$, this does form a filtration. For $A \subseteq B$, it holds that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$. Therefore, for all positive integers i it holds that $\mathcal{P}(\{1, \dots, i\}) \subseteq \mathcal{P}(\{1, \dots, i + 1\})$, which means $\mathcal{F}_i \subseteq \mathcal{F}_{i+1}$. That is the definition of a filtration.

101. If $A \in \mathcal{F}_7$ but $A \notin \mathcal{F}_{10}$, is it possible for the $\mathcal{F}_i$ to form a filtration?

Solution. $\boxed{\text{No}}$, since any element of $\mathcal{F}_i$ will also be an element of $\mathcal{F}_j$ for $j > i$ in a filtration.

103. Say $\mathcal{F}_t$ is the adapted filtration for $X_0, X_1, \dots$. Is X_7 necessarily measurable with respect to $\mathcal{F}_{10}$?

Solution. Note that
$$\mathcal{F}_{10} = \sigma(X_0, \dots, X_{10}),$$
so X_7 $\boxed{\text{is necessarily measurable}}$ with respect to $\mathcal{F}_{10}$.

Chapter 12

105. Suppose $\mathbb{E}(X|R) = R^2$. Also, $\mathbb{P}(R^2 \in [0, 1] = 0.3)$ and $\mathbb{E}(R^2 \mathbb{I}(R^2 \in [0, 1]) = 0.1)$. What is $\mathbb{E}(X \mathbb{I}(R^2 \in [0, 1]))$?

Solution. By the definition of conditional, expectation,

$$\mathbb{E}(X\mathbb{I}(R^2 \in [0,1])) = \mathbb{E}(R^2\mathbb{I}(R^2 \in [0,1])) = \boxed{0.1000}.$$

107. Suppose $\mathbb{E}[X \mid Y] = 4.2Y$. Write $\mathbb{E}[X]$ in terms of the mean of Y.

Solution. Taking the expected value of $\mathbb{E}[X \mid Y]$ returns $\mathbb{E}[X]$. So

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X \mid Y]] \\ &= \mathbb{E}[4.2Y] \\ &= \boxed{4.2\mathbb{E}[Y]}.\end{aligned}$$

109. Suppose $U \sim \textsf{Unif}([0,1])$, and $W = \mathbb{I}(U \geq 0.2)$. Find

$$\mathbb{E}[U \mid W].$$

Prove that your answer is correct.

Solution. First, $[U \mid W = 0] \sim \textsf{Unif}([0, 0.2)$ so $\mathbb{E}[U \mid W = 0] = 0.1$.

Second, $[U \mid W = 1] \sim \textsf{Unif}([0.2, 1])$, so $\mathbb{E}[U \mid W = 1] = 0.6$.

These statements can be combined to say that

$$\boxed{\mathbb{E}[U \mid W] = 0.1 + 0.5W}.$$

Proof. To prove this is the right answer it is necessary to know the elements of $\sigma(W)$. Since W is either 0 or 1, these are

$$\begin{aligned}W^{-1}(\{\}) &= \emptyset \\ W^{-1}(\{0\}) &= [0, 0.2) \\ W^{-1}(\{1\}) &= [0.2, 1] \\ W^{-1}(\{0,1\}) &= [0,1].\end{aligned}$$

Now find the means times the indicators for U:

$$\begin{aligned}\mathbb{E}[U\mathbb{I}(U \in \emptyset)] &= 0 \\ \mathbb{E}[U\mathbb{I}(U \in [0, 0.2)] &= \int_{u \in [0,0.2)} u\, du = 0.02 \\ \mathbb{E}[U\mathbb{I}(U \in [0.2, 1])] &= \int_{u \in [0.2,1]} u\, du = 0.48 \\ \mathbb{E}[U\mathbb{I}(U \in [0, 1])] &= \int_{u \in [0,1]} u\, du = 0.5\end{aligned}$$

Now do the same with the proposed value of $\mathbb{E}[U \mid W]$.

$$\begin{aligned}
\mathbb{E}[(0.1 + 0.5W)\mathbb{1}(U \in \emptyset)] &= 0 \\
\mathbb{E}[(0.1 + 0.5W)\mathbb{1}(U \in [0, 0.2)] &= (0.2)(0.1) = 0.02 \\
\mathbb{E}[(0.1 + 0.5W)\mathbb{1}(U \in [0.2, 1])] &= (0.8)(0.6) = 0.48 \\
\mathbb{E}[(0.1 + 0.5W)\mathbb{1}(U \in [0, 1])] &= (0.2)(0.1) + (0.8)(0.6) = 0.5
\end{aligned}$$

They match, so the proof is complete! □

111. Let $U_1, U_2, U_3, \ldots$ be iid uniform over $[-1, 1]$. Set

$$M_i = \sum_{i=1}^{i} U_i.$$

Prove that $\{M_i\}$ is a martingale with respect to the adapted filtration.

Solution. First, the M_n values will be measurable with respect to the adapted filtration.

Second,

$$\begin{aligned}
|M_n| &= \left|\sum_{i=1}^{n} U_i\right| \\
&\leq \left|\sum_{i=1}^{n} |U_i|\right| \\
&\leq n,
\end{aligned}$$

so $\mathbb{E}[|M_n|] \leq n < \infty$.

Third,

$$\begin{aligned}
\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \quad & \quad = \mathbb{E}[M_n + U_{n+1} \mid \mathcal{F}_n] \\
&= M_n + \mathbb{E}[U_{n+1}] \\
&= M_n.
\end{aligned}$$

So the three properties of a martingale are satisfied!

Chapter 13

113. For a sequence $a_i = (-1)^i$, find

$$\{i \in \{0, 1, 2, 3, \ldots\} : a_i = 1\}.$$

Solution. The value of $(-1)^i = 1$ when i is an even number. Hence this is

$$\boxed{\{0, 2, 4, 6, \ldots\}}.$$

115. Suppose a player is playing a fair game where they either win or lose a dollar on each (independent) play with probability $1/2$. If they start with 4 dollars, and quit when they hit 0 or 16 dollars, what is the chance that the player walks away with 16 dollars?

Solution. Let M_t denote the money the player has after t plays of the game. Set

$$T = \inf\{t : M_t \in \{0, 16\}\}.$$

Then $\mathbb{P}(T < \infty) = 1$, since the chance that there are 16 wins in a row is $(1/2)^{16} > 0$, making $\mathbb{P}(T > 16k) \leq [1 - (1/2)^{16}]^k$ for all k. Note that $(T = \infty) \rightarrow (T > 16k)$ for all k. Therefore

$$\mathbb{P}(T = \infty) \leq \inf_{k \in \{1,2,3,\ldots\}} [1 - (1/2)^{16}]^k = 0.$$

Then $M_{t \wedge T}$ is a martingale so for all t,

$$\mathbb{E}[M_{t \wedge T}] = \mathbb{E}[M_0]$$

and

$$\lim_{t \to \infty} \mathbb{E}[M_{t \wedge T}] = \mathbb{E}[M_0].$$

Since $|M_{t \wedge T}| \leq 16$, the limit can be brought inside the mean by the DCT, and using $\mathbb{P}(T < \infty) = 1$ gives

$$\mathbb{E}[\lim_{t \to \infty} M_{t \wedge T}] = \mathbb{E}[M_T] = \mathbb{E}[M_0].$$

Finally, $M_T \in \{0, 16\}$, so

$$\mathbb{E}[M_T] = 16\mathbb{P}(M_T = 16) + 0\mathbb{P}(M_T = 0) = 4,$$

yielding $\mathbb{P}(M_T = 16) = \boxed{0.2500}$.

117.

Consider a stochastic process where each $X_t \in \{0, 1, 2\}$, with outcome

$$X_1, X_2, \ldots = 1, 1, 2, 1, 2, 2, 2, 2, 1, 0, 1, 2, 0, 0, 2, 1, \ldots$$

a. Let $T_1 = \inf\{t : X_t = 0\}$. What is T_1 for the outcome given above?

b. Let $T_2 = \inf\{t : X_t = 3\}$. What is T_2 for the outcome given above?

Solution.

a. Here the first time $X_t = 0$ is at $t = 10$, so $\boxed{T_1 = 10}$ here.

b. Since there are no 3's in the sequence, $\boxed{T_2 = \infty}$.

119. Suppose $M_0, M_1, M_2, \ldots = 3, 5, 2, 5, 3, 4, 1, 3, 4, -4, 5, \ldots$, and $T = 3$. What does the first eleven terms of the sequence $M_{t \wedge T}$ look like?

Solution. This sequence is the same as M_t up to $t = 3$, then it stays at the value M_3 after that. So

$$\boxed{M_{t \wedge T} = 3, 5, 2, 5, 5, 5, 5, 5, 5, 5, 5, \ldots}.$$

121. Consider two players playing a fair game where one player gives the other player $1 with probability $1/2$. Each play of the game is independent of the previous games. Let M_t be the amount of money owned by the first player after t steps in the game.

Then M_t forms a martingale with respect to the natural filtration. Let $T = \inf\{t : M_t \in \{0, 20\}\}$. If $\mathbb{P}(M_0 = 3) = 1$, prove that $\mathbb{P}(T < \infty) = 1$.

Solution. Every 20 moves, there is a $(1/2)^{20}$ chance that the next 20 games are won by the first player, which would result in $M_t = 20$ sometime during those 20 wins. The next 20 games are independent of the first, and so on, so after $20k$ games (for k a positive integer), the chance of *not* reaching state 20 is at most

$$(1 - (1/2)^{20})^k.$$

Therefore, as $k \to \infty$, $\mathbb{P}(T > 20k) \to 0$, which is equivalent to saying $\mathbb{P}(T < \infty) = 1$.

Chapter 14

123. Suppose that I play an unfair game where I have a 60% chance of winning $1 and a 40% of losing $1. I start with $10, and quit when I reach $20 or $0. What is the probability that when I quit I have $0?

Solution. Let M_t be the amount of money that I have at time t, and let

$$N_t = \left(\frac{40}{60}\right)^{M_t - M_0}.$$

Since N_t is a function of M_t and M_0, it is measurable with respect to the natural filtration F_t.

Next,

$$\mathbb{E}|N_t| \leq (60/40)^t < \infty.$$

and

$$\begin{aligned}\mathbb{E}[N_{t+1}|F_t] &= (0.6)(40/60)^{M_t+1-M_0} + (0.4)(40/60)^{M_t-1-M_0} \\ &= (0.4)(40/60)^{M_t-M_0} + (0.6)(40/60)^{M_t-M_0} \\ &= (40/60)^{M_t-M_0} = N_t.\end{aligned}$$

Hence N_t is a martingale. Let

$$T = \inf\{t : N_t \in \{(40/60)^{-10}, (40/60)^{10}\}\}.$$

Then $N_{t\wedge T}$ is also a martingale, and

$$\lim_{t\to\infty} \mathbb{E}[N_{t\wedge T}] = \mathbb{E}[N_0] = 1.$$

Since winning 20 times in a row results in stopping,

$$\mathbb{P}(T > 20k) \leq (1 - (0.6)^{20})^k,$$

and taking the limit of both sides as k goes to infinity gives $\mathbb{P}(T = \infty) = 0$.

Since $N_{t\wedge T} \in \{(40/60)^{10}, \ldots, (40/60)^{-10}\}$ is bounded, the limit can be brought inside to give

$$\mathbb{E}[N_T] = 1.$$

That means

$$\begin{aligned} 1 &= \mathbb{E}[N_T] \\ &= (40/60)^{10}(1 - \mathbb{P}(M_T = 0)) + (40/60)^{-10}\mathbb{P}(M_T = 0) \\ \mathbb{P}(M_T = 0) &= \frac{1 - (40/60)^{10}}{(40/60)^{-10} - (40/60)^{10}} \\ &\approx \boxed{0.01704} \end{aligned}$$

125. Suppose that I play an unfair game where I have a 60% chance of winning \$1 and a 40% of losing \$1. I start with \$10, and quit when I reach \$20 or \$0. What is the expected number of steps until I reach \$0 or \$20?

Solution. Let M_t be the amount of money that I have at time t, and let

$$R_t = M_t - 0.2t$$

Since R_t is a function of M_t, it is measurable with respect to the natural filtration F_t.

Also,

$$|R_t| \leq |M_t| + |-0.2t| \leq 10 + t + 0.2t,$$

so $\mathbb{E}|R_t| \leq 10 + 1.2t < \infty$.

Next,

$$\begin{aligned} \mathbb{E}[R_{t+1}|F_t] &= \mathbb{E}[M_{t+1} - 0.2(t+1)|F_t] \\ &= -0.2(t+1) + (0.6)(M_t + 1) + (0.4)(M_t - 1) \\ &= -0.2t - 0.2 + M_t + 0.6 - 0.4 \\ &= M_t - 0.2t = R_t. \end{aligned}$$

Hence R_t is a martingale. Let

$$T = \inf\{t : M_t \in \{0, 20\}\}.$$

Then $R_{t\wedge T}$ is also a martingale, and

$$\mathbb{E}[R_{t\wedge T}] = \mathbb{E}[R_0] = 10.$$

On the other hand,

$$\mathbb{E}[R_{t\wedge T}] = \mathbb{E}[M_{t\wedge T}] - 0.2\mathbb{E}[t \wedge T]$$

The first term is bounded and since $\mathbb{P}(T < \infty) = 1$, you can bring $\lim_{t\to\infty}$ inside the mean by the DCT. You can also bring $\lim_{t\to\infty}$ inside the second term using the MCT. That gives

$$\mathbb{E}[M_T] - 0.2\mathbb{E}[T] = 10,$$

so $\mathbb{E}[T] = (\mathbb{E}[M_T] - 10)/0.2$.

From the last problem,

$$\mathbb{E}[M_T] = 20\left(1 - \frac{1-(40/60)^{10}}{(40/60)^{-10} - (40/60)^{10}}\right),$$

which gives $\mathbb{E}[T] \approx \boxed{48.29}$.

