MATH 171, STOCHASTIC PROCESSES, WINTER 2017

STEVEN HEILMAN

Contents

1. Introduction

A **stochastic process** is a collection of random variables. These random variables are often indexed by time, and the random variables are often related to each other by the evolution of some physical procedure. Stochastic processes can then model random phenomena that depend on time.

Date: March 17, 2017 © 2016 Steven Heilman, All Rights Reserved.

A basic question we will always try to answer is: what does the stochastic process "look like" after it runs for along period of time?

We will use conditional probabilities all the time, and the random variables we consider will often not be independent; indeed, the dependence of the random variables on each other makes stochastic processes interesting.

Also, whereas other probability classes focus mostly on equalities, we will additionally deal with inequalities and limits.

2. Review of Probability Theory

2.1. Random Variables, Conditional Probability, Expectation.

Definition 2.1 (**Universal Set**)**.** In a specific problem, we assume the existence of a sample space, or **universal set** $\mathcal{C}$ which contains all other sets. The universal set represents all possible outcomes of some random process. We sometimes call the universal set the **universe**. The universe is always assumed to be nonempty.

Definition 2.2 (**Countable Set Operations**)**.** Let $A_1, A_2, \ldots \subseteq \mathcal{C}$. We define

$$\bigcup_{i=1}^{\infty} A_i = \{x \in \mathcal{C} \colon \exists \text{ a positive integer } j \text{ such that } x \in A_j\}.$$

$$\bigcap_{i=1}^{\infty} A_i = \{x \in \mathcal{C} \colon x \in A_j,\ \forall \text{positive integers}\, j\}.$$

Exercise 2.3. Prove that the set of real numbers $\mathbb{R}$ can be written as the countable union

$$\mathbb{R} = \bigcup_{j=1}^{\infty} [-j, j].$$

(Hint: you should show that the left side contains the right side, and also show that the right side contains the left side.)

Prove that the singleton set $\{0\}$ can be written as

$$\{0\} = \bigcap_{j=1}^{\infty} [-1/j, 1/j].$$

Definition 2.4. A **Probability Law** (or **probability distribution**) $\mathbf{P}$ on a sample space $\mathcal{C}$ is a function whose domain is the set of all subsets of $\mathcal{C}$, and whose range is contained in $[0, 1]$, such that

(i) For any $A \subseteq \mathcal{C}$, we have $\mathbf{P}(A) \geq 0$. (**Nonnegativity**)

(ii) For any $A, B \subseteq \mathcal{C}$ such that $A \cap B = \emptyset$, we have

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

If $A_1, A_2, \ldots \subseteq \mathcal{C}$ and $A_i \cap A_j = \emptyset$ whenever i, j are positive integers with $i \neq j$, then

$$\mathbf{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbf{P}(A_k). \qquad \textbf{(Additivity)}$$

(iii) We have $\mathbf{P}(\mathcal{C}) = 1$. (**Normalization**)

Exercise 2.5 (**Continuity of a Probability Law**). Let $\mathbf{P}$ be a probability law on a sample space $\mathcal{C}$. Let $A_1, A_2, \ldots$ be sets in $\mathcal{C}$ which are increasing, so that $A_1 \subseteq A_2 \subseteq \cdots$. Then

$$\lim_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(\cup_{n=1}^{\infty} A_n).$$

In particular, the limit on the left exists. Similarly, let $A_1, A_2, \ldots$ be sets in $\mathcal{C}$ which are decreasing, so that $A_1 \supseteq A_2 \supseteq \cdots$. Then

$$\lim_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(\cap_{n=1}^{\infty} A_n).$$

Definition 2.6 (**Conditional Probability**). Let A, B be subsets of some sample space $\mathcal{C}$. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Assume that $\mathbf{P}(B) > 0$. We define the **conditional probability of** A **given** B, denoted by $\mathbf{P}(A|B)$, as

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Let $B_1, \ldots, B_n \subseteq \mathcal{C}$. We use the following notation to denote the conditional probability of A given $\cap_{i=1}^{n} B_i$:

$$\mathbf{P}(A|B_1, \ldots, B_n) := \mathbf{P}(A| \cap_{i=1}^{n} B_i).$$

Proposition 2.7 (**A Very Important Proposition**). *Let B be a fixed subset of some sample space $\mathcal{C}$. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Assume that $\mathbf{P}(B) > 0$. Given any subset A in $\mathcal{C}$, define $\mathbf{P}(A|B) = \mathbf{P}(A \cap B)/\mathbf{P}(B)$ as above. Then $\mathbf{P}(A|B)$ is itself a probability law on $\mathcal{C}$, when viewed as a function of subsets A in $\mathcal{C}$.*

Proposition 2.8 (**Multiplication Rule**). *Let n be a positive integer. Let $A_1, \ldots, A_n$ be sets in some sample space $\mathcal{C}$, and let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Assume that $\mathbf{P}(A_i) > 0$ for all $i \in \{1, \ldots, n\}$. Then*

$$\mathbf{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_2 \cap A_1) \cdots \mathbf{P}(A_n| \cap_{i=1}^{n-1} A_i).$$

Theorem 2.9 (**Total Probability Theorem**). *Let $A_1, \ldots, A_n$ be disjoint events in a sample space $\mathcal{C}$. That is, $A_i \cap A_j = \emptyset$ whenever $i, j \in \{1, \ldots, n\}$ satisfy $i \neq j$. Assume also that $\cup_{i=1}^{n} A_i = \mathcal{C}$. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Then, for any event $B \subseteq \mathcal{C}$, we have*

$$\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(B \cap A_i) = \sum_{i=1}^{n} \mathbf{P}(A_i)\mathbf{P}(B|A_i).$$

Theorem 2.10 (**Bayes' Rule**). *Let $A_1, \ldots, A_n$ be disjoint events in a sample space $\mathcal{C}$. That is, $A_i \cap A_j = \emptyset$ whenever $i, j \in \{1, \ldots, n\}$ satisfy $i \neq j$. Assume also that $\cup_{i=1}^{n} A_i = \mathcal{C}$. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Then, for any event $B \subseteq \mathcal{C}$ with $\mathbf{P}(B) > 0$, and for any $j \in \{1, \ldots, n\}$, we have*

$$\mathbf{P}(A_j|B) = \frac{\mathbf{P}(A_j)\mathbf{P}(B|A_j)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_j)\mathbf{P}(B|A_j)}{\sum_{i=1}^{n} \mathbf{P}(A_i)\mathbf{P}(B|A_i)}.$$

Definition 2.11 (**Independent Sets**). Let n be a positive integer. Let $A_1, \ldots, A_n$ be subsets of a sample space $\mathcal{C}$, and let $\mathbf{P}$ be a probability law on $\mathcal{C}$. We say that $A_1, \ldots, A_n$ are **independent** if, for any subset S of $\{1, \ldots, n\}$, we have

$$\mathbf{P}\left(\cap_{i\in S} A_i\right) = \prod_{i\in S} \mathbf{P}(A_i).$$

Definition 2.12 (**Random Variable**)**.** Let $\mathcal{C}$ be a sample space. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. A **random variable** X is a function $X\colon \mathcal{C} \to \mathbb{R}$. (Sometimes we might also consider a random variable to be a function from $\mathcal{C}$ to another set.) A **discrete random variable** is a random variable whose range is either finite or countably infinite. A **probability density function** (PDF) is a function $f\colon \mathbb{R} \to [0,\infty)$ such that $\int_{-\infty}^{\infty} f(x)dx = 1$, and such that, for any $-\infty \leq a \leq b \leq \infty$, the integral $\int_a^b f(x)dx$ exists. A random variable X is called **continuous** if there exists a probability density function f such that, for any $-\infty \leq a \leq b \leq \infty$, we have

$$\mathbf{P}(a \leq X \leq b) = \int_a^b f(x)dx.$$

When this equality holds, we call f the **probability density function of** X.

Let X be any random variable. We then define the **cumulative distribution function** (CDF) $F\colon \mathbb{R} \to [0,1]$ of X by

$$F(x) := \mathbf{P}(X \leq x), \qquad \forall\, x \in \mathbb{R}.$$

We say two random variables X, Y are **identically distributed** if they have the same CDF.

Definition 2.13 (**Probability Mass Function**)**.** Let X be a discrete random variable on a sample space $\mathcal{C}$, so that $X\colon \mathcal{C} \to \mathbb{R}$. The **probability mass function** (or PMF) of X, denote $p_X\colon \mathbb{R} \to [0,1]$ is defined by

$$p_X(x) = \mathbf{P}(X = x) = \mathbf{P}(\{X = x\}) = \mathbf{P}(\{c \in \mathcal{C}\colon X(c) = x\}), \qquad x \in \mathbb{R}.$$

Let $A \subseteq \mathbb{R}$. We denote $\{X \in A\} := \{c \in \mathcal{C}\colon X(c) \in A\}$.

We now give descriptions of some commonly encountered random variables.

Definition 2.14 (**Bernoulli Random Variable**)**.** Let $0 < p < 1$. A random variable X is called a **Bernoulli random variable with parameter** p if X has the following PMF:

$$p_X(x) = \begin{cases} p & \text{, if } x = 1 \\ 1-p & \text{, if } x = 0 \\ 0 & \text{, otherwise.} \end{cases}$$

Definition 2.15 (**Binomial Random Variable**)**.** Let $0 < p < 1$ and let n be a positive integer. A random variable X is called a **binomial random variable with parameters** n **and** p if X has the following PMF. If k is an integer with $0 \leq k \leq n$, then

$$p_X(k) = \mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

For any other x, we have $p_X(x) = 0$.

Recall that a sum of n independent Bernoulli random variables with parameter $0 < p < 1$ is a binomial random variable with parameters n and p.

Definition 2.16 (**Geometric Random Variable**)**.** Let $0 < p < 1$. A random variable X is called a **geometric random variable with parameter** p if X has the following PMF. If k is a positive integer, then

$$p_X(k) = \mathbf{P}(X = k) = (1-p)^{k-1} p.$$

For any other x, we have $p_X(x) = 0$.

Definition 2.17 (**Poisson Random Variable**)**.** Let $\lambda > 0$. A random variable X is called a **Poisson random variable with parameter** λ if X has the following PMF. If k is a nonnegative integer, then

$$p_X(k) = \mathbf{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

For any other x, we have $p_X(x) = 0$.

Example 2.18. We say that a random variable X is **uniformly distributed in** $[c, d]$ when X has the following density function: $f(x) = \frac{1}{d-c}$ when $x \in [c, d]$, and $f(x) = 0$ otherwise.

Example 2.19. Let $\lambda > 0$. A random variable X is called an **exponential random variable with parameter** λ if X has the following density function: $f(x) = \lambda e^{-\lambda x}$ when $x \geq 0$, and $f(x) = 0$ otherwise.

Definition 2.20 (**Normal Random Variable**)**.** Let $\mu \in \mathbb{R}$, $\sigma > 0$. A continuous random variable X is said to be **normal** or **Gaussian** with mean μ and variance σ^2 if X has the following density function:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad \forall\, x \in \mathbb{R}.$$

In particular, a **standard normal** or **standard Gaussian** random variable is defined to be a normal with $\mu = 0$ and $\sigma = 1$.

Definition 2.21 (**Indicator Function**)**.** Let $A \subseteq \mathcal{C}$ be a set. We define the **indicator function of** A, denoted $1_A\colon \mathcal{C} \to \mathbb{R}$ so that $1_A(c) = 0$ if $c \notin A$, and $1_A(c) = 1$ if $c \in A$.

Definition 2.22 (**Expected Value**)**.** Let $\mathcal{C}$ be a sample space, let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Let X be a random variable on $\mathcal{C}$. Assume that $X\colon \mathcal{C} \to [0, \infty)$. We define the **expected value** of X, denoted $\mathbf{E}(X)$, by

$$\mathbf{E}(X) = \int_0^\infty \mathbf{P}(X > t)dt.$$

More generally, if $g\colon [0, \infty) \to [0, \infty)$ is a differentiable function such that g' is continuous and $g(0) = 0$, we define

$$\mathbf{E}g(X) = \int_0^\infty g'(t)\mathbf{P}(X > t)dt.$$

In particular, taking $g(t) = t^n$ for any positive integer n, for any $t \geq 0$, we have

$$\mathbf{E}X^n = \int_0^\infty nt^{n-1}\mathbf{P}(X > t)dt.$$

For a general random variable X, if $\mathbf{E}\max(X, 0) < \infty$ and if $\mathbf{E}\max(-X, 0) < \infty$, we then define $\mathbf{E}(X) = \mathbf{E}\max(X, 0) - \mathbf{E}\max(-X, 0)$. Otherwise, we say that $\mathbf{E}(X)$ is undefined.

Remark 2.23. If we assume that the expected value and the integral on $\mathbb{R}$ can be commuted, then the following derivation of the formula for $\mathbf{E}g(X)$ can be given. From the Fundamental Theorem of Calculus, we have

$$g(X) = \int_0^X g'(t)dt = \int_0^\infty g'(t)1_{\{X>t\}}dt.$$

Therefore, $\mathbf{E}g(X) = \mathbf{E}\int_0^\infty g'(t)1_{\{X>t\}}dt = \int_0^\infty g'(t)\mathbf{E}1_{\{X>t\}}dt = \int_0^\infty g'(t)\mathbf{P}(X > t)dt$.

Remark 2.24. If X only takes positive integer values, then for any $t > 0$, if k is an integer such that $k - 1 < t \leq k$, then $\mathbf{P}(X > t) = \mathbf{P}(X \geq k)$, so

$$\mathbf{E}(X) = \int_0^\infty \mathbf{P}(X > t)dt = \sum_{k=1}^\infty \int_{k-1}^k \mathbf{P}(X > t)dt = \sum_{k=1}^\infty \mathbf{P}(X \geq k) = \sum_{k=0}^\infty \mathbf{P}(X > k).$$

Remark 2.25. If X is positive with density function f that is continuous, then recall that $(d/dt)\mathbf{P}(X \leq t) = f(t)$ for all $t \in \mathbb{R}$. Since $\mathbf{P}(X > t) = 1 - \mathbf{P}(X \leq t)$, we then have $(d/dt)\mathbf{P}(X > t) = -f(t)$. So, we can recover the usual formula for expected value by integrating by parts (assuming $g(0) = 0$ and $|g(t)| \leq 1$ for all $t \geq 0$):

$$\mathbf{E}g(X) = \int_0^\infty g'(t)\mathbf{P}(X > t)dt = -\int_0^\infty g(t)\frac{d}{dt}\mathbf{P}(X > t)dt = \int_0^\infty g(t)f(t)dt.$$

Theorem 2.26 (Fundamental Theorem of Calculus). *Let f be a probability density function. Then the function $g(t) = \int_{-\infty}^t f(x)dx$ is continuous at any $t \in \mathbb{R}$. Also, if f is continuous at a point x, then g is differentiable at $t = x$, and $g'(x) = f(x)$.*

Proposition 2.27. *Let $X_1, \ldots, X_n$ be random variables. Then*

$$\mathbf{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbf{E}(X_i).$$

Unfortunately the above property is not obvious from our definition of expected value.

Definition 2.28 (Convex Function). Let $\phi\colon \mathbb{R} \to \mathbb{R}$. We say that ϕ is **convex** if, for any $x, y \in \mathbb{R}$ and for any $t \in [0, 1]$, we have

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y).$$

Exercise 2.29. Let $\phi\colon \mathbb{R} \to \mathbb{R}$. Show that ϕ is convex if and only if: for any $y \in \mathbb{R}$, there exists a constant a and there exists a function $L\colon \mathbb{R} \to \mathbb{R}$ defined by $L(x) = a(x - y) + \phi(y)$, $x \in \mathbb{R}$, such that $L(y) = \phi(y)$ and such that $L(x) \leq \phi(x)$ for all $x \in \mathbb{R}$. (In the case that ϕ is differentiable, the latter condition says that ϕ lies above all of its tangent lines.)

(Hint: Suppose ϕ is convex. If x is fixed and y varies, show that $\frac{\phi(y)-\phi(x)}{y-x}$ increases as y increases. Draw a picture. What slope a should L have at x?)

Exercise 2.30. Let X, Y be positive random variables on a sample space $\mathcal{C}$. Assume that $X(c) \geq Y(c)$ for all $c \in \mathcal{C}$. Prove that $\mathbf{E}X \geq \mathbf{E}Y$.

More generally, if $X \leq Y$, $\mathbf{E}|X| < \infty$ and $\mathbf{E}|Y| < \infty$, show that $\mathbf{E}X \leq \mathbf{E}Y$.

Proposition 2.31 (Jensen's Inequality). *Let X be a random variable. Let $\phi\colon \mathbb{R} \to \mathbb{R}$ be convex. Then*

$$\phi(\mathbf{E}X) \leq \mathbf{E}\phi(X).$$

Proof. Let $y = \mathbf{E}X$. Then Exercise 2.29 says there exists a linear function $L(x) = a(x - y) + \phi(y)$ such that $L(x) \leq \phi(x)$ for all $x \in \mathbb{R}$. Taking expected values with respect to x and using Exercise 2.30, we get $\mathbf{E}L(X) \leq \mathbf{E}\phi(X)$. But $\mathbf{E}L(X) = a(\mathbf{E}X - y) + \phi(y) = a(y - y) + \phi(y) = \phi(y)$. So, $\phi(y) = \phi(\mathbf{E}X) \leq \mathbf{E}\phi(X)$. □

Definition 2.32 (**Variance**)**.** Let $\mathcal{C}$ be a sample space, let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Let X be a random variable on $\mathcal{C}$. We define the **variance** of X, denoted $\mathrm{var}(X)$, by

$$\mathrm{var}(X) = \mathbf{E}(X - \mathbf{E}(X))^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2.$$

We define the **standard deviation** of X, denoted σ_X, by

$$\sigma_X = \sqrt{\mathrm{var}(X)}.$$

Proposition 2.33. *Let $\mathcal{C}$ be a sample space, let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Let X be a random variable on $\mathcal{C}$. Let a, b be constants. Then*

$$\mathrm{var}(aX + b) = a^2 \mathrm{var}(X).$$

We will review conditional expectation later on in the notes.

Definition 2.34 (**Joint Density Function**)**.** We say that random variables $X_1, \ldots, X_n$ have **joint density function** $f\colon \mathbb{R}^n \to [0, \infty)$ if $\int_{\mathbb{R}^n} f(x)dx = 1$, and if

$$\mathbf{P}((X_1, \ldots, X_n) \in A) = \int_A f(x)dx, \qquad \forall\, A \subseteq \mathbb{R}^n.$$

We define the **marginal density** $f_1\colon \mathbb{R} \to [0, \infty)$ of X_1 so that

$$f_1(x_1) = \int_{\mathbb{R}^{n-1}} f(x_1, \ldots, x_n) dx_2 \cdots dx_n, \qquad \forall\, x_1 \in \mathbb{R}.$$

Similarly, we can define the marginal density $f_{12}\colon \mathbb{R}^2 \to [0, \infty)$ of X_1, X_2 so that

$$f_{12}(x_1, x_2) = \int_{\mathbb{R}^{n-2}} f(x_1, \ldots, x_n) dx_3 \cdots dx_n, \qquad \forall\, x_1, x_2 \in \mathbb{R}.$$

And so on.

Definition 2.35 (**Independence of Random Variables**)**.** Let $X_1, \ldots, X_n$ be random variables on a sample space $\mathcal{C}$, and let $\mathbf{P}$ be a probability law on $\mathcal{C}$. We say that $X_1, \ldots, X_n$ are **independent** if

$$\mathbf{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \leq x_i), \qquad \forall\, x_1, \ldots, x_n \in \mathbb{R}.$$

Exercise 2.36. Let $X_1, \ldots, X_n$ be discrete random variables. Assume that

$$\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i), \qquad \forall\, x_1, \ldots, x_n \in \mathbb{R}.$$

Show that $X_1, \ldots, X_n$ are independent.

Exercise 2.37. Let $X_1, \ldots, X_n$ be continuous random variables with joint PDF $f\colon \mathbb{R}^n \to [0, \infty)$. Assume that

$$f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), \qquad \forall\, x_1, \ldots, x_n \in \mathbb{R}.$$

Show that $X_1, \ldots, X_n$ are independent.

Proposition 2.38. *Let $X_1, \ldots, X_n$ be random variables on a sample space $\mathcal{C}$, and let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Assume that $X_1, \ldots, X_n$ are pairwise independent. That is, X_i and X_j are independent whenever $i, j \in \{1, \ldots, n\}$ with $i \neq j$. Then*

$$\mathrm{var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{var}(X_i).$$

Proposition 2.39. *Let $X_1, \ldots, X_n$ be independent random variables. Then*

$$\mathbf{E}(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} \mathbf{E}(X_i).$$

Proposition 2.40. *Let $0 = n_0 < n_1 < n_2 < \ldots < n_k = n$ be integers. Let $X_1, \ldots, X_n$ be independent random variables. For any $1 \leq i \leq k$, let $g_i \colon \mathbb{R}^{n_i - n_{i-1}} \to \mathbb{R}$. Then the random variables $g_1(X_1, \ldots, X_{n_1})$, $g_2(X_{n_1+1}, \ldots, X_{n_2}), \ldots,$ $g_k(X_{n_{k-1}+1}, \ldots, X_{n_k})$ are independent. Consequently,*

$$\mathbf{E}(\prod_{i=1}^{n} g_i(X_{n_{i-1}+1}, \ldots, X_{n_i})) = \prod_{i=1}^{n} \mathbf{E} g_i(X_{n_{i-1}+1}, \ldots, X_{n_i}).$$

2.2. **Some Linear algebra.**

Definition 2.41 (Eigenvector, Eigenvalue). Let A be an $m \times m$ real matrix, let $x \in \mathbb{R}^m$ be a column vector, and let $y \in \mathbb{R}^m$ be a row vector. We say x is a (right) **eigenvector** of A with eigenvalue $\lambda \in \mathbb{C}$ if $x \neq 0$ and

$$Ax = \lambda x.$$

We say y is a (left) **eigenvector** of A with eigenvalue $\lambda \in \mathbb{C}$ if $y \neq 0$ and

$$yA = \lambda y.$$

Note that x is a right eigenvector for A if and only if x^T is a left eigenvector of A^T.

Definition 2.42. The **null space** (or **kernel**) of an $m \times n$ real matrix A is the set of all column-vectors $x \in \mathbb{R}^n$ such that $Ax = 0$. The **nullity** of A is the number of nonzero vectors that can form a basis of the null space of A

The **column space** is the set of all linear combinations of the columns of the matrix A. The **rank** of A is the number of nonzero vectors that can form a basis of the column space of A.

Theorem 2.43 (Rank-Nullity Theorem). *Let A be an $m \times n$ real matrix. Then the rank of A plus the nullity of A is equal to n.*

2.3. **Law Of Large Numbers.**

Theorem 2.44 (Weak Law of Large Numbers). *Let $X_1, \ldots, X_n$ be independent identically distributed random variables. Assume that $\mu := \mathbf{E} X_1$ is finite. Then for any $\varepsilon > 0$*

$$\lim_{n \to \infty} \mathbf{P}\left(\left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| > \varepsilon \right) = 0.$$

Theorem 2.45 (**Strong Law of Large Numbers**). *Let $X_1, \ldots, X_n$ be independent identically distributed random variables. Assume that $\mu := \mathbf{E}X_1$ is finite. Then*

$$\mathbf{P}\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1.$$

2.4. **Central Limit Theorem.** The following Theorem is a special case of the Central Limit Theorem.

Theorem 2.46 (**De Moivre-Laplace Theorem**). *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with parameter $1/2$. Recall that X_1 has mean $1/2$ and variance $1/4$. Let $a \in \mathbb{R}$. Then*

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{X_1 + \cdots + X_n - (1/2)n}{\sqrt{n}\sqrt{1/4}} \leq a\right) = \int_{-\infty}^{a} e^{-t^2/2} \frac{dt}{\sqrt{2\pi}}.$$

That is, when n is large, the CDF of $\frac{X_1+\cdots+X_n-(1/2)n}{\sqrt{n}\sqrt{1/4}}$ is roughly the same as that of a standard normal. In particular, if you flip n fair coins, then the number of heads you get should typically be in the interval $(n/2 - \sqrt{n}/2, n/2 + \sqrt{n}/2)$, when n is large.

Remark 2.47. The random variable $\frac{X_1+\cdots+X_n-(1/2)n}{\sqrt{n}\sqrt{1/4}}$ has mean zero and variance 1, just like the standard Gaussian. So, the normalizations of $X_1 + \cdots + X_n$ we have chosen are sensible. Also, to explain the interval $(n/2 - \sqrt{n}/2, n/2 + \sqrt{n}/2)$, note that

$$\begin{aligned}
&\lim_{n\to\infty} \mathbf{P}\left(\frac{n}{2} - \frac{\sqrt{n}}{2} \leq X_1 + \cdots + X_n \leq \frac{n}{2} + \frac{\sqrt{n}}{2}\right) \\
&= \lim_{n\to\infty} \mathbf{P}\left(-\frac{\sqrt{n}}{2} \leq X_1 + \cdots + X_n - \frac{n}{2} \leq \frac{\sqrt{n}}{2}\right) \\
&= \lim_{n\to\infty} \mathbf{P}\left(-1 \leq \frac{X_1 + \cdots + X_n - \frac{n}{2}}{\sqrt{n}/2} \leq 1\right) = \int_{-1}^{1} e^{-t^2/2} \frac{dt}{\sqrt{2\pi}} \approx .6827.
\end{aligned}$$

Exercise 2.48. Estimate the probability that 1000000 coin flips of fair coins will result in more than $501,000$ heads, using the De Moivre-Laplace Theorem. (Some of the following integrals may be relevant: $\int_{-\infty}^{0} e^{-t^2/2} dt/\sqrt{2\pi} = 1/2$, $\int_{-\infty}^{1} e^{-t^2/2} dt/\sqrt{2\pi} \approx .8413$, $\int_{-\infty}^{2} e^{-t^2/2} dt/\sqrt{2\pi} \approx .9772$, $\int_{-\infty}^{3} e^{-t^2/2} dt/\sqrt{2\pi} \approx .9987$.)

Casinos do these kinds of calculations to make sure they make money and that they do not go bankrupt. Financial institutions and insurance companies do similar calculations for similar reasons.

In fact, there is nothing special about the parameter $1/2$ in the above theorem.

