Analysis III
Taught by Ewain Gwynne

May 14, 2024

Faye Jackson

Contents

I. Probability 2
- I.1. Motivation and Measure-Theoretic Framework 2
- I.2. Fourier Analysis for Probability 13

II. Conditional Expectation 17

III. Brownian Motion 21
- III.1. Motivation and The Big Idea 21
- III.2. Definition and Properties 22
- III.3. Brownian Motion in $\mathbb{R}^d$ 28
- III.4. Brownian Motion and Harmonic Functions 29
- III.5. Construction of Brownian Motion 35

References 38

I. Probability

I.1. Motivation and Measure-Theoretic Framework

Why should you care about probability? Two good reasons

- Random objects can be easier to analyze than deterministic objects. E.g. three random points on the disk are not collinear.
- Quantum Mechanics informs us that the world is likely to be "inherently probabilistic."
- Probabilistic method in combinatorics. The idea here is to define a random object of the desired type, and show it has the desired property with positive probability, hence an object with this property must exist.
- Heuristically, prime numbers "act like" a random set $X \subseteq \mathbb{N}$ such that

$$\mathbb{P}[n \in X] = \frac{1}{\log n},$$

  and $\{n \in X\}$ are independent.
- Sometimes possible to describe solutions to PDEs in terms of Brownian Motion (B.M.)
- Brownian Motion in complex analysis, which we'll see a lot of in this course.

Definition I.1.1

A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω is a set of "possible states of the world."
- $\mathcal{F}$ is a σ-algebra of subsets of Ω, i.e. it is closed under complements, countable unions, and contains $\emptyset, \Omega$. We think of $\mathcal{F}$ as the collection of "events" (aka things that could happen).
- $\mathbb{P}$ is a measure on the space $(\Omega, \mathcal{F})$ such that $\mathbb{P}(\Omega) = 1$. We think of this as the "probability." Namely $\mathbb{P}(E) \in [0, 1]$ for $E \in \mathcal{F}$ is the "probability that the event E occurs"

Notationally, we say an event $E \in \mathcal{F}$ occurs almost surely (a.s.) if $\mathbb{P}(E) = 1$.

Whenever we talk about probability, we're going to be fixing a set $(\Omega, \mathcal{F}, \mathbb{P})$, so from now on this will be our running notation and always assumed.

Definition I.1.2

Let $\mathcal{X}$ be a topological space (with the Borel σ-algebra). A random variable (R.V.) taking values in $\mathcal{X}$ is a measurable function

$$X : \Omega \to \mathcal{X}.$$

We identify X and $\widetilde{X}$ if $\mathbb{P}[X = \widetilde{X}] = 1$. To specify this exactly, this is

$$\mathbb{P}(\{\omega \in \Omega \mid X(\omega) = \widetilde{X}(\omega)\}) = 1.$$

Similar to the notation $\{X = \widetilde{X}\}$ as above, we'll take the following notation in general for $A \subseteq \mathcal{X}$ a Borel subset of $\mathcal{X}$,

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

Definition I.1.3

Let X be a R.V. taking values in $\mathcal{X}$. The law or distribution of X is the probability measure μ_X on $\mathcal{X}$ defined by

$$\mu_X(A) = \mathbb{P}[X \in A]$$

for all $A \subseteq \mathcal{X}$ Borel. In other words, this is the pushforward measure. Given two R.V.s X, Y, we say they agree in law or agree in distribution provided that $\mu_X = \mu_Y$. We'll abbreviate this as $X \overset{d}{=} Y$.

Given X, Y R.Vs taking values in $\mathcal{X}, \mathcal{Y}$, we say that the joint law of X, Y

The fundamental thing in probability theory is really the laws or distributions of random variables / their joint laws. The underlying probability space Ω is not quite as important. It just has to be large enough to support our random variables and their laws. A really good source is Tao's notes on probability [15]. This will provide some good intuition. We also have the following intuition

Example I.1.1

If $\mathcal{B}$ is the Borel σ-algebra for $\mathcal{X}$, then $(\mathcal{X}, \mathcal{B}, \mu_X$ is a probability space. We can then let $\widetilde{X}(x) = x$ for all $x \in \mathcal{X}$, and then $\widetilde{X}$ is a R.V. with law μ_X.

Note however that a Random Variable is not determined by its law

Example I.1.2

Let $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ (thought of as pairs of dice rolls, with the appropriate probabilities) and consider $X : \Omega \to \mathbb{R}$ to be the first dice roll and $Y : \Omega \to \mathbb{R}$ the second dice roll. Then X, Y are not identical random variables, but they clearly have the same law.

Example I.1.3

If $\mathcal{X}$ is countable, then μ_X is determined by $\mathbb{P}[X = y]$ for all $y \in \mathcal{X}$.

Definition I.1.4

Let $\mathcal{X} = \mathbb{R}$. Then a R.V. X taking values in $\mathbb{R}$ has a cumulative distribution function (cdf)

$$F_X(y) = \mathbb{P}[X \leq y] = \mu_X((-\infty, y]).$$

Note that μ_X is determined by F_X.

For F a cdf, we have the following elementary properties. F is

- Non-decreasing,
- $F : \mathbb{R} \to [0, 1]$.
- Continuous from the right,

  $$\mu_X((-\infty, y]) = \lim_{z \to y^+} \mu_X((-\infty, z])$$

  by the downward continuity of measures.
- $F(y) - \lim_{z \to y^-} F(z) = \mathbb{P}[X = y]$.

Conversely, if $F : \mathbb{R} \to [0,1]$ is non-decreasing, continuous from the right, and $\lim_{y\to\infty} F(y) = 1, \lim_{y\to-\infty} F(y) = 0$, we can define a probability measure on $\mathbb{R}$ by

$$\mu((a,b]) = F(b) - F(a).$$

This is sometimes called a Lebesgue–Stieltjes measure.

Definition I.1.5

Let X be a R.V. taking values in $\mathbb{R}^d$ for $d \geq 1$. If the law μ_X is absolutely continuous with respect to the Lebesgue measureon $\mathbb{R}^d$, then the Radon-Nikodym derivative $f = \frac{\mathrm{d}\mu}{\mathrm{d}\,\mathrm{Leb}}$ is called the probability density function (pdf) or density of X.

If $d = 1$ then $F' = f$ almost everywhere in this case.

Definition I.1.6

Let X be a random variable in $\mathbb{R}$. The expectation (or mean) of X is

$$\mathbb{E}[X] := \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} y\,\mathrm{d}\mu_X(y)$$

You can do a similar definition when X takes values in a Banach space.

Definition I.1.7

If $\mathbb{E}[X]$ exists, then the variance is

$$\operatorname{var} X := \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_\mathbb{R} (y - \mathbb{E}[X])^2\,\mathrm{d}\mu_X(y).$$

Note: even if the expectation exists, the variance may not, for example if $f \in L^1(\mathbb{R})$ is your density but $f \notin L^2(\mathbb{R})$.

The standard deviation is $\sigma(X) = \sqrt{\operatorname{var} X}$.

Example I.1.4

The Gaussian or normal distribution with mean $a \in \mathbb{R}$ and variance $\sigma^2 > 0$ is the probability measure on $\mathbb{R}$ with density

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-a)^2}{2\sigma^2}\right)$$

We'll abbreviate $X \sim N(a, \sigma^2)$ to mean X has Gaussian distribution with mean a and variance σ^2. The standard Gaussian is $a = 0, \sigma^2 = 1$. If $X \sim N(0,1)$, then $\sigma X + a \sim N(a, \sigma^2)$.

This formula looks sort of arbitrary, but later as we'll see with the central limit theorem (see: Theorem I.2.7) the Gaussian is a sort of universal distribution. This makes it a central object in probability theory.

Exercise I.1.5

Show that the density function $f(y)$ is actually a density. In other words $\int_{-\infty}^{\infty} f(y)\,\mathrm{d}y = 1$.

Lemma I.1.1 (The Law of The Unconcious Statistician)

Let X be a R.V. taking values in $\mathcal{X}$ and let $g : \mathcal{X} \to \mathbb{R}$ be a measurable function. Then $g(X)$ is a random variable, and

$$\mathbb{E}[g(X)] = \int_{\mathcal{X}} g(y)\,\mathrm{d}\mu_X(y).$$

Proof. If $g = \sum_{j=1}^N a_j \mathbb{1}_{A_j}$ (the sum of indicators), for $A_j \subseteq X$ Borel and disjoint, then

$$\mathbb{E}[g(X)] = \sum_{j=1}^N a_j \mathbb{P}[X \in A_j] = \int_{\mathcal{X}} g(y)\,\mathrm{d}\mu_X(y).$$

In general, approximate g by linear combinations of indicators. More explicitly

$$\begin{aligned}\int_{\mathcal{X}} g(y)\,\mathrm{d}\mu_X(y) &= \sum_{j=1}^N a_j \int_{\mathcal{X}} \mathbb{1}_{A_j}(y)\,\mathrm{d}\mu_X(y) \\ &= \sum_{j=1}^N a_j \mu_X(A_j) = \sum_{j=1}^N a_j \mathbb{P}[X \in A_j].\end{aligned}$$

We also have

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) = \sum_{j=1}^N a_j \int_{A_j} X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \sum_{j=1}^N a_j \mathbb{P}[X \in A_j].$$

Lets dispel a common mistake that people make when learning probability with an extremely concrete example.

Example I.1.6

Pairwise independence does not imply independence of the whole collection. To explain this, let X_1, X_2 be independent random variables in $\mathbb{Z}/2\mathbb{Z}$ with distribution $\mathbb{P}[X_j = 0] = \mathbb{P}[X_j = 1] = \frac{1}{2}$ (aka coin flips!)

Let $X_3 = X_1 + X_2$ modulo 2. It is not difficult to check that

$$(X_1, X_2) \overset{d}{=} (X_1, X_3) \overset{d}{=} (X_2, X_3)$$

all in distribution. Thus any two of them are independent. But the triple $\overline{X} = (X_1, X_2, X_3)$ are not independent. For example, the probability of the event $\overline{X} = (0, 0, 1)$ is 0, while the pairwise product would be $\frac{1}{8}$.

Example I.1.7

Let X, Y be random variables in $\mathbb{R}$. Assume that (X, Y) has a density with respect to Lebesgue measure on $\mathbb{R}^2$. Call this density $f : \mathbb{R}^2 \to [0, \infty)$.

Then (X, Y) are independent if and only if there exist functions $f_1, f_2 : \mathbb{R} \to [0, \infty)$ such that $f(x, y) = f_1(x) f_2(y)$ Lebesgue-almost everywhere. This is equivalent to the distribution (X, Y) factors as the product measure.

Lemma I.1.2

Let X, Y be independent random variables in $\mathbb{R}$ with $\mathbb{E}[|X|] < \infty, \mathbb{E}[|Y|] < \infty$. Then we have that

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Proof. We see that, by the law of the unconscious statistician

$$\mathbb{E}[XY] = \int_{\mathbb{R}^2} uv \,\mathrm{d}\mu_{(X,Y)}(u,v) = \int_{\mathbb{R}^2} uv \,\mathrm{d}\mu_X(u)\,\mathrm{d}\mu_Y(v) = \int_{\mathbb{R}} u \,\mathrm{d}\mu_X(u) \int_{\mathbb{R}} v \,\mathrm{d}\mu_Y(v) = \mathbb{E}[X]\mathbb{E}[Y],$$

by applying the Fubini theorem.

Again lets emphasize a common mistake. It is possible to have $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ for X, Y not independent random variables.

Lets recall things from last time. Just to fix some terminology, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{X}$ a topological space with the Borel σ-algebra, and $X : \Omega \to \mathcal{X}$ a random variable (aka a measurable function). Some people use the following terminology.

- Ω is the event space, or sometimes called the state of the world.
- $\mathcal{X}$ is the state space for X.

$\mu_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$ is called the law or distribution.

Definition I.1.8

Let X be a random variable in $\mathcal{X}$. The σ-algebra generated by X is

$$\sigma(X) = \{\{X \in A\} \mid A \subseteq \mathcal{X},\ \text{Borel}\} \subseteq \mathcal{F}.$$

In probability, we often think of $\sigma(X)$ as the "information determined by X." This matches intuition because upon seeing X we can only tell if an event like $\{X \in A\}$ happened.

As always, we should fill our head with examples and remember them dearly

Example I.1.8

$X = \mathbb{1}_E$ which is 1 on E, 0 on E^c. Then

$$\sigma(X) = \{\emptyset, E, E^c, \Omega\}.$$

Example I.1.9

Lets say that $\Omega = \mathbb{R}^2$, and $X(a,b) = a$.

$$\sigma(X) = \{A \times \mathbb{R} \mid A \subseteq \mathbb{R} \text{ Borel}\}.$$

Example I.1.10

If $g : \mathcal{X} \to \mathcal{Y}$ is measurable and X is a random variable in $\mathcal{X}$, then

$$\sigma(g(X)) \subseteq \sigma(X).$$

Namely, if $B \subseteq \mathcal{Y}$ is Borel, then $\{g(X) \in B\} = \{X \in g^{-1}(B)\} \in \sigma(X)$. This again matches intuition, if you have a function of X, it contains at most as much information as the value of X itself.

