# MATH 174E, FINANCIAL MATHEMATICS, SPRING 2017

STEVEN HEILMAN

Contents



## 1. Introduction

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

In this class, we will study stochastic processes that model financial securities, such as stocks, bonds, currencies, derivatives, etc. A (financial) **derivative** is a financial security whose value depends on the value of some other financial security. Since the term "derivative" conflicts with the notion of derivative from calculus, we will not explicitly use the term "(financial) derivative" below.

When modeling financial securities, we will always implicitly make the assumptions below, which are **not realistic**. Remember that **a model is only an approximation to reality**. We are reminded of this every time the stock market behaves in strange ways. And Merton

*Date*: June 9, 2017 

and Scholes were reminded of this fact when their hedge fund went bankrupt the year after they won the Nobel prize.

- **Infinite Liquidity**. Market participants can buy or sell a unit of a financial security at any time.
- **Infinite depth**. Buying or selling a financial security does not affect the price of the security for further buying or selling.
- **No transaction costs**. The purchase price and the sale price of a financial security are the same.
- **No arbitrage**. There do not exist risk-free opportunities to instantly make money (such as buying something for \$1 and instantly selling it for \$2.)
- **Infinite credit**. Market participants can borrow or lend any amount of money at a risk-free, continuously compounded interest rate $r > 0$. So, anyone can deposit or lend some amount of money $x > 0$ at time $t_0$ and then receive $xe^{r(t_1-t_0)}$ at time $t_1 > t_0$. And anyone can borrow $x > 0$ at time $t_0$ and then pay back the loan in the amount $xe^{r(t_1-t_0)}$ at time $t_1 > t_0$.

After reviewing probability theory, we will discuss the theory of martingales. In financial terminology, a martingale is a "risk neutral" financial security. Much of mathematical finance can be summarized as: "find a martingale, or compare a financial security to a martingale."

## 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** $\Omega$ 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 $A_1, A_2, \ldots \subseteq \Omega$. We define

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

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

**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 $\Omega$ is a function whose domain is the set of all subsets of $\Omega$, and whose range is contained in $[0,1]$, such that

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

(ii) For any $A, B \subseteq \Omega$ such that $A \cap B = \emptyset$, we have

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

If $A_1, A_2, \ldots \subseteq \Omega$ 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 $\mathbf{P}(\Omega) = 1$. (**Normalization**)

**Definition 2.5** (**Conditional Probability**)**.** Let $A, B$ be subsets of some sample space $\Omega$. Let $\mathbf{P}$ be a probability law on $\Omega$. 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 $B_1, \ldots, B_n \subseteq \Omega$. 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 2.6** (**A Very Important Proposition**)**.** *Let $B$ be a fixed subset of some sample space $\Omega$. Let $\mathbf{P}$ be a probability law on $\Omega$. Assume that $\mathbf{P}(B) > 0$. Given any subset $A$ in $\Omega$, 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 $\Omega$, when viewed as a function of subsets $A$ in $\Omega$.*

**Proposition 2.7** (**Multiplication Rule**)**.** *Let $n$ be a positive integer. Let $A_1, \ldots, A_n$ be sets in some sample space $\Omega$, and let $\mathbf{P}$ be a probability law on $\Omega$. 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 2.8** (**Total Probability Theorem**)**.** *Let $A_1, \ldots, A_n$ be disjoint events in a sample space $\Omega$. That is, $A_i \cap A_j = \emptyset$ whenever $i, j \in \{1, \ldots, n\}$ satisfy $i \neq j$. Assume also that $\cup_{i=1}^n A_i = \Omega$. Let $\mathbf{P}$ be a probability law on $\Omega$. Then, for any event $B \subseteq \Omega$, we have*

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

**Definition 2.9** (**Independent Sets**)**.** Let $n$ be a positive integer. Let $A_1, \ldots, A_n$ be subsets of a sample space $\Omega$, and let $\mathbf{P}$ be a probability law on $\Omega$. 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 2.10** (**Random Variable**)**.** Let $\Omega$ be a sample space. Let $\mathbf{P}$ be a probability law on $\Omega$. A **random variable** $X$ is a function $X\colon \Omega \to \mathbb{R}$. (Sometimes we might also consider a random variable to be a function from $\Omega$ to another set.) A **discrete random variable** is a random variable whose range is either finite or countably infinite. A **probability density function** (PDF) is a function $f\colon \mathbb{R} \to [0,\infty)$ such that $\int_{-\infty}^{\infty} f(x)dx = 1$, and such that, for any $-\infty \leq a \leq b \leq \infty$, the integral $\int_a^b f(x)dx$ exists. A random variable $X$ is called **continuous** if there exists a probability density function $f$ such that, for any $-\infty \leq a \leq b \leq \infty$, we have

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

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

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

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

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

**Definition 2.11** (**Probability Mass Function**)**.** Let $X$ be a discrete random variable on a sample space $\Omega$, so that $X\colon \Omega \to \mathbb{R}$. The **probability mass function** (or PMF) of $X$, denote $p_X\colon \mathbb{R} \to [0,1]$ is defined by

$$p_X(x) = \mathbf{P}(X = x) = \mathbf{P}(\{X = x\}) = \mathbf{P}(\{\omega \in \Omega\colon X(\omega) = x\}), \qquad x \in \mathbb{R}.$$

Let $A \subseteq \mathbb{R}$. We denote $\{X \in A\} := \{\omega \in \Omega\colon X(\omega) \in A\}$.

We now give descriptions of some commonly encountered random variables.

