Probability and Stochastic Processes for PROMYS Counselors

Adam B. Block

Contents

1 The Basics of Probability Theory

1.1 The Measure Theory Background

While measure theory is hardly pleasant, it is an unfortunate prerequisite to the subsequent discussion. We briefly review some relevant concepts. A thorough treatment of the topic can be found in [Dur10, Chu00] and any other introduction to probability theory.

Definition 1. A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a set, $\mathcal{F}$ is a σ-algebra, and $\mathbb{P} : \mathcal{F} \to [0,1]$ is a measure that takes full measure on the entire set, i.e., $\mathbb{P}(\Omega) = 1$. A random variable with state space $(S, \mathcal{S})$ with law $\mathbb{P}$ is a measurable map $X : \Omega \to S$.

The most important probability measures for the first few lectures have $\Omega \subset \mathbb{R}$. Then we may define a distribution function

Definition 2. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\Omega \subset \mathbb{R}$ and $\mathcal{F} \subset \mathcal{B}(\mathbb{R})$, we define the distribution function $F : \mathbb{R} \to [0,1]$ such that if X is a random variable with law $\mathbb{P}$, then $\mathbb{P}(X \leq x) = F(x)$.

Several properties of these functions are immediately apparent. For instance, F is a nondecreasing, right-continuous function such that

$$\lim_{x \downarrow -\infty} F(x) = 0 \qquad \lim_{x \uparrow \infty} F(x) = 1$$

We are now ready for some examples.

Example 3. Let δ_0 denote the point mass at 0, i.e., $\Omega = \{0\}$, $\mathcal{F} = \wp(\Omega)$ and $\mathbb{P}$ is the unique probability measure on $\mathcal{F}$. The distribution function is

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

Similarly, if we take any point $x = (x_1, \ldots, x_n)$ in an $(n+1)$ simplex, and we take any n real numbers $a_1, \ldots, a_n$, we may consider the probability measure that places mass at point a_i with probability x_i.

An important special case of the previous example is a Bernoulli random variable. We say that X is Bernoulli with parameter p if $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$.

Example 4. We call the lebesgue measure on $[0,1]$ the uniform probability distribution on the unit interval. It has distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

More generally, for any borel subset of $\mathbb{R}$ (that is not Lebesgue measure 0), we may define the uniform probability distribution on it by scaling the Lebesgue measure so that the total mass is 1.

The next fundamental concept is that of expectation.

Definition 5. First let X be a random variable with law $\mathbb{P}$ such that $\mathbb{P}(X < 0) = 0$. Then we define $\mathbb{E}X = \int X d\mathbb{P}$. In general, if at least one of $\mathbb{E}X^+$ or $\mathbb{E}X^-$ is finite, then we define $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$.

Clearly, as taking expectation is just integration, the standard integral properties carry over, such as linearity, monotinicity, dominated convergence, and Fatou. Another important property that carries over *mutatis mutandis* is

Theorem 6 (Fubini). *Let μ, ν measures on Ω, Ω' and let $\mu \otimes \nu$ denote the product measure. Let f a measurable function on $\Omega \times \Omega'$ such that $f \geq 0$ or $\int |f| d\mu \otimes \nu < \infty$. Then*

$$\int_\Omega \int_{\Omega'} f d\nu d\mu = \int_{\Omega'} \int_\Omega f d\mu d\nu = \int_{\Omega \times \Omega'} f d\mu \otimes \nu$$

With expected value, we have a powerful new way to create probability measures. Let μ be a measure on some probability space Ω and suppose that $f \geq 0$ is a measurable function such that $\int_\Omega f d\mu = 1$. Then consider the function $\mathbb{P} : \mathcal{F} \to [0,1]$ given by $\mathbb{P}(A) = \int_A f d\mu$. It is easy to check that this is a probability measure.

Example 7. Let $\Omega = \{0,1\}$ and suppose $\mu(\{1\}) = p$. If $f(0) = \frac{1-p'}{1-p}$ and $f(1) = \frac{p'}{p}$ then $\mathbb{P}$ is the law of a Bernoulli p' r.v..

Example 8. The best examples of the above method are what are sometimes called "probability density functions" in undergraduate probability classes. In this case, μ is the lebesgue measure and one can take any nonnegative function that integrates to 1. The most important example of this is the Gaussian. We say that X is Gaussian of mean θ and variance σ^2 if

$$\mathbb{P}(X \in A) = \int_A \left(\sqrt{2\pi\sigma^2}\right)^{-\frac{1}{2}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) \mu(dx)$$

Other examples include the exponential, when $f(x) = e^{-x}\mathbb{1}(x \geq 0)$, the Cauchy distribution when $f(x) = \left(\pi(1+x^2)\right)^{-1}$, and many, many more.

A natural question arises: if $\mathbb{P}$ and μ are probability measures on the same space, when can we find a function f as in the above discussion? One thing to note about the above construction in relating $\mathbb{P}$ to μ is that if $A \in \mathcal{F}$ such that $\mu(A) = 0$, then clearly $\mathbb{P}(A) = 0$ as well. This motivates

Definition 9. Let μ, ν be two measures on the same σ-algebra. We say $\mu \ll \nu$, that μ is absolutely continuous with respect to ν if for all sets A such that $\nu(A) = 0$, then we have $\mu(A) = 0$ as well.

Thus we see that $\mathbb{P} \ll \mu$ is certainly a necessary condition for there to exist such an f. In fact, it is sufficient as well, which is the content of the following theorem, whose proof can be found in [Dur10].

Theorem 10 (Radon-Nikodym)**.** *Let $\mathbb{P} \ll \nu$ be two probability measures on the same space. Then there is a μ-almost everywhere unique measurable function, f, called the Radon-Nikodym derivative, such that for all measurable sets A,*

$$\mathbb{P}(A) = \int_A f d\mu$$

Remark 11. The Radon-Nikodym derivative is often denoted $\frac{d\mathbb{P}}{d\mu}$ and this is not mere coincidence. One can show that many of the laws of normal calculus, such as the chain rule, hold here too.

We now introduce the concept that separates probability from measure theory: that of independence.

Definition 12. A collection of events $\{A_i\}$ is independent if for any finite collection of events $A_{i_1}, \ldots, A_{i_n}$, we have

$$\mathbb{P}\left(\bigcap_j A_{i_j}\right) = \prod_j \mathbb{P}\left(A_{i_j}\right)$$

A collection of σ-algebras is independent if any collection formed by taking an event from each algebra is independent.

Note that the requirement "any finite collection" is stronger then just checking any pair (this is called pair-wise independence). To see this, let X, Y, Z be iid Bernoulli $\frac{1}{2}$ r.v.s and consider the following events

$$A_1 = \{X = Y\} \qquad A_2 = \{Y = Z\} \qquad A_3 = \{X = Z\}$$

Then $\mathbb{P}(A_i \cap A_j) = \frac{1}{4} = \mathbb{P}(A_i)\mathbb{P}(A_j)$, but

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$$

Independence is crucial to the understanding of probability, but it is really a very simple concept. One can think of X, Y being independent as meaning that information about X allows no inference on Y, an intuition that will be explained somewhat in the sequel.

We conclude this sprint through some prerequisites with the following classical result:

Proposition 13 (Markov)**.** *Let $\varphi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ and let $A \in \mathcal{B}(\mathbb{R})$. Let*

$$i_A = \inf_{a \in A} \varphi(a)$$

Then $i_A \mathbb{P}(A) \leq \mathbb{E}\left[\varphi(X)\mathbb{1}_A\right] \leq \mathbb{E}\varphi(X)$.

Proof. By definition we have $i_A \mathbb{1}_A \leq \varphi(X)\mathbb{1}_A$. Taking expected values yields the first inequality. The second follows because $\varphi \geq 0$. ■

Remark 14. This is most commonly used to provide moment bounds. Let $\varphi(x) = |x|^p$. Then applying Proposition 13 to $X - \mathbb{E}X$, we have that for all $\varepsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon) \leq \frac{\mathbb{E}[|X - \mathbb{E}X|^p]}{\varepsilon^p}$$

Where the numerator on the right is called the p^{th} central moment. When $p = 2$ we get the classic Chebyshev inequality that says

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon) \leq \frac{\operatorname{Var} X}{\varepsilon^2}$$

1.2 Types of Convergence

One of the important series of definitions in a first course in probability is the types of convergence that one can witness. Recalling that a random variable X is a function, the easiest type of convergence might be pointwise convergence, i.e., $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$. As we shall see, however, in probability, one often cares little about events of probability zero and often we have random variables only defined up to a null set. Thus this is too stringent a definition. It can be replaced with

Definition 15. If (X_n) is a sequence of random variables we say that $X_n \to X$ $\mathbb{P}$-almost surely if

$$\mathbb{P}\left(\{\omega \in \Omega | X_n(\omega) \to X(\omega)\}\right) = 1$$

While this definition is natural, and often useful, it is at times too stringent. Thus we have

Definition 16. We say that a sequence of random variables converges in probability, $X_n \xrightarrow{p} X$ if for all $\varepsilon > 0$,

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

It is immediate that the first type of convergence implies the second, but the implication is one-way.

Example 17. Let X_n be standard Gaussians. Let $Y_n = \frac{X_n}{n}$. Then $Y_n \xrightarrow{p} 0$. To see this, note that a computation gives $\operatorname{Var} Y_n = \frac{1}{n^2}$ and so Proposition 13 will immediately yield the desired result. In this case, we have almost sure convergence as well, which can be proven using Lemma 27 in the sequel.

If instead of considering pointwise convergence in our analogy with functional analysis, we had considered L^p convergence, we arrive at

Definition 18. A sequence of random variables, (X_n) converges to X in L^p for some $p > 0$ if $\mathbb{E}\left[|X_n - X|^p\right] \to 0$.

This is a useful concept as it is often easy to bound this expected value, but it then becomes subordinate to convergence in probability by the following proposition

Proposition 19. *Suppose there is a $p > 0$ such that $X_n \to X$ in L^p. Then $X_n \xrightarrow{p} X$.*

Proof. This follows directly from Proposition 13. To see this, note that for any $\varepsilon > 0$, we may apply Proposition 13 with $\varphi(x) = |x|^p$ and get

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}\left[|X_n - X|^p\right]}{\varepsilon^p} \to 0$$

as desired. ■

The final type of convergense we need is convergence in distribution:

Definition 20. Let (X_n) a sequence of random variables and let F_n be the distribution functions of their laws. We say that $X_n \xrightarrow{d} X$, or X_n converges to X in distribution if X follows a distribution F and such that for all $x \in \mathbb{R}$ such that F is continuous at x, $F_n(x) \to F(x)$.

This type of convergence is much weaker because it says nothing about the relationship between the random variables themselves; all the convergence treats is the laws of the random variables.

Example 21. Suppose that $X_i \sim N(\mu_i, \sigma^2)$ are independent and suppose that $n^{-\frac{1}{2}} \sum_1^n \mu_i$ converges to some μ as $n \to \infty$. Then $n^{-\frac{1}{2}} \sum_1^n X_i \xrightarrow{d} X$ where $X \sim N(\mu, \sigma^2)$. This is because a simple computation shows that if Z_1, Z_2 are independent Gaussian of means μ_1, μ_2 and variances σ_1, σ_2 then $Z_1 + Z_2$ is gaussian of mean $\mu_1 + \mu_2$ and variance $\sigma_1 + \sigma_2$. The scaling and the hypothesis yield the result.

The most important example of convergence in distribution is the second pillar of classical probability.

Theorem 22 (Central Limit Theorem)**.** *Let $X_1, \dots$ be iid random variables with mean 0 and variance 1. Then $n^{-\frac{1}{2}} \sum_1^n X_i \xrightarrow{d} Z$ where Z is gaussian of mean 0 and variance 1.*

The proof of this theorem was one of the great advances in early 20th century mathematics. While we will not discuss it much here, the concept underlies much of what we do, especially our (optional) construction of Brownian motion and why we care so much about the Gaussian distribution. Note also that the independence condition can be somewhat weakened as explored in, e.g., [Bil61, Bro71].

The general idea for the proof of Theorem 22 lies in the notion of a characteristic function. Given two random variables X, Y one might wonder if there is an easy way to tell if they have the same distribution. Clearly if they have the same distribution then $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$ for all k such that these are well defined. If we restrict to the class of distributions that have moments of all orders, then we might wonder if the fact that $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$ for all k implies that $X \overset{d}{=} Y$. Unfortunately, this is not the case (see [Dur10, vdV12]). However, if these moments grow slowly in k then this is indeed the case. In fact we have

Proposition 23. *Let φ be the moment generating function for X, i.e., if there exists some open interval of 0 such that the following is finite, then we define*

$$\varphi(\lambda) = \mathbb{E}\left[e^{\lambda X}\right] = \sum \frac{\lambda^k}{k!} \mathbb{E}\left[X^k\right]$$

If $\varphi_X = \varphi_Y$ as functions then $X \overset{d}{=} Y$.

However, we note that this condition of finiteness might be hard to check and does not even apply in many instances. A more general technique is

Proposition 24. *Let ψ be the characteristic function of some random variable X, i.e.,*

$$\psi_X(\lambda) = \mathbb{E}\left[e^{i\lambda X}\right]$$

Then ψ is well defined for all distributions on $\mathbb{R}$ and $X \overset{d}{=} Y$ if and only if $\psi_X = \psi_Y$.

For proofs of these results, see [Dur10, Chu00, vdV12]. Moreover, characteristic functions help determine convergence in distribution. In fact, $X_n \xrightarrow{d} X$ if and only if $\psi_{X_n} \to \psi_X$ pointwise. One way to prove Theorem 22 relies upone this fact.

1.3 The Law of Large Numbers and Borel Cantelli

We ignore history and present the law of large numbers after the central limit theorem. The result is very intuitive, and was first conjectured by Cardano in an effort to improve his gambling; the proof, however, took centuries.

Proposition 25 (The "Very" Weak Law of Large Numbers)**.** *Let $X_1, X_2, \dots$ be pairwise independent random varaibles with $\mathbb{E}X_i = \mu$ and $\operatorname{Var} X_i \le \sigma^2 < \infty$. Then*

$$\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{p} \mu$$

Proof. By translation we may assume that $\mu = 0$. By linearity, $\mathbb{E}\overline{X}_n = 0$. By pairwise independence, $\mathbb{E}\left[\overline{X}_n^2\right] = \sum_{i=1}^n \mathbb{E}[X_i^2] \leq n\sigma^2$. But then

$$\mathbb{E}\left[\overline{X}_n^2\right] \leq \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0$$

and so, applying Proposition 19, we are done. ■

While this proof was obviously very easy, it is not a particularly strong result (hence the title of very weak). Without too much trouble, the second moment requirement could be eliminated, but the real progress that can be made is in the type of convergence that is guaranteed. We can significantly strengthen this result as follows

Theorem 26 (Strong Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be pairwise independent, identically distributed random variables with $\mathbb{E}X_1 = \mu$. Then $\overline{X}_n \to \mu$ $\mathbb{P}$-almost surely.*

We will not give a proof of this here, although one may find one in [Dur10, Chu00]. We will, however, need a key technique in the proof that will have great utility for the remainder of the series. If we are given events $A_1, A_2, \ldots$ on a probability space, we can consider the following events

$$\limsup A_n = \bigcap_n \bigcup_{m \geq n} A_m \qquad\qquad \liminf A_n = \bigcup_n \bigcap_{m \geq n} A_m$$

We say in the former case that A_n happens infinitely often, or i.o., while in the latter case we say that A_n happens eventually. The following lemma gives an easy condition on the probability of something happening infinitely often.

Lemma 27 (Borel-Cantelli)**.** *BC I Let $A_1, A_2, \ldots$ be events on a probability space. Suppose*

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$$

Then $\mathbb{P}(A_n \text{ i.o.}) = 0$.

BC II Conversely, suppose that A_1, A_2 are independent events *and that*

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$$

Then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof. (I): We introduce the random variable

$$X = \sum_{i=1}^{\infty} \mathbb{1}_{A_i}$$

the number of A_i that occur. Note that a summation is just integration with respect to the counting measure and, since $X \geq 0$, we may apply Theorem 6 to get that

$$\mathbb{E}X = \mathbb{E}\left[\sum_{i=1}^{\infty} \mathbb{1}_{A_i}\right] = \sum_{i=1}^{\infty} \mathbb{E}[\mathbb{1}_{A_i}] = \sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$$

But then $X < \infty$ $\mathbb{P}$ almost surely. Thus the first Borel-Cantelli lemma is proved.

(II): Let $0 < m < n < \infty$. We show that the probability of the complement of $\{A_n \text{ i.o.}\}$ is zero. Note that

$$\mathbb{P}\left(\bigcap_m^n A_i^c\right) = \prod_m^n (1 - \mathbb{P}(A_i)) \leq \prod_m^n e^{-\mathbb{P}(A_i)} = e^{-\sum \mathbb{P}(A_i)} \to 0$$

as $n \to \infty$. But

$$\mathbb{P}\left(\{A_n \ i.o.\}^c\right) = \mathbb{P}\left(\bigcup_m \bigcap_{n \geq m} A_n^c\right) \leq \sum_m \mathbb{P}\left(\bigcap_{n \geq m} A_n^c\right) = 0$$

as desired. ■

As an easy example we can prove a weak version of Theorem 26.

Example 28. Suppose that $X_1, \ldots$ are independent and identically distributed such that $\mathbb{E}|X_1|^4 \leq K < \infty$. Let $S_n = X_1 + \cdots + X_n$. We may, after translation, assume that $\mathbb{E}X_1 = 0$. Then we see that

$$S_n^4 = \sum X_i^4 + \sum X_i^3 X_j + \sum X_i^2 X_j^2 + \sum X_i^2 X_j X_k + \sum X_i X_j X_k X_\ell$$

Taking expectations, all but the first and third sums drop out. Counting these terms and bounding the square of the variance shows that $\mathbb{E}S_n^4 \leq Kn^2$. Thus we have

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leq \frac{\mathbb{E}[S_n^4]}{n^4 \varepsilon^4} \leq \frac{K}{n^2 \varepsilon^2}$$

Let A_n be the event in the above line. Then $\sum \mathbb{P}(A_n) < \infty$ and so by Lemma 27, we have that $|\overline{X}_n| < \varepsilon$ eventually for all $\varepsilon > 0$, which gives the result.

2 Discrete Time Stochastic Processes

We first review the notion of conditional expectation before giving a brief preview of stochasatic processes in discrete time. Our guiding example is the random walk, as we will turn to Brownian motion following this section. Good introductions and reviews of this theory exist in [Chu00, Dur10, Kar91, Law06].

2.1 Conditional Expectation

In many undergraduate probability courses, conditional probabilities are defined only on discrete spaces or in terms of a probability density function. In the first setting, students often consider expressions of the form

$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x \text{ and } Y = y)}{\mathbb{P}(Y = y)}$$

This expression clearly only makes sense when $\mathbb{P}(Y = y) \neq 0$ and so one runs into problems in the continuous case. The undergraduate approach is to treat a probability density function, f as the probability that $Y = y$, even though of course this is not true. This then gives the expression for the probability density function $f_{X|Y}$ of

$$f_{X|Y}(x, y) = \frac{f(x, y)}{\int f(x, y) d\mathbb{P}(x)}$$

With conditional probability established, these courses then proceed to introduce conditional expectation. The problem with this approach is that it does not generalize to probability spaces that are not discrete or absolutely continuous with respect to the Lebesgue measure. Thus we will adopt Kolmogorov's solution, which is to define conditional expectation using Theorem 10 and then define conditional probability as $\mathbb{P}(X \in A | \mathcal{F}) = \mathbb{E}\left[\mathbb{1}_A | \mathcal{F}\right]$.

Definition 29. Let $\mathcal{G} \subset \mathcal{F}$ be σ-algebras and let X an $\mathcal{F}$-measurable integrable random variable. Then we say that $Y = \mathbb{E}[X|\mathcal{G}]$ if Y is $\mathcal{G}$-measurable and for all $A \in \mathcal{G}$ we have

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}$$

If $A \in \mathcal{F}$ is an event, then $\mathbb{P}(A|\mathcal{G}) = \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$.

Remark 30. Note that even assuming existence, our definition is only up to sets of probability zero and thus technically we may only discuss conditional expectation as an equivalence class of random variables, where two variables are equivalent if they differ only on a null set. In practice this will not come up.

We need to establish that such a variable exists.

Proposition 31. *As in the definition above, $\mathbb{E}[X|\mathcal{G}]$ exists, is unique up to a null set, and is integrable.*

Proof. We first show that if $Y = \mathbb{E}[X|\mathcal{G}]$ in that it satisfies the two conditions then it is integrable. Let $A = \{Y > 0\}$. Then we see

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P} \leq \int_A |X| d\mathbb{P} < \infty$$

$$\int_{A^c} Y d\mathbb{P} = \int_{A^c} X d\mathbb{P} \leq \int_{A^c} |X| < \infty$$

and so $\mathbb{E}[|Y|] \leq \mathbb{E}[|X|] < \infty$ and so integrability has been established.

To show uniqueness, if Y, Y' are two $\mathcal{G}$-measurable random variables satisfying the condiition in the definition, then for all $\varepsilon > 0$, let $B_\varepsilon = \{Y' - Y \geq \varepsilon\}$ and we have

$$0 = \int_B (X - X) d\mathbb{P} = \int_B (Y' - Y) d\mathbb{P} \geq \varepsilon \mathbb{P}(B)$$

Sending $\varepsilon \downarrow 0$ gives $Y \leq Y'$ $\mathbb{P}$-almost surely. Switching the places of Y, Y' gives the result.