127. Suppose M_t is a process where $M_0 = 0$ and

$$\mathbb{P}(M_{t+1} = M_t + 1|\mathcal{F}_t) = 0.30$$
$$\mathbb{P}(M_{t+1} = M_t|\mathcal{F}_t) = 0.10$$
$$\mathbb{P}(M_{t+1} = M_t - 1|\mathcal{F}_t) = 0.60$$

Find a value $r \neq 1$ such that $N_t = r^{M_t}$ is a martingale with respect to the natural filtration generated by M_t.

Solution. Certainly $N_t = r^{M_t}$ is measurable with respect to M_t and hence the natural filtration. Also, $|M_t| \leq t$ which means that $|N_t| \leq \max(r^t, (1/r)^t)$. That means the N_t are bounded, and so are integrable.

The key then is

$$\begin{aligned}\mathbb{E}[N_t|\mathcal{F}_t] &= \mathbb{E}[r^{M_t}|\mathcal{F}_t] \\ &= 0.3r^{M_t+1} + 0.1r^{M_t} + 0.6r^{M_t-1} \\ &= r^{M_t}[0.3r^1 + 0.1 + 0.6r^{-1}].\end{aligned}$$

Since $r^{M_t} = N_t$, it is necessary to find r such that $0.3r + 0.1 + 0.6/r = 1$. This is equivalent to

$$0.3r^2 - 0.9r + 0.6 = 0.$$

One of the solutions is $r = 1$ because this makes N_t identically 1 and a martingale. However, this is unhelpful when it comes to solving problems.

It does give us a way to reduce this to a linear problem by dividing by the factor $(r-1)$ to give

$$0.3r - 0.6 = 0$$

which solves to give $\boxed{r = 2}$.

129. Suppose I play an unfair game where I have a 30% chance of winning \$1, a 20% chance of losing \$1, and a 50% chance of staying at my current dollar amount. If I start with \$10, and quit when I reach \$20 or \$0, what is the chance that I quit when I have \$0?

Solution. Note

$$\mathbb{E}[r^{M_t+1}|\mathcal{F}_t] = r^{M_t}[0.3r + 0.2r^{-1} + 0.5].$$

To make the right hand side 1, set $r = 20/30$, the ratio of the win chance to the lose chance.

Let M_t be the amount of money that I have at time t, and let

$$N_t = \left(\frac{20}{30}\right)^{M_t - M_0}.$$

Since N_t is a function of M_t and M_0, it is measurable with respect to the natural filtration F_t.

Next,

$$\mathbb{E}|N_t| \leq (30/20)^t < \infty.$$

and

$$\begin{aligned}
\mathbb{E}[N_{t+1}|F_t] &= (0.3)(20/30)^{M_t+1-M_0} + (0.5)(20/30)^{M_t-M_0} + (0.2)(20/30)^{M_t-1-M_0} \\
&= (0.2)(20/30)^{M_t-M_0} + (0.5)(20/30)^{M_t-M_0} + (0.3)(20/30)^{M_t-M_0} \\
&= (20/30)^{M_t-M_0} = N_t.
\end{aligned}$$

Hence N_t is a martingale. Let

$$T = \inf\{t : N_t \in \{(20/30)^{-10}, (20/30)^{10}\}\}.$$

Then $N_{t\wedge T}$ is also a martingale, and

$$\lim_{t\to\infty} \mathbb{E}[N_{t\wedge T}] = \mathbb{E}[N_0] = 1.$$

Since winning 20 times in a row results in stopping,

$$\mathbb{P}(T > 20k) \leq (1 - (0.2)^{20})^k,$$

and taking the limit of both sides as k goes to infinity gives $\mathbb{P}(T = \infty) = 0$.

Since $N_{t\wedge T} \in \{(20/30)^{10}, \ldots, (20/30)^{-10}\}$ is bounded, the limit can be brought inside to give

$$\mathbb{E}[N_T] = 1.$$

That means

$$\begin{aligned}
1 &= \mathbb{E}[N_T] \\
&= (20/30)^{10}(1 - \mathbb{P}(M_T = 0)) + (20/30)^{-10}\mathbb{P}(M_T = 0) \\
\mathbb{P}(M_T = 0) &= \frac{1 - (20/30)^{10}}{(20/30)^{-10} - (20/30)^{10}} \\
&\approx \boxed{0.01704}
\end{aligned}$$

Chapter 15

131. Let $G \sim \text{Geo}(1/2)$. What is $\lim_{B\to\infty} \mathbb{E}(G\mathbb{1}(G > B))$?

Solution. Since G is integrable, this is $\boxed{0}$.

133. Let $U \sim \text{Unif}([0,1])$, and $W_n = \sqrt{n} \cdot \mathbb{I}(U \leq 1/n)$. Show that the set of $\{W_n\}$ is uniformly integrable.

Solution. You can show this directly using the definition, or by swapping limits and means.

First show it directly from the definition.

Proof. Let $Z_{B,n} = |W_n|\mathbb{I}(|W_n| > B)$. Then $Z_{B,n} = 0$ if $B > \sqrt{n}$, and $Z_{B,n} = W_n$ if $\sqrt{n} \geq B$. Now $\mathbb{E}[W_n] = \sqrt{n}(1/n) = 1/\sqrt{n}$. So $\mathbb{E}[Z_{B,n}] = (1/\sqrt{n})\mathbb{I}(\sqrt{n} \geq B)$.

As a function of n, $Z_{B,n} = 0$ when $n < B^2$, and is a decreasing function for $n \geq B^2$. Hence $\sup_n \mathbb{E}[Z_{B,n}] = 1/\sqrt{B^2} = 1/B$. Now $\lim_{B\to\infty} 1/B = 0$, so the $\{W_n\}$ are u.i. □

The second proof uses swapping means and limits. Since the random variables are nonnegative, it is uniformly integrable if and only if you can swap the mean and limit as $n \to \infty$.

Proof. Note that $\lim_{n\to\infty} W_n = 0$ with probability 1. So $\mathbb{E}[\lim_{n\to\infty} W_n] = \mathbb{E}[0] = 0$.

Also, $\mathbb{E}[W_n] = \sqrt{n}(1/n) = 1/\sqrt{n}$. So $\lim_{n\to\infty} \mathbb{E}[W_n] = \lim_{n\to\infty} 1/\sqrt{n} = 0$. Hence you can swap the limit and mean, and so the $\{W_n\}$ must be uniformly integrable. □

135. Let $D_1, D_2, \ldots$ be iid $\text{Unif}(\{-1,1\})$, and $X_t = \sum_{i=1}^{t} D_i$. Then X_t is a martingale, and $T = \inf\{t : X_t = 1\}$ is a stopping time with respect to the natural filtration.

If you are given the fact that $\mathbb{P}(T < \infty) = 1$, use this to show that $\{X_{t\wedge T}\}$ are not uniformly integrable.

Solution. *Proof.* Since $\mathbb{P}(T < \infty) = 1$, $\lim_{t\to\infty} X_{t\wedge T} = X_T$ wp 1. Since X_t is a martingale, $X_{t\wedge T}$ is a martingale as well. Hence $\mathbb{E}[X_{t\wedge T}] = \mathbb{E}[X_0] = 0$ for all t. Hence

$$\lim_{t\to\infty} \mathbb{E}[X_{t\wedge T}] = 0.$$

However,

$$\mathbb{E}\left[\lim_{t\to\infty} X_{t\wedge T}\right] = \mathbb{E}[X_T] = 1.$$

Hence $\lim \mathbb{E}[X_{t\wedge T}] \neq \mathbb{E}[\lim X_{t\wedge T}]$, so $\{X_{t\wedge T}\}$ cannot be uniformly integrable. □

137.

Let $D_1, D_2, \ldots$ be iid $\text{Unif}(\{-1,1\})$ and consider the martingale

$$M_n = \sum_{i=1}^{n} D_i.$$

For a fixed positive integer a, let

$$T = \inf\{n : M_n \in \{-a, a\}\}$$

a. Use the Martingale Convergence Theorem to show that $\mathbb{P}(T < \infty) = 1$.

b. Find $\mathbb{P}(M_T = a)$.

Solution. a. Here is the proof.

Proof. The stopped martingale process $M_{t \wedge T}$ lies in $[-a, a]$, and so is bounded, hence uniformly integrable. Therefore the Martignale Convergence Theorem implies that $\lim_{n \to \infty} M_n$ exists.

However, if $T = \infty$, then the stopped martingale is changing by 1 at each step, and does not converge. Because it does converge with probability 1, $\mathbb{P}(T < \infty) = 1$. □

b. Since $\mathbb{P}(T < \infty) = 1$, that means $M_\infty = \lim_{t \to \infty} M_{t \wedge T} = M_T$. So

$$\mathbb{E}[M_T | M_0 = 0] = 0.$$

Again since $\mathbb{P}(T < \infty) = 1$, $M_T \in \{-a, a\}$, hence

$$\mathbb{E}[M_T | M_0 = 0] = \mathbb{P}(M_T = a)a + (1 - \mathbb{P}(M_T = a))(-a) = 0,$$

and solving for $\mathbb{P}(M_T = a)$ gives

$$\mathbb{P}(M_T = a) = \frac{a}{2a} = \frac{1}{2} = \boxed{0.5000}$$

regardless of a.

Chapter 16

139. If $T_{10} = 3$ and $T_{15} = 7$, call $[3, 7]$ an ______________ of $[10, 15]$.

Solution. This is what is means to be an $\boxed{\text{upcrossing}}$

140. Suppose that the B_tUnif$(\{-2, -1, 1, 2\})$ are iid. Show that

$$W_t = \sum_{i=0}^{t} B_i$$

is not uniformly integrable.

Solution. First show that W_t is a martingale with respect to the natural filtration $\mathcal{F}_t$.

By the definition of the natural filtration, W_t is $\mathcal{F}_t$ measurable for all t.

By the triangle inequality,

$$|W_t| \leq \sum_{i=0}^{\infty} |B_t| \leq 2t,$$

so $\mathbb{E}[|W_t|] \leq 2t$ and all the W_t are integrable.

Finally,

$$\mathbb{E}[W_{t+1} | \mathcal{F}_t] = \mathbb{E}[W_t + B_{t+1} | \mathcal{F}_t] = W_t + 0,$$

since B_{t+1} is independent of $\mathcal{F}_t$ and has mean 0.

Hence W_t is a martingale. Moreover, $|W_{t+1} - W_t| \geq 1$, so the W_t does not converge. Hence by the martingale convergence theorem, it cannot be uniformly integrable.

141.

Suppose D is a random variable that is 3 with probability $1/4$, and -1 with probability $3/4$.

a) What is $\mathbb{E}[D]$?

b) Let $D_1, D_2, \ldots$ be iid with the same distribution as D. Then

$$M_n = \sum_{i=1}^{n} D_i$$

is a martingale. Prove that it is not uniformly integrable.

Solution. a) This is

$$3(1/4) + (-1)(3/4) = \boxed{0}.$$

b) Since $|M_{n+1} - M_n| \geq 1$, the sequence can not converge to a single value. Hence it cannot be uniformly integrable.

143. Suppose that $R \in [0,1]$ with probability 1 and that $\mathbb{E}[R] = 0.3$. Then for $R_1, R_2, \ldots$ iid with the same distribution as R,

$$M_n = \sum_{i=1}^{n} (R_i - 0.3)/2^i$$

is a martingale. Show directly that no matter what the values of R_i are, that $\lim_{n\to\infty} M_n$ exists.

Solution. Note that

$$|(R_i - 0.3)/2^i| \leq 0.7/2^i,$$

and

$$\sum_{i=1}^{\infty} 0.7/2^i$$

exists. So by the comparison test,

$$\sum_{i=1}^{\infty} (R_i - 0.3)/2^i$$

converges no matter what the values of the R_i are.

145.

Consider Polya's Urn, and suppose that it starts with 2 red and 1 blue marble.

a) After one step, what is the distribution of the number of blue marbles?

b) After two steps, what is the distribution of the number of blue marbles?

Solution. a) After one step, the number of blue marbles can go $1 \to 1$ or $1 \to 2$. The first happens when a red marble is chosen with probability $2/3$ and the second happens when a blue marble is chosen at the step with probability $1/3$. So if N_1 is the number of blue marbles after 1 step,

$$\boxed{\mathbb{P}(N = 1) = 1/3, \mathbb{P}(N = 2) = 2/3}.$$

b) After two steps, the number of marbles can go $1 \to 2 \to 3$, $1 \to 1 \to 2$, $1 \to 2 \to 2$, or $1 \to 1 \to 1$. Taking these possibilities and adding gives

$$\mathbb{P}(N = 1) = 6/12, \; prob(N = 2) = (2 + 2)/12, \; \mathbb{P}(N = 3) = 2/12,$$

or simplifying

$$\mathbb{P}(N = 1) = 1/2, \; prob(N = 2) = 1/3, \; \mathbb{P}(N = 3) = 1/6,$$

Chapter 17

147. Suppose that M_t is a martingale where $M_0 = 0$ and with stopping time T such that $M_{t \wedge T}$ is uniformly integrable. What can be said about $\mathbb{E}(M_T)$?

Solution. Since the stopped process is uniformly integrable, the Optional Sampling Theorem says that

$$\mathbb{E}(M_T) = \mathbb{E}(\mathbb{E}(M_T | M_0)) = \mathbb{E}(M_0) = \boxed{0}.$$

149. If M_t is a martingale with stopping time T, and $M_{t \wedge T}$ is uniformly integrable, must it be true that $T < \infty$ with probability 1?

Solution. $\boxed{No}$. The random variable M_T exists by the Optional Sampling Theorem, but if $T = \infty$, then $M_T = M_\infty$, the random variable that the martingale converges to.

151. Suppose W_t is a martingale with stopping time S. Say $W_0 = 0$ and $|W_{t \wedge S}| \le 40$. What is $\mathbb{E}[W_S]$?

Solution. Since $|W_{t \wedge S}| \le 40$, the stopped process is bounded and hence is uniformly integrable. Therefore the Optional Sampling Theorem applies to give

$$\mathbb{E}[W_S] = \mathbb{E}[W_0] = \boxed{0}.$$

153. A gambler is trying to develop a betting scheme B_t based on a game that wins B_t with probability 0.4 and loses B_t with probability 0.6 at each play of a game. Is there any betting scheme where $1 \le B_t \le 1000$ at each step where the player quits at the first time T where $B_t > M_t$ or $M_t \ge 1000$ such that $\mathbb{E}[M_T | M_0] > M_0$?