Theorem 2.49 (**De Moivre-Laplace Theorem, Second Version**). *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with parameter p. Recall that X_1 has mean p and variance $p(1-p)$. Let $a \in \mathbb{R}$. Then*

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{X_1 + \cdots + X_n - pn}{\sqrt{n}\sqrt{p(1-p)}} \leq a\right) = \int_{-\infty}^{a} e^{-t^2/2} \frac{dt}{\sqrt{2\pi}}.$$

In fact, there is nothing special about Bernoulli random variables in the above theorem.

Theorem 2.50 (**Central Limit Theorem**). *Let $X_1, \ldots, X_n$ be independent identically distributed random variables. Assume that $\mathbf{E}\,|X_1| < \infty$ and $0 < \mathrm{Var}(X_1) < \infty$.*

Let $\mu = \mathbf{E}X_1$ and let $\sigma = \sqrt{\mathrm{Var}(X_1)}$. Then for any $-\infty \leq a \leq \infty$,

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{X_1 + \cdots + X_n - \mu n}{\sigma\sqrt{n}} \leq a\right) = \int_{-\infty}^{a} e^{-t^2/2}\frac{dt}{\sqrt{2\pi}}.$$

Remark 2.51. The random variable $\frac{X_1+\cdots+X_n-(1/2)n}{\sigma\sqrt{n}}$ has mean zero and variance 1, just like the standard Gaussian.

Theorem 2.52 (**Fubini Theorem for Integrals**). *Let $h\colon \mathbb{R}^2 \to \mathbb{R}$ be a continuous function such that $\iint_{\mathbb{R}^2} |h(x,y)|\,dxdy < \infty$. Then*

$$\iint_{\mathbb{R}^2} h(x,y)dxdy = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} h(x,y)dx\right)dy = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} h(x,y)dy\right)dx.$$

Theorem 2.53 (**Fubini Theorem for Sums**). *Let $\{a_{ij}\}_{i,j\geq 0}$ be a doubly-infinite array of nonnegative numbers. Then*

$$\sum_{i=0}^{\infty}\left(\sum_{j=0}^{\infty} a_{ij}\right) = \sum_{j=0}^{\infty}\left(\sum_{i=0}^{\infty} a_{ij}\right).$$

Exercise 2.54. Find a doubly-infinite array of real numbers $\{a_{ij}\}_{i,j\geq 0}$ such that

$$\sum_{i=0}^{\infty}\left(\sum_{j=0}^{\infty} a_{ij}\right) = 1 \neq 0 = \sum_{j=0}^{\infty}\left(\sum_{i=0}^{\infty} a_{ij}\right).$$

(Hint: the array can be chosen to have all entries either $-1, 0$, or 1. And most of the entries can be chosen to be 0.)

Exercise 2.55. Let X, Y be independent, discrete random variables. Using a total probability theorem-type argument, show that

$$\mathbf{P}(X+Y=z) = \sum_{x\in\mathbb{R}} \mathbf{P}(X=x)\mathbf{P}(Y=z-x), \qquad \forall\, z \in \mathbb{R}.$$

Exercise 2.56. Let X, Y be independent, continuous random variables with densities f_X, f_Y, respectively. Let f_{X+Y} be the density of $X+Y$. Show that

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x)dx, \qquad \forall z \in \mathbb{R}.$$

Using this identity, find the density f_{X+Y} when X and Y are both independent, uniformly distributed on $[0,1]$.

3. Markov Chains

Our first example of a stochastic process will be a Markov chain. Before defining a Markov chain formally, we give an example of one.

Example 3.1 (**Frog on two Lily Pads**). Suppose there are two different lily pads labelled e (for east) and w (for west). Suppose the frog starts on one of the two lily pads. Let $0 < p, q < 1$. There is a coin on the lily pad e which has probability p of landing heads and probability $1-p$ of landing tails. Similarly, there is a coin on the lily pad w which has probability q of landing heads and probability $1-q$ of landing tails. Every day, the frog flips the coin on the lily pad it currently occupies. If the coin lands heads, the frog goes to the other lily pad. If the coin lands tails, the frog stays on its current lily pad.

For any $n \geq 0$, let X_n be the (random) location of the frog at the beginning of day n. Then the sequence of random variables $X_0, X_1, X_2, \ldots$ describes the sequence of positions that the frog takes. Note that if $\mathcal{C}$ is the sample space, then for any $n \geq 0$, $X_n : \mathcal{C} \to \{e, w\}$ is a random variable, taking either the value e or w. We would like to find the probabilities that $X_1, X_2, \ldots$ take the values e and w. To this end, let P be a real 2×2 matrix such that $P(x, y) = \mathbf{P}(X_1 = y \,|\, X_0 = x)$, for all $x, y \in \{e, w\}$. That is,

$$P = \begin{pmatrix} P(e,e) & P(e,w) \\ P(w,e) & P(w,w) \end{pmatrix} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

More generally, note that for any integer $n \geq 1$, $P(x, y) = \mathbf{P}(X_n = y \,|\, X_{n-1} = x)$, since the location of the frog tomorrow only depends on its location today.

Then the random variables $(X_0, X_1, \ldots)$ is a Markov Chain with transition matrix P.

Definition 3.2 (**Finite Markov Chain**)**.** A **finite Markov Chain** is a stochastic process $(X_0, X_1, X_2, \ldots)$ together with a finite set Ω, which is called the **state space** of the Markov Chain, and an $|\Omega| \times |\Omega|$ real matrix P. The random variables $X_0, X_1, \ldots$ take values in the finite set Ω. The matrix P is **stochastic**, that is all of its entries are nonnegative and

$$\sum_{y \in \Omega} P(x, y) = 1, \qquad \forall\, x \in \Omega.$$

And the stochastic process satisfies the following **Markov property**: for all $x, y \in \Omega$, for any $n \geq 1$, and for all events H_{n-1} of the form $H_{n-1} = \cap_{k=0}^{n-1}\{X_k = x_k\}$, where $x_k \in \Omega$ for all $0 \leq k \leq n-1$, such that $\mathbf{P}(H_{n-1} \cap \{X_n = x\}) > 0$, we have

$$\mathbf{P}(X_{n+1} = y \,|\, H_{n-1} \cap \{X_n = x\}) = \mathbf{P}(X_{n+1} = y \,|\, X_n = x) = P(x, y).$$

That is, the next location of the Markov chain only depends on its current location. And the transition probability is defined by $P(x, y)$.

Exercise 3.3. Let P, Q be stochastic matrices of the same size. Show that PQ is a stochastic matrix. Conclude that, if r is a positive integer, then P^r is a stochastic matrix.

Exercise 3.4. Let A, B be events in a sample space. Let $C_1, \ldots, C_n$ be events such that $C_i \cap C_j = \emptyset$ for any $i, j \in \{1, \ldots, n\}$ with $i \neq j$, and such that $\cup_{i=1}^{n} C_i$ is the whole sample space. Show:

$$\mathbf{P}(A|B) = \sum_{i=1}^{n} \mathbf{P}(A|B,\, C_i)\mathbf{P}(C_i|B).$$

(Hint: consider using the Total Probability Theorem (Theorem 2.9) and Proposition 2.7.)

Example 3.5. Returning to the frog example, we have

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

Note that each row of this matrix sums to 1, so P is stochastic. We can then compute the probabilities that X_2 takes various values, by conditioning on the two possible values of X_1. Using Exercise 3.4, the Markov Property, and the definition of P,

$$\begin{aligned}\mathbf{P}(X_2 = w \,|\, X_0 = e) &= \mathbf{P}(X_2 = w \,|\, X_1 = e, X_0 = e)\mathbf{P}(X_1 = e \,|\, X_0 = e) \\ &\quad + \mathbf{P}(X_2 = w \,|\, X_1 = w, X_0 = e)\mathbf{P}(X_1 = w \,|\, X_0 = e) \\ &= \mathbf{P}(X_2 = w \,|\, X_1 = e)\mathbf{P}(X_1 = e \,|\, X_0 = e) + \mathbf{P}(X_2 = w \,|\, X_1 = w)\mathbf{P}(X_1 = w \,|\, X_0 = e) \\ &= P(e,w)P(e,e) + P(w,w)P(e,w) = p(1-p) + (1-q)p. \qquad (1)\end{aligned}$$

More generally, for any $n \geq 1$, define the 1×2 row vector

$$\mu_n := \big(\mathbf{P}(X_n = e \,|\, X_0 = e), \qquad \mathbf{P}(X_n = w \,|\, X_0 = e)\big).$$

Also, assume the frog starts on the lily pad e, so that $\mu_0 = (1,0)$. Then (1) generalizes to

$$\mu_n = \mu_{n-1}P, \qquad \forall\, n \geq 1.$$

Iteratively applying this identity,

$$\mu_n = \mu_0 P^n, \qquad \forall\, n \geq 0.$$

What happens when n becomes large? In this case, we might expect the vector μ_n to converge to something as $n \to \infty$. That is, when n becomes very large, the probability that X_n takes a particular value converges to a number. Suppose the vector μ_n converges to some 1×2 row vector π as $n \to \infty$. Note that the entries of μ_n sum to 1 and are nonnegative, so the same is true for π. We claim that

$$\pi = \pi P.$$

That is, π is a (left)-eigenvector of P with eigenvalue 1. To see why $\pi = \pi P$ should be true, note that

$$\pi = \lim_{n\to\infty} \mu_n = \lim_{n\to\infty} \mu_0 P^n = (\lim_{n\to\infty} \mu_0 P^n)P = (\lim_{n\to\infty} \mu_n)P = \pi P.$$

The equation $\pi = \pi P$ allows us to solve for π, since it says

$$\Big(\pi(e), \;\; \pi(w)\Big) = \Big(\pi(e)(1-p) + \pi(w)q, \;\; \pi(e)p + \pi(w)(1-q)\Big).$$

So, $0 = -p\pi(e) + \pi(w)q$, $\pi(w) = \pi(e)(p/q)$, and $\pi(e) + \pi(w) = 1$, so $\pi(e)(1 + p/q) = 1$, so

$$\pi(e) = \frac{q}{p+q}, \qquad \pi(w) = \frac{p}{p+q}.$$

That is, when n becomes very large, the frog has probability roughly $q/(q+p)$ of being on the e pad, and it has probability roughly $p/(q+p)$ of being on the w pad.

We can actually say something a bit more precise. For any $n \geq 0$, define

$$\Delta_n = \mu_n(e) - \frac{q}{p+q}.$$

Then, using the definition of μ_{n+1}, and $\mu_n(w) = 1 - \mu_n(e)$, we have, for any $n \geq 0$

$$\Delta_{n+1} = (\mu_n P)(e) - \frac{q}{p+q} = \mu_n(e)(1-p) + q(1 - \mu_n(e)) - \frac{q}{p+q} = (1-p-q)\Delta_n.$$

So, iterating this equality, we have

$$\Delta_n = (1-p-q)^n \Delta_0, \qquad \forall\, n \geq 1.$$

Since $0 < p, q < 1$, this means that the quantity Δ_n is converging exponentially fast to 0. In particular,

$$\lim_{n\to\infty} \Delta_n = 0, \qquad \lim_{n\to\infty} \mu_n = \pi.$$

(A similar argument shows that $\mu_n(w) - \frac{p}{p+q}$ converges exponentially fast to zero)

Exercise 3.6. Let $0 < p, q < 1$. Let $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$. Find the (left) eigenvectors of P, and find the eigenvalues of P. By writing any row vector $x \in \mathbb{R}^2$ as a linear combination of eigenvectors of P (whenever possible), find an expression for xP^n for any $n \geq 1$. What is $\lim_{n\to\infty} xP^n$? Is it related to the vector $\pi = (q/(p+q), p/(p+q))$?

3.1. **Examples of Markov Chains.** Unfortunately, not all Markov chains converge when n becomes large, as we now demonstrate.

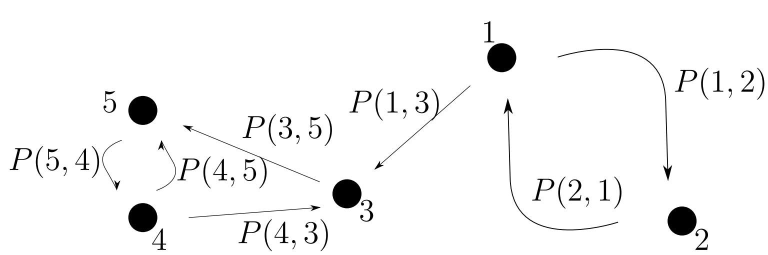
Example 3.7. Consider the Markov chain defined by the matrix $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Note that $P^n = P$ for any positive odd integer n, and $P^n = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ for any positive even integer n. So, if μ is any 1×2 row vector with unequal entries, it is impossible for μP^n to converge as $n \to \infty$.

Example 3.8 (**Random Walk on a Graph**)**.** A (finite, undirected, simple) **graph** $G = (V, E)$ consists of a finite **vertex set** V and an **edge set** E. The edge set consists of unordered pairs of vertices, so that $E \subseteq \{\{x, y\} : x, y \in V, x \neq y\}$. We think of distinct vertices as distinct nodes, where two nodes $x, y \in V$ are joined by an edge if and only if $\{x, y\} \in E$. When $\{x, y\} \in E$, we say that y is a **neighbor** of x (and x is a neighbor of y). The **degree** $\deg(x)$ of a vertex $x \in V$ is the number of neighbors of x. We assume that $\deg(x) > 0$ for every $x \in V$, so that G has no isolated vertices.

Given a graph $G = (V, E)$, we define the **simple random walk** on G to be the Markov chain with state space V and transition matrix

$$P(x, y) = \begin{cases} \frac{1}{\deg(x)} & \text{, if } x \text{ and } y \text{ are neighbors} \\ 0 & \text{, otherwise.} \end{cases}$$

In this Markov chain, starting from any position x, the next state is then any neighbor y of x, each with equal probability. More generally, a **random walk** on a vertex set V is any Markov chain with state space V.



Exercise 3.9. Let $G = (V, E)$ be a graph. Let $|E|$ denote the number of elements in the set E, i.e. $|E|$ is the number of edges of the graph. Prove: $\sum_{x\in V} \deg(x) = 2\,|E|$.

Example 3.10 (**Lazy Random Walk**). Let P be the matrix defined by a simple random walk on a graph $G = (V, E)$. Let I denote the $|V| \times |V|$ identity matrix. The **lazy random walk** is the Markov chain with transition matrix $(P + I)/2$. That is, with probability $1/2$, the next state is your current state, and with probability $1/2$, the next state is any neighbor of the current state, each chosen with equal probability.

Example 3.11 (**Google's PageRank Algorithm**). We can think of the set of all websites on the internet as a graph, where each website is a vertex in V, and $\{x, y\} \in E$ if and only if there is a hyperlink on page x that links to page y (or if there is a hyperlink on page y that links to page x). Let P denote the normalized adjacency matrix, so that $P(x, y) = 1/\deg(x)$ if $\{x, y\} \in E$, and $P(x, y) = 0$ otherwise. Note that P is a stochastic matrix. Let Q be the $|V| \times |V|$ matrix such that all entries of Q are 1. Consider the matrix

$$N := (.85)P + (.15)Q/|V|.$$

Then N is a stochastic matrix. We can think of the Markov chain associated to N as follows: 85% of the time, you move from one website to another by one of the hyperlinks on that site, each with equal probability. And 15% of the time, you go to any website on the internet, uniformly at random. The PageRank vector π is then a $1 \times |V|$ vector with $\pi(x) \geq 0$ for all $x \in V$, and $\sum_{x \in V} \pi(x) = 1$ such that $\pi = \pi N$. That is, the PageRank value of website $x \in V$ is $\pi(x)$. The most "relevant" websites x have the largest values of $\pi(x)$.

The idea here is that if $\pi(x)$ is large, then the Markov chain will often encounter the website x, so we think of x as being an important website. At the moment, π is not guaranteed to exist. We will return to this issue in Theorem 3.33 below.

3.2. **Classification of States.**

Definition 3.12. Suppose we have a Markov chain $(X_0, X_1, X_2, \ldots)$ with state space Ω. Let $x \in \Omega$ be fixed. For any set A in the sample space, define a probability law $\mathbf{P}_x$ such that

$$\mathbf{P}_x(A) := \mathbf{P}(A | X_0 = x).$$

Similarly, we define $\mathbf{E}_x$ to be the expected value with respect to the probability law $\mathbf{P}_x$.

More generally, if μ is a probability distribution on Ω, we let $\mathbf{P}_\mu$ denote the probability law, given that the Markov chain started from the probability distribution μ, so that $\mathbf{P}(X_0 = x_0) = \mu(x_0)$ for any $x_0 \in \Omega$. So, for example,

$$\mathbf{P}_\mu(X_1 = x_1) = \sum_{x_0 \in \Omega} P(x_0, x_1)\mu(x_0), \qquad \forall\, x_1 \in \Omega.$$

Note also that if $x \in \Omega$ is fixed, and if μ is defined so that $\mu(x) = 1$ and $\mu(y) = 1$ for all $y \neq x$, then $\mathbf{P}_\mu = \mathbf{P}_x$.

Definition 3.13 (**Return Time**). Suppose we have a Markov Chain $X_0, X_1, \ldots$ with state space Ω. Let $y \in \Omega$. Define the **first return time** of y to be the following random variable:

$$T_y := \min\{n \geq 1 \colon X_n = y\}.$$

Also, define

$$\rho_{yy} := \mathbf{P}_y(T_y < \infty).$$

That is, ρ_{yy} is the probability that the chain starts at y, and it returns to y in finite time.

Definition 3.14 (**Stopping Time**)**.** A **stopping time** for a Markov chain $X_0, X_1, \ldots$ is a random variable T taking values in $0, 1, 2, \ldots, \cup\{\infty\}$ such that, for any integer $n \geq 0$, the event $\{T = n\}$ is determined by $X_0, \ldots, X_n$. More formally, for any integer $n \geq 1$, there is a set $B_n \subseteq \Omega^{n+1}$ such that $\{T = n\} = \{(X_0, \ldots, X_n) \in B_n\}$. Put another way, the indicator function $1_{\{T=n\}}$ is a function of the random variables $X_0, \ldots, X_n$.

Example 3.15. Fix $y \in \Omega$. The first return time T_y is a stopping time since

$$\begin{aligned}\{T_y = n\} &= \{X_1 \neq y,\, X_2 \neq y, \ldots, X_{n-1} \neq y,\, X_n = y\} \\ &= \{(X_0, \ldots, X_n) \in \Omega \times \{y\}^c \times \cdots \{y\}^c \times \{y\}\}, \qquad \forall\, n \geq 0.\end{aligned}$$

For an intuitive example of a stopping time, suppose $X_0, X_1, \ldots$ is a Markov chain where X_n is the price of a stock at time $n \geq 0$. Then a stopping time could be the first time that the stock price reaches either \$90 or \$100. That is, a stopping time is a stock trading strategy, or a way of "stopping" the random process, but only using information from the past and present. An example of a random variable T that is not a stopping time is to let T be the time that stock price becomes highest, before the price drops to 0. (For example, $\{T = 100\}$ could depend on X_{104}.) So, since T relies on future information, T is not a stopping time.

Theorem 3.16 (**Strong Markov Property**)**.** *Let T be a stopping time for a Markov chain. Let $\ell \geq 1$, and let $A \subseteq \Omega^\ell$. Fix $n \geq 1$. Then, for any $x_0, \ldots, x_n \in \Omega$,*

$$\begin{aligned}\mathbf{P}_{x_0}((X_{T+1}, \ldots, X_{T+\ell}) \in A \,|\, T = n \text{ and } (X_0, \ldots, X_n) = (x_0, \ldots, x_n)) \\ = \mathbf{P}_{x_n}((X_1, \ldots, X_\ell) \in A).\end{aligned}$$

That is, if we know $T = n$, $X_n = x_n$ and if we know the previous n states of the Markov chain, then this is exactly the same as starting the Markov chain from the state x_n.

Proof. By the definition of the stopping time, there exists $B_n \subseteq \Omega^{n+1}$ such that $\{T = n\} = \{(X_0, \ldots, X_n) \in B_n\}$. If $(x_0, \ldots, x_n) \in B_n$, we then have (using Exercise 3.17)

$$\begin{aligned}&\mathbf{P}_{x_0}((X_{T+1}, \ldots, X_{T+\ell}) \in A \,|\, T = n,\, (X_0, \ldots, X_n) = (x_0, \ldots, x_n)) \\ &\quad= \mathbf{P}((X_{T+1}, \ldots, X_{T+\ell}) \in A \,|\, T = n,\, (X_0, \ldots, X_n) = (x_0, \ldots, x_n)) \\ &\quad= \mathbf{P}((X_{n+1}, \ldots, X_{n+\ell}) \in A \,|\, T = n,\, (X_0, \ldots, X_n) = (x_0, \ldots, x_n)) \\ &\quad= \mathbf{P}((X_{n+1}, \ldots, X_{n+\ell}) \in A \,|\, (X_0, \ldots, X_n) = (x_0, \ldots, x_n)) \\ &\quad= \mathbf{P}((X_{n+1}, \ldots, X_{n+\ell}) \in A \,|\, X_n = x_n) \qquad \text{, by Exercise 3.18} \\ &\quad= \mathbf{P}((X_1, \ldots, X_\ell) \in A \,|\, X_0 = x_n) \qquad \text{, by Exercise 3.18} \\ &\quad= \mathbf{P}_{x_n}((X_1, \ldots, X_\ell) \in A), \qquad \text{, by definition of } P_{x_n}.\end{aligned}$$

Finally, if $(x_0, \ldots, x_n) \notin B_n$, then $\{T = n\} \cap \{(X_0, \ldots, X_n) = (x_0, \ldots, x_n)\} = \emptyset$, so the conditional probability of this event is undefined, and there is nothing to prove. □

Exercise 3.17. Let A, B be events such that $B \subseteq \{X_0 = x_0\}$. Then $\mathbf{P}(A|B) = \mathbf{P}_{x_0}(A|B)$.
More generally, if A, B are events, then $\mathbf{P}_{x_0}(A|B) = \mathbf{P}(A|B, X_0 = x_0)$.

Exercise 3.18. Suppose we have a Markov Chain with state space Ω. Let $n \geq 0$, $\ell \geq 1$, let $x_0, \ldots, x_n \in \Omega$ and let $A \subseteq \Omega^\ell$. Using the (usual) Markov property, show that

$$\begin{aligned}\mathbf{P}((X_{n+1}, \ldots, X_{n+\ell}) \in A \,|\, (X_0, \ldots, X_n) = (x_0, \ldots, x_n)) \\ = \mathbf{P}((X_{n+1}, \ldots, X_{n+\ell}) \in A \,|\, X_n = x_n).\end{aligned}$$

Then, show that

$$\mathbf{P}((X_{n+1},\ldots,X_{n+\ell})\in A\,|\,X_n=x_n)=\mathbf{P}((X_1,\ldots,X_\ell)\in A\,|\,X_0=x_n).$$

(Hint: it may be helpful to use the Multiplication Rule (Proposition 2.8).)

Exercise 3.19. Suppose we have a Markov chain $X_0, X_1, \ldots$ with finite state space Ω. Let $y \in \Omega$. Define $L_y := \max\{n \geq 0\colon X_n = y\}$. Is L_y a stopping time? Prove your assertion.

Example 3.20. If y is in the state space of a Markov chain, recall we defined the return time to be $T_y = \min\{n \geq 1\colon X_n = y\}$. We also verified T_y is a stopping time. Let $T_y^{(1)} = T_y$, and for any $k \geq 2$, define a random variable

$$T_y^{(k)} = \min\{n > T_y^{(k-1)}\colon X_n = y\}.$$

So, $T_y^{(k)}$ is the time of the k^{th} return of the Markov chain to state y. Just as before, we can verify that $T_y^{(k)}$ is a stopping time for any $k \geq 1$.

Let $T := T_y^{(k-1)}$. Note that if $T < \infty$, then $T_y^{(k)} - T = \min\{n \geq 1\colon X_{T+n} = y\}$. Let $A \subseteq \Omega^\ell$ such that $A = \{y\}^c \times \cdots \times \{y\}^c \times \{y\}$. From the Strong Markov Property (Theorem 3.16), for any $n \geq 1$,

$$\begin{aligned}\mathbf{P}_{x_0}((X_{T+1},\ldots,X_{T+\ell})\in A\,|\,T=n \text{ and } (X_0,\ldots,X_n)=(x_0,\ldots,x_n))\\ =\mathbf{P}_{x_n}((X_1,\ldots,X_\ell)\in A).\end{aligned}$$

Since $\{T_y^{(k)} - T = \ell\} = \{(X_{T+1},\ldots,X_{T+\ell}) \in A\}$, and $\{T_y = \ell\} = \{(X_1,\ldots,X_\ell) \in A\}$, if we use $x_0 = x_n = y$, we get

$$\mathbf{P}_y(T_y^{(k)}-T=\ell\,|\,T=n,\,X_1=x_1,\ldots,X_{n-1}=x_{n-1},X_n=y)=\mathbf{P}_y(T_y=\ell),\qquad\forall\,\ell,n\geq 1.$$

From the definition of conditional probability,

$$\begin{aligned}&\mathbf{P}_y(T_y^{(k)}-T=\ell,\,T=n,\,X_1=x_1,\ldots,X_{n-1}=x_{n-1},X_n=y)\\ &\quad=\mathbf{P}_y(T=n,\,X_1=x_1,\ldots,X_{n-1}=x_{n-1},X_n=y)\mathbf{P}_y(T_y=\ell)\qquad\forall\,\ell,n\geq 1.\end{aligned}$$

Summing over all $x_1,\ldots,x_{n-1}$ such that $\{X_1=x_1,\ldots,X_{n-1}=x_{n-1},X_n=y\}\subseteq\{T=n\}$,

$$\mathbf{P}_y(T_y^{(k)}-T=\ell,\,T=n)=\mathbf{P}_y(T=n)\mathbf{P}_y(T_y=\ell),\qquad\forall\,\ell,n\geq 1.$$

Taking the union over all $\ell \geq 1$,

$$\mathbf{P}_y(T_y^{(k)}-T<\infty,\,T=n)=\mathbf{P}_y(T=n)\mathbf{P}_y(T_y<\infty)=\mathbf{P}_y(T=n)\rho_{yy},\qquad\forall\,n\geq 1.$$

Then, summing over all $n \geq 1$,

$$\mathbf{P}_y(T_y^{(k)}-T<\infty,\,T<\infty)=\rho_{yy}\mathbf{P}_y(T<\infty).$$

Using the definition of conditional probability again,

$$\mathbf{P}_y(T_y^{(k)}-T<\infty\,|\,T<\infty)=\rho_{yy}.\qquad(*)$$

So, using the multiplication rule (Proposition 2.8) and recalling the definition of T,

$$\begin{aligned}\mathbf{P}_y(T_y^{(k)}<\infty)&=\mathbf{P}_y(T_y^{(k)}-T_y^{(k-1)}<\infty)\\ &=\mathbf{P}_y(T_y^{(k)}-T_y^{(k-1)}<\infty\,|\,T_y^{(k-1)}<\infty)\mathbf{P}_y(T_y^{(k-1)}<\infty)\\ &=\rho_{yy}\mathbf{P}_y(T_y^{(k-1)}<\infty)\qquad\text{, by }(*)\end{aligned}$$

Iterating this equality $k-1$ times, we have shown:

Proposition 3.21. *For any integer $k \geq 1$,*

$$\mathbf{P}_y(T_y^{(k)} < \infty) = [\mathbf{P}_y(T_y < \infty)]^k = \rho_{yy}^k.$$

In particular, if $\rho_{yy} = 1$, then the Markov chain returns to y an infinite number of times. But if $\rho_{yy} < 1$, then eventually the Markov chain will not return to y:

$$\mathbf{P}_y(T_y^{(k)} = \infty \; \forall\, k \geq j) = \mathbf{P}_y(T_y^{(j)} = \infty) = 1 - \rho_{yy}^j \to 1 \text{ as } j \to \infty.$$

For this reason, we make the following definitions.

