AN INTRODUCTION TO STOCHASTIC CALCULUS

VEDANT PATHAK

Abstract. This paper serves as a rigorous introduction to probability theory and stochastic calculus. It will first introduce basic facts about probability over uncountable spaces, and subsequently transition to a discussion of brownian motion and stochastic calculus.

Contents

1. Uncountable Probability Spaces

We begin with the notion of a sample space: an arbitrary set Ω whose elements – outcomes – represent all possible results of the same experiment. To motivate our introduction of measure-theoretic probability, suppose we conducted an experiment whose sample space is $[0,1] \subset \mathbb{R}$, paired with the uniform distribution. The traditional laws of probability give us

$$\sum_{x \in [0,1]} \mathbb{P}(x) = 1 \implies \mathbb{P}(x) = 0,$$

for all $x \in [0,1]$. Hence, if we wanted to calculate the probability that the outcome of our experiment lies in, say, $[1/2, 2/3]$, we would still get 0, as

$$\mathbb{P}\left(\left[\frac{1}{2}, \frac{2}{3}\right]\right) = \sum_{x \in [1/2, 2/3]} \mathbb{P}(x) = 0.$$

Under the uniform distribution, this makes no sense, as it should be 2/3. Instead, we should then define probabilities of subsets of our sample space – events – up front. To represent these events, we introduce the following definition:

Definition 1.1. A **σ-algebra** $\mathcal{F}$ is a collection of subsets of Ω satisfying:

Date: August 17^{th}, 2020.

(i) The empty set is an element of $\mathcal{F}$.

(ii) If $A \in \mathcal{F}$, then $A^c := \Omega \setminus A \in \mathcal{F}$.

(iii) If the sequence $(A_n)_{n\in\mathbb{N}}$ satisfies $A_n \in \mathcal{F}$ for all n, then

$$\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}.$$

Given a rigorous representation of events, we may now define a probability measure:

Definition 1.2. Let Ω be a nonempty set and let $\mathcal{F}$ be a σ-algebra of subsets of Ω. A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is called a **probability measure** if

(i) $\mathbb{P}(\Omega) = 1$

(ii) If $(A_n)_{n\in\mathbb{N}}$ is a sequence of disjoint sets in $\mathcal{F}$, then

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

Given the above definitions, we conclude with the following:

Definition 1.3. Let Ω be a nonempty set, let $\mathcal{F}$ be a σ-algebra of subsets of Ω, and let $\mathbb{P}$ be a probability measure over Ω with respect to $\mathcal{F}$. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**.

Example 1.4. Let $\Omega = [0, 1]$. We would like to construct a probability space from this outcome space in the most intuitive way possible. Let us first construct a σ-algebra $\mathcal{F}$. We begin by taking closed intervals $[a, b] \subset [0, 1]$ and adding them to $\mathcal{F}$. By setting $a = b$, we see that $\mathcal{F}$ must contain singletons. By taking the complements of closed intervals, $\mathcal{F}$ then contains half-open and half-closed intervals. Now, given $0 \le a < b \le 1$, we can see that $\mathcal{F}$ contains open intervals:

$$(a, b) = \bigcup_{n\in\mathbb{N}} \left(a + \frac{1}{n}, b - \frac{1}{n}\right) \in \mathcal{F}.$$

Using these elements, we perform all necessary unions, intersections, and complements in order to generate a σ-algebra. This σ-algebra is typically denoted by $\mathcal{B}([0, 1])$. To generate a probability measure with respect to $\mathcal{B}([0, 1])$, take $[a, b]$ and set $\mathbb{P}([a, b]) = b - a$. We immediately see that the measure of singletons is 0, and thus the measure of (a, b) is also $b - a$. Using the properties of a measure, we can fill in the gaps to obtain $\mathcal{L}$, the Lebesgue measure.

More generally, we can construct a σ-algebra $\mathcal{B}$ over $\mathbb{R}$ in the same way as above, called the Borel σ-algebra. Elements of $\mathcal{B}$ are called Borel sets. The resulting measure, although not a probability measure, is still denoted by $\mathcal{L}$ and referred to as the Lebesgue measure.

2. Random Variables and Expectation

We continue with the notion of a probability space as the set of results of an experiment. A random variable, then, can be considered as an interpretation of the consequence of performing the experiment once. Rigorously, we have the following definition:

Definition 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is called a **random variable** if, for any Borel subset of $\mathbb{R}$,

$$\{X \in B\} := \{\omega \in \Omega \mid X(\omega) \in B\}$$

is an element of $\mathcal{F}$. In other words, X is a random variable if it is $\mathcal{F}$*-measurable.*

Definition 2.2. The **distribution measure** of a random variable X is the probability measure μ_X over $\mathbb{R}$ given by

$$\mu_X(B) = \mathbb{P}\{X \in B\},$$

for any Borel subset $B \subseteq \mathbb{R}$. If there exists a function $f : \mathbb{R} \to \mathbb{R}$ such that

$$\mu_X(B) = \int_B f(x)dx$$

for every Borel subset $B \subseteq \mathbb{R}$, then we say that X has a **density**.

Example 2.3. We say that a random variable X has a normal distribution with mean μ and variance σ if it has a density

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}.$$

In this case, we say $X \sim N(\mu, \sigma)$.

To make the notion of distribution clearer, we can equivalently represent the distribution of a random variable with a cumulative distribution function (CDF) F, given by

$$F(x) = \mathbb{P}\{X \leq x\}.$$

Remark 2.4. It is easy to check that we know F if and only if we know μ_X for a random variable X.

Definition 2.5. The **moment generating function** for a random variable X is given by

$$\mathrm{MGF}_X(t) = \mathbb{E}\left[e^{tX}\right].$$

Remark 2.6. The moment generating function of a random variable uniquely determines its distribution. The proof is beyond the scope of this paper; however, the curious reader can refer to [4].