Solution. $\boxed{\text{No!}}$

Let $D_1, D_2, \ldots$ be iid where $\mathbb{P}(D_i = 1) = 0.4$ and $\mathbb{P}(D_i = -1) = 0.6$. The amount of money the gambler has will be

$$M_0 + \sum_{i=1}^{t} D_i B_i.$$

The stopping time is $T = \inf\{t : B_t > M_t \vee M_t \geq 1000\}$.

First show that M_t is a supermartingale with respect to the natural filtration.

Since

$$|M_t| \leq |M_0| + \sum_{i=1}^{t} |D_i||B_i| \leq |M_0| + 1000t < \infty,$$

each M_t is bounded and so is integrable.

Finally,

$$\begin{aligned}
\mathbb{E}[M_{t+1}|\mathcal{F}_t] &= \mathbb{E}\left[M_t + D_{t+1}B_{t+1}|\mathcal{F}_t\right] \\
&= M_t + B_{t+1}\mathbb{E}[D_{t+1}] \\
&= M_t + B_{t+1}(0.4 - 0.6) \\
&\leq M_t
\end{aligned}$$

so M_{t+1} is a supermartingale, and so is $M_{t\wedge T}$.

For $t < T$, $|M_t| \leq 1000$, and each bet is at most 1000, so $|M_{t\wedge T}| \leq 2000$ for all t. Hence the stopped process is uniformly integrable.

Therefore the Optional Sampling Theorem says that

$$\mathbb{E}[M_T|M_0] \leq M_0.$$

Chapter 18

155. If X_t is a Markov chain, what can you say about the values of

$$\mathbb{P}(X_2 = c|X_0 = a, X_1 = b)$$

and

$$\mathbb{P}(X_2 = c|X_0 = d, X_1 = b)$$

Solution. Because the distribution of X_2 only depends on X_1 and not on X_0, these two values will be $\boxed{\text{equal}}$.

157. If Y_t is a time-homogeneous Markov chain, what can you say about the values

$$\mathbb{P}(Y_1 = d|Y_0 = a) \text{ and } \mathbb{P}(Y_7 = d|Y_6 = a)?$$

Solution. Because of time-homogeneity, these both represent the probability of a particular step in the Markov chain, and will be $\boxed{\text{equal}}$.

159. For $\{X_t\}$ a time-homogeneous Markov chain, what can you say about $\mathbb{P}(X_5 = a|X_3 = b)$ and $\mathbb{P}(X_{12} = a|X_{10} = b)$?

Solution. Since $5 - 3 = 12 - 10 = 2$, in either case we are taking two steps in the Markov chain. Since the probabilities for the chain are the same across time, these probabilities will be the $\boxed{\text{same}}$.

Chapter 19

161.

Suppose a Markov chain with states $\{a, b, c\}$ has transition matrix

$$A = \begin{pmatrix} 0.3 & 0 & 0.7 \\ 0 & 0.5 & 0.5 \\ 0.1 & 0.9 & 0 \end{pmatrix}$$

a. What is $\mathbb{P}(X_1 = c|X_0 = b)$?

b. What is $\mathbb{P}(X_7 = a|X_6 = a)$?

Solution. a. This is the entry for the second row (corresponding to b) and third column (corresponding to c). This is $\boxed{0.5000}$.

b. Moving one step from time 6 to 7 still uses the transition matrix, so this is just the (a, a) entry which is $\boxed{0.3000}$.

163. Suppose that there is a sequence D_t that are iid draws from the uniform distribution over $\{-1, 1\}$. Prove that

$$X_t = \sum_{i=1}^{t} D_t$$

is a Markov chain formed from the filtration $\mathcal{F}_t = \sigma(D_1, \dots, D_t)$.

Solution. Let t be a nonnegative integer. Then $X_0 = 0$, and X_t is a function of $D_1, \dots, D_t$, so is $\sigma(D_1, \dots, D_t)$ measurable.

Also, $X_{t+1} = X_t + D_{t+1}$, so

$$[X_{t+1}|\mathcal{F}_t] \sim \textsf{Unif}(\{X_t - 1, X_t + 1\})$$

which is the same as $[X_{t+1}|X_t]$.

An alternate way to show the second property is to note that for update function

$$f(x, d) = x + d,$$

then $[X_{t+1}|\mathcal{F}_t] \sim f(X_t, D_{t+1})$ where D_{t+1} is independent of $\mathcal{F}_t$. This makes X_t a Markov chain.

165.

Suppose that there is a sequence D_t that are iid draws from the uniform distribution over $\{-1, 1\}$ and

$$X_t = \sum_{i=1}^{t} D_t$$

a. Let $T = \inf\{t : X_t = 2\}$. Prove that $M_t = X_{t\wedge T}$ is a Markov chain with respect to the natural filtration.

b. Prove that $M_t = X_{t\wedge(T+1)}$ is *not* a Markov chain.

Solution. a. Since the natural filtration is being used, M_t is $\mathcal{F}_t$ measurable. Moreover,

$$[M_{t+1}|\mathcal{F}_t]$$

is equal to $\{2\}$ with probability 1 if $M_t = 2$, and uniform over $\{M_t - 1, M_t + 1\}$ if $M_t \neq 2$. Hence the distribution is the same as $[M_{t+1}|M_t]$.

This could also be shown to be a Markov chain by noting that for

$$f(x, d) = x + d\mathbb{1}(x \neq 2)$$

then $[X_{t+1}|\mathcal{F}_t] \sim f(X_t, D_{t+1})$.

b. The M_t process stops at the value one time step past where the stopping time occurs. So example sequences would be $0, 1, 2, 1, 1, 1, 1, \ldots$ or $0, 1, 0, 1, 2, 3, 3, 3, 3, 3, \ldots$ This means

$$\mathbb{P}(M_4 = 1|M_0 = 0, M_1 = 1, M_2 = 2, M_3 = 1) = 1$$

while

$$\mathbb{P}(M_4 = 2|M_0 = 0, M_1 = -1, M_2 = 0, M_3 = 1) = 1/2,$$

so $[M_4|\mathcal{F}_3] \nsim [M_4|M_3]$.

167. If $\mathbb{P}(X_{t+1} = i|X_t = i) = 1$, call the state i of the Markov chain *absorbing*.

Given transition matrix

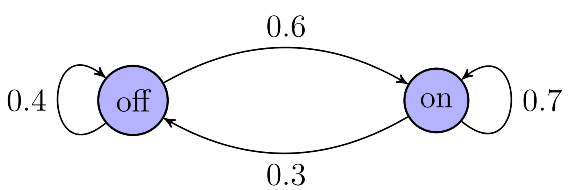
$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0.1 & 0.9 \end{pmatrix},$$

for state space $\{a, b, c, d\}$, which states are absorbing states?

Solution. The absorbing states are $\boxed{\{a, b\}}$

Chapter 20

169. This Markov chain models a telecommunications circuit that is either on or off.



If the state at time 0 is off, what is the chance that in three steps the state is on?

Solution. Putting columns and rows in order (off, on), the transition matrix is

$$A = \begin{pmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{pmatrix}$$

This can be put into R with

```
A <- matrix(c(0.4, 0.6,
              0.3, 0.7),
              byrow = TRUE,
              nrow = 2)
```

Raising it to the 3rd power gives:

```
A %*% A %*% A
```

```
##        [,1]  [,2]
## [1,] 0.334 0.666
## [2,] 0.333 0.667
```

So after three steps, the chance that the chain moves from off (row 1) to on (row 2) is $\boxed{0.6660}$.

171.

Consider a Markov chain with transition matrix over state space $\{1, 2, 3, 4\}$:

$$A = \begin{pmatrix} 0 & 0.1 & 0.9 & 0 \\ 0 & 0.3 & 0.3 & 0.4 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}$$

a. Find the limiting distribution by raising the transition matrix to a high power.

b. Verify that the limiting distribution is a stationary distribution.

Solution. a. First enter the transition matrix into R.

```
A <- matrix(c(0, 0.1, 0.9, 0,
              0, 0.3, 0.3, 0.4,
              0, 0.5, 0.5, 0,
              0, 0, 0.3, 0.7),
            byrow = TRUE,
            nrow = 4)
```

To raise the transition matrix to a high power, the %^% operator can be used in R.

```
A %^% 1000
```

```
##      [,1]      [,2]  [,3]      [,4]
## [1,]    0 0.2678571 0.375 0.3571429
## [2,]    0 0.2678571 0.375 0.3571429
## [3,]    0 0.2678571 0.375 0.3571429
## [4,]    0 0.2678571 0.375 0.3571429
```

From this result, the limiting distribution (to 4 sig figs) is

$$\boxed{\begin{pmatrix} 0 & 0.2678 & 0.3750 & 0.3571 \end{pmatrix}}.$$

b. To verify that this is a stationary distribution, multiply it by the original transition matrix.

```
(A %^% 1000)[1,] %*% A
```

```
##        [,1]      [,2]  [,3]      [,4]
## [1,]      0 0.2678571 0.375 0.3571429
```

The result is once again the limiting distribution, making it also stationary!

173.

Consider a Markov chain with transition matrix over state space $\{a, b, c, d\}$:

$$T = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0 & 0 & 0.4 & 0.6 \end{pmatrix}$$

a. Find $\lim_{t\to\infty} T^t$ by raising T to a high power.

b. Does this chain have a limiting distribution?

Solution. a. Using `{{0.4, 0.6, 0, 0}, {0.6, 0.4, 0, 0}, {0, 0, 0.3, 0.7}, {0, 0, 0.4, 0.6}}^{1000}` in Wolfram Alpha returns

$$\boxed{\begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.3636 & 0.6363 \\ 0 & 0 & 0.3636 & 0.6363 \end{pmatrix}}$$

b. Since the first two rows are different from the last two, starting in one of the first two positions will get you the dist $(0.5, 0.5, 0, 0)$ and starting in the last two positions will get you $(0, 0, 0.3636, 0.6363)$, Since these are different, there is $\boxed{\text{no}}$ single limiting distribution.

175.

A poker player either wins \$1 or loses \$1 with equal probability. The player starts with \$2, and quits when reaching \$0 or \$5.

a. Write this down as a transition matrix where 0 and 5 are absorbing states.

b. Raise the transition matrix to a high power to discover the probability that the poker player ends up with \$5.

c. Was this what you expected from martingale theory?

Solution. a. The transition matrix is

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

b. A matrix in R can be created with the `matrix` command.

```
A <- matrix(
  c(1, 0, 0, 0, 0, 0,
    1 / 2, 0, 1 / 2, 0, 0, 0,
    0, 1 / 2, 0, 1 / 2, 0, 0,
    0, 0, 1 / 2, 0, 1 / 2, 0,
    0, 0, 0, 1 / 2, 0, 1 / 2,
    0, 0, 0, 0, 0, 1),
  byrow = TRUE,
  ncol = 6
)
```

One way the matrix can be raised to a high power by multiplying it by itself several times. Repeated squaring is the fastest way to use the `%*%` operator to raise the matrix to a large power.

```
A2 <- A %*% A
A4 <- A2 %*% A2
A8 <- A4 %*% A4
A16 <- A8 %*% A8
A32 <- A16 %*% A16
A64 <- A32 %*% A32
A128 <- A64 %*% A64
A256 <- A128 %*% A128
A512 <- A256 %*% A256
A1024 <- A512 %*% A512
A1024
```

```
##      [,1]         [,2]         [,3]         [,4]         [,5] [,6]
## [1,]  1.0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00  0.0
## [2,]  0.8 1.549371e-95 0.000000e+00 2.506935e-95 0.000000e+00  0.2
## [3,]  0.6 0.000000e+00 4.056306e-95 0.000000e+00 2.506935e-95  0.4
## [4,]  0.4 2.506935e-95 0.000000e+00 4.056306e-95 0.000000e+00  0.6
## [5,]  0.2 0.000000e+00 2.506935e-95 0.000000e+00 1.549371e-95  0.8
## [6,]  0.0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00  1.0
```

Starting from state 2 (row 3), the probability of ending at state 5 (column 6) is $\boxed{0.4000}$.

c. From the martingale perspective, where

$$M_t = \sum_{i=1}^{t} D_i,$$

with $D_1, D_2, \ldots$ iid Unif($\{-1,1\}$),

$$T = \inf\{t : M_t \in \{0,5\}\}$$

is a stopping time, and $M_{t\wedge T}$ is a uniformly integrable martingale. Since $M_{t\wedge T}$ cannot converge until $M_t \in \{0,5\}$, the Martingale Convergence Theorem gives that $\mathbb{P}(T < \infty) = 1$, and

$$\mathbb{E}[M_\infty | M_0 = 2] = \mathbb{E}[M_T | M_0 = 2] = 2,$$

and

$$2 = \mathbb{P}(M_T = 5 | M_0 = 2)(5) + \mathbb{P}(M_T = 0 | M_0 = 2)(0),$$

hence

$$\mathbb{P}(M_T = 5 | M_0 = 2) = 2/5 = \boxed{0.4000},$$

as with the Markov chain approach.

177.

Consider a Markov chain on state space $\Omega = \{a, b, c\}$ with transition matrix

$$\begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.8 & 0 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$

Raise the matrix to a high power.

a. Does this Markov chain have a limiting distribution?

b. Find a stationary distribution for the chain.

Solution. a. With the expm library in R loaded, the %^% operator raises a matrix to a high power.

```
matrix(c(0.3, 0.4, 0.3,
         0.2, 0.8, 0,
         0, 0.1, 0.9),
       byrow = TRUE,
       nrow = 3) %^% 1000
```

```
##           [,1]      [,2] [,3]
## [1,] 0.1333333 0.4666667  0.4
## [2,] 0.1333333 0.4666667  0.4
## [3,] 0.1333333 0.4666667  0.4
```

So the limiting distribution vector is

$$\boxed{(0.1333\ldots, 0.4666\ldots, 0.4000)}$$

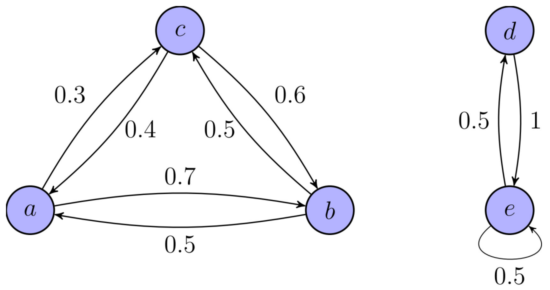
b. Any limiting distribution is a stationary distribution as well, so

$$\boxed{(0.1333\ldots, 0.4666\ldots, 0.4000)}$$

is also stationary.

