Advanced Probability

Jan Olucha Fuentes

May 22, 2025

Dear Reader,

These are a set of lecture notes written for the course Advanced Probability taught at the University of Cambridge in the academic year 2024-2025. The reader is assumed to have a thorough grounding in basic probability theory and a strong will to live. For a refresher on the former, consult the first part of the wonderful book Probability with Martingales, by David Williams [Wil14].

Yours falsely,

JOF.

Contents

0 Review of Measure Theory 5

- 0.1 Basic concepts . . . 5
- 0.2 Integration . . . 5
- 0.3 Product measure spaces . . . 8

1 Conditional Expectation 10

- 1.1 Motivation . . . 10
- 1.2 Agreement with traditional usage . . . 20

2 Discrete-time Martingales 22

- 2.1 Adapted Processes . . . 23
- 2.2 Martingale, superMartingale, subMartingale . . . 23
- 2.3 Stopping times . . . 25
  - 2.3.1 Optional Stopping . . . 28
  - 2.3.2 Gambler's Ruin . . . 31
- 2.4 The Martingale Convergence Theorem . . . 33
- 2.5 $\mathscr{L}^p$ convergence of Martingales . . . 36
- 2.6 UI Martingales . . . 40
- 2.7 Backwards Martingales . . . 44
- 2.8 Application of Martingales . . . 46

3 Continuous-time processes 52

- 3.1 Martingale regularisation . . . 57
- 3.2 Doob's inequalities and convergence in continuous time . . . 62
- 3.3 Kolmogorov's continuity criterion . . . 63

4 Weak convergence 66

- 4.1 Definitions . . . 66
- 4.2 Characteristic functions . . . 74
- 4.3 Large Deviation Theory . . . 79
- 4.4 Cramer's Theorem . . . 79

5 Brownian motion 85

- 5.1 Wiener's Theorem . . . 85
- 5.2 Invariance Properties . . . 88
- 5.3 Strong Markov Property . . . 95
- 5.4 The Reflection Principle . . . 98
- 5.5 Martingales and Brownian Motion . . . 102

5.6 Recurrence and Transience 106
5.7 Dirichlet Problem 111
5.8 Donsker's Invariance Principle 117
5.9 Zeroes of Brownian Motion 123

6 Poisson Random Measures **124**

7 Homework Problems **129**

8 Appendix **132**
8.1 Measure Theory 132
8.1.1 π-systems, extensions and uniqueness 132
8.2 Conditional Expectation 134
8.3 Independence 135

Notation

☕: denotes importance of a concept, with three being maximum.

Ω: a sample space.

$\mathscr{F}$: a σ-algebra.

$\mathbf{P}$: a probability measure.

$\mathrm{m}\mathscr{F}$: the set of $\mathscr{F}$-measurable functions.

$\mathrm{b}\mathscr{F}$: the subset of $\mathrm{m}\mathscr{F}$ of bounded functions.

$\mathbf{E}(X)$: the integral of some $X \in \mathrm{m}\mathscr{F}$ with respect to the measure $\mathbf{P}$. Also denoted by

$$\int_\Omega X \, \mathrm{d}\mathbf{P} \quad \text{or} \quad \int_\Omega X(\omega)\mathbf{P}(d\omega) \quad \text{or} \quad \int_\Omega X(\omega)\,\mathrm{d}\mathbf{P}(\omega)$$

$\mathbf{1}_A$: the indicator function of a set A.

$\sigma(\mathscr{A})$: the σ-algebra generated by a family of sets $\mathscr{A}$.

$n \wedge m$: $\min(n, m)$.

$n \vee m$: $\max(n, m)$.

$C_b(M)$: continuous and bounded functions $M \to \mathbf{R}$ where (M, d) is some metric space.

0 Review of Measure Theory

0.1 Basic concepts

Definition 0.1 Let Ω be a non-empty set. A family of subsets of Ω, denoted by $\mathscr{F}$, is called a σ-algebra if it contains Ω, and it is closed under complements and countable unions. We refer to a pair $(\Omega, \mathscr{F})$ as a measurable space. We refer to elements of $\mathscr{F}$ as measurable sets, or events in the context of probabilities.

Theorem 0.2 Given a subset $S \subseteq \Omega$, there exists a smallest σ-algebra containing S, denoted $\sigma(S)$.

Example 0.3 If (Ω, τ) is a topological space, we call $\sigma(\tau)$ the Borel σ-algebra.

Definition 0.4 A measure μ is a non-negative function $\mu : \mathscr{F} \to [0, \infty]$ such that

1. $\mu(\varnothing) = 0$
2. For a disjoint sequence (A_n) of measurable sets, the measure of the union is the sum of the measures:
$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$$

A triple $(\Omega, \mathscr{F}, \mu)$ is referred to as a measure space. If $\mu(\Omega) = 1$, we refer to this as a probability space and we write $\mathbf{P}$ instead of μ.

Definition 0.5 Given two measurable spaces $(A, \mathscr{A})$, $(B, \mathscr{B})$, a function $f : A \to B$ is said to be measurable if given any $S \in \mathscr{B}$, we have $f^{-1}(S) \in \mathscr{A}$. In the context of probabilities, a measurable function is called a random variable.

0.2 Integration

Definition 0.6 A measurable function $f : \Omega \to \mathbf{R}$ is called simple if it takes finitely-many values. That is to say,
$$f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$$

Where (A_n) are without loss of generality disjoint.

Remark 0.7 Any non-negative measurable function $f:\Omega\to\mathbf{R}$ can be uniformly and monotonically approximated by simple functions, for example:

$$f_n = 2^{-n}\lfloor 2^n f\rfloor \wedge n$$

Definition 0.8 The integral of a simple function f with respect to a measure μ is given by

$$\int_\Omega f\,d\mu := \sum_{i=1}^n a_i\mu(A_i)$$

We also write the integral of f as $\mu(f)$ when convenient.

Remark 0.9 The integral can be shown to be independent of choice of the representation of f as a simple function

Definition 0.10 The integral of a non-negative measurable function f, can be defined in view of Remark 0.7 as

$$\int_\Omega f\,d\mu = \sup\{\mu(g): g\le f \text{ simple}\}$$

We say f is integrable if $\mu(f)<\infty$. We extend this to an integral of general functions in $\mathrm{m}\mathscr{F}$ by writing $f=f^+-f^-$ where $f^+=f\vee 0$ and $f^-=-(f\wedge 0)$ and setting

$$\int_\Omega f\,d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu$$

A general measurable function is said to be integrable if both f^+ and f^- are integrable.

Proposition 0.11 (Properties of the Lebesgue integral) Let $(\Omega,\mathscr{F},\mu)$ be a measure space.

1. The integral is a linear operator $\mathrm{m}\mathscr{F}\to\mathbf{R}$.
2. *Monotone Convergence Theorem:* If (f_n) is a sequence in $\mathrm{m}\mathscr{F}^+$ and $f_n\uparrow f$, then $\mu(f_n)\to\mu(f)$
3. Fatou's Lemma: with (f_n) as before: $\mu(\liminf_n f_n)\le\liminf_n \mu(f_n)$

4. *Dominated Convergence Theorem:* if (f_n) is dominated by an integrable $g \in \mathrm{m}\mathscr{F}^+$, then $\mu(f_n) \to \mu(f)$

0.3 Product measure spaces

Recall the following definition:

Definition 0.12 (σ-algebra generated by a function) Let Ω be a set, and $(A, \mathscr{A})$ be a measurable space. Let $f : \Omega \to A$ be any function. Then the σ-algebra generated by f, denoted $\sigma(f)$, is the smallest σ-algebra, $\mathscr{F}$ that makes f an $\mathscr{F} - \mathscr{A}$ measurable function. That is to say, $\sigma(f)$ is the σ-algebra generated by $f^{-1}(S)$ for all $S \in \mathscr{A}$.

Definition 0.13 (Product σ-algebra) Let $(A, \mathscr{A})$ and $(B, \mathscr{B})$ be two measurable spaces. Consider the projection maps $\pi_A : A \times B \to A$ and $\pi_B : A \times B \to B$ defined in the obvious ways. We define the product sigma algebra, denoted by $\mathscr{A} \otimes \mathscr{B}$ (nothing to do with the tensor product) is the σ-algebra generated by the projection maps, i.e:

$$\mathscr{A} \otimes \mathscr{B} = \sigma(\pi_A, \pi_B)$$

Remark 0.14 Since given any measurable set $S \in \mathscr{A}$, we have that $\pi_A^{-1}(S) = S \times B$, and something identical holds for π_B, we have that $\mathscr{A} \otimes \mathscr{B}$ contains all sets of the form $S_1 \times S_2$, where $S_1 \in \mathscr{A}$ and $S_2 \in \mathscr{B}$. For the case of a countable product of σ-algebras, it can be shown that Cartesian products like that of $S_1 \times S_2$ actually generate the product of algebras, so we may use either characterisation, depending on what's more useful. For uncountable products however, we use the definition given above.

The goal is to construct a measure on $\mathscr{A} \otimes \mathscr{B}$, and we do so by explicit construction.

Lemma 0.15 Let $(A, \mathscr{A}, \mu_A)$ and $(B, \mathscr{B}, \mu_B)$ be measure spaces. Let $\mathscr{E} = \mathscr{A} \otimes \mathscr{B}$ and $f = f(x_1, x_2) \in m\mathscr{E}^+$. Then the function

$$x_1 \mapsto \int_B f(x_1, x_2) \mu_B(dx_2)$$

is $\mathscr{A}$-measurable.

This Lemma ensures that the following result makes sense.

Theorem 0.16 Let $(A, \mathscr{A}, \mu_A)$ and $(B, \mathscr{B}, \mu_B)$ be measure spaces. Let $\mathscr{E} = \mathscr{A} \otimes \mathscr{B}$. Then there exists a unique measure μ which is sometimes denoted as $\mu_A \otimes \mu_B$ on $\mathscr{A} \otimes \mathscr{B}$ such that

$$\mu(S_1 \times S_2) = \mu_A(S_1)\mu_B(S_2)$$

Proof. The measure is constructed as

$$\mu(S) = \int_A \left(\int_B \mathbf{1}_S(x_1, x_2) \mu_B(dx_2) \right) \mu_A(dx_1) \quad S \in \mathscr{A} \otimes \mathscr{B}$$

♡

Theorem 0.17 (Fubini's / Tonelli's Theorem) Let $f \in m(\mathscr{A} \otimes \mathscr{B})^+$. Then

$$\int_{A\times B} f\, d(\mu_A \otimes \mu_B) = \int_A \left(\int_B f\, d\mu_B \right) d\mu_A = \int_B \left(\int_A f\, d\mu_A \right) d\mu_B$$

1 Conditional Expectation

1.1 Motivation

Suppose $(\Omega, \mathscr{F}, \mathbf{P})$ is a probability space, and $X \in \mathrm{m}\mathscr{F}$ is a random variable. Recall from elementary probability that given two events $A, B \in \mathscr{F}$, one may define the *conditional probability of* A *given* B as follows:

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

From this, one may define the **conditional expectation of X given an event B** by:

$$\mathbf{E}[X \mid B] = \frac{\mathbf{E}[X \mathbf{1}_B]}{\mathbf{P}(B)}$$

With this recap done, let us make the following thought experiment: suppose that no information about X is known. Intuitively, our best guess for the value of X would be $\mathbf{E}[X]$. Now suppose that *some information* is known to us, in the form of a sequence of disjoint events (G_n) whose union is the entirety of Ω, that is to say, for a given random outcome ω of our sample space, we do not know what $X(\omega)$ is, but we may know in which G_n ω *falls into*, or in other words, we may know which event has occurred. The finer this sigma algebra, the better information we'll have. Then, if we know $\omega \in G_k$, our best guess for X will be $\mathbf{E}[X \mid G_k]$. This concept is central to the rest of this manuscript.

Definition 1.1 Let (G_n) be a sequence of disjoint events whose union is Ω. Define $\mathscr{G} = \sigma(G_n : n \geq 0)$. Then the **conditional expectation of** X given $\mathscr{G}$, $\mathbf{E}[X \mid \mathscr{G}]$ is defined by

$$\mathbf{E}[X \mid G] = \sum_{i=1}^{\infty} \mathbf{E}[X \mid G_i] \mathbf{1}_{G_i}$$

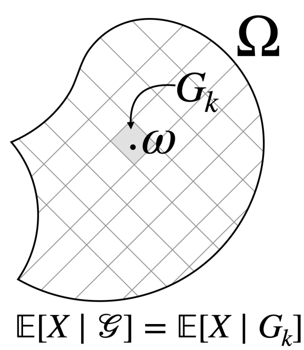


Figure 1: A visual picture of conditional expectation

We now see a first result about conditional expectation:

Lemma 1.2 (Discrete Conditional Expectation, ♩) Given $X \in L^1(\mathscr{F})$, and $\mathscr{G}$ as above, $\mathbf{E}[X \mid \mathscr{G}]$ satisfies the following:

1. $\mathbf{E}[X \mid \mathscr{G}]$ is $\mathscr{G}$-measurable.
2. $\mathbf{E}[X \mid \mathscr{G}] \in L^1(\mathscr{G})$, and for any $A \in \mathscr{G}$:

$$\mathbf{E}[\mathbf{E}[X \mid G]\,\mathbf{1}_A] = \mathbf{E}[X\,\mathbf{1}_A]$$

Main idea: The first statement is obvious, showing integrability is a straightforward computation involving the convergence Theorems. For the last part, we use the fact that $\{G_n\}$ is a disjoint cover of Ω, and so a union or intersection of these "atoms" remains some sort of union of said "atoms".

