# Markov Chains and Martingales

David M. McClendon

Department of Mathematics
Ferris State University

2021 edition

# Contents

*Chapter 1*

# Markov chains

## 1.1 What is 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 the language and theory of probability we learned in MATH 414.

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 chapter of these notes focuses on discrete-time processes; Chapter 2 contains ideas useful in both settings, and Chapters 3 and 4 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. Sums and 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* **Markov chain** *is a discrete-time stochastic process with finite or countable state space that has the Markov property.*

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.

## The three ingredients of a Markov chain

**Question:** What are the "ingredients" of 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 $\pi_0$:
   $$\pi_0(x) = P(X_0 = x) \text{ for all } x \in \mathcal{S}$$
   $\pi_0(x)$ 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 $t - 1$, that you are in state $y$ at the next time (which is time $t$).

   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.

## Simulating a Markov chain

To get used to how Markov chains work, let's simulate one using a computer. Let's suppose:

- the state space is $\mathcal{S} = \{1, 2, 3\}$;
- the initial distribution $\pi_0$ satisfies $\pi_0(1) = \frac{1}{2}$, $\pi_0(2) = \frac{1}{6}$ and $\pi_0(3) = \frac{1}{3}$. We shorthand this by writing $\pi_0$ as

$$\pi_0 = \left(\frac{1}{2}, \frac{1}{6}, \frac{1}{3}\right).$$

- the transition probabilities are $P(1,1) = \frac{1}{4}$, $P(1,2) = \frac{1}{2}$, $P(1,3) = \frac{1}{4}$, $P(2,1) = \frac{3}{4}$, $P(2,2) = 0$, $P(2,3) = \frac{1}{4}$, $P(3,1) = \frac{1}{4}$, $P(3,2) = \frac{1}{4}$, $P(3,3) = \frac{1}{2}$. A shorthand way of writing all these is by treating them as entries of a matrix:

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

We can capture the state space and the transition probabilities with the following picture:

To simulate this Markov chain, we first have to select the state $X_0$ in which we start. This is done using $\pi_0$: we pick state $1$ with probability $\frac{1}{2}$, state $2$ with probability $\frac{1}{6}$, and state $3$ with probability $\frac{1}{3}$.

One way to perform this random choice on a computer is to have the computer generate a "uniformly random" real number in $[0, 1]$ (in *Mathematica*, you use the RandomReal[ ] command to do this). If the number is less than $\frac{1}{2}$, let $X_0 = 1$; if the number is between $\frac{1}{2}$ and $\frac{1}{2} + \frac{1}{6}$, let $X_0 = 2$; otherwise $X_0 = 3$:

Suppose we picked $X_0 = 3$. Now, since $X_0 = 3$, we pick the next state $X_1$ using Row 3 of $P$. By this, I mean $X_1 = 1$ with probability $\frac{1}{4}$, $X_1 = 2$ with probability $\frac{1}{4}$, and $X_1 = 3$ with probability $\frac{1}{2}$ (if you did this on a computer by selecting a random real number in $[0, 1]$, then $X_1$ would be determined as follows:

Let's suppose our random choice led to $X_1 = 1$. The next thing to do is to pick the state $X_2$, which is done by using Row 1 of $P$ (so $X_2 = 1$ with probability $\frac{1}{2}$, etc.). The idea expressed in the Markov property is that so long as we know $X_1$, the fact $X_0$ was 3 is no longer relevant to the computation of $X_2$, i.e. that $X_0 = 3$ is "old news".

Similarly, once you've figured $X_2$, the fact that $X_1 = 1$ doesn't influence how $X_3$ is generated, etc.

To get the rest of the chain $\{X_t\}$, you pick each state $X_t$ from the previous one $X_{t-1}$: if $X_{t-1} = j$, $X_t$ is chosen using Row $j$ of $P$ as described above.

## 1.2 Basic examples of Markov chains

EXAMPLE 1: I.I.D PROCESS (OF DISCRETE R.V.S)

**State space:** $\mathcal{S} =$

**Initial distribution:**

**Transition probabilities:**

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

EXAMPLE 2: BERNOULLI PROCESS

**State space:** $\mathcal{S} = \mathbb{N} = \{0, 1, 2, 3, ...\}$

**Initial distribution:**

**Transition probabilities:**

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

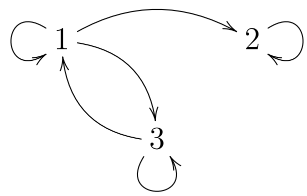
We represent these transition probabilities with the following picture:

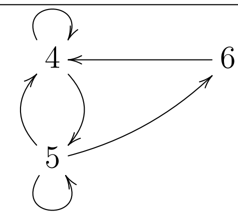
*The above picture generalizes:* Every Markov chain can be thought of as a random walk on a graph as follows:

**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 $y$.*

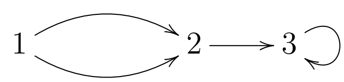
EXAMPLES OF DIRECTED GRAPHS

MORE DIRECTED GRAPHS




NOT A DIRECTED GRAPH:




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 the transition probabilities of 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: BASIC URN MODEL

An urn initially holds $2$ red and $2$ green marbles. Every minute, you choose a marble uniformly from the urn. If you draw a red marble, you put the red marble back in the urn, and add two green marbles from the urn. If you draw a green marble, you leave it out of the urn. Let $X_t$ be the number of green marbles in the jar after $t$ draws.

EXAMPLE 4: GAMBLER'S RUIN

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.

## 1.3 Matrix theory applied to Markov chains

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.*

**Why do we put the transition probabilities in a matrix?** We will see that we can answer almost any question about a finite state space Markov chain by performing some simple matrix algebra associated to the transition matrix $P$.

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.*

## $n$-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*

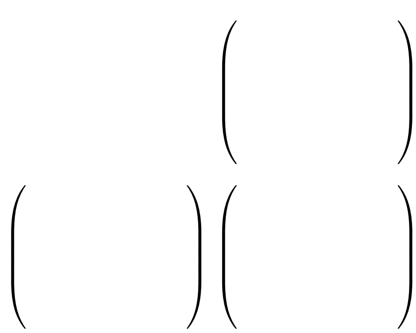
$$P^n(x, y) = (P^n)_{x,y},$$

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

PROOF I'm going to prove this only when $n = 2$ (the proof for general $n$ uses a proof technique called "induction", for which $n = 2$ constitutes the base case). By time homogeneity,

$$P^2(x, y) = P(X_2 = y \,|\, X_0 = x)$$

Now, recall how matrix multiplication works:



## 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 $\pi$ on $(\mathcal{S}, 2^{\mathcal{S}})$, i.e. a function $\pi : \mathcal{S} \to [0,1]$ such that $\sum_{x \in \mathcal{S}} \pi(x) = 1$.*

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

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

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

This is unusual, as normally one would represent a vector in $\mathbb{R}^d$ as a column matrix, but this convention makes upcoming formulas easier.

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

$$\pi_0 = (\pi_0(1), \pi_0(2), ..., \pi_0(d))_{1 \times d}$$

*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:

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

EXAMPLE 5

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 $P(X_8 = 1 \,|\, X_7 = 0)$.
5. Find $P(X_7 = 0 \,|\, X_4 = 0)$.

## Markov chains with infinite state space

Although the formulas for $n-$step transitions and time $n$ distributions are motivated by those obtained earlier in this 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 $\pi_n$ 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 $\pi_0$. 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 $\pi_n$ satisfies, for all $y \in \mathcal{S}$,*

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

## 1.4 The Fundamental Theorem of Markov chains

Many areas of mathematics have a central result which is key to understanding the ideas of the subject. These central results are called "Fundamental Theorems":

**Fundamental Theorem of Arithmetic:** every integer greater than $1$ can be factored uniquely into a product of prime numbers.

**Fundamental Theorem of Algebra:** every polynomial whose coefficients are in $\mathbb{C}$ has a root in $\mathbb{C}$.

**Fundamental Theorem of Calculus:** if $f : \mathbb{R} \to \mathbb{R}$ is cts and $F(x) = \int_a^x f(t)\,dt$, then $F'(x) = f(x)$. (Also, if $f : \mathbb{R} \to \mathbb{R}$ is cts with antiderivative $F$, then $\int_a^b f(x)\,dx = F(b) - F(a)$.)

**Fundamental Theorem of Line Integrals:** If $\mathbf{f} = \nabla f$ is a conservative vector field on $\mathbb{R}^n$, then for any piecewise $C^1$ curve $\gamma$ with initial point $\mathbf{a}$ and terminal point $\mathbf{b}$, $\int_\gamma \mathbf{f} \cdot d\mathbf{s} = f(\mathbf{b}) - f(\mathbf{a})$.

**Fundamental Theorem of Linear Algebra:** If $A \in M_{mn}(\mathbb{R})$, then $\dim C(A) = \dim R(A)$, $[R(A)]^\perp = N(A)$ and $[C(A)]^\perp = N(A^T)$.

This section is about the Fundamental Theorem of Markov Chains (FTMC). To get an idea of what this theorem is about, we'll do some experimentation.

What we (almost assuredly) saw in our experiment is that the Markov chain we invented had a special distribution $\pi$, so that as $n \to \infty$, the time $n$ distributions $\pi_n$ approached this distribution $\pi$, no matter what the initial distribution was. The FTMC says that for most (not all) Markov chains, this phenomenon holds:

**Theorem 1.17 (Fundamental Theorem of Markov Chains (FTMC))** *Let $\{X_t\}$ be an irreducible, aperiodic, positive recurrent Markov chain. Then $\{X_t\}$ has a unique stationary distribution $\pi$, such that $\pi$ is steady-state, meaning*

$$\lim_{n\to\infty} \pi_n(x) = \pi(x)$$

*for all $x \in \mathcal{S}$, no matter the initial distribution $\pi_0$.*

To understand this theorem, we need to learn the meaning of its vocabulary: "irreducible", "aperiodic", "positive recurrent", "stationary", "steady-state". Learning this vocabulary is the goal of the next four sections.

## 1.5 Stationary distributions

**Recall:** A Markov chain is determined by two things:

-
-

From this, you get time $n$ distributions $\pi_n$ 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 the transition matrix)

We are investigating the FTMC, which says that if $\{X_t\}$ is "irreducible", "aperiodic" and "positive recurrent", then there is a "stationary, steady-state" distribution $\pi$ such that $\pi_n(x)$ approaches $\pi(x)$ for all $x \in \mathcal{S}$. This means that for large $n$, $\pi_n(x)$ can be approximated by $\pi(x)$.

**Question:** What do "stationary" and "steady-state" mean?

## Stationary distributions

Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Suppose $\pi$ is a distribution on $\mathcal{S}$ so that, if the initial distribution $\pi_0$ is $\pi$, the time $1$ distribution $\pi_1$ is also $\pi$. Then $\pi$ is called "stationary" (because it didn't change as time passed). More precisely:

**Definition 1.18** *Let $\{X_t\}$ be a Markov chain. A distribution $\pi$ on $\mathcal{S}$ is called* **stationary** *(with respect to $\{X_t\}$) if for all $y \in \mathcal{S}$,*

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

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

$$\pi \, P = \pi$$

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

**Lemma 1.19** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $\pi$ 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 1.20** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. An initial distribution $\pi_0$ is stationary if and only if the time $n$ distributions are the same for every $n$.*

PROOF ($\Rightarrow$) Assume $\pi_0$ 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*.

## Steady-state distributions

A steady-state distribution for a Markov chain is like the special one in our experiment: if $\pi$ is steady-state for $\{X_t\}$, then no matter the initial distribution $\pi_0$, $\pi_n(x) \to \pi(x)$ as $n \to \infty$, so for large $n$, $\pi_n(x) \approx \pi(x)$. More precisely:

**Definition 1.21** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. A distribution $\pi$ on $\mathcal{S}$ is called* **steady-state** *(with respect to $\{X_t\}$) if*

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

**Theorem 1.22** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Suppose $\pi$ is a steady-state distribution for $\{X_t\}$. Then for any initial distribution $\pi_0$,*

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

PROOF By Theorem 1.16 (3), we get the top equation below. Then take the limit of both sides as $n \to \infty$:

$$\begin{array}{ccccc}
\pi_n(y) & = P(X_n = y) & = \sum_{x\in\mathcal{S}} \pi_0(x) P^n(x, y) \\
\Big\downarrow {\scriptstyle n\to\infty} & & \Big\downarrow {\scriptstyle n\to\infty} \\
\lim_{n\to\infty} \pi_n(y) & & \sum_{x\in\mathcal{S}} \pi_0(x)\pi(y)
\end{array}$$

So *steady-state distributions are those which "attract" the time $n$ distribution as $n$ increases, no matter the initial distribution.*

EXAMPLE 6

Let $p, q \in (0,1)$ (there is no relationship between $p$ and $q$). Consider a Markov chain with $\mathcal{S} = \{0,1\}$ whose transition matrix is

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

Find all stationary distributions of this Markov chain (there might not be any).

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.

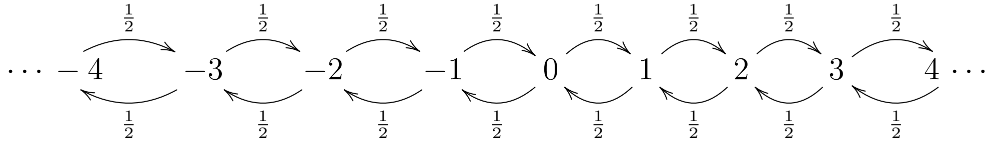
EXAMPLE 7

Find all stationary distributions of $\{X_t\}$, if $\{X_t\}$ has transition matrix

$$\begin{pmatrix} \frac{1}{7} & \frac{4}{7} & \frac{2}{7} \\ 0 & \frac{5}{7} & \frac{2}{7} \\ \frac{3}{7} & \frac{1}{7} & \frac{3}{7} \end{pmatrix}.$$

EXAMPLE 8

Let $\{X_t\}$ be simple, unbiased random walk on $\mathbb{Z}$ (this means $\mathcal{S} = \mathbb{Z}$, and for every $x \in \mathcal{S}$, $P(x, x+1) = P(x, x-1) = \frac{1}{2}$). Find all stationary distributions of $\{X_t\}$.



## Uniqueness of stationary and steady-state distributions

**Big picture questions:** Given Markov chain $\{X_t\}$:

1. Does $\{X_t\}$ have a stationary distribution?

2. If so, how many stationary distributions does it have?

3. Does $\{X_t\}$ have a steady-state distribution?

4. If so, how many steady-state distributions does it have?

In the rest of this section, we are going to run through some theorems addressing these questions. We'll start with ideas related to Question 2 above.

**Definition 1.23** *Suppose $\pi_1, \pi_2, \pi_3, ...$ 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 $\alpha_j$ are nonnegative numbers satisfying $\sum_j \alpha_j = 1$.*

EXAMPLE

Let $\pi_1 = (.1, .5, .4)$, $\pi_2 = (0, 1, 0)$ and $\pi_3 = (.7, .2, .1)$. The distribution

$$\begin{aligned} .5\pi_1 + .3\pi_2 + .2\pi_3 &= .5(.1, .5, .4) + .3(0, 1, 0) + .2(.7, .2.1) \\ &= (.05, .25, .2) + (0, .3, 0) + (.14, .04, .02) \\ &= (.19, .59, .22) \end{aligned}$$

is a convex combination of $\pi_1, \pi_2$ and $\pi_3$ with $\alpha_1 = .5, \alpha_2 = .3$ and $\alpha_3 = .2$.

**Lemma 1.24** *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 $\alpha_j$ are nonnegative, then $\pi(x) \geq 0$ for all $x$ as well, so $\pi$ is a distribution. □

**Special case:** A convex combination of two distributions $\pi_1$ and $\pi_2$ is a distribution of the form

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

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

**Theorem 1.25 (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 $\pi_j$ is also a stationary distribution for $\{X_t\}$.*

PROOF HW (you have to check that the stationarity equation $\pi(y) = \sum_{x \in \mathcal{S}} \pi(x) P(x, y)$ holds for the convex combination)

**Corollary 1.26 (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 $\pi_1$ and $\pi_2$. 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 $\alpha$, the Markov chain will have infinitely many stationary distributions. □

Now we turn to Question 4 from earlier (how many steady-state distributions can a Markov chain have?).

To handle this, we first need to introduce a theorem that we won't prove but will use frequently; this theorem comes from a branch of mathematics called *real analysis*:

**Theorem 1.27 (Bounded Convergence Theorem for Sums (BCT))** *Let a(z) be non-negative numbers such that* $\sum_z a(z) < \infty$. *Fix* $B > 0$ *and let* $b_n(z)$ *be numbers such that* $|b_n(z)| \leq B$ *for all* $z$ *and* $n$ *and*

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

*Then*

$$\sum_z a(z) b_n(z) \overset{n\to\infty}{\longrightarrow} \sum_z a(z) b(z),$$

*in other words*

$$\lim_{n\to\infty} \sum_z a(z) b_n(x) = \sum_z \lim_{n\to\infty} a(z) b_n(z) = \sum_z a(z) \lim_{n\to\infty} b_n(z).$$

**Why this theorem isn't obvious:** Suppose $z \in \{1,2,3,...\}$, $b_n(z) = \frac{z^2}{n^2}$ and $a(z) = \frac{1}{z^2}$. Then $b(z) = \lim_{n\to\infty} b_n(z) = 0$ so

$$\lim_{n\to\infty} \sum_{z=1}^{\infty} a(z) b_n(z) = \lim_{n\to\infty} \sum_{z=1}^{\infty} \frac{1}{z^2} \left( \frac{z^2}{n^2} \right) = \lim_{n\to\infty} \sum_{z=1}^{\infty} \frac{1}{n^2} = \lim_{n\to\infty} \infty = \infty$$

but

$$\sum_{z=1}^{\infty} a(z) \lim_{n\to\infty} b_n(z) = \sum_{z=1}^{\infty} \frac{1}{z^2} b(z) = \sum_{z=1}^{\infty} \frac{1}{z^2}(0) = \sum_{z=1}^{\infty} 0 = 0.$$

Observe that the $b_n(z)$ in this example don't work in the BCT, since they are not bounded by any $B$.

**Moral:** You cannot interchange an infinite sum over $z$ (or over $x$ or $y$) and a limit as $n \to \infty$ without the BCT (or some other theorem).

(However, if there are only finitely many $z$s in the sum, then you can always interchange the limit with the sum.)

**Theorem 1.28 (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 $\pi$, then*

1. *$\pi$ is stationary for $\{X_t\}$; and*
2. *$\pi$ is the only stationary distribution for $\{X_t\}$.*

PROOF Let's start with statement (1). Suppose $\pi$ is steady-state. Let $y \in \mathcal{S}$.

$$\begin{aligned}
\pi(y) &= \lim_{n\to\infty} P^n(x,y) \\
&= \lim_{n\to\infty} P^{n+1}(x,y) \quad \text{(def'n of steady-state)} \\
&= \lim_{n\to\infty} \sum_{z\in\mathcal{S}} P^n(x,z)P(z,y) \quad \text{(LTP)}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{z\in\mathcal{S}} \lim_{n\to\infty} P^n(x,z)P(z,y) \\
&= \sum_{z\in\mathcal{S}} \pi(z)P(z,y) \quad \text{(def'n of steady-state).}
\end{aligned}$$

Since $\pi(y) = \sum_{z\in\mathcal{S}} \pi(z)P(z,y)$, $\pi$ is stationary by definition.

Now for statement (2). 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 $\pi_0$ 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 $\pi$ is steady-state. □

EXAMPLE 6, REVISITED

For the Markov chain whose state space is $\mathcal{S} = \{0,1\}$ and whose transition matrix is $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$, we saw that the stationary distribution was

$$\pi = \left( \frac{q}{p+q}, \frac{p}{p+q} \right).$$

Is this distribution steady-state?

*Solution:* $\pi$ is steady-state if $\lim_{n\to\infty} P^n(x,y) = \pi(y)$ for all $x, y \in \mathcal{S}$, i.e.

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}.$$

**Q:** How might we compute powers $P^n$ of the matrix $P$?

**A:**

If you did all that for this matrix $P$, you'd find

$$\lambda = 1 \leftrightarrow (1,1) \qquad \lambda = 1-p-q \leftrightarrow (-p,q)$$

so

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix} \qquad S = \begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix}$$

and therefore (after some calculation)

$$P^n = S\Lambda^n S^{-1} = \begin{pmatrix} \frac{q}{p+q} + \frac{p}{p+q}(1-p-q)^n & \frac{p}{p+q} - \frac{q}{p+q}(1-p-q)^n \\ \frac{q}{p+q} - \frac{p}{p+q}(1-p-q)^n & \frac{p}{p+q} + \frac{q}{p+q}(1-p-q)^n \end{pmatrix}.$$

Since $-1 < 1-p-q < 1$, $\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}$, so $\lim_{n\to\infty} P^n(x,y) = \pi(y)$ and $\pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ is indeed steady-state.

**Remark:** There will be a better way of showing $\pi$ is steady-state, based on theory we will develop in this chapter.

## 1.6 Class structure and periodicity

**What this section is about:** The FTMC says that if $\{X_t\}$ is irreducible, aperiodic and positive recurrent, then it has a steady-state distribution. In this section, we discuss what is meant by "irreducible" and "aperiodic".

To get started, we need to establish some notation that we'll use frequently.

**Definition 1.29** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$.*

1. *Given an event $E$, define $P_x(E) = P(E \mid X_0 = x)$. This is the probability of event $E$, given that you start at $x$.*
2. *Given a r.v. $Z$, define $E_x(Z) = E(Z \mid X_0 = x)$. This is the expected value of $Z$, given that you start at $x$.*

**Definition 1.30** *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 $\sum_{n=1}^{\infty} P(T_A = n) = 1 - P(T_A = \infty) \leq 1$.

### Class structure

**Definition 1.31** *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. *We say $x$ and $y$* **communicate** *(and write $x \leftrightarrow y$) if $x \to y$ and $y \to x$.*

**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 all members of $C$ communicate.*
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.

**Remark:** whether or not a Markov chain is irreducible depends only on its transition probabilities, and not on its initial distribution.

**Lemma 1.33** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Then*

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

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

Notice

$$P_x(T_y < \infty) = \sum_{n=1}^{\infty} P_x(T_y = n),$$

so if this sum is $> 0$, there must be at least one $N$ such that $P_x(T_y = N) > 0$.

Since $P^N(x,y) \geq P_x(T_y = N)$, we can conclude $P^N(x,y) > 0$ as wanted.

($\Leftarrow$) Suppose $P^N(x,y) > 0$ for one or more $N$. Take the smallest such $N$; for this $N$, we have

$$P_x(T_y = N) = P^N(x,y) > 0.$$

Therefore

$$f_{x,y} = P_x(T_y < \infty) \geq P_x(T_y = N) > 0,$$

so $x \to y$ as wanted. □

**Lemma 1.34** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Then*

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

PROOF Apply Lemma 1.33 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 1.33 $x \to z$. □

EXAMPLE 9

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

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

Find all closed sets and all communicating classes of $\{X_t\}$.

**Remark:** To solve Example 9, the only thing relevant is whether the entries of $P$ are zero or nonzero. So long as an entry is nonzero, whether it is $\frac{1}{2}$ or $\frac{1}{4}$ or whatever doesn't affect the closed sets and communicating classes of $\{X_t\}$.

EXAMPLE 10

Each matrix below is the transition matrix of a Markov chain with state space $\{1,2,3,4\}$. The "+" in the matrices represent arbitrary positive numbers. For each Markov chain, find all its communicating classes and determine if the chain is irreducible.

$$\begin{pmatrix} + & 0 & + & 0 \\ 0 & + & + & 0 \\ 0 & + & + & + \\ + & 0 & 0 & + \end{pmatrix} \qquad \begin{pmatrix} + & 0 & + & 0 \\ 0 & + & + & 0 \\ + & 0 & + & 0 \\ + & + & 0 & + \end{pmatrix}$$

**Recall:** One of the necessary ingredients in the FTMC is that the chain is irreducible. In the next example, we see why irreducibility is important to ensuring the existence of a steady-state distribution.

EXAMPLE 11

Suppose $\{X_t\}$ is a Markov chain with state space $\{0, 1\}$ whose transition matrix is the $2 \times 2$ identity matrix ($P = I$).

1. Sketch the directed graph of this Markov chain, and find its communicating classes. Is $\{X_t\}$ irreducible?

2. Find all stationary distributions of this Markov chain.

3. Does $\{X_t\}$ have a steady-state distribution? Explain.

## Periodicity

To explain the concept of periodicity, let's start with this simple example, which illustrates why "aperiodicity" is important in the FTMC:

EXAMPLE 12

Suppose $\{X_t\}$ is a Markov chain with state space $\{0,1\}$ whose transition matrix is

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

1. Sketch the directed graph of this Markov chain, and find its communicating classes. Is $\{X_t\}$ irreducible?
2. Find all stationary distributions of this Markov chain.
3. Suppose $\pi_0 = (1,0)$. Compute $\pi_n$ for every $n$. Does $\lim\limits_{n\to\infty} \pi_n$ exist?
4. Does $\{X_t\}$ have a steady-state distribution? Explain.

The problem with the Markov chain in Example 12 (i.e. what causes its stationary distribution to not be steady-state) is that it is "periodic"... if you start in a certain state, you can only return to that state at times that are a multiple of $2$. This means the chain has period $2$. More generally:

**Definition 1.35** *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

$$6\,|\,42 \qquad 5 \nmid 42 \qquad 3\,|\,180$$

$$\gcd\{12,36\} = 12 \qquad \gcd\{18,27,15\} = 3 \qquad \gcd\{8,17\} = 1$$

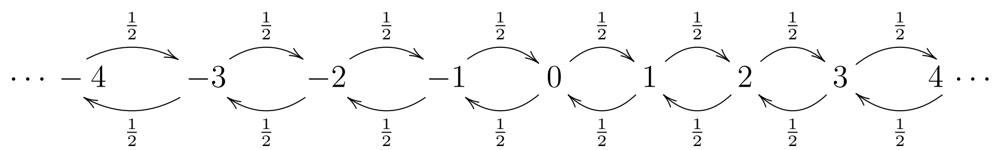
**Definition 1.36** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $x \in \mathcal{S}$ be such that $f_x > 0$ (equivalently, $P^n(x,x) > 0$ for some $n \geq 1$; equivalently, $x \to x$). The* **period** *of $x$, denoted by $d_x$, is the largest integer which divides every $n$ for 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 13

Let $\{X_t\}$ be simple, unbiased random walk on $\mathbb{Z}$ (this means $\mathcal{S} = \mathbb{Z}$, and for every $x \in \mathcal{S}$, $P(x,x+1) = P(x,x-1) = \frac{1}{2}$).




Find the period of each state.

**Theorem 1.37 (Communicating states have the same period)** *Suppose $\{X_t\}$ is a Markov chain with state space $\mathcal{S}$. Let $x, y \in \mathcal{S}$ be such that $x \leftrightarrow y$. 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 $d_x \,|\, n$.

A symmetric argument shows $d_y \leq d_x$, so $d_x = d_y$ as wanted. □

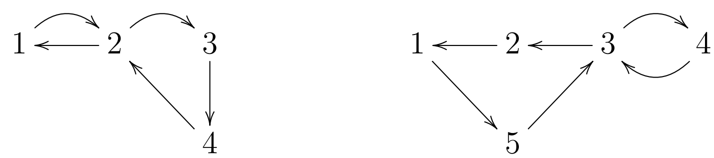
Theorem 1.37 shows that period is a **class property**, meaning that it is a property shared by all members of a communicating class. This implies:

**Corollary 1.38** *If $\{X_t\}$ is an irreducible Markov chain, all states have the same period.*

**Definition 1.39** *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 $x \in \mathcal{S}$.*

EXAMPLE 14

Find the period of each Markov chain whose directed graph is given below.