Now, for a random variable X in a countable probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the traditional notion of expectation, given by

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega),$$

reduces to a well-defined sum. In an uncountable probability space, however, this makes no sense at all. To account for this, we must use the Lebesgue integral.

Definition 2.7. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. The **expectation** of X is the quantity

$$\mathbb{E}[X] := \int_\Omega X(\omega)d\mathbb{P}(\omega).$$

Using this definition, we can prove some interesting properties of expectation.

Theorem 2.8. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable.*

(i) If X takes on finitely many values $x_1, ..., x_k$, then

$$\mathbb{E}[X] = \sum_{n=1}^{k} x_n \mathbb{P}\{X = x_n\}.$$

(ii) If Y is another random variable, and $X \leq Y$ almost surely, i.e.

$$\mathbb{P}\left\{\omega \in \Omega \mid X(\omega) \leq Y(\omega)\right\} = 1$$

then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

(iii) Given $\alpha, \beta \in \mathbb{R}$,

$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y].$$

Proof. See section 1.3 of [1]. □

The following theorem is key to proving useful facts about expectation.

Lemma 2.9 (Fatou)**.** *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables. Then,*

$$\mathbb{P}(\liminf_{n \to \infty} X_n) \leq \liminf_{n \to \infty} \mathbb{P}(X_n).$$

Proof. See section 5.4 of [2]. □

Theorem 2.10 (monotone convergence)**.** *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables converging almost surely to a random variable X, i.e.*

$$\mathbb{P}\left\{\omega \in \Omega \mid \lim_{n \to \infty} X_n(\omega) = X(\omega)\right\} = 1$$

If for all n, $X_n \leq X_{n+1}$ almost surely, then

$$\lim_{n \to \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Proof. Since $X_n \leq X_{n+1}$ almost surely for all n, it follows that $\mathbb{E}[X_n] \leq \mathbb{E}[X_{n+1}]$, for all n. Hence,

$$\limsup_{n \to \infty} \mathbb{E}[X_n] \leq \mathbb{E}[X].$$

By Fatou's lemma and properties of $\limsup$,

$$\mathbb{E}[X] \leq \liminf_{n \to \infty} \mathbb{E}[X_n] \leq \limsup_{n \to \infty} \mathbb{E}[X_n].$$

By the two inequalities above, it is clear that $\mathbb{E}[X] < \infty$, and $\mathbb{E}[X] = \lim_{n \to \infty} \mathbb{E}[X_n]$. □

With the above theorem, we can prove an important fact about computing expectations.

Theorem 2.11. *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let g be a Borel-measurable function on $\mathbb{R}$. If X has a density f, then*

$$\mathbb{E}\left[|g(X)|\right] = \int_{-\infty}^{\infty} |g(x)| f(x) dx.$$

If $\mathbb{E}\left[|g(X)|\right] < \infty$, then

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f(x) dx,$$

as well.

Proof. First, suppose $g(x) = \mathbb{I}_A(x)$ for some subset $A \subset \mathbb{R}$, i.e.

$$g(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Then,

$$\mathbb{E}[|g(X)|] = \mathbb{P}\{X \in A\} = \mu_X(A) = \int_A f(x)dx = \int_{-\infty}^{\infty} |g(x)|f(x)dx.$$

Now, suppose $g(x)$ is a simple function, i.e.

$$g(x) = \sum_{k=1}^{n} \alpha_k \mathbb{I}_{A_k},$$

for constants $\alpha_1, ..., \alpha_n \in \mathbb{R}$ and subsets $A_1, ..., A_n \subset \mathbb{R}$. Our claim clearly holds by the above proof for indicator functions, and by linearity of expectation. If $g(x)$ is a nonnegative Borel-measurable function, we may pick a sequence of simple functions converging to g, and apply the monotone convergence theorem to prove our claim. Finally, if $g(x)$ is a general Borel-measurable function, we may write $g(x)$ as $g^+(x) - g^-(x)$, where g^+ and g^- represent the positive and negative parts of g, respectively. Then, we can apply the result for nonnegative Borel-measurable functions to both g^+ and g^-, which proves our claim. □

Remark 2.12. The above proof technique is used often in proving facts about Lebesgue integrals – it is known as the standard machine, and it will be used later in this paper.

3. Conditional Expectation

Consider a stock price as a sequence of random variables (we will, of course, rigorously define this in the next section). When considering the price at time T, the prices at all times in $[0, T)$ are not redundant by any means. Each prior price generates new information with which one can narrow down all possibilities of the subsequent price. In particular, the event space we begin with changes with respect to time t. This motivates the following definitions:

Definition 3.1. Let $T > 0$, and let Ω be a nonempty set. Suppose that for all $t \in [0, T]$, there is a σ-algebra $\mathcal{F}(t)$. If $0 \leq s \leq t \leq T$ implies $\mathcal{F}(s) \subseteq \mathcal{F}(t)$, then we call $\{\mathcal{F}(t) \mid t \in [0, T]\}$ a **filtration**.

Definition 3.2. Let X be a random variable on a sample space Ω. Let $\sigma(X)$ denote the **σ-algebra generated by** X, i.e.

$$\sigma(X) := \left\{X^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\right\}.$$

Definition 3.3. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G}$ be another σ-algebra. We say that X is **$\mathcal{G}$-measurable** if $\sigma(X) \subseteq \mathcal{G}$.

Definition 3.4. Let $\mathcal{F}(t)$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If X_t is a collection of random variables indexed by time such that for each t, X_t is $\mathcal{F}(t)$-measurable, then we say that X_t is **adapted** to $\mathcal{F}(t)$.

Definition 3.5. Let X and Y be random variables. We say that X and Y are **independent** if $\sigma(X)$ and $\sigma(Y)$ are independent, i.e.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \forall A \in \sigma(X), B \in \sigma(Y).$$

To describe the correlation between two random variables, we use covariance.