**Definition 2.12** (**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 2.13** (**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 2.14** (**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 2.15** (**Poisson Random Variable**)**.** Let $\lambda > 0$. A random variable $X$ is called a **Poisson random variable with parameter** $\lambda$ 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 2.16.** 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 2.17.** Let $\lambda > 0$. A random variable $X$ is called an **exponential random variable with parameter** $\lambda$ if $X$ has the following density function: $f(x) = \lambda e^{-\lambda x}$ when $x \geq 0$, and $f(x) = 0$ otherwise.

**Definition 2.18** (**Normal Random Variable**)**.** Let $\mu \in \mathbb{R}$, $\sigma > 0$. A continuous random variable $X$ is said to be **normal** or **Gaussian** with mean $\mu$ and variance $\sigma^2$ 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 2.19** (**Indicator Function**)**.** Let $A \subseteq \Omega$ be a set. We define the **indicator function of** $A$, denoted $1_A \colon \Omega \to \mathbb{R}$ so that $1_A(\omega) = 0$ if $\omega \notin A$, and $1_A(\omega) = 1$ if $\omega \in A$.

**Definition 2.20** (**Expected Value**)**.** Let $\Omega$ be a sample space, let $\mathbf{P}$ be a probability law on $\Omega$. Let $X$ be a random variable on $\Omega$. Assume that $X \colon \Omega \to [0, \infty)$. 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 2.21.** 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 2.22.** 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 2.23.** 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 2.24 (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 2.25.** *Let $X_1, \ldots, X_n$ be random variables. Then*

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

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

**Definition 2.26 (Convex Function).** Let $\phi\colon \mathbb{R} \to \mathbb{R}$. We say that $\phi$ 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 2.27.** Let $\phi\colon \mathbb{R} \to \mathbb{R}$. Show that $\phi$ 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 $\phi$ is differentiable, the latter condition says that $\phi$ lies above all of its tangent lines.)

(Hint: Suppose $\phi$ 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 2.28.** Let $X, Y$ be positive random variables on a sample space $\Omega$. Assume that $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$. Prove that $\mathbf{E}X \geq \mathbf{E}Y$.

**Proposition 2.29 (Jensen's Inequality).** *Let $X$ be a random variable. Let $\phi\colon \mathbb{R} \to \mathbb{R}$ be convex. Then*

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

*Proof.* Let $y = \mathbf{E}X$. Then Exercise 2.27 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 2.28, 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 2.30 (Variance).** Let $\Omega$ be a sample space, let $\mathbf{P}$ be a probability law on $\Omega$. Let $X$ be a random variable on $\Omega$. 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 $\sigma_X$, by

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

**Proposition 2.31.** *Let $\Omega$ be a sample space, let $\mathbf{P}$ be a probability law on $\Omega$. Let $X$ be a random variable on $\Omega$. Let $a, b$ be constants. Then*

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

We will review conditional expectation later on in the notes.

**Definition 2.32** (**Joint Density Function**)**.** We say that random variables $X_1, \ldots, X_n$ have **joint density function** $f\colon \mathbb{R}^n \to [0, \infty)$ if $\int_{\mathbb{R}^n} f(x)dx = 1$, and if

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

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

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

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

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

And so on.

**Definition 2.33** (**Independence of Random Variables**)**.** Let $X_1, \ldots, X_n$ be random variables on a sample space $\Omega$, and let $\mathbf{P}$ be a probability law on $\Omega$. We say that $X_1, \ldots, X_n$ are **independent** if

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

**Exercise 2.34.** Let $X_1, \ldots, X_n$ be discrete random variables. Assume that

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

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

**Exercise 2.35.** 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 2.36.** *Let $X_1, \ldots, X_n$ be random variables on a sample space $\Omega$, and let $\mathbf{P}$ be a probability law on $\Omega$. 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*

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

**Proposition 2.37.** *Let $X_1, \ldots, X_n$ be independent random variables. Then*

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

**Proposition 2.38.** *Let $0 = n_0 < n_1 < n_2 < \ldots < n_k = n$ be integers. Let $X_1, \ldots, X_n$ be independent random variables. For any $1 \leq i \leq k$, let $g_i \colon \mathbb{R}^{n_i - n_{i-1}} \to \mathbb{R}$. Then the random variables $g_1(X_1, \ldots, X_{n_1})$, $g_2(X_{n_1+1}, \ldots, X_{n_2}), \ldots,$ $g_k(X_{n_{k-1}+1}, \ldots, X_{n_k})$ are independent. Consequently,*

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

2.2. **Law Of Large Numbers.**

**Theorem 2.39** (**Weak Law of Large Numbers**)**.** *Let $X_1, \ldots, X_n$ be independent identically distributed random variables. Assume that $\mu := \mathbf{E}X_1$ is finite. Then for any $\varepsilon > 0$*

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

**Theorem 2.40** (**Strong Law of Large Numbers**)**.** *Let $X_1, \ldots, X_n$ be independent identically distributed random variables. Assume that $\mu := \mathbf{E}X_1$ is finite. Then*

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

2.3. **Central Limit Theorem.**

**Theorem 2.41** (**Central Limit Theorem**)**.** *Let $X_1, \ldots, X_n$ be independent identically distributed random variables. Assume that $\mathbf{E}\left|X_1\right| < \infty$ and $0 < \mathrm{Var}(X_1) < \infty$.*

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

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

**Remark 2.42.** The random variable $\frac{X_1 + \cdots + X_n - (1/2)n}{\sigma\sqrt{n}}$ has mean zero and variance 1, just like the standard Gaussian.

**Exercise 2.43.** Estimate the probability that 1000000 coin flips of fair coins will result in more than 501,000 heads, using the Central Limit Theorem. (Some of the following integrals may be relevant: $\int_{-\infty}^{0} e^{-t^2/2} dt/\sqrt{2\pi} = 1/2$, $\int_{-\infty}^{1} e^{-t^2/2} dt/\sqrt{2\pi} \approx .8413$, $\int_{-\infty}^{2} e^{-t^2/2} dt/\sqrt{2\pi} \approx .9772$, $\int_{-\infty}^{3} e^{-t^2/2} dt/\sqrt{2\pi} \approx .9987$.) (Hint: use Bernoulli random variables.)