One important consequence of aperiodicity is that in an irreducible, aperiodic Markov chain, for every pair of states you can get from one to the other in any sufficiently large amount of time. This is made precise in Theorem 1.40:

**Theorem 1.40** *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 of claim:** 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

$$n - n_1^2 = n - N = mn_1 + r$$

where $m \in \mathbb{N}$ and $r \in \{0, 1, ..., n_1 - 1\}$. Rewriting this, we get

$$n = mn_1 + r + n_1^2.$$

Rewriting this again, we get

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

which is in $I$ since $n_1 + 1 \in I$ and $n_1 \in I$. $\square$

## 1.7 Recurrence and transience

**What this section is about:** We are going to divide the states of a Markov chain into different "types". There will be general laws which govern the behavior of each "type" of state, and the types of states of the chain gives you information about whether the chain has stationary distributions and/or a steady-state distribution.

**Definition 1.41** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$.*

1. *For each $x \in \mathcal{S}$, set $f_x = f_{x,x} = P_x(T_x < \infty)$.*
2. *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.*
3. *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 two of the "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".

## Elementary properties of recurrent and transient states

The rest of this section is devoted to developing properties of recurrent and transient states. The key to deriving these properties is to introduce random variables which count the number of times a Markov chain "visits" each state.

**Definition 1.42** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. For each $y \in \mathcal{S}$, define*

$$V_y = \# \text{ of times } t \geq 1 \text{ such that } X_t = y.$$

*$V_y$ is a r.v. called the* **number of visits** *to $y$.*

*For each $x \in \mathcal{S}$ and each $N \in \{1, 2, 3, ...\}$, define*

$$V_{y,N} = \# \text{ of times } t \in \{1, 2, ..., N\} \text{ such that } X_t = y.$$

*$V_{y,N}$ is a r.v. called the* **number of visits** *to $y$* **up to time** *$N$.*

**Note:** $V_y : \Omega \to \{0, 1, 2, 3, ...\} \bigcup \{\infty\}$, but $V_{y,N} : \Omega \to \{0, 1, 2, 3, ..., N-1\}$.

**Lemma 1.43 (Formula for expected number of visits)** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Then, for any $x, y \in \mathcal{S}$, we have*

$$E_x(V_y) = \sum_{n=1}^{\infty} P^n(x, y) \quad \text{and} \quad E_x(V_{y,N}) = \sum_{n=1}^{N} P^n(x, y).$$

PROOF Let $f : \mathcal{S} \to \{0, 1\}$ be defined by

$$f(s) = \begin{cases} 1 & \text{if } s = y \\ 0 & \text{else} \end{cases}$$

Then, the first equation follows as a direct calculation (the second statement is proved the same way, with an $N$ in place of the $\infty$):

$$\begin{aligned} E_x(V_y) &= E_x\left[\sum_{n=1}^{\infty} f(X_n)\right] \\ &= \sum_{n=1}^{\infty} E_x(f(X_n)) \\ &= \sum_{n=1}^{\infty} [1 \cdot P_x(X_n = y) + 0 \cdot P_x(X_n \neq y)] \\ &= \sum_{n=1}^{\infty} P_x(X_n = y) \\ &= \sum_{n=1}^{\infty} P^n(x, y). \ \square \end{aligned}$$

**Theorem 1.44 (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 may be possible that you 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 $P_x(V_y = n) =$

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

$$P_x(V_y = \infty) = \lim_{n\to\infty} P_x(V_y \geq n) = \lim_{n\to\infty} f_{x,y} f_y^{n-1} = 0$$

so $P_x(V_y < \infty) = 1$ as wanted. Also,

$$\begin{aligned} E_x(V_y) &= \sum_{n=0}^{\infty} n \cdot P_x(V_y = n) \\ &= \sum_{n=1}^{\infty} n \cdot P_x(V_y = n) \\ &= \sum_{n=1}^{\infty} n f_{x,y} f_y^{n-1}(1 - f_y) \qquad \text{(from above)} \\ &= f_{x,y}(1 - f_y) \sum_{n=1}^{\infty} n f_y^{n-1} \\ &= f_{x,y}(1 - f_y) \frac{1}{(1 - f_y)^2} \qquad \text{(pink sheet)} \\ &= \frac{f_{x,y}}{1 - f_y}. \end{aligned}$$

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

$$P_x(V_y = \infty) = \lim_{n\to\infty} P_x(V_y \geq n) = \lim_{n\to\infty} f_{x,y} f_y^{n-1} = f_{x,y}.$$

So if $f_{x,y} > 0$, then $E_x(V_y) = \infty$, since $P_x(V_y = \infty) = f_{x,y} > 0$.
If $f_{x,y} = 0$, then $P^n(x, y) = 0$ for all $n \geq 1$, so by Lemma 1.43,

$$E_x(V_y) = \sum_{n=1}^{\infty} P^n(x, y) = \sum_{n=1}^{\infty} 0 = 0. \square$$

The theorem we just proved leads to these criteria, which can be useful in some situations to determine if a state is recurrent or transient:

**Corollary 1.45 (Recurrence criterion I)** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $x \in \mathcal{S}$. Then*

$$x \in \mathcal{S}_R \iff E_x(V_x) = \infty.$$

**Corollary 1.46 (Recurrence criterion II)** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $x \in \mathcal{S}$. Then*

$$x \in \mathcal{S}_R \iff \sum_{n=1}^{\infty} P^n(x, x) \text{ diverges.}$$

PROOF

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

**Corollary 1.47 (Recurrence criterion III)** *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 $\lim_{n \to \infty} P^n(x,y) = 0$. □

EXAMPLE 15

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.

Our next result shows that recurrence and transience are class properties:

**Theorem 1.48 (Recurrent states lead only to recurrent states)** *Suppose that* $\{X_t\}$ *is 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:

First, we prove statement (3). **Suppose not, i.e. that** $f_{y,x} < 1$**.** Then

$$1 - f_x \geq P(x, y_1)P(y_1, y_2)P(y_2, y_3) \cdots P(y_{N-1}, y_N)\,[1 - f_{y,x}] > 0$$

so $1 - f_x > 0$, so $f_x < 1$, contradicting $x \in \mathcal{S}_R$. Therefore $f_{y,x} = 1$, proving (3).

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

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

We'll prove $y$ is recurrent by showing $E_y(V_y) = \infty$:

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

$$= P^{N'}(y, x)P^N(x, y) \sum_{n=1}^{\infty} P^n(x, x)...$$

From the previous page,

$$\begin{aligned}E_y(V_y) &\geq P^{N'}(y,x)P^N(x,y)\sum_{n=1}^{\infty}P^n(x,x)\\ &= P^{N'}(y,x)P^N(x,y)E_x(V_x) \quad \text{(Formula for expected number of visits)}\\ &= \infty \quad \text{(Recurrence criterion I).}\end{aligned}$$

Therefore by recurrence criterion I, $y \in \mathcal{S}_R$, proving (1).

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

**Corollary 1.49 (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 third recurrence criterion,

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

$$\begin{aligned}\Rightarrow 0 &= \lim_{n\to\infty} \sum_{y\in\mathcal{S}} P^n(x,y)\\ \Rightarrow 0 &= \lim_{n\to\infty} 1.\end{aligned}$$

This is a contradiction! Therefore there must be at least one recurrent state. □

**Theorem 1.50 (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 $\mathcal{S}_R = \bigcup_{x \in \mathcal{S}_R} C(x)$. It is left to show the $C(x)$ are disjoint or coincide for different $x$.

To verify this, suppose $z \in C(x) \cap C(y)$.

To summarize, we have:

**Theorem 1.51 (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.*

**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

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

where the $C_j$ are recurrent communicating classes (there might be communicating classes in $\mathcal{S}_T$, but we don't care so much about those).

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.

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

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

## 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 1.52** *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 1.53** *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 16

Consider a Markov chain with transition matrix

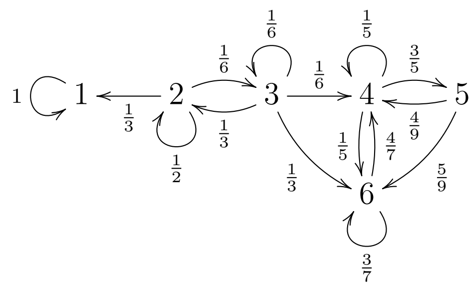
$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Determine which states of the chain are recurrent and which states are transient. For every $x \in \mathcal{S}_T$, compute $f_{x,1}$.

EXAMPLE 17

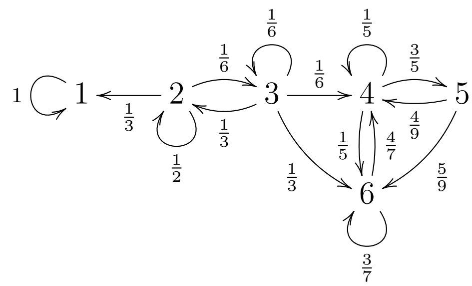
Let $\{X_t\}$ be a Markov chain with state space $\{1, 2, 3, 4, 5, 6\}$ whose transition matrix and associated directed graph are

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \\ 0 & 0 & 0 & \frac{4}{9} & 0 & \frac{5}{9} \\ 0 & 0 & 0 & \frac{4}{7} & 0 & \frac{3}{7} \end{pmatrix}$$




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

(repeated for convenience)



## 1.8 Positive and null recurrence

The crux of this section deals with big picture question (1) from earlier: when does a Markov chain have a stationary distribution?

We begin with a couple of results telling us when there is *no* stationary distribution:

> **Theorem 1.54** *Let $\pi$ be a stationary distribution of Markov chain $\{X_t\}$. If $y \in \mathcal{S}_T$, then $\pi(y) = 0$.*

PROOF By stationarity, for all $n \geq 1$,

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

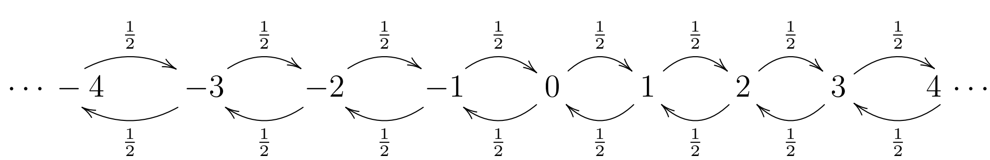
Take limits on both sides as $n \to \infty$. By the third recurrence criterion, since $y \in \mathcal{S}_T$, $\lim_{n \to \infty} P^n(x, y) = 0$, so the equation above becomes $0 = \pi(y)$. □

> **Corollary 1.55** *If an irreducible Markov chain has a stationary (or steady-state) distribution, then the chain is recurrent.*

We'd like the converse of this corollary to be true (it would be great if every irreducible, recurrent Markov chain had a stationary distribution). Unfortunately, it isn't. To see, why, consider this example, which we've seen before:

EXAMPLE OF RECURRENT MARKOV CHAIN WITH NO STATIONARY DISTRIBUTION

Let $\{X_t\}$ be simple, unbiased random walk on $\mathbb{Z}$:




Earlier, we saw that $\{X_t\}$ has no stationary distribution (because such a distribution would have to be uniform on $\mathbb{Z}$, and no such distribution exists).

Now, let's show that this chain is recurrent. Since $\{X_t\}$ is irreducible, it is sufficient to show that state $0$ is recurrent. To do this, we'll use the second recurrence criterion, and show that $\sum_{n=1}^{\infty} P^n(0, 0)$ diverges.

To show $\sum_{n=1}^{\infty} P^n(0,0)$ diverges, notice first that

$$P^n(0,0) = \begin{cases} & \text{if } n \text{ is odd} \\ & \text{if } n = 2k \text{ is even} \end{cases}$$

By a HW problem from MATH 414 (that used Stirling's Formula), $\binom{2k}{k} \approx 4^k\sqrt{\pi k}$ for large $k$. So

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

This series diverges by the ________________________________, so by the second recurrence criterion $0 \in \mathcal{S}_R$, and by irreducibility the entire chain is recurrent.

**Punchline:** Simple unbiased random walk is an example of a Markov chain which is recurrent, but has no stationary distribution.

**What's "wrong" in this example?** Simple, unbiased random walk is recurrent, meaning that every state eventually returns to itself with probability 1. But it's only "barely" recurrent, because the expected amount of time it takes to return to your initial value is infinite. The technical term for this kind of recurrence is "null recurrence".

To have a stationary distribution, not only does an irreducible Markov chain need to be recurrent (meaning every state returns to itself with probability 1), but the expected amount of time it takes to return to each state must be finite. The term for this is "positive recurrence", and this is the last ingredient in the FTMC.

## Detour: a new type of convergence

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

EXAMPLES

- $\frac{1}{n} \to 0$.
- $\frac{n+1}{n-1} \to 1$.
- 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 1.56** *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*

$$b_n = \frac{1}{n}\sum_{k=1}^{n} a_k$$

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

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

*We write $a_n \stackrel{Ces}{\longrightarrow} L$ to represent this.*

EXAMPLE 18

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

EXAMPLE 19

The Strong Law of Large Numbers (MATH 414) says

**Facts about Cesàro convergence:**

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

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

"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, and with positive and null recurrence).

**Recall:** $\sum_{k=1}^{n} P^k(x,y) =$

Therefore, the Cesàro averages of the sequence $\{P^n(x,y)\}$ are actually

$$\frac{1}{n}\sum_{k=1}^{n} P^k(x,y) =$$

## Positive and null recurrence

**Definition 1.57** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$ and transition function $P$.*

1. *Given $y \in \mathcal{S}_R$, define $m_y = E_y(T_y)$. $m_y$ is a number (possibly $\infty$) called the* **mean return time** *to $y$.*

2. *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**.

3. *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**.

**Note:** The mean return times of a transient state is trivially $\infty$, because if $y \in \mathcal{S}_T$,

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

**Theorem 1.58** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $y \in \mathcal{S}$.*

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

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

*(These limits hold with probability 1.)*

PROOF Statement (2) is obvious. To prove (1), assume WLOG that you start in state $y$ (since by hypothesis you must hit $y$ at some point). Define the following random variables:

- $T_y^r = \min\{n \geq 1 : V_{y,n} = r\} =$ 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 $m_y$. So by the Strong Law of Large Numbers, $W_y^j \overset{Ces}{\longrightarrow} m_y$ with probability 1. This means

$$
\begin{aligned}
&P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}W_y^j = m_y\right) = 1\\
&\Rightarrow P\left(\lim_{n\to\infty}\frac{T_y^n}{n} = m_y\right) = 1 \quad (*)
\end{aligned}
$$

**Theorem 1.59** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. Let $x, y \in \mathcal{S}$.*

1. $\lim_{n\to\infty} \frac{E_x(V_{y,n})}{n} = \frac{f_{x,y}}{m_y}$.
2. $P^n(x,y) \xrightarrow{Ces} \frac{f_{x,y}}{m_y}$.

*(These limits hold with probability 1.)*

PROOF From the previous discussion, (1) implies (2), so it is sufficient to prove (1). To do this, note

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

**Corollary 1.60** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$.*

1. *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 $\lim_{n\to\infty} \frac{V_{y,n}}{n} = \frac{1}{m_y}\ \forall\, y \in C$.*

2. *If $y \in \mathcal{S}_T \cup \mathcal{S}_N$, then for all $x \in \mathcal{S}$, $P^n(x,y) \xrightarrow{Ces} 0$.*
3. *If $y \in \mathcal{S}_P$, then $P^n(y,y) \xrightarrow{Ces} \frac{1}{m_y}$.*

PROOF (1) follows immediately from Theorem 1.59.

For (2), notice that if $y \in \mathcal{S}_T \cup \mathcal{S}_N$, $m_y = \infty$ so $P^n(x,y) \xrightarrow{Ces} \frac{f_{x,y}}{m_y} = \frac{f_{x,y}}{\infty} = 0$.

For (3), since $y$ is recurrent, $f_y = f_{y,y} = 1$ so $P^n(y,y) \xrightarrow{Ces} \frac{f_{y,y}}{m_y} = \frac{1}{m_y}$. □

**Note:** Corollary 1.60 provides a new distinction between positive recurrent and null recurrent states. If $y \in \mathcal{S}$ is null recurrent (or transient), then $P^n(y,y) \xrightarrow{Ces} 0$ but if $y \in \mathcal{S}$ is positive recurrent, then $P^n(y,y) \xrightarrow{Ces} \frac{1}{m_y} > 0$.

**Theorem 1.61 (Positive recurrent states lead only to positive recurrent states)** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $x \in \mathcal{S}_P$ and $x \to y$, then $y \in \mathcal{S}_P$.*

PROOF $x$ is recurrent, so 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

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

**Corollary 1.62 (Null recurrent states lead only to null recurrent states)** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$. If $x \in \mathcal{S}_N$ and $x \to y$, then $y \in \mathcal{S}_N$.*

PROOF $x$ is recurrent, so by previous result, $y$ is recurrent and $y \to x$. If $y$ is positive recurrent, then by the above theorem $x$ is positive recurrent, a contradiction. Thus $y$ must be null recurrent. □

**Corollary 1.63** *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 1.64** *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 For every $x \in C$ and $k \in \{1, 2, 3, ...\}$, we have $\sum_{y \in C} P^k(x, y) = 1$. So

Therefore there must be some $y \in C$ such that $m_y < \infty$, i.e. $y \in \mathcal{S}_P$. Since positive recurrence is a class property, every $x \in C$ is positive recurrent. □

**Corollary 1.65** *Any irreducible Markov chain with a finite state space is positive recurrent.*

## Existence and uniqueness of stationary distributions

We begin by showing that for an irreducible Markov chain, values of any of its stationary distributions are determined by mean return times:

**Theorem 1.66** *Let $\{X_t\}$ be a Markov chain with state space $\mathcal{S}$.*

1. *If $x \in \mathcal{S}$ is either transient or null recurrent, then for any stationary distribution $\pi$, $\pi(x) = 0$.*
2. *If $\{X_t\}$ is irreducible, then for any stationary distribution $\pi$ of $\{X_t\}$, $\pi(x) = \frac{1}{m_x}$.*

PROOF Suppose $\pi$ is stationary. Then, for all $k \in \{1, 2, 3, ...\}$ and all $z \in \mathcal{S}$, we have

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

**Corollary 1.67 (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 $\pi$ 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 $\pi$ would not be a distribution. □

**Theorem 1.68 (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 $\pi$ defined by $\pi(x) = \frac{1}{m_x}$ for all $x \in \mathcal{S}$.*

PROOF What's left to show is that for an irreducible, positive recurrent Markov chain $\{X_t\}$, the formula $\pi(x) = \frac{1}{m_x}$ defines a stationary distribution.

**Case 1:** $\mathcal{S}$ is finite. In this situation,

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

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

so $\pi(x) = \frac{1}{m_x}$ defines a distribution.

We're still in case 1 ($\mathcal{S}$ is finite); what's left is to show that the distribution defined by $\pi(x) = \frac{1}{m_x}$ is in fact stationary (we have to verify that $\sum_{x \in \mathcal{S}} \pi(x)P(x,y) = \pi(y)$):

$$\begin{aligned} P^{k+1}(z,y) &= \sum_{x \in \mathcal{S}} P^m(z,x)P(x,y) \\ \frac{1}{n}\sum_{k=1}^{n} P^{k+1}(z,y) &= \frac{1}{n}\sum_{k=1}^{n}\sum_{z \in \mathcal{S}} P^m(z,x)P(x,y) \end{aligned}$$

**Case 2:** $\mathcal{S}$ is infinite. In this situation, 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

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

Since $\mathcal{S}'$ is arbitrary, it must be that, setting $\pi(x) = \frac{1}{m_x}$ for every $x$, we have

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

If $\sum_{x \in \mathcal{S}} \pi(x)P(x,y) < \pi(y)$, then

$$\begin{aligned} 1 = \sum_{y \in \mathcal{S}} \pi(y) > \sum_{y \in \mathcal{S}}\sum_{x \in \mathcal{S}} \pi(x)P(x,y) &= \sum_{x \in \mathcal{S}}\sum_{y \in \mathcal{S}} \pi(x)P(x,y) \\ &= \sum_{x \in \mathcal{S}} \pi(x) \sum_{y \in \mathcal{S}} P(x,y) \\ &= \sum_{x \in \mathcal{S}} \pi(x) \cdot 1 = 1 \end{aligned}$$

which is a contradiction. Therefore, $\sum_{x \in \mathcal{S}} \pi(x)P(x,y) = \pi(y)$ so some multiple of $\pi$, say $M\pi$, is stationary, but by Theorem 1.66 $M$ must be $1$, so $\pi$ is stationary. □

**Corollary 1.69** *Any irreducible Markov chain on a finite state space has a unique stationary distribution.*

**Theorem 1.70 (Ergodic Theorem for Markov chains)** *Let $\{X_t\}$ be an irreducible, positive recurrent Markov chain with state space $\mathcal{S}$ and let $\pi$ be its unique stationary distribution. Then for all $x \in \mathcal{S}$,*

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

PROOF We've seen that $\pi(x) = \frac{1}{m_x}$; the result follows from Theorem 1.59. □

**A picture to explain the ergodic theorem:**

EXAMPLE 20

Suppose $\{X_t\}$ is a Markov chain with $\mathcal{S} = \{1,2,3,4\}$ whose stationary distribution is $\left(\frac{1}{9}, \frac{2}{9}, \frac{4}{9}, \frac{2}{9}\right)$. Suppose $X_0 = 1$. Estimate the number of times $t$ in the interval $[1, 900]$ such that $X_t = 2$. (P.S. What is $m_2$?)

## Stationary distributions for non-irreducible Markov chains

**Definition 1.71** *A distribution* $\pi$ *on* $\mathcal{S}$ *is* **supported** *or* **concentrated** *on a subset* $C \subseteq \mathcal{S}$ *if* $\pi(x) = 0$ *for all* $x \notin C$.

EXAMPLE 21

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

To summarize:

**Existence and 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 $\pi$ 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 that class (call it $\pi_C$) 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 $\pi_C$ are also stationary, so $\{X_t\}$ has infinitely many stationary distributions. (All stationary distributions are convex combinations of these $\pi_C$.)

EXAMPLE 22

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 23

Let $\{X_t\}$ be the Markov chain with state space $\mathcal{S} = \{0, 1, 2, 3, ...\}$ and transition function $P$ defined by

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

Show that $\{X_t\}$ is positive recurrent, and find $\pi(0)$, $\pi(1)$, $\pi(2)$ and $\pi(3)$ for the stationary distribution $\pi$ of $\{X_t\}$.

## 1.9 Proving the Fundamental Theorem

**Theorem 1.72 (FTMC)** *Let $\{X_t\}$ be an irreducible, aperiodic, pos. recurrent Markov chain. Then the unique stationary distribution of this chain, defined by $\pi(x) = \frac{1}{m_x}$ is steady-state, meaning*

$$\lim_{n\to\infty} \pi_n(x) = \pi(x)$$

*for all $x \in \mathcal{S}$, no matter the initial distribution $\pi_0$.*

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 $\pi$.

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 $T = \infty$). $T$ is called the **coupling time** in this argument.

**Claim:** $P(T < \infty) = 1$.
**Proof of Claim:** HW (this requires aperiodicity of $\{X_t\}$, because it uses Theorem 1.40 which says that for any two states, it is possible to get from one to the other in all times greater than or equal to some $N$).
*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 $\pi_0$ 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 $\pi$ is steady-state. □

## What if the Markov chain is periodic?

**Theorem 1.73** *Let $\{X_t\}$ be an irreducible, positive recurrent Markov chain with state space $\mathcal{S}$, whose period is $d \geq 2$. Let $\pi$ denote its unique stationary distribution. Then:*

1. *$P^n(x,y) = 0$ unless $n = md + r$ for some $m \in \mathbb{N}$ (i.e. unless $n \equiv r \mod d$)*
2. *$\lim_{m\to\infty} P^{md+r}(x,y) = d \cdot \pi(y)$.*

PROOF 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 the FTMC; 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

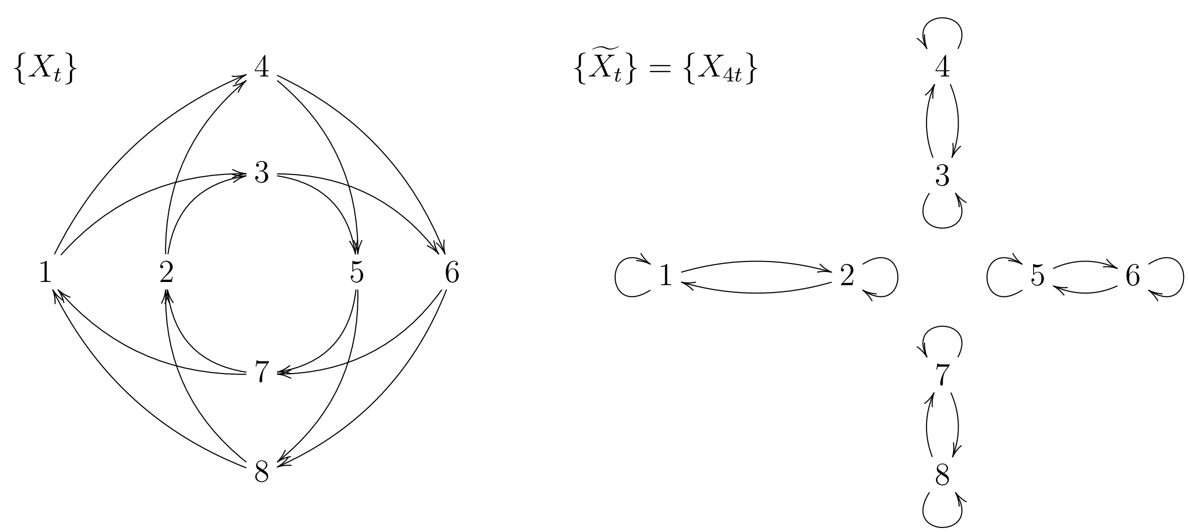
$$\lim_{m\to\infty} P^{md}(z,x) = d\pi(x).$$

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. □

## A picture to explain the periodic case




So, for instance, $P^n(3,2)$ looks like

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

## 1.10 Example computations

DIRECTIONS

For each given Markov chain in Examples 24-27:

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) and determine which (if any) of these distributions are steady-state. (If you can't compute the entire stationary distribution, find as many values of the stationary distribution as you can.)
5. Find the mean return time to state $2$.

EXAMPLE 24

The Ehrenfest chain with $d = 4$.

EXAMPLE 25

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 26

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 27

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

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

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

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

**Alternate solution:**

*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 |
|---|---|---|---|---|---|
| $P(X_4 = x)$ | $\frac{1}{16}$ | $\frac{4}{16}$ | $\frac{6}{16}$ | $\frac{4}{16}$ | $\frac{1}{16}$ |

Notice that your expected bankroll is

$$\begin{aligned} EX_4 &= \frac{1}{16}(6) + \frac{4}{16}(8) + \frac{6}{16}(10) + \frac{4}{16}(12) + \frac{1}{16}(14) \\ &= \frac{6 + 32 + 60 + 48 + 14}{16} \\ &= \frac{160}{16} \\ &= 10. \end{aligned}$$

Notice that the expected amount you have after $4$ rolls is the amount you started with:

$$EX_4 = X_0.$$

## 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).

## The idea of a martingale

Let's return to the setup of the previous section, where you were wagering $\$1$ on each flip of a fair coin. We saw that in this setting, $E[X_4] = X_0$.