For existence, if $X \geq 0$ then we may define

$$\mu(A) = \int_A X d\mathbb{P}$$

for $A \in \mathcal{G}$ and by Theorem 10 we have that there exists Y which is $\mathcal{G}$-measurable such that

$$\int_A X d\mathbb{P} = \mu(A) = \int_A Y d\mathbb{P}$$

for all $A \in \mathcal{G}$. To treat the general case, apply the above to the decomposition $X = X^+ - X^-$. ■

While this definition is certainly sufficiently general, it is possible that the motivation and intuition behind it is slightly lacking at this point. To recount, we now think of conditional expectation as a random variable, which might seem a little bit odd. Of course, to be given the name "expectation" it should satisfy certain properties of integrals, such as monotonicity, lineartiy, and convergence properties; it is easy to see that this definition does indeed satisfy these. The following examples should clarify some of the intuition.

Example 32. It should be clear from the definition that $\mathbb{E}[X|X] := \mathbb{E}[X|\sigma(X)] = X$ and, more generally, if X is $\mathcal{F}$-measurable, then $\mathbb{E}[X|\mathcal{F}] = X$. This makes sense because if X is $\mathcal{F}$-measurable already, then information in $\mathcal{F}$ should not improve our knowledge of X.

On the other hand, if X is independent from X then $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$. To see this, note that clearly $\mathbb{E}[X]$ is $\mathcal{F}$-measurable and that if $A \in \mathcal{F}$ then

$$\int_A X d\mathbb{P} = \mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbb{1}_A] = \int_A \mathbb{E}[X] d\mathbb{P}$$

Example 33. Now suppose that $\{A_i\}$ is a countable collection of disjoint events such that $\mathbb{P}(A_i) > 0$ and $\mathbb{P}\left(\bigcup_i A_i\right) = 1$. Let $\mathcal{F} = \sigma(A_1, A_2, \dots)$. Then on A_i,

$$\mathbb{E}[X|\mathcal{F}] = \frac{\mathbb{E}\left[X \mathbb{1}_{A_i}\right]}{\mathbb{P}(A_i)}$$

To see this, it is enough to check this on A_i for some i. But then it is clear that

$$\int_{A_i} X d\mathbb{P} = \mathbb{E}\left[X \mathbb{1}_{A_i}\right] = \int_{A_i} \frac{\mathbb{E}\left[X \mathbb{1}_{A_i}\right]}{\mathbb{P}(A_i)} d\mathbb{P}$$

Setting $X = \mathbb{1}_B$ we recover

$$\mathbb{P}(X \in B|A_i) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(A_i)}$$

the identity from an undergraduate class.

Example 34. Let X, Y be random variables on $\mathbb{R}$ and let their joint distribution be absolutely continuous with respect to Lebesgue with Radon-Nikodym derivative $f(x, y)$, i.e., if $A \in \mathcal{B}(\mathbb{R}^2)$ then

$$\mathbb{P}((X, Y) \in A) = \int_A f dx dy$$

Let g be a function such that $\mathbb{E}[|X|] < \infty$ and define

$$\widetilde{g}(y) = \frac{\int_{\mathbb{R}} g(x) f(x, y) dx}{\int_{\mathbb{R}} f(x, y) dx}$$

Then we claim that $\mathbb{E}[g(X)|Y] = \widetilde{g}(Y)$. To see this, let $A \in \sigma(Y)$, then $A = Y^{-1}(A')$ for some $A' \in \mathcal{B}(\mathbb{R})$ and so

$$\mathbb{E}\left[g(X)\mathbb{1}_A\right] = \mathbb{E}\left[g(X)\mathbb{1}_{A'}(Y)\right] = \int_{A'} \int g(x) f(x, y) dx dy = \int_{A'} \widetilde{g}(y) \int f(x, y) dx dy = \int_{A'} \int \widetilde{g}(y) f(x, y) dx dy = \mathbb{E}[\widetilde{g}(Y)\mathbb{1}_A]$$

by definition of $\widetilde{g}$. It is clear that this recovers the probability density function approach to conditional probability if we set $g(x) = x$.

Example 35. Finally, we will derive the classic Bayes' theorem from statistics. Let $B \in \mathcal{G}$. Suppose that B is an event of positive probaility. We may then define

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

The classic Bayes formula asserts that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

To show this, we suppose that $A \in \mathcal{G}$ and we note that

$$\int_\Omega \mathbb{E}\left[\mathbb{1}_B|\mathcal{G}\right] (\mathbb{P}(A \cap B) - \mathbb{P}(B)\mathbb{1}_A) d\mathbb{P} = 0$$

Rearranging gives

$$\mathbb{P}(A|B) = \frac{\int_A \mathbb{P}(B\mathcal{G}) d\mathbb{P}}{\int_\Omega \mathbb{P}(B|\mathcal{G}) d\mathbb{P}}$$

Now, if $\mathcal{G} = \sigma(A_1, A_2, \dots)$ is a countable partition of Ω, then this reduces, by Example 33 to the standard Bayes theorem of

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)}$$

2.2 Martingales and the Random Walk

Now that we have defined conditional expectation, we are able to push forward and get to one of the most important concepts in probability: martingales. Our introdcution of martingales will primarily be accomplished with a single motivating example: the random walk.

Definition 36. We say that $(M_n, \mathcal{F}_n)$ is a martingale if $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all n are σ-algebras, M_n is $\mathcal{F}_n$-measurable and integrable, and $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ for all n. If instead we have $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \geq M_n$ this is called a submartingale and if we have $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n$ this is called a supermartingale. If no filtration is specified, we will assume that $\mathcal{F}_n = \sigma(M_1, \ldots, M_n)$.

The standard example that we will use for a martingale is the random walk. Let $X_1, X_2, \ldots$ be iid random variables such that $\mathbb{E}[X_1] = 0$. Then we call $S_n = X_1 + \cdots + X_n$ a random walk. It is easy to see that this is a martingale. If $0 < \mathbb{E}[X_1] < \infty$ then this is a submartingale and the third case yields a supermartingale. A standard random walk that we will be considering throughout the lecture is called the simple symmetric random walk where X_1 is ± 1 with equal probability. Another example of a martingale is given if $\varphi(\theta) = \mathbb{E}\left[e^{\theta X_1}\right] < \infty$ for some $\theta > 0$. Then we define $W_n = (\varphi(\theta))^{-n} e^{\theta S_n}$ the *Wald martingale.* It is an easy exercise to check that this is a martingale.

One of the concepts that will recur again and again is that of a stopping time.

Definition 37. Let $(\mathcal{F}_n)$ be a filtration of σ-algebras. Then τ is an $\mathcal{F}_n$-stopping time if τ is a random time, $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ such that $\{\tau \leq n\} \in \mathcal{F}_n$ for all n. A stopped stochastic process, (M_n^τ) is defined for a stopping time τ as $M_n^\tau = M_{n \wedge \tau}$.

This definition makes intuitive sense if we consider $\mathcal{F}_n$ to be the aggregate information up until time n. Then for $\{\tau \leq n\} \in \mathcal{F}_n$ means that the information of whether or not we have reached τ is known at each time. The term "martingale" originally described a betting strategy; if we adopt this, we can imagin that a stopping time is a time at which a gambler (who cannot see into the future) can decide to stop playing.

Example 38. The easiest examples of stopping times are constants, $\tau = c$. Another classic example is the hitting time. If $A \subset \mathbb{R}$ we may define $\tau = \inf\{n | M_n \in A\}$. This will be used to great effect with the random walk, where we let $M_n = S_n$ and we consider $\tau = \inf\{n | S_n = a\}$ for some fixed a. In this case, we have $S_n^\tau = a$ for all $n > \tau$ (we will see that τ is almost surely finite in certain cases).

We are now ready to introduce the discrete stochastic integral and give an example of its applications.

Definition 39. We say that (H_n) is a predictable sentence for $\mathcal{F}_n$ if H_n is $\mathcal{F}_{n-1}$-measurable for all n. We then define the discrete stochastic integral as

$$(H \cdot M)_n = \sum_{k=1}^{n} H_k (M_k - M_{k-1})$$

We note that this definition of predictable is perfectly intuitive. Once again, we consider $\mathcal{F}_n$ to be the information that we have at time n. Thus a sequence is predictable if and only if we can predict H_n from time $n-1$. In the gambler analogy, we may consider H_n to be bets that are placed; since a gambler cannot see into the future, they may only place bets based on information that they already have, prior to the next role of the dice.

Proposition 40. *If M_n is a martingale and H_n is a predictable sequence such that each H_n is bounded then $(H \cdot M)_n$ is a martingale. If instead M_n is either a sub- or supermartingale and additionally $H_n \geq 0$ then $(H \cdot M)_n$ is still a sub- or supermartingale.*

Proof. We prove only the martingale case as the others are identical. If H_n is bounded then clearly $(H \cdot M)_n$ is integrable so it suffices to compute $\mathbb{E}[(H \cdot M)_n | \mathcal{F}_{n-1}]$. We compute

$$\begin{aligned}
\mathbb{E}\left[(H \cdot M)_n | \mathcal{F}_{n-1}\right] &= \sum_{k=0}^{n} \mathbb{E}[H_k (M_k - M_{k-1}) | \mathcal{F}_{k-1}] = (H \cdot M)_{n-1} + \mathbb{E}\left[H_n M_n - H_n M_{n-1} | \mathcal{F}_{n-1}\right] \\
&= (H \cdot M)_{n-1} + H_n \mathbb{E}[M_n | \mathcal{F}_{n-1}] - H_n M_{n-1} = (H \cdot M)_{n-1}
\end{aligned}$$

■

As an example, we can show that a stopped martingale is still a martingale. To do this, let $H_n = 1 - \mathbb{1}_{\tau \leq n-1}$ for some stopping time τ. Clearly H_n is predictable. But we note that $H_n = 1 - \mathbb{1}_{\tau < n} = \mathbb{1}_{\tau \geq n}$ and a quick computation reveals that $M_n^\tau = (H \cdot M)_n$.

One strong application of the discrete stochastic integral is to the question of convergence. Fix $a < b$ and let M_n be a submartingale (we could do this equally well for supermartingales of course), and let

$$\begin{aligned} \tau_{2k-1} &= \inf\{n > \tau_{2k-2} | M_n \leq a\} \\ \tau_{2k} &= \inf\{n > \tau_{2k-1} | M_n \geq b\} \\ U_n &= \sup\{k | \tau_{2k} \leq n\} \end{aligned}$$

The τ_{2k-1} and τ_{2k} are stopping times that are struck when our stochastic process either goes below a or above b. The U_n are the number of *upcrossings*, the times that our process goes from below a to above b. We have

Lemma 41 (Upcrossings Inequality)**.** *If M_n is a submartingale and U_n is as above then we have*

$$(b-a)\mathbb{E}U_n \leq \mathbb{E}\left[(M_n - a)^+\right] - \mathbb{E}\left[(M_0 - a)^+\right]$$

Proof. Let $X_n = a + (X_n - a)^+$. Let

$$H_n = \begin{cases} 1 & \tau_{2k-1} \leq n \leq \tau_{2k} \text{ for some } k \\ 0 & \text{otherwise} \end{cases}$$

Then H_n is a predictable sequence because τ_n are stopping times. We may think of H_n as a betting strategy where if M_n is the price of a stock, then we buy one share as soon as the price drops below a and we hold on to this share until the price rises above b, when we sell it. With this intuition, an easy computation shows that $(b-a)U_n \leq (H \cdot X)_n$, because each upcrossing adds at least $b - a$ to our capital and because of the definition of X, anything after the last upcrossing adds something nonnegative to our portfolio. If we let $K_n = 1 - H_n$ then we see $(H \cdot Y)_n + (K \cdot Y)_n = Y_n - Y_0$. By Proposition 40, $(K \cdot Y)_n$ is a submartingale so $\mathbb{E}[(K \cdot Y)_n] \geq \mathbb{E}[(K \cdot Y)_0] = 0$ and so $\mathbb{E}[(H \cdot Y)_n] \leq \mathbb{E}[Y_n - Y_0]$. Putting this all together yields the desired result. ■

The above result is incredibly useful and clever application of Lemma 41 gets one quite far. An important such application is the following

Theorem 42. *If M_n is a submartingale such that $\sup \mathbb{E}\left[M_n^+\right] < \infty$ then there exists an integrable random variable M_∞ such that $M_n \to M_\infty$ $\mathbb{P}$-almost surely.*

Remark 43. Of course, we could have established a downcrossing inequality for supermartingales and then shown Theorem 42 and gotten the same theory. The above clearly also applies to supermartingales by considering $M_n' = -M_n$ and then the condition becomes $\sup \mathbb{E}\left[M_n^-\right] < \infty$. Moreover, this result establishes submartingales as the analogue of increasing sequences. Thus, this theorem is the probabilistic analogue of the statement that increasing sequences that are bounded above converge to a finite limit.

Proof. Note that for any $x \in \mathbb{R}$, $(x-a)^+ \leq |x| + |a|$. Thus Lemma 41 yields

$$\mathbb{E}[U_n] \leq \frac{|a| + \mathbb{E}\left[M_n^+\right]}{b-a}$$

for any $a < b$. By hypothesis, $\mathbb{E}\left[M_n^+\right] \leq C < \infty$ for some $C \in \mathbb{R}$. As U_n is increasing, we may take $U = \sup U_n < \infty$ almost surely. But then this suggests, because $\mathbb{Q}$ is countable, that by countable subadditivity,

$$\mathbb{P}\left(\bigcup_{a,b\in\mathbb{Q}} \{\liminf M_n < a < b < \limsup M_n\}\right) \leq \sum_{a,b\in\mathbb{Q}} \mathbb{P}\left(\{\liminf M_n < a < b < \limsup M_n\}\right) = 0$$

Thus $M_n \to M_\infty$ almost surely for some random variable M_∞.

To show that M_∞ is integrable, we use Fatou's lemma. Note that

$$\mathbb{E}\left[M_\infty^+\right] \leq \liminf \mathbb{E}\left[M_n^+\right] < \infty$$

and similarly

$$\mathbb{E}\left[M_n^-\right] = \mathbb{E}\left[M_n^+\right] - \mathbb{E}[M_n] \leq \mathbb{E}\left[M_n^-\right] - \mathbb{E}[M_0]$$

and so, again by Fatou, we have

$$\mathbb{E}\left[M_\infty^-\right] \leq \liminf \mathbb{E}\left[M_n^-\right] \leq \sup \mathbb{E}\left[M_n^+\right] - \mathbb{E}[M_0] < \infty$$

Thus M_∞ is integrable. ■

Aside from Martingale convergence, another major theorem in the theory of martingales is the optional stopping theorem:

Theorem 44 (Optional Stopping)**.** *Let M be a submartingale and suppose that $\sigma \leq \tau$ are stopping times. Then $\mathbb{E}[M_\tau|\mathcal{F}_\sigma] \geq M_\sigma$ if either of the following is true:*

i) There is some $M < \infty$ such that $\mathbb{P}(\sigma \leq \tau \leq M) = 1$

ii) We have $\mathbb{E}[\tau] < \infty$ and there is some $C < \infty$ such that $\mathbb{E}[|M_{n+1} - M_n||\mathcal{F}_n] \leq C$ almost surely on $\{\tau \geq n\}$

Clearly if M is a supermartingale or martingale we have the corresponding (in)equality in the statement of the theorem. It should also be noted that there are a number of other conditions that guarantee Theorem 44; we omit them for the sake of brevity.

Example 45. The intuition for Theorem 44 should be obvious. If M is a martingale then we know that $\mathbb{E}[M_n|\mathcal{F}_m] = M_m$ for all $m < n$ fixed. Thus if one is to go from fixed times to more general stopping times, one might naïvely expect a similar result to hold. To give an example of why we need these conditions, consider S_n the symmetric simple random walk. We may consider each X_i as a flip of a coin where we win if we get 1 and we can consider a predictable sequence H_n as a betting strategy. We may start by betting 1 and if we win we quit. If we lose, we double our bet. We can iterate this to get a predictable sequence. We have seen that S_n is a martingale and so by Proposition 40, we have that $(H \cdot S)_n$ is also a martingale. Let $\tau = \inf\{n|(H \cdot S)_n > 0\}$. We will see below (and it is easy to take on faith) that $\mathbb{P}(\tau < \infty) = 1$, and an easy analysis shows that $(H \cdot S)_\tau = 1$ almost surely and so $\mathbb{E}[(H \cdot S)_\tau] = 1$. Thus, this strategy guarantees that we make a dollar! As we shall see below, $\mathbb{E}[\tau] = \infty$ and so we might have to wait a while (and go into serious debt) before we win.

Proof. We prove only the first case for martingales and refer the reader to [Dur10, Chu00] for the other case (and other conditions) and merely state that the proofs for sub- and supermartingales are essentially the same. In the first case we have

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_M^\tau] = \mathbb{E}[M_0^\tau] = \mathbb{E}[M_0]$$

by the corollary of Proposition 40 giving that stopped martingales are martingales and the definition of a martingale. ■

We are now ready to begin our analysis of the random walk. Recall that X_i are iid such that $\mathbb{E}[X_i] = \mu < \infty$. Consider $M_n = S_n - n\mu$. It is easy to see that M_n is a martingale. Let $-a < 0 < b$ and let $\tau = \inf\{n|S_n \notin (-a, b)\}$. If $\mathbb{E}[\tau] < \infty$, then note that we are in the second case of Theorem 44 and so $\mathbb{E}[M_\tau] = \mathbb{E}[M_0] = 0$. By linearity we have

$$0 = \mathbb{E}[M_\tau] = \mathbb{E}[S_\tau] - \mathbb{E}[\tau\mu] = \mathbb{E}[S_\tau] - \mu\mathbb{E}[\tau]$$

This is known as Wald's first equation. Let us apply this in the case of the simple symmetric random walk. Note that $\mathbb{P}(S_{b+a} \notin (-a, b)) \geq 2^{-(a+b)}$ because the path that just goes up has at least this probability. But then we have $\mathbb{P}(\tau > n(a+b)) \leq \left(1 - 2^{-(a+b)}\right)^n$ and so by Theorem 6, we have

$$\mathbb{E}\tau = \sum_i \mathbb{P}(\tau > i) \leq C \sum_n \mathbb{P}(\tau > n(a+b)) \leq \sum_n \left(1 - 2^{-(a+b)}\right)^n < \infty$$

Thus we are in the situation of Theorem 44 and so $\mathbb{E}[S_\tau] = \mathbb{E}[\tau]\mathbb{E}[X_1] = 0$ because $\mathbb{E}[X_1] = 0$. But S_n moves exactly 1 at each time step so $S_\tau = -a$ with probability p_a and $S_\tau = b$ with probability $p_b = 1 - p_a$. Thus we have $p_b b - p_a a = 0$ and solving this and $p_b + p_a = 1$ yields

$$\mathbb{P}(S_\tau = -a) = \frac{b}{a+b} \qquad\qquad \mathbb{P}(S_\tau = b) = \frac{a}{a+b}$$

We can do more, however. Let $\mathbb{E}[X_1^2] = \sigma^2 < \infty$ and let $\widetilde{M_n} = S_n^2 - n\sigma^2$. Then

$$\mathbb{E}\left[\widetilde{M_{n+1}}|\mathcal{F}_n\right] = \mathbb{E}\left[(S_n + X_{n+1})^2 - (n+1)\sigma^2|\mathcal{F}_n\right] = S_n^2 - n\sigma^2 + \mathbb{E}\left[X_{n+1}^2|\mathcal{F}_\setminus\right] - \sigma^2 = \widetilde{M_n}$$

and so $\widetilde{M_n}$ is a martingale. Letting τ be as above, we note that we are still in the situation of Theorem 44 and so we have $\mathbb{E}\left[\widetilde{M_\tau}^2\right] = 0$. Thus $\mathbb{E}[S_\tau^2] = \mathbb{E}[\tau]\sigma^2$. But $\sigma^2 = 1$ and $S_\tau \in \{-a, b\}$, so, plugging in, we get

$$\mathbb{E}[\tau] = a^2\frac{b}{a+b} + b^2\frac{a}{a+b} = ab$$

Thus the expected time to leave the interval (a, b) is ab. Now, if $c > 0$, we may ask what the expected time to reach c is. Let $\tau' = \inf\{n|S_n = c\}$. By the fact that we may only get to c if we have first reached $c-1$, we have that $\tau' = \tau_{(-\infty,c-1)}$. Now note that $\tau_{(-a,b)} > \tau_{(-a+1,b)}$ and we may apply monotone convergence to get that

$$\mathbb{E}\tau' = \lim_{a\to\infty} \mathbb{E}[\tau_{(-a,c)}] = \lim_{a\to\infty} ab = \infty$$

Recalling our betting strategy above, we might have a long time to wait before we win our dollar!