179.

Consider the following Markov chain:



(a) Write down the transition matrix for the chain. [Notice that the matrix you find is in block form.]

(b) What does the transition matrix look like after taking a lot of steps in the Markov chain?

Solution. (a) The transition matrix is:

$$\boxed{\begin{pmatrix} 0 & 0.7 & 0.3 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}.}$$

(b) The limiting transition matrix is:

$$\boxed{\begin{pmatrix} 0.3139 & 0.3946 & 0.2915 & 0 & 0 \\ 0.3139 & 0.3946 & 0.2915 & 0 & 0 \\ 0.3139 & 0.3946 & 0.2915 & 0 & 0 \\ 0 & 0 & 0 & 0.3333 & 0.6666 \\ 0 & 0 & 0 & 0.3333 & 0.6666 \end{pmatrix}.}$$

Chapter 21

181.

Suppose a Markov chain has communication classes $\{a, b\}, \{c\}, \{d, e\}$. States a, c, d are all recurrent.

a. Is state b recurrent or transient?

b. Is state e recurrent or transient?

Solution. a. Because b is in the same communication class as state a it must also be $\boxed{\text{recurrent}}$.

b. Because e is in the same communication class as state d it must also be $\boxed{\text{recurrent}}$.

183. Suppose a Markov chain over state space $\{1, 2, 3\}$ has transition matrix

$$A = \begin{pmatrix} 0.3 & 0.7 & 0 \\ 0.6 & 0.4 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

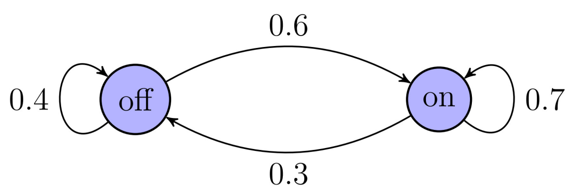
Classify each state as either recurrent or transient.

Solution. States 1 and 2 communicate with each other, and 3 communicates with itself. So the communication classes are $\{1, 2\}, \{3\}$.

Since each communication class has no outgoing edges, all the classes are recurrent. That is, $\boxed{1, 2, 3 \text{ are recurrent}}$.

185.

This Markov chain models a telecommunications circuit that is either on or off.



a. What is the transition matrix for the chain?

b. Find the limiting distribution by raising the transition matrix to a high power.

c. Verify that the limiting distribution is a stationary distribution.

Solution. In R, use the `expm` library to raise matrices to a high power. (Use `install.packages("expm")` if the library is not installed on your system.)

```
library(expm)
```

(a) Putting columns and rows in order (off, on), the transition matrix is

$$\boxed{A = \begin{pmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{pmatrix}}$$

(b) Raising this to a high power gives

```
A <- matrix(c(0.4, 0.6,
              0.3, 0.7),
            byrow = TRUE,
            nrow = 2)
A %^% 100
```

```
##          [,1]      [,2]
## [1,] 0.3333333 0.6666667
## [2,] 0.3333333 0.6666667
```

(c) The numerical method cuts off after 7 sig figs, but the real limiting distribution is $(1/3, 2/3)$.

To verify that this distribution is also stationary, multiply on the right by the transition matrix:

$$(1/3 \quad 2/3)\begin{pmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{pmatrix} = ([4/30 + 6/30] \quad [3/30 + 14/30]) = (1/3 \quad 2/3)$$

so the limiting distribution is also stationary!

187.

Consider a Markov chain with transition matrix over state space $\{a, b, c, d\}$:

$$T = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0 & 0 & 0.4 & 0.6 \end{pmatrix}$$

a. Find $\lim_{t\to\infty} T^t$ by raising T to a high power.

b. Does this chain have a limiting distribution?

Solution. a. Using `{{0.4, 0.6, 0, 0}, {0.6, 0.4, 0, 0}, {0, 0, 0.3, 0.7}, {0, 0, 0.4, 0.6}}^{1000}` in Wolfram Alpha returns

$$\boxed{\begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.3636 & 0.6363 \\ 0 & 0 & 0.3636 & 0.6363 \end{pmatrix}}$$

b. Since the first two rows are different from the last two, starting in one of the first two positions will get you the dist $(0.5, 0.5, 0, 0)$ and starting in the last two positions will get you $(0, 0, 0.3636, 0.3636)$, Since these are different, there is $\boxed{\text{no}}$ single limiting distribution.

189.

A poker player either wins \$1 or loses \$1 with equal probability. The player starts with \$2, and quits when reaching \$0 or \$5.

a. Write this down as a transition matrix where 0 and 5 are absorbing states.

b. Raise the transition matrix to a high power to discover the probability that the poker player ends up with \$5.

c. Was this what you expected from martingale theory?

Solution. a. The transition matrix is

$$\boxed{A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}$$

b. A matrix in R can be created with the `matrix` command. Entries are placed by columns, so the `t()` function transposes the matrix to rows.

```
A <- t(matrix(
  c(1, 0, 0, 0, 0, 0,
    1 / 2, 0, 1 / 2, 0, 0, 0,
    0, 1 / 2, 0, 1 / 2, 0, 0,
    0, 0, 1 / 2, 0, 1 / 2, 0,
    0, 0, 0, 1 / 2, 0, 1 / 2,
    0, 0, 0, 0, 0, 1),
  ncol = 6
))
```

The matrix can be raised to a high power by multiplying it by itself several times.

```
A2 <- A %*% A
A4 <- A2 %*% A2
A8 <- A4 %*% A4
A16 <- A8 %*% A8
A32 <- A16 %*% A16
A64 <- A32 %*% A32
A128 <- A64 %*% A64
A256 <- A128 %*% A128
A512 <- A256 %*% A256
A1024 <- A512 %*% A512
A1024
```

```
##      [,1]         [,2]         [,3]         [,4]         [,5] [,6]
## [1,]  1.0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00  0.0
## [2,]  0.8 1.549371e-95 0.000000e+00 2.506935e-95 0.000000e+00  0.2
## [3,]  0.6 0.000000e+00 4.056306e-95 0.000000e+00 2.506935e-95  0.4
## [4,]  0.4 2.506935e-95 0.000000e+00 4.056306e-95 0.000000e+00  0.6
## [5,]  0.2 0.000000e+00 2.506935e-95 0.000000e+00 1.549371e-95  0.8
## [6,]  0.0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00  1.0
```

Starting from state 2 (row 3), the probability of ending at state 5 (row 6) is $\boxed{0.4000}$.

c. From the martingale perspective, where

$$M_t = \sum_{i=1}^{t} D_i,$$

with $D_1, D_2, \ldots$ iid Unif($\{-1, 1\}$),

$$T = \inf\{t : M_t \in \{0, 5\}\}$$

is a stopping time, and $M_{t \wedge T}$ is a uniformly integrable martingale. Since $M_{t \wedge T}$ cannot converge until $M_t \in \{0, 5\}$, the Martingale Convergence Theorem gives that $\mathbb{P}(T < \infty) = 1$, and

$$\mathbb{E}[M_\infty | M_0 = 2] = \mathbb{E}[M_T | M_0 = 2] = 2,$$

and

$$2 = \mathbb{P}(M_T = 5 | M_0 = 2)(5) + \mathbb{P}(M_T = 0 | M_0 = 2)(0),$$

hence

$$\mathbb{P}(M_T = 5 | M_0 = 2) = 2/5 = \boxed{0.4000},$$

as with the Markov chain approach.

191.

Consider a Markov chain on state space $\Omega = \{a, b, c\}$ with transition matrix

$$\begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.8 & 0 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$

Raise the matrix to a high power.

a. Does this Markov chain have a limiting distribution?

b. Find a stationary distribution for the chain.

Solution. a. With the `expm` library in R loaded, the `%^%` operator raises a matrix to a high power. {r}

```
matrix(c(0.3, 0.4, 0.3,          0.2, 0.8, 0,          0,
0.1, 0.9),          byrow = TRUE,          nrow = 3) %^% 1000
```

So the limiting distribution vector is

$$\boxed{(0.1333\ldots, 0.4666\ldots, 0.4000)}$$

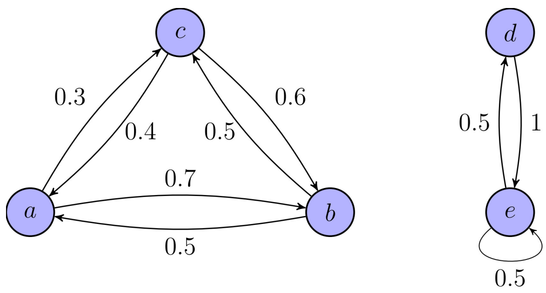
b. Any limiting distribution is a stationary distribution as well, so

$$\boxed{(0.1333\ldots, 0.4666\ldots, 0.4000)}$$

is also stationary.

193.

Consider the following Markov chain:



(a) Write down the transition matrix for the chain. [Notice that the matrix you find is in block form.]

(b) What does the transition matrix look like after taking a lot of steps in the Markov chain?

Solution. (a) The transition matrix is:

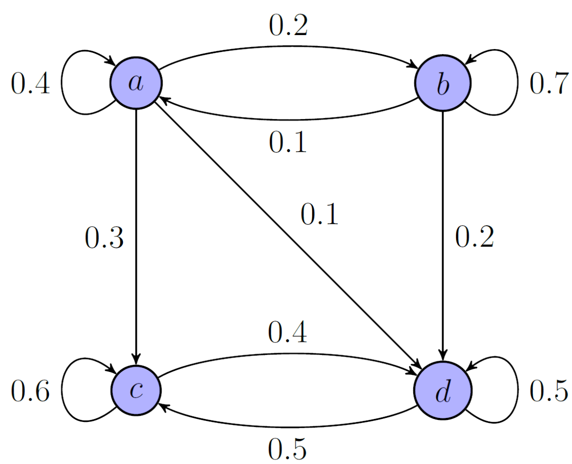
$$\begin{pmatrix} 0 & 0.7 & 0.3 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}.$$

(b) The limiting transition matrix is:

$$\begin{pmatrix} 0.3139 & 0.3946 & 0.2915 & 0 & 0 \\ 0.3139 & 0.3946 & 0.2915 & 0 & 0 \\ 0.3139 & 0.3946 & 0.2915 & 0 & 0 \\ 0 & 0 & 0 & 0.3333 & 0.6666 \\ 0 & 0 & 0 & 0.3333 & 0.6666 \end{pmatrix}.$$

195.

Consider this Markov chain:



(a) What are the communication classes?

(b) Which communication classes are transient?

(c) What is the limiting distribution π?

(d) What is $\pi(i)$ for the transient states i?

Solution. (a) Since there is a path from a to b and b to a, they are in the same class. Similarly a path from c to d and d to c puts them in the same class. Since there is no path from c to a they are in different classes. The resulting communication classes are:

$$\boxed{\{a,b\},\{c,d\}}.$$

(b) Since only the class $\boxed{\{a,b\}}$ has edges leaving it, this is the only class that is transient.

(c) To find the limiting distribution, raise the transition matrix to a high power:

$$A = \begin{pmatrix} 0.4 & 0.2 & 0.3 & 0.1 \\ 0.1 & 0.7 & 0 & 0.2 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix},$$

so

$$A^{100} = \begin{pmatrix} 9.89157 \times 10^{-14} & 3.52293 \times 10^{-13} & 0.555556 & 0.444444 \\ 1.76147 \times 10^{-13} & 6.27356 \times 10^{-13} & 0.555556 & 0.444444 \\ 0 & 0 & 0.555556 & 0.444444 \\ 0 & 0 & 0.555556 & 0.444444 \end{pmatrix}$$

and

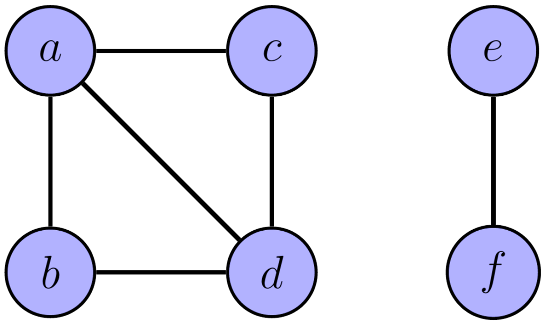
$$A^{1000} = \begin{pmatrix} 5.56104 \times 10^{-123} & 1.98059 \times 10^{-122} & 0.555556 & 0.444444 \\ 9.90297 \times 10^{-123} & 3.52699 \times 10^{-122} & 0.555556 & 0.444444 \\ 0 & 0 & 0.555556 & 0.444444 \\ 0 & 0 & 0.555556 & 0.444444 \end{pmatrix}.$$

So the limiting distribution is

$$\boxed{(0.000 \quad 0.000 \quad 0.5555 \quad 0.4444)}.$$

(d) The transient states $\{a,b\}$ both have limiting distribution $\boxed{0}$.

197. In a *random walk on an undirected graph* you start at a node on the graph, and then take a step uniformly at random to move to a node that there is an edge to. For instance, consider the following graph.



The node a is connected by an edge to three other nodes, b, c, and d. Hence a has a $1/3$ chance of moving to b, a $1/3$ chance of moving to c, and a $1/3$ chance of moving to d.

Write down the transition matrix for this Markov chain.

Solution. This would be

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

Chapter 22

199. Give an example of a finite state aperiodic recurrent Markov chain (so exactly 1 recurrent communication class) with at least 4 states and stationary measure $\mu(i) = 1$ for all $i \in \Omega$.