What happens if we condition on some additional information? For example, suppose that the first flip is heads (so that you win your first bet, so that $X_1 = 11$). Given this, what is $E[X_4]$? In other words, what is $E[X_4 \mid X_1 = 11]$?

Repeating the argument from the previous section, 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

$$E[X_t \,|\, X_0] = X_0.$$

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 $EX_t = EX_s$. Therefore, if we know the value of $X_s$, we obtain

$$E[X_t \,|\, X_s] = X_s.$$

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.

## 2.2 Filtrations

### $\sigma$-algebras

**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 $\Omega$ be a set. A nonempty collection $\mathcal{F}$ of subsets of $\Omega$ 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 $\bigcup\limits_j A_j$ and $\bigcap\limits_j A_j$ belong to $\mathcal{F}$ as well.*

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

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

EXAMPLES OF $\sigma$-ALGEBRAS

1. Let $\Omega$ be any set. Let $\mathcal{F} = \{\emptyset, \Omega\}$. This is called the **trivial $\sigma$-algebra** of $\Omega$.
2. Let $\Omega$ be any set. Let $\mathcal{F} = 2^\Omega$ be the set of all subsets of $\Omega$. This is called the **power set** of $\Omega$.
3. Let $\Omega$ be any set and let $\mathcal{P} = \{P_1, P_2, ..., P_n\}$ be any partition of $\Omega$ (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 **$\sigma$-algebra generated by** $\mathcal{P}$.

4. Let $\Omega = [0,1] \times [0,1]$.
   - Let $\mathcal{F}_1$ be the trivial $\sigma$-algebra of $\Omega$.
   - Let $\mathcal{F}_2$ be the collection of all subsets of $\Omega$ of the form $A \times [0,1]$ where $A \subset [0,1]$.
   - Let $\mathcal{F}_3$ be the power set of $\Omega$.

Suppose $\omega = (x, y) \in \Omega$.

1. If you know all the sets in $\mathcal{F}_1$ to which $\omega$ belongs, what do you know about $\omega$?

2. If you know all the sets in $\mathcal{F}_1$ to which $\omega$ belongs, what do you know about $\omega$?

3. If you know all the sets in $\mathcal{F}_1$ to which $\omega$ belongs, what do you know about $\omega$?

## Measurability

**Definition 2.3** *Let $\Omega$ be a set and let $\mathcal{F}$ be a $\sigma$-algebra on $\Omega$. A subset $E$ of $\Omega$ 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 $\sigma$-algebra generated by the partition

$$\mathcal{P} = \{[0, 1/3), [1/3, 1/2), [1/2, 1]\}.$$

Determine whether each of these functions $X$ is $\mathcal{F}$-measurable:

1. $X : \Omega \to \mathbb{R}$ defined by $X(\omega) = 2\omega$.

2. $X : \Omega \to \mathbb{R}$ defined by $X(\omega) = 2$.

3. $X : \Omega \to \mathbb{R}$ defined by $X(\omega) = \begin{cases} 1 & \text{if } x < \frac{1}{2} \\ 0 & \text{else} \end{cases}$.

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 $\omega$ but only on which partition element $\omega$ belongs to.

This idea illustrates the point of measurability in general: think of a $\sigma$-algebra $\mathcal{F}$ as revealing some partial information about an $\omega$ (i.e. it tells you which sets in $\mathcal{F}$ to which $\omega$ belongs, but not necessarily exactly what $\omega$ 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 $\Omega = [0,1] \times [0,1]$, so $\omega \in \Omega \leftrightarrow \omega = (x, y)$.

| $\sigma$-algebra $\mathcal{F}$ | information $\mathcal{F}$ reveals about $\omega$ | description of $\mathcal{F}$-measurable functions |
|---|---|---|
| trivial $\sigma$-algebra $\mathcal{F} = \{\emptyset, \Omega\}$ | nothing | |
| $\mathcal{F}_x =$ sets of form $A \times [0,1]$ | the $x$-coordinate of $\omega$ | |
| $\mathcal{F}_y =$ sets of form $[0,1] \times A$ | | |
| power set $\mathcal{F} = 2^\Omega$ | everything ($x$ and $y$) | |

## Filtrations

**Definition 2.4** *Let $\Omega$ be a set and let $\mathcal{I} \subseteq [0, \infty)$. A* **filtration** $\{\mathcal{F}_t\}_{t \in \mathcal{I}}$ *on $\Omega$ is a sequence of $\sigma$-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 $\Omega$.

**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 $\Omega$ as a list which records the outcome of every bet you make. $\mathcal{F}_t$ is the $\sigma$-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 $t \in \mathcal{I} = \{0, 1, 2\}$). Describe the filtration $\{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\}$.

## 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 OF STRATEGIES

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 $\{X_t\}_{t\in\mathcal{I}}$ 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 $0 \le t \le 8$.

| time $t$ | bet size $B_t$ | "W-L" record $X_t$ | result of the $t^{th}$ game $S_t$ | bankroll using strategy $\{B_t\}$ $(B\cdot X)_t$ |
|---|---|---|---|---|
| 0 | DNE | 50 | DNE | 50 |
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

## Stopping times

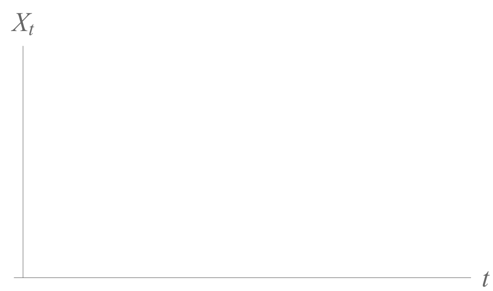
**Definition 2.8** *Let $\{X_t\}_{t \in \mathcal{I}}$ be a stochastic process with standard filtration $\{\mathcal{F}_t\}$. A r.v. $T : \Omega \to \mathbb{R} \cup \{\infty\}$ is called a* **stopping time (for $\{X_t\}$)** *if for every $a \in \mathbb{R}$, the set of sample functions satisfying $T \leq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

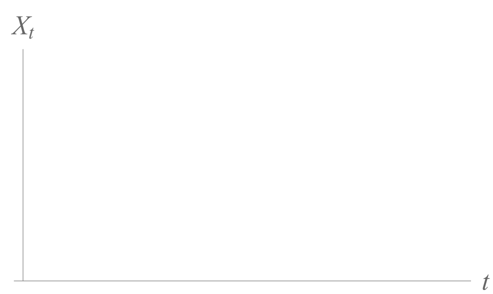
In the context of playing a game over and over, think of $T$ as a "trigger" which causes you to stop playing the game. Thus you would walk away from the table with winnings given by $X_T$ (or, if you are employing strategy $\{B_t\}$, your winnings would be $(B \cdot X)_T$).

EXAMPLES

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

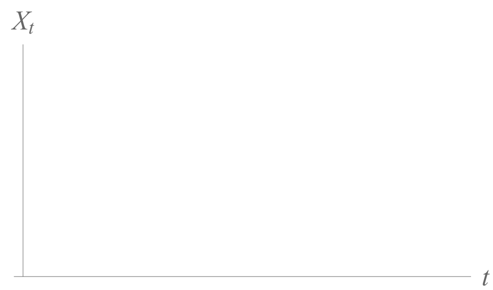



- $T = \min\{t > 0 : X_t = X_0\}$



NON-EXAMPLE

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




**<u>Recall our big picture question</u>:** is there a strategy under which you can beat a fair game?

**<u>Restated in mathematical terms</u>:** 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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $E[X] = E[E[X|\mathcal{G}]]$. 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 1

Let $\Omega = \{A, B, C, D\}$; let $\mathcal{F} = 2^{\Omega}$; let $P$ be the uniform distribution on $\Omega$. Let $\mathcal{G}$ be the $\sigma$-algebra generated by $\mathcal{P} = \{\{A, B\}, \{C, D\}\}$. Let $X : \Omega \to \mathbb{R}$ be defined by $X(A) = 2$; $X(B) = 6$; $X(C) = 3$; $X(D) = 1$. Compute $E[X|\mathcal{G}]$.

EXAMPLE 2

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 $X(A) = X(B) = X(D) = 2$; $X(C) = 0$; $X(E) = 1$. Compute $E[X|\mathcal{G}]$.

EXAMPLE 3

Let $\Omega = [0, 1] \times [0, 1]$; let $\mathcal{F} = 2^{\Omega}$; let $P$ be the uniform distribution. Let $\mathcal{G}$ be the $\sigma$-algebra of vertical sets (i.e. sets of the form $A \times [0, 1]$). Let $X : \Omega \to \mathbb{R}$ be $X(x, y) = x + y$. Compute $E[X|\mathcal{G}]$.

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. (we write $X \perp \mathcal{G}$ to represent this), 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. Law of total expectation: $E[E(X|\mathcal{G})] = EX$.
7. Constants: $E[a|\mathcal{G}] = a$.

*(These statements hold with probability one.)*

## 2.4 Martingales and optional stopping

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 We use the tower property of conditional expectation. Let $s \le t$. Then

$$\begin{aligned}
E[X_t|\mathcal{F}_s] &= E[E[\cdots E[E[E[X_t|\mathcal{F}_{t-1}]|\mathcal{F}_{t-2}]|\mathcal{F}_{t-3}]\cdots|\mathcal{F}_{s+1}]|\mathcal{F}_s] \\
&= E[E[\cdots E[E[X_{t-1}|\mathcal{F}_{t-2}]|\mathcal{F}_{t-3}]\cdots|\mathcal{F}_{s+1}]|\mathcal{F}_s] \\
&= E[E[\cdots E[X_{t-2}|\mathcal{F}_{t-3}]\cdots|\mathcal{F}_{s+1}]|\mathcal{F}_s] \\
&= \ldots \\
&= E[X_{s+1}|\mathcal{F}_s] \\
&= X_s.
\end{aligned}$$

By definition, $\{X_t\}$ is a martingale. □

**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):

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

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

Next, statement (3):

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

(4): By the law of total expectation and part (3), $E[S_t] = E[E[S_t|\mathcal{F}_{t-1}]] = E[0] = 0$.

(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 2.13, $\{(B\cdot X)_t\}$ 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

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

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 on $\mathbb{Z}$

**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 $S_1, S_2, S_3, \ldots$ 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:

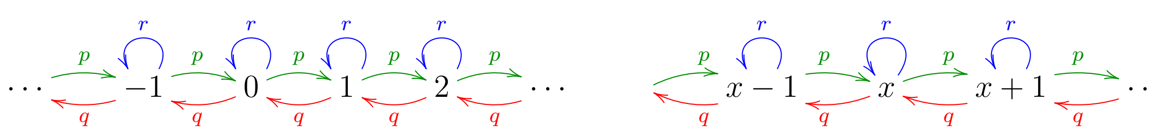
**Definition 2.22** *A random walk on $\mathbb{Z}$ is called* **simple** *if the steps $S_j$ take values only in $\{-1, 0, 1\}$. For a simple random walk, we define*

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

*For a simple random walk, we let*

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

A simple random walk models a repeated game where you bet \$1 on each play; simple random walk is a Markov chain which has the following directed graph:




**Lemma 2.23** *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

**Definition 2.24** *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 $q > 0$.

**Theorem 2.25** *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. □

As a special case of this, 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 compute escape probabilities, 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 $\theta$ 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 $Y_t = X_t - t\mu$. Since the index set is discrete, to show $\{Y_t\}$ is a martingale, we need to show $E[Y_{t+1}|\mathcal{F}_t] = Y_t$. Then

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

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

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

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

From the previous page,

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

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

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

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

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

**Idea of the proof:** To prove formulas like these, you follow this procedure:

1. Cleverly choose some martingale associated to the process $\{X_t\}$. Let's call that martingale $\{M_t\}$.
2. Let $T = \min(T_a, T_b) = \min\{t : X_t \in \{a, b\}\}$. $T$ is a finite stopping time, and
$$X_T = \begin{cases} b & \text{if } T_b < T_a \\ a & \text{if } T_a < T_b \end{cases}.$$
Work out $E[M_T]$ based on this information. You should get an expression which contains the probability you want to compute.
3. By the (tweaked version of the) OST $E[M_T] = E[M_0]$. Set your formula from step (3) equal to the value of $E[M_0]$ (which is usually known), and solve for the probability you want.

Now for the details.

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

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 4

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 4, CONTINUED

Recall that 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?

*Solution:* 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

**Follow-up 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 (this follows the same pattern as all the proofs we have been doing, but with a different choice of martingale).

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*

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

**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

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

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

$$= \begin{cases} 1 & \text{if } p = q \\ \frac{1-0}{1-0} & \text{if } p > q \\ \lim_{n\to-\infty} \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^n}{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^n} & \text{if } p < q \end{cases}$$

$$= \begin{cases} 1 & \text{if } p \geq q \\ \lim_{n\to-\infty} \frac{\left(\frac{q}{p}\right)^x - 0}{\left(\frac{q}{p}\right)^a - 0} & \text{if } p < q \end{cases}$$

$$= \begin{cases} 1 & \text{if } p \geq q \\ \left(\frac{q}{p}\right)^{x-a} & \text{if } p < q \end{cases}$$

$$= \begin{cases} 1 & \text{if } p \geq q \\ \left(\frac{p}{q}\right)^{a-x} & \text{if } p < q \end{cases}$$

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

**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, i.e. if and only if $f_0 = 1$. By direct calculation,

$$\begin{aligned} f_0 = P_0(T_0 < \infty) &= P_0(T_0 < \infty \mid X_1 = -1) P_0(X_1 = -1) \\ &\quad + P_0(T_0 < \infty \mid X_1 = 0) P_0(X_1 = 0) \qquad \text{(Law of Total Prob.)} \\ &\quad + P_0(T_0 < \infty \mid X_1 = 1) P_0(X_1 = 1) \\ &= P_1(T_0 < \infty) q + 1 \cdot r + P_1(T_0 < \infty) p \\ &= \begin{cases} 1 \cdot q & +r & + \left(\frac{q}{p}\right) p & \text{if } p > q \\ 1 \cdot q & +r & +1 \cdot p & \text{if } p = q \\ \left(\frac{p}{q}\right) q & +r & +1 \cdot p & \text{if } p < q \end{cases} \qquad \text{(Gambler's Ruin)} \\ &= \begin{cases} 2q + r & \text{if } p > q \\ 1 & \text{if } p = q \\ 2p + r & \text{if } p < q \end{cases} \end{aligned}$$

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

## 2.6 Birth and death chains

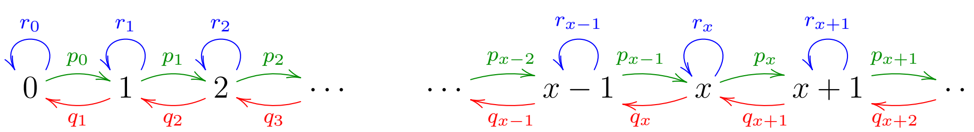
If we start with a random walk, but relax the requirement that the "$p$"s and "$q$"s of the chain are the same at every state, then we get a class of Markov chains called "birth-death" chains:

**Definition 2.34** *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*

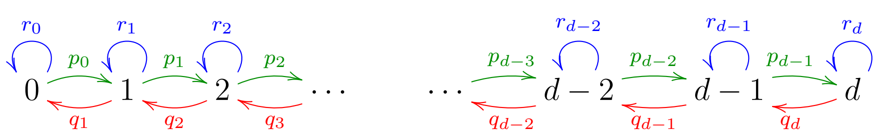
1. *For all $x \in \mathcal{S}$, $p_x + q_x + r_x = 1$;*
2. *$q_0 = 0$;*
3. *If $\mathcal{S} = \{0, 1, ..., d\}$, then $p_d = 0$; and*
4. *For all $x \in \mathcal{S}$,* $\begin{cases} P(x, x+1) = p_x \\ P(x, x) = r_x \\ P(x, x-1) = q_x \end{cases}$

**Examples of birth-death chains include:** gambler's ruin, Ehrenfest chain.

Every birth-death chain has a directed graph that looks like




if $\mathcal{S} = \{0, 1, 2, ...\}$, or




if $\mathcal{S} = \{0, 1, ..., d\}$.

**First observation:** 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

**Big picture 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.

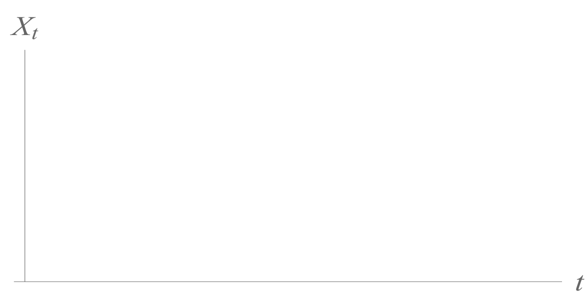
**Refined question:** Under what circumstances is an irreducible birth-death chain with $\mathcal{S} = \{0, 1, 2, 3, ...\}$ recurrent? When is such a chain transient?

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 2.35** *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$) = 0$. □

**Important intermediate 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 an unbiased random walk, $\{X_t\}$ is a martingale, and for biased random walk, the process $\left\{\left(\frac{q}{p}\right)^{X_t}\right\}$ is a martingale). You can do something similar for birth-death chains, but you need a more complicated martingale:

**Lemma 2.36** *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*

$$\begin{aligned}\widetilde{\gamma}(y) &= 1 + \frac{q_1}{p_1} + \frac{q_2 q_1}{p_2 p_1} + \ldots + \frac{q_{y-1} q_{y-2} \cdots q_2 q_1}{p_{y-1} p_{y-2} \cdots p_2 p_1} \\ &= \gamma_0 + \gamma_1 + \gamma_2 + \ldots + \gamma_{y-1} \\ &= \sum_{j=0}^{y-1} \gamma_j.\end{aligned}$$

*Then the stochastic process $\{Y_t\}$ is a martingale, where $Y_t = \widetilde{\gamma}(X_t)$.*

PROOF HW

**Theorem 2.37 (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 the $\gamma_y$ are as defined in Lemma 2.36.*

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 2.38** *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 2.39 (Recurrence/transience of birth-death chains)** *Let $\{X_t\}$ be an irreducible birth-death chain with $\mathcal{S} = \{0, 1, 2, ...\}$. Then defining $\gamma_y$ 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

$$\begin{aligned}
f_{1,0} &= P_1(T_0 < \infty) \\
&= P_1\left(\bigcup_{n=2}^{\infty} (T_0 < T_n)\right) \\
&= \lim_{n\to\infty} P_1(T_0 < T_n) \quad \text{by monotonocity (chapter 1 of Math 414)} \\
&= \lim_{n\to\infty} \left( \frac{\sum_{y=1}^{n-1} \gamma_y}{\sum_{y=0}^{n-1} \gamma_y} \right) \quad \text{by Theorem 2.37 with } x = 1, a = 0, b = n \\
&= \lim_{n\to\infty} \left( \frac{\sum_{y=0}^{n-1} \gamma_y - \gamma_0}{\sum_{y=0}^{n-1} \gamma_y} \right) \\
&= \lim_{n\to\infty} \left( \frac{\sum_{y=0}^{n-1} \gamma_y - 1}{\sum_{y=0}^{n-1} \gamma_y} \right) \\
&= \lim_{n\to\infty} \left( 1 - \frac{1}{\sum_{y=0}^{n-1} \gamma_y} \right) \\
&= 1 - \frac{1}{\sum_{y=0}^{\infty} \gamma_y} \\
&= \begin{cases} & \text{if } \sum_{y=0}^{\infty} \gamma_y \text{ diverges} \\ & \text{if } \sum_{y=0}^{\infty} \gamma_y \text{ converges to } C \end{cases}
\end{aligned}$$

By the preceding lemma, $\{X_t\}$ is recurrent if and only if $f_{1,0} = 1$, so this proves the theorem. □

EXAMPLE 5

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).

$$\pi \text{ stationary} \Rightarrow \sum_{x=0}^{d} \pi(x)P(x,y) = \pi(y) \text{ and } \sum_{y\in\mathcal{S}} \pi(y) = 1$$
$$\Rightarrow \begin{cases} \pi(0)r_0 + \pi(1)q_1 = \pi(0) & (y=0) \\ \pi(y-1)p_{y-1} + \pi(y)r_y + \pi(y+1)q_{y+1} = \pi(y) & (y>0) \\ \sum_{y=0}^{d} \pi(y) = 1 \end{cases}$$

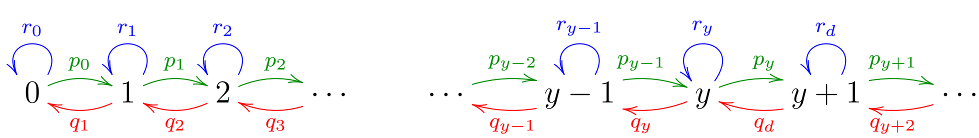
Since $p_y + q_y = 1 - r_y$ for all $y$, these equations yield (after some significant algebra)

$$\pi(y+1) = \frac{p_y}{q_{y+1}}\pi(y)\,\forall\, y \geq 0$$
$$\Rightarrow \pi(y) = \frac{p_0 p_1 p_2 \cdots p_{y-1}}{q_1 q_2 \cdots q_y}\pi(0)\,\forall\, y \geq 1$$

Define

$$\zeta_y = \begin{cases} \frac{p_0 p_1 \cdots p_{y-1}}{q_1 q_2 \cdots q_y} & \text{if } y > 0 \\ 1 & \text{if } y = 0 \end{cases};$$

think of $\zeta_y$ as "the product of all the $p$s to the left of $y$ over the product of all the $q$s to the left of $y$ in the directed graph:




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 2.40 (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 $\pi$ 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 6

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.

*Solution:* First, compute the $\zeta_y$: $\zeta_0 = 1$ and for $y \geq 1$,

$$\zeta_y = \frac{p_0 p_1 \cdots p_{y-1}}{q_1 q_2 \cdots q_y} =$$

Then apply Theorem 2.40:

$$\sum_{y \in \mathcal{S}} \zeta_y = \sum_{y=0}^{\infty} \zeta_y =$$

So the stationary distribution $\pi$ satisfies

$$\pi(x) = \frac{\zeta_x}{\sum_y \zeta_y} = \frac{\zeta_x}{2e} = \frac{x+1}{2ex!}.$$

## 2.7 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 at each step, you move one unit in one of the coordinate directions, choosing the direction you move in uniformly.

These random walks are all irreducible and have period 2.

EXAMPLE: SIMPLE RANDOM WALK ON $\mathbb{Z}$

When $d = 1$, this is a description of simple, unbiased random walk on $\mathbb{Z}$ with $p = q = \frac{1}{2}$. This Markov chain is

EXAMPLE: DRUNKARD'S WALK (RANDOM WALK ON $\mathbb{Z}^2$)

EXAMPLE: DRUNK BIRD'S FLIGHT (RANDOM WALK ON $\mathbb{Z}^3$)

**Main question:** Will the drunk person ever make it home? Will they make it back to the bar? What about the inebriated bird? In other words, **is unbiased random walk on $\mathbb{Z}^d$ recurrent or transient**?

**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 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

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

Therefore (P.S. you may have had to do this computation in an activity)

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

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 (and you have to or had to do this in an activity), 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 2.41 (Pólya'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 (null) recurrent.*
2. *If $d > 2$, then $\{X_t\}$ is transient.*

Mathematician Shizuo Kakutani famously described this theorem by saying "A drunk man will find his way home, but a drunk bird may be lost forever."

*Chapter 3*

# Continuous-time Markov chains

## 3.1 Motivation

Our goal in this chapter is to study analogues of Markov chains (including random walk) where time is measured continuously rather than discretely. (The state space will still be finite or countable.)

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

| | (DISCRETE-TIME) MARKOV CHAIN | CTMC (CONTINUOUS-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 $t$ <br> $t \in \{0, 1, 2, ...\}$ or $t \in \mathbb{Z}$ | $X_t =$ state at time $t$ <br> $t \in [0, \infty)$ or $t \in \mathbb{R}$ |
| initial distribution | $\pi_0 : \mathcal{S} \to [0, 1]$; $\sum_{x \in \mathcal{S}} \pi_0(x) = 1$ <br> $\pi_0(x) = P(X_0 = x)$ | $\pi_0 : \mathcal{S} \to [0, 1]$; $\sum_{x \in \mathcal{S}} \pi_0(x) = 1$ <br> $\pi_0(x) = P(X_0 = x)$ <br> (same) |

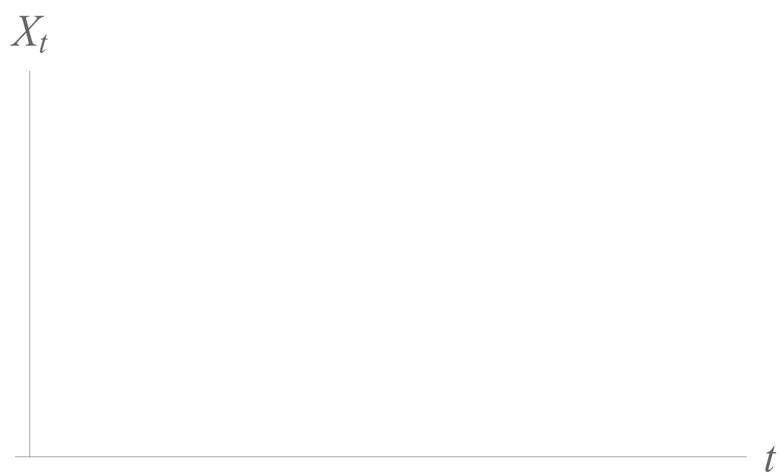
(continued on next page)

| | MARKOV CHAIN | CTMC |
| --- | --- | --- |
| transition probabilities | we specify time 1 transitions: $P : \mathcal{S} \times \mathcal{S} \to [0,1]$ $\sum_{y \in \mathcal{S}} P(x,y) = 1 \, \forall \, x \in \mathcal{S}$ $P(x,y) = P(X_{t+1} = y \| X_t = x)$ (we assume these are $\perp$ of $t$)<br><br>If $\mathcal{S}$ is finite, write $P$ as a matrix: $P(x,y) \leftrightarrow P_{x,y} = P_{xy}$<br><br>From the time 1 transitions, we calculate transition probabilities for any time $n$: $P^n(x,y) = P(X_{t+n} = y \| X_t = x)$ $= \sum_{z \in \mathcal{S}} P(x,z) P^{n-1}(z,y)$<br><br>If $\mathcal{S}$ finite, $P^n(x,y) = (P^n)_{xy}$. | |
| Markov property | $P(X_t = x_t \| X_0 = x_0, ..., X_{t-1} = x_{t-1})$ $= P(X_t = x_t \| X_{t-1} = x_{t-1})$ $= P(x_{t-1}, x_t)$ | |

**Definition 3.1** *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 $J_1 < J_2 < J_3 < \ldots$ (these are r.v.s, not constants) and states $x_0, x_1, x_2, \ldots \in \mathcal{S}$ 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 (we'll see one of these reasons later).

**Definition 3.2** *A* **continuous-time Markov chain (CTMC)** $\{X_t\}$ *is a jump process satisfying the Markov property .*

## 3.2 More matrix theory: CTMCs with finite state space

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

**Definition 3.3** *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 3.4 (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 (1) and (2) are obvious; (3) is essentially the Law of Total Probability (see the chart two pages ago). □

**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 function (or functions) can $f$ be?

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

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

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 the hypothesis that $f(0) = 1$. Thus $f(t) > 0$ for all $t$.

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

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

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 3.5** *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

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

(This is not a problem if you remember your MATH 322 and/or MATH 330.)