This statement is really interesting, and really captures the intuition. Lets make it even more precise.

Proposition I.1.3

Conversely to the example above let X be a R.V. in $\mathcal{X}$, Y a R.V. in $\mathbb{R}$ and suppose that Y is $\sigma(X)$ measurable (aka Y is measurable and $\sigma(Y) \subseteq \sigma(X)$). Then in fact there exists $g : \mathcal{X} \to \mathbb{R}$ such that $Y = g(X)$ almost surely.

Proof. First assume $Y = \sum_{j=1}^{N} a_j \mathbb{1}_{A_i}(X)$ for $A_i \subseteq \mathcal{X}$ Borel measurable. Then $g(x) = \sum_{j=1}^{n} a_i \mathbb{1}_{A_j}(x)$. Sine Y is $\sigma(X)$-measurable, there exists Y_n random variables such that each is a linear combination of indicators and $Y_n \to Y$ almost surely.

For each n, there exists $g_n : \mathcal{X} \to \mathbb{R}$ measurable such that $Y_n = g_n(X)$. Let $g(x) = \liminf_{n\to\infty} g_n(x)$. We then have that

$$Y = \lim_{n\to\infty} g_n(X) = g(X).$$

almost surely.

Perfect! This demonstrates that the information-theoretic perspective on $\sigma(X)$ is rigorous.

We'll now switch to the notion of independence.

Definition I.1.9

Events E, F are independent if

$$\mathbb{P}(E \cap F) = \mathbb{P}(E) \cdot \mathbb{P}(F).$$

An indexed family of events $\mathcal{E} \subseteq \mathcal{F}$ is independent provided that

$$\mathbb{P}[E_1 \cap \cdots \cap E_n] = \prod$$

for any $E_1, \ldots, E_n \in \mathcal{E}$ (the fact this is an indexed family instead of a set is important). We also define if $\mathbb{P}[F] \neq 0$, the conditional probability

$$\mathbb{P}[E \mid F] = \frac{\mathbb{P}[E \cap F]}{\mathbb{P}[F]}.$$

We then have that E, F are independent if and only if $\mathbb{P}[E \mid F] = \mathbb{P}[E]$.

It's fruitful to extend this discussion to independence of σ-algebra, and then to random variables.

Definition I.1.10

A collection of σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are independent if for all events $E_j \in \mathcal{F}_j, j = 1, \ldots, n$ we have

$$\mathbb{P}[E_1 \cap \cdots \cap E_n] = \prod_{j=1}^{n} \mathbb{P}(E_j).$$

Likewise, a collection of R.V.s $X_1, \ldots, X_n$ are independent if $\sigma(X_1), \ldots, \sigma(X_n)$ are independent σ-algebras.

The interpretation of this is that having any information about $n-1$ of these random variables doesn't tell you any information about the remaining random variable.

The following lemma provides a directly measure-theoretic equivalent definition to independence.

Lemma I.1.4

Let $X_1, \ldots, X_n$ be random variables with state spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$ respectively. Then the following are equivalent

(i) $X_1, \ldots, X_n$ are independent.

(ii) For all events $A_j \in \mathcal{X}_j$ Borel for $j = 1, \ldots, n$ we have

$$\mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n] = \prod_{j=1}^{n} \mathbb{P}[X_j \in A_j]$$

(iii) Let $\overline{X} = (X_1, \ldots, X_n)$ be a variable in $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$. Then

$$\mu_{\overline{X}} = \mu_{X_1} \times \cdots \times \mu_{X_n}$$

as the product measure.

Proof. Assertion (i) is equivalent to assertion (ii) just by definition of $\sigma(X_j)$ and independence.

Assertion (ii) implies assertion (iii) since a measure on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ is determined by its values on rectangular sets $A_1 \times \cdots \times A_n$.

We'll now show assertion (iii) implies assertion (ii). We find that

$$\begin{aligned}\mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n] = \mathbb{P}[\overline{X} \in A_1 \times \cdots \times A_n] &= \mu_{\overline{X}}(A_1 \times \cdots \times A_n) \\ &= \prod_{j=}^{n} \mu_{X_j}(A_j) = \prod_{j=1}^{n} \mathbb{P}(X_j \in A_j)\end{aligned}$$

Great!

Last time: we introduced independent events and independent random variables.

Lemma I.1.5

Let $X_1, \ldots, X_n$ be pairwise independent random variables in $\mathbb{R}$, and assume $\mathbb{E}[|X_j|] < \infty$. Then

$$\mathrm{var}(X_1 + \cdots + X_n) = \sum_{j=1}^{n} \mathrm{var}(X_j).$$

Proof. Assume without loss of generality that $\mathbb{E}[X_j] = 0$, by subtracting off the means from each X_j. This works because for c constant

$$\mathbb{E}[X - c] = \mathbb{E}[X] - c \qquad\qquad \mathrm{var}(X - c) - \mathrm{var}(X).$$

Thus we can just take $X_j' = X_j - \mathbb{E}[X_j]$. Then

$$\begin{aligned}\mathrm{var}\left(\sum_{j=1}^{n} X_j\right) &= \mathbb{E}\left[\left(\sum_{j=1}^{n} X_j\right)^2\right] \\ &= \sum_{i,j=1}^{n} \mathbb{E}[X_i X_j] \qquad = \sum_{j=1}^{n} \mathbb{E}[X_j^2] + \sum_{i \neq j} \mathbb{E}[X_i]\mathbb{E}[X_j]\end{aligned}$$

$$= \sum_{j=1}^{n} \mathbb{E}[X_j]^2 = \sum_{j=1}^{n} \operatorname{var}(X_j).$$

Perfect!

Example I.1.11

$\operatorname{var}(X + X) = \operatorname{var}(2X) = 4 \operatorname{var}(X)$.

Definition I.1.11

A sequence of random variables $\{X_j\}_{j \geq 1}$ are independent and identically distributed (abbreviated i.i.d.) if for all $n \geq 1$, $X_1, \ldots, X_n$ are independent and $X_j \overset{d}{=} X_i$ for all i, j (they all have the same distribution).

Theorem I.1.6

Let μ be a Borel probability measure on $\mathbb{R}$ and let $(\Omega, \mathcal{F}, \mathbb{P})$ be the unit interval with Lebesgue measure on the Borel σ-algebra.

Then there exists a sequence of iid random variables X_n on $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution μ.

Proof. See the homework!

Theorem I.1.7

Let μ be a probability measure on a topological space $\mathcal{X}$. Then there exists some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which is defined a sequence of iid random variables with distribution μ.

We'll now swap over to notions of convergence. Much of the content here is similar to measure theory, but there are some new things, a lot of new language, and different things are generally emphasized.

Definition I.1.12

Let $(X_n)_{n \geq 1}$ be random variables in a metric space $(\mathcal{X}, D)$. We say that X_n converges to X,

(1) Almost surely provided that

$$\mathbb{P}[\lim_{n \to \infty} X_n = X] = 1,$$

which is equivalent to almost everywhere convergence of measurable functions.

(2) In probability provided that for all $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}[D(X_n, X) < \varepsilon] = 1,$$

which is equivalent to convergence in measure in measure theory (weaker than almost sure convergence).

(3) In distribution provided that $\mu_{X_n} \to \mu_X$ weakly (aka weak convergence of measures on $\mathcal{X}$). In other words for all bounded continuous functions $f : \mathcal{X} \to \mathbb{R}$ we have $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$. In measure theory language this is

$$\int_{\mathcal{X}} f(y) \, \mathrm{d}\mu_{X_n}(y) \to \int_{\mathcal{X}} f(y) \, \mathrm{d}\mu_X(y).$$

The relationship between these is summarized by

Almost sure convergence $\implies$ convergence in probability $\implies$ convergence in distribution.

Example I.1.12

Take $\{X_n\}_{n\geq 1}$, independent identically distributed. Then $X_n \to X_1$ in distribution, trivially, as $\mu_{X_n} = \mu_{X_1}$. However, independence will forbid these random variables from converging in probability (so long as the X_n are not constant).

Lemma I.1.8

Let $\{X_n\}_{n\geq 1}$ and X be random variables taking values in $\mathbb{R}$, then the following are equivalent

(i) $X_n \to X$ in distribution.

(ii) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all smooth, compactly supported f.

(iii) $\mathbb{P}[X_n \in [a, b]] \to \mathbb{P}[X \in [a, b]]$ for all $a < b$ with $\mathbb{P}[X \in \{a, b\}] = 0$.

Proof. Lets complete the cycle

- (i) $\implies$ (ii) because smooth compactly supported functions are in particular bounded and continuous.
- (ii) $\implies$ (iii). The idea is to construct a bump function approximating the indicator for $[a, b]$. Let $a < b$ and $\varepsilon > 0$. Let $g_\varepsilon : \mathbb{R} \to [0, 1]$ be smooth such that $g_\varepsilon \equiv 1$ on $[a + \varepsilon, b - \varepsilon]$ and $g_\varepsilon \equiv 0$ on $\mathbb{R} \setminus [a, b]$. Then we compute, for any Y, that

  $$\mathbb{P}[Y \in [a + \varepsilon, b - \varepsilon]] \leq \mathbb{E}[g_\varepsilon(Y)] \leq \mathbb{P}[Y \in [a, b]].$$

  If $\mathbb{P}[X \in \{a, b\}] = 0$, then

  $$\mathbb{P}[X \in [a + \varepsilon, b - \varepsilon]] = \mathbb{P}[X \in [a, b]] - o(1),$$

  where $o(1) \to 0$ as $\varepsilon \to 0$ (from the upper continuity of the measure). Thus

  $$\begin{aligned} \liminf_{n\to\infty} \mathbb{P}[X_n \in [a, b]] &\geq \liminf_{n\to\infty} \mathbb{E}[g_\varepsilon(X_n)] \\ &\geq \mathbb{E}[g_\varepsilon(X)] \\ &\geq \mathbb{P}[X \in [a + \varepsilon, b - \varepsilon]] \\ &= \mathbb{P}[X \in [a, b]] - o(1). \end{aligned}$$

  Thus $\liminf_{n\to\infty} \mathbb{P}[X_n \in [a, b]] \geq \mathbb{P}[X \in [a, b]]$. We can do a similar argument to get that $\limsup_{n\to\infty} \mathbb{P}[X_n \in [a, b]] \leq \mathbb{P}[X \in [a, b]]$. This time we'd approximate by a function whose support is slightly larger than the interval.
- (iii) $\implies$ (i) by a similar approximation argument. Namely we approximate a bounded continuous function by finite linear combinations of indicators of closed intervals. Notably μ_X, μ_{X_n} are finite measures, and so "most" of the measure will be in some compact set, so the behavior at ∞ is not too important.

Theorem I.1.9 (Weak Law of Large Numbers (LLN))

Let $\{X_n\}_{n\geq 1}$ be iid random variables taking values in $\mathbb{R}$ with finite variance $\operatorname{var}(X_n) < \infty$. Then

$$\frac{1}{n}\sum_{j=1}^{n} X_j \to \mathbb{E}[X_1]$$

in probability.

Proposition I.1.10 (Markov's Inequality)

Let X be a non-negative random variable in $\mathbb{R}$. Then for all $R > 0$,

$$\mathbb{P}[X > R] \leq \frac{\mathbb{E}[X]}{R}.$$

Applying this to X^p, for $p > 0$, we obtain "Chebyshev's inequality"

$$\mathbb{P}[X > R] = \mathbb{P}[X^p > R^p] \leq \frac{\mathbb{E}[X^p]}{R^p}.$$

Proof of Markov's Inequality. We see that $X \geq R\mathbb{1}_{X>R}$, hence

$$\mathbb{E}[X] \geq R\mathbb{E}[\mathbb{1}_{X>R}] = R\mathbb{P}[X > R].$$

Proof of Weak LLN. Write $a \equiv \mathbb{E}[X_1]$. We see that

$$\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n} X_j\right] = a.$$

Consequently, we have that

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n} X_j - a\right)^2\right] &= \operatorname{var}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)\\ &= \frac{1}{n^2}\operatorname{var}\left(\sum_{j=1}^{n} X_j\right)\\ &= \frac{1}{n^2}\sum_{j=1}^{n}\operatorname{var}(X_j).\end{aligned}$$

Great! Now since these are identically distributed, we obtain,

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n} X_j - a\right)^2\right] = \frac{1}{n^2}\cdot(n\operatorname{var}(X_1)) = \frac{\operatorname{var}(X_1)}{n}$$

Now by Markov's inequality

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{j=1}^{n} X_j - a\right| > \varepsilon\right] = \mathbb{P}\left[\left(\frac{1}{n}\sum_{j=1}^{n} X_j - a\right)^2 > \varepsilon^2\right)$$

$$\leq \frac{1}{\varepsilon^2} \cdot \frac{1}{n} \operatorname{var}(X_1).$$

Perfect! For fixed $\varepsilon > 0$, this goes to 0 as $n \to \infty$, and this is exactly the definition of convergence of $\frac{1}{n} \sum_{j=1}^{n} X_j \to a = \mathbb{E}[X_1]$ in probability.