Solution. There are many ways to do this. One way is to have every state communicate with every other state. If each state sends out its probability equally to every other state, then $(1, 1, 1, 1)$ will be a left (and right eigenvector). Hence one such transition matrix is

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

201. Consider a Markov chain $\mathcal{M}$ on state space $\Omega = \{a, b, c\}$ with transition matrix

$$\begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.8 & 0 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$

Consider the measure

$$\mu(a) = 3, \ \mu(b) = 2, \ \mu(c) = 5.$$

What is $\mathcal{M}(\mu)$?

Solution. $\mathcal{M}(\mu)$ assigns $3(0.3) + 2(0.2) + 5(0) = 1.3$ for state a, $3(0.4) + 2(0.8) + 5(0.1) = 3.3$ to state b, and $3(0.3) + 2(0) + 5(0.9) = 5.4$. So the measure is

$$\boxed{\mathcal{M}(\mu)(a) = 1.3,\ \mathcal{M}(\mu)(b) = 3.3,\ \mathcal{M}(\mu)(c) = 5.4}.$$

203. Let M_t be a Markov chain on state space x, y, z. Suppose that $X_0 = x$, and $X_{R_x} = x$ where $R_x = \inf\{t > 0 : X_t = x\}$. During the trip from x back to x, y is visited an average of 3.2 times, and z is visited an average of 2.7 times.

Give a stationary measure for the Markov chain.

Solution. The number of vists of x during this loop (not counting the last visit) is 1. Hence the stationary measure is

$$\boxed{\mu(x) = 1,\ \mu(y) = 3.2,\ \mu(z) = 2.7}.$$

205.

Suppose a Markov chain has a single communication class $\{1, 2, 3, 4\}$. If you start at state 1, the expected number of visits to the other states before returning to 1 are

$$\mathbb{E}[N_2] = 4.3$$
$$\mathbb{E}[N_3] = 1.2$$
$$\mathbb{E}[N_4] = 0.6$$

a. What is $\mathbb{E}[R_1]$?

b. Give at least one stationary measure for this chain.

Solution. a. Including the last step that takes us back to 1, this is

$$1 + 4.3 + 1.2 + 0.6 = \boxed{7.100}.$$

b. One stationary measure has vector

$$\boxed{(1, 4.3, 1.2, 0.6)}$$

where the entries are just the expected number of visits to each state in between visits to 1 (including zero steps).

Chapter 23

207. A finite state Markov chain has one communication class, $\{a, b, c\}$. How many stationary distributions π for the chain have $\pi(\{a, b, c\}) = 1$?

Solution. There is exactly $\boxed{\text{one}}$ stationary distribution for the one recurrent class.

209.

State if state x is recurrent (but not positive recurrent), positive recurrent, or transient.

a. $\mathbb{E}[R_x] = 4.2$.

b. $\mathbb{E}[R_x] = \infty, \mathbb{P}(R_x < \infty) = 1$.

c. $\mathbb{P}(R_x = \infty) = 0.3$.

Solution. a. Here the mean return time is finite, so it is $\boxed{\text{positive recurrent.}}$

b. Here the mean return time is infinity, but the return time is finite with probability 1, so it is $\boxed{\text{recurrent}}$.

c. Here the probability it does not return to the state is positive, so it is $\boxed{\text{transient}}$.

211. Suppose that a Markov chain has recurrent communication class $\{a, b, c\}$ with stationary distribution $(0.3, 0.2, 0.5)$. What is $\mathbb{E}[R_c]$?

Solution. This is the inverse of the stationary distribution at c, so

$$\frac{1}{0.5} = \boxed{2}.$$

Chapter 24

213. Suppose a Markov chain has state space $\{a, b, c\}$, where $\{a, b\}$ is a recurrent communication class and $\{c\}$ is a transient communication class. Does the chain have a limiting distribution?

Solution. $\boxed{\text{Yes.}}$ Because the chain has exactly one recurrent communication class, by the ergodic theorem it is guaranteed to have one stationary distribution that is also the limiting distribution.

215.

Suppose $X \sim \textsf{Unif}(\{1, 2, 3\})$ and $Y \sim \textsf{Unif}(\{1, 2, 3, 4, 5\})$.

a. By adding in all elements with $\mathbb{P}(X = a) > \mathbb{P}(Y = a)$, find the set $A \subseteq \{1, 2, 3, 4, 5\}$ such that $\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)$ is as large as possible?

b. What is the set $B \subseteq \{1, 2, 3, 4, 5\}$ such that $\mathbb{P}(Y \in B) - \mathbb{P}(X \in B)$ is as large as possible?

c. What is the relationship between A and B?

d. What is the total variation distance between X and Y?

Solution. a. Note that

$$\mathbb{P}(X \in A) - \mathbb{P}(Y \in A) = \sum_{a \in A} \mathbb{P}(X = a) - \mathbb{P}(Y = a).$$

So for A to be as large as possible, all a such that $\mathbb{P}(X = a) > \mathbb{P}(Y = a)$ should be included. This makes $\boxed{A = \{1, 2, 3\}}$.

b. Similarly, including all elements $b \in B$ such that $\mathbb{P}(X = b) > \mathbb{P}(Y = b)$ gives $\boxed{B = \{4, 5\}}$.

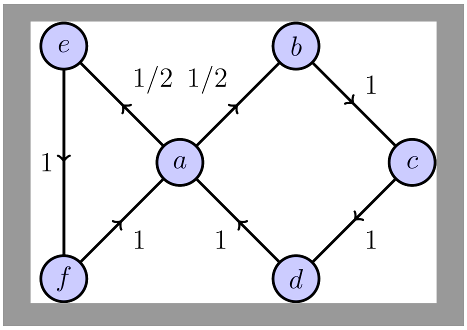
c. The sets A and B are $\boxed{\text{complements}}$ of each other.

d. The total variation distance then is

$$\mathbb{P}(X \in A) - \mathbb{P}(Y \in A) = 1 - 0.6 = \boxed{0.4000} = \mathbb{P}(Y \in B) - \mathbb{P}(X \in B) = 2/5 - 0.$$

217.

Consider the Markov chain with transition graph:



If the chain has period k, then it holds that $\{t : \mathbb{P}(X_t = a|X_0 = a) > 0\} \subseteq \{k, 2k, 3k, 4k, ...\}$.

(a) Find enough elements in the set $\{t : \mathbb{P}(X_t = a|X_0 = a) > 0\}$ to show that the chain is aperiodic.

(b) Find the limiting distribution of the chain.

Solution. (a) Consider: $\mathbb{P}(X_3 = a|X_0 = a) = \mathbb{P}(X_4 = a|X_0 = a) = 1/2 > 0$. Hence $\{3, 4\} \subseteq \{k, 2k, 3k, ...\}$. But the only k for which this is true is $k = 1$, so the chain is aperiodic.

(b) Since the chain is recurrent and aperiodic, the stationary distribution is also the limiting distribution by the Ergodic Theorem for finite state Markov chains, and so is also

$$\boxed{(0.2857 \quad 0.1428 \quad 0.1428 \quad 0.1428 \quad 0.1428 \quad 0.1428)}$$

219. Suppose that transition matrix A has eigenvalues 1 and -1, but there are no other eigenvalues that when raised to a nonnegative integer power equals 1. What is the period of A?

Solution. Since $(-1)^2 = 1$, the period is a multiple of 2. Since there are no other eigenvalues that when raised to a power equals 1, the period must be $\boxed{2}$.

Chapter 25

221.

Consider a two state Markov chain with transition matrix.

$$A = \begin{pmatrix} 0.4 & 0.6 \\ 0.5 & 0.5 \end{pmatrix}.$$

a. What is the expected travel time from b to a?

b. What is the expected return time to a?

c. What is the stationary distribution of a?

Solution. a. Starting at b, at each step there is a $1/2$ chance of moving to a. Hence $\boxed{\mathbb{E}[T_{ba} = 2]}$.

b. From state a, there is a 0.4 chance of returning to a in one step, and a 0.6 chance of moving to b, then requiring $\mathbb{E}[T_{ba}]$ more steps. Now, since the state stays at b before moving to a, T_{ba} is a geometric random variable with parameter $1/2$. That makes the mean value $1/(1/2) = 2$, so

$$\mathbb{E}[R_a] = 0.4(1) + 0.6(2) = \boxed{1.600}.$$

c. The stationary distribution of a is $1/1.6 = 1/(8/5) = 5/8 = \boxed{0.6250}$.

223. Consider the following four state Markov chain transition matrix.

$$A = \begin{pmatrix} 0.1 & 0.3 & 0.4 & 0.2 \\ 0 & 0.5 & 0 & 0.5 \\ 0.3 & 0.1 & 0.2 & 0.4 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

What is the expected time needed to travel from state a to state d? Calculate this through the artificial node method.

Solution. Add an artificial node e where $p(d, e) = p(e, a) = 1$ to get transition matrix

$$B = \begin{pmatrix} 0.1 & 0.3 & 0.4 & 0.2 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0.3 & 0.1 & 0.2 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then

```
B <- matrix(c(0.1, 0.3, 0.4, 0.2, 0,
              0, 0.5, 0, 0.5, 0,
              0.3, 0.1, 0.2, 0.4, 0,
              0, 0, 0, 0, 1,
              1, 0, 0, 0, 0),
            byrow = TRUE,
            nrow = 5)
```

In this case, the chain has a limiting distribution.

```r
B1000 <- B %^% 1000
B1000
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 0.2702703 0.1891892 0.1351351 0.2027027 0.2027027
## [2,] 0.2702703 0.1891892 0.1351351 0.2027027 0.2027027
## [3,] 0.2702703 0.1891892 0.1351351 0.2027027 0.2027027
## [4,] 0.2702703 0.1891892 0.1351351 0.2027027 0.2027027
## [5,] 0.2702703 0.1891892 0.1351351 0.2027027 0.2027027
```

So one over the stationary distribution for e minus 2 gives the expected travel time from a to d.

```r
1 / B1000[1, 5] - 2
```

```
## [1] 2.933333
```

So the answer to four sig figs is $\boxed{2.933}$.

225.

Consider simple symmetric random walk with partially reflecting boundaries on states $\{0, 1, \ldots, 10\}$.

a. What is the expected time needed to return to state 0 starting from state 0?

b. What is the expected time needed to return to state 7 starting from state 7?

Solution. a. From our calculations above, this will be $1/11 = \boxed{0.09090\ldots}$.

b. This is still $1/11 = \boxed{0.09090\ldots}$.

Chapter 26

227. Suppose $A \sim \text{Unif}(\{1, \ldots, 90\})$ and $B \sim \text{Unif}(\{10, 11, \ldots, 110\})$ have a coupling where there is a 0.4 chance the states are equal. What can be said about the total variation distance?

Solution. If there is a 0.4 chance of the states being equal, there is a 0.6 chance that they are unequal. Hence the total variation distance is at most $\boxed{0.6000}$.

229. Suppose $X \sim \text{Exp}(1/2)$ and $Y \sim \text{Unif}([0, 1])$ have a 0.1 chance of being equal with a certain coupling. What can be said about the total variation distance?

Solution. By the coupling lemma, $\boxed{\text{dist}_{\text{TV}}(X, Y) \leq 0.9000}$.

231.

Consider $X \sim \text{Unif}([0, 2])$ and $Y \sim \text{Unif}([1, 3])$. Note that $\mathbb{P}(X \in [1, 2])$ and $\mathbb{P}(Y \in [1, 2])$ both are $1/2$. Also recall that for W a uniform random variable over set B, for $A \subseteq B$, $[W | W \in A] \sim \text{Unif}(A)$. This gives a way to couple uniform random variables.

Let $B \sim \mathsf{Bern}(1/2)$ and $U \sim \mathsf{Unif}([0,1])$

a. Let $X = 2U$ and $Y = 1 + 2U$. What is $\mathbb{P}(X \neq Y)$?

b. Let $X = B(1+U) + (1-B)(U)$ and $Y = B(1+U) + (1-B)(2+U)$. What is $\mathbb{P}(X \neq Y)$?

Solution. a. This is $\boxed{1}$, since the values are always different.

b. Note if $B = 1$, then $X = Y$, and if $B = 0$, then $X \neq Y$. So the answer is $\boxed{0.5000}$.

233.

Suppose that $X_0 = x_0$ and $Y_0 \sim \pi$ are two copies of a Markov chain which has π as the stationary distribution. Suppose $\mathbb{P}(Y_0 \neq X_0) \leq 10\exp(-t/5)$.

a. What can be said about $\text{dist}_{\mathsf{TV}}(X_t, Y_t)$ when $t = 20$?

b. What can be said about $\text{dist}_{\mathsf{TV}}(X_t, Y_t)$ when $t = 30$?

Solution. a. The coupling lemma says that the total variation distance can be bounded above by $\boxed{0.1831\ldots}$.

b. The coupling lemma says that the total variation distance can be bounded above by $\boxed{0.0.2478\ldots}$.

Chapter 27

235. Suppose two copies of a Markov chain $\{X_t\}$ and $\{Y_t\}$ are coupled together so that each follows the same transition matrix.

Further, say that $X_0 = x_0$ where $x_0 \in \Omega$, $Y_0 \sim \pi$, and $\mathbb{P}(X_{100} = Y_{100} | X_0 = x_0) = 0.99$. Bound the total variation distance between X_{100} and π.

Solution. By negation, the chance that X_{100} does not equal Y_{100} is $1 - 0.99 = 0.01$. By the Coupling Lemma, it holds that

$$\text{dist}_{\mathsf{TV}}([X_{100} | X_0 = x_0], \pi) \leq \boxed{1\%}$$

237. Suppose that for all states x and y,

$$\mathbb{P}(X_{100} = Y_{100} | X_0 = x, Y_0 = y) = 0.1,$$

and that if $X_t = Y_t$, then $X_{t'} = Y_{t'}$ for all $t' \geq t$. Create an upper bound for

$$\mathbb{P}(X_{200} = Y_{200})$$

Solution. In order for $X_{200} \neq Y_{200}$ it needs to be true that the two chains failed to meet in the first 100 steps, and then they failed to meet again in the second set of 100 steps. Therefore, the chance that they *don't* meet is at most $0.9^2 = 0.81$. So the complement gives $\mathbb{P}(X_{200} = Y_{200}) \geq \boxed{0.1900}$.