## Exponentiation of matrices

**Recall** the Taylor series of $e^t$:

$$e^t = \sum_{n=0}^{\infty} \frac{t^n}{n!} = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots$$

This series converges for all $t$ (by the Ratio Test), so we can use this formula as a definition of the function $e^t$. In fact, $e^t$ can also be obtained in a second, useful way:

**Lemma 3.6** *For any $t \in \mathbb{R}$,*

$$e^t = \lim_{n\to\infty} \left(1 + \frac{t}{n}\right)^n.$$

PROOF HW (use L'Hôpital's Rule on the limit).

**Definition 3.7** *Given a square matrix $A$, define the* **matrix exponential** *of $A$ to be the matrix $e^A$ (also denoted $\exp(A)$) defined by*

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

There's an issue here with what it means for an infinite series of matrices to converge. Take my word for it: this series converges for all square matrices $A$, to a matrix $e^A = \exp(A)$ which is the same size as $A$.

Like the function $e^t$, we can also define matrix exponentials by a limit:

**Lemma 3.8** *For any square matrix $A$,*

$$e^A = \lim_{n\to\infty} \left(I + \frac{1}{n}A\right)^n.$$

PROOF We can use the binomial theorem to multiply out the expression $\left(I + \frac{1}{n}A\right)^n$ to get

$$\begin{aligned}
\left(I + \frac{1}{n}A\right)^n &= \left(\frac{1}{n}A + I\right)^n \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{1}{n^k} A^k I^{n-k} \\
&= \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \frac{1}{n^k} A^k \\
&= \sum_{k=0}^{n} \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \frac{1}{k!} A^k \\
&= \sum_{k=0}^{n} \left[\frac{n^k + \text{lower power terms of } n}{n^k}\right] \frac{1}{k!} A^k \\
&= \sum_{k=0}^{n} \left[1 + \text{some negative power terms of } n\right] \frac{1}{k!} A^k
\end{aligned}$$

As $n \to \infty$, each of the terms inside the square brackets goes to $1$, so

$$\lim_{n\to\infty} \left(I + \frac{1}{n}A\right)^n = \sum_{k=0}^{\infty} \frac{1}{k!} A^k = e^A. \square$$

**WARNING:** 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:** if $A$ is a square matrix, then for any $t \in \mathbb{R}$, we have

$$e^{tA} = e^{At} = \sum_{k=0}^{\infty} \frac{1}{k!}(At)^k = I + At + \frac{A^2}{2}t^2 + \frac{A^3}{3!}t^3 + \ldots$$

and

$$e^{tA} = e^{At} = \lim_{n\to\infty} \left(I + \frac{t}{n}A\right)^n$$

**Theorem 3.9 (Properties of matrix exponentials)** *Let $A$, $B$ and $S$ be square matrices of the same size, where $S$ is invertible. Let $n \in \{0, 1, 2, 3, ...\}$. 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 $B^n = \exp(An)$.*
4. *For any matrix $A$, $(e^A)^n = e^{An} = e^{nA}$.*
5. *$\exp($zero matrix$) = I$.*
6. *$\exp(SAS^{-1}) = Se^A S^{-1}$.*

PROOF MATH 322 or MATH 330. □

**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 $\Lambda$ are the corresponding eigenvalues); then $e^A = Se^\Lambda S^{-1}$.

**Theorem 3.10** *Let $P(t)$ be a family of square matrices, indexed by $t$. Then, the following are equivalent:*

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

**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) ⇒ (2) begins with an **as-yet unproven lemma:** which says that (1) implies that each $P(t)$ is a continuous and differentiable function of $t$ (we'll prove this lemma later).

Assuming this lemma, let $h > 0$ be small, and define $Q = P'(0)$. From calculus 1, we know we can approximate any differentiable function near $x = a$ by its linear approximation (a.k.a. tangent line) $L$:

$$f(x) \approx L(x) = f(a) + f'(a)(x - a).$$

In this context, since $P$ is differentiable, we can approximate $P(h)$ for $h$ near $0$ by

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

So for any $t > 0$,

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

(2) $\Rightarrow$ (3), (4): suppose $P(t) = e^{Qt} = \sum\limits_{n=0}^{\infty} \frac{1}{n!}Q^n$. Clearly

$$P(0) = e^{Q0} = \exp(\text{zero matrix}) = I.$$

Also,

$$\frac{d}{dt}P(t) = \frac{d}{dt}e^{Qt} = \frac{d}{dt}\sum_{n=0}^{\infty}\frac{Q^n}{n!}t^n = \sum_{n=1}^{\infty}\frac{Q^n}{(n-1)!}t^{n-1}$$
$$= \begin{cases} Q\left(\sum\limits_{n=0}^{\infty}\frac{Q^n}{n!}\right) = Qe^{Pt} = Q\,P(t). \\ \left(\sum\limits_{n=0}^{\infty}\frac{Q^n}{n!}\right)Q = e^{Pt}Q = P(t)\,Q. \end{cases}$$

(2) $\Rightarrow$ (5):

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

(3) $\Rightarrow$ (2); (4) $\Rightarrow$ (2) follow from the *Existence-Uniqueness Theorem* of differential equations (MATH 330), which says that a system of (ordinary) differential equations with given initial condition has a unique solution (under natural hypotheses

that hold here). Since $e^{Qt}$ is a solution of $P'(t) = QP(t)$; $P(0) = I$, it must be the only solution.

(5) $\Rightarrow$ (1) by writing the Taylor series of $P(t)$:

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

**Definition 3.11** *Let $\{X_t\}$ be a CTMC with finite state space. Then, by the preceding theorem, the time $t$ transition function $P(t)$ satisfies both of these differential equations:*

1. *the* **forward equation**
$$\begin{cases} P'(t) = P(t)\,Q \\ P(0) = I \end{cases}$$

2. *the* **backward equation**
$$\begin{cases} P'(t) = Q\,P(t) \\ P(0) = I \end{cases}$$

**Corollary 3.12** *If $P(t)$ is the time $t$ transition function for a CTMC with finite state space, then $P(t) = \exp(Qt)$ for some matrix $Q$ (and $Q$ 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 3.13** *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 OF A Q-MATRIX

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

> **Theorem 3.14** *A square matrix $Q$ is a Q-matrix if and only if for every $t$, $P(t) = \exp(Qt)$ is a stochastic matrix.*

PROOF We'll prove this by establishing two claims.

**Claim 1:** $q_{xx} < 0$ and $q_{xy} > 0 \, \forall \, x \neq y \iff P_{xy}(t) \in [0,1] \, \forall \, x, y \in \mathcal{S}, t \geq 0$.

**Claim 2:** $\sum_{y \in \mathcal{S}} q_{xy} = 0 \iff \sum_{y \in \mathcal{S}} P_{xy}(t) = 1 \, \forall \, x \in \mathcal{S}, t \geq 0$.

**Proof of Claim 1:** For small $h$, $P(h) \approx P(0) + P'(0)(h - 0) = I + Qh$.

**Proof of Claim 2:** ($\Rightarrow$) Assume $\sum_{y \in \mathcal{S}} q_{xy} = 0$ and let $P(t) = \exp(Qt)$.

Now let $q_{xy}^{(n)}$ denote the $xy$-entry of the matrix $Q^n$.

Having shown this, we see that for fixed $x \in \mathcal{S}$ and $t \geq 0$,

$$\begin{aligned}\sum_{y\in\mathcal{S}} P_{xy}(t) &= \sum_{y\in\mathcal{S}} [\exp(Qt)]_{xy} \\ &= \sum_{y\in\mathcal{S}} \left[\sum_{n=0}^{\infty} \frac{1}{n!} Q^n t^n\right]_{xy} \\ &= \sum_{y\in\mathcal{S}} I_{xy} + \sum_{n=1}^{\infty} \frac{t^n}{n!} q_{xy}^{(n)} \\ &= 1 + 0 = 1.\end{aligned}$$

($\Leftarrow$) Assume $\sum_{y\in\mathcal{S}} P_{xy}(t) = 1$. Then

**Corollary 3.15** *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 $t$.*

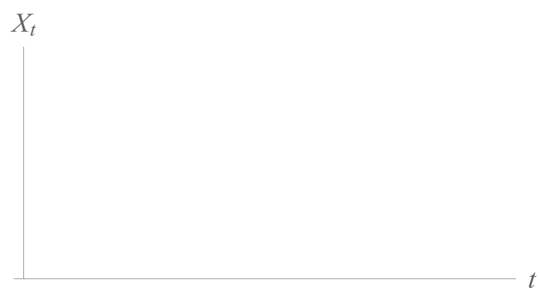
**Definition 3.16** *Let $\{X_t\}$ be a CTMC with finite state space $\mathcal{S}$. Then the matrix $Q = P'(0)$ is called the* **infinitesimal matrix** *or the* **generating matrix** *of the CTMC.*

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

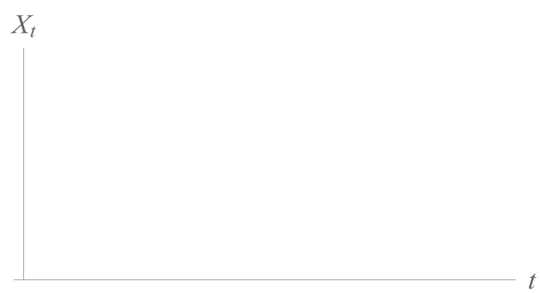
**Question:** Do the entries of $Q$ have any significance?

## Waiting times

**Definition 3.17** *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 $X_0 = x$.*



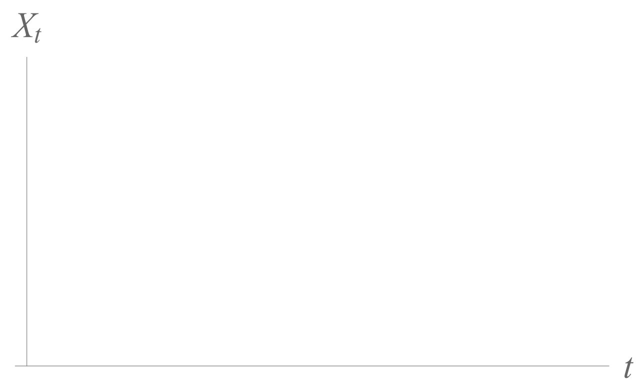

**Note:** One of the reasons we assume that the sample functions in a jump process are right-continuous is to make sure that $W_x$ is well-defined. We don't want, for example




**Theorem 3.18 (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 So let's compute $P(W_x > t)$.

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




$$=$$

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 that since $P'_{xx}(0) = q_{xx}$, we have

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

Therefore

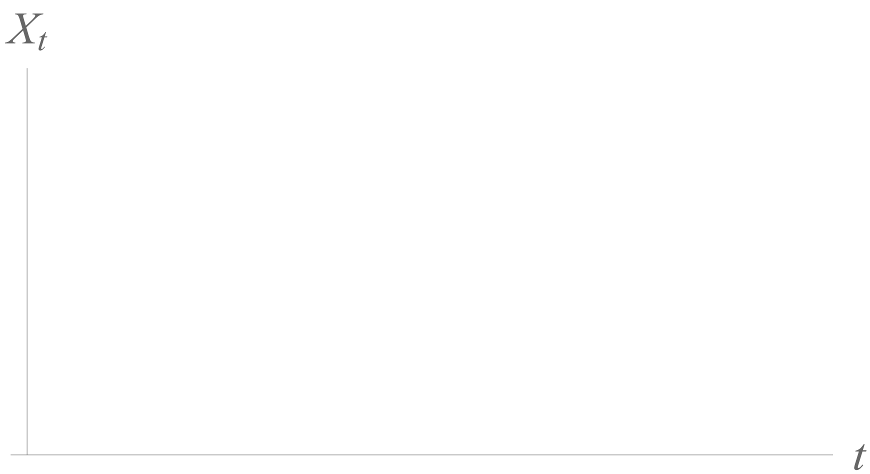
$$F_{W_x}(t) = P(W_x \leq t) = 1 - e^{q_{xx}t}$$

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

**Definition 3.19** *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




## Jump probabilities

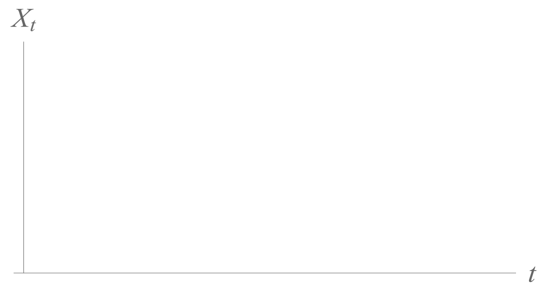
**Definition 3.20** *Let* $\{X_t\}$ *be a CTMC with finite state space* $\mathcal{S}$ *and initial distribution* $\pi_0$*. For each* $x, y \in \mathcal{S}$*, define the* **jump probability from** $x$ **to** $y$ *to be*

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

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

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

*The* **jump chain** *of* $\{X_t\}$ *is the discrete-time Markov chain with initial distribution* $\pi_0$ *and transition matrix* $\Pi$.




**Theorem 3.21 (Formula for jump probabilities)** *Let $\{X_t\}$ be a CTMC with finite state space $\mathcal{S}$ whose infinitesimal matrix is $Q$. Then for all $x, y \in \mathcal{S}$,*

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

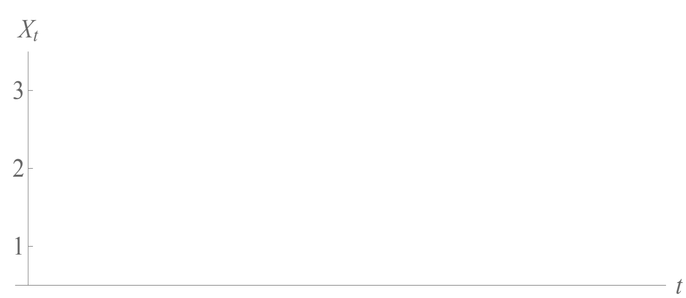
PROOF Later

EXAMPLE 1

Suppose the infinitesimal matrix of some CTMC $\{X_t\}$ is

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

1. Describe the waiting times for each state. In which state, on the average, would you expect to stay for the longest times before jumping?
2. Compute the jump matrix of the CTMC.



EXAMPLE 2

Consider a CTMC with state space $\{1,2,3\}$ and infinitesimal matrix

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -5 & 4 \\ 2 & 1 & -3 \end{pmatrix}.$$

1. Compute 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. Recall from the previous page that $Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -5 & 4 \\ 2 & 1 & -3 \end{pmatrix}$. Compute $P(t)$.

(To verify your work, you can check that $P(0) = I$, $P'(0) = Q$, the rows sum to $1$, all entries are always at least $0$, and the constant terms are the same in each entry of a fixed column.)

5. Recall from the previous page that

$$P(t) = \begin{pmatrix} \frac{11}{24} - \frac{1}{12}e^{-6} + \frac{5}{8}e^{-4} & \frac{1}{6} - \frac{1}{6}e^{-6} & \frac{3}{8} + \frac{1}{4}e^{-6} - \frac{5}{8}e^{-4} \\ \frac{11}{24} + \frac{5}{12}e^{-6} - \frac{7}{8}e^{-4} & \frac{1}{6} + \frac{5}{6}e^{-6} & \frac{3}{8} - \frac{5}{4}e^{-6} + \frac{7}{8}e^{-4} \\ \frac{11}{24} - \frac{1}{12}e^{-6} - \frac{3}{8}e^{-4} & \frac{1}{6} - \frac{1}{6}e^{-6} & \frac{3}{8} + \frac{1}{4}e^{-6} + \frac{3}{8}e^{-4} \end{pmatrix}.$$

Compute $P(X_{3/4} = 0 \,|\, X_{1/2} = 1)$.

6. If the initial distribution is $\pi_0 = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, find the distribution at time $t = \ln 2$.

## 3.3 General theory of CTMCs

| Henceforth we are no longer assuming that the state space $\mathcal{S}$ is finite. |
|---|

**<u>Recall</u>** 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 discrete-time 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 C-K 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 $\pi_{x,y} = 0$.)

What we don't know (that we knew in the finite state space case) is that $q_x = -q_{xx} = -P'_{xx}(0)$; in fact, we haven't even proved yet that $P_{xy}$ is a differentiable function of $t$. We'll address that issue first.

Throughout this chapter, let $\delta_{x,y} = \delta_{xy}$ be the **Kronecker delta**, i.e.

$$\delta_{x,y} = \delta_{xy} = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{else} \end{cases}$$

**Theorem 3.22 (Integral Equation)** *Let $\{X_t\}$ be a CTMC. Then for all $t \geq 0$,*

$$P_{x,y}(t) = \delta_{x,y} e^{-q_x t} + \int_0^t q_x e^{-q_x s} \left[ \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t-s) \right] ds.$$

PROOF

$$\begin{aligned} P_{x,y}(t) = P_x(X_t = y) &= P_x(X_t = y \,\cap\, W_x > t) + P_x(X_t = y \,\cap\, W_x \leq t) \\ &= P_x(X_t = y \,|\, W_x > t) P(W_x > t) + P_x(X_t = y \,\cap\, W_x \leq t) \\ &= \delta_{x,y} e^{-q_x t} + \int_0^t P(X_t = y \,|\, W_x = s) f_{W_x}(s)\, ds \\ &\qquad \text{(Law of Total Probability, continuous version)} \end{aligned}$$

Thus

$$\begin{aligned} P_{x,y}(t) &= \delta_{x,y} e^{-q_x t} + \int_0^t f_{W_x}(s) \sum_{z \in \mathcal{S}} P(X_s = z \,\cap\, X_t = y \,|\, W_x = s)\, ds \\ &= \delta_{x,y} e^{-q_x t} + \int_0^t q_x e^{-q_x s} \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t-s)\, ds. \;\square \end{aligned}$$

**Theorem 3.23 (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

$$P_{x,y}(t) = \delta_{xy} e^{-q_x t} + - \int_t^0 q_x e^{-q_x(t-u)} \left[ \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(u) \right] du$$
$$= \delta_{xy} e^{-q_x t} + q_x e^{-q_x t} \int_0^t e^{q_x u} \left[ \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(u) \right] du \qquad (\star)$$

**Theorem 3.24 (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 3.23, 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 earlier in this chapter.) Now

$$P'_{x,y}(t) = \frac{d}{dt} \left[ \delta_{xy} e^{-q_x t} + q_x e^{-q_x t} \int_0^t e^{q_x u} \left[ \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(u) \right] du \right]$$
$$= -q_x \left[ e^{-q_x t} \left( \delta_{xy} + q_x \int_0^t e^{q_x u} \left[ \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(u) \right] du \right) \right] + e^{-q_x t} q_x e^{q_x t} \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t)$$
$$= -q_x P_{x,y}(t) + q_x \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t). \ \square$$

**Corollary 3.25** *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 3.24,

$$\begin{aligned} P'_{x,y}(0) &= -q_x P_{x,y}(0) + q_x \sum_{z\in\mathcal{S}} \pi_{x,z} P_{z,y}(0) \\ &= -q_x \delta_{xy} + q_x \left[0 + 0 + \ldots + 0 + \pi_{x,y} \cdot 1 + 0 + \ldots + 0\right] \\ &= -q_x \delta_{xy} + q_x \pi_{x,y} \\ \\ &= \end{aligned}$$

**Definition 3.26** *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 3.25 we immediately see

**Theorem 3.27 (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 by linear approximation we have

$$P_{x,y}(t) \approx P_{x,y}(0) + P'_{x,y}(0)t = \delta_{x,y} + q_{xy}t.$$

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 3.28** *Let $\{X_t\}$ be a CTMC and let $x \in \mathcal{S}$. Then*

$$\sum_{y\in\mathcal{S}} q_{xy} = 0.$$

PROOF

$$\sum_{y\in\mathcal{S}} q_{xy} = q_{xx} + \sum_{y\neq x} q_{xy} = -q_x + \sum_{y\neq x} q_x \pi_{x,y} = -q_x + q_x \sum_{y\neq x} \pi_{x,y} = -q_x + q_x = 0. \square$$

**Theorem 3.29 (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 3.24,

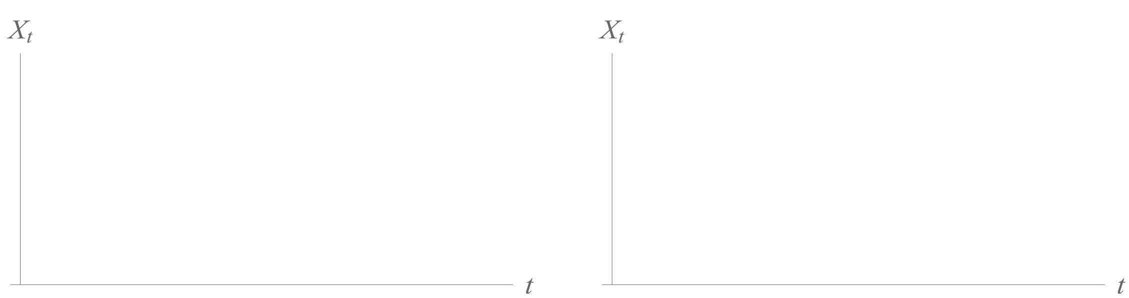
$$\begin{aligned} P'_{x,y}(t) &= -q_x P_{x,y}(t) + q_x \sum_{z \in \mathcal{S}} \pi_{x,z} P_{z,y}(t) \\ &= q_{xx} P_{x,y}(t) + q_x \sum_{z \neq x \in \mathcal{S}} \pi_{x,z} P_{z,y}(t) \\ &= \sum_{z \in \mathcal{S}} q_{x,z} P_{z,y}(t). \ \Box \end{aligned}$$

The transition functions of a CTMC also satisfy the forward equation given in Theorem 3.31. Deriving this equation follows the same line of argument as what we just went through, but instead of conditioning on the first jump in the process, you condition on the last jump before time $t$. To make this argument go through, we first need this lemma:

**Lemma 3.30 (Time reversal identity)** *Let $\{X_t\}$ be a CTMC. Then*

$$\begin{aligned} &q_{x_n} P\left(J_n \leq t < J_{n+1} \,|\, X_0 = x_0, X_{J_1} = x_1, X_{J_2} = x_2, ..., X_{J_n} = x_n\right) \\ &= q_{x_0} P\left(J_n \leq t < J_{n+1} \,|\, X_0 = x_n, X_{J_1} = x_{n-1}, X_{J_2} = x_{n-2}, ..., X_{J_n} = x_0\right). \end{aligned}$$

**What this lemma says:** Here's a picture when $n = 2$:




PROOF The event $J_n \leq t < J_{n+1}$ corresponds exactly to

$$J_n = W_{x_0} + W_{x_1} + ... + W_{x_{n-1}} + W_{x_n} > t,$$

i.e.

$$W_{x_n} > t - W_{x_0} - W_{x_1} - ... - W_{x_{n-1}}.$$

Since $W_{x_n} \sim Exp(q_{x_n})$, given values $s_0, ..., s_{n-1}$ of $W_{x_0}, ..., W_{x_{n-1}}$, the probability of this is

$$e^{-q_{x_n}(t-s_0-s_1-...-s_{n-1})} = e^{-q_{x_n}\left(t-\sum_{k=0}^{n-1} s_k\right)}.$$

So by the continuous LTP, the conditional probability of $J_n \leq t < J_{n+1}$ given $X_{J_j} = x_j$ for $j \in \{1, ..., n\}$ is therefore

$$\int \cdots \int_\Delta e^{-q_{x_n}\left(t-\sum_{k=0}^{n-1} s_k\right)} f_{W_{x_0}, W_{x_1}, ..., W_{x_n}}(s_0, s_1, ..., s_{n-1})\, dV$$

and since $W_{x_1}, ..., W_{x_n}$ are independent, this is

$$\int \cdots \int_\Delta e^{-q_{x_n}\left(t-\sum_{k=0}^{n-1} s_k\right)} f_{W_{x_0}}(s_0) f_{W_{x_1}}(s_1) \cdots f_{W_{x_{n-1}}}(s_{n-1})\, dV$$

and since $W_{x_k} \sim Exp(q_{x_k})$, this is

$$\int \cdots \int_\Delta e^{-q_{x_n}\left(t-\sum_{k=0}^{n-1} s_k\right)} \prod_{k=0}^{n-1} q_{x_k} e^{-q_{x_k} s_k}\, dV \tag{3.1}$$

where $\Delta$ is the set of $(s_1, ..., s_n) \in \mathbb{R}^n$ with $s_j \geq 0$.

In this last integral, perform a change of variables (with Jacobians) from the variables $(s_0, ..., s_{n-1})$ to $(u_0, ..., u_{n-1})$ by setting $u_0 = t - s_0 - s_1 - ... - s_{n-1}$, $u_1 = s_{n-1}$, $u_2 = s_{n-2}$, $u_3 = s_{n-3}$, ..., $u_{n-1} = s_1$ in the integral above to rewrite it as

$$\int \cdots \int_\Delta e^{-q_{x_0} u_0} \prod_{k=0}^{n-1} q_{x_{n-k}} e^{-q_{x_{n-k}} u_k}\, dV; \tag{3.2}$$

this gives the conditional probability in the second expression in the lemma. Notice that in (4.1), the integral has $q_{x_0}, ..., q_{x_{n-1}}$ but in (4.2), the integral has $q_{x_n}, ..., q_{x_n}$. So multiplying (4.1) by $q_{x_n}$ is equal to what you get when you multiply (4.2) by $q_{x_1}$, proving the lemma. □

**Theorem 3.31 (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 HW (maybe)

## 3.4 Class structure, recurrence and transience of CTMCs

**Definition 3.32** *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.)*

**Definition 3.33** *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 3.34** *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 3.35** *Let $\{X_t\}$ be a CTMC with state space $\mathcal{S}$. A distribution $\pi$ 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 3.36 (Stationarity equation for CTMCs)** *Let $\{X_t\}$ be a CTMC with state space $\mathcal{S}$. A distribution $\pi$ 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 3.37 (Stationarity and jump probabilities)** *Let $\{X_t\}$ be a CTMC with state space $\mathcal{S}$.*

1. *Suppose $\pi$ is a stationary distribution for $\{X_t\}$. For each $x \in \mathcal{S}$, set $\pi_{jump}(x) = \pi(x) q_x$. Then*
$$\sum_{x \in \mathcal{S}} \pi_{jump}(x) \pi_{x,y} = \pi_{jump}(y).$$
2. *Suppose $\pi_{jump} : \mathcal{S} \to [0, \infty)$ is a function such that*
$$\sum_{x \in \mathcal{S}} \pi_{jump}(x) \pi_{x,y} = \pi_{jump}(y).$$
*Then if we set, for each $y \in \mathcal{S}$, $\pi^*(y) = \frac{\pi_{jump}(y)}{q_y}$, then for all $y \in \mathcal{S}$,*
$$\sum_{x \in \mathcal{S}} \pi^*(x) P_{x,y}(t) = \pi^*(y).$$

PROOF HW

**WARNING:** The $\pi_{jump}$ defined in this theorem may not be a distribution on $\mathcal{S}$ (because its values may not sum to 1. But it satisfies the "stationarity equation" for the jump probabilities, which in matrix language would be

$$\pi_{jump} \Pi = \pi_{jump}.$$

However, if $\sum_{y \in \mathcal{S}} \pi_{jump}(y) = C$, then by setting $\pi'(y) = \frac{1}{C} \pi_{jump}(y)$ for all $y \in \mathcal{S}$ we get a stationary distribution $\pi'$ for the jump chain. That means that if irreducible CTMC $\{X_t\}$ has a stationary distribution, so does its jump chain, so the jump chain is positive recurrent, so the original CTMC is positive recurrent.