Definition 3.22 (Recurrent State, Transient State). If $\rho_{yy} = 1$, we say the state $y \in \Omega$ is **recurrent**. If $\rho_{yy} < 1$, we say the state $y \in \Omega$ is **transient**.

Example 3.23 (Gambler's Ruin). Consider the Markov Chain defined by the following 5×5 stochastic matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .6 & 0 & .4 & 0 & 0 \\ 0 & .6 & 0 & .4 & 0 \\ 0 & 0 & .6 & 0 & .4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We label the rows and columns of this matrix as $\{1, 2, 3, 4, 5\}$, so that we consider the Markov chain to have state space $\{1, 2, 3, 4, 5\}$. We think of state 1 as a Gambler going bankrupt, state 5 as a Gambler reaching a high amount of money and cashing out. And at each of the states $2, 3, 4$, the gambler can either win a round of some game with probability .4, or lose a round of the game with probability .6.

We will show that states 1 and 5 are recurrent, whereas states $2, 3, 4$ are transient.

Since $P(1, 1) = 1$, $\mathbf{P}_1(T_1 = 1) = 1$, so $\mathbf{P}_1(T_1 < \infty) = 1$. Similarly, $P(5, 5) = 1$, so $\mathbf{P}_5(T_5 = 1)$ and $\mathbf{P}_5(T_5 < \infty) = 1$. So, states 1 and 5 are recurrent.

Now, $P(2, 1) = .6$, and since $P(1, 1) = 1$, if the Markov chain reaches 1 it will never return to 2. So, using the Multiplication rule and the Markov property,

$$\begin{aligned} \mathbf{P}_2(T_2 = \infty) &\geq \mathbf{P}_2(X_1 = 1, X_2 = 1, X_3 = 1, \ldots) \\ &= \mathbf{P}(X_1 = 1 \,|\, X_0 = 2)\mathbf{P}(X_2 = 1 \,|\, X_1 = 1)\mathbf{P}(X_3 = 1 \,|\, X_2 = 1) \cdots \\ &= \lim_{n \to \infty} P(2, 1)P(1, 1)^n = P(2, 1) = .6 > 0. \end{aligned}$$

That is, $\mathbf{P}_2(T_2 < \infty) = 1 - \mathbf{P}(T_2 = \infty) \leq 1 - .6 < 1$, so that state 2 is transient. Similarly, $P(4, 5) = .4$, and $P(5, 5) = 1$, so $\mathbf{P}_4(T_4 = \infty) \geq P(4, 5) > 0$, so $\mathbf{P}_4(T_4 < \infty) < 1$, so state 4 is transient. Using similar reasoning again,

$$\mathbf{P}_3(T_3 = \infty) \geq \lim_{n \to \infty} P(3, 2)P(2, 1)P(1, 1)^n = P(3, 2)P(2, 1) > 0.$$

So, $\mathbf{P}_3(T_3 < \infty) < 1$, so state 3 is transient.

We defined the transition matrix P so that $P(x, y) = \mathbf{P}(X_1 = y \,|\, X_0 = x)$, for any x, y in the state space of the Markov chain. Powers of the matrix P have a similar interpretation. For any $n \geq 1$, $x, y \in \Omega$, define $p^{(n)}(x, y) := \mathbf{P}(X_n = y \,|\, X_0 = x)$.

Proposition 3.24 (**Chapman-Kolmogorov Equation**)**.** *Let $n, m \geq 1$. Let $x, y \in \Omega$ be states of a finite (or countable) Markov chain. Then*

$$p^{(m+n)}(x,y) = \sum_{z\in\Omega} p^{(m)}(x,z)p^{(n)}(z,y)$$

So, for any $x, y, z \in \Omega$, $p^{(m+n)}(x,y) \geq p^{(m)}(x,z)p^{(n)}(z,y)$.

Corollary 3.25. *Let $m \geq 1$. Let $x, y \in \Omega$ be states of a finite Markov chain. Then*

$$P^m(x,y) = p^{(m)}(x,y).$$

Proof of Corollary 3.25. We induct on m. The case $m = 1$ follows since by definition, $p^{(1)}(x,y) = P(x,y)$ for all $x, y \in \Omega$. We now perform the inductive step. From Proposition 3.24 with $n = 1$,

$$p^{(m+1)}(x,y) = \sum_{z\in\Omega} p^{(m)}(x,z)p^{(1)}(z,y) = \sum_{z\in\Omega} P^m(x,z)P(z,y) = P^{m+1}(x,y).$$

The second equality is the inductive hypothesis, and the last equality is the definition of matrix multiplication. □

Proof of Proposition 3.24. Let $x, y \in \Omega$. Using the Total Probability Theorem, we have

$$\begin{aligned}
p^{(m+n)}(x,y) = \mathbf{P}(X_{m+n} = y \,|\, X_0 = x) &= \sum_{z\in\Omega} \mathbf{P}(X_{m+n} = y, X_m = z \,|\, X_0 = x) \\
&= \sum_{z\in\Omega} \frac{\mathbf{P}(X_{m+n} = y, X_m = z, X_0 = x)}{\mathbf{P}(X_0 = x)} \\
&= \sum_{z\in\Omega} \frac{\mathbf{P}(X_{m+n} = y, X_m = z, X_0 = x)}{\mathbf{P}(X_m = z, X_0 = x)} \frac{\mathbf{P}(X_m = z, X_0 = x)}{\mathbf{P}(X_0 = x)} \\
&= \sum_{z\in\Omega} \mathbf{P}(X_{m+n} = y \,|\, X_m = z, X_0 = x)\mathbf{P}(X_m = z \,|\, X_0 = x).
\end{aligned}$$

Finally, the Markov property and Exercise 3.18 imply that

$$\begin{aligned}
p^{(m+n)}(x,y) &= \sum_{z\in\Omega} \mathbf{P}(X_{m+n} = y \,|\, X_m = z)\mathbf{P}(X_m = z \,|\, X_0 = x) \\
&= \sum_{z\in\Omega} \mathbf{P}(X_n = y \,|\, X_0 = z)\mathbf{P}(X_m = z \,|\, X_0 = x) = \sum_{z\in\Omega} p^{(n)}(z,y)p^{(m)}(x,z).
\end{aligned}$$

(Since we only condition on events with positive probability, we did not divide by zero.) □

Definition 3.26 (**Irreducible**)**.** A Markov chain with state space Ω and with transition matrix P is called **irreducible** if, for any $x, y \in \Omega$, there exists an integer $n \geq 1$ (which is allowed to depend on x, y) such that $P^n(x,y) > 0$. That is the Markov chain is irreducible if any state can reach any other state, with some positive probability, if the chain runs long enough.

Lemma 3.27. *Suppose we have a finite irreducible Markov chain with state space Ω. Then there exists $0 < \alpha < 1$ and there exists an integer $j > 0$ such that, for any $x, y \in \Omega$,*

$$\mathbf{P}_x(T_y > kj) \leq \alpha^k, \qquad \forall\, k \geq 1.$$

Proof. As a consequence of irreducibility, there exists $\varepsilon > 0$ and integer $j > 0$ such that, for any $x, y \in \Omega$, there exists $r(x, y) \leq j$ such that $P^{r(x,y)}(x, y) > \varepsilon$. That is, after at most j steps of the Markov chain, the chain will move from x to y with some positive probability.

$$\begin{aligned}
\mathbf{P}_x(T_y > kj) &= \mathbf{P}_x(T_y > kj \,|\, T_y > (k-1)j)\mathbf{P}_x(T_y > (k-1)j) \\
&\leq \max_{z\in\Omega} \mathbf{P}_z(T_y > j)\mathbf{P}_x(T_y > (k-1)j), \qquad \text{by Exercise 3.28} \\
&\leq \max_{z\in\Omega} \mathbf{P}_z(T_y > r(z, y))\mathbf{P}_x(T_y > (k-1)j), \qquad \text{since } r(z, y) \leq j \\
&= \max_{z\in\Omega}(1 - \mathbf{P}_z(T_y \leq r(z, y)))\mathbf{P}_x(T_y > (k-1)j) \\
&\leq \max_{z\in\Omega}(1 - P^{r(z,y)}(z, y))\mathbf{P}(T_y > (k-1)j), \qquad \text{by Exercise 3.29} \\
&\leq (1 - \varepsilon)\mathbf{P}(T_y > (k-1)j).
\end{aligned}$$

Iterating this inequality $k - 1$ times concludes the Lemma with $\alpha := 1 - \varepsilon$. □

Exercise 3.28. Let x, y be points in the state space of a finite Markov Chain $(X_0, X_1, \ldots)$. Let $T_y = \min\{n \geq 1 \colon X_n = y\}$ be the first arrival time of y. Let j, k be positive integers. Show that

$$\mathbf{P}_x(T_y > kj \,|\, T_y > (k-1)j) \leq \max_{z\in\Omega} \mathbf{P}_z(T_y > j).$$

(Hint: use Exercise 3.18)

Exercise 3.29. Let x, y be points in the state space of a finite Markov Chain $(X_0, X_1, \ldots)$ with transition matrix P. Let $T_y = \min\{n \geq 1 \colon X_n = y\}$ be the first arrival time of y. Let j be a positive integer. Show that

$$P^j(x, y) \leq \mathbf{P}_x(T_y \leq j).$$

(Hint: can you induct on j?)

Example 3.30. Consider the Markov Chain with state space $\Omega = \{1, 2, 3\}$ and transition matrix

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

Then for any x, y in the state space of the Markov chain, $P(x, y) \geq .1$. So, we can use $j = r = 1$ and $\varepsilon = .1$, $\alpha = .9$ in Lemma 3.27 to get

$$\mathbf{P}_x(T_y > k) \leq (.9)^k, \qquad \forall\, k \geq 1, \forall\, x, y \in \Omega.$$

In particular, $\mathbf{P}_y(T_y < \infty) = 1$, so all states are recurrent.

Exercise 3.31. Let x, y be any states in a finite irreducible Markov chain. Show that $\mathbf{E}_x T_y < \infty$. In particular, $\mathbf{P}_y(T_y < \infty) = 1$, so all states are recurrent.

3.3. **Stationary Distribution.**

Definition 3.32 (**Stationary Distribution**). Let P be the $m \times m$ transition matrix of a finite irreducible Markov chain with state space Ω. Let π be a $1 \times m$ row vector. We say that π is a **stationary distribution** if $\pi(x) \geq 0$ for every $x \in \Omega$, $\sum_{x\in\Omega} \pi(x) = 1$, and if π satisfies

$$\pi = \pi P.$$

As discussed above, if a stationary distribution exists, we can think of $\pi(x)$ as roughly the fraction of time that the Markov chain spends in x, when the Markov chain runs for a long period of time. Put another way, after the Markov chain has run for a long period of time, $\pi(x)$ is the probability that the Markov chain is in state x. In fact, π defines a probability law on the state space Ω: for any $A \subseteq \Omega$, define $\pi(A) := \sum_{x\in A} \pi(x)$. Then π is a probability law on Ω.

Unfortunately, even if the stationary distribution exists, it may not be unique! If there is more than one stationary distribution, then there may not be a sensible way of describing where the Markov chain could be, after a long time has passed.

In this section, we address the existence and uniqueness of a stationary distribution π.

Theorem 3.33 (**Existence**). *Suppose we have a finite irreducible Markov chain $(X_0, X_1, \ldots)$ with state space Ω and transition matrix P. Then there exists a stationary distribution π such that $\pi = \pi P$ and $\pi(x) > 0$ for all $x \in \Omega$.*

Proof. Let $y, z \in \Omega$. Let let $T_z = \min\{n \geq 1 \colon X_n = z\}$. We define $\widetilde{\pi}(y)$ to be the expected number of times the chain visits y before returning to z. That is, define

$$\widetilde{\pi}(y) = \mathbf{E}_z\left(\sum_{n=0}^{\infty} 1_{\{X_n=y,\, T_z>n\}}\right) = \sum_{n=0}^{\infty} \mathbf{P}_z(X_n = y,\, T_z > n). \qquad (*)$$

First, note that since the Markov chain is irreducible, there is always some probability that the chain starts at z and visits y before returning to z. Therefore, $\widetilde{\pi}(y) > 0$ for any $y \in \Omega$. Now, using Remark 2.24, and then Exercise 3.31,

$$\widetilde{\pi}(y) \leq \sum_{n=0}^{\infty} \mathbf{P}_z(T_z > n) = \mathbf{E}_z T_z < \infty, \qquad \forall\, y \in \Omega.$$

We now show that $\widetilde{\pi}$ satisfies $\widetilde{\pi} = \widetilde{\pi}P$. By definition of $\widetilde{\pi}$,

$$\sum_{x\in\Omega} \widetilde{\pi}(x)P(x,y) = \sum_{x\in\Omega}\sum_{n=0}^{\infty} \mathbf{P}_z(X_n = x,\, T_z > n)P(x,y). \qquad (**)$$

Consider the event $\{T_z > n\} = \{T_z \geq n+1\} = \{T_z \leq n\}^c$. That is, $\{T_z > n\}$ only depends on $X_0, \ldots, X_n$. So, the usual Markov property (rearranged a bit) says

$$\mathbf{P}_z(X_{n+1} = y,\, X_n = x,\, T_z \geq n+1) = \mathbf{P}_z(X_n = x,\, T_z \geq n+1)P(x,y).$$

Substituting this into $(**)$ and first changing the order of summation,

$$\begin{aligned}
\sum_{x\in\Omega} \widetilde{\pi}(x)P(x,y) &= \sum_{n=0}^{\infty}\sum_{x\in\Omega} \mathbf{P}_z(X_{n+1} = y,\, X_n = x,\, T_z \geq n+1) \\
&= \sum_{n=0}^{\infty} \mathbf{P}_z(X_{n+1} = y,\, T_z \geq n+1) = \sum_{n=1}^{\infty} \mathbf{P}_z(X_n = y,\, T_z \geq n) \\
&= \widetilde{\pi}(y) - \mathbf{P}_z(X_0 = y,\, T_z > 0) + \sum_{n=1}^{\infty} \mathbf{P}_z(X_n = y,\, T_z = n), \qquad \text{by } (*) \\
&= \widetilde{\pi}(y) - \mathbf{P}_z(X_0 = y) + \mathbf{P}_z(X_{T_z} = y), \qquad \text{substituting } n = T_z.
\end{aligned}$$

We now split into two cases. If $y = z$, then $\mathbf{P}_z(X_0 = y) = 1$ by definition of $\mathbf{P}_z$, and also $X_{T_z} = z = y$ by definition of T_z, so $\mathbf{P}_z(X_{T_z} = y) = 1$. If $y \neq z$, then by similar reasoning, $\mathbf{P}_z(X_0 = y) = \mathbf{P}_z(X_{T_z} = y) = 0$ In any case $-\mathbf{P}_z(X_0 = y, T_z > 0) + \mathbf{P}_z(X_{T_z} = y) = 0$. In conclusion, we have shown that

$$\widetilde{\pi} = \widetilde{\pi}P.$$

Finally, to get a stationary distribution π also satisfying $\pi = \pi P$, we just define $\pi(x) := \widetilde{\pi}(x)/\sum_{y\in\Omega}\widetilde{\pi}(y)$ for any $x \in \Omega$. □

Remark 3.34. We note in passing the following identity. By $(*)$ and Remark 2.24,

$$\sum_{y\in\Omega}\widetilde{\pi}(y) = \sum_{n=0}^{\infty}\sum_{y\in\Omega}\mathbf{P}_z(X_n = y,\, T_z > n) = \sum_{n=0}^{\infty}\mathbf{P}_z(T_z > n) = \mathbf{E}_z T_z.$$

Lemma 3.35. *Let P be the transition matrix of a finite irreducible Markov chain with state space Ω. Let $f\colon \Omega \to \mathbb{R}$ be a* ***harmonic*** *function, so that*

$$f(x) = \sum_{y\in\Omega} P(x,y)f(y), \qquad \forall\, x \in \Omega.$$

Then f is a constant function.

Proof. Since Ω is finite, there exists $x_0 \in \Omega$ such that $M := \max_{x\in\Omega} f(x) = f(x_0)$. Let $z \in \Omega$ with $P(x_0, z) > 0$, and assume that $f(z) < M$. Then since f is harmonic,

$$f(x_0) = P(x_0,z)f(z) + \sum_{y\in\Omega\colon\, y\neq z} P(x_0,y)f(y) < M\sum_{y\in\Omega}P(x_0,y) = M,$$

a contradiction. Thus, $f(z) = M$ for any $z \in \Omega$ with $P(x_0, z) > 0$.

Finally, for any $z \in \Omega$, irreducibility of P implies that there is a sequence of points $x_0, x_1, \ldots, x_k = z$ in Ω such that $P(x_i, x_{i+1}) > 0$ for every $0 \leq i < k$. So, by repeating the above argument $k-1$ times, $M = f(x_0) = f(x_1) = \cdots = f(x_k) = f(z)$. That is, $f(z) = M$ for every $z \in \Omega$. □

Theorem 3.36 (Uniqueness). *Let P be the transition matrix of a finite irreducible Markov chain. Then there exists a unique stationary distribution π such that $\pi = \pi P$.*

Proof. By Theorem 3.33, there exists at least one stationary distribution π such that $\pi = \pi P$. Let I denote the $|\Omega| \times |\Omega|$ identity matrix. Lemma 3.35 implies that the null-space of $P - I$ has dimension 1. So, by the rank-nullity theorem, the column rank of $P - I$ is $|\Omega| - 1$. Since row rank and column rank are equal, the row rank of $P - I$ is $|\Omega| - 1$. That is, the space of solutions of the row-vector equation $\mu = \mu P$ is one-dimensional (where μ denotes a $1 \times |\Omega|$ row vector.) Since this space is one-dimensional, it has only one vector whose entries sum to 1. □

The following Corollary gives a sensible way of computing the stationary distribution of an irreducible Markov chain.

Corollary 3.37. *Let P be the transition matrix of a finite irreducible Markov chain with state space Ω. If π is the unique solution to $\pi = \pi P$, then*

$$\pi(x) = \frac{1}{\mathbf{E}_x T_x}, \qquad \forall\, x \in \Omega.$$

Proof. Let $y, z \in \Omega$ and define $\widetilde{\pi}_z(y) := \widetilde{\pi}(y)$, where $\widetilde{\pi}(y)$ is defined in $(*)$ in Theorem 3.33. Also, define $\pi_z(y) := \widetilde{\pi}_z(y)/\mathbf{E}_z T_z$. Theorem 3.33 and Remark 3.34 imply that π_z is a stationary distribution such that $\pi_z = \pi_z P$. Theorem 3.36 implies that π_z does not depend on z. That is, for any $x \in \Omega$, if we define $\pi(x) := \pi_z(x)$ (for any particular $z \in \Omega$, since the expression does not depend on z), then we have $\pi = \pi P$, and

$$\pi(x) = \pi_x(x) = \frac{\widetilde{\pi}_x(x)}{\mathbf{E}_x T_x} = \frac{1}{\mathbf{E}_x T_x}.$$

In the last equality, we used $\widetilde{\pi}_x(x) = 1$, which follows by the definition of $\widetilde{\pi}_x$. (The $n = 0$ term in $\sum_{n=0}^{\infty} \mathbf{P}_x(X_n = x, T_x > n)$ is 1, and all other terms in the sum are zero.) $\square$

Exercise 3.38 (Knight Moves)**.** Consider a standard 8×8 chess board. Let V be a set of vertices corresponding to each square on the board (so V has 64 elements). Any two vertices $x, y \in V$ are connected by an edge if and only if a knight can move from x to y. (The knight chess piece moves in an L-shape, so that a single move constitutes two spaces moved along the horizontal axis followed by one move along the vertical axis (or two spaces moved along the vertical axis, followed by one move along the horizontal axis.) Consider the simple random walk on this graph. This Markov chain then represents a knight randomly moving around a chess board. For every space x on the chessboard, compute the expected return time $\mathbf{E}_x T_x$ for that space. (It might be convenient to just draw the expected values on the chessboard itself.)

Exercise 3.39 (Simplified Monopoly)**.** Let $\Omega = \{1, 2, \ldots, 10\}$. We consider Ω to be the ten spaces of a circular game board. You move from one space to the next by rolling a fair six-sided die. So, for example $P(1, k) = 1/6$ for every $2 \leq k \leq 7$. More generally, for every $j \in \Omega$ with $j \neq 5$, $P(j, k) = 1/6$ if $k = (j + i) \bmod 10$ for some $1 \leq i \leq 6$. Finally, the space 5 forces you to return to 1, so that $P(5, 1) = 1$. (Note that $\bmod 10$ denotes arithmetic modulo 10, so e.g. $7 + 5 = 2 \bmod 10$.)

Using a computer, find the unique stationary distribution of this Markov chain. Which point has the highest stationary probability? The lowest?

Compare this stationary distribution to the stationary distribution that arises from the doubly stochastic matrix: for all $j \in \Omega$, $P(j, k) = 1/6$ if $k = (j+i) \bmod 10$ for some $1 \leq i \leq 6$. (See Exercise 3.42.)

Exercise 3.40. Give an example of a Markov chain where there are at least two different stationary distributions.

Exercise 3.41. Is there a finite Markov chain where no stationary distribution exists? Either find one, or prove that no such finite Markov chain exists.

(If you want to show that no such finite Markov chain exists, you are allowed to just prove the weaker assertion that: for every stochastic matrix P, there always exists a nonzero vector π with $\pi = \pi P$.)

Exercise 3.42. Let P be the transition matrix for a finite Markov chain with state space Ω. We say that the matrix P is **doubly stochastic** if the columns of P each sum to 1. (Since P is a transition matrix, each of its rows already sum to 1.) Let π such that $\pi(x) = 1/|\Omega|$ for all $x \in \Omega$. That is, π is uniform on Ω. Show that $\pi = \pi P$.

Remark 3.43. If a finite Markov chain is not irreducible, we can divide the state space into pieces, each of which is irreducible (or transient), and then study how the Markov chain acts on each individual piece. (For a precise statement, see Theorem 1.8 in the Durrett book.)

Definition 3.44 (Reversible). Let P be the transition matrix of a finite Markov chain with state space Ω. We say that the Markov chain is **reversible** if there exists a probability distribution π on Ω satisfying the following **detailed balance condition**:

$$\pi(x)P(x,y) = \pi(y)P(y,x), \qquad \forall\, x,y \in \Omega.$$

Exercise 3.45. Give an example of a random walk on a graph that is not reversible.

Proposition 3.46 (Reversible Implies Stationary). *Let π be a probability distribution satisfying the detailed balance condition for a finite Markov chain. Then π is a stationary distribution.*

Proof. We sum both sides of the detailed balance condition over y, and use that P is stochastic to get

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

□

Exercise 3.47. Let P be the transition matrix of a finite, irreducible, reversible Markov chain with state space Ω and stationary distribution π. Let $f,g \in \mathbb{R}^{|\Omega|}$ be column vectors. Consider the following bilinear function on f,g, which is referred to as an inner product (or dot product):

$$\langle f,g\rangle := \sum_{x\in\Omega} f(x)g(x)\pi(x).$$

Show that P is self-adjoint (i.e. symmetric) in the sense that

$$\langle f,Pg\rangle = \langle Pf,g\rangle.$$

In particular (for those that have taken 115A), the spectral theorem implies that all eigenvalues of P are real.

Finally, find a transition matrix P such that at least one eigenvalue of P is not real.

Proposition 3.48. *Suppose we have a finite irreducible Markov chain with state space Ω, transition matrix P and stationary distribution π. Fix $n \geq 1$, and for any $0 \leq m \leq n$, define $\widehat{X}_m = X_{n-m}$. Then $\widehat{X}_m$ is a Markov chain with transition probabilities given by*

$$\widehat{P}(x,y) = \frac{\pi(y)P(y,x)}{\pi(x)}, \qquad \forall\, x,y \in \Omega.$$

Moreover, π is stationary for $\widehat{P}$, and we have

$$\mathbf{P}_\pi(X_0 = x_0,\ldots,X_n = x_n) = \mathbf{P}_\pi(\widehat{X}_0 = x_n,\ldots,\widehat{X}_n = x_0), \qquad \forall\, x_0,\ldots,x_n \in \Omega.$$

Proof. First, from Theorem 3.33, $\pi(x) > 0$ for all x in the state space of the Markov chain, so we have not divided by zero. Now, we first check π is stationary for $\widehat{P}$:

$$\sum_{y\in\Omega} \pi(y)\widehat{P}(y,x) = \sum_{y\in\Omega} \pi(y)\frac{\pi(x)P(x,y)}{\pi(y)} = \pi(x).$$

Using similar reasoning, we know that $\sum_{y\in\Omega}\widehat{P}(x,y)=1$, so that $\widehat{P}$ is itself a stochastic matrix. Finally, noting that $P(x_{i-1},x_i)=\pi(x_i)\widehat{P}(x_i,x_{i-1})/\pi(x_{i-1})$ for each $1\leq i\leq n$,

$$\begin{aligned}\mathbf{P}_{\pi}(X_0=x_0,\ldots,X_n=x_n)&=\pi(x_0)P(x_0,x_1)\cdots P(x_{n-1},x_n)\\&=\pi(x_n)\widehat{P}(x_n,x_{n-1})\cdots\widehat{P}(x_1,x_0)\\&=\mathbf{P}_{\pi}(\widehat{X}_0=x_n,\ldots,\widehat{X}_n=x_0).\end{aligned}$$

□

Remark 3.49. If the Markov chain is reversible, then $\widehat{P}=P$. So, being reversible means that the Markov chain can be run backwards or forwards in the same way, if we start the Markov chain from the stationary distribution.