Remark 1.3 An alternative way to phrase the second result is that for any $A \in \mathscr{G}$,

$$\mathbf{E}[\mathbf{E}[X \mid \mathscr{G}]|A] = \mathbf{E}[X \mid A]$$

or also

$$\int_A \mathbf{E}[X \mid \mathscr{G}]\,\mathrm{d}\mathbf{P} = \int_A X\,\mathrm{d}\mathbf{P}$$

An informal interpretation is that since $\mathbf{E}[X \mid \mathscr{G}]$ is a "best guess" for X given the information encoded in $\mathscr{G}$, when we average this "best guess" over sets already contained in $\mathscr{G}$, we obtain the same average as we would by using the original X. In other words, $\mathbf{E}[X \mid \mathscr{G}]$ serves as a surrogate for X that captures the essential features of X in a way that is consistent with the information available in $\mathscr{G}$.

Proof. The first statement is obvious. To show the second, we first note that

$$\begin{aligned}
\mathbf{E}[|\mathbf{E}[X \mid \mathscr{G}]|] &= \mathbf{E}\left[\left|\sum_{i=1}^{\infty} \mathbf{E}[X \mid G_i]\mathbf{1}_{G_i}\right|\right] \\
&\leq \mathbf{E}\left[\sum_{i=1}^{\infty} \mathbf{E}[|X| \mid G_i]\mathbf{1}_{G_i}\right] \\
&= \sum_{i=1}^{\infty} \mathbf{E}(\mathbf{E}[|X| \mid G_i]\mathbf{1}_{G_i}) && \text{(Monotone convergence)} \\
&= \sum_{i=1}^{\infty} \mathbf{E}[|X| \mid G_i]\mathbf{P}(G_i) \\
&= \sum_{i=1}^{\infty} \mathbf{E}[|X|\mathbf{1}_{G_i}] && \text{(Elementary conditional probability)} \\
&= \mathbf{E}\left[\sum_{i=1}^{\infty} |X|\mathbf{1}_{G_i}\right] && \text{(Monotone convergence)} \\
&= \mathbf{E}[|X|] && (G_n) \text{ partitions } \Omega \\
&< \infty
\end{aligned}$$

Showing the preservation of integrals first involves analysing what a typical element of $\mathscr{G}$ looks like. The key is that since (G_n) are disjoint and their union is all of Ω, any countable union, intersection of G_n's or their complements will remain being a countable union of G_n's, i.e: any $A \in \mathscr{G}$ has the shape

$$A = \bigcup_{n \in I} G_n \qquad \text{for some } I \subseteq \mathbf{N}$$

Now we can simply perform calculations, and by repeated use of the Dominated Convergence Theorem, the proof follows. Indeed:

$$\begin{aligned}
\mathbf{E}[\mathbf{E}[X \mid \mathscr{G}]\mathbf{1}_A] &= \int_\Omega \sum_{i=1}^{\infty} \mathbf{E}[X \mid G_i]\mathbf{1}_{G_i}\mathbf{1}_A \, \mathrm{d}\mathbf{P} \\
&= \int_\Omega \sum_{i \in I} \mathbf{E}[X \mid G_i]\mathbf{1}_{G_i} \, \mathrm{d}\mathbf{P} && \text{(structure of } A) \\
&= \sum_{i \in I} \int_\Omega \mathbf{E}[X \mid G_i]\mathbf{1}_{G_i} \, \mathrm{d}\mathbf{P} && \text{(Dominated Convergence Theorem)} \\
&= \sum_{i \in I} \mathbf{E}[X\mathbf{1}_{G_i}] \\
&= \mathbf{E}[X\mathbf{1}_A] && \text{(Dominated Convergence Theorem)}
\end{aligned}$$

♡

The beauty of these two conditions is that they in fact fully characterise conditional expectation. More precisely, we can show that up to a set of measure zero, there is only one random variable that satisfies the two properties in the above Lemma, and moreover, we can now generalise conditional expectations to more general sub-sigma algebras. This is formalised in this theorem

Theorem 1.4 (Existence and uniqueness of conditional expectation, ♩) Let $X \in L^1(\Omega, \mathscr{F}, \mathbf{P})$ and $\mathscr{G} \subseteq \mathscr{F}$ a sigma algebra. Then there exists a random variable $Y \in L^1(\Omega, \mathscr{F}, \mathbf{P})$ that is $\mathscr{G}$-measurable and given any $A \in \mathscr{G}$,

$$\mathbf{E}[Y\,\mathbf{1}_A] = \mathbf{E}[X\,\mathbf{1}_A]$$

Moreover, if Y' is another random variable with these two properties, we have that $Y = Y'$ almost surely.

Main idea: For uniqueness, use a general argument of considering $\mathscr{G}$-measurable sets of the type $\{Y > Y'\}$ and $\{Y < Y'\}$ and show that they have zero measure. For existence first use the geometry of the Hilbert spaces $\mathscr{L}(\mathscr{G}) \subseteq \mathscr{L}(\mathscr{F})$ and then extend the result to $\mathscr{L}^1$ random variables by suitable monotone approximations, i.e:

$$X_n = X \wedge n$$

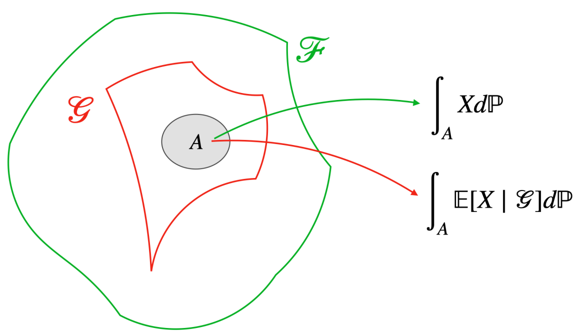


Figure 2: X and its conditional expectation have the same average on each $\mathscr{G}$-measurable set.

Definition 1.5 Given $X \in L^1(\Omega, \mathscr{F}, \mathbf{P})$ and $\mathscr{G} \subseteq \mathscr{F}$ a sigma algebra, we refer to the random variable Y of Theorem 1.4 as the conditional expectation of X given $\mathscr{G}$.

Proof of Theorem 1.4. We first prove the almost-sure uniqueness of a version of $X' = \mathbf{E}[X \mid \mathscr{G}]$. We then prove the existence of X' when $X \in \mathscr{L}^2$, and finally, we prove existence in general. Suppose that $X \in \mathscr{L}^1$ and that Y and Y' are both versions of $\mathbf{E}[X \mid \mathscr{G}]$. Then $Y, Y' \in \mathscr{L}^1$ (we are assuming existence at the moment), and

$$\mathbf{E}\big[(Y - Y')\mathbf{1}_G\big] = 0 \qquad \forall G \in \mathscr{G}$$

Clearly the set $G = \{Y > Y'\}$ is in $\mathscr{G}$, but since on G we have that $(Y - Y') > 0$ almost surely, and $\mathbf{E}[(Y - Y')\mathbf{1}_G] = 0$ it must be the case that $\mathbf{P}(G) = 0$. Which means that $Y \leq Y'$ almost surely. Symmetry finishes the argument.

We now show existence of $\mathbf{E}[X \mid \mathscr{G}]$ for $X \in \mathscr{L}^2$. We know that $\mathscr{L}^2(\mathscr{G}) \subseteq \mathscr{L}^2(\mathscr{F})$ is a complete subspace, so we know (from linear analysis, or other general measure theory) that there exists an orthogonal projection, i.e: a $Y \in \mathscr{L}^2(\mathscr{G})$ such that

$$\mathbf{E}\left[(X - Y)^2\right] = \inf\left\{\mathbf{E}\left[(X - W)^2\right] : W \in \mathscr{L}^2(\mathscr{G})\right\}$$

and

$$\langle X - Y, Z \rangle = 0 \quad \forall Z \in \mathscr{L}^2(\mathscr{G})$$

Then taking any $G \in \mathscr{G}$, we clearly see that $\mathbf{1}_G \in \mathscr{L}^2(\mathscr{G})$ and since $\langle X - Y, \mathbf{1}_G \rangle = \mathbf{E}[(X - Y)\mathbf{1}_G] = 0$ then we have shown existence of conditional expectation in $\mathscr{L}^2$. We now move to the case of $\mathscr{L}^1$. Without loss of generality we can assume $X \geq 0$ because otherwise we can do the usual splitting into positive and negative parts. Take $X \in \mathscr{L}^1(\mathscr{F})$. Then by constructing $X_n = X \wedge n$, we have a sequence of bounded random variables, which means they are in L^2, so we have the existence of a version Y_n of $\mathbf{E}[X_n \mid \mathscr{G}]$. In a moment we will show that $0 \leq Y_n \uparrow$ almost surely, but supposing it is true, we can set $Y(\omega) = \limsup Y_n(\omega)$, and then we see that $Y \in \mathrm{m}\mathscr{G}$ and $Y_n \uparrow Y$ almost surely, so from the monotone convergence theorem for regular expectations, we see that for any $G \in \mathscr{G}$, $\mathbf{E}[Y\mathbf{1}_G] = \mathbf{E}[X\mathbf{1}G]$ where we have used MCT both for X_n and Y_n. To prove the fact that we used here, we can show something better, namely that when U is non-negative and bounded, then $\mathbf{E}[U \mid \mathscr{G}] \geq 0$ almost surely. Indeed: let W be a version of $\mathbf{E}[U \mid \mathscr{G}]$. If $\mathbf{P}(W < 0) > 0$, then for some n, the set

$$G := \{W < -n^{-1}\} \in \mathscr{G}$$

has positive probability, which means that

$$0 \leq \mathbf{E}[U\mathbf{1}_G] = \mathbf{E}[W\mathbf{1}_G] < -n^{-1}\mathbf{P}(G) < 0$$

thus contradicting our assumption. ♡

Proposition 1.6 Let $X, \mathscr{G}$ be as before.

1. $\mathbf{E}[X \mid \mathscr{G}] = X$ if and only if X is $\mathscr{G}$-measurable.
2. $\mathbf{E}(\mathbf{E}[X \mid \mathscr{G}]) = \mathbf{E}[X]$
3. If $X \geq 0$ a.s, then $\mathbf{E}[X \mid \mathscr{G}] \geq 0$ a.s.

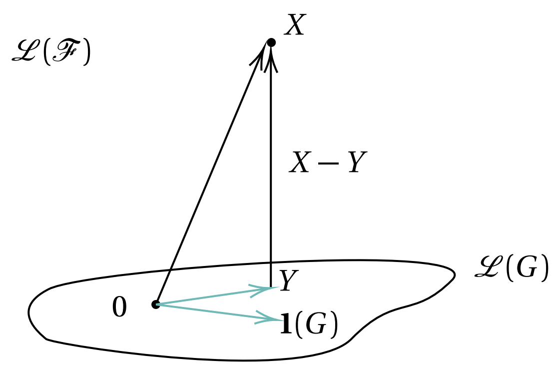


Figure 3: The diagram that says it all: existence of Conditional Expectation in $\mathscr{L}^2$

4. If $\sigma(X)$ and $\mathcal{G}$ are independent σ-algebras, then $\mathbf{E}[X \mid \mathcal{G}] = \mathbf{E}[X]$

5. If $X_1, X_2 \in L^1(\Omega, \mathcal{F}, \mathbf{P})$, and $\alpha, \beta \in \mathbf{R}$, then

$$\mathbf{E}[\alpha X_1 + \beta X_2 \mid \mathcal{G}] = \alpha \mathbf{E}[X_1 \mid \mathcal{G}] + \beta \mathbf{E}[X_2 \mid \mathcal{G}]$$

6. *Monotone Convergence Theorem:* If $X_n \uparrow X$, then

$$\mathbf{E}[X_n \mid \mathcal{G}] \to \mathbf{E}[X \mid \mathcal{G}]$$

7. *Fatou's Lemma:* If $X_n \in \mathrm{m}\mathcal{F}^+$

$$\mathbf{E}[\liminf_{n\to\infty} X_n \mid \mathcal{G}] \le \liminf_{n\to\infty} \mathbf{E}[X_n \mid \mathcal{G}]$$

8. *Dominated Convergence Theorem:* If $X_n \to X$ a.s. and $Y \in L^1$ with $|X_n| \le Y$ for all n a.s, then

$$\mathbf{E}[X_n \mid \mathcal{G}] \to \mathbf{E}[X \mid \mathcal{G}]$$

9. *Jensen's Inequality:* given $g : \mathbf{R} \to \mathbf{R}$ convex, then

$$\mathbf{E}[g(X) \mid \mathcal{G}] \ge g(\mathbf{E}[X \mid \mathcal{G}])$$

   And for $p \ge 1$

$$||\mathbf{E}[X \mid \mathcal{G}]||_p \le ||X||_p$$

10. *Tower Law:* given a sub sigma algebra $\mathcal{H} \subseteq \mathcal{G}$,

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

11. *Taking out what's known:* If $Z \in m\mathscr{G}$ is bounded, then

$$\mathbf{E}[ZX \mid \mathscr{G}] = Z\mathbf{E}[X \mid \mathscr{G}]$$

Proof of properties $(1) \to (3)$. The proof of the first three properties is almost immediate. For the first one, we note that by definition, a random varible Y is a version of $\mathbf{E}[X \mid \mathscr{G}]$ if it's $\mathscr{G}$-measurable and if the integrals coincide on $\mathscr{G}$-measurable sets. Therefore if we assume X is already $\mathscr{G}$-measurable, then the claim follows since obviously the integrals will coincide. For property (2) we recall that by definition, we have that

$$\mathbf{E}[\mathbf{E}[X \mid \mathscr{G}]\,\mathbf{1}(A)] = \mathbf{E}[X\,\mathbf{1}(A)]$$