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.

**Theorem 2.44** (**Fubini Theorem for Integrals**)**.** *Let $h \colon \mathbb{R}^2 \to \mathbb{R}$ be a continuous function such that $\iint_{\mathbb{R}^2} \left|h(x, y)\right| dxdy < \infty$. 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 2.45** (**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 2.46.** 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 2.47.** 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 2.48.** 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 $[0,1]$.

**Exercise 2.49** (**Confidence Intervals**)**.** Among 625 members of a bank chosen uniformly at random among all bank members, it was found that 25 had a savings account. Give an interval of the form $[a,b]$ where $0\leq a,b\leq 625$ are integers, such that with about 95% certainty, the number of any set of 625 bank members with savings accounts chosen uniformly at random lies in the interval $[a,b]$. (Hint: if $Y$ is a standard Gaussian random variable, then $\mathbf{P}(-2\leq Y\leq 2)\approx .95$.)

**Exercise 2.50** (**Hypothesis Testing**)**.** Suppose we run a casino, and we want to test whether or not a particular roulette wheel is biased. Let $p$ be the probability that red results from one spin of the roulette wheel. Using statistical terminology, "$p=18/38$" is the null hypothesis, and "$p\neq 18/38$" is the alternative hypothesis. (On a standard roulette wheel, 18 of the 38 spaces are red.) For any $i\geq 1$, let $X_i=1$ if the $i^{th}$ spin is red, and let $X_i=0$ otherwise.

Let $\mu:=\mathbf{E}X_1$ and let $\sigma:=\sqrt{\mathrm{var}(X_1)}$. If the null hypothesis is true, and if $Y$ is a standard Gaussian random variable

$$\lim_{n\to\infty}\mathbf{P}\left(\left|\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\right|\geq 2\right)=\mathbf{P}(|Y|\geq 2)\approx .05.$$

To test the null hypothesis, we spin the wheel $n$ times. In our test, we reject the null hypothesis if $|X_1+\cdots+X_n-n\mu|>2\sigma\sqrt{n}$. Rejecting the null hypothesis when it is true is called a type $I$ error. In this test, we set the type $I$ error percentage to be 5%. (The type $I$ error percentage is closely related to the p-value.)

Suppose we spin the wheel $n = 3800$ times and we get red 1868 times. Is the wheel biased? That is, can we reject the null hypothesis with around 95% certainty?

**Exercise 2.51** (**Numerical Integration**)**.** In computer graphics in video games, etc., various integrations are performed in order to simulate lighting effects. Here is a way to use random sampling to integrate a function in order to quickly and accurately render lighting effects. Let $\Omega = [0, 1]$, and let $\mathbf{P}$ be the uniform probably law on $\Omega$, so that if $0 \leq a < b \leq 1$, we have $\mathbf{P}([a, b]) = b - a$. Let $X_1, \ldots, X_n$ be independent random variables such that $\mathbf{P}(X_i \in [a, b]) = b - a$ for all $0 \leq a < b \leq 1$, for all $i \in \{1, \ldots, n\}$. Let $f\colon [0, 1] \to \mathbb{R}$ be a continuous function we would like to integrate. Instead of integrating $f$ directly, we instead compute the quantity

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

Show that

$$\lim_{n\to\infty} \mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right) = \int_0^1 f(t)dt.$$

$$\lim_{n\to\infty} \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right) = 0.$$

That is, as $n$ becomes large, $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$ is a good estimate for $\int_0^1 f(t)dt$.

**Exercise 2.52** (Optional; Numerical Integration, Continued)**.** Let $\mathbf{P}$ denote the uniform probability law on $[0, 1]$, and let $X\colon [0, 1] \to \mathbb{R}$ be a random variable. This exercise discusses how to numerically compute expected values on a computer, as in Exercise 2.51. The procedure below is an example of **Monte Carlo simulation**.

Consider the function $X(t) := t$ for all $t \in [0, 1]$. We know that $\mathbf{E}X = 1/2$. To approximate $\mathbf{E}X$ with Matlab, we can use `sum(rand(1,1000))/1000`, which sums 1000 independent, random samples from the uniform probability law on $[0, 1]$, and averages them (by dividing by 1000). Enter the term `sum(rand(1,1000))/1000` a few times in the command line of Matlab, to get a few different results.

Consider the function $X(t) := t^2$ for all $t \in [0, 1]$. Using Matlab, approximate $\mathbf{E}X$ by averaging 1000 random samples from the uniform probability law on $[0, 1]$.

Now, let $\mathbf{P}$ denote the standard Gaussian probability law on $\mathbb{R}$, so that

$$\mathbf{E}X := \int_{-\infty}^{\infty} X(t)e^{-t^2/2}dt/\sqrt{2\pi}$$

for any function $X\colon \mathbb{R} \to \mathbb{R}$. Using the Matlab function `randn`, approximate $\mathbf{E}X$ for $X(t) := t$ and $X(t) := t^2$ by averaging 1000 random samples from the standard Gaussian probability law.

**Remark 2.53.** When Matlab or other computer programs generate "random numbers" using e.g. `rand` or `randn`, these numbers are not actually random or independent. These numbers are **pseudorandom**. That is, functions such as `rand` output numbers in a deterministic way, but these numbers behave as if they were random. All "random" numbers generated by computers are actually pseudorandom, and this includes slot machines at casinos, video games, etc. So, when using Monte Carlo simulation as we did above, we should be careful about interpreting our results, since it is generally impossible to take random samples from a probability law on a computer.