Example 3.50. We return to Example 3.8. Let $G=(V,E)$ be a graph with at least one edge, and let P correspond to the simple random walk on G. So, $P(x,y)=1/\deg(x)$ if x and y are neighbors, and $P(x,y)=0$ otherwise. For any $x\in V$, define $\pi(x):=\deg(x)/(2\,|E|)$. We show π is stationary. From Proposition 3.46, it suffices to show the detailed balance condition holds.

If x and y are not neighbors, then $P(x,y)=P(y,x)=0$, and both sides of the detailed balance condition are equal. If x and y are neighbors, then

$$\pi(x)P(x,y)=\frac{\deg(x)}{2\,|E|}\frac{1}{\deg(x)}=\frac{1}{2\,|E|}=\frac{\deg(y)}{2\,|E|}\frac{1}{\deg(y)}=\pi(y)P(y,x).$$

Exercise 3.51 (Ehrenfest Urn Model). Suppose we have two urns and n spheres. Each sphere is in either of the first or the second urn. At each step of the Markov chain, one of the spheres is chosen uniformly at random and moved from its current urn to the other urn. Let X_n be the number of spheres in the first urn at time n. A state of the Markov chain is an integer in $\{0,1,\ldots,n\}$, which represents the number of spheres in the first urn. Then for any $j,k\in\{1,\ldots,n\}$, the transition matrix defining the Markov chain is

$$P(j,k)=\begin{cases}\frac{n-j}{n} & \text{, if } k=j+1\\ \frac{j}{n} & \text{, if } k=j-1\\ 0 & \text{, otherwise.}\end{cases}$$

Show that the unique stationary distribution for this Markov chain is a binomial PMF with parameters n and $1/2$.

Exercise 3.52. Let $V=\{0,1\}^n$ be a set of vertices. We construct a graph from V as follows. Let $x=(x_1,\ldots,x_n),y=(y_1,\ldots,y_n)\in\{0,1\}^n$. Then x and y are connected by an edge in the graph if and only if $\sum_{i=1}^n|x_i-y_i|=1$. That is, x and y are connected if and only if they differ by a single coordinate.

For any $x\in V$, define $f(x)=\sum_{i=1}^n x_i$, $f\colon V\to\{0,1,\ldots,n\}$. Given $x\in V$, we identify x with the state in the Ehrenfest urn model where the first urn has exactly $f(x)$ spheres. Show that the Ehrenfest urn model is a **projection** of the simple random walk on V in the following sense. The probability that $x\in V$ transitions to any state $z\in V$ such that $y=f(z)$ is equal to: the probability that Ehrenfest model with state $f(x)$ transitions to state y.

Moreover, the unique stationary distribution for the simple random walk on V can be projected to give the unique stationary distribution in the Ehrenfest model. That is, if π is the unique stationary distribution for the simple random walk on V, and if for any $A \subseteq \{0, 1, \ldots, n\}$, we define $\mu(A) := \pi(f^{-1}(A))$, then μ is a Binomial PMF with parameters n and $1/2$. (Here $f^{-1}(A) = \{x \in V : f(x) \in A\}$.)

Exercise 3.53 (**Birth-and-Death Chains**)**.** A birth-and-death chain can model the size of some population of organisms. Fix a positive integer k. Consider the state space $\Omega = \{0, 1, 2, \ldots, k\}$. The current state is the current size of the population, and at each step the size can increase or decrease by at most 1. We define $\{(p_n, r_n, q_n)\}_{n=0}^{k}$ such that $p_n + r_n + q_n = 1$ and $p_n, r_n, q_n \geq 0$ for each $0 \leq n \leq k$, and

- $P(n, n+1) = p_n > 0$ for every $0 \leq n < k$.
- $P(n, n-1) = q_n > 0$ for every $0 < n \leq k$.
- $P(n, n) = r_n \geq 0$ for every $0 \leq n \leq k$.
- $q_0 = p_k = 0$.

Show that the birth-and-death chain is reversible.

3.4. **Limiting Behavior.** From Theorem 3.36, we know an irreducible Markov chain has a unique stationary distribution, and Corollary 3.37 gives a sensible way of computing that stationary distribution. But what does this distribution tell us about the Markov chain's behavior? In general, it might not say anything! For example, recall Example 3.7, where we considered the transition matrix $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. If $\mu = (\mu(1), \mu(2))$ is any 1×2 row vector, then $\mu P^n = \mu$ for n even, and $\mu P^n = (\mu(2), \mu(1))$ for n odd. So, if the Markov chain starts at the probability distribution μ where $\mu(1) \neq \mu(2)$, then it is impossible for $\lim_{n\to\infty} \mu P^n$ to exist. That is, there is no sensible way of talking about the limiting behavior of this Markov chain.

Put another way, we need to eliminate this "periodic" behavior to hope to get convergence of the Markov chain. Thankfully, if an irreducible Markov chain has no "periodic" behavior as in the above example, then it does actually converge as $n \to \infty$. In fact, we will be able to give an exponential rate of convergence of the Markov chain. Before doing so, we formally define periodic behavior, and we formally define periodicity and how the Markov chain converges.

Definition 3.54 (**Period, Aperiodic**)**.** Let P be the transition matrix of a finite Markov chain with state space Ω. For any $x \in \Omega$, let $\mathcal{N}(x) := \{n \geq 1 : P^n(x, x) > 0\}$. The **period** of state $x \in \Omega$ is the largest integer that divides all of the integers in $\mathcal{N}(x)$. That is, the period of x, denoted $\gcd \mathcal{N}(x)$, is the greatest common divisor of $\mathcal{N}(x)$. (If $\mathcal{N}(x) = \emptyset$, we leave $\gcd \mathcal{N}(x)$ undefined.) (We say an integer m divides an integer n if there exists an integer k such that $n = km$.)

A Markov chain is called **aperiodic** if all $x \in \Omega$ have period 1.

Exercise 3.55. Give an explicit example of a Markov chain where every state has period 100.

Lemma 3.56. *Let P be the transition matrix of an irreducible, finite Markov chain with state space Ω. Then $\gcd \mathcal{N}(x) = \gcd \mathcal{N}(y)$ for all $x, y \in \Omega$.*

Proof. Let $x, y \in \Omega$. Since the Markov chain is irreducible, there exist $r, \ell \geq 1$ such that $P^r(x,y) > 0$ and $P^\ell(y,x) > 0$. Let $m = r + \ell$. Then $m \in \mathcal{N}(x) \cap \mathcal{N}(y)$ (since $P^m(x,x) \geq P^r(x,y)P^\ell(y,x) > 0$, and $P^m(y,y) \geq P^\ell(y,x)P^r(x,y) > 0$), and $\mathcal{N}(x) \subseteq \mathcal{N}(y) - m$. (If $P^k(x,x) > 0$, then $P^{k+m}(y,y) \geq P^\ell(y,x)P^k(x,x)P^r(x,y) > 0$.) Since $\gcd \mathcal{N}(y)$ divides m and all elements of $\mathcal{N}(y)$, we conclude that $\gcd \mathcal{N}(y)$ divides all elements of $\mathcal{N}(x)$. In particular, $\gcd \mathcal{N}(y) \leq \gcd \mathcal{N}(x)$. Reversing the roles of x and y in the above argument, $\gcd \mathcal{N}(x) \leq \gcd \mathcal{N}(y)$. □

Lemma 3.57. *Let P be the transition matrix of an aperiodic, irreducible, finite Markov chain with state space Ω. Then there exists an integer $r > 0$ such that $P^r(x,y) > 0$ for all $x, y \in \Omega$. (That is, we can choose the r to not depend on x, y.)*

Proof. Since the Markov chain is aperiodic, $\gcd \mathcal{N}(x) = 1$. The set $\mathcal{N}(x)$ is closed under addition, since if $n, m \in \mathcal{N}(x)$, then $P^{n+m}(x,x) \geq P^n(x,x)P^m(x,x) > 0$, so that $n + m \in \mathcal{N}(x)$. From Lemma 3.58 with $g = 1$, there exists $n(x)$ such that if $n \geq n(x)$, then $n \in \mathcal{N}(x)$. Since the Markov chain is irreducible, for any $y \in \Omega$ there exists $r = r(x,y)$ such that $P^r(x,y) > 0$. So, if $n \geq n(x) + r$, we have

$$P^n(x,y) \geq P^{n-r}(x,x)P^r(x,y) > 0.$$

So, if $n \geq n'(x) := n(x) + \max_{x,y\in\Omega} r(x,y)$, then $P^n(x,y) > 0$ for all $y \in \Omega$. Then, if $n \geq \max_{x\in\Omega} n'(x)$, then $P^n(x,y) > 0$ for all $x, y \in \Omega$. □

Lemma 3.58. *Let S be a nonempty subset of the positive integers. Let $g = \gcd(S)$. Then there exists some integer n_S such that, for all $m \geq n_S$, the product mg can be written as a linear combination of elements of S, with nonnegative integer coefficients.*

Proof. Let g^* be the smallest positive integer which is an integer combination of elements of S. Then $g^* \leq s$ for every $s \in S$. Also, g^* divides every element of S (if $s \in S$ and if g^* does not divide s, then the remainder obtained by dividing s by g^* would be smaller than g^*, while being an integer combination of elements of S). So, $g^* \leq g$. Since g divides every element of S as well, g divides g^*, and $g \leq g^*$. So, $g = g^*$.

Now, without loss of generality, we can assume S is finite, since the case that S is infinite follows from the case that S is finite. The case when S has one element is clear. As a base case, we consider when $S = \{a, b\}$, where a, b are distinct positive integers. Let $m > 0$. Since $g = g^*$ and $mg \geq g^*$, we can write $mg = ca + db$ for some integers c, d. Since $mg = ca + db$, we can also write $mg = (c+kb)a + (d-ka)b$ for any k. That is, we can write $mg = ca + db$ for integers c, d with $0 \leq c \leq b-1$. If $mg > (b-1)a - b$, then $db = mg - ca \geq mg - a(b-1) > -b$. So, $d \geq 0$ as well. That is, we can choose n_S such that $n_S \geq ((ab - a - b)/g) + 1$.

We now induct on the size of S, by adding one element a to S. Let $g_S := \gcd(S)$ and let $g := \gcd(\{a\} \cup S)$. For any positive integer a, the definition of gcd implies that $\gcd(\{a\} \cup S) = \gcd(a, g_S)$. Suppose m satisfies $mg \geq n_{\{a,g_S\}}g + n_S g_S$. Then we can write $mg - n_S g_S = ca + dg_S$ for integers $c, d \geq 0$, from the case when S could be $\{a, g_S\}$. Therefore, $mg = ca + (d + n_S)g_S = ca + \sum_{s\in S} c_s s$ for some integers $c_s \geq 0$, by definition of n_S, and using $d + n_S \geq n_S$. In conclusion, we can choose $n_{\{a\}\cup S} = n_{\{a,g_S\}} + n_S g_S/g$, completing the inductive step. □

Definition 3.59 (**Total Variation Distance**). Let μ, ν be probability distributions on a finite state space Ω. We define the **total variation distance** between μ and ν to be

$$\|\mu - \nu\|_{\mathrm{TV}} := \max_{A \subseteq \Omega} |\mu(A) - \nu(A)| .$$

Exercise 3.60. Let Ω be a finite state space. This exercise demonstrates that the total variation distance is a metric. That is, the following three properties are satisfied:

- $\|\mu - \nu\|_{\mathrm{TV}} \geq 0$ for all probability distributions μ, ν on Ω, and $\|\mu - \nu\|_{\mathrm{TV}} = 0$ if and only if $\mu = \nu$.
- $\|\mu - \nu\|_{\mathrm{TV}} = \|\nu - \mu\|_{\mathrm{TV}}$
- $\|\mu - \nu\|_{\mathrm{TV}} \leq \|\mu - \eta\|_{\mathrm{TV}} + \|\eta - \nu\|_{\mathrm{TV}}$ for all probability distributions μ, ν, η on Ω.

(Hint: you may want to use the triangle inequality for real numbers: $|x - y| \leq |x - z| + |z - y|$, $\forall\, x, y, z \in \mathbb{R}$.)

Exercise 3.61. Let μ, ν be probability distributions on a finite state space Ω. Then

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| .$$

(Hint: consider the set $A = \{x \in \Omega \colon \mu(x) \geq \nu(x)\}$.)

Theorem 3.62 (**The Convergence Theorem**). *Let P be the transition matrix of a finite, irreducible, aperiodic Markov chain, with state space Ω and with (unique) stationary distribution π. Then there exist constants $\alpha \in (0, 1)$ and $C > 0$ such that*

$$\max_{x \in \Omega} \|P^n(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leq C\alpha^n, \qquad \forall\, n \geq 1.$$

Proof. Since the Markov chain is irreducible and aperiodic, Lemma 3.57 implies there exists $r > 0$ such that all entries of P^r are positive. Let Π be the matrix with $|\Omega|$ rows, each of which is the row vector π (so $\Pi = (1, \ldots, 1)^T \pi$). From Theorem 3.33 (and Theorem 3.36), $\min_{z \in \Omega} \pi(z) > 0$. So, there exists $0 < \delta < 1$ such that

$$P^r(x, y) \geq \delta \pi(y), \qquad \forall\, x, y \in \Omega.$$

From Exercise 3.3, P^r is a stochastic matrix. Also, Π is a stochastic matrix. Let $\theta := 1 - \delta$. Define $Q := \theta^{-1}(P^r - (1 - \theta)\Pi)$. Then Q is a stochastic matrix, and

$$P^r = (1 - \theta)\Pi + \theta Q.$$

If M is an $|\Omega| \times |\Omega|$ stochastic matrix, then $M\Pi = \Pi$ (since $M\Pi = M(1, \ldots, 1)^T \pi = (1, \ldots, 1)^T \pi = \Pi$.) Similarly, if M satisfies $\pi M = \pi$, then $\Pi M = \Pi$. We now prove by induction that, for all $k \geq 1$,

$$P^{rk} = (1 - \theta^k)\Pi + \theta^k Q^k. \qquad (*)$$

We already know $k = 1$ holds, by the definition of Q. Assume $(*)$ holds for all $1 \leq k \leq n$. Then using $(*)$ twice,

$$\begin{aligned} P^{r(n+1)} &= P^{rn} P^r = [(1 - \theta^n)\Pi + \theta^n Q^n] P^r \\ &= (1 - \theta^n)\Pi P^r + (1 - \theta)\theta^n Q^n \Pi + \theta^{n+1} Q^{n+1} \\ &= (1 - \theta^n)\Pi + (1 - \theta)\theta^n \Pi + \theta^{n+1} Q^{n+1}, \qquad \text{since } \pi P = \pi, \text{ so } \pi P^n = \pi, \text{ and } Q^n \text{ is stochastic} \\ &= (1 - \theta^{n+1})\Pi + \theta^{n+1} Q^{n+1}. \end{aligned}$$

So, we have completed the inductive step, i.e. we have shown $(*)$ holds for all $k \geq 1$.

Let $j \geq 1$. Multiplying $(*)$ by P^j on the right and rearranging,

$$P^{rk+j} - \Pi = \theta^k (Q^k P^j - \Pi). \qquad (**)$$

From Exercise 3.3, $Q^k P^j$ is a stochastic matrix. Fix $x \in \Omega$. Sum up the absolute values of all the entries in row x of both sides of $(**)$ and divide by 2. By Exercise 3.61, the term on the right is then θ^k multiplied by the total variation distance between two probability distributions, which is at most 1, by definition of total variation distance. That is, the right side is at most θ^k. So, using Exercise 3.61 for the left side as well,

$$\left\| P^{rk+j}(x, \cdot) - \pi(\cdot) \right\|_{\mathrm{TV}} \leq \theta^k, \qquad \forall\, j, k \geq 1.$$

Taking the maximum of both sides over $x \in \Omega$, and writing an arbitrary positive integer n as $n = rk + j$ where $0 \leq j < r$ by Euclidean division of n by r (so that $k = (n/r) - (j/r) \geq (n/r) - 1$), we get the bound

$$\max_{x \in \Omega} \left\| P^n(x, \cdot) - \pi(\cdot) \right\|_{\mathrm{TV}} \leq \theta^{-1} (\theta^{1/r})^n.$$

Setting $C := \theta^{-1}$ and $\alpha := \theta^{1/r}$ completes the proof. □

3.5. Infinite State Spaces.

Definition 3.63 (**Markov Chain, Countable State Space**). Let Ω be a countable set. A Markov chain on a countable state space Ω is defined, as before, by its transition matrix $P\colon \Omega \times \Omega \to [0, 1]$, where $\sum_{y \in \Omega} P(x, y) = 1$ for all $x \in \Omega$. The remaining defining properties are stated in the same way as in the finite case. We can still think of P as a matrix, albeit one with countably many rows and columns.

Unfortunately, the Convergence Theorem (Theorem 3.62), may not hold for all irreducible, aperiodic Markov chains on infinite state spaces. So, studying the existence/non-existence of stationary distributions is not as meaningful for infinite state spaces. However, we can still try to understand where the Markov chain "typically" lies after the chain runs for a long time.

To see why the Convergence Theorem cannot hold for all irreducible, aperiodic Markov chains, just note that *all* states of the Markov chain could be transient. (We will show this below in Exercise 3.70; all states are transient for the nearest neighbor simple random walk on $\mathbb{Z}^3$, though we will not show this.) And if all states in the chain are transient, then $\lim_{n \to \infty} P^n(x, x)$ must converge to 0.

Note that by Exercise 3.31, all states in a finite irreducible Markov chain are recurrent, so having all transient states can only happen for an irreducible Markov chain when the state space is infinite.

Rather than delving into a general theory of infinite state space Markov chains (which can become a bit more complicated than the finite case), we focus on some classic examples.

Example 3.64 (**Nearest-Neighbor Random Walk on $\mathbb{Z}$**). Let $\Omega = \mathbb{Z}$. Let $p, r, q \geq 0$ such that $p + r + q = 1$. We define the transition matrix P so that

$$P(k, k+1) = p, \qquad P(k, k) = r, \qquad P(k, k-1) = q.$$

The case $p = q = 1/2$ and $r = 0$ corresponds to the simple random walk on $\mathbb{Z}$. Let $k \in \mathbb{Z}$ and let $n \geq 0$. If $X_n = k$, then $\sum_{j=1}^{n} (X_j - X_{j-1}) = k$, and each term in the sum is an independent random variable, each with probability 1/2 of being 1 and probability 1/2 of being -1. To sum to k, there must be $(n+k)/2$, 1's and $n-(n+k)/2=(n-k)/2$, -1's. There are $\binom{n}{(n+k)/2}$ different ways to choose the 1's and -1's to sum to k. So,

$$\mathbf{P}_0(X_n=k)=\begin{cases}\binom{n}{(n+k)/2}2^{-n} & \text{, if } n-k \text{ is even}\\ 0 & \text{, otherwise.}\end{cases}$$

The case $p=q=1/4$ and $r=1/2$ is the lazy simple random walk on $\mathbb{Z}$.

Exercise 3.65. Let $(X_0,X_1,\ldots)$ be the simple random walk on $\mathbb{Z}$. Show that $\mathbf{P}_0(X_n=0)$ decays like $1/\sqrt{n}$ as $n\to\infty$. That is, show

$$\lim_{n\to\infty}\sqrt{2n}\,\mathbf{P}_0(X_{2n}=0)=\sqrt{\frac{2}{\pi}}.$$

Also, show the upper bound

$$\mathbf{P}_0(X_n=k)\leq\frac{10}{\sqrt{n}},\qquad\forall\, n\geq 0,\ k\in\mathbb{Z}.$$

(Hint 1: first consider the case $n=2r$ for $r\in\mathbb{Z}$. It may be helpful to show that $\binom{2r}{r+j}$ is maximized when $j=0$. To eventually deal with k odd, just condition on the first step of the walk.)

(Hint 2: you can freely use **Stirling's formula**:

$$\lim_{n\to\infty}\frac{n!}{\sqrt{2\pi n}(n/e)^n}=1.$$

Or, there is a more precise estimate: for any $n\geq 3$, there exists $1/(12n+1)\leq\varepsilon_n\leq 1/(12n)$ such that

$$n!=\sqrt{2\pi}e^{-n}n^{n+1/2}e^{\varepsilon_n}.)$$

We can get an upper bound matching Exercise 3.65 even when the simple random walk starts away from 0.

Theorem 3.66. *Let $(X_0,X_1,\ldots)$ be the simple random walk on $\mathbb{Z}$. Let $k,r>0$ be integers. We will start the Markov chain at k and upper bound $T_0:=\min\{n>0\colon X_n=0\}$, the first time the random walk hits 0.*

$$\mathbf{P}_k(T_0>r)\leq\frac{20k}{\sqrt{r}}.$$

Before proving Theorem 3.66, we prove some lemmas.

Lemma 3.67 (Reflection Principle). *Let $(X_0,X_1,\ldots)$ be the simple random walk on $\mathbb{Z}$ or the lazy simple random walk on $\mathbb{Z}$. For any positive integers j,k,r,*

$$\mathbf{P}_k(T_0<r,\,X_r=j)=\mathbf{P}_k(X_r=-j).$$

$$\mathbf{P}_k(T_0<r,\,X_r>0)=\mathbf{P}_k(X_r<0).$$

Proof. From the Strong Markov property, if the walk hits zero, then the walk is independent of its previous movements, and we can then treat the walk as if it started at 0. That is, for any integers $0<s<r$ and j,

$$\mathbf{P}_k(X_{T_0+(r-s)}=j\,|\,T_0=s,\,X_s=0)=\mathbf{P}_0(X_{r-s}=j).$$

Rearranging and simplifying,

$$\mathbf{P}_k(T_0 = s,\, X_r = j) = \mathbf{P}_k(T_0 = s)\mathbf{P}_0(X_{r-s} = j). \qquad (*)$$

When the Markov chain starts at zero, it has equal probability of reaching j or $-j$ (that is, the random walk is symmetric with respect to zero). So, the right side is equal to

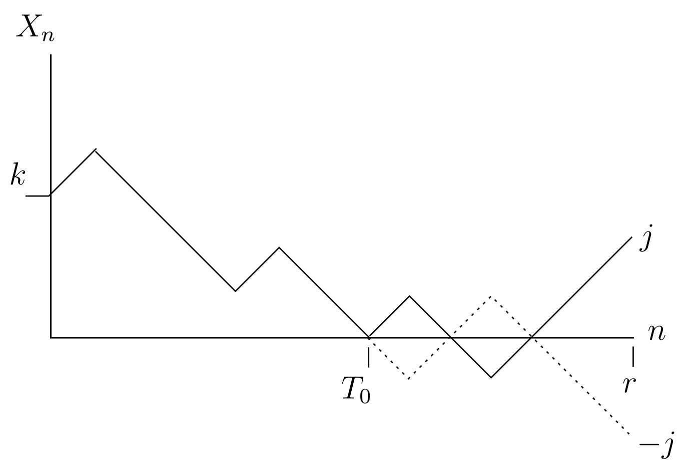
$$\mathbf{P}_k(T_0 = s)\mathbf{P}_0(X_{r-s} = -j) \overset{(*)}{=} \mathbf{P}_k(T_0 = s,\, X_r = -j).$$

Summing over all $1 \leq s < r$, and combining this equality with $(*)$ (with $j > 0$),

$$\mathbf{P}_k(T_0 < r,\, X_r = j) = \mathbf{P}_k(T_0 < r,\, X_r = -j) = \mathbf{P}_k(X_r = -j).$$

The last equality follows since a random walk started from $k > 0$ must pass through 0 before reaching a negative integer $-j$. That is, given $X_0 = k$, the event $X_r = -j$ is contained in the event $T_0 < r$.

Finally, summing over $j > 0$ gives the final equality of the Lemma. □



Remark 3.68. We can interpret Lemma 3.67 combinatorially as follows. We plot the sequence of points visited by the Markov chain in the plane as $(n, X_n) \in \mathbb{R}^2$, $n \geq 0$. Then there is a bijection from the set of paths starting at $k > 0$ which hit 0 before time r and are positive at time r, and the set of paths starting at $k > 0$ which are negative at time r. To create the bijection, reflect a path across the line $y = 0$ after the first time it hits 0.

Lemma 3.69. *Let $(X_0, X_1, \ldots)$ be the simple random walk on $\mathbb{Z}$ or the lazy simple random walk on $\mathbb{Z}$. For any $r, k > 0$*

$$\mathbf{P}_k(T_0 > r) = \mathbf{P}_0(-k < X_r \leq k).$$

Proof. First, write

$$\mathbf{P}_k(X_r > 0) = \mathbf{P}_k(X_r > 0,\, T_0 \leq r) + \mathbf{P}_k(T_0 > r) = \mathbf{P}_k(X_r > 0,\, T_0 < r) + \mathbf{P}_k(T_0 > r).$$

Applying the Reflection Principle (Lemma 3.67), we then get

$$\mathbf{P}_k(X_r > 0) = \mathbf{P}_k(X_r < 0) + \mathbf{P}_k(T_0 > r).$$

Since the walk is symmetric, $\mathbf{P}_k(X_r < 0) = \mathbf{P}_k(X_r > 2k)$, so rearranging and then using translation invariance of the Markov chain,

$$\mathbf{P}_k(T_0 > r) = \mathbf{P}_k(X_r > 0) - \mathbf{P}_k(X_r > 2k) = \mathbf{P}_k(0 < X_r \leq 2k) = \mathbf{P}_0(-k < X_r \leq k).$$

□

Proof of Theorem 3.66. Summing the upper bound of Exercise 3.65, we have

$$\mathbf{P}_0(-k < X_r \leq k) \leq \frac{20k}{\sqrt{r}}.$$

Then Lemma 3.69 completes the proof. □

Exercise 3.70. Show that every state in the simple random walk on $\mathbb{Z}$ is recurrent. (You should show this statement for any starting location of the Markov chain.)