For all $A \in \mathscr{G}$, from this we gather that by taking $A = \Omega$, we get the claim. Part (3) was proven in the existence and uniqueness Theorem but for completeness we repeat the argument, take the event $A = \{\mathbf{E}[X \mid \mathscr{G}] < 0\}$. Assume towards a contradiction that $\mathbf{P}[A] > 0$, then we have that for some n large enough, the event $A_n = \{\mathbf{E}[X \mid \mathscr{G}] < -n^{-1}\}$ also has positive probability. Then we have that

$$0 \overset{(1)}{\leq} \mathbf{E}[X\,\mathbf{1}(A_n)] \overset{(2)}{=} \mathbf{E}[\mathbf{E}[X \mid \mathscr{G}]\,\mathbf{1}(A_n)] \overset{(3)}{<} -n^{-1}\mathbf{P}[A_n] < 0$$

Where step (1) is due to assumption of X being non-negative, (2) is due to the definition of conditional expectation, step (3) is due to a fundamental estimate. This is a contradiction. ♡

Proof of Property (4). Proving Property (4) as it is is quite trivial, indeed: if $\sigma(X)$ is independent of $\mathscr{G}$, then in particular, random variables with respect to these distinct sigma algebras will be independent to each other, so in particular X is independent to $\mathbf{1}(G)$ for any $G \in \mathscr{G}$. Therefore:

$$\mathbf{E}[X\,\mathbf{1}(G)] = \mathbf{E}[X]\mathbf{E}[\mathbf{1}(G)] = \mathbf{E}[\mathbf{E}[X]\,\mathbf{1}(G)]$$

However, due to a lack of self-respect I will state and prove a stronger statement. If $\mathscr{H}$ is a sigma algebra that is independent of $\sigma(\sigma(X), \mathscr{G})$, then we have that

$$\mathbf{E}[X \mid \sigma(\mathscr{H}, \mathscr{G})] = \mathbf{E}[X \mid \mathscr{G}]$$

Of course our goal is to show that

$$\mathbf{E}[X\,\mathbf{1}(F)] = \mathbf{E}[\mathbf{E}[X \mid \mathscr{G}]\,\mathbf{1}(F)]$$

For all $F \in \sigma(\mathscr{G}, \mathscr{H})$. To do so we first notice that the maps

$$F \mapsto \mathbf{E}[X\,\mathbf{1}(F)] \qquad F \mapsto \mathbf{E}[\mathbf{E}[X \mid \mathscr{G}]\,\mathbf{1}(F)]$$

are two measures, so to ask whether they coincide on the complicated σ-algebra $\sigma(\mathcal{G},\mathcal{H})$, a much more tractable task is to ask if they agree on a simpler generating π system. Such a π-system is the sets of the form $\{G\cap H : G\in\mathcal{G}, H\in\mathcal{H}\}$. Then we have that

$$\begin{aligned}\mathbf{E}[X\,\mathbf{1}(G\cap H)] &= \mathbf{E}[X\,\mathbf{1}(G)\,\mathbf{1}(H)]\\ &\overset{(!)}{=} \mathbf{E}[X\,\mathbf{1}(G)]\mathbf{P}[H]\\ &= \mathbf{E}[\mathbf{E}[X\mid\mathcal{G}]\,\mathbf{1}(G)]\mathbf{P}[H]\\ &= \mathbf{E}[\mathbf{E}[X\mid\mathcal{G}]\,\mathbf{1}(G\cap H)]\end{aligned}$$

Where the only questionable step is (!), and this is due to the assumption of $\mathcal{H}$ being independent to the sigma-algebra $\sigma(\sigma(X),\mathcal{G})$. Thus we see that the two measures agree on a generating π-system and so they agree on the whole sigma algebra as required. ♡

Proof of Property (5). Proving property (5) (linearity) is trivial. Indeed, let $G\in\mathcal{G}$, and for simplicity let $Y_1=\mathbf{E}[X_1\mid\mathcal{G}]$ and define Y_2 in the same manner. Then

$$\begin{aligned}\mathbf{E}[(\alpha X_1+\beta X_2)\,\mathbf{1}(G)] &= \alpha\mathbf{E}[X_1\,\mathbf{1}\,G]+\beta\mathbf{E}[X_2\,\mathbf{1}(G)]\\ &= \alpha\mathbf{E}[Y_1\,\mathbf{1}(G)]+\beta\mathbf{E}[Y_2\,\mathbf{1}(G)]\\ &= \mathbf{E}[(\alpha Y_1+\beta Y_2)\,\mathbf{1}(G)]\end{aligned}$$

♡

Proof of the Conditional Monotone Convergence Theorem. Let $Y_n=\mathbf{E}[X_n\mid\mathcal{G}]$. Since $\{X_n\}$ is monotone increasing, it follows from monotonicity of the conditional expectation that $Y_n\le Y_{n+1}$ and so $Y_n\uparrow\limsup_n Y_n$, which we label by Y. Letting $G\in\mathcal{G}$ we have that

$$\begin{aligned}\mathbf{E}[X\,\mathbf{1}(G)] &= \mathbf{E}[\lim_{n\to\infty}X_n\,\mathbf{1}(G)]\\ &= \lim_{n\to\infty}\mathbf{E}[X_n\,\mathbf{1}(G)]\\ &= \lim_{n\to\infty}\mathbf{E}[Y_n\,\mathbf{1}(G)]\\ &= \mathbf{E}[\lim_{n\to\infty}Y_n\,\mathbf{1}(G)]\\ &= \mathbf{E}[Y\,\mathbf{1}(G)]\end{aligned}$$

Where the swapping of limits occurs because of monotonicity. Hence showing that

$$\mathbf{E}[X\mid\mathcal{G}]=\limsup\mathbf{E}[X_n\mid\mathcal{G}]=\lim\mathbf{E}[X_n\mid\mathcal{G}]$$

We note that here we have written limit instead of limsup at the very end because X_n is a monotone increasing sequence. ♡

Proof of the Conditional Fatou Lemma. The idea is to apply the Conditional Monotone Convergence Theorem to the fact that the sequence of functions $\left\{\inf_{k\geq n} X_k\right\}_n$ is monotone increasing (making the set smaller makes the infimum no smaller). Therefore:

$$\begin{aligned}\mathbf{E}\left[\liminf_{n\to\infty} X_n \mid \mathscr{G}\right] &:= \mathbf{E}\left[\lim_{n\to\infty}\inf_{k\geq n} X_k \mid \mathscr{G}\right] \\ &\overset{(1)}{=} \lim_{n\to\infty}\mathbf{E}\left[\inf_{k\geq n} X_k \mid \mathscr{G}\right] \\ &\overset{(2)}{\leq} \lim_{n\to\infty}\inf_{k\geq n}\mathbf{E}[X_k \mid \mathscr{G}] \\ &= \liminf_{n\to\infty}\mathbf{E}[X_n \mid \mathscr{G}]\end{aligned}$$

Where (1) comes from the Conditional MCT and (2) comes from the fact that conditional expectation is monotone and so since $\inf_{k\geq n} X_k \leq X_j$ for all $j \geq n$, then the inequality on conditional expectations also holds. ♡

Proof of the Conditional Dominated Convergence Theorem. Since we assume $|X_n| \leq Y$ we then have that both $Y + X_n$ and $Y - X_n$ are non-negative. Therefore we may write

$$\mathbf{E}[Y + X] = \mathbf{E}\left[\liminf_n (Y + X_n)\right] \leq \liminf_{n\to\infty}\mathbf{E}[Y + X_n]$$

and

$$\mathbf{E}[Y - X] = \mathbf{E}\left[\liminf_n (Y - X_n)\right] \leq \liminf_{n\to\infty}\mathbf{E}[Y - X_n]$$

which rearranging gives that

$$\limsup_n X_n \leq \mathbf{E}[X] \leq \liminf_n X_n$$

as required. ♡

Proof of Conditional Jensen Inequality. The key of this proof is that any convex function f may be expressed as the supremum of countably many affine functions:

$$f(x) = \sup_i (a_i x + b_i)$$

Then

$$\begin{aligned}\mathbf{E}[f(X) \mid \mathscr{G}] &= \mathbf{E}\left[\sup_n a_n X + b_n \mid \mathscr{G}\right] \\ &\overset{(!)}{\geq} \sup_n a_n \mathbf{E}[X \mid \mathscr{G}] + b_n = f(\mathbf{E}[X \mid \mathscr{G}])\end{aligned}$$

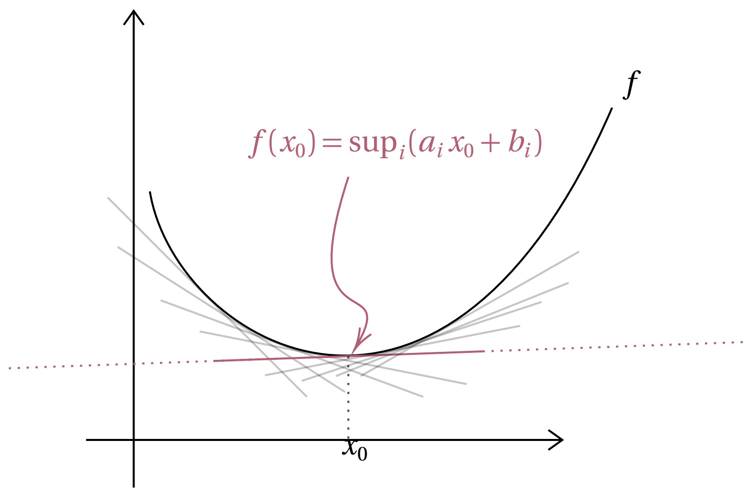


Figure 4: The key idea in the proof of Conditional Jensen

Where step (!) is not that it's unclear but rather there is a big subtlety that's easy to overlook: to do this we needed the supremum to be countable, so that the mechanism in the proof that conditional expectation is monotone (i.e: respects inequalities) works. To prove the "non-lengthening" property, we use a direct application of Jensen's Inequality. Note that it is important that $p \geq 1$ so that $x \mapsto |x|^p$ is convex:

$$\begin{aligned} \|\mathbf{E}[X \mid \mathscr{G}]\|_p^p &= \mathbf{E}[|\mathbf{E}[X \mid \mathscr{G}]|^p] \\ &\leq \mathbf{E}[\mathbf{E}[|X|^p \mid \mathscr{G}]] \\ &= \mathbf{E}[|X|^p]. \end{aligned}$$

♡

The proof of the Tower Law is truly trivial, the only thing left to prove is Taking What's Known:

Proof of Taking Out What's Known. This is just a matter of knowing that given a $\mathscr{G}$-measurable random variable Y, one may express this as a monotone increasing limit of simple functions, i.e:

$$\sum_{i=1}^{n} a_i \mathbf{1}(A_i) \uparrow Y$$

With each $A_i \in \mathscr{G}$. Then we can just use the conditional MCT and linearity and a tiny bit of algebra and get the claim. ♡

Remark 1.7 Note that we can make sense of things like $\mathbf{E}[X \mid Z]$ for some other random variable Z. Simply, we set $\mathbf{E}[X \mid Z] = \mathbf{E}[X \mid \sigma(Z)]$, similarly, $\mathbf{E}[X \mid Z_1, Z_2, \cdots] = \mathbf{E}[X \mid \sigma(Z_1, Z_2, \cdots)]$.

1.2 Agreement with traditional usage

Example 1.8 (Two RVs with joint density) Suppose that X and Z are two random variables with a joint density function

$$f_{X,Z}(x,z).$$

Then

$$f(z) = \int_{\mathbf{R}} f_{X,Z}(x,z)\,dx$$

acts as a probability density function for Z. We can define the elementary conditional density function

$$f_{X|Z}(x\mid z) := \frac{f_{X,Z}(x,z)}{f_Z(z)} \mathbf{1}_{\{f_Z(z)\neq 0\}}$$

Let $h : \mathbf{R} \to \mathbf{R}$ be a Borel function such that $h(X) \in \mathscr{L}^1$. Set

$$g(z) := \int_{\mathbf{R}} h(x) f_{X|Z}(x\mid z)\,dx$$

Then the claim is that $g(Z)$ is a version of $\mathbf{E}[h(X) \mid Z]$.

Proof. The typical element of $\sigma(Z)$ takes the form $\{\omega : Z(\omega) \in B\}$ for $B \in \mathscr{B}(\mathbf{R})$. Hence, we wish to show that

$$\mathbf{E}[h(X)\mathbf{1}_B(Z)] = \mathbf{E}\big[g(Z)\mathbf{1}_B(Z)\big]$$

But the left hand side simply equates to

$$\iint h(x)\mathbf{1}_B(z) f_{X,Z}(x,z)\,dx\,dz$$

and the right hand size equals

$$\int g(z)\mathbf{1}_B(z) f_Z(z)dz = \int \left(\int h(x) f_{X|Z}(x\mid z)dx\right)\mathbf{1}_B(z) f_Z(z)dz$$

thus Fubini's Theorem and the definition of the elementary conditional density finishes the claim.
♡

Example 1.9 (Gaussian random variables) Let (X,Y) be a random Gaussian vector, that is to say, $aX + bY$ is a one-dimensional Gaussian for all $a, b \in \mathbf{R}$. Let $\mathscr{G} = \sigma(Y)$, and let us compute $X' = \mathbf{E}[X \mid \mathscr{G}]$.