Definition 3.6. Let X and Y be two random variables. The **covariance** of X and Y is given by

$$\begin{aligned}\mathrm{Cov}(X,Y) &:= \mathbb{E}\left[(X-\mathbb{E}[X])\,(Y-\mathbb{E}[Y])\right]\\ &= \mathbb{E}\left[XY-\mathbb{E}[Y]X-\mathbb{E}[X]Y+\mathbb{E}[X]\mathbb{E}[Y]\right]\\ &= \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

Note that we can get the **variance** from this definition:

$$\mathrm{Cov}(X,X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2=\mathrm{Var}(X).$$

We now have the tools required to discuss conditional expectation. Simply put, the conditional expectation of a random variable X with respect to a σ-algebra $\mathcal{G}$ is our best guess for X given the information in $\mathcal{G}$. More rigorously,

Definition 3.7. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G}$ be a sub-algebra of $\mathcal{F}$. The **conditional expectation of X with respect to $\mathcal{G}$**, denoted by $\mathbb{E}[X|\mathcal{G}]$, is the unique random variable satisfying

- $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable.
- $\mathbb{E}[X|\mathcal{G}]$ satisties the partial-averaging property:

$$\int_A \mathbb{E}[X|\mathcal{G}](\omega)d\mathbb{P}(\omega)=\int_A X(\omega)d\mathbb{P}(\omega),$$

  for all $A\in\mathcal{G}$.

Remark 3.8. If Y is another random variable, we denote the conditional expectation of X with respect to $\sigma(Y)$ by $\mathbb{E}[X|Y]$.

We can now deduce certain key properties of conditional expectation from the above definition.

Theorem 3.9. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ be a sub-algebra of $\mathcal{F}$. Then,*

(i) If X and Y are random variables and $\alpha, \beta \in \mathbb{R}$, then

$$\mathbb{E}[\alpha X+\beta Y|\mathcal{G}]=\alpha\mathbb{E}[X|\mathcal{G}]+\beta\mathbb{E}[Y|\mathcal{G}].$$

(ii) If X and Y are random variables, Y and XY are integrable, and X is $\mathcal{G}$-measurable, then

$$\mathbb{E}[XY|\mathcal{G}]=X\mathbb{E}[Y|\mathcal{G}].$$

(iii) If $\mathcal{H}$ is a sub-algebra of $\mathcal{G}$, and X is an integrable random variable, then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]=\mathbb{E}[X|\mathcal{H}].$$

(iv) If X is integrable and independent of $\mathcal{G}$, then

$$\mathbb{E}[X|\mathcal{G}]=\mathbb{E}[X].$$

Proof. We begin by proving (i). As a sum of $\mathcal{G}$-measurable random variables, the right-hand side of the desired equality is also $\mathcal{G}$-measurable. Now, take $A\in\mathcal{G}$. We have

$$\begin{aligned}\int_A \alpha\mathbb{E}[X|\mathcal{G}](\omega) + \beta\mathbb{E}[Y|\mathcal{G}](\omega)d\mathbb{P}(\omega) &= \alpha\int_A \mathbb{E}[X|\mathcal{G}](\omega)d\mathbb{P}(\omega) + \beta\int_A \mathbb{E}[Y|\mathcal{G}](\omega)d\mathbb{P}(\omega)\\ &= \alpha\int_A X(\omega)d\mathbb{P}(\omega) + \beta\int_A Y(\omega)d\mathbb{P}(\omega)\\ &= \int_A (\alpha X + \beta Y)(\omega)d\mathbb{P}(\omega).\end{aligned}$$

Since this identity is unique to $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}]$ by definition, the desired equality in (i) holds.

Now, suppose X and Y are random variables, Y and XY are integrable, and X is $\mathcal{G}$-measurable. Note that the right-hand side of (ii) is $\mathcal{G}$-measurable, being the product of $\mathcal{G}$-measurable random variables. We are left to show that

$$\int_A X\mathbb{E}[Y|\mathcal{G}](\omega)d\mathbb{P}(\omega) = \int_A (XY)(\omega)d\mathbb{P}(\omega).$$

First suppose $X = \mathbb{I}_B(\omega)$. Then,

$$\begin{aligned}\int_A X\mathbb{E}[Y|\mathcal{G}](\omega)d\mathbb{P}(\omega) &= \int_{A\cap B} \mathbb{E}[Y|\mathcal{G}](\omega)d\mathbb{P}(\omega)\\ &= \int_{A\cap B} Y(\omega)d\mathbb{P}(\omega)\\ &= \int_A \mathbb{I}_B Y(\omega)d\mathbb{P}(\omega) = \int_A (XY)(\omega)d\mathbb{P}(\omega).\end{aligned}$$

We may proceed via the standard machine (see remark 2.10) to show (ii).

Now, suppose X is integrable, and let $\mathcal{H}$ be a sub-algebra of $\mathcal{G}$. Since the right-hand side of (iii) is $\mathcal{H}$-measurable, it must be $\mathcal{G}$-measurable as well. Now, note that for any $A \in \mathcal{H}$,

$$\int_A \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}](\omega)d\mathbb{P}(\omega) = \int_A \mathbb{E}[X|\mathcal{G}](\omega)d\mathbb{P}(\omega) = \int_A X(\omega)d\mathbb{P}(\omega),$$

so we can use uniqueness of conditional expectation to prove (iii).

