Lectures on Markov Chains

David M. McClendon

Department of Mathematics
Ferris State University

2016 edition

Contents

Chapter 1

Markov chains

1.1 The definition of a Markov chain

In Math 416, our primary goal is to describe probabilistic models which simulate real-world phenomena. As with all modeling problems, there is a "Goldilocks" issue:

- If the model is too simple,

- if the model is too complex,

In applied probability, we want to model phenomena which evolve randomly. The mathematical object which describes such a situation is a "stochastic process":

Definition 1.1 *A* **stochastic process** $\{X_t : t \in \mathcal{I}\}$ *is a collection of random variables indexed by t. The set $\mathcal{I}$ of values of t is called the* **index set** *of the stochastic process, and members of $\mathcal{I}$ are called* **times**. *We assume that each X_t has the same range, and we denote this common range by $\mathcal{S}$. $\mathcal{S}$ is called the* **state space** *of the process, and elements of $\mathcal{S}$ are called* **states**.

Remark: $\{X_t\}$ refers to the entire process (i.e. at all times t), whereas X_t is a single random variable (i.e. refers to the state of the process at a fixed time t).

Remark: Think of X_t as recording your " position" or "state" at time t. As t changes, you think of "moving" or "changing states". This process of "moving" will be random, and modeled using probability theory.

Almost always, the index set is $\{0, 1, 2, 3, ...\}$ or $\mathbb{Z}$ (in which case we call the stochastic process a **discrete-time** process, or the index set is $[0, \infty)$ or $\mathbb{R}$ (in which case we call the stochastic process a **continuous-time** process) . The first three chapters of these notes focus on discrete-time processes; chapters 4 and 5 center on continuous-time processes.

In Math 414, we encountered the three most basic examples of stochastic processes:

1. The **Bernoulli process**, a discrete-time process $\{X_t\}$ with state space $\mathbb{N}$ where X_t is the number of successes in the first t trials of a Bernoulli experiment. Probabilities associated to a Bernoulli process are completely determined by a number $p \in (0, 1)$ which gives the probability of success on any one trial.

2. The **Poisson process**, a continuous-time process $\{X_t\}$ with state space $\mathbb{N}$ where X_t is the number of successes in the first t units of time. Probabilities associated to a Poisson process are completely determined by a number $\lambda > 0$ called the **rate** of the process.

3. **i.i.d. processes** are discrete-time processes $\{X_t\}$ where each X_t has the same density and all the X_t are mutually independent. Averages of random variables from these processes are approximately normal by the Central Limit Theorem.

We now define a class of processes which encompasses the three examples above and much more:

Definition 1.2 *Let $\{X_t\}$ be a stochastic process with state space $\mathcal{S}$. $\{X_t\}$ is said to have the* **Markov property** *if for any times $t_1 < t_2 < ... < t_n$ and any states $x_1, ..., x_n \in \mathcal{S}$,*

$$P(X_{t_n} = x_n \,|\, X_{t_1} = x_1, X_{t_2} = x_2, ..., X_{t_{n-1}} = x_{n-1}) = P(X_{t_n} = x_n \,|\, X_{t_{n-1}} = x_{n-1}).$$

A discrete-time stochastic process with finite or countable state space that has the Markov property is called a **Markov chain**.

To understand this definition, think of time t_n as the "present" and the times $t_1 < ... < t_{n-1}$ as being times in the "past". If a process has the Markov property, then given some values of the process in the past, the probability of the present value of the process **depends only on the most recent given information**, i.e. on $X_{t_{n-1}}$.

Note: Bernoulli processes, Poisson processes and i.i.d. processes all have the Markov property.

Question: What determines a Markov chain? In other words, what makes one Markov chain different from another one?

Answer:

1. The **state space** $\mathcal{S}$ of the Markov chain
   (Usually $\mathcal{S}$ is labelled $\{1, ..., d\}$ or $\{0, 1\}$ or $\{0, 1, 2, ...\}$ or $\mathbb{N}$ or $\mathbb{Z}$, etc.)

2. The **initial distribution** of the r.v. X_0, denoted π_0:

   $$\pi_0(x) = P(X_0 = x) \text{ for all } x \in \mathcal{S}$$

   $\pi_0(x)$ is the probability the chain starts in state x.

3. **Transition probabilities**, denoted $P(x, y)$ or $P_{x,y}$ or P_{xy}:

   $$P(x, y) = P_{xy} = P_{x,y} = P(X_t = y \,|\, X_{t-1} = x)$$

   $P(x, y)$ is the probability, given that you are in state x at a certain time, that you are in state y at the next time.

   Technically, transition probabilities depend not only on x and y but on t, but throughout our study of Markov chains we will assume (often without stating it) that the transition probabilities do not depend on t; that is, that they have the following property:

Definition 1.3 *Let $\{X_t\}$ be a Markov chain. We say the transition probabilities of $\{X_t\}$ are* **time homogeneous** *if for all $s, t \in \mathcal{S}$,*

$$P(X_t = y \,|\, X_{t-1} = x) = P(X_s = y \,|X_{s-1} = x),$$

i.e. that the transition probabilities depend only on x and y (and not on t).

The reason the transition probabilities are sufficient to describe a Markov chain is that by the Markov property,

$$P(X_t = x_t \,|\, X_0 = x_0, ..., X_{t-1} = x_{t-1}) = P(X_t = x_t \,|\, X_{t-1} = x_{t-1}) = P(x_{t-1}, x_t).$$

In other words, conditional probabilities of this type **depend only on the most recent transition** and ignore any past behavior in the chain.

1.2 Basic examples of Markov chains

1. **i.i.d. process** (of discrete r.v.s)
   - *State space:* $\mathcal{S}$
   - *Initial distribution:*
   - *Transition probabilities:*

     $P(x, y) = P(X_t = y \,|X_{t-1} = x) =$

2. **Bernoulli process**
   - *State space:* $\mathcal{S} = \mathbb{N} = \{0, 1, 2, 3, ...\}$.
   - *Initial distribution:*
   - *Transition probabilities:*

$$P(x, y) = P(X_t = y \,|X_{t-1} = x) = \begin{cases} \\ \\ \end{cases}$$

We represent these transition probabilities with the following picture:

The above picture generalizes: Every Markov chain can be thought of as a random walk on a graph as follows:

> **Definition 1.4** *A* **directed graph** *is a finite or countable set of points called* **nodes***, usually labelled by integers, together with "arrows" from one point to another, such that given two nodes x and y, there is either zero or one arrow going directly from x to y.*

Example:

Example:

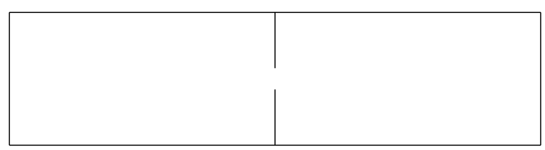
Nonexample:

If one labels the arrow from x to y with a number $P(x, y)$ such that for each node x, $\sum_y P(x, y) = 1$, then the directed graph represents a Markov chain, where the nodes are the states and the arrows represent the transitions. If you are in state x at time $t - 1$ (i.e. if $X_{t-1} = x$), then to determine your state X_t at time t, you follow one of the arrows starting at x (with probabilities as indicated on the arrows which start at x).

Example 3: Make a series of $\$1$ bets in a casino, where you are 60% likely to win and 40% likely to lose each game. Let X_t be your bankroll after the t^{th} bet.

3. **Ehrenfest chain**

Suppose you start with a container that looks like this:



Notice that the container has two "chambers", with only a small slit open between them. Suppose there are a total of d objects (think of the objects as molecules of gas) in the container. Over each unit of time, one and only one of these objects (chosen uniformly from the objects) moves through the slit to the opposite chamber. For each t, let X_t be the number of objects in the left-hand chamber. $\{X_t\}$ is a Markov chain called the **Ehrenfest chain**; it models the diffusion of gases across a semi-permeable membrane.

As an example, let's suppose $d = 3$, and suppose all the objects start in the left-hand chamber.

time t	object which switches chambers (chosen uniformly)	list of objects in the left-hand chamber after the switch	X_t

- *State space of the Ehrenfest chain:*

- *Transition probabilities:*

$$P(x,y) = P(X_t = y \,|\, X_{t-1} = x) = \begin{cases} \\ \\ \end{cases}$$

- *Directed graph:*

1.3 Markov chains with finite state space

Suppose $\{X_t\}$ is a Markov chain with state space $\mathcal{S} = \{1, ..., d\}$. Let $\pi_0 : \mathcal{S} \to [0,1]$ give the initial distribution (i.e. $\pi_0(x) = P(X_0 = x)$) and let the transition probabilities be $P_{x,y}$ ($P_{x,y}$ is the same thing as $P(x,y)$).

If the state space is finite, the most convenient representation of the chain's transition probabilities is in a matrix:

Definition 1.5 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S} = \{1, ..., d\}$. The $d \times d$ matrix of transition probabilities*

$$P = \begin{pmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,d} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ P_{d,1} & P_{d,2} & \cdots & P_{d,d} \end{pmatrix}_{d \times d}$$

is called the **transition matrix** *of the Markov chain.*

A natural question to ask is what matrices can be transition matrices of a Markov chain. Notice that all the entries of P must be nonnegative, and that the rows of P must sum to 1, since they represent the probabilities associated to all the places x can go in 1 unit of time.

Definition 1.6 *A $d \times d$ matrix of real numbers P is called a* **stochastic matrix** *if*

1. *P has only nonnegative entries, i.e. $P_{x,y} \geq 0$ for all $x, y \in \{1, ..., d\}$; and*
2. *each row of P sums to 1, i.e. for every $x \in \{1, ..., d\}$, $\sum_{y=1}^{d} P_{x,y} = 1$.*

Theorem 1.7 (Transition matrices are stochastic) *A $d \times d$ matrix of real numbers P is the transition matrix of a Markov chain if and only if it is a stochastic matrix.*

We can answer almost any question about a finite state space Markov chain by performing some calculation related to the transition matrix.

n-step transition probabilities

Definition 1.8 *Let $\{X_t\}$ be a Markov chain and let $x, y \in \mathcal{S}$. Define the* **n-step transition probability** *from x to y by*

$$P^n(x, y) = P(X_{t+n} = y \,|\, X_t = x).$$

(Since we are assuming the transition probabilities are time homogeneous, these numbers will not depend on t.)

$P^n(x, y)$ measures the probability, given that you are in state x, that you are in state y exactly n units of time from now.

Theorem 1.9 *Let $\{X_t\}$ be a Markov chain with finite state space $\mathcal{S} = \{1, ..., d\}$. If P is the transition matrix of $\{X_t\}$, then for every $x, y \in \mathcal{S}$ and every $n \in \{0, 1, 2, 3, ...\}$, we have*

$$P^n(x, y) = (P^n)_{x,y},$$

the $(x, y)-$entry of the matrix P^n.

PROOF By time homogeneity,

$$P^n(x, y) = P(X_n = y \,|\, X_0 = x)$$

.

Time n distributions

Definition 1.10 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. A* **distribution** *on $\mathcal{S}$ is a probability measure π on $(\mathcal{S}, 2^{\mathcal{S}})$, i.e. a function $\pi : \mathcal{S} \to [0,1]$ such that $\sum_{x \in \mathcal{S}} \pi(x) = 1$.*

We often denote distributions as row vectors, i.e. if $\mathcal{S} = \{1, 2, ..., d\}$ then

$$\pi = \begin{pmatrix} \pi(1) & \pi(2) & \cdots & \pi(d) \end{pmatrix}$$

The coordinates of any distribution must be nonnegative and sum to 1.

Definition 1.11 *Let $\{X_t\}$ be a Markov chain. The* **time n distribution** *of the Markov chain, denoted π_n, is the distribution π_n defined by*

$$\pi_n(x) = P(X_n = x).$$

$\pi_n(x)$ gives the probability that at time n, you are in state x.

Theorem 1.12 *Let $\{X_t\}$ be a Markov chain with finite state space $\mathcal{S} = \{1, ..., d\}$. If*

$$\pi_0 = \begin{pmatrix} \pi_0(1) & \pi_0(2) & \cdots & \pi_0(d) \end{pmatrix}_{1 \times d}$$

is the initial distribution of $\{X_t\}$ (written as a row vector), and if P is the transition matrix of $\{X_t\}$, then for every $x, y \in \mathcal{S}$ and every $n \in \mathcal{I}$, we have

$$\pi_n(y) = (\pi_0 P^n)_y,$$

the $y^{th}-$entry of the $(1 \times d)$ row vector $\pi_0 P^n$.

PROOF This is a direct calculation:

$$\begin{aligned}\pi_n(y) = P(X_n = y) &= \sum_{x\in\mathcal{S}} P(X_n = y \mid X_0 = x)P(X_0 = x) \quad \text{(LTP)}\\ &= \sum_{x\in\mathcal{S}} (P^n)_{x,y}\, \pi_0(x) \quad \text{(Theorem 1.9)}\\ &= \sum_{x\in\mathcal{S}} \pi_0(x)\, (P^n)_{x,y}\\ &= [\pi_0 P^n]_y \quad \text{(def'n of matrix multiplication)} \quad \square\end{aligned}$$

Example: Consider the Markov chain with state space $\{0, 1\}$ whose transition matrix is

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

and whose initial distribution is uniform.

1. Sketch the directed graph representing this Markov chain.
2. Find the distribution of X_2.
3. Find $P(X_3 = 0)$.
4. Find $P(X_8 = 1 \mid X_7 = 0)$.
5. Find $P(X_7 = 0 \mid X_4 = 0)$.

1.4 Markov chains with infinite state space

Although the formulas for $n-$step transitions and time n distributions are motivated by those obtained in the previous section, the big difference if $\mathcal{S}$ is infinite is that the transitions $P(x,y)$ **cannot be expressed in a matrix** (since the matrix would have to have infinitely many rows and columns). The proper notation is to use functions:

Definition 1.13 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$.*

1. *The* **transition function** *of the Markov chain is the function*

$$P : \mathcal{S} \times \mathcal{S} \to [0,1] \text{ defined by } P(x,y) = P(X_t = y \,|\, X_{t-1} = x).$$

2. *The* **initial distribution** *of the Markov chain is the function*

$$\pi_0 : \mathcal{S} \to [0,1] \text{ defined by } \pi_0(x) = P(X_0 = x).$$

3. *The* $n-$**step transition function** *of the Markov chain is the function* $P^n : \mathcal{S} \times \mathcal{S} \to [0,1]$ *defined by*

$$P^n(x,y) = P(X_{t+n} = y \,|\, X_t = x).$$

4. *The* **time** n **distribution** *of the Markov chain is the function*

$$\pi_n : \mathcal{S} \to [0,1] \text{ defined by } \pi_n(x) = P(X_n = x).$$

As with finite state spaces, the transition functions must be "stochastic":

Lemma 1.14 *$P : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ is the transition function of a Markov chain with state space $\mathcal{S}$ if and only if*

1. *for every $x, y \in \mathcal{S}$, $P(x,y) \geq 0$, and*

2. *for every $x \in \mathcal{S}$, $\sum_{y \in \mathcal{S}} P(x,y) = 1$.*

Lemma 1.15 *If π_n is the time n distribution of a Markov chain with state space $\mathcal{S}$, then $\sum_{x \in \mathcal{S}} \pi_n(x) = 1$.*

Theorem 1.16 *Let $\{X_t\}$ be a Markov chain with transition function P and initial distribution π_0. Then:*

1. *For all $x_0, x_1, ..., x_n \in \mathcal{S}$,*

$$P(X_0 = x_0, X_1 = x_1, ..., X_n = x_n) = \pi_0(x_0) \prod_{j=1}^{n} P(x_{j-1}, x_j)$$

2. *For all $x, y \in \mathcal{S}$,*

$$P^n(x,y) = \sum_{z_1,...,z_{n-1} \in \mathcal{S}} P(x, z_1)P(z_1, z_2) \cdots P(z_{n-2}, z_{n-1})P(z_{n-1}, y)$$

3. *The time n distribution π_n satisfies, for all $y \in \mathcal{S}$,*

$$\pi_n(y) = \sum_{x \in \mathcal{S}} \pi_0(x)P^n(x,y).$$

1.5 Recurrence and transience

Goal: Determine the long-term behavior of a Markov chain.

General technique: Divide the states of the Markov chain into various "types"; there will be general laws which govern the behavior of each "type" of state.

Definition 1.17 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$.*

1. *Given an event E, define $P_x(E) = P(E \,|\, X_0 = x)$. This is the probability of event E, given that you start at x.*

2. *Given a r.v. Z, define $E_x(Z) = E(Z \,|\, X_0 = x)$. This is the expected value of Z, given that you start at x.*

Definition 1.18 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$.*

1. *Given a set $A \subseteq \mathcal{S}$, let T_A be the r.v. defined by*

$$T_A = \min\{t \geq 1 : X_t \in A\}.$$

   ($T_A = \infty$ if $X_t \notin A$ for all t.) T_A is called the **hitting time** *or* **first passage time** *to A.*

2. *Given a state $a \in \mathcal{S}$, denote by T_a the r.v. $T_{\{a\}}$.*

Note: $T_A : \Omega \to \mathbb{N} \bigcup \{\infty\}$, so $\sum_{n=1}^{\infty} P(T_A = n) = 1 - P(T_A = \infty) \leq 1$.

Definition 1.19 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. A state $a \in \mathcal{S}$ is called* **absorbing** *if $P(a, a) = 1$ (i.e. once you hit a, you never leave).*

Examples of absorbing states:

Definition 1.20 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$.*

1. *For each $x, y \in \mathcal{S}$, define*

$$f_{x,y} = P_x(T_y < \infty).$$

   This is the probability you get from x to y in some finite (positive) time.

2. *We say x* **leads to** *y (and write $x \to y$) if $f_{x,y} > 0$. This means that if you start at x, there is some positive probability that you will eventually hit y.*

3. *For each $x \in \mathcal{S}$, set $f_x = f_{x,x} = P_x(T_x < \infty)$.*

4. *A state $x \in \mathcal{S}$ is called* **recurrent** *if $f_x = 1$. The set of recurrent states of the Markov chain is denoted $\mathcal{S}_R$. The Markov chain $\{X_t\}$ is called* **recurrent** *if $\mathcal{S}_R = \mathcal{S}$, i.e. all of its states are recurrent.*

5. *A state $x \in \mathcal{S}$ is called* **transient** *if $f_x < 1$. The set of transient states of the Markov chain is denoted $\mathcal{S}_T$. The Markov chain $\{X_t\}$ is called* **transient** *if all its states are transient.*

Recurrent and transient states are the two "types" of states referred to earlier.

- a recurrent state (by definition) is "a state to which you ***must*** return" (with probability 1)
- a transient state is (by definition) "a state to which you ***might not*** return".

Definition 1.21 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. For each $x \in \mathcal{S}$, define*

$$V_x = \#\text{ of times } t \geq 0 \text{ such that } X_t = x.$$

V_x is a r.v. called the **number of visits** *to x.*

Note: $V_x : \Omega \to \{0, 1, 2, 3, ...\} \bigcup \{\infty\}$.

Elementary properties of recurrent and transient states

The rest of this section is devoted to developing properties of recurrent and transient states.

Lemma 1.22 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Then*

$$x \to y \iff P^n(x, y) > 0 \text{ for some } n \geq 1.$$

PROOF (⇒) Assume $x \to y$, i.e. $f_{x,y} = P_x(T_y < \infty) > 0$.

(⇐) Suppose $P^N(x, y) > 0$.

Lemma 1.23 *Let* $\{X_t\}$ *be a Markov chain with state space* $\mathcal{S}$. *Then*

$$(x \to y \text{ and } y \to z) \Rightarrow x \to z.$$

PROOF Apply Lemma 1.22 twice:

$$x \to y \Rightarrow \exists\, n_1 \text{ such that } P^{n_1}(x,y) > 0.$$

$$y \to z \Rightarrow \exists\, n_2 \text{ such that } P^{n_2}(y,z) > 0.$$

Thus

$$P^{n_1+n_2}(x,z) \geq P^{n_1}(x,y)P^{n_2}(y,z) > 0,$$

so by Lemma 1.22 $x \to z$. □

Lemma 1.24 *Let* $\{X_t\}$ *be a Markov chain with state space* $\mathcal{S}$. *Then for all* $x, y \in \mathcal{S}$ *and all* $n \geq 1$,

$$P^n(x,y) = \sum_{m=1}^{n} P_x(T_y = m)P^{n-m}(y,y).$$

PROOF

Application: We know that $P_x(T_y = 1) = P(x,y)$. Similarly,

$$P_x(T_y = 2) = \sum_{z \neq y} P(X_0 = x, X_1 = z, X_2 = y) = \sum_{z \neq y} P(x,z)P(z,y);$$

$$P_x(T_y = n) = \sum_{z \neq y} P(x,z)P_z(T_y = n-1) \quad \text{for } n \geq 2$$

so the numbers inside the summation in Lemma 1.24 could all be computed inductively.

Corollary 1.25 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If a is an absorbing state, then*

$$P^n(x,a) = P_x(T_a \le n) = \sum_{m=1}^{n} P_x(T_a = m).$$

PROOF If a is absorbing, then $P(a,a) = 1$ so $P^{n-m}(a,a) = 1$ for all $m \le n$ as well. Therefore by Lemma 1.24,

$$P^n(x,a) = \sum_{m=1}^{n} P_x(T_a = m)P^{n-m}(a,a) = \sum_{m=1}^{n} P_x(T_a = m). \square$$

Theorem 1.26 (Properties of recurrent and transient states) *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Then:*

1. *If $y \in \mathcal{S}_T$, then for all $x \in \mathcal{S}$,*

$$P_x(V_y < \infty) = 1 \text{ and } E_x(V_y) = \frac{f_{x,y}}{1 - f_y}.$$

2. *If $y \in \mathcal{S}_R$, then*

$$P_x(V_y = \infty) = P_x(T_y < \infty) = f_{x,y}$$

(in particular $P_y(V_y = \infty) = 1$) and

   (a) *if $f_{x,y} = 0$, then $E_x(V_y) = 0$;*
   (b) *if $f_{x,y} > 0$, then $E_x(V_y) = \infty$.*

What this theorem says in English:

1. If y is transient, then no matter where you start, you only visit y a finite number of times (and the expected number of times you visit is $\frac{f_{x,y}}{1-f_y}$).

2. If y is recurrent, then
   - it is possible to never hit y, but
   - if you hit y, then you must visit y infinitely many times.

PROOF First, observe that $V_y \geq 1 \iff T_y < \infty$, because both statements correspond to hitting y in a finite amount of time.

Therefore $P_x(V_y \geq 1) = P_x(T_y < \infty) = f_{x,y}$.

Now $P_x(V_y \geq 2) =$

Similarly $P_x(V_y \geq n) =$

Therefore, for all $n \geq 1$ we have $P_x(V_y = n) =$

Therefore $P_x(V_y = 0) =$

First situation: y is transient (i.e. $f_y = f_{y,y} < 1$). Then

$$P_x(V_y = \infty) = \lim_{n\to\infty} P_x(V_y \geq n) =$$

Also,

$$\begin{aligned} E_x(V_y) = E(V_y \,|\, X_0 = x) &= \sum_{m=1}^{\infty} m \cdot P(V_y = m \,|\, X_0 = x) \\ &= \end{aligned}$$

Second situation: y is recurrent (i.e. $f_y = f_{y,y} = 1$). Then

One offshoot of the theorem above are these criteria, which can be useful in some situations to determine if a state is recurrent or transient:

Corollary 1.27 (Recurrence criterion I) *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $x \in \mathcal{S}$. Then*

$$x \text{ is recurrent} \iff \sum_{n=1}^{\infty} P^n(x,x) \text{ diverges.}$$

PROOF

$$x \in \mathcal{S}_R \iff E_x(V_x) = \infty \iff \sum_{n=1}^{\infty} P^n(x,x) = \infty. \; \square$$

Corollary 1.28 (Recurrence criterion II) *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $y \in \mathcal{S}_T$, then for all $x \in \mathcal{S}$,*

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

The reason this is called a "recurrence criterion" is that the contrapositive says that if $P^n(x,y)$ does not converge to 0, then y is recurrent.

PROOF y being transient implies $E_x(V_y) < \infty$ which implies $\sum_{n=1}^{\infty} P^n(x,y) < \infty$. By the n^{th}-term Test for infinite series (Calculus II), that means $\lim_{n\to\infty} P^n(x,y) = 0$. $\square$

Corollary 1.29 (Finite state space Markov chains are not transient) *Let $\{X_t\}$ be a Markov chain with* **finite** *state space $\mathcal{S}$. Then the Markov chain is not transient (i.e. there is at least one recurrent state).*

PROOF Suppose not, i.e. all states are transient. Then by the second recurrence criterion,

$$0 = \lim_{n\to\infty} P^n(x,y) \quad \forall\, x,y \in \mathcal{S}$$
$$\Rightarrow 0 = \sum_{y\in\mathcal{S}} \lim_{n\to\infty} P^n(x,y)$$

Example: Consider a Markov chain with state space $\{1,2,3\}$ and transition matrix

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

where $p \in (0,1)$.

1. Which states are recurrent? Which states are transient?
2. Find $f_{x,y}$ for all $x, y \in \mathcal{S}$.
3. Find the expected number of visits to each state, given that you start in any of the states.

Theorem 1.30 (Recurrent states lead only to other recurrent states) *Let $\{X_t\}$ be a Markov chain. If $x \in \mathcal{S}$ is recurrent and $x \to y$, then*

1. *y is recurrent;*
2. *$f_{x,y} = 1$; and*
3. *$f_{y,x} = 1$.*

Proof: If $y = x$, this follows from the definition of "recurrent", so assume $y \neq x$. We are given $x \to y$, so $P^n(x, y) > 0$ for some $n \geq 1$. Let N be the smallest $n \geq 1$ such that $P^n(x, y) > 0$. Then we have a picture like this:

Suppose now that $f_{y,x} < 1$**.** Then

Now since $f_{y,x} = 1$, $y \to x$ so there exists a number N' so that $P^{N'}(y, x) > 0$.

So for every $n \geq 0$, $P^{N'+n+N}(y, y) \geq P^{N'}(y, x)P^n(x, x)P^N(x, y)$.

Therefore

$$
\begin{aligned}
E_y(V_y) = \sum_{n=1}^{\infty} P^n(y, y) \geq \sum_{n=N'+N+1}^{\infty} P^n(y, y) &= \sum_{n=1}^{\infty} P^{N'+n+N}(y, y) \\
&\geq \sum_{n=1}^{\infty} P^{N'}(y, x)P^n(x, x)P^N(x, y) \\
&= P^{N'}(y, x)P^N(x, y) \sum_{n=1}^{\infty} P^n(x, x)
\end{aligned}
$$

Finally, as $y \in \mathcal{S}_R$ and $y \to x$, $f_{x,y} = 1$ by statement (3) of this Theorem. This proves (2). □

Corollary 1.31 *Let $\{X_t\}$ be a Markov chain. If $y \in \mathcal{S}$ is absorbing and $x \neq y$ leads to y, then x is transient.*

PROOF If $x \in \mathcal{S}_R$, then $f_{y,x} = 1$ by the previous theorem. But $f_{y,x} = 0$ since $y \neq x$ and y is absorbing. □

Closed sets and communicating classes

Definition 1.32 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$, and let C be a subset of $\mathcal{S}$.*

1. *C is called* **closed** *if for every $x \in C$, if $x \to y$, then y must also be in C.*
2. *C is called a* **communicating class** *if C is closed, and if for every x and y in C, $x \to y$ (thus by symmetry $y \to x$).*
3. *$\{X_t\}$ is called* **irreducible** *if $\mathcal{S}$ is a communicating class.*

- closed sets are those which are like the Hotel California: "you can never leave".
- A set is a communicating class if you never leave, and you can get from anywhere to anywhere within the class.
- A Markov chain is irreducible if you can get from any state to any other state.