Solution. Since X' is $\mathscr{G}$-measurable and $\mathscr{G} = \sigma(Y)$, then we can write $X' = f(Y)$ for some Borel function f. We make an Ansatz that $f(y) = ay + b$, i.e: $X' = aY + b$, and then solve for the constants a and b and finally check whether this satisfies the properties of conditional expectation. Since $\mathbf{E}[X'] = \mathbf{E}[X]$ we have that by linearity,

$$a\mathbf{E}[Y] + b = \mathbf{E}[X]$$

By the orthogonality property of conditional expectation in $\mathscr{L}^2$ (note that Gaussians have finite variance so they are indeed in this space):

$$0 = \mathsf{Cov}\big(X - X', Y\big)$$

so that

$$\mathsf{Cov}(X, Y) = a\mathsf{Var}(Y)$$

Then, since $(X - X', Y)$ is a Gaussian vector, by elementary probability we know that $X - X'$ and Y are independent, so for any $A \in \sigma(Y)$, we will have that

$$\mathbf{E}[(X - X')\mathbf{1}_A] = \mathbf{E}[(X - X')]\mathbf{E}[\mathbf{1}_A] = 0$$

thus showing that with this choice of a and b, we have that X' is indeed a version of the conditional expectation.

♡

2 Discrete-time Martingales

In this chapter, we explore the concept of discrete-time Martingales, which are mathematical models that capture the idea of fair games or unbiased processes over time. Intuitively, a Martingale represents a sequence of random variables, where each value is the best prediction of the next value given all the past information. This means that, on average, the future value is expected to be the same as the current one, reflecting the notion that there is no predictable gain or loss. Martingales are powerful tools in probability theory, with applications in finance, gambling strategies, and stochastic processes, helping to model situations where outcomes evolve over time in a manner that is "fair" and memoryless. The first concept we need is that of a filtration

Definition 2.1 Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space. An increasing family of sub-σ-algebras $\{\mathscr{F}_n : n \geq 0\}$, i.e:

$$\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}$$

We define

$$\mathscr{F}_\infty = \sigma\left(\bigcup_n \mathscr{F}_n\right)$$

Remark 2.2 The significance of a filtration is that they model the information available to us at a given stage of the random process. This information comes in the shape of events, which one may think of being able to know whether they have occurred. Naturally, as n grows larger, the filtration becomes *finer*, and as such more detailed information is available to us. The choice of wording indicates that as n grows, the information gets filtered, say through a sieve, and we get finer details. Another way to think about this is that for a given random outcome $\omega \in \Omega$ of our sample space, the only information known to us about ω at time n, is the value $Z(\omega)$ of all $\mathscr{F}_n$-measurable functions Z.

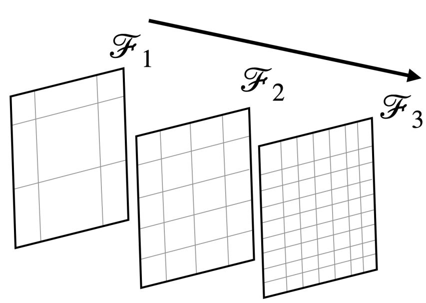


Figure 5: Illustration of a filtration

Example 2.3 If (X_n) is a sequence of random variables, the filtration $\mathscr{F}_n^X := \sigma(X_1, \cdots, X_n)$ is called the natural filtration.

2.1 Adapted Processes

Definition 2.4 A family of random variables, usually called a process, (X_n) is called adapted to a filtration $(\mathscr{F}_n)$ is X_k is $\mathscr{F}_k$-measurable.

Remark 2.5 The idea is that if (X_n) is an adapted process, then the value $X_k(\omega)$ is known to us at time k. An alternative way to think about this is that a process is adapted to $(\mathscr{F}_n)$ if and only if $\mathscr{F}_n^X \subseteq \mathscr{F}_n$.

2.2 Martingale, superMartingale, subMartingale

Definition 2.6 (☕☕☕, Martingale) A process (X_n) is called a Martingale relative to $((\mathscr{F}_n), \mathbf{P})$ if:

1. (X_n) is $(\mathscr{F}_n)$-adapted.
2. (X_n) is integrable, that is to say, $\mathbf{E}[|X_n|] < \infty$ for all n.
3. $\mathbf{E}[X_n \mid \mathscr{F}_{n-1}] = X_{n-1}\ a.s$

We distinguish between two other kinds of Martingales:

Definition 2.7 With conditions 1. and 2. from Definition 2.6, we have that if

- $\mathbf{E}[X_n \mid \mathscr{F}_{n-1}] \leq X_{n-1}$, then (X_n) is referred to as a super-Martingale.
- $\mathbf{E}[X_n \mid \mathscr{F}_{n-1}] \geq X_{n-1}$, then (X_n) is referred to as a sub-Martingale.

Remark 2.8 (Interpretation) It is useful to think of Martingales in terms of betting games. See the introductory paragraph of this chapter again.

Remark 2.9 (Equivalent characterisations) It is easily shown that an adapted integrable process (X_n) is a Martingale if and only if $\mathbf{E}[X_n \mid \mathscr{F}_m] = X_m$ for $m < n$. Indeed: one direction is obvious, but supposing this property holds, we can use the Tower Law, and see that

$$\mathbf{E}[X_n \mid \mathscr{F}_m] = \mathbf{E}[\mathbf{E}[X_n \mid \mathscr{F}_{m+1}] \mid \mathscr{F}_m]$$

which for convenience we write as $\mathbf{E}[X \mid \mathscr{F}_{m+1} \mid \mathscr{F}_m]$. We can repeatedly use the Tower Law until we reach the conclusion

$$\mathbf{E}[X_n \mid \mathscr{F}_m] = \mathbf{E}[X_n \mid \mathscr{F}_{n-1} \mid \mathscr{F}_{n-2} \mid \cdots \mid \mathscr{F}_m] = X_m$$

Example 2.10 (Three examples of Martingales) Let us look at three examples

1. **Sums of independent zero-mean RVs:** Let (X_n) be an L^1 sequence of independent RVs with zero-mean. Define $(S_0 := 0)$

$$S_n := \sum_{i=1}^{n} X_i \qquad \mathscr{F}_n = \sigma(X_1, X_2, \cdots, X_n)$$

   Observe that

$$\mathbf{E}[S_n \mid \mathscr{F}_{n-1}] = \mathbf{E}[X_n \mid \mathscr{F}_{n-1}] + \mathbf{E}[S_{n-1} \mid \mathscr{F}_{n-1}]$$

   Since the RVs are independent, $\sigma(X_n)$ and $\mathscr{F}_{n-1}$ are independent σ-algebras and as such $\mathbf{E}[X_n \mid \mathscr{F}_{n-1}] = \mathbf{E}[X_n] = 0$. Since S_{n-1} is $\mathscr{F}_{n-1}$-measurable, the whole quantity above is equal to S_{n-1}, thus showing (S_n) is a Martingale.

2. **Product of non-negative independent RVs of mean 1:** Let (X_n) be a sequence of independent non-negative RVs with $\mathbf{E}[X_n] = 1$ for all n. Define $(M_0 := 1, \mathscr{F}_0 := \{\varnothing, \Omega\})$

$$M_n := \prod_{i=1}^{n} X_i \quad \mathscr{F}_n = \sigma(X_1, X_2, \cdots, X_n)$$

   Then we see that

$$\mathbf{E}[M_n \mid \mathscr{F}_{n-1}] = \mathbf{E}[X_n M_{n-1} \mid \mathscr{F}_{n-1}]$$

   Recalling property 12 of Proposition 1.6, we can rewrite this quantity as $M_{n-1}\mathbf{E}[X_n \mid \mathscr{F}_{n-1}]$. But the fact that once again, the random variables are independent makes this quantity equal to $M_{n-1}\mathbf{E}[X_n] = M_{n-1}$, showing (M_n) is a Martingale.

3. **Accumulating data about a random variable:** Let $(\mathscr{F}_n)$ be a given filtration and $\xi \in L^1(\Omega, \mathscr{F}, \mathbf{P})$ be a random variable we are extracting information about. Define $M_n := \mathbf{E}[\xi \mid \mathscr{F}_n]$.

By the Tower Law,

$$\mathbf{E}[M_n \mid \mathscr{F}_{n-1}] = \mathbf{E}[\xi \mid \mathscr{F}_n \mid \mathscr{F}_{n-1}] = \mathbf{E}[\xi \mid \mathscr{F}_{n-1}] = M_{n-1}$$

2.3 Stopping times

With the analogy in mind that Martingales model fair betting games, we may now add a layer of complexity which enables us to *stop playing* the game at a given time. This idea is given by the notion of a stopping time. A real life example of a stopping time would be the time in which a trader wishes to exercise an American option.

Definition 2.11 (☕☕☕, Stopping Time) Let $(\Omega, \mathscr{F}, \{\mathscr{F}_n\}, \mathbf{P})$ be as usual. A map $T : \Omega \to \{0, 1, 2, \cdots, \infty\}$ is called a stopping time if given any $n \in \{0, 1, 2, \cdots, \infty\}$, the set

$$\{T \leq n\} := \{\omega : T(\omega) \leq n\} \in \mathscr{F}_n$$

Remark 2.12 (Intuition) If $T(\omega)$ is the time we stop playing our game according to some strategy, we may know whether to keep playing or stop just by looking at the information available to us at the moment, but not in the future.

Proposition 2.13 (Equivalent characterisation for Stopping Times in discrete time) A map $T : \Omega \to \{0, 1, 2, \cdots, \infty\}$ is a Stopping Time if and only if for any $n \in \{0, 1, 2, \cdots, \infty\}$,

$$\{T = n\} \in \mathscr{F}_n$$

Proof. Observe that $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$. If T is a stopping time, then since $\{T \leq n-1\} \in \mathscr{F}_{n-1} \subseteq \mathscr{F}_n$, we have that $\{T = n\} \in \mathscr{F}_n$. Conversely, if $\{T = n\} \in \mathscr{F}_n$, then one has that

$$\{T \leq n\} = \bigcup_{k=0}^{n} \{T = k\} \in \mathscr{F}_n$$

♡

Example 2.14 Let us see some (non) examples of Stopping Times.

- **First Passage Time:** Let (A_n) be an adapted process, and let $B \in \mathscr{B}$ be some Borel set, then the least time in which (A_n) enters B,

$$T = \inf\{n : A_n \in B\}$$

is a stopping time, this is because

$$\{T \leq n\} = \bigcup_{k=0}^{n} \{A_k \in B\} \in \mathscr{F}_n.$$

- However, if we set

$$L = \sup\{n : A_n \in B\}$$

then L is not a stopping time, because in some sense, it is looking into the future. Indeed, if it were, $\{L \leq k\} \in \mathscr{F}_k$, i.e: if we can know that k is largest value for which $A_n \in B$, then we know that A_{k+1} does not belong to B, so $\{A_{k+1} \in B^c\} \in \mathscr{F}_k$, which could be true in some cases, but not generally.

With the notion of a stopping time we can talk about *stopped processes*.

Definition 2.15 Let (X_n) be a process adapted to a filtration $\mathscr{F}_n$, and let τ be a stopping time. The process $X^\tau \equiv X_{n\wedge\tau}$ is referred to as a **stopped process**.

Naturally, a stopped process $X_{n\wedge\tau}$ agrees with the underlying process X_n until the stopping time occurs, after which the process stays constant and equal to X_τ. This should be interpreted as some strategy by which the gambler decides to stop playing.

Something else we are interested in, is to know with respect to what σ-algebra is the random variable X_τ measurable. I.e: we would like to make sense of something of the shape $\mathscr{F}_\tau$. Of course, since τ is a random variable, $\mathscr{F}_\tau$ doesn't make much sense yet.

Definition 2.16 Let $(\Omega, \mathscr{F}, \{\mathscr{F}_n\}, \mathbf{P})$ be a filtered space, (X_n) an adapted process and τ a stopping time for $(\mathscr{F}_n)$. Then we define

$$\mathscr{F}_\tau = \{A \in \mathscr{F}_\infty : A \cap \{\tau \leq n\} \in \mathscr{F}_n\}$$

Lemma 2.17 $\mathscr{F}_\tau$ as above is a σ-algebra.

Proof. It is clear that $\Omega \in \mathscr{F}_\tau$. Let (A_n) be a sequence of sets in $\mathscr{F}_\tau$, then

$$\left(\bigcup_n A_n\right) \cap \{\tau \leq n\} = \bigcup_n A_n \cap \{\tau \leq n\}$$

By definition, each $A_n \cap \{\tau \leq n\} \in \mathscr{F}_n$, and as such this union is also in $\mathscr{F}_n$, showing that $\mathscr{F}_\tau$ is closed under countable unions. Suppose now that $A \in \mathscr{F}_\tau$, then $A^c \cap \{\tau \le n\} = \{\tau \le n\} \setminus (A \cap \{\tau \le n\}) \in \mathscr{F}_n$. ♡

Example 2.18 Let us give an example of some $\mathscr{F}_\tau$ to illustrate the choice of that definition. Suppose $\tau = k$ for some $k \in \mathbf{N}$, i.e: τ is a constant random variable. Then the claim is that $\mathscr{F}_\tau = \mathscr{F}_k$. Indeed, if $A \in \mathscr{F}_\tau$, then

$$A \cap \{k \le n\} \in \mathscr{F}_n$$

if $k \le n$, then this is equivalent to saying that $A \cap \Omega \in \mathscr{F}_n$, i.e: $A \in \mathscr{F}_n$. If $k > n$, then $A \cap \{k \le n\} = A \cap \varnothing = \varnothing \in \mathscr{F}_n$, therefore $\mathscr{F}_\tau \subseteq \mathscr{F}_n$. It is also easily seen that $\mathscr{F}_n \subseteq \mathscr{F}_\tau$

Proposition 2.19 Let $S, T, (T_n)$ be stopping times for a filtration $(\mathscr{F}_m)$. Show the following properties:

1. $T \wedge S, T \vee S, \sup_n T_n, \inf_n T_n, \limsup T_n, \liminf T_n$ are all stopping times.
2. If $T \le S$, then $\mathscr{F}_T \subseteq \mathscr{F}_S$.
3. $X_T \mathbf{1}_{\{T<\infty\}} \in \mathrm{m}\mathscr{F}_T$.
4. If (X_n) is an adapted process, then so is $(X_{n \wedge T})$.
5. If (X_n) is an integrable process, then so is $(X_{n \wedge T})$.