239. Suppose (X_t, Y_t) is a coupling for two copies of a Markov chain such that

$$\mathbb{P}(X_t \neq Y_t) \leq 100 \exp(-t/10).$$

Give an upper bound on $\tau_{0.05}$.

Solution. The total variation distance is at most $100 \exp(-t/10)$, so the goal is to get this number to be at most 0.05. This happens when

$$\begin{aligned} 100 \exp(-t/10) &= 0.05 \\ \exp(-t/10) &= 0.0005 \\ -t/10 &= \ln(0.0005) \\ t &= 10 \ln(1/0.0005) \\ t &= 76.009 \ldots, \end{aligned}$$

so the best that can be said is that $\tau_{0.05} \leq \boxed{77}$.

241.

Suppose that

$$(\forall x_0 \in \Omega)(\text{dist}_{\text{TV}}([X_t | X_0 = x_0], \pi) \leq 1000 \exp(-t/100)).$$

a. Give an upper bound on $\tau_{0.01}$.

b. Give an upper bound on $\tau_{0.000001}$.

Solution. a. We need

$$1000 \exp(-t/100) \leq 0.01,$$

which solves to give

$$t \geq 100 \ln(1000/0.01) = 1151.29 \ldots,$$

which means $\boxed{1152}$ is an upper bound on $\tau_{0.01}$.

b. Here the need is to find t such that

$$1000 \exp(-t/100) \leq 10^{-6},$$

which solves to give

$$t \geq 100 \ln(1000/0.000001) = 2072.32 \ldots,$$

which means $\boxed{2073}$ is an upper bound on $\tau_{10^{-6}}$.

Note that by doubling our number of steps, the total variation distance has gone down by a factor of 10000.

Chapter 28

243.

State if the following are true or false. (You do not need to justify your answer.)

a. In finite state Markov chains, recurrent communication classes are also positive recurrent.

b. In countable state space Markov chains, recurrent communication classes are also positive recurrent.

c. Positive recurrent communication classes are always aperiodic.

d. In a countable state space Markov chain with one recurrent communication class, there is always a stationary measure.

Solution. a. This is $\boxed{\text{T}}$.

b. This is $\boxed{\text{F}}$.

c. This is $\boxed{\text{F}}$.

d. This is $\boxed{\text{T}}$.

245. Give an example of a Markov chain with a countably infinite state space that has one recurrent communication class of period 2.

Solution. There are of course an infinite number of ways to do this. One way is to just use simple symmetric random walk on the integers:

$$(\forall t \in \mathbb{Z})(\mathbb{P}(X_t = X_{t-1} + 1 | X_{t-1}) = \mathbb{P}(X_t = X_{t-1} - 1 | X_{t-1}) = 1/2).$$

247.

Consider a Markov chain on the nonnegative integers that has $\pi(i) = (1/2)^{i+1}$ as a stationary distribution.

a. Is this enough information to find $\mathbb{E}[R_0]$? If so, what is it?

b. What if we add the extra condition that $(\forall i, j)(\exists t)(\mathbb{P}(X_t = j | X_0 = i) > 0)$, and $\mathbb{P}((\exists t)(X_t = 0 | X_0 = 0)) = 1$? Now do we have enough information to find $\mathbb{E}[R_0]$, and if so, what is it?

Solution. a. $\boxed{\text{No!}}$ To find $\mathbb{E}[R_0]$ it is necessary to use the ergodic theorem for countable state spaces, which in turn requires that there be a single communication class. It could be that $\mathbb{P}(X_t = i | X_{t-1} = i) = 1$, and there are an infinite number of communication classes.

b. $\boxed{\text{Yes!}}$ With this information, the state space has a single recurrent communication class, and since π is stationary over this single class, $\mathbb{E}[R_0] = 1/\pi(0) = \boxed{2}$.

249. Suppose a Markov chain on $\Omega = \{2, 4, 6, \ldots\}$ has stationary distribution

$$\pi(i) = \frac{C}{i^2},$$

where C is a constant.

Suppose

$$(\forall i \in \Omega)(\mathbb{P}(X_1 = i | X_0 = 2) > 0),$$

and

$$(\forall i \in \Omega)(\mathbb{P}(X_1 = 2|X_0 = i) > 0),$$

Prove that this chain has one positive recurrent communication class.

Solution. First consider any two states $i \neq j$ in Ω. Then

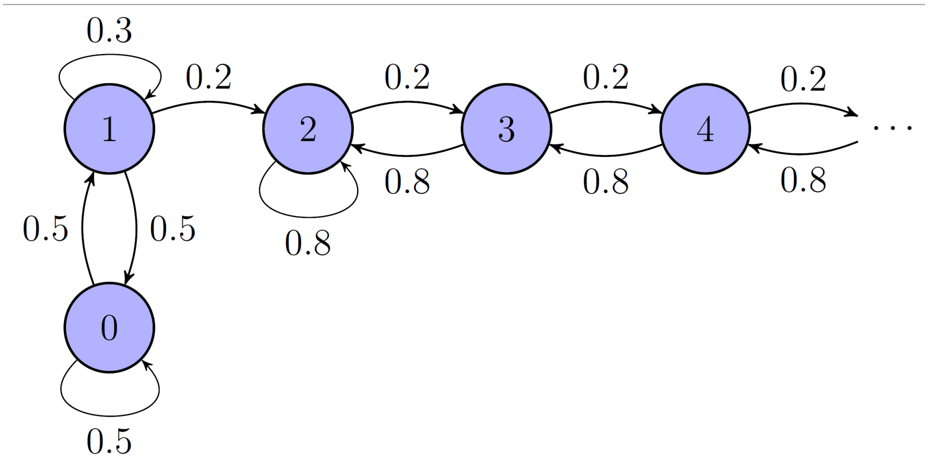
$$\mathbb{P}(X_2 = j|X_0 = i) \geq \mathbb{P}(X_1 = 2|X_0 = i)\mathbb{P}(X_2 = j|X_1 = 2) > 0,$$

so every pair of states communicates, and there is only one communication class.

Therefore, because there is only one communication class with a stationary distribution, that class is positive recurrent by the Ergodic Theorem, and the proof is complete. □

251.

Consider the following Markov chain:



a. What are the communication classes of the chain?

b. Write down the balance equations for the chain.

c. Does a stationary distribution for this chain exist? If so, what is it?

d. For each communication class, state if it is recurrent or transient.

Solution. a. States 0 and 1 communicate, and every state i and j with $i \geq 2$ and $j \geq 2$ communicate, so their are two communication classes:

$$\boxed{\{0, 1\}, \{2, 3, 4, ...\}}.$$

b. States 0, 1, and 2 are different from the rest:

$$\pi(0) = 0.5\pi(0) + 0.5\pi(1)$$
$$\pi(1) = 0.5\pi(0) + 0.3\pi(1)$$
$$\pi(2) = 0.2\pi(1) + 0.8\pi(2) + 0.8\pi(3)$$

and for all $i > 2$:

$$\pi(i) = 0.2\pi(i-1) + 0.8\pi(i+1).$$

c. Yes, it does! First, since 0 and 1 are transient, set $\pi(0) = \pi(1) = 0$. Next, we have

$$\pi(2) = 0.8\pi(2) + 0.8\pi(3) \Rightarrow \pi(2) = 4\pi(3).$$

That in turn gives:

$$\pi(3) = 0.2\pi(2) + 0.8\pi(4) \Rightarrow \pi(3) = 4\pi(4).$$

This pattern continues, in general

$$\pi(i) = 4\pi(i+1).$$

Put another way, for $i \geq 2$, $\pi(i+1) = (1/4)\pi(i)$. Since

$$\sum_{i=2}^{\infty} \pi(2)(1/4)^{i-1} = \frac{\pi(2)}{1 - 1/4} = (4/3)\pi(2) = 1,$$

we know that $\pi(2) = 3/4$. Hence the stationary distribution is

$$\boxed{\pi(i) = \frac{3}{4}\left(\frac{1}{4}\right)^{i-2} \mathbb{I}(i \geq 2).}$$

d. Because the class $\{0, 1\}$ has edges leaving the class, this class is transient. Because the class $\{2, 3, \ldots\}$ has a stationary distribution with probability 1 of falling in the class, it is recurrent. so the answer is

$$\boxed{\{0,1\} \text{ is transient, } \{2,3,\ldots\} \text{ is recurrent.}}$$

Chapter 29

253.

State whether or not the following are true. You do not have to justify your answer.

a. All connected countable state space chains are Harris.

b. All Harris chains have a countable state space.

c. Periodicity is not necessary for a Harris chain to have a limiting distribution.

d. Harris chains always return to the set A in the definition with probability 1.

Solution. a. This is $\boxed{\text{true}}$.

b. This is $\boxed{\text{false}}$.

c. This is $\boxed{\text{false}}$.

d. This is $\boxed{\text{false}}$.

255. Show that if $U_1, \ldots, U_{10}$ are iid $\text{Unif}([0,1])$, then $\mathbb{P}(U_1 + \cdots + U_{10} \in [6, 6.1]) > 0$.

Solution. If $U_i \in [6/10, 6.1/10]$ for all i from 1 to 10, then $\sum_{i=1}^{10} U_i \in [6, 6.1]$. Since $\mathbb{P}(U_i \in [6/10, 6.1/10]) = 6.1/10 - 6/10 = 0.01$, the chance that $\sum U_i$ is in $[6, 6.1]$ is at least $(0.01)^{10} > 0$.

257. Let $a \leq a' \leq b' \leq b$. Show that if $X \sim \text{Unif}([a,b])$, then $[X|X \in [a', b']] \sim \text{Unif}([a', b'])$.

Solution. *Proof.* One way to show that a distribution is correct is to find the cdf of the random variable. Let $\alpha \in [a', b']$. Then

$$\begin{aligned}\mathbb{P}(X \leq \alpha | X \in [a', b']) &= \frac{\mathbb{P}(X \leq \alpha, X \in [a', b'])}{\mathbb{P}(X \in [a', b'])} \\ &= \frac{(\alpha - a')/(b-a)}{(b' - a')/(b-a)} = \frac{\alpha - a'}{b' - a'}.\end{aligned}$$

which is exactly the cdf of a uniform over $[a', b']$. Therefore $[X|X \in [a', b'] \sim \text{Unif}([a', b'])$. □

Chapter 30

259. Suppose that $p(a,a) = 0.3$ and the communication class containing a is recurrent. What can you say about the period of the class containing a?

Solution. Because it has a holding probability at a, the communication class containing a must be aperiodic. That is, it has period $\boxed{1}$.

261. Suppose that $\{1, 2, 3, \ldots\}$ is a recurrent communication class in a Markov chain with stationary distribution π where $\pi(4) = 0.3$. What can you say about $\mathbb{E}[R_4]$?

Solution. Because the communication class is recurrent with a stationary distribution, it is positive recurrent, and $\mathbb{E}[R_4] = 1/\pi(4) = \boxed{3.333\ldots}$.

263.

Suppose that the state space is $\{0, 1, 2, \ldots\}$ with (for $i \geq 1$) $p(0,0) = p(i, i-1) = 0.9$ and $p(i, i+1) = 0.1$.

a. Find the unique solution to the balance equations.

b. Find $\mathbb{E}[R_1]$ exactly.

Solution. a. The balance equations are

$$\begin{aligned}\pi(0) &= 0.9\pi(1) + 0.9\pi(0)\\ \pi(1) &= 0.9\pi(2) + 0.1\pi(0)\\ \pi(2) &= 0.9\pi(3) + 0.1\pi(1)\\ \vdots &= \vdots\,.\end{aligned}$$

The solution to the first equation is $\pi(0) = 9\pi(1)$. This then plugged into the second equation is $\pi(1) = 9\pi(2)$. This continues to give

$$\boxed{\pi(i) = (8/9)(1/9)^i}.$$

b. This will be

$$1/pi(1) = 81/8 = \boxed{10.125}.$$

Chapter 31

265. Suppose $X_1, X_2, \ldots$ are iid uniform over $\{0,1,2\}$, and $\mathbb{P}(W = 0) = 0.2,\ \mathbb{P}(W = 2) = 0.8$. What is the generating function of

$$S = \sum_{i=1}^{W} X_i?$$

Solution. This will be

$$\text{gf}_S(a) = \text{gf}_W(\text{gf}_X(a)),$$

where $X \sim \text{Unif}(\{0,1,2\})$, so

$$\text{gf}_X(a) = (1/3)(1 + a + a^2),\ \text{gf}_W(a) = 0.2 + 0.8a^2.$$

Hence

$$\begin{aligned}\text{gf}_W(\text{gf}_X(a)) &= 0.2 + 0.8([(1/3)(1 + a + a^2)]^2)\\ &= 0.2 + 0.8/9 + 2(0.8/9)a + 3(0.8/9)a^2 + 2(0.8/9)a^3 + (0.8/9)a^4\\ &= \boxed{(1/45)[4a^4 + 8a^3 + 12a^2 + 8a + 13]}.\end{aligned}$$

267. Suppose a branching process has a number of children that is uniform over $\{0,1,2,3\}$. What is the extinction probability of this branching process?

Solution. The number of children has the uniform distribution. That is, $Y \sim \text{Unif}(\{0,1,2,3\})$. So the generating function is

$$\text{gf}_Y(a) = (1/4)[a^0 + a^1 + a^2 + a^3].$$

Setting $a = \text{gf}_Y(a)$ gives

$$0 = (1/4)a^3 + (1/4)a^2 - (3/4)a + 1/4 \to 0 = a^3 + a^2 - 3a + 1.$$

Recall $a = 1$ is always a solution, giving

$$0 = (a-1)(a^2 + 2a - 1).$$

Then the quadratic formula gives $r_1 = (-2 + \sqrt{4+4})/2$ and $r_2 = (-2 - \sqrt{4+4})/2$. Since $r_2 < 0$, only $r_1 \in (0,1)$, making the answer

$$\frac{-2+\sqrt{8}}{2} = -1 + \sqrt{2} = \boxed{0.4142\ldots}.$$

269. Suppose in a branching process where each individual has Y children

$$\mathbb{P}(Y=0) = 0.1,\ \mathbb{P}(Y=1) = 0.6,\ \mathbb{P}(Y=3) = 0.3.$$

What is the extinction probability of this branching process?