And, theoretically, if you knew enough about the random number generator that a slot machine is using, you could predict its output.

**Exercise 2.54.** Suppose you begin at the lower left corner of an $8 \times 8$ chess board. Every day, you are allowed to move either up or right to a consecutive board space (unless you are waiting). When you land on a new space, you have to wait a number of days specified by the number sitting on that board space, until you move again. The numbers on the board spaces appear below.

$$\begin{pmatrix} 1 & 2 & 2 & 1 & 3 & 2 & 6 & 0 \\ 4 & 7 & 3 & 2 & 4 & 8 & 3 & 4 \\ 3 & 4 & 4 & 4 & 5 & 5 & 4 & 2 \\ 4 & 7 & 5 & 3 & 4 & 4 & 5 & 5 \\ 4 & 5 & 4 & 2 & 3 & 3 & 7 & 3 \\ 4 & 6 & 6 & 4 & 3 & 4 & 3 & 2 \\ 5 & 4 & 6 & 3 & 4 & 3 & 4 & 1 \\ 0 & 3 & 6 & 2 & 7 & 2 & 7 & 5 \end{pmatrix}.$$

Your goal is to reach the top right corner of the chess board in the shortest amount of time. Find the path that takes the shortest amount of time, and also find the shortest amount of time that it takes to reach the top right corner. (Hint: Use recursion. That is, solve a more general problem. For *any* square on the board, find the least number of days it takes to reach that square starting from the bottom left corner, using only up and right moves. If you are still stuck, read a bit about dynamic programming.)

## 3. Martingales

### 3.1. Review of Conditional Expectation.

**Definition 3.1 (Conditional Expectation).** Let $X$ be a random variables on a sample space $\Omega$. Let $A \subseteq \Omega$ 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 $B \subseteq \Omega$. 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 3.2.** Suppose a random variable $X$ and a set $A \subseteq \Omega$ 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 2.37, $\mathbf{E}(X 1_A) = (\mathbf{E}X)(\mathbf{E}1_A) = \mathbf{P}(A)\mathbf{E}X$. 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 2.25

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

**Remark 3.3.** 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 3.4.** *Let $X, Y$ be random variables on a sample space $\Omega$. Let $A \subseteq \Omega$ and let $c \in \mathbb{R}$. If $X$ is a random variable such that $X = c$ on the set $A$, then*

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

*Proof.* Since $X = c$ on $A$, $XY1_A = cY1_A$, so $\mathbf{E}(XY1_A) = c\mathbf{E}(Y1_A)$. Dividing by $\mathbf{P}(A)$ concludes the Lemma. □

As stated in Definition 3.1, conditional expectation is itself an expected value with respect to a conditional probability. In particular, Jensen's inequality (Proposition 2.29) applies to conditional expectation

**Lemma 3.5 (Jensen's Inequality).** *Let $X$ be a random variable on a sample space $\Omega$. Let $A \subseteq \Omega$. Let $\phi\colon \mathbb{R} \to \mathbb{R}$ be convex. Then*

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

**Lemma 3.6.** *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 $B = \Omega$, 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 3.7 (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 3.8.** A stochastic process is a martingale if and only if it is both a submartingale and a supermartingale.

**Remark 3.9.** It follows from the Total Expectation Theorem that $\mathbf{E}(M_{n+1} - M_n) = 0$ for a martingale, for every $n \geq 0$. Consequently,

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

That is, a martingale does not change in expectation.

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

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

**Definition 3.10 (Conditional Expectation).** Suppose we have a partition of a sample space $\Omega$. That is, we have sets $A_1, \ldots, A_k \subseteq \Omega$ such that $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, \ldots, k\}$ with $i \neq j$, and $\cup_{i=1}^k A_i = \Omega$. Denote $\mathcal{A} = \{A_1, \ldots, A_k\}$. Define $\mathbf{E}(X|\mathcal{A})$ to be a random variable that takes the value $\mathbf{E}(X|A_i)$ on the set $A_i$. That is, $\mathbf{E}(X|\mathcal{A})$ is itself a function on the sample space $\Omega$.

**Remark 3.11.** Lemma 3.6 with $B = \Omega$ (i.e. the Total Expectation Theorem) can be rewritten as

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

Also, Lemma 3.4 says: if for each $1 \leq i \leq k$, $X$ is constant on $A_i$, then

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

**Exercise 3.12.** Let $\Omega = [0, 1]$. Let $\mathbf{P}$ be the uniform probability law on $\Omega$. Let $X : [0, 1] \to \mathbb{R}$ be a random variable such that $X(t) = t^2$ for all $t \in [0, 1]$. Let

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

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

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

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

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

**Exercise 3.13.** Let $X$ be a random variable with finite variance, and let $t \in \mathbb{R}$. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(t) = \mathbf{E}(X - t)^2$. Show that the function $f$ is uniquely minimized when $t = \mathbf{E}X$. That is, $f(\mathbf{E}X) < f(t)$ for all $t \in \mathbb{R}$ such that $t \neq \mathbf{E}X$. Put another way, setting $t$ to be the mean of $X$ minimizes the quantity $\mathbf{E}(X - t)^2$ uniquely.