Proof. .

1. Let us first show $T \vee S$ is a stopping time. If $n \in \mathbf{N} \cup \{\infty\}$, then

   $$\{T \vee S \le n\} = \{T \le n\} \cap \{S \le n\}$$

   since both of these belong to $\mathscr{F}_n$, then $T \vee S$ is a stopping time. For $T \wedge S$ we show that the complement belongs to $\mathscr{F}_n$, and since this is a sigma algebra, the defining property will follow.

   $$(\{T \wedge S \le n\})^c = \{T \wedge S > n\} = \{T > n\} \cap \{S > n\} = \{T \le n\}^c \cap \{S \le n\}^c \in \mathscr{F}_n$$

   Next, notice that

   $$\left\{\inf_m T_m \ge n\right\} = \bigcap_m \{T_m \ge n\}$$

   and the sup follows similarly, and $\liminf$ and $\limsup$ follow immediately now.

2. Let $A \in \mathcal{F}_T$, then since $\{S \le n\} \subseteq \{T \le n\}$

$$A \cap \{S \le n\} \subseteq A \cap \{T \le n\} \in \mathcal{F}_n$$

Thus $A \in \mathcal{F}_S$.

3. To show that $X_T \mathbf{1}(T < \infty)$ is $\mathcal{F}_T$-measurable we have to show that for a Borel set A, we have that $\{X_T \mathbf{1}(T < \infty) \in A\} \in \mathcal{F}_T$, that is to say, for $n \in \mathbf{N}$:

$$\{X_T \mathbf{1}(T < \infty) \in A\} \cap \{T \le n\} \in \mathcal{F}_n$$

But this event can be seen as

$$\bigcup_{s=0}^{n} \{X_s \in A\} \cap \{T = s\} \in \mathcal{F}_n$$

as required.

4. To show that the stopped process is adapted, we first note that $X_{T\wedge t}$ is $\mathcal{F}_{T\wedge t}$-measurable, as this new stopping time is always finite, and since $T \wedge t \le t$, then this sigma algebra is contained in $\mathcal{F}_t$.

5. To show that the stopped process is integrable, we have that

$$\begin{aligned}
\mathbf{E}[|X_{T\wedge t}|] &= \mathbf{E}\left[\sum_{s=0}^{t-1} |X_s| \mathbf{1}(T = s)\right] + \mathbf{E}\left[\sum_{s=t}^{\infty} |X_t| \mathbf{1}(T = s)\right] \\
&\le \mathbf{E}\left[\sum_{s=0}^{t-1} |X_s| \mathbf{1}(T = s)\right] + \mathbf{E}|X_t| \\
&\le \sum_{s=0}^{t} \mathbf{E}|X_s| < \infty
\end{aligned}$$

♡

2.3.1 Optional Stopping

Theorem 2.20 (The Optional Stopping Theorem) Let (X_n) be a Martingale. Then

1. If T is a stopping time, then the stopped process X^T is also a Martingale, which in particular implies that $\mathbf{E}\left[(X^T)_n\right] = \mathbf{E}[X_0]$.

2. If $S \le T$ are bounded stopping times, then $\mathbf{E}[X_T \mid \mathcal{F}_S] = X_S$ almost surely.

3. If $S \le T$ are bounded stopping times, then $\mathbf{E}[X_T] = \mathbf{E}[X_S]$.

4. If there is some $Y \in L^1$ so that $|X_n| \le Y$ almost surely for all n, then if $T < \infty$ almost surely, we have $\mathbf{E}[X_T] = \mathbf{E}[X_0]$

5. If X has bounded increments, i.e: $|X_{n+1} - X_n| \le M$ for all n, and T is a stopping time with $\mathbf{E}[T] < \infty$, then $\mathbf{E}[X_T] = \mathbf{E}[X_0]$

We will add the intuition sections as the proofs appear

Main idea: Use the standard decomposition of the stopped process and then perform a routine computation.

Proof of 1. For the first claim, we will use the following decomposition trick, which will be used a lot in the future:

$$(X^T)_n \equiv X_{T\wedge n} = \sum_{k=0}^{n-1} X_k \mathbf{1}(T=k) + X_n \mathbf{1}(T > n-1)$$

By linearity of conditional expectation

$$\begin{aligned}\mathbf{E}[X_{T\wedge n} \mid \mathscr{F}_{n-1}] &= \mathbf{E}\left[\sum_{k=0}^{n-1} X_k \mathbf{1}(T=k) \mid \mathscr{F}_{n-1}\right] + \mathbf{E}[X_n \mathbf{1}(T > n-1) \mid \mathscr{F}_{n-1}] \\ &= \sum_{k=0}^{n-1} \mathbf{E}[X_k \mathbf{1}(T=k) \mid \mathscr{F}_{n-1}] + X_{n-1} \mathbf{1}(T > n-1) \stackrel{!}{=}\end{aligned}$$

in this last step, we have used the fact that $\mathbf{1}(T > n-1) \in \mathrm{m}\mathscr{F}_{n-1}$ to take $\mathbf{1}(T > n-1)$ out of the expectation as well as the Martingale property on X_n. Notice as well that $X_k \mathbf{1}(T=k)$ is $\mathscr{F}_{n-1}$ measurable for $k \in [n-1]$, which means that the expression above becomes

$$\stackrel{!}{=} \sum_{k=0}^{n-1} X_k \mathbf{1}(T=k) + X_{n-1} \mathbf{1}(T > n-1) = X_{T\wedge(n-1)}$$

by using the decomposition again in reverse. ♡

Main idea: This is also a decomposition argument, but since we want to talk about both stopping times S and T, the decomposition we will use is

$$X_T = X_S + (X_T - X_{T-1}) + (X_{T-1} + X_{T-2}) + \cdots + (X_{S+1} - X_S)$$

Then we write this decomposition as a sum and verify that the definition of conditional expectation holds.

Proof of 2. The trick for this proof is to consider the increments

$$X_T = (X_T - X_{T-1}) + (X_{T-1} - X_{T-2}) + \cdots + (X_{S+1} - X_S) + X_S$$

By assumption T is bounded, say $T \le n$, so we can write this decomposition as

$$X_T = X_S + \sum_{k=0}^{n} (X_{k+1} - X_k)\,\mathbf{1}(S \le k < T)$$

(We are summing all the way to n from zero and only keeping the relevant summands). Now we let $A \in \mathscr{F}_S$. Then

$$\mathbf{E}[X_T\,\mathbf{1}_A] = \mathbf{E}[X_S\,\mathbf{1}_A] + \sum_{k=0}^{n} \mathbf{E}[(X_{n+1} - X_n\,\mathbf{1}(S \le k < T)\,\mathbf{1}_A]$$

Naturally:

$$\{S \le k\} \cap A \cap \{T > k\} \in \mathscr{F}_k$$

by definition of the stopped sigma algebra, so by $\{X_n\}$ being a Martingale $\mathbf{E}[(X_{k+1} - X_k\,\mathbf{1}(S \le k < T)\,\mathbf{1}_A]$ vanishes. This gives $\mathbf{E}[X_T\,\mathbf{1}_A] = \mathbf{E}[X_S\,\mathbf{1}_A]$ as required.

♡

Proof of 3. Follows from 2 by taking expectations of both sides. ♡

Main idea: This one follows from the first one and using DCT.

Proof of 4. We prove the case where X is bounded almost surely and T is almost-surely finite. We start by noting that since T is almost surely finite, then $X_{T\wedge n} \to X_T$ almost surely as $n \to \infty$. Moreover, since X is bounded, we can apply the Dominated Convergence Theorem and see that

$$\mathbf{E}[X_T] = \mathbf{E}\left[\lim_{n\to\infty} X_{T\wedge n}\right] = \lim_{n\to\infty} \mathbf{E}[X_{T\wedge n}] = \mathbf{E}[X_0]$$

♡

Main idea: Show that the sequence $X_{T\wedge n} - X_0$ has bounded expectation and use the DCT.

Proof of 5. For the case where $\mathbf{E}[T] < \infty$ and bounded increments, say bounded by $M \ge 0$. We consider the following decomposition of $X_{T\wedge n}-X_0$:

$$|X_{T\wedge n}-X_0|=\left|\sum_{i=1}^{T\wedge n}X_i-X_{i-1}\right|\leq\sum_{i=1}^{T\wedge n}|X_i-X_{i-1}|\leq M(T\wedge n)\leq MT$$

By taking expectations we see that $\mathbf{E}[|X_{T\wedge n}-X_0|]\leq M\mathbf{E}[T]<\infty$, and so by the DCT, followed by the first part of the OST:

$$\mathbf{E}[X_T-X_0]=\lim_{n\to\infty}\mathbf{E}[X_{T\wedge n}-X_0]=0$$

As required. ♡

Example 2.21 (Non-example) In the conditions of the Optional Stopping Theorem, to deduce that $\mathbf{E}[X_T]=\mathbf{E}[X_0]$, we needed that T was deterministically bounded, i.e: for all $\omega\in\Omega$, $T(\omega)<\infty$. To show that we cannot relax this assumption to being almost surely bounded, we consider the following example:

Let $(\xi_k)\overset{iid}{\sim}\begin{cases}+1 & wp=1/2\\ -1 & wp=1/2\end{cases}$ Then setting $X_0=0$ and $X_n=\sum_{k\leq n}\xi_k$, gives that (X_n) is a Martingale, because each ξ_k is iid and centered. Define now the stopping time of the first passage time of 1, i.e:

$$T=\inf\{t\geq 0:X_t=1\}$$

It is a known fact that a SSRW on $\mathbf{Z}$ is recurrent, so the probability that T is finite is of 1. However, it is also clear that T is not deterministically bounded. Therefore the assumptions are not satisfied. This is why we have that

$$\mathbf{E}[X_T]=1\neq 0=\mathbf{E}[X_0]$$

(Just to be thorough, could not be in the last case either because even though our increments are bounded, it is not the case that $\mathbf{E}T<\infty$)

Of course the inequalities in the above Theorem can be adjusted for super and sub Martingales respectively. Now we see a classic result of Martingale Theory

2.3.2 Gambler's Ruin

Let us use the Optional Stopping Theorem to show a more general version of a classical result. Suppose you are a gambler with initial wealth 0 and your wealth increases or decreases by 1 with equal probability at each time step t. Your wealth $X=(X_t)$ is clearly a Martingale. Define $T_c=\inf\{t\geq:X_t=c\}$ be the hitting time for c. This is obviously also a stopping time. For fixed $a, b > 0$, we can now understand what's the probability of hitting $-a$ before b (i.e: going broke before getting rich in gambling terms)

Theorem 2.22 (Gambler's Ruin)

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

Main idea: T_{-a} and T_b are both stopping times, you are interested in $T = T_{-a} \wedge T_b$. Use the fact that X has bounded increments and show that $\mathbf{E}[T] < \infty$.

Proof. It is clear that X has bounded increments, and moreover, note that we can bound above T by the first time that $(a+b)$ consecutive $+1$s appear. If we think of looking at consecutive $(a+b)$ games, since all are games are iid, each $(a+b)$ string of games has probability $2^{-(a+b)}$ of occurring, so the waiting time until the first one with all $+1$s appears, is distributed geometrically, so the expected number of $(a+b)$ strings needed is precisely 2^{a+b}, however, we need to take into account that each of these strings takes $a+b$ turns to finish so in total we have that

$$T \leq (a+b)2^{a+b}$$

which in particular means that the expectation of T is finite. Now that we have established bounded increments and finite expectation of T, we can use the Optional Stopping Theorem and conclude that

$$\mathbf{E}[X_T] = \mathbf{E}[X_0] = 0$$

But $X_T = -a\,\mathbf{1}(T_{-a} < T_b) + b\,\mathbf{1}(T_b > T_{-a})$, so

$$\mathbf{E}[X_T] = -a\mathbf{P}(T_{-a} < T_b) + b\,\mathbf{1}(T_b > T_{-a}) = 0$$

We also have the set of equations

$$1 = \mathbf{P}(T_b > T_{-a}) + \mathbf{P}(T_b > T_{-a})$$

which finishes the claim. ♡

2.4 The Martingale Convergence Theorem

Let us now study the problem of when does a Martingale (X_n) converge, i.e. under which conditions, do we have that almost surely

$$X_n \to X$$

And what properties can be attribute to X: what integrability properties does X satisfy? With respect to which σ-algebra is it measurable? To understand this convergence let us talk about a characterisation of convergence of sequences in $\mathbf{R}$ that perhaps the reader didn't think about before:

Let $a < b$ be two real numbers, and $(x_n) \in \mathbf{R}^\infty$, let us define the number of upcrossings between a and b. Let us define the following times: $T_0 = 0$, and recursively for $n \geq 1$

$$S_n = \inf\{k \geq T_{n-1} : x_k \leq a\} \qquad T_n = \inf\{k \geq S_n : x_k \geq b\}$$

We note that each T_n corresponds to an upcrossing: we started below a and we traveled all the way above b. We thus define the total number of upcrossings up to time n as

$$U_n([a,b],x) = \sup_{k\geq 0}\{T_k \leq n\} \uparrow \sup_{k\geq 0}\{T_k < \infty\} := U([a,b],x)$$

And similarly we have the total number of upcrossings in the infinite life of the sequence. A clear result is that

Lemma 2.23 A sequence $x \in \mathbf{R}^\infty$ converges if and only if for each pair $a < b$ of real numbers, $U([a,b],x) < \infty$.

