MATH 505B, GRADUATE APPLIED PROBABILITY II, SPRING 2021

STEVEN HEILMAN

Contents

Date: April 21, 2021

1. Introduction

This course is an introduction to graduate stochastic processes, assuming no previous background in measure theory.

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.

A basic question we will often 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**)

For technical reasons, it might be impossible to define a probability law $\mathbf{P}$ on all subsets of a sample space $\mathcal{C}$. For example, if $\mathcal{C} = [0, 1]$ and if $\mathbf{P}$ is the probability law such that $\mathbf{P}([a, b]) := b - a$ for all $0 \leq a \leq b \leq 1$, then $\mathbf{P}$ cannot be defined on all subsets of $[0, 1]$, due the existence of non-measurable sets. The most natural set of subsets of $\mathcal{C}$ on which a probability law $\mathbf{P}$ can be defined is a **σ-algebra** (or **σ-field**). Since this course assumes no background in measure theory, we will not discuss this issue further.

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 and let Ω be a set. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. A **random variable** X is a function $X \colon \mathcal{C} \to \Omega$. (Some textbooks exclusively refer to function $X \colon \mathcal{C} \to \mathbb{R}$ as random variables, but we do not use this convention.) 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 real-valued 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 real-valued 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 $X \colon \mathcal{C} \to \Omega$. The **probability mass function** (or PMF) of X, denote $p_X \colon \mathbb{R} \to [0, 1]$ is defined by

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

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 real-valued 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 real-valued 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 real-valued 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 real-valued 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 real-valued 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 real-valued 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** $f_{X_1}\colon\mathbb{R}\to[0,\infty)$ of X_1 so that

$$f_{X_1}(x_1)=\int_{\mathbb{R}^{n-1}}f(x_1,\ldots,x_n)dx_2\cdots dx_n,\qquad\forall\,x_1\in\mathbb{R}.$$

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

$$f_{X_1,X_2}(x_1,x_2)=\int_{\mathbb{R}^{n-2}}f(x_1,\ldots,x_n)dx_3\cdots dx_n,\qquad\forall\,x_1,x_2\in\mathbb{R}.$$

And so on.

Definition 2.35 (**Independence of Random Variables**). Let $X_1,\ldots,X_n\colon \mathcal{C}\to\Omega$ 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

$$\mathbf{P}(X_1\in A_1,\ldots,X_n\in A_n)=\prod_{i=1}^{n}\mathbf{P}(X_i\in A_i),\qquad \forall\, A_1,\ldots,A_n\subseteq\mathcal{C}.$$

In the case that $X_1,\ldots,X_n$ are real-valued, this condition implies the following:

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

Exercise 2.36. Let $X_1,\ldots,X_n\colon\mathcal{C}\to\Omega$ be discrete random variables. Assume that

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

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 real-valued 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 $X,Y\colon\Omega\to\mathbb{C}$ be independent random variables. If either condition $X,Y\geq 0$ or $\mathbf{E}\left|XY\right|<\infty$ or $\mathbf{E}\left|X\right|,\mathbf{E}\left|Y\right|<\infty$ holds, then*

$$\mathbf{E}(XY)=\mathbf{E}X\mathbf{E}Y.$$

More generally, if $X,Y\colon\Omega\to S$ are independent random variables, if $F,G\colon S\to\mathbb{C}$, and if either $F(X),G(Y)\geq 0$ or $\mathbf{E}\left|F(X)G(Y)\right|<\infty$ or $\mathbf{E}\left|F(X)\right|,\mathbf{E}\left|G(Y)\right|<\infty$, then

$$\mathbf{E}(F(X)G(Y))=\mathbf{E}F(X)\mathbf{E}G(Y).$$

Proposition 2.40. *Let $0=n_0<n_1<n_2<\ldots<n_k=n$ be integers. Let $X_1,\ldots,X_n$ be real-valued, 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,*

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

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.**

Definition 2.44 (**Almost Sure Convergence**)**.** We say random variables $Y_1, Y_2, \ldots : \mathcal{C} \to \mathbb{R}$ converge **almost surely** (or **with probability one**) to a random variable $Y : \mathcal{C} \to \mathbb{R}$ if

$$\mathbf{P}(\lim_{n\to\infty} Y_n = Y) = 1.$$

That is, $\mathbf{P}(\{c \in \mathcal{C} \colon \lim_{n\to\infty} Y_n(c) = Y(c)\}) = 1$

Definition 2.45 (**Convergence in Probability**)**.** We say that a sequence of random variables $Y_1, Y_2, \ldots : \mathcal{C} \to \mathbb{R}$ **converges in probability** to a random variable $Y : \mathcal{C} \to \mathbb{R}$ if: for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}(|Y_n - Y| > \varepsilon) = 0.$$

That is, $\forall\, \varepsilon > 0$, $\lim_{n\to\infty} \mathbf{P}(c \in \mathcal{C} \colon |Y_n(c) - Y(c)| > \varepsilon) = 0$.

Definition 2.46 (**Convergence in Distribution**)**.** We say that real-valued random variables $Y_1, Y_2, \ldots$ **converge in distribution** to a real-valued random variable Y if, for any $t \in \mathbb{R}$ such that $s \mapsto \mathbf{P}(Y \leq s)$ is continuous at $s = t$,

$$\lim_{n\to\infty} \mathbf{P}(Y_n \leq t) = \mathbf{P}(Y \leq t).$$

Note that the random variables are allowed to have different domains.

Definition 2.47 (**Convergence in L_p**)**.** Let $0 < p \leq \infty$. We say that random variables $Y_1, Y_2, \ldots : \mathcal{C} \to \mathbb{R}$ **converge in** L_p to $Y : \mathcal{C} \to \mathbb{R}$ if $\|Y\|_p < \infty$ and

$$\lim_{n\to\infty} \|Y_n - Y\|_p = 0.$$

(Recall that $\|Y\|_p := (\mathbf{E}\,|Y|^p)^{1/p}$ if $0 < p < \infty$ and $\|X\|_\infty := \inf\{c > 0 \colon \mathbf{P}(|X| \leq c) = 1\}$.)

Exercise 2.48. Let $Y_1, Y_2, \ldots \colon \mathcal{C} \to \mathbb{R}$ be random variables that converge almost surely to a random variable $Y \colon \mathcal{C} \to \mathbb{R}$. Show that $Y_1, Y_2, \ldots$ converges in probability to Y in the following way.

- For any $\varepsilon > 0$ and for any positive integer n, let

$$A_{n,\varepsilon} := \bigcup_{m=n}^{\infty} \{c \in \mathcal{C} \colon \ |Y_m(c) - Y(c)| > \varepsilon\}.$$

  Show that $A_{n,\varepsilon} \supseteq A_{n+1,\varepsilon} \supseteq A_{n+2,\varepsilon} \supseteq \cdots$.
- Show that $\mathbf{P}(\cap_{n=1}^{\infty} A_{n,\varepsilon}) = 0$.
- Using Continuity of the Probability Law, deduce that $\lim_{n\to\infty} \mathbf{P}(A_{n,\varepsilon}) = 0$.

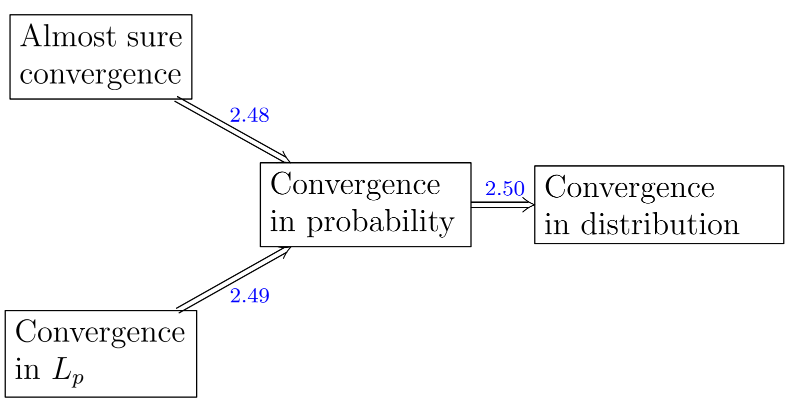
Now, show that the converse is false. That is, find random variables $Y_1, Y_2, \ldots$ that converge in probability to Y, but where $Y_1, Y_2, \ldots$ do not converge to Y almost surely.

Exercise 2.49. Let $0 < p \leq \infty$. Show that, if $Y_1, Y_2, \ldots \colon \mathcal{C} \to \mathbb{R}$ converge to $Y \colon \mathcal{C} \to \mathbb{R}$ in L_p, then $Y_1, Y_2, \ldots$ converges to Y in probability. Then, show that the converse is false.

Exercise 2.50. Suppose random variables $Y_1, Y_2, \ldots \colon \mathcal{C} \to \mathbb{R}$ converge in probability to a random variable $Y \colon \mathcal{C} \to \mathbb{R}$. Prove that $Y_1, Y_2, \ldots$ converge in distribution to Y. Then, show that the converse is false.

Exercise 2.51. Prove the following statement. Almost sure convergence does not imply convergence in L_2, and convergence in L_2 does not imply almost sure convergence. That is, find random variables that converge in L_2 but not almost surely. Then, find random variables that converge almost surely but not in L_2.

Remark 2.52. The following table summarizes our different notions of convergence of random variables, i.e. the following table summarizes the implications of Exercises 2.49, 2.50 and 2.48.



Theorem 2.53 (**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 $\varepsilon > 0$*

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

That is, $\frac{X_1+\cdots+X_n}{n}$ converges in probability to μ as $n \to \infty$.

Theorem 2.54 (**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*

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

That is, $\frac{X_1+\cdots+X_n}{n}$ converges almost surely to μ as $n\to\infty$.

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

Theorem 2.55 (**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 2.56. 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 2.57. 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 2.58 (**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 2.59 (**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 $-\infty \leq a \leq \infty$,

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

That is, $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ converges in distribution to a standard Gaussian as $n \to \infty$.

Theorem 2.60 (**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 2.61 (**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 2.62. 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 2.63. 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 2.64. 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 that has probability p of landing heads and probability $1 - p$ of landing tails. Similarly, there is a coin on the lily pad w that 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.

The set of 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,

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

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.

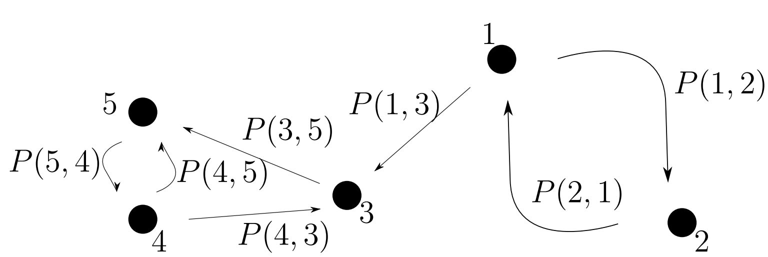
Example 3.7. Consider the Markov chain defined by the matrix $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Note that $P^n = P$ for any positive odd integer n, and $P^n = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ for any positive even integer n. So, if μ is any 1×2 row vector with unequal entries, 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. And a **random walk on a graph** G is any Markov chain with state space V such that $P(x, y) = 0$ whenever $\{x, y\} \notin E$.



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 3.34 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 $\mathbf{P}_\mu(X_0 = x_0) = \mu(x_0)$ 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 $\mu(y) = 0$ 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.

More generally, if $A \subseteq \Omega$, the **hitting time** of A is defined as

$$N := \min\{n \geq 1 \colon X_n \in A\}.$$

And N is a stopping time since, for any $n \geq 1$,

$$\{N = n\} = \{X_1 \in A^c, \ldots, X_{n-1} \in A^c, X_n \in A\} = \{(X_1, \ldots, X_n) \in A^c \times \cdots \times A^c \times A\}.$$

Exercise 3.16. Let M, N be stopping times for a Markov chain $X_0, X_1, \ldots$. Show that $\max(M, N)$ and $\min(M, N)$ are stopping times. In particular, if $n \geq 0$ is fixed, then $\max(M, n)$ and $\min(M, n)$ are stopping times

Theorem 3.17 (**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 3.18)

$$\begin{aligned}&\mathbf{P}_{x_0}((X_{T+1}, \ldots, X_{T+\ell}) \in A \,|\, T = n,\, (X_0, \ldots, X_n) = (x_0, \ldots, x_n))\\ &\quad = \mathbf{P}((X_{T+1}, \ldots, X_{T+\ell}) \in A \,|\, T = n,\, (X_0, \ldots, X_n) = (x_0, \ldots, x_n))\\ &\quad = \mathbf{P}((X_{n+1}, \ldots, X_{n+\ell}) \in A \,|\, T = n,\, (X_0, \ldots, X_n) = (x_0, \ldots, x_n))\\ &\quad = \mathbf{P}((X_{n+1}, \ldots, X_{n+\ell}) \in A \,|\, (X_0, \ldots, X_n) = (x_0, \ldots, x_n))\\ &\quad = \mathbf{P}((X_{n+1}, \ldots, X_{n+\ell}) \in A \,|\, X_n = x_n) \qquad \text{, by Exercise 3.19}\\ &\quad = \mathbf{P}((X_1, \ldots, X_\ell) \in A \,|\, X_0 = x_n) \qquad \text{, by Exercise 3.19}\\ &\quad = \mathbf{P}_{x_n}((X_1, \ldots, X_\ell) \in A), \qquad \text{, by definition of } P_{x_n}.\end{aligned}$$

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. $\square$

Exercise 3.18. 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 3.19. 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

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

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 3.20. 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 3.21. 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 3.17), for any $n\geq 1$,

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

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,

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

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 3.22. *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 3.23 (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**.

Recall that ρ_{yy} is defined in Definition 3.13.

Example 3.24 (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 $P(x,y) = \mathbf{P}(X_1 = y \mid X_0 = x)$, 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 $p^{(n)}(x,y) := \mathbf{P}(X_n = y \mid X_0 = x)$.

Proposition 3.25 (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 3.26. *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 3.26. 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 3.25 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 3.25. Let $x, y \in \Omega$. Using the Total Probability Theorem, we have

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

Finally, the Markov property and Exercise 3.19 imply that

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

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

Definition 3.27 (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 3.28. *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.

$$\begin{aligned}
\mathbf{P}_x(T_y > kj) &= \mathbf{P}_x(T_y > kj \,|\, T_y > (k-1)j)\mathbf{P}_x(T_y > (k-1)j) \\
&\leq \max_{z\in\Omega} \mathbf{P}_z(T_y > j)\mathbf{P}_x(T_y > (k-1)j), \qquad \text{by Exercise 3.29} \\
&\leq \max_{z\in\Omega} \mathbf{P}_z(T_y > r(z, y))\mathbf{P}_x(T_y > (k-1)j), \qquad \text{since } r(z, y) \leq j \\
&= \max_{z\in\Omega}(1 - \mathbf{P}_z(T_y \leq r(z, y)))\mathbf{P}_x(T_y > (k-1)j) \\
&\leq \max_{z\in\Omega}(1 - P^{r(z,y)}(z, y))\mathbf{P}(T_y > (k-1)j), \qquad \text{by Exercise 3.30} \\
&\leq (1 - \varepsilon)\mathbf{P}(T_y > (k-1)j).
\end{aligned}$$

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

Exercise 3.29. 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 3.19)

Exercise 3.30. 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 3.31. 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 3.28 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 3.32. 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 3.33 (**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 3.34 (**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 3.32,

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

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

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

Lemma 3.36. *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 3.37 (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 3.34, there exists at least one stationary distribution π such that $\pi=\pi P$. Let I denote the $|\Omega|\times|\Omega|$ identity matrix. Lemma 3.36 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 3.38. *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 3.34. Also, define $\pi_z(y) := \widetilde{\pi}_z(y)/\mathbf{E}_z T_z$. Theorem 3.34 and Remark 3.35 imply that π_z is a stationary distribution such that $\pi_z = \pi_z P$. Theorem 3.37 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.) □

Exercise 3.39 (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 3.40 (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 3.43.)

Exercise 3.41. Give an example of a Markov chain where there are at least two different stationary distributions.

Exercise 3.42. 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 3.43. 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 3.44. 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 3.89 below.)

Definition 3.45 (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 3.46. Give an example of a random walk on a graph that is not reversible.

Proposition 3.47 (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 3.48. 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 inner product function:

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

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

$$\langle f, Pg\rangle_\pi = \langle Pf, g\rangle_\pi.$$

In particular, 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 3.49. *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 3.34, $\pi(x) > 0$ 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 3.50. 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 3.51. 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 3.47, 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 3.52 (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 3.53. 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 3.54 (**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 3.37, we know an irreducible Markov chain has a unique stationary distribution, and Corollary 3.38 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 3.55 (**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 3.56. Give an explicit example of a Markov chain where every state has period 100.

Lemma 3.57. *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 3.58. *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 3.59 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 3.59. *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 3.60 (**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 3.61. 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 3.62. 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 3.63 (**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 3.58 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 3.34 (and Theorem 3.37), $\min_{z\in\Omega} \pi(z) > 0$. 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 3.62, 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 3.62 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 3.64 (**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 3.63), 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 3.71; all states are transient for the nearest neighbor simple random walk on $\mathbb{Z}^{3}$, see Theorem 3.81 below.) 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 3.32, 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 3.65 (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 3.66. 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 3.66 even when the simple random walk starts away from 0.

Theorem 3.67. *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 3.67, we prove some lemmas.

Lemma 3.68 (**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

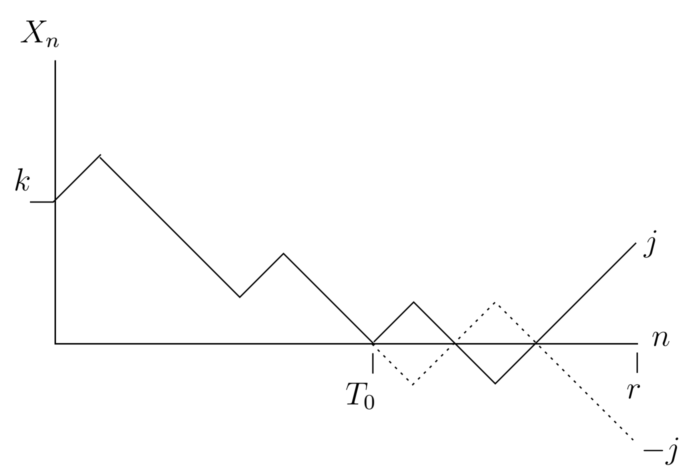
$$\mathbf{P}_k(T_0 = s)\mathbf{P}_0(X_{r-s} = -j) \overset{(*)}{=} \mathbf{P}_k(T_0 = s,\, X_r = -j).$$

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 3.69. We can interpret Lemma 3.68 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 3.70. *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 3.68), 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 3.67. Summing the upper bound of Exercise 3.66, we have

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

Then Lemma 3.70 completes the proof. □

Exercise 3.71. 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 3.72. 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 3.73. 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 $\mathbb{Z}^2$.

3.6. **Random Walks on Integer Lattices.**

Definition 3.74 (Random Walk). Let $X : \mathbb{R}^d \to \mathbb{R}^d$ be a random variable. Let $X_1, X_2, \ldots$ be i.i.d. copies of X. Let $x \in \mathbb{R}^d$. Let $X_0 := x$. For any $n \geq 0$, let $S_n := X_0 + \cdots + X_n$. We call the sequence of random variables $S_0, S_1, \ldots$ a **random walk** on $\mathbb{R}^d$ started at x.

3.6.1. *Limiting Behavior.*

Theorem 3.75. *Let $S_0, S_1, \ldots$ be a random walk on $\mathbb{R}$ with $S_0 = 0$. Exactly one of the following four conditions holds with probability one.*

(i) $S_n = 0$ *for all* $n \geq 1$.
(ii) $\lim_{n\to\infty} S_n = \infty$.
(iii) $\lim_{n\to\infty} S_n = -\infty$.
(iv) $-\infty = \liminf_{n\to\infty} S_n$ *and* $\limsup_{n\to\infty} S_n = \infty$.

Exercise 3.76. Let $X_1, X_2, \ldots \colon \Omega \to \mathbb{R}$ be i.i.d. In each of the cases below, show that with probability one, $-\infty = \liminf_{n\to\infty} S_n$ and $\limsup_{n\to\infty} S_n = \infty$.

- The distribution μ_{X_1} is symmetric about 0 (i.e. $\mu_{-X_1} = \mu_{X_1}$) and $\mathbf{P}(X_1 = 0) < 1$.
- $\mathbf{E}X_1 = 0$ and $\mathbf{E}X_1^2 \in (0, \infty)$. (Hint: use the Central Limit Theorem.)

For example, when $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$ and $S_0 = 0$, show that with probability one, $S_0, S_1, \ldots$ takes every integer value infinitely many times.

The reasoning of Proposition 3.22 implies the following.

Exercise 3.77. Let $S_0, S_1, \ldots$ be a random walk with $S_0 = 0$. Let Y be the number of times the random walk takes the value 0. Let $T_0 := \min\{n \geq 1 \colon S_n = 0\}$.