Similarly, the $\pi^*$ obtained in (2) must be a multiple of a stationary distribution for $\{X_t\}$.

**Theorem 3.38 (Steady-state distribution of CTMCs)** *Let $\{X_t\}$ be an irreducible, positive recurrent CTMC with a stationary distribution $\pi$. Then the stationary distribution is steady-state, i.e.*

- $\lim_{t\to\infty} P_{x,y}(t) = \pi(y)$ *for all* $x, y \in \mathcal{S}$*; and*
- $\lim_{t\to\infty} P(X_t = y) = \pi(y)$ *for all* $y \in \mathcal{S}$*, regardless of the initial distribution.*

**Why is the stationary distribution always steady-state**? The short answer is

PROOF Fix $h > 0$ and consider the discrete-time Markov chain $\{Z_n\} = \{X_{hn}\}$ for $n \in \{0, 1, 2, ...\}$. $\{Z_n\}$ has transition functions $P(x, y) = P_{x,y}(h)$, and since these functions are always positive for $h > 0$, $\{Z_n\}$ is irreducible and aperiodic. So the FTMC applied to $\{Z_n\}$ gives a stationary distribution $\pi$ for $\{Z_n\}$ (which must be the stationary distribution for $\{X_t\}$) which is steady-state for $\{Z_n\}$, i.e.

$$\lim_{n\to\infty} P^n(x, y) = \lim_{n\to\infty} P_{x,y}(hn) = \pi(y)$$

for all $x, y \in \mathcal{S}$. So for $t$ that are multiples of $h$, $P_{x,y}(t) \to \pi(y)$. Since $h$ can be chosen arbitarily small and since $t \mapsto P_{x,y}(t)$ is (uniformly) continuous, it follows (by a MATH 430 argument) that $\lim_{t\to\infty} P_{x,y}(t) = \pi(y)$. □

**Corollary 3.39** *An irreducible, positive recurrent CTMC cannot have more than one stationary distribution.*

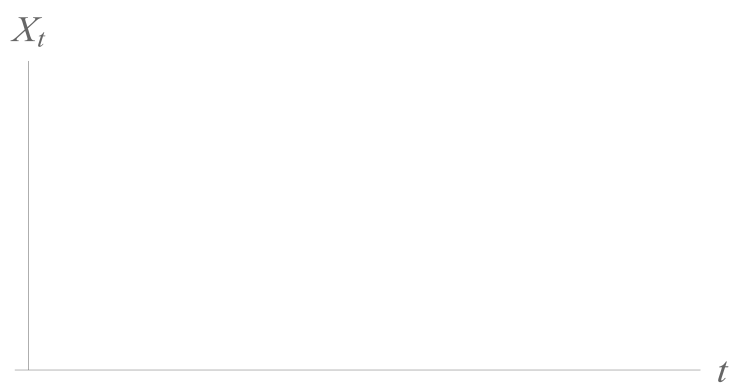
PROOF If $\pi$ and $\pi'$ are both stationary, then they would both be steady-state for $\{Z_n\}$ as described in the previous theorem. This is impossible. □

**Definition 3.40** *Let $E$ be an event. The symbol $\mathbb{1}_E$ refers to what is called the* **indicator function** *or* **characteristic function** *of $E$. This is a function that takes the value 1 if $E$ holds, and takes the value 0 if $E$ does not hold.*

**Theorem 3.41 (Stationary distributions of CTMCs)** *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 $\pi$ given by $\pi(x) = \frac{1}{m_x q_x}$ for all $x \in \mathcal{S}$.*

**Why should** $\pi(x) = \frac{1}{m_x q_x}$? Some motivation:




PROOF The only thing left to prove is the formula for the stationary distribution in the positive recurrent case. Suppose $\{X_t\}$ is irreducible and positive recurrent, and fix $x \in \mathcal{S}$. For each $y \in \mathcal{S}$, define

$$\tau_x(y) = E_x\left[\int_0^{T_x} \mathbb{1}_{\{X_s = y\}}\, ds\right];$$

this is the expected amount of time the chain spends in state $y$ before it first returns to $x$. Notice

$$\sum_{y \in \mathcal{S}} \tau_x(y) = E_x(T_x) = m_x,$$

so by setting $\pi_x(y) = \frac{1}{m_x}\tau_x(y)$, we get a distribution $\pi_x$ on $\mathcal{S}$.

Now let $\{Y_n\}$ denote the jump chain associated to $\{X_t\}$ and let $T_x^{jump}$ be the first return time to $x$ in $\{Y_n\}$ (this is the number of jumps it takes $x$ to return to itself in the CTMC). We have

$$\begin{aligned}
\tau_x(y) &= E_x\left[\sum_{n=0}^{\infty} W_y \mathbb{1}_{\{Y_n=y, n<T_x^{jump}\}}\right] \\
&= \frac{1}{q_y} E_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{Y_n=y, n<T_x^{jump}\}}\right] \\
&= \frac{1}{q_y} E_x\left[\sum_{n=0}^{T_x^{jump}-1} \mathbb{1}_{\{Y_n=y\}}\right]
\end{aligned}$$

Define

$$\gamma_x(y) = E_x\left[\sum_{n=0}^{T_x^{jump}-1} \mathbb{1}_{\{Y_n=y\}}\right],$$

so that $\tau_x(y) = \frac{1}{q_y}\gamma_x(y)$.

**Claim:** "$\gamma_x \Pi = \gamma_x$", i.e.

$$\sum_{y\in\mathcal{S}} \gamma_x(y)\, \pi(y,z) = \gamma_x(z).$$

**Proof of claim:** Since the jump chain is positive recurrent, $T_x^{jump} < \infty$) with probability $1$, so

$$\begin{aligned}
\gamma_x(z) &= E_x\left[\sum_{n=1}^{T_x^{jump}} \mathbb{1}_{\{Y_n=z\}}\right] \\
&= E_x\left[\sum_{n=1}^{\infty} \mathbb{1}_{\{Y_n=z \text{ and } n<T_x^{jump}\}}\right] \\
&= \sum_{n=1}^{\infty} E_x\left[\mathbb{1}_{\{Y_n=z \text{ and } n<T_x^{jump}\}}\right] \\
&= \sum_{n=1}^{\infty} P(Y_n = z \text{ and } n < T_x^{jump}) \\
&= \sum_{y\in\mathcal{S}} \sum_{n=1}^{\infty} P(Y_n = z, Y_{n-1} = y \text{ and } n < T_x^{jump}) \\
&= \sum_{y\in\mathcal{S}} \pi(y,z) \sum_{n=1}^{\infty} P(Y_{n-1} = y \text{ and } n < T_x^{jump}) \\
&= \sum_{y\in\mathcal{S}} \pi(y,z) E_x\left[\sum_{m=0}^{\infty} \mathbb{1}_{\{Y_m=y \text{ and } n<T_x^{jump}-1\}}\right]
\end{aligned}$$

From the previous page,

$$\begin{aligned}\gamma_x(z) &= \sum_{y\in\mathcal{S}} \pi(y,z) E_x\left[\sum_{m=0}^{\infty} \mathbb{1}_{\{Y_m=y \text{ and } n<T_x^{jump}-1\}}\right] \\ &= \sum_{y\in\mathcal{S}} \pi(y,z) E_x\left[\sum_{m=0}^{T_x^{jump}-1} \mathbb{1}_{\{Y_m=y\}}\right] \\ &= \sum_{y\in\mathcal{S}} \pi(y,z)\gamma_x(y).\end{aligned}$$

Having proven the claim, by (2) of Theorem 3.37, $\tau_x$ is a multiple of a stationary distribution of $\{X_t\}$. But the only multiple of $\tau_x$ which is a distribution is the $\pi_x$ we defined earlier, i.e.

$$\pi_x(y) = \frac{1}{m_x}\tau_x(y) = \frac{1}{m_x q_y}\gamma_x(y).$$

So for each $x \in \mathcal{S}$, this $\pi_x$ must be stationary, and since there is at most one stationary distribution, we know $\pi_x = \pi$ for all $x \in \mathcal{S}$. In particular,

$$\begin{aligned}\pi(x) = \pi_x(x) = \frac{1}{m_x}\tau_x(x) = \frac{1}{m_x q_x}\gamma_x(x) &= \frac{1}{m_x q_x} E_x\left[\sum_{n=0}^{T_x^{jump}-1} \mathbb{1}_{\{Y_n=x\}}\right] \\ &= \frac{1}{m_x q_x}(1) \\ &= \frac{1}{m_x q_x}. \ \square\end{aligned}$$

**Corollary 3.42** *Let $\{X_t\}$ be an irreducible, positive recurrent CTMC and let $x, y \in \mathcal{S}$. If we define $\tau_x(y)$ to be the expected amount of time spent in state $y$ before the first return to $x$, given that $X_0 = x$, then*

$$\tau_x(y) = m_x \pi(y).$$

PROOF From above, $\frac{1}{m_y q_y} = \pi_y(y) = \pi_x(y) = \frac{1}{m_x q_y}\gamma_x(y)$, it must be that $\gamma_x(y) = \frac{m_x}{m_y}$, so

$$\tau_x(y) = \frac{1}{q_y}\gamma_x(y) = \frac{m_x}{q_y m_y} = m_x \pi(y). \ \square$$

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 3.43 (Ergodic theorem for CTMCs)** *Let $\{X_t\}$ be an irreducible, positive recurrent CTMC, and let $\pi$ be the stationary distribution of $\{X_t\}$. Then for all $y \in \mathcal{S}$,*

$$P\left[\lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{1}_{\{X_s = y\}}\, ds = \pi(y)\right] = 1.$$

**A picture to explain:**

PROOF (Really just a sketch of the proof) Let $x = X_0$. The expected length of each block of time in $[0, t]$ between successive returns to $x$ is $m_x$, so by the Strong Law of Large Numbers, the average length of the blocks approaches $m_x$ with probability $1$. In each block, the expected amount of time spent in state $y$ is $\tau_x(y) = m_x \pi(y)$, so by the SLLN the average time spent in $y$ in each block approaches $m_x \pi(y)$ with probability $1$. Therefore the proportion of time spent in state $y$ in each block approaches $\frac{m_x \pi(y)}{m_x} = \pi(y)$ with probability $1$. The result follows. □

## 3.5 Birth and death CTMCs

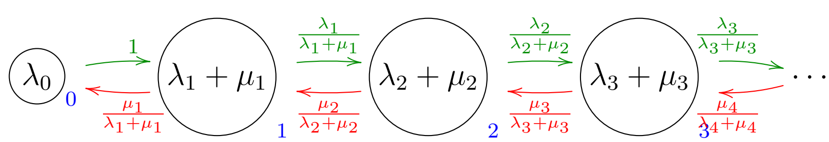
A birth-death CTMC is a CTMC where all the jumps are of size $\pm 1$. More formally:

**Definition 3.44** *A* **birth and death CTMC** *(or* **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:

$$\underline{\text{Given}}: \begin{cases} \text{birth rates} & \lambda_x = q_{x,x+1} \\ \text{death rates} & \mu_x = q_{x,x-1} \end{cases}$$

So the directed graph of a birth-death CTMC looks like:




**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.

$$\text{“}p_x\text{”} = \frac{\lambda_x}{q_x} \quad \text{and} \quad \text{“}q_x\text{”} = \frac{\mu_x}{q_x} :$$

Recall from Chapter 2 that the jump chain (and hence the birth-death CTMC) is transient if and only if

$$\sum_{x=1}^{\infty} \gamma_x < \infty$$

which happens if and only if

$$\begin{aligned} \sum_{x=1}^{\infty} \frac{``q_1" \quad ``q_2" \quad \cdots \quad ``q_x"}{``p_1" \quad ``p_2" \quad \cdots \quad ``p_x"} &< \infty \\ \text{i.e. } \sum_{x=1}^{\infty} \frac{\frac{\mu_1}{q_1} \cdot \frac{\mu_2}{q_2} \cdots \frac{\mu_x}{q_x}}{\frac{\lambda_1}{q_1} \frac{\lambda_2}{q_2} \cdots \frac{\lambda_x}{q_x}} &< \infty \\ \text{i.e. } \sum_{x=1}^{\infty} \frac{\mu_1 \mu_2 \cdots \mu_x}{\lambda_1 \lambda_2 \cdots \lambda_x} &< \infty. \end{aligned}$$

We have proven:

**Theorem 3.45** *An irreducible birth-death CTMC with state space $\mathcal{S} = \{0, 1, ..., \}$ 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 classify birth-death CTMCs as positive recurrent or not, and compute their stationary distribution, using the following machinery:

**Definition 3.46** *Let $\{X_t\}$ be an irreducible birth-death CTMC. Define*

$$\phi_0 = 1 \;\; \textit{and} \;\; \phi_y = \frac{\lambda_0 \cdots \lambda_{y-1}}{\mu_1 \cdots \mu_y}$$

*for every $y > 0$ in $\mathcal{S}$.*

Think of $\phi_y$ as "the product of all the $\lambda$s to the left of $y$ over the product of all the $\mu$s to the left of $y$".

**Theorem 3.47** *An irreducible birth-death CTMC on $\mathcal{S} = \{0, 1, ..., \}$ is positive recurrent if and only if*

$$\sum_{y \in \mathcal{S}} \phi_y < \infty.$$

**Theorem 3.48** *The stationary distribution of an irreducible, positive recurrent birth-death CTMC is given by*

$$\pi(x) = \frac{\phi_x}{\sum\limits_{y \in \mathcal{S}} \phi_y}.$$

EXAMPLE 3

Show that the birth-death CTMC $\{X_t\}$ with $\lambda_x = 1$ and $\mu_x = 2$ for all $x$ is positive recurrent.

*Solution:* Compute $\phi_y$ for each $y$. $\phi_0 = 1$ and for $y \geq 1$, we have

$$\phi_y = \frac{\lambda_0 \cdots \lambda_{y-1}}{\mu_1 \cdots \mu_y} = \frac{1(1) \cdots 1}{2(2) \cdots 2} = \frac{1}{2^y}.$$

Since

$$\sum_{y=0}^{\infty} \phi_y = \sum_{y=0}^{\infty} \frac{1}{2^y} = 2 < \infty,$$

$\{X_t\}$ is positive recurrent. The stationary distribution is therefore given by

$$\pi(x) = \frac{\phi_x}{\sum_y \phi_y} = \frac{\phi_x}{2} = \frac{1}{2^{x+1}}.$$

(In other words, $\pi = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, ...\right)$.)

EXAMPLE 4: PURE BIRTH PROCESS

Let $\{X_t\}$ be a **pure birth process**, meaning that $\{X_t\}$ is a birth-death CTMC on $\mathcal{S} = \{0, 1, 2, 3, ...\}$ with $\mu_x = 0$ for all $x$.

1. Sketch the directed graph of $\{X_t\}$.
2. Compute $P_{x,y}(t)$ if $y < x$.
3. Compute $P_{x,x}(t)$.
4. Use the forward equation to derive a recursive formula for $P_{x,y+1}(t)$ in terms of $P_{x,y}(t)$.
5. Compute $P_{x,x+1}(t)$.

EXAMPLE 5: TWO-STATE (BIRTH-DEATH) CTMC

Let $\{X_t\}$ be a birth-death CTMC on $\mathcal{S} = \{0, 1\}$ = {OFF, ON}.

1. Sketch the directed graph of $\{X_t\}$.
2. Compute the infinitesimal matrix of $\{X_t\}$.
3. Compute the stationary distribution of $\{X_t\}$.
4. Compute the transition matrices $P_{x,y}(t)$.

EXAMPLE 6: POISSON PROCESS

A **Poisson process** is a pure birth process on $\mathcal{S} = \{0, 1, 2, ...\}$ with $\lambda_x = \lambda$ for all $x$. $\lambda$ is called the **rate** of the Poisson process.

In Example 4, we derived the following recursive formula for $P_{x,y+1}(t)$ in terms of $P_{x,y}(t)$, which holds for any pure birth process:

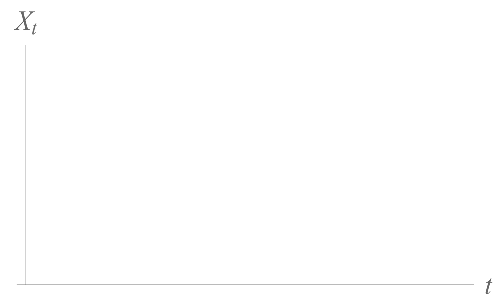
$$P_{x,y+1}(t) = e^{-\lambda_{y+1}t} \int \lambda_y e^{\lambda_{y+1}t} P_{x,y}(t)\, dt.$$

Use this recursive formula to compute $P_{x,x+1}(t)$, $P_{x,x+2}(t)$ and $P_{x,x+3}(t)$ for a Poisson process with rate $\lambda$, and use your answers to conjecture a general formula for $P_{x,x+a}(t)$ for $a \geq 0$.

## 3.6 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 3.49 (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, let $M = \min(A_1, ..., A_d)$). Compute the cdf of $M$ using transformation methods from MATH 414). $\square$

**Corollary 3.50** *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 CTMC).*

(Henceforth, all branching processes are assumed to have $A$ exponential, and $\lambda$ 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

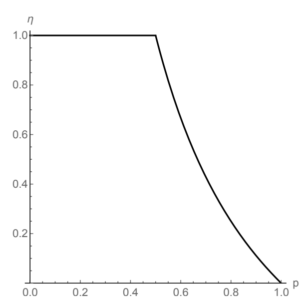
**Observations:** In a branching process,

1. $0$ is absorbing (meaning that $P_{0,0}(t) = 1$ for every $t \geq 0$); and
2. every nonzero state in $\mathcal{S}$ is transient (because that state leads to $0$ with positive probability);
3. the jump chain of a branching process is a ____________________ with

**Theorem 3.51** *Let $\{X_t\}$ be a branching process. Then the extinction probability $\eta = f_{1,0}$ satisfies*

$$\eta = \begin{cases} \frac{1-p}{p} & \text{if } p > \frac{1}{2} \\ 1 & \text{if } p < \frac{1}{2} \end{cases}.$$

**Note:** As with a Galton-Watson branching chain, $f_{x,0} = \eta^x$ for all $x \in \{0, 1, 2, ...\}$.




PROOF Notice that $\eta = f_{1,0}$ in the branching process is the same as $\eta = f_{1,0}$ in the associated jump chain. Now use the formulas derived in the proof of Theorem 2.39. First, $\gamma_0 = 1$ and if $y > 0$,

$$\gamma_y = \frac{q_1 q_2 \cdots q_y}{p_1 \cdots p_y} = \frac{(1-p)(1-p)\cdots(1-p)}{p\, p \cdots p} = \left(\frac{1-p}{p}\right)^y,$$

so

$$\begin{aligned} f_{1,0} &= 1 - \frac{1}{\sum\limits_{y=0}^{\infty} \gamma_y} \quad \text{(from the proof of Thm 2.39 on p. 112)} \\ &= 1 - \left[ \sum_{y=0}^{\infty} \left( \frac{1-p}{p} \right)^y \right]^{-1} \\ &= \begin{cases} 1 - \left[ \frac{1}{1 - \left(\frac{1-p}{p}\right)} \right]^{-1} & \text{if } \frac{1-p}{p} < 1 \\ 1 - [\infty]^{-1} & \text{else} \end{cases} \\ &= \begin{cases} 1 - \left[ 1 - \frac{1-p}{p} \right] & \text{if } \frac{1-p}{<} p \\ 1 & \text{else} \end{cases} \\ &= \begin{cases} \frac{1-p}{p} & \text{if } p > \frac{1}{2} \\ 1 & \text{else} \end{cases} \quad \square \end{aligned}$$

## 3.7 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 $\lambda$ (i.e. that the number of arrivals in line follows a Poisson process with rate $\lambda$) and that the time it takes each customer to be served is exponential with parameter $\mu$. 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 CTMC with birth and death rates

$$\begin{cases} \lambda_x = \\ \\ \mu_x = \end{cases}$$

| |
|---|
| **Question 1:** 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

$$\begin{aligned}
P(X_t^{new} = n) &= \sum_{k=0}^{\infty} P(X_t^{new} = n \text{ and } C_t = k) \\
&= \sum_{k=n}^{\infty} P(X_t^{new} = n \text{ and } C_t = k) \quad (\text{since } X_t^{new} \leq C_t) \\
&= \sum_{k=n}^{\infty} P(X_t^{new} = n \,|\, C_t = k) P(C_t = k) \\
&= \sum_{k=n}^{\infty} \left[ \begin{pmatrix} k \\ n \end{pmatrix} p_t^n (1 - p_t)^{k-n} \right] \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\
&= \frac{p_t^n e^{-\lambda t}}{n!} \sum_{k=n}^{\infty} \frac{(1 - p_t)^{k-n} (\lambda t)^k}{(k-n)!} \\
&= \frac{(\lambda t p_t)^n e^{-\lambda t}}{n!} \sum_{k=n}^{\infty} \frac{[\lambda t (1 - p_t)]^{k-n}}{[k-n]!}
\end{aligned}$$

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 $X_t^{new} \sim Pois(\lambda t p_t)$.

Now let $X_t^{orig} = \#$ of customers present initially that are still being served at time $t$.

$X_t^{orig}$ is with parameters $\Bigg\{$

Since $X_t = X_t^{new} + X_t^{orig}$, we have

$$\begin{aligned}
P_{x,y}(t) &= P_x(X_t = y) \\
&= \sum_{k=0}^{\min(x,y)} P_x(X_t^{orig} = k) P_x(X_t^{new} = y - k) \\
&= \sum_{k=0}^{\min(x,y)} \left( \left[ \begin{pmatrix} x \\ k \end{pmatrix} e^{-\mu k t} (1 - e^{-\mu t})^{x-k} \right] \left[ \frac{\left[ \frac{\lambda}{\mu} (1 - e^{-\mu t}) \right]^{y-k}}{(y-k)!} \exp\left( \frac{-\lambda}{\mu} (1 - e^{-\mu t}) \right) \right] \right).
\end{aligned}$$

On the previous page, we showed

$$P_{x,y}(t) = \sum_{k=0}^{\min(x,y)} \left( \left[ \binom{x}{k} e^{-\mu k t}(1-e^{-\mu t})^{x-k} \right] \left[ \frac{\left[\frac{\lambda}{\mu}(1-e^{-\mu t})\right]^{y-k}}{(y-k)!} \exp\left( \frac{-\lambda}{\mu}(1-e^{-\mu t}) \right) \right] \right).$$

**Question 2:** Is the infinite server queue positive recurrent, null recurrent or transient?

**Answer:** We could use the standard formulas for classifying birth-death CTMCs discussed earlier, but instead, let's use our formula for $P_{x,y}(t)$ and see what happens when $t \to \infty$. It turns out that this will give us a steady-state distribution, which will tell us that the infinite server queue is positive recurrent:

$$\begin{aligned}
\lim_{t\to\infty} P_{x,y}(t) &= \lim_{t\to\infty} (k = 0 \text{ term of the above sum}) \\
&= \lim_{t\to\infty} \binom{x}{0} e^{-0} \left(1-e^{-\mu t}\right)^x \left[ \frac{\left[\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\right]^{y}}{y!} \exp\left( \frac{-\lambda}{\mu}\left(1-e^{-\mu t}\right) \right) \right] \\
&= (1)1(1)^x \left[ \frac{\left[\frac{\lambda}{\mu}(1)\right]^{y}}{y!} \exp\left( \frac{-\lambda}{\mu}(1) \right) \right] \\
&= \frac{\left(\frac{\lambda}{\mu}\right)^y}{y!} e^{-(\lambda/\mu)}.
\end{aligned}$$

We have proven:

**Theorem 3.52 (Steady-state distribution of the infinite server queue)** *The steady-state distribution of the infinite server queue where the customers arrive exponentially with parameter $\lambda$ and are served exponentially with parameter $\mu$ is Poisson with parameter $\frac{\lambda}{\mu}$.*

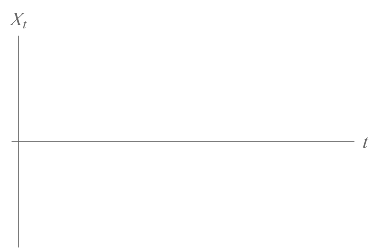
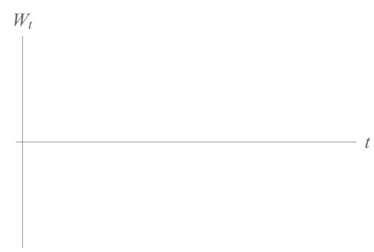
*Chapter 4*

# Brownian motion

## 4.1 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 | (SIMPLE, UNBIASED) 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 increment property | $\forall\, 0 \le t_1 \le t_2 \le ... \le t_n \in \mathbb{Z}$, the r.v.s $X_{t_2} - X_{t_1}$, $X_{t_3} - X_{t_2}, ..., X_{t_n} - X_{t_{n-1}}$ are mutually $\perp$   |   |
| stationarity property (time homoegenity) | 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 4.1** *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* $\sigma^2$ *if*

1. $W_0 = 0$;
2. *For all* $0 \leq t_1 \leq t_2 \leq ... \leq t_n \in \mathbb{R}$, *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* $0 \leq t_1 \leq t_2$ *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 4.2 (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 $n \geq 1$, 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 $[0, \infty)$

Next, for all $t \in \mathbb{D}_n^{new}$, set

$$t^+ = \min\{s \in \mathbb{D}_{n-1} : s > t\}$$

and

$$t^- = \max\{s \in \mathbb{D}_{n-1} : s < t\}.$$

As an example of this notation,

$$\mathbb{D}_1 = \{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, ...\}$$
$$\mathbb{D}_2 = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}, ...\}$$
$$\mathbb{D}_2^{new} = \{\frac{1}{4}, \frac{3}{4}, \frac{5}{4}, ...\}$$

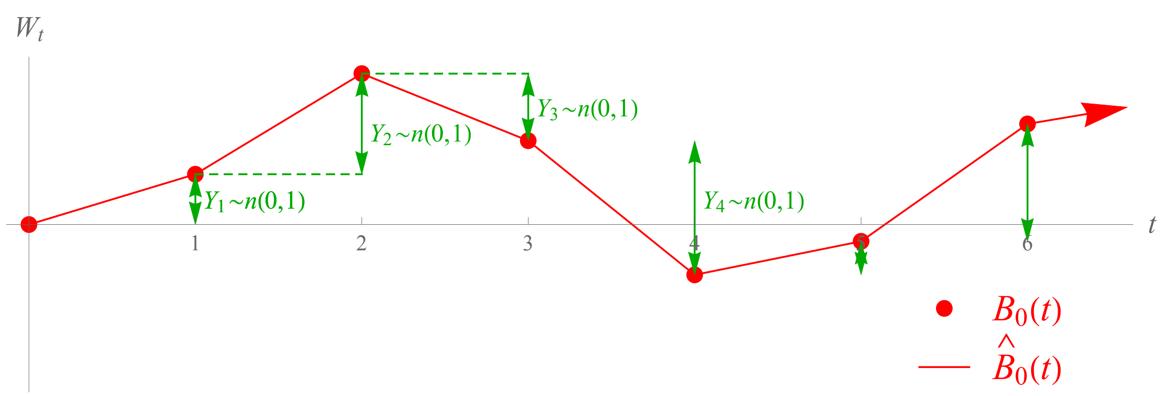
So, for example,

$$\left(\frac{3}{4}\right)^- = \frac{1}{2} \text{ and } \left(\frac{3}{4}\right)^+ = 1$$