This is intuitively clear because a sequence that converges will not oscillate forever between any pair of real numbers, and a sequence that oscillates infinitely between some pair of real numbers will not converge.

This is precisely the strategy we will use to prove a convergence Theorem for Martingales. This ingenious argument, due to Doob, first proceeds by finding a bound on the expected number of upcrossings that the Martingale does in an interval $[a,b]$. We will then see that under mild integrability conditions, this expectation can be bounded above by a finite number, which intuitively explains why we will have convergence. Let us present the aforementioned bound on the upcrossings:

Theorem 2.24 (Doob's Upcrossing Lemma ☕) Let X be a non-negative superMartingale. Then

$$(b-a)\mathbf{E}[U_N[a,b]] \leq \mathbf{E}\big[(X_N - a)^-\big]$$

Main idea: The idea is to count the "total winnings" made by our process between upcrossings.

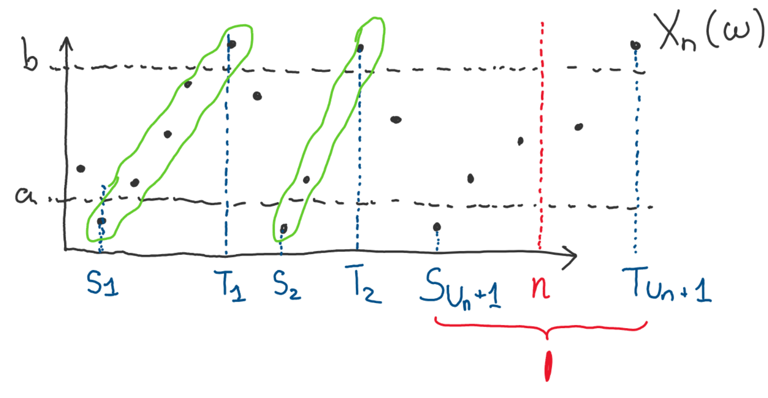
Proof. We have the first trivial observation:

$$\underbrace{X_{T_k}}_{\geq b} - \underbrace{X_{S_k}}_{\leq a} \geq b-a$$

And now We wish to count the overall "net winnings if you wish" obtained by all the upcrossings we have done up to time n. If we let $U_n = U_n([a,b],x)$, then we claim that

$$\sum_{k=1}^{n}(X_{T_k\wedge n} - X_{S_k\wedge n}) = \sum_{k=1}^{U_n}(X_{T_k} - X_{S_k}) + (X_n - X_{S_{U_n+1}})\mathbf{1}(S_{U_n+1} \leq n)$$

This is because up to time n, it might be the case that S_{U_n+1} occurs before time n, and T_{U_n+1} occurs after time n, in which case, the sum on the left hand side will include a $X_n - X_{S_{U_n+1}}$ but the big sum on the right hand side will exclude these two quantities, therefore we add them on the right hand side with an appropriate indicator function (see diagram for clarification).



Moving on, it is intuitively clear (or alternatively by induction) that $(T_k)_k$ and $(S_k)_k$ are both sequences of stopping times. Hence, for all n, $S_k \wedge n \leq T_k \wedge n$ are bounded stopping times so we can use the Optional Stopping Theorem to deduce that since (X_n) is a superMartingale, then

$$\mathbf{E}\left[X_{S_k\wedge n}\right] \geq \mathbf{E}\left[X_{T_k\wedge n}\right]$$

(The sooner you stop playing the higher your earnings), so taking expectations we see that by combining the facts above:

$$0 \geq \sum_{k=1}^{n}\mathbf{E}\left[X_{T_k\wedge n}\right] - \mathbf{E}\left[X_{S_k\wedge n}\right] \geq (b-a)\mathbf{E}[U_n] - \mathbf{E}\left[(X_n-a)^-\right]$$

We have used in this very last line, the fact here that $(X_n - S_{U_n+1})\mathbf{1}(S_{U_n+1} \le n) \ge -(X_n - a)^-$ ♡

With this in mind we can finally say something about convergence of Martingales.

Theorem 2.25 (Doob's Martingale Convergence Theorem ♩) Let (X_n) be a superMartingale that is bounded in $\mathscr{L}^1$, that is to say

$$\sup_n \mathbf{E}[|X_n|] < \infty$$

Then, almost surely $X_\infty := \lim X_n$ exists. Moreover X_∞ is $\mathscr{F}_\infty$ measurable, and $X_\infty \in \mathscr{L}^1(\mathscr{F}_\infty)$.

Main idea: The set $\Lambda := \{\omega : X_n(\omega) \text{ does not converge to a limit in } [-\infty, \infty]\}$ can be expressed as a countable union of sets involving an infinite number of upcrossings for some interval. Use Doob's upcrossing inequality to study the probability of these sets. This is a nice proof!

Proof. We write

$$\begin{aligned}
\Lambda &:= \{\omega : X_n(\omega) \text{ does not converge to a limit in } [-\infty, \infty]\} \\
&= \{\omega : \liminf X_n(\omega) < \limsup X_n(\omega)\} \\
&= \bigcup_{a,b \in \mathbf{Q}\, a<b} \{\omega : \liminf X_n(\omega) < a < b < \limsup X_n(\omega)\} \\
&:= \bigcup \Lambda_{a,b}
\end{aligned}$$

Note also that $\Lambda_{a,b} \subseteq \{\omega : U_\infty[a,b](\omega) = \infty\}$. However, by Doob's upcrossing's inequality, we have that

$$(b-a)\mathbf{E}[U_n[a,b]] \le |a| + \mathbf{E}[|X_n|] \le |a| + \sup_n \mathbf{E}[|X_n|] < \infty$$

By assumption of $\mathscr{L}^1$ boundedness. Now using the monotone convergence theorem, we can take $n \to \infty$ and obtain that

$$\mathbf{E}[U_\infty[a,b]] < \infty$$

which means that

$$\mathbf{P}(U_\infty[a,b] = \infty) = 0$$

And as such $\mathbf{P}(\Lambda_{a,b}) = 0$, but since Λ is a countable union of such sets, we have that $\mathbf{P}(\Lambda) = 0$. Thus we have that

$$X_\infty := \lim X_n$$

exists almost surely in $[-\infty, \infty]$ where we of course define $X_\infty(\omega)$ as $\lim X_n(\omega)$. This construction makes it clear that $X_\infty \in m\mathscr{F}_\infty$, and moreover, using Fatou's Lemma:

$$\mathbf{E}[|X_\infty|] = \mathbf{E}[\liminf |X_n|] \le \liminf \mathbf{E}[|X_n|] \le \sup \mathbf{E}[|X_n|] < \infty$$

Showing that $X_\infty \in \mathscr{L}^1(\mathscr{F}_\infty)$. ♡

2.5 $\mathscr{L}^p$ convergence of Martingales

We have just seen how for a superMartingale (X_n) (and hence also for a Martingale), we have that under certain regularity conditions,

$$X_n \to X_\infty \qquad a.s$$

Where X_∞ is some $\mathscr{F}_\infty = \sigma(\mathscr{F}_n : n \geq 0)$-measurable function. We would like now to generalise this statement of convergence in the $\mathscr{L}^p$ norms: when can we say that $\|X_n - X_?\|_p \to 0$? What can we say about the mystery random variable $X_?$. We need some prior results for this

Theorem 2.26 (Doob's maximal inequality , ♩) Let $X = (X_n)$ be a non-negative sub-Martingale. Let $X_n^* = \sup\{|X_t| : 0 \leq t \leq n\}$ be the running maximum. Let $\lambda \geq 0$. Then

$$\lambda \mathbf{P}(X_n^* \geq \lambda) \leq \mathbf{E}\left[X_n \mathbf{1}(X_n^* \geq \lambda)\right] \leq \mathbf{E}[X_n]$$

Main idea: Construct a stopping time T that tracks the first time the Martingale surpasses λ. Then use the fact that $T \leq n$ if and only if $X_n^* \geq \lambda$. Then use the Optional Stopping Theorem with the bounded stopping times $n \geq T \wedge n$ and decompose $X_{T\wedge n}$ as $X_T \mathbf{1}(T \leq n) + X_n \mathbf{1}(T > n)$.

Proof. Let $T = \inf\{t \geq 0 : X_t \geq \lambda\}$. This is obviously a stopping time. Then $T \wedge n$ is a bounded stopping time. Clearly the constant time n is also a bounded stopping time with $n \geq T \wedge n$ so by the third point of the Optional Stopping Theorem we have that since (X_n) is a sub-Martingale, then

$$\mathbf{E}[X_n] \geq \mathbf{E}[X_{T\wedge n}]$$

We can use the usual trick of decomposing $X_{T\wedge n}$ to see that

$$\begin{aligned}
\mathbf{E}[X_n] &\geq \mathbf{E}[X_T \mathbf{1}(T \leq n)] + \mathbf{E}[X_n \mathbf{1}(T > n)] \\
&\geq \lambda \mathbf{P}(T \leq n) + \mathbf{E}[X_n \mathbf{1}(T > n)] \\
&= \lambda \mathbf{P}(X_n^* \geq n) + \mathbf{E}[X_n(1 - \mathbf{1}(T \leq n))]
\end{aligned}$$

Subtracting $\mathbf{E}[X_n]$ from both sides and rearranging gives that

$$\lambda \mathbf{P}(X_n^* \geq n) \leq \mathbf{E}\left[X_n \mathbf{1}(X_n^* \geq \lambda)\right] \leq \mathbf{E}[X_n]$$

where the last inequality comes from the fact that X_n is non-negative. ♡

Example 2.27 Doob's Maximal Inequality can be used to show a version of the Azuma-Hoeffding Inequality, See the Example Sheet 2.

Theorem 2.28 (Doob's $\mathscr{L}^p$ inequality ☕) Let X be a Martingale or a non-negative sub-Martingale. Then for all $p > 1$ we have that

$$\left\|X_n^*\right\|_p \leq \frac{p}{p-1} \|X_n\|_p$$

Main idea:

Let X be a non-negative random variable and let $k < \infty$, then

$$\begin{aligned}
\mathbf{E}[(X \wedge k)^p] &= \int_0^\infty \mathbf{P}(X \wedge k \geq x^{1/p}) dx \\
&= \int_0^\infty p u^{p-1} \mathbf{P}(X \wedge k \geq u) du \qquad \text{substitute } u = x^{1/p} \\
&= \int_0^\infty p u^{p-1} \mathbf{E}[\mathbf{1}(X \wedge k \geq u)]\, du \\
&= \mathbf{E}\left[\int_0^\infty p u^{p-1} \mathbf{1}(X \wedge k \geq u) du\right] \qquad \text{Fubini} \\
&= \mathbf{E}\left[\int_0^k p u^{p-1} \mathbf{1}(X \geq u) du\right] = \int_0^k p u^{p-1} \mathbf{P}(X \geq u) du \qquad \text{Fubini}
\end{aligned}$$

Proof.

$$\begin{aligned}
\mathbf{E}\left[(X_n^* \wedge k)^p\right] &= \int_0^k p x^{p-1} \mathbf{P}(X_n^* \geq x) dx \\
&\leq \int_0^k p x^{p-2} \mathbf{E}\left[X_n \mathbf{1}(X_n^* \geq x)\right] dx \qquad \text{(Doob's Maximal Inequality)} \\
&= \frac{p}{p-1} \mathbf{E}\left[X_n (X_n^* \wedge k)^{p-1})\right] \qquad \text{(Above calculation in reverse)}
\end{aligned}$$

Now we are going to apply Hölder's inequality as follows:

$$\begin{aligned}
\mathbf{E}\left[X_n (X_n^* \wedge k)^{p-1}\right] &\leq \mathbf{E}\left[X_n^p\right]^{1/p} \mathbf{E}\left[(X_n^* \wedge k)^{p-1\cdot\frac{p}{p-1}}\right]^{\frac{p-1}{p}} \\
&= \|X_n\|_p \left\|(X_n^* \wedge k)\right\|_p^{p-1}
\end{aligned}$$

In summary we have that

$$\left\|X_n^* \wedge k\right\|_p^p = \mathbf{E}\left[(X_n^* \wedge k)^p\right] \leq \frac{p}{p-1} \|X_n\|_p \left\|X_n^* \wedge k\right\|_p^{p-1}$$

Rearranging gives

$$\left\|X_n^* \wedge k\right\|_p \leq \frac{p}{p-1} \|X_n\|_p$$

Taking $k \to \infty$ and using monotone convergence finishes the proof. ♡

Armed with these two results, we are now ready to state and prove the goal of this section:

Theorem 2.29 ($\mathscr{L}^p$ convergence of Martingales) Let X be a Martingale and $p > 1$. Then the following are equivalent:

- X is $\mathscr{L}^p$ bounded. I.e:

$$\sup_n \|X_n\|_p < \infty$$

- X converges almost surely and in $\mathscr{L}^p$ to some X_∞
- There exists some $Z \in \mathscr{L}^p(\mathscr{F})$ such that almost surely:

$$X_n = \mathbf{E}[Z \mid \mathscr{F}_n]$$

Main idea: By Jensen, $\mathscr{L}^p$-boundedness implies $\mathscr{L}^1$-boundedness, so we can extract a limit almost surely. Then prove convergence in $\mathscr{L}^p$ to this same object by using $\mathscr{L}^p$-Inequality, and the fact that $|X_n - X_\infty| \leq |X_n| + |X_\infty| \leq 2X_\infty^*$ along with the DCT.