- Y is a geometric random variable with success probability $\mathbf{P}(T_0 = \infty)$.
- $\mathbf{E}Y = \frac{1}{\mathbf{P}(T_0 = \infty)}$. (Here we interpret $1/0$ as ∞.)

(Hint: $\{Y = k\} = \{T_0^{(k-1)} < \infty, T_0^{(k)} = \infty\} = \{T_0^{(k-1)} < \infty, T_0^{(k)} - T_0^{(k-1)} = \infty\}$.)

Theorem 3.78. *Let $S_0, S_1, \ldots$ be a random walk in $\mathbb{R}^d$ started at $x = 0$. Let $T := \min\{n \geq 1 \colon S_n = 0\}$. Then the following are equivalent*

(i) $\mathbf{P}(T < \infty) = 1$.
(ii) $\mathbf{P}(S_n = 0 \text{ for infinitely many } n \geq 1) = 1$.
(iii) $\sum_{n=0}^{\infty} \mathbf{P}(S_n = 0) = \infty$.

If additionally $S_1, S_2, \ldots$ only takes values in $\mathbb{Z}^d$, then (i),(ii),(iii) are equivalent to:

(iv)
$$\infty = \lim_{s\to 1^-} \int_{[-\pi,\pi]^d} \frac{1}{1 - s\phi(y)} dy.$$

Here $i = \sqrt{-1}$, $\phi(y) := \mathbf{E}e^{i\langle y, X_1\rangle}$, $\forall\, y \in \mathbb{R}^d$, and for any $x = (x_1, \ldots, x_d), y = (y_1, \ldots, y_d) \in \mathbb{R}^d$, we define $\langle x, y\rangle := \sum_{j=1}^d x_j y_j$.

Proof. Let Y be the number of times the random walk takes the value 0. Let $T^{(n)}$ be the n^{th} return time to 0 for any $n \geq 1$. Then

$$Y = \sum_{n=0}^{\infty} 1_{S_n = 0} = \sum_{n=0}^{\infty} 1_{T^{(n)} < \infty}, \qquad T^{(0)} := 0,$$

$$\mathbf{E}Y = \sum_{n=0}^{\infty} \mathbf{P}(S_n = 0) = \sum_{n=0}^{\infty} \mathbf{P}(T^{(n)} < \infty). \qquad (*)$$

If (i) occurs, then $\mathbf{P}(T^{(n)} < \infty) = 1$ for every $n \geq 1$ by Proposition 3.22, so that (ii) occurs. If (ii) occurs, then $\mathbf{P}(T^{(n)} < \infty) = 1$ for every $n \geq 1$, so the second equality of $(*)$ shows that (iii) occurs. If (iii) occurs, then (i) occurs by Exercise 3.77. So, (i),(ii),(iii) are equivalent.

By $(*)$, it remains to show that the right side of (iv) is equal to $(2\pi)^d\mathbf{E}Y$. Recall that $\int_{-\pi}^{\pi} e^{im\theta}d\theta = 0$ for any nonzero $m \in \mathbb{Z}$, while $\int_{-\pi}^{\pi} e^{i0\theta}d\theta = 2\pi$. Therefore, for any $n \geq 0$,

$$1_{S_n=0} = \int_{[-\pi,\pi]^d} e^{i\langle y,S_n\rangle}\frac{dy}{(2\pi)^d}.$$

Taking expected values of both sides,

$$\mathbf{P}(S_n = 0) = \int_{[-\pi,\pi]^d} \mathbf{E}e^{i\langle y,S_n\rangle}\frac{dy}{(2\pi)^d}. \qquad (**)$$

Recalling $S_n = X_1 + \cdots + X_n$ and using that $X_1, \ldots, X_n$ are i.i.d., we have $\mathbf{E}e^{i\langle y,S_n\rangle} = \prod_{j=1}^{n} \mathbf{E}e^{i\langle y,X_j\rangle} = (\phi(y))^n$. So, multiplying both sides of $(**)$ by s^n and summing over $n \geq 0$,

$$\sum_{n=0}^{\infty} s^n\mathbf{P}(S_n = 0) = \int_{[-\pi,\pi]^d} \sum_{n=0}^{\infty}(s\phi(y))^n\frac{dy}{(2\pi)^d} = \int_{[-\pi,\pi]^d} \frac{1}{1 - s\phi(y)}\frac{dy}{(2\pi)^d}.$$

(Since $|\phi(y)| \leq 1\ \forall\ y \in \mathbb{R}^d$, if $|s| < 1$, then $|s\phi(y)| < 1\ \forall\ y \in \mathbb{R}^d$.) Letting $s \to 1^-$, the left side increases monotonically to $\mathbf{E}Y$ by $(*)$, so the limit of the right side exists as well. $\square$

Exercise 3.79. Let $(X_0, X_1, \ldots)$ be a finite, irreducible Markov chain with transition matrix P and state space Ω. For any $x, y \in \Omega$, define

$$G(x, y) := \mathbf{E}_x \sum_{n=0}^{\infty} 1_{\{X_n=y\}} = \sum_{n=0}^{\infty} \mathbf{P}^n(x, y)$$

to be the expected number of visits to y starting from x. Show that the following are equivalent:

(i) $G(x, x) = \infty$ for some $x \in \Omega$.
(ii) $G(x, y) = \infty$ for all $x, y \in \Omega$.
(iii) $\mathbf{P}_x(T_x < \infty)$ for some $x \in \Omega$.
(iii) $\mathbf{P}_x(T_y < \infty)$ for all $x, y \in \Omega$.

So, in an irreducible finite Markov chain, a single state is recurrent if and only if all states are recurrent.

Definition 3.80 (**Simple Random Walk**)**.** For any $1 \leq j \leq d$, let $e_j \in \mathbb{R}^d$ be the vector with a 1 in the j^{th} entry and zeros in all other entries, so that $e_1, \ldots, e_d$ is the standard basis of $\mathbb{R}^d$. Let X be a random variable so that $\mathbf{P}(X = e_j) = \mathbf{P}(X = -e_j) = 1/(2d)$ for all $1 \leq j \leq d$. Let $X_1, X_2, \ldots$ be i.i.d. copies of X. The random walk $S_n := X_1 + \cdots + X_n$, $\forall$ $n \geq 1$ with $S_0 := 0$ is called the **simple random walk** on $\mathbb{Z}^d$.

The Simple Random Walk is the most basic random walk. It may be surprising that the transcience/recurrence of this random walk depends on d. Note that each point in the integer grid $\mathbb{Z}^d$ has $2d$ locations to move to at each step of the walk. And when d is large, there are more ways for the random walk to wander away from the origin.

Theorem 3.81. *Simple Random Walk is recurrent when $d \leq 2$ and transient when $d \geq 3$.*

Proof. It suffices to check whether or not condition (iv) holds of Theorem 3.78. For any $y \in \mathbb{R}^d$, we have

$$\phi(y) = \mathbf{E}e^{i\langle y, X_1\rangle} = \frac{1}{2d}\sum_{j=1}^{d}[e^{iy_j} + e^{-iy_j}] = \frac{1}{d}\sum_{j=1}^{d}\cos(y_j) = 1 + \frac{1}{d}\sum_{j=1}^{d}[-1+\cos(y_j)].$$

For any $z \in [-\pi, \pi]$, we have $z^2/4 \leq 1 - \cos(z) \leq z^2$ by e.g. taking derivatives and using the Fundamental Theorem of Calculus. Therefore, for any $y \in \mathbb{R}^d$,

$$-\frac{1}{d}\sum_{j=1}^{d} y_j^2 \leq \frac{1}{d}\sum_{j=1}^{d}[-1+\cos(y_j)] \leq -\frac{1}{4d}\sum_{j=1}^{d} y_j^2.$$

So, for any $y \in \mathbb{R}^d$, and for any $0 < s < 1$,

$$1 - s + s\frac{1}{4d}\sum_{j=1}^{d} y_j^2 \leq 1 - s\phi(y) \leq 1 - s + s\frac{1}{d}\sum_{j=1}^{d} y_j^2.$$

Letting $s \to 1^-$, and noting that the integrand increases monotonically in a neighborhood of $y = 0$ while remaining bounded outside this neighborhood,

$$(d/4)\int_{[-\pi,\pi]^d} \frac{1}{\sum_{j=1}^{d} y_j^2} dy \leq \lim_{s\to 1^-}\int_{[-\pi,\pi]^d} \frac{1}{1 - s\phi(y)} dy \leq d\int_{[-\pi,\pi]^d} \frac{1}{\sum_{j=1}^{d} y_j^2} dy.$$

And $\int_{[-\pi,\pi]^d} \frac{1}{\sum_{j=1}^{d} y_j^2} dy = \infty$ if and only if $d \leq 2$, by e.g. changing to polar coordinates. □

Exercise 3.82. Give a combinatorial proof that the simple random walk $S_0, S_1, \ldots$ on $\mathbb{Z}^d$ is recurrent for $d \leq 2$. That is, estimate $\mathbf{P}(S_n = 0) \approx n^{-d/2}$ when n is large and $d \leq 2$, and conclude $\sum_{n=0}^{\infty} \mathbf{P}(S_n = 0) = \infty$ for $d \leq 2$. (Hint: use Stirling's Formula, Proposition 3.93)

Exercise 3.83. Show that if the Simple Random Walk on $\mathbb{Z}^d$ is recurrent, then this random walk takes every value in $\mathbb{Z}^d$ infinitely many times. And if the Simple Random Walk on $\mathbb{Z}^d$ is transient, then this random walk takes any fixed value in $\mathbb{Z}^d$ only finitely many times.

Exercise 3.84. Let $0 < p < 1$. Consider the random walk on $\mathbb{Z}$ such that $\mathbf{P}(X_1 = 1) = p$ and $\mathbf{P}(X_1 = -1) = 1 - p$. Show that the corresponding random walk $S_0, S_1, \ldots$ is transient when $p \neq 1/2$.

Exercise 3.85. Let $S_0, S_1, \ldots$ and $S_0', S_1', \ldots$ be independent simple random walks on $\mathbb{Z}^d$. Let $N := \sum_{n,m\geq 0} 1_{S_n = S_m'}$ be the number of pairs of intersections of these two random walks. For any $y \in \mathbb{R}^d$, let $\phi(y) := \mathbf{E}e^{i\langle y, X_1\rangle}$.

- Show $\mathbf{E}N = \lim_{s\to 1^-}\int_{[-\pi,\pi]^d} \frac{1}{|1-s\phi(y)|^2}\frac{dy}{(2\pi)^d}$. (Hint: consider $\mathbf{E}e^{i\langle y, (S_n - S_m')\rangle}$.)
- For what $d \geq 1$ is $\mathbf{E}N < \infty$?
- Let $C := \{S_n : n \geq 0\} \cap \{S_n' : n \geq 0\}$ be the intersection set of the two independent random walks. Let $|C|$ denote the cardinality of C. Show that if the simple random walk on $\mathbb{Z}^d$ is transient, then $\mathbf{P}(N = \infty) = 1$ if and only if $\mathbf{P}(|C| = \infty) = 1$. (Hint: $N = \sum_{x\in C} N_x N_x'$ where $N_x := \sum_{n\geq 0} 1_{S_n = x}$ is the number of visits of the first random walk to x.) In the recurrent case $d = 1, 2$, Exercise 3.83 implies that $\mathbf{P}(|C| = \infty) = 1$. For any $d \geq 1$, note that $N < \infty$ implies $|C| < \infty$. It can also be shown that $\mathbf{P}(N < \infty) \in \{0, 1\}$, $\mathbf{P}(|C| = \infty) \in \{0, 1\}$, and that $\mathbf{P}(N < \infty) = 1$ if and only if

$\mathbf{E}N < \infty$ (you don't have to show these things). In summary, $\mathbf{P}(|C| = \infty) = 1$ if and only if $\mathbf{E}N = \infty$.

- Hypothesize what happens to $\mathbf{E}N$ when we instead consider the tuples of intersections of $k > 2$ independent simple random walks in $\mathbb{R}^d$. (You don't have to prove your hypothesis.)

The following proposition will be derived from a more general result, Theorem 4.22 below.

Proposition 3.86 (**Wald's Equations**). *Let $X_1, X_2, \ldots\colon \mathcal{C} \to \mathbb{R}$ be i.i.d. Let N be a stopping time. Let $S_0, S_1, \ldots$ be the corresponding random walk with $S_0 := 0$.*

- *If $\mathbf{E}N < \infty$, and $\mathbf{E}|X_1| < \infty$, then $\mathbf{E}S_N = \mathbf{E}X_1\mathbf{E}N$.*
- *If $\mathbf{E}X_1 = 0, \mathbf{E}X_1^2 < \infty$ and $\mathbf{E}N < \infty$, then $\mathbf{E}S_N^2 = \mathbf{E}X_1^2\mathbf{E}N$.*

Example 3.87. Suppose $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$. Let $a, b \in \mathbb{Z}$ with $a < 0 < b$. Let $N := \min\{n \geq 1\colon S_n \notin (a, b)\}$. We first check that $\mathbf{E}N < \infty$. This follows from Lemma 3.28. The first part of Proposition 3.86 says $\mathbf{E}S_N = 0$. Note that S_N only takes two values, a and b, so $\mathbf{E}S_N$ is straightforward to compute directly. Let $c := \mathbf{P}(S_N = a)$. Then

$$0 = \mathbf{E}S_N = ca + (1 - c)b.$$

Solving for c we get

$$c = \mathbf{P}(S_N = a) = \frac{b}{b - a}, \qquad \mathbf{P}(S_N = b) = \frac{-a}{b - a}.$$

The second part of Proposition 3.86 says $\mathbf{E}S_N^2 = \mathbf{E}N$. Once again, S_N^2 only takes two values, so

$$\mathbf{E}N = \mathbf{E}S_N^2 = ca^2 + (1 - c)b^2 = \frac{a^2 b - ab^2}{b - a} = ab\frac{a - b}{b - a} = -ab.$$

Exercise 3.88. Let $1/2 < p < 1$. Consider the random walk on $\mathbb{Z}$ such that $\mathbf{P}(X_1 = 1) = p$ and $\mathbf{P}(X_1 = -1) = 1 - p$. Let $S_0, S_1, \ldots$ be the corresponding random walk with $S_0 := 0$. Let $N := \min\{n \geq 1\colon S_n > 0\}$. Using Wald's equation for $\min(N, n)$ and then letting $n \to \infty$, show that $\mathbf{E}N = 1/\mathbf{E}X_1 = 1/(2p - 1)$.

3.7. **Additional Comments.** The term "random walk" was first proposed by Karl Pearson in 1905 in a letter to *Nature*. In this letter, Pearson proposed model of mosquito infestation of a forest. At each time step, a single mosquito moves a fixed length at a randomly chosen angle. Pearson asked for the distribution of the mosquitoes in the forest after a long time has passed. Rayleigh answered the letter, since he had solved a similar problem in 1880 for the modeling of sound waves in a heterogeneous material. A sound wave traveling through a material can be modeled as summing a sequence of vectors of constant amplitude but random phase, i.e. a sum of the form $\sum_{j=1}^{n} e^{iY_j}$, where $Y_1, Y_2, \ldots$ are real-valued and independent.

In 1900, Bachelier proposed random walks as a model for stock prices, and he also related random walks to the continuous diffusion of heat. Apparently unaware of other related works, around 1905 Einstein published his work on Brownian motion, i.e. the path of a dust particle in the air pushed in random directions by collision with gas molecules. Einstein modeled this behavior with a random walk. Smoluchowski published results similar to Einstein in 1906.

Random Walks are some of the most basic stochastic processes. They are used to model random phenomena in many scientific fields. The Simple Random Walk is essentially a discrete version of Brownian Motion.

In the case that a finite Markov chain is not irreducible, the state space can be partitioned into irreducible "pieces" plus transient states, in the following way.

Theorem 3.89 (**Decomposition Theorem**). *Let $(X_0, X_1, \ldots)$ be a finite Markov chain with (finite) state space Ω. Then Ω can be written uniquely as the disjoint union*

$$\Omega = T \cup I_1 \cup \cdots \cup I_k$$

where T is the set of transient states of the Markov chain, and for each $1 \leq i \leq k$, $P|_{I_i}$ is an irreducible Markov chain on the state space I_i.

In the case that a finite, irreducible Markov chain is not aperiodic, then all states have the same period by Lemma 3.57, and a version of the convergence theorem holds.

Lemma 3.90. *Let $(X_0, X_1, \ldots)$ be a finite, irreducible Markov chain with (finite) state space Ω. Assume that all states in the Markov chain are recurrent with period $j \geq 1$. For any $x, y \in \Omega$, let $\mathcal{N}(x, y) := \{n \geq 1 \colon P^n(x, y) > 0\}$. Fix $x \in \Omega$. Then*

- *$\exists\, m_y \in \{0, 1, \ldots, j-1\}$ such that, for all $n \in \mathcal{N}(x, y)$, we have $n = m_y \bmod j$.*
- *For any $0 \leq m < j$, let $\Omega_m := \{y \in \Omega \colon m_y = m\}$. Let $0 \leq m \leq m' < j$. If $y \in \Omega_m$ and $y' \in \Omega_{m'}$, and $P^n(y, y') > 0$, then $n = (m' - m) \bmod j$. Also, Ω is the disjoint union of $\Omega_0, \ldots, \Omega_{j-1}$.*
- *For each $0 \leq m < j$, $P^j|_{\Omega_m}$ is an irreducible Markov chain where all states have period 1.*

Theorem 3.91 (**Convergence Theorem, Periodic Case**). *Let $(X_0, X_1, \ldots)$ be a finite, irreducible Markov chain with (finite) state space Ω. Assume that all states in the Markov chain are recurrent with period $j \geq 1$. Assume that a stationary distribution π exists for the Markov chain. Fix $x \in \Omega$. As in Lemma 3.90, let $\Omega_0, \ldots, \Omega_{j-1}$ be a decomposition of the state space Ω. Let $0 \leq m < j$ and let $y \in \Omega_m$. Then*

$$\lim_{n \to \infty} P^{nj+m}(x, y) = j \cdot \pi(y).$$

Our presentation above focused on random walks where X_1 is discrete. In the case that X_1 is not discrete, if $S_0, S_1, \ldots$ is a random walk with $S_0 := 0$, then $x \in \mathbb{R}^d$ is called a **recurrent value** for the random walk if, for any $\varepsilon > 0$, $\mathbf{P}(\|S_n - x\| < \varepsilon \text{ for infinitely many } n \geq 1) = 1$. Here $\|(x_1, \ldots, x_d)\| := (\sum_{j=1}^d x_j^2)^{1/2}$. And $x \in \mathbb{R}^d$ is called a **possible value** for the random walk if, for any $\varepsilon > 0$, $\exists\, n \geq 0$ such that $\mathbf{P}(\|S_n - x\| < \varepsilon) > 0$. The random walk is said to be **transient** if it has no recurrent values. Otherwise, the random walk is called **recurrent**. If the random walk is recurrent, it can be shown that the set of recurrent values is equal to the set of possible values, as in Exercise 3.83.

Theorem 3.78 can then be generalized as follows.

Theorem 3.92. *Let $S_0, S_1, \ldots$ be a random walk on $\mathbb{Z}^d$ with $S_0 := 0$. For any $y \in \mathbb{R}^d$, let $\phi(y) := \mathbf{E} e^{i\langle y, X_1 \rangle}$, where $i = \sqrt{-1}$.*

(a) *The convergence (or divergence) of $\sum_{n \geq 0} P(\|S_n\| < \varepsilon)$ for a single $\varepsilon > 0$ is sufficient to prove transience (or recurrence) of the random walk.*

(b) *Let $\delta > 0$. Then the random walk is recurrent if and only if*

$$\sup_{0 < s < 1} \int_{(-\delta, \delta)^d} \mathrm{Re} \frac{1}{1 - s\phi(y)} dy = \infty.$$

Lemma 3.93 (Stirling's Formula). *Let $n \in \mathbb{N}$. Then $n! \sim \sqrt{2\pi n} n^n e^{-n}$. That is,*

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

Proof. We prove the weaker estimate that $\exists\, c \in \mathbb{R}$ such that

$$n! = (1 + O(1/n)) e^{1-c} \sqrt{n} n^n e^{-n}. \qquad (*)$$

Note that $\log(n!) = \sum_{m=1}^{n} \log m$. We use integral comparison for this sum. On the interval $[m, m+1]$ the function $x \mapsto \log x$ has second derivative $O(1/m^2)$. So, Taylor expansion (i.e. the trapezoid rule) gives

$$\int_m^{m+1} \log x dx = \frac{1}{2}\log(m+1) + \frac{1}{2}\log m + O(1/m^2).$$

$$\int_1^n \log x dx = \sum_{m=1}^{n-1} \int_m^{m+1} \log x dx = \sum_{m=1}^{n-1} \log m + \frac{1}{2}\log n + c + O(1/n).$$

Since $\int_1^n \log x dx = n(\log(n) - 1) + 1$, $\log(n!) = \sum_{m=1}^{n} \log m$, exponentiating proves $(*)$. $\square$

4. Martingales

We begin by reviewing 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

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

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, $\mathbf{E}(X 1_A) = (\mathbf{E}X)(\mathbf{E}1_A) = \mathbf{P}(A)\mathbf{E}X$. 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.