**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 $Y_t$s. Let $\{B_0(t)\}_{t\in\mathbb{N}}$ be a discrete-time stochastic process defined by setting

$$B_0(t) = \sum_{j=1}^{t} Y_j.$$

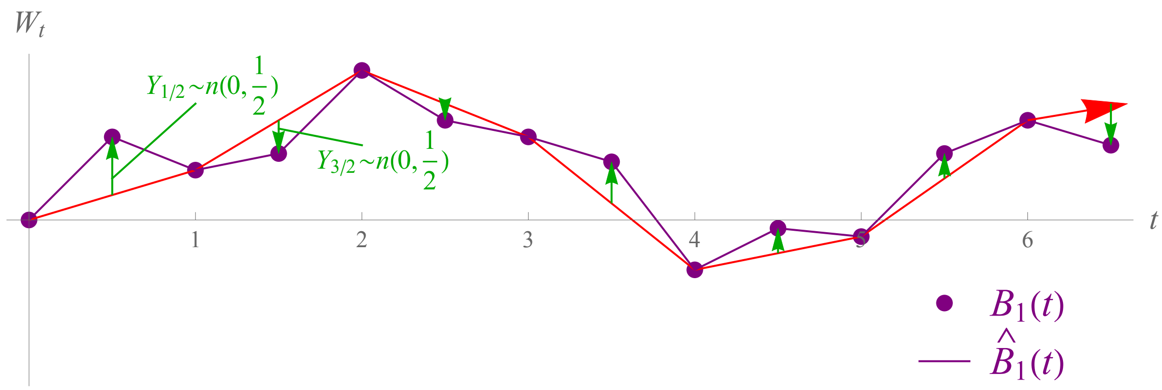
Then let $\{\widehat{B_0}(t)\}_{t\in[0,\infty)}$ be the continuous-time stochastic process obtained by interpolating linearly between the points of $\{B_0(t)\}$:




**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 $Y_t$'s defined either here or earlier. Let $\{B_1(t)\}_{t\in\mathbb{D}_1}$ be a discrete-time stochastic process defined by setting

$$B_1(t) = \begin{cases} B_0(t) & \text{if } t \in \mathbb{D}_0 \\ \frac{1}{2}\left(B_0(t^-) + B_0(t^+)\right) + Y_t & \text{if } t \in \mathbb{D}_1^{new} \end{cases}.$$

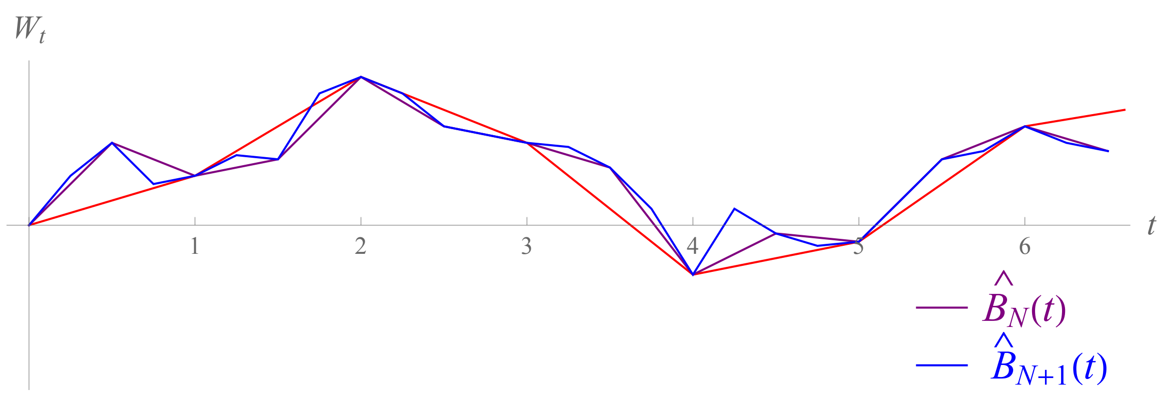
Then let $\{\widehat{B_1}(t)\}_{t\in[0,\infty)}$ be the continuous-time stochastic process obtained by interpolating linearly between the points of $\{B_1(t)\}$:




**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 $Y_t$'s 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

$$B_{N+1}(t) = \begin{cases} B_N(t) & \text{if } t \in \mathbb{D}_N \\ \frac{1}{2}\left(B_N(t^-) + B_N(t^+)\right) + Y_t & \text{if } t \in \mathbb{D}_{N+1}^{new} \end{cases}.$$

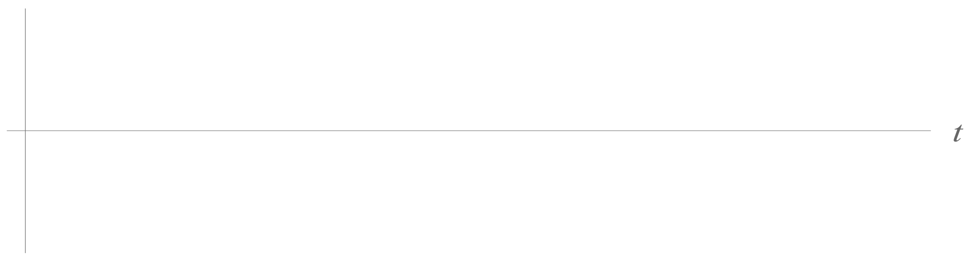
Then let $\{\widehat{B_{N+1}}(t)\}_{t \in [0,\infty)}$ be the continuous-time stochastic process obtained by interpolating linearly between the points of $\{B_{N+1}(t)\}$:




Now define $W_t = \lim_{n\to\infty} \widehat{B_N}(t)$. One can show that $\{W_t\}$ satisfies all the properties necessary to be a Brownian motion (I'm omitting the details). $\square$

Brownian motion arises commonly in real-world situations, including movements of particles suspended in a liquid, fluctuations in the stock market, the path-integral formulation of quantum mechanics, option pricing models (the Black-Scholes equations) and 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 1

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 $W_{14} - W_{12} > -3$.

## 4.2 Markov properties of Brownian motion

Let $\{W_t\}$ be a BM. The discrete version of the Markov property would say something like this:

$$P(W_t = y \,|\, W_{t_1} = x_1, W_{t_2} = x_2, ..., W_{t_n} = x_n) = P(W_t = y \,|\, W_{t_n} = x_n)$$
$$\forall\, 0 \le t_1 \le t_2 \le t_3 \le ... \le t_n \le t, \forall\, x_1, ..., x_n, y \in \mathbb{R}$$

But since $W_t$ is continuous, a better formulation of the same idea in this setting is in terms of conditional densities:

This holds because of the independent increment property in the definition of BM.

**Definition 4.3** *Let $\{W_t\}$ be a BM. Given $x, y \in \mathbb{R}$ and $t \ge 0$, the* **time $t$ transition density** *for the BM is*

$$p_{x,y}(t) = f_{W_t|W_0}(y|x) \quad (= f_{W_{s+t}|W_s}(y|x)\,\forall\, s \textit{ by time homogeneity}).$$

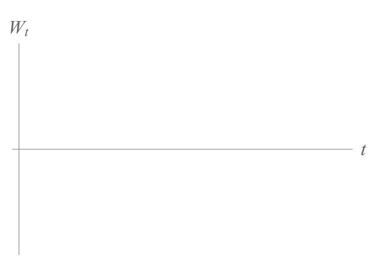
**Theorem 4.4 (Markov property for Brownian motion)** *Let $\{W_t\}$ be a BM with parameter $\sigma^2$. Then for any $s, t \ge 0$, $W_{s+t} - W_s$ is independent of $W_s$ and $W_{s+t} - W_s \sim n(0, \sigma^2 t)$.*

In other words, if $W_s = x$, then $W_{s+t}$ is a continuous r.v. with density function

$$f(y) = \frac{1}{\sigma\sqrt{2\pi t}} \exp\left[\frac{-(y-x)^2}{2\sigma^2 t}\right].$$

A stronger version of the Markov property is this result, whose proof is beyond the scope of this class:

**Theorem 4.5 (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 $\{W_t : t \le T\}$.*




**Note:** The strong Markov property also holds for Markov chains and CTMCs.

Now we come to an important result, which gives the distribution function for the hitting time to state $b$ in a BM:

**Theorem 4.6 (Reflection Principle)** *Let $\{W_t\}$ be a BM with parameter $\sigma^2$. 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 4.7** *Let $\{W_t\}$ be a BM with parameter $\sigma^2$. Fix $b > 0$ and let $T_b = \min\{t \geq 0 : W_t = b\}$. Then $T_b$ has density*

$$f_{T_b}(t) = \frac{b}{\sigma\sqrt{2\pi t^3}} \exp\left[\frac{-b^2}{2t\sigma^2}\right].$$

PROOF Differentiate $F_{T_b}$ with respect to $t$. □

**Corollary 4.8** *Let $\{W_t\}$ be a BM with parameter $\sigma^2$. Then $\{W_t\}$ is irreducible, i.e. for any $b \in \mathbb{R}$,*

$$P(W_t = b \text{ for some } t \geq 0) = 1.$$

PROOF This can be calculated using the Reflection Principle:

$$\begin{aligned} P(W_t = b \text{ for some } t \geq 0) &= P(T_b < \infty) \\ &= \lim_{t\to\infty} P(T_b < t) \end{aligned}$$

**Theorem 4.9 (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

$$\begin{aligned} P_0(W_s = 0 \text{ for some } s \geq 1) &= \lim_{t\to\infty} P_0(W_s = 0 \text{ for some } s \in [1,t]) \\ &= \lim_{t\to\infty} \int_{-\infty}^{\infty} f_{W_1}(b) P_0(W_s = 0 \text{ for some } s \in [1,t] \,|\, W_1 = b)\, db \end{aligned}$$

$$= \lim_{t\to\infty} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-b^2}{2\sigma^2}\right) \left[2 - 2\Phi\left(\frac{b}{\sigma\sqrt{t-1}}\right)\right] db$$

From the previous page,

$$P_0(W_s = 0 \text{ for some } s \geq 1) = \lim_{t\to\infty} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-b^2}{2\sigma^2}\right) \left[2 - 2\Phi\left(\frac{b}{\sigma\sqrt{t-1}}\right)\right] db$$

$$= \lim_{t\to\infty} \frac{2}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-b^2}{2\sigma^2}\right) \int_{\frac{b}{\sigma\sqrt{t-1}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx\, db$$

$$= \lim_{t\to\infty} \frac{2}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-b^2}{2\sigma^2}\right) \int_{\frac{b}{\sqrt{t-1}}}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-u^2}{2\sigma^2}\right) du\, db$$

$$= \frac{1}{\pi\sigma^2} \int_{-\infty}^{\infty} \int_{0}^{\infty} \exp\left(\frac{-(u^2+b^2)}{2\sigma^2}\right) du\, db$$

(change to polar coordinates)

$$= \frac{1}{\pi\sigma^2} \int_{r=0}^{\infty} \int_{0}^{\pi} \exp\left(\frac{-r^2}{2\sigma^2}\right) r\, d\theta\, dr$$

$$= \frac{1}{\sigma^2} \int_{0}^{\infty} e^{-r^2/2\sigma^2} r\, dr$$

(let $v = -r^2/2\sigma^2$ so that $dv = -\frac{r}{\sigma^2} dr$, i.e. $-\sigma^2\, dv = r\, dr$)

$$= \frac{1}{\sigma^2}(-\sigma^2) \int_{0}^{-\infty} e^v\, dv$$

$$= \int_{-\infty}^{0} e^v\, dv = e^0 - e^{-\infty} = 1. \square$$

## 4.3 Martingales associated to Brownian motion

**Theorem 4.10** *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 $0 < s < t$:

$$\begin{aligned} E[W_t \,|\, \mathcal{F}_s] &= E[W_s + (W_t - W_s) \,|\, \mathcal{F}_s] \\ &= E[W_s \,|\, \mathcal{F}_s] + E[W_t - W_s \,|\, \mathcal{F}_s] \\ &= W_s + E[W_t - W_s \,|\, \mathcal{F}_s] \quad \text{(since } W_s \text{ is } \mathcal{F}_s\text{-mble)} \\ &= W_s + E[W_t - W_s] \quad \text{(since } W_t - W_s \perp \mathcal{F}_s) \\ &= W_s + 0 \quad \text{(since } W_t - W_s \text{ is } n(0, \sigma^2(t-s))) \\ &= W_s. \end{aligned}$$

Thus $\{W_t\}$ is a martingale by definition.

The proof that $\{W_t^2 - t\}$ is a martingale is a HW problem.

To prove $\left\{\exp\left(\theta W_t - \frac{\theta^2 t}{2}\right)\right\}$ is a martingale, let $U_t = \exp\left(\theta W_t - \frac{\theta^2 t}{2}\right)$, again let $\{\mathcal{F}_t\}$ be the natural filtration of $\{W_t\}$, and let $0 < s < t$:

$$\begin{aligned} E[U_t \,|\, \mathcal{F}_s] &= E\left[\exp\left(\theta W_t - \frac{1}{2}\theta^2 t\right) \,|\, \mathcal{F}_s\right] \\ &= \end{aligned}$$

**Theorem 4.11** *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 4.12** *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 4.8, $P(T < \infty) = 1$, so $P_x(T_b < T_a) = 1 - P_x(T_a < T_b)$.

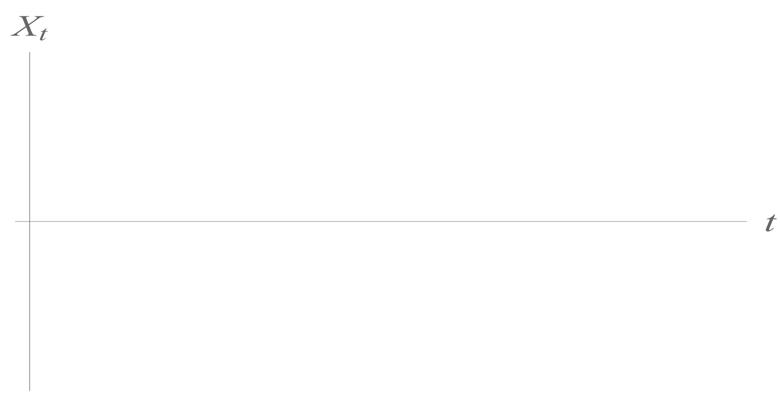
EXAMPLE 2

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 $30$?

## Brownian motion with drift

**Q:** Unbiased simple random walk : BM :: Biased simple random walk : ?

**Definition 4.13** *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.* $\mu$ *is called the* **drift parameter** *of* $\{X_t\}$*.*




**Theorem 4.14 (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 For statement (1),

$$X_t = W_t + \mu_t \sim n(0, \sigma^2 t) + \mu t = n(\mu t, \sigma^2 t).$$

For statement (2),

$$\begin{aligned} X_{t_2} - X_{t_1} &= (W_{t_2} + \mu t_2) - (W_{t_1} + \mu t_1) \\ &= (W_{t_2} - W_{t_1}) + \mu(t_2 - t_1) \end{aligned}$$

and similarly,

$$X_{t_4} - X_{t_3} =$$

Since $\{W_t\}$ has the independent increment property,

For statement (3),

For statement (4),

$$X_{T+t} - X_T = (W_{T+t} + \mu(T+t)) - (W_T - \mu T) = W_{T+t} - W_T + \mu t.$$

Since $\{W_t\}$ has the strong Markov property, $W_{T+t} - W_T$ is a BM, independent of $W_T$. So $X_{T+t} - X_T$ is a BM with drift, independent of $W_T$ (hence independent of $X_T$). $\square$

**Theorem 4.15** *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 4.16 (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

EXAMPLE 2

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 $\$60$?

## 4.4 Gaussian processes

**Definition 4.17** *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 $\Sigma$ (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 4.18** *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 4.19** *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 3

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 $X_3$.

*Solution:* (1) To prove $\{X_t\}$ is Gaussian, we have to prove that any finite linear combination of the $\{X_t\}$ is normal. To do this, let $t_1, ..., t_n \in [0, \infty)$ and let

$a_1, ..., a_n \in \mathbb{R}$. Then

$$\begin{aligned}
\sum_{j=1}^{n} a_j X_{t_j} &= \sum_{j=1}^{n} a_j \left( Z_1 \cos \lambda t_j + Z_2 \sin \lambda t_j \right) \\
&= \sum_{j=1}^{n} (a_j \cos \lambda t_j) Z_1 + \sum_{j=1}^{n} (a_j \sin \lambda t_j) Z_2 \\
&\sim \sum_{j=1}^{n} (a_j \cos \lambda t_j) n(0,1) + \sum_{j=1}^{n} (a_j \sin \lambda t_j) n(0,1) \\
&\sim n \left( 0, \left( \sum_{j=1}^{n} (a_j \cos \lambda t_j) \right)^2 \right) + n \left( 0, \left( \sum_{j=1}^{n} (a_j \sin \lambda t_j) \right) \right) \\
&\sim n \left( 0, \left( \sum_{j=1}^{n} (a_j \cos \lambda t_j) \right)^2 + \left( \sum_{j=1}^{n} (a_j \sin \lambda t_j) \right)^2 \right)
\end{aligned}$$

since $Z_1 \sim n(0,1)$, $Z_2 \sim n(0,1)$ and $Z_1 \perp Z_2$.

Since this arbitrary linear combination is normal, $\{X_t\}$ is Gaussian.

(2) The mean function is

$\mu_X(t) = E[X_t] = E\left[Z_1 \cos \lambda t + Z_2 \sin \lambda t\right] =$

The covariance function is

$$\begin{aligned}
r_X(s,t) = Cov(X_s, X_t) &= Cov\left(Z_1 \cos \lambda s + Z_2 \sin \lambda s, Z_1 \cos \lambda t + Z_2 \sin \lambda t\right) \\
&= Cov\left(Z_1 \cos \lambda s, Z_1 \cos \lambda t\right) + Cov\left(Z_1 \cos \lambda s, Z_2 \sin \lambda t\right) \\
&\quad + Cov\left(Z_2 \sin \lambda s, Z_1 \cos \lambda t\right) + Cov\left(Z_2 \sin \lambda s, Z_2 \sin \lambda t\right) \\
&= \cos \lambda s \cos \lambda t Cov(Z_1, Z_1) + 0 + 0 + \sin \lambda s \sin \lambda t Cov(Z_2, Z_2) \\
&= \cos \lambda s \cos \lambda t Var(Z_1) + \sin \lambda s \sin \lambda t Var(Z_2) \\
&= \cos \lambda s \cos \lambda t + \sin \lambda s \sin \lambda t \quad \text{(since } Z_1, Z_2 \sim n(0,1)\text{)}.
\end{aligned}$$

(3) $Var(X_t) = Cov(X_t, X_t) = r_X(t,t) =$

EXAMPLE 4

Let $\{X_t\}$ be a Poisson process with rate $\lambda$. Find the mean and covariance functions of $\{X_t\}$. Is $\{X_t\}$ Gaussian?

*Solution:* The mean function is

$$\mu_X(t) = E[X_t] = E[\qquad\qquad\qquad] =$$

For the covariance function, assume first that $s \leq t$. Then

$$\begin{aligned} r_X(s,t) = Cov(X_s, X_t) &= Cov(X_s, X_s + (X_t - X_s)) \\ &= Cov(X_s, X_s) + Cov(X_s, X_t - X_s) \\ &= \end{aligned}$$

By the same argument with $s$ and $t$ reversed, if $t \leq s$ then $r_X(s,t) =$

So in general, $r_X(s,t) = \lambda \min(s,t)$.

Last, notice that $X_1 \sim Pois(\lambda)$.

**Theorem 4.20** *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 4.21** *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 4.22** *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

$$\sum_{j=1}^{n} b_j Y_{t_j} = \sum_{j=1}^{n} b_j f(t_j) X_{g(t_j)} = \sum_{j=1}^{n} (b_j f(t_j))\, X_{g(t_j)}$$

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. □

## 4.5 Symmetries and scaling laws

The upshot of the preceding theorem (Theorem 4.22) 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 4.23** *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.

First, let $Y_t = -W_t = f(t)W_{g(t)}$ where $f(t) = -1$ and $g(t) = t$.

The middle two are left as homework exercises (but be warned, Theorem 4.22 doesn't apply to the second process).

For the last one, let $Y_t = aW_{t/a^2} = f(t)W_{g(t)}$ where $f(t) = a$ and $g(t) = \frac{t}{a^2}$.

**Corollary 4.24 (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 4.25 (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**.

## 4.6 Zero sets of Brownian motion

**Definition 4.26** *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 4.27 (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 $y$.

## 4.7 Brownian motion in higher dimensions

**Definition 4.28** *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 $0 < r < R < \infty$.

Define the sets

$$\begin{aligned} A_r &= \{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}|| = r\}; \\ A_R &= \{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}|| = R\}; \\ A &= \{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}|| \in (r, R)\}; \end{aligned}$$

and also let

$$\begin{aligned} T_1 &= \min\{t \geq 0 : W_t \in A_r\}; \\ T_2 &= \min\{t \geq 0 : W_t \in A_R\}; \\ T &= \min\{T_1, T_2\}. \end{aligned}$$

Finally, for $\mathbf{x} \in A$, define $f(\mathbf{x}) = P_{\mathbf{x}}(T_{A_R} < T_{A_r})$ and set $f(\mathbf{x}) = 0$ if $\mathbf{x} \in A_r$ and set $f(\mathbf{x}) = 1$ if $\mathbf{x} \in A_R$.

By symmetry, $f(\mathbf{x}) = g(||\mathbf{x}||)$ for some function $g : \mathbb{R} \to \mathbb{R}$ such that $g(r) = 0$ and $g(R) = 1$.

$f$ has another important property: the value of $f$ at $\mathbf{x}$ is equal to the average value of $f$ along any circle of small radius centered at $\mathbf{x}$:

Therefore $f : A \to \mathbb{R}$ is what is called a **harmonic** function, meaning it satisfies the following equation, which is called the **heat equation** (Google the "Dirichlet problem" or "heat equation" for more on this):

$$\sum_{j=1}^{d} \frac{\partial^2}{\partial x_j^2} f(\mathbf{x}) = 0 \quad \text{for all } \mathbf{x} \in A.$$

To analyze this equation, first observe that for any $x_j$, we can use the Chain Rule to obtain

$$\frac{\partial}{\partial x_j}(||\mathbf{x}||) = \frac{\partial}{\partial x_j}\left(\sqrt{x_1^2 + \ldots + x_d^2}\right) = \frac{1}{2\sqrt{x_1^2 + \ldots + x_d^2}} \cdot 2x_j = \frac{x_j}{\sqrt{x_1^2 + \ldots + x_d^2}} = \frac{x_j}{||\mathbf{x}||}.$$

Therefore

$$\begin{aligned}
0 = \Delta f(\mathbf{x}) &= \sum_{j=1}^{d} \frac{\partial^2}{\partial x_j^2} f(\mathbf{x}) \\
&= \sum_{j=1}^{d} \frac{\partial^2}{\partial x_j^2} g(||\mathbf{x}||) \\
&= \sum_{j=1}^{d} \frac{\partial}{\partial x_j}\left[g'(||\mathbf{x}||)\frac{x_j}{||\mathbf{x}||}\right] \\
&\quad \text{(using the Chain Rule with the above computation)} \\
&= \sum_{j=1}^{d}\left[g''(||\mathbf{x}||)\frac{x_j}{||\mathbf{x}||} \cdot \frac{x_j}{||\mathbf{x}||} + g'(||\mathbf{x}||)\frac{1 \cdot ||\mathbf{x}|| - \frac{x_j}{||\mathbf{x}||}x_j}{||\mathbf{x}||^2}\right] \\
&\quad \text{(Product and Quotient Rules)} \\
&= \sum_{j=1}^{d}\left[\frac{x_j^2 g''(||\mathbf{x}||)}{||\mathbf{x}||^2} + \frac{g'(||\mathbf{x}||)}{||\mathbf{x}||} - \frac{g'(||\mathbf{x}||)x_j^2}{||\mathbf{x}||^3}\right] \\
&= g''(||\mathbf{x}||) + d\frac{g'(||\mathbf{x}||)}{||\mathbf{x}||} - \frac{g'(||\mathbf{x}||)}{||\mathbf{x}||} \\
&\quad \left(\text{since } \sum_{j=1}^{d} x_j^2 = ||\mathbf{x}||^2\right).
\end{aligned}$$

Multiply through by $||\mathbf{x}||$ to obtain

$$0 = ||\mathbf{x}||g''(||\mathbf{x}||) + (d-1)g'(||\mathbf{x}||).$$

Thinking of $||\mathbf{x}||$ as "$t$", this is the second-order ODE

$$0 = tg''(t) + (d-1)g'(t).$$

which has no $g$ in it (only $t$, $g'$ and $g''$); therefore it can be solved with MATH 330 methods:

Integrate $g'(t) = Ct^{1-d}$ to get

$$g(t) = \begin{cases} & \text{if } d = 2 \\ & \text{if } d \geq 3 \end{cases}$$

If you plug in the known values of $g$ (i.e. $g(r) = 0$ and $g(R) = 1$) and solve for the constants (HW), you will obtain:

**Theorem 4.29 (Annular hitting times for higher-dimensional BM)** *Let $\{W_t\}$ be a standard, $d$-dimensional BM. Suppose $X_0 = \mathbf{x}$ where $r \leq ||\mathbf{x}|| \leq R$. Then, if $A_r$ and $A_R$ are the spheres of radius $r$ and $R$ centered at the origin, we have*

$$P_{\mathbf{x}}(T_{A_R} < T_{A_r}) = \begin{cases} \frac{x-r}{R-r} & \text{if } d = 1 \\ \frac{\ln ||\mathbf{x}|| - \ln r}{\ln R - \ln r} & \text{if } d = 2 \\ \frac{r^{2-d} - ||\mathbf{x}||^{2-d}}{r^{2-d} - R^{2-d}} & \text{if } d \geq 3 \end{cases}$$

*In all cases, $P_{\mathbf{x}}(T_{A_r} < T_{A_R}) = 1 - P_{\mathbf{x}}(T_{A_R} < T_{A_r})$.*

## What does this have to do with recurrence and/or transience?

**Dimension 3 (or higher):**

Suppose $r > 0$ is the radius of a small sphere centered at the origin. If a 3-dimensional BM travels to $\mathbf{x}$ with $||\mathbf{x}|| > r$, then

$$\begin{aligned} P_{\mathbf{x}}(T_{A_r} < \infty) &= \lim_{R\to\infty} P_{\mathbf{x}}(T_{A_r} < T_{A_R}) \\ &= 1 - \lim_{R\to\infty} P_{\mathbf{x}}(T_{A_R} < T_{A_r}) \\ &= \\ &= \\ &= \\ &= (\text{something bigger than } 1)^{\text{negative number}} \\ &< 1. \end{aligned}$$

So there is a chance that the BM never comes back to within $r$ of the origin. Thus we say that in dimension $3$ or higher, BM is transient.

**Dimension 2:**

(more interesting) Repeating the above calculation when $d = 2$, we get

$$\begin{aligned} P_{\mathbf{x}}(T_{A_r} < \infty) &= \lim_{R\to\infty} P_{\mathbf{x}}(T_{A_r} < T_{A_R}) \\ &= \\ &= \\ &= \\ &= \end{aligned}$$

This time, it is assured that the BM will return to within $r$ of the origin, so in dimension 2, BM is "neighborhood recurrent", because it returns to any "neighborhood" (i.e. within any positive distance) of where it was.