Theorem I.1.11 (Strong Law of Large Numbers)

Let $\{X_n\}_{n \geq 1}$ be independent random variables with the same mean $a \in \mathbb{R}$. Assume there exists a $c > 0$ so that $\mathbb{E}[X_j^4] \leq c$ for all j. Then $\frac{1}{n} \sum_{j=1}^{n} X_j \to a$ almost surely.

Remark I.1.1

If $\{X_n\}_{n \geq 1}$ are iid with $\mathbb{E}[|X_1|] < \infty$ then $\frac{1}{n} \sum_{j=1}^{n} X_j \to \mathbb{E}[X_1]$ almost surely. See [DD19, Theorem 2.4.1].

Lemma I.1.12 (Borel-Cantelli)

Let $\{E_j\}_{j \geq 1}$ be events. If $\sum_{j=1}^{\infty} \mathbb{P}[E_j] < \infty$, then $\mathbb{P}[\text{infinitely many } E_j\text{'s occur}] = 0$.

Proof. The event that infinitely many E_j occur is the same as

$$\{\text{infinitely many } E_i \text{ occur}\} = \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} E_j.$$

But let $G_n = \bigcup_{j=n}^{\infty} E_j$. But then for all $m \in \mathbb{N}$.

$$\mathbb{P}[\bigcap_{n=1}^{\infty} G_n] \leq \mathbb{P}[G_m] \leq \sum_{j=m}^{\infty} \mathbb{P}[E_j],$$

and the right hand side goes to zero as $m \to \infty$ since $\sum_{j=1}^{\infty} \mathbb{P}[E_j] < 1$. Thus $\mathbb{P}[\bigcap_{n=1}^{\infty} G_n] = 0$, just as desired.

Proof. Proof of the law of large numbers. Assume without loss of generality that $a = 0$.

Claim

There exists a $c_1 > 0$ such that $\mathbb{E}\left[\left(\sum_{j=1}^{n} X_j\right)^4\right] \leq c_1 n^2$.

Alright, lets just expand

$$\mathbb{E}\left[\left(\sum_{j=1}^{n} X_j\right)^4\right] = \sum_{i,j,k,\ell=1}^{n} \mathbb{E}[X_i X_j X_k X_\ell].$$

By independence and having mean zero, any term where one of i, j, k, ℓ is different from the rest is zero. There are then two kinds of terms

(a) $\mathbb{E}[X_j^4]$ for some j,

(b) $\mathbb{E}[X_i^2 X_j^2] = \mathbb{E}[X_i^2]\mathbb{E}[X_j^2]$ for some i, j.

There are n terms of type (a), and by hypothesis $\mathbb{E}[X_j^4] \le c$. The total contribution then of terms of type (a) is nc. By Cauchy-Schwarz,

$$\mathbb{E}[X_j^2] \le \mathbb{E}[X_j^4]^{1/2} \le c^{1/2}.$$

Hence each term of type (b) is $\le c$. The total contribution is then $\le Cn^2$ for some constant depending on the combinatorics (the number of permutations of 4 copies of 2 distinct objects). Thus

$$\mathbb{E}\left[\left(\sum_{j=1}^n X_j\right)^4\right] = \sum_{i,j,k,\ell=1}^n \mathbb{E}[X_i X_j X_k X_\ell] \le cn + Cn^2,$$

and the result follows.

Great! Now consider

$$\begin{aligned} E_n &= \left\{\frac{1}{n}\sum_{j=1}^n X_j > n^{-1/8}\right\} \\ &= \left\{\sum_{j=1}^n X_j > n^{7/8}\right\}. \end{aligned}$$

By Markov, we see that

$$\begin{aligned} \mathbb{P}[E_n] &\le \frac{\mathbb{E}\left[\left(\sum_{j=1}^n X_j\right)^4\right]}{n^{7/2}} \\ &\le c_1 n^2 \cdot n^{-7/2} = c_1 n^{-3/2}. \end{aligned}$$

Great! By Borel-Cantelli, this tells us that almost surely E_n occurs for only finitely many n. This is equivalent to the statement that almost surely

$$\frac{1}{n}\sum_{j=1}^n X_j \le n^{-1/8}$$

for all n large enough. This implies that it goes to zero almost surely.

I.2. Fourier Analysis for Probability

We'll now begin the study of Fourier analysis in probability. In this context, the fourier transform is called a characteristic function. Namely

Definition I.2.1

Let X be a random variable in $\mathbb{R}$. The characteristic function $\varphi_X : \mathbb{R} \to \mathbb{C}$ is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{ity}\, \mathrm{d}\mu_X(y),$$

which is the Fourier transform of μ_X.

Lemma I.2.1

$t \mapsto \varphi_X(t)$ is continuous.

Proof. Let $t_n \to t$. Then

$$\left|e^{it_nX} - e^{itX}\right| \to 0$$

pointwise, and $\left|e^{it_nX} - e^{itX}\right| \le 2$, so the dominated convergence theorem applies to give $\varphi_X(t_n) \to \varphi_X(t)$.

Lemma I.2.2

If $X_1, \dots, X_n$ are independent, then

$$\varphi_{X_1+\cdots+X_n}(t) = \prod_{j=1}^{n} \varphi_{X_j}(t).$$

In the language of Fourier transforms, this is equivalent to the fourier transform of a convolution is the product of the fourier transforms.

Proof. We see that

$$\varphi_{X_1+\cdots+X_n}(t) = \mathbb{E}[e^{it(X_1+\cdots+X_n)}] = \mathbb{E}[e^{itX_1}\cdots e^{itX_n}] = \prod_j \mathbb{E}[e^{itX_j}] = \prod_j \varphi_{X_j}(t).$$

Lemma I.2.3

Suppose X is a normal distribution $X \sim N(a, \sigma^2)$. Then $\varphi_X(t) = e^{iat-\sigma^2t^2/2}$.

Proof. See Problem Set 2.

Lemma I.2.4

Let X be a random variable in $\mathbb{R}$, and let $n \in \mathbb{N}$ such that $\mathbb{E}[|X|^n] < \infty$. Then

$$\varphi_X^{(n)}(0) = i^n \mathbb{E}[X^n].$$

Proof. First we need an elementary inequality for exponentials. Namely, for all $y \in \mathbb{R}$, we have

$$\left|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}\right| \le \min\left(\frac{|y|^{n+1}}{(n+1)!}, \frac{2|y|^n}{n!}\right).$$

The proof of this is just elementary calculus (e.g. integration by parts). For a proof see Lemma 3.3.7 in [DD19].

Set $y = tX$ and take expectations, then

$$\begin{aligned}\left|\varphi_X(t) - \sum_{k=0}^{n} \frac{i^k\mathbb{E}[X^k]}{k!}t^k\right| &\le \mathbb{E}\left[\min\left\{\frac{t^{n+1}|X|^{n+1}}{(n+1)!}, \frac{2t^n|X|^n}{n!}\right\}\right] \\ &\le t^n\mathbb{E}\left[\min\left\{\frac{t|X|^{n+1}}{(n+1)!}, \frac{2|X|^n}{n!}\right\}\right].\end{aligned}$$

Now wait, the left hand side of this minimum goes to 0 pointwise as $t \to 0$, and the right hand side is bounded by assumption. Thus the dominated convergence theorem applies, and we find that

$$\left|\varphi_X(t) - \sum_{k=0}^{n} \frac{i^k \mathbb{E}[X^k]}{k!} t^k\right| = o(t^n)$$

as $t \to 0$. Perfect! This implies that $\sum_{k=0}^{n} \frac{i^k \mathbb{E}[X^k]}{k!} t^k$ is the degree n Taylor polynomial for φ_X at $t = 0$. Hence the result follows! $\varphi_X^{(n)}(0) = i^n \mathbb{E}[X^n]$.

Last time: We defined the characteristic function $\varphi_X(t) = \mathbb{E}[e^{itX}]$..

Proposition I.2.5

Let $\{X_j\}_{j \geq 1}$ and X be random variables in $\mathbb{R}$. Assume that $\varphi_{X_j}(t) \to \varphi_X(t)$ Lebesgue almost everywhere. Then $X_j \to X$ in distribution.

Proof. Let $g : \mathbb{R} \to \mathbb{R}$ be a smooth compactly supported function We need to show that

$$\mathbb{E}[g(X_j)] \to \mathbb{E}[g(X)].$$

Well, let

$$\widehat{g}(y) = \int_{\mathbb{R}} e^{-ixy} g(x) \, \mathrm{d}x,$$

and so since g is a Schwartz function (being compactly supported), so is $\widehat{g}$ by basic Fourier analysis. Thus we can apply the Fourier inversion formula to find that

$$g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{ixy} \widehat{g}(y) \, \mathrm{d}y.$$

Now we can look at the following

$$\begin{aligned}
\mathbb{E}[g(X_j)] &= \frac{1}{2\pi} \mathbb{E}\left[\int_{\mathbb{R}} e^{iX_j y} \widehat{g}(y) \, \mathrm{d}y\right] \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \mathbb{E}[e^{iX_j y}] \widehat{g}(y) \, \mathrm{d}y \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_{X_j}(y) \widehat{g}(y) \, \mathrm{d}y.
\end{aligned}$$

Here we've used that everything converges absolutely to exchange the expectation and the integral. Similarly

$$\mathbb{E}[g(X)] = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(y) \widehat{g}(y) \, \mathrm{d}y.$$

Note now that $\|\varphi_{X_j}\|_\infty, \|\varphi_X\|_\infty = 1$ and $\widehat{g}$ is Schwartz (so in particular L^1). Thus by the dominated convergence theorem (with $|\widehat{g}|$ dominating), we then see that $\mathbb{E}[g(X_j)] \to \mathbb{E}[g(X)]$.

Corollary I.2.6

If X, Y are random variables with $\varphi_X = \varphi_Y$ almost everywhere, then X, Y have the same distribution.

Proof. Take $X_j = Y$, then $\varphi_{X_j} = \varphi_X$, and so $X = X_j \to X$ in distribution. And thus $Y = X$ in distribution.

Theorem I.2.7 (Central Limit Theorem)

Let $\{X_j\}_{j\geq 1}$ be independent identically distributed random variables in $\mathbb{R}$ with $\mathbb{E}[X_j] = 0, \operatorname{var}(X_j) = 1$. Now let

$$Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j.$$

Then $Z_n \to N(0,1)$ (the standard normal distribution) in distribution.

Remark I.2.1

This is telling us some interesting information. We know $\mathbb{E}[Z_n] = 0$ from the law of large numbers, which is sort of first order information. The second order information, that $\sum_{j=1}^n X_j$ has size $\approx \sqrt{n}$ is very interesting, and is a phenomenon referred to as "square root cancellation." This is also telling us that $N(0,1)$ is actually very special, and not some random distribution. In some sense it is universal.

Proof. Let $Z \sim N(0,1)$. Then we know from the Complex Analysis homework that

$$\varphi_Z(t) = e^{-t^2/2}.$$

We want to show that $\varphi_{Z_n}(t) \to \varphi_Z(t)$ Lebesgue almost everywhere, so that we can use the proposition. Let $\varphi = \varphi_{X_1} = \varphi_{X_j}$ (since these all have the same distribution).

Then, using independent of $e^{itX_j/\sqrt{n}}$ we may write

$$\varphi_{Z_n}(t) = \mathbb{E}\left[\exp\left(it \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_j\right)\right] = \prod_{j=1}^{n} \mathbb{E}[e^{itX_j/\sqrt{n}}] = (\varphi(t/\sqrt{n}))^n.$$

Great! Now we know that $\varphi(0) = \mathbb{E}[e^0] = 1$. We also know that $\varphi'(0) = i\mathbb{E}[X_1] = 0$ and $\varphi''(0) = i^2\mathbb{E}[X_1^2] = -1$ by the mean and variance assumptions on the X_j. We can now do a Taylor expansion

$$\varphi(t/\sqrt{n}) = 1 - \frac{t^2}{2n} + o_t(1/n).$$

Where $o_t(1/n) \to 0$ possibly dependent on t. So now we see that

$$\begin{aligned} \operatorname{Log} \varphi_{Z_n}(t) = n \operatorname{Log} \varphi(t/\sqrt{n}) &= n \operatorname{Log}\left(1 - \frac{t^2}{2n} + o_t(1/n)\right) \\ &= n\left(-\frac{t^2}{2n} + o(1/n)\right) \to -\frac{t^2}{2}. \end{aligned}$$

Here we've used the taylor expansion of the logarithm... but note that this is a complex logarithm, which is scary! Well we can choose the branch cut on the negative real axis, because we only care about the logarithm of something close to 1. Hence

$$\varphi_{Z_n}(t) \to e^{-t^2/2} = \varphi_Z(t),$$

just as desired! An alternative proof could use that for all $a \in \mathbb{R}$,

$$\lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = e^a,$$

and explicit errors on this estimate.