Proof of $1 \implies 2$. If X is bounded in $\mathscr{L}^p$, then by a straightforward application of Jensen's inequality, it follows that X is $\mathscr{L}^1$ bounded, and so by the Martingale Convergence Theorem we have that X converges almost surely to an almost surely finite limit X_∞. Now we prove convergence in $\mathscr{L}^p$.

By Doob's $\mathscr{L}^p$ inequality, we have that

$$\left\|X_n^*\right\|_p \leq \frac{p}{p-1} \|X_n\| - p$$

By using the MCT on the fact that $X_n^* \uparrow X_\infty^*$, we get the bound that

$$\left\|X_\infty^*\right\|_p \le \frac{p}{p-1} \sup_{n\ge 0} \|X_n\|_p < \infty$$

Therefore, noting that $|X_n - X_\infty| \le 2X_\infty^* \in \mathscr{L}^p$, and the DCT we get that $X_n \to X_\infty$ in $\mathscr{L}^p$. ♡

Main idea: Use the Martingale property of X to show that for $Z = X_\infty$, the $\mathscr{L}^p$ distance between X_n and $\mathbf{E}[X_\infty \mid \mathscr{F}_n]$ is zero, and hence they are the same thing almost surely.

Proof of $2 \implies 3$. We claim that $X_n = \mathbf{E}[X_\infty \mid \mathscr{F}_n]$, where X_∞ is the $\mathscr{L}^p$ random variable to which X converges to by hypothesis. Then, using the Martingale property (wlog $m \ge n$):

$$\begin{aligned}\|X_n - \mathbf{E}[X_\infty \mid \mathscr{F}_n]\|_p &= \|\mathbf{E}[X_m - X_\infty \mid \mathscr{F}_n]\|_p \\ &\le \|X_m - X_\infty\|_p \to 0 \qquad (m \to \infty)\end{aligned}$$

Therefore $\|X_n - \mathbf{E}[X_\infty \mid \mathscr{F}_n]\|_p = 0$ and so the two random variables coincide almost surely by definition of the $\mathscr{L}^p$ space. ♡

Main idea: Conditional Jensen's Inequality

Proof of $3 \implies 1$. First of course, we note that $\|Z\|_p < \infty$. Now:

$$\begin{aligned}\|X_n\|_p^p &= \mathbf{E}[|X_n|^p] \\ &= \mathbf{E}[|\mathbf{E}[Z|\mathscr{F}_n]|^p] \\ &\le \mathbf{E}[\mathbf{E}[|Z|^p \mid \mathscr{F}_n]] \\ &= \mathbf{E}[|Z|^p] = \|Z\|_p^p\end{aligned}$$

Where we have used Conditional Jensen's Inequality on the convex function $x \mapsto |x|^p$ and the Tower Property. This finishes the proof noting that the right hand side of the bound is independent of n. ♡

Remark 2.30 A Martingale $X = (X_n)$ of the form $X_n = \mathbf{E}[Z \mid \mathscr{F}_n]$ for $Z \in \mathscr{L}^p(\mathscr{F})$ is said to be a Martingale closed in $\mathscr{L}^p$.

In the case where the Martingale X is closed in $\mathscr{L}^p$, we can also tell the form of X_∞:

Corollary 2.31 Let $Z \in \mathscr{L}^p$ and let $X_n = \mathbf{E}[Z \mid \mathscr{F}_n]$ be a Martingale closed in $\mathscr{L}^p$. Then we have

$$X_n \to X_\infty = \mathbf{E}[Z \mid \mathscr{F}_\infty] \qquad \text{a.s and in } \mathscr{L}^p$$

Where just as a reminder, $\mathscr{F}_\infty = \sigma(\mathscr{F}_n : n \geq 0)$.

Main idea: The convergence to X_∞ is guaranteed by the previous Theorem. One just needs to show that $X_\infty = \mathbf{E}[Z \mid \mathscr{F}_\infty]$ almost surely. For this, just note that $\bigcup_n \mathscr{F}_n$ is a π-system generating $\mathscr{F}_\infty$.

Proof. Let us show that $X_\infty = \mathbf{E}[Z \mid \mathscr{F}_\infty]$. For this, we let $A \in \bigcup_n \mathscr{F}_n$, which is a π-system generating $\mathscr{F}_\infty$. We now show that $\mathbf{E}[X_\infty \mathbf{1}(A)] = \mathbf{E}[X_\infty \mathbf{1}(A)]$, and as such we will get the desired equality. Before doing the calculation, just note that if $A \in \bigcup_n \mathscr{F}_n$, then $A \in \mathscr{F}_N$ for some N. So for all $n \geq N$

$$\begin{aligned} \mathbf{E}[X_\infty \mathbf{1}(A)] &= \mathbf{E}\left[\lim_{n\to\infty} X_n \mathbf{1}(A)\right] \\ &\overset{(1)}{=} \lim_{n\to\infty} \mathbf{E}[X_n \mathbf{1}(A)] \\ &\overset{(2)}{=} \mathbf{E}[Z \mathbf{1}(A)] \end{aligned}$$

Where (1) comes from the fact that $\mathscr{L}^p$ convergence allows us to swap limit and integral, and (2) comes from the fact that since $n \geq N$, then A is also contained in $\mathscr{F}_n$, and so we may use the definition of $X_n = \mathbf{E}[Z \mid \mathscr{F}_n]$ to obtain the last equality. ♡

2.6 UI Martingales

Definition 2.32 (Uniform integrability) Let $(X_\alpha)_{\alpha \in A}$ be a family of random variables. We say (X_α) is uniformly integrable if given any $\epsilon > 0$, there exists some $K > 0$ such that

$$\sup_{\alpha \in A} \mathbf{E}[|X_\alpha| \mathbf{1}\{|X_\alpha| > K\}] < \epsilon.$$

An equivalent definition that perhaps fits more with the definition is (X_i) is UI, if it is bounded in $\mathscr{L}^1$ and for any $\epsilon > 0$, there exists some $\delta > 0$ such that whenever an event A has $\mathbf{P}[A] < \delta$, then $\sup_i \mathbf{E}[|X_i| \mathbf{1}(A)] < \epsilon$.

I.e: a family of random variables is uniformly integrable if the contribution of the tails of each random variable can be made uniformly small across the entire family. Recall that the main point of wanting uniform integrability is that it allows us to pass certain limits inside expectations.

Theorem 2.33 Let X and (X_n) be random variables. Then the following are equivalent:

1. $X_n, X \in L^1$ for all n, and $X_n \to X$ in L^1, i.e: $\mathbf{E}[|X_n - X|] \to 0$
2. (X_n) is uniformly integrable, and $X_n \to X$ in probability.

Proof. Consult [Wil14, Page 131] ♡

We now have a result about uniform integrability in the context of conditional expectation

Theorem 2.34 (UI Property of Conditional Expectation) Let $X \in L^1$, then the family

$$\{\mathbf{E}[X \mid \mathscr{G}] : \mathscr{G} \text{ is a sub-}\sigma\text{-algebra of } \mathscr{F}\}$$

is uniformly integrable.

To prove this Theorem we first need this Lemma:

Lemma 2.35 (Absolute continuity property) Suppose $X \in L^1$, then for any $\epsilon > 0$ there exists some $\delta > 0$ such that whenever $A \in \mathscr{F}$ is such that $\mathbf{P}(A) < \delta$, one has $\mathbf{E}[|X|\mathbf{1}_A] < \epsilon$

Proof of Lemma 2.35. Suppose to the contrary that there exists an $\epsilon_0 > 0$ such that we may find a sequence (F_n) of events with decreasing probabilities, $\mathbf{P}(F_n) = 2^{-n}$, yet $\mathbf{E}[|X|\mathbf{1}_{F_n}] \geq \epsilon_0$. By Borel-Cantelli, since

$$\sum_n \mathbf{P}(F_n) < \infty$$

we have that $\mathbf{P}(\limsup F_n) = 0$. Recall the *Reverse Fatou Lemma*: if $(f_n) \in \mathrm{m}\mathscr{F}^+$ is dominated by some integrable $g \in \mathrm{m}\mathscr{F}^+$, then

$$\mathbf{E}[\limsup f_n] \geq \limsup \mathbf{E}[f_n]$$

Thus we would have

$$\mathbf{E}[\limsup |X|\mathbf{1}_{F_n}] = \mathbf{E}[|X|\mathbf{1}_{\limsup F_n}] \geq \limsup \mathbf{E}[|X|\mathbf{1}_{F_n}] \geq \epsilon_0$$

But $\mathbf{E}[|X|\mathbf{1}_{\limsup F_n}] = 0$. For a justification on the first equality, see 8.7. ♡

Main idea: Set K large enough so that $K^{-1}\mathbf{E}|X| < \delta$, then show that $\mathbf{P}[|\mathbf{E}[X \mid \mathscr{G}]| > K] < \delta$ and use the absolute continuity property.

Proof of Theorem 2.34. Let $\epsilon > 0$ be given and as per Lemma 2.35, choose some $\delta > 0$ so that whenever an event F has $\mathbf{P}(F) < \delta$, we have $\mathbf{E}[|X|\mathbf{1}_F] < \epsilon$. Now choose K large enough so that $K^{-1}\mathbf{E}[|X|] < \delta$. Let $\mathcal{G}$ be any sub-σ-algebra of $\mathcal{F}$. For simplicity let Y be a version of $\mathbf{E}[X \mid \mathcal{G}]$, then by Jensen's inequality

$$|Y| \leq \mathbf{E}[|X| \mid \mathcal{G}]$$

And as such, by the Tower Law:

$$\mathbf{E}[|Y|] \leq \mathbf{E}[|X|]$$

And as such we have the chain of inequalities

$$K\mathbf{P}(\{|Y| > K\}) \leq \mathbf{E}[|Y|] \leq \mathbf{E}[|X|]$$

Therefore $\mathbf{P}(|Y| > K) < \delta$. And as such, we note that

$$\mathbf{E}[|Y|\mathbf{1}_{\{|Y|>K\}}] \leq \mathbf{E}[|X|\mathbf{1}_{\{|Y|>K\}}] < \epsilon$$

♡

Theorem 2.36 (UI Martingale Convergence Theorem) Let X be a Martingale. The following are equivalent:

- X is a UI family.
- X_n converges almost surely and in $\mathcal{L}^1$ to a limit in X_∞.
- There exists some $Z \in \mathcal{L}^1$ such that $X_n = \mathbf{E}[Z \mid \mathcal{F}_n]$

Proof of $1 \implies 2$. Since X is UI, by definition it is bounded in $\mathcal{L}^1$, and so we have by the standard Martingale Convergence Theorem, that it convergence almost surely to an almost surely finite limit X_∞. By Theorem 2.33 we also have convergence in $\mathcal{L}^1$. ♡

Main idea: Prove that the $\mathcal{L}^1$ distance is zero. To do this employ Martingale property and the "non-expansive" corollary of conditional Jensen.

Proof of $2 \implies 3$. The proof is identical to a previous argument. We use the Martingale property and the hypothesis of convergence in $\mathcal{L}^1$ to a limit X_∞. For any $m \geq n$, we have that

$$\|X_n - \mathbf{E}[X_\infty \mid \mathcal{F}_n]\|_1 = \|\mathbf{E}[X_m - X_\infty \mid \mathcal{F}_n]\|_1 \leq \|X_m - X_\infty\|_1 \to 0 \quad (m \to \infty)$$

And so $\|X_n - \mathbf{E}[X_\infty \mid \mathscr{F}_n]\| = 0$ which means they are almost surely equal. ♡

Main idea: UI Property of conditional expectation.

Proof of $3 \implies 1$. We have shown that $\{\mathbf{E}[X \mid \mathscr{G}] : \mathscr{G} \subseteq \mathscr{F}\}$ is a UI family, so if we have the sub-family where the sigma algebras are $\{\mathscr{F}_n\}$, the result still holds. ♡

Remark 2.37 If X is UI Martingale, then $X_\infty = \mathbf{E}[Z \mid \mathscr{F}_\infty]$. If X is a superMartingale, then $\mathbf{E}[X_\infty \mid \mathscr{F}_n] \le X_n$ (Respectively for subMartingales).

Theorem 2.38 (Optional Stopping for UI Martingales) Let X be a UI Martingale and S and T be stopping times with $S \le T$. Then

$$\mathbf{E}[X_T \mid \mathscr{F}_S] = X_S$$

Remark 2.39 (Reality check) Note that we have not required S and T to be bounded. To make sense of X_T and X_S, we use the hypothesis of UI and simply set

$$X_T = \sum_{n \ge 0} X_n \mathbf{1}(T = n) + X_\infty \mathbf{1}(T = \infty)$$

where $X_\infty = \lim_n X_n$.

Main idea: Show that for any stopping time T we have that $\mathbf{E}[X_\infty \mid \mathscr{F}_T] = X_T$. Then use this and the tower property to extend to the general case.