Theorem 1.33 (Main Recurrence and Transience Theorem) *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$.*

1. *If $C \subseteq \mathcal{S}$ is a communicating class, then every state in C is recurrent (i.e. $C \subseteq \mathcal{S}_R$), or every state in C is transient (i.e. $C \subseteq \mathcal{S}_T$).*
2. *If $C \subseteq \mathcal{S}$ is a communicating class of recurrent states, then $f_{x,y} = 1$ for all $x, y \in C$.*
3. *If $C \subseteq \mathcal{S}$ is a finite communicating class, then $C \subseteq \mathcal{S}_R$.*
4. *If $\{X_t\}$ is irreducible, then $\{X_t\}$ is either recurrent or transient.*
5. *If $\{X_t\}$ is irreducible and $\mathcal{S}$ is finite, then $\{X_t\}$ is recurrent.*

Example: Let $\{X_t\}$ be a Markov chain with state space $\{1, 2, 3, 4, 5, 6\}$ whose transition matrix is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & 0 & \frac{1}{8} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

Determine which states of the chain are recurrent and which states are transient. Identify all communicating classes. For each $x, y \in \mathcal{S}$, compute $f_{x,y}$.

Example: Describe the closed subsets of a Bernoulli process. Do Bernoulli processes have any communicating classes?

Theorem 1.34 (Decomposition Theorem) *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $\mathcal{S}_R \neq \emptyset$, then we can write*

$$\mathcal{S}_R = \bigcup_j C_j$$

where the C_j are disjoint communicating classes (the union is either finite or countable).

PROOF $\mathcal{S}_R \neq \emptyset \Rightarrow$ let $x \in \mathcal{S}_R$. Define $C(x) = \{y \in \mathcal{S} : x \to y\}$.

Observe that $x \in C(x)$ since x is recurrent. Thus $C(x) \neq \emptyset$.

Claim: $C(x)$ is closed.

Claim: $C(x)$ is a communicating class.

This shows $\mathcal{S}_R = \bigcup_{x \in \mathcal{S}_R} C(x)$. Left to show the $C(x)$ are disjoint or coincide for different x:

.

We can summarize all the qualitative results regarding recurrence and transience in the following block.

One catch: in this block, the phrase "you will" really means "the probability that you will is 1".

State space decomposition of a Markov chain

Given a Markov chain with state space $\mathcal{S}$, we can write $\mathcal{S}$ as a disjoint union

$$\mathcal{S} = \mathcal{S}_R \bigcup \mathcal{S}_T = \left(\underset{j}{\cup}\, C_j \right) \bigcup \mathcal{S}_T .$$

1. If you start in one of the C_j, you will stay in that C_j forever and visit every state in that C_j infinitely often.

2. If you start in $\mathcal{S}_T$, you either

(a) stay in $\mathcal{S}_T$ forever (but hit each state in $\mathcal{S}_T$ only finitely many times)

or

(b) eventually enter a C_j, in which case you subsequently stay in that C_j forever and visit every state in that C_j infinitely often.

Remark: 2 (a) above is only possible if $\mathcal{S}_T$ is infinite.

1.6 Absorption probabilities

Question: Suppose you have a Markov chain with state space decomposition as described above. Suppose you start at $x \in \mathcal{S}_T$. What is the probability that you eventually enter recurrent communicating class C_j?

Definition 1.35 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $x \in \mathcal{S}_T$ and let C_j be a communicating class of recurrent states. The* **probability** x **is absorbed by** *C_j, denoted f_{x,C_j}, is*

$$f_{x,C_j} = P_x(T_{C_j} < \infty).$$

Lemma 1.36 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $x \in \mathcal{S}_T$ and let C be a communicating class of recurrent states. Then for any $y \in C$, $f_{x,C_j} = f_{x,y}$.*

In the situation where $\mathcal{S}_T$ is finite, we can solve for these probabilities by solving a system of linear equations. Here is the method:

Suppose $\mathcal{S}_T = \{x_1, ..., x_n\}$.

Since $\mathcal{S}_T$ is finite, each x_j must eventually be absorbed by a C_j, so we have

$$\sum_i f_{x_j,C_i} = 1 \text{ for all } j.$$

Fix one of the C_i; then

$$f_{x_j,C_i} = P_{x_j}(T_{C_i} = 1) + P_{x_j}(T_{C_i} > 1)$$

If you write this equation for each $x_j \in \mathcal{S}_T$, you get a system of n equations in the n unknowns $f_{x_1,C_i}, f_{x_2,C_i}, f_{x_3,C_i}, ..., f_{x_n,C_i}$. This can be solved for the absorption probabilities for C_i; repeating this procedure for each i yields all the absorption probabilities of the Markov chain.

Example: Consider a Markov chain with transition matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}..$$

For every $x, y \in \mathcal{S}$, compute $f_{x,y}$.

Chapter 2

Martingales

2.1 Motivation: betting on fair coin flips

Let's suppose you are playing a game with your friend where you bet \$1 on each flip of a fair coin (fair means the coin flips heads with probability $\frac{1}{2}$ and tails with probability $\frac{1}{2}$). If the coin flips heads, you win, and if the coin flips tails, you lose (mathematically, this is the same as "calling" the flip and winning if your call was correct).

Suppose you come to this game with \$10. What will happen after four plays of this game?

To set up some notation, we will let X_t be your bankroll after playing the game t times; this gives a stochastic process $\{X_t\}_{t\in\mathbb{N}}$. We know $X_0 = 10$, for example.

Sequence of flips (in order)	Probability of that sequence	X_4 = bankroll after four flips
H H H H	$\frac{1}{16}$	14
H H H T	$\frac{1}{16}$	12
H H T H	$\frac{1}{16}$	12
H H T T	$\frac{1}{16}$	10
H T H H	$\frac{1}{16}$	12
H T H T	$\frac{1}{16}$	10
H T T H	$\frac{1}{16}$	10
H T T T	$\frac{1}{16}$	8
T H H H	$\frac{1}{16}$	12
T H H T	$\frac{1}{16}$	10
T H T H	$\frac{1}{16}$	10
T H T T	$\frac{1}{16}$	8
T T H H	$\frac{1}{16}$	10
T T H T	$\frac{1}{16}$	8
T T T H	$\frac{1}{16}$	8
T T T T	$\frac{1}{16}$	6

To summarize, your bankroll after four flips, i.e. X_4, has the following density:

x	6	8	10	12	14
$P(X_4 = x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

What happens if we assume some additional information? For example, suppose that the first flip is heads. Given this, what is $E[X_4]$? In other words, what is $E[X_4 \mid X_1 = 11]$?

Repeating the argument from above, we see

Sequence of flips (in order)	Probability of that sequence	Resulting bankroll after four flips
H H H H	$\frac{1}{8}$	14
H H H T	$\frac{1}{8}$	12
H H T H	$\frac{1}{8}$	12
H H T T	$\frac{1}{8}$	10
H T H H	$\frac{1}{8}$	12
H T H T	$\frac{1}{8}$	10
H T T H	$\frac{1}{8}$	10
H T T T	$\frac{1}{8}$	8

Therefore $X_4 \mid X_1 = 11$ has conditional density

x	6	8	10	12	14
$P(X_4 = x \mid X_1 = 11)$	0	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

and

$$E[X_4 \mid X_1 = 11] = 0(6) + \frac{1}{8}(8) + \frac{3}{8}(10) + \frac{3}{8}(12) + \frac{1}{8}(14) = 11.$$

A similar calculation would show that if the first flip was tails, then we would have

$$E[X_4 \mid X_1 = 9] = 9.$$

From the previous two statements, we can conclude:

In fact, something more general holds. For this Markov chain $\{X_t\}$, we have for any $s \leq t$ that

$$E[X_t \mid X_s] = X_s.$$

To see why, let's define another sequence of random variables coming from the process $\{X_t\}$. For each $t \in \{1, 2, 3, ...\}$, define

$$S_t = X_t - X_{t-1} = \begin{cases} +1 & \text{if the } t^{th} \text{ flip is H (i.e. you win \$1 on the } t^{th} \text{ game)} \\ -1 & \text{if the } t^{th} \text{ flip is T (i.e. you lose \$1 on the } t^{th} \text{ game).} \end{cases}$$

Note that $E[S_t] = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0$, and also note that

$$\begin{aligned} X_t &= X_0 + (X_1 - X_0) + (X_2 - X_1) + ... + (X_t - X_{t-1}) \\ &= X_0 + S_1 + ... + S_t \\ &= X_0 + \sum_{j=1}^{t} S_j \end{aligned}$$

so therefore

$$EX_t = E\left[X_0 + \sum_{j=1}^{t} S_j\right] = E[X_0] + \sum_{j=1}^{t} E[S_j] = E[X_0] + 0 = EX_0.$$

If we are given the value of X_0 (and we usually are, given that X_0 represents the initial bankroll), we have

$$E[X_t \mid X_0] = X_0.$$

More generally, for any $s \leq t$, we have

$$X_t = X_s + S_{s+1} + S_{s+2} + ... + S_t = X_s + \sum_{j=s+1}^{t} S_j$$

so by a similar calculation as above, we have $EX_t = EX_s$ so if we know the value of X_s, we obtain

$$E[X_t \mid X_s] = X_s.$$

What we have proven is that the process $\{X_t\}$ defined by this game is something called a "martingale". Informally, a process is a martingale if, given the state(s) of the process up to and including some time s (you think of time s as the "present time"), the expected state of the process at a time $t \geq s$ (think of t as a "future time") is equal to X_s.

Unfortunately, to define this formally in a way that is useful for deriving formulas, proving theorems, etc., we need quite a bit of additional machinery.

The major question: can you beat a fair game?

Suppose that instead of betting $1 on each flip, that you varied your bets from one flip to the next. Suppose you think of a method of betting as a "strategy". Here are some things you might try:

Strategy 1: Bet $1 on each flip.

Strategy 2: Alternate between betting $1 and betting $2.

Strategy 3: Start by betting $1 on the first flip. After that, bet $2 if you lost the previous flip, and bet $1 if you won the previous flip.

Strategy 4: Bet $1 on the first flip. If you lose, double your bet after each flip you lose until you win once. Then go back to betting $1 and repeat the procedure.

Is there a strategy (especially one with bounded bet sizes) you can implement such that your expected bankroll after the 20^{th} flip is greater than your initial bankroll X_0? If so, what is it? If not, what about if you flip 100 times? Or 1000 times? Or any finite number of times?

Furthermore, suppose that instead of planning beforehand to flip a fixed number of times, decide that you will stop at a random time depending on the results of the flips. For instance, you might stop when you win five straight bets. Or you might stop when you are ahead $3.

The big picture question: All told, what we want to know is whether or not there is a betting strategy and a time you can plan to stop so that if you implement that strategy and stop when you plan to, you will expect to have a greater bankroll than what you start with (even though you are playing a fair game).

2.2 Filtrations

Goal: define what is meant in general by a "strategy", and what is meant in general by a "stopping time".

Recall the following definition from Math 414:

Definition 2.1 *Let Ω be a set. A nonempty collection $\mathcal{F}$ of subsets of Ω is called a $\sigma-$**algebra** (a.k.a. $\sigma-$**field**) if*

1. *$\mathcal{F}$ is "closed under complements", i.e. whenever $E \in \mathcal{F}$, $E^C \in \mathcal{F}$.*
2. *$\mathcal{F}$ is "closed under finite and countable unions and intersections", i.e. whenever $E_1, E_2, E_3, ... \in \mathcal{F}$, both $\bigcup_j A_j$ and $\bigcap_j A_j$ belong to $\mathcal{F}$ as well.*

Theorem 2.2 *Let $\mathcal{F}$ be a $\sigma-$algebra on set Ω. Then $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$.*

(In Math 414, I used $\mathcal{A}$ rather than $\mathcal{F}$ to denote σ-algebras.)

Examples of σ-algebras

Example 1: Let Ω be any set. Let $\mathcal{F} = \{\emptyset, \Omega\}$. This is called the **trivial σ-algebra** of Ω.

Example 2: Let Ω be any set. Let $\mathcal{F} = 2^\Omega$ be the set of all subsets of Ω. This is called the **power set** of Ω.

Example 3: Let $\Omega = [0,1] \times [0,1]$. Let $\mathcal{F}$ be the collection of all subsets of Ω of the form $A \times [0,1]$ where $A \subset [0,1]$.

Example 4: Let Ω be any set and let $\mathcal{P} = \{P_1, P_2, ..., P_n\}$ be any partition of Ω (that is, that $P_i \cap P_j = \emptyset$ for all $i \neq j$ and $\bigcup_j P_j = \Omega$). Then let $\mathcal{F}$ be the collection of all sets which are unions of some number of the P_j. This $\mathcal{F}$ is called the **σ-algebra generated by** $\mathcal{P}$.

Measurability

Definition 2.3 *Let Ω be a set and let $\mathcal{F}$ be a σ-algebra on Ω. A subset E of Ω is called $\mathcal{F}-$**measurable** (or just **measurable**) if $E \in \mathcal{F}$. A function (i.e. a random variable) $X : \Omega \to \mathbb{R}$ is called $\mathcal{F}$-**measurable** if for any open interval $(a, b) \subseteq \mathbb{R}$, the set*

$$X^{-1}(a, b) = \{\omega \in \Omega : X(\omega) \in (a, b)\}$$

is $\mathcal{F}$-measurable.

Example: Let $\Omega = [0, 1]$ and let $\mathcal{F}$ be the σ-algebra generated by the partition $\mathcal{P} = \{[0, 1/3), [1/3, 1/2), [1/2, 1]\}$.

More generally, if $\mathcal{F}$ is generated by a partition $\mathcal{P}$, a r.v. X is measurable if and only if it is constant on each of the partition elements; in other words, if $X(\omega)$ depends not on ω but only on which partition element ω belongs to.

This idea illustrates the point of measurability in general: think of a σ-algebra $\mathcal{F}$ as revealing some partial information about an ω (i.e. it tells you which sets in $\mathcal{F}$ to which ω belongs, but not necessarily exactly what ω is); to say that a function X is $\mathcal{F}$-measurable means that the evaluation of $X(\omega)$ depends only on the information contained in $\mathcal{F}$.

Throughout this chart, let $\Omega = [0,1] \times [0,1]$.

σ-algebra $\mathcal{F}$	information $\mathcal{F}$ reveals about ω	description of $\mathcal{F}$-measurable functions

Filtrations

> **Definition 2.4** *Let Ω be a set and let $\mathcal{I} \subseteq [0, \infty)$. A* **filtration** $\{\mathcal{F}_t\}_{t \in \mathcal{I}}$ *on Ω is a sequence of σ-algebras indexed by elements of $\mathcal{I}$ which is increasing, i.e. if $s, t \in \mathcal{I}$, then*
>
> $$s \leq t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t.$$

Idea: for any filtration $\{\mathcal{F}_t\}$, when $s \leq t$, each $\mathcal{F}_s$-measurable set is also $\mathcal{F}_t$-measurable, so *as t increases, there are more $\mathcal{F}_t$-measurable sets*.

Put another way, as t increases you get more information about the points in Ω.

> **Definition 2.5** *Let $\{X_t\}_{t \in \mathcal{I}}$ be a stochastic process with index set $\mathcal{I}$. The* **natural filtration** *of $\{X_t\}$ is described by setting*
>
> $$\mathcal{F}_t = \{\text{events which are characterized only by the values of } X_s \text{ for } 0 \leq s \leq t\}.$$

Every natural filtration is clearly a filtration. To interpret this in the context of gambling, think of points in Ω as a list which records the outcome of every bet you make. $\mathcal{F}_t$ is the σ-algebra that gives you the result of the first t bets; as t increases, you get more information about what happens.

Example: Flip a fair coin twice, start with \$10 and bet \$1 on the first flip and \$3 on the second flip. Let X_t be your bankroll after the t^{th} flip (where $t \in \mathcal{I} = \{0, 1, 2\}$).

Strategies

> **Definition 2.6** *Let $\{X_t\}_{t\in\mathcal{I}}$ be a stochastic process and let $\{\mathcal{F}_t\}$ be its natural filtration. A* **predictable sequence** *(a.k.a.* **strategy***) for $\{X_t\}$ is another stochastic process $\{B_t\}$ such that for all $s < t$, B_t is $\mathcal{F}_s$-measurable.*

Idea: Suppose you are betting on repeated coin flips and you decide to implement a strategy where B_t is the amount you are going to bet on the t^{th} flip.

- If you own a time machine, you would just go forward in time to see what the coin flips to, bet on that, and win.
- But if you don't own a time machine, the amount B_t you bet on the t^{th} flip is **only allowed to depend on information coming from flips before the t^{th} flip**, i.e. B_t is only allowed to depend on information coming from X_s for $s < t$, i.e. B_t must be $\mathcal{F}_s$-measurable for all $s < t$.

Remark: If the index set $\mathcal{I}$ is discrete, then a process $\{B_t\}$ is a strategy for $\{X_t\}$ if for every t, B_t is $\mathcal{F}_{t-1}$-measurable.

Examples: Suppose you are betting on repeated coin flips. Throughout these examples, let's use the following notation to keep track of whether you win or lose each game:

$$X_0 = \text{your initial bankroll}$$
$$X_t = \begin{cases} X_{t-1}+1 & \text{if you win the } t^{th} \text{ game} \\ X_{t-1}-1 & \text{if you lose the } t^{th} \text{ game} \end{cases}$$
$$S_t = X_t - X_{t-1} = \begin{cases} 1 & \text{if you win the } t^{th} \text{ game} \\ -1 & \text{if you lose the } t^{th} \text{ game} \end{cases}$$

So $\{X_t\}$ would measure your bankroll after t games, **if you are betting $\$1$ on each game**. However, you may want to bet more or less than $\$1$ on each game (varying your bets according to some "strategy"). The idea is that B_t will be the amount you bet on the t^{th} game.

Strategy 1: Bet \$1 on each flip.

Strategy 2: Alternate between betting \$1 and betting \$2.

Strategy 3: Start by betting \$1 on the first flip. After that, bet \$2 if you lost the previous flip, and bet \$1 if you won the previous flip.

Strategy 4: Bet \$1 on the first flip. If you lose, double your bet after each flip you lose until you win once. Then go back to betting \$1 and repeat the procedure.

"Strategy" 5: Bet \$5 on the n^{th} flip if you are going to win the n^{th} flip, and bet \$1 otherwise.

Suppose we implement arbitrary strategy $\{B_t\}$ when playing this game. Then our bankroll after t games isn't measured by $\{X_t\}$ any longer; it is

Definition 2.7 *Let* $\{X_t\}_{t\in\mathbb{I}}$ *be a discrete-time stochastic process; let* $S_t = X_t - X_{t-1}$ *for all* t. *Given a strategy* $\{B_t\}$ *for* $\{X_t\}$, *the* **transform of** $\{X_t\}$ **by** $\{B_t\}$ *is the stochastic process denoted* $\{(B \cdot X)_t\}_{t\in\mathbb{N}}$ *defined by*

$$(B \cdot X)_t = X_0 + B_1S_1 + B_2S_2 + ... + B_tS_t = X_0 + \sum_{j=1}^{t} B_jS_j.$$

The point: If you use strategy $\{B_t\}$ to play game $\{X_t\}$, then your bankroll after t games is $(B \cdot X)_t$.

Note: $(B \cdot X)_0 = X_0$.

Example: Suppose you implement Strategy 4 as described above. If your initial bankroll is \$50, and the results of the first eight flips are H T T H T T T H, give the values of B_t, X_t, S_t and $(B \cdot X)_t$ for $0 \le t \le 8$.

Stopping times

Definition 2.8 *Let $\{X_t\}_{t\in\mathcal{I}}$ be a stochastic process with standard filtration $\{\mathcal{F}_t\}$. A r.v. $T : \Omega \to \mathbb{R} \cup \{\infty\}$ is called a* **stopping time (for $\{X_t\}$)** *if for every $a \in \mathbb{R}$, the set of sample functions satisfying $T \leq a$ is $\mathcal{F}_a$-measurable.*

In other words, T is a stopping time if you can determine whether or not $T \leq a$ solely by looking at the values of X_t for $t \leq a$.

In the context of playing a game over and over, think of T as a "trigger" which causes you to stop playing the game. Thus you would walk away from the table with winnings given by X_T (or, if you are employing strategy $\{B_t\}$, your winnings would be $(B \cdot X)_T$).

Example: $T = T_y = \min\{t \geq 0 : X_t = y\}$

Example: $T = \min\{t > 0 : X_t = X_0\}$

Nonexample: $T = \min\{t \geq 0 : X_t = \max\{X_s : 0 \leq s \leq 100\}\}$

Recall our big picture question: is there a strategy under which you can beat a fair game?

Restated in mathematical terms: Suppose stochastic process $\{X_t\}$ represents a fair game (i.e. $E[X_t|X_s] = X_s$ for all $s \leq t$). Is there a predictable sequence $\{B_t\}$ for this process, and a stopping time T for this process such that $E[(B \cdot X)_T] > X_0$? (If so, what $\{B_t\}$ and what T maximizes $E[(B \cdot X)_T]$?)

2.3 Conditional expectation with respect to a σ-algebra

Recall from Math 414: Conditional expectation of one r.v. given another:

Here is a useful theorem that follows from this definition:

Theorem 2.9 *Given any bounded, continuous function $\phi : \mathbb{R} \to \mathbb{R}$,*

$$E[X\,\phi(Y)] = E\left[E(X|Y)\,\phi(Y)\right].$$

PROOF (when X, Y continuous):

$$\begin{aligned}
E[X\,\phi(Y)] &= \int\int x\,\phi(y) f_{X,Y}(x,y)\,dA \\
&= \int\int x\,\phi(y) f_{X|Y}(x|y) f_Y(y)\,dA \\
&= \int\int x f_{X|Y}(x|y)\phi(y) f_Y(y)\,dx\,dy \\
&= \int \left(\int x f_{X|Y}(x|y)\,dx\right)\,\phi(y) f_Y(y)\,dy \\
&= \int E(X|Y)(y)\,\phi(y) f_Y(y)\,dy \\
&= E\left[E(X|Y)\,\phi(Y)\right].
\end{aligned}$$

The proof when X, Y are discrete is similar, but has sums instead of integrals. □

To define the conditional expectation of a random variable given a σ-algebra, we use Theorem 2.9 to motivate a definition:

Definition 2.10 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a $\mathcal{F}$-measurable r.v. and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub σ-algebra. The* **conditional expectation of X given $\mathcal{G}$** *is a function $E(X|\mathcal{G}) : \Omega \to \mathbb{R}$ with the following two properties:*

1. *$E(X|\mathcal{G})$ is $\mathcal{G}$-measurable, and*
2. *for any bounded, $\mathcal{G}$-measurable r.v. $Z : \Omega \to \mathbb{R}$, $E[XZ] = E\left[E(X|\mathcal{G})\,Z\right]$.*

Facts about conditional expectation given a σ-algebra:

1. Conditional expectations always exist.
2. Conditional expectations are unique up to sets of probability zero.
3. By setting $Z = 1$, we see that $E[X] = E[X|\mathcal{G}]$. This gives you the idea behind this type of conditional expectation: *$E(X|\mathcal{G})$ is a $\mathcal{G}$-mble r.v. with the same expected value(s) as the original r.v. X.*

Example: Let $\Omega = \{A, B, C, D\}$; let $\mathcal{F} = 2^{\Omega}$; let P be the uniform distribution on Ω.

Let $\mathcal{G}$ be the σ-algebra generated by $\mathcal{P} = \{\{A, B\}, \{C, D\}\}$.

Let $X : \Omega \to \mathbb{R}$ be defined by $X(A) = 2$; $X(B) = 6$; $X(C) = 3$; $X(D) = 1$.

Example: Let $\Omega = \{A, B, C, D, E\}$; let $\mathcal{F} = 2^\Omega$; let $P(A) = \frac{1}{4}$; $P(B) = P(C) = P(E) = \frac{1}{8}$; $P(D) = \frac{3}{8}$.

Let $\mathcal{G}$ be generated by the partition $\mathcal{P} = \{\{A, B\}, \{C, D\}, \{E\}\}$.

Let $X(A) = X(B) = X(D) = 2$; $X(C) = 0$; $X(E) = 1$.

Example: Let $\Omega = [0, 1] \times [0, 1]$; let $\mathcal{F} = L(\Omega)$; let P be the uniform distribution. Let $\mathcal{G}$ be the σ-algebra of vertical sets (i.e. sets of the form $E \times [0, 1]$). Let $X : \Omega \to \mathbb{R}$ be $X(x, y) = x + y$. Find $E[X|\mathcal{G}]$.

The following properties of conditional expectation are widely used (their proofs are beyond the scope of this class):

Theorem 2.11 (Properties of conditional expectation) *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Suppose $X, Y : \Omega \to \mathbb{R}$ are $\mathcal{F}$-measurable r.v.s. Let a, b, c be arbitrary real constants. Then:*

1. Positivity: *If $X \geq c$, then $E(X|\mathcal{G}) \geq c$.*
2. Linearity: $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$.
3. Stability: *If X is $\mathcal{G}$-measurable, then $E[X|\mathcal{G}] = X$ and $E[XY|\mathcal{G}] = X\, E[Y|\mathcal{G}]$.*
4. Independence: *If X is independent of any $\mathcal{G}$-measurable r.v., then $E[X|\mathcal{G}] = EX$.*
5. Tower property: *If $\mathcal{H} \subseteq \mathcal{G}$ then $E[E(X|\mathcal{G})|\mathcal{H}] = E[X|\mathcal{H}]$.*
6. Preservation of expectation: $E[E(X|\mathcal{G})] = EX$.
7. Constants: $E[a|\mathcal{G}] = a$.

(These statements hold with probability one.)

2.4 Martingales and the Optional Sampling Theorem

A "martingale" is a mathematical formulation of a fair game:

Definition 2.12 *Let $\{X_t\}_{t\in\mathcal{I}}$ be a stochastic process with natural filtration $\{\mathcal{F}_t\}$.*

- *The process $\{X_t\}$ is called a* **martingale** *if for every $s \le t$ in $\mathcal{I}$,*

$$E[X_t|\mathcal{F}_s] = X_s.$$

- *The process $\{X_t\}$ is called a* **submartingale** *if for every $s \le t$ in $\mathcal{I}$,*

$$E[X_t|\mathcal{F}_s] \ge X_s.$$

- *The process $\{X_t\}$ is called a* **supermartingale** *if for every $s \le t$ in $\mathcal{I}$,*

$$E[X_t|\mathcal{F}_s] \le X_s.$$

Theorem 2.13 (Characterization of discrete-time martingales) *A discrete-time process $\{X_t\}_{t\in\mathbb{N}}$ is a martingale if and only if $E[X_{t+1}|\mathcal{F}_t] = X_t$ for every $t \in \mathbb{N}$.*

PROOF

Theorem 2.14 (Properties of discrete-time martingales) *Suppose that the stochastic process* $\{X_t\}_{t \in \mathbb{N}}$ *is a martingale whose natural filtration is* $\{\mathcal{F}_t\}$. *Define* $S_t = X_t - X_{t-1}$ *for all* t. *Then, for all* t:

1. $X_t = X_0 + \sum_{j=1}^{t} S_t$;
2. S_t *is* $\mathcal{F}_t$*-measurable;*
3. $E[S_{t+1}|\mathcal{F}_t] = 0$;
4. $E[S_t] = 0$;
5. $E[X_t] = E[X_0]$.

PROOF First, statement (1):

$$\begin{aligned} X_t &= X_0 + (X_1 - X_0) + (X_2 - X_1) + \ldots + (X_t - X_{t-1}) \\ &= X_0 + S_1 + \ldots + S_t \\ &= X_0 + \sum_{j=1}^{t} S_j \end{aligned}$$

Statement (2) is obvious, since both X_t and X_{t-1} are $\mathcal{F}_t$-measurable.