II. Conditional Expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Here as usual we think of $\mathcal{G}$ as a "packet of information." If X is a random variable taking values in $\mathbb{R}$, then we can think of $\mathbb{E}[X]$ as our "best guess" for the value of X. We wish to define $\mathbb{E}[X \mid \mathcal{G}]$, the conditional expectation given $\mathcal{G}$. In other words, our "best guess for X given information in $\mathcal{G}$." This will in fact be a random variable.

Remark II.0.1

We shouldn't think of $\mathbb{E}[X \mid \mathcal{G}]$ as saying we know that events in $\mathcal{G}$ happened. But rather, as saying that we can tell whether or not events in $\mathcal{G}$ happen with our measurement devices.

Now lets translate this idea into abstract measure theory. We'll give it in terms of its properties and then show it exists and is unique.

Definition II.0.1

Let X be a random variable in $\mathbb{R}$, $\mathbb{E}[|X|] < \infty$. Then the conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}[X \mid \mathcal{G}]$, is the unique $\mathcal{G}$-measurable random variable such that for all $G \in \mathcal{G}$ we have

$$\mathbb{E}[\mathbb{1}_G \mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X \mathbb{1}_G].$$

In other words,

$$\int_G \mathbb{E}[X \mid \mathcal{G}](\omega) \, \mathrm{d}\mathbb{P}(\omega) = \int_G X(\omega) \, \mathrm{d}\mathbb{P}(\omega).$$

Another way to think of this is that $\mathbb{E}[X \mid \mathcal{G}]$ is the "projection (in L^2) of X onto $\{\mathcal{G}$-measurable random variables$\}$." This definition doesn't work quite if $X \notin L^2$, but you can do some approximation arguments.

Proposition II.0.1

In the setting of the above definition, $\mathbb{E}[X \mid \mathcal{G}]$ exists and is unique (up to events of probability 0).

Proof of Existence. We're going to apply the Radon-Nikodym theorem. To start, lets do some reductions

(1) If $X = X_1 - X_2$ and $\mathbb{E}[X_1 \mid \mathcal{G}], \mathbb{E}[X_2 \mid \mathcal{G}]$ exist we can just set $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X_1 \mid \mathcal{G}] - \mathbb{E}[X_2 \mid \mathcal{G}]$.

(2) We can write $X = X \mathbb{1}_{X \geq 0} - (-X)\mathbb{1}_{X<0}$. Thus we can assume that $X \geq 0$ (is nonnegative).

Now for $G \in \mathcal{G}$ we let

$$\nu(G) = \mathbb{E}[\mathbb{1}_G X] = \int_G X(\omega) \, \mathrm{d}\mathbb{P}(\omega).$$

This is a measure on G, and if $\mathbb{P}[G] = 0$, we see that $\nu(G) = 0$. Therefore ν is absolutely continuous with respect to $\mathbb{P}$ defined on $\mathcal{G}$. Thus, by the Radon-Nikodym theorem, there exists a nonnegative $\mathcal{G}$-measurable function $Z = \frac{\mathrm{d}\nu}{\mathrm{d}\mathbb{P}}$ so that

$$\nu(G) = \int_G Z(\omega) \, \mathrm{d}\mathbb{P}(\omega)$$

Thus $\mathbb{E}[\mathbb{1}_G X] = \mathbb{E}[\mathbb{1}_G Z]$, and Z is $\mathcal{G}$-measurable, so Z can be $\mathbb{E}[X \mid \mathcal{G}]$.

Proof of Uniqueness. Assume that $Z, \widetilde{Z}$ are $\mathcal{G}$-measurable and they both satisfy the definition of $\mathbb{E}[X \mid \mathcal{G}]$. Let $G = \{Z > \widetilde{Z}\} \in \mathcal{G}$. Hence

$$\mathbb{E}[(Z - \widetilde{Z})\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G] - \mathbb{E}[X\mathbb{1}_G] = 0.$$

Because $(Z - \widetilde{Z})$ is positive on $\mathbb{1}_G$, we see this can only happen if $\mathbb{P}[Z > \widetilde{G}] = 0$. Likewise $\mathbb{P}[Z < \widetilde{G}] = 0$.

Last Time: We defined the conditional expectation of a random variable X with $\mathbb{E}[|X|] < \infty$ given a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ is defined as a $\mathcal{G}$-measurable random variable $\mathbb{E}[X \mid \mathcal{G}]$ satisfying

$$\mathbb{E}[X \mathbb{1}_G] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_G],$$

where $G \in \mathcal{G}$. We'll now develop some lemmas that show that thinking of this as the "best guess for X given $\mathcal{G}$" is correct. All the lemmas will be trivial to prove, motivating this definition as the right one.

Lemma II.0.2

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

Proof. Take $G = \Omega$.

Lemma II.0.3

For any constants $a, b \in \mathbb{R}$, we have

$$\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$$

Proof. Check that $a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$ satisfies the definition of $\mathbb{E}[aX + bY \mid \mathcal{G}]$ using the linearity of expectation.

Lemma II.0.4

If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X \mid \mathcal{G}] = X$.

Proof. By definition.

Lemma II.0.5

If $\sigma(X)$ is independent from $\mathcal{G}$, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$.

Proof. Observe that $\mathbb{E}[X]$ is $\mathcal{G}$-measurable (being a constant) and

$$\begin{aligned} \mathbb{E}[\mathbb{1}_G X] &= \mathbb{E}[\mathbb{1}_G]\mathbb{E}[X] \\ &= \mathbb{E}[\mathbb{E}[X]\mathbb{1}_G]. \end{aligned}$$

by independence and linearity of expectation.

Lemma II.0.6

If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ are σ-algebras, then

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

Proof. $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}]$ satisfies the defining property of $\mathbb{E}[X \mid \mathcal{H}]$.

Lemma II.0.7

If Y is $\mathcal{G}$-measurable and $\mathbb{E}[|XY|] < \infty$, then

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

Proof. First we prove this for $Y = \mathbb{1}_G$ for $G \in \mathcal{G}$, then for all $H \in \mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}[\mathbb{1}_H Y \mathbb{E}[X \mid \mathcal{G}]] &= \mathbb{E}[\mathbb{1}_{G\cap H}\mathbb{E}[X \mid \mathcal{G}]] \\ &= \mathbb{E}[\mathbb{1}_{G\cap H}X] = \mathbb{E}[\mathbb{1}_G Y X].\end{aligned}$$

Hence $Y\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[XY \mid \mathcal{G}]$ for Y an indicator function. In general, we approximate Y by a linear combination of indicator functions and use linearity.

Definition II.0.2

Let X be a random variable in $\mathbb{R}$ with $\mathbb{E}[|X|] < \infty$, and let Y be any random variable. Then we define the conditional expectation of X given Y by $\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)]$.

In the aim of being concrete, lets link this directly with a very traditional way of thinking about conditional expectation from probability.

Example II.0.1

Let Y be a random variable taking values in a countable set A, X be a random variable taking values in $\mathbb{R}$ with $\mathbb{E}[|X|] < \infty$. We'll now compute $\mathbb{E}[X \mid Y]$.

Claim

$\mathbb{E}[X \mid Y] = \sum_{a\in A} \frac{\mathbb{E}[X\mathbb{1}_A]}{\mathbb{P}[Y=a]}\mathbb{1}_{Y=a}$.

Proof. Let Z be the right hand side. Notably Z is $\sigma(Y)$ measurable. We must show for any $G \in \sigma(Y)$ that

$$\mathbb{E}[Z\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G].$$

Since A is countable, we can write $\mathbb{1}_G = \sum_{b\in B}\mathbb{1}_{Y=b}$ for some subset $B \subseteq A$, because every event $G = \{Y \in B\}$ for $G \in \sigma(Y), B \subseteq A$. By linearity, we can assume that $G = \{Y = b\}$ for some $b \in A$.

Then we find that

$$\begin{aligned}\mathbb{E}[Z\mathbb{1}_{Y=b}] &= \sum_{a\in A}\frac{\mathbb{E}[X\mathbb{1}_{Y=a}]}{\mathbb{P}[Y=a]}\mathbb{E}[\mathbb{1}_{Y=a}\mathbb{1}_{Y=b}] \\ &= \frac{\mathbb{E}[X\mathbb{1}_{Y=b}]}{\mathbb{P}[Y=b]}\cdot\mathbb{P}[Y=b] = \mathbb{E}[X\mathbb{1}_{Y=b}].\end{aligned}$$

Perfect! This is just what we wanted!

To go even more concrete.

Example II.0.2

Consider $Y = \mathbb{1}_F, F \in \mathcal{F}$, $\mathbb{P}[F] > 0$. We find that

$$\mathbb{E}[X \mid Y] = \frac{\mathbb{E}[X\mathbb{1}_F]}{\mathbb{P}[F]}\mathbb{1}_F + \frac{\mathbb{E}[X\mathbb{1}_{F^c}]}{\mathbb{P}[F^c]}\mathbb{1}_{F^c}.$$

And if $X = \mathbb{1}_E$ for some event E, we have

$$\mathbb{E}[X \mid Y] = \frac{\mathbb{P}[E\cap F]}{\mathbb{P}[F]}\mathbb{1}_F + \frac{\mathbb{E}[X\cap F^c]}{\mathbb{P}[F^c]}\mathbb{1}_{F^c}.$$

So, on F, this is exactly $\mathbb{P}[E \mid F]$, the conditional probability.

Definition II.0.3

For $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra, and $E \in \mathcal{F}$ we can take

$$\mathbb{P}[E \mid \mathcal{G}] = \mathbb{E}[\mathbb{1}_E \mid \mathcal{G}].$$

This is an okay definition, but really we want a conditional distribution, being able to tell the distribution of all events given the information of a σ-algebra.

Definition II.0.4

Let $\mathcal{G} \subseteq \mathcal{F}$. Let $X : \Omega \to \mathcal{X}$ be a random variable. Let $M : \Omega \to \mathrm{Prob}(\mathcal{X})$ be a $\mathcal{G}$-measurable random variable taking values in

$$\mathrm{Prob}(\mathcal{X}) = \{\text{Borel Probability measures on } \mathcal{X}\},$$

where we place the topology of weak convergence on $\mathrm{Prob}(\mathcal{X})$. We say that M is the (regular) conditional distribution of X given $\mathcal{G}$ if for all Borel sets $A \subseteq \mathcal{X}$ almost surely

$$\mathbb{P}[X \in A \mid \mathcal{G}] = M(A).$$

Note: both sides of this equation are $\mathcal{G}$-measurable random variables taking values in $[0, 1]$.

Theorem II.0.8

If $\mathcal{X}$ is a separable metric space, then for all σ-algebras $\mathcal{G} \subseteq \mathcal{F}$, and for all random variables X taking values in $\mathcal{X}$, the conditional distribution of X given $\mathcal{G}$ exists and is unique.

Proof. Section 5.1.3 of Durrett's book [DD19].

Remark II.0.2

In general in math, a good goal is to break down complicated things into simpler things and understand each smaller thing more easily. Here are some examples

- A function f, break it down as $f = g \circ h$ and understand g, h.
- To understand a group G, find a normal subgroup H and try to understand $H, G/H$.
- ...

In probability, the analogy is

- To understand a random variable X, find some random variable Y so that the distribution of Y and the conditional distribution of X given Y are somehow simpler to understand than the distribution of X.

Example II.0.3

Let X and Y be independent. Lets compute the conditional distribution of X given $\sigma(Y)$. This is just constant at μ_X, the unconditional distribution of X. Namely

$$\mathbb{P}[X \in A \mid \sigma(Y)] = \mathbb{P}[X \in A] = \mu_X(A).$$

Example II.0.4

X is $\sigma(Y)$-measurable, then the conditional distribution of X given $\sigma(Y)$ is a point mass at X, because

$$\mathbb{P}[X \in A \mid \sigma(Y)] = \begin{cases} 1 & \text{if } X \in A \\ 0 & \text{if } X \notin A \end{cases},$$

this is a special case of $\mathbb{E}[Z \mid \mathcal{G}] = Z$ if Z is $\mathcal{G}$-measurable.