**BUT:** does a 2-dim'l BM get back exactly to where it was? Suppose a 2-dim'l BM starts at **0** and then travels distance $||\mathbf{x}||$ away. The probability that it returns to **0** is

$$\begin{aligned} P_{\mathbf{x}}(T_{A_0} < \infty) &= \lim_{r \to 0} P_{\mathbf{x}}(T_{A_r} < T_{A_R}) \\ &= 1 - \lim_{r \to 0} P_{\mathbf{x}}(T_{A_R} < T_{A_r}) \\ &= 1 - \lim_{r \to 0} \frac{\ln ||\mathbf{x}|| - \ln r}{\ln R - \ln r} \\ &= \\ &= \\ &= \end{aligned}$$

Therefore, with probability 1, 2-dim'l BMs do <u>not</u> return to where they start, so 2-dim'l BM is "point transient".

**Dimension 1:**

We already proved 1-dim'l BM is point recurrent in Theorem 4.9.

Putting this together, we have shown the following set of facts:

**Theorem 4.30** *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.*

EXAMPLE 6

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?

*Chapter 5*

# Homework exercises

## 5.1 Exercises from Chapter 1

### Exercises from Section 1.3

1. Consider a Markov chain with state space $\mathcal{S} = \{0, 1\}$, where $p = P(0, 1)$ and $q = P(1, 0)$; compute the following in terms of $p$ and $q$:

   a) $P(X_2 = 0 \,|\, X_1 = 1)$
   b) $P(X_3 = 0 \,|\, X_2 = 0)$
   c) $P(X_2 = 1 \,|\, X_0 = 0)$
   d) $P(X_1 = 0 \,|\, X_0 = X_2 = 0)$

2. Continuing with the Markov chain described in Problem 1, suppose the initial distribution is $(\pi_0(0), \pi_0(1)$. In terms of the entries of $\pi_0$, $p$ and $q$, compute $P(X_0 = 0 \,|\, X_1 = 0)$.

3. The weather in a city is always one of two types: rainy or dry. If it rains on a given day, then it is 25% likely to rain again on the next day. If it is dry on a given day, then it is 10% likely to rain the next day. If it rains today, what is the probability it rains the day after tomorrow?

4. Suppose we have two boxes and $2d$ marbles, of which $d$ are black and $d$ are red. Initially, $d$ of the balls are placed in Box 1, and the remainder are placed in Box 2. At each trial, a ball is chosen uniformly from each of the boxes; these two balls are put back in the opposite boxes. Let $X_0$ denote the number of black balls initially in Box 1, and let $X_t$ denote the number of black balls in Box 1 after the $t^{th}$ trial. Find the transition function of the Markov chain $\{X_t\}$.

5. A Markov chain on the state space $\mathcal{S} = \{1, 2, 3, 4, 5\}$ has transition matrix

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

and initial uniform distribution on $\mathcal{S}$.

a) Sketch the directed graph associated to this Markov chain.
b) Find the distribution of $X_2$.
c) Find $P(X_3 = 5 \,|\, X_2 = 4)$.
d) Find $P(X_4 = 2 \,|\, X_2 = 3)$.
e) Find $P(X_4 = 5, X_3 = 2, X_1 = 1)$.
f) Find $P(X_8 = 3 \,|\, X_7 = 1 \text{ and } X_9 = 5)$

6. A dysfunctional family has six members (named Al, Bal, Cal, Dal, Eal, and Fal) who have trouble passing the salt at the dinner table. The family sits around a circular table in clockwise alphabetical order. This family has the following quirks:

- Al is twice as likely to pass the salt to his left than his right.
- Cal and Dal alway pass the salt to their left.
- All other family members pass the salt to their left half the time and to their right half the time.

a) Sketch the directed graph associated to this Markov chain.
b) If Al has the salt now, what is the probability Bal has the salt 3 passes from now?
c) If Al has the salt now, what is the probability that the first time he gets it back is on the 4th pass?
d) If Bal has the salt now, what is the probability that Eal can get it in at most 4 passes?

7. Consider the Markov chain with $\mathcal{S} = \{1, 2, 3\}$ whose transition matrix is

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

where $p \in (0, 1)$ is a constant.

a) Find $P^2$.

b) Show $P^4 = P^2$.

c) Find $P^n$ for all $n \geq 1$.

d) If the initial distribution is $(\frac{2}{5}, \frac{2}{5}, \frac{1}{5})$, find the time $200$ distribution.

e) If the initial distribution is $(\frac{2}{5}, \frac{2}{5}, \frac{1}{5})$, find the time $111111$ distribution.

8. Consider a Markov chain with state space $\{1, 2, 3, 4, 5\}$ and transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{2}{7} & 0 & \frac{5}{7} & 0 & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

a) Compute $P^2$ and $P^3$.

b) If the initial distribution is uniform, find the distributions at times $1, 2$ and $3$.

9. For the Markov chain given in Problem 8, find a distribution $\pi$ on $\mathcal{S}$ with the property that if the initial distribution is $\pi$, then the time $1$ distribution is also $\pi$.

10. Consider Markov chain with $\mathcal{S} = \{0, 1, 2, ...\}$, where for all $x \in \mathcal{S}$, $P(x, x + 1) = \frac{1}{2^x}$ and $P(x, 0) = 1 - \frac{1}{2^x}$.

a) Compute $P(X_8 = 9 \,|\, X_7 = 8)$.

b) Compute $P(X_4 = 7 \,|\, X_2 = 4)$.

c) Compute $P(X_4 = 7 \,|\, X_2 = 5)$.

d) Compute $P(X_6 = 4 \,|\, X_0 = 2)$

e) If the initial distribution $\pi_0$ is uniform on $\{0, 1\}$, compute $\pi_2$.

## Exercises from Section 1.5

11. Consider a Markov chain with state space $\{1, 2, 3\}$ whose transition matrix is

$$\begin{pmatrix} .4 & .4 & .2 \\ .3 & .4 & .3 \\ .2 & .4 & .4 \end{pmatrix}.$$

Find all stationary distributions of this Markov chain.

12. Let $\{X_t\}$ be a Markov chain that has a stationary distribution $\pi$. Prove that if $\pi(x) > 0$ and $x \to y$, then $\pi(y) > 0$.

13. Find all stationary distributions of the Markov chain with transition function

$$P = \begin{pmatrix} 1/9 & 0 & 4/9 & 4/9 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 2/9 & 2/9 & 2/9 & 1/3 & 0 \\ 0 & 0 & 1/9 & 8/9 & 0 \\ 0 & 0 & 7/9 & 2/9 & 0 \end{pmatrix}.$$

*Hint:* The answer is $\pi = \left(\frac{9}{221}, \text{something}, \text{something}, \text{something}, \frac{8}{221}\right)$.

14. Compute all stationary distributions of the Markov chain described in Problem 4, in the situation where $d = 4$.

15. a) Show that the Markov chain introduced in Problem 7 has a unique stationary distribution (and compute this stationary distribution, in terms of $p$).

    b) Is this stationary distribution steady-state? Why or why not?
    *Hint:* The work you did in Problem 7 should be useful in answering this.

16. A transition matrix of a Markov chain is called **doubly stochastic** if its columns add to 1 (recall that for any transition matrix, the rows must add to 1). Find a stationary distribution of a finite state-space Markov chain with a doubly stochastic transition matrix (the way you do this is by "guessing" the answer, and then showing your guess is stationary).

    **Note:** It is useful to remember the fact you prove in this question.

17. Prove Theorem 1.25 from the notes, which goes like this: let $\pi_1, \pi_2, ...,$ be a finite or countable list of stationary distributions for a Markov chain $\{X_t\}$. Let $\alpha_1, \alpha_2, ...$ be nonnegative numbers whose sum is 1, and let $\pi = \sum_j \alpha_j \pi_j$. Prove that the distribution $\pi$ is stationary for $\{X_t\}$.

18. Show that for any $d \times d$ stochastic matrix $P$, 1 is an eigenvalue of $P$ corresponding to eigenvector $(1, 1, 1, ..., 1) \in \mathbb{R}^d$.

    *Hint:* the crux of this question is to get you to remember what eigenvalues and eigenvectors are (you learned about these creatures in MATH 322).

19. Let

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

a) Find the eigenvalues of $P$ by solving $\det(P - \lambda I) = 0$.

b) For each eigenvalue you found in part (a), find a corresponding eigenvector (by finding $\mathbf{v} \neq \mathbf{0}$ such that $P\mathbf{v} = \lambda \mathbf{v}$).

c) Diagonalize $P$ (i.e. write $P = S\Lambda S^{-1}$ where $\Lambda$ is a diagonal matrix whose entries are eigenvalues of $P$, and $S$ is a matrix whose columns are corresponding eigenvectors of $P$).

d) Compute $P^n$ (by multiplying out the formula $P^n = S\Lambda^n S^{-1}$).

e) Compute $\lim_{n\to\infty} P^n$.

20. Let $\{X_t\}$ be a Markov chain with state space $\{1, 2, 3\}$ whose transition matrix is the matrix $P$ given in Problem 19. Based on your work in Problem 19, what do you know about stationary and/or steady-state distributions of $\{X_t\}$?

21. Let $\{X_t\}$ be a Markov chain with state space $\{1, 2, 3, 4\}$ whose transition matrix is

$$P = \begin{pmatrix} \frac{1}{7} & \frac{6}{7} & 0 & 0 \\ \frac{11}{14} & \frac{3}{14} & 0 & 0 \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} \\ 0 & 0 & \frac{2}{5} & \frac{3}{5} \end{pmatrix}.$$

a) Find all stationary distributions of $\{X_t\}$.

b) Does $\{X_t\}$ have a steady-state distribution? Explain.

## Exercises from Section 1.6

22. Consider a Markov chain with state space $\mathcal{S} = \{0, 1\}$, where $p = P(0, 1)$ and $q = P(1, 0)$. (Assume that neither $p$ nor $q$ are either $0$ or $1$.) Compute, for each $n$, the following in terms of $p$ and $q$:

a) $P_0(T_0 = n)$
*Hint:* There are two cases: one for $n = 1$, and one for $n > 1$.

b) $P_1(T_0 = n)$

c) $P_0(T_1 = n)$

d) $P_1(T_1 = n)$

23. For the same Markov chain described in Problem 22, compute these quantities (in terms of $p$ and $q$):

a) $f_{0,1}$
*Hint:* Recall that $f_{0,1} = P_0(T_1 < \infty)$. There are two ways to do this: first, you can add up the values of $P_0(T_1 = n)$ from $n = 1$ to $\infty$; second, you can compute $P_0(T_1 = \infty)$ and use the complement rule.

b) $f_{1,0}$

c) $f_0$ (this means $f_{0,0}$)

d) $f_1$ (this means $f_{1,1}$)

24. For the Markov chain in Problem 8:

a) For each $x \in \mathcal{S}$, compute $P_x(T_0 = 1)$.

b) For each $x \in \mathcal{S}$, compute $P_x(T_0 = 2)$.

c) For each $x \in \mathcal{S}$, compute $P_x(T_0 = 3)$.

25. Consider a Markov chain whose state space is $\mathcal{S} = \{1, 2, 3, 4, 5, 6, 7\}$ and whose transition matrix is

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

a) Find all closed subsets of $\mathcal{S}$.

b) Find all communicating classes.

c) Find the period of each state that belongs to a communicating class.

26. Let $p \in (0, 1)$ be a constant. Consider a Markov chain with state space $\mathcal{S} = \{0, 1, 2, 3, ...\}$ such that $P(x, x+1) = p$ for all $x \in \mathcal{S}$ and $P(x, 0) = 1 - p$ for all $x \in \mathcal{S}$. Explain why this chain is irreducible by showing, for arbitrary states $x$ and $y$, a sequence of steps which could be followed to get from $x$ to $y$.

## Problems from Section 1.7

27. For the Markov chain introduced in Problem 1, compute $E_0(V_{1,3})$.

28. For the Markov chain given in Problem 25:

    a) Determine which states are recurrent and which states are transient.
    b) Compute $f_{x,y}$ for all $x, y \in \mathcal{S}$.
    c) Compute $E_1(V_1)$.

29. Determine whether the Markov chain given in Problem 26 is recurrent or transient.

30. Consider a Markov chain with state space $\mathcal{S} = \{0, 1, 2, 3, ...\}$ and transition function defined by

$$P(x,y) = \begin{cases} \frac{1}{2} & \text{if } x = y \\ \frac{1}{2} & \text{if } x > 0 \text{ and } y = x - 1 \\ \left(\frac{1}{2}\right)^{y+1} & \text{if } x = 0 \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases}.$$

    a) Explain why this Markov chain is irreducible.
    b) Is this chain recurrent or transient?

31. Consider a Markov chain with state space $\mathcal{S} = \{0, 1, 2, 3, ...\}$ and transition function defined by

$$P(x,y) = \begin{cases} \frac{1}{7} & \text{if } y = 0 \\ \frac{2}{7} & \text{if } y \in \{x+2, x+4, x+6\} \\ 0 & \text{otherwise} \end{cases}.$$

    Classify the states of this Markov chain as recurrent or transient, and find all communicating classes (if any).

32. Consider a Markov chain with state space $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ whose transition matrix is

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 & \frac{7}{8} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \end{pmatrix}$$

a) Determine which states are recurrent and which states are transient.

b) Find $f_{x,1}$ for all $x \in \mathcal{S}$.

c) Compute $E_x(V_y)$ for all $x, y \in \mathcal{S}_T$.

33. a) What is the period of the Ehrenfest chain?

    b) What is the period of the Markov chain introduced in Problem 4?

34. Compute the stationary distribution of the Ehrenfest chain (introduced in one of the group presentations), in the situation where $d = 5$.

35. Let $\{X_t\}$ be the Wright-Fisher chain (introduced in one of the group presentations) with $d = 3$. Compute $f_{x,0}$ for all $x \in \mathcal{S}$.

36. Let $\{X_t\}$ be the Wright-Fisher chain with $d = 4$. Compute $E_1(V_2)$.

37. Let $\{X_t\}$ be a Galton-Watson branching chain where each individual has either $0$ or $3$ offspring, each with probability $\frac{1}{2}$. Compute the extinction probability $\eta$.

38. Let $\{X_t\}$ be a Galton-Watson branching chain where the number of offspring of each individual is $Geom(p)$. Compute the extinction probability $\eta$.

    *Hint:* There are two cases, depending on $p$.

39. Let $X_t$ denote the number of people waiting for service at a fast-food restaurant at time $t$. Assume $\{X_t\}$ is modeled by a discrete queuing chain where with probability $\frac{2}{3}$, two customers enter the queue in each time period, and with probability $\frac{1}{3}$, no customers enter the queue in each time period.

    a) If there is initially $1$ person being served, what is the probability that at some point in the future, there will be no one in line?

    b) If there are initially $4$ people in the queue, what is the probability that the queue never empties?

## Exercises from Section 1.8

40. Find the Cesàro limit of the sequence of numbers $\{0, 1, 0, 1, 0, 1, ...\}$ (justify your answer).

41. Compute (directly, without appealing to any stationary distribution), in terms of $p$ and $q$, the mean return time to each state for the Markov chain given in Problem 1.

## Exercises from Section 1.9

42. ($\star$) Complete the proof of Theorem 1.72 by explaining why $P(T < \infty) = 1$.

## Exercises from Section 1.10

43. Consider the irreducible Markov chain with state space $\mathcal{S} = \{1, 2, 3, 4, 5\}$ whose transition matrix is

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

a) Compute the period of this Markov chain.
b) Compute the stationary distribution. Is this distribution steady-state?
c) Describe $P^n$ for $n$ large (there is more than one answer depending on the relationship $n$ and the period $d$).
d) Suppose the initial distribution is uniform on $\mathcal{S}$. Estimate the time $n$ distribution for large $n$ (there are cases depending on the value of $n$).
e) Find $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k$.
f) Find $m_1$ and $m_2$.

44. Fix nonnegative constants $p_0, p_1, ...$ such that $\sum_{y=0}^{\infty} p_y = 1$ and let $X_t$ be a Markov chain on $\mathcal{S} = \{0, 1, 2, \cdots\}$ with transition function $P$ defined by

$$P(x,y) = \begin{cases} p_y & \text{if } x = 0 \\ 1 & \text{if } x > 0, y = x - 1 \\ 0 & \text{else} \end{cases}$$

a) Show this chain is recurrent.
b) Calculate, in terms of the $p_y$, the mean return time to $0$.
c) Under what conditions on the $p_y$ is the chain positive recurrent?
d) Suppose this chain is positive recurrent. Find $\pi(0)$, the value that stationary distribution assigns to state $0$.

e) Suppose this chain is positive recurrent. Find the value the stationary distribution $\pi$ assigns to an arbitrary state $x$.

45. Let $\{X_t\}$ be the Ehrenfest chain with $d = 4$ and $X_0 = 0$ (i.e. there are no particles in the left-hand chamber).

a) Estimate the distribution of $X_t$ when $t$ is large and even.

b) Estimate the distribution of $X_t$ when $t$ is large and odd.

c) Compute the expected amount of time until there are again no particles in the left-hand chamber.

46. Consider a Markov chain on $\mathcal{S} = \{0, 1, 2, 3\}$ with transition matrix

$$P = \begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{5} & 0 & \frac{4}{5} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{5} & 0 & \frac{4}{5} & 0 \end{pmatrix}.$$

a) Compute the Cesàro limit of $P^n$.

b) Compute $m_0$ and $m_2$.

47. Consider a Markov chain $\{X_t\}$ on $\mathcal{S} = \{0, 1, 2, ...\}$ with transition function

$$P(x, y) = \begin{cases} 2^{-y-1} & \text{if } x \leq 3 \\ 1/4 & \text{if } x > 3 \text{ and } y \leq 3 \\ 0 & \text{if } x > 3 \text{ and } y > 3 \end{cases}$$

a) Show the chain is positive recurrent.
*Hint:* Consider a Markov chain $\{Y_t\}$ defined by $Y_t = X_t$ if $X_t \leq 3$ and $Y_t = 4$ if $X_t \geq 4$. Show $\{Y_t\}$ is positive recurrent; why does this imply $\{X_t\}$ is positive recurrent?

b) Find all stationary distributions of $\{X_t\}$.
*Hint:* The stationary distribution of $Y_t$ (from part (a)) tells you something about the stationary distribution of $X_t$.

c) Suppose you start in state 2. How long would you expect it to take for you to return to state 2 for the fifth time?

48. ($\star$) Suppose a fair die is thrown repeatedly. Let $S_n$ represent the sum of the first $n$ throws. Compute

$$\lim_{n\to\infty} P(S_n \text{ is a multiple of } 13),$$

justifying your reasoning.

49. ($\star$) Your professor owns $3$ umbrellas, which at any time may be in his office or at his home. If it is raining when he travels between his home and office, he carries an umbrella (if possible) to keep him from getting wet.

    a) If on every one of his trips, the probability that it is raining is $p$, what is the long-term proportion of journeys on which he gets wet?
    b) What $p$ as in part (a) causes the professor to get wet most often?
    c) In the worst-case scenario described in part (b), on what fraction of his trips will he get wet?

50. ($\star$) A knight is placed in one corner of a chess board. At each step, the knight chooses a square uniformly from the squares that the knight can legally move to (i.e. two squares in one direction, and one to the side). Compute the expected number of moves the knight will make before returning to its starting position.

## 5.2 Exercises from Chapter 2

### Exercises from Section 2.2

51. In each part of this problem, you are given a set $\Omega$, a $\sigma$-algebra $\mathcal{F}$, and a r.v. $X : \Omega \to \mathbb{R}$. Determine if the given r.v. $X$ is $\mathcal{F}$-measurable.

    a) $\Omega = \{1, 2, 3, 4\}$; $\mathcal{F}$ is generated by the partition of $\Omega$ into even and odd numbers; $X(\omega) = \omega^2$.
    b) $\Omega = \{1, 2, 3, 4\}$; $\mathcal{F}$ is generated by the partition of $\Omega$ into even and odd numbers; $X(\omega) = \frac{1}{4}\omega^4 - 10\omega^2 + 30\omega$.
    c) $\Omega = [0, 1] \times [0, 1]$; $\mathcal{F}$ is the $\sigma-$algebra of vertical sets (i.e. sets of the form $A \times [0, 1]$); $X(x, y) = xy$.
    d) $\Omega = [0, 1] \times [0, 1]$; $\mathcal{F}$ is the $\sigma-$algebra of vertical sets (i.e. sets of the form $A \times [0, 1]$); $X(x, y) = y^2 - y + 3$.
    e) $\Omega = [0, 1] \times [0, 1]$; $\mathcal{F}$ is the $\sigma-$algebra of vertical sets (i.e. sets of the form $A \times [0, 1]$); $X(x, y) = x^3 - x$.

52. Suppose you are betting on fair coin flips (as usual, you win if you flip heads, and lose if you flip tails), and that you implement Strategy 3 as described in the notes (bet $\$1$ on the first flip; afterwards, bet $\$2$ if you lost the previous

flip and $\$1$ if you won the previous flip). If the first eight flips are H T T H T T H H, compute the amount you have won or lost in the first eight flips.

53. Suppose you are betting on fair coin flips (as usual, you win if you flip heads, and lose if you flip tails), and that you implement Strategy 4 as described in the notes. If your initial bankroll is $\$100$, compute the expected amount of your bankroll after $3$ flips.

54. Suppose you are betting on fair coin flips (as usual, you win if you flip heads, and lose if you flip tails), and you implement a strategy described as follows: on the first flip, bet $1$. On even numbered flips (the second, fourth, sixth, etc.), bet $3$ if you won the previous flip, and bet $1$ if you lost the previous flip. On odd numbered flips (other than the first flip), bet $2$ if the preceding two flips were the same, and bet $1$ if the preceding two flips were different.

    a) Let $B_t$ be the size of your bet on the $t^{th}$ flip. Define $B_t$ using mathematical notation.
    b) Suppose the results of the first ten flips are H T T H H T H H H T. Assuming $X_0 = 0$, compute $(B \cdot X)_t$ for $0 \leq t \leq 10$.

## Exercises from Section 2.3

55. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ have the uniform distribution and suppose $\mathcal{F}$ is the $\sigma-$algebra generated by the partition $\{\{1, 2\}, \{3, 4, 5\}, \{6\}$ of $\Omega$. Let $X$ be the random variable defined by $X(1) = 5$, $X(2) = X(3) = X(4) = 1$, $X(5) = X(6) = 9$. Compute $E[X|\mathcal{F}]$.

56. Let $\Omega = [0, 1] \times [0, 1]$ have the uniform distribution, and let $X : \Omega \to \mathbb{R}$ be $X(x, y) = x^2 y + x$. Compute $E[X|\mathcal{F}]$, where $\mathcal{F}$ is the $\sigma$-algebra of horizontal sets (i.e. sets of the form $[0, 1] \times B$).

57. Suppose $\{X_t\}_{t \in \{0,1,2,...\}}$ is a stochastic process in which you flip a coin that flips heads with probability $\frac{1}{3}$ and tails with probability $\frac{2}{3}$. Let $\{\mathcal{F}_t\}$ be the natural filtration of $\{X_t\}$. Let $X$ be a random variable defined by setting

$$X = \begin{cases} 0 & \text{if the first three flips are heads} \\ 10 & \text{if the first two flips are heads but the third is tails} \\ 4 & \text{if the first flip is heads but the second flip is tails} \\ -7 & \text{if the first flip is tails but the second and third flips are heads} \\ -1 & \text{if the first flip is tails and the second and third flips have} \\ & \qquad \text{opposite results} \\ 3 & \text{if the first three flips are tails} \end{cases}$$

Compute $E[X|\mathcal{F}_1]$ and $E[X|\mathcal{F}_2]$.

## Exercises from Section 2.4

58. Let $\{X_t\}$ be the Wright-Fisher chain (introduced in a group presentation). Prove that $\{X_t\}$ is a martingale.

59. Let $\{X_t\}$ be the Pólya urn model. For each $t$, let $M_t$ be the fraction of balls in the urn which are red. Prove that $\{M_t\}$ is a martingale.

60. ($\star$) Modify the Pólya urn model so that you add $c \geq 2$ balls of the color you most recently drew to the urn after each draw (instead of adding one marble of the color you drew). Is the $\{M_t\}$ described in Problem 59 still a martingale?

## Exercises from Section 2.5

61. Prove Lemma 2.23 from the notes, which says that for a simple, unbiased random walk, $\mu = 0$ and $\sigma^2 = p + q$.

62. ($\star$) Prove the second part of Lemma 2.27 from the notes, which says that if $\{X_t\}$ is an irreducible, simple random walk and $Z_t = (X_t - t\mu)^2 - t\sigma^2$, then $\{Z_t\}$ is a martingale.

63. ($\star$) Prove Wald's Third Identity (this is Theorem 2.31 in the notes).

64. ($\star$) Finish the proof of Gambler's Ruin by writing out the case where $a < x$.

65. A gambler makes a series of independent $\$1$ bets. He decides to quit betting as soon as his net winnings reach $\$25$ or his net losses reach $\$50$. Suppose the probabilities of his winning and losing each bet are each equal to $\frac{1}{2}$.

    a) Find the probability that when he quits, he will have lost $\$50$.
    b) Find the expected amount he wins or loses.
    c) Find the expected number of bets he will make before quitting.

66. A typical roulette wheel has 38 numbered spaces, of which 18 are black, 18 are red, and 2 are green. A gambler makes a series of independent $\$1$ bets, betting on red each time (such a bet pays him $\$1$ if the ball in the roulette wheel ends up on a red number). He decides to quit betting as soon as his net winnings reach $\$25$ or his net losses reach $\$50$.

    a) Find the probability that when he quits, he will have lost $\$50$.
    b) Find the expected amount he wins or loses.
    c) Find the expected number of bets he will make before quitting.

67. Suppose two friends, George the Genius and Ichabod the Idiot, play a game that has some elements of skill and luck in it. Because George is better at the game than Ichabod, George wins $55\%$ of the games they play and Ichabod wins the other $45\%$ (the result of each game is independent of each other game). Suppose George and Ichabod both bring $\$100$ to bet with, and they agree to play until one of them is broke.

   a) Suppose George and Ichabod wager $\$1$ on each game. What is the probability that George ends up with all the money?
   b) Suppose George and Ichabod wager $\$5$ on each game. What is the probability that George ends up with all the money?
   c) Suppose George and Ichabod wager $\$25$ on each game. What is the probability that George ends up with all the money?
   d) Suppose George and Ichabod wager $\$100$ on each game. What is the probability that George ends up with all the money?
   e) Based on the answers to parts (a),(b) and (c), determine which of the following statements is true:

      **Statement I:** The more skilled player benefits when the amount wagered on each game increases.

      **Statement II:** The more skilled player is harmed when the amount wagered on each game increases.
   f) Suppose you had $\$1000$ and needed $\$2000$ right away, and you therefore decided to go to a casino and turn your $\$1000$ into $\$2000$ by gambling on roulette. In light of your answer to the previous question, which of these strategies gives you the highest probability of ending up with $\$2000$: betting $\$1000$ on red on one spin of the wheel, or betting $\$1$ on red repeatedly, trying to work your way up to $\$2000$ without going broke first?

68. Consider an irreducible, simple random walk $X_t$ starting at zero, where $r = 0$.

   a) Find the probability that $X_t = -2$ for some $t > 0$.
   b) Find $p$ such that $P(X_t = 4 \text{ for some } t > 0) = \frac{1}{2}$.

## Exercises from Section 2.6

69. Prove Lemma 2.36 from the notes.

70. Let $\{X_t\}$ be an irreducible birth-death chain with $\mathcal{S} = \{0, 1, 2, 3, ...\}$. Show that if for all $x \geq 1$, $p_x \leq q_x$, then the chain is recurrent.