Next, statement (3):

$$\begin{aligned} E[S_{t+1}|\mathcal{F}_t] &= E[X_{t+1} - X_t|\mathcal{F}_t] \\ &= E[X_{t+1}|\mathcal{F}_t] - E[X_t|\mathcal{F}_t] \\ &= E[X_t] - E[X_t|\mathcal{F}_t] \quad \text{(since } \{X_t\} \text{ is a martingale)} \\ &= E[X_t] - E[X_t] \quad \text{(by stability)} \\ &= 0. \end{aligned}$$

(4): $E[S_t] = E[X_t] - E[X_{t-1}] = E[X_0] - E[X_0]$ (since $\{X_t\}$ is a martingale) which simplifies to 0.

(5) follows from (1) and (4). □

Theorem 2.15 (Transforms of martingales are martingales) *Let* $\{X_t\}_{t \in \mathbb{N}}$ *be a martingale and suppose that* $\{B_t\}$ *is a strategy for* $\{X_t\}$. *Then the transform* $\{(B \cdot X)_t\}$ *is also a martingale.*

PROOF

$$\begin{aligned}
E\left[(B\cdot X)_{t+1}|\mathcal{F}_t\right] &= E\left[X_0+\sum_{j=1}^{t+1} B_jS_j|\mathcal{F}_t\right] \quad \text{(by the definition of } (B\cdot X))\\
&= E[X_0|\mathcal{F}_t]+\sum_{j=1}^{t} E[B_jS_j|\mathcal{F}_t]+E[B_{t+1}S_{t+1}|\mathcal{F}_t] \quad \text{(by linearity)}\\
&= X_0+\sum_{j=1}^{t} B_jS_j+B_{t+1}E[S_{t+1}|\mathcal{F}_t] \quad \text{(by stability)}\\
&= X_0+\sum_{j=1}^{t} B_jS_j+B_{t+1}0 \quad \text{(by (3) of Thm 2.14)}\\
&= X_0+\sum_{j=1}^{t} B_jS_j\\
&= (B\cdot X)_t.
\end{aligned}$$

By Theorem 1.12 above, $\{(B\cdot X)_t\}$ is a discrete-time martingale. □

Theorem 2.16 (Optional Stopping Theorem (OST)) *Let $\{X_t\}$ be a martingale. Let T be a bounded stopping time for $\{X_t\}$. (To say T is bounded means there is a constant n such that $P(T\le n)=1$.) Then*

$$E[X_T]=E[X_0].$$

PROOF Let $B_t=\begin{cases}1 & \text{if } T\ge t\\ 0 & \text{else}\end{cases}$.

$$\begin{aligned}
T \text{ is a stopping time} \quad &\Rightarrow \quad G=\{T\le t-1\}=\{B_t=0\} \text{ is } \mathcal{F}_{t-1}\text{-measurable } \forall t\\
&\Rightarrow \quad G^C=\{T\ge t\}=\{B_t=1\} \text{ is } \mathcal{F}_{t-1}\text{-measurable } \forall t\\
&\Rightarrow \quad \text{each } B_t \text{ is } \mathcal{F}_{t-1}\text{-measurable}\\
&\Rightarrow \quad \{B_t\} \text{ is a predictable sequence for } \{X_t\}.
\end{aligned}$$

Now, we are assuming T is bounded; let n be such that $P(T\le n)=1$. Now for any $t\ge n$, we have

$$\begin{aligned}
(B\cdot X)_t &= X_0+\sum_{j=1}^{t} B_tS_t\\
&= X_0+\sum_{j=1}^{t} B_t(X_t-X_{t-1})\\
&= X_0+(X_1-X_0)+1(X_2-X_1)+\ldots+1(X_T-X_{T-1})\\
&\qquad +0(X_{T+1}-X_t)+0(X_{T+2}-X_{T+1})+\ldots\\
&= X_T.
\end{aligned}$$

Finally,

$$\begin{aligned} EX_T &= E[(B \cdot X)_t] \\ &= E[(B \cdot X)_0] \quad \text{(since } \{(B \cdot X)_t\} \text{ is a martingale)} \\ &= EX_0. \ \square \end{aligned}$$

Note: The OST is also called the **Optional Sampling Theorem** because of its applications in statistics.

We will need the following "tweaked version" of the OST, which requires a little less about T (it only has to be finite rather than bounded) but a little more about $\{X_t\}$ (the values of X_t have to be bounded until T hits):

Theorem 2.17 (OST (tweaked version)) *Let $\{X_t\}$ be a martingale. Let T be a stopping time for $\{X_t\}$ which is finite with probability one. If there is a fixed constant C such that for sufficiently large n, $T \geq n$ implies $|X_n| \leq C$, then*

$$E[X_T] = E[X_0].$$

PROOF Choose a sufficiently large n and let $\overline{T} = \min(T, n)$. $\overline{T}$ is a stopping time which is bounded by n, so the original OST applies to $\overline{T}$, i.e.

$$EX_{\overline{T}} = EX_0.$$

Now

$$|EX_T - EX_0| = |EX_T - EX_{\overline{T}}|$$

Recall: Our big picture question is whether one can beat a fair game by varying their strategy and/or stopping time. The OST implies that the answer is <u>**NO**</u>:

Corollary 2.18 (You can't beat a fair game) *Let $\{X_t\}$ be a martingale. Let T be a finite stopping time for $\{X_t\}$ and let $\{B_t\}$ be any bounded strategy for $\{X_t\}$. Then*

$$E(B \cdot X)_T = EX_0.$$

PROOF If $\{X_t\}$ is a martingale, so is $(B \cdot X)_t$. Therefore by the tweaked OST,

$$E(B \cdot X)_T = E(B \cdot X)_0 = EX_0. \square$$

Catch: If you are willing to play forever, and/or you are willing to lose a possibly unbounded amount of money first, the OST doesn't apply, and you can beat a fair game using Strategy 4 described several pages ago. But this isn't realistic if you are a human with a finite lifespan and finite wealth.

Application: Suppose a gambler has $\$50$ and chooses to play a fair game repeatedly until either the gambler's bankroll is up to $\$100$, or until the gambler is broke.

If the gambler bets all $\$50$ on one game, then the probability he leaves a winner is $\frac{1}{2}$. What if the gambler bets in some other way?

The results of this section also apply to sub- and supermartingales:

Corollary 2.19 *Suppose that $\{X_t\}_{t \in \mathbb{N}}$ is a submartingale and that $\{B_t\}$ is a strategy for $\{X_t\}$. Then:*

1. *The transform $\{(B \cdot X)_t\}$ is also a submartingale.*
2. *If T is a bounded stopping time for $\{X_t\}$. Then $E[X_T] \geq E[X_0]$ (and $E[(B \cdot X)_T] \geq E[X_0]$).*
3. *If T is a finite stopping time for $\{X_t\}$ and there is a fixed constant C such that for sufficiently large n, $T \geq n$ implies $|X_n| \leq C$, then $E[X_T] \geq E[X_0]$ (and $E[(B \cdot X)_T] \geq E[X_0]$).*

Corollary 2.20 *Suppose that* $\{X_t\}_{t\in\mathbb{N}}$ *is a supermartingale and that* $\{B_t\}$ *is a strategy for* $\{X_t\}$*. Then:*

1. *The transform* $\{(B \cdot X)_t\}$ *is also a supermartingale.*
2. *If* T *is a bounded stopping time for* $\{X_t\}$*. Then* $E[X_T] \leq E[X_0]$ *(and* $E[(B \cdot X)_T] \leq E[X_0]$*).*
3. *If* T *is a finite stopping time for* $\{X_t\}$ *and there is a fixed constant* C *such that for sufficiently large* n*,* $T \geq n$ *implies* $|X_n| \leq C$*, then* $E[X_T] \leq E[X_0]$ *(and* $E[(B \cdot X)_T] \leq E[X_0]$*).*

2.5 Random walk in dimension 1

Definition 2.21 *A discrete-time stochastic process* $\{X_t\}$ *with state space* $\mathbb{Z}$ *is called a* **random walk (on** $\mathbb{Z}$**)** *if there exist*

1. *i.i.d. r.v.s* $S_1, S_2, S_3, ...$ *taking values in* $\mathbb{Z}$ *(*S_j *is called the* j^{th} **step** *or* j^{th} **increment** *of the random walk), and*
2. *a r.v.* X_0 *taking values in* $\mathbb{Z}$ *which is independent of all the* S_j*,*

such that for all t*,* $X_t = X_0 + \sum_{j=1}^{t} S_j$*.*

In this setting:

- X_0 is your starting position;
- $S_j = X_j - X_{j-1}$ is the amount you walk between times $j-1$ and j;
- and X_t is your position at time t.

Note: A random walk on $\mathbb{Z}$ is a Markov chain:

- *State space:* $\mathcal{S} = \mathbb{Z}$
- *Initial distribution:* X_0
- *Transition function:* $P(x, y) = P(S_j = y - x)$.

Note: Random walk models a gambling problem where you make the same bet on the same game over and over. The amount you win/lose on the j^{th} game is S_j.

Example: Make a series of bets (each bet is of of size B) which you win with probability p and lose with probability $1-p$. Then:

Theorem 2.22 *Let $\{X_t\}$ be a random walk. Then:*

1. *$\{X_t\}$ is a martingale if $ES_j = 0$;*
2. *$\{X_t\}$ is a submartingale if $ES_j \geq 0$;*
3. *$\{X_t\}$ is a supermartingale if $ES_j \leq 0$.*

PROOF Applying properties of conditional expectation, we see

$$E[X_{t+1}|\mathcal{F}_t] = E[X_t + S_{t+1}|\mathcal{F}_t] = E[X_t|\mathcal{F}_t] + E[S_{t+1}|\mathcal{F}_t] = X_t + E[S_{t+1}].$$

If $ES_j = 0$, then this reduces to $X_t + 0 = X_t$, so the process is a martingale by Theorem 2.13. If $ES_j \geq 0$, then the last expression above is $\geq X_t$ so the process is a submartingale, and if $ES_j \leq 0$, then the expression is $\leq X_t$ so the process is a supermartingale. □

Definition 2.23 *A random walk on $\mathbb{Z}$ is called* **simple** *if the steps S_j take values only in $\{-1, 0, 1\}$. For a simple random walk, we define*

$$p = P(S_j = 1) \qquad q = P(S_j = -1) \qquad r = P(S_j = 0).$$

For a simple random walk, we let

$$\mu = ES_j \quad \text{and} \quad \sigma^2 = Var(S_j).$$

Lemma 2.24 *For a simple random walk, $\mu = ES_j = p - q$. If the simple random walk is unbiased, then $\mu = 0$ and $\sigma^2 = Var(S_j) = p + q$.*

PROOF HW

A simple random walk models a repeated game where you bet $1 on each play.

Note: A simple random walk is a Markov chain which has the following directed graph:

Definition 2.25 *A simple random walk on $\mathbb{Z}$ is called* **unbiased** *if $p = q$ and is called* **biased** *if $p \neq q$. A biased random walk is called* **positively biased** *if $p > q$ and* **negatively biased** *if $p < q$.*

Note: A simple random walk is irreducible if and only if $p > 0$ and $q > 0$.

Note: Unbiased simple random walks are martingales; positively biased simple random walks are submartingales; negatively biased simple random walks are supermartingales.

Analysis of hitting times for simple random walk

Question: Under what circumstances is a simple random walk recurrent? When is it transient?

To approach this question, we are going to solve a class of problems related to hitting times. Recall that for a set $A \subseteq \mathcal{S}$, $T_A = \min\{t \geq 1 : X_t \in A\}$. T_A is called the **hitting time to** A.

First, for a simple random walk, if $a, b \in A$ and $a < x < b$ but $A \bigcap (a, b) = \emptyset$, then if you start at x, then $T_A = T_{\{a,b\}}$, because you cannot hit A at any point other than a or b (that would require "jumping over" a or b). So we will restrict to hitting times for sets consisting of two points: $A = \{a, b\}$.

First, we start with a result which says that if your initial state in a simple random walk between two numbers a and b, you will definitely hit a or b (or both) in the future:

Lemma 2.26 *Let $\{X_t\}$ be an irreducible simple random walk. Let $A = \{a, b\} \subseteq \mathbb{Z}$ and suppose $X_0 = x$ where $a < x < b$. Then $P(T_A < \infty) = 1$.*

PROOF Since $\{X_t\}$ is irreducible, $p > 0$. Now let G_n be the event that between times $(n-1)(b-a)$ and $n(b-a)$, the chain always steps in the positive direction. In precise math notation,

$$G_n = \{S_j = 1 \, \forall j \in \{(n-1)(b-a)+1, (n-1)(b-a)+2, ..., n(b-a)\}\}.$$

Note that

1. $P(G_n) \geq p^{b-a} > 0$.
2. since G_j and G_k refer to disjoint blocks of time in the chain, $G_j \perp G_k$.

Thus

$$\begin{aligned}
P(\text{no } G_n \text{ occurs}) &= P\left(\bigcap_{n=1}^{\infty} G_n^C\right) \\
&= \prod_{n=1}^{\infty} P(G_n^C) \quad \text{(since the } G_n\text{s are } \perp) \\
&= \lim_{N\to\infty} \prod_{n=1}^{N} P(G_n^C) \\
&= \lim_{N\to\infty} (1-p^{b-a})^N \\
&= 0 \quad (\text{since } 1-p^{b-a} \in (0,1))
\end{aligned}$$

Therefore with probability 1, at least one G_n occurs. This means that with probability 1, at some time in the future there will be $b-a$ consecutive steps in the positive direction, and that means that unless T_a has already occurred, after those $b-a$ consecutive steps, X_t will be $\geq b$. Thus either T_a or T_b is finite, and therefore $P(T_A < \infty) = 1$. □

At this point, we know that in an irreducible, simple random walk, if you start at x and $a < x < b$, you will hit at least one of a or b in the future (with probability one).

Question: what is the probability that you will hit a before b (as opposed to hitting b before a)?

$$P_x(T_a < T_b) = ?$$

Probabilities like $P_x(T_a < T_b)$ are called **escape probabilities** or **first passage-time probabilities**.

To approach this question we will use martingales and the Optional Stopping Theorem.

Lemma 2.27 *Let $\{X_t\}$ be an irreducible simple random walk. Then the following three processes are martingales:*

- *$\{Y_t\}$, where $Y_t = X_t - t\mu$;*
- *$\{Z_t\}$, where $Z_t = (X_t - t\mu)^2 - t\sigma^2$;*
- *$\{U_t\}$, where $U_t = \left(\frac{q}{p}\right)^{X_t}$;*
- *$\{V_t\}$, where $V_t = \frac{e^{\theta X_t}}{[M_{S_j}(\theta)]^t}$ (here the θ can be any arbitrary constant).*

PROOF Throughout this proof, $\{\mathcal{F}_t\}$ is the natural filtration of $\{X_t\}$ (thus also the natural filtration of $\{Y_t\}$, $\{Z_t\}$ and $\{V_t\}$ since they are formulas of $\{X_t\}$). First, let $Y_t = X_t - t\mu$. Then

$$\begin{aligned}
E[Y_{t+1}|\mathcal{F}_t] &= E[X_{t+1} - (t+1)\mu|\mathcal{F}_t] \\
&= E[X_t + S_{t+1} - (t+1)\mu|\mathcal{F}_t] \\
&= X_t + E[S_{t+1}|\mathcal{F}_t] - (t+1)\mu \\
&= X_t + E[S_{t+1}] - (t+1)\mu \\
&= X_t + \mu - t\mu - \mu \\
&= X_t - t\mu = Y_t.
\end{aligned}$$

By Theorem 2.13, $\{Y_t\}$ is a martingale.

Next, let $U_t = \left(\frac{q}{p}\right)^{X_t}$. Here is the calculation:

$$\begin{aligned}
E[U_{t+1}|\mathcal{F}_t] &= E\left[\left(\frac{q}{p}\right)^{X_{t+1}} \middle| \mathcal{F}_t\right] \\
&= E\left[\left(\frac{q}{p}\right)^{X_t+S_{t+1}} \middle| \mathcal{F}_t\right] \\
&= E\left[\left(\frac{q}{p}\right)^{X_t} \left(\frac{q}{p}\right)^{S_{t+1}} \middle| \mathcal{F}_t\right] \\
&= \left(\frac{q}{p}\right)^{X_t} E\left[\left(\frac{q}{p}\right)^{S_{t+1}} \middle| \mathcal{F}_t\right] \qquad \text{(stability)} \\
&= \left(\frac{q}{p}\right)^{X_t} E\left[\left(\frac{q}{p}\right)^{S_{t+1}}\right] \qquad \text{(independence)} \\
&= \left(\frac{q}{p}\right)^{X_t} \left[\left(\frac{q}{p}\right)^{1} p + \left(\frac{q}{p}\right)^{0} r + \left(\frac{q}{p}\right)^{-1} q\right] \\
&= \left(\frac{q}{p}\right)^{X_t} [q + r + p] \\
&= \left(\frac{q}{p}\right)^{X_t} \\
&= U_t.
\end{aligned}$$

By Theorem 2.13, $\{U_t\}$ is a martingale.

Next, let $V_t = \frac{e^{\theta X_t}}{[M_{S_j}(\theta)]^t}$.

$$
\begin{aligned}
E[V_{t+1}|\mathcal{F}_t] &= E\left[\left.\frac{e^{\theta X_{t+1}}}{[M_{S_j}(\theta)]^{t+1}}\right|\mathcal{F}_t\right] \\
&= E\left[\left.\frac{e^{\theta(X_t+S_{t+1})}}{[M_{S_j}(\theta)]^{t+1}}\right|\mathcal{F}_t\right] \\
&= E\left[\left.\frac{e^{\theta X_t}e^{\theta S_{t+1}}}{[M_{S_j}(\theta)]^{t+1}}\right|\mathcal{F}_t\right] \\
&= \frac{e^{\theta X_t}}{[M_{S_j}(\theta)]^{t+1}}E\left[\left.e^{\theta S_{t+1}}\right|\mathcal{F}_t\right] \quad \text{(stability)} \\
&= \frac{e^{\theta X_t}}{[M_{S_j}(\theta)]^{t+1}}E\left[e^{\theta S_{t+1}}\right] \quad \text{(independence)} \\
&= \frac{e^{\theta X_t}}{[M_{S_j}(\theta)]^{t+1}}M_{S_{t+1}}(\theta) \quad \text{(def'n of MGF)} \\
&= \frac{e^{\theta X_t}}{[M_{S_j}(\theta)]^{t}} \quad \text{(since } \{S_j\} \text{ i.i.d.)} \\
&= V_t.
\end{aligned}
$$

By Theorem 2.13, $\{V_t\}$ is a martingale.

The proof for $\{Z_t\}$ is left as a homework exercise. □

Theorem 2.28 (Escape probabilities for random walk) *Let $\{X_t\}$ be an irreducible, simple random walk on $\mathbb{Z}$. Let $a < x < b$ be integers. Then:*

- *if $p = q$ (i.e the random walk is unbiased), then*
  1. $P_x(T_a < T_b) = \frac{b-x}{b-a}$
  2. $P_x(T_b < T_a) = \frac{x-a}{b-a}$
- *if $p \neq q$ (i.e. the random walk is biased), then*
  1. $P_x(T_a < T_b) = \frac{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^x}{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^a}$
  2. $P_x(T_b < T_a) = \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^a}{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^a}$

PROOF *Case 1:* Suppose the random walk is unbiased. That means $\{X_t\}$ is a martingale. Let $T = \min(T_a, T_b) = \min\{t : X_t \in \{a, b\}\}$. T is a finite stopping time, and

$$X_T = \begin{cases} b & \text{if } T_b < T_a \\ a & \text{if } T_a < T_b \end{cases}.$$

That means that

$$\begin{aligned} EX_T &= b\,P_x(T_b < T_a) + a\,P_x(T_a < T_b) \\ &= b[1 - P_x(T_a < T_b)] + a\,P_x(T_a < T_b) \\ &= b + (a-b)P_x(T_a < T_b). \end{aligned}$$

By the "tweaked version" of the OST, we have

$$x = EX_0 = EX_T = b + (a-b)P_x(T_a < T_b).$$

Solve for $P_x(T_a < T_b)$ to get

$$P_x(T_a < T_b) = \frac{x-b}{a-b} = \frac{b-x}{b-a}$$

as desired ($P_x(T_b < T_a) = 1 - \frac{b-x}{b-a} = \frac{x-a}{b-a}$ by the complement rule).

Case 2: Suppose the random walk is biased. Now $\{X_t\}$ is no longer a martingale, but from the preceding lemma, $\{U_t\}$ is a martingale, where $U_t = \left(\frac{q}{p}\right)^{X_t}$. Note first that $EU_0 = \left(\frac{q}{p}\right)^x$ and note second that $U_T = \begin{cases} \left(\frac{q}{p}\right)^b & \text{if } T_b < T_a \\ \left(\frac{q}{p}\right)^a & \text{if } T_a < T_b \end{cases}$. Therefore

$$\begin{aligned} EU_T &= \left(\frac{q}{p}\right)^b P_x(T_b < T_a) + \left(\frac{q}{p}\right)^a P_x(T_a < T_b) \\ &= \left(\frac{q}{p}\right)^b [1 - P_x(T_a < T_b)] + \left(\frac{q}{p}\right)^a P_x(T_a < T_b) \\ &= \left(\frac{q}{p}\right)^b + \left[\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^b\right] P_x(T_a < T_b). \end{aligned}$$

Again, let $T = \min(T_a, T_b) = \min\{t : X_t \in \{a, b\}\}$; by the OST we have

$$\left(\frac{q}{p}\right)^x = EU_0 = EU_T = \left(\frac{q}{p}\right)^b + \left[\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^b\right] P_x(T_a < T_b).$$

Solving for $P_x(T_a < T_b)$, we get

$$P_x(T_a < T_b) = \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^b}{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^b} = \frac{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^x}{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^a}.$$

The last statement follows from the complement rule. □

Note: $P_x(T_a < T_b) + P_x(T_b < T_a) = 1$ (so you really only need to remember formulas for one of these two quantities).

Example: I have $20 and you have $15. We each make a series of $1 bets until one of us goes broke.

1. If we are equally likely to win each bet, what is the probability that you go broke? What amount of money should I expect to end up with?
2. Suppose you are twice as likely as me to win each bet (assume no ties are possible). In this setting, what is the probability you go broke?

A new kind of question: In the previous example, how long will it take for one of us to go broke?

Theorem 2.29 (Wald's First Identity) *Let $\{X_t\}$ be an irreducible, simple random walk. Let $a < x < b$ be integers and suppose $X_0 = x$. Let $T = \min\{T_a, T_b\} = T_{\{a,b\}}$. Then*

$$E[X_T] = x + \mu ET = x + (p - q)ET.$$

PROOF By Lemma 2.27, we know that $\{Y_t\}$ is a martingale, where $Y_t = X_t - t\mu$. By the Optional Stopping Theorem,

$$x = E[X_0] = E[X_0 - 0\mu] = EY_0 = EY_T = E[X_T - T\mu] = EX_T - \mu ET.$$

Solve for EX_T to get the result. □

Usefulness of Wald's First Identity: From the escape probability theorem, we know that if the walk is biased,

$$P(X_T = a) = P_x(T_a < T_b) = \frac{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^x}{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^a}$$

$$P(X_T = b) = P_x(T_b < T_a) = \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^a}{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^a}$$

so

$$E[X_T] =$$

and therefore, since $EX_T = x + (p-q)ET$,

$$ET = \frac{E[X_T] - x}{p - q} =$$

Example: Return to the earlier example (I have \$20 and you have \$15. We each make a series of \$1 bets until one of us goes broke.) How long will it take one of us to go broke, if you are twice as likely as I am to win each bet?

Recall: We previously showed that the amount of money I expect to end up with is $E[X_T] = 35\left(\frac{1-2^{20}}{1-2^{35}}\right) \approx .001$. Thus

Question: What if we are equally likely to win each bet?