Last time: Defined conditional distribution of X given $\mathcal{G}$ via

$$M(A) = \mathbb{P}[X \in A \mid \mathcal{G}]$$

Example II.0.5

Let X be a random variable taking values in $\mathcal{X}$ and Y be a random variable taking values in a countable set $\mathcal{Y}$. For each $y \in \mathcal{Y}$ such that $\mathbb{P}[Y = y] > 0$ we can define a measure

$$M_y(A) = \frac{\mathbb{P}[X \in A, Y = y]}{\mathbb{P}[Y = y]} = \mathbb{P}[X \in A \mid Y = y].$$

For all $A \subseteq \mathcal{X}$ Borel.

Claim

M_Y is the conditional distribution of X given $\sigma(Y)$. To be clear, this is the function

$$M_Y(A) : \omega \mapsto M_{Y(\omega)}(A) = \frac{\mathbb{P}[X \in A, Y = Y(\omega)]}{\mathbb{P}[Y = Y(\omega)]}.$$

We want to show that $M_Y(A) = \mathbb{P}[X \in A \mid Y]$. Well this is if and only if for all $B \subseteq \mathcal{Y}$ we have

$$\mathbb{E}[M_Y(A)\mathbb{1}_{Y \in B}] = \mathbb{E}[\mathbb{1}_{X \in A}\mathbb{1}_{Y \in B}] = \mathbb{P}[X \in A, Y \in B],$$

by the definition of conditional probability. Well, lets evaluate

$$\mathbb{E}[M_Y(A)\mathbb{1}_{Y \in B}] = \sum_{y \in B} \mathbb{P}[Y = y] M_y(A) = \sum_{y \in B} \mathbb{P}[X \in A, Y = y] = \mathbb{P}[X \in A, Y \in B].$$

Great! This is just as desired!

There is a similar expression for when the random variables X, Y take values in $\mathbb{R}$, with (X, Y) having a density. There the sum is replaced by an integral.

III. Brownian Motion

III.1. Motivation and The Big Idea

Brownian Motion is the answer to the following question: What is the most natural random continuous function $[0, \infty) \to \mathbb{R}$? Equivalently, what is the most natural probability measure on (or random variable in) $C([0, \infty), \mathbb{R})$ with the topology of locally uniform convergence.

The idea will be to discretize the problem. What is the most natural random function on $\mathbb{N}_0 \to \mathbb{Z}$? To be precise, let $\{\xi_j\}_{j \geq 1}$ be identically distributed random variables with

$$\mathbb{P}[\xi_j = 1] = \mathbb{P}[\xi_j = -1] = \frac{1}{2}.$$

Let $S_0 = 0$ and $S_n = \sum_{j=1}^n \xi_j$. THen $n \mapsto S_n$ is a random function from $\mathbb{N}_0 \to \mathbb{Z}$.

Definition III.1.1

Such a function $\mathbb{N}_0 \to \mathbb{Z}$ is called a random walk on $\mathbb{Z}$. We can extend to $S : [0, \infty) \to \mathbb{R}$ by piecewise linear interpolation.

The idea to get a continuous function, is to now rescale. Namely we'll intuitively take the limit $t \mapsto S_{nt}$ as $n \to \infty$. The problem here first is that this won't converge, we'll need a good rescaling. By the Central Limit Theorem, we know that $n^{-1/2} S_n$ converges in distribution to a standard normal distribution $N(0, 1)$. This tells us the right scaling factor is $n^{-1/2} S_{nt}$.

We see in fact that

$$n^{-1/2} S_{nt} = \frac{(tn)^{1/2}}{n^{1/2}} (tn)^{-1/2} S_{nt} \to t^{1/2} N(0, 1) = N(0, t),$$

in distribution. We also want to know the information of the joint distribution of t. Namely, for any $t > 0$, we have that

$$\{S_{u+\lfloor tn \rfloor} - S_{\lfloor tn \rfloor}\}_{u \geq 0}$$

is independent from $S\big|_{0,\lfloor tn \rfloor}$ and has the same distribution as $\{S_u\}_{u \geq 0}$. Why? Well this is obtained in the same way as $\{S_u\}_{u \geq 0}$ but using $\{\xi_{j+\lfloor tn \rfloor}\}_{j \geq 0}$.

Iterating this for $0 \leq t_0 < t_1 < \cdots < t_N$ we find that

$$(S_{\lfloor t_1 n \rfloor} - S_{\lfloor t_0 n \rfloor}, \ldots, S_{\lfloor t_N n \rfloor} - S_{\lfloor t_{N-1} n \rfloor}),$$

are all independent, and $S_{\lfloor t_j n \rfloor} - S_{\lfloor t_{j-1} n \rfloor} \overset{d}{=} S_{\lfloor (t_j - t_{j-1}) n \rfloor}$. What this actually shows is that, if we look at this N-tuple and rescale it we get

$$\frac{1}{n^{1/2}} (S_{\lfloor t_1 n \rfloor} - S_{\lfloor t_0 n \rfloor}, \ldots, S_{\lfloor t_N n \rfloor} - S_{\lfloor t_{N-1} n \rfloor}) \xrightarrow{d} (X_1, \ldots, X_N)$$

of independent X_j with $X_j \sim N(0, t_j - t_{j-1})$.

This really pins down the limit of the process $t \mapsto n^{-1/2} S_{nt}$, and tells us something about its properties. We'll use this to make a definition.

III.2. Definition and Properties

Ok! Using the motivation from last section, we'll now make a definition via properties, and discuss its existence.

Definition III.2.1

Brownian Motion is the random continuous function $B : [0, \infty) \to \mathbb{R}, t \mapsto B_t$ so that $B_0 = 0$ and

For all $s < t$, $B_t - B_s \sim N(0, t - s)$.

(ii) For all $0 \leq t_0 < \cdots < t_N$, the increments $B_{t_j} - B_{t_{j-1}}$ for $j = 1, \ldots, N$ are independent.

By this, we mean that B is a random variable taking values in $C([0, \infty), \mathbb{R})$.

It is a nontrivial theorem that...

Theorem III.2.1

Brownian Motion Exists.

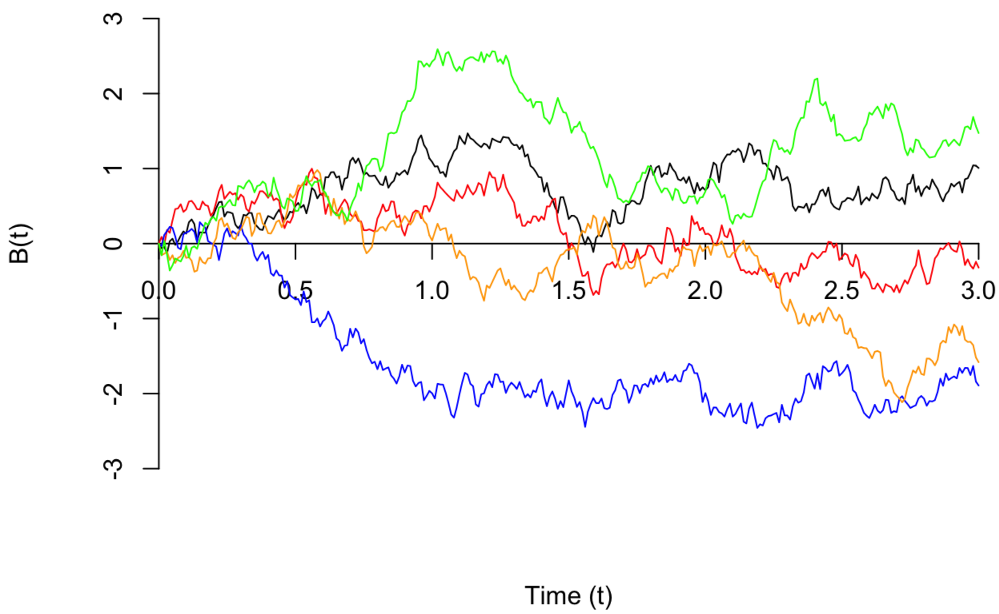


Figure 1. Brownian Motion Graphs

For a proof using the motivation from last section, see [DD19, Section 8]. We'll see a different proof in Week 9 from a different lecturer while Professor Gwynne is at a conference ☺.

Remark III.2.1

Lets discuss some applications / realities to Brownian Motion

- The motion of particles in a dust cloud, e.g. motion of a single particle is tracked by three independent brownian motions. This phenomenon was studied by Brown and Einstein.
- Fluctuations of Stock Prices. Keyword: Black-Scholes equation.
- Feynman Path Integrals.
- Solutions of PDEs, and analyzing behavior of conformal mappings in complex analysis.

Last Time: We defined Brownian motion as a random continuous function $B : [0, \infty) \to \mathbb{R}$ satisfying $B_0 = 0$ and

(i) For all $s < t$, $B_t - B_s \sim N(0, t - s)$.

(ii) For all $0 \leq t_0 \leq t_1 \leq \cdots \leq t_N$, $B_{t_j} - B_{t_{j-1}}$ are independent.

Lemma III.2.2

If $B, \widetilde{B}$ satisfy the definition of Brownian Motion then $B \overset{d}{=} \widetilde{B}$.

Proof. By definition, for all $t_1 \leq \cdots \leq t_N$, we have

$$(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_N} - B_{t_{N-1}}) \overset{d}{=} (\widetilde{B}_{t_1}, \widetilde{B}_{t_2} - \widetilde{B}_{t_1}, \ldots, \widetilde{B}_{t_N} - \widetilde{B}_{t_{N-1}}).$$

Then this implies that

$$(B_{t_1}, \dots, B_{t_N}) \overset{d}{=} (\widetilde{B}_{t_1}, \dots, \widetilde{B}_{t_N}).$$

Because Brownian motion is continuous, we can chop up any finite time interval into $\varepsilon > 0$ size intervals. Thus the joint distribution being the same for any finite collection of times being the same implies $B \overset{d}{=} \widetilde{B}$.

Lemma III.2.3 (Markov Property)

Let B be Brownian motion. Then for all $t > 0$, $s \mapsto B_{s+t} - B_t$ is a Brownian Motion independent from $B\big|_{[0,t]}$.

Proof. Let $\widetilde{B}_s = B_{s+t} - B_t$. It's easy to check that $\widetilde{B}_s$ satisfies the defining properties for Brownian motion. By independent increments, we have for any $0 \le t_1 \le \cdots \le t_N \le t, 0 \le s_1 \le \cdots \le s_M$ that the random vectors

$$(B_{t_1}, \dots, B_{t_N}), (\widetilde{B}_{s_1}, \dots, \widetilde{B}_{s_M})$$

are independent. By an approximation argument, it follows that $B_t, \widetilde{B}_s$ are independent.

Remark III.2.2

An equivalent formulation is that the conditional distribution of $s \mapsto B_{s+t} - B_t$ given $B\big|_{[0,t]}$ is just the distribution of Brownian Motion.

Lemma III.2.4 (Brownian Scaling)

If B is a Brownian Motion, then for any constant $c > 0$, we have that $\{c^{-1/2} B_{ct}\}_{t \ge 0}$ is a Brownian Motion.

Proof. Clearly $t \mapsto c^{-1/2} B_{ct}$ is continuous with independent increments. Furthermore, if $X \sim N(0, t)$ then $c^{-1/2} X \sim N(0, t/c)$. Hence, $c^{-1/2} B_{ct} \sim c^{-1/2} N(0, ct) = N(0, t)$ as desired. Thus $t \mapsto c^{-1/2} B_{ct}$ is Brownian Motion.

Now we'll mention a theorem we won't prove but is important to be aware of.

Theorem III.2.5

Let $\widetilde{B} : [0, \infty) \to \mathbb{R}$ be a random continuous function such that $\widetilde{B}_0 = 0$ and

(i) For all $s < t$, we have $\widetilde{B}_t - \widetilde{B}_s \overset{d}{=} \widetilde{B}_{t-s}$.

(ii) For any $0 \le t_0 \le \cdots \le t_N$ it holds that $\widetilde{B}_{t_j} - \widetilde{B}_{t_{j-1}}$ are independent.

Then $\widetilde{B}$ is "essentially" a Brownian motion. Namely there exists $a \ge 0, b \in \mathbb{R}$ such that $\widetilde{B}_t = aB_t + bt$ where B is a Brownian Motion. Notably here, condition (i) does not give us that these are normals.

Remark III.2.3

The idea of the proof is to apply the Central Limit Theorem to the increments. However, actually carrying this out is difficult in practice, since e.g. it is not immediately clear that the increments have finite variance.

Okay! So now lets think more about properties of Brownian Motion to see why this is so incredible. We're going to analyze $s \mapsto B_{s+\tau}$ where τ is a random time. To make notational convenience, let

Definition III.2.2

$\mathcal{F}_t = \sigma(B\big|_{[0,t]}$

And so that we can actually analyze these, we only consider a special type of time τ.