4.1. **Examples of Martingales.**

Example 4.11 (**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,

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

Example 4.12 (**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,

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

4.2. **Gambling Strategies.**

Example 4.13. 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 2.54) 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 4.11 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 4.14. *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 4.15. 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 4.16. 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 4.14. 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

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

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

Definition 4.17 (**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 4.18. 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 4.11, $M_0, M_1, \ldots$ 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 4.19. 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 4.20 (**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 4.14, $W_0, W_1, \ldots$ is a martingale. By definition of H_m,

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

$\square$

Theorem 4.21 (**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 4.20, 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 $n\geq 1$,

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

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

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

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

Theorem 4.22 (**Optional Stopping Theorem, Version 3**). *Let $(M_0,M_1,\ldots)$ be a martingale, and let T be a stopping time such that $\mathbf{E}T<\infty$. Let $d\in\mathbb{R}$. Assume that $\mathbf{E}\,|M_{n+1}-M_n|\leq d$ for all $n\geq 0$. Then $\mathbf{E}M_T=\mathbf{E}M_0$.*

Exercise 4.23. Prove Wald's Equation, Proposition 3.86, using Theorem 4.22.

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 4.24. The assumptions of the Optional Stopping Theorem cannot be abandoned, as shown in Example 4.18. 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 4.25 (**Gambler's Ruin**). We return to Example 4.12. 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 4.12 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 3.28 implies that $\mathbf{P}(T<\infty)=1$. (Consider the corresponding Markov chain on the state space $\{a,a+1,a+2,\ldots,b\}$ that "reflects at its boundary," so that $P(a,a+1)=1$ and $P(b,b-1)=1$.) From Theorem 4.21,

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

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 4.26. 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 4.27. 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 4.28. 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 4.26 and the Optional Stopping Theorem to get $0 = \mathbf{E}(Y_T^2 - T)$, then using Example 4.25,

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

Strictly speaking, the Optional Stopping Theorem, Version 2, does not apply, since the martingale is not bounded. But Theorem 4.20 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 4.29. 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 4.30. 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 4.31 (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 4.32. 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 $0 < k \leq m$,

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

Exercise 4.33. Let $X_1, X_2, \ldots$ be i.i.d. 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 $a \in \mathbb{Z}$.

4.3. **Concentration for Product Measures.** In certain cases, we can make rather strong conclusions about the distribution of sums of i.i.d. random variables, improving upon estimates from either the Markov or Chevyshev inequalities.

Theorem 4.34 (**Hoeffding Inequality/ Large Deviation Estimate**)**.** *Let $X_1, X_2, \ldots$ be independent identically distributed random variables with $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$. Let $a_1, a_2, \ldots \in \mathbb{R}$. Then, for any $n \geq 1$,*

$$\mathbf{P}\Big(\sum_{i=1}^{n} a_i X_i \geq t\Big) \leq e^{-\frac{t^2}{2\sum_{i=1}^n a_i^2}}, \qquad \forall\, t \geq 0.$$

Consequently,

$$\mathbf{P}\Big(\Big|\sum_{i=1}^{n} a_i X_i\Big| \geq t\Big) \leq 2e^{-\frac{t^2}{2\sum_{i=1}^n a_i^2}}, \qquad \forall\, t \geq 0.$$

Proof. By dividing $a_1, \ldots, a_n$ by a constant, we may assume $\sum_{i=1}^n a_i^2 = 1$. Let $\alpha > 0$. Using the (exponential) moment method, and $\alpha t \geq 0$,

$$\mathbf{P}(\sum_{i=1}^{n} a_i X_i \geq t) = \mathbf{P}(e^{\alpha \sum_{i=1}^n a_i X_i} \geq e^{\alpha t}) \leq e^{-\alpha t}\mathbf{E}e^{\alpha \sum_{i=1}^n a_i X_i} = e^{-\alpha t} \prod_{i=1}^{n} \mathbf{E}e^{\alpha a_i X_i}.$$

The last equality used independence of $X_1, X_2, \ldots$ and Proposition 2.39. Using an explicit computation and Exercise 4.35,

$$\mathbf{E}e^{\alpha a_i X_i} = (1/2)(e^{\alpha a_i} + e^{-\alpha a_i}) = \cosh(\alpha a_i) \leq e^{\alpha^2 a_i^2/2}, \qquad \forall\, i \geq 1.$$

In summary, for any $t \geq 0$

$$\mathbf{P}(\sum_{i=1}^{n} a_i X_i \geq t) \leq e^{-\alpha t} e^{\alpha^2 \sum_{i=1}^n a_i^2/2} = e^{-\alpha t + \alpha^2/2}.$$

Since $\alpha > 0$ is arbitrary, we choose α to minimize the right side. This minimum occurs when $\alpha = t$, so that $-\alpha t + \alpha^2/2 = -t^2/2$, giving the first desired bound. The final bound follows by writing $\mathbf{P}(|\sum_{i=1}^n a_i X_i| \geq t) = \mathbf{P}(\sum_{i=1}^n a_i X_i \geq t) + \mathbf{P}(-\sum_{i=1}^n a_i X_i \geq t)$ and then applying the first inequality twice. □

Exercise 4.35. Show that $\cosh(x) \leq e^{x^2/2}$, $\forall\, x \in \mathbb{R}$.

In particular, Hoeffding's inequality implies that

$$\mathbf{P}\Big(\frac{1}{n}\Big|\sum_{i=1}^{n} X_i\Big| \geq t\Big) \leq 2e^{-nt^2/2}, \qquad \forall\, t \geq 0.$$

This inequality is much stronger than either Markov's or Cheyshev's inequality, since they only respectively imply that

$$\mathbf{P}\Big(\frac{1}{n}\Big|\sum_{i=1}^{n} X_i\Big| \geq t\Big) \leq \frac{1}{t}, \quad \mathbf{P}\Big(\frac{1}{n}\Big|\sum_{i=1}^{n} X_i\Big| \geq t\Big) \leq \frac{1}{nt^2}, \qquad \forall\, t \geq 0.$$

Note also that Hoeffding's inequality gives a quantitative bound for any fixed $n \geq 1$, unlike the (non-quantitative) limit theorems which only hold as $n \to \infty$.

Exercise 4.36 (Chernoff Inequality). Let $0 < p < 1$. Let $X_1, X_2, \ldots$ be independent identically distributed random variables with $\mathbf{P}(X_1 = 1) = p$ and $\mathbf{P}(X_1 = 0) = 1 - p$ for any $i \geq 1$. Then for any $n \geq 1$

$$\mathbf{P}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i \geq t\Big) \leq e^{-np}\Big(\frac{ep}{t}\Big)^{tn}, \qquad \forall\, t \geq p.$$

Prove the same estimate for $\mathbf{P}(\frac{1}{n}\sum_{i=1}^n X_i \leq t)$ for any $t \leq p$. (Hint: $1 + x \leq e^x$ for any $x \in \mathbb{R}$, so $1 + (e^\alpha - 1)p \leq e^{(e^\alpha - 1)p}$.)

Exercise 4.37. For any natural number n and a parameter $0 < p < 1$, define an Erdös-Renyi graph on n vertices with parameter p to be a random graph (V, E) on a (deterministic) vertex set V of n vertices (thus (V, E) is a random variable taking values in the discrete space of all $2^{\binom{n}{2}}$ possible undirected graphs one can place on V) such that the events $\{i, j\} \in E$ for unordered pairs with $i, j \in V$ are independent and each occur with probability p.

For each $n \geq 1$, let (V_n, E_n) be an Erdös-Renyi graph on n vertices with parameter $p = 1/2$ (we do not require the graphs to be independent of each other). Define $d := p(n-1)$.

- Show that d is the expected degree of each vertex in G. (The degree of a vertex $v \in V$ is the number of vertices connected to v by an edge in E.)
- Show that there exists a constant $c > 0$ such that the following holds. Assume $p \geq \frac{c \log n}{n}$. Then with probability larger than .9, all vertices of G have degrees in the range $(.9d, 1.1d)$. (Hint: first consider a single vertex, then use the union bound over all vertices.)

Exercise 4.38 (**Azuma's Inequality**). In this exercise, we prove a generalization of the Hoeffding inequality to martingales. Let $c_1, c_2, \ldots > 0$. Let $(X_0, X_1, \ldots)$ be a martingale. Assume that $|X_n - X_{n-1}| \leq c_n$ for all $n \geq 1$. Then for any $t > 0$,

$$\mathbf{P}(|X_n - X_0| > t) \leq 2e^{-\frac{t^2}{2\sum_{i=1}^n c_i^2}}.$$

Prove this inequality using the following steps.

- Let $\alpha > 0$. Show that $\mathbf{E}e^{\alpha(X_n - X_0)} = \mathbf{E}[e^{\alpha(X_{n-1}-X_0)}\mathbf{E}(e^{\alpha(X_n - X_{n-1})}|\mathcal{F}_{n-1})]$. (When Y is a random variable, we denote $\mathbf{E}(Y|\mathcal{F}_n) := g(X_0, \ldots, X_n)$ where $g(x_0, \ldots, x_n) := \mathbf{E}(Y|X_0 = x_0, \ldots, X_n = x_n)$ for any $x_0, \ldots, x_n \in \mathbb{R}$.)
- For any $y \in [-1, 1]$, show that $e^{\alpha c_n y} \leq \frac{1+y}{2}e^{\alpha c_n} + \frac{1-y}{2}e^{-\alpha c_n}$.
- Take the conditional expectation of this inequality when $y = (X_n - X_{n-1})/c_n$.
- Now argue as in Hoeffding's inequality.

Using Azuma's inequality, deduce **McDiarmid's Inequality**. Let $X_1, \ldots, X_n$ be independent real-valued random variables. Let $c_1, c_2, \ldots > 0$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a measurable function such that, for any $1 \leq m \leq n$,

$$\sup_{x_1,\ldots,x_{m-1},x_m,x_m',x_{m+1},\ldots,x_n \in \mathbb{R}} |f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{m-1}, x_m', x_{m+1}, \ldots, x_n)| \leq c_m.$$

Then, for any $t > 0$,

$$\mathbf{P}(|f(X_1, \ldots, X_n) - \mathbf{E}f(X_1, \ldots, X_n)| > t) \leq 2e^{-\frac{t^2}{2\sum_{i=1}^n c_i^2}}.$$

(Note that a linear function f recovers Hoeffding's inequality, Theorem 4.34.)

5. Poisson Process

Before introducing the Poisson Process, we review 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**

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

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

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

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.

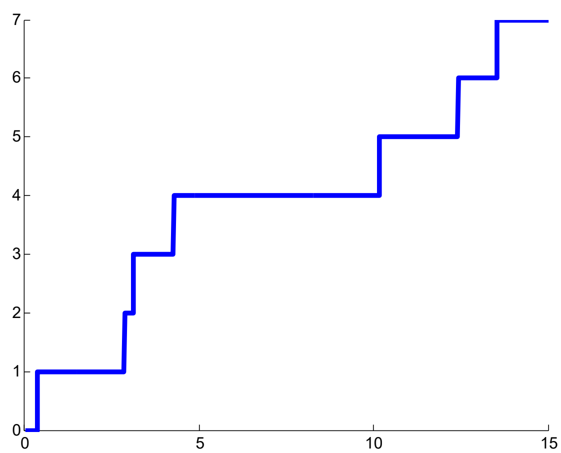


FIGURE 1. One Sample Path of a Poisson Process. The horizontal axis is the s-axis.

5.1. **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.

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

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 $N(s) := \max\{n \geq 0 \colon T_n \leq s\}$.

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.*

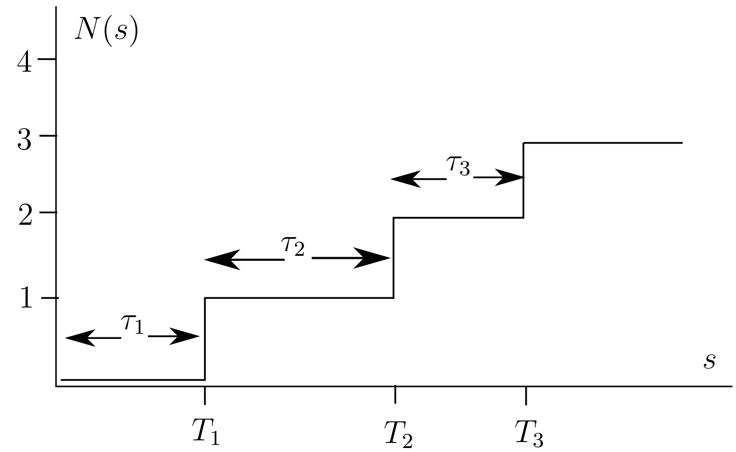


FIGURE 2. One Sample Path of a Poisson Process.

Proof. Let n be a nonnegative integer. Then

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

□

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),

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

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

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

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

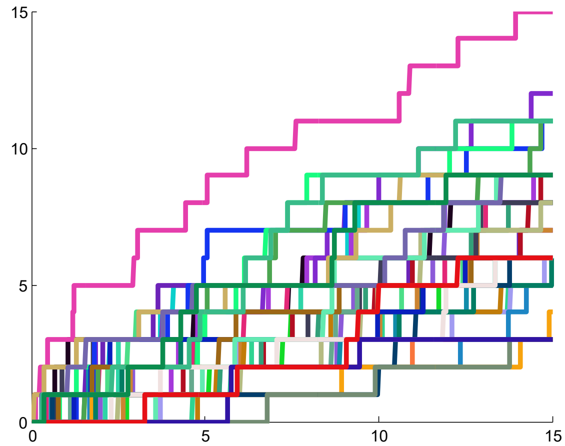


FIGURE 3. Several Sample Paths of a Poisson Process. The horizontal axis is the s-axis.

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

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

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

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

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

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

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. $\square$

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 $r \geq 0$.

5.2. **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 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.)

5.3. **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

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

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 $\forall\, j \geq 1$. □

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. □

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?

5.4. **Continuous-Time Chains and Semigroups.** In a discrete-time Markov chain, the chain changes its state at nonnegative integer times. In a continuous-time Markov chain, the chain changes its state at the transition times of a rate one Poisson process. That is, the times between changes of state are independent exponential random variables with parameter $\lambda = 1$.

For some applications, a continuous-time Markov chain could be more natural than its discrete-time counterpart.

Definition 5.25 (Continuous Time Markov Chain). Let $(Y_0, Y_1, \ldots)$ be a finite (discrete-time) Markov chain with (finite) state space Ω and transition matrix P. Let $\{N(t)\}_{t\geq 0}$ be a Poisson process with parameter $\lambda = 1$ that is independent of $(Y_0, Y_1, \ldots)$. The (finite) **continuous-time Markov chain** $\{X_t\}_{t\geq 0}$ with transition matrix P and state space Ω is defined by

$$X_t := Y_{N(t)}, \qquad \forall\, t \geq 0.$$

Let m be a positive integer and let P be an $m \times m$ real matrix. Define the matrix e^P by

$$e^P := \sum_{k=0}^{\infty} \frac{P^k}{k!}.$$

That is, the (x, y) entry of e^P, denoted $e^P(x, y)$ is

$$e^P(x, y) := \sum_{k=0}^{\infty} \frac{P^k(x, y)}{k!}.$$

For example, if I denotes the $m \times m$ identity matrix, we have $e^I = \sum_{k=0}^{\infty} \frac{I^k}{k!} = e \cdot I$.

Exercise 5.26. Let m be a positive integer and let P be an $m \times m$ real matrix.

- Show that the sum
$$\sum_{k=0}^{\infty} \frac{P^k}{k!}$$
converges. That is, e^P is well-defined.
- Show that
$$e^{P+I} = e^P e^I.$$
- Find $m \times m$ matrices P, Q such that $e^{P+Q} \neq e^P e^Q$.

Proposition 5.27 (**Markov Property, Continuous-Time**). *A (finite) continuous-time Markov chain satisfies the following Markov property: for all $x, y \in \Omega$, and for any $s, t > 0$,*

$$\mathbf{P}(X_{t+s} = y \,|\, X_s = x) = \mathbf{P}(X_t = y \,|\, X_0 = x) = e^{t(P-I)}(x, y).$$

Proof. Let n be a positive integer. From Definition 5.25, Theorem 5.13, the Markov property for $(Y_0, Y_1, \ldots)$, and Exercise 5.26,

$$\begin{aligned}
&\mathbf{P}(X_{t+s} = y \,|\, X_s = x, N(s) = n) = \mathbf{P}(Y_{N(t+s)} = y \,|\, Y_{N(s)} = x, N(s) = n) \\
&\quad = \mathbf{P}(Y_{N(t+s)-N(s)+n} = y \,|\, Y_n = x) = \mathbf{P}(Y_{N(t)+n} = y \,|\, Y_n = x) \\
&\quad = \sum_{k=0}^{\infty} \mathbf{P}(Y_{k+n} = y \,|\, Y_n = x)\mathbf{P}(N(t) = k) = \sum_{k=0}^{\infty} \mathbf{P}(Y_k = y \,|\, Y_0 = x)\mathbf{P}(N(t) = k) \\
&\quad = \sum_{k=0}^{\infty} P^k(x, y)\mathbf{P}(N(t) = k) = \sum_{k=0}^{\infty} P^k(x, y)e^{-t}\frac{t^k}{k!} = e^{-t}e^{tP}(x, y) = e^{t(P-I)}(x, y).
\end{aligned}$$

By averaging over all values of $N(s)$, we conclude that

$$\mathbf{P}(X_{t+s} = y \,|\, X_s = x) = e^{t(P-I)}(x, y).$$

Choosing in particular $s = 0$ concludes the proof. □

Definition 5.28 (**Semigroup, Heat Kernel**). Let $\{X_t\}_{t \geq 0}$ be a (finite) continuous-time Markov chain with transition matrix P and state space Ω. The **heat kernel** $H_t(x, y)$ of the Markov chain is defined as

$$H_t(x, y) := \mathbf{P}(X_t = y \,|\, X_0 = x) = e^{t(P-I)}(x, y), \qquad \forall\, t \geq 0,\ \forall\, (x, y) \in \Omega.$$

The set of matrices $\{H_t\}_{t \geq 0} := \{e^{t(P-I)}\}_{t \geq 0}$ is sometimes called a **semigroup with generator** $P - I$.

Exercise 5.29. Let m be a positive integer and let P be an $m \times m$ real matrix. Denote $H_t := e^{t(P-I)}$ for all $t \geq 0$. Let $f \in \mathbb{R}^m$ be a column vector. Then $H_t f$ denotes multiplying the matrix H_t against the vector f. Show the following:

- $H_0 = I$.
- $H_{s+t} = H_s H_t$ for all $s, t \geq 0$. (This identity is an analogue of the Chapman-Kolmogorov equation from Proposition 3.25.)
- $H_t 1 = 1$ for all $t \geq 0$. (Moreover, H_t is a stochastic matrix, for all $t \geq 0$. Here 1 denotes the vector of all ones.)
- $\frac{d}{dt}|_{t=0} H_t = \lim_{t \to 0^+} \frac{H_t - H_0}{t} = (P - I)$.
- For any $f \in \mathbb{R}^m$, we have

$$\frac{d}{dt} H_t f = (P - I) H_t f, \qquad \forall\, t \geq 0.$$

Exercise 5.30 (Markov Property, Continuous-Time). Show that a (finite) continuous-time Markov chain satisfies the following Markov property: for all $x, y \in \Omega$, for any $n \geq 1$, $t > 0$ and for any $s > s_{n-1} > \cdots > s_0 > 0$ and for all events H_{n-1} of the form $H_{n-1} = \cap_{k=0}^{n-1} \{X_{s_k} = x_k\}$, where $x_k \in \Omega$ for all $0 \leq k \leq n-1$, such that $\mathbf{P}(H_{n-1} \cap \{X_s = x\}) > 0$, we have

$$\mathbf{P}(X_{t+s} = y \,|\, H_{n-1} \cap \{X_s = x\}) = \mathbf{P}(X_t = y \,|\, X_0 = x).$$

Theorem 5.31 (The Convergence Theorem, Continuous-Time). *Let P be the transition matrix of a finite, irreducible Markov chain, with state space Ω and with (unique) stationary distribution π. Let H_t be the corresponding heat kernel. Then $\pi H_t = \pi$ for all $t \geq 0$ and*

$$\lim_{t \to \infty} \max_{x \in \Omega} \|H_t(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} = 0.$$

Unlike in the discrete-time case of Theorem 3.63, we do not need to assume aperiodicity of the Markov chain in Theorem 5.31. Also, since $\pi = \pi P$, we have

$$\pi H_t = \pi e^{t(P-I)} = e^{-t} \pi \sum_{k=0}^{\infty} \frac{t^k P^k}{k!} = e^{-t} \pi \sum_{k=0}^{\infty} \frac{t^k}{k!} = \pi.$$

Exercise 5.32. Let P be the transition matrix of a finite Markov chain. Show the following.