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

**Exercise 3.14.** Let $\Omega = [0,1]$. Let $\mathbf{P}$ be the uniform probability law on $\Omega$. Let $X\colon [0,1] \to \mathbb{R}$ be a random variable such that $X(t) = t^2$ for all $t \in [0,1]$. For every integer $k > 1$, let $s = 2^{-k}$, let $\mathcal{A}_k := \{[0,s), [s,2s), [2s,3s), \ldots, [1-2s, 1-s), [1-s,1)\}$, and let $M_k := \mathbf{E}(X|\mathcal{A}_k)$. Show that the increments $M_2 - M_1, M_3 - M_2, \ldots$ are orthogonal in the following sense. For any $i, j \geq 1$ with $i \neq j$,

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

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

3.2. **Examples of Martingales.**

**Example 3.15** (**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 3.2,

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

**Example 3.16** (**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,

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

**Exercise 3.17** (**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.)

**Exercise 3.18** (MFE Sample Question)**.** For a two-period binomial model (i.e. the binomial option pricing model with $n = 2$), you are given:

(i) Each period is one year.
(ii) The current price for a nondividend-paying stock is 20.
(iii) $u = 1.2840$.
(iv) $d = 0.8607$.
(v) The continuously compounded risk-free interest rate is 5%. (That is, $1 + r = e^{.05}$.)

Calculate the price of an American call option on the stock with a strike price of 22. (That is, compute $(1+r)^{-2}\mathbf{E}\max(S_2 - 22, 0)$. Here $S_n$ is the stock price at time $n$.)

## 3.3. **Gambling Strategies.**

**Example 3.19.** 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 2.40) 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 3.15 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 $f_n : \mathbb{R}^n \to \mathbb{R}$ 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 3.20.** *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 3.21.** 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 3.22.** 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 3.20.* 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 3.4 implies

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

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

**Definition 3.23 (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 3.9, 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 3.24.** 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 3.15, $M_0, M_1, \ldots$ 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 3.25.** 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 3.26** (**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 3.20, $W_0, W_1, \ldots$ is a martingale. By definition of $H_m$,

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

□

**Theorem 3.27** (**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 $c \in \mathbb{R}$. Assume that $|M_{n\wedge T}| \leq c$ for all $n \geq 0$. Then $\mathbf{E}M_T = \mathbf{E}M_0$.*

*Proof.* From Theorem 3.26, 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

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

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

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

Therefore, for any $n \geq 1$,

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

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

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

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

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

**Remark 3.28.** The assumptions of the Optional Stopping Theorem cannot be abandoned, as shown in Example 3.24. 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 $\mathbf{E}M_T=1\neq 0=\mathbf{E}M_0$.

The following Lemma is proven in Math 171, so we omit the proof.

**Lemma 3.29.** *Let $X_0 := x_0$. Let $0<p<1$. Let $0\leq a<x_0<b$. 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 $T=\min\{n\geq 1\colon Y_n\in\{a,b\}\}$. Then there exists $0<\alpha<1$ and there exists an integer $j>0$ such that, for any*

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

**Example 3.30 (Gambler's Ruin).** 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 $(X_0, X_1, \ldots)$ 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 3.16 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 3.29 implies that $\mathbf{P}(T<\infty)=1$. From Theorem 3.27,

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

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 by Theorem 3.27,

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

(It follows from Lemma 3.29 that $\mathbf{P}(T<\infty)=1$.) Solving for $c$, we get

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

**Exercise 3.31.** 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 3.32.** 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 $T_0 = \min\{n \geq 1 \colon Y_n = 0\}$. If $X_0 = 1$, prove that $\mathbf{P}(T_0 = \infty) > 0$. Then, deduce that, if $X_0 = 0$, then $\mathbf{P}(T_0 = \infty) > 0$. That is, there is a positive probability that the biased random walk never returns to 0, even though it started at 0.

**Example 3.33.** 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 3.31 and the Optional Stopping Theorem to get $0 = \mathbf{E}(Y_T^2 - T)$, then using Example 3.30,

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

Strictly speaking, the Optional Stopping Theorem, Version 2, does not apply, since the martingale is not bounded. But Theorem 3.26 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 3.34.** 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 3.35.** 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 $\mathbf{P}(T < \infty) = 1$. 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 3.36 (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. 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 3.37.** 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 $0 < k \leq m$, if $X_0 = k$, then

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

(You can apply the Optional stopping theorem without verifying that the martingale is bounded.)

**Exercise 3.38.** Let $X_1, X_2, \ldots$ be independent random variables with $\mathbf{E}X_i = 0$ for every $i \geq 1$. Suppose there exists $\sigma > 0$ such that $\mathrm{Var}(X_i) = \sigma^2$ for all $i \geq 1$. For any $n \geq 1$, let $S_n = X_1 + \cdots + X_n$. Show that $S_n^2 - n\sigma^2$ is a martingale with respect to $X_1, X_2, \ldots$. (We let $X_0 = 0$.)

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

(You can apply the Optional stopping theorem without verifying that the martingale is bounded.)

## 4. Brownian Motion

4.1. **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 2.41), 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 2.18, 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 4.1** (**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 4.2** (**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.