Finally, suppose X is integrable and independent of $\mathcal{G}$. Let us begin with the case where $X = \mathbb{I}_B(\omega)$. Then, for $A \in \mathcal{G}$,

$$\begin{aligned}\int_A \mathbb{E}[X|\mathcal{G}](\omega)d\mathbb{P}(\omega) &= \int_A X(\omega)d\mathbb{P}(\omega)\\ &= \mathbb{P}(A\cap B)\\ &= \mathbb{P}(A)\mathbb{P}(B)\\ &= \mathbb{P}(A)\mathbb{E}[X] = \int_A \mathbb{E}[X]d\mathbb{P}(\omega).\end{aligned}$$

We proceed via the standard machine to complete the proof. □

4. The Symmetric Random Walk

To ease into our study of brownian motion, we begin with a discussion of stochastic processes in the discrete case. Consider the infinite coin toss space $(\Omega, \mathcal{F}, \mathbb{P})$, and let

$$\omega := \prod_{k=1}^{\infty} \omega_k,$$

where $\omega_k \in \{H, T\}$. Consider random variables $(X_k)_{k\in\mathbb{N}}$, where

$$X_k = \begin{cases} 1 & \text{if } \omega_k = H \\ 0 & \text{if } \omega_k = T \end{cases}.$$

We can then define $M_0 = 0$ and

$$M_k = \sum_{n=0}^{k} X_k.$$

Definition 4.1. The process $(M_k)_{k\in\mathbb{N}}$ is called a **symmetric random walk**.

Intuitively, the symmetric random walk starts at zero, and at each integer time step, either increases or decreases by one, based on the flip of a coin. It is worth noting that the intervals of a symmetric random walk satisfy certain key properties.

Proposition 4.2. *Let $(M_k)_{k\in\mathbb{N}}$ be a symmetric random walk, and let m, n be integers such that $m < n$. Then,*

(i) $\mathbb{E}[M_n - M_m] = 0$.
(ii) $Var(M_n - M_m) = n - m$.
(iii) If k, ℓ are integers such that $k < \ell$ and $(k, \ell) \cap (m, n) = \varnothing$, then $M_n - M_m$ and $M_\ell - M_k$ are independent.

Proof. Linearity of expectation gives us the first property:

$$\begin{aligned}
\mathbb{E}\left[M_n - M_m\right] &= \mathbb{E}\left[\sum_{i=m+1}^{n} X_i\right] \\
&= \sum_{i=m+1}^{n} \mathbb{E}[X_i] \\
&= \sum_{i=m+1}^{n} \left(\mathbb{P}\{X_i = 1\} - \mathbb{P}\{X_i = -1\}\right) \\
&= \sum_{i=m+1}^{n} 0 = 0.
\end{aligned}$$

To prove the second property, we need only compute $\mathbb{E}[(M_n - M_m)^2]$. Again, we can use linearity of expectation.

$$\begin{aligned}
\mathbb{E}\left[(M_n - M_m)^2\right] &= \mathbb{E}\left[\left(\sum_{i=m+1}^{n} X_i\right)^2\right] \\
&= \sum_{i=m+1}^{n}\sum_{j=m+1}^{n} \mathbb{E}[X_i X_j] \\
&= \sum_{\substack{m+1 \le i,j \le n \\ i \neq j}} \mathbb{E}[X_i]\mathbb{E}[X_j] + \sum_{i=m+1}^{n} \mathbb{E}[X_i^2] \\
&= \sum_{i=m+1}^{n} \mathbb{E}[X_i^2] \\
&= \sum_{i=m+1}^{n} (\mathbb{P}\{X_i = 1\} + \mathbb{P}\{X_i = -1\}) \\
&= \sum_{i=m+1}^{n} 1 = n - m.
\end{aligned}$$

Note the the third equality holds due to the fact that for $i \neq j$, X_i and X_j are independent, since their values are determined by different coin tosses. Hence,

$$\mathrm{Var}(M_n - M_m) = \mathbb{E}\left[(M_n - M_m)^2\right] - \mathbb{E}\left[M_n - M_m\right]^2 = n - m,$$

as desired.

We can use this independence trick once again to prove the third property.

$$\begin{aligned}
\mathbb{E}[(M_n - M_m)(M_\ell - M_k)] &= \mathbb{E}\left[\sum_{i=m+1}^{n}\sum_{j=k+1}^{\ell} X_i X_j\right] \\
&= \sum_{i=m+1}^{n}\sum_{j=k+1}^{\ell} \mathbb{E}[X_i]\mathbb{E}[X_j] \\
&= 0.
\end{aligned}$$

This, however, is exactly equal to $\mathbb{E}[M_n - M_m]\mathbb{E}[M_\ell - M_k]$, which is what we wanted to show. □

The most important property of the symmetric random walk is that, if we are observing the value of M_n, but are only given information up to time $m < n$, then our best guess for M_n can only be M_m. Let us formalize this notion.

Definition 4.3. Let $(X_k)_{k \in \mathbb{N}}$ be a sequence of random variables. We say that (X_k) satisfies the **martingale property** (or, that (X_k) is a martingale) if, for $m \le n$

$$\mathbb{E}[X_n | \mathcal{F}_m] = X_m.$$

Proposition 4.4. *The symmetric random walk is a martingale.*

Proof. We proceed by computation. Given integers $m \leq n$,

$$\begin{aligned}\mathbb{E}[M_n|\mathcal{F}_m] &= \mathbb{E}[(M_n - M_m) + M_m|\mathcal{F}_m] \\ &= \mathbb{E}[M_n - M_m|\mathcal{F}_m] + \mathbb{E}[M_m|\mathcal{F}_m] \\ &= \mathbb{E}[M_n - M_m|\mathcal{F}_m] + M_m = M_m.\end{aligned}$$

□

Although this model somewhat clears up the intuition behind stock prices as a stochastic process, it offers a limited depiction, in that real stock prices do not change in discrete time. We need to find a way of narrowing down changes in both t and M_t. The intuitive next step, then, is to *scale* a random walk.