Solution. The extinction probability satisfies the equation

$$a = \phi_Y(a),$$

or in this problem:

$$a = 0.1 + 0.6a + 0.3a^3.$$

Rearranging gives

$$0.3a^3 - 0.4a + 0.1 = 0.$$

Plugging in $a = 1$ works, so use that to factor the equation using polynomial long division:

$$0.3a^3 - 0.4a + 0.1 = (a-1)(0.3a^2 + 0.3a - 0.1).$$

So the quadratic formula gives the other roots as:

$$r_1 = \frac{-0.3 + \sqrt{0.21}}{0.6}, \quad r_2 = \frac{-0.3 - \sqrt{0.21}}{0.6}.$$

Only $r_1 \in (0,1)$, and is approximately $\boxed{0.2637}$.

Chapter 32

271. Suppose

$$\mathbb{P}(W=0) = 0.5,\ \mathbb{P}(W=2) = 0.5$$

What is the generating function of W?

Solution. This would be

$$\boxed{\phi_W(x) = 0.5 + 0.5x^2}.$$

273. Suppose

$$\mathbb{P}(W=0) = 0.5,\ \mathbb{P}(W=2) = 0.5$$

is the distribution of the number of children for an individual in a branching process. Starting with 1 person, what is the probability that the population is extinct at the 3rd generation?

Solution. This would be

$$(\phi \circ \phi \circ \phi \circ \phi)(0).$$

This is

$$\begin{aligned}(\phi \circ \phi \circ \phi)(0) &= (\phi \circ \phi)(0.5) \\ &= \phi(0.5 + 0.5 \cdot 0.5^2) \\ &= \phi(0.625) \\ &= 0.5 + 0.5 \cdot 0.625^2 \\ &\approx \boxed{0.6953}\end{aligned}$$

275. Suppose $\mathbb{P}(R = 0) = 0.3$ and $\mathbb{P}(R = 2) = 0.7$ is the distribution for the number of children in a branching process. Find the probability of extinction starting with 1 person in the population.

Solution. To find this, solve

$$x = 0.3 + 0.7x^2$$

to get the root in $(0, 1)$, which is $\boxed{0.4285}$

277.

Suppose that in a branching process, each individual has a number of children that is Poisson distributed with parameter λ. Such a distribution has generating function equal to $\exp(-\lambda(1-x))$.

a) If $\lambda = 1$, find the probability that the population is extinct at the 2nd generation.

b) If $\lambda = 0.6$, find the probability that the population is extinct at the 2nd generation.

Solution. a) Using the generating function

$$\begin{aligned}(\phi \circ \phi)(0) &= \phi(\exp(-1(1-0))) \\ &= \exp(-1(1 - exp(-1))) \\ &= \boxed{0.5314\ldots}\end{aligned}$$

b) Using the generating function

$$\begin{aligned}(\phi \circ \phi)(0) &= \phi(\exp(-0.6(1-0))) \\ &= \exp(-0.6(1 - \exp(-0.6))) \\ &= \boxed{0.7628\ldots}\end{aligned}$$

Chapter 33

279.

Let B_t be a standard Brownian motion.

a) What is the mean of $B_8 - B_4$?

b) What is the variance of $B_8 - B_4$?

Solution. a) Increments of Brownian motion have mean $\boxed{0}$.

b) The variance of an increment of Brownian motion is the length of time of the interval, so in this case $8 - 4 = \boxed{4}$.

281. Let B_t be a standard Brownian motion. What is the distribution of $B_4 - B_{1.5}$?

Solution. Increments of Brownian motion have a normal distribution with mean 0 and variance equal to the length of the interval. So in this case, the result is $\boxed{\mathsf{N}(0, 2.5)}$.

283. Let B_t be a standard Brownian motion. What is the probability that B_t is continuous?

Solution. Standard Brownian motion is continuous with probability 1?

285. Suppose B_t is standard Brownian motion. Find $\mathbb{P}(B_5 > 2)$.

Solution. Since $B_0 = 0$,

$$\mathbb{P}(B_5 > 2) = \mathbb{P}(B_5 - B_0 > 2) = 1 - \mathbb{P}(\sqrt{5} \cdot Z \leq 2),$$

where $Z \sim \mathsf{N}(0, 1)$ is a standard normal random variable.

This can be found in R with `r 1 - pnorm(2 / sqrt(5))`

So the answer is $\boxed{0.1855\ldots}$.

Chapter 34

287. Suppose $B_4 = -2.4$ for B_t standard Brownian motion. What is the distribution of B_5?

Solution. Because $B_5 - B_4 \sim \mathsf{N}(0, 1)$, $\boxed{B_5 \sim \mathsf{N}(-2.4, 1)}$.

289. Suppose $B_1 = 1.3$ and $B_4 = -2.4$ where B_t is standard Brownian motion. What is the distribution of B_2?

Solution. This is an interpolation problem. Here

$$\lambda = \frac{2-1}{4-1} = 1/3$$

Hence the mean is

$$1.3 + (1/3)(-2.4 - 1.3) = 0.06666\ldots.$$

The variance is

$$\frac{1}{3} \cdot \frac{2}{3}(4 - 1) = 0.6666\ldots.$$

The distribution is normal, so

$$[B_2|B_1 = 1.3, B_4 = -2.4] \boxed{\mathsf{N}(0.06666, 0.6666)}.$$

291. Suppose $B_4 = -2.4$ for B_t standard Brownian motion. What is the distribution of B_1?

Solution. Because $B_0 = 0$, this is an interpolation problem. First find λ:

$$\lambda = \frac{1-0}{4-0} = 1/4.$$

This makes $\lambda B_0 + (1-\lambda)B_4 = (3/4)(-2.4) = -1.8$ and $\lambda(1-\lambda)(4-0) = 0.75$.

$$[B_1|B_4 = -2.4, B_0 = 0] \sim \boxed{\mathsf{N}(-1.8, 0.75)}.$$

Chapter 35

293. Suppose that a continuous time Markov chain has infinitesimal generator

$$\begin{pmatrix} -3.4 & 3.4 \\ 1.2 & -1.2 \end{pmatrix}.$$

Approximate $\mathbb{P}(X_{0.0007} = 2|X_0 = 1)$ using the first order approximation.

Solution. The first order approximation is

$$A(1,2)(0.0007) = 3.4(0.0007) = \boxed{0.00238}.$$

295. Suppose that a continuous time Markov chain has infinitesimal generator

$$\begin{pmatrix} -3.4 & 3.4 \\ 1.2 & -1.2 \end{pmatrix}.$$

What is $\mathbb{P}(X_{0.7} = 2|X_0 = 1)$?

Solution. The transition matrix at time 4.1 is

$$\exp(0.7A),$$

which can be found in R using the `expm` package.

```
library(expm)
```

The function call is then

```
expm(0.7 * matrix(c(-3.4, 3.4,
                    1.2, -1.2),
                    nrow = 2,
                    byrow = TRUE))
```

```
##             [,1]      [,2]
## [1,] 0.2904016 0.7095984
## [2,] 0.2504465 0.7495535
```

The $(1, 2)$ entry then gives the probability: $\boxed{0.7095\dots}$.

297. Suppose that a continuous time Markov chain has infinitesimal generator

$$\begin{pmatrix} -3.4 & 3.4 \\ 1.2 & -1.2 \end{pmatrix}.$$

Find the eigenvector associated with eigenvalue 0 and normalize to get the unique stationary distribution

Solution. In R, the left eigenvectors and their eigenvalues can be found with:

```
A <- matrix(c(-3.4, 3.4,
              1.2, -1.2),
              nrow = 2,
              byrow = TRUE)
eigen(t(A))
```

```
## eigen() decomposition
## $values
## [1] -4.600000e+00 -2.220446e-16
##
## $vectors
##             [,1]       [,2]
## [1,] -0.7071068 -0.3328201
## [2,]  0.7071068 -0.9429903
```

One eigenvalue is -4.6 while the second is below machine epsilon, so that is the one being sought. Normalize:

```
v <- eigen(t(A))$vectors[, 2]
v / sum(v)
```

```
## [1] 0.2608696 0.7391304
```

So the stationary distribution is about

$$\boxed{(0.2608\dots, 0.7391\dots)}.$$

Use the Ergodic Theorem to check!

```r
expm(20 * A)
```

```
##              [,1]      [,2]
## [1,] 0.2608696 0.7391304
## [2,] 0.2608696 0.7391304
```

Looking good!

299.

Suppose a CTMC has infinitesimal generator

$$B = \begin{pmatrix} -10.3 & 4.1 & 6.2 \\ 5.0 & -9.3 & 4.3 \\ 6.2 & 3.7 & -9.9 \end{pmatrix}$$

and states $\{a, b, c\}$.

a. If the current state is a, what is the chance that the next jump is to b instead of c?

b. At what rate is the chain jumping away from state a?

Solution. a. The probability of where the state jumps to is proportional to the rate of moving to that state. So the answer is

$$\frac{4.1}{4.1 + 6.2} = \frac{4.1}{10.3} = \boxed{0.3980\ldots}.$$

b. The rate at which a chain jumps away from a state is the sum of the rates leaving the state. For state a, this is $4.1 + 6.2 = \boxed{10.30}$.

Chapter 36

301. Arrivals to a queue are modeled as a Poisson point process with constant rate. If on average their are 3 arrivals in the first two hours, what is the rate of the PPP?

Solution. This is 3 arrivals per 2 hours, or $3/2 = \boxed{1.500 \text{ per hour}}$.

303. Arrivals to a queue are modeled as a Poisson point process of rate 2.3/hour. What is the chance that the third arrival occurs more than 10 minutes after the second arrival?

Solution. Ten minutes is one sixth of an hour, so in terms of arrival times, the question is asking to find

$$\mathbb{P}(T_3 - T_2 > 1/6 \text{ hours})?$$

Since $T_3 - T_2 \sim \text{Exp}(2.3/\text{hour})$, the survival function gives

$$\mathbb{P}(T_3 - T_2 > 1/6 \text{ hours}) = \exp(-(2.3/\text{hour})(1/6) \text{ hour}) = \exp(-2.3/6) = \boxed{0.6815\ldots}.$$

(Note that the time units cancel out in the survival function.)

305. Arrivals to a queue are modeled as a Poisson point process of rate 2.3/hour. What is the chance that there is exactly 1 arrival in the first five minutes?

Solution. This will be the probability that a Poisson with parameter $(2.3)(1/20)$ equals 1. Note that 5 minutes equals $1/20$ of an hour. This will be

$$(2.3/20)^1 \exp(-2.3/20)/1! = \boxed{0.1025\ldots}.$$

307.

Suppose that the arrivals of airport shuttles at a particular stop follow a Poisson process of rate $1/[10 \text{ min}]$.

a. On average, how many shuttles will arrive in an hour?

b. What is the chance that there are no shuttles in the first 20 minutes?

c. What is the chance that the second shuttle arrives somewhere in $[15, 25]$ minutes?

Solution. a. One hour is 60 minutes, so on average there are $60/10 = \boxed{6}$ arrivals.

b. In the first 20 minutes, there are on average $20/10 = 2$ arrivals. Therefore, the chance of zero is

$$\exp(-2) = \boxed{0.1353\ldots}.$$

c. Use the gamma distribution with parameter 2 and 0.1 to get

$$\int_{15}^{25} 0.1^2 t \exp(-0.1t)/(2-1)! = \boxed{0.2705\ldots}.$$

309. A webpage receives hits at rate 4 per minute. Suppose that five hits are received in the first five minutes. What is the chance that exactly three of them arrived in the first two minutes?

Solution. Each hit is uniform over $[0, 5]$, and so has a $(2-0)/(5-0) = 0.4$ chance of falling into $[0, 2]$. Each hit is independent of the others, and so the number that fall in $[0, 2]$ is binomial with parameters 5 and 0.4. Therefore, the answer is

$$\binom{5}{3}(0.4)^3(0.6)^2 = \boxed{0.2304}.$$

Chapter 37

311. The stationary distribution π is a left eigenvector for the intfitesimal generator with what eigenvalue?

Solution. This is $\boxed{0}$.

312. For a continuous time Markov chain with states $\{a, b, c\}$ and infinitesimal generator

$$\begin{pmatrix} -1 & 1 & 0 \\ 0.5 & -2 & 1.5 \\ 0.4 & 0.2 & -0.6 \end{pmatrix},$$

what is $p(b, c)$ in the discrete time underlying Markov chain?

Solution. This will be

$$\frac{1.5}{2} = \boxed{0.7500}$$

313. Consider a Markov chain on $\{0, 1, 2, \ldots\}$ which $\lambda(i, i+1) = 2$ for $i \geq 0$ and $\lambda(i, i-1) = 3$ for $i \geq 1$. Show that this is an irreducible continuous time Harris chain.

Solution. Note $\lambda(i, i) \neq 0$ for every state i in the chain.

Consider two states $i < j$ in the chain. Then there is a path from i to j that is $(i, i+1, \ldots, j)$ where the λ value is positive for every step in the path, so $\lambda(k, k+1)/-\lambda(k, k) > 0$ as well.

Similarly, $(j, j-1, \ldots, i)$ is a path from j down to i. Hence i and j communicate for all $i < j$, making the chain irreducible.

315. Consider the continuous time Markov chain with infinitesimal generator

$$A = \begin{pmatrix} -2.3 & 1.1 & 1.2 & 0 \\ 3.2 & -5.0 & 1.0 & 0.8 \\ 0.6 & 0.6 & -3.2 & 2.0 \\ 1.1 & 0.5 & 2.1 & -3.7 \end{pmatrix}$$

Find the Jordan Normal Form of this generator.