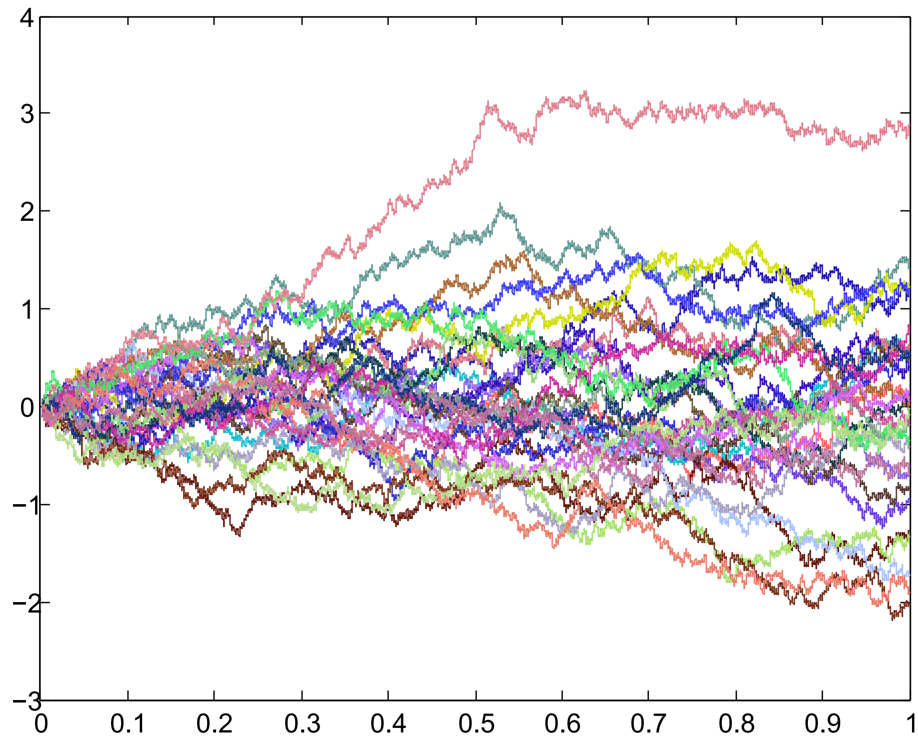


FIGURE 1. Sample Paths of Standard Brownian Motion. The horizontal axis is the $t$-axis.

In fact, by the independent increments property of Brownian Motion, if $t > s > 0$, if $x_1, \ldots, x_n \in \mathbb{R}$, and if $s > s_n > \cdots > s_1 > 0$, then

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

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

**Remark 4.3.** 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 3.27.

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 4.4** (**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 4.5.** 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 4.6.** 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 4.7.** 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 4.8.** *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,

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

$\square$

### 4.2. **Properties of Brownian Motion.**

**Proposition 4.9.** *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. $\square$

**Exercise 4.10.** 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 4.11.** *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 4.10 and the Optional Stopping Theorem, we get $0 = \mathbf{E}((B(T))^2 - T)$, then using Proposition 4.9,

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

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 3.33. □

**Exercise 4.12.** 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 4.13** (**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 4.14.** 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 4.15.**