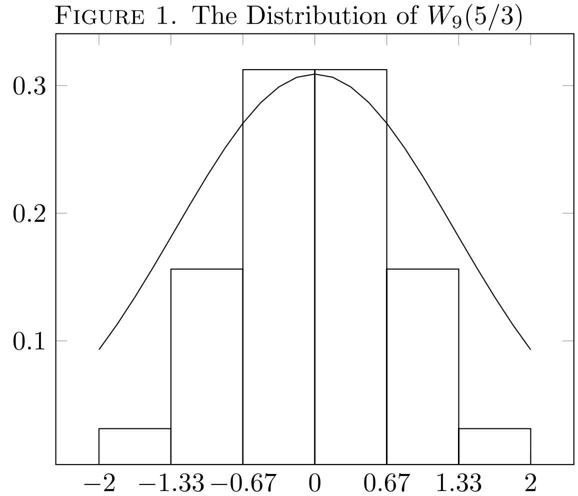
Definition 4.5. Let $(M_k)_{k\in\mathbb{N}}$ be a symmetric random walk. Fix $n \in \mathbb{N}$. For $t \in \mathbb{R}$, whenever $nt \in \mathbb{N}$, define

$$W_n(t) = \frac{1}{\sqrt{n}} M_{nt},$$

and for all other values of t, $W_n(t)$ can be computed via linear interpolation. The process $W_n(t)$ is called a **scaled symmetric random walk**.

It turns out that the scaled symmetric random walk also has independent increments, and is also a martingale. In fact, we may reuse the same proofs in propositions 4.2 and 4.4 to show this. As opposed to the standard symmetric random walk, the scaled symmetric random walk gives rise to a very interesting property regarding the distribution of $W_n(t)$ for any time t. Consider $W_9(5/3)$. This random variable can only take on values in $V = \{-5/3, -1, -1/3, 1/3, 1, 5/3\}$. What is important to note here is that the distribution of this random variable looks quite like a normal distribution with mean 0 and variance 5/3.

FIGURE 1. The Distribution of $W_9(5/3)$



Since we chose a rather small n the connection is less apparent. Nevertheless, this connection exists, and it is no coincidence. The central limit theorem actually tells us that the distribution of $M_n(t)$ *converges* to $N(0, t)$.

Theorem 4.6 (Central limit)**.** *As $n \to \infty$, $W_n(t) \sim N(0, t)$.*

Proof. By remark 2.6, it suffices to show that, as $n \to \infty$,

$$\text{MGF}_{W_n(t)}(u) \to \exp\left\{\frac{1}{2}u^2t^2\right\},$$

which is the moment generating function for $N(0,t)$. Fix $n \in \mathbb{N}$. We need only consider the case where nt is an integer. Noting that all X_i are independent of each other, we have

$$\begin{aligned}
\text{MGF}_{W_n(t)}(u) &= \mathbb{E}\left[\exp\left\{uW_n(t)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{\frac{u}{\sqrt{n}}\sum_{i=1}^{nt} X_i\right\}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{nt} \exp\left\{\frac{u}{\sqrt{n}}X_i\right\}\right] \\
&= \prod_{i=1}^{nt} \mathbb{E}\left[\exp\left\{\frac{u}{\sqrt{n}}X_i\right\}\right] \\
&= \prod_{i=1}^{nt}\left(\frac{1}{2}e^{\frac{u}{\sqrt{n}}} - \frac{1}{2}e^{\frac{-u}{\sqrt{n}}}\right) = \left(\frac{1}{2}e^{\frac{u}{\sqrt{n}}} - \frac{1}{2}e^{\frac{-u}{\sqrt{n}}}\right)^{nt}
\end{aligned}$$

The above is the main intuition behind the proof. What remains is to take the logarithm of $\text{MGF}_{W_n(t)}(u)$ and show, by iterative use of L'Hôpital's rule, that it converges to $u^2t^2/2$ – a process that is not as enlightening as the idea itself. However, the detail-oriented reader can refer to section 3.2 of [1]. □

As n tends to infinity, the limit B_t of $W_n(t)$ inherits all properties that $W_n(t)$ has. Furthermore, section 2.5 of [3] shows that this limit is, in fact, continuous. In particular, we have constructed a brownian motion.

5. Properties of Brownian Motion

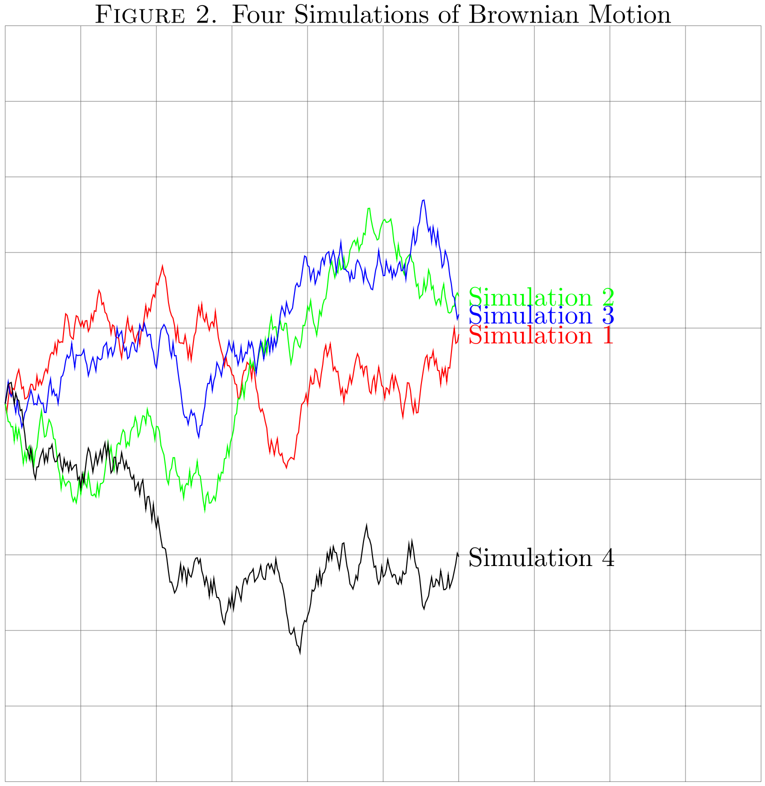
To summarize our construction in the previous chapter, we present the following definition:

Definition 5.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $B_t := B(t)$ be a continuous function such that $B_0 = 0$, that depends on $\omega \in \Omega$. We say that B_t is a **brownian motion** (or a **Wiener process**) if

- If $a < b < c < d$, then $B_b - B_a$ and $B_d - B_c$ are independent of each other.
- For all t, $B_t \sim N(0,t)$.