Definition III.2.3

A random time τ in $[0, +\infty]$ is a stopping time for B if for all $t \geq 0$ if $\{\tau \leq t\} \in \mathcal{F}_t$.

Example III.2.1

Any non-random time is a stopping time. For more interesting examples:

- $\tau = \min\{t \geq 0 \mid B_t \in A\}$ where $A \subseteq \mathbb{R}$ is closed. Intuitively, you can tell if you've hit A before time t by looking at the behavior of B from time 0 to t. Formally

$$\{\tau \leq t\} = \{B[0,t] \cap A \neq \emptyset\}$$

- $\tau = \min\{\tau_1, \tau_2\}$ or $\tau = \max\{\tau_1, \tau_2\}$ where τ_1, τ_2 are stopping times. You can do this even for countably many stopping times as

$$\{\min(\tau_1, \tau_2) \leq t\} = \{\tau_1 \leq t\} \cup \{\tau_2 \leq t\}$$
$$\{\max(\tau_1, \tau_2) \leq t\} = \{\tau_1 \leq t\} \cap \{\tau_2 \leq t\}$$

Non-Example III.2.2

Let $A \subseteq \mathbb{R}$ be closed. Then let

$$\tau = \sup\{\tau \in [0,1] \mid B_t \in A\}.$$

The last time that Brownian motion hits something is not a stopping time!

Remark III.2.4

An intuitive way to think of stopping times is that they make good directions! Turn right at the first stoplight is a good direction. Turn right at the last stoplight is a bad direction (hence not a stopping time).

Using this definition, we can upgrade the Markov Probability!

Theorem III.2.6 (Strong Markov Property)

Let τ be a stopping time such that $\mathbb{P}[\tau < \infty] = 1$. Then $s \mapsto B_{s+\tau} - B_\tau$ is a Brownian Motion which is independent from $B\big|_{[0,\tau]}$.

Remark III.2.5

If $\mathbb{P}[\tau = \infty] \in (0,1)$ then instead on the event $\{\tau < \infty\}$, the conditional distribution of $s \mapsto B_{s+\tau} - B_\tau$ given $B\big|_{[0,\tau]}$ is Brownian Motion.

Also, to see why you stopping times are necessary. Consider $\tau = \sup\{\tau \in [0,1] \mid B_t \in A\}$. We then see that $B_{s+\tau}$ cannot hit A, and so $B_{s+\tau} - B_\tau$ will not be independent from $B\big|_{[0,\tau]}$!

Proof. For $t \geq 0$, consider

$$\underline{B}_s^t = B_{\min\{s,t\}}$$

$$\overline{B}^t_s = B_{s+t} - B_t.$$

Note that $\sigma(\underline{B}^t) = \sigma(B\big|_{[0,t]})$. Now we want to show that $\overline{B}^\tau \overset{d}{=} B$ and $\overline{B}^\tau$ is independent from $\underline{B}^\tau$.

Claim

The result holds when there exists a countable set $T \subseteq \mathbb{R}$ such that $\mathbb{P}[\tau \in T] = 1$.

Let $U, V \subseteq C([0,\infty), \mathbb{R})$ be measurable. We need to show that

$$\mathbb{P}[\overline{B}^\tau \in U, \underline{B}^\tau \in V] = \mathbb{P}[B \in U]\mathbb{P}[\underline{B}^\tau \in V],$$

This is sufficient since if we take V to be the whole space, this reduces to $\overline{B}^\tau \overset{d}{=} B$. The fact that it splits as a product is then independence. Now we have

$$\begin{aligned}\mathbb{P}[\overline{B}^\tau \in U, \underline{B}_\tau \in V] &= \sum_{t \in T} \mathbb{P}[\overline{B}^\tau \in U, \underline{B}^\tau \in V, \tau = t] \\ &= \sum_{t \in T} \mathbb{P}[\overline{B}^t \in U, \underline{B}^t \in V, \tau = t].\end{aligned}$$

Great! Now $\{\tau = t\} \in \mathcal{F}_t = \sigma(\underline{B}^t)$. Furthermore, by the ordinary Markov property, $\overline{B}^t \overset{d}{=} B$ and is independent from $\underline{B}^t$. Notably, it's also independent from the event $\{\tau = t\}$. Hence

$$\begin{aligned}\sum_{t \in T} \mathbb{P}[\overline{B}^t \in U, \underline{B}^t \in V, \tau = t] &= \sum_{t \in T} \mathbb{P}[B \in U]\mathbb{P}[\underline{B}^t \in V, \tau = t] \\ &= \mathbb{P}[B \in U] \sum_{t \in T} \mathbb{P}[\underline{B}^t \in V, \tau = t] \\ &= \mathbb{P}[B \in U]\mathbb{P}[\underline{B}^\tau \in V].\end{aligned}$$

Perfect! Combining this with the equation above, we have the proof!

In general, we take a general stopping time τ, and then define

$$\tau_n \coloneqq 2^{-n} \lceil 2^n \tau \rceil$$

This is the first integer multiple of 2^{-n} coming after τ. This is a stopping time, since τ is a stopping time. Thus the result holds for $\tau_n \in 2^{-n}\mathbb{Z}$, as there are only countably many values. Furthermore $\lim_{n\to\infty} \tau_n = \tau$. Thus, for all n, $\overline{B}^{\tau_n} \overset{d}{=} B$ and is independent from $\underline{B}_{\tau_n}$.

Furthermore $\overline{B}^{\tau_n} \to \overline{B}^\tau$ and $\underline{B}^{\tau_n} \to \underline{B}^\tau$ converge almost surely. Hence $\overline{B}^\tau \overset{d}{=} B$ and is independent from $\underline{B}^\tau$, by basic checks with almost sure convergence.

Last time: We defined what a stopping time τ for a Brownian Motion B is, and we showed the Strong Markov Property, namely that if τ is a stopping time then $B_{s+\tau} - B_\tau$ is a Brownian Motion independent from $B\big|_{[0,\tau]}$.

As a consequence

Proposition III.2.7 (Reflection Principle)

Let $a > 0$. Then for all $t > 0$, both of the following equivalent statements are true

$$\mathbb{P}\left[\max_{0\leq s\leq t} B_s \geq a\right] = 2\mathbb{P}[B_t \geq a] \iff \max_{0\leq s\leq t} B_s \overset{d}{=} |B_t|.$$

Note: this equality does not hold in distribution as functions of t. One can see this since the left hand side always goes up, while the right hand side can increase or decrease.

Proof. Let $\tau = \min\{t \mid B_t = a\}$ be a stopping time. Then

$$\{\max_{0\leq s\leq t} B_s \geq a\} = \{\tau \leq t\}.$$

In particular, if $B_t \geq a$ then $\tau \leq t$. Hence, by the definition of conditional probability,

$$\mathbb{P}[B_t \geq a] = \mathbb{P}[\tau \leq t]\mathbb{P}[B_t \geq a \mid \tau \leq t].$$

By the strong Markov property, the conditional distribution of $B_t - B_\tau$ given $B\big|_{[0,\tau]}$ is $N(0, t-\tau)$. Of course $B_\tau = a$, and so

$$\mathbb{P}[B_t \geq a \mid B\big|_{[0,\tau]}] = \mathbb{P}[N(0, t-\tau) \geq 0] = \frac{1}{2}.$$

If we multiply both sides of this by $\mathbb{1}_{\tau \leq t}$ we obtain

$$\mathbb{P}[B_t \geq a \mid B\big|_{[0,\tau]}]\mathbb{1}_{\tau\leq t} = \frac{1}{2}\tau \leq \mathbb{t}.$$

Hence, taking expectations on both sides yields

$$\mathbb{P}[B_t \geq a] = \frac{1}{2}\mathbb{P}[\tau \leq t] = \frac{1}{2}\mathbb{P}\left[\max_{0\leq s\leq t} B_s \geq a\right].$$

Great! Why is this equivalent to the second relation? Well

$$\mathbb{P}[|B_t| \geq a] = \mathbb{P}[B_t \geq a] + \mathbb{P}[B_t \leq -a] = 2\mathbb{P}[B_t \geq a],$$

by symmetry of the normal distribution. Hence $|B_t|$ and $\max_{0\leq s\leq t} B_s$ have the same ("reversed") cumulative distribution functions.

Proposition III.2.8

Almost surely, we have

$$\limsup_{t\to\infty} B_t = +\infty \qquad \liminf_{t\to\infty} B_t = -\infty.$$

Proof. By the reflection principle,

$$\max_{0\leq s\leq t} B_s \overset{d}{=} |B|_t \overset{d}{=} t^{1/2}|B_1|.$$

Furthermore, for any $a > 0$,

$$\lim_{t\to\infty} \mathbb{P}[t^{1/2}|B_1| > a] = \lim_{t\to\infty} \mathbb{P}[|B_1| > t^{-1/2}a] = 1.$$

Hence we see that

$$\lim_{t\to\infty} \mathbb{P}[\max_{0\le s\le t} B_s > a] = 1.$$

This implies that $\limsup_{t\to\infty} B_t = +\infty$ almost surely. Hence by symmetry ($B \overset{d}{=} -B$, $\liminf_{t\to\infty} B_t = -\infty$.

III.3. Brownian Motion in $\mathbb{R}^d$

Now lets look at Brownian Motion in higher dimensions. The definition is rather simple!

Definition III.3.1

Let $d \in \mathbb{N}$. The Brownian Motion in $\mathbb{R}^d$ is the random continuous function $B = (B^1, \dots, B^d) : [0,\infty) \to \mathbb{R}^d$ where $B^1, \dots, B^d$ are independent 1-dimensional brownian motions.

Proposition III.3.1

Brownian Motion in $\mathbb{R}^d$ is the unique (in distribution) random continuous function $B : [0,\infty) \to \mathbb{R}^d$ satisfying $B_0 = 0$ and

(i) For any time $s < t$, we have $B_t - B_s$ is a vector of d independent Gaussians $N(0, t-s)$.
(ii) For $0 \le t_0 \le t_1 \le \cdots \le t_N$, we have that $B_{t_j} - B_{t_j - 1}$ are independent.

The proof of this is exactly the 1-dimensional proof. We also have many properties similar to the one-dimensional case.

Lemma III.3.2 (Brownian Scaling in $\mathbb{R}^d$)

For any $c > 0$, the function $t \mapsto c^{-1/2} B_{ct}$ is a d-dimensional Brownian Motion.

Lemma III.3.3 (Strong Markov Property)

Let τ be a stopping time for B. Then $s \mapsto B_{s+\tau} - B_\tau$ is a d-dimensional Brownian Motion independent from $B\big|_{[0,\tau]}$.

Proposition III.3.4 (Rotational Invariance)

Let Q be a $d \times d$ orthogonal matrix. Then $Q \cdot B \overset{d}{=} B$

Proof. QB is continuous, $QB_0 = 0$, and QB has independent increments (they are Q times the increments for B). Now for $s < t$ we have

$$QB_t - QB_s = Q(B_t - B_s),$$

now $B_t - B_s$ is a vector of d independent $N(0, t-s)$ Gaussians. By Problem Set 5, $Q(B_t - B_s)$ is in fact a vector of d independent $N(0, t-s)$ Gaussians.

We'll now state some geometric properties without proof. If you want to see proofs see Brownian Motion by Mörters-Peres [MP10].

- For $d = 2$, almost surely we have
  - For all $a < b$, there exists $s, t \in [a, b]$, $s \neq t$ so that $B_s = B_t$.

  - Neighborhood recurrence: For all $z \in \mathbb{R}$, for all $\varepsilon > 0$, there exists arbitrarily large t such that $B_t = D_\varepsilon(z)$, where $D_\varepsilon(z)$ is a disk of radius ε about z (we don't use the usual B notation for ball, because it would be needlessly confusing here).
  - For all $z \in \mathbb{R}^2$, $\mathbb{P}[\exists\, t > 0 \text{ s.t. } B_t = z] = 0$. Thus the range is dense with zero Lebesgue measure.
- For $d = 3$, almost surely we have,
  - Self-intersections in every time interval.
  - Transient: $\lim_{t\to\infty} |B_t| = \infty$. Sort of the opposite of neighborhood recurrence.
- For $d \geq 4$, almost surely we have,
  - For all $s \leq t$, $B_s \neq B_t$.
  - Transient, as in $\mathbb{R}^3$.
- In any dimension, almost surely the Hausdorff dimension of the range of B_t is 2. This is related to 1D Brownian Motion being 1/2-Hölder continuous, which we'll see in the construction next week. We'll also prove something about Hölder continuity on the homework this week.

  Additional intuition: A Lipschitz function preserves Hausdorff dimension from the domain, so a $\frac{1}{2}$-Hölder continuous function can double the Hausdorff dimension of the domain.

We're now going to move from Brownian Motion to its implications on analysis.