As another example of an application of Theorem 44, let X_i, S_n be as in the symmetric simple random walk above. Let $\varphi(\theta) = \mathbb{E}\left[e^{\theta X_1}\right] = \cosh\theta$. We let

$$W_n = \frac{e^{\theta S_n}}{\varphi(\theta)^n} \qquad\qquad W_0 = 1$$

we saw earlier that (W_n) is a martingale. Let $\tau = \inf\{n > 0|S_n = 1\}$ and let $\tau^n = \tau \wedge n$. Then τ^n is a stopping time placing us in the situation of Theorem 44 so

$$1 = \mathbb{E}[W_0] = \mathbb{E}\left[W_{\tau^n}\right] = \mathbb{E}\left[\frac{e^{\theta S_{\tau\wedge n}}}{(\cosh\theta)^{\tau\wedge n}}\right]$$

but

$$0 < \frac{e^{\theta S_{\tau^n}}}{(\cosh\theta)^{(\tau\wedge n)}} < \frac{e^\theta}{\cosh\theta}$$

So by dominated convergence we have

$$1 = \mathbb{E}\left[\lim_{n\to\infty} W_{\tau^n}\right] = \mathbb{E}\left[\frac{e^\theta}{\cosh\theta}\mathbb{1}_{\{\tau<\infty\}} + \lim_{n\to\infty} W_n\mathbb{1}_{\{\tau=\infty\}}\right]$$

But $0 < W_n < (\cosh\theta)^{-n}e^\theta$ on $\{\tau = \infty\}$ for all n so $\lim W_n = 0$ almost surely on $\{\tau = \infty\}$. Thus we have for all $\theta > 0$ that

$$\mathbb{E}\left[\frac{e^\theta\mathbb{1}_{\{\tau<\infty\}}}{\cosh\theta}\right] = 1$$

Sending $\theta \downarrow 0$ and applying dominated convergence shows that $\mathbb{P}(\tau < \infty) = 1$ and so the probability that S_n eventually visits 1 is 1. By symmetry this is true for -1 as well. The random walk satisfies $S_n - S_k$ is independent of $\mathcal{F}_k$ for all $n \geq k \geq 0$ so once we have visited 1, we move either to 2 or 0 and repeat the argument. By induction, we may apply this same argument to any integer in $\mathbb{Z}$. We have thus proved:

Proposition 46. *Let S_n be the simple symmetric random walk. For any integer $a \in \mathbb{Z}$, let $\tau = \inf\{n | S_n = a\}$. Then both of the following are true:*

$$\mathbb{P}(S_n = a \text{ i.o.}) = 1$$
$$\mathbb{E}[\tau] = \infty$$

Thus S_n visits every integer infinitely many times but the expected time to visit any integer is infinite.

As a slight digression, we could prove the first of these statements without martingales pretty easily. We call an event A exchangeable if for all $X = (X_1, X_2, \dots)$ discrete time stochastic processes on our probability space and π a permutation of $\mathbb{N}$ that fixes all but finitely many values, we have $(\pi X)(A) = X(A)$ almost surely. We let $\mathcal{E}$ be the σ-algebra of exchangeable events. The following is true, and a proof can be found in [Chu00, Dur10]:

Theorem 47 (Hewitt-Savage 0-1 Theorem)**.** *If $X_1, X_2, \dots$ are iid and $A \in \mathcal{E}$ is an exchangeable event, then $\mathbb{P}(A) \in \{0, 1\}$*

We may use this theorem to prove

Proposition 48. *Let S_n be a random walk on $\mathbb{R}$. Then exactly one of the four events occurs with probability 1:*

1. $-\infty = \liminf S_n < \limsup S_n = \infty$
2. $\lim S_n \to -\infty$
3. $S_n \to \infty$
4. *For all n, we have $S_n = 0$*

Proof. Notice that if $A = \{\limsup S_n = c\}$ then A is exchangeable so by Theorem 47, we have $\mathbb{P}(A) \in \{0, 1\}$ and so $\limsup S_n = a$ is constant. But note that $S'_n = S_{n+1} - X_1$ has the same distribution as S_n so if a is finite then we have $a = a - X_1$ or $X_1 = 0$. By the fact that X_i are iid, we have that $S_n = 0$ for all n. Thus if we are not in the last case, then $\limsup \in \{\infty, -\infty\}$. The same analysis can be applied to $\liminf S_n$ and the possibilities are listed above. ∎

Now note that if $X_1, X_2, \dots$ are iid and $\mathbb{P}(X_1 = 0) < 1$ then we are not in the last case. If the distribution of the X_i is symmetric about 0, in that $-X_1 \stackrel{d}{=} X_1$, then $-S_n \stackrel{d}{=} S_n$ and so cases (2) and (3) are impossible. Thus the first case holds. If we are in the case of the simple symmetric random walk, then because S_n has to hit all integers on its traversal of $\mathbb{Z}$, this implies the first statement of Proposition 46.

The above statement is one of recurrence. Let τ_a be the hitting time of a. We say that S_n is recurrent if $\mathbb{P}(\tau_a < \infty) = 1$ for some (and hence all) a. Otherwise it is transient. We state the following theorem for the interest of the reader (a proof can be found in [Dur10]):

Theorem 49. *Let S_n be a random walk in $\mathbb{R}$. If $n^{-1} S_n \xrightarrow{p} 0$ then S_n is recurrent. If S_n is a random walk in $\mathbb{R}^2$ and $n^{-\frac{1}{2}} S_n \xrightarrow{d} X$ for some random variable X then S_n is recurrent.*

Examples of how this theorem can be applied abound. For the second case, note that Theorem 22 yields that if $\mathbb{E}[|X_i|^2] < \infty$ then the random walk S_n is recurrent.

3 Brownian Motion and the Random Walk

3.1 An Introduction to Brownian Motion

We have as our canonical example of a stochastic process in discrete time considered a random walk. A natural question of how we might turn this into a continuous time setting arises. A naïve way would be to take a random walk and linearly interpolate between the discrete times. One might then wonder, however, at the arbitrary nature of when these discrete jumps are taking place. Thus we might take a limit of the time between these discrete “jumps” going to zero. Donsker’s invariance theorem formalizes this notion and says that the (scaled) limit of these random walks is given precisely by Brownian motion.

The theory of stochastic processes in discrete time is largely devoid of technicalities, but the same cannot be said for that in continuous time. In the sequel, we will often cite proofs of results that require a standard of rigor that is too technical for this series and thus the goal of the talk is more to give a taste of the relevant concepts than to provide a rigorous introduction to the foundations. Books which do the latter include [Kar91, RY99, Dur10, Mö10, Oks19] in descending order of rigor.

Definition 50. A filtration in continuous time $(\mathcal{F}_t)_{t\geq 0}$ is a collection of σ-algebras such that if $s < t$ then $\mathcal{F}_s \subset \mathcal{F}_t$. A stochastic process $(X_t)_{t\geq 0}$ is a collection of random variables adapted to $\mathcal{F}_t$ in that X_t is $\mathcal{F}_t$-measurable. We call the random functions $t \mapsto X_t$ the sample paths of X_t. A continuous martingale is $(M_t, \mathcal{F}_t)$ is a stochastic process with continuous sample paths such that if $s < t$ then $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$. Continuous analogues of sub- and supermartingales are defined similarly.

Remark 51. One of the technicalities of note is the continuity of filtrations. For instance, we may define

$$\mathcal{F}_{t^+} = \bigcap_{s>t} \mathcal{F}_s$$

and note that this is a σ-algebra. We may define $\mathcal{F}_{t^-}$ similarly. When $\mathcal{F}_t \neq \mathcal{F}_{t^+}$ there can be technical issues that require more sophisticated techniques such as augmentation. This is certainly important as the natural filtration of Brownian motion, $\mathcal{F}_t = \sigma((B_s)_{s\leq t})$ is not even right continuous. We will elide over this by assuming that there is some sufficiently nice filtration and think of this no further.

We may now define Brownian motion

Definition 52. A (standard one-dimensional) Brownian motion is a stochastic process $(B_t)_{t\geq 0}$ satisfying

i) The sample paths are almost surely continuous

ii) For all n and all $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, the random variables $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent

iii) (B_t) is stationary in the sense that for all $0 \leq s < t$, $B_{t+s} - B_t \overset{d}{=} B_s$

iv) For all $t > 0$, B_t is distributed as a gaussian with mean 0 and variance t.

It is not at all obvious that such a stochastic process exists and, in fact, it was not until Norbert Wiener constructed it in 1923 that Brownian motion was treated in a rigorous matter. We will assume the existence of Brownian motion for the remainder of the series, though. An immediate consequence of the definition is that the covariance of B_s and B_t is given by $s \wedge t$. To see this, we know that $\mathbb{E}[B_s] = \mathbb{E}[B_t] = 0$ so it suffices to compute $\mathbb{E}[B_s B_t]$. But we have, assuming that $s \leq t$,

$$\mathbb{E}\left[B_s B_t\right] = \mathbb{E}\left[\mathbb{E}\left[B_s B_t|\mathcal{F}_s\right]\right] = \mathbb{E}\left[\mathbb{E}\left[(B_t - B_s)B_s|\mathcal{F}_s\right] + \mathbb{E}\left[B_s^2|\mathcal{F}_s\right]\right] = \mathbb{E}\left[B_s^2\right] = s$$

by the fact that B_s is $\mathcal{F}_s$- measurable and $B_t - B_s$ is independent of $\mathcal{F}_s$ and is distributed as B_{t-s} which has mean 0. Two very useful properties of Brownian motion, scaling and time inversion, are contained in the following proposition.

Proposition 53. *Let B_t be a standard Brownian motion.*

i) If $\lambda > 0$ then $\lambda B_{\frac{t}{\lambda^2}}$ is again standard Brownian motion.

ii) Let

$$X_t = \begin{cases} 0 & t = 0 \\ tB_{\frac{1}{t}} & t > 0 \end{cases}$$

then X_t is standard Brownian motion.

Proof. (i): Continuity, independence, and stationarity are obvious from these properties of B_t. To check that the distribution is correct, note that $\lambda B_{\frac{t}{\lambda^2}}$ is still gaussian with zero mean and has variance $(\lambda)^2 \frac{t}{\lambda^2} = t$. Thus we are done.

(ii): Independence and stationarity are obvious. A similar computation to the above gives that we have the correct distributions so all that is required is continuity. In fact, continuity is obvious for $t > 0$ so we need only check continuity at 0. By the fact that X_t has the same distribution as B_t, we have because $\mathbb{Q}_{>0}$ countable that

$$\lim_{\substack{t \to 0 \\ t \in \mathbb{Q}}} X_t = 0$$

But $\mathbb{Q}$ is dense in $\mathbb{R}$ and X_t is continuous for $t > 0$ so $X_t \to 0$ as $t \to 0$ and we have continuity. ∎

These two simple properties have surprisingly far reaching consequences. For instance, they allow us to bound the growth of Brownian motion pretty well.

Proposition 54. *Let B_t be standard Brownian motion. Then, almost surely*

$$\lim_{t \to \infty} \frac{B_t}{t} = 0$$

and

$$\limsup \frac{B_t}{\sqrt{t}} = \infty$$

Proof. For the first part, let X_t be the time inversion of B_t. Then we have

$$\lim_{t \to \infty} \frac{B_t}{t} = \lim_{t \to \infty} X_{\frac{1}{t}} = X_0 = 0$$

by Proposition 53.

For the second part, note that by Fatou's lemma, we have for any $c > 0$,

$$\mathbb{P}\left(B_n > c\sqrt{n} \text{ i.o.}\right) = \mathbb{E}\left[\limsup \mathbb{1}_{\{B_n > c\sqrt{n}\}}\right] \geq \limsup \mathbb{E}\left[\mathbb{1}_{\{B_n > c\sqrt{n}\}}\right] = \limsup \mathbb{P}\left(B_n > c\sqrt{n}\right)$$

But by Proposition 53, $\mathbb{P}(B_n > c\sqrt{n}) = \mathbb{P}(B_1 > c) > 0$. But this is an exchangeable event so by Theorem 47 (or by Proposition 60), this probability is 1. ∎

We note that under the interpretation of Brownain motion as a limit of a random walk both of these growth results make sense. The first is the equivalent of Theorem 26, the Strong Law of Large Numbers, while the second is the analogue of Theorem 22, the Central Limit Theorem. Putting these results together, we know that $|B_t|$ grows more quickly than $\sqrt{t}$ and less quickly than t; in the sequel, we will develop a much finer result in this direction. Moreover, by Proposition 54, and symmetry we have that $\liminf B_t = -\infty$. Because B_t is continuous, this means that B_t passes through every real number infinitely many times almost surely, another result that should be familiar from the simple symmetric random walk.

3.2 Brownian Motion as a Stochastic Process

We consider the symmetric random walk to be the standard example of a martingale in discrete time. Similarly, we would like to consider Brownian motion as a martingale in continuous time. Indeed, we have

Proposition 55. *Let B_t be a Brownian motion. Then B_t is a martingale with respect to its natural filtration $\mathcal{F}_t$.*

Proof. We compute for $s < t$,

$$\mathbb{E}\left[B_t | \mathcal{F}_s\right] = \mathbb{E}\left[B_t - B_s + B_s | \mathcal{F}_s\right] = \mathbb{E}\left[B_t - B_s | \mathcal{F}_s\right] + \mathbb{E}\left[B_s | \mathcal{F}_s\right] = 0 + B_s = B_s$$

∎

Example 56. For another example of a martingale, we could consider $B_t^2 - t$. Indeed we have

$$\mathbb{E}[B_t^2 - t|\mathcal{F}_s] = \mathbb{E}[(B_t - B_s)^2 + 2(B_t - B_s)B_s + B_s^2 - t|\mathcal{F}_s] = B_s^2 - s$$

We can think of this martingale as the continuous analogue of $S_n^2 - \sigma^2 n$ of the random walk in discrete time. This will be an important martingale when we talk about stochastic integration.

Just as we have the definition of a stopping time in discrete time, so, too, do we have one in continuous time

Definition 57. A random time τ is a stopping time for the filtration $\mathcal{F}_t$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t > 0$.

Remark 58. In continuous time we are forced to make a choice between considering sets $\{\tau < t\}$ and sets $\{\tau \leq t\}$ for the definition of a stopping time, a difficulty that does not appear in discrete time. The former are called optional times, which is where the name optional stopping theorem comes from. Note that if the filtration is not right continuous, then these notions are distinct. In order to avoid this (rather important) technicality, we will augment our filtrations so as to be right continuous whenever we talk about stopping times; in particular this means we are *not* using the natural Brownian filtration.

Similarly to Theorem 44, we have an optional stopping theorem in continuous time.

Theorem 59. *Let $(M_t, \mathcal{F}_t)$ be a continuous martingale and $S \leq T$ be $\mathcal{F}_t$-stopping times. and suppose that $|M_{t\wedge T}| \leq X$ for some random variable X such that $\mathbb{E}[X] < \infty$. Then $\mathbb{E}[M_T|\mathcal{F}_S] = M_S$.*

Rigorous proofs of this theorem can be found in [Kar91, RY99, Mö10], but the idea is simple. We have already shown this theorem in the case of discrete time, so we take a sequence of stopping times (T_n) such that $T_n \downarrow T$ and apply Theorem 44. We will eventually want to apply this to get a continuous version of Wald's equations, but before we can do this, we need some facts about the running maximum.

Brownian motion is a martingale, as has been established, but it is also a Markov process. We do not wish to go into the details of defining such, but we call the fact that $B_{t+s} - B_t$ is independent of $\mathcal{F}_t$ for all $0 \leq t, s$ to be the Markov property. Applying this property to $s = 0$ gives the Bluementhal 0-1 law.

Proposition 60. *If $\mathcal{F}_{0+}$ is the augmentation of the natural filtration associated to B_t then it is trivial in the sense of Theorem 47. Similarly, if $\mathcal{T}$ is the tail σ-algebra, then it too is trivial.*

Proof. The first statement follows from the markov property applied to $s = 0$. The second statement follows from the first applied to the time inverted Brownian motion $X_t = tB_{\frac{1}{t}}$. ∎

Remark 61. The second statement above could be treated as a special case of the Kolmogorov 0-1 law which states that if $X_1, X_2, \ldots$ are independent and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ then the tail σ-algebra $\mathcal{T} = \bigcap_n \mathcal{F}_n$ is trivial.

While the markov property is useful, the following generalization is even more useful.

Theorem 62 (Strong Markov Property)**.** *Let τ be a stopping time for a standard brownian motion B_t such that $\mathbb{P}(\tau < \infty) = 1$. Then for all $t > 0$, $B_{\tau+t} - B_\tau$ is distributed as a standard brownian motion independent of $\mathcal{F}_\tau$.*

We will not prove this here, and instead cite [Kar91, Mö10, RY99, Dur10], although we remark that the proof method is the same discrete approximation idea mentioned following Theorem 59.

One of the key applications of Theorem 62 is the reflection principle. Given a stopping time τ, we may define reflected Brownian motion as

$$\widetilde{B}_t = \begin{cases} B_t & t \leq \tau \\ 2B_\tau - B_t & t > \tau \end{cases}$$

An example is in Figure 1 The key result regarding such reflections is

Theorem 63 (Schwarz Reflection Principle)**.** *Let B_t be a standard Brownian motion, τ a stopping time, and $\widetilde{B}_t$ the reflection at τ. Then $\widetilde{B}_t$ is a standard Brownian motion.*

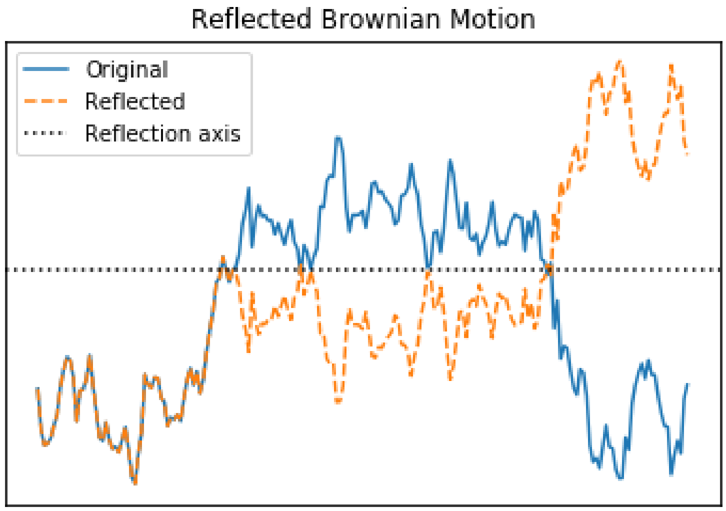

Figure 1: A Brownian motion reflected where the grey represents the path before reflection, the blue is the continuation of the original, and the dotted orange is the reflection.

Proof. On $\{\tau = \infty\}$ this is trivial so we restrict to $\{\tau < \infty\}$. In this case we note that $B_{\tau+t} - B_\tau$ is a standard Brownian motion by Theorem 62. By symmetry, $B_\tau - B_{\tau+t} = -(B_{\tau+t} - B_t)$ is also a standard Brownian motion. We note that concatenating independent paths is measurable and so we see that gluing $(B_{t+\tau} - B_\tau)$ to B_t at τ is the same in distribution as gluing $(B_\tau - B_{\tau+t})$. The first of these is just B_t and the second is $\widetilde{B}_t$. Thus $(\widetilde{B}_t) \overset{d}{=} (B_t)$ as stochastic processes. ■

Remark 64. Note that this is stronger than saying that $B_t \overset{d}{=} \widetilde{B}_t$ for all t. For instance, if we consider $X_t = \sqrt{t}Z$ where Z is a standard Gaussian then $B_t \overset{d}{=} X_t$ for all t, but their distributions are clearly different as stochastic processes.

As an example application of this powerful principle, we can find the distribution of the maximum process and the intimately related first passage time. In particular, for some $a \in \mathbb{R}$, we may define $\tau_a = \inf\{t | B_t = a\}$. Note that τ_a is a stopping time. Then we have

Proposition 65. *Let $M_t = \max_{s \le t} B_s$ and let τ_a as above. We have*

$$\mathbb{P}(M_t > a) = \mathbb{P}(\tau_a < t) = 2\mathbb{P}(B_t > a) = \mathbb{P}(|B_t| > a) = \sqrt{\frac{2}{\pi}} \int_{at^{-\frac{1}{2}}}^{\infty} e^{-\frac{x^2}{2}} dx$$

Remark 66. Similar to the previous remark, we have that the distributions of M_t and $|B_t|$ are the same for all t but they are not equal in distribution as stochastic processes, as M_t is always increasing. There is another stochastic process, $M_t - B_t$, however, that does have the same distribution as a stochastic process as $|B_t|$. This is illustrated in Figure 2.

Proof of Proposition 65. Notice that $-B_t \overset{d}{=} B_t$ so without loss of generality, $a > 0$. But if $a > 0$ then $\tau_a < t$ if and only if $M_t \ge a$, thus we have the first equality. Now let $\widetilde{B}_t$ be the reflection through τ_a. We note that if $\tau_a < t$ then $\widetilde{B}_t < a$ if and only if $\widetilde{B}_t > a$. We thus have

$$\begin{aligned}
\mathbb{P}(\tau_a < t) &= \mathbb{P}(\{\tau_a < t\} \cap \{B_t < a\}) + \mathbb{P}(\{\tau_a < t\} \cap \{B_t > a\}) \\
&= \mathbb{P}(\{\tau_a < t\} \cap \{B_t > a\}) + \mathbb{P}\left(\{\tau_a < t\} \cap \{\widetilde{B}_t > a\}\right) \\
&= 2\mathbb{P}(\{\tau_a < t\} \cap \{B_t > a\}) = 2\mathbb{P}(B_t > a)
\end{aligned}$$

Figure 2: The first figure shows the original B_t and $|B_t|$. The second figure shows M_t and $M_t - B_t$. Note that M_t is increasing.

where the penultimate equality comes from Theorem 63 and the last equality comes from the fact that $\{\tau_a < t\} \supset \{B_t > a\}$ by continuity. But by symmetry, $2\mathbb{P}(B_t > a) = \mathbb{P}(|B_t| > a)$. For the last equality, we note that

$$\mathbb{P}\left(|B_t| > a\right) = \mathbb{P}\left(\left|\frac{B_t}{\sqrt{t}}\right| > \frac{a}{\sqrt{t}}\right) = \sqrt{\frac{2}{\pi}} \int_{bt^{-\frac{1}{2}}}^{\infty} e^{-\frac{x^2}{2}} dx$$

by Proposition 53. ■

The above result is of independent interest, because the maximum process has many important applications. One easy one is to verify a Brownian motion analogue of Wald's equations.