Note that, by construction, B_t is a martingale. After having answered the question about the continuity of brownian motion, the natural next step is to consider its differentiability. However, consider the simulations below, adapted from [6]:

FIGURE 2. Four Simulations of Brownian Motion



Intuitively speaking, none of these random curves appear to be *smooth*. In particular, these curves do not immediately appear to be differentiable. That's because they aren't – at least, almost surely.

Theorem 5.2. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space upon which* B_t *is a brownian motion. Then the function* $t \mapsto B_t$ *is almost surely nowhere differentiable.*

Proof. See section 37 of [5]. □

A reasonable question to ask upon seeing this result is – why does this even matter? The answer lies in quadratic variation, which we will explore in the next section.

6. Quadratic Variation

We begin this discussion with the following definition:

Definition 6.1. Fix $T \in \mathbb{R}$, and let $f(t)$ be a function defined on $[0, T]$ The **quadratic variation** of f up to time T is given by

$$[f, f](T) = \lim_{\|\Pi\| \to 0} \sum_{i=1}^{n-1} \left(f(x_{i+1}) - f(x_i)\right)^2,$$

where $\Pi = \{x_1, ..., x_n\}$ is a partition of $[0, T]$.

Suppose, in the above definition, that f is continuously differentiable. Then,

$$\begin{aligned}
[f,f](T) &= \lim_{\|\Pi\|\to 0} \sum_{i=1}^{n-1} \left(f(x_{i+1}) - f(x_i)\right)^2 \\
&= \lim_{\|\Pi\|\to 0} \sum_{i=1}^{n-1} f'^2(y_i)\left(x_{i+1} - x_i\right)^2 \\
&\leq \lim_{\|\Pi\|\to 0} \sum_{i=1}^{n-1} f'^2(y_i)\left(x_{i+1} - x_i\right) \cdot \|\Pi\| \\
&= \lim_{\|\Pi\|\to 0} \|\Pi\| \sum_{i=1}^{n-1} f'^2(y_i)\left(x_{i+1} - x_i\right) \\
&= \lim_{\|\Pi\|\to 0} \|\Pi\| \cdot \int_0^T f'^2(t)dt \\
&= \lim_{\|\Pi\|\to 0} \|\Pi\| \cdot I = 0,
\end{aligned}$$

where the last equality holds because f' is continuous and thus I is finite. This property, however, does not hold for a brownian motion.

Proposition 6.2. *Let B_t be a brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and fix $T > 0$. Then $[B_t, B_t](T) = T$.*

Proof. Define

$$Q_\Pi = \sum_{i=1}^{n-1} \left(B_{x_{i+1}} - B_{x_i}\right)^2,$$

where $\Pi = \{x_1 = 0, ..., x_n = T\}$ is a partition of $[0, T]$. It suffices to show that, as $\|\Pi\| \to 0$,

$$\mathbb{E}[Q_\Pi] \to T \quad \text{and} \quad \operatorname{Var}(Q_\Pi) \to 0.$$

The first limit is rather easy to show:

$$\mathbb{E}[Q_\Pi] = \sum_{i=1}^{n-1} \mathbb{E}\left[\left(B_{x_{i+1}} - B_{x_i}\right)^2\right] = \sum_{i=1}^{n-1} (x_{i+1} - x_i) = x_n = T.$$

The second equality was actually computed in proposition 4.2, and inherited by brownian motion as a limit of the scaled symmetric random walk. For the second limit, since all increments of B_t are independent, the variance of the sum becomes the sum of the individual variances. We then have

$$\begin{aligned}\operatorname{Var}(Q_\Pi) &= \sum_{i=1}^{n-1} \operatorname{Var}\left((B_{x_{i+1}} - B_{x_i})^2\right) \\ &= \sum_{i=1}^{n-1} \mathbb{E}\left[\left(B_{x_{i+1}} - B_{x_i}\right)^4\right] - \mathbb{E}\left[\left(B_{x_{i+1}} - B_{x_i}\right)^2\right]^2 \\ &= \sum_{i=1}^{n-1} \mathbb{E}\left[\left(B_{x_{i+1}} - B_{x_i}\right)^4\right] - (x_{i+1} - x_i)^2 \\ &= \sum_{i=1}^{n-1} 3(x_{i+1} - x_i)^2 - (x_{i+1} - x_i)^2 \\ &\le 2\,\|\Pi\| \sum_{i=1}^{n-1} (x_{i+1} - x_i) \\ &= 2T\,\|\Pi\| \to 0,\end{aligned}$$

as $\|\Pi\| \to 0$. □

What does this all build up to? Well, suppose one invested in a stock whose price was modeled by a deterministic, continuously differentiable function $S(t)$. If this person's position in the stock was given by $a(t)$, then the total returns up to a time T would be given by

$$\int_0^T a(t)dS(t).$$

Unfortunately, life is unfair and stock prices are neither deterministic, nor continuously differentiable. Suppose we, instead, used a model more true to reality – i.e. brownian motion – to model $S(t)$. We run into another problem: $dS(t)$ is not well-defined based on traditional calculus, since the variation of $S(t)$ is unbounded. If we want the above integral to be well-defined, then we need to deviate from traditional calculus.

7. Itô Calculus

The solution to our problem at the end of the previous section is the Itô integral, which we seek to define in this section.