III.4. **Brownian Motion and Harmonic Functions**

Let $U \subseteq \mathbb{R}^d$ be open. Recall that $u : U \to \mathbb{R}$ is *harmonic* if $\Delta u = 0$ where $\Delta = \sum_{j=1}^d \partial_{x_j}^2$. We're now interested in the Dirichlet Problem.

Let $\phi : \partial U \to \mathbb{R}$ be continuous, can we find $u : \overline{U} \to \mathbb{R}$ continuous with

$$\begin{cases} \Delta u = 0 & \text{in U} \\ u = \phi & \text{on } \partial U. \end{cases}$$

We'll see that this is related to Brownian motion. Namely, let B^z for $z \in U$ be a d-dimensional Brownian Motion with $B_0^z = z$, and set

$$\tau = \min\{t \geq 0 \mid B_t^z \notin U\}.$$

Then the solution to the Dirichlet problem is actually

$$u(z) = \mathbb{E}[\varphi(B_\tau^z)].$$

Plan for the Rest of Class

- Today: Brownian Motion and Harmonic Functions
- Friday: Applications of Brownian Motion to Complex Analysis and Vice Versa
- Next Week: Construction of Brownian Motion, lectures by Sayan Das, since Professor Gwynne is away at a conference.

Last time, we talked about d-dimensional Brownian Motion. Today we're going to continue exploring the Dirichlet problem, using Brownian Motion. For notation, for $z \in \mathbb{R}^d$ we'll write $\mathbb{P}_z$ for the probability measure (or distribution) on the space of functions, where B starts at z instead of 0. This is the distribution of $B^0 + z$.

Another way to think about it is that

$$\mathbb{E}_z[F(B)] = \mathbb{E}[F(B^0 + z)],$$

where F is any functional on paths.

Remark III.4.1

If τ is a stopping time, then the conditional distribution of $s \mapsto B_{s+\tau}$ given $B\big|_{[0,\tau]}$ is just $\mathbb{P}_{B_\tau}$. This is an equivalent formulation of the strong Markov property.

Proposition III.4.1

Let $U \subseteq \mathbb{R}^d$ be open, let B be a d-dimensional Brownian Motion, and let $\tau = \min\{t \mid B_t \notin U\}$. Assume that for all $z \in U$, we have that $\mathbb{P}_z[\tau < \infty] = 1$. Note this is automatic if U is bounded.

Let $\varphi : U \to \mathbb{R}$ be a bounded, measurable function. For all $z \in U$, set

$$u(z) = \mathbb{E}_z[\varphi(B_\tau)].$$

Then u is harmonic.

Proof. Recall from harmonic / complex analysis, that a function is harmonic if and only if it satisfies the mean value property. Let $r > 0, z \in U$ such that $\overline{D_r(z)} \subseteq U$ where $D_r(z)$ is the disk of radius r about z. We want to show that

$$\int_{\partial D_r(z)} u(w)\,\mathrm{d}\sigma(w) = u(z),$$

where σ is the uniform measure on $\partial D_r(z)$. Let

$$T = \min\{t \geq 0 \mid B_t \in \partial D_r(z)\}.$$

In particular, T is a stopping time, and $T \leq \tau$. We know by basic properties of conditional expectation that

$$u(z) = \mathbb{E}_z[\varphi(B_t)] = \mathbb{E}_z[\mathbb{E}_z[\varphi(B_\tau) \mid B\big|_{[0,T]}]]$$

The strong markov property implies that the conditional distribution of $s \mapsto B_{s+T}$ given $B\big|_{[0,T]}$ is just $\mathbb{P}_{B_T}$. Furthermore,

$$\tau - T = \min\{s \geq 0 \mid B_{s+T} \notin U\}.$$

By combining these two facts we see that

$$\mathbb{E}_z[\varphi(B_\tau) \mid B\big|_{[0,T]}] = \mathbb{E}_{B_T}[\varphi(\widetilde{B}_\tau)] = u(B_T),$$

where $\widetilde{B}_\tau$ is being used to distinguish that it starts at B_T. To take the expectation of this, we need to know something about the distribution of B_T. By rotational invariance, B_T has the uniform distribution σ on $\partial D_r(z)$. Hence

$$u(z) = \mathbb{E}_z[\mathbb{E}_z[\varphi(B_\tau) \mid B\big|_{[0,T]}]] = \mathbb{E}_z[u(B_T)] = \int_{\partial D_r(z)} u(w)\,\mathrm{d}\sigma(w).$$

Perfect! This finishes the proof!

The rest of the solution of the Dirichlet problem is a bit more technical. We need to show that if $\varphi : \partial U \to \mathbb{R}$ is continuous, then the constructed function u extends continuously to the boundary

Essentially, we want to show that if z is close to the boundary, then our Brownian motion exits U on the boundary close to z. This property is not actually true for every domain. Consider $D_1(0) \setminus \{0\}$. Then if we start the Brownian motion close to the origin, it will never hit the origin... oops. So we need some reasonable conditions on the domain.

We'll restrict our life to $d = 2$, since we can impose topological conditions on the domain in this case. In higher dimensions we need geometric conditions on the domain, not just topological conditions. For $d = 2$, it will be enough for U to be simply connected. We'll do this in several lemmas. We won't give detailed proofs here for all of these lemmas, but they can be found in Professor Gwynne's notes.

Lemma III.4.2

Let B be a 2D Brownian Motion and let $z \in \partial D_{1/2}(0)$. If $B_0 = z$, then with positive probability, B makes a loop around $D_1(0) \setminus D_{1/2}(0)$ before exiting $D_1(0)$.

Proof by Picture. Construct the loop piece by piece, using that Brownian motion is uniform on the boundary of a disk. Make the loop almost close up, then make it cross itself. You can do each piece with positive probability.

Lemma III.4.3

Let B be a Brownian Motion with $B_0 = 0$. Then there exists $C, \alpha > 0$ such that for all $R > r > 0$ we have that

$$\mathbb{P}[B \text{ disconnects } \partial D_R(0) \text{ from } \partial D_r(0) \text{ before exiting } D_R(0)] \geq 1 - C\left(\frac{r}{R}\right)^{\alpha}.$$

Proof by Picture. Take the big annulus, and divide it up into dyadic sub-annuli. All the sub-annuli have aspect ratio ≈ 2. Namely $r, 2r, 4r, \ldots$ until we reach R, the number of annuli is $\approx \log_2(R/r)$. There's probability at least $p > 0$ that it will make a loop between $r, 2r$, by the previous lemma. The probability it fails to make a loop there is $\leq 1 - p$. By scale invariance, it has probability at least p to succeed and make a loop. By strong markov, the chance for each annulus is independent. Thus we have $\log_2(R/r)$ trials, and the probability we fail to make a loop is $1 - p$. Thus the probability that we fail to make a loop

$$\mathbb{P}[\text{fail to make a loop in each annulus}] \leq (1-p)^{\lfloor \log_2(R/r) \rfloor} \leq C\left(\frac{r}{R}\right)^{\alpha},$$

where $1 - p = 2^{-\log_2 \frac{1}{1-p}}$ and we just simplify to get $\alpha = \log_2 \frac{1}{1-p} > 0$. The C comes from the floor function.

Theorem III.4.4

Let $U \subseteq \mathbb{R}^2$ be open, simply connected, $U \neq \mathbb{R}^2$. Let $\tau = \min\{t > 0 \mid B_t \notin U\}$. Now let $\varphi : \partial U \to \mathbb{R}$ be bounded, continuous. Let

$$u(z) = \begin{cases} \mathbb{E}_z[\varphi(B_\tau)] & z \in U \\ \varphi(z) & z \in \partial U, \end{cases}$$

then $\Delta u = 0$ on U, and u is continuous on $\overline{U}$, solving the Dirichlet problem in this case.

Proof by Picture. We've already shown that it's harmonic. We need to show that it's continuous on $\overline{U}$. It essentially suffices to show that if we start Brownian motion at z close to ∂U, it's very likely for z to exit U close to z.

Let $\operatorname{dist}(z, \partial U) = \delta < 1$. Now cosnder $D_\delta(z) \subseteq D_{\delta^{1/2}}(z)$. Then

$$\mathbb{P}_z[B \text{ disconnects } \partial D_\delta(z) \text{ from } \partial D_{\delta^{1/2}}(z) \text{ before exiting } D_{\delta^{1/2}}(z)] \geq 1 - C\delta^{\alpha/2}.$$

In this case, the Brownian Motion B has to exit U before exiting $D_{\delta^{1/2}}(z)$, by basic topological considerations. This is where we use that U is simply connected. Essentially, we know that $\mathbb{R}^2 \setminus U$ is connected, and if we created a loop without exiting U, this would provide a disconnection of $\mathbb{R}^2 \setminus U$ (notably this is also where we use $U \neq \mathbb{R}^2$). Hence

$$\mathbb{P}_z[\operatorname{dist}(z, B_\tau) \leq \delta^{1/2}] \geq 1 - C\delta^{\alpha/2}.$$

Once we have this statement, the continuity of φ and basic expectation estimates will give continuity of u.

Last Time: We used Brownian Motion to solve the Dirichlet problem on simply connected domains $U \subsetneq \mathbb{R}^2$. This time we're going to show uniqueness of our solution. Recall the construction, let $\mathcal{B}$ be a 2D Brownian motion

$$\tau = \min\{t \geq 0 \mid \mathcal{B}_t \notin U\},$$

$\varphi : \partial U \to \mathbb{R}$ bounded and continuous. We let

$$u(z) = \begin{cases} \mathbb{E}_z[\varphi(\mathcal{B}_\tau)] & z \in U \\ \varphi(z) & z \in \partial U. \end{cases}$$

Then u is continuous on $\overline{U}$ and $\Delta u = 0$ on U.

Proposition III.4.5

Suppose $v_1, v_2 : \overline{U} \to \mathbb{R}$ is continuous such that $v_i\big|_{\partial U} = \varphi$ and $\Delta v_i = 0$. Assume U is bounded/simply connected, then $v_1(z) = v_2(z)$ for all $z \in U$.

Proof. $v_1 - v_2$ is harmonic on U, continuous on $\overline{U}$, and zero on ∂U. By the maximum principle, $v_1 - v_2$ achieves its maximum / minimum on ∂U, and hence $v_1 - v_2 = 0$.

Remark III.4.2

The same construction of a solution u to the Dirichlet problem works if $\mathbb{R}^2 \setminus U$ has finitely many components, all non-singleton.

Remark III.4.3

For $d \geq 3$ Brownian Motion cannot disconnect $\partial D_R(0)$ from $\partial D_r(0)$. Thus the proof that we described cannot work in higher dimensions. The hypotheses on the domain are more complicated. A sufficient condition is that ∂U is a C^1-submanifold of $\mathbb{R}^d$ of dimension $d - 1$. But C^0 is not sufficient.

A good way to see our construction does not work is to run an increasingly thin spike through a sphere.

Now identify $\mathbb{C}$ with $\mathbb{R}^2$, $x + iy \sim (x, y)$.

Definition III.4.1

Let $U \subseteq \mathbb{C}$ be open, and let $\mathcal{B}$ be a Brownian motion in $\mathbb{C}$, $\mathcal{B} = \mathcal{B}^1 + i\mathcal{B}^2$. Define τ by

$$\tau = \min\{t \geq 0 \mid \mathcal{B}_t \notin U\}.$$

Also, assume $\mathbb{P}_z[\tau < \infty] = 1$ for all $z \in U$. We define the harmonic measure on ∂U viewed from z as

$$\operatorname{hm}_U^z(A) = \mathbb{P}[\mathcal{B}_\tau \in A]$$

for all $A \subseteq \partial U$ Borel. In other words, this is the distribution of $\mathcal{B}_\tau$.

Remark III.4.4

In this language, we have

$$u(z) = \mathbb{E}_z[\varphi(\mathcal{B}_\tau)] = \int_{\partial U} \varphi(x) \, \mathrm{d}\operatorname{hm}_U^z(x).$$

Remark III.4.5

The most elementary harmonic measure: If $U = D_r(0)$, $z = 0$, the $\operatorname{hm}_{D_r(0)}^0$ is the uniform measure on $\partial D_r(0)$.

Theorem III.4.6

Let $U, V \subseteq \mathbb{C}$ be open and simply connected. Assume also that U is bounded. Let $f : U \to V$ be a biholomorphism, and assume also that f extends continuously to $\overline{U} \to \overline{V}$. Then we have that

$$\operatorname{hm}_V^z(A) = \operatorname{hm}_U^{f^{-1}(z)}(f^{-1}(A))$$

for all $z \in V$ and $A \subseteq \partial V$ Borel.