Proposition 67. *Let B_t be Brownian motion and τ a stopping time such that $\mathbb{E}\tau < \infty$. Then $\mathbb{E}[B_\tau] = 0$ and $\mathbb{E}\left[B_\tau^2\right] = \mathbb{E}\tau$.*

Proof. For the first part, we consider

$$M_k = \max_{0 \le s \le 1} |B_{k+s} - B_k| \qquad\qquad M = \sum_{k=0}^{\lceil T \rceil} M_k$$

Note that M_k are iid random variables by the definition of Brownian motion. Moreover, by the tail sum formula, recall that

$$\mathbb{E}[M_1] = \int_0^\infty \mathbb{P}(M_1 > x) dx \le 1 + \int_1^\infty C \frac{e^{-\frac{x^2}{2}}}{x} dx < \infty$$

where the first inequality follows from Proposition 65 and the gaussian tail bound. Thus we have

$$\mathbb{E}[M] = \mathbb{E}\left[\sum_{k=0}^{\lceil T \rceil} M_k\right] = \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbb{1}_{\{\tau > k-1\}} M_k\right] = \sum_{k=0}^{\infty} \mathbb{P}(T > k-1)\mathbb{E}[M_k] = \mathbb{E}[M_1]\mathbb{E}[T] < \infty$$

Thus we have $|B_{t\wedge\tau}| \le M$ and $\mathbb{E}M < \infty$ so we may apply Theorem 59 to get the result.

Now for the second statement, let $T_n = \inf\{t : |B_t| = n\}$. Let $X_t = B_t^2 - t$, which we have already seen is a martingale. Then we can stop X_t and consider $X_t^{\tau \wedge T_n}$ and we note that $\left|X_t^{\tau\wedge T_n}\right| \le n^2 + \tau$ which is integrable by assumption. By Theorem 59, we have $\mathbb{E}\left[B^2_{\tau\wedge T_n}\right] = \mathbb{E}[\tau \wedge T_n]$. Now, by Fatou's lemma, we have

$$\mathbb{E}[B^2_\tau] \leq \liminf \mathbb{E}\left[B^2_{\tau\wedge T_n}\right] = \liminf \mathbb{E}\left[\tau \wedge T_n\right] \leq \mathbb{E}[\tau]$$

Conversely, we notice that

$$\begin{aligned}\mathbb{E}[B^2_\tau] &= \mathbb{E}\left[B^2_{\tau\wedge T_n}\right] + 2\mathbb{E}\left[B_{\tau\wedge T_n}(B_\tau - B_{\tau\wedge T_n})\right] + \mathbb{E}\left[(B_\tau - B_{\tau\wedge T_n})^2\right] \\ &= \mathbb{E}\left[B^2_{\tau\wedge T_n}\right] + \mathbb{E}\left[(B_\tau - B_{\tau\wedge T_n})^2\right] \geq \mathbb{E}\left[B^2_{\tau\wedge T_n}\right]\end{aligned}$$

Finally, by monotone convergence,

$$\mathbb{E}\left[B^2_\tau\right] \geq \lim \mathbb{E}\left[B^2_{\tau\wedge T_n}\right] = \lim \mathbb{E}[\tau \wedge T_n] = \mathbb{E}[\tau]$$

Putting the inequalities together yields the result. ■

As an example, we may consider $\tau = \inf\{t : B_t \notin (-a, b)\}$ for $0 < a, b$. Clearly $\tau \wedge n$ is an integrable stopping time, so we may apply Theorem 59 to get $\mathbb{E}[\tau \wedge n] = \mathbb{E}[W^2_{\tau\wedge n}] \leq \min(a^2, b^2)$. By the monotone convergence theorem and the fact that $\tau \wedge n \uparrow \tau$, we get $\mathbb{E}[\tau] \leq \min(a^2, b^2) < \infty$. Now we may apply Proposition 67 in the same manner as we did for the simple symmetric random walk to get

$$\mathbb{P}(B_\tau = -a) = \frac{b}{a+b} \qquad \mathbb{P}(B_\tau = b) = \frac{a}{a+b} \qquad \mathbb{E}[\tau] = ab$$

Just as in the symmetric random walk, we may take $b \to \infty$ and prove that if $\tau_a = \inf\{t : B_t = a\}$ then $\mathbb{E}[\tau_a] = \infty$. Of course, we could have proved this using our explicit characterization of the distribution of τ_a instead.

3.3 The Law of the Iterated Logarithm

In this section we formalize a connection between the random walk and Brownian motion and illustrate its utility with the celebrated law of the iterated logarithm. We seek to answer the question of how quickly can a random walk grow. We know from Theorems 22 and 26 that

$$\limsup \frac{S_n}{n} = 0 \qquad \limsup \frac{S_n}{\sqrt{n}} = \infty$$

almost surely. Similarly we know by Proposition 54 that

$$\limsup \frac{B_t}{t} = 0 \qquad \limsup \frac{B_t}{\sqrt{t}} = \infty$$

Thus we know that the rate of growth is greater than $\sqrt{t}$ but less than t for Brownian motion. We will use Brownian scaling to find the rate of growth of Brownian motion. We will then embed a random walk in Brownian motion and use this embedding to find the growth rate of a general random walk. We begin by

Theorem 68 (Law of the Iterated Logarithm)**.** *Let B_t be standard brownian motion. Then almost surely*

$$\limsup \frac{B_t}{\sqrt{2t\log\log t}} = 1$$

Proof. Let $\varphi(t) = \sqrt{2t\log\log t}$. Fix $\varepsilon > 0$ and $r > 1$. Consider the events

$$A_n = \left\{\max_{t\leq r^n} B_t \geq (1+\varepsilon)\varphi(r^n)\right\}$$

Letting M_t be the running maximum and applying Proposition 65, we have

$$\mathbb{P}\left(A_n\right) = \mathbb{P}\left(\frac{|B_{r^n}|}{\sqrt{r^n}} \geq (1+\varepsilon)\frac{\varphi(r^n)}{\sqrt{r^n}}\right)$$

Applying Proposition 53 and the Gaussian tail bound from the problem set, we have

$$\mathbb{P}(A_n) \le 2\exp\left(-(1+\varepsilon)^2 \log\log r^n\right) = \frac{2}{(n\log r)^{(1+\varepsilon)^2}}$$

But this is summable, so applying Lemma 27 we have that $\mathbb{P}(A_n \text{ i.o.}) = 0$. For $t \gg 0$ we have $r^{n-1} \le t < r^n$ and we note that $\frac{\varphi(t)}{t} \downarrow 0$ so we have

$$\frac{B_t}{\varphi(t)} = \frac{B_t}{\varphi(r^n)} \frac{\varphi(r^n)}{r^n} \frac{t}{\varphi(t)} \le (1+\varepsilon)r$$

eventually by the previous computation. Thus we have almost surely that

$$\limsup \frac{B_t}{\varphi(t)} \le (1+\varepsilon)r$$

Sending ε to 0 and r to 1 gives that $\limsup \frac{B_t}{\varphi(t)} \le 1$.

For the opposite inequality, we obviously wish to use the second Borel-Cantelli, but for this we need to make the events independent. Again we fix an $r > 1$. Now we consider the events

$$A_n = \left\{B_{r^n} - B_{r^{n-1}} \ge \varphi(r^n - r^{n-1})\right\}$$

We know from the lower gaussian tail bound that if Z is standard gaussian then there is a constant c such that

$$\mathbb{P}(Z > x) \ge c\frac{e^{-\frac{x^2}{2}}}{x}\left(1 - \frac{1}{x^2}\right)$$

Thus, for $n \gg 0$ we have

$$\mathbb{P}(A_n) = \mathbb{P}\left(Z \ge \frac{\varphi(r^n - r^{n-1})}{\sqrt{r^n - r^{n-1}}}\right) \ge C\frac{e^{-\log\log(r^n - r^{n-1})}}{\sqrt{\log\log(r^n - r^{n-1})}} > \frac{C}{n\log n}$$

We know that the A_n are independent so we may apply Lemma 27 to get that $\mathbb{P}(A_n \text{ i.o.}) = 1$. We now need to translate this back into a statement for general t. To do this, we note that because $\limsup \frac{B_t}{\varphi(t)} \le 1$, we have that

$$B_{r^n} \ge B_{r^{n-1}} + \varphi(r^n - r^{n-1}) \ge -2\varphi(r^{n-1}) + \varphi(r^n - r^{n-1})$$

infinitely often. Dividing by $\varphi(r^n)$, and noting that

$$\frac{\varphi(r^{n-1})}{\varphi(r^n)} = \frac{\varphi(r^{n-1})}{\sqrt{r^{n-1}}} \frac{\sqrt{r^n}}{\varphi(r^n)} \frac{1}{\sqrt{r}} \le \frac{1}{\sqrt{r}}$$

and that $\frac{\varphi(t)}{t} \downarrow 0$, we have

$$\frac{B_{r^n}}{\varphi(r^n)} \ge \frac{-2\varphi(r^{n-1}) + \varphi(r^n - r^{n-1})}{\varphi(r^n)} \ge -\frac{2}{\sqrt{r}} + \frac{r^n - r^{n-1}}{r^n} = 1 - \frac{2}{\sqrt{r}} - \frac{1}{r}$$

infinitely often. Thus we know that

$$\limsup \frac{B_t}{\varphi(t)} \ge 1 - \frac{2}{\sqrt{r}} - \frac{1}{r}$$

Sending $r \to \infty$ we get the other inequality and we are done. ■

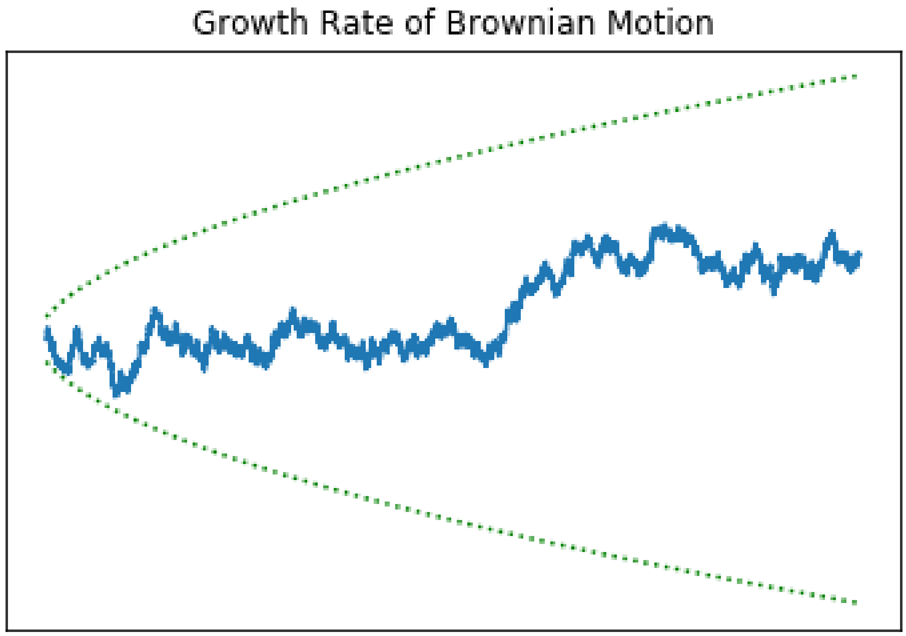

Figure 3: A sample Brownian Path and the envelope given by Theorem 68 in dashed lines.

This gives a very fine result for Brownian motion, in that we know how quickly it can grow. By symmetry, we have that almost surely, for all $t \gg 0$, B_t is contained in the envelope given by $\pm\varphi(t)$, an example is seen in Figure 3 We have not yet related this to a random walk. The trick will be "embedding" the random walk in Brownian motion. To do this we will want to choose stopping times $\tau_1, \tau_2, \dots$ such that B_{τ_n} has the same distribution as S_n. Then we will hope that this will be enough to coerce our random walk into satisfying the conclusion of Theorem 68. This requires two steps, the first being the embedding and the second showing that the limit supremum is the same along subsequences. Because of its similarity to the previous proof, we do the second step first.

Proposition 69. *Let $\tau_1 \le \tau_2 \le \dots$ be a sequence of random times such that $\tau_n \to \infty$ and $\frac{\tau_{n+1}}{\tau_n} \to 1$ almost surely. Then almost surely,*

$$\limsup \frac{B_{\tau_n}}{\varphi(\tau_n)} = 1$$

Similarly, if τ_n is a sequence of random times such that $\frac{\tau_n}{n} \to a$ almost surely for some $a > 0$ then almost surely,

$$\limsup \frac{B_{\tau_n}}{\varphi(an)} = 1$$

Where $\varphi(n) = \sqrt{2n \log\log n}$.

Remark 70. The upper bound is obvious, but the restrictions for the lower bound are less obvious. Intuitively, they ensure that the stopping times are sufficiently dense so as to be representative. For instance, if we took $\tau_n = \inf\{\tau_{n-1} + 1 || B_{\tau_n}| \le 1\}$ then τ_n are almost surely finite by the recurrence of Brownian motion, but the conclusion clearly does not hold. This is because these times are relatively rare and so are not representative.

Proof. This proof proceeds in much the same way as the proof of Theorem 68. For the first statement, the upper bound is clear because if $\frac{B_{\tau_n}}{\varphi(\tau_n)} > 1$ infinitely often then $\frac{B_t}{\varphi(t)} > 1$ infinitely often, contradicting Theorem 68. For the lower bound, let $r > 4$ and let

$$A_k = \left\{B_{r^k} - B_{r^{k-1}} \geq \varphi(r^k - r^{k-1})\right\} \qquad D_k = \left\{\min_{r^k \leq t \leq r^{k+1}} B_t - B_{r^k} \geq -\sqrt{r^k}\right\}$$

Moreover, let $A'_k = A_k \cap D_k$. Note that by the definition of Brownian motion, A_k and D_k are independent. Using the Gaussian tail bound and Proposition 53, we have

$$\mathbb{P}(A_k) = \mathbb{P}\left(\frac{B_{r^k} - B_{r^{k-1}}}{\sqrt{r^k - r^{k-1}}} \geq \frac{\varphi(r^k - r^{k-1})}{\sqrt{r^k - r^{k-1}}}\right) = \mathbb{P}\left(B_1 \geq \frac{\varphi(r^k - r^{k-1})}{\sqrt{r^k - r^{k-1}}}\right) \geq \frac{c}{k \log k}$$

Now note that $\mathbb{P}(D_k) = p_r > 0$ independent of k, by the markov property. By independence, we have $\mathbb{P}(A'_k) = \mathbb{P}(A_k)p_r$. Thus $\mathbb{P}(A'_k)$ is summable and they are independent so by Lemma 27 we have that infinitely often

$$\min_{r^k \leq t \leq r^{k+1}} B_t \geq B_{r^{k-1}} + \varphi(r^k - r^{k-1}) - \sqrt{r^k}$$

By $B_r^{k-1} \geq -2\varphi(r^{k-1})$ for $k \gg 0$ by Theorem 68, and the computation

$$\varphi(r^k) - \varphi(r^{k-1}) \geq \varphi(r^k)\left(1 - \frac{1}{r}\right)$$

We have for infinitely many k,

$$\min_{r^k \leq t \leq r^{k+1}} B_t \geq \varphi(r^k - r^{k-1}) - 2\varphi(r^{k-1}) - \sqrt{r^k} \geq \varphi(r^k)\left(1 - \frac{1}{r} - \frac{2}{\sqrt{r}}\right) - \sqrt{r^k} > 0$$

by $r > 4$ Let $n_k = \inf\{n | \tau_n > r^k\}$. By $\frac{\tau_{n+1}}{\tau_n} \to 1$ we have for all $\varepsilon > 0$ and $k \gg 0$ that $r^k \leq \tau_{n_k} < r^k(1+\varepsilon)$ so

$$\frac{B_{\tau_{n_k}}}{\varphi(\tau_{n_k)}} \geq \frac{\varphi(r^k)}{\varphi(r^k(1+\varepsilon))}\left(1 - \frac{1}{r} - \frac{2}{\sqrt{r}}\right) - \frac{\sqrt{r^k}}{\varphi(r^k)}$$

but we know that $\frac{\sqrt{r^k}}{\varphi(r^k)} \to 0$ and that $\frac{\varphi(r^k)}{\varphi(r^k(1+\varepsilon))} \to \frac{1}{\sqrt{1+\varepsilon}}$ so

$$\limsup \frac{B_{\tau_n}}{\varphi(\tau_n)} \geq \frac{1}{\sqrt{1+\varepsilon}}\left(1 - \frac{1}{r} - \frac{2}{\sqrt{r}}\right)$$

almost surely. Sending $r \to \infty$ and $\varepsilon \to 0$ yields the first result.

For the second statement, note that if $\frac{\tau_n}{n} \to a$ almost surely, then $\frac{\varphi(\tau_n)}{\varphi(an)} \to 1$ almost surely and so we have

$$\limsup \frac{B_{\tau_n}}{\varphi(an)} = \limsup \frac{B_{\tau_n}}{\varphi(\tau_n)} \frac{\varphi(\tau_n)}{\varphi(an)} = \limsup \frac{B_{\tau_n}}{\varphi(\tau_n)} = 1$$

by the first statement so we are done. ■

Thus we need to find stopping times such that B_{τ_n} has the same distribution as S_n and $\frac{\tau_n}{n} \to 1$. Suppose we had a way of constructing a stopping time τ_1 such that $B_\tau \stackrel{d}{=} X_1$ and $\mathbb{E}[\tau] < \infty$. Then we could consider $B_{\tau+t} - B_\tau$ and by Theorem 62 we have that this is distributed again as Brownian motion. Thus we could define τ_2 as equal in distribution to τ_1 and run this on the new Brownian motoin. In this way we can construct $\tau_1 \leq \tau_2 \leq \ldots$ stopping times and we have

$$B_{\tau_n} = \sum_{k=1}^{n} B_{\tau_k} - B_{\tau_{k-1}} \stackrel{d}{=} \sum_{k=1}^{n} X_k = S_n$$

Moreover, we note that $\frac{\tau_n}{n} \to \mathbb{E}[\tau_1]$ almost surely. Thus we have reduced the problem to embedding X into Brownian motion such that $\mathbb{E}[\tau] < \infty$. In this case we would have by Proposition 67 that $\mathbb{E}[X] = \mathbb{E}[B_\tau] = 0$ and $\mathbb{E}[X^2] = \mathbb{E}[\tau] < \infty$. Thus we have necessary conditions on X in order to embed X in Brownian motion. For convenience, we assume that $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1^2] = 1$, which by translation and scaling is no loss of generality. The surprising fact is that the above conditions are sufficient to embed X_1 into Brownian motion. This is captured in

Theorem 71 (Skorokhod Embedding Theorem)**.** *Let X be a random variable such that $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = 1$. Then there exists a stopping time τ for Brownian motion such that B_τ is distributed as X and $\mathbb{E}[\tau] = 1$.*

Our proof will follow the strategy of [Dub68] as quoted in [Mö10], but there are many other approaches. A particularly slick proof appears in [CW76], and [Obł04] is a survey of all known constructions and their various merits. That said, we favor Dubins' proof for its brevity and introduce binary splitting martingales for this reason. We begin with an example

Example 72. Suppose that X is a random variable that takes on only two values, i.e., $X \in \{-a, b\}$. If we want $\mathbb{E}[X] = 0$ then we need $a, b > 0$ and we must have

$$\mathbb{P}(X = a) = \frac{b}{a+b} \qquad \mathbb{P}(X = b) = \frac{a}{a+b} \qquad \mathbb{E}[X^2] = ab$$

We have seen in Proposition 67 that if $\tau = \inf\{t | B_t \notin (-a, b)\}$ then B_τ is distributed as X and $\mathbb{E}[\tau] = ab$. Thus if we want to embed the simple symmetric random walk in Brownian motion, we could let $\tau = \inf\{t | B_t \notin (-1, 1)\}$.

While this example is easy, it is the basis for the proof of the more general case. The above example motivates the following definition:

Definition 73. Let (X_n) be a martingale in discrete time such that $\mathbb{E}[X_0] = 0$. For $x_0, \ldots, x_n \in \mathbb{R}$, define the event

$$A(x_0, \ldots, x_n) = \{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\}$$

We say that X_n is binary splitting if for all $x_0, \ldots, x_n$ such that $\mathbb{P}(A(x_0, \ldots, x_n)) > 0$ we have that $\mathbb{E}[X_{n+1} | A(x_0, \ldots, x_n)]$ is supported on at most two values.

While the definition might seem a little complicated, the intuition is quite simple in that X_n is a binary splitting martingale if and only if at each time step, the stochasatic process acts like the X in the example above. In fact, using the above example, we may embed a binary splitting martingale in Brownian motion. This is because X_0 is supported on two values and has zero mean so the example above provides a stopping time τ_0 such that B_{τ_0} is an embedding. Now, given this information, X_1 is supported at two values and so we may again apply the example to $B_{\tau_0 + t} - B_{\tau_0}$ and get a stopping time τ_1 by Theorem 62. We may continue this process and note that B_{τ_n} is distributed as X_n and $\mathbb{E}[\tau_n] = \mathbb{E}[X_n^2]$. This is all very well, but we still need to connect the notion of binary splitting martingales to the more general Theorem 71. This is the content of the following lemma.