Repeating the same logic doesn't work:

So in this setting, we need another fact to answer the question:

Theorem 2.30 (Wald's Second Identity) *Let $\{X_t\}$ be a simple, irreducible unbiased random walk. Let $a < x < b$ be integers and suppose $X_0 = x$. Let $T = \min\{T_a, T_b\} = T_{\{a,b\}}$. Then*

$$Var(X_T) = Var(S_j) \cdot ET = \sigma^2 ET.$$

PROOF By Wald's First Identity, we have $EX_T = x + \mu ET = x + 0ET = x$. Therefore by the variance formula we see

$$Var(X_T) = E[X_T^2] - (EX_T)^2 = E[X_T^2] - x^2,$$

and therefore

$$x^2 = E[X_T^2] - Var(X_T). \tag{2.1}$$

By Lemma 2.27, we know that $\{Z_t\}$ is a martingale, where

$$Z_t = (X_t - t\mu)^2 - t\sigma^2 = X_t^2 - t\sigma^2.$$

Observe that $EZ_0 = E[(X_0 - 0\mu)^2 - 0\sigma^2] = E[X_0^2] = x^2$. Therefore, applying the OST, we have

$$x^2 = EZ_0 = EZ_T = E[X_T^2 - T\sigma^2] \tag{2.2}$$
$$= E[X_T^2] - \sigma^2 ET. \tag{2.3}$$

In Equations (2.1) and (2.3) above we have found x^2 two different ways. This means

$$E[X_T^2] - Var(X_T) = x^2 = E[X_T^2] - \sigma^2 ET.$$

Subtract $E[X_T^2]$ from both sides and multiply through by (-1) to obtain Wald's Second Identity. □

Usefulness of Wald's Second Identity: Suppose $\{X_t\}$ is a simple, unbiased, random walk with $r \neq 1$. From the escape probability theorems, we know

$$P(X_T = a) = P_x(T_a < T_b) = \frac{b-x}{b-a} \qquad P(X_T = b) = P_x(T_b < T_a) = \frac{x-a}{b-a}$$

so

$$E[X_T] =$$

$$E[X_T^2] =$$

$$Var(X_T) = E[X_T^2] - (E[X_T])^2 =$$

Also,

$$Var(S_j) = E[S_j^2] - E[S_j] = E[S_j^2] =$$

and therefore

$$ET = \frac{Var(X_T)}{Var(S_j)} =$$

Theorem 2.31 (Wald's Third Identity) *Let $\{X_t\}$ be an irreducible, simple random walk. Let $a < x < b$ be integers and suppose $X_0 = 0$. Let $T = \min\{T_a, T_b\} = T_{\{a,b\}}$. Then*

$$E\left[\frac{e^{\theta X_T}}{[M_{S_j}(\theta)]^T}\right] = 1.$$

PROOF HW

Changing gears, we are now in a position to derive formulas for $f_{x,y}$ when $\{X_t\}$ is a random walk. These formulas are rather famous and known by the name "Gambler's Ruin":

Theorem 2.32 (Gambler's Ruin) *Let $\{X_t\}$ be an irreducible, simple random walk on $\mathbb{Z}$. Let a and x be distinct integers. Then*

- *if $p = q$ (i.e. the walk is unbiased), then $f_{x,a} = P_x(T_a < \infty) = 1$.*
- *if $p > q$ (i.e. the walk is positively biased), then*

$$f_{x,a} = P_x(T_a < \infty) = \begin{cases} 1 & \text{if } a > x \\ \left(\frac{q}{p}\right)^{x-a} & \text{if } a < x \end{cases}$$

- *if $p < q$ (i.e. the walk is negatively biased),, then*

$$f_{x,a} = P_x(T_a < \infty) = \begin{cases} 1 & \text{if } a < x \\ \left(\frac{p}{q}\right)^{a-x} & \text{if } a > x \end{cases}$$

PROOF *Case 1:* Suppose that $a > x$. To say that $X_t = a$ for some t means that there must be some number n (n is probably very, very negative) so that the walk hits a before n. That means

$$\begin{aligned} f_{x,a} = P_x(T_a < \infty) &= \lim_{n\to-\infty} P_x(T_a < T_n) \\ &= \begin{cases} \lim_{n\to-\infty} \frac{x-n}{a-n} & \text{if } p = q \\ \lim_{n\to-\infty} \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^n}{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^n} & \text{if } p \neq q \end{cases} \\ &= \begin{cases} 1 & \text{if } p = q \\ \frac{1-0}{1-0} & \text{if } p > q \\ \lim_{n\to-\infty} \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^n}{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^n} & \text{if } p < q \end{cases} \\ &= \begin{cases} 1 & \text{if } p \geq q \\ \lim_{n\to-\infty} \frac{\left(\frac{q}{p}\right)^x - 0}{\left(\frac{q}{p}\right)^a - 0} & \text{if } p < q \end{cases} \\ &= \begin{cases} 1 & \text{if } p \geq q \\ \left(\frac{q}{p}\right)^{x-a} & \text{if } p < q \end{cases} \\ &= \begin{cases} 1 & \text{if } p \geq q \\ \left(\frac{p}{q}\right)^{a-x} & \text{if } p < q \end{cases} \end{aligned}$$

Case 2: Now suppose that $a < x$. This is similar (HW problem). □

Why is this called "Gambler's Ruin"? Suppose a gambler brings $\$50$ to a casino and makes a series of $\$1$ bets in a game where he has a 50% chance of winning each bet, and a 50% chance of losing each bet. The Gambler's Ruin Theorem says

> **Theorem 2.33 (Recurrence/transience of random walk on $\mathbb{Z}$)** *Let $\{X_t\}$ be an irreducible, simple random walk on $\mathbb{Z}$. Then $\{X_t\}$ is recurrent if and only if the random walk is unbiased.*

PROOF Since $\{X_t\}$ is irreducible, $\{X_t\}$ is irreducible if and only if 0 is recurrent. By direct calculation,

$$f_0 = P_0(T_0 < \infty)$$

$$\begin{aligned} &= P_0(T_0 < \infty \,|\, X_1 = -1)P_0(X_1 = -1) \\ &\quad + P_0(T_0 < \infty \,|\, X_1 = 0)P_0(X_1 = 0) \qquad \text{(Law of Total Prob.)} \\ &\qquad + P_0(T_0 < \infty \,|\, X_1 = 1)P_0(X_1 = 1) \end{aligned}$$

$$= P_1(T_0 < \infty)q + 1 \cdot r + P_1(T_0 < \infty)p$$

$$= \begin{cases} 1 \cdot q \quad +r \quad + \left(\frac{q}{p}\right) p & \text{if } p > q \\ 1 \cdot q \quad +r \quad +1 \cdot p & \text{if } p = q \\ \left(\frac{p}{q}\right) q \quad +r \quad +1 \cdot p & \text{if } p < q \end{cases} \qquad \text{(Gambler's Ruin)}$$

$$= \begin{cases} 2q + r & \text{if } p > q \\ 1 & \text{if } p = q \\ 2p + r & \text{if } p < q \end{cases}$$

Therefore 0 is recurrent iff $f_{0,0} = 1$ iff $p = q$. $\square$

2.6 Random walk in higher dimensions

Notation: The vector $\mathbf{e}_j \in \mathbb{R}^d$ is the vector $(0, 0, 0, ..., 0, 1, 0, ..., 0)$ which has a 1 in the j^{th} place and zeros everywhere else. (Thus $-\mathbf{e}_j$ is $(0, 0, ..., 0, -1, 0, ..., 0)$.)

In this section we consider simple, unbiased random walks in $\mathbb{Z}^d$. This means that we assume $\{X_t\}$ is a Markov chain taking values in $\mathbb{Z}^d$ with

- $X_0 = (0, 0, ..., 0) = \mathbf{0}$;
- $P(\mathbf{x}, \mathbf{y}) = \begin{cases} \frac{1}{2d} & \text{if } \mathbf{x} - \mathbf{y} = \pm \mathbf{e}_j \text{ for some } j \\ 0 & \text{else} \end{cases}$.

In other words, you start at the origin and move one unit in one of the coordinate directions (the direction is chosen uniformly) at each step.

These random walks are all irreducible and have period 2.

Example: ($d = 2$) "Drunkard's walk"

Questions: Will the drunk person ever make it home? Will they make it back to the bar? (i.e. is the random walk recurrent?)

Recall the recurrence criterion from Chapter 1: A state $x \in \mathcal{S}$ in any Markov chain is recurrent if and only if $\sum_{n=0}^{\infty} P^n(x, x)$ diverges. So to determine whether a random walk as set up above is recurrent, it is sufficient to check whether or not $\sum_{n=0}^{\infty} P^n(\mathbf{0}, \mathbf{0})$ converges or diverges.

Dimension 1: unbiased random walk on $\mathbb{Z}$

Here, $P(x, x+1) = P(x, x-1) = \frac{1}{2}$ for all x.

Now

$$P^n(\mathbf{0},\mathbf{0}) = \begin{cases} \binom{2k}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^k & \text{if } n = 2k \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$$

So

$$\begin{aligned}\sum_{n=0}^{\infty} P^n(\mathbf{0},\mathbf{0}) = \sum_{k=0}^{\infty} P^{2k}(\mathbf{0},\mathbf{0}) &= \sum_{k=0}^{\infty} \binom{2k}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^k \\ &\approx \sum_{k=0}^{\infty} \frac{4^k}{\sqrt{\pi k}}\left(\frac{1}{4^k}\right) \quad \text{(by a HW problem from 414)} \\ &= \sum_{k=0}^{\infty} \frac{1}{\sqrt{\pi k}}\end{aligned}$$

which diverges. Hence unbiased simple random walk in dimension 1 is recurrent.

Dimension 2: unbiased random walk on $\mathbb{Z}^2$

Here, the probability of moving in any particular direction on any one step is $\frac{1}{4}$.

Now

$$P^n(\mathbf{0},\mathbf{0}) = \begin{cases} \sum_{l=0}^{k} \frac{(2k)!}{l!^2(k-l)!^2}\left(\frac{1}{4}\right)^{2k} & \text{if } n = 2k \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$$

So

$$
\begin{aligned}
\sum_{n=0}^{\infty} P^n(\mathbf{0},\mathbf{0}) = \sum_{k=0}^{\infty} P^{2k}(\mathbf{0},\mathbf{0}) &= \sum_{k=0}^{\infty}\sum_{l=0}^{k} \frac{(2k)!}{l!^2(k-l)!^2}\left(\frac{1}{4}\right)^{2k} \\
&= \sum_{k=0}^{\infty} \frac{1}{16^k} \sum_{l=0}^{k} \frac{(2k)!}{(k!)^2} \cdot \binom{k}{l}^2 \\
&= \sum_{k=0}^{\infty} \frac{1}{16^k} \binom{2k}{k} \sum_{l=0}^{k} \cdot \binom{k}{l}^2 \\
&= \sum_{k=0}^{\infty} \frac{1}{16^k} \binom{2k}{k}^2 \\
&\approx \sum_{k=0}^{\infty} \frac{1}{16^k} \cdot \left(\frac{4^k}{\sqrt{\pi k}}\right)^2 \\
&= \sum_{k=0}^{\infty} \frac{1}{\pi k}
\end{aligned}
$$

which diverges. Hence unbiased simple random walk in dimension 2 is recurrent.

Dimension 3: unbiased random walk on $\mathbb{Z}^3$

Here the picture looks like

If you did the same kind of stuff as was done in dimensions 1 and 2, you'd get

$$
\sum_{n=0}^{\infty} P^n(\mathbf{0},\mathbf{0}) \approx \sum_{k=0}^{\infty} \frac{1}{(\pi k)^{3/2}}
$$

which converges. Hence unbiased simple random walk in dimension 3 is transient.

To summarize, we have the following characterization of simple, unbiased random walk as recurrent or transient:

Theorem 2.34 (Polya's Theorem) *Let $\{X_t\}$ be simple, unbiased random walk on $\mathbb{Z}^d$ as described in this section. Then:*

1. *If $d = 1$ or 2, then $\{X_t\}$ is recurrent.*
2. *If $d > 2$, then $\{X_t\}$ is transient.*

Chapter 3

Stationary distributions

3.1 Stationary and steady-state distributions

Recall: A Markov chain is determined by two things:

-
-

From this, you get time n distributions π_n which give the probability of each state at time n:

$$\pi_n(y) = P(X_n = y) = \sum_{x \in \mathcal{S}} \pi_{n-1}(x) P(x,y) = \sum_{x \in \mathcal{S}} \pi_0(x) P^n(x,y)$$

(i.e. $\pi_n = \pi_0 P^n$ if $\mathcal{S}$ is finite and P is transition matrix)

Motivating question: Can you predict/approximate π_n (for large n) <u>without knowing π_0</u>?

> **Definition 3.1** *A distribution π on $\mathcal{S}$ is called* **stationary** *(with respect to $\{X_t\}$) if for all $y \in \mathcal{S}$,*
> $$\sum_{x \in \mathcal{S}} \pi(x) P(x,y) = \pi(y).$$

Note: If $\mathcal{S}$ is finite (say $\mathcal{S} = \{1, 2, 3, ..., d\}$, to say π is stationary means (in matrix multiplication terminology)

$$\pi P = \pi$$

if we write $\pi = \begin{pmatrix} \pi(1) & \pi(2) & \cdots & \pi(d) \end{pmatrix}$.

Lemma 3.2 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If π is a stationary distribution, then for all $n > 0$ and all $y \in \mathcal{S}$, we have*

$$\pi(y) = \sum_{x \in \mathcal{S}} \pi(x) P^n(x, y).$$

(So if $\mathcal{S}$ is finite, this means $\pi = \pi\, P^n$ for all n.)

PROOF Definition of "stationary" + induction on n.

Lemma 3.3 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. An initial distribution π_0 is stationary if and only if the time n distributions are the same for every n.*

PROOF ($\Rightarrow$) Assume π_0 is stationary. Then

($\Leftarrow$) Assume the time n distributions are the same for every n. Then

Put another way, this lemma says that *stationary distributions are those which do not change as time passes*.

Definition 3.4 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. A distribution π on $\mathcal{S}$ is called* **steady-state** *(with respect to $\{X_t\}$) if*

$$\lim_{n \to \infty} P^n(x, y) = \pi(y) \text{ for all } x, y \in \mathcal{S}.$$

Idea:

The idea on the previous page is made precise in the following theorem:

Theorem 3.5 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Suppose π is a steady-state distribution for $\{X_t\}$. Then for any initial distribution π_0,*

$$\lim_{n\to\infty} \pi_n(y) = \lim_{n\to\infty} P(X_n = y) = \pi(y) \; \forall \, y \in \mathcal{S}.$$

PROOF

$$\pi_n(y) = P(X_n = y) = \sum_{x\in\mathcal{S}} \pi_0(x) P^n(x, y)$$

Big picture questions related to stationary and steady-state distributions: Given Markov chain $\{X_t\}$ with transition function P,

1.

2.

3.

4.

Theorem 3.6 (Uniqueness of steady-state distributions) *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If the Markov chain has a steady-state distribution π, then π is the only possible stationary distribution for $\{X_t\}$.*

Proof: Suppose $\pi_0 \neq \pi$ is stationary. Since $\pi_0 \neq \pi$, there is $y \in \mathcal{S}$ such that $\pi_0(y) \neq \pi(y)$.

Use π_0 as the initial distribution; then the time n distribution of state y is $\pi_n(y) = \pi_0(y)$ by stationarity. Thus

$$\lim_{n\to\infty} \pi_n(y) = \lim_{n\to\infty} \pi_0(y) = \pi_0(y) \neq \pi(y);$$

this contradicts the preceding proposition since π is steady-state. $\square$

Example: Consider a Markov chain with $\mathcal{S} = \{1, 2\}$ whose transition matrix is

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

(There is no relationship between p and q in this example.) Find all stationary distributions of this Markov chain.

In general, you find stationary distributions for finite state-space Markov chains by solving a system of linear equations corresponding to $\pi P = \pi$ as above.

Example: Let $\{X_t\}$ be simple, unbiased random walk on $\mathbb{Z}$ ($p = q = \frac{1}{2}$). Find all stationary distributions of $\{X_t\}$.

Example: Find all stationary distributions of $\{X_t\}$ if the Markov chain has transition matrix

$$\left(\begin{array}{ccc} \frac{1}{7} & \frac{4}{7} & \frac{2}{7} \\ 0 & \frac{5}{7} & \frac{2}{7} \\ \frac{3}{7} & \frac{1}{7} & \frac{3}{7} \end{array} \right).$$

3.2 Positive and null recurrence

A new type of convergence

Recall: A sequence $\{a_n\}$ is said to **converge** to limit L if $\lim\limits_{n\to\infty} a_n = L$. (We write $a_n \to L$ to represent this.)

Example: $\frac{1}{n} \to 0$.

Example: $\frac{n+1}{n-1} \to 1$.

Example: The sequence $\{0, 1, 2, 0, 1, 2, 0, 1, 2, 0, 1, 2, ...\}$ does not converge. However, this sequence does have some regular behavior:

Definition 3.7 *Let $\{a_n\}$ be a sequence of real numbers. The* **sequence of Cesàro averages** *of $\{a_n\}$ is the sequence $\{b_n\}$ defined by setting*

$$b_n = \frac{1}{n}\sum_{j=1}^{n} a_j$$

for all n. We say $\{a_n\}$ **converges in the Cesàro sense** *to L if the Cesàro averages converge to L, i.e. if*

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} a_j = L.$$

We write $a_n \overset{Ces}{\to} L$ to represent this.

Example: the sequence $\{a_n\} = \{0, 1, 2, 0, 1, 2, ...\}$ converges in the Cesàro sense to 1.

Example: Strong Law of Large Numbers

Restated, this says says that the Cesàro averages of i.i.d. r.v.s with finite mean converge to the mean with probability 1.

Facts:

$$a_n \to L \text{ in the usual sense} \Rightarrow a_n \overset{Ces}{\to} L$$

$$a_n \overset{Ces}{\to} L \text{ and } \{a_n\} \text{ converges } \Rightarrow a_n \to L$$

"Cesàro convergence is weaker than usual convergence"

Application to Markov chains: For any Markov chain, we will see that although $\lim_{n\to\infty} P^n(x,y)$ may not exist, the sequence $P^n(x,y)$ converges in the Cesàro sense for any $x,y \in \mathcal{S}$ (and the value to which the Cesáro averages converge has a lot to do with stationary and steady-state distributions).

Definition 3.8 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$ and transition function P.*

1. *Given $y \in \mathcal{S}$, define $V_{y,n} = \#(t \in \{1,2,...,n\} : X_t = y)$. This is a r.v. taking values in $\{0,1,2,...,n\}$ called the* **number of visits to** y **up to time** n.
2. *Given $y \in \mathcal{S}_R$, define $m_y = E_y(T_y)$. m_y is a number (possibly ∞) called the* **mean return time** *to y.*
3. *A recurrent state y is called* **null recurrent** *if $m_y = \infty$. The set of null recurrent states of $\{X_t\}$ is denoted $\mathcal{S}_N$. If all the states of $\{X_t\}$ are null recurrent, $\{X_t\}$ is called* **null recurrent**.
4. *A recurrent state y is called* **positive recurrent** *if $m_y < \infty$. The set of positive recurrent states of $\{X_t\}$ is denoted $\mathcal{S}_P$. If all the states of $\{X_t\}$ are positive recurrent, $\{X_t\}$ is called* **positive recurrent**.

Thus $E_x(V_{y,n})$ is the expected number of visits to state y in the time interval $[1,n]$, given that you start at x.

Note: It makes no sense to talk about mean return times of transient states, because if $y \in \mathcal{S}_T$,

$$P_y(T_y = \infty) > 0 \Rightarrow E_y(T_y) = \infty \text{ automatically.}$$

Lemma 3.9 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$ and transition function P. Then for any $x,y \in \mathcal{S}$ and any $n > 0$,*

$$E_x(V_{y,n}) = \sum_{m=1}^{n} P^m(x,y).$$

PROOF In the context of proving Theorem 1.26 earlier (see p. 21), we proved this statement with $n = \infty$; the proof is the same (just replace all the ∞ with n). □

Theorem 3.10 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $y \in \mathcal{S}$.*

1. (a) *If $T_y < \infty$ (i.e. if the chain hits y), $\lim_{n\to\infty} \frac{V_{y,n}}{n} = \frac{1}{m_y}$.*

   (b) *If $T_y = \infty$ (i.e. the chain never hits y), then $\lim_{n\to\infty} \frac{V_{y,n}}{n} = 0$.*

2. $\lim_{n\to\infty} \frac{E_x(V_{y,n})}{n} = \frac{f_{x,y}}{m_y}$ *for all $x \in \mathcal{S}$.*

3. $P^n(x,y) \stackrel{Ces}{\to} \frac{f_{x,y}}{m_y}$ *for all $x \in \mathcal{S}$.*

(These limits hold with probability 1.)

Proof: Statement 1 (b) is obvious. Also, (2 ⇒ 3) follows from Lemma 3.10.

Next, we prove that Statement 1 implies Statement 2:

$$\lim_{n\to\infty} \frac{E_x(V_{y,n})}{n} = \lim_{n\to\infty} E_x\left[\frac{V_{y,n}}{n}\right]$$

Last, we prove Statement 1 (a):

Assume WLOG that you start at state y (since you must hit y at some point). Define the following:

- $T_y^r = \min\{n \geq 1 : V_{y,n} = r\} =$ time of r^{th} return to y
- $W_y^1 = T_y^1$
- $W_y^j = T_y^j - T_y^{j-1}$ for all $j \geq 2$

Notice that the W_y^j are i.i.d., each with mean m_y.

$$\text{Strong Law of Large Numbers} \Rightarrow P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} W_y^j = m_y\right) = 1$$
$$\Rightarrow P\left(\lim_{n\to\infty} \frac{T_y^n}{n} = m_y\right) = 1 \quad (*)$$

Corollary 3.11 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$ and let $C \subseteq \mathcal{S}$ be a communicating class of recurrent states. Then for all $x, y \in C$,*

$$\lim_{n\to\infty} \frac{E_x(V_{y,n})}{n} = \frac{1}{m_y}.$$

Furthermore, if $P(X_0 \in C) = 1$, then $\lim_{n\to\infty} \frac{V_{y,n}}{n} = \frac{1}{m_y}\ \forall\, y \in C$.

Corollary 3.12 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $y \in \mathcal{S}$ is null recurrent, then $P^n(x,y) \overset{Ces}{\to} 0$ for all $x \in \mathcal{S}$.*

PROOF

$$P^n(x,y) \overset{Ces}{\to} \frac{f_{x,y}}{m_y} = \frac{f_{x,y}}{\infty} = 0. \ \square$$

Corollary 3.13 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $y \in \mathcal{S}$ is positive recurrent, then $P^n(y,y) \overset{Ces}{\to} \frac{1}{m_y}$.*

PROOF

$$P^n(y,y) \overset{Ces}{\to} \frac{f_{y,y}}{m_y} = \frac{1}{m_y}. \square$$

Note: The previous two corollaries provide a new distinction between positive recurrent and null recurrent states. If $y \in \mathcal{S}$ is null recurrent (or transient), then $P^n(y,y) \overset{Ces}{\to} 0$ but if $y \in \mathcal{S}$ is positive recurrent, then $P^n(y,y) \overset{Ces}{\to} \frac{1}{m_y} > 0$.

Theorem 3.14 (Positive recurrent states lead only to positive recurrent states) *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $x \in \mathcal{S}$ is positive recurrent and $x \to y$, then y is also positive recurrent.*

PROOF By previous result, $y \to x$. Thus there are n_1 and n_2 such that $P^{n_1}(x,y) > 0$ and $P^{n_2}(y,x) > 0$. Therefore

$$P^{n_1+m+n_2}(y,y) \geq P^{n_1}(x,y)P^m(x,x)P^{n_2}(y,x) \quad \text{for all } m \geq 0$$
$$\Rightarrow \frac{1}{n}\sum_{m=1}^{n} P^{n_1+m+n_2}(y,y) \geq \frac{1}{n}P^{n_1}(x,y)P^{n_2}(y,x)\sum_{m=1}^{n} P^m(x,x)$$

Corollary 3.15 (Null recurrent states lead only to null recurrent states) *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $x \in \mathcal{S}$ is null recurrent and $x \to y$, then y is also null recurrent.*

PROOF By previous result, y is recurrent. If y is positive recurrent, then by the above theorem x is positive recurrent, a contradiction. Thus y must be null recurrent. □

Corollary 3.16 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $C \subseteq \mathcal{S}$ is a communicating class, then (every $x \in C$ is transient) or (every $x \in C$ is null recurrent) or (every $x \in C$ is positive recurrent).*

Theorem 3.17 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $C \subseteq \mathcal{S}$ is a finite communicating class, then every $x \in C$ is positive recurrent.*

PROOF

Corollary 3.18 *Any irreducible Markov chain with a finite state space is positive recurrent.*

Existence and uniqueness of stationary distributions

The next result will not be proven; it is a fact from a branch of mathematics called *real analysis*.

Theorem 3.19 (Bounded Convergence Theorem for Sums) *Let a(x) be nonnegative numbers such that* $\sum_x a(x) < \infty$. *Fix* $B > 0$ *and let* $b_n(x)$ *be numbers such that* $|b_n(x)| \leq B$ *for all* x *and* n *and*

$$\lim_{n\to\infty} b_n(x) = b(x) \text{ for all } x.$$

Then

$$\sum_x a(x)b_n(x) \to \sum_x a(x)b(x).$$

Theorem 3.20 *Let* $\{X_t\}$ *be a Markov chain with state space* $\mathcal{S}$. *If* $x \in \mathcal{S}$ *is either transient or null recurrent, then for any stationary distribution* π, $\pi(x) = 0$.

PROOF

$$x \in \mathcal{S}_T \cup \mathcal{S}_N \Rightarrow \lim_{n\to\infty} \frac{1}{n} E_z(V_{x,n}) = 0 \text{ for all } z \in \mathcal{S}.$$

If π is stationary, then

Corollary 3.21 (Nonexistence of stationary distributions) .

1. *A transient Markov chain has no stationary distributions.*
2. *A null recurrent Markov chain has no stationary distributions.*

PROOF By the preceding theorem, a stationary distribution π for such a Markov chain would have to satisfy $\pi(x) = 0$ for all $x \in \mathcal{S}$. But then $\sum_{x \in \mathcal{S}} \pi(x) = 0 \neq 1$ so π would not be a distribution. □

Theorem 3.22 (Existence/uniqueness of stationary distributions) *Let $\{X_t\}$ be an irreducible Markov chain with state space $\mathcal{S}$. $\{X_t\}$ has a stationary distribution if and only if $\{X_t\}$ is positive recurrent, in which case the Markov chain has a unique stationary distribution π defined by $\pi(x) = \frac{1}{m_x}$ for all $x \in \mathcal{S}$.*

PROOF ($\Leftarrow$) Assume $\{X_t\}$ is positive recurrent. First, observe that for any $x, z \in \mathcal{S}$,

$$\lim_{n\to\infty} \frac{1}{n} E_z(V_{x,n}) \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m(z,x) = \frac{f_{z,x}}{m_x} = \frac{1}{m_x}.$$

Second, we show that if π is stationary, then $\pi(x)$ must equal $\frac{1}{m_x}$ for all x. On the previous page, we showed that if π is stationary, then for all x we have

$$\pi(x) = \sum_{z \in \mathcal{S}} \pi(z) \frac{1}{m_x}$$

Third, we show that $\pi(x) = \frac{1}{m_x}$ is actually a distribution on $\mathcal{S}$:

$$\begin{aligned}
\sum_{x \in \mathcal{S}} P^m(z,x) &= 1 \qquad \forall z \in \mathcal{S}, \forall m > 0 \\
\Rightarrow \frac{1}{n} \sum_{m=1}^{n} \sum_{x \in \mathcal{S}} P^m(z,x) &= \frac{1}{n} \sum_{m=1}^{n} 1 = 1 \\
\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \sum_{x \in \mathcal{S}} P^m(z,x) &= 1 \\
\sum_{x \in \mathcal{S}} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m(z,x) &= 1 \\
\sum_{x \in \mathcal{S}} \frac{1}{m_x} &= 1.
\end{aligned}$$

Fourth, we show that the distribution defined by $\pi(x) = \frac{1}{m_x}$ is in fact stationary (we have to verify that $\sum_{x \in \mathcal{S}} \pi(x)P(x,y) = \pi(y)$):

Case 1: $\mathcal{S}$ is finite:

Case 2: $\mathcal{S}$ is infinite:

Let $\mathcal{S}' \subseteq \mathcal{S}$ be an arbitrary finite subset of $\mathcal{S}$. Repeating Case 1 with $\mathcal{S}'$ instead of $\mathcal{S}$, we get

$$\begin{aligned} &\sum_{x \in \mathcal{S}'} P^m(z,x)P(x,y) \leq P^{m+1}(x,y) \\ &\Rightarrow \sum_{x \in \mathcal{S}'} \pi(x)P(x,y) \leq \pi(y). \end{aligned}$$

Last, the $(\Rightarrow)$ direction was proven earlier.

Corollary 3.23 *Any irreducible Markov chain on a finite state space has a unique stationary distribution.*

Theorem 3.24 (Ergodic Theorem for Markov chains) *Let $\{X_t\}$ be an irreducible, positive recurrent Markov chain with state space $\mathcal{S}$ and let π be its unique stationary distribution. Then for all $x \in \mathcal{S}$,*

$$P\left(\lim_{n\to\infty} \frac{V_{x,n}}{n} = \pi(x)\right) = 1.$$

A picture to explain the ergodic theorem:

Stationary distributions for non-irreducible Markov chains

Definition 3.25 *A distribution π on $\mathcal{S}$ is* **supported** *or* **concentrated** *on a subset $C \subseteq \mathcal{S}$ if $\pi(x) = 0$ for all $x \notin C$.*

Example: If $\mathcal{S} = \{1, 2, 3, 4\}$ and $\pi = (\frac{1}{2}, 0, \frac{1}{2}, 0)$, we say π is supported on $\{1, 3\}$.

Definition 3.26 *Suppose $\pi_1, \pi_2, \pi_3, \ldots$ are all distributions on a set $\mathcal{S}$ (there could be finitely or countably many distributions). A* **convex combination** *of these distributions is another distribution of the form*

$$\sum_j \alpha_j \pi_j$$

where the α_j are nonnegative numbers satisfying $\sum_j \alpha_j = 1$.

Lemma 3.27 *A convex combination of distributions is a distribution.*

PROOF If

$$\pi = \sum_j \alpha_j \pi_j,$$

then

$$\sum_{x \in \mathcal{S}} \pi(x) = \sum_{x \in \mathcal{S}} \sum_j \alpha_j \pi_j(x) = \sum_j \alpha_j \sum_{x \in \mathcal{S}} \pi_j(x) = \sum_j \alpha_j \cdot 1 = 1.$$

Since all the α_j are nonnegative, then $\pi(x) \geq 0$ for all x as well, so π is a distribution.
□

Special case: A convex combination of two distributions π_1 and π_2 is a distribution of the form

$$\alpha\pi_1 + (1-\alpha)\pi_2$$

where $\alpha \in [0, 1]$.