- If $\lambda \in \mathbb{C}$ is an eigenvalue of P, then $|\lambda| \leq 1$.
- If P is irreducible, then the vector space of eigenvectors of P corresponding to the eigenvalue 1 is the one-dimensional space spanned by the all ones vector $(1, \ldots, 1)^T$.
- if P is irreducible and aperiodic, then -1 is not an eigenvalue of P.

Assume that the Markov chain is finite, reversible, and irredicuble, and let π be the corresponding (unique) stationary distribution. For any $f, g \in \mathbb{R}^\Omega$, define the inner product

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

From Exercise 3.48, the inner product space $(\mathbb{R}^\Omega, \langle \cdot, \cdot \rangle_\pi)$ has an orthonormal basis of real-valued eigenfunctions $g_1, \ldots, g_{|\Omega|} \in \mathbb{R}^\Omega$ of P with real eigenvalues. So, if P is the transition matrix of a reversible, irreducible Markov chain, we can label the corresponding eigenvalues of P in decreasing order:

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\Omega|} \geq -1.$$

Exercise 5.32 implies that $|\lambda_j| \leq 1$ for all $1 \leq j \leq |\Omega|$ and $\lambda_1 > \lambda_2$.

$$H_t g_j = e^{t(P-I)} g_j = e^{-t} \sum_{k=0}^{\infty} \frac{t^k P^k}{k!} g_j = e^{-t} \sum_{k=0}^{\infty} \frac{t^k \lambda_j^k}{k!} g_j = e^{t(\lambda_j - 1)} g_j.$$

So, if $f \in \mathbb{R}^\Omega$, then we can write $H_t f$ in terms of its action on the basis vectors:

$$H_t f = H_t \sum_{j=1}^{|\Omega|} \langle f, g_j \rangle_\pi g_j = \sum_{j=1}^{|\Omega|} e^{t(\lambda_j - 1)} \langle f, g_j \rangle_\pi g_j, \qquad \forall\, t \geq 0.$$

$$\langle H_t f, g_i \rangle_\pi = \sum_{j=1}^{|\Omega|} e^{t(\lambda_j - 1)} \langle f, g_j \rangle_\pi \langle g_j, g_i \rangle_\pi = e^{t(\lambda_i - 1)} \langle f, g_i \rangle_\pi, \qquad \forall\, t \geq 0, \quad \forall\, 1 \leq i \leq |\Omega|. \qquad (*)$$

The **spectral gap** of P is defined as

$$\gamma := 1 - \lambda_2.$$

Also, define

$$\mathbf{E}_\pi f := \langle f, 1 \rangle, \qquad \|f\|_{2,\pi}^2 := \langle f, f \rangle_\pi, \qquad \mathrm{Var}_\pi(f) := \|f - \mathbf{E}_\pi f\|_{2,\pi}^2 .$$

In the reversible case, the following variant of Theorem 5.31 holds:

Proposition 5.33. *Let P be the transition matrix of a finite, reversible, irreducible Markov chain, with state space Ω and with (unique) stationary distribution π. Let H_t be the corresponding heat kernel. Let $\gamma := 1 - \lambda_2$ be the spectral gap of P. Then, for any $f \in \mathbb{R}^\Omega$,*

$$\|H_t f - \mathbf{E}_\pi f\|_{2,\pi}^2 \leq e^{-2\gamma t} \mathrm{Var}_\pi f, \qquad \forall\, t \geq 0.$$

Proof. From Exercise 5.29, for any $f \in \mathbb{R}^m$ and for any $t \geq 0$, we have

$$\frac{d}{dt} H_t f = (P - I) H_t f.$$

For any $t \geq 0$, let $u(t) := \|H_t f\|_{2,\pi}^2$. From the product rule, we have

$$u'(t) = 2 \langle H_t f, \frac{d}{dt} H_t f \rangle_\pi . = 2 \langle H_t f, (P - I) H_t f \rangle_\pi .$$

From Exercises 5.32 and 3.48, we can uniquely write $H_t f$ as a linear combination of orthonormal eigenvectors $g_1, \ldots, g_{|\Omega|} \in \mathbb{R}^\Omega$ of P, so that

$$u'(t) = 2 \sum_{i=1}^{|\Omega|} |\langle H_t f, g_i \rangle_\pi|^2 \langle g_i, (P - I) g_i \rangle_\pi = 2 \sum_{i=1}^{|\Omega|} |\langle H_t f, g_i \rangle_\pi|^2 (\lambda_i - 1).$$

Assume for now that $\mathbf{E}_\pi f = 0$, so that $\langle H_t f, g_1 \rangle_\pi \overset{(*)}{=} \mathbf{E}_\pi f = 0$ since $\lambda_1 = 1$ and g_1 is the constant vector. Then

$$u'(t) \leq -2\gamma \sum_{i=1}^{|\Omega|} |\langle H_t f, g_i \rangle_\pi|^2 = -2\gamma \|H_t f\|_{2,\pi}^2 = -2\gamma \cdot u(t).$$

That is, $\frac{d}{dt}[u(t) e^{2\gamma t}] = 0$ for all $t \geq 0$, so there exists $c > 0$ such that $u(t) = c e^{-2\gamma t}$. Since $u(0) = c = \|f\|_{2,\pi}^2$, we have shown that

$$\|H_t f\|_{2,\pi}^2 \leq e^{-2\gamma t} \|f\|_{2,\pi}^2 , \qquad \forall\, t \geq 0,$$

in the case that $\mathbf{E}_\pi f = 0$. For the general case, we apply the above result to $f - \mathbf{E}_\pi f$. $\square$

Exercise 5.34. Prove the following discrete-time version of the above spectral gap inequality.

Let P be the transition matrix of a finite, irreducible, reversible Markov chain, with state space Ω and with (unique) stationary distribution π. Let

$$\gamma_* := 1 - \max\{|\lambda| : \lambda \text{ is an eigenvalue of } P \text{ with } \lambda \neq 1\}$$

be the absolute spectral gap of P. Then, for any $f \in \mathbb{R}^\Omega$ and for any integer $k \geq 1$,

$$\mathrm{Var}_\pi(P^k f) \leq (1 - \gamma_*)^{2k} \mathrm{Var}_\pi f.$$

6. Renewal Theory

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 2.54, 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 $t \in [j, j+1)$. So, 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 $t - s$.

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 2.59), 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

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

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.

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.

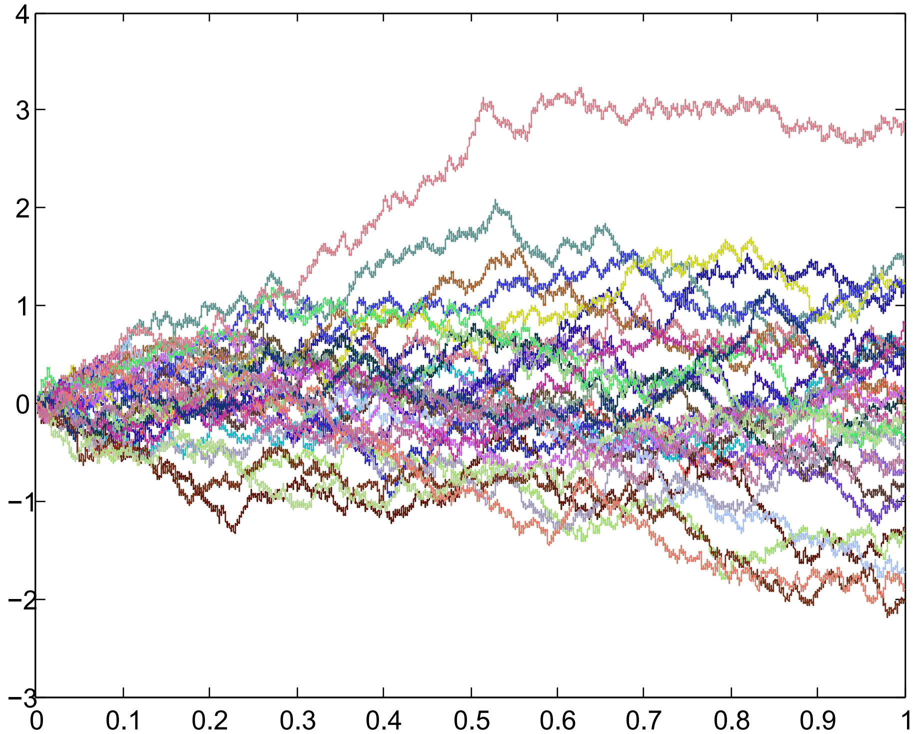


FIGURE 4. Sample Paths of Standard Brownian Motion. The horizontal axis is the t-axis.

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 4.21.

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 $0 \leq u \leq s$. $\square$

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.**

Proposition 7.9. *Let $\{B(t)\}_{t \geq 0}$ be a standard Brownian motion. Let $a, b > 0$. Let $T_a := \min\{t \geq 0 \colon B(t) = a\}$. Then*

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

Proof. Let $c := \mathbf{P}(T_a < T_{-b})$. Let $T := \min\{t \geq 0 \colon B(t) \in \{a, -b\}\}$. 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 7.10. 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

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

More generally, for any $\alpha \in \mathbb{R}$, let $Y(t) := e^{\alpha B(t) - \alpha^2 t/2}$. Show that $\{Y(t)\}_{t\geq 0}$ is a martingale.

Then, using the power series expansion of the exponential function, we have $Y(t) = \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} M_n(t)$ for some random variables $M_1(t), M_2(t), \ldots$, for any $\alpha \in \mathbb{R}$. It follows that $\{M_1(t)\}_{t\geq 0}$ is a martingale, $\{M_2(t)\}_{t\geq 0}$ is a martingale, etc. (Starting with the following sentence, you do not have to prove anything.) It turns out that

$$M_n(t) = t^{n/2} p_n(B(t)/\sqrt{t}), \quad \forall\, t \in \mathbb{R}, \quad \forall\, n \geq 1,$$

where p_n is the n^{th} Hermite polynomial, so that

$$p_n(x) = e^{x^2/2}(-1)^n \frac{d^n}{dx^n} e^{-x^2/2}, \quad \forall\, x \in \mathbb{R}, \quad \forall\, n \geq 1.$$

For example, using $n = 3$, we know that $\{(B(t))^3 - 3B(t)\}_{t\geq 0}$ is a martingale.

Proposition 7.11. *Let $a, b > 0$. Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Let $T = \min\{t \geq 0 \colon B(t) \notin (-b, a)\}$. Then*

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

Proof. Using Exercise 7.10 and the Optional Stopping Theorem, we get $0 = \mathbf{E}((B(T))^2 - T)$, then using Proposition 7.9,

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

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 4.28. □

Exercise 7.12. 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?

Proposition 7.13 (**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 7.14. 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 $\mathbf{E}T_x = \infty$.

Corollary 7.15.

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

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 7.13. □

Remark 7.16. 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 7.17 (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 7.18. 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 4.25, 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 7.19. 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 7.20. *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 $T_a := \min\{t > 0 \colon X(t) = a\}$. 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 7.19, $e^{\alpha X(t)}$ is a martingale. Let $T := \min\{t \geq 0 \colon X(t) \in \{a, b\}\}$. 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 $a < X(t) = \sigma B(t) + \mu t \leq \sigma B(t)$ for all $t \geq 0$. So, if we define $T_a' := \min\{t \geq 0 \colon B(t) = a/\sigma\}$, then $T_a = \infty$ implies $T_a' = \infty$, by property (i) of Brownian motion. So, $\mathbf{P}(T_a = \infty) \leq \mathbf{P}(T_a' = \infty)$, and $\mathbf{P}(T_a' = \infty) = 0$ by Proposition 7.13, since $\mathbf{P}(T_a' = \infty) = \lim_{s\to\infty} \int_{-a/\sigma}^{a/\sigma} e^{-\frac{y^2}{2s}} \frac{dy}{\sqrt{2\pi s}} = 0$.)

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 7.21. 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 negative drift $\mu < 0$. Let $a < 0 < b$. Let $T := \min\{t \geq 0 \colon X(t) \in \{a, b\}\}$. Let $\alpha := 2\,|\mu|\,/\sigma^2$. Show that

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

(If you use a martingale, you do not have to verify that it is bounded.)

Exercise 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 $a < 0$. Let $T_a := \min\{t \geq 0 \colon X(t) = a\}$. Let $\alpha := 2\,|\mu|\,/\sigma^2$. Show that

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

(If you use a martingale, you do not have to verify that it is bounded.)