71. Let $\{X_t\}$ be an irreducible birth-death chain with $\mathcal{S} = \{0, 1, 2, 3, ...\}$ such that

$$\frac{q_x}{p_x} = \left(\frac{x}{x+1}\right)^2 \qquad \text{for all } x \geq 1.$$

   a) Is this chain recurrent or transient?

   b) Compute $f_{x,0}$ for all $x \geq 1$. *Hint:* $\sum\limits_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.

72. Consider a birth and death chain on $\mathcal{S} = \{0, 1, 2, ...\}$ with

$$p_x = \frac{1}{2^{x+1}}\,\forall x;\; q_x = \frac{1}{2^{x-1}}\,\forall x > 1;\; q_1 = \frac{1}{2}.$$

Show this chain is positive recurrent, find the stationary distribution, and find the mean return time to state 2.

73. Compute the stationary distribution of the Ehrenfest chain, for arbitrary $d$.

74. Compute all stationary distributions of the Markov chain described in Problem 4, for arbitrary $d$.

*Hint:* You will need the following identity, which can be assumed without proof:

$$\sum_{j=0}^{d} \begin{pmatrix} d \\ j \end{pmatrix}^2 = \begin{pmatrix} 2d \\ d \end{pmatrix}.$$

## 5.3 Exercises from Chapter 3

### Exercises from Section 3.2

75. Without using the binomial theorem, prove that

$$\lim_{n\to\infty} \left(1 + \frac{t}{n}\right)^n = e^t.$$

*Hint:* Rewrite the expression inside the limit using natural exponentials and logarithms, then use L'Hôpital's Rule.

76. Consider a continuous-time Markov chain $\{X_t\}$ with with state space $\{1, 2, 3\}$ and infinitesimal matrix

$$Q = \begin{pmatrix} -5 & 3 & b \\ 4 & -6 & 2 \\ 2 & 1 & -3 \end{pmatrix}.$$

where $b$ is a constant.

a) What is $b$?

b) Compute the jump matrix $\Pi$.

c) Compute the holding rate of state $2$.

d) Suppose $X_0 = 2$. What is the probability that when the chain jumps, the next state is $3$?

e) Suppose $X_0 = 3$. What is the probability that $X_t = 3$ for all $t \in [0, 4]$?

f) Suppose $X_0 = 3$. What is the expected amount of time before the first jump?

77. Let $\{X_t\}$ be the CTMC of Problem 76.

a) Compute the time $t$ transition matrix $P(t)$.

b) Compute the probability that $X_5 = 2$ given that $X_2 = 2$.

78. Consider a continuous-time Markov chain $\{X_t\}$ with state space $\mathcal{S} = \{1, 2, 3\}$ with holding rates $q_1 = q_2 = 1, q_3 = 3$ and jump probabilities $\pi_{13} = \frac{1}{3}$, $\pi_{23} = \frac{1}{4}$ and $\pi_{31} = \frac{2}{5}$.

a) Use linear approximation to estimate $P_{31}(.001)$ and $P_{22}(.06)$.

b) What is the probability that $X_t = 2$ for all $t < 4$, given that $X_0 = 2$?

c) What is the probability that your first two jumps are first to state $3$ and then to state $2$, given that you start in state $1$?

79. Let $\{X_t\}$ be a CTMC with time $t$ transition matrix

$$P(t) = \frac{1}{22}\begin{pmatrix} 12 - 11e^{-13t/2} + 21e^{-11t/2} & 4 + 11e^{-13t/2} - 15e^{-11t/2} & b(t) \\ 12 - 33e^{-13t/2} + 21e^{-11t/2} & 4 + 33e^{-13t/2} - 15e^{-11t/2} & 6 - 6e^{-11t/2} \\ 12 + 44e^{-13t/2} - Ke^{-11t/2} & 4 - 44e^{-13t/2} + 40e^{-11t/2} & 6 + 16e^{-11t/2} \end{pmatrix},$$

where $b(t)$ is a function and $K$ is a constant.

a) Compute $b(t)$.

b) Compute $K$.

c) Compute the infinitesimal matrix $Q$.

d) Compute the jump matrix $\Pi$.

## Exercises from Section 3.3

80. Consider the CTMC $\{X_t\}$ from Problem 76.

    a) Write out the system of differential equations which constitute the backward equation of $\{X_t\}$.
    b) Write out the system of differential equations which constitute the forward equation of $\{X_t\}$.

81. ($\star$) In this problem, we prove Theorem 3.31, which asserts that a CTMC $\{X_t\}$ satisfies the forward equation.

    a) Take a look at this equation (I hope you can convince yourself that this equation is true):
    $$P_{x,y}(t) = P_x(X_t = y) = \sum_{n=0}^{\infty} \sum_{z \neq y} P_x(J_n \leq t < J_{n+1}, X_{J_{n-1}} = z, X_{J_n} = y).$$
       i. In this equation, describe in English what the $n$ is referring to.
       ii. In this equation, describe in English what the $z$ is referring to.
    b) Using the time reversal identity proven in Lemma 3.30, prove
    $$\begin{aligned} &P_x(J_n \leq t < J_{n+1} \,|\, X_{J_{n-1}} = z, X_{J_n} = y) \\ &= q_x \int_0^t e^{-q_y s} \frac{q_z}{q_x} P_x(J_{n-1} \leq t - s < J_n \,|\, X_{J_{n-1}} = z) \, ds. \end{aligned}$$
    c) Use the multiplication principle and substitute in the formula you found in part (b) to the equation from part (a) to derive the following "forward integral equation":
    $$P_{x,y}(t) = \delta_{x,y} e^{-q_x t} + \int_0^t \sum_{z \neq x} P_{x,z}(t-s) q_{zy} e^{-q_y s} \, ds$$
    d) Perform the $u$-sub $u = t - s$ in the forward integral equation of part (c) and simplify what you get to obtain
    $$P_{x,y}(t) = \delta_{x,y} e^{-q_y t} + e^{-q_y t} \int_0^t \sum_{z \neq x} P_{x,z}(u) q_{zy} e^{q_y u} \, du.$$
    e) Explain why you know from the formula of part (d) that $P_{x,y}$ is a differentiable function of $t$.
    f) Differentiate both sides of the equation in (d), and rewrite the equation you obtain to get the forward equation
    $$P'_{x,y}(t) = \sum_{z \in \mathcal{S}} P_{x,z}(t) q_{zy}.$$
    *Hint:* This should resemble the computation done in the proof of Theorems 3.24 and 3.29.

## Exercises from Section 3.4

82. Consider the CTMC $\{X_t\}$ from Problem 76.
    a) Compute the stationary distribution of $\{X_t\}$.
    b) Compute the mean return time to each state.
83. Compute the stationary distribution of the CTMC described in Problem 78.
84. ($\star$) (In this problem, we prove Theorem 3.36 from the lecture notes.) Suppose $\{X_t\}$ is a continuous-time Markov chain with finite state space $\mathcal{S}$ and infinitesimal matrix $Q$.
    a) Prove that if $\pi$ is stationary (i.e. $\pi P(t) = \pi$ for all $t \geq 0$), then $\pi Q = \mathbf{0}$.
    b) Prove that if $\pi Q = \mathbf{0}$, then $\pi$ is stationary.
85. Prove Theorem 3.37 from the lecture notes.
86. Suppose $\{X_t\}$ is a continuous-time Markov chain with state space $\{1, 2, 3, 4\}$ and time $t$ transition matrix

$$P(t) = \frac{1}{9}\begin{pmatrix} 1 + 6te^{-3t} + 8e^{-3t} & 6 - 6e^{-3t} & 2 - 6te^{-3t} - 2e^{-3t} \\ 1 - 3te^{-3t} - e^{-3t} & 6 + 3e^{-3t} & 2 + 3te^{-3t} - 2e^{-3t} \\ 1 + 6te^{-3t} - e^{-3t} & 6 - 6e^{-3t} & 2 - 6te^{-3t} + 7e^{-3t} \end{pmatrix}.$$

    a) Compute the infinitesimal matrix of this process.
    b) What is the probability that $X_2 = 1$, given that $X_0 = 1$?
    c) What is the probability that $X_t = 1$ for all $t < 2$, given that $X_0 = 1$?
    d) Compute the steady-state distribution $\pi$.
    e) Compute the mean return time to each state.
    f) Suppose you let time pass from $t = 0$ to $t = 1,200,000$. What is the expected amount of time in this interval for which $X_t = 3$?
    g) Suppose $X_0 = 2$. What is the expected amount of time spent in state $3$ before the first time the chain returns to state $2$?
    h) Suppose $X_0 = 2$. What is the expected amount of time spent in state $3$ before the eleventh time the chain returns to state $2$?
87. Consider a continuous-time Markov chain $\{X_t\}$ with with state space $\{1, 2, 3\}$ and infinitesimal matrix

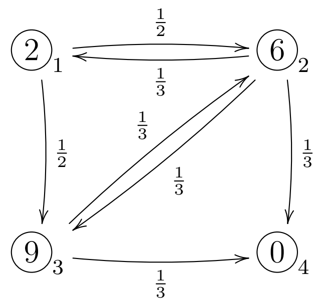
$$Q = \begin{pmatrix} -4 & 1 & 3 \\ 0 & -1 & 1 \\ 0 & 2 & -2 \end{pmatrix}.$$

a) Classify the states as recurrent or transient.

b) Are the recurrent states positive recurrent or null recurrent? Explain.

c) Find all stationary distributions of $\{X_t\}$. Are any of them steady-state?

88. Consider a CTMC $\{X_t\}$, whose directed graph is as given below (the fact that the holding rate of state $4$ is $0$ means that once you are in state $4$, you stay in state $4$ forever):




Compute $E_1(T_4)$.

*Hint:* For $i = 1, 2, 3$, let $k_i = E_i(T_4)$. Set up a system of equations that will enable you to solve for all of the $k_i$.

## Exercises from Section 3.5

89. ($\star$) Consider a pure death process on $\{0, 1, 2, ...\}$ (i.e. a birth-death CTMC with $\lambda_x = 0$ for all $x \in \mathcal{S}$).

a) Write the forward equation of this process.

b) Find $P_{x,x}(t)$.

c) Solve the differential equation from part (a) to obtain a recursive formula for $P_{x,y}(t)$ in terms of $P_{x,y+1}(t)$.

d) Find $P_{x,x-1}(t)$.

90. Consider a birth-death process $\{X_t\}$ with $\mathcal{S} = \{0, 1, 2, 3, ...\}$ where $\lambda_x = \lambda x$ and $\mu_x = \mu x$ for constants $\lambda, \mu \geq 0$.

a) Write the forward equation of this process.

b) Let $g_x(t) = E_x(X_t)$. Use the forward equation to show $g_x'(t) = (\lambda - \mu)g_x(t)$.

c) Based on part (b), derive a formula for $g_x(t)$.

d) Compute $E_0(X_8)$.

91. Consider a birth-death process $\{X_t\}$ on $\{0, 1, 2, 3, ...\}$ whose death rates are given by $\mu_x = x$ for all $x \in \mathcal{S}$.

a) Determine whether the process is transient, null recurrent or positive recurrent, if the birth rates are $\lambda_x = x + 1$ for all $x \in \mathcal{S}$.

b) Determine whether the process is transient, null recurrent or positive recurrent, if the birth rates are $\lambda_x = x + 2$ for all $x \in \mathcal{S}$.

92. Suppose $A_1, ..., A_d$ are independent exponential r.v.s with respective parameters $\lambda_1, ..., \lambda_d$. Prove that $M = \min(A_1, ..., A_d)$ is exponential with parameter $\lambda_1 + ... + \lambda_d$.

    *Hint:* This is a transformation problem; the first step is to compute $F_M(m) = P(M \leq m)$. It is easiest to compute this probability by computing the probability of its complement.

93. ($\star$) Suppose $d$ particles are distributed into two boxes, A and B. Each particle in box A remains in that box for a random length of time that is exponentially distributed with parameter $\mu$ before moving to box B. Each particle in box B remains in that box for a random length of time that is exponentially distributed with parameter $\lambda$ before moving to box A. All particles act independently of one another. For each $t \geq 0$, let $X_t$ be the number of particles in box A at time $t$. Then $\{X_t\}$ is a birth-death process on $\mathcal{S} = \{0, 1, 2, ..., d\}$.

    a) This setup be thought of as a continuous version of what discrete-time Markov chain?

    b) Find the birth and death rates.

    c) Find $P_{x,d}(t)$ for all $x \in \mathcal{S}$. *Hint:* Think of each particle as generating its own CTMC, where state zero corresponds to being in box B and state 1 corresponds to being in box A. This is a two-state CTMC, so its transition probabilities were derived in class. From these transition probabilities, you can get the probability that any one fixed particle is in box A at time $t$. Multiply these together to get $P_{x,d}(t)$.

    d) Find $E_x(X_t)$. *Hint:* Write $X_t = A_t + B_t$ where $A_t$ is the number of particles in box A that started in box A and $B_t$ is the number of particles in box A at time $t$ that started in box B. If $X_0 = x$, then $A_t$ and $B_t$ are both binomial, defined in terms of $x$ and the transition function of the two-state birth-death process described in the hint for part (c).

    e) Compute the steady-state distribution for this process; identify this distribution as a common r.v. (stating the parameters).

    f) Verify that as $t \to \infty$, $E_x(X_t)$ converges to the expected value of the steady-state distribution.

## Exercises from Section 3.6

94. Let $\{X_t\}$ be a continuous-time branching process with $p = \frac{2}{3}$.

a) Compute the extinction probability $\eta$.

b) Compute $f_{5,0}$.

95. If $\{X_t\}$ is a continuous-time branching process with extinction probability $\eta = \frac{1}{2}$, what is the probability that each particle splits into two "children" (as opposed to dying)?

96. Suppose customers call a technical support line according to a Poisson process with parameter $\lambda > 0$. They are provided with technical support by $N$ agents where $N$ is a positive integer ($N$ is a constant, not a r.v.). Suppose that the amount of time it takes an agent to solve a customer's problem is exponentially distributed with parameter $\mu$ (and that these times are independent of the Poisson process and all independent of one another). Last, assume that whenever there are more than $N$ customers calling the technical support line, the excess customers get placed on hold until one of the $N$ agents is available. Let $X_t$ represent the number of people on the phone with technical support (including those on hold) at time $t$. $\{X_t\}$ is called the $N-$*server queue* or the $(M/M/N)-$*queue*.

    a) Explain why $\{X_t\}$ is a birth and death process.

    b) Find the birth and death rates of $\{X_t\}$.

    c) Show that $\lambda < N\mu$ if and only if $\{X_t\}$ is positive recurrent.

    d) Show that $\lambda > N\mu$ if and only if $\{X_t\}$ is transient.

## 5.4 Exercises from Chapter 4

### Exercises from Section 4.1

95. Suppose $\{W_t\}$ is a Brownian motion with parameter $\sigma^2 = 3$.

    a) Find $P(W_4 \geq 1)$.

    b) Find $P(W_9 - W_2 \leq -2)$.

    c) Find $P(W_7 > W_5)$.

    d) Find the variance of $W_8$.

    e) Find $Cov(W_3, W_7)$.

    f) Find $Var(W_8 + W_9)$.

## Exercises from Section 4.2

96. Suppose $\{W_t\}$ is a Brownian motion with parameter $\sigma^2 = 5$. Compute $P(W_8 < 2 \,|\, W_2 = W_1 = 3)$.

97. Let $\{W_t\}$ be a Brownian motion with parameter $\sigma^2$ and let $M(t) = \max\{W_s : 0 \le s \le t\}$. Show $M(t)$ is a continuous r.v. (this implies $M(t) > 0$ with probability one) and find the density function of $M(t)$.

98. ($\star$) Let $\{W_t\}$ be a standard Brownian motion, and let $0 < t_0 < t_1$. Show

$$P(W_t = 0 \text{ for some } t \in (t_0, t_1)) = \frac{2}{\pi} \arctan \sqrt{\frac{t_1 - t_0}{t_0}}.$$

*Hint:* Condition on the value of $W_{t_0}$ and use the result of Problem 97.

99. Let $\{W_t\}$ be a standard Brownian motion, and let $L$ be the largest time $t \in [0, 1]$ such that $W_t = 0$.

    a) Compute the density function of $L$, and use a computer or graphing calculator to graph this density function.
    *Hint:* Use the result of Problem 99.

    b) Based on the graph you see, describe qualitatively what is true about $L$ (i.e. which values of $L$ are most likely)?

## Exercises from Section 4.3

100. Let $\{W_t\}$ be a standard Brownian motion. For each $t \ge 0$, let $X_t = W_t^2 - t$. Prove $\{X_t\}$ is a martingale.

101. You own one share of stock whose price is approximated by a Brownian motion with parameter $\sigma^2 = 10$ (time $t = 1$ here corresponds to the passage of one day). You bought the stock when its price was $\$15$, but now it is worth $\$25$.

    a) Suppose you decide to sell the stock when the price of the stock next reaches either $\$28$ or $\$20$:

        i. What is the probability you sell the stock for $\$28$?
        ii. What is the expected amount you will sell the stock for?
        iii. How much longer should you expect to hold the stock before selling?

b) Suppose instead that you decide to sell the stock the next time its price hits $\$15$ or after ten days, whichever happens first. What is the probability that when you sell your stock, you will have to sell it for $\$15$?

102. ($\star$) Prove Theorem 4.15 in the lecture notes, which says that if $\{X_t\}$ is a BM with drift, then $\{M_t\}$ is a martingale, where $M_t = \exp\left(\frac{-2\mu}{\sigma^2} X_t\right)$.

103. ($\star$) Prove Corollary 4.16 in the lecture notes, in which escape probabilities for BM with drift are derived.

104. Suppose $\{X_t\}$ is a Brownian motion with parameter $\sigma^2 = 4$ and drift parameter $\mu = 5$.

a) Find $P(X_1 \geq 6)$.
b) Find $P(X_9 - X_7 \leq 3)$.
c) Find $P(X_4 > 15 \,|\, X_2 = 7, X_1 = -1)$.
d) Find $P(X_7 > X_5)$.
e) Find the mean and variance of $X_8$.
f) Find $Cov(X_{11}, X_{16})$.
g) Find $Var(X_2 + X_5)$.

105. Suppose that you own a collectible item whose value at time $t$ is modeled by a Brownian motion with drift with $\sigma^2 = 2$ and $\mu = \frac{1}{5}$. The item is presently valued at $\$30$, and you plan to sell the item when the value of the item reaches $\$45$ or $\$20$, whichever happens first.

a) What is the probability that you sell the item for $\$45$?
b) What is the expected value at which you will sell the item?

## Exercises from Section 4.4

106. Prove Theorem 4.21 in the lecture notes (which says that the mean and covariance functions of a BM with drift are $\mu_X(t) = \mu t$ and $r_X(s,t) = \sigma^2 \min(s,t)$).

107. Let $\{W_t\}$ be standard Brownian motion and let $X_t = (W_t)^2$ for all $t$.

a) Is $\{X_t\}$ a Gaussian process? Explain your answer.
b) Find the mean function of $\{X_t\}$.
c) Find the joint moment generating function of $W_s$ and $W_t$.
d) Use your answer to part (b) to find $E[W_s^2 W_t^2]$.
e) Find the covariance function of $\{X_t\}$.

108. Let $\{W_t\}$ be a Brownian motion with parameter $\sigma^2$ and let $a \leq s$ and $a \leq t$. Prove that
$$E[(W_s - W_a)(W_t - W_a)] = \sigma^2 \min(s-a, t-a).$$

## Exercises from Section 4.5

109. Let $\{W_t\}$ be a standard Brownian motion. Let $V_0 = 0$ and for each $t > 0$, let $V_t = tW_{1/t}$. Prove that $\{V_t\}$ is a standard Brownian motion.

110. Fix $s \geq 0$ and suppose that $\{W_t\}$ is a standard Brownian motion. For each $t \geq 0$, let $X_t = W_{t+s} - W_s$. Prove that $\{X_t\}$ is a standard Brownian motion.

111. Prove that if $\{W_t\}$ and $\{\widehat{W}_t\}$ are independent Brownian motions with respective parameters $\sigma^2$ and $\widehat{\sigma}^2$, then for any constants $b_1$ and $b_2$, the process $\{b_1 W_t + b_2 \widehat{W}_t\}$ is also a Brownian motion. Find its parameter in terms of $b_1$, $b_2$, $\sigma$ and $\widehat{\sigma}$.

*Note:* The result of this problem generalizes: any linear combination of a finite number of independent BMs is also a BM (although you don't have to prove this).

## Exercises from Section 4.7

112. In the lecture, we saw that for a 2-dimensional Brownian motion, the function $g$ described in Section 4.7 had the form
$$g(t) = C \ln t + D$$
for unknown constants $C$ and $D$. Use the fact that $g(r) = 0$ and $g(R) = 1$ to solve for $C$ and $D$, and therefore write $g$ in terms of $r$ and $R$. (You should get the formula stated in Theorem 4.29.)

113. In the lecture, we saw that for a $d$-dimensional Brownian motion where $d \geq 3$, the function $g$ described in Section 4.7 had the form
$$g(t) = \frac{C}{2-d} t^{2-d} + D$$
for unknown constants $C$ and $D$. Use the fact that $g(r) = 0$ and $g(R) = 1$ to solve for $C$ and $D$, and therefore write $g$ in terms of $r$ and $R$. (You should get the formula stated in Theorem 4.29.)

114. Let $\{W_t\}$ be a standard 2-dimensional Brownian motion with $W_0 = (3, 4)$.

a) What is the probability that $W_t = (0, 0)$ for some $t > 0$?

b) What is the probability that $W_t \in \{(x,y) : x^2 + y^2 < 1\}$ for some $t > 0$?

c) What is the probability that $W_t$ strikes the circle $\{(x,y) : x^2 + y^2 = 4\}$ before it strikes the circle $\{(x,y) : x^2 + y^2 = 49\}$?

115. Suppose that the position of a particle of pollen suspended in a liquid is modeled by a standard 3-dimensional Brownian motion, and that at time 4, the pollen is at position $(1, 2, 3)$.

a) What is the probability that the pollen particle eventually reaches $(0, 4, -1)$?

b) What is the probability that the pollen particle strikes the sphere of radius 4 centered at the origin before it strikes the sphere of radius 2 centered at the origin?

116. Suppose $\{W_t\}$ is a standard 5-dimensional Brownian motion with

$$W_0 = (1, 2, 1, -3, 1).$$

What is the probability that $||W_t|| = 2||W_0||$ before $||W_t|| = \frac{1}{2}||W_0||$?

117. ($\star$) Let $\{W_t\}$ and $\{\widehat{W}_t\}$ be independent, standard Brownian motions and let $a$ be a positive constant.

a) Prove that $P\left(W_t = a\widehat{W}_t \text{ for infinitely many } t\right) = 1$.

b) What is the probability that $W_t = \widehat{W}_t + a$ for infinitely many $t$? Prove your answer.

118. ($\star$) Let $\{W_t\}, \{\widehat{W}_t\}, \{\widetilde{W}_t\}$ be independent, standard BMs.

a) Is $P\left(W_t = \widehat{W}_t \text{ for infinitely many } t\right) = 1$? Why or why not?

b) Is $P\left(W_t = \widehat{W}_t = \widetilde{W}_t \text{ for infinitely many } t\right) = 1$? Why or why not?

119. ($\star$) Let $\{(X_t, Y_t)\}$ be a standard 2-dimensional Brownian motion. Let $T = \min\{t : X_t = 1\}$. Compute the density function of $Y_T$, and identify $Y_T$ as a common random variable.

*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 $f_X(x)$ | $EX$ | Var($X$) | PGF $G_X(t)$ <br> MGF $M_X(t)$ |
|---|---|---|---|---|
| uniform on $\{1, ..., n\}$ | $f(x) = \frac{1}{n}$ for $x = 1, 2, ..., n$ | $\frac{n+1}{2}$ | $\frac{n^2-1}{12}$ | $G_X(t) = \frac{t(t^n-1)}{n(t-1)}$ <br> $M_X(t) = \frac{e^t(e^{nt}-1)}{n(e^t-1)}$ |
| binomial$(n, p)$ | $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, ..., n$ | $np$ | $np(1-p)$ | $G_X(t) = (1-p+pt)^n$ <br> $M_X(t) = (1-p+pe^t)^n$ |
| $Geom(p)$ <br> $0 < p < 1$ | $f(x) = p(1-p)^x$ for $x = 0, 1, 2, ...$ | $\frac{1-p}{p}$ | $\frac{1-p}{p^2}$ | $G_X(t) = \frac{p}{1-(1-p)t}$ <br> $M_X(t) = \frac{p}{1-(1-p)e^t}$ |
| negative binomial $NB(r, p)$ | $f(x) = \binom{r+x-1}{x} p^r (1-p)^x$ for $x = 0, 1, 2, ...$ | $r\frac{1-p}{p}$ | $r\frac{1-p}{p^2}$ | $G_X(t) = \left(\frac{p}{1-(1-p)t}\right)^r$ <br> $M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$ |
| $Pois(\lambda)$ | $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, 2, ...$ | $\lambda$ | $\lambda$ | $G_X(t) = e^{\lambda(t-1)}$ <br> $M_X(t) = e^{\lambda(e^t-1)}$ |
| 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 $f_X(x)$ <br> DISTRIBUTION FUNCTION $F_X(x)$ | EXPECTED VALUE $EX$ <br> VARIANCE $Var(X)$ <br> MGF $M_X(t)$ |
|---|---|---|
| uniform on $[a,b]$ | $f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{else} \end{cases}$ <br> $F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$ | $EX = \frac{a+b}{2}$ <br> $Var(X) = \frac{(b-a)^2}{12}$ <br> $M_X(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}$ |
| exponential with parameter $\lambda > 0$ | $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$ <br> $F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$ | $EX = \frac{1}{\lambda}$ <br> $Var(X) = \frac{1}{\lambda^2}$ <br> $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$ |
| Cauchy | $f(x) = \frac{1}{\pi(1+x^2)}$ <br> $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x$ | $EX = \infty$ <br> $Var(X)$ DNE <br> $M_X(t)$ DNE |
| std. normal $n(0,1)$ | $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ <br> $F(x) = \Phi(x)$ | $EX = 0$ <br> $Var(X) = 1$ <br> $M_X(t) = e^{t^2/2}$ |
| normal $n(\mu, \sigma^2)$ | $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$ <br> $F(x) = \Phi(\frac{x-\mu}{\sigma})$ | $EX = \mu$ <br> $Var(X) = \sigma^2$ <br> $M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$ |
| gamma $\Gamma(r, \lambda)$ | $f(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$ <br> $F_X$ not given here | $EX = \frac{r}{\lambda}$ <br> $Var(X) = \frac{r}{\lambda^2}$ <br> $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$ for $t < \lambda$ |
| joint normal with mean vector $\mu$; covariance matrix $\Sigma$ | $f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left[\frac{-1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right]$ | $E\mathbf{X}$ and $Var(\mathbf{X})$ DNE <br> $M_{\mathbf{X}}(\mathbf{t}) = \exp\left(\mathbf{t} \cdot \mu + \frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}\right)$ |

## 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 $x, y \in \mathbb{R}$,

$$\sum_{k=0}^{n} \left( \begin{array}{c} n \\ k \end{array} \right) x^k y^{n-k} = (x+y)^n.$$

**Vandermonde Identity:** for all $n, k, r \in \mathbb{N}$,

$$\sum_{x=0}^{n} \left( \begin{array}{c} r \\ x \end{array} \right) \left( \begin{array}{c} n-r \\ k-x \end{array} \right) = \left( \begin{array}{c} n \\ k \end{array} \right).$$

**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 | 0.9992 | 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 |