Lemma 74. *Let X be a random variable such that $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$. Then there exists a binary splitting martingale $(X_n, \mathcal{F}_n)$ such that $X_n \to X$ almost surely and in L^2.*

Proof. We define $(X_n, \mathcal{F}_n)$ recursively. Let $\mathcal{F}_0$ be trivial and let $X_0 = \mathbb{E}[X]$. We define

$$Y_0 = \begin{cases} 1 & X \geq X_0 \\ -1 & X < X_0 \end{cases}$$

Now, let $\mathcal{F}_n = \sigma(Y_0, \ldots, Y_{n-1})$ and let $X_n = \mathbb{E}[X | \mathcal{F}_{n-1}]$ and finally

$$Y_n = \begin{cases} 1 & X \geq X_n \\ -1 & X < X_n \end{cases}$$

Now, we note that $\mathcal{F}_n = \sigma(A_i)$ where there are 2^n such A_i and each A_i splits into two sets to make $\mathcal{F}_{n+1}$; thus it is clear that X_n is a binary splitting martingale. Thus, it suffices to show that $X_n \to X$ almost surely and in L^2. Note that

$$\mathbb{E}[X^2] = \mathbb{E}[(X - X_n)^2] + \mathbb{E}[X_n^2] + 2\mathbb{E}[X(X - X_n)] \geq \mathbb{E}[X_n^2]$$

Thus, X_n is bounded in L^2 and so $X_n \to X_\infty = \mathbb{E}[X|\mathcal{F}_\infty]$ almost surely and in L^2. Thus it suffices to show that $X_\infty = X$ almost surely. To do this, we claim that

$$\lim_{n\to\infty} Y_n(X - X_{n+1}) = |X - X_\infty|$$

If $X(\omega) < X_n(\omega)$ then for sufficiently large n we have $Y_n = -1$ and if $X(\omega) > X_n(\omega)$ then for sufficiently large n we have that $Y_n = 1$. Thus the claim holds. Recalling that $Y_n \in \mathcal{F}_n$, we have

$$\mathbb{E}[Y_n(X - X_{n+1})] = \mathbb{E}[\mathbb{E}[Y_n(X - X_{n+1})|\mathcal{F}_n]] = \mathbb{E}[Y_n\mathbb{E}[X - X_{n+1}|\mathcal{F}_n]] = 0$$

by the fact that $X_{n+1} := \mathbb{E}[X|\mathcal{F}_n]$. But we have integrability of X and so by the dominated convergence theorem, we have

$$\mathbb{E}[|X - X_\infty|] = 0$$

and so $X = X_\infty$. Thus we are done. ∎

With Lemma 74 in hand, we are ready to prove the embedding theorem.

Proof of Theorem 71. Let X be a random variable with mean zero and $\mathbb{E}[X^2] < \infty$. By Lemma 74 there is a sequence of binary splitting martingales $X_n \to X$ almost surely and in L^2. But if X_n is supported on 2 points then we have already seen how to construct a stopping time that satisfies the theorem in Example 72. Thus we can find a sequence of stopping times $\tau_1 \leq \tau_1 \leq \dots$ such that τ_n is an embedding of X_n in Brownian motion. Because they are increasing $\tau_n \uparrow \tau$ a stopping time, we have

$$\mathbb{E}[\tau] = \lim \mathbb{E}[\tau_n] = \lim \mathbb{E}[X_n^2] = \mathbb{E}[X^2] < \infty$$

by the fact that $X_n \to X$ in L^2 and thus $\tau < \infty$ almost surely. But then we also have

$$B_\tau = \lim_{n\to\infty} B_{\tau_n} \stackrel{d}{=} \lim_{n\to\infty} X_n = X$$

and thus we have an embedding. ∎

Armed with all of the machinery that we have just built up, we are finally ready to prove the major result of the section.

Theorem 75 (Law of the Iterated Logarithm, II)**.** *Let $X_1, \dots$ be iid random variables of zero mean and $\mathbb{E}[X_i^2] = 1$. Let $S_n = X_1 + \cdots + X_n$ be the random walk generated by the X_i. Then, almost surely,*

$$\limsup \frac{S_n}{\sqrt{2n \log\log n}} = 1$$

Proof. By Theorem 71, we may take stopping times $\tau_1 \leq \tau_2 \leq \dots$ such that B_{τ_n} is distributed as S_n and $\mathbb{E}[\tau_n] = \mathbb{E}[S_n^2] = n$. Note that $(\tau_k - \tau_{k-1})$ are iid and $\mathbb{E}[\tau_k - \tau_{k-1}] = \mathbb{E}[X_i^2] = 1$. Thus by Theorem 26,

$$\lim_{n\to\infty} \frac{\tau_n}{n} = \lim_{n\to\infty} \frac{\sum \tau_k - \tau_{k-1}}{n} = 1$$

almost surely. By Proposition 69 we are done. ∎

The above is a beautiful application of some of the power of Brownian motion. The method is important: we first took advantage of the fractal nature of Brownian motion to prove a longtime property and then leveraged this result to show a result in the discrete case. The bridge was, of course, Theorem 71. The reverse, too, can be accomplished with a companion theorem. Let S_n be the random walk in Theorem 75. Then we may make a continuous version by defining

$$S_t = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor) X_{\lfloor t \rfloor + 1}$$

We may scale the above to define a sequence of functions on $[0, 1]$ so that

$$S_t^{(n)} = \frac{S_{nt}}{\sqrt{n}}$$

Then we have

Theorem 76 (Donsker Invariance Principle)**.** *Under the metric induced by the sup-norm, $S_t^{(n)}$ converges in distribution to the standard Brownian motion on* $[0, 1]$*.*

This gives rigorous backing to the earlier statement that Brownian motion should be viewed as a random walk in continuous time; in a sense it can be viewed as the limit of random walks, as prescribed by the above theorem. We could then do some combinatorial analysis in discrete time and apply the result to Brownian motion.

Example 77. An example from [vdV12] involves the theory of empirical processes. Suppose that $X_1, \ldots$ are iid random variables from a distribution F. If nothing is known about F, a natural estimate for this function is the empirical distribution function given by

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq t\}}$$

A theorem of Glivenko and Cantelli says that

$$||F_n - F||_\infty \to 0$$

as $n \to \infty$ (for proof, see [Dur10, vdV12]). One might wonder if we can have finer information on the rate of convergence. Indeed we note that $nF_n(t)$ can be seen to be a random walk and using Theorem 75 and some more sophisticated machinery, it can be seen that

$$\limsup \sqrt{\frac{2n}{\log \log n}} ||F_n - F||_\infty \leq 1$$

almost surely.

Many more beautiful examples of the application of Theorems 71 and 76 can be found in [Mö10, Kar91, RY99, vdV12].

4 Stochastic Calculus

4.1 Can We Use Normal Calculus?

We have seen some of the basics of stochastic processes in continuous time, but it is now time to introduce the tool that really separates the continuous case from the discrete time case. Stochastic calculus allows one to do many computations and to analyze in depth some of the properties of the Brownian sample paths; [Mö10] contains many applications. More generally, stochastic calculus allows for the introduction of randomness into models given by deterministic differential equations. Applications of this are everywhere from finance to engineering; more about these can be found in [Kar91, Oks19, Law06].

Normal calculus begins with defining a derivative. For stochastic calculus, this will not work. Brownian motion has quite rough sample paths and they almost surely not differentiable anywhere (see [Mö10, Kar91] for proof); as Brownian motion is the most important example of a stochastic process in continuous time, we will need to look somewhere else. Thus, we define instead a stochastic integral. This is not quite as easy as it first appears either however. A naïve suggestion might be to integrate pathwise. That is to say, could we define for some continuous martingale M,

$$\left(\int_0^t X_s dM_s\right)(\omega) = \int_0^t X_s(\omega)dM_s(\omega)$$

where the integral on the right hand side is just a standard Stieltjes integral? The answer is no. To see this, recall that a Stieltjes integral is defined as the limit of the integral of simple approximations on some partition of the interval $[0,t]$. The proof that this integral is well defined in that the points where one takes the approximate value of the integrand are irrelevant to the final result requires that we have bounded first variation in M_s. Unfortunately, we have

Proposition 78. *A continuous square integral martingale with bounded first variation is constant.*

Proof. Without loss of generality, we may assume that $M_0 = 0$. Letting $V(B_t)$ be the first variation, we have $\tau_n = \inf\{t|V(B_t) > n\}$ and considering the stopped martingale M^{τ_n}, we may assume that $|M_t|, V(M_t) \leq n$ for some n. We have

$$\begin{aligned}\mathbb{E}[M_t^2] &= \mathbb{E}\left[\sum M_{t_{i+1}}^2 - M_{t_i}^2\right] = \mathbb{E}\left[\sum \mathbb{E}\left[M_{t_{i+1}}^2 - M_{t_i}^2|\mathcal{F}_{t_i}\right]\right] \\ &= \mathbb{E}\left[\sum \mathbb{E}\left[(M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_{t_i}\right]\right] = \mathbb{E}\left[\sum(M_{t_{i+1}} - M_{t_i})^2\right]\end{aligned}$$

By our assumption, then we have

$$\mathbb{E}[M_t^2] \leq \mathbb{E}\left[V\left(\sup\left|M_{t_i} - M_{t_{i-1}}\right|\right)\right] \leq \mathbb{E}\left[\sup\left|M_{t_i} - M_{t_{i-1}}\right|\right] \to 0$$

Thus $M_t = 0$ almost surely. This holds for all t and so we are done. ∎

Thus we see that we cannot define pathwise integrals and need to come up with an entirely new theory of integration for continuous martingales, a theory that is somehow more global and probability theoretic. We do this in the next section.

4.2 The Ito Integral

We saw in the last section that normal calculus is insufficient for our purposes. The question remains as to how we might construct a new integration theory. This question occupied the minds of many famous probabilists in the mid-20th century and there have been several formulations proposed. We stick with the formulation of Ito. In [PWZ33], Paley, Wiener, and Zygmund introduce the concept of integrating nonrandom functions with respect to Brownian motion and Ito introduced randomness into the integral that bears his name in [Itô44]. Look to [Kar91, Pro04] for historical accounts of this development. The intuition is simple and identical to that of the Riemann-Stieltjes integral: define an integral for simple functions and approximate. The issues here were first that it was unclear exactly which functions should be termed simple and second that it was unclear as to which functions could be approximated. We adopt a relatively nonrigorous treatment of the subject due to time constraints; look to [Kar91, RY99, Pro04, Law06] for more in depth analyses.

Suppose we had some discrete time process H_n where there are times $t_1, t_2, \ldots$ such that H jumps at t_n, i.e. if $t_n \leq t < t_{n+1}$ then $H_t = H_n$. Then it is obvious how we define the integral: we use Proposition 40 to get

$$\int_0^t H_t dB_t = \sum_k H_{t_k}(B_{t_k} - B_{t_{k-1}})$$

In particular, we would like H_t to be predictable because in this case we have that the integral is a martingale. This leads to the following definition:

Definition 79. We say that a random function is simple if there are $0 = t_0 < t_1 < \dots$ such that

$$f(t,\omega) = \sum f_j(\omega)\mathbb{1}_{[t_j,t_{j+1})}(t)$$

The above discussion immediately yields a definition for the integral of a simple function. The question is how well does this definition behave after taking the limits that we will need to take to make this nontrivial? The answer is not well without further restrictions as the following example from [Oks19] demonstrates:

Example 80. A natural way to approximate a continuous function $g_\omega(t)$ in the above way is to choose $f_j(\omega) = g_\omega(t_j^*)$ where $t_j^* \in [t_j, t_{j+1}]$. When we are integrating in the Stieltjes sense, we are restricted to integrating against functions of bounded first variation, where the choice of the endpoint does not matter. We will see that that is not the case here. We consider two sequences of simple functions

$$\begin{aligned} f_1(t,\omega) &= \sum_{k>0} B_{k2^{-n}}(\omega)\mathbb{1}_{[k2^{-n},(k+1)2^{-n})}(t) \\ f_2(t,\omega) &= \sum_{k>0} B_{(k+1)2^{-n}}(\omega)\mathbb{1}_{[k2^{-n},(k+1)2^{-n})}(t) \end{aligned}$$

The first function approximates Brownian motion from the left endpoint and the second approximates Brownian motion from the right endpoint. Thus as $n \to \infty$ we see that both functions approach Brownian motion. Ideally, that should mean that their integrals should be the same. Now, if we take the expected value, we see

$$\mathbb{E}\left[\int_0^t f_1 dB_s\right] = \sum \mathbb{E}\left[B_{k2^{-n}}(B_{(k+1)2^{-n}} - B_{k2^{-n}})\right] = 0$$

by the markov property. Note that this is expected because, as B_t is continuous, it must be predictable and so by Proposition 40 we have that the integral is a martingale. But we have

$$\mathbb{E}\left[\int_0^t f_2 dB_s\right] = \sum \mathbb{E}\left[B_{(k+2)2^{-n}}\left(B_{(k+1)2^{-n}} - B_{k2^{-n}}\right)\right] = t$$

Thus we see that these "integrals" certainly do not agree!

Thus if we wish to approximate a function we must choose a convention for which point in the interval $[t_j, t_{j+1})$ we wish to evaluate our function. We follow Ito and choose the leftmost point. The reason for this is because we would like the integral of a martingale to remain a martingale (at least in nice cases, in less nice cases we end up with a local martingale). Now we need to consider a class of functions that are integrable.

Definition 81. We say that $f \in \mathcal{I}$ is integrable if $(t,\omega) \to f(t,\omega)$ is $\mathcal{B}\times\mathcal{F}$ measurable, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}^+$ and $\mathcal{F}$ is the filtration on which our Brownian motion exists and $f(t,\omega)$ is $\mathcal{F}_t$ adapted and finally for all $t > 0$

$$\mathbb{E}\left[\int_0^t f^2 ds\right] < \infty$$

Note that clearly bounded simple functions are integrable. While the first two conditions make perfect sense, the last might be surprising. The motivation for this lies in the fact that Brownian motion has finite quadratic variation instead of finite first variation and thus we should naturally look to limits in L^2 as opposed to L^1. This intuition is on display in the following lemma:

Lemma 82 (Ito Isometry). *Let f be bounded and simple. Then for all $t > 0$,*

$$\mathbb{E}\left[\int_0^t f^2 dt\right] = \mathbb{E}\left[\left(\int_0^t f dB_t\right)^2\right]$$

Proof. Note that from the definition of Brownian motion and the markov property

$$\mathbb{E}\left[f_{i-1}\left(B_{t_i}-B_{t_{i-1}}\right)f_{j-1}\left(B_{t_j}-B_{t_{j-1}}\right)\right]=\begin{cases}0 & i\neq j\\ \mathbb{E}[f_{i-1}^2](t_i-t_{i-1})^2 & i=j\end{cases}$$

follow from the definition of Brownian motion. The result follows from the construction stochastic integral for simple processes. ■

We call Lemma 82 the Ito isometry because the equality (more precisely its general form when f is any integrable process) implies the Ito integral on $[0,t]$ is an isometry from the L^2 space associated to adapted square-integrable processes and square-integrable random variables. In this form, the proof of the equality follows trivially from the fact that The key is now to find an approximation of an integrable process by simple ones. The proof of the following proposition can be found, with varying degrees of generality, in [Oks19, Kar91, Pro04, RY99].

Proposition 83. *Let $X_t \in \mathfrak{I}$ be an integrable process. Then there exists a sequence of bounded, simple processes $X_t^{(n)}$ such that*

$$\sup_{t>0}\lim_{n\to\infty}\mathbb{E}\left[\int_0^t\left(X_t-X_t^{(n)}\right)^2dt\right]=0$$

Thus we may define

Definition 84. Let $X_t \in \mathfrak{I}$ be an integrable process. Letting $X_t^{(n)}$ be the process in Proposition 83, we define the stochastic integral

$$\int_0^t X_s dB_s := \lim_{n\to\infty}\int_0^t X_s^{(n)}dB_s$$

To see that this limit exists, we note that by Lemma 82, we have that $\left\{\int_0^t X_s^{(n)}dB_s\right\}$ is a cauchy sequence in L^2, which is complete and thus the limit exists. The fact that this limit does not depend on the choice of approximating $X^{(n)}$ follows from Lemma 82 as well. Thus we have constructed the Ito integral.

Example 85. As an example of calculating a stochastic integral, we will show that

$$\int_0^t 2B_s dB_s = B_t^2 - t$$

As our simple function approximations, we let

$$f_n(s,\omega)=\sum_k B_{\frac{kt}{n}}\mathbb{1}_{\left[\frac{kt}{n},\frac{(k+1)t}{n}\right)}(s)$$

Let $B_k = B_{\frac{kt}{n}}$ and let $\Delta_k = B_{k+1}-B_k$. Then we have

$$\mathbb{E}\left[\int_0^t(f_n-B_s)^2ds\right]=\mathbb{E}\left[\sum_k\int_{\frac{kt}{n}}^{\frac{(k+1)t}{n}}(B_s-B_k)^2ds\right]=\sum_k\int_{\frac{kt}{n}}^{\frac{(k+1)t}{n}}\left(s-\frac{kt}{n}\right)ds$$
$$=\frac{1}{2}\sum_k\left(\frac{(k+1)t}{n}-\frac{kt}{n}\right)^2=\frac{1}{2}\sum_k\frac{1}{n^2t}=\frac{1}{2nt}\to 0$$

Thus $f_n \to B_t$ in the appropriate sense. Now we note that

$$B_{k+1}^2-B_k^2=\Delta_k^2+2B_k\Delta_k$$

and so we have

$$B_t^2=\sum_k B_{k+1}^2-B_k^2=\sum_k\Delta_k^2+2B_k\Delta_k$$

and thus, rearranging, we see that

$$\sum_k 2B_k\Delta_k = B_t^2 - \sum_k \Delta_k^2$$

Taking limits, we see that the left hand side converges to the Ito integral and the right hand side convergest to $B_t^2 - t$.

We see that it is not necessarily easy to compute an Ito integral directly from the definition. Fortunately we have another way of doing this that we shall explore later on.

4.3 The Ito Formula and First Applications

Note that our construction of the Ito integral immediately yields some basic properties. We first note that we can define

$$\int_{t_0}^{t_1} X_s dB_s := \int_0^{t_1} X_s dB_s - \int_0^{t_0} X_s dB_s$$

Moreover, the integral is clearly linear and is mean zero. The fact that the integral (or a continuous modification at least) is a continuous square integrable martingale is not completely obvious, but the reader will be referred to the usual suspects ([Kar91, Pro04, RY99]) for proof in the interest of time.

We now extend our focus slightly to integrals that take time evolution into account as well.

Definition 86. We say that X_t is an Ito process if there are integrable and possibly random b, σ such that

$$X_t = X_0 + \int_0^t b(s,\omega)ds + \int_0^t \sigma(s,\omega)dB_s$$

We will often have cause to denote this as

$$dX_t = b(t)dt + \sigma(t)dB_t \qquad\qquad X_0 = Y$$

where the dependence on ω of the random functions b, σ is implicit.

While we have defined a stochasatic integral, it is still far from clear how we might actually compute anything with it. To do this we will need the stochastic equivalent of the chain rule. We need the following definiton:

Definition 87. Let X_t be an Ito process. We define its quadratic variation as

$$\langle X\rangle_t = V^{(2)}(X_t) = \lim_{n\to\infty}\sum_{k=1}^n \left|X_{\frac{(k+1)}{n}t} - X_{\frac{k}{n}t}\right|^2$$

Remark 88. While it was a major hole to omit discussion of the Doob decomposition theorem, this is the motivation for why the quadratic variation is important. In particular, if M is a continuous square integrable martingale, then $\langle M\rangle_t$ is the unique increasing, predictable process such that $M_t^2 - \langle M\rangle_t$ is a martingale. In the discrete case, there is a good discussion in [Chu00, Dur10]. For the continuous case, see [RY99, Kar91, Pro04].

Example 89. We have seen in the problem set that $\langle B\rangle_t = t$. If we suppose that $Y_n = \int X_s dB_s$ and X is simple then we have

$$\langle Y\rangle_t = \sum_i X_{t_j}^2\left(B_{t_{j+1}} - B_{t_j}\right)^2$$

Taking limits we arrive at

Proposition 90. *Let $b \in C^1$ and sufficiently nice*[1] *almost surely and let*

$$Y_t = Y_0 + \int_0^t b(s)ds + \int_0^t X_s dB_s$$

Then

$$\langle Y \rangle_t = \int_0^t X_s^2 ds$$

Proof. Translation does not affect quadratic variation so without loss of generality $Y_0 = 0$. The process $\int_0^t b(s)ds$ has finite first variation so adding it does not affect quadratic variation and thus only the last term contributes. We note, however, that if X is simple, we have

$$\begin{aligned}
\mathbb{E}\left[\left(\int_s^t X_s^2 dBs\right)|\mathcal{F}_s\right] &= \mathbb{E}\left[\left(X_0(B_{t_1} - B_s) + \sum X_i(B_{t_{i+1}} - B_{t_i}) + X_n(B_t - B_{t_n}\right)\middle|\mathcal{F}_s\right] \\
&= \mathbb{E}\left[X_1^2(B_{t_1} - B_s)^2 + \sum X_i^2(B_{t_{i+1}} - B_{t_i})^2 + X_n^2(B_t - B_{t_n})^2\middle|\mathcal{F}_s\right] \\
&= \mathbb{E}\left[X_1^2(t_1 - s) + \sum X_i^2(t_{i+1} - t_i) + X_n^2(t - t_n)\middle|\mathcal{F}_s\right] \\
&= \mathbb{E}\left[\int_s^t X_u^2 dt\middle|\mathcal{F}_s\right]
\end{aligned}$$

Thus by the time additivity and Lemma 82, we have equality in L^2 which suffices. Everything above passes through limits in L^2 so we are done. ∎

We are now ready to set up the product rule of stochastic calculus:

Lemma 91 (Product Rule)**.** *Let X_t, Y_t be Ito processes*

$$\begin{aligned}
X_t &= X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s \\
Y_t &= Y_0 + \int_0^t b_s' ds + \int_0^t \sigma_s' dB_s
\end{aligned}$$

Then

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \int_0^t \sigma_s \sigma_s' ds$$

Remark 92. Note that this is the normal product rule with an extra term. Where does the extra integral with respect to s come in? The answer is the quadratic variation! In the proof below we will see a sum that includes terms with a $(\Delta B_t)^2$ in them. In normal calculus this would be a $(\Delta t)^2$ and disappears when we take limits. In this case, because B_t has finite quadratic variation, one can think of $(\Delta B_t)^2 \approx \Delta t$ in the limit sense and the extra term appears as we take the limit.