Exercise 7.23 (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 7.24 (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)')
```

7.3. **Geometric Brownian Motion, Options, Black-Scholes.** Below, we let log denote the natural logarithm.

Exercise 7.25. Let $\mu \in \mathbb{R}$ and let $\sigma > 0$. Let X be a Gaussian random variable with mean μ and variance σ^2. Let $Y := e^X$. We then say Y has a **lognormal distribution with parameters** μ and σ^2. Show that Y has density

$$f(y) := \begin{cases} \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}} & \text{, if } y > 0 \\ 0 & \text{, if } y \leq 0. \end{cases}$$

Then, show that

$$\mathbf{E}Y = e^{\mu+\sigma^2/2}.$$
$$\mathbf{E}Y^2 = e^{2\mu+2\sigma^2}.$$

Recall that if $\{B(t)\}_{t\geq 0}$ is a standard Brownian motion and if $t > 0$ is fixed, then $B(t)$ is a mean zero Gaussian random variable. In particular, $B(t)$ has an equal chance of being above or below 0. For this reason, Brownian motion is perhaps not the best model for certain stocks or commodities. For example, stocks often go up or down by an amount proportional to their value. To better model this situation, we can instead model a stock price by $\{e^{B(t)}\}_{t\geq 0}$ where $\{B(t)\}_{t\geq 0}$ is a standard Brownian motion. More generally, we can also incorporate a drift.

Definition 7.26 (**Geometric Brownian Motion**). 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 $\mu \in \mathbb{R}$. Let $S_0 > 0$. We then define **geometric Brownian motion** with parameters $\sigma > 0$ and $\mu \in \mathbb{R}$ to be a stochastic process of the form $\{S(t)\}_{t\geq 0} = \{S_0 e^{X(t)}\}$.

Definition 7.27 (**European Call Option**). Let $\{S(t)\}_{t\geq 0}$ be a geometric Brownian motion with parameters $\sigma > 0$ and $\mu \in \mathbb{R}$. Let t, k be positive real numbers. In a **European call option**, we model a stock price as a geometric Brownian motion, and there is a payoff of

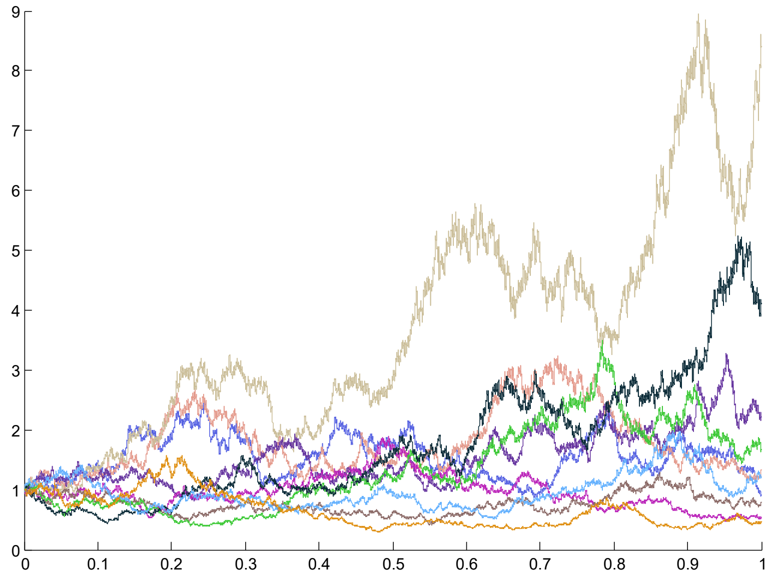


FIGURE 5. Sample Paths of Geometric Brownian Motion with $S_0 = 1$, $\sigma = 1$ and $\mu = 0$. The horizontal axis is the t-axis.

$\max(S(t) - k, 0)$. That is, at some future time t, we have the option to purchase the stock for a **strike price** k. If the price of the stock goes below k, i.e. if $S(t) < k$ we do not buy the stock (so the payoff is 0). And if the stock price goes above k (so that $S(t) > k$), we buy the stock at the price k, so the payoff is $S(t) - k$.

If the option has positive value at time $t = 0$, the option is called **in the money**. If the option has no value at time $t = 0$, the option is called **out of the money**. If $S(0) = k$, the option is called **at the money**.

From Exercise 7.19 with $\lambda = 1$, and $r := \mu + \sigma^2/2$,

$$\{e^{-rt}S(t)\}_{t\geq 0}$$

is a martingale. So, at time t, it is sensible to value the European call option at the price

$$c := e^{-(\mu+\sigma^2/2)t}\mathbf{E}\max(S(t) - k, 0).$$

Below, if $d \in \mathbb{R}$, we define $\Phi(d_1) := \int_{-\infty}^{d_1} e^{-y^2/2}\frac{dy}{\sqrt{2\pi}}$. Note: $1 - \Phi(-d_1) = \Phi(d_1)$.

Theorem 7.28 (**Black-Scholes Option Pricing Formula**)**.** *Let $\{S(t)\}_{t\geq 0}$ be a geometric Brownian motion with parameters $\sigma > 0$ and $\mu \in \mathbb{R}$ which models the price of a stock. Fix $t, k > 0$. Define $r := \mu + \sigma^2/2$. The value of the European call option with expiration time t and strike price k is*

$$c = S_0\Phi(d_1) - e^{-rt}k\Phi(d_1 - \sigma\sqrt{t}),$$

where

$$d_1 := \frac{\log(S_0/k) + (r + \sigma^2/2)t}{\sigma\sqrt{t}}.$$

Proof. We compute the quantity

$$c = e^{-rt}\mathbf{E}\max(S(t) - k, 0)$$

Note that $X(t) = \sigma B(t) + \mu t$ is a Gaussian random variable with variance $\sigma^2 t$ and mean μt. That is, $X(t)$ has density

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

Therefore

$$\begin{aligned}
&e^{rt}c = \mathbf{E}\max(S(t)-k,0) = \int_{-\infty}^{\infty} \max(S_0e^x - k, 0)\frac{1}{\sigma\sqrt{2\pi t}} e^{-\frac{(x-\mu t)^2}{2t\sigma^2}}dx\\
&= \frac{1}{\sigma\sqrt{2\pi t}}\int_{\log(k/S_0)}^{\infty}(S_0e^x - k)e^{-\frac{(x-\mu t)^2}{2t\sigma^2}}dx\\
&= \frac{1}{\sigma\sqrt{2\pi t}}S_0\int_{\log(k/S_0)}^{\infty} e^x e^{-\frac{(x-\mu t)^2}{2t\sigma^2}}dx - \frac{k}{\sigma\sqrt{2\pi t}}\int_{\log(k/S_0)}^{\infty} e^{-\frac{(x-\mu t)^2}{2t\sigma^2}}dx\\
&= \frac{e^{\mu t}}{\sqrt{2\pi}}S_0\int_{\frac{\log(k/S_0)-\mu t}{\sigma\sqrt{t}}}^{\infty} e^{y\sigma\sqrt{t}}e^{-y^2/2}dy - \frac{k}{\sqrt{2\pi}}\int_{\frac{\log(k/S_0)-\mu t}{\sigma\sqrt{t}}}^{\infty} e^{-y^2/2}dy, \quad \text{substituting } y = \tfrac{(x-\mu t)}{\sigma\sqrt{t}}\\
&= \frac{e^{\mu t+\sigma^2 t/2}}{\sqrt{2\pi}}S_0\int_{\frac{\log(k/S_0)-\mu t}{\sigma\sqrt{t}}}^{\infty} e^{-(y-\sigma\sqrt{t})^2/2}dy - k\left[1-\Phi\left(\frac{\log(k/S_0)-\mu t}{\sigma\sqrt{t}}\right)\right]\\
&= \frac{e^{\mu t+\sigma^2 t/2}}{\sqrt{2\pi}}S_0\int_{\frac{\log(k/S_0)-\mu t}{\sigma\sqrt{t}}}^{\infty} e^{-(y-\sigma\sqrt{t})^2/2}dy - k\Phi(d_1-\sigma\sqrt{t}), \qquad \text{using } 1-\Phi(-d_1)=\Phi(d_1)\\
&= \frac{e^{rt}}{\sqrt{2\pi}}S_0\int_{\frac{\log(k/S_0)-\mu t}{\sigma\sqrt{t}}-\sigma\sqrt{t}}^{\infty} e^{-z^2/2}dz - k\Phi(d_1-\sigma\sqrt{t}), \quad \text{substituting } z = y-\sigma\sqrt{t}\\
&= e^{rt}S_0[1-\Phi(-d_1)] - k\Phi(d_1-\sigma\sqrt{t}) = e^{rt}S_0\Phi(d_1) - k\Phi(d_1-\sigma\sqrt{t}).
\end{aligned}$$

□

Remark 7.29. Since $e^{\sigma B(t)+\mu t}$ has a log-normal density, we could have also used the formula

$$e^{-rt}\mathbf{E}\max(S(t)-k,0) = e^{-rt}\int_0^{\infty}\max(S_0z-k,0)\frac{1}{\sigma\sqrt{2\pi t}}\frac{1}{z}e^{-\frac{(\log(z)-\mu t)^2}{2t\sigma^2}}dz.$$

Exercise 7.30 (Binomial Option Pricing Model). Let $u, d > 0$. Let $0 < p < 1$. Let $(X_1, X_2, \ldots)$ be independent random variables such that $\mathbf{P}(X_n = \log u) =: p$ and $\mathbf{P}(X_n = \log d) = 1-p$ $\forall\, n \geq 1$. Let X_0 be a fixed constant. Let $Y_n := X_0 + \cdots + X_n$, and let $S_n := e^{Y_n}$ $\forall\, n \geq 1$. In general, $S_0, S_1, \ldots$ will not be a martingale, but we can still compute $\mathbf{E}S_n$, by modifying $S_0, S_1, \ldots$ to be a martingale.

First, note that if $n \geq 1$, then Y_n has a binomial distribution, in the sense that

$$\mathbf{P}(Y_n = X_0 + i\log u + (n-i)\log d) = \binom{n}{i}p^i(1-p)^{n-i}, \qquad \forall\, 0 \leq i \leq n.$$

Now define

$$r := p(u-d) - 1 + d.$$

Here we chose r so that $p = \frac{1+r-d}{u-d}$. For any $n \geq 1$, define

$$M_n := (1+r)^{-n}S_n.$$

Show that $M_0, M_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$. Consequently,

$$(1+r)^{-n}\mathbf{E}S_n = \mathbf{E}S_0, \qquad \forall\, n \geq 0.$$

(This presentation might be a bit backwards from the financial perspective. Typically, r is a fixed interest rate, and then you choose p such that $p = \frac{1+r-d}{u-d}$. That is, you adjust how the random variables behave in order to get a martingale.)

Remark 7.31. In Exercise 7.30, we considered a discrete version of geometric Brownian motion as follows. Let $u, d > 0$. Let $0 < p < 1$. Let $(X_1, X_2, \ldots)$ be independent random variables such that $\mathbf{P}(X_n = \log u) = p$ and $\mathbf{P}(X_n = \log d) = 1 - p \;\forall\, n \geq 1$. Let X_0 be a fixed constant. Let $Y_n := X_0 + \cdots + X_n$, and let $S_n := e^{Y_n} \;\forall\, n \geq 1$. Let $r := p(u-d) - 1 + d$. Let $M_n := (1+r)^{-n} S_n$ for any $n \geq 1$. Then $M_0, M_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$. So, using this discrete version of geometric Brownian motion and Remark 4.10, we can similarly price a European call option at time n at the price

$$(1+r)^{-n}\mathbf{E}\max(S_n - k, 0) = (1+r)^{-n} \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \max(u^i d^{n-i} S_0 - k, 0).$$

Remark 7.32. Using the Black-Scholes theory to model a stock incurs the following unrealistic assumptions.

- **Infinite Divisibility**. The stock can be bought and sold in arbitrary non-integer amounts.
- **Short selling**. Market participants can borrow arbitrary amounts of stock at no interest for an arbitrary amount of time. In a short sale, you borrow the stock and instantly sell it to someone else, and you can then buy back the stock at any later time.
- **No storage costs**. Market participants can hold arbitrary amounts of stock at no cost for an arbitrary amount of time.

Remark 7.33 (Implied Volatility). Theorem 7.28 is often used in the following way. It is given that r is an interest rate, σ is the unknown **volatility**, and we then define $\mu := r - \sigma^2/2$. (In mathematical finance, volatility and standard deviation are nearly synonymous.) Then the only unknown quantity in Theorem 7.28 is σ. We then choose σ so that c is equal to the actual observed price of the European call option. The σ found in this way is referred to as **implied volatility**. (From Exercise 7.48, c is an increasing function of σ, so a unique solution exists.)

If the volatility σ of a fixed stock is known, and if Theorem 7.28 accurately models this stock price, then European call options based on this stock should use the same volatility, regardless of the strike price k. In practice, this is not true. In practice, it is observed that σ has a U-shaped graph, as a function of k. That is, σ does seem to depend on k. This graph of σ as a function of k is known as the **volatility smile**, and it is one way of demonstrating that Theorem 7.28 is not an accurate model of a stock price. Alternatively, a firm believer in the Black-Scholes theory could argue that the pricing of options with very low or high volatility is irrational, as demonstrated by the volatility smile.

Remark 7.34. It follows by property (i) of Brownian motion in Definition 7.1 that the sample paths of Geometric Brown motion are continuous, with probability 1. Since Theorem 7.28 models a stock price as Geometric Brownian motion, we are implicitly assuming that a stock price is continuous, i.e. it does not "jump." This assumption is unrealistic. It is possible to model stocks using jumps, but doing so is fairly complicated.

Definition 7.35 (**American Call Option**)**.** Let $\{S(t)\}_{t\geq 0}$ be a geometric Brownian motion with parameters $\sigma > 0$ and $\mu \in \mathbb{R}$. Let t_0, k be positive real numbers. In an **American call option**, we model a stock price as a geometric Brownian motion, and there is a payoff of $c_t = \max(S(t) - k, 0)$ if the stock is purchased at any time $0 \leq t \leq t_0$. That is, at any time $0 \leq t \leq t_0$, we have the option to purchase the stock for a **strike price** k.

Remark 7.36. For a stock that does not pay dividends, even though we can exercise the call option at any time $0 \leq t \leq t_0$, it is always optimal to choose $t = t_0$. That is, it is optimal to treat this option as a European call option, so the American call option has the same value as the European call option. To see this, note that it never makes sense to buy the stock at time t if $S(t) \leq k$, so we assume that $S(t) > k$. That is, suppose we purchase the stock at time $t < t_0$ for price k, and $S(t) > k$. But instead of exercising the option, we could have just waited until time t_0; in the case $S(t_0) > k$, we could have purchased the stock for the price k, so the profit $S(t_0) - k$ would be the same at time t_0, no matter when we purchased the stock. However, if $S(t_0) \leq k$, then it would have been better if we never exercised the option at all. So, in any case, it is better to exercise the option at time t_0.

(If $S(t) > k$ with $t < t_0$, you might be tempted to exercise the option at time t, since this seems to be a profit that may be lost in the future. However, if you genuinely believe the profit will be lost in the future, then instead of exercising the option at $t < t_0$, consider short selling the stock at time $t < t_0$. Doing so guarantees a profit of at least $S(t) - k$ at time t_0, and you will even increase your profit in the case that $S(t_0) < k$.)

The above argument no longer holds when the stock does pay dividends, since it that case it may be more sensible to e.g. buy the stock a day before it pays out a dividend, instead of waiting until time t_0 is reached.

Definition 7.37 (**European Put Option**)**.** Let $\{S(t)\}_{t\geq 0}$ be a geometric Brownian motion with parameters $\sigma > 0$ and $\mu \in \mathbb{R}$. Let t, k be positive real numbers. In a **European put option**, we model a stock price as a geometric Brownian motion, and there is a payoff of $p = \max(k - S(t), 0)$. That is, at some future time t, we have the option to *sell* the stock for a **strike price** k.

The value of the European Put Option can be computed from the value of the European Call Option, via the following formula.

Proposition 7.38 (**Put-Call Parity**)**.** *Let c be the price of a European call option for a fixed stock with strike price k, with an option to exercise it at time t. Let p be the price of a European put option with strike price k for the same stock, with an option to exercise it at time t. Let S_0 be the price of the stock at time 0. Suppose money can be borrowed at a continuously-compounded nominal interest rate $r \geq 0$ (i.e. the rate of interest before adjusting for inflation). Then, assuming no arbitrage opportunity exists,*

$$S_0 + p - c = ke^{-rt}.$$

Proof. Assume that $S_0 + p - c < ke^{-rt}$. We will demonstrate an arbitrage opportunity. At the present time, buy one share of stock, buy one put option, and sell one call option. We then initially borrow $S_0 + p - c$ and pay this amount to complete the purchase. We now break into two cases according to the price $S(t)$ of the stock at time t.

Case 1. $S(t) \leq k$. Then the call option has no value, so it will not be exercised, and we exercise the put option to sell the stock we own for the price k.

Case 2. $S(t) > k$. Then the put option has no value, and the call option we sold will be exercised, so that we have to sell the stock we own for the price k.

In either case, we earned k at time t. Since $e^{rt}(S_0 + p - c) < k$, we can pay off the loan and earn a profit $k - e^{rt}(S_0 + p - c) > 0$.

Now, assume that $S_0 + p - c > ke^{-rt}$. We can demonstrate an arbitrage opportunity by reversing the above procedure. □

Exercise 7.39. In the context of Put-Call parity, show that an arbitrage opportunity exists if $S_0 + p - c > ke^{-rt}$. (That is, fill in the omitted details from the notes in this case.)

Exercise 7.40 (MFE Sample Question)**.** Consider a European call option and a European put option on a nondividend-paying stock. The following things are given

- The current price of the stock is 60.
- The call option currently sells for 0.15 more than the put option.
- Both the call option and put option will expire in 4 years.
- Both the call option and put option have a strike price of 70.

Calculate the continuously compounded risk-free interest rate. (That is, compute the interest rate r that ensures that no arbitrage opportunity exists.)

Exercise 7.41 (MFE Sample Question)**.** Near market closing time on a given day, you lose access to stock prices, but some European call and put prices for a stock are available as follows:

Strike Price	Call Price	Put Price
\$40	\$11	\$3
\$50	\$6	\$8
\$55	\$3	\$11

All six options have the same expiration date.

After reviewing the information above, John tells Mary and Peter that no arbitrage opportunities can arise from these prices.

Mary disagrees with John. She argues that one could use the following portfolio to obtain arbitrage profit: Long one call option with strike price 40; short three call options with strike price 50; lend \$1; and long some calls with strike price 55. Peter also disagrees with John. He claims that the following portfolio, which is different from Mary's, can produce arbitrage profit: Long 2 calls and short 2 puts with strike price 55; long 1 call and short 1 put with strike price 40; lend \$2; and short some calls and long the same number of puts with strike price 50.

Which of the following statements is true?

(A) Only John is correct.
(B) Only Mary is correct.
(C) Only Peter is correct.
(D) Both Mary and Peter are correct.
(E) None of them is correct.

Definition 7.42 (**American Put Option**)**.** Let $\{S(t)\}_{t \geq 0}$ be a geometric Brownian motion with parameters $\sigma > 0$ and $\mu \in \mathbb{R}$. Let t_0, k be positive real numbers. In an **American put option**, we model a stock price as a geometric Brownian motion, and there is a payoff of $p_t = \max(k - S(t), 0)$ if the stock is sold at any time $0 \leq t \leq t_0$. That is, at any time $0 \leq t \leq t_0$, we have the option to sell the stock for a **strike price** k.

If $0 < t < t_0$ and if $S(t) < k$, then it can make sense to exercise the American put option at time t. In this case, your profit $k - S(t)$ would then earn interest at a rate $r > 0$ until time t_0. And this profit could be more than the profit obtained by waiting to exercise until time t_0. Also, the profit $(k - S(t))e^{r(t_0 - t)}$ at time t_0 could be more than the profit obtained by buying the stock at time t and waiting to exercise the option until time t_0.

Recall that, for an American call option, the "reverse" of this argument does not apply. If $S(t) > k$, and if you believe the stock has reached a high value at time $t < t_0$, then rather than exercising the American call option early, you should short sell the stock at time t. Short selling the stock at time t means you borrow the stock for zero interest, and you instantly sell it, so you earn the price $S(t)$ of the short sold stock at time t upon completion of the short sale.

Since it can make sense to exercise the American put option early, the Put-call parity as stated in Proposition 7.38 no longer holds.

Exercise 7.43. There are many ways of buying and selling American put and call options on the same underlying asset, in order to make profits while minimizing risk. These strategies are known as **spreads**. (Every put and call option below will be an American option.) Describe the pros and cons of creating each spread specified below.