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

*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 4.13. □

**Remark 4.16.** 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 4.17 (Brownian Motion with Drift).** Let $\sigma > 0$ and let $\mu \in \mathbb{R}$. A **standard Brownian motion with drift** $\mu$ and variance $\sigma^2$ 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 4.18.** Let $\{X(s)\}_{s\geq 0}$ be a standard Brownian motion with drift $\mu$ and variance $\sigma^2$. 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 3.30, 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 4.19.** 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 4.20.** *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 4.19, $e^{\alpha X(t)}$ 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 4.13, 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 4.21.** 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 4.22.** 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 4.23** (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 4.24** (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)')
```

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

**Exercise 4.25.** Let $\mu \in \mathbb{R}$ and let $\sigma > 0$. Let $X$ be a Gaussian random variable with mean $\mu$ and variance $\sigma^2$. Let $Y := e^X$. We then say $Y$ has a **lognormal distribution with parameters** $\mu$ and $\sigma^2$. 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 4.26** (**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 4.27** (**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

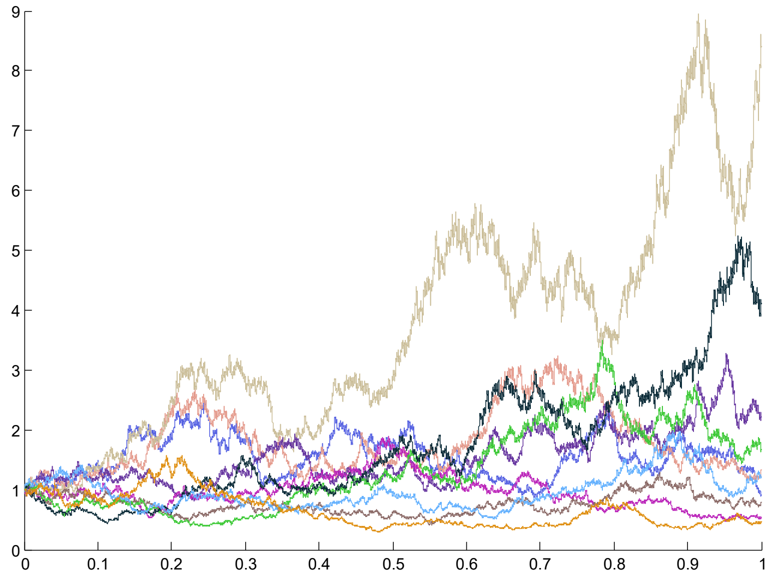


FIGURE 2. Sample Paths of Geometric Brownian Motion with $S_0 = 1$, $\sigma = 1$ and $\mu = 0$. The horizontal axis is the $t$-axis.

$\max(S(t) - k, 0)$. 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 4.19 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 4.28** (**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 $\mu t$. 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

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

□

**Remark 4.29.** Since $e^{\sigma B(t) + \mu t}$ has a log-normal density, we could have also used the formula

$$e^{-rt} \mathbf{E} \max(S(t) - k, 0) = e^{-rt} \int_0^{\infty} \max(S_0 z - k, 0) \frac{1}{\sigma\sqrt{2\pi t}} \frac{1}{z} e^{-\frac{(\log(z) - \mu t)^2}{2t\sigma^2}} dz.$$

**Remark 4.30.** In Exercise 3.17, 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 3.9, 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 4.31.** 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 4.32** (**Implied Volatility**)**.** Theorem 4.28 is often used in the following way. It is given that $r$ is an interest rate, $\sigma$ 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 4.28 is $\sigma$. We then choose $\sigma$ so that $c$ is equal to the actual observed price of the European call option. The $\sigma$ found in this way is referred to as **implied volatility**. (From Exercise 4.47, $c$ is an increasing function of $\sigma$, so a unique solution exists.)

If the volatility $\sigma$ of a fixed stock is known, and if Theorem 4.28 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 $\sigma$ has a U-shaped graph, as a function of $k$. That is, $\sigma$ does seem to depend on $k$. This graph of $\sigma$ as a function of $k$ is known as the **volatility smile**, and it is one way of demonstrating that Theorem 4.28 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 4.33.** It follows by property (i) of Brownian motion in Definition 4.1 that the sample paths of Geometric Brown motion are continuous, with probability 1. Since Theorem 4.28 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 4.34** (**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 4.35.** 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 4.36** (**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 4.37** (**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 4.38.** 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 4.39** (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 4.40** (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 4.41** (**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 4.37 no longer holds.

**Exercise 4.42.** 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 4.43.** 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 4.44.** 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

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

(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 4.15, 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?

4.4. **Black-Scholes Statistics.** Recall that in Theorem 4.28, 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 4.45** (**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 4.46.** 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 4.47.** 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 4.46.) (To make these exercises easier, write $c = \mathbf{E}(e^{-rt}\max(S(t)-k, 0))$, use the $S(t)$ formula from Exercise 4.46, 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 4.48** (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 4.49.** Suppose $f: \mathbb{R} \to \mathbb{R}$ 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 4.50** (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 4.51.** 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.)

## 5. 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 5.1** (**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 5.2.** 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 5.3** (**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 5.4.** 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

$$\begin{aligned}\mathbf{E}(\int_0^b B(t)dt)^2 &= \mathbf{E}\int_0^b B(t)dt\int_0^b B(s)ds = \int_0^b\int_0^b \mathbf{E}B(t)B(s)dsdt \\ &= \int_0^b\int_0^b \min(s,t)dsdt \qquad \text{, by Proposition 4.8} \\ &= 2\int_{t=0}^{t=b}\int_{s=0}^{s=t} sdsdt = \int_{t=0}^{t=b} t^2dt = \frac{1}{3}b^3.\end{aligned}$$

**Example 5.5.** 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 5.6.** Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Let $f: \mathbb{R} \to \mathbb{R}$. 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 5.7 (Stochastic Integral).** Let $f: \mathbb{R} \to \mathbb{R}$ 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 $f: \mathbb{R}^2 \to \mathbb{R}$ 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 3.20. 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 $f: \mathbb{R} \to \mathbb{R}$ 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 5.8.** 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 5.9.** 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 4.7.)

**Theorem 5.10 (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 5.11.** 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 5.12.** Theorem 5.10 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 5.13.** Using Itô's formula, write an expression for $\int_0^1 (B(s))^2 dB(s)$.

**Exercise 5.14.** 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 5.15** (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 5.16.** *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 5.10.* 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 $0 \leq i \leq n-1$, let $x := B(b(i+1)/n)$, let $y := B(bi/n)$, and sum over $i$ to get

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

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

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