Proof. Note that Because $(X+Y)^2 - X^2 - Y^2 = 2XY$ and integrals are linear, it suffices to prove the result for $X = Y$. Now we note that

$$\sum \left(X_{t_{j+1}} - X_{t_j}\right)^2 = X_t^2 - X_0^2 + 2\sum X_{t_j}(X_{t_{j+1}} - X_{t_j})$$

Taking limits we see that

$$X_t^2 = X_0^2 + 2\int_0^t X_s dX_s + \langle X \rangle_t$$

We are done by Proposition 90. ∎

[1]See [Kar91, Pro04, RY99] for details of what we mean by "nice."

In fact, the product rule in some sense captures all of the difference between normal calculus and stochastic calculus. Thus we see a simplified version of Ito's rule as follows:

Proposition 93. *Let X_t be an Ito process and let $f \in C^2$. Then*

$$f(X_t) = f(X_0) + \frac{1}{2}\int_0^t f''(X_s)ds + \int_0^t f'(X_s)dX_s$$

This will often be written as

$$df(X_t) = \frac{1}{2}f''(X_t)dt + f'(X_t)dX_t$$

Proof. Because we can stop X at any time $T > 0$, it suffices to prove the above for times lying in a compact set in $\mathbb{R}$. By the well known fact that continuous functions on compact sets of $\mathbb{R}$ can be uniformly approximated by polynomials (Problem 12), we have that it suffices to show the result on polynomials. By linearity it suffices to show the result for monomials. But an easy computation with Lemma 91 gives that it holds for all functions $f(x) = x^k$ for $k \in \mathbb{N}$. Thus we are done. ■

A more general version of the statement of Ito's rule is the following:

Theorem 94 (Ito's Rule)**.** *Let X_t be a (d-dimensional) Ito process and let $f(t, x)$ be a function that is C^1 in t and C^2 in each coordinate of x. Then*

$$f(t, X_t) = f(0, X_0) + \sum_{i=1}^{d}\int_0^t \frac{\partial f}{\partial x_i}(s, X_s)dX_i + \int_0^s \frac{\partial f}{\partial t}(s, X_s)ds + \frac{1}{2}\int_0^t \sum \frac{\partial^2 f}{\partial X_i^2}d\langle X_i\rangle$$

In differential form this becomes

$$df(X_t) = \frac{\partial f}{\partial t}dt + \sum \frac{\partial f}{\partial X_i}dX_i + \frac{1}{2}\sum \frac{\partial^2 f}{\partial x_i^2}dt$$

A proof of the more general result can be found in any of the standard references, although [RY99] is nice because it follows the method developed above. Another proof uses Taylor's formula and can be found in [Kar91, Law06]. We are now ready for some examples.

Example 95. Suppose we wanted to know what $\int_0^t B_s dB_s$ was. To evaluate this, we need to introduce a function such that $f'(x) = x$. From calculus then we find

$$d(B_t^2) = dt + 2B_t dB_t$$

or

$$\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - t)$$

Note that this provides another proof that $B_t^2 - t$ is a martingale.

Example 96. Note that if B_t is d-dimensional standard Brownian motion then $dB_t = dB_t$ and so by Theorem 94, we have

$$df(B_t) = \sum \frac{\partial f}{\partial x_i}dB_t^{(i)} + \frac{1}{2}\Delta f dt = (\nabla f) \cdot dB_t + \frac{1}{2}\Delta f dt$$

Thus we note that if f is harmonic then $f(B_t)$ is a martingale. Similar facts hold if f is super- or subharmonic.

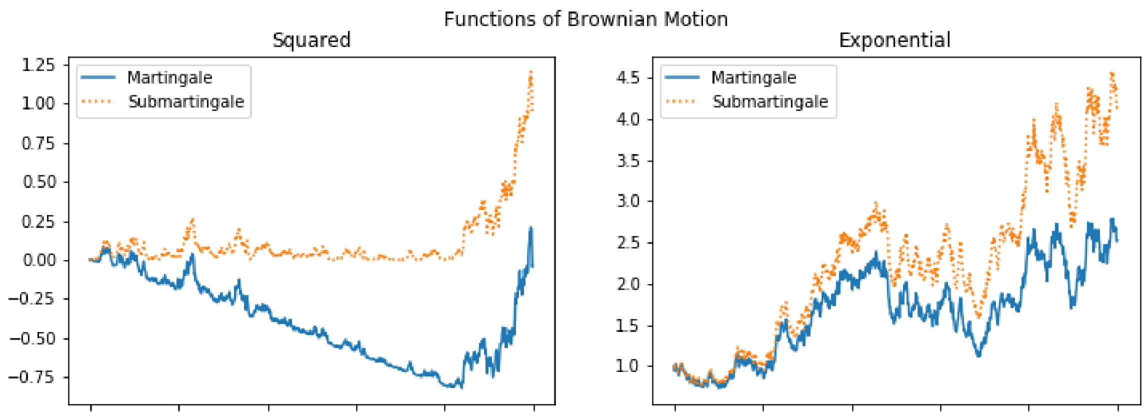

Figure 4: Sample paths for $f(B_t)$ where f is a function and B_t is a standard Brownian motion. The martingale is "corrected" function and the submartingale (the dashed line) is the "uncorrected" function. On the left $f(x) = x^2$ corrected by subtracting t and on the right $f(x) = e^x$, corrected by a factor of $e^{-\frac{1}{2}t}$.

Example 97. In normal calculus we define the exponential function as the unique function whose derivative is itself. What if we try to do this for Brownian motion? We will define $\mathcal{E}(B_t) = X_t$ as a process that satisfies $dX_t = X_t dB_t$. We have not explored any questions relating to stochastic differential equations, so it is far from clear that this equation actually admits a solution. If we were trying to guess a solution, though, we might guess that $X_t = f(t, B_t)$ for some function f. If we further assume that f is C^1 in t and C^2 in B_t, then by Theorem 94, we have

$$X_t dB_t = f(t, B_t)dB_t = dX_t = df(t, B_t) = \frac{\partial f}{\partial x}(t, B_t)dB_t + \left(\frac{\partial f}{\partial t}(t, B_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, B_t)\right) dt$$

Thus we have by equating terms

$$\frac{\partial f}{\partial x} = f \qquad\qquad \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0$$

The first equation implies that there is some function $g(t)$ such that $f(t, x) = e^{x+g(t)}$. Plugging this into the second equation, we see that $\frac{1}{2} + g'(t) = 0$. Thus we get

$$\mathcal{E}(B_t) = e^{B_t - \frac{1}{2}t}$$

Checking now, we see that $\mathcal{E}$ is indeed C^1 in t and C^2 in B_t and thus we may apply Theorem 94 and conclude that X_t satisfies the above differential equation. We could generalize this process to some continuous square integrable martingale M_t and define

$$\mathcal{E}(M_t) = \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right)$$

the stochastic exponential. This exponential function is incredibly important in the study of stochastic differential equations.

Example 98. A generalization of the previous example is geometric Brownian motion. If we have constants a, b then we want to consider a process X_t such that

$$\frac{dX_t}{X_t} = adt + bdB_t$$

Once again, it is not obvious at all from the theory developed thus far that there exists such a process. Integrating, we see that

$$\int_0^t \frac{dX_s}{X_s} = at + bB_t$$

But we see also that

$$d(\log X_t) = \frac{dX_t}{X_t} - \frac{1}{2X_t^2} d\langle X\rangle_t = \frac{dX_t}{X_t} - \frac{1}{2X_t^2}(b^2X_t^2 dt) = \frac{dX_t}{X_t} - \frac{b^2}{2}dt$$

by Proposition 90. But then plugging in we have

$$\int_0^t \frac{dX_s}{X_s} = \log\left(\frac{X_t}{X_0}\right) + \frac{b^2}{2t} = at + bB_t$$

and thus

$$X_t = X_0 \exp\left(\left(a - \frac{1}{2}b^2\right)t + bB_t\right)$$

Note that if $a = 0$ and $b = 1$ then we have recovered $\mathcal{E}(B_t)$. Our intuition might suggest that $\mathbb{E}[X_t] = \mathbb{E}[X_0]e^{at}$. To see that this is indeed the case, let $Y_t = e^{bB_t}$ so that $X_t = Y_t e^{at-\frac{1}{2}b^2 t}$. By Theorem 94, we have

$$Y_t = 1 + b\int_0^t Y_s dB_s + \frac{1}{2}b^2 \int_0^t Y_s ds$$

Taking expectations, applying Theorem 6, and recalling that the stochastic integral is a martingale, we have

$$\mathbb{E}[Y_t] = 1 + \frac{1}{2}b^2 \int_0^t \mathbb{E}[Y_s]ds$$

Solving this differential (integral?) equation, we get that

$$\mathbb{E}[Y_t] = e^{\frac{1}{2}b^2 t}$$

Thus we have

$$\mathbb{E}[X_t] = \mathbb{E}[Y_t e^{at-\frac{1}{2}b^2 t}] = e^{at}$$

which is as expected.

The stochastic exponenetial is important in many contexts, not least the martingale representation theorem. We have noted that the stochastic integral with respect to Brownian motion is a continuous square-integrable martingale, but in fact the converse is true as well. For a rigorous proof see [Kar91, Oks19, Pro04, RY99]. The general idea is that for fixed $t > 0$, the linear span of random variables of the form

$$\exp\left(\int_0^t f dB_s - \frac{1}{2}\int_0^t f^2 ds\right)$$

with f non-random is dense in the L^2 space associated to the probability measure associated to Brownian motion on $[0, t]$ and thus the stochastic exponential provides a way of representing M_t as an integral with respect to B_t. This is not the only way of representing martingales in terms of Brownian motion; random time change is another important way.

Much like Theorem 71, we wish to "embed" a continuous martingale in Brownian motion using stopping times. In this case, however, we need some random continuous "clock" that changes the distribution of the Brownian motion. Before we can do this, however, we need an important theorem of P. Lèvy.

Theorem 99 (Lèvy Characterization of Brownian Motion)**.** *Let M_t be a continuous martingale (with respect to some filtration $\mathcal{F}_t$) such that $\langle M\rangle_t = t$ and $M_0 = 0$. Then M_t is Brownian motion.*

Proof. Applying the argument in Example 98, we note that

$$\mathbb{E}\left[e^{i\lambda M_t}\right] = e^{-\frac{\lambda^2 t}{2}}$$

because $\langle M\rangle_t = t$. Thus we note that

$$\mathbb{E}\left[e^{i\lambda(M_t - M_s)}|\mathcal{F}_s\right] = e^{-\frac{la^2}{2}(t-s)}$$

and so by Proposition 24, we have that $M_t - M_s$ is distributed as $N(0, t-s)$ independent of $\mathcal{F}_s$. Thus we have stationarity, independence, and Gaussianity. By assumption, M_t is continuous so M_t is Brownian motion. ∎

We are now able to prove our representation theorem.

Theorem 100 (Martingale Representation Theorem)**.** *Let B_t be a standard linear Brownian motion and let M_t be a continuous, square integrable martingale such that $M_0 = 0$. Then there exists a (possibly random) function $\alpha(t)$ such that M_t and $B_{\alpha(t)}$ are identically distributed as stochastic processes.*

Remark 101. Note that once again, just as we remarked after Theorem 63 and Proposition 65, this is a stronger statement than saying that for all t, $M_t \overset{d}{=} B_{\alpha(t)}$.

Proof. For $t \geq 0$, let

$$\tau_t = \inf\{s | \langle M\rangle_s = t\}$$

Because quadratic variation is increasing and continuous, the function $t \mapsto \tau_t$ is increasing and by Theorem 99, we have that M_{τ_t} is standard Brownian motion. Let $\alpha(t) = \langle M\rangle_t$. Then we have

$$B_{\alpha(t)} = B_{\langle M\rangle_t} = M_{\tau_{\langle M\rangle_t}} = M_t$$

as desired. ∎

The above fact is fundamental to the study of mathematical finance. It also provides evidence for the intuition that Brownian motion is the "fundamental continuous stochastic process" as, in the sense described above, other (square-integrable) continuous martingales are just Brownian motions on different clocks. For more information on time change see [Oks19, Kar91]. For a more general treatment of stochastic integration, see [Kar91, Pro04, RY99].

5 Brownian Motion and Partial Differential Equations

There are many connections between the study of Brownian motion and PDEs. These connections are explained beautifully and in great depth in [Kar91] and our treatment of the subject is taken almost entirely from that source. In the interest of time, we seek to accomplish only two goals: first to illustrate how the study of Brownian motion can help solve a problem from the study of PDEs and second to demonstrate how the study of PDEs can do the same for Brownian motion.

5.1 The Dirichlet Problem

For the remainder of the section we let B_t be a standard Brownian motion in d dimensions. If $A \subset \mathbb{R}^d$ is some (open and bounded) domain then we fix

$$\tau_A = \inf\{t \geq 0 | B_t \notin A\}$$

We have seen that this is a stopping time. Moreover, because A is bounded, we know that $\mathbb{P}(\tau_A < \infty) = 1$. We fix the notation $\mathbb{P}_x(\cdot)$ as the probability measure of B_t started such that $B_0 = x$ and $\mathbb{E}_x[\cdot]$ as the expectation with respect to this probability measure. Finally, we define a probability measure on ∂A for $x_0 \in A$, which is given by

$$\mu_{A,x_0}(dx) = \mathbb{P}_{x_0}\left(B_{\tau_A} \in dx\right)$$

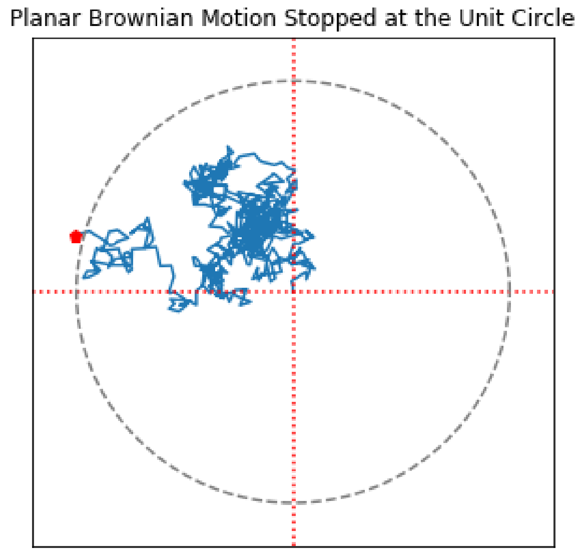

Figure 5: A planar Brownian motion started at the origin and stopped when it hits the unit circle.

Example 102. We let $D_r = D_r(0)$ be the disk of radius r around the origin. In this case, we note that $\mu_{D_r(x_0),x_0}$ is just the uniform measure on the sphere. To see this, we just note that Brownian motion is invariant under the action of the orthogonal group. An example is seen in Figure 5.

This section is focused on solutions to the classical Dirichlet problem. Namely, given a domain A and a function f on ∂A, we want to find a continuous function u such that

$$\begin{cases} \Delta u = 0 & \text{in } A \\ u = f & \text{on } \partial A \end{cases}$$

More specifically, we want $u \in C(\overline{A}) \cap C^2(A)$ that is harmonic in the interior and continuous up to the boundary with prescribed boundary values given by f. At first sight, this has nothing to do with Brownian motion. The first connection comes from Example 96: namely, a harmonic function of Brownian motion is a martingale. This leads to the following result:

Proposition 103. *A function u is harmonic, i.e., $\Delta u = 0$, in A if and only if it satisfies the mean value property for all $x \in A$, namely that for all $x \in A$ and for all $r > 0$ such that $D_r(x) \subset A$, we have*

$$u(x) = \frac{1}{|\partial D_r|} \int_{\partial D_r(x)} u(y)dy$$

Proof. We prove the if direction. The "only if" direction is purely analytic and thus less relevant to the current discussion. We refer to [Kar91, Gil01] for those interested.

Let u be harmonic. By Theorem 94, we have for all $t > 0$ that

$$u\left(B_{t\wedge\tau_{D_r}}\right) = u(B_0) + \sum_{i=1}^{d} \int_0^{t\wedge\tau_{D_r}} \frac{\partial f}{\partial x_i}(B_s)dB_s + \int_0^{t\wedge\tau_{D_r}} \Delta u(B_s)ds = u(B_0) + M_t$$

where M_t is a continuous martingale by the fact that it is the stochastic integral of a (bounded) continuous martingale and the fact that $\Delta u = 0$. Taking expectations and applying bounded convergence, we have

$$u(x) = u(B_0) = \mathbb{E}_x[u(B_0)] = \mathbb{E}_x\left[u\left(B_{t\wedge\tau_{D_r}}\right)\right] \to \mathbb{E}_x\left[u\left(B_{\tau_{D_r}}\right)\right] = \mathbb{E}_{\mu_{D_r(x),x}}\left[u\left(y\right)\right] = \frac{1}{|\partial D_r|}\int_{\partial D_r(x)} u(y)dy$$

and thus we have the mean value property. ■

This proposition immediately yields the maximum principle:

Corollary 104. *Let u be harmonic on A, a bounded domain. Then*

$$\sup_{\overline{A}} u = \sup_{\partial A} u$$

the maximum of a harmonic function is attained on the boundary.

Notice in this proof that we relied on the function $\mathbb{E}_x\left[u\left(B_{\tau_A}\right)\right]$. Now note that $B_{\tau_A} \in \partial A$ and so if u is a solution to the Dirichlet problem then we would have $u\left(B_{\tau_A}\right) = f\left(B_{\tau_A}\right)$ and so the above function would become $u(x) = \mathbb{E}_x\left[f\left(B_{\tau_A}\right)\right]$. We might hope that, given the argument in Proposition 103, we have that u is a harmonic function. In fact, modulo integrability concerns, this is exactly what we have.

Proposition 105. *Let A be a domain and let $f : \partial A \to \mathbb{R}$ be continuous. Suppose that for all $x \in A$ that we have $u(x) = \mathbb{E}_x\left[f\left(B_{\tau_A}\right)\right] < \infty$ and well-defined. Then u is harmonic in A and $u|_{\partial A} = f$.*

Proof. Note that if $x \in \partial A$ then $\tau_A = 0$ and so

$$u(x) = \mathbb{E}_x\left[f\left(B_{\tau_A}\right)\right] = \mathbb{E}_x\left[f(B_0)\right] = f(x)$$

and so the second statement is clear. To prove the first statement, we note that

$$u(x) = \mathbb{E}_x\left[f\left(B_{\tau_A}\right)\right] = \mathbb{E}_x\left[\mathbb{E}_x\left[f\left(B_{\tau_A}\right)|\mathcal{F}_{\tau_{D_r(x)}}\right]\right] = \mathbb{E}_x\left[u\left(B_{\tau_{D_r}}\right)\right] = \frac{1}{|\partial D_r|}\int_{\partial D_r(x)} u(y)dy$$

which is the mean value property. By Proposition 103 we are done. ■

Thus we have one harmonic function on A that agrees with f on the boundary. One might wonder if there are others, however. In the case of bounded f, there are not.

Proposition 106. *Suppose A is a domain such that for all $x \in A$, $\mathbb{P}_x(\tau_A < \infty) = 1$ and $f : \partial A \to \mathbb{R}$ is continuous and bounded. Then if u solves the Dirichlet problem for (A, f) then $u(x) = \mathbb{E}_x\left[f\left(B_{\tau_A}\right)\right]$.*

Proof. Let u be a solution to the Dirichlet problem and let

$$A_n = \left\{x \in A \middle| \inf_{y \in \partial A} |x - y| > \frac{1}{n}\right\}$$

By Theorem 94 and the fact that $\Delta u = 0$, we have

$$u\left(B_{t \wedge \tau_{D_n(x)} \wedge \tau_{A_n}}\right) = u(B_0) + \sum_{i=1}^{d} \int_0^{t \wedge \tau_{D_n(x)} \wedge \tau_{A_n}} \frac{\partial u}{\partial x_i}(B_s) dB_s^{(i)} = u(B_0) + X_s$$

Note that $\frac{\partial u}{\partial x_i}$ is bounded in $\overline{B_n \cap A_n}$ so X_s is a martingale and since the integral at $t = 0$ vanishes, after we take expectations we get

$$\mathbb{E}_x\left[u\left(B_{t \wedge \tau_{D_n(x)} \wedge \tau_{A_n}}\right)\right] = \mathbb{E}_x[u(B_0)] = u(x)$$

Now we may take $t, n \to \infty$ but we see that f is bounded and so by Corollary 104 we have that u is bounded as well, so we may apply bounded convergence and so the result holds. ■

It seems like we should be done now. With some not very strong conditions on f, conditions moreover that are obviously satisfied if A is bounded and so ∂A is compact, we have at most one harmonic function u that agrees with f on the boundary. We do not have, however, that u is continuous necessarily. One might argue that u is certainly continuous on A, after all it is harmonic and it can be shown without too much effort that Proposition 103 implies that u is continuous. The problem, however is that it is not at all clear that for all $x_0 \in \partial A$ we have

$$\lim_{\substack{x \to x_0 \\ x \in A}} u(x) = u(x_0)$$

In fact, this is not always true. To see this, consider the following example:

Example 107. Let $d \geq 2$ and consider $A = D_1(0) \setminus \{0\}$ and let f be a continuous function on ∂A. Note that $\partial A = \partial D_1 \cup \{0\}$. If there does not exist a solution to the Dirichlet problem for $f|_{\partial D_1}$ and D_1 then we are done (there is a solution, as we shall see below). If there does exist a solution call it $\widetilde{u}$. By Proposition 106 this solution is unique. In particular if $f(0) \neq \widetilde{u}(0)$ then there certainly can not be a solution to the Dirichlet problem given by (A, f)!