- In the **collar** spread, you own a stock which has a variable price s, you buy a put option for that same stock with strick price k_1, and you short a call option with strike price k_2, where $k_1 < k_2$. So, the revenue you will make by exercising all of these options (and selling the stock) is
$$s + \max(k_1 - s, 0) - \max(s - k_2, 0).$$
Plot this function as a function of s. The **zero-cost collar** occurs when k_1 is equal to the current price of the stock.
- In the **straddle** spread, you buy a call and a put option for the same stock and with the same strike price k. So, the revenue you can make by exercising both options simultaneously is
$$\max(k - s, 0) + \max(s - k, 0).$$
Plot this function as a function of s.
- In the **strangle** spread, you buy a call option with strike price k_1, and you buy a put option with strike price k_2, where $k_1 < k_2$. Plot your revenue from exercising both options simultaneously, as a function of s, the price of the underlying asset.
- Let $c > 0$. In the **butterfly** spread, you buy a call option with strike price k, you short two call options with strike price $k + c$, and you buy a call option with strike price $k + 2c$. Plot your revenue from exercising these options simultaneously, as a function of s, the price of the underlying asset.

Exercise 7.44. There are many ways to try to value an American Put Option. One way is to emulate the formula for a European Put Option which is exercised at time $0 \leq t \leq t_0$:
$$e^{-(\mu + \sigma^2/2)t}\mathbf{E}\max(k - S(t), 0)$$
We would like to simply take the maximum of the above quantity over all $t \in [0, t_0]$. However, this would be equivalent to knowing the future price of the stock at all times, which is

unrealistic. So, we instead consider replacing the variable t by a stopping time. Suppose T is a stopping time. That is, $T(t) \geq 0$ is only allowed to depend on values of $S(t')$ where $t' < t$. Then we could try to maximize the quantity

$$\mathbf{E}e^{-(\mu+\sigma^2/2)T}\max(k - S(T), 0)$$

over all stopping times T where $0 \leq T \leq t_0$. To approximate that quantity, let $0 \leq t_1 \leq t_0$ and just consider stopping times T of the form $T = \min\{t_1 \leq t \leq t_0 \colon S(t) < S(t')\,\forall\, 0 \leq t' \leq (3/4)t_1\}$, or $T = t_0$ if the set of t inside the minimum is empty. Then, using a computer, compute the maximum over all $0 \leq t_1 \leq t_0$ of

$$\mathbf{E}e^{-(\mu+\sigma^2/2)T}\max(k - S(T), 0)$$

when $\mu = 0$, $= \sigma = t_0 = S_0 = 1$ and $k = 2$.

This procedure is analogous to the solution of the Secretary Problem.

In order to compute the expected value, use a Monte Carlo simulation of Brownian motion, and take the average value over many runs of the simulation.

Exercise 7.45. In each of the following examples, choose a few parameters (e.g. use $\mu = 0$, $\sigma = S_0 = t = 1$ and $k = 2$.), and value the option using several runs of a Monte Carlo simulation of Brownian motion. In each case, we multiply by an exponential term in order to emulate the Black-Scholes formula.

(i) (**Asian Call Option**) The value of an Asian option with strike price $k > 0$ at time $t > 0$ is computed using the average value of the stock from time 0 to time t. That is, if the option is exercised at time $t > 0$, then its value is

$$e^{-(\mu+\sigma^2/2)t}\mathbf{E}\max\left((\frac{1}{t}\int_0^t S(r)dr) - k, 0\right).$$

(ii) (**Lookback Call Option**) The value of a lookback call option with strike price $k > 0$ at time $t > 0$ is computed using the maximum value of the stock between time 0 and time t. That is, if the option is exercised at time $t > 0$, then its value is

$$e^{-(\mu+\sigma^2/2)t}\mathbf{E}\max\left(\max_{0\leq r\leq t} S(r) - k, 0\right).$$

In other words, you can "look back" over the past behavior of the stock, and choose the best price possible over the past.

(iii) (**Lookback Put Option**) The value of a lookback put option with strike price $k > 0$ at time $t > 0$ is computed using the minimum value of the stock between time 0 and time t. That is, if the option is exercised at time $t > 0$, then its value is

$$e^{-(\mu+\sigma^2/2)t}\mathbf{E}\max\left(k - \min_{0\leq r\leq t} S(r), 0\right).$$

Finally, using Corollary 7.15, give an exact formula for the value of the Lookback Call Option. (And check that this formula agrees with the results of your simulation.)

Can you also give an explicit formula for the value of the Lookback Put Option?

7.4. **Black-Scholes Statistics.** Recall that in Theorem 7.28, we modeled a stock price in the following way. Let $\{S(t)\}_{t\geq 0}$ be a geometric Brownian motion with parameters $\sigma > 0$ and $\mu \in \mathbb{R}$. Fix $t, k > 0$. Define $r := \mu + \sigma^2/2$. The value of the European call option with expiration time t and strike price k is

$$c = S_0\Phi(d_1) - e^{-rt}k\Phi(d_1 - \sigma\sqrt{t}),$$

where $\Phi(d_1) := \int_{-\infty}^{d_1} e^{-y^2/2}dy/\sqrt{2\pi}$, and

$$d_1 := \frac{\log(S_0/k) + (r + \sigma^2/2)t}{\sigma\sqrt{t}}.$$

In this section, we consider c as defined above to be a function of its input parameters, so that

$$c = c(S_0, t, k, \sigma, r).$$

Several statistics of the stock are studied by taking derivatives of c with respect to its input parameters. That is, these statistics measure how the price changes as the underlying parameters change.

Definition 7.46 (**Greeks**)**.**

- $\Delta := \partial c/\partial S_0$.
- $\Gamma := \partial^2 c/\partial^2 S_0$.
- $\Theta := -\partial c/\partial t$.
- $\nu := \partial c/\partial\sigma$. (This quantity is called vega, but it is denoted by the Greek letter nu.)
- $\rho := \partial c/\partial r$.
- $\lambda := \Delta \cdot \frac{S_0}{c} = \frac{\partial c}{\partial S_0} \cdot \frac{S_0}{c} = S_0 \cdot \frac{\partial}{\partial S_0}\log c$. (This quantity is called the **elasticity**.)

Exercise 7.47. Let Z be a standard normal random variable. Recall that we can express a geometric Brownian motion as

$$S(t) = S_0 e^{\sigma\sqrt{t}Z + (r-\sigma^2/2)t}, \qquad t > 0.$$

Show that

$$e^{-rt}\mathbf{E}[S(t) \cdot Z \cdot 1_{\{S(t)>k\}}] = S_0(\Phi'(d_1) + \sigma\sqrt{t}\Phi(d_1)).$$
$$e^{-rt}\mathbf{E}[S(t) \cdot 1_{\{S(t)>k\}}] = S_0\Phi(d_1).$$

Exercise 7.48. Show the following (using the notation from the Black-Scholes Formula)

- $\Delta = \Phi(d_1)$.
- $\rho = kte^{-rt}\Phi(d_1 - \sigma\sqrt{t})$.
- $\nu = S_0\sqrt{t}\Phi'(d_1)$.
- $-\Theta = \frac{\sigma}{2\sqrt{t}}S_0\Phi'(d_1) + kre^{-rt}\Phi(d_1 - \sigma\sqrt{t})$.

(Hint: use Exercise 7.47.) (To make these exercises easier, write $c = \mathbf{E}(e^{-rt}\max(S(t) - k, 0))$, use the $S(t)$ formula from Exercise 7.47, and pretend that you can apply the chain rule to the max function, so that $(d/dx)\max(x, 0) = 1_{\{x>0\}}$ for any $x \in \mathbb{R}$, even though technically the max function is not differentiable at 0.)

Exercise 7.49 (MFE Sample Question)**.** You are considering the purchase of a 3-month 41.5-strike American call option on a nondividend-paying stock.

You are given:

(i) The Black-Scholes framework holds.

(ii) The stock is currently selling for 40.
(iii) The stock's volatility is 30%.
(iv) The current call option delta is 0.5.

Determine the current price of the option.

(A) $20 - 20.453\int_{-\infty}^{.15} e^{-x^2/2}dx$.
(B) $20 - 16.138\int_{-\infty}^{.15} e^{-x^2/2}dx$.
(C) $20 - 40.453\int_{-\infty}^{.15} e^{-x^2/2}dx$.
(D) $-20.453 + 16.138\int_{-\infty}^{.15} e^{-x^2/2}dx$.
(E) $-20.453 + 40.453\int_{-\infty}^{.15} e^{-x^2/2}dx$.

Remark 7.50. Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is a piecewise linear function that is zero outside of $[0,1]$. Let $y_1 < \cdots < y_{n-1}$. Suppose $f(i/n) = y_i$ for every $1 \leq i \leq n-1$, and that f is linear between the points $0, 1/n, 2/n, \ldots, 1$. Then it is theoretically possible to buy and short call options so that, if the stock has a price $0 \leq s \leq 1$, then your revenue from exercising all of the options is $f(s)$. To see this, recall that for any $1 \leq i \leq n-1$, a butterfly spread can be constructed so that the revenue $r_i(s)$ from exercising it as a function of s is a piecewise linear function of s which is zero outside of $[(i-1)/n, (i+1)/n]$ and which takes the value 1 at i/n. So, consider the portfolio obtained by constructing the i^{th} such butterfly spread in the amount y_i, for every $1 \leq i \leq n-1$. Then the revenue from exercising all of the options (as a function of s) is

$$r(s) := \sum_{i=1}^{n-1} y_i r_i(s).$$

Then r is linear between the points $0, 1/n, 2/n, \ldots, 1$, r is zero outside of $[0,1]$, and $r(j/n) = y_j$ for every $1 \leq j \leq n-1$. Therefore, $r = f$.

Exercise 7.51 (Put-Call Parity for American Options)**.** As we mentioned above, Put-call parity does not hold for American Options, as an equality. However, we can still obtain upper and lower bounds on the difference of the American put and call option, as stated below.

Let c be the price of an American call option for a fixed stock with strike price k, with an option to exercise it at any time $0 \leq t \leq t_0$. Let p be the price of an American put option with strike price k for the same stock, with an option to exercise it at any time $0 \leq t \leq t_0$. Let S_0 be the price of the stock at time 0. Suppose money can be borrowed at a continuously-compounded nominal interest rate $r \geq 0$ (i.e. the rate of interest before adjusting for inflation). Then, assuming no arbitrage opportunity exists,

$$S_0 - k \leq c - p \leq S_0 - ke^{-rt_0}.$$

(Hint: first, show that $p \geq c - S_0 + ke^{-rt_0}$, since p is larger or equal to the value of a European put option, and then apply the Put-Call parity for European options. Then, show that $c \geq p + S_0 - k$ in the following way. Consider the portfolio of buying one call, shorting one put, shorting the stock and borrowing k dollars. If all of the options are exercised at any time $0 \leq t \leq t_0$, show that you obtain a nonnegative profit. That is, the value of this portfolio at time 0 is nonnegative.)

Exercise 7.52. In the discrete binomial model, we can find a price for an American put option using dynamic programming.

Recall this model. Let $u, d > 0$. Let $0 < p < 1$. Let $(X_1, X_2, \ldots)$ be independent random variables such that $\mathbf{P}(X_n = \log u) =: p$ and $\mathbf{P}(X_n = \log d) = 1 - p \;\forall\, n \geq 1$. Let X_0 be a fixed constant. Let $Y_n := X_0 + \cdots + X_n$, and let $S_n := e^{Y_n} \;\forall\, n \geq 1$. Let $r := p(u-d) - 1 + d$. For any $n \geq 1$, define $M_n := (1+r)^{-n} S_n$. Recall that $M_0, M_1, \ldots$ is a martingale.

Note that, at time n, the random variable S_n has $n+1$ possible values. Label these values as $S_{n,1} \leq \cdots \leq S_{n,m}$. Let $k > 0$. Let $V_{n,m}$ be the value of the American put option at time $n > 0$ with strike price k, when S_n has its m^{th} value. Then

$$V_{n,m} = \max\Big(\max(k - S_{n,m}, 0)\,,\ (1+r)^{-1}(pV_{n+1,m+1} + (1-p)V_{n+1,m})\Big), \qquad \forall\, 1 \leq m \leq n+1.$$

This recursion formula holds since, at step n, you can either exercise the option at time n, or you can wait and see what happens at time $n+1$. The quantity $\max(k - S_{n,m}, 0)$ is your revenue from exercising at time n, and the second quantity $(1+r)^{-1}(pV_{n+1,m+1} + (1-p)V_{n+1,m})$ is your expected revenue from waiting until time $n+1$ to exercise the option. So, at time n, you choose the maximum of these two quantities.

Let's solve this recursion in the following example. Suppose $S_0 = 8$, $p = 1/2$, $u = 2$, $d = 1/2$ (so that $r = 1/4$), and $k = 10$. And suppose the option expires at time $n = 3$ (so that $V_{3,m} = \max(k - S_{3,m}, 0)$ is known for each $1 \leq m \leq 4$.) Then, working backwards, eventually find $V_{0,1}$, the price of the option.

Compare your result in this example with the price of the European put option with the same parameters. (It should be smaller.)

8. Stochastic Integration, Itô's Formula

Let $-\infty < a < b < \infty$. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Recall that a continuous function f is Riemann integrable on $[a, b]$. That is, there is a real number, denoted by $\int_a^b f(x)dx$ such that

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(a + (b-a)i/n)\frac{b-a}{n} = \int_a^b f(x)dx.$$

Note that we are using the "left endpoint" Riemann sum. We will continue to do so below.

Since the sample paths of Brownian motion are continuous with probability 1, we can also integrate a standard Brownian motion $\{B(t)\}_{t\geq 0}$ using a Riemann integral. For any $n \geq 1$, consider the Riemann sum

$$X_n := \sum_{i=0}^{n-1} B(a + (b-a)i/n)\frac{b-a}{n}.$$

We would like to say that this quantity converges in some sense as $n \to \infty$. However, since this Riemann sum no longer has a meaning as a real number, we need to change the meaning of the limit as $n \to \infty$.

Definition 8.1 (**Convergence in Probability**)**.** Let $X_1, X_2, \ldots$ be random variables, and let X be a random variable. We say that $X_1, X_2, \ldots$ **converges in probability** to X if: for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}(|X_n - X| > \varepsilon) = 0.$$

Exercise 8.2. Let $\mathbf{P}$ be the uniform probability law on $[0,1]$. Let $X(t) = 0$ for any $t \in [0,1]$. For any $n \geq 1$, define $X_n(t) = n \cdot 1_{\{0 \leq t < 1/n\}}$. Show that $X_1, X_2, \ldots$ converges in probability to X. However, $\mathbf{E}X = 0$ whereas $\mathbf{E}X_n = 1$ for all $n \geq 1$. So, convergence in probability does not imply that expected values converge.

Also, note that $X_n(0)$ does not converge to $X(0)$ as $n \to \infty$. So, convergence in probability does not imply pointwise convergence.

Exercise 8.3 (Uniqueness of the Limit). Suppose $X_1, X_2, \ldots$ converges in probability to X. Also, suppose $X_1, X_2, \ldots$ converges in probability to Y. Show that $\mathbf{P}(X \neq Y) = 0$.

Example 8.4. Returning to the above example, there exists a random variable, which we denote by $\int_a^b B(t)dt$ such that

$$X_n := \sum_{i=0}^{n-1} B(a + (b-a)i/n)\frac{b-a}{n}$$

converges to $\int_a^b B(t)dt$ in probability as $n \to \infty$.

For example, choosing $a = 0$, $\sum_{i=0}^{n-1} B(bi/n)\frac{b}{n}$ converges to $\int_0^b B(t)dt$ in probability as $n \to \infty$. To compute the variance of the Riemann sum, we first rearrange the sum and then use a telescoping sum to get

$$\begin{aligned}
\sum_{i=0}^{n-1} B(bi/n)\frac{b}{n} &= \sum_{i=0}^{n-1} B(bi/n)\left(\frac{b(i+1)}{n} - \frac{bi}{n}\right) = \sum_{i=0}^{n-1} B(bi/n)\frac{b(i+1)}{n} - \sum_{i=0}^{n-1} B(bi/n)\frac{bi}{n} \\
&= \sum_{i=1}^{n} B(b(i-1)/n)\frac{bi}{n} - \sum_{i=0}^{n-1} B(bi/n)\frac{bi}{n} \\
&= \sum_{i=1}^{n-1} \big(B(b(i-1)/n) - B(bi/n)\big)(bi/n) + bB(b(n-1)/n) \\
&= \sum_{i=1}^{n-1} \big(B(b(i-1)/n) - B(bi/n)\big)\left(\frac{bi}{n} - b\right).
\end{aligned}$$

From the independent and stationary increment properties of standard Brownian motion, $\sum_{i=1}^{n} B(bi/n)\frac{b}{n}$ is then a Gaussian random variable with mean zero, and variance

$$\mathbf{E}\left(\sum_{i=0}^{n-1} B(bi/n)\frac{b}{n}\right)^2 = \sum_{i=1}^{n-1} \mathbf{E}(B(b(i-1)/n) - B(bi/n))^2\left(\frac{bi}{n} - b\right)^2 = \sum_{i=1}^{n-1}(b/n)\left(\frac{bi}{n} - b\right)^2.$$

Letting $n \to \infty$, this becomes

$$\int_0^b (s-b)^2 ds = -\frac{1}{3}(s-b)^3|_{s=0}^{s=b} = \frac{1}{3}b^3.$$

So, we anticipate that $\int_0^b B(t)dt$ is a Gaussian random variable with mean zero and variance $(1/3)b^3$. And indeed, we can compute the variance as follows

$$\begin{aligned}\mathbf{E}(\int_0^b B(t)dt)^2 &= \mathbf{E}\int_0^b B(t)dt \int_0^b B(s)ds = \int_0^b \int_0^b \mathbf{E}B(t)B(s)dsdt \\ &= \int_0^b \int_0^b \min(s,t)dsdt \qquad \text{, by Proposition 7.8} \\ &= 2\int_{t=0}^{t=b}\int_{s=0}^{s=t} sdsdt = \int_{t=0}^{t=b} t^2dt = \frac{1}{3}b^3.\end{aligned}$$

Example 8.5. Similarly, if $-\infty < a < b < \infty$, and if $f\colon \mathbb{R} \to \mathbb{R}$ is continuous, we define

$$\int_a^b f(B(t))dt$$

to be the random variable such that, as $n \to \infty$, $\sum_{i=0}^{n-1} f(B(a + (b-a)i/n))\frac{b-a}{n}$ converges in probability to $\int_a^b f(B(t))dt$. We can think of $\int_a^b f(B(t))dt$ as the area under the random curve $f(B(t))$.

Exercise 8.6. Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Let $f\colon \mathbb{R} \to \mathbb{R}$. Assume that $\int_{\mathbb{R}} |f(x)|\, dx < \infty$ and $\int_{\mathbb{R}} f(x)dx = 1$. For any $s > 0$, define

$$X(s) := \frac{1}{\sqrt{s}}\int_0^s f(B(t))dt.$$

Show that $\lim_{s\to\infty} \mathbf{E}X(s) = \sqrt{2/\pi}$. For an optional challenge, show that $\lim_{s\to\infty} \mathbf{E}(X(s))^2 = 1$. (Hint: for the second part, look up the formula for a multivariate normal random variable.)

The Stochastic integral is a slightly different object, where instead of integrating against the "infinitesimal width" of a rectangle, we integrating against the "infinitesimal increment" of a Brownian motion.

Definition 8.7 (**Stochastic Integral**)**.** Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous. Let $b > 0$. For any $n \geq 1$, consider the Riemann sum on $[0, b]$:

$$X_n := \sum_{i=0}^{n-1} f\left(B\left(\frac{bi}{n}\right)\right)\left(B\left(\frac{b(i+1)}{n}\right) - B\left(\frac{bi}{n}\right)\right).$$

We define the **stochastic integral** of f on $[0, b]$ with respect to Brownian motion, denoted

$$\int_0^b f(B(s))dB(s).$$

to be the random variable X such that X_n converges to X in probability, as $n \to \infty$.

More generally, if $f\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous, we define

$$\int_0^b f(s, B(s))dB(s)$$

as the limit as $n \to \infty$ of the following Riemann sums (in the sense of convergence in probability):

$$\sum_{i=0}^{n-1} f\left(\frac{bi}{n}, B\left(\frac{bi}{n}\right)\right)\left(B\left(\frac{b(i+1)}{n}\right) - B\left(\frac{bi}{n}\right)\right)$$

So, the stochastic integral is itself a random variable, unlike the Riemann integral of a real-valued function, which is a fixed number. We can think of $\int_0^b f(B(s))dB(s)$ as the randomly measured area under the random curve $f(B(s))$.