Theorem 3.28 (Convex combinations of stationary distributions are stationary) *Suppose $\pi_1, \pi_2, \pi_3, \ldots$ are all stationary distributions for a Markov chain $\{X_t\}$. Then any convex combination of the π_j is also a stationary distribution for $\{X_t\}$.*

PROOF HW

Corollary 3.29 (Number of stationary distributions) *A Markov chain must have either zero, one, or infinitely many stationary distributions.*

PROOF Suppose the Markov chain has two different stationary distributions, say π_1 and π_2. Then for any $\alpha \in [0, 1]$,

$$\alpha\pi_1 + (1-\alpha)\pi_2$$

is also a stationary distribution. Since there are infinitely many choices for α, the Markov chain will have infinitely many stationary distributions. □

Summary of existence/uniqueness of stationary distributions for Markov chains

Consider a Markov chain $\{X_t\}$ with state space $\mathcal{S}$. We can write

$$\mathcal{S} = \mathcal{S}_T \bigcup \mathcal{S}_R = \mathcal{S}_T \bigcup \left(\mathcal{S}_N \bigcup \mathcal{S}_P\right) \quad \text{(disjoint union)}$$

- If $\mathcal{S}_P = \emptyset$, then $\{X_t\}$ has no stationary distribution.
- If $\mathcal{S}_P \neq \emptyset$ consists of one communicating class, then $\{X_t\}$ has a unique stationary distribution π defined by

$$\pi(x) = \begin{cases} \frac{1}{m_x} & \text{if } x \in \mathcal{S}_P \\ 0 & \text{else} \end{cases}$$

- If $\mathcal{S}_P \neq \emptyset$ consists of more than one communicating class, then for each communicating class $C \subseteq \mathcal{S}_P$ there is a unique (stationary distribution supported on C) defined by

$$\pi_C(x) = \begin{cases} \frac{1}{m_x} & \text{if } x \in C \\ 0 & \text{else} \end{cases}$$

  Convex combinations of these π_C are also stationary, so $\{X_t\}$ has infinitely many stationary distributions. (All stationary distributions are convex combinations of these π_C.)

(This solves the big picture questions 1 and 2 from the beginning of the chapter.)

Example: Find all stationary distributions of the Markov chain with transition matrix

$$\begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 & \frac{3}{4} & 0 \\ \frac{1}{8} & \frac{1}{2} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{3}{4} \end{pmatrix}$$

Example: Find all stationary distributions of the Markov chain $\{X_t\}$ with state space $\mathcal{S} = \{0, 1, 2, 3, ...\}$ and transition function P defined by

$$P(x,y) = \begin{cases} \frac{1}{2} & \text{if } y = 0 \\ \frac{1}{4} & \text{if } y = x+1 \\ \frac{1}{4} & \text{if } y = x+2 \\ 0 & \text{else} \end{cases}$$

3.3 Periodicity and convergence issues

Question: Which stationary distributions are steady-state?

First observation: We saw earlier that if π is steady-state, then π is the **only** stationary distribution of $\{X_t\}$. Thus if $\mathcal{S}_P$ contains more than one communicating class, $\{X_t\}$ has infinitely many stationary distributions, and none of these can be steady-state.

We also know that for any transient or null recurrent state x, $\pi(x) = 0$ for any stationary (hence any steady-state distribution).

So henceforth we assume $\{X_t\}$ is an irreducible, positive recurrent Markov chain. Given this,

we know $\quad P^n(x, y) \overset{Ces}{\to} \pi(y) \quad \forall x, y \in \mathcal{S}.$

When does $\quad P^n(x, y) \to \pi(y) \quad \forall x, y \in \mathcal{S}?$

(Wouldn't it be great if this was always true?)

Unfortunately, even for irreducible, positive recurrent Markov chains, the stationary distribution may not be steady-state.

Example: Consider the Markov chain with state space $\{1, 2\}$ and transition matrix

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

Definition 3.30 *Let a and b be integers. We say a* **divides** *b (and write $a|b$) if b is a multiple of a. The* **greatest common divisor** *of a set E of integers, denoted $\gcd E$, is the largest integer dividing every number in that set.*

Examples:

Definition 3.31 *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $x \in \mathcal{S}$ be such that $f_x > 0$ (equivalently, $P^n(x,x) > 0$ for some $n \geq 1$; equivalently, $x \to x$). The* **period** *of x, denoted by d_x, is the largest integer which divides every n for which $P^n(x,x) > 0$. More formally,*

$$d_x = \gcd\{n : P^n(x,x) > 0\}.$$

Note: If $P(x,x) > 0$, then $d_x|1$, so $d_x = 1$.

Example: In simple random walk on $\mathbb{Z}$ with $r = 0$, $d_x = 2$ for all x.

Theorem 3.32 (States lead only to states of the same period) *Suppose $\{X_t\}$ is a Markov chain with state space $\mathcal{S}$. Let $x, y \in \mathcal{S}$ be such that $x \to y$ and $y \to x$. Then $d_x = d_y$.*

PROOF

$$x \to y \Rightarrow \exists\, n_1 \text{ s.t. } P^{n_1}(x,y) > 0$$
$$y \to x \Rightarrow \exists\, n_2 \text{ s.t. } P^{n_2}(y,x) > 0.$$

Therefore
$$P^{n_1+n_2}(x,x) \geq P^{n_1}(x,y)P^{n_2}(y,x) > 0 \Rightarrow d_x \,|\, (n_1+n_2).$$

Let n be such that $P^n(y,y) > 0$. Then

$$P^{n_1+n+n_2}(x,x) \geq P^{n_1}(x,y)P^n(y,y)P^{n_2}(y,x) > 0 \Rightarrow d_x \,|\, (n_1+n+n_2).$$

Now if d_x divides both $n_1 + n_2$ and $n_1 + n + n_2$, then d_x divides the difference, so $d_x \,|\, n$.

Corollary 3.33 *If $\{X_t\}$ is an irreducible Markov chain, all states have the same period.*

Definition 3.34 *An irreducible Markov chain with state space $\mathcal{S}$ is called* **aperiodic** *if $d_x = 1$ for all $x \in \mathcal{S}$ and is called* **periodic with period** *d if $d_x = d > 1$ for all $x \in \mathcal{S}$.*

Examples:

Theorem 3.35 *Suppose $\{X_t\}$ is an irreducible, aperiodic Markov chain. Then, for every $x, y \in \mathcal{S}$, there is a number N such that $P^n(x, y) > 0$ for all $n \geq N$.*

PROOF Let $I \subset \mathbb{N}$ be defined by $I = \{n : P^n(x, y) > 0\}$; I is the set of times that you can get from state x to state y. We know $1 = d = \gcd I$.

Claim: There is a number n_1 such that $n_1 \in I$ and $n_1 + 1 \in I$.

Proof: Suppose not; then there is an integer $k \geq 2$ which is the smallest gap between two consecutive numbers in I. Since $\{X_t\}$ is aperiodic, k is not the period of $\{X_t\}$ so k cannot divide some number in I. Let $n_1 \in I$ be such that $n_1 + k \in I$. Now let $m_1 \in I$ be a number which is not divisible by k. Write $m_1 = mk + r$ where $r \in \{1, 2, ..., k - 1\}$. We know

$$(m+1)(n_1 + k) \in I \quad \text{and} \quad m_1 + (m+1)n_1 \in I$$

but the difference of these numbers is

$$mk + k - m_1 = k - r \in \{1, 2, ..., k - 1\}.$$

This contradicts the definition of k, so $k = 1$, proving the claim (as the smallest gap between two consecutive numbers in I is 1).

Now, we know there is an n_1 such that $n_1 \in I$, $n_1 + 1 \in I$. Let $N = n_1^2$. Then if $n \geq N$, we can divide $n - N$ by n_1 and write

$$n - N = mn_1 + r$$

where $m \in \mathbb{N}$ and $r \in \{0, 1, ..., n_1 - 1\}$. Now

$$n = r(n_1 + 1) + (n_1 - r + m)n_1$$

which is in I since $n_1 + 1 \in I$, $n_1 \in I$. □

Theorem 3.36 (Existence of steady-state distributions) *Let $\{X_t\}$ be an irreducible, positive recurrent Markov chain with state space $\mathcal{S}$. Let π denote its unique stationary distribution. Then:*

1. *If $\{X_t\}$ is aperiodic, then π is steady-state, i.e.*

$$\lim_{n\to\infty} P^n(x, y) = \pi(y) \text{ for all } x, y \in \mathcal{S}.$$

2. *If $\{X_t\}$ has period $d \geq 2$, then for all $x, y \in \mathcal{S}$ there exists an integer $r = r(x, y) \in [0, d)$ such that*
   - *(a) $P^n(x, y) = 0$ unless $n = md + r$ for some $m \in \mathbb{N}$ (i.e. unless $n \equiv r \mod d$)*
   - *(b) $\lim_{m\to\infty} P^{md+r}(x, y) = d \cdot \pi(y)$.*

PROOF Let $\{Y_t\}$ be a Markov chain, independent of $\{X_t\}$, with the same state space and transition function as $\{X_t\}$, where the initial distribution of $\{Y_t\}$ is the stationary distribution π.

Pick $b \in \mathcal{S}$ arbitrarily and set $T = \min\{t \geq 1 : X_t = Y_t = b\}$ (if there is no such t, set $T = \infty$).

Claim: $P(T < \infty) = 1$.

Proof of Claim: HW (this requires aperiodicity of $\{X_t\}$ because it uses Theorem 3.35).

Hint: Consider a Markov chain with state space $\mathcal{S} \times \mathcal{S}$ where the first coordinate is X_t and the second coordinate is Y_t. Explain why this Markov chain is irreducible and positive recurrent; it follows that $P(T < \infty) = 1$ (why?).

Now, define for each t, r.v.s Z_t by

$$Z_t = \begin{cases} X_t & \text{if } t < T \\ Y_t & \text{if } t \geq T \end{cases}$$

$\{Z_t\}$ is a Markov chain with the same initial distribution as $\{X_t\}$ and the same transition function as $\{X_t\}$, therefore $\{Z_t\} = \{X_t\}$. Therefore

$$\begin{aligned} |P(X_t = y) - \pi(y)| &= |P(Z_t = y) - P(Y_t = y)| \\ &= |P(X_t = y \text{ and } t < T) + P(Y_t = y \text{ and } t \geq T) - P(Y_t = y)| \\ &= |P(X_t = y \text{ and } t < T) - P(Y_t = y \text{ and } t < T)| \\ &\leq P(t < T) \to 0 \text{ as } t \to \infty \text{ by the Claim above.} \end{aligned}$$

Therefore $|P(X_t = y) - \pi(y)| \to 0$ as $t \to \infty$, so

$$\lim_{t\to\infty} \pi_t(y) = \lim_{t\to\infty} \sum_{x \in \mathcal{S}} \pi_0(x) P^t(x, y) = \pi(y)$$

for all x and y. By choosing π_0 to be

$$\pi_0(x) = \begin{cases} 1 & \text{if } x = z \\ 0 & \text{else} \end{cases},$$

we see that

$$\lim_{t\to\infty} P^t(z, y) = \pi(y)$$

for all $z \in \mathcal{S}$; thus π is steady-state. This proves statement 1 of the theorem.

To prove statement 2, let m_x be the mean return time of each state x with respect to the Markov chain $\{X_t\}$. Now consider the Markov chain $\{\widetilde{X_t}\}$ with the same initial distribution as $\{X_t\}$ whose transition function is P^d, i.e. let

$$P(\widetilde{X_t} = x) = P(X_{dt} = x).$$

Note that the mean return time for each state with respect to $\{\widetilde{X_t}\}$ is $\frac{m_x}{d}$.

$\{\widetilde{X_t}\}$ is not irreducible; it has d disjoint, positive recurrent communicating classes. Restricting the Markov chain $\{\widetilde{X_t}\}$ to each of these classes gives an aperiodic, positive recurrent, irreducible chain to which we can apply part 1 of this theorem; this gives

$$\lim_{m\to\infty} (P^d)^m(x,x) = \frac{1}{m_x/d} = \frac{d}{m_x},$$

i.e.

$$\lim_{m\to\infty} P^{md}(x,x) = d\pi(x).$$

More generally, if $z \in \mathcal{S}$ is such that $P^d(z,x) > 0$, then z and x belong to the same communicating class of $\{\widetilde{X_t}\}$, so

$$\lim_{m\to\infty} P^{md}(z,x) = d\pi(x).$$

Now let $x, y \in \mathcal{S}$. If r is such that $P^r(x,y) > 0$, then

$$\lim_{m\to\infty} P^{md+r}(x,y) = \lim_{m\to\infty} \sum_{z\in\mathcal{S}} P^r(x,z)P^{md}(z,y) = \sum_{z\in\mathcal{S}} P^r(x,z)d\pi(y) = d\pi(y)\cdot 1 = d\pi(y)$$

as desired. $\Box$

A picture to explain this theorem in the periodic case:

So $P^n(x,y)$ looks like

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$
$P^n(x,y)$	0	0	0		0	0	0		0	0	0		0	0	0		$\cdots$

3.4 Examples

Directions: For each given Markov chain:

1. Classify the states as transient, positive recurrent or null recurrent;
2. Find all communicating classes of the Markov chain;
3. Find the period of each state;
4. Find all stationary distribution(s) of the Markov chain (if any exist);
5. Find the steady-state distribution of the Markov chain (if it exists).

Example 1: The Ehrenfest chain with $d = 4$

Example 2: The Markov chain whose transition matrix is

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Example 3: Let $\{X_t\}$ be a Markov chain with $\mathcal{S} = \{0, 1, 2, 3, 4, 5, 6\}$ such that $P(0, y) = \frac{1}{6}$ for all $y \neq 0$; $P(x, 0) = \frac{1}{2}$ if $x \neq 0$; $P(x, x+1) = \frac{1}{2}$ if $x \in \{1, 2, 3, 4, 5\}$; and $P(6, 1) = \frac{1}{2}$.

Example 4: Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S} = \{0, 1, 2, 3, ...\}$ whose transition function is

$$P(0, y) = \begin{cases} 0 & \text{if } y \text{ is odd or } y = 0 \\ \left(\frac{1}{2}\right)^{y/2} & \text{if } y \geq 2 \text{ is even} \end{cases}$$

$$P(1, y) = \begin{cases} 0 & \text{if } y = 1 \text{ or } y \text{ is even} \\ \left(\frac{1}{2}\right)^{(y-1)/2} & \text{if } y \geq 3 \text{ is odd} \end{cases}$$

$$x \geq 2 \Rightarrow P(x, y) = \begin{cases} \frac{1}{2} & \text{if } y = 0 \\ \frac{1}{2} & \text{if } y = 1 \\ 0 & \text{else} \end{cases}$$

Chapter 4

Continuous-time Markov chains

4.1 Motivation

Our goal in this chapter is to study analogues of Markov chains (including random walk) where time is measured continuously rather than discretely.

First Question: What "should" a continuous-time Markov chain look like?

	MARKOV CHAIN	CTS-TIME MARKOV CHAIN
state space $\mathcal{S}$	finite or countable; usually $\mathcal{S} = \{0, 1, ..., d\}$ or $\mathcal{S} = \{0, 1, 2, ...\}$ or $\mathcal{S} = \mathbb{Z}$.	finite or countable; usually $\mathcal{S} \subseteq \mathbb{Z}$ (same)
index set $\mathcal{I}$	X_t = state at time t $t \in \{0, 1, 2, ...\}$ or $t \in \mathbb{Z}$	X_t = state at time t $t \in [0, \infty)$ or $t \in \mathbb{R}$
initial distribution	$\pi_0 : \mathcal{S} \to [0, 1]$; $\sum_{x \in \mathcal{S}} \pi_0(x) = 1$ $\pi_0(x) = P(X_0 = x)$	$\pi_0 : \mathcal{S} \to [0, 1]$; $\sum_{x \in \mathcal{S}} \pi_0(x) = 1$ $\pi_0(x) = P(X_0 = x)$ (same)
transition probabilities	we specify time 1 transitions: $P : \mathcal{S} \times \mathcal{S} \to [0, 1]$ $\sum_{y \in \mathcal{S}} P(x, y) = 1 \,\forall\, x \in \mathcal{S}$ $P(x, y) = P(X_{t+1} = y \mid X_t = x)$ (we assume these are $\perp$ of t) If $\mathcal{S}$ is finite, write P as a matrix: $P(x, y) \leftrightarrow P_{x,y} = P_{xy}$ From the time 1 transitions, we calculate transition probabilities for any time n: $P^n(x, y) = P(X_{t+n} = y \mid X_t = x)$ $= \sum_{z \in \mathcal{S}} P(x, z) P^{n-1}(z, y)$ If $\mathcal{S}$ finite, $P^n(x, y) = (P^n)_{xy}$.	
Markov property	$P(X_t = x_t \mid X_0 = x_0, ..., X_{t-1} = x_{t-1})$ $= P(X_t = x_t \mid X_{t-1} = x_{t-1})$ $= P(x_{t-1}, x_t)$	

Definition 4.1 *A* **jump process** $\{X_t : t \in \mathcal{I}\}$ *is a stochastic process with index set* $\mathcal{I} = [0, \infty)$ *or* $\mathbb{R}$ *and finite or countable state space* $\mathcal{S}$ *such that with probability 1, the functions* $t \mapsto X_t$ *(these functions are called* **sample functions** *of the process) are right-continuous and piecewise constant.*

That is, there exist times $J_1 < J_2 < J_3 < ...$ (these are r.v.s, not constants) and states $x_0, x_1, x_2, ... \in \mathcal{S}$ such that

$$X_t = \begin{cases} x_0 & \text{if } 0 \leq t < J_1 \\ x_1 & \text{if } J_1 \leq t < J_2 \\ x_2 & \text{if } J_2 \leq t < J_3 \end{cases}$$

The assumption that the sample functions are right-continuous is necessary for technical reasons (see p. 67 Norris).

Definition 4.2 *A* **continuous-time Markov chain (CTMC)** $\{X_t\}$ *is a jump process satisfying the Markov property .*

4.2 CTMCs with finite state space

In this section, $\mathcal{S}$ is assumed finite; we will write $\mathcal{S} = \{1, 2, ..., d\}$.

Definition 4.3 *Let $\{X_t\}$ be a CTMC with finite state space. For each t, set $P_{xy}(t) = P(X_{s+t} = y \,|\, X_s = x)$ (we assume that $\{X_t\}$ is* **time homogeneous** *so that these probabilities do not depend on s). Then let*

$$P(t) = \begin{pmatrix} P_{11}(t) & \cdots & P_{1d}(t) \\ \vdots & \ddots & \vdots \\ P_{d1}(t) & \cdots & P_{dd}(t) \end{pmatrix};$$

$P(t)$ is called the time t transition function *or* **time t transition matrix** *of the CTMC.*

Theorem 4.4 (Properties of transition matrices) *Let $\{X_t\}$ be a CTMC with index set $\mathcal{I}$ and finite state space $\mathcal{S}$, and let $P(t)$ be the transition matrices of this CTMC. Then:*

1. *Every transition matrix is stochastic (it has nonnegative entries and the rows sum to 1), i.e. for all $t \in \mathcal{I}$,*

$$P_{xy}(t) \in [0,1] \text{ and } \sum_{y \in \mathcal{S}} P_{xy}(t) = 1 \text{ for all } x \in \mathcal{S}.$$

2. *$P(0) = I$, the $d \times d$ identity matrix;*

3. *The* **Chapman-Kolmogorov (C-K) equation** *holds: for all $s, t \in \mathcal{I}$,*

$$P(s)P(t) = P(s+t).$$

PROOF See page 99. □

Question: Which families $P(t)$ of matrices satisfy the four conditions of the preceding theorem?

Related Question: Suppose $f : [0, \infty) \to \mathbb{R}$ satisfies the analogue of (2) and (3) above, i.e. $f(0) = 1$ and $f(s)f(t) = f(s+t)$ for all $s, t \geq 0$. If f is continuous, what must f be?

Answer to related question: Let $t > 0$;

$$f(t) = f\left(\frac{t}{n} + \frac{t}{n} + \frac{t}{n} + ... + \frac{t}{n}\right) = \left[f\left(\frac{t}{n}\right)\right]^n$$

so $f\left(\frac{t}{n}\right) = [f(t)]^{1/n}$.

Therefore if $f(t) = 0$ for any $t > 0$, $f(\frac{t}{n}) = 0$ for all n so $f(0) = \lim_{n \to \infty} f\left(\frac{t}{n}\right) = 0$ as well, contradicting a hypothesis. Thus $f(t) > 0$ for all t.

Now let $C = f(1) > 0$. Then for any $m \in \mathbb{N}$,

$$f(m) = f(1 + 1 + ... + 1) = [f(1)]^m = C^m$$

and for any $\frac{m}{n} \in \mathbb{Q}$,

$$f\left(\frac{m}{n}\right) = [f(m)]^{1/n} = C^{m/n}.$$

By continuity, it must be that $f(t) = C^t = e^{t \ln C} = e^{qt}$ for all $t \geq 0$. We have proven:

Lemma 4.5 *If $f : [0, \infty) \to \mathbb{R}$ is a continuous function satisfying $f(0) = 1$ and $f(s)f(t) = f(s + t)$ for all $s, t \geq 0$, then $f(t) = e^{qt}$ for some constant q.*

Back to matrices: the idea is that

$$(P(s + t) = P(s)P(t)\,\forall s, t \text{ and } P(0) = I) \quad \Rightarrow$$

where Q is some matrix. This makes sense because

Problem: What does $e^{Qt} = \exp(Qt)$ mean? What is $e^Q = \exp(Q)$ for a matrix Q?

Exponentiation of matrices

Recall the Taylor series of e^t:

$$\begin{aligned} e^t &= \sum_{n=0}^{\infty} \frac{t^n}{n!} = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + ... \\ &= \lim_{n\to\infty} \left(1 + \frac{t}{n}\right)^n. \end{aligned}$$

The fact that the limit above equals e^t is a homework problem that uses L'Hôpital's Rule.

Definition 4.6 *Given a square matrix A, define the* **matrix exponential** *of A to be the matrix e^A (also denoted $\exp(A)$) defined by*

$$\begin{aligned} e^A = \exp(A) &= \sum_{n=0}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + ... \\ &= \lim_{n\to\infty} \left(I + \frac{1}{n}A\right)^n. \end{aligned}$$

That the two definitions (the one with the series and the one with the limit) are equal will not be proven here; the proof is similar to the HW problem described above.

Note: If $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $e^A \neq \begin{pmatrix} e^1 & e^2 \\ e^3 & e^4 \end{pmatrix}$.

Observe:

$$\begin{aligned} e^{At} &= I + At + \frac{A^2}{2}t^2 + \frac{A^3}{3!}t^3 + ... \\ &= \lim_{n\to\infty} \left(I + \frac{t}{n}A\right)^n \end{aligned}$$

Theorem 4.7 (Properties of matrix exponentials) *Let A, B and S be square matrices of the same size, where S is invertible. Then:*

1. *If A is diagonal (i.e. $A = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix}$), then*
$$e^A = \begin{pmatrix} e^{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_d} \end{pmatrix}.$$
2. *If $AB = BA$, then $\exp(A + B) = \exp(A)\exp(B)$.*
3. *If $B = \exp(A)$, then $B^n = \exp(An)$ for all $n \in \{0, 1, 2, 3, ...\}$.*
4. *For any matrix A, $(e^A)^n = e^{An}$.*
5. *$\exp($zero matrix$) = I$.*
6. *$\exp(SAS^{-1}) = Se^AS^{-1}$.*

PROOF Math 322 □

Importance: Property (6) above suggests a method to compute the exponential of a matrix A. Diagonalize A (this means write $A = S\Lambda S^{-1}$ where the columns of S are eigenvectors of A and the entries of the diagonal matrix Λ are the corresponding eigenvalues); then $e^A = Se^\Lambda S^{-1}$.

Theorem 4.8 *Let $P(t)$ be a family of square matrices, indexed by t. Then, the following are equivalent:*

1. *$P(t) = e^{Qt}$ for some square matrix Q.*
2. *$\frac{d}{dt}P(t) = P(t)\,Q$ and $P(0) = I$.*
3. *$\frac{d}{dt}P(t) = Q\,P(t)$ and $P(0) = I$;*
4. *$\frac{d^k}{dt^k}P(t)\Big|_{t=0} = Q^k$ for all k;*
5. *$P(0) = I$ and $P(s + t) = P(s)P(t)$ for all $s, t \geq 0$.*

Note: In the theorem above, $\frac{d}{dt}P(t)$ means differentiate each entry of $P(t)$ with respect to t, i.e.

$$\frac{d}{dt}\begin{pmatrix} t^2 & 2 \\ \sin t & t \end{pmatrix} = \begin{pmatrix} 2t & 0 \\ \cos t & 1 \end{pmatrix};$$

PROOF (1) ⇒ (5) follows from properties of matrix exponentials.

(1) ⇒ (2), (3):

$$\frac{d}{dt}P(t) = \frac{d}{dt}e^{Qt} = \frac{d}{dt}\sum_{n=0}^{\infty}\frac{Q^n}{n!}t^n = \sum_{n=1}^{\infty}\frac{Q^n}{(n-1)!}t^{n-1}$$

(1) ⇒ (4):

$$\frac{d^k}{dt^k}P(t)\bigg|_{t=0} = \frac{d^k}{dt^k}\sum_{n=0}^{\infty}\frac{Q^n}{n!}t^n\bigg|_{t=0} = \sum_{n=k}^{\infty}\frac{Q^n}{(n-k)!}t^{n-k}\bigg|_{t=0} = Q^k.$$

(2) ⇒ (1); (3) ⇒ (1) follow from the fact that a system of (ordinary) differential equations with given initial condition has a unique solution (under natural hypotheses that hold here).

(4) ⇒ (1) by Taylor series:

$$P(t) = \sum_{k=0}^{\infty}\frac{1}{k!}\left[\frac{d^k}{dt^k}P(t)\bigg|_{t=0}\right]t^k = \sum_{k=0}^{\infty}\frac{Q^k}{k!}t^k = \exp(Qt).$$

(5) $\Rightarrow$ (1): **As-yet unproven lemma:** (5) implies that $P(t)$ is a continuous and differentiable function of t.

Assuming this lemma, let $h > 0$ be small, and define $Q = P'(0)$. By linear approximation,

$$P(h) \approx P(0) + hP'(0) = I + Qh$$

So for any $t > 0$,

$$P(t) = P\left(\frac{t}{n} + \frac{t}{n} + ... + \frac{t}{n}\right) = \left[P\left(\frac{t}{n}\right)\right]^n \approx \left[I + Q\frac{t}{n}\right]^n$$

Definition 4.9 *Let $\{X_t\}$ be a CTMC with finite state space. Then, by the preceding theorem, the time t transition function $P(t)$ satisfies these differential equations:*

1. *the* **forward equation** $P'(t) = P(t)\,Q$; $P(0) = I$;
2. *the* **backward equation** $P'(t) = Q\,P(t)$; $P(0) = I$.

Corollary 4.10 *If $P(t)$ is the time t transition function for a CTMC with finite state space, then $P(t) = \exp(Qt)$ for some matrix Q (in fact, Q must be equal to $P'(0)$).*

Q-matrices

Next question: What matrices are possible for the Q, if $P(t) = \exp(Qt)$ are the transition matrices of a CTMC?

Definition 4.11 *A square matrix* $Q = \begin{pmatrix} q_{11} & \cdots & q_{1d} \\ \vdots & \ddots & \vdots \\ q_{d1} & \cdots & q_{dd} \end{pmatrix}$ *is called a* **Q-matrix** *if*

1. $q_{ii} \leq 0$ *for all* i*; that is, the diagonal entries are nonpositive;*
2. $q_{ij} \geq 0$ *for all* $i \neq j$*; that is, the off-diagonal entries are nonnegative; and*
3. $\sum_{j=1}^{d} q_{ij} = 0$ *for all* i*; that is, the rows sum to zero.*

Example:

$$Q = \begin{pmatrix} -3 & 2 & 1 \\ 4 & -6 & 2 \\ 0 & 7 & -7 \end{pmatrix}$$

Theorem 4.12 *A square matrix* Q *is a Q-matrix if and only if for every* t, $P(t) = \exp(Qt)$ *is a stochastic matrix.*

PROOF

.