The question is, what went wrong? Intuitively, it seems like the u defined in Proposition 105 should be continous as when we change where we start by a little bit, it should not greatly affect where we end up on average. This intuition is close to being true, but misses an important point. We saw in the problem sets that given a Brownian motion in $d \geq 2$ dimensions started at x and a point $y \in \mathbb{R}^d$ then the probability that $B_t = y$ for some t is zero. However, clearly, the probability that $B_t = x$ for some t is 1 as $B_0 = x$. Thus if we have an isolated point on the boundary, then we can never expect to be able to guarantee the continuity of u. This intuition motivates the following definition.

Definition 108. Let A be a domain and define $\sigma_A = \inf\{t > 0 | B_t \notin A\}$. Note that this is in contrast to τ_A where the infimum is taken over $t \geq 0$. We say that a point $x \in \partial A$ is regular if $\mathbb{P}_x(\sigma_A = 0) = 1$. Otherwise a point is irregular.

Remark 109. Note that the event $\{\sigma_A = 0\} \in \mathcal{F}_0^+$ and so by Proposition 60 we have that if x is irregular then $\mathbb{P}_x(\sigma_A = 0) < 1$ and so it is zero.

Note that a point $x \in \partial A$ is regular if and only if a Brownian motion started at x does not immediately enter A and stay there for some nonzero amount of time. Thus an isolated boundary point is certainly not regular.

Example 110. In dimension $d = 1$ we see that all points are regular. By translation it suffices to suppose that $0 \in \partial A$ and show that 0 is regular. Let $T = \inf\{t > 0 | B_t > 0\}$. Note that for all $\varepsilon > 0$ we have that $\mathbb{P}(T < \varepsilon) \geq \frac{1}{2}$ because $\mathbb{P}(B_\varepsilon > 0) = \frac{1}{2}$. Thus $\mathbb{P}(T = 0) \geq \frac{1}{2}$; but we also have $\{T = 0\} \in \mathcal{F}_0^+$ and so by Proposition 60 $T = 0$ almost surely. By symmetry, the result holds if T is replaced by the first time that B_t is negative. But then we have, because $0 \in \partial A$ that $\mathbb{P}_0(\sigma_A = 0) = 1$.

The above example suggests that the linear case is somewhat boring; of course, we could have guessed this because the linear case can be solved directly by integrating and shown that the only solutions are affine functions, which is hardly the most interesting result. In higher dimensions, however, there is something interesting.

Theorem 111. *Let $d \geq 2$ and consider a Dirichlet problem (A, f). Then the following are equivalent:*

i) For $x_0 \in \partial A$ and for all bounded, measurable $f : \partial A \to \mathbb{R}$, we have

$$\lim_{\substack{x \to x_0 \\ x \in A}} \mathbb{E}_x \left[f \left(B_{\tau_A} \right) \right] = f(x_0)$$

ii) $x_0 \in \partial A$ is regular for A

iii) For all $\varepsilon > 0$, we have

$$\lim_{\substack{x \to x_0 \\ x \in A}} \mathbb{P}_x \left[\tau_A > \varepsilon \right] = 0$$

Proof. By translation we may assume that $x_0 = 0$. We begin by proving that (i) implies (ii). Suppose that 0 is irregular; we will show that (i) cannot hold. By irregularity and Proposition 60, we have $\mathbb{P}(\sigma_A = 0) = 0$ and thus

$$\lim_{r \to 0} \mathbb{P}_0 \left(B_{\sigma_A} \in D_r \right) = \mathbb{P}_0 \left(B_{\sigma_A} = 0 \right) = 0$$

because (in dimension d) Brownian motion misses points. Thus let us fix an $r > 0$ such that $\mathbb{P}_0 \left(B_{\sigma_A} \in D_r \right) < \frac{1}{4}$ and a sequence δ_n such that $0 < \delta_n < r$ and $\delta_n \downarrow 0$ and $\tau_n = \inf\{t | |B_t| > \delta_n\}$. We will show that for sufficiently large n, there exists an $x_n \in A \cap D_{\delta+n}$ such that $\mathbb{P}_{x_n} \left(B_{\tau_A} \in D_r \right) \leq \frac{1}{2}$. To see this note that

$\tau_n \downarrow 0$ almost surely and so $\mathbb{P}_0(\tau_n < \sigma_A) \to 1$ almost surely. On $\{\tau_n < \sigma_A\}$ we have $B_{\tau_n} \in A$ and so for $n \gg 0$, we have $\mathbb{P}_0(\tau_n < \sigma_A) > \frac{1}{2}$. But then we have

$$\begin{aligned}\frac{1}{4} &> \mathbb{P}_0\left(B_{\sigma_A} \in D_r\right) \geq \mathbb{P}_0\left(\{B_{\sigma_A} \in D_r\} \cap \{\tau_n < \sigma_A\}\right) = \mathbb{E}_0\left[\mathbb{1}_{\{\tau_n<\sigma_A\}}\mathbb{P}_0\left(B_{\sigma_A} \in D_r \middle| \mathcal{F}_{\tau_n}\right)\right] \\ &= \int_{A\cap B_{\delta_n}} \mathbb{P}_x\left(B_{\sigma_A} \in D_r\right)\mathbb{P}_0\left(\{\tau_n < \sigma_A\} \cap \{B_{\tau_n} \in dx\}\right) \geq \frac{1}{2}\inf_{A\cap D_{\delta_n}} \mathbb{P}_x\left(B_{\sigma_A} \in D_r\right)\end{aligned}$$

Thus there is some x_n such that $\mathbb{P}_{x_n}\left(B_{\tau_A} \in D_r\right) \leq \frac{1}{2}$. Now suppose that f is bounded and continuous on $\partial A \to \mathbb{R}$ such that f vanishes outside of D_r and $f \leq 1$ in D_r and $f(0) = 1$. Then if (i) holds, we have

$$f(0) \leq \limsup \mathbb{E}_{x_n}\left[f\left(B_{\sigma_A}\right)\right] \leq \limsup \mathbb{P}_{x_n}\left(B_{\sigma_A} \in D_r\right) \leq \frac{1}{2} < 1 = f(0)$$

which is a contradiction.

We now show that (ii) implies (iii). If $0 < \delta < \varepsilon$ then define

$$f_\delta(x) = \mathbb{P}_x\left(\{B_s \in A | \delta \leq s \leq \varepsilon\}\right) = \mathbb{E}\left[\mathbb{P}_{B_\delta}(\tau_A > \varepsilon - \delta)\right] = \int_{\mathbb{R}^d} \mathbb{P}_y(\tau_A > \varepsilon - \delta)\mathbb{P}_x(B_\delta \in dy)$$

Note that f_δ is continuous for $\delta > 0$ because B_t is continuous. Let

$$f(x) = \mathbb{P}_x\left(\{B_s \in A | 0 < s \leq \varepsilon\}\right) = \mathbb{P}(\sigma_A > \varepsilon)$$

Then we see that $f_\delta \downarrow f$ pointwise and so f is the infimum of a family of continuous functions and so f is upper semi-continuous. Now, $\sigma_A \geq \tau_A$ so we have by the upper semi-continuity of f that

$$\limsup_{x\to 0} \mathbb{P}_x(\tau_A > \varepsilon) \leq \limsup_{x\to 0} f(x) \leq f(0) = 0$$

where the last equality comes from our assumption of (ii).

Finally we show that (iii) implies (i). We know that for all $r > 0$, by continuity we have

$$\mathbb{P}_x\left(\max_{t\in[0,\varepsilon]} |B_t - B_0| < r\right) \to 1$$

as $\varepsilon \to 0$ and that this probability is independent of x. Thus we have

$$\mathbb{P}_x(|B_{\tau_A} - B_0| < r) \geq \mathbb{P}_x\left(\left\{\max_{[0,\varepsilon]} |B_t - B_0| < r\right\} \cap \{\tau_A \leq \varepsilon\}\right) \geq \mathbb{P}_0\left(\max_{t\in[0,\varepsilon]} |B_t| < r\right) - \mathbb{P}_x(\tau_A > \varepsilon)$$

Letting $x \to 0$ and $\varepsilon \downarrow 0$ we get

$$\lim_{\substack{x\to 0\\ x\in A}} \mathbb{P}_x(|B_{\tau_A} - x| < r) = 1$$

But f is continuous at the origin and bounded on ∂A so we may apply the bounded convergence theorem to get

$$\lim_{\substack{x\to 0\\ x\in A}} \mathbb{E}_x\left[f\left(B_{\tau_A}\right)\right] = f(0)$$

and so we are done. ■

A trivial corollary of Theorem 111 and Proposition 105 is

Corollary 112. *If A is a domain such that all boundary points are regular and $f : \partial A \to \mathbb{R}$ is bounded and continuous, then there exists a unique solution to the Dirichlet problem (A, f).*

Remark 113. While such a result is also available in the theory of PDEs, the use of probability is much nicer as all of the technical computations are done under the hood after the establishment of Theorem 94. Moreover, the above method provides a natural way to numerically approximate the value $u(x)$ of a solution at a particular point; just take a large number of independent Brownian motions started at x and stop them when they leave A, then take the empirical average of $f(B_{\tau_A})$. By Theorem 26, this converges to the mean. In fact, this method is an incredibly fast approach to the problem and is widely used.

Now the only problem is to determine which points are regular. We have already seen one example of irregular points (namely isolated boundary points), but in dimensions $d \geq 3$ there are nonisolated boundary points that can be irregular (a classic example is Lebesgue's Thorn; for more information see [Kar91, Mö10]). There is a strict condition for when a boundary point is regular, called Wiener's Test, which lies outside the scope of this talk (see [Mö10, Theorem 8.37] for more information), but we present a slightly easier condition, called the exterior cone condition.

Definition 114. For $y \in \mathbb{R}^d \setminus \{0\}$ and $0 \leq \theta \leq \pi$, we define the cone

$$C_\theta(y) = \{x \in \mathbb{R}^d | x \cdot y \geq |x||y| \cos\theta\}$$

the set of x whose planar angle relative to y is at most θ. We say that a point $x_0 \in \partial A$ satisfies the exterior cone condition if there is some $y \neq 0$ and some $\theta \in (0, \pi)$ such that $x_0 + C_\theta(y) \subset \mathbb{R}^d \setminus A$.

The intuition behind this definition might be a little bit vague, but it comes from the idea that the existence of an exterior cone at this point makes the point smooth in a certain sense which will then imply regularity. As one might expect, we have

Proposition 115. *If $x_0 \in \partial A$ satisfies the exterior cone condition then it is regular for A.*

Proof. After translation, we may assume that $x_0 = 0$. Let $y \neq 0$ and $\theta \in (0, \pi)$ such that $C_\theta(y) \subset \mathbb{R}^3 \setminus A$. Now we note that the the map $x \mapsto \frac{x}{\sqrt{t}}$ is a bijection on $C_y(\theta)$. Thus we have by Proposition 53 that

$$\mathbb{P}(B_t \in C_\theta(y)) = \mathbb{P}\left(\frac{B_t}{\sqrt{t}} \in C_\theta(y)\right) = \mathbb{P}(B_1 \in C_\theta(y)) = p > 0$$

because the gaussian measure and the lebesgue measure are mutually absolutely continuous. But note that p does not depend on t. Thus we have

$$\mathbb{P}_0(\sigma_A \leq t) \geq \mathbb{P}_0(B_t \in C_\theta(y)) = p$$

and thus, sending $t \downarrow 0$ we have that $\mathbb{P}_0(\sigma_A = 0) \geq p > 0$. By Proposition 60, we have that $\mathbb{P}_0(\sigma_A = 0) = 1$ and so we are done. ■

In some sense, the proof helps explain why we chose a cone; after all, we need some set that is invariant under scaling by any positive factor.

We now have a wealth of information about the Dirichlet problem and can demonstrate (and approximate) solutions for a large class of sets A and functions f. We now turn to a way that the theory of differential equations can help us solve a problem in probability.

5.2 Feynman-Kac and an Application

We now turn to a different kind of differential equation, a homogeneous parabolic one. If f is continuous on $\mathbb{R}^d$ and $k : \mathbb{R}^d \to [0, \infty)$ continuous, then we say that u solves a parabolic differential equation with coefficient k and initial condition f if $u : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\begin{cases} \frac{1}{2}\Delta u = \frac{\partial u}{\partial t} + ku & \text{on } (0, \infty) \times \mathbb{R}^d \\ u(0, x) = f(x) & \text{on } \{0\} \times \mathbb{R}^d \end{cases}$$

We may proceed in a manner related to, but more involved than, our solution of the Dirichlet problem in the previous section. Were we to finish the computation, we would arrive at the famous theorem of Feynman and Kac:

Theorem 116. *Suppose $f : \mathbb{R}^d \to \mathbb{R}$, $k : \mathbb{R}^d \to [0, \infty)$ are continuous and $u : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ exists such that for each $0 < T < \infty$ there exist constants $K > 0$ and $a < \frac{1}{2Td}$ such that for all $x \in \mathbb{R}^d$, we have*

$$\max_{[0,T]} |u(t, x)| \leq K e^{a|x|^2}$$

Then u satisfies

$$u(t, x) = \mathbb{E}_x \left[f(B_t) \exp \left(- \int_0^t k(B_s) ds \right) \right]$$

We will not prove this theorem here, but [Kar91] has a wonderful section about this topic on which this talk is based.

In order to make our analysis more tenable, we introduce the Laplace transform.

Definition 117. Suppose f is an integrable function on the positive real line. We define the Laplace transform as a function on the positive real line

$$\mathcal{L}f(\alpha) = \int_0^\infty e^{-\alpha t} f(t) dt$$

whenever this is finite for all $\alpha > 0$.

Remark 118. The Laplace transform is an important tool in the study of ordinary differential equations and can be found in any book on the subject. The special fact that we will use is that they are unique; this holds because of the existence of an inverse Laplace transform. To clarify, if $\mathcal{L}f = \mathcal{L}g$ then f and g agree.

We now assume the growth condition

$$\lim_{t \to \infty} e^{-\alpha t} u(t, x) = 0$$

on any solution to the parabolic PDE above. Then let $z_\alpha(x)$ be the Laplace transform of $u(t, x)$ and we compute

$$\begin{aligned} \frac{1}{2} \Delta z_\alpha &= \frac{1}{2} \int_0^\infty e^{-\alpha t} \Delta u dt = \int_0^\infty e^{-\alpha t} \left(ku + \frac{\partial u}{\partial t} \right) dt \\ &= k \int_0^\infty e^{-\alpha t} u dt + \int_0^\infty d \left(e^{-\alpha t} u \right) + \int_0^\infty \alpha e^{-\alpha t} u dt = (\alpha + k) z_\alpha - f \end{aligned}$$

Thus by the uniqueness of the Laplace transform, solving the first equation is equivalent to solving

$$\frac{1}{2} \Delta z_\alpha = (\alpha + k) z_\alpha - f$$

With Theorem 116, we know that we must have a similar result for the Laplace transformed expression of the problem. Indeed, in one dimension, we have

Theorem 119. *Let $f : \mathbb{R} \to \mathbb{R}$, $k : \mathbb{R} \to [0, \infty)$ be piecewise continuous and suppose that for all x,*

$$\mathbb{E}_x \left[\int_0^\infty e^{-\alpha t} |f(B_t)| dt \right] < \infty$$

Then if we let

$$z(x) = \mathbb{E}_x \left[\int_0^\infty f(B_t) \exp \left(-\alpha t - \int_0^t k(B_s) ds \right) \right]$$

we have that z is piecewise C^2 and satisfies

$$\frac{1}{2} \frac{d^2 z}{dx^2} + f = (\alpha + k) z$$

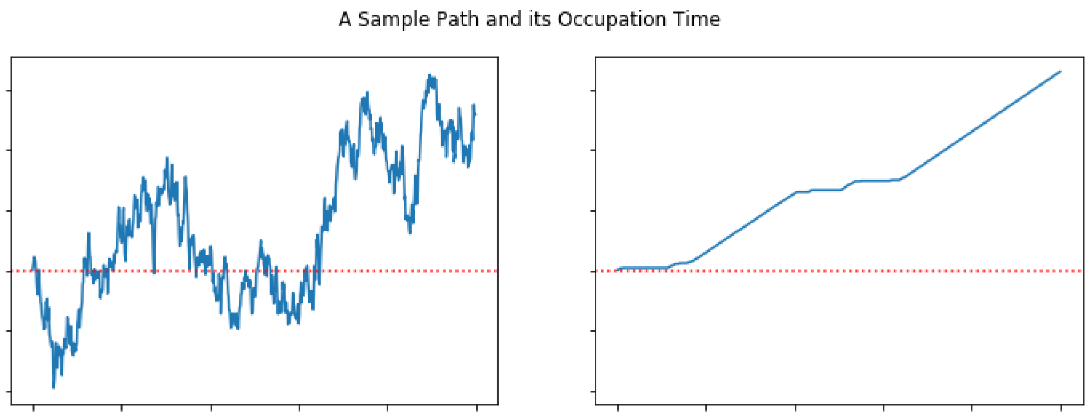

Figure 6: A sample path of linear Brownian motion and its corresponding occupation time of the positive half line.

Once again, a proof of this theorem (which follows mostly by computation) can be found in [Kar91].

The above seems like it has, once again, been applying the theory of Brownian motion to solve problems in PDEs. We can, however, apply the theory in the opposite direction as well. The following application due to Lèvy is a good example of this powerful technique, using Theorem 116 to find the distribution for the occupation time.

Theorem 120 (Lèvy's Arcsine Law)**.** *Let B_t be a standard linear Brownian motion. We define the occupation time of the positive halfline as the random function*

$$\Gamma_+(t) = \int_0^t \mathbb{1}_{(0,\infty)}(B_s)ds$$

the amount of time such that $B_t > 0$. Then for all fixed $t > 0$ and for all $0 \le a \le t$, we have

$$\mathbb{P}_0\left(\Gamma_+(t) \le a\right) = \frac{2}{\pi}\arcsin\left(\sqrt{\frac{a}{t}}\right)$$

the arcsine distribution with density given by

$$\frac{ds}{\pi\sqrt{s(t-s)}}\mathbb{1}_{[0,t]}$$

Proof. The method of the proof is as follows. We use Theorem 119 to find the (Laplace transform of) the moment generating function of $\Gamma_+(t)$. We then calculate and show that the Laplace transform of the moment generating function of the arcsine law is the same. By the uniqueness of the Laplace transform, then, we are done by Proposition 23

Fix $\alpha, \lambda > 0$ and let

$$z(x) = \mathbb{E}_x\left[\int_0^\infty \exp\left(-\alpha t - \lambda\Gamma_+(t)\right)\right]$$

Letting $f(x) = 1$ and $k(x) = \lambda\mathbb{1}_{(0,\infty)}(x)$ we see by Theorem 119 that z satisfies the differential equation

$$\begin{cases} \alpha z = \frac{1}{2}\frac{d^2z}{dx^2} - \lambda z + 1 & x > 0 \\ \\ \alpha z = \frac{1}{2}\frac{d^2z}{dx^2} + 1 & x < 0 \end{cases}$$

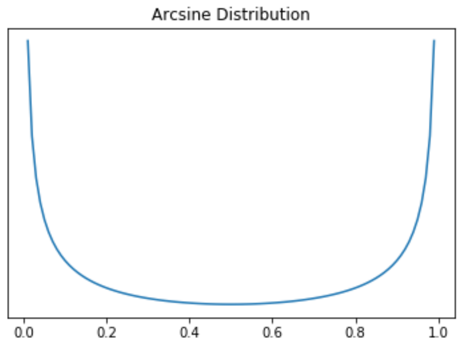

Figure 7: The density of the Arcsine distribution.

and z and z' are continuous. From the theory of ordinary differential equations, we can solve the above problem and note that there is a unique (bounded) solution given by

$$z(x) = \begin{cases} Ae^{-x\sqrt{2(\alpha+\lambda)}} + \frac{1}{\alpha+\lambda} & x > 0 \\ \\ Be^{x\sqrt{2\alpha}} + \frac{1}{\alpha} & x < 0 \end{cases}$$

where A, B are constants determined to make z and z' continuous at 0. But then we may solve to get

$$A = \frac{\sqrt{\alpha+\lambda} - \sqrt{\alpha}}{(\alpha+\lambda)\sqrt{\alpha}} \qquad\qquad B = \frac{\sqrt{\alpha} - \sqrt{\alpha+\lambda}}{\alpha\sqrt{\alpha+\lambda}}$$

and so we have

$$z(0) = \frac{1}{\sqrt{\alpha(\alpha+\lambda)}} = \int_0^\infty e^{-\alpha t}\mathbb{E}_0\left[e^{-\lambda\Gamma_+(t)}\right] dt$$

This is the Laplace transform of the moment generating function $\varphi(\lambda)$ of $\Gamma_+(t)$.