Also, if $W_m := \sum_{i=0}^{m} f\left(\frac{bi}{n}, B\left(\frac{bi}{n}\right)\right)\left(B\left(\frac{b(i+1)}{n}\right) - B\left(\frac{bi}{n}\right)\right)$ for any $m \geq 1$, then $W_1, W_2, \ldots$ is a martingale by Theorem 4.14. So $\int_0^b f(s, B(s))dB(s)$ should be a martingale as well. (We can think of the integrand as some function of the stock price, such as a stock trading strategy, and we multiplying the integrand by the change in the stock price.)

Okay, we now have a stochastic integral, so we should discuss how to manipulate this integral. In real variable calculus, the most important way to compute integrals is via the Fundamental Theorem of calculus. Recall that if $f\colon \mathbb{R} \to \mathbb{R}$ has one continuous derivative, and if $b > 0$, then the Fundamental Theorem of Calculus says

$$f(b) - f(0) = \int_0^b f'(s)ds.$$

So, if $g\colon \mathbb{R} \to \mathbb{R}$ is another function with one continuous derivative, then the chain rule implies

$$f(g(b)) - f(g(0)) = \int_0^b \frac{d}{ds}(f(g(s)))ds = \int_0^b f'(g(s))g'(s)ds.$$

Or, using the notation $dg(s) := g'(s)ds$, we have

$$f(g(b)) - f(g(0)) = \int_0^b f'(g(s))dg(s).$$

This equality *almost* holds if we replace $g(s)$ by a standard Brownian motion $B(s)$, but we need to add an additional term on the right side, for reasons we will explain below.

Exercise 8.8. Let $t > 0$ and let $\{B(s)\}_{s\geq 0}$ be a standard Brownian motion. Compute the mean and variance of

$$\int_0^t B(s)dB(s).$$

(Hint: start with the Riemann sum, then take a limit.)

Exercise 8.9. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function. Let $t > 0$ and let $\{B(s)\}_{s\geq 0}$ be a standard Brownian motion. Find the distribution of

$$\int_0^t f(s)dB(s).$$

That is, find the CDF of $\int_0^t f(s)dB(s)$. (Hint: use Exercise 7.7.)

Theorem 8.10 (Itô's Formula). *Let $f\colon \mathbb{R} \to \mathbb{R}$ have two continuous derivatives. Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Then, with probability 1, for all $b \geq 0$,*

$$f(B(b)) - f(B(0)) = \int_0^b f'(B(s))dB(s) + \frac{1}{2}\int_0^b f''(B(s))ds.$$

Remark 8.11. Choosing $f(x) = x$ for all $x \in \mathbb{R}$ shows that

$$B(b) = B(b) - B(0) = \int_0^b dB(s).$$

Choosing $f(x) = x^2$ for all $x \in \mathbb{R}$ shows that $(B(b))^2 = 2\int_0^b B(s)dB(s) + \int_0^b ds$, so that

$$\int_0^b B(s)dB(s) = \frac{1}{2}(B(b))^2 - \frac{1}{2}b.$$

Remark 8.12. Theorem 8.10 can be stated in the equivalent **differential form** as

$$df(B(s)) = f'(B(s))dB(s) + \frac{1}{2}f''(B(s))ds.$$

We can almost interpret this expression as a chain rule, except that the second derivative has no analogue in real variable calculus.

Exercise 8.13. Using Itô's formula, write an expression for $\int_0^1 (B(s))^2 dB(s)$.

Exercise 8.14. Let $b > 0$. We know from calculus that $\int_0^b e^s ds = e^b - 1$.

Use $f(x) = e^x$, $x \in \mathbb{R}$, in Itô's formula to find a similar expression for $\int_0^b e^{B(s)} dB(s)$. (Note that $e^{B(s)}$ is a Geometric Brownian motion, so now we know how to take the stochastic integral of Geometric Brownian motion.)

Exercise 8.15 (MFE Sample Question, from an old exam)**.** Let $\{Z(t)\}_{t\geq 0}$ be a standard Brownian motion. You are given:

(i) $U(t) := 2Z(t) - 2$, for all $t \geq 0$.
(ii) $V(t) := (Z(t))^2 - t$, for all $t \geq 0$.
(iii) $W(t) := t^2 Z(t) - 2\int_0^t sZ(s)ds$, for all $t \geq 0$.

Which of the processes defined above has/have zero drift? (A stochastic process $\{U(t)\}_{t\geq 0}$ has zero drift if $dU(t) = f(Z(t), t)dZ(t)$ for some function $f\colon \mathbb{R}^2 \to \mathbb{R}$.)

Recall that if $t > s > 0$, then $\mathbf{E}(B(t) - B(s))^2 = t - s$. So, intuitively, $(B(t) - B(s))^2$ behaves like $(t - s)$. More specifically, we have the following lemma, which we will not prove.

Lemma 8.16. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous. Let $b > 0$. Consider the following sum (which is not quite a Riemann sum, since the increment is squared):*

$$X_n := \sum_{i=0}^{n-1} f\left(B\left(\frac{bi}{n}\right)\right)\left(B\left(\frac{b(i+1)}{n}\right) - B\left(\frac{bi}{n}\right)\right)^2.$$

Then, as $n \to \infty$, X_n converges in probability to

$$\int_0^b f(B(s))ds.$$

Proof Sketch of Theorem 8.10. From Taylor's Theorem, if $x, y \in \mathbb{R}$, then there exists an error term $R(x, y)$ such that

$$f(x) - f(y) = f'(y)(x - y) + \frac{1}{2}f''(y)(x - y)^2 + R(x, y).$$

For any $0 \le i \le n-1$, let $x := B(b(i+1)/n)$, let $y := B(bi/n)$, and sum over i to get

$$f(B(b)) - f(B(0)) = \sum_{i=0}^{n-1} (f(B(b(i+1)/n)) - f(B(bi/n)))$$

$$= \sum_{i=1}^{n-1} f'(B(bi/n))\big(B(b(i+1)/n) - B(bi/n)\big)$$

$$+ \frac{1}{2}\sum_{i=1}^{n-1} f''(B(bi/n))\big(B(b(i+1)/n) - B(bi/n)\big)^2 + \sum_{i=1}^{n-1} R(B(b(i+1)/n), B(bi/n)).$$

We now let $n \to \infty$. The first term converges in probability to $\int_0^b f'(B(s))dB(s)$ by the definition of the stochastic integral. The second term converges in probability to $\frac{1}{2}\int_0^b f''(B(s))ds$ by Lemma 8.16. Treating the last term as an error term concludes the proof. □

There is also a version of Itô's formula for a function f both of time and of the Brownian motion.

Theorem 8.17 (**Itô's Formula, Version 2**). *Let $f: \mathbb{R}^2 \to \mathbb{R}$ have two continuous derivatives in each coordinate. We write $f = f(x, y)$, $x, y \in \mathbb{R}$. Then, with probability 1, for all $b \geq 0$,*

$$f(b, B(b)) - f(0, B(0)) = \int_0^b \frac{\partial f}{\partial x}(s, B(s))ds + \int_0^b \frac{\partial f}{\partial y}(s, B(s))dB(s) + \frac{1}{2}\int_0^b \frac{\partial^2 f}{\partial y^2}(s, B(s))ds.$$

Remark 8.18. Theorem 8.17 can be stated in the equivalent **differential form** as

$$df(s, B(s)) = \frac{\partial f}{\partial y}(s, B(s))dB(s) + \left(\frac{\partial f}{\partial x}(s, B(s)) + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(s, B(s))\right) ds.$$

Exercise 8.19. Let $f: \mathbb{R} \times [0, \infty) \to \mathbb{R}$. We write $f = f(x, t)$, where $(x, t) \in \mathbb{R} \times [0, \infty)$. Let $g: \mathbb{R} \to \mathbb{R}$ be a continuous and bounded function with $\int_{\mathbb{R}} |g(x)|\, dx < \infty$. We say that f satisfies the one-dimensional **heat equation** if

$$\frac{\partial f}{\partial t}(x, t) = \frac{\partial^2 f}{\partial x^2}(x, t), \quad \forall\, (x, t) \in \mathbb{R} \times [0, \infty),$$
$$f(x, 0) = g(x), \qquad \forall\, x \in \mathbb{R}.$$

Show that f defined by

$$f(x, t) := \frac{1}{\sqrt{4\pi t}} \int_{\mathbb{R}} e^{-\frac{(x-y)^2}{4t}} g(y)dy = \mathbf{E}(g(B(2t) + x)), \qquad \forall\, (x, t) \in \mathbb{R} \times [0, \infty),$$

satisfies the heat equation. (Just check the first condition. You do not have to show that $\lim_{t \to 0^+} f(x, t) = g(x)$ for all $x \in \mathbb{R}$.)

Using a computer, plot the function $f(x, t)$ as a function of x for several different values of $t > 0$, using $g = 1_{[0,1]}$. Lastly, verify that $\int_{\mathbb{R}} \frac{1}{\sqrt{4\pi t}} e^{-x^2/(4t)} dx = 1$ for any $t > 0$.

Exercise 8.20. Let $f: \mathbb{R} \times [0, \infty) \to \mathbb{R}$. We write $f = f(x, t)$, where $(x, t) \in \mathbb{R} \times [0, \infty)$. Let $g: \mathbb{R} \to \mathbb{R}$ be a continuous and bounded function. We say that f satisfies the one-dimensional **heat equation with forcing term** $h: \mathbb{R} \times [0, \infty) \to \mathbb{R}$ if

$$\frac{\partial f}{\partial t}(x, t) = \frac{\partial^2 f}{\partial x^2}(x, t) + h(x, t), \quad \forall\, (x, t) \in \mathbb{R} \times [0, \infty),$$

$$f(x,0)=g(x), \qquad \forall\, x\in\mathbb{R}.$$

For any $(x,t)\in\mathbb{R}\times[0,\infty)$, define $f(x,t)$ so that

$$f(x,t):=\frac{1}{\sqrt{4\pi t}}\int_{\mathbb{R}}e^{-\frac{(x-y)^2}{4t}}g(y)dy+\int_0^t\frac{1}{\sqrt{4\pi(t-s)}}\int_{\mathbb{R}}e^{-\frac{(x-y)^2}{4(t-s)}}h(y,s)dyds.$$

Show that f satisfies the heat equation with forcing term h. (Just check the first condition.)

Exercise 8.21. Let $t_0>0$. Let $V\colon\mathbb{R}\times[0,t_0]\to\mathbb{R}$. We write $V=V(s,t)$, $s\in\mathbb{R}$, $t\in[0,t_0]$. Let $F\colon\mathbb{R}\to\mathbb{R}$. Let $r\in\mathbb{R}$, let $\sigma>0$. We say that V satisfies the **Black-Scholes** equation if $V(s,t_0)=F(s)$ for all $s\in\mathbb{R}$, and if

$$\frac{\partial V}{\partial t}+rs\frac{\partial V}{\partial s}+\frac{\sigma^2s^2}{2}\frac{\partial^2V}{\partial s^2}-rV=0.$$

Show that a solution of this equation is

$$V(s,t):=\frac{e^{-r(t_0-t)}}{\sqrt{2\pi\sigma^2(t_0-t)}}\int_0^\infty\frac{1}{z}e^{-\frac{(\log(s/z)+(r-\sigma^2/2)(t_0-t))^2}{2\sigma^2(t_0-t)}}F(z)dz.$$

(This formula should be nearly identical to the Black-Scholes Option Pricing formula from Remark 7.29, where we take $F(z):=\max(S_0z-k,0)$.) Instead of differentiating V directly, use the following strategy.

First, show that the Black-Scholes equation reduces to the one-dimensional heat equation

$$\frac{\partial U}{\partial\tau}=\frac{\partial^2U}{\partial x^2},$$

where $V(s,t)=e^{ax+b\tau}U(x,\tau)$, $x=\log s$, $\tau=(\sigma^2/2)(t_0-t)$, $a=(1/2)-r/\sigma^2$, and $b=-(1/2+r/\sigma^2)^2$, and U satisfies the initial condition $U(x,0)=e^{-ax}F(e^x)$ for all $x\in\mathbb{R}$. (Start by differentiating V with respect to s and t, etc.) That is, the Black-Scholes equation is the heat equation, run backwards in time.

Finally, use the formula for U using Exercise 8.20.

8.1. **Vasicek Interest Rate Model/Ornstein-Uhlenbeck Model.** Let $f\colon\mathbb{R}\to\mathbb{R}$ be a differentiable function. We write $f=f(t)$ so that $t\in\mathbb{R}$. Let $a,b>0$. Suppose f satisfies the following ordinary differential equation

$$\frac{df}{dt}(t)=a(b-f(t)), \qquad \forall\, t\in\mathbb{R}. \qquad (*)$$

We can solve this equation using the method of integrating factors. Note that

$$\frac{d}{dt}\big(e^{at}f(t)\big)=e^{at}\frac{df}{dt}(t)+ae^{at}f(t)\overset{(*)}{=}ae^{at}(b-f(t)+f(t))=abe^{at}.$$

Integrating both sides with respect to t,

$$e^{at}f(t)=f(0)+\int_0^t abe^{as}ds=f(0)+b(e^{at}-1).$$

In summary, we can solve $(*)$ using the formula

$$f(t)=e^{-at}f(0)+b-be^{-at}=b+e^{-at}(f(0)-b), \qquad \forall\, t\in\mathbb{R}.$$

Note that $\lim_{t\to\infty} f(t) = b$ since $a > 0$. Also, by $(*)$, if $f(t) > b$, then f will decrease, and if $f(t) < b$, then f will increase. And from our explicit formula for f, we see that f converges exponentially fast to b.

The Vasicek model uses the same differential equation, with an added stochastic noise, which together form a stochastic differential equation.

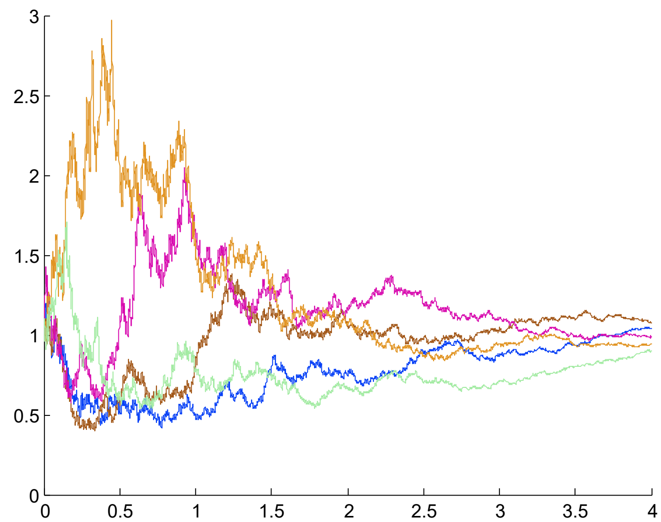


FIGURE 6. Sample Paths of the Vasicek model with $a = b = \sigma = f(0) = 1$. The horizontal axis is the t-axis.

Definition 8.22 (Vasicek model/Ornstein-Uhlenbeck model). Let $a, b, \sigma > 0$. Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. The **Vasicek model** models an interest rate as a (random) function $f\colon \mathbb{R} \to \mathbb{R}$ satisfying the following stochastic differential equation for any $t > 0$:

$$df(t) = a(b - f(t))dt + \sigma dB(t).$$

(Since f is a random function, f is also a function of the sample space, but we omit this dependence from our notation here and below.)

Proposition 8.23. *A solution of the Vasicek model can be written as*

$$f(t) = b + e^{-at}(f(0) - b) + \sigma \int_0^t e^{a(s-t)} dB(s), \qquad \forall\, t > 0.$$

Proof. As in the case $\sigma = 0$, we use the method of integrating factors. Using Itô's formula Version 2, Theorem 8.17, for the function $g(x, y) = e^{ax} f(x)$, and the usual product rule,

$$\begin{aligned} d(e^{at} f(t)) &= \frac{d}{dt}[e^{at} f(t)]dt = ae^{at} f(t)dt + e^{at}\frac{df}{dt}(t)dt \\ &= ae^{at} f(t)dt + e^{at} df(t) = ae^{at} f(t)dt + e^{at}[a(b - f(t))dt + \sigma dB(t)] \\ &= abe^{at}dt + \sigma e^{at} dB(t). \end{aligned}$$

Or, written in integral form,

$$e^{at} f(t) = f(0) + ab\int_0^t e^{as} ds + \sigma \int_0^t e^{as} dB(s) = f(0) + b(e^{at} - 1) + \sigma \int_0^t e^{as} dB(s).$$

Multiplying both sides by e^{-at} completes the proof. $\square$

Exercise 8.24. Let $a,b,\sigma>0$. Let $f\colon\mathbb{R}\to\mathbb{R}$ satisfying the Vasicek stochastic differential equation for any $t\in\mathbb{R}$.

$$df(t)=a(b-f(t))dt+\sigma dB(t).$$

Show that, for any $t>0$,

$$\mathbf{E}f(t)=b+e^{-at}(f(0)-b),\qquad \mathrm{var}(f(t))=\frac{\sigma^2}{2a}(1-e^{-2at}).$$

More generally, for any $s,t>0$, show that

$$\mathrm{cov}(f(t),f(u))=\mathbf{E}\big((f(t)-\mathbf{E}f(t))(f(u)-\mathbf{E}f(u))\big)=\frac{\sigma^2}{2a}(e^{-a|t-u|}-e^{-a(t+u)}).$$

Conclude that $\lim_{t\to\infty}\mathbf{E}f(t)=b$ and $\lim_{t\to\infty}\mathrm{var}(f(t))=\frac{\sigma^2}{2a}$.

Exercise 8.25. Using a Monte Carlo simulation, plot several sample paths of the Vasicek stochastic differential equation, with $a=b=\sigma=f(0)=1$.

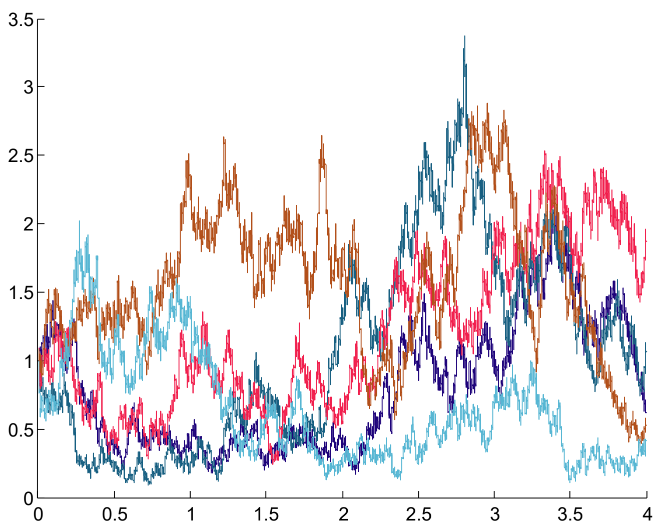


FIGURE 7. Sample Paths of the CIR model with $a=b=\sigma=f(0)=1$. The horizontal axis is the t-axis.

Exercise 8.26 (**Cox-Ingersoll-Ross (CIR) model**). Let $a,b,\sigma>0$. Let $\{B(t)\}_{t\geq0}$ be a standard Brownian motion. The **Cox-Ingersoll-Ross model** models an interest rate as a (random) function $f\colon\mathbb{R}\to\mathbb{R}$ satisfying the following stochastic differential equation for any $t>0$:

$$df(t)=a(b-f(t))dt+\sqrt{f(t)}\sigma dB(t).$$

(Since f is a random function, f is also a function of the sample space, but we omit this dependence from our notation here and below.)

A priori, this stochastic differential equation is not rigorously defined, since $\sqrt{f(t)}$ will not be a real number when $f(t)<0$. In this exercise, we ignore this issue. (In actuality, if $f(0)>0$, then $f(t)<0$ occurs with probability 0.)

Unlike the Vasicek model, we might not be able to get a closed form solution of this equation. Nevertheless, we can still run a Monte Carlo simulation of this stochastic differential equation as follows. Let $f(0) = 1$. Let $i, n > 0$ be integers. Suppose we have inductively determined $f(i/n)$ using a Monte Carlo simulation, and we would like the determine $f((i+1)/n)$. The stochastic differential equation then suggests that

$$f((i+1)/n) \approx f(i/n) + a(b - f(i/n))(1/n) + \sqrt{f(i/n)}\sigma(B((i+1)/n) - B(i/n)).$$

This approximation is known as a **finite difference scheme**.

Using this approximation, plot several sample paths of the CIR model with $a = b = f(0) = \sigma = 1$.

What would be the corresponding finite difference scheme for the Vasicek model?

8.2. **Stochastic Heat Equation.** In Exercise 8.20, we showed that a solution f of the heat equation with forcing term

$$\frac{\partial f}{\partial t}(x,t) = \frac{\partial^2 f}{\partial x^2}(x,t) + h(x,t), \quad \forall\,(x,t) \in \mathbb{R} \times [0,\infty),$$
$$f(x,0) = g(x), \quad \forall\, x \in \mathbb{R}.$$

can be written so that, for any $(x,t) \in \mathbb{R} \times [0,\infty)$,

$$f(x,t) := \frac{1}{\sqrt{4\pi t}} \int_{\mathbb{R}} e^{-\frac{(x-y)^2}{4t}} g(y)dy + \int_0^t \frac{1}{\sqrt{4\pi(t-s)}} \int_{\mathbb{R}} e^{-\frac{(x-y)^2}{4(t-s)}} h(y,s)dyds.$$

The heat equation is a partial differential equation that is used to model the flow of heat, given an initial distribution of heat defined by the function $g\colon \mathbb{R} \to \mathbb{R}$. We can think of h as supplying a "source" of heat in this equation. The stochastic heat equation is the same equation, where the function h becomes a random variable.

Definition 8.27 (**Stochastic Heat Equation**). Let $\{Z(x,t)\}_{x\in\mathbb{R},t\geq 0}$ be a set of independent, standard Gaussian random variables. Let $g\colon \mathbb{R} \to \mathbb{R}$ be a bounded continuous function. The **stochastic heat equation** is the following stochastic partial differential equation for a (random) function $f\colon \mathbb{R} \times [0,\infty) \to \mathbb{R}$:

$$\frac{\partial f}{\partial t}(x,t) = \frac{\partial^2 f}{\partial x^2}(x,t) + Z(x,t), \quad \forall\,(x,t) \in \mathbb{R} \times [0,\infty),$$
$$f(x,0) = g(x), \qquad \forall\, x \in \mathbb{R}.$$

(Since f is a random function, f is also a function of the sample space, but we omit this dependence from our notation here and below.)

Exercise 8.28. Let $\{Z(x,t)\}_{x\in\mathbb{R},t\geq 0}$ be a set of independent, standard Gaussian random variables. Suppose $f\colon \mathbb{R} \times [0,\infty) \to \mathbb{R}$ satisfies the stochastic heat equation.

$$\frac{\partial f}{\partial t}(x,t) = \frac{\partial^2 f}{\partial x^2}(x,t) + h(x,t), \quad \forall\,(x,t) \in \mathbb{R} \times [0,\infty),$$
$$f(x,0) = 0, \qquad \forall\, x \in \mathbb{R}.$$

We can explicitly solve this equation by its analogy with Exercise 8.20. That is,

$$f(x,t) := \int_0^t \frac{1}{\sqrt{4\pi(t-s)}} \int_{\mathbb{R}} e^{-\frac{(x-y)^2}{4(t-s)}} Z(y,s)dyds, \qquad \forall\,(x,t) \in \mathbb{R} \times [0,\infty),$$

satisfies the stochastic heat equation. Show that f has the following covariance for any $s, t > 0$:

$$\mathbf{E}[f(0,s)f(0,t)] = \frac{1}{2\sqrt{\pi}}(|s+t|^{1/2} - |s-t|^{1/2}).$$

8.3. **Itô Processes.** Let $S_0, \sigma > 0$ and let $\mu \in \mathbb{R}$. Let $\{S(t)\}_{t\geq 0} = \{S_0 e^{\sigma B(t)+\mu t}\}_{t\geq 0}$ be a geometric Brownian motion. Using the function $f(x,y) := S_0 e^{\sigma y+\mu x}$ $\forall\ x, y \in \mathbb{R}$ in Itô's Lemma, Theorem 8.10, we get

$$dS(t) = \sigma S(t)dB(t) + (\mu + \sigma^2/2)S(t)dt, \qquad \forall\, t \geq 0.$$

If $X(t) := \sigma B(t) + \mu t$, $\forall\ t \geq 0$, we use $f(x,y) = \sigma y + \mu x$ $\forall\ x, y \in \mathbb{R}$ in Theorem 8.10 to get

$$dX(t) = \sigma dB(t) + \mu dt, \qquad \forall\, t \geq 0.$$

The term in front of $dB(t)$ is called the **diffusion** of the stochastic process, and the term in front of dt is called the **drift** of the stochastic process. In mathematical finance, the diffusion term is interpreted as the volatility of the stock. An Itô processes is a stochastic process satisfying a general stochastic differential equation of the above form.

Definition 8.29 (**Itô Process**)**.** Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. We say that a stochastic process $\{Y(t)\}_{t\geq 0}$ is an **Itô process** if $\exists$ functions $\sigma, \mu\colon \mathbb{R}^2 \to \mathbb{R}$ such that $Y(t)$ satisfies a stochastic differential equation of the following form:

$$dY(t) = \sigma(t, Y(t))dB(t) + \mu(t, Y(t))dt, \qquad \forall\, t \geq 0.$$

As we have shown above, Brownian motion and geometric Brownian motion are examples of Itô processes. Itô processes can model stock prices and other financial securities. Itô processes themselves satisfy a version of Itô's Lemma. The difference between Theorem 8.10 and this new version of Itô's Lemma is that the dt term in Theorem 8.10 might be different. We now describe this new term in Itô's Lemma.