Then, find a nearest-neighbor random walk on $\mathbb{Z}$ such that every state is transient.

Exercise 3.71. For the simple random walk on $\mathbb{Z}$, show that $\mathbf{E}_0 T_0 = \infty$. Conclude that, for any $x, y \in \mathbb{Z}$, $\mathbf{E}_x T_y = \infty$.

Exercise 3.72. Let $(X_0, X_1, \ldots)$ be the "corner walk" on $\mathbb{Z}^2$. The transitions are described as follows. From any point $(x, y) \in \mathbb{Z}^2$, the Markov chain adds any of the following four vector to (x, y) each with probability $1/4$: $\{(1, 1,), (1, -1), (-1, 1), (-1, -1)\}$. Using that the coordinates of this walk are each independent simple random walks on $\mathbb{Z}$, conclude that there exists $c > 0$ such that

$$\lim_{n\to\infty} n\mathbf{P}_{(0,0)}(X_{2n} = (0, 0)) = c.$$

That is, $\mathbf{P}_{(0,0)}(X_{2n} = (0, 0))$ is about c/n, when n is large.

Now, note that the usual nearest-neighbor simple random walk on $\mathbb{Z}^2$ is a rotation of the corner walk by an angle of $\pi/4$. So, the above limiting statement also holds for the simple random walk on $\mathbb{Z}^2$.

4. Martingales

4.1. Review of Conditional Expectation.

Definition 4.1 (Conditional Expectation). Let X be a random variables on a sample space $\mathcal{C}$. Let $A \subseteq \mathcal{C}$ with $\mathbf{P}(A) > 0$. Then the **conditional expectation of X given A**, denoted $\mathbf{E}(X|A)$ is

$$\mathbf{E}(X|A) := \frac{\mathbf{E}(X \cdot 1_A)}{\mathbf{P}(A)}.$$

Equivalently, $\mathbf{E}(X|A)$ is the expectation of X with respect to the conditional probability $\mathbf{P}(B|A) := \mathbf{P}(B \cap A)/\mathbf{P}(A)$, for any $B \subseteq \mathcal{C}$. To see the equivalence, note that the expectation of $X \geq 0$ with respect to $\mathbf{P}(\cdot|A)$ is

$$\int_0^\infty \mathbf{P}(X > t|A)dt = \frac{1}{\mathbf{P}(A)} \int_0^\infty \mathbf{P}(X > t,\ A)dt = \frac{1}{\mathbf{P}(A)} \int_0^\infty \mathbf{P}(X1_A > t)dt = \frac{\mathbf{E}(X \cdot 1_A)}{\mathbf{P}(A)}.$$

Example 4.2. Suppose a random variable X and a set $A \subseteq \mathcal{C}$ are independent. That is, $\mathbf{P}(X \in B, A) = \mathbf{P}(X \in B)\mathbf{P}(A)$ for all $B \subseteq \mathbb{R}$. Then $\mathbf{P}(X \in B, A^c) = \mathbf{P}(X \in B)\mathbf{P}(A^c)$ for all $B \subseteq \mathbb{R}$. Consequently, X and 1_A are independent as random variables. So, from Proposition 2.39, $\mathbf{E}(X1_A) = (\mathbf{E}X)(\mathbf{E}1_A) = \mathbf{P}(A)\mathbf{E}X$. That is, if X, A are independent, then

$$\mathbf{E}(X|A) = \mathbf{E}X.$$

Also, if X, Y are random variables, then since $\mathbf{E}(X|A)$ is expectation of X with respect to a conditional probability, we immediately have from Proposition 2.27

$$\mathbf{E}(X + Y|A) = \mathbf{E}(X|A) + \mathbf{E}(Y|A).$$

Remark 4.3. Let $A_1, \ldots, A_k$ be sets and let X be a random variables. We use the notation

$$\mathbf{E}(X \mid A_1, \ldots, A_k) = \mathbf{E}(X \mid A_1 \cap \cdots \cap A_k).$$

Lemma 4.4. *Let X, Y be random variables on a sample space $\mathcal{C}$. Let $A \subseteq \mathcal{C}$ and let $d \in \mathbb{R}$. If X is a random variable such that $X = d$ on the set A, then*

$$\mathbf{E}(XY|A) = d\mathbf{E}(Y|A).$$

Proof. Since $X = d$ on A, $XY1_A = dY1_A$, so $\mathbf{E}(XY1_A) = d\mathbf{E}(Y1_A)$. Dividing by $\mathbf{P}(A)$ concludes the Lemma. □

As stated in Definition 4.1, conditional expectation is itself an expected value with respect to a conditional probability. In particular, Jensen's inequality (Proposition 2.31) applies to conditional expectation

Lemma 4.5 (Jensen's Inequality). *Let X be a random variable on a sample space $\mathcal{C}$. Let $A \subseteq \mathcal{C}$. Let $\phi\colon \mathbb{R} \to \mathbb{R}$ be convex. Then*

$$\phi(\mathbf{E}(X|A)) \leq \mathbf{E}(\phi(X)|A).$$

Lemma 4.6. *Let $A_1, \ldots, A_k$ be disjoint events such that $\cup_{i=1}^k A_i = B$. Let X be a random variable. Then*

$$\mathbf{E}(X|B) = \sum_{i=1}^k \mathbf{E}(X|A_i)\frac{\mathbf{P}(A_i)}{\mathbf{P}(B)}.$$

In particular, if $B = \mathcal{C}$, we get the Total Expectation Theorem: $\mathbf{E}X = \sum_{i=1}^k \mathbf{E}(X|A_i)\mathbf{P}(A_i)$.

Proof. By assumption, $1_B = \sum_{i=1}^k 1_{A_i}$. So,

$$\mathbf{E}(X|B) = \frac{1}{\mathbf{P}(B)}\mathbf{E}(X1_B) = \sum_{i=1}^k \frac{1}{\mathbf{P}(B)}\mathbf{E}(X1_{A_i}) = \sum_{i=1}^k \mathbf{E}(X|A_i)\frac{\mathbf{P}(A_i)}{\mathbf{P}(B)}$$

□

Definition 4.7 (Martingale). Let $(X_0, X_1, \ldots)$ be a real-valued stochastic process. A **real-valued martingale with respect to** $(X_0, X_1, \ldots)$ is a stochastic process $(M_0, M_1, \ldots)$ such that $\mathbf{E}\,|M_n| < \infty$ for all $n \geq 0$, and for any $m_0, x_0, \ldots, x_n \in \mathbb{R}$,

$$\mathbf{E}(M_{n+1} - M_n|X_n = x_n, \ldots, X_0 = x_0, M_0 = m_0) = 0.$$

We say $(M_0, M_1, \ldots)$ is a **supermartingale** with respect to $(X_0, X_1, \ldots)$ if

$$\mathbf{E}(M_{n+1} - M_n|X_n = x_n, \ldots, X_0 = x_0, M_0 = m_0) \leq 0.$$

We say $(M_0, M_1, \ldots)$ is a **submartingale** with respect to $(X_0, X_1, \ldots)$ if

$$\mathbf{E}(M_{n+1} - M_n | X_n = x_n, \ldots, X_0 = x_0, M_0 = m_0) \geq 0.$$

Remark 4.8. Some martingales are not Markov chains. Some Markov chains are not martingales. Some Markov chains are martingales. And some martingales are Markov chains.

Remark 4.9. A stochastic process is a martingale if and only if it is both a submartingale and a supermartingale.

Remark 4.10. It follows from the Total Expectation Theorem that $\mathbf{E}(M_{n+1} - M_n) = 0$ for a martingale, for every $n \geq 0$. Consequently,

$$\mathbf{E}M_n = \mathbf{E}M_0, \qquad \forall n \geq 0.$$

That is, a martingale does not change in expectation.

Similarly, a supermartingale decreases in expectation, and a submartingale increases in expectation. This terminology may then seem a bit backwards, but it is standard.

For many purposes, it is more natural to think of a conditional expectation as another random variable, rather than just a number.

Definition 4.11 (**Conditional Expectation**)**.** Suppose we have a partition of a sample space $\mathcal{C}$. That is, we have sets $A_1, \ldots, A_k \subseteq \mathcal{C}$ such that $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, \ldots, k\}$ with $i \neq j$, and $\cup_{i=1}^k A_i = \mathcal{C}$. Denote $\mathcal{A} = \{A_1, \ldots, A_k\}$. Define $\mathbf{E}(X|\mathcal{A})$ to be a random variable that takes the value $\mathbf{E}(X|A_i)$ on the set A_i. That is, $\mathbf{E}(X|\mathcal{A})$ is itself a function on the sample space $\mathcal{C}$.

Remark 4.12. Lemma 4.6 with $B = \mathcal{C}$ (i.e. the Total Expectation Theorem) can be rewritten as

$$\mathbf{E}[\mathbf{E}(X|\mathcal{A})] = \mathbf{E}(X)$$

Also, Lemma 4.4 says: if for each $1 \leq i \leq k$, X is constant on A_i, then

$$\mathbf{E}(XY|\mathcal{A}) = X\mathbf{E}(X|\mathcal{A}).$$

Exercise 4.13. Let $\mathcal{C} = [0, 1]$. Let $\mathbf{P}$ be the uniform probability law on $\mathcal{C}$. Let $X : [0, 1] \to \mathbb{R}$ be a random variable such that $X(t) = t^2$ for all $t \in [0, 1]$. Let

$$\mathcal{A} = \{[0, 1/4), [1/4, 1/2), [1/2, 3/4), [3/4, 1]\}$$

Compute explicitly the function $\mathbf{E}(X|\mathcal{A})$. (It should be constant on each of the partition elements.) Draw the function $\mathbf{E}(X|\mathcal{A})$ and compare it to a drawing of X itself.

Now, for every integer $k > 1$, let $s = 2^{-k}$, and let $\mathcal{A}_k := \{[0, s), [s, 2s), [2s, 3s), \ldots, [1 - 2s, 1 - s), [1 - s, 1)\}$. Try to draw $\mathbf{E}(X|\mathcal{A}_k)$. Convince yourself of the following fact (you can prove it if you want, but you do not have to): for every $t \in [0, 1]$

$$\lim_{k \to \infty} \mathbf{E}(X|\mathcal{A}_k)(t) = X(t).$$

The purpose of this exercise is to demonstrate that $\mathbf{E}(X|\mathcal{A})$ is given by averaging X over each partition element, such that $\mathbf{E}(X|\mathcal{A})$ is constant on each partition element of $\mathcal{A}$.

Exercise 4.14. Let X be a random variable with finite variance, and let $t \in \mathbb{R}$. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(t) = \mathbf{E}(X-t)^2$. Show that the function f is uniquely minimized when $t = \mathbf{E}X$. That is, $f(\mathbf{E}X) < f(t)$ for all $t \in \mathbb{R}$ such that $t \neq \mathbf{E}X$. Put another way, setting t to be the mean of X minimizes the quantity $\mathbf{E}(X-t)^2$ uniquely.

The conditional expectation, being a piecewise version of taking an average, has a similar property. Let $A_1, \ldots, A_k \subseteq \mathcal{C}$ such that $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, \ldots, k\}$ with $i \neq j$, and $\cup_{i=1}^k A_i = \mathcal{C}$. Write $\mathcal{A} = \{A_1, \ldots, A_k\}$. By definition, for each $1 \leq i \leq k$, $\mathbf{E}(X|\mathcal{A})$ is constant on A_i. Now, let Y be any other random variable such that, for each $1 \leq i \leq k$, Y is constant on A_i. Show that the quantity $\mathbf{E}(X-Y)^2$ is uniquely minimized by such a Y only when $Y = \mathbf{E}(X|\mathcal{A})$.

Exercise 4.15. Let $\mathcal{C} = [0,1]$. Let $\mathbf{P}$ be the uniform probability law on $\mathcal{C}$. Let $X\colon [0,1] \to \mathbb{R}$ be a random variable such that $X(t) = t^2$ for all $t \in [0,1]$. For every integer $k > 1$, let $s = 2^{-k}$, let $\mathcal{A}_k := \{[0,s), [s,2s), [2s,3s), \ldots, [1-2s, 1-s), [1-s, 1)\}$, and let $M_k := \mathbf{E}(X|\mathcal{A}_k)$. Show that the increments $M_2 - M_1, M_3 - M_2, \ldots$ are orthogonal in the following sense. For any $i, j \geq 1$ with $i \neq j$,

$$\mathbf{E}(M_{i+1} - M_i)(M_{j+1} - M_j) = 0.$$

This property is sometimes called **orthogonality of martingale increments**. This property holds for many martingales, but we will not prove this.

4.2. **Examples of Martingales.**

Example 4.16 (Random Walk). Let $X_1, X_2, \ldots$ be independent identically distributed random variables. Assume also that $\mathbf{E}\,|X_1| < \infty$. Let $\mu := \mathbf{E}X_1$. For any $n \geq 1$, define $M_n := X_1 + \cdots + X_n - \mu n$. Let $M_0 := 0$ and let $X_0 := 0$. Then $(M_0, M_1, \ldots)$ is a martingale with respect to $(X_0, X_1, \ldots)$. Indeed, for any $m_0, x_0, \ldots, x_n$, using Example 4.2,

$$\begin{aligned}&\mathbf{E}(M_{n+1} - M_n | X_n = x_n, \ldots, X_0 = x_0, M_0 = m_0)\\ &\quad= \mathbf{E}(X_{n+1} - \mu | X_n = x_n, \ldots, X_0 = x_0, M_0 = m_0) = \mathbf{E}(X_{n+1}) - \mu = 0.\end{aligned}$$

Example 4.17 (Gambler's Ruin). Let $0 < p < 1$. Suppose you are playing a game of chance. For each round of the game, with probability p you win \$1 and with probability $1-p$ you lose \$1. Suppose you start with \$50 and you decide to quit playing when you reach either \$0 or \$100. With what probability will you end up with \$100?

Later on, we will answer this question using Martingales and Stopping Times.

Let $(X_1, X_2, \ldots)$ be independent random variables such that $\mathbf{P}(X_n = 1) =: p$ and $\mathbf{P}(X_n = -1) = 1 - p =: q$ $\forall\, n \geq 1$. Let $X_0 := 50$. Let $Y_n = X_0 + \cdots + X_n$, and let $M_n := (q/p)^{Y_n}$ $\forall\, n \geq 1$. Then Y_n denotes the amount of money you have at time $n \leq 50$. We claim that $M_0, M_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$. Indeed,

$$\begin{aligned}&\mathbf{E}((q/p)^{Y_{n+1}} - (q/p)^{Y_n} \,|\, X_n = x_n, \ldots, X_0 = x_0, M_0 = m_0)\\ &\quad= (q/p)^{x_0 + \cdots + x_n}\mathbf{E}((q/p)^{X_{n+1}} - 1 \,|\, X_n = x_n, \ldots, X_0 = x_0, M_0 = m_0)\\ &\quad= (q/p)^{x_0 + \cdots + x_n}\mathbf{E}((q/p)^{X_{n+1}} - 1) = (q/p)^{x_0 + \cdots + x_n}(p(q/p) + q(p/q) - 1) = 0.\end{aligned}$$

4.3. **Gambling Strategies.**

Example 4.18. Suppose you can bet any amount of money you want on a fair coin flip. And the coin can be flipped any number of times, i.e. you can play this game any number of times. If you bet \$$d$ with $d > 0$ and the coin lands heads, then you win \$$d$, but if the coin lands tails, then you lose \$$d$. A naive strategy to make money off of this game is the following. Just keep doubling your bet until you win. For example, start by betting \$1. If you lose, bet \$2. If you lose that, bet \$4. Then let's say you finally won, then in total you won \$4 and you lost \$3, so you gained \$1 in total. We know that the probability of losing $k > 0$ rounds of this game in a row is 2^{-k}, so it seems like this strategy must win money. However, there are some caveats to this analysis.

First, if your starting bet is \$1, and if you lose twenty rounds of the game in a row, you will be betting over one million dollars. More generally, if you lose k times in a row, you will have to bet \$$2^k$. So, when $k \geq 20$, most people would not be able to continue playing the game, i.e. they would lose all of their money.

Second, *your expected gain from every round of the game is zero*. At each round of the game, no matter what your bet is, your expected earnings are zero. So, it is impossible to win money in this game, in expectation. And indeed, the Law of Large Numbers (Theorem 2.45) assures us that when the game is repeated many times, we will earn zero dollars on average, with probability 1.

It turns out that, no matter what betting strategy is chosen in this game, there is still no way to make any money. We will prove this using martingale methods. And indeed, these gambling strategies are the first studied examples of martingales.

Let $X_1, X_2, \ldots$ each be independent random variables such that $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = 1/2$ for every $i \geq 0$. For any $n \geq 1$, let $M_n = X_1 + \cdots + X_n$. Let $M_0 = 0$. If someone bets one dollar at every round of the game, then their profit is M_n after the n^{th} round of the game. Since $\mathbf{E}X_1 = 0$, Example 4.16 implies that $M_0, M_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$. A gambling strategy for the n^{th} round of the game can use any information from the previous rounds of the game. Let H_n be the amount of money we bet in the n^{th} round of the game. We assume that H_n is a function of $X_{n-1}, \ldots, X_1, M_0$, and we call the random variables $H_1, H_2, \ldots$ a **predictable process**. That is, for every $n \geq 1$, there exists a function $f_n \colon \mathbb{R}^n \to \mathbb{R}$ such that $H_n = f_n(X_{n-1}, \ldots, X_1, M_0)$. When the m^{th} round of the game occurs, we earn $H_m(M_m - M_{m-1})$ dollars. In summary, our wealth W_n at time $n \geq 1$ is then

$$W_n := M_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}).$$

We will now prove that we cannot make money from this game.

Theorem 4.19. *Let $(X_0, X_1, \ldots)$ be a stochastic process. Assume that $(M_0, M_1, \ldots)$ is a (super)martingale with respect to $X_0, X_1, \ldots$. Let $c_1, c_2, \ldots$ be constants. Let $H_1, H_2, \ldots$ be a predictable process. Assume that $0 \leq H_n \leq c_n$ for all $n \geq 1$. Then*

$$W_n := M_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}).$$

is also a (super)martingale with respect to $(X_0, X_1, \ldots)$.

That is, you cannot make money by trying to bet on a (super)martingale.

Remark 4.20. The quantity $M_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1})$ is a finite version of a stochastic integral. And in fact, there is a corresponding statement to be made about stochastic integrals, namely that you cannot make money off of (continuous time) supermartingales.

Remark 4.21. Allowing $H_n < 0$ would correspond to betting negative amounts, so that the gambler could assume the position of the "house." So, we do not allow this to happen. Also, requiring the predictable process to be bounded is only assumed so that the expected values involved are finite; the boundedness assumption can in fact be weakened.

Proof of Theorem 4.19. First, observe that

$$W_{n+1} - W_n = H_{n+1}(M_{n+1} - M_n)$$

Also, from the triangle inequality, and since $M_0, M_1, \ldots$ is a (super)martingale, so that $\mathbf{E}\,|M_m| < \infty$ for all $m \geq 0$,

$$\mathbf{E}\,|W_n| \leq \mathbf{E}\,|M_0| + \sum_{m=1}^{n} c_m(\mathbf{E}\,|M_m| + \mathbf{E}\,|M_{m-1}|) < \infty.$$

So, the sequence $W_0, W_1, \ldots$ satisfies the first condition of being a (super)martingale. Now, let $m_0, x_0, \ldots, x_n \in \mathbb{R}$. Let $A := \{X_n = x_n, \ldots, X_0 = x_0, M_0 = m_0\}$. Since H_{n+1} is predictable, H_{n+1} is constant on A, so Lemma 4.4 implies

$$\mathbf{E}(W_{n+1} - W_n \,|\, A) = \mathbf{E}(H_{n+1}(M_{n+1} - M_n) \,|\, A) = H_{n+1}\mathbf{E}(M_{n+1} - M_n \,|\, A) \leq 0.$$

The last inequality follows since $M_0, M_1, \ldots$ is a (super)martingale and $H_{n+1} \geq 0$. □

Definition 4.22 (**Stopping Time**)**.** A **stopping time** for a martingale $M_0, M_1, \ldots$ is a random variable T taking values in $0, 1, 2, \ldots, \cup\{\infty\}$ such that, for any integer $n \geq 0$, the event $\{T = n\}$ is determined by $M_0, X_0, \ldots, X_n$. More formally, for any integer $n \geq 1$, there is a set $B_n \subseteq \mathbb{R}^{n+2}$ such that $\{T = n\} = \{(M_0, X_0, \ldots, X_n) \in B_n\}$. Put another way, the indicator function $1_{\{T=n\}}$ is a function of the random variables $M_0, X_0, \ldots, X_n$.

From Remark 4.10, a martingale satisfies $\mathbf{E}M_n = \mathbf{E}M_0$ for all $n \geq 0$. In some cases, we can replace n with a stopping time T in this equality. However, this cannot always hold.

Example 4.23. Let $(X_1, X_2, \ldots)$ be a sequence of independent random variables such that $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = 1/2$ for all $i \geq 0$. Let $M_0 = 0$ and let $M_n = X_0 + \cdots + X_n$ for all $n \geq 0$. Note that $\mathbf{E}X_0 = 0$. So, from Example 4.16, $M_0, M_1, \ldots$ is a martingale. Let $T := \min\{n \geq 1 \colon M_n = 1\}$ be the return time to 1. Then $M_T = 1$, so $\mathbf{E}M_T = 1 \neq 0 = \mathbf{E}M_0$.

Remark 4.24. Let $a, b \in \mathbb{R}$. We use the notation $a \wedge b := \min(a, b)$. Note that if T is a stopping time, then $a \wedge T$ is a stopping time, for any fixed $a \in \mathbb{R}$.

Theorem 4.25 (**Optional Stopping Theorem, Version 1**)**.** *Let $(M_0, M_1, \ldots)$ be a martingale with respect to $X_0, X_1, \ldots$, and let T be a stopping time. Then $(M_{0\wedge T}, M_{1\wedge T}, \ldots)$ is a martingale. In particular, $\mathbf{E}M_{n\wedge T} = \mathbf{E}M_0$ for all $n \geq 0$.*

Proof. Let $n \geq 1$. Let $H_n = 1_{\{T \geq n\}}$. Then

$$H_n = 1 - 1_{\{T \leq n-1\}} = 1 - \sum_{m=0}^{n-1} 1_{\{T=m\}}.$$

Since T is a stopping time, we know that H_n can be written as a function of $X_0, \ldots, X_{n-1}$. That is, $H_1, H_2, \ldots$ is a predictable process. For any $n \geq 0$, define

$$W_n := M_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}).$$

By Theorem 4.19, $W_0, W_1, \ldots$ is a martingale. By definition of H_m,

$$W_n = M_0 + \sum_{m=1}^{n}(1_{\{T\geq m\}})(M_m - M_{m-1}) = M_0 + \sum_{m=1}^{n}(M_{T\wedge m} - M_{T\wedge(m-1)}) = M_{T\wedge n}.$$

□

Theorem 4.26 (**Optional Stopping Theorem, Version 2**). *Let $(M_0, M_1, \ldots)$ be a martingale, and let T be a stopping time such that $\mathbf{P}(T < \infty) = 1$. Let $d \in \mathbb{R}$. Assume that $|M_{n\wedge T}| \leq d$ for all $n \geq 0$. Then $\mathbf{E}M_T = \mathbf{E}M_0$.*

Proof. From Theorem 4.25, for any $n \geq 1$,

$$\mathbf{E}M_0 = \mathbf{E}M_{n\wedge T} = \mathbf{E}M_{n\wedge T}(1_{\{T\leq n\}} + 1_{\{T>n\}}) = \mathbf{E}M_{n\wedge T}1_{\{T\leq n\}} + \mathbf{E}M_{n\wedge T}1_{\{T>n\}}.$$

We bound each term separately. We have

$$\left|\mathbf{E}M_{n\wedge T}1_{\{T>n\}}\right| \leq \mathbf{E}\left|M_{n\wedge T}\right|1_{\{T>n\}} \leq d\cdot\mathbf{E}1_{\{T>n\}} = d\cdot\mathbf{P}(T>n). \qquad (*)$$

Also, since $\mathbf{P}(T<\infty) = 1$, we have

$$\mathbf{P}(\lim_{n\to\infty} M_{n\wedge T} = M_T) = 1, \qquad \mathbf{P}(|M_T| \leq d) = 1.$$

Therefore, for any $n \geq 1$,

$$\begin{aligned}\left|\mathbf{E}M_{n\wedge T}1_{\{T\leq n\}} - \mathbf{E}M_T\right| &= \left|\mathbf{E}M_T 1_{\{T\leq n\}} - \mathbf{E}M_T(1_{\{T\leq n\}} + 1_{\{T>n\}})\right| \\ &= \left|\mathbf{E}M_T 1_{\{T>n\}}\right| \leq \mathbf{E}\left|M_T\right|1_{\{T>n\}} \leq d\cdot\mathbf{E}1_{\{T>n\}} = d\cdot\mathbf{P}(T>n). \qquad (**)\end{aligned}$$

So, subtracting $\mathbf{E}M_T$ from both sides of the above equality and using the triangle inequality,

$$\begin{aligned}\left|\mathbf{E}M_T - \mathbf{E}M_0\right| &= \left|\mathbf{E}M_T - \mathbf{E}M_{n\wedge T}1_{\{T\leq n\}} - \mathbf{E}M_{n\wedge T}1_{\{T>n\}}\right| \\ &\leq \left|\mathbf{E}M_T - \mathbf{E}M_{n\wedge T}1_{\{T\leq n\}}\right| + \left|\mathbf{E}M_{n\wedge T}1_{\{T>n\}}\right| \overset{(*),(**)}{\leq} 2d\cdot\mathbf{P}(T>n), \quad \forall\, n\geq 1.\end{aligned}$$

Letting $n\to\infty$ and using $\mathbf{P}(T<\infty) = 1$ concludes the proof. (By continuity of the probability law, $\lim_{n\to\infty}\mathbf{P}(T>n) = \mathbf{P}(T=\infty) = 0$.) □

For a real-world example, suppose $M_0, M_1, \ldots$ is a martingale which describes the price of a stock. Suppose the stock is currently priced at $M_0 = 100$ and you instruct your stock broker to sell the stock when its price reaches either \$110 or \$90. That is, define the stopping time $T = \min\{n \geq 1\colon M_n \geq 110 \text{ or } M_n \leq 90\}$. Then T is a stopping time. From the Optional Stopping Theorem Version 2, $\mathbf{E}M_T = \mathbf{E}M_0$. That is, you cannot make money off of this stock (if it is a martingale).