Corollary 4.13 *If $\{X_t\}$ is a CTMC with finite state space $\mathcal{S}$, then the time t transition matrices must satisfy $P(t) = \exp(Qt)$ for some Q-matrix Q. Conversely, every Q-matrix Q defines a CTMC by setting $P(t) = \exp(Qt)$ for all t.*

Definition 4.14 *Let $\{X_t\}$ be a CTMC with finite state space $\mathcal{S}$. Then the matrix $Q = P'(0)$ is called the* **infinitesimal matrix** *of the CTMC.*

Consequence: A CTMC with finite state space is *completely determined* by its infinitesimal matrix Q (and its initial distribution).

Question: Do the entries of Q have any significance?

Definition 4.15 *Let $\{X_t\}$ be a CTMC with finite state space $\mathcal{S}$. Given each state $x \in \mathcal{S}$, define the* **waiting time** *W_x to be the smallest $t \geq 0$ such that $X_t \neq x$, given that $X_0 = x$.*

Note: We assume the sample functions are right-continuous in part to make sure that W_x is well-defined. We don't want, for example

Theorem 4.16 (Waiting times in a CTMC are exponential) *Let $\{X_t\}$ be a CTMC with finite state space $\mathcal{S}$ and Q-matrix Q. Then for each state $x \in \mathcal{S}$, the waiting time W_x is exponential with parameter $q_x = -q_{xx} = -$(the x, x-entry of Q).*

PROOF

$$\begin{aligned} P(W_x > t) &= P(X_s = x \,\forall\, s \in [0,t] \,|\, X_0 = x) \\ &= \lim_{n\to\infty} P\left(X_s = x \,\forall\, s \in \left\{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{tn}{n}\right\} \,|\, X_0 = x\right) \end{aligned}$$

Recall from calculus that for a differentiable function f, if n is large, then $\frac{1}{n}$ is small so $f(\frac{1}{n})$ is approximately equal to $L(\frac{1}{n})$ where L is the tangent line to f at 0, i.e. $L(x) = f(0) + f'(0)x$. Thus $f(\frac{1}{n}) \approx f(0) + f'(0)\frac{1}{n}$. Applying this where $f = P_{xx}$, we see

$$\begin{aligned} P(W_x > t) &= \lim_{n\to\infty} \left[P_{xx}\left(\frac{1}{n}\right)\right]^{tn} \\ &= \lim_{n\to\infty} \left[I_{xx} + q_{xx}\frac{1}{n}\right]^{tn} \\ &= \left[\lim_{n\to\infty} \left(1 + \frac{q_{xx}}{n}\right)^n\right]^t \\ &= e^{q_{xx}t}. \end{aligned}$$

Therefore

$$F_{W_x}(t) = P(W_x \le t) = 1 - e^{q_{xx}t}$$

so W_x is exponential with parameter $q_x = -q_{xx}$ as desired. □

Definition 4.17 *Let $\{X_t\}$ be a CTMC with finite state space $\mathcal{S}$. For each $x \in \mathcal{S}$, define the* **holding rate of** x *to be the nonnegative number q_x satisfying all of the following:*

- $q_x = -P'_{xx}(0)$;
- $q_x = -q_{xx}$ *where q_{xx} is the $(x, x)-$entry of the Q-matrix of the CTMC;*
- $q_x =$ *parameter of the waiting time W_x;*
- $\frac{1}{q_x} = E[W_x] =$ *expected amount of time you stay in state x before leaving/jumping.*

This theory tells you that in a CTMC, your position (state) as time passes is

Definition 4.18 *Let $\{X_t\}$ be a CTMC with finite state space $\mathcal{S}$. For each $x, y \in \mathcal{S}$, define the* **jump probability from** x **to** y *to be*

$$\pi_{x,y} = P(X_{W_x} = y \mid X_0 = x).$$

The **jump matrix** *of the CTMC is the matrix Π whose entries are the jump probabilities, i.e.*

$$\Pi = \begin{pmatrix} \pi_{1,1} & \cdots & \pi_{1,d} \\ \vdots & \ddots & \vdots \\ \pi_{d,1} & \cdots & \pi_{d,d} \end{pmatrix}.$$

Theorem 4.19 (Formula for jump probabilities) *Let $\{X_t\}$ be a CTMC with finite state space $\mathcal{S}$ whose infinitesimal matrix is Q. Then for all $x, y \in \mathcal{S}$,*

$$\pi_{x,y} = \begin{cases} 0 & \text{if } x = y \\ \frac{q_{xy}}{q_x} = \frac{-q_{xy}}{q_{xx}} & \text{if } x \neq y \end{cases}$$

PROOF Later.

Recall the Q-matrix we wrote down a few pages ago:

$$Q = \begin{pmatrix} -3 & 2 & 1 \\ 4 & -6 & 2 \\ 0 & 7 & -7 \end{pmatrix}$$

If this is the infinitesimal matrix of some CTMC $\{X_t\}$ then:

Example: Consider a CTMC with state space $\{1, 2, 3\}$ and infinitesimal matrix

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}.$$

1. Find the jump matrix of this CTMC.

2. Suppose you start in state 1. What is the probability you stay in state 1 for at least three units of time before jumping?

3. What is the probability that the first three jumps are from state 1 to state 3, then state 3 to state 2, then state 2 to state 3 (given that you start in state 1)?

4. Find $P(t)$.

5. Find $P(X_{3/4} = 0 \,|\, X_{1/2} = 1)$.

6. If the initial distribution is $\pi_0 = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, find the distribution at time $t = \ln 2$.

4.3 General theory of CTMCs

At this point we are no longer assuming that the state space $\mathcal{S}$ is finite.

Recall that a CTMC is a jump process that satisfies the Markov property. As before, we can define a time t transition function, i.e. for every $x, y \in \mathcal{S}$ and every $t \in \mathcal{I}$, set

$$P_{x,y}(t) = P(X_{s+t} = y \,|\, X_s = x)$$

and assume that these numbers do not depend on s (i.e that the process is **time homogeneous**).

As with Markov chains, the difference if $\mathcal{S}$ is infinite is that one cannot think of these transition functions as matrices.

However, one can still derive the Chapman-Kolmogorov equation for a general CTMC:

$$P_{x,y}(s+t) = \sum_{z \in \mathcal{S}} P_{x,z}(s) P_{z,y}(t)$$

and from the Markov property, one can deduce that the waiting times W_x must be memoryless, hence exponential. For each $x \in \mathcal{S}$, we can define q_x to be the parameter of the waiting time W_x, and then we can define jump probabilities as before: for every $x \neq y \in \mathcal{S}$,

$$\pi_{x,y} = P(X_{W_x} = y \,|\, X_0 = x).$$

(If $x = y$, we set $\pi_{x,y} = 0$.)

Let δ_{xy} be the **Kronecker delta**, i.e.

$$\delta_{xy} = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{else} \end{cases}$$

Theorem 4.20 (Integral Equation) *Let $\{X_t\}$ be a CTMC. Then for all $t \geq 0$,*

$$P_{x,y}(t) = \delta_{xy} e^{-q_x t} + \int_0^t q_x e^{-q_x s} \left[\sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t-s) \right] ds.$$

PROOF

$$\begin{aligned} P_{x,y}(t) = P_x(X_t = y) &= P_x(X_t = y \cap W_x > t) + P_x(X_t = y \cap W_x \leq t) \\ &= P_x(X_t = y \,|\, W_x > t) P(W_x > t) + P_x(X_t = y \cap W_x \leq t) \\ &= \delta_{x,y} e^{-q_x t} + \int_0^t P(X_t = y \,|\, W_x = s) f_{W_x}(s) \, ds \end{aligned}$$

(Law of Total Probability, continuous version)

Thus

$$\begin{aligned} P_{x,y}(t) &= \delta_{x,y} e^{-q_x t} + \int_0^t f_{W_x}(s) \sum_{z \in \mathcal{S}} P(X_s = z \cap X_t = y \,|\, W_x = s) \, ds \\ &= \delta_{x,y} e^{-q_x t} + \int_0^t q_x e^{-q_x s} \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t-s) \, ds. \ \Box \end{aligned}$$

Theorem 4.21 (Continuity of transition probabilities) *Let $\{X_t\}$ be a CTMC. Then for any $x, y \in \mathcal{S}$, the function $t \mapsto P_{x,y}(t)$ is a continuous function of t.*

PROOF In the integral equation, set $u = t - s$ so that $du = -ds$. Then

$$\begin{aligned} P_{x,y}(t) &= \delta_{xy} e^{-q_x t} + - \int_t^0 q_x e^{-q_x (t-u)} \left[\sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(u) \right] du \\ &= \delta_{xy} e^{-q_x t} + q_x e^{-q_x t} \int_0^t e^{q_x u} \left[\sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(u) \right] du \qquad (\star) \end{aligned}$$

Theorem 4.22 (Differentiability of transition probabilities) *Let $\{X_t\}$ be a CTMC. Then for any $x, y \in \mathcal{S}$, the function $t \mapsto P_{x,y}(t)$ is a differentiable function of t, and*

$$P'_{x,y}(t) = -q_x P_{x,y}(t) + q_x \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t).$$

PROOF By Theorem 4.21, the integrand of the integral in ($\star$) is continuous. Therefore

$$P_{x,y}(t) = \delta_{xy} e^{-q_x t} + q_x e^{-q_x t} \int_0^t e^{q_x u)} \left[\sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(u) \right] du$$

Therefore $P_{x,y}(t)$ is differentiable. (By the way, this proves the "As-yet unproven lemma" from page 89.) Now

$$\begin{aligned} P'_{x,y}(t) &= \frac{d}{dt} \left[\delta_{xy} e^{-q_x t} + q_x e^{-q_x t} \int_0^t e^{q_x u)} \left[\sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(u) \right] du \right] \\ &= -q_x \left[e^{-q_x t} \left(\delta_{xy} + q_x \int_0^t e^{q_x u)} \left[\sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(u) \right] du \right) \right] + e^{-q_x t} q_x e^{q_x t} \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t) \\ &= -q_x P_{x,y}(t) + q_x \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t). \ \Box \end{aligned}$$

Corollary 4.23 *Let $\{X_t\}$ be a CTMC. Then for any $x, y \in \mathcal{S}$,*

$$P'_{x,y}(0) = -q_x \delta_{xy} + q_x \pi_{x,y}.$$

Proof: From Theorem 4.22,

$$\begin{aligned} P'_{x,y}(0) &= -q_x P_{x,y}(0) + q_x \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(0) \\ &= -q_x \delta_{xy} + q_x [0 + 0 + ... + 0 + \pi_{x,y} \cdot 1 + 0 + ... + 0] \\ &= -q_x \delta_{xy} + q_x \pi_{x,y}. \end{aligned}$$

Definition 4.24 *Let $\{X_t\}$ be a CTMC. For any $x, y \in \mathcal{S}$, define the* **infinitesimal parameters** *$q_{xy} = q_{x,y}$ to be $q_{xy} = P'_{x,y}(0)$.*

From Corollary 4.23 we immediately see

Theorem 4.25 (Formula for infinitesimal parameters) *Let $\{X_t\}$ be a CTMC whose infinitesimal parameters are q_{xy}. Then*

$$q_{xy} = \begin{cases} -q_x & \text{if } x = y \\ q_x \pi_{x,y} & \text{if } x \neq y \end{cases}$$

Note: $q_{xx} \leq 0$ for all x, and if $x \neq y$ then $q_{xy} \geq 0$.

Note: If $\mathcal{S}$ is finite, then these are the entries of the Q-matrix (a.k.a. infinitesimal matrix) of the CTMC.

Why are they called infinitesimal parameters? If t is very small (i.e. infinitesimally small), then

$$P_{x,y}(t) \approx P_{x,y}(0) + P'_{x,y}(0)t = \delta_{x,y} + q_{xy}t.$$

The next theorem says that the property of rows of a Q-matrix summing to zero generalizes, even when the state space is infinite:

Theorem 4.26 *Let $\{X_t\}$ be a CTMC and let $x \in \mathcal{S}$. Then*

$$\sum_{y \in \mathcal{S}} q_{xy} = 0.$$

PROOF

$$\sum_{y \in \mathcal{S}} q_{xy} = q_{xx} + \sum_{y \neq x} q_{xy} =$$

Theorem 4.27 (Backward equation) *Let $\{X_t\}$ be a CTMC. Then for all $x, y \in \mathcal{S}$,*

$$P'_{x,y}(t) = \sum_{z \in \mathcal{S}} q_{x,z} P_{z,y}(t) \text{ and } P_{x,y}(0) = \delta_{xy}.$$

Note: If $\mathcal{S}$ is finite, this is equivalent to $P'(t) = Q\,P(t); P(0) = I$.

PROOF By Theorem 4.22,

$$\begin{aligned} P'_{x,y}(t) &= -q_x P_{x,y}(t) + q_x \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t) \\ &= q_{xx} P_{x,y}(t) + q_x \sum_{z \neq x \in \mathcal{S}} \pi_{x,z} P_{z,y}(t) \\ &= \sum_{z \in \mathcal{S}} q_{x,z} P_{z,y}(t). \ \square \end{aligned}$$

Theorem 4.28 (Forward equation) *Let $\{X_t\}$ be a CTMC. Then for all $x, y \in \mathcal{S}$,*

$$P'_{x,y}(t) = \sum_{z \in \mathcal{S}} P_{x,z}(t) q_{zy} \text{ and } P_{x,y}(0) = \delta_{xy}.$$

Note: If $\mathcal{S}$ is finite, this is equivalent to $P'(t) = P(t)\,Q; P(0) = I$.

PROOF Omitted (see Norris text).

4.4 Class structure, recurrence and transience of CTMCs

Definition 4.29 *Let $\{X_t\}$ be a CTMC and let $y \in \mathcal{S}$. Define the* **hitting time to** y *to be*

$$T_y = \min\{t \geq J_1 : X_t = y\}.$$

(Recall that J_1 is the time of the first jump.) (For convenience, set $T_y = 1$ if y is absorbing and $X_0 = y$.)

Definition 4.30 *Let $\{X_t\}$ be a CTMC and let $x, y \in \mathcal{S}$.*

- *Define $f_{x,y} = P_x(T_y < \infty)$. We say $x \to y$ if $f_{x,y} > 0$.*
- *x is called* **recurrent** *if $f_{x,x} = 1$ and* **transient** *otherwise.*
- *x is called* **positive recurrent** *if x is recurrent $m_x = E_x(T_x) < \infty$.*
- *x is called* **null recurrent** *if x is recurrent and $m_x = E_x(T_x) = \infty$.*
- *$\{X_t\}$ is* **irreducible** *if $x \to y$ for all $x, y \in \mathcal{S}$.*

Definition 4.31 *Let $\{X_t\}$ be a CTMC with state space $\mathcal{S}$. The* **embedded chain** *or* **jump chain** *of the CTMC is the (discrete-time) Markov chain whose transition probabilities are $P(x, y) = \pi_{x,y}$.*

Notice that $f_{x,y}$ for the embedded chain is the same as $f_{x,y}$ for the CTMC; so a CTMC is recurrent, transient, etc. if and only if its embedded chain is recurrent, transient, etc., respectively.

Furthermore, irreducible CTMCs are either positive recurrent, null recurrent, or transient (and must be positive recurrent if their state space is finite). All the same theorems regarding class structure for discrete-time Markov chains hold for CTMCs.

Definition 4.32 *Let $\{X_t\}$ be a CTMC with state space $\mathcal{S}$. A distribution π on $\mathcal{S}$ is called* **stationary** *if for all $y \in \mathcal{S}$ and all $t \geq 0$,*

$$\sum_{x \in \mathcal{S}} \pi(x) P_{x,y}(t) = \pi(y).$$

Note: If $\mathcal{S}$ is finite, this means $\pi P(t) = \pi$ in matrix multiplication language.

Theorem 4.33 (Stationarity equation for CTMCs) *Let $\{X_t\}$ be a CTMC with state space $\mathcal{S}$. A distribution π on $\mathcal{S}$ is stationary if and only if*

$$\sum_{x \in \mathcal{S}} \pi(x) q_{xy} = 0 \text{ for all } y \in \mathcal{S}.$$

Note: If $\mathcal{S}$ is finite, this means $\pi Q = \mathbf{0}$ in matrix multiplication language. This gives you a good way to find stationary distributions of CTMCs.

PROOF HW

Theorem 4.34 *Let $\{X_t\}$ be an irreducible CTMC with state space $\mathcal{S}$.*

1. *If $\{X_t\}$ is transient or null recurrent, then it has no stationary distributions.*
2. *If $\{X_t\}$ is positive recurrent, then it has one stationary distribution π given by $\pi(x) = \frac{1}{m_x q_x}$ for all $x \in \mathcal{S}$, and this distribution is steady-state, i.e.*
    - $\lim\limits_{t \to \infty} P_{x,y}(t) = \pi(y)$ *for all $x, y \in \mathcal{S}$; and*
    - $\lim\limits_{t \to \infty} P(X_t = y) = \pi(y)$ *for all $y \in \mathcal{S}$, regardless of the initial distribution.*

PROOF Omitted (see chapter 3, sections 5 and 6 of Norris).

Why does $\pi(x) = \frac{1}{m_x q_x}$? Some motivation:

Why is the stationary distribution always steady-state?

We finish this section with a theorem that says the proportion of time spent in state x in a CTMC converges to the value that the stationary distribution gives x.

Theorem 4.35 (Ergodic theorem for CTMCs) *Let $\{X_t\}$ be an irreducible, positive recurrent CTMC, and let π be the stationary distribution of $\{X_t\}$. Then for all $x \in \mathcal{S}$,*

$$P\left[\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X_s=x\}}\, ds = \pi(x)\right] = 1.$$

A picture to explain:

Chapter 5

Branching and queuing models

5.1 Birth-death Markov chains

Definition 5.1 *A Markov chain with state space $\mathcal{S} = \{0, 1, 2, ...\}$ or $\mathcal{S} = \{0, 1, 2, ..., d\}$ is called a* **birth-death chain** *if for every $x \in \mathcal{S}$, there are three nonnegative numbers p_x, q_x and r_x such that*

1. *For all $x \in \mathcal{S}$, $p_x + q_x + r_x = 1$;*
2. *$q_0 = 0$;*
3. *If $\mathcal{S} = \{0, 1, ..., d\}$, then $p_d = 0$; and*
4. *For all $x \in \mathcal{S}$,* $\begin{cases} P(x, x+1) = p_x \\ P(x, x) = r_x \\ P(x, x-1) = q_x \end{cases}$

Examples: gambler's ruin, Ehrenfest chain.

Every birth-death chain has a directed graph that looks like this:

Observe: A birth-death chain is irreducible if and only if no p_x nor q_x is 0 (other than q_0 or p_d). If a birth-death chain is not irreducible, then the communicating classes of the chain are themselves birth-death chains (after perhaps relabeling the state space).

Analysis of hitting times for birth-death chains

Question: Under what circumstances is an irreducible birth-death chain recurrent? When is such a chain transient?

Partial answer: If $\mathcal{S} = \{0, 1, 2, ..., d\}$, then since $\mathcal{S}$ is finite, the chain is recurrent.

Refined Question: Under what circumstances is an irreducible birth-death chain with $\mathcal{S} = \{0, 1, 2, 3, ...\}$ recurrent? When is such a chain transient?

We will approach this question similar to how we approached the question for random walks (by analyzing hitting times to sets consisting of two points a and b).

Lemma 5.2 *Let $\{X_t\}$ be an irreducible birth-death chain. Let $A = \{a, b\} \subseteq \mathcal{S}$ and suppose $X_0 = x$ where $a < x < b$. Then $P(T_A < \infty) = 1$.*

PROOF This proof is essentially the same as the proof of the similar statement given for random walk. Let $p = \min\{p_a, p_{a+1}, ..., p_b\}$. Since $\{X_t\}$ is irreducible, $p > 0$. Now let G_n be the event that between times $(n-1)(b-a)$ and $n(b-a)$, there are only births in the birth-death chain. Note that $P(G_n) \geq p^{b-a} > 0$, so by repeating the rest of the proof given for random walk, we see that $P(T_A = \infty) \leq P$(no G_n occurs) $= 0$. □

Question: $P_x(T_a < T_b) = ?$

We solved this question for random walks using the OST, by setting up an appropriate martingale related to the random walk (the key idea was that for a random walk, the process $\left\{\left(\frac{q}{p}\right)^{X_t}\right\}$ was a martingale). You can do something similar for birth-death chains, but you need a more complicated definition of the martingale:

Lemma 5.3 *Let $\{X_t\}$ be an irreducible birth-death chain. Then define $\gamma_0 = 1$ and for each $y > 0$, set*

$$\gamma_y = \frac{q_y q_{y-1} q_{y-2} \cdots q_2 q_1}{p_y p_{y-1} p_{y-1} \cdots p_2 p_1}.$$

Define the function $\widetilde{\gamma} : \mathcal{S} \to \mathbb{R}$ by setting $\widetilde{\gamma}(0) = 1$, $\widetilde{\gamma}(1) = 1$ and for $y \geq 2$, setting

$$\begin{aligned}\widetilde{\gamma}(y) &= 1 + \frac{q_1}{p_1} + \frac{q_2 q_1}{p_2 p_1} + ... + \frac{q_{y-1} q_{y-2} \cdots q_2 q_1}{p_{y-1} p_{y-2} \cdots p_2 p_1} \\ &= \gamma_0 + \gamma_1 + \gamma_2 + ... + \gamma_{y-1} \\ &= \sum_{j=0}^{y-1} \gamma_j.\end{aligned}$$

Then the stochastic process $\{Y_t\}$ is a martingale, where $Y_t = \widetilde{\gamma}(X_t)$.

PROOF HW

Theorem 5.4 (Escape probabilities for birth-death chains) *Let $\{X_t\}$ be an irreducible birth-death chain with infinite state space. Then if $a < x < b$,*

$$P_x(T_a < T_b) = \frac{\sum_{y=x}^{b-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y} \quad \text{and} \quad P_x(T_b < T_a) = \frac{\sum_{y=a}^{x-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}$$

where $\gamma_0 = 1$ and for all $y > 0$,

$$\gamma_y = \frac{q_y q_{y-1} q_{y-2} \cdots q_2 q_1}{p_y p_{y-1} p_{y-1} \cdots p_2 p_1}.$$

PROOF Let $\{Y_t\} = \{\widetilde{\gamma}(X_t)\}$ be as in the preceding lemma; we see

$$Y_0 = \widetilde{\gamma}(X_0) = \widetilde{\gamma}(x).$$

Let $T = \min(T_a, T_b) = T_{\{a,b\}}$; T is a finite stopping time and $\{X_t\}$ is bounded (by a and b) until T occurs, so the tweaked version of the OST applies to give

$$\begin{aligned}\widetilde{\gamma}(x) = E[Y_0] = E[Y_T] &= \widetilde{\gamma}(a) P(X_T = a) + \widetilde{\gamma}(b) P(X_t = b) \\ &= \widetilde{\gamma}(a) P_x(T_a < T_b) + \widetilde{\gamma}(b)[1 - P_x(T_a < T_b)] \\ &= \widetilde{\gamma}(b) + [\widetilde{\gamma}(a) - \widetilde{\gamma}(b)] P_x(T_a < T_b).\end{aligned}$$

Solve for $P_x(T_a < T_b)$ to get

$$P_x(T_a < T_b) = \frac{\widetilde{\gamma}(b) - \widetilde{\gamma}(x)}{\widetilde{\gamma}(b) - \widetilde{\gamma}(a)} = \frac{\sum_{y=0}^{b-1} \gamma_y - \sum_{y=0}^{x-1} \gamma_y}{\sum_{y=0}^{b-1} \gamma_y - \sum_{y=0}^{a-1} \gamma_y} = \frac{\sum_{y=x}^{b-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}.$$

The other result follows from the complement rule. □

Using this theorem, we can determine under which circumstances an irreducible birth-death chain on an infinite state space is recurrent:

Lemma 5.5 *Let $\{X_t\}$ be an irreducible birth-death chain with infinite state space. Then $\{X_t\}$ is recurrent if and only if $f_{1,0} = 1$.*

PROOF $\{X_t\}$ is irreducible, so $\{X_t\}$ is recurrent $\iff$ 0 is recurrent $\iff$ $f_{0,0} = 1$. Now

$$\begin{aligned} f_{0,0} &= P_0(T_0 < \infty) \\ &= P_0(T_0 = 1) + P_0(T_0 \in [2, \infty)) \\ &= \end{aligned}$$

Theorem 5.6 (Recurrence/transience of birth-death chains) *Let $\{X_t\}$ be an irreducible birth-death chain with infinite state space. Then defining γ_y as in the previous theorem,*

$$\{X_t\} \text{ is recurrent} \iff \sum_{y=0}^{\infty} \gamma_y = \infty.$$

PROOF Suppose $X_0 = 1$. Since $\{X_t\}$ is a birth-death chain,

$$1 \leq T_2 < T_3 < T_4 < ... < T_n < ...$$

so

$$(T_0 < T_2) \subseteq (T_0 < T_3) \subseteq (T_0 < T_4) \subseteq ...$$

and consequently

$$\begin{aligned}
f_{1,0} &= P_1(T_0 < \infty) \\
&= P_1\left(\bigcup_{n=2}^{\infty} (T_0 < T_n)\right) \\
&= \lim_{n\to\infty} P_1(T_0 < T_n) \quad \text{by monotonocity (chapter 1 of Math 414)} \\
&= \lim_{n\to\infty} \left(\frac{\sum_{y=1}^{n-1} \gamma_y}{\sum_{y=0}^{n-1} \gamma_y}\right) \quad \text{by Theorem 5.4 with } x = 1, a = 0, b = n \\
&= \lim_{n\to\infty} \left(\frac{\sum_{y=0}^{n-1} \gamma_y - \gamma_0}{\sum_{y=0}^{n-1} \gamma_y}\right) \\
&= \lim_{n\to\infty} \left(\frac{\sum_{y=0}^{n-1} \gamma_y - 1}{\sum_{y=0}^{n-1} \gamma_y}\right) \\
&= \lim_{n\to\infty} \left(1 - \frac{1}{\sum_{y=0}^{n-1} \gamma_y}\right) \\
&= \left\{ \vphantom{\begin{matrix} a \\ b \\ c \\ d \end{matrix}} \right.
\end{aligned}$$

By the preceding lemma, $\{X_t\}$ is recurrent if and only if $f_{1,0} = 1$, so this proves the theorem. □

Example: Let $\{X_t\}$ be a birth-death chain on $\mathcal{S} = \{0, 1, 2, 3, ...\}$ such that

$$p_x = \frac{x+2}{2(x+1)} \quad \text{and} \quad q_x = \frac{x}{2(x+1)}.$$

Is this chain recurrent or transient?