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 5.16. 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 5.17** (**Itô's Formula, Version 2**). *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ have two continuous derivatives in each coordinate. We write $f = f(x, y)$, $x, y \in \mathbb{R}$. Then, with probability 1, for all $t \geq 0$,*

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

**Remark 5.18.** Theorem 5.17 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 5.19.** Let $f\colon \mathbb{R} \times [0, \infty) \to \mathbb{R}$. We write $f = f(x, t)$, where $(x, t) \in \mathbb{R} \times [0, \infty)$. Let $g\colon \mathbb{R} \to \mathbb{R}$ 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 5.20.** Let $f\colon \mathbb{R} \times [0, \infty) \to \mathbb{R}$. We write $f = f(x, t)$, where $(x, t) \in \mathbb{R} \times [0, \infty)$. Let $g\colon \mathbb{R} \to \mathbb{R}$ be a continuous and bounded function. We say that $f$ satisfies the one-dimensional **heat equation with forcing term** $h\colon \mathbb{R} \times [0, \infty) \to \mathbb{R}$ 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 5.21.** 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

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

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 4.29, where we take $F(z) := \max(S_0 z - k, 0)$.) Instead of differentiating $V$ directly, use the following strategy.

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

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

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 5.20.

5.1. **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.

**Definition 5.22** (**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 5.23.** *A solution of the Vasicek model can be written as*

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

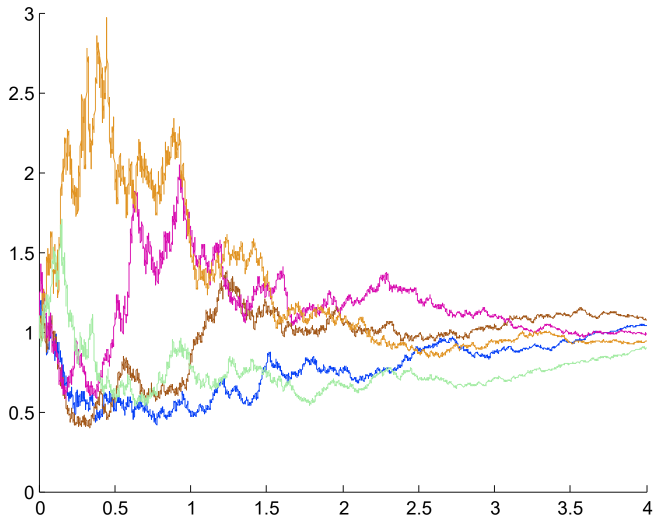


FIGURE 3. Sample Paths of the Vasicek model with $a = b = \sigma = f(0) = 1$. The horizontal axis is the $t$-axis.

*Proof.* As in the case $\sigma = 0$, we use the method of integrating factors. Using Itô's formula Version 2, Theorem 5.17, 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. □

**Exercise 5.24.** Let $a, b, \sigma > 0$. Let $f: \mathbb{R} \to \mathbb{R}$ 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 5.25.** Using a Monte Carlo simulation, plot several sample paths of the Vasicek stochastic differential equation, with $a = b = \sigma = f(0) = 1$.

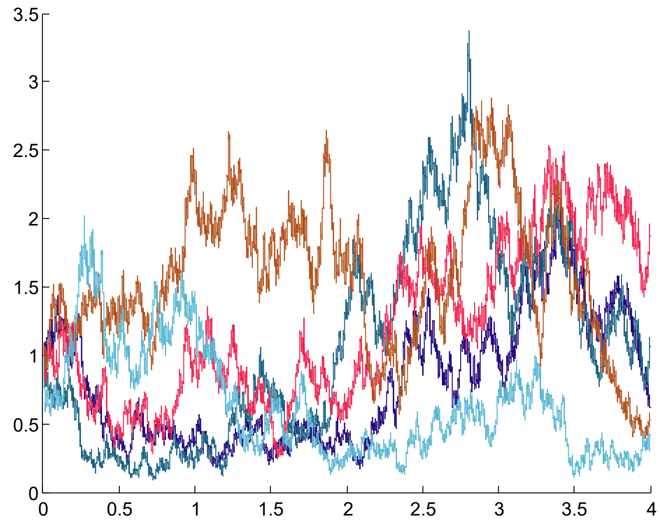


FIGURE 4. Sample Paths of the CIR model with $a = b = \sigma = f(0) = 1$. The horizontal axis is the $t$-axis.

**Exercise 5.26** (**Cox-Ingersoll-Ross (CIR) model**)**.** Let $a, b, \sigma > 0$. Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. The **Cox-Ingersoll-Ross model** models an interest rate as a (random) function $f: \mathbb{R} \to \mathbb{R}$ 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

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

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?

5.2. **Stochastic Heat Equation.** In Exercise 5.20, 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 5.27** (**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 5.28.** 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 5.20. 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 $s,t>0$:

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

5.3. **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 3.27, 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 3.27 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 5.29** (**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 3.27 and this new version of Itô's Lemma is that the $dt$ term in Theorem 3.27 might be different. We now describe this new term in Itô's Lemma.

**Definition 5.30** (**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 5.31.** Some books use a fairly deceptive and informal notation of "$(dY(t))^2$" instead of $d[Y]_t$ for the quadratic variation process.

**Remark 5.32.** In Lemma 5.16, 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 5.33.** *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

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

As $n \to \infty$, the first term converges in probability to $\int_0^b (\sigma(s, B(s)))^2 dB(s)$ by Lemma 5.16. 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, $[Y]_t = \int_0^t (\sigma(s, Y(s)))^2 ds$. $\square$

**Theorem 5.34** (**Itô's Formula, Version 3**). *Let $f : \mathbb{R} \to \mathbb{R}$ 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 5.35.** 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 5.10, as a special case of Theorem 5.34. Also, from Lemma 5.33,

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

*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. $\square$

Lemma 5.33 simplifies computations of Theorem 5.34.

**Theorem 5.36** (**Itô's Formula, Version 4**). *Let $f : \mathbb{R}^2 \to \mathbb{R}$ have two continuous derivatives in each coordinate. We write $f = f(x, y)$, $x, y \in \mathbb{R}$. 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 5.37.** 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 5.33, 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 5.38** (**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 5.39** (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 5.34 and Lemma 5.33, 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 5.38,

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

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

**Remark 5.40.** For more details on stochastic integration, see e.g. Section 7 of https://www.stat.berkeley.edu/~peres/bmbook.pdf

## 6. Appendix: Notation

Let $n, m$ be a positive integers. Let $A, B, B_1, \ldots, B_n$ be sets contained in a universal set $\Omega$.

$\mathbb{R}$ 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 $B$

$A \smallsetminus B := \{a \in A \colon a \notin B\}$

$A^c := \Omega \smallsetminus A$, the complement of $A$ in $\Omega$

$A \cap B$ denotes the intersection of $A$ and $B$

$A \cup B$ denotes the union of $A$ and $B$

$\mathbf{P}$ denotes a probability law on $\Omega$

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

$1_A \colon \Omega \to \{0, 1\}$, denotes the indicator function of $A$, so that

$$1_A(\omega) = \begin{cases} 1 & \text{, if } \omega \in A \\ 0 & \text{, otherwise.} \end{cases}$$

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 $\Omega$, so that $Y\colon\Omega\to\mathbb{R}$. Let $\mathbf{P}$ be a probability law on $\Omega$. Let $x\in\mathbb{R}$. Let $A\subseteq\Omega$. Let $Y$ be another discrete random variable

$$p_Y(x)=\mathbf{P}(Y=x)=\mathbf{P}(\{\omega\in\Omega\colon Y(\omega)=x\}),\ \forall\, x\in\mathbb{R}$$

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 $\Omega$, so that $Y,Y\colon\Omega\to\mathbb{R}$. Let $-\infty\leq a\leq b\leq\infty$, $-\infty\leq c\leq d\leq\infty$. Let $\mathbf{P}$ be a probability law on $\Omega$. Let $A\subseteq\Omega$.

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

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

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

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

$f_{Y|A}$ 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 $\Omega$.

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

$$F_Y(x)=\mathbf{P}(Y\leq x)=\mathbf{P}(\{\omega\in\Omega\colon Y(\omega)\leq x\})$$

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 $\{B(s)\}_{s\geq0}$ 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 $\{Y(s)\}_{s\geq 0}$

UCLA Department of Mathematics, Los Angeles, CA 90095-1555
*E-mail address*: heilman@math.ucla.edu