Remark 4.27. The assumptions of the Optional Stopping Theorem cannot be abandoned, as shown in Example 4.23. Let $(M_0, M_1, \ldots)$ be the symmetric simple random walk on $\mathbb{Z}$ with $M_0 = 0$. Let $T = \min\{n \geq 1\colon M_n = 1\}$. Then $\mathbf{E}M_0 = 0$ but $M_T = 1$, so $\mathbf{E}M_T = 1 \neq 0 = \mathbf{E}M_0$.

Example 4.28 (**Gambler's Ruin**). We return to Example 4.17. Let $0 < p < 1$ with $p \neq 1/2$, and let $q := 1-p$. Let $0 \leq a < x_0 < b$. Let $X_0 := x_0$. Let $(X_0, X_1, \ldots)$ be independent random variables such that $\mathbf{P}(X_i = 1) = p$ and $\mathbf{P}(X_i = -1) = 1-p$ for all $i \geq 1$. For any $n \geq 0$, let $Y_n = X_0 + \cdots + X_n$. Let $T = \min\{n \geq 1\colon Y_n \in \{a, b\}\}$. That is, T is the first time the simple random walk Y_n hits either a or b. We showed in Example 4.17

that $(q/p)^{Y_n}$ is a Martingale. Let $c := \mathbf{P}(Y_T = a)$ be the probability that the random walk hits a before it hits b. Lemma 3.27 implies that $\mathbf{P}(T < \infty) = 1$. From Theorem 4.26,

$$(q/p)^{x_0} = \mathbf{E}(q/p)^{Y_0} = \mathbf{E}(q/p)^{Y_T} = c(q/p)^a + (1-c)(q/p)^b.$$

Solving for c, we get

$$c = \frac{(q/p)^{x_0} - (q/p)^b}{(q/p)^a - (q/p)^b}.$$

In the case $p = 1/2$, Y_n itself is a martingale, so

$$x_0 = \mathbf{E}Y_0 = \mathbf{E}Y_T = ca + (1-c)b.$$

Solving for c, we get

$$c = \frac{x_0 - b}{a - b}.$$

Exercise 4.29. Let $X_0 = 0$. Let $(X_0, X_1, \ldots)$ such that $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = 1/2$ for all $i \geq 1$. For any $n \geq 0$, let $Y_n = X_0 + \cdots + X_n$. So, $(Y_0, Y_1, \ldots)$ is a symmetric simple random walk on $\mathbb{Z}$. Show that $Y_n^2 - n$ is a martingale (with respect to $(X_0, X_1, \ldots)$).

Exercise 4.30. Let $1/2 < p < 1$. Let $(X_0, X_1, \ldots)$ such that $\mathbf{P}(X_i = 1) = p$ and $\mathbf{P}(X_i = -1) = 1-p$ for all $i \geq 1$. For any $n \geq 0$, let $Y_n = X_0 + \cdots + X_n$. Let $T_0 = \min\{n \geq 1 \colon Y_n = 0\}$. Prove that $\mathbf{P}_1(T_0 = \infty) > 0$. Then, deduce that $\mathbf{P}_0(T_0 = \infty) > 0$. That is, there is a positive probability that the biased random walk never returns to 0, even though it started at 0.

Example 4.31. Continuing the Gambler's Ruin example with $p = 1/2$, let $a < 0 < b$ be integers, and let $x_0 = 0$ and let $T := \min\{n \geq 0 \colon Y_n \notin (a, b)\}$. We claim that $\mathbf{E}T = -ab$. To see this, we use Exercise 4.29 and the Optional Stopping Theorem to get $0 = \mathbf{E}(Y_T^2 - T)$, then using Example 4.28,

$$\begin{aligned}
\mathbf{E}T = \mathbf{E}Y_T^2 &= a^2\mathbf{P}(S_T = a) + b^2\mathbf{P}(S_T = b) \\
&= a^2\frac{b}{b-a} + b^2\frac{(-a)}{b-a} = ab\frac{a-b}{b-a} = -ab.
\end{aligned}$$

Strictly speaking, the Optional Stopping Theorem, Version 2, does not apply, since the martingale is not bounded. But Theorem 4.25 does apply, and we can then let $n \to \infty$ to get $\mathbf{E}T = -ab$. Filling in the details is beyond the scope of this course.

Exercise 4.32. Let $X_1, \ldots$ be independent identically distributed random variables with $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = 1/2$ for every $i \geq 1$. For any $n \geq 1$, let $M_n := X_1 + \cdots + X_n$. Let $M_0 = 0$. For any $n \geq 1$, define

$$W_n := M_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}).$$

Show that if you have an infinite amount of money, then you *can* make money by using the double-your-bet strategy in the game of coinflips (where if you bet \$$d$, then you win \$$d$ with probability 1/2, and you lose \$$d$ with probability 1/2). For example, show that if you start by betting \$1, and if you keep doubling your bet until you win (which should define some betting strategy $H_1, H_2, \ldots$ and a stopping time T), then $\mathbf{E}W_T = 1$, for a suitable stopping time T.

Exercise 4.33. Prove the following variant of the Optional Stopping Theorem. Assume that $(M_0, M_1, \ldots)$ is a submartingale, and let T be a stopping time such that $\mathbf{P}(T < \infty)$. Let $c \in \mathbb{R}$. Assume that $|M_{n\wedge T}| \leq c$ for all $n \geq 0$. Then $\mathbf{E}M_T \geq \mathbf{E}M_0$. That is, you can make money by stopping a submartingale.

Exercise 4.34 (**Ballot Theorem**)**.** Let a, b be positive integers. Suppose there are c votes cast by c people in an election. Candidate 1 gets a votes and candidate 2 gets b votes. (So $c = a + b$.) Assume $a > b$. The votes are counted one by one. The votes are counted in a uniformly random ordering, and we would like to keep a running tally of who is currently winning. (News agencies seem to enjoy reporting about this number.) Suppose the first candidate eventually wins the election. We ask: with what probability will candidate 1 always be ahead in the running tally of who is currently winning the election? As we will see, the answer is $\frac{a-b}{a+b}$.

To prove this, for any positive integer k, let S_k be the number of votes for candidate 1, minus the number of votes for candidate 2, after k votes have been counted. Then, define $X_k := S_{c-k}/(c-k)$. Show that $X_0, X_1, \ldots$ is a martingale with respect to $S_c, S_{c-1}, S_{c-2}, \ldots$. Then, let T such that $T = \min\{0 \leq k \leq c \colon X_k = 0\}$, or $T = c-1$ if no such k exists. Apply the Optional Stopping theorem to X_T to deduce the result.

Exercise 4.35. Let $(X_0, X_1, \ldots)$ be the simple random walk on $\mathbb{Z}$. For any $n \geq 0$, define $M_n = X_n^3 - 3nX_n$. Show that $(M_0, M_1, \ldots)$ is a martingale with respect to $(X_0, X_1, \ldots)$

Now, fix $m > 0$ and let T be the first time that the walk hits either 0 or m. Show that, for any $0 < k \leq m$,

$$\mathbf{E}_k(T \,|\, X_T = m) = \frac{m^2 - k^2}{3}.$$

Exercise 4.36. Let $X_1, X_2, \ldots$ be independent random variables with $\mathbf{E}X_i = 0$ for every $i \geq 1$. Suppose there exists $\sigma > 0$ such that $\mathrm{Var}(X_i) = \sigma^2$ for all $i \geq 1$. For any $n \geq 1$, let $S_n = X_1 + \cdots + X_n$. Show that $S_n^2 - n\sigma^2$ is a martingale with respect to $X_1, X_2, \ldots$. (We let $X_0 = 0$.)

Let $a > 0$. Let $T = \min\{n \geq 1 \colon |S_n| \geq a\}$. Using the Optional Stopping Theorem, show that $\mathbf{E}T \geq a^2/\sigma^2$. Observe that a simple random walk on $\mathbb{Z}$ has $\sigma^2 = 1$ and $\mathbf{E}T = a^2$ when $a \in \mathbb{Z}$.

5. Poisson Process

5.1. **Review of Conditional Expectation for Continuous Random Variables.**

Definition 5.1 (**Conditioning one Random Variable on Another**)**.** Let X and Y be continuous random variables with joint PDF $f_{X,Y}$. That is, for any $A \subseteq \mathbb{R}^2$,

$$\mathbf{P}((X,Y) \in A) = \iint_A f_{X,Y}(x,y)dxdy.$$

Fix some $y \in \mathbb{R}$ with $f_Y(y) > 0$. For any $x \in \mathbb{R}$, define the **conditional PDF** of X, given that $Y = y$ by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad \forall\, x \in \mathbb{R}.$$

We also define the **conditional expectation**

$$\mathbf{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx.$$

And for any $-\infty \leq a < b \leq \infty$, define the **conditional probability**

$$\mathbf{P}(a \leq X \leq b \mid Y = y) = \int_{a}^{b} f_{X|Y}(x|y) dx.$$

More generally, if $X_1, \ldots, X_n$ have joint PDF $f_{X_1,\ldots,X_n}$, we define

$$f_{X_1|X_2,\ldots,X_n}(x_1|x_2, \ldots, x_n) = \frac{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}{f_{X_2,\ldots,X_n}(x_2, \ldots, x_n)}, \qquad \forall x_1, \ldots, x_n \in \mathbb{R}.$$

Here the marginal $f_{X_2,\ldots,X_n}$ is defined by

$$f_{X_2,\ldots,X_n}(x_2, \ldots, x_n) = \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1, \qquad \forall x_2, \ldots, x_n \in \mathbb{R}.$$

We can similarly define conditional probability and conditional expectations.

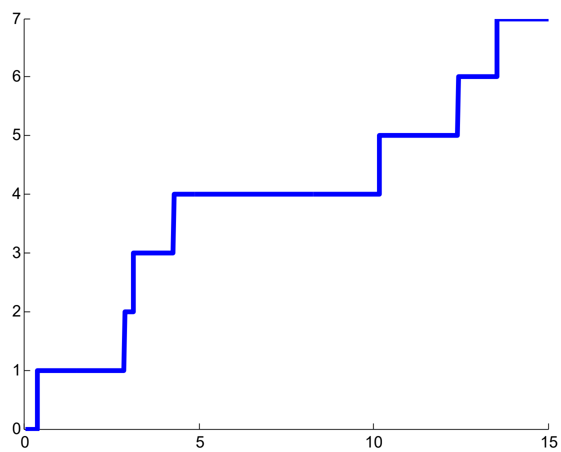


FIGURE 1. One Sample Path of a Poisson Process. The horizontal axis is the s-axis.

5.2. **Construction of the Poisson Process.** Up until this point, we have focused on discrete time stochastic processes. That is, we have discussed sequences $(X_0, X_1, X_2, \ldots)$ of random variables, indexed by the nonnegative integers. In theory and in applications, it is often beneficial to consider *continuous time* stochastic processes. That is, it is often helpful to consider sets of random variables $\{X_s\}_{s\geq 0}$. Here, s ranges over all nonnegative real numbers.

The Poisson Process is our first example of a continuous time stochastic process. This process will be integer-valued.

Let $\lambda > 0$. Recall that a random variable T is **exponential with parameter** λ if T has the density function given by $f_T(x) = \lambda e^{-\lambda x}$ for all $x \geq 0$, and $f_T(x) = 0$ otherwise.

Moreover,

$$\mathbf{P}(T \leq t) = \int_{-\infty}^{t} f_T(x)dx = \int_{0}^{t} f_T(x)dx = \begin{cases} 1 - e^{-\lambda t}, & \text{if } t \geq 0 \\ 0 & \text{if } t < 0. \end{cases}$$

Lemma 5.2. *Let τ be an exponential random variable with parameter $\lambda > 0$. Let $t, s > 0$. Then*

$$\mathbf{P}(\tau > t + s \,|\, \tau > t) = \mathbf{P}(\tau > s).$$

*That is, T has the **memoryless** property, or **lack of memory** property. Moreover,*

$$\mathbf{P}(\tau \leq t + s \,|\, \tau > t) = \mathbf{P}(\tau \leq s).$$

Proof.

$$\begin{aligned}\mathbf{P}(\tau > t + s \,|\, \tau > t) &= \frac{\mathbf{P}(\tau > t + s, \tau > t)}{\mathbf{P}(\tau > t)} = \frac{\mathbf{P}(\tau > t + s)}{\mathbf{P}(\tau > t)} \\ &= \frac{\lambda \int_{t+s}^{\infty} e^{-\lambda x} dx}{\lambda \int_{t}^{\infty} e^{-\lambda x} dx} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \lambda \int_{s}^{\infty} e^{-\lambda x} dx.\end{aligned}$$

Then, note that $\mathbf{P}(\tau \leq t+s \,|\, \tau > t) = 1 - \mathbf{P}(\tau > t+s \,|\, \tau > t) = 1 - \mathbf{P}(\tau > s) = \mathbf{P}(\tau \leq s)$. □

Lemma 5.3. *Let $\lambda > 0$. Let $\tau_1, \ldots, \tau_n$ be independent exponential random variables with parameter λ. Define $T_n := \tau_1 + \cdots + \tau_n$. Then T_n is a **gamma distributed random variable** with parameters n and λ. That is, T_n has density*

$$f_{T_n}(t) := \begin{cases} \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, & \text{if } t \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Proof. We induct on n. The case $n = 1$ follows (using $0! = 1$) since $T_1 = \tau_1$ is an exponential random variable. We now do the inductive step. Suppose the assertion holds for n and consider the case $n + 1$. Then $T_{n+1} = T_n + \tau_{n+1}$. So, for any $s > 0$, using that T_n and τ_{n+1} are independent,

$$\mathbf{P}(T_{n+1} < s) = \mathbf{P}(T_n + \tau_{n+1} < s) = \int_{-\infty}^{\infty} \int_{-\infty}^{s-t} f_{\tau_{n+1}}(y) f_{T_n}(t) dy dt.$$

Taking the derivative with respect to $s > 0$,

$$f_{T_{n+1}}(s) = \frac{d}{ds}\mathbf{P}(T_{n+1} < s) = \int_{-\infty}^{\infty} \frac{d}{ds} \int_{-\infty}^{s-t} f_{\tau_{n+1}}(y) f_{T_n}(t) dy dt. = \int_{-\infty}^{\infty} f_{\tau_{n+1}}(s - t) f_{T_n}(t) dt.$$

Applying the inductive hypothesis,

$$f_{T_{n+1}}(s) = \int_{0}^{s} \lambda e^{-\lambda(s-t)} \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} dt = \lambda e^{-\lambda s} \frac{\lambda^n}{(n-1)!} \int_{0}^{s} t^{n-1} dt = \lambda e^{-\lambda s} \frac{\lambda^n s^n}{n!}.$$

□

Definition 5.4 (**Poisson Process**). Let $\lambda > 0$. Let $\tau_1, \tau_2, \ldots$ be independent exponential random variables with parameter λ. Let $T_0 = 0$, and for any $n \geq 1$, let $T_n := \tau_1 + \cdots + \tau_n$. A **Poisson Process** with parameter $\lambda > 0$ is a set of integer-valued random variables $\{N(s)\}_{s \geq 0}$ defined by $N(s) := \max\{n \geq 0 \colon T_n \leq s\}$.

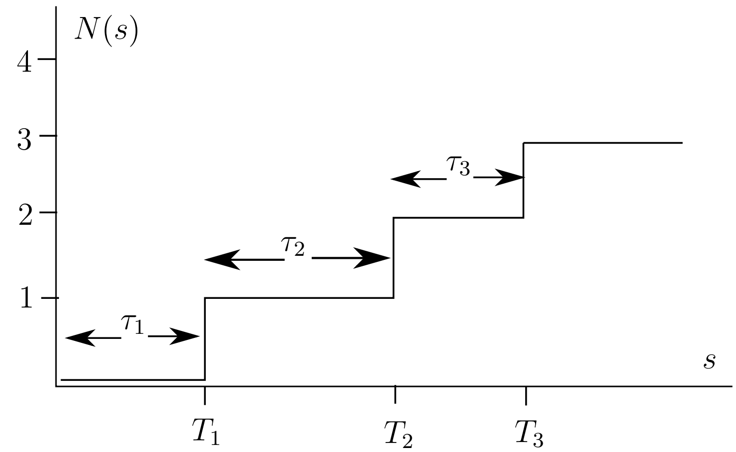


FIGURE 2. One Sample Path of a Poisson Process.

We can think of the Poisson Process intuitively, so that τ_k is the time between the arrival of the $(k-1)^{st}$ person and the k^{th} person at a bank, and $N(s)$ is the number of people who have arrived by time $s \geq 0$.

Recall that a discrete random variable X is a **Poisson random variable with mean** $\lambda > 0$ if $\mathbf{P}(X = n) = e^{-\lambda} \cdot \frac{\lambda^n}{n!}$ for all nonnegative integers n.

Lemma 5.5. *Let $\{N(s)\}_{s\geq 0}$ be a Poisson Process with parameter $\lambda > 0$. Then, for any $s \geq 0$, $N(s)$ is a Poisson random variable with parameter λs.*

Proof. Let n be a nonnegative integer. Then

$$\begin{aligned}
\mathbf{P}(N(s) = n) &= \mathbf{P}(\max\{m \geq 0 \colon T_m \leq s\} = n) = \mathbf{P}(T_n \leq s,\, T_{n+1} > s) \\
&= \mathbf{P}(T_n \leq s,\, T_n + \tau_{n+1} > s) \\
&= \int_{-\infty}^{s} \int_{s-t}^{\infty} f_{\tau_{n+1}}(y) f_{T_n}(t)\, dy dt, \qquad \text{since } T_n \text{ and } \tau_{n+1} \text{ are independent} \\
&= \int_{-\infty}^{s} \mathbf{P}(\tau_{n+1} > s - t) f_{T_n}(t) dt = \int_0^s e^{-\lambda(s-t)} \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} dt, \qquad \text{by Lemma 5.3} \\
&= e^{-\lambda s} \frac{\lambda^n}{(n-1)!} \int_0^s t^{n-1} dt = e^{-\lambda s} \frac{\lambda^n s^n}{n!}.
\end{aligned}$$

□

Exercise 5.6. Let $\lambda > 0$. Let $\tau_1, \tau_2, \ldots$ be independent exponential random variables with parameter λ. For any $n \geq 1$, let $T_n = \tau_1 + \cdots + \tau_n$. Fix positive integers $n_k > \cdots > n_1$ and positive real numbers $t_k > \cdots > t_1$. Then

$$f_{T_{n_k}, \ldots, T_{n_1}}(t_k, \ldots, t_1) = f_{T_{(n_k - n_{k-1})}}(t_k - t_{k-1}) \cdots f_{T_{(n_2 - n_1)}}(t_2 - t_1) f_{T_{n_1}}(t_1).$$

(Hint: just try to case $k = 2$ first, and use a conditional density function.)

Exercise 5.7. Let $s, t > 0$ and let m, n be nonnegative integers. Let $0 < t_m < t_{m+1} < t_{m+n} < t_{m+n+1}$, and define (using the notation of Exercise 5.6),

$$g(t_m, t_{m+1}, t_{m+n}, t_{m+n+1}) := f_{T_1}(t_{m+n+1} - t_{m+n}) f_{T_{n-1}}(t_{m+n} - t_{m+1}) f_{T_1}(t_{m+1} - t_m) f_{T_m}(t_m).$$

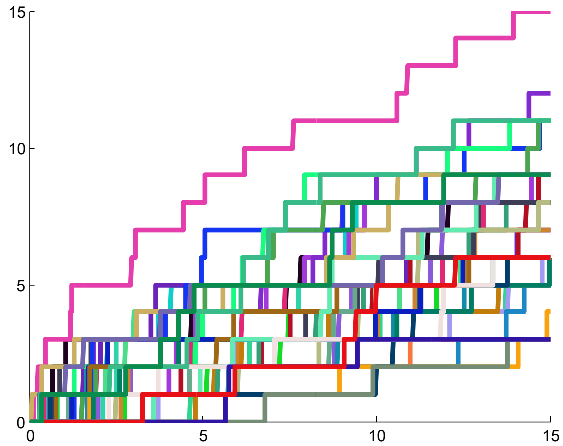


FIGURE 3. Several Sample Paths of a Poisson Process. The horizontal axis is the s-axis.

Let $\{N(s)\}_{s\geq 0}$ be a Poisson Process with parameter $\lambda > 0$. Show that

$$\begin{aligned}&\mathbf{P}(N(s+t) = m+n,\, N(s) = m)\\ &= \int_0^s \left(\int_s^{s+t} \left(\int_{t_{m+1}}^{s+t} \left(\int_{s+t}^{\infty} g(t_m, t_{m+1}, t_{m+n}, t_{m+n+1}) dt_{m+n+1} \right) dt_{m+n} \right) dt_{m+1} \right) dt_m.\end{aligned}$$

(Hint: use the joint density, and then use Exercise 5.6.)

Lemma 5.8. *Let $\{N(s)\}_{s\geq 0}$ be a Poisson Process with parameter $\lambda > 0$. Let $s, t > 0$ and let m, n be nonnegative integers. Then*

$$\mathbf{P}(N(s+t) = m+n,\, N(s) = m) = \mathbf{P}(N(s) = m)\mathbf{P}(N(t) = n).$$

Proof. Suppose $n > 1$. From Lemma 5.3 and Exercise 5.7 we have

$$g(t_m, t_{m+1}, t_{m+n}, t_{m+n+1}) = \lambda^{m+n+1} e^{-\lambda t_{m+n+1}} \frac{(t_{m+n} - t_{m+1})^{n-2}}{(n-2)!} \frac{t_m^{m-1}}{(m-1)!}.$$

$$\begin{aligned}\mathbf{P}(N(s+t) = m+n,\, N(s) = m) &= \frac{\lambda^{m+n+1}}{(n-2)!(m-1)!} \int_{s+t}^{\infty} e^{-\lambda t_{m+n+1}} dt_{m+n+1}\\ &\cdot \left(\int_s^{s+t} \int_{t_{m+1}}^{s+t} (t_{m+n} - t_{m+1})^{n-2} dt_{m+n} dt_{m+1} \right) \int_0^s t_m^{m-1} dt_m\\ &= \lambda^{m+n} e^{-\lambda(s+t)} \frac{s^m}{m!(n-1)!} \left(\int_s^{s+t} (s+t-t_{m+1})^{n-1} dt_{m+1} \right)\\ &= \lambda^{m+n} e^{-\lambda(s+t)} \frac{s^m t^n}{m!n!} = \mathbf{P}(N(s) = m)\mathbf{P}(N(t) = n).\end{aligned}$$

In the last line, we used Lemma 5.5. The cases $n = 0$ and $n = 1$ are treated similarly. $\square$

Lemma 5.9. *Let $\{N(s)\}_{s\geq 0}$ be a Poisson Process with parameter $\lambda > 0$. Fix $s > 0$. Then $N(t+s) - N(s)$ is a Poisson random variable which is independent of $N(s)$. (In fact, $\{N(t+s) - N(s)\}_{t\geq 0}$ is a Poisson process with parameter λ which is independent of the random variable $N(s)$, but we cannot prove this yet.)*

Proof. Let $s, t > 0$ and let m, n be nonnegative integers. From Lemma 5.8,

$$\begin{aligned}\mathbf{P}(N(s+t) - N(s) = n,\, N(s) = m) &= \mathbf{P}(N(s+t) = m+n,\, N(s) = m)\\ &= \mathbf{P}(N(t) = n)\mathbf{P}(N(s) = m). \qquad (*)\end{aligned}$$

Summing over all $m \geq 0$ gives $\mathbf{P}(N(s+t) - N(s) = n) = \mathbf{P}(N(t) = n)$, for all $s, t > 0$, for all $n \geq 0$. That is, $N(t+s) - N(s)$ is a Poisson random variable with parameter λ. For the independence property, we can just rewrite $(*)$ as

$$\mathbf{P}(N(s+t)-N(s) = n,\, N(s) = m) = \mathbf{P}(N(s+t)-N(s) = n)\mathbf{P}(N(s) = m), \qquad \forall\, m, n \geq 0.$$

□

Lemma 5.10. *The Poisson Process has* ***independent increments****. That is, for any* $0 < u_0 < \cdots < u_k$*, the following random variables are independent:*

$$N(u_1) - N(u_0), \ldots, N(u_k) - N(u_{k-1}).$$

Proof. In Lemma 5.9, we showed that $N(s+t)-N(s)$ is independent of $N(s)$. By generalizing the arguments of Exercise 5.7 and Lemma 5.8, we have: if $1 < n_1, \ldots, n_k$, and if $m_i := n_1 + \cdots + n_i$, for all $1 \leq i \leq k$,

$$g(t_{m_1}, t_{m_1+1}, \ldots, t_{m_k}, t_{m_k+1}) := \lambda^{m_k+1} e^{-\lambda t_{m_k+1}} \frac{t_{m_1}^{n_1-1}}{(n_1-1)!} \prod_{i=2}^{k} \frac{(t_{m_i} - t_{m_{(i-1)}+1})^{n_i-2}}{(n_i-2)!}.$$

If $0 < s_1, \ldots, s_k$, and if $u_i := s_1 + \cdots + s_i$ for all $1 \leq i \leq k$,

$$\begin{aligned}&\mathbf{P}(N(u_k) = m_k, \ldots, N(u_1) = m_1)\\ &= \int_0^{u_1} \int_{u_1}^{u_2} \int_{t_{m_1+1}}^{u_2} \int_{u_2}^{u_3} \int_{t_{m_2+1}}^{u_3} \cdots \int_{t_{m_k}+1}^{u_k} \int_{u_k}^{\infty}\\ &\qquad\qquad g(t_{m_1}, t_{m_1+1}, \ldots, t_{m_k}, t_{m_k+1}) dt_{m_k+1} dt_{m_k} \cdots dt_{m_1+1} dt_{m_1}\\ &= \lambda^{m_k} e^{-\lambda u_k} \prod_{i=1}^{k} \frac{s_i^{n_i}}{n_i!} = \prod_{i=1}^{k} \frac{\lambda^{n_i} e^{-\lambda s_i} s_i^{n_i}}{n_i!} = \prod_{i=1}^{k} \mathbf{P}(N(s_k) = n_k).\end{aligned}$$

So, using this equality and Lemma 5.9,

$$\begin{aligned}&\mathbf{P}(N(u_k) - N(u_{k-1}) = n_k, \ldots, N(u_2) - N(u_1) = n_2, N(u_1) = n_1)\\ &\quad= \mathbf{P}(N(u_k) = m_k, \ldots, N(u_2) = m_2, N(u_1) = n_1)\\ &\quad= \prod_{i=1}^{k} \mathbf{P}(N(s_k) = n_k) = \prod_{i=1}^{k} \mathbf{P}(N(u_k) - N(u_{k-1}) = n_k).\end{aligned}$$

□

We summarize the above discussion.