Stationary distributions of irreducible birth-death chains

Let the state space be $\mathcal{S} = \{0, 1, 2, 3, ..., d\}$ or $\mathcal{S} = \{0, 1, 2, 3, ...\}$ (in the second situation, $d = \infty$ in what follows).

$$\pi \text{ stationary} \Rightarrow \sum_{x=0}^{d} \pi(x)P(x, y) = \pi(y) \text{ and } \sum_{y \in \mathcal{S}} \pi(y) = 1$$
$$\Rightarrow \begin{cases} \pi(0)r_0 + \pi(1)q_1 = \pi(0) & (y = 0) \\ \pi(y-1)p_{y-1} + \pi(y)r_y + \pi(y+1)q_{y+1} = \pi(y) & (y > 0) \\ \sum\limits_{y=0}^{d} \pi(y) = 1 \end{cases}$$

Since $p_y + q_y = 1 - r_y$ for all y, these equations yield (after some significant algebra)

$$\pi(y+1) = \frac{p_y}{q_{y+1}} \pi(y) \,\forall\, y \geq 0$$
$$\Rightarrow \pi(y) = \frac{p_0 p_1 p_2 \cdots p_{y-1}}{q_1 q_2 \cdots q_y} \pi(0) \,\forall\, y \geq 1$$

Define

$$\zeta_y = \begin{cases} \frac{p_0 p_1 \cdots p_{y-1}}{q_1 q_2 \cdots q_y} & \text{if } y > 0 \\ 1 & \text{if } y = 0 \end{cases}$$

Then $\pi(y) = \zeta_y \pi(0)$ for all $y \in \mathcal{S}$.
This means

$$1 = \sum_{y \in \mathcal{S}} \pi(y) = \sum_{y \in \mathcal{S}} \zeta_y \pi(0);$$

this can only be true if

$$\sum_{y \in \mathcal{S}} \zeta_y \text{ converges (this is always true if } d < \infty).$$

in which case

$$\pi(0) \cdot \sum_{y \in \mathcal{S}} \zeta_y = 1 \Rightarrow \pi(0) = \left[\sum_{y \in \mathcal{S}} \zeta_y \right]^{-1}.$$

We have essentially proven:

Theorem 5.7 (Stationary distribution for irred. birth-death chains) *Let $\{X_t\}$ be an irreducible birth-death chain. Define $\zeta_0 = 1$ and for each $y > 0$ in $\mathcal{S}$, define $\zeta_y = \frac{p_0 p_1 \cdots p_{y-1}}{q_1 q_2 \cdots q_y}$. Then:*

1. *If $\sum_{y \in \mathcal{S}} \zeta_y$ converges, then $\{X_t\}$ is positive recurrent and has one stationary distribution π defined by*
$$\pi(x) = \frac{\zeta_x}{\sum_{y \in \mathcal{S}} \zeta_y}.$$
(This includes all situations where $\mathcal{S}$ is finite.)

2. *If $\sum_{y \in \mathcal{S}} \zeta_y$ diverges, then $\{X_t\}$ has no stationary distributions (so it is either null recurrent or transient).*

Example: Let $\{X_t\}$ be a birth-death chain on $\{0, 1, 2, 3, ...\}$ with $p_0 = 1$; $p_x = \frac{1}{x+1}$ for all $x \geq 1$; $q_x = \frac{x}{x+1}$ for all $x \geq 1$. Find the stationary distribution of $\{X_t\}$, if one exists.

5.2 Birth-death CTMCs

A birth-death CTMC is a CTMC where all the jumps are of size ± 1. More formally:

Definition 5.8 *A* **birth-death CTMC** *is a CTMC* $\{X_t\}$ *whose state space is either* $\mathcal{S} = \{0, 1, ..., d\}$ *or* $\mathcal{S} = \{0, 1, 2, ...\}$ *or* $\mathcal{S} = \mathbb{Z}$*, such that* $q_{x,y} = 0$ *whenever* $|x - y| > 1$*. The numbers* $\lambda_x = q_{x,x+1}$ *are called the* **birth rates** *of the process and the numbers* $\mu_x = q_{x,x-1}$ *are called the* **death rates***. A birth-death CTMC is called a* **pure birth process** *if* $\mu_x = 0$ *for all* x*, and is called a* **pure death process** *if* $\lambda_x = 0$ *for all* x*.*

In a birth-death CTMC, we have

Observe: An irreducible birth-death CTMC on $\mathcal{S} = \{0, 1, ..., d\}$ or $\mathcal{S} = \{0, 1, 2, ...\}$ is transient if and only if its embedded jump chain is transient.

This jump chain is a (discrete-time) birth-death chain with transition function $\pi_{x,y}$, i.e.

$$“p_x” = \frac{\lambda_x}{q_x} \quad \text{and} \quad “q_x” = \frac{\mu_x}{q_x}:$$

Recall from Packet 416-2 that the jump chain (and hence the birth-death CTMC) is transient if and only if

$$\sum_{x=1}^{\infty} \gamma_x < \infty$$

We have proven:

Theorem 5.9 *An irreducible birth-death CTMC on* $\mathcal{S} = \{0, ..., d\}$ *or* $\mathcal{S} = \{0, 1, ..., \}$ *is transient if and only if*

$$\sum_{x \in \mathcal{S}} \frac{\mu_1 \cdots \mu_x}{\lambda_1 \cdots \lambda_x} < \infty.$$

Similarly, one can show:

Theorem 5.10 *An irreducible birth-death CTMC on* $\mathcal{S} = \{0, ..., d\}$ *or* $\mathcal{S} = \{0, 1, ..., \}$ *is positive recurrent if and only if*

$$\sum_{x \in \mathcal{S}} \frac{\lambda_0 \cdots \lambda_{x-1}}{\mu_1 \cdots \mu_x} < \infty,$$

in which case the stationary distribution π *satisfies*

$$\pi(x) = \frac{\phi_x}{\sum_{x \in \mathcal{S}} \phi_x}$$

where $\phi_0 = 1$ *and for all* $y > 0$*,*

$$\phi_y = \frac{\lambda_0 \cdots \lambda_{y-1}}{\mu_1 \cdots \mu_y}$$

Example: (Pure birth process) Consider a birth-death CTMC on $\mathcal{S} = \{0, 1, 2, 3, ...\}$ with $\mu_x = 0$ for all x.

.

Example: (Two-state birth-death CTMC) Consider a birth-death CTMC on $\mathcal{S} = \{0, 1\} = \{\text{OFF, ON}\}$.

Example: A **Poisson process** is a pure birth process on $\mathcal{S} = \{0, 1, 2, ...\}$ with $\lambda_x = \lambda$ for all x.

5.3 Galton-Watson chain (branching in discrete time)

Setup: Consider lifeforms that reproduce asexually. Each organism has N offspring, where $N : \Omega \to \{0, 1, 2, 3, ...\}$ is a r.v. with density $f = f_N$ (so that $f_N(n) = P(N = n)$, the probability that any parent has exactly n offspring). Let X_t be the number of organisms in the t^{th} generation for $t = 0, 1, 2, 3, ...$ $\{X_t\}$ is a Markov chain called a **branching chain** or a **Galton-Watson chain**.

Note: If $f(1) = 1$, then $P(N = 1) = 1$ so the number of organisms is constant from generation to generation. In this setting, every state is absorbing, and the branching chain is called "trivial".

In a nontrivial branching chain, the transition function is

$$P(x, y) =$$

Associated picture:

Theorem 5.11 *Let $\{X_t\}$ be a nontrivial branching chain. Then 0 is absorbing (hence recurrent), but all other states are transient.*

PROOF Earlier (group presentations).

In a branching chain, we are most interested in "extinction probabilities", i.e. the probability that you eventually hit 0:

Set $\eta = f_{1,0} = P_1(T_0 < \infty)$. Then $f_{x,0} = \eta^x$ for all $x \geq 1$.

Theorem 5.12 *Let $\{X_t\}$ be a nontrivial branching chain and let $\eta = f_{1,0}$ (η is called the* **extinction probability** *of the chain). Then $\eta = f_{1,0}$ is the solution of the equation $t = G_N(t)$, where G_N is the pgf of the number of offspring N. (Of course this solution has to be in $[0,1]$ for this to make sense.)*

PROOF Earlier (group presentations).

Corollary 5.13 *Let $\{X_t\}$ be a nontrivial branching chain and let $\eta = f_{1,0}$. Then, if N is the number of offspring,*

1. $EN \leq 1 \iff \eta = 1$.
2. $EN > 1 \iff \eta < 1$.

PROOF First, if $f(0) = 0$, then $X_{t+1} \geq X_t$ for all t, so $\eta = 0$; in this case $EN \geq 1$ since the branching chain is nontrivial.

Henceforth assume that $f(0) > 0$. In this setting, from facts about pgfs in Math 414, the equation $G_N(t) = t$ has a solution in $(0,1)$ if and only if $EN > 1$. The result then follows from Theorem 5.12. □

5.4 Continuous-time branching processes

Setup: Suppose that you start at time $t = 0$ with a population of X_0 particles (X_0 is a random variable taking values in $\{0, 1, 2, ...\}$). Each particle does nothing for time A ($A : \Omega \to [0, \infty)$ is a cts r.v.) and the either splits into two particles (with probability p) or dies (with probability $1 - p$). For $t \in [0, \infty)$, let X_t be the number of particles at time t. $\{X_t\}$ is called a **branching process**.

"population picture" process $\{X_t\}$

Theorem 5.14 (Minimum of $\perp$ exponential r.v.s is exponential) *Let $A_1, ..., A_d$ be $\perp$ exponential r.v.s with respective parameters $\lambda_1, ..., \lambda_d$. Then $\min(A_1, ..., A_d)$ is exponential with parameter $\sum_{j=1}^{d} \lambda_j$.*

PROOF HW (as a hint, start by computing the distribution function of $\min(A_1, ..., A_d)$). □

Corollary 5.15 *Let $\{X_t\}$ be a branching process with the waiting time A exponential. Then $\{X_t\}$ is a CTMC (in fact, it is a birth-death process).*

(Henceforth, all branching processes are assumed to have A exponential, and λ is the parameter of the exponential waiting time.)

Recall that a birth-death process is determined by birth and death rates. In a branching process, we have

Observations: In a branching process,

1. 0 is absorbing;
2. Every nonzero state in $\mathcal{S}$ is transient.

(Proofs are similar to the discrete-time case.)

Theorem 5.16 *Let $\{X_t\}$ be a branching process. Then the extinction probability $\eta = f_{1,0}$ satisfies*

Note: As with a branching chain, $f_{x,0} = \eta^x$ for all $x \in \{0, 1, 2, ...\}$.

PROOF Notice that $f_{1,0}$ in the branching process is the same as $f_{1,0}$ in the associated jump chain. Now use the formulas from earlier:

5.5 Discrete-time queuing chains

Setup: Consider a line at a supermarket checkout counter where one person is checked out per unit of time. Take a r.v. $Z : \Omega \to \{0, 1, 2, 3...\}$ with density f_Z, and assume in the j^{th} unit of time, Z_j people get in line, where $Z_1, Z_2, ...$ are i.i.d. r.v.s, each having the density of Z. Let X_0 be the number of customers initially in line, and let X_t be the number of customers in line after the t^{th} unit of time (where $t \in \mathbb{N}$).

This is a Markov chain called a **queuing chain** with $\mathcal{S} = \{0, 1, 2, ...\}$ and transition function

Theorem 5.17 *A queuing chain is irreducible if and only if ($f_Z(0) > 0$ and $f_Z(0) + f_Z(1) < 1$).*

PROOF Earlier (group presentations).

Theorem 5.18 (Recurrence/transience of queuing chains) *Let $\{X_t\}$ be a queuing chain with Z customers arriving in each unit of time. Then:*

1. *If $EZ > 1$, then $\{X_t\}$ is transient.*
2. *If $EZ = 1$ and $\{X_t\}$ is irreducible, then $\{X_t\}$ is null recurrent.*
3. *If $EZ < 1$ and $\{X_t\}$ is irreducible, then $\{X_t\}$ is positive recurrent, and the mean return time to state 0 is $m_0 = (1 - EZ)^{-1}$.*

PROOF The transience/recurrence was earlier (group presentations). The proof of the positive recurrence/null recurrence issues are omitted.

5.6 The infinite server queue

Setup: Let X_t denote the number of people in line for some service (including those being served). Assume that the people arrive at rate λ (i.e. that the number of arrivals in line follows a Poisson process with rate λ) and that the time it takes each customer to be served is exponential with parameter μ. Assume that there are an infinite number of servers (so no one has to wait in line before being served). The resulting CTMC $\{X_t\}$ is called the **infinite server queue**.

The infinite server queue is also called the $M/M/\infty$ queue.

Observe: $\{X_t\}$ is a birth-death process with birth and death rates

$$\begin{cases} \lambda_x = \\ \\ \mu_x = \end{cases}$$

Question: What is the time t transition function for the infinite server queue?

Answer: Let $C_t = \#$ of customers arriving in $[0, t]$. Suppose for now that $C_t = c$. The first thing we want to know is how the arrival times of these c customers are distributed. To determine this, choose a partition $0 = t_0 < t_1 < ... < t_m = t$ of $[0, t]$.

Then let $V_j = \#$ of customers arriving in $(t_{j-1}, t_j]$.

Now

$$P(V_j = x_j \,\forall j \,|\, C_t = x_1 + ... + x_m) =$$

so the times when customers arrive (given a fixed total number of arriving customers in an interval of length t) are i.i.d. uniform on $[0, t]$.

Notice that if a customer arrives at time $s \in (0, t]$, the probability he is still being served at time t is

So if a customer arrives at a uniformly chosen time in $(0, t]$, we have

$p_t = P(\text{customer is still being served at time } t) =$

Let $X_t^{new} = \#$ of customers arriving in $(0, t]$ still being served at time t.

$$P(X_t^{new} = n \,|\, C_t = k) =$$

Therefore

$$\begin{aligned}
P(X_t^{new} = n) &= \sum_{k=0}^{\infty} P(X_t^{new} = n \text{ and } C_t = k) \\
&= \sum_{k=n}^{\infty} P(X_t^{new} = n \text{ and } C_t = k) \quad (\text{since } X_t^{new} \leq C_t) \\
&= \sum_{k=n}^{\infty} P(X_t^{new} = n \,|\, C_t = k) P(C_t = k) \\
&= \sum_{k=n}^{\infty} \left[\binom{k}{n} p_t^n (1-p_t)^{k-n} \right] \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\
&= \frac{p_t^n e^{-\lambda t}}{n!} \sum_{k=n}^{\infty} \frac{(1-p_t)^{k-n} (\lambda t)^k}{(k-n)!} \\
&= \frac{(\lambda t p_t)^n e^{-\lambda t}}{n!} \sum_{k=n}^{\infty} \frac{[\lambda t(1-p_t)]^{k-n}}{[k-n]!}
\end{aligned}$$

Now change indices in the series by setting $s = k - n$:

$$\begin{aligned}
&= \frac{(\lambda t p_t)^n e^{-\lambda t}}{n!} \sum_{s=0}^{\infty} \frac{[\lambda t(1-p_t)]^s}{s!} \\
&= \frac{(\lambda t p_t)^n e^{-\lambda t}}{n!} e^{\lambda t(1-p_t)} \\
&= \frac{(\lambda t p_t)^n e^{-\lambda t p_t}}{n!}
\end{aligned}$$

This proves that X_t^{new} is Poisson with parameter $\lambda t p_t$.

Now let $X_t^{orig} = \#$ of customers present initially that are still being served at time t.

X_t^{orig} is with parameters $\left\{ \right.$

Since $X_t = X_t^{new} + X_t^{orig}$, we have

$$\begin{aligned}
P_{x,y}(t) &= P_x(X_t = y) \\
&= \sum_{k=0}^{\min(x,y)} P_x(X_t^{orig} = k) P_x(X_t^{new} = y-k) \\
&= \sum_{k=0}^{\min(x,y)} \left(\left[\binom{x}{k} e^{-\mu k t} (1-e^{-\mu t})^{x-k} \right] \left[\frac{\left[\frac{\lambda}{\mu}(1-e^{-\mu t})\right]^{y-k}}{(y-k)!} \exp\left(\frac{-\lambda}{\mu}(1-e^{-\mu t}) \right) \right] \right).
\end{aligned}$$

This is a miserable formula, but it can be used to find the steady-state distribution:

$$\begin{aligned}
\lim_{t\to\infty} P_{x,y}(t) &= \lim_{t\to\infty}(k=0 \text{ term of the above sum}) \\
&= \lim_{t\to\infty} \begin{pmatrix} x \\ 0 \end{pmatrix} e^{-0}\left(1-e^{-\mu t}\right)^x \left[\frac{\left[\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\right]^y}{y!}\exp\left(\frac{-\lambda}{\mu}\left(1-e^{-\mu t}\right)\right)\right] \\
&= (1)1(1)^x \left[\frac{\left[\frac{\lambda}{\mu}(1)\right]^y}{y!}\exp\left(\frac{-\lambda}{\mu}(1)\right)\right] \\
&= \frac{\left(\frac{\lambda}{\mu}\right)^y}{y!}e^{-(\lambda/\mu)}.
\end{aligned}$$

We have proven:

Theorem 5.19 (Steady-state distribution of the infinite server queue) *The steady-state distribution of the infinite server queue where the customers arrive exponentially with parameter λ and are served exponentially with parameter μ is Poisson with parameter $\frac{\lambda}{\mu}$.*

Note: The existence of a steady-state distribution means that the infinite server queue is positive recurrent (this could also be derived using the facts from earlier in this chapter about birth-death CTMCs).

Chapter 6

Brownian motion

6.1 Definition and construction

Goal: Develop a model for "continuous random movement", i.e. a continuous version of simple, unbiased random walk. This stochastic process will be called $\{W_t\}$.

First Question: What properties should such a process have?

PROPERTY	RANDOM WALK	$\{W_t\}$
index set $\mathcal{I}$ (times)	$\mathbb{Z} \cap [0, \infty)$	
state space $\mathcal{S}$ (positions)	$\mathbb{Z}$	
initial distribution	$X_0 = 0$	
independent increments	$\forall\, 0 \le t_1 \le t_2 \le ... \le t_n \in \mathbb{Z}$, $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, ..., X_{t_n} - X_{t_{n-1}}$ mutually $\perp$	
stationarity property (time homogeneity)	The distribution of $X_t - X_s$ (for $0 \le s \le t$) depends only on $t - s$ (and not on X_s, s or t) and is binomial $b(t - s, \frac{1}{2})$.	
continuity	trivial (or none)	

Definition 6.1 *A stochastic process* $\{W_t : t \in [0, \infty)\}$ *taking values in* $\mathbb{R}$ *(or* $\mathbb{R}^d$*) is called a* **Brownian motion (BM)** *or a* **Weiner process** *with parameter* σ^2 *if*

1. $W_0 = 0$;
2. *For all* $0 \le t_1 \le t_2 \le ... \le t_n \in \mathbb{R}$*, the random variables* $W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, ..., W_{t_n} - W_{t_{n-1}}$ *are mutually* $\perp$*;*
3. *For any* $0 \le t_1 \le t_2$ *in* $\mathbb{R}$*,* $W_{t_2} - W_{t_1}$ *is* $n(0, \sigma^2(t_2 - t_1))$*; and*
4. *with probability 1, the functions* $t \mapsto W_t$ *are continuous in* t*.*

If $\sigma^2 = 1$*, then* W_t *is called a* **standard Brownian motion**. *A* **Brownian motion starting at** x *is a process satisfying 2,3 and 4 above but having* $X_0 = x$*.*

Theorem 6.2 (Weiner's Theorem) *There is a process which is a Brownian motion.*

PROOF (really just a sketch of the proof)

For each $n \in \mathbb{N}$, let $\mathbb{D}_n$ be the **dyadic rationals of order** n, i.e.

$$\mathbb{D}_n = \left\{ \frac{m}{2^n} : m \in \mathbb{N} \right\} = \left\{ 0, \frac{1}{2^n}, \frac{2}{2^n}, \frac{3}{2^n}, ... \right\}.$$

For $n \ge 1$, let $\mathbb{D}_n^{new} = \mathbb{D}_n - \mathbb{D}_{n-1}$. These are the numbers which are expressible as an integer over 2^n, but not expressible as an integer over 2^{n-1}; equivalently these are numbers which are an odd integer divided by 2^n.

Quick observations about the dyadic rationals of order n*:*

- $\mathbb{D}_0 = \mathbb{N} = \{0, 1, 2, 3, ...\}$
- $\mathbb{D}_0 \subseteq \mathbb{D}_1 \subseteq \mathbb{D}_2 \subseteq \mathbb{D}_3 \subseteq ...$
- $\bigcup_{n=0}^{\infty} \mathbb{D}_n$ is countable and dense in $[0, \infty)$

Notation: For all $t \in \mathbb{D}_n^{new}$, set $t^+ = \min\{s \in \mathbb{D}_{n-1} : s > t\}$ and set $t^- = \max\{s \in \mathbb{D}_{n-1} : s < t\}$.

Step 0: For each $t \in \mathbb{D}_0 = \mathbb{Z}$, let Y_t be a $n(0,1)$ r.v. independent of the other Y_ts. Let $\{B_0(t)\}_{t\in\mathbb{N}}$ be a discrete-time stochastic process defined by setting

$$B_0(t) = \sum_{j=1}^{t} Y_j.$$

Then let $\{\widehat{B_0}(t)\}_{t\in[0,\infty)}$ be the continuous-time stochastic process obtained by interpolating linearly between the points of $\{B_0(t)\}$:

Step 1: For each $t \in \mathbb{D}_1^{new}$, let Y_t be a $n(0, \frac{1}{2})$ r.v. independent of the Y_ts defined either here or earlier. Let $\{B_1(t)\}_{t\in\mathbb{D}_1}$ be a discrete-time stochastic process defined by setting

$$B_1(t) = \begin{cases} B_0(t) & \text{if } t \in \mathbb{D}_0 \\ \frac{1}{2}\left(B_0(t^-) + B_0(t^+)\right) + Y_t & \text{if } t \in \mathbb{D}_1^{new} \end{cases}.$$

Then let $\{\widehat{B_1}(t)\}_{t\in[0,\infty)}$ be the continuous-time stochastic process obtained by interpolating linearly between the points of $\{B_1(t)\}$:

Step $N+1$**:** Suppose the processes $\{B_N(t)\}$ and $\{\widehat{B_N}(t)\}$ have been constructed. Here is how we define $\{B_{N+1}(t)\}$: for each $t \in \mathbb{D}_{N+1}^{new}$, let Y_t be a $n(0, \frac{1}{2^{N+1}})$ r.v. independent of the Y_ts defined either here or earlier. Let $\{B_{N+1}(t)\}_{t \in \mathbb{D}_{N+1}}$ be a discrete-time stochastic process defined by setting

$$B_{N+1}(t) = \begin{cases} B_N(t) & \text{if } t \in \mathbb{D}_N \\ \frac{1}{2}\left(B_N(t^-) + B_N(t^+)\right) + Y_t & \text{if } t \in \mathbb{D}_{N+1}^{new} \end{cases}.$$

Then let $\{\widehat{B_{N+1}}(t)\}_{t \in [0,\infty)}$ be the continuous-time stochastic process obtained by interpolating linearly between the points of $\{B_{N+1}(t)\}$:

Now define $W_t = \lim\limits_{n \to \infty} \widehat{B_N}(t)$. One can show that $\{W_t\}$ satisfies all the properties necessary to be a Brownian motion (details are on pages 161-163 of Norris). $\Box$

Brownian motion arises commonly in real-world situations:

1. Movements of particles suspended in a liquid
2. Fluctuations in the stock market
3. Quantum mechanics (path-integral formulation)
4. Option pricing models (Black-Scholes) (quackery according to some)
5. Cosmology models

Why is BM so prevalent? Because it arises as a "limit of rescaled random walks":

Brownian motions approximate random walks with small but frequent jumps (so long as the size of the jump is proportional to the square root of the time between jumps).

What do we know about Brownian motion so far?

Example: Suppose $\{W_t\}$ is a BM with parameter $\sigma^2 = 9$.

1. Describe the random variable W_3.
2. Describe the random variable $W_8 - W_2$.
3. Find the probability that $W_8 > 1$.
4. Find the probability that $W_7 - W_5 \leq 2$.
5. Find the probability that $W_8 - W_7 < 1$ and $W_{14} - W_{12} > -3$.

6.2 Markov properties of Brownian motion

Let $\{W_t\}$ be a BM. The discrete version of the Markov property would say something like this:

$$P(W_t = y \,|\, W_{t_1} = x_1, W_{t_2} = x_2, ..., W_{t_n} = x_n) = P(W_t = y \,|\, W_{t_n} = x_n)$$

$$\forall\, 0 \leq t_1 \leq t_2 \leq t_3 \leq ... \leq t_n \leq t, \forall\, x_1, ..., x_n, y \in \mathbb{R}$$

A better formulation of the same idea in this setting is this:

This holds because of the independent increment property in the definition of BM.

Definition 6.3 *Let $\{W_t\}$ be a BM. Given $x, y \in \mathbb{R}$ and $t \geq 0$, the* **time t transition density** *for the BM is*

$$p_{x,y}(t) = f_{W_t|W_0}(y|x) \quad (= f_{W_{s+t}|W_s}(y|x)\, \forall\, s \text{ by time homogeneity}).$$

Theorem 6.4 (Markov property for Brownian motion) *Let $\{W_t\}$ be a BM with parameter σ^2. Then the time t transition densities are $n(x, \sigma^2 t)$.*

In other words, if $W_s = x$, then W_{s+t} is a continuous r.v. with density function

$$f(y) = \frac{1}{\sigma\sqrt{2\pi t}} \exp\left[\frac{-(y-x)^2}{2\sigma^2 t}\right].$$

Theorem 6.5 (Strong Markov property) *Let $\{W_t\}$ be a BM and let T be a stopping time for $\{W_t\}$. Define $Y_t = W_{T+t} - W_T$. Then Y_t is a BM, independent of $\{W_t : t \leq T\}$.*

PROOF Hard analysis.

Note: This property also holds for Markov chains and CTMCs; for a proof, read p. 19-20 and p. 223-228 of Norris.

Theorem 6.6 (Reflection Principle) *Let* $\{W_t\}$ *be a BM with parameter* σ^2*. Fix* $b > 0$ *and let* $T_b = \min\{t \geq 0 : W_t = b\}$*. Then*

$$F_{T_b}(t) = P(T_b \leq t) = 2 - 2\Phi\left(\frac{b}{\sigma\sqrt{t}}\right).$$

PROOF

$$P(W_t \geq b) = P(W_t \geq b \,|\, T_b \leq t)P(T_b \leq t)$$

$$\Rightarrow F_{T_b}(t) = P(T_b \leq t) = \frac{P(W_t \geq b)}{P(W_t \geq b \,|\, T_b \leq t)}$$

Corollary 6.7 *Let $\{W_t\}$ be a BM with parameter σ^2. Fix $b > 0$ and let $T_b = \min\{t \geq 0 : W_t = b\}$. Then T_b has density*

$$f_{T_b}(t) = \frac{b}{\sigma\sqrt{2\pi t^3}} \exp\left[\frac{-b^2}{2t\sigma^2}\right]$$

(where Φ is the cdf of the standard normal).