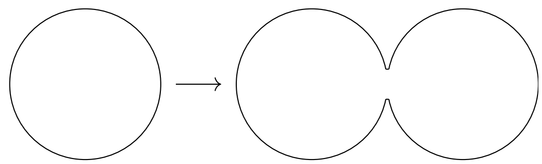
Proof. Let $\varphi : \partial V \to \mathbb{R}$ be any bounded and continuous function. Let $v : \overline{V} \to \mathbb{R}$ solve the Dirichlet problem for φ. We see that $v \circ f$ is continuous on $\overline{U}$. We know that $v \circ f$ is harmonic since f is holomorphic, so it solves the Dirichlet problem for $\varphi \circ f$ on ∂U. By uniqueness, this implies that, for $z \in U$,

$$\int_{\partial V} \varphi \, \mathrm{d}\operatorname{hm}_V^z = v(z) = v(f(f^{-1}(z))) = \int_{\partial U} (\varphi \circ f) \, \mathrm{d}\operatorname{hm}_U^{f^{-1}(z)}.$$

This shows the required claim, by approximating an indicator $\mathbb{1}_A$ by a bounded continuous function φ.

Example III.4.1

For $\varepsilon > 0$ let $U_\varepsilon = D_1(0) \cup D_\varepsilon(1) \cup D_1(2)$. In other words something like a bowtie ☺



Let $f_\varepsilon : D_1(0) \to U_\varepsilon$ be a biholomorphism with $f_\varepsilon(0) = 0$. What does $f_\varepsilon^{-1}(\partial U_{\mathbb{E}} \cap \partial D_1(2))$ look like? If $\mathcal{B}$ is a Brownian Motion started from B is unlikely to enter $D_\varepsilon(1)$ before leaving $D_1(0)$. Hence

$$\mathrm{hm}_{U_\varepsilon}^0(\partial U_\varepsilon \cap \partial D_1(2)) = O(\varepsilon).$$

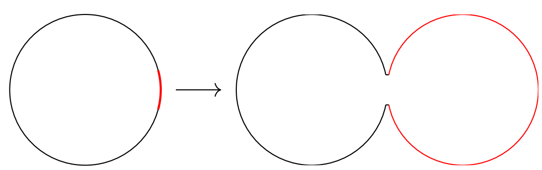
Hence

$$\mathrm{hm}_{D_1(0)}^0(f_\varepsilon^{-1}(\partial U_\varepsilon \cap \partial D_1(2))) = O(\varepsilon).$$

But from our knowledge of the disk we know that

$$\mathrm{Leb}(f_\varepsilon^{-1}(\partial U_\varepsilon \cap \partial D_1(2))).$$

We know also that this preimage is an interval, so it looks absolutely tiny in the disk. In other words, we can draw things like



Example III.4.2

Let $\varepsilon > 0$, $U_\varepsilon = D_1(0) \setminus [\varepsilon, 1]$. Let $f_\varepsilon : D_1(0) \to U_\varepsilon$ be a biholomorphism with $f_\varepsilon(0) = 0$. But what does $f_\varepsilon^{-1}([\varepsilon, 1])$ look like? Well it looks like almost all of the disk!

Brownian motion starting at 0 in U_ε is unlikely to exit $D_1(0)$ before hitting $[\varepsilon, 1]$. Hence

$$\lim_{\varepsilon \to 0} \mathrm{hm}_{U_\varepsilon}^0([\varepsilon, 1]) = 1$$
$$\lim_{\varepsilon \to 0} \mathrm{Leb}(f_\varepsilon^{-1}([\varepsilon, 1])) = 1.$$

Last time: We defined and talked about the harmonic measure. This was defined for $U \subseteq \mathbb{C}$ open, $\mathcal{B}$ a 2D Brownian motion, with

$$\tau := \min\{t \geq 0 \mid \mathcal{B}_t \notin U\},$$

and assuming $\mathbb{P}_z[\tau < \infty] = 1$. The harmonic measure on ∂U was then

$$\mathrm{hm}_U^z(A) = \mathbb{P}_z[\mathcal{B}_\tau \in A]$$

for $A \subseteq \partial U$ Borel.

Theorem III.4.7

For $U, V \subseteq \mathbb{C}$ open, simply connected, $U, V \neq \mathbb{C}$. A biholomorphism $f : U \to V$ extending continuously to $\overline{U} \to \overline{V}$ has

$$\text{hm}_V^z(A) = \text{hm}_U^{f^{-1}(z)}(f^{-1}(A)).$$

We were able to use this last time to understand the geometry of such biholomorphisms supplied by the Riemann Mapping Theorem

Example III.4.3

We'll use the harmonic measure to construct an explicit Riemannian mapping $U \to \mathbb{D}$ for $U \subsetneq \mathbb{C}$ a simply connected open domain. For simplicity we'll do this for ∂U a simple C^0 loop. Then fix $z_0 \in U, x_0 \in \partial U$. We'll produce an explicit biholomorphism $f : \overline{U} \to \overline{\mathbb{D}}$ with $f(z_0) = 0, f(x_0) = 1$.

We start by defining the action on $y \in \partial U$,

$$\begin{aligned} p(y) &= \text{hm}_U^{z_0}(\text{counterclockwise arc of } \partial U \text{ from } x_0 \text{ to } y) \\ g &: \partial U \to \partial \mathbb{D} \\ g(y) &= e^{2\pi i p(y)}. \end{aligned}$$

This is essentially *forced* on us by harmonic measure being a conformal invariant. We can now solve the Dirichlet problem with this function. Let $f : U \to \mathbb{C}$ solve

$$f = g \text{ on } \partial U \qquad \Delta f = 0 \text{ on } U,$$

where we can just solve the Dirichlet problem on the real and imaginary parts separately. To be more explicit

$$f(z) = \mathbb{E}_z[g(\mathcal{B}_\tau)] \qquad \tau \coloneqq \min\{t \geq 0 \mid \mathcal{B}_\tau \notin U\}.$$

In general, it's not clear that f will be holomorphic (i.e. that its real and imaginary parts are harmonic conjugates). We have only solved two a priori unrelated Dirichlet problems for some boundary data. Let $\widetilde{f} : U \to B_1(0)$ be the actual biholomorphism with $\widetilde{f}(z_0) = 0$ and $\widetilde{f}(x_0) = 1$ from the Riemann mapping theorem.

By the conformal invariance of harmonic measure, we see that $\widetilde{f}, f$ agree on the boundary ∂U. By uniqueness of the solution to the Dirichlet problem, we then see that $f = \widetilde{f}$.

III.5. Construction of Brownian Motion

Recall that a 1D Brownian Motion is a random continuous $\mathcal{B} : [0, \infty) \to \mathbb{R}$ such that $\mathcal{B}_0 = 0$ and

(i) For all $s < t$, $\mathcal{B}_t - \mathcal{B}_s \sim N(0, t - s)$.

(ii) For all $t_0 < \cdots < t_N$ we have $\{\mathcal{B}_{t_j} - \mathcal{B}_{t_{j-1}}\}_{j=1,\ldots,n}$ are independent.

Theorem III.5.1

Brownian Motion exists.

Notice it suffices to construct $\mathcal{B}\big|_{[0,1]}$, since we can just concatenate countably many iid $\mathcal{B}\big|_{[0,1]}$. Here's the two main steps which we'll follow.

(i) Define $\mathcal{B}$ for dyadic times.

(ii) Show that $\mathcal{B}$ extends to a continuous function on $[0,1]$, by proving it is uniformly continuous on the dyadic rationals.

For each $n \in \mathbb{N}$ we let

$$D_n = \left\{ \frac{j}{2^n} \mid j = 1, \dots, 2^n \right\},$$

and we set $D = \bigcup_{n=1}^{\infty} D_n \subseteq [0,1]$. We're also going to let $\{Z_t \mid t \in D\}$ be iid $N(0,1)$ random variables (countably many!). We'll define $\{B_t \mid t \in D\}$ inductively. Define $\mathcal{B}_0 = 0, \mathcal{B}_1 = Z_1$. Inductively, now assume $\mathcal{B}_t$ is defined for all $t \in D_{n-1}$. Note that if $t \in D_n \setminus D_{n-1}$, then $t \pm 2^{-n} \in D_{n-1}$. We set

$$\mathcal{B}_t = \frac{1}{2}\left(\mathcal{B}_{t-2^{-n}} + \mathcal{B}_{t+2^{-n}}\right) + 2^{-(n+1)/2} \cdot Z_t.$$

Great! Now we need to know some things...

Lemma III.5.2

For all $n \in \mathbb{N}$ we have $\mathcal{B}_{j/2^n} - \mathcal{B}_{(j-1)/2^n}$ for $\frac{j}{2^n} \in D_n$ are iid $N(0, 2^{-n})$.

Proof. To prove this, we'll induct on n. For $n = 0$, this is immediate, $\mathcal{B}_1 - \mathcal{B}_0 = Z_1 \sim N(0,1)$. Now let $n \geq 1$ and assume the lemma for $n-1$. If $\frac{j}{2^n} \in D_n \setminus D_{n-1}$, then $(j-1)/2^n \in D_{n-1}$. Hence

$$\mathcal{B}_{j/2^n} - \mathcal{B}_{(j-1)/2^n} = 2^{-(n+1)/2} Z_{j/2^n} + \frac{1}{2}\left(\mathcal{B}_{(j+1)/2^n} - \mathcal{B}_{(j-1)/2^n}\right).$$

Similarly if $j/2^n \in D_{n-1}$ then we instead have

$$\mathcal{B}_{j/2^n} - \mathcal{B}_{(j-1)/2^n} = 2^{-(n+1)/2} Z_{(j-1)/2^n} + \frac{1}{2}(\mathcal{B}_{j/2^n} - \mathcal{B}_{(j-2)/2^n}).$$

These are normally distributed and independent by induction and the fact that the Z_t are independently distributed and normal.

To be more precise, $(\mathcal{B}_{j/2^n} - \mathcal{B}_{(j-1)/2^n})_{j=1,\dots,2^n}$ is the image under a linear map of

$$\left\{ (\mathcal{B}_{i/2^{n-1}} - \mathcal{B}_{(i-1)/2^{n-1}})_{i=1,\dots,2^{n-1}}, (2^{-(n-1)/2} Z_t)_{t \in D_n \setminus D_{n-1}} \right\}.$$

All of these are independent, the first is iid with distribution $N(0, 2^{-(n-1)})$, as are the second iid with distribution $N(0, 2^{-(n-1)})$. Furthermore the left is independent from the right, since $\mathcal{B}_{i/2^{n-1}}$ is a function of the Z_t for $t \in D_{n-1}$. In particular, each $\mathcal{B}_{j/2^n} - \mathcal{B}_{(j-1)/2^n}$ is the sum of two independent $N(0, 2^{-(n-1)})$ random variables times $1/2$.

It is not difficult to see then that $\mathcal{B}_{j/2^n} - \mathcal{B}_{(j-1)/2^n}$ is distributed as $N(0, 2^{-n})$. One can compute that

$$\mathbb{E}\left[\left(\mathcal{B}_{i/2^n} - \mathcal{B}_{(i-1)/2^n}\right)\left(\mathcal{B}_{j/2^n} - \mathcal{B}_{(j-1)/2^n}\right)\right] = 0,$$

if $i \neq j$. On Problem Set 5, we showed that a linear function $M \cdot A$ of independent identically distributed Gaussian random variables $A_1, \dots, A_m$ has independent components if and only if $\mathbb{E}[MA_i \cdot MA_j] = 0$ for $i \neq j$. Notably, this is essentially checking that M is a scalar multiple of orthogonal matrix.

Now lets go further

Lemma III.5.3

We have the properties of Brownian Motion on the dyadics

(i) For all $s, t \in D$ with $s < t$ we have $\mathcal{B}_t - \mathcal{B}_s \sim N(0, t-s)$.

(ii) For all $t_0 < \cdots < t_N$ in D we have $B_{t_j} - B_{t_{j-1}}$ are independent for $j = 1, \ldots, N$.

Proof. Let's go!

(i) Choose n such that $s, t \in D_n$. Then

$$\mathcal{B}_t - \mathcal{B}_s = \sum_{j, j/2^n \in (s,t]} \mathcal{B}_{j/2^n} - \mathcal{B}_{(j-1)/2^n}.$$

These are iid with distribution $N(0, 2^{-n})$ and there are $2^n(t-s)$ of them. Hence $\mathcal{B}_t - \mathcal{B}_s \sim N(0, t-s)$.

(ii) Choose n such that $t_0, \ldots, t_N \in D_n$ and proceed similarly.

The second step, uniform continuity, will come Wednesday!

References

[15] *275A, Notes 0: Foundations of Probability Theory*. Sept. 2015. (Visited on 04/15/2024).

[DD19] Richard Durrett and Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, Apr. 2019. ISBN: 978-1-108-47368-2.

[MP10] P. Mörters and Y. Peres. *Brownian Motion*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010. ISBN: 9781139486576. URL: https://books.google.com/books?id=e-TbA-dSrzYC.