Solution. Using

```
jordan form of {{-2.3, 1.1, 1.2, 0},
{3.2, -5.0, 1.0, 0.8},
{0.6, 0.6, -3.2, 2.0},
{1.1, 0.5, 2.1, -3.7}}
```

at wolframalpha.com gives for this matrix A

$$A = SJS^{-1},$$

where

$$S \approx \begin{pmatrix} -0.9040 & 1 & -0.3267 & -1.351 \\ 4.661 & 1 & 2.455 & -0.9528 \\ -1.636 & 1 & -1.340 & 1.385 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$J \approx \begin{pmatrix} -5.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -5.645 & 0 \\ 0 & 0 & 0 & -2.754 \end{pmatrix}$$

$$S^{-1} \approx \begin{pmatrix} -0.2905 & 0.4358 & 0.4358 & -0.5810 \\ 0.3624 & 0.1360 & 0.3064 & 0.1950 \\ 0.2708 & -0.5024 & -0.8934 & 1.124 \\ -0.3428 & -0.06941 & 0.1512 & 0.2610 \end{pmatrix}$$

317. For the continuous time Markov chain with infinitesimal generator

$$A = \begin{pmatrix} -2.3 & 1.1 & 1.2 & 0 \\ 3.2 & -5.0 & 1.0 & 0.8 \\ 0.6 & 0.6 & -3.2 & 2.0 \\ 1.1 & 0.5 & 2.1 & -3.7 \end{pmatrix},$$

find the limiting distribution by calculating $\exp(tA)$ for t large.

Solution. In R, this can be done with the `expm()` function,

```
A <- matrix(c(-2.3, 1.1, 1.2, 0,
              3.2, -5.0, 1.0, 0.8,
              0.6, 0.6, -3.2, 2.0,
              1.1, 0.5, 2.1, -3.7),
              byrow = TRUE,
              nrow = 4)
expm::expm(100 * A)
```

```
##           [,1]      [,2]     [,3]      [,4]
## [1,] 0.3624792 0.1360239 0.306442 0.1950549
## [2,] 0.3624792 0.1360239 0.306442 0.1950549
## [3,] 0.3624792 0.1360239 0.306442 0.1950549
## [4,] 0.3624792 0.1360239 0.306442 0.1950549
```

So the limiting distribution is about

$$\boxed{(0.3624, 0.1360, 0.3064, 0.1950)}.$$

Chapter 38

319. For the CTMC model of exponential growth $y' = ky$, if $\lambda(2,3) = 10$, what is k?

Solution. Here $k(2) = 10$, so $\boxed{k = 5}$.

321.

For the state space $\{0, 1, 2, \ldots\}$, and CTMC where $\lambda(i, i+1) = 3i$ for all i, answer the following.

a) What is the expected time spent in state 4 before jumping?

b) Will the state of the chain converge?

Solution. a) Since the time until jump is an exponetial of rate $(3)(4) = 12$, the expected time until jumping is $1/12 = \boxed{0.08333}$.

322. Suppose that from state (i, j) there are positive rate edges to $(i+1, j-1)$ and $(i-1, j+1)$. What simple linear function of the i and j is being conserved here?

Solution. In this case $\boxed{i + j}$ is not changing.

Chapter 39

324. Suppose $\mathbb{E}[X_i^2] \leq 5$ for all $i \in \{0, 1, 2, \ldots\}$. Are the $\{X_i\}$ uniformly integrable?

Solution. $\boxed{\text{Yes}}$, because they are square integrable, and square integrability implies uniform integrability.

325. Suppose $\mathbb{E}[Y_i^2] = i$ for all $i \in \{0, 1, 2, \ldots\}$. Are the $\{Y_i\}$ uniformly integrable?

Solution. $\boxed{\text{No}}$, because for any value of M, $\mathbb{E}[Y_{M+1}^2] > M$.

326. Suppose $\mathbb{E}[A^2] \leq 10$. Show that $\mathbb{E}[|A|\mathbb{I}(|A| > 4)] \leq 10/4 = 2.5$.

Solution. If $\mathbb{I}(|A| > 4) = 1$, then $|A|/4 \geq 1$, and it is always true that $|A|/4 \geq 0$ so even when $\mathbb{I}(|A| > 4) = 0$ it holds that

$$\mathbb{I}(|A| > 4) \leq \frac{|A|}{4}.$$

Therefore

$$|A|\mathbb{I}(|A| > 4) \leq |A|\frac{|A|}{4} = A^2/4.$$

Hence

$$\mathbb{E}[|A|\mathbb{I}(|A| > 4)] \leq \mathbb{E}[A^2/4] \leq 10/4 = 2.5,$$

completing the proof.

Chapter 40

327. Suppose $\mathbb{P}(X = 0) = 0.3$ and $\mathbb{P}(X = 2) = 0.7$. Say $X_1, X_2, \ldots$ with the same distribution as X. What is $\mathbb{E}[T]$ for

$$T = \inf\{t : X_1 + X_2 + \cdots + X_t = 10\}?$$

Solution. Here

$$\sum_{i=1}^{T} X_i = 10,$$

and the $X_i \geq 0$, which means Wald's Identity can be used to say that

$$\mathbb{E}[X]\mathbb{E}[T] = [(2)(0.7) + (0)(0.3)]\mathbb{E}[T] = 10,$$

and

$$\mathbb{E}[T] = \frac{10}{1.4} = \boxed{7.142\ldots}.$$

329. Suppose $X \in [0, 1]$, and $\mathbb{E}[X] = 0.7$. Say $X_1, X_2, \ldots$ with the same distribution as X. Give the tightest bound you can on $\mathbb{E}[T]$ for

$$T = \inf\{t : X_1 + X_2 + \cdots + X_t \geq 10\}?$$

Solution. Because $X \in [0,1]$, it holds that

$$\sum_{i=1}^{T} X_i \in [10, 11).$$

Therefore, because the X_i are positive Wald's Identity applies, and

$$\mathbb{E}\left[\sum_{i=1}^{T} X_i\right] = \mathbb{E}[T]\mathbb{E}[X],$$

so

$$0.7\mathbb{E}[T] \in [10, 11).$$

and

$$\mathbb{E}[T] \in \left[\frac{100}{7}, \frac{110}{7}\right].$$

331. Let $B_1, B_2, \ldots$ be iid Bernoulli random variables with mean p. Given that a geometric random variable can be defined as

$$G = \inf\{t : B_1 + \cdots B_t = 1\},$$

use Wald's Equality to find $\mathbb{E}[G]$.

Solution. Here the $B_i \geq 0$, so Wald can be applied. Note that

$$\sum_{i=1}^{G} B_i = 1,$$

so

$$\mathbb{E}[B_i]\mathbb{E}[T] = 1,$$

and since $\mathbb{E}[B_i] = p$, the result is $\boxed{\mathbb{E}[T] = 1/p}$.

333. Suppose $X = -1$ with probability 0.6 and $X = 1$ with probability 0.4. For $X_1, X_2, \ldots$ iid X, find the expected value of T, where

$$T = \inf\{t : X_1 + \cdots + X_t = -10\}.$$

Solution. The X_i are not nonnegative, so to use Wald it is first necessary to show that $\mathbb{E}[T] < \infty$.

Let

$$T_a = \inf\{t : X_1 + \cdots + X_t \in \{-10, a\}\}.$$

Then $|T_a| \leq \max\{10, a\}$, so T_a is integrable and Wald's Equation gives

$$\mathbb{E}\left(\sum_{i=1}^{T_a} X_i\right) = \mathbb{E}[T_a]\mathbb{E}[X_i].$$

Since the left hand side is at least -10 and $\mathbb{E}[X_i] = -1(0.6) + 1(0.4) = -0.2$,

$$-10 \leq \mathbb{E}[T_a](-0.2),$$

or

$$\mathbb{E}[T_a] \leq 50.$$

The T_a are increasing and converge to T, hence the Monotone Convergence Theorem gives that $\mathbb{E}[T] \leq 50$. Therefore, Wald can be applied to say:

$$\mathbb{E}\left(\sum_{i=1}^{T} X_i\right) = \mathbb{E}[T]\mathbb{E}[X_i].$$

This gives

$$-10 = \mathbb{E}[T]\mathbb{E}[X_i],$$

or

$$\mathbb{E}[T] = -10/(-0.2) = \boxed{50}.$$

Chapter 41

335. What is

$$\int_{t=0}^{10} 4 \, dS_t$$

where $S_t = 50 + 60\mathbb{1}(t \geq 4)$.

Solution. The 4 shares cost $(4)(50) = 200$ at time 0 and $(4)(110) = 440$ at time 10, so the profit is $\boxed{220}$.

337. What is the distribution of

$$W = \int_{t=0}^{20} [3\mathbb{1}(t \in [0, 10)) + 5\mathbb{1}(t \in [10, 20])] \, dB_t.$$

Solution. This is a simple strategy, so the result is

$$3(B_{10} - B_0) + 5(B_{20} - B_{10}).$$

Since $B_{10} - B_0 \sim \mathsf{N}(0, 10)$ and $B_{20} - B_{10} \sim \mathsf{N}(0, 10)$ are independent, this linear combination will also be normally distributed with mean $3(0) + 5(0) = 0$ and variance $3^2(10) + 5^2(10) = 340$. So

$$W \sim \boxed{\mathsf{N}(0, 340)}.$$

339. What is

$$\mathbb{E}\left(\int_{t=0}^{30} (t + B_t) \, dB_t\right)^2.$$

Solution. Following the rules for second moments of Ito integrals:

$$\begin{aligned}\mathbb{E}\left(\int_{t=0}^{30}(t+B_t)\,dB_t\right)^2 &= \int_{t=0}^{30}\mathbb{E}(t+B_t)^2\,dt\\ &= \int_{t=0}^{30}\mathbb{E}(t^2+2tB_t+B_t^2)\,dt\\ &= \int_{t=0}^{30} t^2+0+t)\,dt\\ &= (t^3/3+t^2/2)|_0^{30}\\ &= \boxed{9450}.\end{aligned}$$

Chapter 42

341.

Say $f(t,x) = 4\exp(3t-x)$.

a) Find $\partial f/\partial t$.

b) Find $\partial f/\partial x$.

c) Find $\partial^2 f/\partial x^2$.

Solution. a) This is $\boxed{12\exp(3t-x)}$.

b) This is $\boxed{-4\exp(3t-x)}$.

c) This is $\boxed{4\exp(3t-x)}$.

342. Say $f(t,x) = 4\exp(3t-x)$.

Find $df(t,B_t)$ in terms of dt and dB_t.

Solution. Using Ito's Formula:

$$df(t,B_t) = \left[\frac{\partial f}{\partial t}(t,B_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,B_t)\right]\,dt + \frac{\partial f}{\partial x}(t,B_t)\,dB_t,$$

and plugging in these partial derivatives gives

$$\boxed{df(t,B_t) = 14\exp(3t-B_t)\,dt - 4\exp(3t-B_t)\,dB_t.}$$

344. For $R_t = B_t^2$, write

$$\int_{t=0}^{T} Y_t\,dR_t$$

as a stochastic integral with respect to dB_t using Ito's Lemma.

Solution. Here $f(t,b) = b^2$, so

$$\frac{\partial f}{\partial t} = 0, \ \frac{\partial f}{\partial b} = 2b, \ \frac{\partial^2 f}{\partial b^2} = 2,$$

so

$$dR_t = (1/2)2B_t\,dt + 2dB_t$$

and

$$\int_{t=0}^{T} Y_t\,dR_t = \boxed{\int_{t=0}^{T} Y_t B_t\,dt + \int_{t=0}^{T} 2Y_t dB_t}.$$

Chapter 43

346. What is the stationary distribution of a Markov chain whose transition matrix is symmetric?

Solution. This is $\boxed{\text{uniform}}$ over the state space.

348. Suppose a Markov chain has two states a and b, and that $p(a,b) = 0.6$ while $p(b,a) = 0.4$. Show that this chain is reversible with respect to the probability vector $(0.4, 0.6)$.

Solution. Here

$$\pi(a)p(a,b) = (0.6)(0.4) = p(b,a)\pi(b),$$

so the chain is reversible.

349. Create a Markov chain whose limiting distribution is uniform over $\{a,b,c,d,e\}$.

Solution. There are many solutions, but not every chain works. From the ergodic theorem the limiting distribution will be stationary if the chain is irreducible and aperiodic. Aperiodicity can be enforced by putting a holding probability of $1/2$ on each state. The stationary distribution will be uniform if the transition matrix is symmetric. So one example is

$$\boxed{\begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}}$$

Chapter 44

351.

Suppose a chain has $p(a,b) = p(b,a) = 0.4$. The goal is to create a chain with stationary measure $w(a) = 4$ and $w(b) = 5$ using Metropolis-Hastings.

a) What is the chance of accepting a move from a to b?

b) What is the chance of accepting a move from b to a?

Solution. a) This is

$$\min\left(\frac{5}{4}, 1\right) = \boxed{1}$$

b) This is

$$\min\left(\frac{4}{5}, 1\right) = \boxed{0.8000}$$

352.

Suppose a chain has $p(a, b) = 0.6$ and $p(b, a) = 0.5$. The goal is to create a chain with $w(a) = 4$ and $w(b) = 5$ using Metropolis-Hastings.

a) What is the chance of accepting a move from a to b?

b) What is the chance of accepting a move from b to a?

Solution. a) This is

$$\min\left(\frac{(5)(0.5)}{(4)(0.6)}, 1\right) = \boxed{1}$$

b) This is

$$\min\left(\frac{(4)(0.6)}{(5)(0.5)}, 1\right) = \boxed{0.9600}$$

Chapter 45

353. Consider an $M/M/1$ queue with $\lambda = 2$ and $\mu = 3$. What is the stationary distribution for the chain?

Solution. Using the formula for stationary distribution gives

$$\pi(i) = \frac{(\lambda/\mu)^i}{1 - \lambda/\mu} = \boxed{\frac{1}{3}\left(\frac{2}{3}\right)^i}.$$

355. Prove that

$$\sum_{i=1}^{\infty} \frac{1}{i^{1.5}} < \infty.$$

Solution. The integral test says that if

$$\int_{x=1}^{\infty} \frac{1}{x^{1.5}}\, dx$$

is finite, then so is the sum. An antiderivative of $x^{-1.5}$ is $x^{-0.5}/(-0.5)$. Hence

$$\int_{x=1}^{\infty} \frac{1}{x^{1.5}}\, dx = \lim_{b\to\infty} 2 \cdot 1^{-0.5} - 2 \cdot b^{-0.5} < \infty.$$

Therefore the sum $\sum_{i=1}^{\infty} i^{-1.5} < \infty$ as well.