PROOF Differentiate F_{T_b} with respect to t. □

Corollary 6.8 *Let $\{W_t\}$ be a BM with parameter σ^2. Then $\{W_t\}$ is irreducible, i.e. for any $b \in \mathbb{R}$,*

$$P(W_t = b \text{ for some } t \geq 0) = 1.$$

Theorem 6.9 (Recurrence of BM) *Brownian motion (in dimension 1, starting at any value) is recurrent (i.e. with probability 1, there is an unbounded set of times t such that $W_t = W_0$).*

PROOF It is sufficient to show $P_0(W_s = 0 \text{ for some } s \geq 1) = 1$. We have

$$
\begin{aligned}
P_0(W_s = 0 \text{ for some } s \geq 1) &= \lim_{t\to\infty} P_0(W_s = 0 \text{ for some } s \in [1,t]) \\
&= \lim_{t\to\infty} \int_{-\infty}^{\infty} f_{W_1}(b) P_0(W_s = 0 \text{ for some } s \in [1,t] \,|\, W_1 = b)\, db
\end{aligned}
$$

$$
= \lim_{t\to\infty} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-b^2}{2\sigma^2}\right) \left[2 - 2\Phi\left(\frac{b}{\sigma\sqrt{t-1}}\right)\right] db
$$

$$
= \lim_{t\to\infty} \frac{2}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-b^2}{2\sigma^2}\right) \int_{\frac{b}{\sigma\sqrt{t-1}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx\, db
$$

From the previous page, we have

$$P_0(W_s = 0 \text{ for some } s \geq 1) = \lim_{t\to\infty} \frac{2}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-b^2}{2\sigma^2}\right) \int_{\frac{b}{\sqrt{t-1}}}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-u^2}{2\sigma^2}\right) du\, db$$

$$= \frac{1}{\pi\sigma^2} \int_{-\infty}^{\infty} \int_0^{\infty} \exp\left(\frac{-(u^2+b^2)}{2\sigma^2}\right) du\, db$$

$$= \frac{1}{\pi\sigma^2} \int_{r=0}^{\infty} \int_0^{\pi} \exp\left(\frac{-r^2}{2\sigma^2}\right) r\, d\theta\, dr$$

$$= \frac{1}{\sigma^2} \int_0^{\infty} e^{-r^2/2\sigma^2} r\, dr$$

Let $v = -r^2/2\sigma^2$ so that $dv = -\frac{r}{\sigma^2} dr$, i.e. $-\sigma^2\, dv = r\, dr$;

$$= \frac{1}{\sigma^2}(-\sigma^2) \int_0^{-\infty} e^v\, dv$$

$$= \int_{-\infty}^{0} e^v\, dv = e^0 - e^{-\infty} = 1.$$

6.3 Martingales associated to Brownian motion

Theorem 6.10 *Let $\{W_t\}$ be a standard Brownian motion. Then each of these is a martingale:*

- $\{W_t\}$
- $\{W_t^2 - t\}$
- $\left\{\exp\left(\theta W_t - \frac{\theta^2 t}{2}\right)\right\}$ *(for any constant $\theta \in \mathbb{R}$)*

PROOF We start with the proof that $\{W_t\}$ is a martingale. Let $\{\mathcal{F}_t\}$ be the natural filtration of $\{W_t\}$, and let $0 < s < t$:

$$\begin{aligned} E[W_t \,|\, \mathcal{F}_s] &= E[W_s + (W_t - W_s) \,|\, \mathcal{F}_s] \\ &= E[W_s \,|\, \mathcal{F}_s] + E[W_t - W_s \,|\, \mathcal{F}_s] \\ &= W_s + E[W_t - W_s \,|\, \mathcal{F}_s] \quad \text{(since } W_s \text{ is } \mathcal{F}_s\text{-mble)} \\ &= W_s + E[W_t - W_s] \quad \text{(since } W_t - W_s \perp \mathcal{F}_s\text{)} \\ &= W_s + 0 \quad \text{(since } W_t - W_s \text{ is } n(0, \sigma^2(t-s))\text{)} \\ &= W_s. \end{aligned}$$

Thus $\{W_t\}$ is a martingale by definition.

Theorem 6.11 *Let $\{W_t\}$ be a standard Brownian motion. Then for any stopping time T for which the OST holds, we have the following:*

- (Wald's First Identity for BM) $EW_T = EW_0$;
- (Wald's Second Identity for BM) $E[W_T^2] = ET$;
- (Wald's Third Identity for BM) $E\left[\exp\left(\theta W_T - \frac{\theta^2 T}{2}\right)\right] = 1$.

PROOF First, we prove Wald's First Identity. From the previous theorem, we know that $\{W_t\}$ is a martingale. By the OST, this means that $EW_T = EW_0$.

For the second identity, we know from the previous theorem that $\{W_t^2 - t\}$ is a martingale. By the OST, this means

$$\begin{aligned} 0 = E[W_0^2 - 0] &= E[W_T^2 - T] \\ &= E[W_T^2] - ET. \end{aligned}$$

Add ET to both sides to get Wald's Second Identity.

For the last identity, we know from the previous theorem that $\left\{\exp\left(\theta W_t - \frac{\theta^2 t}{2}\right)\right\}$ is a martingale, so by the OST we have

$$1 = E\left[\exp\left(\theta W_0 - \frac{\theta^2 \cdot 0}{2}\right)\right] = E\left[\exp\left(\theta W_T - \frac{\theta^2 T}{2}\right)\right]. \square$$

Theorem 6.12 *Let $\{W_t\}$ be a standard Brownian motion starting at x. Let $a, b \in \mathbb{R}$ with $a < x < b$ and let $T = \min(T_a, T_b) = T_{a,b}$. Then*

$$P_x(T_a < T_b) = \frac{b - x}{b - a}$$

and

$$ET = bx + ax - ab.$$

PROOF By Theorem 6.8, $P(T < \infty) = 1$, so $P_x(T_b < T_a) = 1 - P_x(T_a < T_b)$.

Brownian motion with drift

Idea: Unbiased simple random walk : BM :: Biased simple random walk : ?

Definition 6.13 *A* **Brownian motion with drift** *is a stochastic process* $\{X_t\}_{t\geq 0}$ *satisfying* $X_t = W_t + \mu t$*, where* $\mu \in \mathbb{R}$ *is a constant and* $\{W_t\}$ *is a BM.* μ *is called the* **drift parameter** *of* $\{X_t\}$*.*

Theorem 6.14 (Properties of BM with drift) *Suppose* $\{X_t\}$ *is a BM with drift. Then:*

1. *For each* t*,* X_t *is* $n(\mu t, \sigma^2 t)$*.*
2. (Independent increment property) *if* $t_1 < t_2 < t_3 < t_4$*, then* $X_{t_2} - X_{t_1} \perp X_{t_4} - X_{t_3}$*.*
3. (Time homogeneity) *For all* $s < t$*,* $X_t - X_s$ *is* $n(\mu(t-s), \sigma^2(t-s))$*.*
4. (Strong Markov property) *If* T *is any stopping time, then* $\{X_{T+t} - X_T\}$ *is a BM with drift (with the same parameters as* $\{X_t\}$*), independent of* $\{X_t\}_{t\leq T}$*.*

PROOF

Theorem 6.15 *Suppose $\{X_t\}$ is a BM with drift. Then the process $\{M_t\}$ defined by*

$$M_t = \exp\left(\frac{-2\mu}{\sigma^2} X_t\right)$$

is a martingale.

PROOF HW

Corollary 6.16 (Escape probabilities for BM with drift) *Suppose $\{X_t\}$ is a BM with drift (starting at 0). Then for all $a < 0$ and $b > 0$,*

$$P(T_b < T_a) = \frac{1 - \exp\left(\frac{-2\mu a}{\sigma^2}\right)}{\exp\left(\frac{-2\mu b}{\sigma^2}\right) - \exp\left(\frac{-2\mu a}{\sigma^2}\right)}$$

and

$$P(T_a < T_b) = \frac{\exp\left(\frac{-2\mu b}{\sigma^2}\right) - 1}{\exp\left(\frac{-2\mu b}{\sigma^2}\right) - \exp\left(\frac{-2\mu a}{\sigma^2}\right)}$$

PROOF HW

Application: Suppose the price of a stock is currently $\$70$. If the price is modeled with a BM with drift with $\mu = \frac{1}{2}$ and $\sigma^2 = 8$, what is the probability the price of the stock hits $\$80$ before it hits $\$60$?

6.4 Gaussian processes

Definition 6.17 *A stochastic process* $\{X_t : t \in \mathcal{I}\}$ *is called* **Gaussian** *if for any* $t_1, ..., t_n \in \mathcal{I}$*, the collection of random variables*

$$\mathbf{X} = (X_{t_1}, ..., X_{t_n})$$

has a joint normal distribution (i.e. every finite linear combination of the X_j *is normal).*

Recall from Math 414: Joint normal distributions are determined by a mean vector $\overrightarrow{\mu}$ and a covariance matrix Σ (see Math 414). Therefore, we see that a Gaussian process is completely determined if you know the mean of X_t for each t and the covariances between s and t for all s and t. Toward that end, we make the following definitions:

Definition 6.18 *Let $\{X_t\}$ be a stochastic process where $EX_t^2 < \infty$ for all $t \in \mathcal{I}$. The* **mean function** *of $\{X_t\}$ is the function $\mu_X : \mathcal{I} \to \mathbb{R}$ is defined by*

$$\mu_X(t) = E[X_t].$$

The **covariance function** *of $\{X_t\}$ is the function $r_X : \mathcal{I} \times \mathcal{I} \to \mathbb{R}$ is defined by*

$$r_X(s,t) = Cov(X_s, X_t).$$

Theorem 6.19 *A Gaussian process is determined completely by its mean and covariance functions, i.e. if two Gaussian processes have the same mean and covariance functions, then they are the same process.*

Example 1: Let Z_1 and Z_2 be i.i.d. $n(0, \sigma^2)$ r.v.s and let $\lambda > 0$. Define, for each $t \in [0, \infty)$, X_t by $X_t = Z_1 \cos \lambda t + Z_2 \sin \lambda t$.

1. Prove that $\{X_t\}$ is Gaussian.
2. Find the mean and covariance functions of $\{X_t\}$.
3. Find the variance of X_3.

Example 2: Let $\{X_t\}$ be a Poisson process with rate λ. Find the mean and covariance functions of $\{X_t\}$. Is $\{X_t\}$ Gaussian?

Theorem 6.20 *Brownian motion is a Gaussian process with $\mu_W(t) = 0$ and $r_W(s,t) = \sigma^2 \min(s,t)$.*

PROOF First, we show that $\{W_t\}$ is Gaussian: let $b_1, ..., b_n \in \mathbb{R}$ and let $t_1, ..., t_n \in [0, \infty)$; without loss of generality $t_1 < t_2 < ... < t_n$. Let $t_0 = 0$ (for notational purposes only). Then

$$\begin{aligned}
\sum_{j=1}^{n} b_j W_{t_j} &= b_1 W_{t_1} + b_2 W_{t_2} + ... + b_n W_{t_n} \\
&= b_1 W_{t_1} + b_2 \left[W_{t_1} + (W_{t_2} - W_{t_1})\right] + b_3 \left[W_{t_1} + (W_{t_2} - W_{t_1}) + (W_{t_3} - W_{t_2})\right] + ... \\
&= (b_1 + ... + b_n) W_{t_1} + (b_2 + ... + b_n)(W_{t_2} - W_{t_1}) + (b_3 + ... + b_n)(W_{t_3} - W_{t_2}) + ... \\
&= \left[\sum_{j=1}^{n} b_j\right] W_{t_j} + \left[\sum_{j=2}^{n} b_j\right] (W_{t_2} - W_{t_1}) + \left[\sum_{j=3}^{n}\right] (W_{t_3} - W_{t_2}) + ... \\
&= \sum_{i=1}^{n} \left[\sum_{j=i}^{n} b_j\right] (W_{t_i} - W_{t_{i-1}}).
\end{aligned}$$

All the terms inside the parentheses are normal (by the Markov property) and independent (by the independent increment property). Therefore any linear combination of them is normal, so $\sum_{j=1}^{n} b_j W_{t_j}$ is normal, so $\{W_t\}$ is Gaussian by definition.

Now for the mean function:

$$\mu_W(t) = E[W_t] = E[n(0, \sigma^2 t) = 0.$$

Finally, the covariance function: suppose first that $s \leq t$. Then

$$\begin{aligned}
r_W(s,t) = Cov(W_s, W_t) &= Cov(W_s, W_s + (W_t - W_s)) \\
&= Cov(W_s, W_s) + Cov(W_s, W_t - W_s) \\
&= Var(W_s) \\
&= \sigma^2 s.
\end{aligned}$$

If $t \leq s$, a symmetric computation gives $r_W(s,t) = \sigma^2 t$, so in general $r_W(s,t) = \sigma^2 \min(s,t)$ as desired. □

Theorem 6.21 *BM with drift is a Gaussian process with* $\mu_X(t) = \mu t$ *and* $r_W(s,t) = \sigma^2 \min(s,t)$.

PROOF HW

Theorem 6.22 *Let* $\{X_t\}$ *be a Gaussian process, and let* f *and* g *be functions from* $\mathbb{R}$ *to* $\mathbb{R}$. *Then, if for each t we set* $Y_t = f(t)X_{g(t)}$, $\{Y_t\}$ *is a Gaussian process whose mean and covariance functions are*

$$\mu_Y(t) = f(t)\mu_X(g(t))$$

$$r_Y(s,t) = f(s)f(T)r_X(g(s),g(t))$$

PROOF First, we will prove $\{Y_t\}$ is Gaussian. Let $t_1, ..., t_n \in \mathcal{I}$ and let $b_1, ..., b_n \in \mathbb{R}$. Then

$$\sum_{j=1}^{n} b_j Y_{t_j} = \sum_{j=1}^{n} b_j f(t_j) X_{g(t_j)} = \sum_{j=1}^{n} \left(b_j f(t_j)\right) X_{g(t_j)}$$

Since $\{X_t\}$ is assumed Gaussian, the linear combination above is therefore normal so $\{Y_t\}$ is Gaussian. Now for the mean function:

$$\mu_Y(t) = E[Y_t] = E[f(t)X_{g(t)}] = f(t)E[X_{g(t)}] = f(t)\mu_X(g(t)).$$

Finally, the covariance function:

$$\begin{aligned} r_Y(s,t) = Cov(Y_s, Y_t) = Cov(f(s)X_{g(s)}, f(t)X_{g(t)}) &= f(s)f(t)Cov(X_{g(s)}, X_{g(t)}) \\ &= f(s)f(t)r_X(g(s),g(t)). \end{aligned}$$

This completes the proof. $\square$

6.5 Symmetries and scaling laws

The upshot of the preceding theorem is that you take some process of the form $f(t)W_{g(t)}$, where $\{W_t\}$ is a BM, then you know that $\{X_t\}$ is Gaussian and you can work out the mean and covariance functions of $\{X_t\}$ using these formulas. It turns out that sometimes these mean and covariance functions are of the form $\mu_X(t) = 0$ and $r_X(s,t) = \sigma^2 \min(s,t)$, in which case you can conclude that $\{X_t\}$ is the same as $\{W_t\}$!

Theorem 6.23 *Let $\{W_t\}$ be a standard BM. Then each of the following processes are also standard BMs:*

- $-W_t$
- $W_{t+s} - W_s$ *(for any $s \geq 0$)*
- $tW_{1/t}$
- aW_{t/a^2} *(for any $a > 0$)*

The fact that aW_{t/a^2} is also a BM is called the **universal scaling law** *of BM.*

PROOF The idea behind the proof is that we can show these processes are Gaussian, and if we compute their mean and covariance functions and observe that those are the same as the mean and covariance functions of a BM, then we can conclude that they must be BMs.

The last two are left as homework exercises.

Corollary 6.24 (Nondifferentiability of paths) *Let $\{W_t\}$ be a BM, and fix $t_0 \geq 0$. With probability 1, the "Brownian path" $t \mapsto W_t$ is not differentiable at t_0.*

PROOF WLOG $t_0 = 0$; otherwise apply the second bullet of the previous theorem. Now

$$\begin{aligned}
\left.\frac{d}{dt}W_t\right|_{t=0} \text{ exists } &\iff \lim_{h\to 0}\frac{W_h - W_0}{h} \text{ exists} \\
&\iff \lim_{h\to 0}\frac{W_h}{h} \text{ exists} \\
&\Rightarrow \frac{W_h}{h} < A \text{ for some fixed constant } A\ \forall h \in (0,\epsilon) \\
&\iff W_h < Ah\ \forall h \in (0,\epsilon).
\end{aligned}$$

But by the Reflection Principle,

$$P(W_h < Ah) = 1 - \left(2 - 2\Phi\left(\frac{Ah}{\sqrt{h}}\right)\right) = 2\Phi(A\sqrt{h}) - 1$$

which goes to zero as $h \to 0$. Therefore

$$P\left(\left.\frac{d}{dt}W_t\right|_{t=0} \text{ exists}\right) = 0. \ \square$$

In fact, something stronger holds:

Theorem 6.25 (Nondifferentiability of paths) *Let $\{W_t\}$ be a BM. With probability 1, a Brownian path is nowhere differentiable (i.e. not differentiable at any time t).*

What this means is that with probability 1, the trajectory of a Brownian motion is "infinitely jagged", i.e. it is nowhere smooth. Furthermore, the universal scaling law tells us that if we take a trajectory of a BM, and zoom in on part of it (zooming in faster horizontally than we do vertically), we will see the same thing no matter how much we zoom in, i.e. the trajectories are "self-similar". Thus the trajectories in a BM are objects called **fractals**.

6.6 Zero sets of Brownian motion

Definition 6.26 *Let $\{W_t\}$ be a standard BM. The set $Z = \{t : W_t = 0\}$ (this is a subset of $\mathbb{R}$, not a r.v.) is called the* **zero set of** $\{W_t\}$.

Theorem 6.27 (Properties of zero sets) *Let $\{W_t\}$ be a standard BM. With probability one, the zero set Z has these properties:*

1. *Z is unbounded.*
2. *Z is closed, i.e. if $z_1, ..., z_n \in \mathbb{Z}$, then $\lim_{n\to\infty} z_n \in Z$.*
3. *Z is* **totally disconnected** *(i.e. Z does not contain an interval of positive length).*
4. *Z is* **perfect** *(i.e. for all $y \in Z$, there are points $z_1, z_2, ... \in Z$ with $z_j \neq y$ for all j but $\lim_{n\to\infty} z_j = y$)*
5. *$Z \cap (0, \epsilon)$ is infinite for any $\epsilon > 0$.*

Therefore Z is infinite, closed, perfect and totally disconnected. This makes Z something called a **Cantor set**. What do Cantor sets "look like"? A classical example of a Cantor set is the **middle-thirds Cantor set**:

PROOF Statement (1) follows from the fact that $\{W_t\}$ is recurrent.

Statement (2) follows from the fact that the sample functions $t \to W_t$ are continuous, hence preserve limits.

(3): Note that if $W_t = 0$ for all $t \in [0, \epsilon)$, then an infinite number of normal random variables would all have to be zero. The probability of this is zero (because among other things, normal r.v.s are cts so they take any individual value with probability zero).

(5): From Theorem 6.15, we see that $\{X_t\}$ defined by $X_t = tW_{1/t}$ is also a BM. By the recurrence of BM, there is an unbounded set of times $t_1, t_2, ...$ such that $X_t = 0$. But that means $W_{1/t_1}, W_{1/t_2}, ...$ must also all be zero. Now given any $\epsilon > 0$, there will be infinitely many of the times $\frac{1}{t_1}, \frac{1}{t_2}, ...$ in the interval $[0, \epsilon)$ (since the t_j are unbounded), so $\{W_t\}$ will have infinitely many zeros in $[0, \epsilon)$.

(4): **Case 1:** There is an increasing sequence of numbers $\{z_n\}$ in Z such that $z_n \to y$.

Case 2: There is not an increasing sequence of numbers in Z which converge to y.

6.7 Brownian motion in higher dimensions

Definition 6.28 *A stochastic process taking values in $\mathbb{R}^d$ is called* **standard d-dim'l Brownian motion** *if each coordinate of the process is a standard BM, and the coordinates are independent.*

Let $\{W_t\}$ be a standard $d-$dim'l BM and fix $0 < R_1 < R_2 < \infty$.

Let $A_1 = \{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}|| = R_1\}$;
let $A_2 = \{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}|| = R_2\}$;
let $A = \{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}|| \in (R_1, R_2)\}$;

let $T_1 = \min\{t \geq 0 : W_t \in A_1\}$;
let $T_2 = \min\{t \geq 0 : W_t \in A_2\}$;
let $T = \min\{T_1, T_2\}$.

Finally, for $\mathbf{x} \in A$, define $f(\mathbf{x}) = P_{\mathbf{x}}(T_2 < T_1)$ and set $f(\mathbf{x}) = 0$ if $\mathbf{x} \in A_1$ and set $f(\mathbf{x}) = 1$ if $\mathbf{x} \in A_2$.

By symmetry, $f(x) = g(||\mathbf{x}||)$ for some function $g : \mathbb{R} \to \mathbb{R}$.

In dimension 2:

$P(W_t$ never returns to within ϵ of $\mathbf{0}$, once it goes a distance $> \epsilon$ from $\mathbf{0})$

$=$

$P(W_t$ returns to $\mathbf{0}$ before going distance $> R_2$ from $\mathbf{0})$

$=$

<u>Conclusion:</u>

In dimension ≥ 3:

$P(W_t$ returns to within ϵ of $\mathbf{0}$, once it goes a distance $> \epsilon$ from $\mathbf{0})$

$=$

<u>Conclusion:</u>

We have shown the following set of facts:

Theorem 6.29 *Let $\{W_t\}$ be a standard, d-dim'l BM.*

1. *If $d = 1$, then $\{W_t\}$ is point recurrent.*
2. *If $d = 2$, then $\{W_t\}$ is point transient, but neighborhood recurrent.*
3. *If $d \geq 3$, then $\{W_t\}$ is transient.*

The results of this section can be used to solve some hitting time problems for Brownian motion:

Example: Suppose the price of a stock is modeled by a standard BM. If the price of the stock is initially 40, what is the probability that the stock price hits 60 before it hits 30?

Example: Suppose a 3-dimensional BM starts at the point $(1, 1, 1)$. What is the probability that the point strikes the sphere of radius 1 centered at the origin before it strikes the sphere of radius 2 centered at the origin?

Appendix A

Tables

A.1 Charts of properties of common random variables

The next page has a chart listing relevant properties of the common discrete random variables.

The following page has a chart listing relevant properties of the common continuous random variables.

Discrete Distribution X	Density Function $f_X(x)$	$E(X)$	Var(X)	PGF $G_X(t)$ MGF $M_X(t)$
uniform on $\{1, ..., n\}$	$f(x) = \frac{1}{n}$ for $x = 1, 2, ..., n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$G_X(t) = \frac{t(t^n-1)}{n(t-1)}$ $M_X(t) = \frac{e^t(e^{nt}-1)}{n(e^t-1)}$
binomial(n, p)	$f(x) = \binom{n}{x} p^x(1-p)^{n-x}$ for $x = 0, 1, ..., n$	np	$np(1-p)$	$G_X(t) = (1-p+pt)^n$ $M_X(t) = (1-p+pe^t)^n$
$Geom(p)$ $0 < p < 1$	$f(x) = p(1-p)^x$ for $x = 0, 1, 2, ...$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$G_X(t) = \frac{p}{1-(1-p)t}$ $M_X(t) = \frac{p}{1-(1-p)e^t}$
negative binomial $NB(r, p)$	$f(x) = \binom{r+x-1}{x} p^r(1-p)^x$ for $x = 0, 1, 2, ...$	$r\frac{1-p}{p}$	$r\frac{1-p}{p^2}$	$G_X(t) = \left(\frac{p}{1-(1-p)t}\right)^r$ $M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$
$Pois(\lambda)$	$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, 2, ...$	λ	λ	$G_X(t) = e^{\lambda(t-1)}$ $M_X(t) = e^{\lambda(e^t-1)}$
hypergeometric $Hyp(n, r, k)$	$f(x) = \frac{\binom{r}{x}\binom{n-r}{k-x}}{\binom{n}{k}}$ for $x = 0, 1, ..., k$	$\frac{kr}{n}$	$\frac{kr}{n}\left(\frac{n-r}{n}\right)\frac{n-k}{n-1}$	not given here
$d-$dimensional hypergeometric with parameters $n, (n_1, ..., n_d), k$	$f(x_1, ..., x_d) = \frac{\binom{n_1}{x_1}\binom{n_2}{x_2}\cdots\binom{n_d}{x_d}}{\binom{n}{k}}$ for $x_1 + x_2 + ... + x_d = k$	N/A	N/A	N/A
multinomial $n, (p_1, ..., p_d)$	$f(x_1, ..., x_d) = \frac{n!}{x_1!x_2!\cdots x_d!}p_1^{x_1}p_2^{x_2}\cdots p_d^{x_d}$ for $x_1 + x_2 + ... + x_d = n$	N/A	N/A	N/A

CONTINUOUS DISTRIBUTION X	DENSITY FUNCTION $f_X(x)$ DISTRIBUTION FUNCTION $F_X(x)$	EXPECTED VALUE EX VARIANCE $Var(X)$ MGF $M_X(t)$
uniform on (a,b)	$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{else} \end{cases}$ $F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b) \\ 1 & x \geq b \end{cases}$	$EX = \frac{a+b}{2}$ $Var(X) = \frac{(b-a)^2}{12}$ $M_X(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}$
exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$ $F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$EX = \frac{1}{\lambda}$ $Var(X) = \frac{1}{\lambda^2}$ $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$
Cauchy	$f(x) = \frac{1}{\pi(1+x^2)}$ $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$	$EX = \infty$ $Var(X)$ DNE $M_X(t)$ DNE
std. normal $n(0,1)$	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ $F(x) = \Phi(x)$	$EX = 0$ $Var(X) = 1$ $M_X(t) = e^{t^2/2}$
normal $n(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$ $F(x) = \Phi(\frac{x-\mu}{\sigma})$	$EX = \mu$ $Var(X) = \sigma^2$ $M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$
gamma $\Gamma(r, \lambda)$	$f(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$ F_X not given here	$EX = \frac{r}{\lambda}$ $Var(X) = \frac{r}{\lambda^2}$ $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$ for $t < \lambda$
joint normal with mean vector μ; covariance matrix Σ	$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left[\frac{-1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right]$	$E\mathbf{X}$ and $Var(\mathbf{X})$ DNE $M_{\mathbf{X}}(\mathbf{t}) = \exp\left(\mathbf{t} \cdot \mu + \frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}\right)$

A.2 Useful sum and integral formulas

Triangular Number Formula: For all $n \in \{1, 2, 3, ...\}$,

$$1 + 2 + 3 + ... + n = \sum_{j=0}^{n} j = \frac{n(n+1)}{2}.$$

Finite Geometric Series Formula: for all $r \in \mathbb{R}$,

$$\sum_{n=0}^{N} r^n = \frac{1 - r^{N+1}}{1 - r}.$$

Infinite Geometric Series Formulas: for all $r \in \mathbb{R}$ such that $|r| < 1$,

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1 - r} \qquad \sum_{n=N}^{\infty} r^n = \frac{r^N}{1 - r}.$$

Derivative of the Geometric Series Formula: for all $r \in \mathbb{R}$ such that $|r| < 1$,

$$\sum_{n=0}^{\infty} n r^n = \frac{r}{(1 - r)^2}.$$

Exponential Series Formula: for all $r \in \mathbb{R}$,

$$\sum_{n=0}^{\infty} \frac{r^n}{n!} = e^r.$$

Binomial Theorem: for all $n \in \mathbb{N}$, and all $x, y \in \mathbb{R}$,

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x + y)^n.$$

Vandermonde Identity: for all $n, k, r \in \mathbb{N}$,

$$\sum_{x=0}^{n} \binom{r}{x} \binom{n - r}{k - x} = \binom{n}{k}.$$

Gamma Integral Formula: for all $r > 0$, $\lambda > 0$,

$$\int_0^{\infty} x^{r-1} e^{-\lambda x}\, dx = \frac{\Gamma(r)}{\lambda^r}.$$

Normal Integral Formula: for all $\mu \in \mathbb{R}$ and all $\sigma > 0$,

$$\int_{-\infty}^{\infty} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right) dx = \sigma\sqrt{2\pi}.$$

Beta Integral Formula: for all $r > 0$, $\lambda > 0$,

$$\int_0^1 x^{\alpha-1}(1 - x)^{\beta-1}\, dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

A.3 Table of values for the cdf of the standard normal

Entries represent $\Phi(z) = P(n(0,1) \leq z)$. The value of z to the first decimal is in the left column. The second decimal place is given in the top row.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8436	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Index