Definition 7.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let B_t be a brownian motion defined on this space. Further, let $\mathcal{F}(t)$ be a filtration for B_t, and let $A(t)$ be a simple process adapted to $\mathcal{F}(t)$. In particular, for any time T, there exists a partition $\Pi = \{t_0 = 0, ..., t_n\}$ with $t_n < T$ such that for all $k \le n$, $A(t)$ is constant on $[t_k, t_{k+1})$. Then the **Itô integral** of $A(t)$ with respect to B_t is given by

$$\int_0^T A(t)dB(t) := \sum_{i=1}^{n-1} A(t_i)(B_{i+1} - B_i) + A(t_n)(B_T - B_{t_n}).$$

Immediately, we can derive some useful properties from this definition.

Proposition 7.2. *Let B_t be a brownian motion, and let $\mathcal{F}(t)$ be a filtration for B_t. Suppose $A(t)$ is a simple process adapted to $\mathcal{F}(t)$. Then,* *(i) If $C(t)$ is another simple process adapted to $\mathcal{F}(t)$ and $\alpha, \beta \in \mathbb{R}$ are constants, then*

$$\int_0^T \alpha A(t) + \beta C(t) dB(t) = \alpha \int_0^T A(t)dB(t) + \beta \int_0^T C(t)dB(t).$$

(ii) If $S \leq t$, then

$$\mathbb{E}\left[\int_0^t A(t)dB(t)\middle|\mathcal{F}(S)\right] = \int_0^S A(t)dB(t).$$

(iii) The Itô integral satisfies

$$\mathbb{E}\left[\left(\int_0^T A(t)dB(t)\right)^2\right] = \mathbb{E}\left[\int_0^T A^2(t)dt\right]$$

Proof. The first claim simply holds by the definition of the Itô integral – one only needs to break up the sum, which is trivial.

To prove the second property, we begin by expanding the Itô integral. Suppose $S < t$, and let $\Pi = \{t_1 = 0, ..., t_n\}$ be a partition of $[0, t]$ such that for all i, $A(t)$ is constant on $[t_i, t_{i+1})$, and $t_n < t$. Then, if $t_{k-1} < S < t_k$,

$$\begin{aligned}\mathbb{E}\left[\int_0^t A(t)dB(t)\middle|\mathcal{F}(S)\right] &= \int_0^S A(t)dB(t) + \mathbb{E}\left[A(t_k)(B_{t_k} - B_s)|\mathcal{F}(S)\right] \\ &+ \sum_{i=k}^{n-1} \mathbb{E}\left[A(t_i)(B_{t_{i+1}} - B_{t_i})\middle|\mathcal{F}(S)\right] \\ &+ \mathbb{E}\left[A(t_k)(B_{t_k} - B_t)|\mathcal{F}(S)\right].\end{aligned} \tag{7.3}$$

We are left to show that the three expectations in the equation are all zero. Beginning with the first,

$$\begin{aligned}\mathbb{E}\left[A(t_k)(B_{t_k} - B_s)|\mathcal{F}(S)\right] &= A(S)\mathbb{E}\left[B_{t_k} - B_s|\mathcal{F}(S)\right] \\ &= A(S)\left(\mathbb{E}\left[B_{t_k}|\mathcal{F}(S)\right] - B_s\right) \\ &= A(S)(B_s - B_s) = 0.\end{aligned}$$

We need only show that the third expectation in equation 7.3 is zero. The second expectation can be shown to be zero by the same process. We have

$$\begin{aligned}\mathbb{E}\left[A(t_k)(B_{t_k} - B_t)|\mathcal{F}(S)\right] &= \mathbb{E}\left[\mathbb{E}\left[A(t_n)(B_t - B_{t_n})|\mathcal{F}(t_n)\right]|\mathcal{F}(S)\right] \\ &= \mathbb{E}\left[A(t_n)\left(\mathbb{E}\left[B_t|\mathcal{F}(t_n)\right] - B_{t_n}\right)|\mathcal{F}(S)\right] \\ &= \mathbb{E}\left[A(t_n)\left(B_{t_n} - B_{t_n}\right)|\mathcal{F}(S)\right] \\ &= \mathbb{E}\left[0|\mathcal{F}(S)\right] = 0,\end{aligned}$$

and thus the Itô integral is a martingale.

Finally, we are left with the last property, known informally as the *Itô isometry.* Take the partition Π as above, and for the sake of convenience fix $t_{n+1} := T$. Then,

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T A(t)dB(t)\right)^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^n A(t_i)(B_{t_{i+1}} - B_{t_i})\right)^2 t\right] \\ &= \sum_{0 \leq i,j \leq n} \mathbb{E}\left[A(t_i)A(t_j)(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right].\end{aligned}$$

Observe the case where $i \neq j$. In particular, suppose $i < j$. Then,

$$\mathbb{E}\left[A(t_i)A(t_j)(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right] = \mathbb{E}\left[\mathbb{E}\left[A(t_i)A(t_j)(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\middle|\mathcal{F}(t_j)\right]\right]. \tag{7.4}$$

Almost every random variable in this expression is $\mathcal{F}(t_j)$-measurable, with the exception of $B_{t_{j+1}}$. In this case, conditional expectation shows us that

$$\mathbb{E}[B_{t_{j+1}}|\mathcal{F}(t_j)] = B_{t_j},$$

so equation 7.4 evaluates to zero. We are left to consider the case where $i = j$. Then,

$$\begin{aligned}\mathbb{E}\left[A(t_i)A(t_j)(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right] &= \mathbb{E}[A^2(t_i)(B_{t_{i+1}} - B_{t_i})^2] \\ &= \mathbb{E}\left[A^2(t_i)\mathbb{E}\left[(B_{t_{i+1}} - B_{t_i})^2\middle|\mathcal{F}(t_i)\right]\right] \\ &= \mathbb{E}\left[A^2(t_i)(t_{i+1} - t_i)\right].\end{aligned}$$

Putting these two results together, we see that

$$\mathbb{E}\left[\left(\int_0^T A(t)dB(t)\right)^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[A^2(t_i)(t_{i+1} - t_i)\right] = \mathbb{E}\left[\int_0^T A^2(t)dt\right],$$

where the last equality holds by linearity of expectation, yielding the definition of the Itô integral. □

Now, we proceed to extend the Itô integral to general adapted processes – not just simple processes.