Definition 5.11 (**Right-Continuous Function**)**.** Let $f\colon [0,\infty) \to \mathbb{R}$. We say that f is **right-continuous** if: for any $s \geq 0$, $\lim_{t \to s^+} f(t) = f(s)$.

Exercise 5.12. Give an example of a right-continuous function. Then give an example of a function that is not right-continuous.

Theorem 5.13. *Let* $\{N(s)\}_{s \geq 0}$ *be a Poisson process with parameter* $\lambda > 0$*. Then* $N(0) = 0$*,*

(i) *With probability* 1, $s \mapsto N(s)$ *is right-continuous.*
(ii) $N(t+s) - N(s)$ *is a Poisson random variable with parameter* λt *for all* $s, t > 0$.
(iii) $\{N(s)\}_{s\geq 0}$ *has independent increments.*

Conversely, if $N(0) = 0$ *and if (i), (ii) and (iii) hold, then* $\{N(s)\}_{s\geq 0}$ *is a Poisson process with parameter* $\lambda > 0$.

Remark 5.14. In particular, we could use (i), (ii) and (iii) as an alternate definition of a Poisson process.

Proof. Property (ii) follows from Lemma 5.9, and Property (iii) follows from Lemma 5.10. Property (i) follows from Definition 5.4. For the converse direction, suppose $\{N(s)\}_{s\geq 0}$ is a stochastic process satisfying (i), (ii) and (iii) and $N(0) = 0$. For any $n \geq 1$, define $T_n = \min\{s \geq 0\colon N(s) \geq n\}$. Note that $N(s)$ is valued in the nonnegative integers and increasing by Property (ii). Also, by Property (i), $\min\{s \geq 0\colon N(s) \geq n\}$ exists and $N(T_n) = n$ for any $n \geq 1$. To see that $N(T_n) = n$, note that, $N(T_n) \geq n$ by definition of T_n, and if $N(T_n) > n$, then $N(T_n) - N(T_n - \varepsilon) > 1$ for all $0 < \varepsilon < T_n$. Then, for any $s \geq 0, j \geq 1$, we have by the union bound

$$\begin{aligned}
&\mathbf{P}(N(T_n) > n, T_n < s)\\
&\leq \mathbf{P}\left(\exists\, 1 \leq i \leq j,\, N\left(\left(s\Big(1 - \frac{i}{j}\Big)\right) - N\left(\left(s\Big(1 - \frac{i-1}{j}\Big)\right)\right) > 1\right)\right)\\
&\leq \sum_{i=1}^{j} \mathbf{P}\left(N\Big(s\Big(1 - \frac{i}{j}\Big)\Big) - N\Big(s\Big(1 - \frac{i-1}{j}\Big)\Big) > 1\right) \overset{(ii)}{=} \sum_{i=1}^{j}(1 - e^{-\lambda/j}[1 + \lambda/j])
\end{aligned}$$

By Taylor expansion, $e^{-\lambda/j}(1 + \lambda/j) = 1 - \lambda^2/j^2 + c(j)$, where $|c(j)| \leq 10\lambda^3/j^3$. So,

$$\mathbf{P}(N(T_n) > n, T_n < s) \leq \sum_{i=1}^{j} \frac{\lambda^2}{j^2} + \frac{10\lambda^3}{j^3} = \frac{\lambda^2}{j} + \frac{10\lambda^3}{j^2}.$$

Letting $j \to \infty$, we get $\mathbf{P}(N(T_n) > n, T_n < s) = 0$. Letting $s, n \to \infty$, we see that $N(T_n) = n$ with probability 1, as desired.

Now, for any $t > 0$, property (ii) says

$$\mathbf{P}(T_1 > t) = \mathbf{P}(N(t) = 0) = e^{-\lambda t}.$$

That is, T_1 is an exponential random variable with parameter λ.

Also, if $\tau_1 := T_1$ and $\tau_2 := T_2 - T_1$, then property (iii) implies

$$\begin{aligned}
\mathbf{P}(\tau_2 > t \,|\, \tau_1 = s) &= \mathbf{P}(T_2 > t + s \,|\, N(s) - N(r) = 1 \,\text{for all}\, 0 < r < s)\\
&= \mathbf{P}(N(t+s) - N(s) = 0 \,|\, N(s) - N(r) = 1 \,\text{for all}\, 0 < r < s)\\
&= \mathbf{P}(N(t+s) - N(s) = 0) = e^{-\lambda t}, \qquad \text{by Property (ii)}.
\end{aligned}$$

Since this equality holds for any $s > 0$, we conclude that τ_2 is an exponential random variable with parameter λ, and τ_1, τ_2 are independent.

More generally, if $k > 1$ and $\tau_k := T_k - T_{k-1}$, then for any $0 < s_1 < \cdots < s_{k-1}$,

$$\begin{aligned}
&\mathbf{P}(\tau_k > t \,|\, \tau_{k-1} = s_{k-1}, \ldots, \tau_1 = s_1) \\
&= \mathbf{P}(N(t + s_{k-1}) - N(s_{k-1}) = 0 \,| \\
&\qquad N(s_{k-1}) - N(r_{k-1}) = 1, \forall\, s_{k-2} < r_{k-1} < s_{k-1}, \ldots, N(s_1) - N(r_1) = 1, \forall\, 0 < r_1 < s_1) \\
&= \mathbf{P}(N(t + s_{k-1}) - N(s_{k-1}) = 0) = e^{-\lambda t}, \qquad \text{by Property (ii).}
\end{aligned}$$

Since this equality holds for any $0 < s_1 < \cdots < s_{k-1}$, we conclude that τ_k is an exponential random variable with parameter λ, and by induction on k, $\tau_1, \ldots, \tau_k$ are independent. □

Remark 5.15. Let $\{N(s)\}_{s\geq 0}$ be a Poisson Process with parameter $\lambda > 0$. Fix $s > 0$. Then $\{N(t + s) - N(s)\}_{t\geq 0}$ is a Poisson process with parameter λ which is independent of the random variable $N(s)$. This follows from Theorem 5.13.

Proposition 5.16 (Poisson Approximation to the Binomial). *Let $\lambda > 0$. For each positive integer n, let $0 < p_n < 1$, and let X_n be a binomial distributed random variable with parameters n and p_n. Assume that $\lim_{n\to\infty} p_n = 0$ and $\lim_{n\to\infty} np_n = \lambda$. Then, for any nonnegative integer k, we have*

$$\lim_{n\to\infty} \mathbf{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

From the Poisson Approximation to the Binomial, we can use a Poisson random variable to model any low probability event with many chances of happening. For example, the Poisson random variable can model the number of people who win the lottery, the number of magnetic defects in a hard drive, the number of typos per page in a book, etc.

The Poisson Process can be treated in the same way, with an added time variable. That is, we can use the Poisson Process to model any kind of low probability event with many chances of happening over time. For example, this process can model the number of people arriving at a restaurant during a week, the number of car accidents over the course of a day, the number of visitors to a website over the course of a year, etc.

Definition 5.17 (Inhomogeneous Poisson Process). Let $\lambda\colon [0, \infty) \to [0, \infty)$ be a function. We say a stochastic process $\{N(s)\}_{s\geq 0}$ is a **inhomogeneous Poisson Process with rate** λ if $N(0) = 0$ and if

(i) With probability 1, $s \mapsto N(s)$ is right-continuous.
(ii) $N(t) - N(s)$ is a Poisson random variable with parameter $\int_s^t \lambda(r)dr$ for all $t > s > 0$.
(iii) $\{N(s)\}_{s\geq 0}$ has independent increments.

We recover the usual Poisson process by choosing $\lambda(r) := \lambda$ for all $r \geq 0$.

5.3. **Compound Poisson Process.**

Exercise 5.18. Let $Y_1, Y_2, \ldots$ be independent identically distributed random variables. Let N be an independent, nonnegative integer-valued random variable. Let $S = Y_1 + \cdots + Y_N$, where $S := 0$ if $N = 0$.

- If $\mathbf{E}\,|Y_1| < \infty$ and $\mathbf{E}N < \infty$, then $\mathbf{E}S = (\mathbf{E}N)(\mathbf{E}Y_1)$.
- If $\mathbf{E}Y_1^2 < \infty$ and $\mathbf{E}N^2 < \infty$, then $\mathrm{var}(S) = (\mathbf{E}N)(\mathrm{var}(Y_1)) + (\mathbf{E}Y_1)^2(\mathrm{var}(N))$.
- If N is a Poisson random variable with parameter $\lambda > 0$, then $\mathrm{var}(S) = \lambda\mathbf{E}Y_1^2$.

(Hint: for the second part, use $\mathbf{E}(S^2|N=n) = n\cdot\mathrm{var}(Y_1) + (n\mathbf{E}Y_1)^2$. Use this to compute $\mathbf{E}S^2$. Then compute $\mathrm{var}(S)$.)

Exercise 5.19. Suppose the number of students going to a restaurant in Ackerman in a single day has a Poisson distribution with mean 500. Suppose each student spends an average of \$10 with a standard deviation of \$5. What is the average revenue of the restaurant in one day? What is the standard deviation of the revenue in one day? (The amounts spent by the students are independent identically distributed random variables.)

5.4. **Transformations.**

Theorem 5.20 (Splitting). *Let $Y_1, Y_2, \ldots$ be independent identically distributed positive integer-valued random variables. Let $\{N(s)\}_{s\geq 0}$ be a Poisson process with parameter $\lambda > 0$ that is independent of $Y_1, Y_2, \ldots$. For any $s > 0$, $j \geq 1$, let $N_j(s)$ be the number of integers $i \leq N(s)$ such that $Y_i = j$. Then $\{N_1(s)\}_{s\geq 0}, \{N_2(s)\}_{s\geq 0}, \ldots$ are independent Poisson processes with rates $\lambda\mathbf{P}(Y_1 = 1), \lambda\mathbf{P}(Y_1 = 2), \ldots$.*

Proof. Fix an integer $k > 0$ and assume that $Y_1 \leq k$. Note that $N_j(s) = \sum_{i=1}^{N(s)} 1_{\{Y_i = j\}}$, and $N_1(s) + \cdots + N_k(s) = N(s)$. Let $n := n_1 + \cdots + n_k$. We first consider the case $k = 2$. Then

$$\begin{aligned}
\mathbf{P}(N_1(s) = n_1, N_2(s) = n_2 \,|\, N(s) = n) &= \mathbf{P}(\sum_{i=1}^{n} 1_{\{Y_i=1\}} = n_1, \sum_{i=1}^{n} 1_{\{Y_i=2\}} = n_2 \,|\, N(s) = n) \\
&= \mathbf{P}(\sum_{i=1}^{n} 1_{\{Y_i=1\}} = n_1, \sum_{i=1}^{n} 1_{\{Y_i=2\}} = n_2) \qquad \text{, by independence} \\
&= \frac{n!}{n_1! n_2!}\mathbf{P}(Y_1 = 1)^{n_1}\mathbf{P}(Y_1 = 2)^{n_2}.
\end{aligned}$$

So, since $\{N(s) = n\} \supseteq \{N_1(s) = n_1, N_2(s) = n_2\}$, we get from Lemma 5.5,

$$\begin{aligned}
\mathbf{P}(N_1(s) = n_1, N_2(s) = n_2) &= \mathbf{P}(N_1(s) = n_1, N_2(s) = n_2, N(s) = n) \\
&= \mathbf{P}(N_1(s) = n_1, N_2(s) = n_2 \,|\, N(s) = n)\mathbf{P}(N(s) = n) \\
&= \frac{n!}{n_1! n_2!}\mathbf{P}(Y_1 = 1)^{n_1}\mathbf{P}(Y_1 = 2)^{n_2} e^{-\lambda s}\frac{\lambda^n s^n}{n!} \\
&= e^{-\lambda s(\mathbf{P}(Y_1=1))}\frac{[\lambda s\mathbf{P}(Y_1 = 1)]^{n_1}}{n_1!} e^{-\lambda s(\mathbf{P}(Y_1=2))}\frac{[\lambda s\mathbf{P}(Y_1 = 2)]^{n_2}}{n_2!}.
\end{aligned}$$

So, $N_1(s)$ and $N_2(s)$ are independent Poisson random variables with parameters $\lambda s\mathbf{P}(Y_1 = 1)$ and $\lambda s\mathbf{P}(Y_1 = 2)$, respectively. So, one part of condition (ii) of Theorem 5.13 holds. Conditions (iii) follows since $\{N(s)\}_{s\geq 0}$ itself has independent increments. (If we condition on the values of $Y_1, Y_2, \ldots$, then N_1 has (conditionally) independent increments. Then the Total Probability Theorem implies that N_1 has independent increments.)

We now handle the more general case, where we verify the full condition (ii). Let $s, t > 0$, and for any $1 \leq i \leq k$, let $X_i := N_i(s+t) - N_i(s)$, and let $X := N(s+t) - N(s)$. Then

$$\begin{aligned}
&\mathbf{P}(X_1 = n_1, \ldots, X_k = n_k \,|\, X = n) \\
&\qquad = \mathbf{P}(\sum_{i=1}^{n} 1_{\{Y_i=1\}} = n_1, \ldots, \sum_{i=1}^{n} 1_{\{Y_i=k\}} = n_k \,|\, X = n) \qquad \text{, since } X = \sum_{j=1}^{k} X_j \\
&\qquad = \mathbf{P}(\sum_{i=1}^{n} 1_{\{Y_i=1\}} = n_1, \ldots, \sum_{i=1}^{n} 1_{\{Y_i=k\}} = n_k) \qquad \text{, by independence} \\
&\qquad = \frac{n!}{n_1! \cdots n_k!} \mathbf{P}(Y_1 = 1)^{n_1} \cdots \mathbf{P}(Y_1 = k)^{n_k}.
\end{aligned}$$

So, since $\{X = n\} \supseteq \{X_1 = n_1, \ldots, X_k = n_k\}$, we get from Lemma 5.5,

$$\begin{aligned}
\mathbf{P}(X_1 = n_1, \ldots, X_k = n_k) &= \mathbf{P}(X_1 = n_1, \ldots, X_k = n_k, X = n) \\
&= \mathbf{P}(X_1 = n_1, \ldots, X_k = n_k \,|\, X = n)\mathbf{P}(X = n) \\
&= \frac{n!}{n_1! \cdots n_k!} \mathbf{P}(Y_1 = 1)^{n_1} \cdots \mathbf{P}(Y_1 = k)^{n_k} e^{-\lambda s} \frac{\lambda^n s^n}{n!} \\
&= \prod_{i=1}^{n} e^{-\lambda s(\mathbf{P}(Y_1 = i))} \frac{[\lambda s \mathbf{P}(Y_1 = i)]^{n_i}}{n_i!}.
\end{aligned}$$

The Theorem now follows since conditions (i), (ii) and (iii) of Theorem 5.13 hold $\forall\, j \geq 1$. $\square$

Exercise 5.21. Suppose you run a (busy) car wash, and the number of cars that come to the car wash between time 0 and time $s > 0$ is a Poisson poisson with rate $\lambda = 1$. Suppose each car is equally likely to have one, two, three, or four people in it. What is the average number of cars with four people that have arrived by time $s = 100$?

Proposition 5.22 (**Superposition**). *Let $\{N_1(s)\}_{s \geq 0}, \ldots, \{N_k(s)\}_{s \geq 0}$ be independent Poisson processes with rates $\lambda_1, \ldots, \lambda_k > 0$, respectively. Then $\{N_1(s) + \cdots + N_k(s)\}_{s \geq 0}$ is a Poisson process with rate $\lambda_1 + \cdots + \lambda_k$*

Proof. It suffices to check the three conditions of Theorem 5.13. The first condition is clear. The second condition follows by repeated application of Exercise 5.23. The third condition follows by assumption. $\square$

Exercise 5.23. Let X be a Poisson random variable with parameter $\lambda > 0$. Let Y be a Poisson random variable with parameter $\delta > 0$. Assume that X, Y are independent. Then $X + Y$ is a Poisson random variable with parameter $\lambda + \delta$.

Exercise 5.24. Suppose you are still running a (busy) car wash. The number of red cars that come to the car wash between time 0 and time $s > 0$ is a Poisson poisson with rate 2. The number of blue cars that come to car wash between time 0 and time $s > 0$ is a Poisson poisson with rate 3. Both Poisson processes are independent of each other. All cars are either red or blue. With what probability will five blue cars arrive, before three red cars have arrived?

The Poisson Process can be generalized by replacing the exponential random variables by more general random variables. This generalized process is called a renewal process. We can still think of a renewal process in the same way that we think of the Poisson process, e.g. by modeling the number of people visiting a restaurant over time, or the number of lightbulbs that need to be installed in a single socket, up to a certain time, etc. However, a general renewal process will no longer have the independent increment property, as we had in the case of the Poisson process. Indeed, the independent increment property was a crucial ingredient in Theorem 5.13, where we uniquely characterized the Poisson process.

Definition 6.1 (**Renewal Process**)**.** Let $\tau_1, \tau_2, \ldots$ be nonnegative independent identically distributed variables. Let $T_0 = 0$, and for any $n \geq 1$, let $T_n := \tau_1 + \cdots + \tau_n$. A **Renewal process** is a set of integer-valued random variables $\{N(s)\}_{s\geq 0}$ defined by $N(s) := \max\{n \geq 0 \colon T_n \leq s\}$.

Example 6.2. Let $X_0, X_1, \ldots$ be a Markov chain with $X_0 := x \in \Omega$. Let $T_1 := \min\{k \geq 1 \colon X_k = x\}$, and for any $n \geq 2$, inductively define $T_n := \min\{k > T_{n-1} \colon X_k = x\}$. Let $\tau_n := T_{n+1} - T_n$ for any $n \geq 1$. The Strong Markov property implies that $\tau_1, \tau_2, \ldots$ are independent and identically distributed. Therefore, $\{N(s)\}_{s\geq 0}$, as defined above is a renewal process. Note that $N(s)$ is the number of times the Markov chain returns to x up to time s.

6.1. **Law of Large Numbers.**

Theorem 6.3 (**Law of Large Numbers for Renewal Process**)**.** *Suppose we have a renewal process $\{N(s)\}_{s\geq 0}$ with arrival increments $\tau_1, \tau_2, \ldots$. Let $\mu := \mathbf{E}\tau_1$. Assume that $0 < \mu < \infty$. Then*

$$\mathbf{P}\left(\lim_{s\to\infty} \frac{N(s)}{s} = \frac{1}{\mu}\right) = 1.$$

That is, if one light bulb lasts μ years on average, then after s years, we will have replaced about s/μ light bulbs (when s is large).

Proof. Let $T_n := \tau_1 + \cdots + \tau_n$. Recall that $\tau_1, \tau_2, \ldots$ are independent and identically distributed, by the definition of a renewal process. So, the Strong Law of Large Numbers, Theorem 2.45, implies that

$$\mathbf{P}\left(\lim_{n\to\infty} \frac{T_n}{n} = \mu\right) = 1. \qquad (*)$$

By the definition of $N(s) := \max\{n \geq 0 \colon T_n \leq s\}$, we have

$$T_{N(s)} \leq s < T_{N(s)+1}.$$

Dividing by $N(s) > 0$, we get

$$\frac{T_{N(s)}}{N(s)} \leq \frac{s}{N(s)} < \frac{T_{N(s)+1}}{N(s)+1}\frac{N(s)+1}{N(s)}. \qquad (**)$$

Also by definition of $N(s)$, for any fixed integer $m > 0$, we have $\mathbf{P}(N(s) < m) = \mathbf{P}(T_m > s) \leq \mathbf{E}T_m/s = m\mu/s \to 0$ as $s \to \infty$. So, using this fact and $(*)$, the left and right sides of $(**)$ converge to μ with probability 1. The Theorem follows. $\square$

Exercise 6.4. Prove the following two facts, which we used in the proof of the Law of Large Numbers for Renewal Processes.

Let $X_1, X_2, \ldots, Y_1, Y_2, \ldots, Z_1, Z_2, \ldots$ be random variables. Let $a, b \in \mathbb{R}$.

- Assume that $X_n \leq Y_n \leq Z_n$ for any $n \geq 1$. Assume that $\mathbf{P}(\lim_{n\to\infty} X_n = a) = 1$ and $\mathbf{P}(\lim_{n\to\infty} Z_n = a) = 1$. Prove that $\mathbf{P}(\lim_{n\to\infty} Y_n = a) = 1$.
- Assume that $\mathbf{P}(\lim_{n\to\infty} X_n = a) = 1$ and $\mathbf{P}(\lim_{n\to\infty} Y_n = b) = 1$. Prove that $\mathbf{P}(\lim_{n\to\infty} X_n Y_n = ab) = 1$.

7. Brownian Motion

7.1. **Construction of Brownian Motion.** Let $X_1, X_2, \ldots$ be independent random variables such that $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = 1/2$ for all $i \geq 1$. Define

$$B_1(t) := \sum_{i=1}^{\lfloor t \rfloor} X_i, \qquad \forall\, t \geq 0.$$

Note that if j is an integer such that $j \leq t < j+1$, then $\lfloor t \rfloor := j$ and $B_1(t)$ is constant when $t \in [j, j+1)$, and then the value of $B_1(t)$ changes at $t = j$, according to the value of X_j. That is, the value of $B_1(t)$ changes at each positive integer value according to one of the random variables $X_1, X_2, \ldots$. Put another way, $B_1(t)$ plots the path of a simple random walk on the integers, if we imagine that the random walker stops for one second before each of their random movements. Note also that, for any integers $t > s > 0$, $B_1(t) - B_1(s)$ has mean zero and variance $t - s$.

Now, let k be a positive integer. We now consider changing the time between the random walker's movements to $1/k$. To keep the same variance property as before, we also multiply the sum by $1/\sqrt{k}$:

$$B_k(t) := \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor tk \rfloor} X_i, \qquad \forall\, t \geq 0.$$

Note that $B_k(t)$ is only constant on intervals of length $1/k$ now. Also, as promised, if $t > s > 0$ are integers divided by k, then $B_k(t) - B_k(s)$ has mean zero and variance $(tk - sk)/k = t - s$. Finally, observe that the process $\{B_k(t)\}_{t\geq 0}$ has the **independent increments** property. So, for example, if $0 < t_1 < t_2 < t_3 < t_4$ are integers divided by k, then $B_k(t_4) - B_k(t_3)$ and $B_k(t_2) - B_k(t_1)$ are independent.

If k is large, i.e. something like $k = 1000$, already $B_k(t)$ can model various random phenomena that depend on time, e.g. a stock price, the position of a randomly moving particle, etc. However, just as we let Riemann sums converge to integrals to create a useful theory of integration, it is also helpful for us to take a certain limit of the continuous-time process $\{B_k(t)\}_{t\geq 0}$ as $k \to \infty$. The resulting limiting stochastic process $\{B(t)\}_{t\geq 0}$ is called **Brownian motion**. The precise meaning of this limit as $k \to \infty$ is beyond this course material. However, we can still make some observations about Brownian motion.

Fix $t > 0$. From the Central Limit Theorem (Theorem 2.50), observe that

$$\lim_{k\to\infty} \mathbf{P}\left(\frac{1}{\sqrt{t}} B_k(t) \leq a\right) = \lim_{k\to\infty} \mathbf{P}\left(\frac{1}{\sqrt{tk}} \sum_{i=1}^{\lfloor tk \rfloor} X_i \leq a\right) = \int_{-\infty}^{a} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}, \qquad \forall\, a \in \mathbb{R}.$$

Replacing a by $a/\sqrt{t}$ and changing variables, we get

$$\lim_{k\to\infty}\mathbf{P}\left(B_k(t)\leq a\right)=\int_{-\infty}^{a/\sqrt{t}}e^{-x^2/2}\frac{dx}{\sqrt{2\pi}}=\int_{-\infty}^{a}e^{-\frac{x^2}{2t}}\frac{dx}{\sqrt{2\pi t}},\qquad\forall\, a\in\mathbb{R}.$$

That is, from Definition 2.20, as $k\to\infty$, $B_k(t)$ has the same CDF as a Gaussian random variable with mean zero and variance t.

Arguing similarly, if $t>s>0$, then as $k\to\infty$, $B_k(t)-B_k(s)$ has the same CDF as a Gaussian random variable with mean zero and variance $t-s$. Moreover, we could believe that the stationary increments property is also preserved as $k\to\infty$. We are therefore led to the following definition.

Definition 7.1 (**Brownian Motion**). Standard Brownian motion is a stochastic process $\{B(t)\}_{t\geq 0}$ which is the limit (in a sense we will not make precise) of the processes $\{B_k(t)\}_{t\geq 0}$ as $k\to\infty$. Standard Brownian motion with $B(0)=0$ is uniquely characterized by the following properties:

(i) (Continuous Sample Paths) With probability 1, the function $t\mapsto B(t)$ is continuous.

(ii) (Stationary Gaussian increments) for any $0<s<t$, $B(t)-B(s)$ is a Gaussian random variable with mean zero and variance $t-s$.

(iii) (Independent increments) For any $0<t_1<\cdots<t_n$, the random variables $B(t_2)-B(t_1),\ldots,B(t_n)-B(t_{n-1})$ are all independent.