Definition 8.30 (**Quadratic Variation**)**.** Let $\{Y(t)\}_{t\geq 0}$ be an Itô process. We define another stochastic process $\{[Y]_t\}_{t\geq 0}$ so that, for any $b > 0$, $[Y]_b$ is the limit in probability as $n \to \infty$ of

$$\sum_{i=0}^{n-1}(Y(b(i+1)/n) - Y(bi/n))^2.$$

Remark 8.31. Some books use a fairly deceptive and informal notation of "$(dY(t))^2$" instead of $d[Y]_t$ for the quadratic variation process.

Remark 8.32. In Lemma 8.16, if we choose f to be a constant function, and if $\{B(t)\}_{t\geq 0}$ is standard Brownian motion, it follows that $[B]_t = t$, for all $t \geq 0$. In general, the quadratic variation will be a random variable. Though, in this special case, the quadratic variation of Brownian motion itself is not random.

Lemma 8.33. *Let $\{Y(t)\}_{t\geq 0}$ be an Itô process, so that $\exists$ $\sigma, \mu\colon \mathbb{R}^2 \to \mathbb{R}$ with $dY(t) = \sigma(t, Y(t))dB(t) + \mu(t, Y(t))dt$ $\forall\ t \geq 0$. Then*

$$d[Y]_t = (\sigma(t, Y(t)))^2 dt, \qquad \forall\, t \geq 0.$$

Proof Sketch. By the definition of the Itô process, for any $b > 0$, we have the approximation

$$\begin{aligned}
&\sum_{i=0}^{n-1}(Y(b(i+1)/n) - Y(bi/n))^2 \\
&\quad\approx \sum_{i=0}^{n-1}\Big(\sigma(bi/n, B(bi/n))\Big(B(b(i+1)/n) - B(bi/n)\Big) + \mu(bi/n, B(bi/n))\frac{b}{n}\Big)^2 \\
&\quad= \sum_{i=0}^{n-1}\Big(\sigma(bi/n, B(bi/n))\Big)^2\Big(B(b(i+1)/n) - B(bi/n)\Big)^2 + \frac{b}{n}\sum_{i=0}^{n-1}(\mu(bi/n, B(bi/n)))^2\frac{b}{n} \\
&\qquad\qquad + 2\sum_{i=0}^{n-1}\frac{b}{n}(\sigma(bi/n, B(bi/n))\Big(B(b(i+1)/n) - B(bi/n)\Big)\mu(bi/n, B(bi/n)).
\end{aligned}$$

As $n \to \infty$, the first term converges in probability to $\int_0^b(\sigma(s, B(s)))^2 ds$ by Lemma 8.16. The second term is a Riemann sum divided by n, so it converges to 0. The final term is a mean zero Gaussian with variance

$$4\sum_{i=0}^{n-1}\frac{b^3}{n^3}\mathbf{E}(\sigma(bi/n, B(bi/n))\mu(bi/n, B(bi/n))^2.$$

So, the final term converges to 0 as $n \to \infty$. That is, $[Y]_t = \int_0^t(\sigma(s, Y(s)))^2 ds$. □

Theorem 8.34 (**Itô's Formula, Version 3**)**.** *Let $f\colon \mathbb{R} \to \mathbb{R}$ have two continuous derivatives. Let $\{Y(t)\}_{t\geq 0}$ be an Itô process. Then, with probability 1, for all $b \geq 0$,*

$$f(Y(b)) - f(Y(0)) = \int_0^b f'(Y(s))dY(s) + \frac{1}{2}\int_0^b f''(Y(s))d[Y]_s.$$

Remark 8.35. Or, written in its differential form, we get

$$df(Y(s)) = f'(Y(s))dY(s) + \frac{1}{2}f''(Y(s))d[Y]_s, \qquad \forall\, s \geq 0.$$

In the case that $Y(t)$ is a standard Brownian motion, $[Y]_s = s$ for any $s \geq 0$, so we recover Itô's formula, Theorem 8.10, as a special case of Theorem 8.34. Also, from Lemma 8.33,

$$df(Y(s)) = f'(Y(s))dY(s) + \frac{1}{2}f''(Y(s))(\sigma(s, Y(s)))^2 ds, \qquad \forall\, s \geq 0.$$

Proof Sketch. From Taylor's Theorem, if $x, y \in \mathbb{R}$, then there exists an error term $R(x, y)$ such that

$$f(x) - f(y) = f'(y)(x - y) + \frac{1}{2}f''(y)(x - y)^2 + R(x, y).$$

For any $0 \leq i \leq n-1$, let $x := Y(b(i+1)/n)$, let $y := Y(bi/n)$, and sum over i to get

$$f(Y(b)) - f(Y(0)) = \sum_{i=0}^{n-1} (f(Y(b(i+1)/n)) - f(Y(bi/n)))$$

$$= \sum_{i=1}^{n-1} f'(Y(bi/n))\big(Y(b(i+1)/n) - Y(bi/n)\big)$$

$$+ \frac{1}{2}\sum_{i=1}^{n-1} f''(Y(bi/n))\big(Y(b(i+1)/n) - Y(bi/n)\big)^2 + \sum_{i=1}^{n-1} R(Y(b(i+1)/n), Y(bi/n)).$$

We now let $n \to \infty$. The first term converges in probability to $\int_0^b f'(Y(s))dY(s)$ by the definition of the stochastic integral. The second term converges to $\frac{1}{2}\int_0^b f''(Y(s))d[Y]_s$ in probability. Treating the last term as an error term concludes the proof. □

Lemma 8.33 simplifies computations of Theorem 8.34.

Theorem 8.36 (**Itô's Formula, Version 4**). *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ have two continuous derivatives in each coordinate. We write $f = f(x, y)$, $x, y \in \mathbb{R}$. Let $\{Y(t)\}_{t\geq 0}$ be an Itô process. Then, with probability 1, for all $b \geq 0$,*

$$f(b, Y(b)) - f(0, Y(0)) = \int_0^b \frac{\partial f}{\partial x}(s, Y(s))ds + \int_0^b \frac{\partial f}{\partial y}(s, Y(s))dY(s) + \frac{1}{2}\int_0^b \frac{\partial^2 f}{\partial y^2}(s, Y(s))d[Y]_s.$$

Remark 8.37. Or, written in its differential form, for all $s \geq 0$,

$$df(s, Y(s)) = \frac{\partial f}{\partial x}(s, Y(s))ds + \frac{\partial f}{\partial y}(s, Y(s))dY(s) + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(s, Y(s))d[Y]_s.$$

Also, from Lemma 8.33, we have for all $s \geq 0$,

$$df(s, Y(s)) = \frac{\partial f}{\partial x}(s, Y(s))ds + \frac{\partial f}{\partial y}(s, Y(s))dY(s) + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(s, Y(s))(\sigma(s, Y(s)))^2 ds.$$

Proposition 8.38 (**The Sharpe Ratio**). *Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Let $\mu_1, \mu_2 \in \mathbb{R}, \sigma_1, \sigma_2 > 0$. Suppose the prices $\{S_1(t)\}_{t\geq 0}$ and $\{S_2(t)\}_{t\geq 0}$ of two (non-dividend paying) stocks satisfy the following (coupled) stochastic differential equations for any $t \geq 0$:*

$$dS_1(t) = \mu_1 S_1(t)dt + \sigma_1 S_1(t)dB(t),$$

$$dS_2(t) = \mu_2 S_2(t)dt + \sigma_2 S_2(t)dB(t).$$

If no arbitrage opportunity exists, and if money can be borrowed at a risk-free interest rate $r > 0$, then the ***Sharpe ratios*** *of the stocks are the same:*

$$\frac{\mu_1 - r}{\sigma_1} = \frac{\mu_2 - r}{\sigma_2}.$$

Proof. We argue by contradiction. Without loss of generality, assume that $\frac{\mu_1 - r}{\sigma_1} > \frac{\mu_2 - r}{\sigma_2}$. We will then create an arbitrage opportunity. At time 0, buy $1/(\sigma_1 S_1(0))$ shares of the first stock for the price $1/\sigma_1$, short $1/(\sigma_2 S_2(0))$ shares of the second stock for the price $1/\sigma_2$, and lend the price difference $(1/\sigma_2) - (1/\sigma_1)$ at the interest rate r. (If this quantity is negative, we borrow this amount.) At time 0, the instantaneous revenue from this investment is

$$\frac{1}{\sigma_1 S_1(0)} dS_1(0) - \frac{1}{\sigma_2 S_2(0)} dS_2(0) + \Big(\frac{1}{\sigma_2} - \frac{1}{\sigma_1}\Big) r dt = \Big(\frac{\mu_1 - r}{\sigma_1} - \frac{\mu_2 - r}{\sigma_2}\Big) dt.$$

Note that the $dB(t)$ terms cancelled. By assumption, the instantaneous revenue is positive, a contradiction. □

Example 8.39 (MFE Sample Question, from an old exam)**.** Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Let $a, b, c \in \mathbb{R}$. Suppose the prices $\{S_1(t)\}_{t\geq 0}$ and $\{S_2(t)\}_{t\geq 0}$ of two (non-dividend paying) stocks satisfy the following (coupled) stochastic differential equations for any $t \geq 0$:

$$dS_1(t) = (.07)S_1(t)dt + (.12)S_1(t)dB(t),$$
$$dS_2(t) = aS_2(t)dt + bS_2(t)dB(t).$$

It is also given that $r = .04$ is a risk-free interest rate, and

$$d\log(S_2(t)) = cdt + (.08)dB(t), \qquad \forall\, t \geq 0.$$

What is a?

Applying Theorem 8.34 and Lemma 8.33, for any $t \geq 0$,

$$d\log(S_2(t)) = \frac{1}{S_2(t)} dS_2(t) - \frac{1}{2(S_2(t))^2} d[S_2]_t = \frac{1}{S_2(t)} dS_2(t) - \frac{b^2}{2} dt.$$

Using the assumed formula for $dS_2(t)$, for any $t \geq 0$,

$$d\log(S_2(t)) = (a - b^2/2)dt + bdB(t).$$

Then, using the given formula for $d\log(S_2(t))$, we get $b = .08$, and $c = a - b^2/2$. Then, using Proposition 8.38,

$$\frac{.07 - .04}{.12} = \frac{a - .04}{.08}.$$

Solving for a, we get $a = (2/3)(.03) + .04 = .06$.

8.4. **Partial Differential Equations and Brownian Motion.** Some (deterministic) partial differential equations can be solved by taking expected values with respect to Brownian motion. This probabilistic perspective offers an alternative to analytical approaches to solving (deterministic) PDEs.

Let $U \subseteq \mathbb{R}^n$ be a connected open set with boundary ∂U, such that ∂U can be locally written as the graph of a twice continuously differentiable function. Let $f\colon U \to \mathbb{R}$ be a twice continuously differentiable function. We say that f is **harmonic** in U if

$$\Delta f(x) := \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}(x) = 0, \qquad \forall\, x \in U.$$

We say that f is **subharmonic** in U if $\Delta f(x) \geq 0$ for all $x \in U$. We say that f is **superharmonic** in U if $\Delta f(x) \leq 0$ for all $x \in U$.

Let $\phi\colon \partial U \to \mathbb{R}$ be a continuous function. We say that $f\colon \overline{U} \to \mathbb{R}$ solves the **Dirichlet problem** with boundary value ϕ if f is harmonic in U and $f(x) = \phi(x)$ for all $x \in \partial U$. Let

$\{B(t)\}_{t\geq 0}$ be a Brownian motion. For any $x \in U$, we use the notation $\mathbf{E}_x$ to denote that the Brownian motion is started at x (so that $B(0) = x$). Define a stopping time

$$T := \inf\{t > 0 \colon B(t) \in \partial U\}.$$

Theorem 8.40 (**Solution of Dirichlet Problem**). *Given ϕ, define $f \colon \overline{U} \to \mathbb{R}$ by*

$$f(x) := \mathbf{E}_x \phi(B(T)).$$

Then f is the unique continuous function that is harmonic in U such that $f(x) = \phi(x)$ for all $x \in \partial U$.

Proof Sketch. Let $\varepsilon > 0$. Denote $D(x,\varepsilon) := \{y \in \mathbb{R}^n \colon \|x - y\| < \varepsilon\}$. Choose $\varepsilon > 0$ so that $D(x,\varepsilon) \subseteq U$ and define $\widetilde{T} := \inf\{t > 0 \colon B(t) \notin D(x,\varepsilon)\}$. Let $\{\overline{B}_t\}_{t\geq 0}$ be a Brownian motion that is independent of $\{B(t)\}_{t\geq 0}$. Using the Strong Markov property for Brownian motion (i.e. Proposition 7.4 for a stopping time: $\{\overline{B}(\widetilde{T} + s) - \overline{B}(s)\}_{s\geq 0}$ is a standard Brownian motion that is independent of $\widetilde{T}$), and then using the definition of f,

$$\begin{aligned} f(x) &= \mathbf{E}_x \phi(B(T)) = \mathbf{E}\Big(\phi(B(T)) \,\Big|\, B(0) = x\Big) \\ &= \mathbf{E}\Big(\phi(B(T)) \,\Big|\, \overline{B}(0) = x,\, B(0) = \overline{B}(\widetilde{T})\Big) = \mathbf{E}\Big(f(\overline{B}(\widetilde{T})) \,\Big|\, \overline{B}(0) = x\Big) = \mathbf{E}_x f(\overline{B}(\widetilde{T})). \end{aligned}$$

That is, $f(x)$ is the average value of f on $\partial D(x,\varepsilon)$, for any $\varepsilon > 0$, for any $x \in U$ such that $D(x,\varepsilon) \subseteq U$. This mean value property implies that f is harmonic. (Hint: define $h(\varepsilon)$ to be $\mathbf{E}_x f(\overline{B}(\widetilde{T}))$, as defined above. Then show $h'(\varepsilon) = 0$, let $\varepsilon \to 0$, and apply the divergence theorem.) □

Brownian motion can be used to prove the following well-known (deterministic) result for harmonic functions.

Theorem 8.41 (**Liouville's Theorem**). *Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a bounded harmonic function. Then f is a constant function.*

Proof. Let $x, y \in \mathbb{R}^n$ with $x \neq y$, and let $H \subseteq \mathbb{R}^n$ be the unique hyperplane such that reflection across H interchanges x and y. Let $\{B(t)\}_{t\geq 0}$ be a Brownian motion with $B(0) = x$ and let $\{\overline{B}(t)\}_{t\geq 0}$ be the reflection of $\{B(t)\}_{t\geq 0}$ across H. Let $T := \inf\{t > 0 \colon B(t) \in H\}$. Then the following two events are equal in distribution

$$\{B(t) \colon t \geq T\}, \qquad \{\overline{B}(t) \colon t \geq T\}. \qquad (*)$$

Since f is harmonic and bounded, $f(x) = \mathbf{E} f(B(T))$. Fix $t > 0$. Then

$$f(x) = \mathbf{E} f(B(T)) = \mathbf{E} f(B(T)) 1_{t<T} + \mathbf{E} f(B(T)) 1_{t\geq T},$$

and similarly,

$$f(y) = \mathbf{E} f(\overline{B}(T)) = \mathbf{E} f(\overline{B}(T)) 1_{t<T} + \mathbf{E} f(\overline{B}(T)) 1_{t\geq T}.$$

So, by $(*)$,

$$|f(x) - f(y)| = \left|\mathbf{E} f(B(T)) 1_{t<T} + \mathbf{E} f(\overline{B}(T)) 1_{t<T}\right| \leq \sup_{z\in\mathbb{R}^n} |f(z)| \cdot 2 \cdot \mathbf{P}(t < T).$$

Letting $t \to \infty$, we have $\mathbf{P}(T > t) \to 0$ as $t \to \infty$, so that $f(x) = f(y)$. □

Remark 8.42. For more details on stochastic integration, see e.g. Section 7 of https://people.bath.ac.uk/maspm/book.pdf

One can solve other partial differential equations in a similar way to Theorem 8.40.

Example 8.43. A solution of the equation

$$\frac{1}{2}\Delta f(x) = -g(x), \quad \forall\, x \in U, \qquad f(x) = 0, \quad \forall\, x \in \partial U$$

can be written as

$$f(x) := \mathbf{E}_x \int_0^T g(B(t))dt,$$

where $T := \inf\{t > 0\colon B(t) \notin U\}$.

Example 8.44. A solution of the equation

$$\frac{1}{2}\Delta f(x) + g(x)f(x) = 0, \quad \forall\, x \in U, \qquad f(x) = \phi(x), \quad \forall\, x \in \partial U$$

can be written as

$$f(x) := \mathbf{E}_x(\phi(B(T))e^{g(T)}),$$

Example 8.45. A solution of the equation

$$\frac{1}{2}\Delta f(x,t) = \frac{d}{dt}f(x,t), \quad \forall\, (x,t) \in \mathbb{R}^n \times (0,\infty), \qquad f(x,0) = \phi(x), \quad \forall\, x \in \partial U$$

can be written as

$$f(x,t) := \mathbf{E}_x\phi(B(t)), \qquad \forall\, (x,t) \in \mathbb{R}^n \times (0,\infty).$$

Example 8.46. A solution of the equation

$$\frac{1}{2}\Delta f(x,t) + g(x,t) = \frac{d}{dt}f(x,t), \quad \forall\, (x,t) \in \mathbb{R}^n \times (0,\infty), \qquad f(x,0) = \phi(x), \quad \forall\, x \in \partial U$$

can be written as

$$f(x,t) := \mathbf{E}_x\Big(\phi(B(t)) + \int_0^t g(B(s), t-s)ds\Big), \qquad \forall\, (x,t) \in \mathbb{R}^n \times (0,\infty).$$

Example 8.47. A solution of the equation

$$\frac{1}{2}\Delta f(x,t) + g(x,t)f(x,t) = \frac{d}{dt}f(x,t), \quad \forall\, (x,t) \in \mathbb{R}^n \times (0,\infty), \qquad f(x,0) = \phi(x), \quad \forall\, x \in \partial U$$

can be written as

$$f(x,t) := \mathbf{E}_x\Big(\phi(B(t)) \exp\Big(\int_0^t g(B(s), t-s)ds\Big)\Big), \qquad \forall\, (x,t) \in \mathbb{R}^n \times (0,\infty).$$

9. 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 Y be a discrete random variable on a sample space $\mathcal{C}$, so that $Y\colon \mathcal{C} \to \mathbb{R}$. 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

$$p_Y(x) = \mathbf{P}(Y = x) = \mathbf{P}(\{c \in \mathcal{C} \colon Y(c) = x\}),\ \forall\, x \in \mathbb{R}$$

the Probability Mass Function (PMF) of Y

$\mathbf{E}(Y)$ denotes the expected value of Y

$\mathrm{var}(Y) = \mathbf{E}(Y - \mathbf{E}(Y))^2$, the variance of Y

$\sigma_Y = \sqrt{\mathrm{var}(Y)}$, the standard deviation of Y

$Y|A$ denotes the random variable Y conditioned on the event A.

$\mathbf{E}(Y|A)$ denotes the expected value of Y conditioned on the event A.

$\mathbf{E}(Y|B_1,\ldots,B_n) := \mathbf{E}(Y|\cap_{i=1}^n B_i)$ denotes the conditional expectation of Y, given $\cap_{i=1}^n B_i$.

Let Y, Y be a continuous random variables on a sample space $\mathcal{C}$, so that $Y, Y\colon \mathcal{C} \to \mathbb{R}$. 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 $A \subseteq \mathcal{C}$.

$f_Y\colon \mathbb{R} \to [0,\infty)$ denotes the Probability Density Function (PDF) of Y, so

$$\mathbf{P}(a \le Y \le b) = \int_a^b f_Y(x)dx$$

$f_{Y,Y}\colon \mathbb{R} \to [0,\infty)$ denotes the joint PDF of Y and Y, so

$$\mathbf{P}(a \le Y \le b, c \le Y \le d) = \int_c^d \int_a^b f_{Y,Y}(x,y)dxdy$$

$f_{Y|A}$ denotes the Conditional PDF of Y given A

$\mathbf{E}(Y|A)$ denotes the expected value of Y conditioned on the event A.

$\mathbf{E}(Y|\mathcal{A})$ denotes the expected value of Y given a partition $\mathcal{A} = \{A_1,\ldots,A_k\}$ of $\mathcal{C}$.

Let Y be a random variable on a sample space $\mathcal{C}$, so that $Y\colon \mathcal{C} \to \mathbb{R}$. Let $\mathbf{P}$ be a probability law on $\mathcal{C}$. Let $x \in \mathbb{R}$.

$$F_Y(x) = \mathbf{P}(Y \le x) = \mathbf{P}(\{c \in \mathcal{C} \colon Y(c) \le x\})$$

the Cumulative Distibution Function (CDF) of Y.

Let $(Y_0, Y_1, \ldots)$ be a real valued stochastic process. Let $x, y \in \mathbb{R}$.

$\mathbf{P}_x$ denotes the conditional probability such that

$$\mathbf{P}_x(A) = \mathbf{P}(A \,|\, Y_0 = x), \forall\, A \text{ in the sample space}$$

$\mathbf{E}_x$ denotes expectation with respect to $\mathbf{P}_x$

$$\Phi(y) := \int_{-\infty}^{y} e^{-t^2/2} \frac{dt}{\sqrt{2\pi}}.$$

Let $\{B(s)\}_{s\ge 0}$ be a standard Brownian motion. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuous function. Let $g\colon \mathbb{R}^2 \to \mathbb{R}$ be a continuous function. Let $b > 0$. Let $S_0, \sigma > 0$, $\mu \in \mathbb{R}$.

$\{S(t)\}_{t\geq 0} = \{S_0 e^{\sigma B(t)+\mu t}\}_{t\geq 0}$ denotes a geometric Brownian motion

$c := S_0\Phi(d_1) - e^{-rt}k\Phi(d_1 - \sigma\sqrt{t})$ denotes the Black-Scholes pricing formula for a European call option with strike price $k > 0$ and expiration time $t > 0$, where

$$d_1 := \frac{\log(S_0/k) + (r + \sigma^2/2)t}{\sigma\sqrt{t}}, \qquad r := \mu + \sigma^2/2$$

$\int_0^b f(s)ds = \lim_{n\to\infty} \sum_{i=0}^{n-1} f\Big(\frac{bi}{n}\Big)\frac{b}{n}$, denotes the Riemann integral of f on $[0,b]$

$\int_0^b f(B(s))ds =_{\substack{\text{limit in probability}\\ \text{as } n\to\infty}} \sum_{i=0}^{n-1} f\Big(B\Big(\frac{bi}{n}\Big)\Big)\frac{b}{n}$, denotes the Riemann integral of $f(B(s))$ on $[0,b]$

$\int_0^b f(B(s))dB(s) =_{\substack{\text{limit in probability}\\ \text{as } n\to\infty}} \sum_{i=0}^{n-1} f\Big(B\Big(\frac{bi}{n}\Big)\Big)\Big(B\Big(\frac{b(i+1)}{n}\Big) - B\Big(\frac{bi}{n}\Big)\Big)$, denotes the stochastic integral of f on $[0,b]$

$\int_0^b g(s,B(s))dB(s) =_{\substack{\text{limit in probability}\\ \text{as } n\to\infty}} \sum_{i=0}^{n-1} g\Big(\frac{bi}{n}, B\Big(\frac{bi}{n}\Big)\Big)\Big(B\Big(\frac{b(i+1)}{n}\Big) - B\Big(\frac{bi}{n}\Big)\Big)$, denotes the stochastic integral of g on $[0,b]$

Let $\{Y(s)\}_{s\geq 0}$ be an Itô process.

$\int_0^b f(Y(s))dY(s) =_{\substack{\text{limit in probability}\\ \text{as } n\to\infty}} \sum_{i=0}^{n-1} f\Big(Y\Big(\frac{bi}{n}\Big)\Big)\Big(Y\Big(\frac{b(i+1)}{n}\Big) - Y\Big(\frac{bi}{n}\Big)\Big)$, denotes the stochastic integral of f on $[0,b]$, with respect to $\{Y(s)\}_{s\geq 0}$

$\int_0^b g(s,Y(s))dY(s) =_{\substack{\text{limit in probability}\\ \text{as } n\to\infty}} \sum_{i=0}^{n-1} g\Big(\frac{bi}{n}, Y\Big(\frac{bi}{n}\Big)\Big)\Big(Y\Big(\frac{b(i+1)}{n}\Big) - Y\Big(\frac{bi}{n}\Big)\Big)$, denotes the stochastic integral of g on $[0,b]$, with respect to $\{Y(s)\}_{s\geq 0}$

USC Mathematics, Los Angeles, CA
E-mail address: stevenmheilman@gmail.com