Definition 7.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space upon which B_t is a brownian motion. If $\mathcal{F}(t)$ is a filtration for B_t and $A(t)$ is adapted to $\mathcal{F}(t)$, then we may pick a sequence of simple processes $A_n(t)$ converging to $A(t)$, and construct the well-defined **Itô integral** of $A(t)$ as

$$\int_0^T A(t)dB(t) = \lim_{n\to\infty} \int_0^T A_n(t)dB(t).$$

As a limit of the Itô integral of simple processes, the Itô integral for a general adapted process inherits all the properties of the former.

At this point, it should be relatively clear that stochastic calculus is similar to ordinary calculus, at least with respect to what the results for integration have shown us. We have seen that differentiation of brownian motion itself is not possible; however, Itô's formula presents a profound result for functions that *depend* on time and a brownian motion, similar to the fundamental theorem of calculus. We shall now formally state and prove this theorem.

Theorem 7.6 (Itô's Formula)**.** *Suppose $f(t, x)$ is C^1 in t and C^2 in x. Then if B_t is a brownian motion,*

$$f(T, B_T) - f(0, B_0) = \int_0^T f_t(t, B_t)dt + \int_0^T f_x(t, B_t)dB(t) + \frac{1}{2}\int_0^T f_{xx}(t, B_t)dt.$$

Proof. Let $\{t_1 = 0, ..., t_n = T\}$ be a partition of $[0, T]$. Then, we can write the Taylor expansion of f as a telescoping sum:

$$\begin{aligned} f(T, B_T) - f(0, B_0) &= \sum_{i=1}^{n-1} f(t_{i+1}, B_{t_{i+1}}) - f(t_i, B_{t_i}) \\ &= \sum_{i=1}^{n-1} f_x(t_i, B_{t_i})\Delta_x + \frac{1}{2}\sum_{i=1}^{n-1} f_{xx}(t_i, B_{t_i})\Delta_x^2 + \sum_{i=1}^{n-1} f_t(t_i, B_{t_i})\Delta_t, \end{aligned}$$

where $\Delta_x = \left(B_{t_{i+1}} - B_{t_i}\right)$ and $\Delta_t = (t_{i+1} - t_i)$. Note that there are higher-order terms in this sum, both in terms of t and B_t; however, as $\|\Pi\| \to 0$, these terms vanish. Now, as $\|\Pi\| \to 0$, the first term in this sum approaches

$$\int_0^T f_x(t, B_t)dB(t),$$

simply by definition. Now, observe the second sum. In particular, consider Δ_x^2 as $\|\Pi\| \to \infty$. As previously shown,

$$\mathbb{E}[\Delta_t^2] = \mathbb{E}\left[\left(B_{t_{i+1}} - B_{t_i}\right)^2\right] = t_{i+1} - t_i.$$

In addition, $\mathrm{Var}(\Delta_t^2) \to 0$ as $\|\Pi\| \to 0$, so $\Delta_2^t \to t_{i+1} - t_i$. As a result,

$$\frac{1}{2}\sum_{i=1}^{n-1} f_{xx}(t_i, B_{t_i})\Delta_x^2 \to \frac{1}{2}\sum_{i=1}^{n-1} f_{xx}(t_i, B_{t_i})(t_{i+1} - t_i) \to \frac{1}{2}\int_0^T f_{xx}(t, B_t)dt.$$

We may treat the third sum similarly to how we treated the first – this time, rather than the Itô integral, it gives us the Riemann integral as $\|\Pi\| \to 0$. Putting these results together, we see that

$$\begin{aligned} &\sum_{i=1}^{n-1} f_x(t_i, B_{t_i})\Delta_x + \frac{1}{2}\sum_{i=1}^{n-1} f_{xx}(t_i, B_{t_i})\Delta_x^2 + \sum_{i=1}^{n-1} f_t(t_i, B_{t_i})\Delta_t \to \\ &\qquad \int_0^T f_x(t, B_t)dB(t) + \frac{1}{2}\int_0^T f_{xx}(t, B_t)dt + \int_0^T f_t(t, B_t)dt \end{aligned}$$

as $\|\Pi\| \to 0$, which is the desired result. □

Remark 7.7. It is often more convenient to write the above result in differential form, i.e.

$$df(t, B_t) = f_x(t, B_t)dB(t) + \left(\frac{1}{2}f_{xx}(t, B_t) + f_t(t, B_t)\right)dt.$$

Acknowledgements

I would like to thank my mentor Nixia Chen for aiding me through the writing process, as well as answering any questions I had regarding probability theory. I would also like to thank professor Gregory Lawler for presenting an incredibly interesting series of lectures on random walks. Finally, I would like to thank professor Peter May for organizing this year's REU despite the numerous setbacks due to these uncertain times.

References

[1] Steven E. Shreve. Stochastic Calculus for Finance II, Continuous Time Models. Springer; 1st ed. 2004. Corr. 2nd printing 2010 edition (19 Jun. 2008).
[2] David Williams. Probability With Martingales. Cambridge University Press 1991.
[3] Gregory F. Lawler. Stochastic Calculus: An Introduction With Applications. Gregory F. Lawler 2014.
[4] J. H. Curtiss. The Annals of Mathematical Statistics Vol. 13, No. 4 (Dec., 1942). Institute of Mathematical Statistics 1942.
[5] Patrick Billingsley. Probability and Measure. John Wiley & Sons, Inc. 1995.
[6] https://tex.stackexchange.com/a/59927