Now we note that the moment generating function of the arcsine distribution is given by

$$\varphi(\lambda) = \mathbb{E}\left[e^{-\lambda X}\right] = \int_0^t \frac{e^{-\lambda s}}{\pi\sqrt{s(t-s)}} ds$$

Thus we have that the Laplace transform of the moment generating function is given by

$$\begin{aligned}\mathcal{L}\varphi(\alpha) &= \int_0^\infty e^{-\alpha t} \int_0^t \frac{e^{-\lambda t}}{\pi\sqrt{s(t-s)}} dsdt = \int_0^\infty \frac{e^{-\lambda s}}{\pi\sqrt{s}} \int_s^\infty \frac{e^{-\alpha t}}{\sqrt{t-s}} \\ &= \frac{1}{\pi}\int_0^\infty \frac{e^{-(\alpha+\lambda)s}}{\sqrt{s}} \int_0^\infty \frac{e^{-\alpha t}}{\sqrt{t}} dtds = \frac{1}{\sqrt{\alpha(\alpha+\lambda)}}\end{aligned}$$

by the identity

$$\int_0^\infty \frac{e^{-at}}{\sqrt{t}} dt = \sqrt{\frac{\pi}{a}}$$

Thus the Laplace transforms are the same and so the moment generating functions are the same and so the distributions are the same by Proposition 23 and we are done. ■

We have illustrated in this section a powerful technique of relating PDEs to Brownian motion. Many more applications can be found in the references below.

6 References

[Bil61] Patrick Billingsley. The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12(5):788, October 1961.

[Bro71] B. M. Brown. Martingale central limit theorems. *The Annals of Mathematical Statistics*, 42(1):59–66, February 1971.

[Chu00] Kai Lai Chung. *A Course in Probability Theory, Third Edition*. Academic Press, oct 2000.

[CW76] Rafael V. Chacon and John B. Walsh. One-dimensional potential embedding. *Séminaire de probabilités de Strasbourg*, 10:19–23, 1976.

[Dub68] Lester E. Dubins. On a theorem of skorohod. *The Annals of Mathematical Statistics*, 39(6):2094–2097, December 1968.

[Dur10] Rick Durrett. *Probability: Theory and Examples (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, aug 2010.

[Gil01] David Gilbarg. *Elliptic partial differential equations of second order*. Springer, Berlin New York, 2001.

[Itô44] Kiyosi Itô. Stochastic integral. *Proceedings of the Imperial Academy*, 20(8):519–524, 1944.

[Kar91] Ioannis Karatzas. *Brownian Motion and Stochastic Calculus (Graduate Texts in Mathematics) (Volume 113)*. Springer, aug 1991.

[Law06] Gregory F. Lawler. *Introduction to Stochastic Processes*. Chapman and Hall/CRC, may 2006.

[Mö10] Peter Mörters. *Brownian Motion (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, mar 2010.

[Obł04] Jan Obłój. The skorokhod embedding problem and its offspring. *Probability Surveys*, 1(0):321–392, 2004.

[Oks19] Bernt Oksendal. *Stochastic Differential Equations*. Springer Verlag, jul 2019.

[Pro04] Philip Protter. *Stochastic integration and differential equations*. Springer, Berlin New York, 2004.

[PWZ33] R. E. A. C. Paley, N. Wiener, and A. Zygmund. Notes on random functions. *Mathematische Zeitschrift*, 37(1):647–668, December 1933.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Springer, Berlin New York, 1999.

[vdV12] A. W. van der Vaart. *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, jun 2012.

7 Problem Sets

Some problems are adapted from [Dur10, Kar91, Mö10, Oks19]. Some are classic problems in the field. Some are either my own creation or remembered from too long ago for me to remember an attribution.

7.1 Problem Set 1

Problem 1. *Give an example of*

a) *a sequence of random variables X_n such that $X_n \xrightarrow{d} X$ but it does* NOT *hold that $X_n \xrightarrow{p} X$*

b) *a sequence of random variables X_n such that $X_n \xrightarrow{p} X$ but we do* NOT *have $X_n \to X$ almost surely.*

Problem 2. *Let $X_1, X_2, \dots$ be iid (independent identically distributed) random variables taken from the uniform distribution on the unit interval. Consider the running maximum*

$$M_n := \max_{i \le n} X_i$$

Show that there is a random variable X such that $n\,(1 - M_n) \xrightarrow{d} X$ and find the distribution of X.

Problem 3. *In this problem we introduce the Gaussian distribution. We say that a random variable has a Gaussian distribution with mean μ and variance σ^2 if its law is given by*

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

A standard Gaussian has mean 0 and variance 1.

a) *Show the Gaussian tail-bounds, i.e., if X is a standard gaussian, then for any $a > 0$*

$$\frac{e^{-\frac{a^2}{2}}}{\sqrt{2\pi}} \left(\frac{1}{a} - \frac{1}{a^3}\right) \le \mathbb{P}(X > a) \le \frac{1}{a} \frac{e^{-\frac{a^2}{2}}}{\sqrt{2\pi}}$$

b) *Let $X_1, X_2, \dots$ be standard Gaussian variables (not necessarily independent). Let M_n be the running maximum as in the last problem. Show that, almost surely,*

$$\limsup \frac{M_n}{\sqrt{2 \log n}} \le 1$$

c) *Now assume that the X_i in the previous part are independent and show that the limit supremum is in fact equal to 1 in this case.*

Problem 4. *This problem gives a probabilistic proof of the well known fact that continuous functions on compact intervals can be uniformly approximated by polynomials. Thus, let f be a continous function on $[0,1]$ and define*

$$f_n(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} f\left(\frac{i}{n}\right)$$

a) *Let $X_1, X_2, \dots$ be iid Bernoulli(p) r.v.s (i.e., $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1 - p$). Let $S_n = X_1 + \dots + X_n$. Show*

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \le \frac{1}{4n\varepsilon^2}$$

b) Because continuous functions are uniformly continuous on compact sets, for all $\varepsilon > 0$, there is some $\varepsilon' > 0$ such that if $|x - y| < \varepsilon'$ we have $|f(x) - f(y)| < \varepsilon$ for all $x, y \in [0,1]$. Using this and the above part, show that for all $\varepsilon > 0$,

$$\left| \mathbb{E} f\left(\frac{S_n}{n}\right) - f(p) \right| \leq \varepsilon + 2\left(\max_{[0,1]} f\right) \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon'\right)$$

c) Show that $\mathbb{E} f\left(\frac{S_n}{n}\right) = f_n(p)$ and conclude that $\sup_{[0,1]} |f_n(x) - f(x)| \to 0$ as $n \to \infty$.

Problem 5. *Let $X_1, X_2, \ldots$ be iid such that $0 < X_i < \infty$. Let $S_n = X_1 + \cdots + X_n$ and let $N_s = \sup_{S_n < s} n$. Show that, almost surely,*

$$\frac{N_s}{s} \to \frac{1}{\mathbb{E}X_1}$$

(One interpretation for this is to view the X_i as random lifespans of a battery with the battery being instantaneously replaced by the next everytime the last one dies. Then S_n is the amount of time that the first n batteries last in total and N_s is the total number of batteries used up until time s.)

7.2 Problem Set 2

Problem 6. *Suppose you are playing a game as follows. There are $X_1, X_2, \ldots$ iid random variables such that $X_1 \geq 0$ and $\mathbb{E}[X_1] < \infty$. At each step you have the choice to continue or to stop. If you continue, you pay c dollars and receive another X_i. If you leave, you take the M_n dollars, where $M_n = \max_{m \leq n} X_m$. A "strategy" is any stopping time τ (adapted to $\sigma(X_1, \ldots, X_n)$ of course) with finite expectation. Your goal is to choose a strategy to maximize your winnings.*

a) One strategy is to pick some $a > 0$ and let $\tau = \inf\{n | X_n > a\}$. Let W_n be your winnings at time n were you to leave, i.e., $W_n = M_n - cn$. Compute $\mathbb{E}[W_\tau]$ in terms of a, $\mathbb{E}[(X_1 - a)^+]$, and $p = \mathbb{P}(X_1 > a)$. (Hint: Use the tail sum formula which says for $X \geq 0$, $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)dt$).

b) Show that there exists a unique solution $a = \alpha$ to the equation $\mathbb{E}[(X_1 - a)^+] = c$. Show that

$$W_n \leq \alpha + \sum_{k=1}^{n} \left((X_k - \alpha)^+ - c\right)$$

c) Show that if τ is any stopping time such that $\tau < \infty$ almost surely then $\mathbb{E}[W_\tau] \leq \alpha$. As a bonus, one might consider the intuition behind restricting focus to stopping times with finite expectation.

Problem 7. *Let $0 < p < 1$, $q = 1 - p$ and suppose that $p \neq \frac{1}{2}$. Let $X_1, X_2, \ldots$ be iid random variables such that $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = -1) = q$. Let $S_n = X_1 + \cdots + X_n$ be a random walk and let $-a < 0 < b$ and $\tau = \inf\{n | S_n \notin (-a, b)\}$ be the stopping time corresponding to leaving the intervale (a, b).*

a) What is $\mathbb{P}(S_\tau = -a)$? What is $\mathbb{P}(S_\tau = b)$?

b) What is $\mathbb{E}[\tau]$?

c) A random walk is recurrent if there exists a number a such that $\mathbb{P}(S_n = a \text{ i.o.}) = 1$. Is the random walk described above recurrent?

Problem 8. *A function $f : \mathbb{R}^d \to \mathbb{R}$ is called superharmonic if for all x and all small $r > 0$, it satisfies the following inequality:*

$$f(x) \geq \frac{1}{|B_r(x)|} \int_{B_r(x)} f(y)dy$$

(If we have the reverse inequality such a function is called subharmonic and if we have equality, the function is called harmonic. These functions are very important in the study of elliptic PDEs.)

a) *Let $X_1, X_2, \dots$ be iid random variables taken uniformly from the unit ball $B_1(0)$ and let $S_n = X_1 + \cdots + X_n$ be the random walk generated. If f is a superharmonic function, show that $f(S_n)$ is a supermartingale.*

b) *Using a result from lecture, note that S_n is recurrent in dimensions 1 and 2. Using this fact, show that if $f : \mathbb{R}^d \to \mathbb{R}$ is a nonnegative superharmonic function where $d \in \{1, 2\}$, then f is constant.*

c) *Find a superharmonic function $f : \mathbb{R}^d \to \mathbb{R}$ that is nonnegataive where $d \geq 3$. Show that the random walk S_n above is transient (not recurrent) in dimensions $d \geq 3$.*

Problem 9. *For a martingale $(M_n, \mathcal{F}_n)$ such that $\mathbb{E}\left[M_n^2\right] < \infty$ for all n we define*

$$\langle M \rangle_n = \sum_{k=1}^{n} \mathbb{E}\left[(M_k - M_{k-1})^2 | \mathcal{F}_{k-1}\right]$$

the quadratic variation of M.

a) *Show that X_n is a martingale such that $\mathbb{E}[X_n^2] < \infty$ where*

$$X_n = \sum_{k=1}^{n} \frac{M_k - M_{k-1}}{1 + \langle M \rangle_k}$$

b) *Show that $\langle X \rangle_n$ is increasing and that*

$$\langle X \rangle_n \leq \frac{1}{1 + \langle M \rangle_0} - \frac{1}{1 + \langle M \rangle_n} \leq 1$$

and deduce that $\langle X \rangle_n \to \langle X \rangle_\infty$ almost surely, where $\langle X \rangle_\infty \leq 1$.

c) *Using the conclusion of the part above and the fact that if M_n is a martingale such that we have $\mathbb{E}[M_n^2] < \infty$, then M_n is bounded if and only if $\mathbb{E}[\langle M \rangle_\infty] < \infty$, show that $X_n \to X$ almost surely for some $X < \infty$.*

d) *Using the previous part, show that*

$$\lim_{n \to \infty} \frac{M_n}{\langle M \rangle_n} = 0$$

on the event $\{\langle M \rangle_\infty = \infty\}$. Hint: the Kronecker lemma might be useful, where the Kronecker lemma states that if b_n is a nonnegative increasing sequence such that $b_n \uparrow \infty$, and x_n is a sequence such that $\sum \frac{x_n}{b_n} < \infty$ then

$$\frac{1}{b_n} \sum_{j=1}^{n} x_j \to 0$$

e) *Put the above to use to prove the "weak" strong law of large numbers: If $X_1, X_2, \dots$ are independent such that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] \leq K < \infty$, then show that $n^{-1} S_n \to 0$ almost surely, where $S_n = X_1 + \cdots + X_n$.*

7.3 Problem Set 3

Problem 10. *Let S_n be a random walk whose steps are all ± 1.*

a) *Show the reflection principle in discrete time. More precisely show that the number of paths from $(0, x)$ to (n, y) that touch 0 is the same as the number of paths from $(0, -x)$ to (n, y).*

b) *Suppose people are voting on which they like better, PROMYS or Ross, and there are p votes for PROMYS and r votes for Ross, with $p > r$ (PROMYS obviously wins). If we are counting each vote one at a time and we pick each vote uniformly from the remainder, then what is the probability that PROMYS is always winning?*

Problem 11. *Let B_t be a standard Brownian motion.*

a) Show that for all $t > 0$, we have

$$\lim_{n\to\infty}\sum_{k=1}^{n}\left(B_{\frac{kt}{n}} - B_{\frac{(k-1)t}{n}}\right)^2 = t$$

b) Define the p-variation of a function $f : \mathbb{R} \to \mathbb{R}$ as

$$V_t^p(f) = \lim_{n\to\infty}\sum_{k=1}^{n}\left|f\left(\frac{kt}{n}\right) - f\left(\frac{(k-1)t}{n}\right)\right|^p$$

If we denote by B the (random) function that takes $t \mapsto B_t$, show that almost surely

$$V_t^p(B) = \begin{cases}\infty & p < 2\\ 0 & p > 2\end{cases}$$

c) Is there a reasonable class $\mathcal{C}$ of functions on which we could almost surely make sense of

$$\int_0^t f(s)dB_s(\omega)$$

for all $f \in \mathcal{C}$?

d) Can you suggest a reasonable way to define

$$\int_0^t f(s)(dB_s(\omega))^2$$

for f in some class of functions $\mathcal{C}$?

Problem 12. *We call $B_t = \left(B_t^1, B_t^2, \ldots, B_t^d\right)$ standard d-dimensional Brownian motion if the coordinates are independent standard (linear) Brownian motions. Recall that if $f \in C_0^2$ is a twice differentiable function with compact support then we define $\Delta f = \sum \frac{\partial^2 f}{\partial x_i^2} = \operatorname{div}\nabla f$. Recall also the notation of $\mathbb{E}_x[\cdot]$ meaning expectation with respect to Brownian motion started at $x \in \mathbb{R}^d$.*

a) Let B_t be a standard Brownian motion in d dimensions. Show that if $\mathbb{P}_x$ is the probability measure with respect to B_t started at x, i.e., $B_0 = x$, then we have

$$\mathbb{P}_x(B_t \in A) = \int_A p_t(x, y)dy$$

where

$$p_t(x, y) = (2\pi t)^{-\frac{d}{2}} e^{-\frac{|x-y|^2}{2t}}$$

b) Show that $p_t(x, y)$ satisfies the heat equation, i.e.,

$$\frac{\partial p_t}{\partial t} - \frac{1}{2}\Delta_y p_t = 0$$

where Δ_y is the Laplacian in the y-coordinates:

$$\Delta_y p_t = \sum \frac{\partial^2 p_t}{\partial y_j^2}$$

c) Let $f \in C_0^2$. Show that

$$f(B_t) - \int_0^t \frac{1}{2}\Delta f(B_s)ds$$

is a martingale (with respect to its natural filtration). In particular, if f is harmonic then $f(B_t)$ is a martingale.

d) Show that $|B_t|^2 - dt$ *is a martingale.*

e) Let $\tau_a = \inf\{t||B_t| = a\}$*. Show that if* $|x| < R$ *then* $\mathbb{E}_x[\tau_R] = \frac{R^2-|x|^2}{d}$

Problem 13. *This is a continuation of Problem 12. Let* $0 < r < R$ *and let* $T_{rR} = \tau_r \wedge \tau_R$*. We now restrict to the case of* $d = 2$*.*

a) Show that $\mathbb{E}_x[\log|B_{T_{rR}}|] = \log|x|$*.*

b) Show that

$$\mathbb{P}_x(\tau_r < \tau_R) = \frac{\log R - \log|x|}{\log R - \log r}$$

where $\mathbb{P}_x$ *is the probability measure associated to Brownian motion starting at* x*.*

c) Show that if $x \neq 0$ *then* $\mathbb{P}_x(B_t = 0 \text{ for some } t < \infty) = 0$*. Conclude that Brownian motion in the plane is not recurrent.*

d) Show that B_t *is* neighborhood *recurrent in that if* $U \subset \mathbb{R}^2$ *is any nonempty open set then we have* $\mathbb{P}_x(B_t \in U \text{ i.o.}) = 1$*.*

Problem 14. *We now do a similar analysis as Problem 13 but in higher dimensions. Fix* $d > 2$, $0 < r < R$*, and* B_t *a* d*-dimensional Brownian motion.*

a) Show that $\mathbb{E}_x\left[|B_{T_{rR}}|^{2-d}\right] = |x|^{2-d}$

b) Show that

$$\mathbb{P}(\tau_r < \tau_R) = \frac{R^{2-d} - |x|^{2-d}}{R^{2-d} - r^{2-d}}$$

c) Is B_t *recurrent? Is* B_t *neighborhood recurrent?*

7.4 Problem Set 4

Problem 15. *Let* $a_k(t) = \mathbb{E}[B_t^{2k}]$ *and let* $b_k(t) = \mathbb{E}[B_t^{2k+1}]$*.*

i) Find recurrence relations for $a_k(t), b_k(t)$*.*

ii) Let $Z \sim N(0,1)$ *and let* $m_n = \mathbb{E}[Z^n]$*. Find* m_n *for all* n*.*

iii) Show that $\mathbb{E}\left[e^{tZ}\right] = e^{\frac{t^2}{2}}$*.*

Problem 16. *Let* B_t *be standard Brownian motion in* d *dimensions with filtration* $\mathcal{F}_t$*, let* $a \geq \frac{1}{2}$*, and let*

$$dX_t = \frac{a}{X_t}dt + dB_t$$

if such a process exists. Let $T = \inf\{t|X_t = 0\}$ *and consider* X_t *only on the time interval* $[0,T)$*. Fix* $0 < r < R$ *and let* $\tau = \inf\{t|X_t \in \{r,R\}\}$*. For* $r \leq x \leq R$ *let* $\varphi_a(x) = \mathbb{P}_x(X_\tau = R)$ *where* $\mathbb{P}_x$ *is the probability measure associated to* X_t *where* $X_0 = x$*.*

i) Let $Y_t = \mathbb{E}\left[\mathbb{1}_{\{X_\tau = R\}}|\mathcal{F}_t\right]$*. Show that* Y_t *is a martingale and show that* $Y_t = \varphi_a(X_{t\wedge\tau})$*.*

ii) Find a stochastic differential equation for $\varphi(X_t)$*, i.e., find functions* u_1, u_2 *such that*

$$d(\varphi(X_t)) = u_1 dt + u_2 dB_t$$

iii) Show that $u_1 = 0$ *and use this fact to find an ordinary differential equation for* φ_a*. Show that* $\varphi_a(r) = 0$ *and* $\varphi_a(R) = 1$ *and use this to find the function* $\varphi_a(x)$*.*

iv) Let $M_t = |B_t| = \sqrt{(B_t^{(1)})^2 + \cdots + (B_t^{(d)})^2}$. *Show that if if* $a = \frac{d-1}{2}$ *then* $M_t = X_t$. *Provide another proof that Brownian motion is neighborhood recurrent in dimension 2 and transient in dimensions* $d \geq 3$.

Problem 17. *In this problem we explore the so-called "Brownian Bridge"*

i) Fix x *and show that*

$$X_t = xt + (1-t)\int_0^t \frac{dB_s}{1-s}$$

satisfies

$$dX_t = \frac{x - X_t}{1-t}dt + dB_t$$

on $[0,1)$.

ii) Show that $X_t \to x$ *almost surely as* $t \to 1$.

iii) One interpretation of the above process is that X_t *is Brownian motion "conditioned" on the event that* $B_1 = x$. *Explain why this makes sense intuitively. Explain why this does not make sense in a rigorous manner.*

Problem 18. *Let* B_t *be standard linear Brownian motion and consider*

$$dX_t = -X_t dt + dB_t$$

with $X_0 = x_0$.

i) Explicitly find a process X_t *that satisfies the above differential equation.*

ii) Now let $x_0 = 0$ *and find* $\alpha(t)$ *such that* $B_{\alpha(t)}$ *is the same process as* X_t *above. Find* $\langle X \rangle_t$.

iii) Show that no matter what x_0 *is, that if* $Z \sim N(0,1)$ *then* $X_t \xrightarrow{d} Z$.