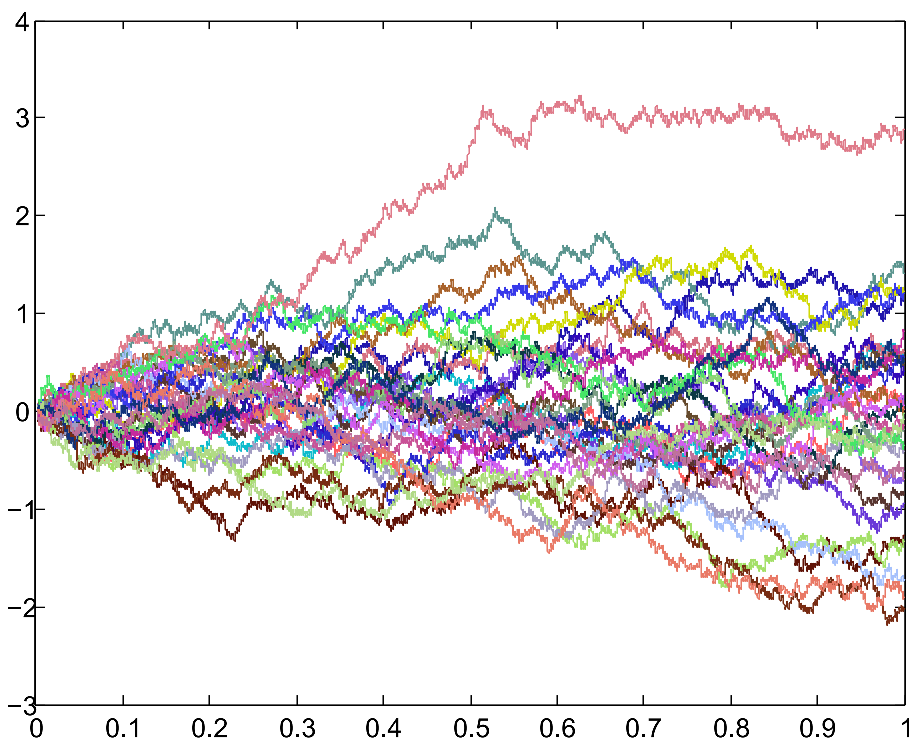


FIGURE 4. Sample Paths of Standard Brownian Motion. The horizontal axis is the t-axis.

Exercise 7.2 (**Scaling Invariance**)**.** Let $a > 0$. Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. For any $t > 0$, define $X(t) := \frac{1}{\sqrt{a}}B(at)$. Then $\{X(t)\}_{t\geq 0}$ is also a standard Brownian motion.

Dealing rigorously with Brownian motion is beyond our course material. So, we will occasionally ignore some details when dealing with Brownian motion, and when doing your homework, it is okay to do the same. However, we will always try to provide as many details as possible, and you should try your best to do the same.

Below, we will not formally define a stopping time, and we will not formally state an Optional Stopping Theorem. However, since we know that $\{B_k(t)\}_{t\in\{0,1/k,2/k,3/k,\ldots\}}$ is a martingale for every $k \geq 1$, then it seems that $\{B(t)\}_{t\geq 0}$ should be a martingale in some sense. In fact, by the independent increments property of Brownian Motion, if $t > s > 0$, if $x_1, \ldots, x_n \in \mathbb{R}$, and if $s > s_n > \cdots > s_1 > 0$, then

$$\mathbf{E}(B(t) - B(s) \,|\, B(s_n) = x_n, \ldots, B(s_1) = x_1) = \mathbf{E}(B(t) - B(s)) = 0.$$

The last equality follows since $B(t) - B(s)$ is a mean zero Gaussian random variable.

Remark 7.3. Just as we have seen for random walks, we cannot apply an Optional Stopping Theorem to every stopping time. For example, let $\{B(t)\}_{t\geq 0}$ be standard Brownian motion, and let $T = \min\{t > 0\colon B(t) = 1\}$. Then $\mathbf{E}B(0) = \mathbf{E}(0) = 0$ but $B(T) = 1$, so $\mathbf{E}B(T) = 1 \neq 0 = \mathbf{E}B(0)$.

Below, whenever we apply an Optional Stopping Theorem to a stochastic process $\{X(t)\}_{t\geq 0}$ and stopping time T, we will always verify that there exists a constant $c > 0$ such that $|X(t \wedge T)| \leq c$ for all $t \geq 0$, as in the statement of Theorem 4.26.

We will not formally define a stopping time T in these notes for continuous time stochastic processes.

Brownian Motion satisfies a Markov property, in the following sense

Proposition 7.4 (**Markov Property**)**.** *Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Let $s > 0$. Then the stochastic process $\{B(t+s)-B(s)\}_{t\geq 0}$ is itself a standard Brownian motion, which is independent of the set of random variables $\{B(u)\}_{0\leq u\leq s}$.*

Proof. Properties (i), (ii) and (iii) for $\{B(t + s) - B(s)\}_{t\geq 0}$ in the definition of Brownian Motion all follow from properties (i), (ii) and (iii) for $\{B(t)\}_{t\geq 0}$. To see the independence property, note that the independent increments property for $\{B(t)\}_{t\geq 0}$ implies that $B(t) - B(s)$ is independent of $B(u) - B(0) = B(u)$, for all $0 \leq u \leq s$. □

Remark 7.5. Standard Brownian motion is also a martingale in the following sense: if $t > s > 0$, and if $s > s_n > \cdots > s_1 > 0$, and $x_1, \ldots, x_n \in \mathbb{R}$, then

$$\mathbf{E}(B(t) - B(s) \,|\, B(s_n) = x_n, \ldots, B(s_1) = x_1) = \mathbf{E}(B(t) - B(s)) = 0.$$

The first equality follows from property (iii) and the second equality follows from (ii).

Exercise 7.6. Let $x_1, \ldots, x_n \in \mathbb{R}$, and if $t_n > \cdots > t_1 > 0$. Using the independent increment property, show that the event

$$\{B(t_1) = x_1, \ldots, B(t_n) = x_n\}$$

has a multivariate normal distribution. That is, the joint density of $(B(t_1), \ldots, B(t_n))$ is

$$f(x_1, \ldots, x_n) = f_{t_1}(x_1) f_{t_2-t_1}(x_2 - x_1) \cdots f_{t_n-t_{n-1}}(x_n - x_{n-1})$$

where

$$f_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}, \qquad \forall x \in \mathbb{R},\, t > 0.$$

Exercise 7.7. Let X be a Gaussian random variable with mean 0 and variance $\sigma_X^2 > 0$. Let Y be a Gaussian random variable with mean 0 and variance $\sigma_Y^2 > 0$. Assume that X and Y are independent. Show that $X + Y$ is also a Gaussian random variable with mean 0 and variance $\sigma_X^2 + \sigma_Y^2$.

(Hint: write an expression for $\mathbf{P}(X + Y \leq t)$, $t \in \mathbb{R}$, then take a derivative in t.)

The covariances of Brownian motion can be computed from the definition of Brownian motion.

Proposition 7.8. *Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Let $0 < s < t$. Then*

$$\mathbf{E}B(s)B(t) = s.$$

Proof. Using that $B(s)$ has variance s, and using the independent increment property,

$$\begin{aligned}\mathbf{E}B(s)B(t) &= \mathbf{E}B(s)(B(t) - B(s) + B(s)) = \mathbf{E}(B(s))^2 + \mathbf{E}[B(s)(B(t) - B(s))] \\ &= s + [\mathbf{E}B(s)][\mathbf{E}(B(t) - B(s))] = s.\end{aligned}$$

□

7.2. **Properties of Brownian Motion.** Standard Brownian motion has some counterintuitive properties: for example with probability 1, there is no smallest $t > 0$ such that $B(t) = 0$. That is, with probability 1, the function $t \mapsto B(t)$ crosses the t-axis an infinite number of times. For an intuitive explanation of this fact, recall that the simple random walk on the integers has all states recurrent. That is, a random walk on the integers returns to $t = 0$ an infinite number of times. And Brownian motion is constructed by re-scaling the random walk on the integers.

Since there is no smallest $t > 0$ when $B(t) = 0$, in order to ask when standard Brownian motion takes the value zero, we need to replace the minimum over $t > 0$ by the infimum over $t > 0$.

Definition 7.9 (**Infimum**)**.** Let A be a nonempty set of nonnegative real numbers. The **infimum of** A, denote $\inf(A)$ is the largest number $y \in \mathbb{R}$ such that $y \leq a$ for all $a \in A$. That is, $\inf(A)$ is the greatest lower bound of A.

The existence of $\inf(A)$ is proven in Math 131A, so we will not justify its existence here.

Exercise 7.10. Let $A := \{1, 1/2, 1/3, 1/4, 1/5, \ldots\}$. Find $\inf(A)$. Note that $\inf(A)$ exists, but A has no minimum element. The infimum is better to work with for this reason.

Proposition 7.11. *Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Let $a, b > 0$. Let $T_a := \inf\{t \geq 0 \colon B(t) = a\}$. Then*

$$\mathbf{P}(T_a < T_{-b}) = \frac{b}{a + b}$$

Proof. Let $c := \mathbf{P}(T_a < T_{-b})$. Let $T := \inf\{t \geq 0 \colon B(t) \in \{a, -b\}\}$. From the Optional Stopping Theorem (for continuous-time martingales) (noting that $|B(t \wedge T)| \leq \max(a, b)$ for all $t \geq 0$)

$$0 = \mathbf{E}B(0) = \mathbf{E}B(T) = ac - b(1 - c).$$

Solving for c proves the result. □

Exercise 7.12. Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Then $\{(B(t))^2 - t\}_{t\geq 0}$ is a (continuous-time) martingale in the following sense: it $t > s > 0$, and if $s > s_n > \cdots > s_1 > 0$, and $x_1, \ldots, x_n \in \mathbb{R}$, then

$$\mathbf{E}((B(t))^2 - t - ((B(s))^2 - s) \mid B(s_n) = x_n, \ldots, B(s_1) = x_1) = 0.$$

Proposition 7.13. *Let $a, b > 0$. Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Let $T = \inf\{t \geq 0 \colon B(t) \notin (-b, a)\}$. Then*

$$\mathbf{E}T = ab.$$

Proof. Using Exercise 7.12 and the Optional Stopping Theorem, we get $0 = \mathbf{E}((B(T))^2 - T)$, then using Proposition 7.11,

$$\begin{aligned}\mathbf{E}T = \mathbf{E}(B(T))^2 &= a^2\mathbf{P}(B(T) = a) + b^2\mathbf{P}(B(T) = -b)\\ &= a^2\frac{b}{a+b} + b^2\left(1 - \frac{b}{a+b}\right) = a^2\frac{b}{a+b} + b^2\frac{a}{a+b} = ab\frac{a+b}{a+b} = ab.\end{aligned}$$

Strictly speaking, the Optional Stopping Theorem, Version 2, does not apply, since the martingale is not bounded. But Optional Stopping Version 1 does apply to $(B(T\wedge t))^2 - T\wedge t$, and we can then let $t \to \infty$ to get $\mathbf{E}T = -ab$. Filling in the details is beyond the scope of this course, as in Example 4.31. □

Exercise 7.14. Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion.

- Given that $B(1) = 10$, what is the expected length of time after $t = 1$ until $B(t)$ hits either 8 or 12?
- Now, let $\sigma = 2$, and $\mu = -5$. Suppose a commodity has price $X(t) = \sigma B(t) + \mu t$ for any time $t \geq 0$. Given that the price of the commodity is 4 at time $t = 8$, what is the probability that the price is below 1 at time $t = 9$?
- Suppose a stock has a price $S(t) = 4e^{B(t)}$ for any $t \geq 0$. That is, the stock moves according to Geometric Brownian Motion. What is the probability that the stock reaches a price of 7 before it reaches a price of 2?

A reflection principle holds for standard Brownian motion, similar to Lemma 3.69 for the simple random walk on $\mathbb{Z}$.

Proposition 7.15 (**Reflection Principle**)**.** *Let $x > 0$. Then*

$$\mathbf{P}(T_x > t) = \mathbf{P}(-x < B(t) < x) = \int_{-x}^{x} e^{-\frac{y^2}{2t}}\frac{dy}{\sqrt{2\pi t}}, \qquad \forall\, t > 0.$$

The final equality above follows since $B(t)$ is a Gaussian random variable with mean 0 and variance t.

Exercise 7.16. Fix $x > 0$

- Show the bound $\mathbf{P}(-x < B(t) < x) \geq \frac{x}{20\sqrt{t}}$ holds for all $t > x^2$.
- Show that $\mathbf{E}T_x = \infty$. (Recall we observed something similar for the simple random walk on $\mathbb{Z}$ in Exercise 3.71.)

Corollary 7.17.

$$\mathbf{P}(\max_{0\leq s\leq t} B(s) \geq x) = \mathbf{P}(T_x \leq t) = 1 - \mathbf{P}(-x < B(t) < x) = 1 - \int_{-x}^{x} e^{-\frac{y^2}{2t}}\frac{dy}{\sqrt{2\pi t}}.$$

Proof. The first equality follows since $\max_{0\leq s\leq t} B(s) \geq x$ occurs if and only if $T_x \leq t$ (by property (i) of Brownian motion). Finally, apply Proposition 7.15. ☐

Remark 7.18. Property (i) of Brownian motion and the Extreme Value Theorem ensure that $\max_{0\leq s\leq t} B(s)$ exists with probability 1.

Definition 7.19 (Brownian Motion with Drift). Let $\sigma > 0$ and let $\mu \in \mathbb{R}$. A **standard Brownian motion with drift** μ and variance σ^2 is a stochastic process of the form

$$\{\sigma B(t) + \mu t\}_{t\geq 0}$$

where $\{B(t)\}_{t\geq 0}$ is a standard Brownian motion.

Exercise 7.20. Let $\{X(s)\}_{s\geq 0}$ be a standard Brownian motion with drift μ and variance σ^2. For any $t > s > 0$, show that $X(t) - X(s)$ is a Gaussian random variable with mean $\mu(t-s)$ and variance $\sigma^2(t-s)$.

In the Gambler's ruin problem (i.e. for a biased random walk on $\mathbb{Z}$), in Example 4.28, we computed the probabilities that the random walk hits a certain value before another. We can do a similar computation for the standard Brownian motion with drift.

Exercise 7.21. Let $\{X(t)\}_{t\geq 0} = \{\sigma B(t) + \mu t\}_{t\geq 0}$ be a standard Brownian motion with variance $\sigma^2 > 0$ and drift $\mu \in \mathbb{R}$. Fix $\lambda \in \mathbb{R}$. Then $\{Y(t)\}_{t\geq 0} = \{e^{\lambda X(t) - (\lambda\mu + \lambda^2\sigma^2/2)t}\}_{t\geq 0}$ is a (continuous-time) martingale in the following sense: it $t > s > 0$, and if $s > s_n > \cdots > s_1 > 0$, and $x_1, \ldots, x_n \in \mathbb{R}$, then

$$\mathbf{E}(Y(t) - Y(s) \,|\, B(s_n) = x_n, \ldots, B(s_1) = x_1) = 0.$$

Proposition 7.22. *Let $\{X(t)\}_{t\geq 0} = \{\sigma B(t) + \mu t\}_{t\geq 0}$ be a standard Brownian motion with variance $\sigma^2 > 0$ and negative drift $\mu < 0$. Let $a < 0 < b$. Let $\alpha := 2\,|\mu|\,/\sigma^2$. Let $T_a := \inf\{t > 0 \colon X(t) = a\}$. Then*

$$\mathbf{P}(T_b < T_a) = \frac{1 - e^{\alpha a}}{e^{\alpha b} - e^{\alpha a}}.$$

Letting $a \to -\infty$, we then get

$$\mathbf{P}(\max_{t\geq 0} X(t) \geq b) = e^{-\alpha b}, \qquad \forall\, b \geq 0.$$

That is, $\max_{t\geq 0} X(t)$ is an exponential random variable with mean $\sigma^2/(2\,|\mu|)$.

Proof. Let $c := \mathbf{P}(T_b < T_a)$. Choose $\lambda := \alpha = -2\mu/\sigma^2$. Then, by Exercise 7.21, $e^{\alpha X(t)}$ is a martingale. Let $T := \inf\{t \geq 0 \colon X(t) \in \{a, b\}\}$. From the Optional Stopping Theorem

$$1 = \mathbf{E}e^{\alpha X(0)} = \mathbf{E}e^{\alpha X(T)} = ce^{\alpha b} + (1-c)e^{\alpha a}.$$

Solving for c proves the first statement. (We verify the assumptions of the Optional Stopping Theorem, Version 2. Note that $|e^{\alpha X(t\wedge T)}| \leq \max\{e^{\alpha a}, e^{\alpha b}\}$ for all $t \geq 0$. Also, $\mathbf{P}(T < \infty) \geq \mathbf{P}(T_a < \infty)$, and if $T_a < \infty$, then $a = X(T_a) = \sigma B(T_a) + \mu T_a \leq \sigma B(T_a)$. So, if we define $T_a' := \inf\{t \geq 0 \colon B(t) = a/\sigma\}$, then $T_a < \infty$ implies $T_a' < \infty$, by property (i) of Brownian motion. So, $\mathbf{P}(T < \infty) \geq \mathbf{P}(T_a < \infty) \geq \mathbf{P}(T_a' < \infty)$, and $\mathbf{P}(T_a' < \infty) = 1$ by Proposition 7.15, since $\mathbf{P}(T_a' < \infty) = 1 - \lim_{s\to\infty} \int_{-a/\sigma}^{a/\sigma} e^{-\frac{y^2}{2s}} \frac{ds}{\sqrt{2\pi s}} = 1$.)

For the second statement, letting $a \to -\infty$ gives $\mathbf{P}(T_b < \infty) = e^{-\alpha b}$ (assuming that $T_a \to \infty$ as $a \to -\infty$). Then, note that $\{T_b < \infty\} = \{\max_{t\geq 0} X(t) \geq b\}$. ☐

For example, there is some chance that the standard Brownian motion with negative drift will never take the value $b = 1$.

Exercise 7.23. Let $\{X(t)\}_{t\geq 0} = \{\sigma B(t) + \mu t\}_{t\geq 0}$ be a standard Brownian motion with variance $\sigma^2 > 0$ and negative drift $\mu < 0$. Let $a < 0 < b$. Let $T := \inf\{t \geq 0 \colon X(t) \in \{a, b\}\}$. Let $\alpha := 2\,|\mu|\,/\sigma^2$. Show that

$$\mathbf{E}T = \frac{1}{\mu} \cdot \frac{b(1 - e^{\alpha a}) + a(e^{\alpha b} - 1)}{e^{\alpha b} - e^{\alpha a}}$$

Exercise 7.24. Let $\{X(t)\}_{t\geq 0} = \{\sigma B(t) + \mu t\}_{t\geq 0}$ be a standard Brownian motion with variance $\sigma^2 > 0$ and negative drift $\mu < 0$. Let $a < 0$. Let $T_a := \inf\{t \geq 0 \colon X(t) = a\}$. Let $\alpha := 2\,|\mu|\,/\sigma^2$. Show that

$$\mathbf{E}T_a = \frac{a}{\mu}.$$

Exercise 7.25 (Optional)**.** Write a computer program to simulate standard Brownian motion. More specifically, the program should simulate a random walk on $\mathbb{Z}$ with some small step size such as .002. (That is, simulate $B_k(t)$ when $k = 500^2$ and, say, $0 \leq t \leq 1$.)

Exercise 7.26 (Optional)**.** The following exercise assumes familiarity with Matlab and is derived from Cleve Moler's book, Numerical Computing with Matlab.

The file `brownian.m` plots the evolution of a cloud of particles that starts at the origin and diffuses in a two-dimensional random walk, modeling the Brownian motion of gas molecules.

(a) Modify `brownian.m` to keep track of both the average and the maximum particle distance from the origin. Using loglog axes, plot both sets of distances as functions of n, the number of steps. You should observe that, on the log-log scale, both plots are nearly linear. Fit both sets of distances with functions of the form $cn^{1/2}$. Plot the observed distances and the fits, using linear axes.

(b) Modify `brownian.m` to model a random walk in three dimensions. Do the distances behave like $n^{1/2}$?

The program `brownian.m` appears below.

```
% BROWNIAN   Two-dimensional random walk.
%   What is the expansion rate of the cloud of particles?

shg
clf
set(gcf,'doublebuffer','on')
delta = .002;
x = zeros(100,2);
h = plot(x(:,1),x(:,2),'.');
axis([-1 1 -1 1])
axis square
stop = uicontrol('style','toggle','string','stop');
while get(stop,'value') == 0
   x = x + delta*randn(size(x));
   set(h,'xdata',x(:,1),'ydata',x(:,2))
   drawnow
```

```
end
set(stop,'string','close','value',0,'callback','close(gcf)')
```

8. Appendix: Notation

Let n, m be a positive integers. Let $A, B, B_1, \ldots, B_n$ be sets contained in a universal set $\mathcal{C}$.

$\mathbb{R}$ denotes the set of real numbers

$\mathbb{Z}$ denotes the set of integers

$\in$ means "is an element of." For example, $2 \in \mathbb{R}$ is read as "2 is an element of $\mathbb{R}$."

$\forall$ means "for all"

$\exists$ means "there exists"

$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \colon x_i \in \mathbb{R} \,\forall\, 1 \leq i \leq n\}$

$f \colon A \to B$ means f is a function with domain A and range B. For example,

$f \colon \mathbb{R}^2 \to \mathbb{R}$ means that f is a function with domain $\mathbb{R}^2$ and range $\mathbb{R}$

$\emptyset$ denotes the empty set

$A \subseteq B$ means $\forall\, a \in A$, we have $a \in B$, so A is contained in B

$A \smallsetminus B := \{a \in A \colon a \notin B\}$

$A^c := \mathcal{C} \smallsetminus A$, the complement of A in $\mathcal{C}$

$A \cap B$ denotes the intersection of A and B

$A \cup B$ denotes the union of A and B

$\mathbf{P}$ denotes a probability law on $\mathcal{C}$

$\mathbf{P}(A|B)$ denotes the conditional probability of A, given B.

$\mathbf{P}(A|B_1, \ldots, B_n) := \mathbf{P}(A| \cap_{i=1}^n B_i)$ denotes the conditional probability of A, given $\cap_{i=1}^n B_i$.

$|A|$ denotes the number of elements in the (finite) set A.

$1_A \colon \mathcal{C} \to \{0, 1\}$, denotes the indicator function of A, so that

$$1_A(c) = \begin{cases} 1 & \text{, if } c \in A \\ 0 & \text{, otherwise.} \end{cases}$$

Let $a_1, \ldots, a_n$ be real numbers. Let n be a positive integer.

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_{n-1} + a_n.$$

$$\prod_{i=1}^{n} a_i = a_1 \cdot a_2 \cdots a_{n-1} \cdot a_n.$$

$$\min(a_1, a_2) = a \wedge b \text{ denotes the minimum of } a_1 \text{ and } a_2.$$

$$\max(a_1, a_2) = a \vee b \text{ denotes the maximum of } a_1 \text{ and } a_2.$$

Let A be a set and let $f \colon A \to \mathbb{R}$ be a function. Then $\max_{x \in A} f(x)$ denotes the maximum value of f on A (if it exists). Similarly, $\min_{x \in A} f(x)$ denotes the minimum value of f on A (if it exists).

Let X be a discrete random variable on a sample space $\mathcal{C}$, so that $X\colon \mathcal{C} \to \mathbb{R}$. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Let $x \in \mathbb{R}$. Let $A \subseteq \mathcal{C}$. Let Y be another discrete random variable

$$p_X(x) = \mathbf{P}(X = x) = \mathbf{P}(\{c \in \mathcal{C}\colon X(c) = x\}),\ \forall\, x \in \mathbb{R}$$
the Probability Mass Function (PMF) of X

$\mathbf{E}(X)$ denotes the expected value of X

$\mathrm{var}(X) = \mathbf{E}(X - \mathbf{E}(X))^2$, the variance of X

$\sigma_X = \sqrt{\mathrm{var}(X)}$, the standard deviation of X

$X|A$ denotes the random variable X conditioned on the event A.

$\mathbf{E}(X|A)$ denotes the expected value of X conditioned on the event A.

$\mathbf{E}(X|B_1, \ldots, B_n) := \mathbf{E}(X| \cap_{i=1}^n B_i)$ denotes the conditional expectation of X, given $\cap_{i=1}^n B_i$.

Let X, Y be a continuous random variables on a sample space $\mathcal{C}$, so that $X, Y\colon \mathcal{C} \to \mathbb{R}$. Let $-\infty \le a \le b \le \infty$, $-\infty \le c \le d \le \infty$. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Let $A \subseteq \mathcal{C}$.

$f_X\colon \mathbb{R} \to [0, \infty)$ denotes the Probability Density Function (PDF) of X, so

$$\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)dx$$

$f_{X,Y}\colon \mathbb{R} \to [0, \infty)$ denotes the joint PDF of X and Y, so

$$\mathbf{P}(a \le X \le b, c \le Y \le d) = \int_c^d \int_a^b f_{X,Y}(x, y)dxdy$$

$f_{X|A}$ denotes the Conditional PDF of X given A

$\mathbf{E}(X|A)$ denotes the expected value of X conditioned on the event A.

$\mathbf{E}(X|\mathcal{A})$ denotes the expected value of X given a partition $\mathcal{A} = \{A_1, \ldots, A_k\}$ of $\mathcal{C}$.

Let X be a random variable on a sample space $\mathcal{C}$, so that $X\colon \mathcal{C} \to \mathbb{R}$. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Let $x \in \mathbb{R}$.

$$F_X(x) = \mathbf{P}(X \le x) = \mathbf{P}(\{c \in \mathcal{C}\colon X(c) \le x\})$$
the Cumulative Distibution Function (CDF) of X.

Let $(X_0, X_1, \ldots)$ be a finite Markov chain with state space Ω. Let $x \in \Omega$. Let C be a subset of positive integers. Let $y \in \mathbb{R}$.

$\mathbf{P}_x$ denotes the conditional probability such that

$$\mathbf{P}_x(A) = \mathbf{P}(A \,|\, X_0 = x)\, \forall\, A \text{ in the sample space}$$

$\mathbf{E}_x$ denotes expectation with respect to $\mathbf{P}_x$

$\gcd C$ is the greatest common divisor of C

$$\Theta(y) := \int_{-\infty}^{y} e^{-t^2/2} \frac{dt}{\sqrt{2\pi}}.$$

Let A be a nonempty set of nonnegative real numbers. Then $\inf(A)$ denotes the infimum of A, the greatest lower bound of A.

UCLA Department of Mathematics, Los Angeles, CA 90095-1555
E-mail address: heilman@math.ucla.edu