Proof. To show that $\mathbf{E}[X_\infty \mid \mathscr{F}_T] = X_T$, we do to things:

- Check X_T is integrable: we first note that $\mathbf{E}[|X_\infty| \mid \mathscr{F}_n] \ge |\mathbf{E}[X_\infty \mid \mathscr{F}_n]| = |X_n|$, which in turn implies that for any $A \in \mathscr{F}_n$, $\mathbf{E}[|X_n| \mathbf{1}(A)] \le \mathbf{E}[|X_\infty| \mathbf{1}(A)]$. Armed with this:

$$\begin{aligned}
\mathbf{E}[|X_T|] &= \sum_{n \ge 0} \mathbf{E}[|X_n| \mathbf{1}(T = n)] + \mathbf{E}[|X_\infty| \mathbf{1}(T = \infty)] \\
&\le \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbf{E}[|X_\infty| \mathbf{1}(T = n)] \\
&= \mathbf{E}[|X_\infty|] < \infty
\end{aligned}$$

  Because X being UI means in particular that $X_\infty \in \mathscr{L}^1$.

- Check the conditional expectation property. Let $B \in \mathcal{F}_T$. Then

$$\begin{aligned}\mathbf{E}[\mathbf{1}(B)X_T] &= \sum_{n\in\mathbf{N}\cup\{\infty\}} \mathbf{E}[\mathbf{1}(\underbrace{B\cap\{T=n\}}_{\in\mathcal{F}_n})X_n] \\ &\overset{(!)}{=} \sum_{n\in\mathbf{N}\cup\{\infty\}} \mathbf{E}[\mathbf{1}(B\cap\{T=n\})X_\infty] \\ &= \mathbf{E}[\mathbf{1}(B)X_\infty]\end{aligned}$$

Where step (!) comes from the fact that $X_n = \mathbf{E}[X_\infty \mid \mathcal{F}_n]$ almost surely (UI MG Convergence Theorem). Now that we have established the Optional Stopping Property for X_∞, we extend this to the general case as follows quite easily:

$$\begin{aligned}\mathbf{E}[X_T \mid \mathcal{F}_S] &= \mathbf{E}[\mathbf{E}[X_\infty \mid \mathcal{F}_T] \mid \mathcal{F}_S] \\ &\overset{(!)}{=} \mathbf{E}[X_\infty \mid \mathcal{F}_S] \\ &= X_S\end{aligned}$$

Where we have used the property proved above in the first and third steps, and on step (!) we have used the Tower Law.

♡

2.7 Backwards Martingales

Definition 2.40 Let $\cdots \subseteq \mathcal{G}_{-2} \subseteq \mathcal{G}_{-1} \subseteq \mathcal{G}_0$ be a decreasing family of σ-algebras. An integrable process $(X_n)_{n\leq 0}$ adapted to this filtration is called a backwards Martingale if

$$\mathbf{E}[X_{n+1} \mid \mathcal{G}_n] = X_n \qquad n \leq -1$$

A different way to think about this (I don't like the minus signs, is that if you have a decreasing sequence of sigma algebras $(\mathcal{G}_n)_{n\geq 0}$, an integrable process $\{X_n\}_{n\geq 0}$ is a backwards Martingale if

$$\mathbf{E}[X_n \mid \mathcal{G}_{n+1}] = X_{n+1}$$

which when contrasted with the usual Martingale property:

$$\mathbf{E}[X_n \mid \mathcal{F}_{n-1}] = X_{n-1}$$

gives perhaps a clearer vision of what this means.

Remark 2.41 By iterating the tower property we have that $\mathbf{E}[X_0 \mid \mathscr{G}_n] = X_n$, and since $X_0 \in \mathscr{L}^1$ by assumption we automatically get that X_n is a uniformly integrable family.

Theorem 2.42 (Convergence of Backwards Martingales) Let X be a backwards Martingale with $X_0 \in \mathscr{L}^p$ for some $p \in [1, \infty)$. Then $X_n \to X_{-\infty}$ as $n \to -\infty$ to the random variable $X_{-\infty} = \mathbf{E}[X_0 \mid \mathscr{G}_{-\infty}]$ almost surely and in $\mathscr{L}^p$, where $\mathscr{G}_{-\infty} = \bigcap_{n \le 0} \mathscr{G}_n$

Main idea: Adapt Doob's Upcrossing Lemma to work backwards in time and then just delegate the rest of the proof to the usual Almost Sure MG Convergence Theorem, and then by showing Uniform Integrability of the Martingale, upgrade the convergence to $\mathscr{L}^p$.

Proof. To extract a limiting random variable and its almost sure convergence, we essentially replay the proof for the Almost Sure MG Convergence Theorem, and all we need to do is justify that Doob's Upcrossing Lemma works backwards in time. To do this, we simply note that for a fixed n, the process (X_{-n+k}) is a usual Martingale with respect to the filtration $\mathscr{F}_k = \mathscr{G}_{-n+k}$ and so the usual Doob's Upcrossing Lemma works:

$$\mathbf{E}[U_{-n}([a,b],X)] \le \frac{1}{b-a}\mathbf{E}[(X_0 - a)^-]$$

Then passing the limit $n \to \infty$ (noting that the RHS of the bound does not depend on n), we have that

$$\mathbf{E}[\text{total upcrossings from a to b}] < \infty$$

and so one may call upon the proof of the A.S MG Convergence Theorem to finish showing that $X_m \to X_{-\infty}$ as $m \to -\infty$ a.s, where $X_{-\infty} \in m\mathscr{G}_{-\infty}$.

To upgrade this to $\mathscr{L}^p$ convergence. We note that since $X_n = \mathbf{E}[X_0 \mid \mathscr{G}_n]$ and $X_0 \in \mathscr{L}^p$, then $X_n \in \mathscr{L}^p$ and by a usual Fatou Lemma argument we have that $X_{-\infty} \in \mathscr{L}^p$. Now consider the sequence $\{|X_n - X_{-\infty}|^p\}_n$. Since we have that

$$|X_n - X_{-\infty}|^p = |\mathbf{E}[X_0 - X_{-\infty} \mid \mathscr{G}_n]| \le \mathbf{E}[|X_0 - X_{-\infty}|^p \mid \mathscr{G}_n]$$

and the latter is UI by the UI Property of Conditional Expectation, it follows that the sequence $\{|X_n - X_{-\infty}|^p\}_n$ is UI, and since it converges a.s to 0, it will also converge in $\mathscr{L}^1$ to zero (this is one of the equivalent characterisations of UI), which means that

$$\mathbf{E}[|X_n - X_\infty|^p] \to 0$$

showing the $\mathscr{L}^p$ convergence we wanted.

To show the final part of the Theorem, namely that $X_{-\infty} = \mathbf{E}[X_0 \mid \mathscr{G}_{-\infty}]$, we simply let $A \in \mathscr{G}_{-\infty}$, which by decreasingness means that $A \in \mathscr{G}_n$ for all $n \leq 0$. And so by the Martingale Property on X, we have that

$$\mathbf{E}[X_0 \mathbf{1}(A)] = \mathbf{E}[X_n \mathbf{1}(A)]$$

and by $\mathscr{L}^1$ convergence to $X_{-\infty}$ we may pass the limit $n \to -\infty$ and finish the proof. ♡

2.8 Application of Martingales

Theorem 2.43 (Kolmogorov 0−1 Law) Let (X_i) be IID random variables, and letting $\mathscr{F}_n = \sigma(X_k : k \geq n)$, and $\mathscr{F}_\infty = \bigcap_n \mathscr{F}_n$. Then if $A \in \mathscr{F}_\infty$, we have that $\mathbf{P}[A] \in \{0, 1\}$.

Main idea: Let $A \in \mathscr{F}_\infty$ and letting $\mathscr{G}_n = \sigma(X_k : k \leq n)$, consider the Martingale $\mathbf{E}[\mathbf{1}(A) \mid \mathscr{G}_n]$. Using UI Martingale Convergence Theorem, and the fact that $\mathscr{F}_\infty \subseteq \mathscr{G}_\infty$ finish the claim.

Proof. Consider the sigma algebra $\mathscr{G}_n = \sigma(X_k : k \leq n)$ generated by the history of the process up to time n. This sigma algebra is clearly independent from the future $\mathscr{F}_{n+1}$ because $\{X_i\}$ are independent. And so in particular, if we let $A \in \mathscr{F}_\infty$, the random variable $\mathbf{1}(A)$ is independent of the sigma algebra $\mathscr{G}_n$ for any n. From this we gather that

$$\mathbf{E}[\mathbf{1}(A) \mid \mathscr{G}_n] = \mathbf{P}[A]$$

Moreover, since the process $\{\mathbf{E}[\mathbf{1}(A) \mid \mathscr{G}_n]\}_n$ is a Martingale and also UI by the UI Property of conditional expectation, it follows that it converges almost surely to $\mathbf{E}[\mathbf{1}(A) \mid \mathscr{G}_\infty]$, where $\mathscr{G}_\infty = \bigwedge_n \mathscr{G}_n := \sigma(G_k : k \geq 0)$. However, it is also clear that $\mathscr{F}_\infty \subseteq \mathscr{G}_\infty$, and so $\mathbf{1}(A) \in m\mathscr{G}_\infty$, which means that $\mathbf{E}[\mathbf{1}(A) \mid \mathscr{G}_\infty] = \mathbf{1}(A)$. Putting all this together:

$$\begin{aligned} \mathbf{1}(A) &= \mathbf{E}[\mathbf{1}(A) \mid \mathscr{G}_\infty] \\ &= \lim_{n \to \infty} \mathbf{E}[\mathbf{1}(A) \mid \mathscr{G}_n] \quad a.s \\ &= \mathbf{P}[A] \end{aligned}$$

Thus $\mathbf{P}[A] \in \{0, 1\}$. ♡

Theorem 2.44 (Strong Law of Large Numbers) Let $\{X_i\}$ be IID integrable random variables and let μ be their common mean. Then letting $S_n = X_1 + \cdots + X_n$ gives that

$$\frac{S_n}{n} \to \mu \qquad \text{almost surely and in } \mathscr{L}^1$$

Main idea: The proof is via backwards Martingales. The decreasing sequence of sigma algebras is $\{\mathscr{G}_n\}$ with $\mathscr{G}_n = \sigma(S_k : k \geq n)$. The backwards Martingale is just the process $\{S_n/n\}_{n\geq 0}$. Once it is shown that this is indeed a backwards Martingale, we have convergence a.s and in $\mathscr{L}^1$ to some random variable Y. But then by index shifting in the sum, it can be shown that Y is measurable with respect to a tail sigma algebra and so by Kolmogorov, Y must be constant, and from here computing the value of Y is straightforward.

Proof. We start by constructing the following decreasing sequence of sigma algebras: $\mathscr{G}_n = \sigma(S_k : k \geq n) = \sigma(S_n, S_{n+1}, \cdots) = \sigma(S_n, X_{n+1}, X_{n+2}, \cdots)$. Now we claim that the process $\{S_n/n\}_{n\geq 0}$ is a backwards Martingale. To show this we need to prove that

$$\mathbf{E}\left[\frac{S_{n-1}}{n-1}\middle|\mathscr{G}_n\right] = \frac{S_n}{n}$$

Its a simple calculation:

$$\begin{aligned}\mathbf{E}\left[\frac{S_{n-1}}{n-1}\middle|\mathscr{G}_n\right] &= \mathbf{E}\left[\frac{S_n - X_n}{n-1}\middle|\mathscr{G}_n\right] \\ &= \frac{S_n}{n-1} - \frac{1}{n-1}\mathbf{E}[X_n \mid \mathscr{G}_n]\end{aligned}$$

and since $\mathbf{E}[X_n \mid \mathscr{G}_n] = \mathbf{E}[X_n \mid \sigma(S_n, X_{n+1}, \cdots)] = \mathbf{E}[X_n \mid S_n]$ by the fact that X_n is independent to X_k for $k > n$, and by symmetry $\mathbf{E}[X_n \mid S_n] = \mathbf{E}[X_k]$ for $k \in \{1, \cdots, n\}$, we conclude that

$$S_n = \mathbf{E}[S_n \mid S_n] = \mathbf{E}[X_1 \mid S_n] + \cdots + \mathbf{E}[X_n \mid S_n] = n\mathbf{E}[X_n \mid S_n]$$

and so $\mathbf{E}[X_n \mid S_n] = \frac{S_n}{n}$. Plugging this into our previous computation gives that

$$\mathbf{E}\left[\frac{S_{n-1}}{n-1}\middle|\mathscr{G}_n\right] = \frac{S_n}{n-1} + \frac{S_n}{n(n-1)} = \frac{S_n}{n}$$

as required. Now we have guaranteed by the backwards Martingale convergence theorem the convergence, almost surely and in $\mathscr{L}^1$, to a random variable Y. However, by shifting the indices (!), we see that for all $k \geq 0$

$$Y = \lim_{n\to\infty} \frac{S_n}{n} \overset{(!)}{=} \lim_{n\to\infty} \frac{X_{k+1} + \cdots + X_{k+n}}{n}$$

we see that Y is measurable with respect to $\mathscr{T}_k = \sigma(X_{k+1}, X_{k+2}, \cdots)$ for all k, i.e: its measurable with respect to the tail sigma algebra $\mathscr{T} = \bigcap_k \mathscr{T}_k$, and so by Kolmogorov's $0-1$ law, it Y must be constant, because $\mathscr{T}$ is a trivial sigma algebra. But now that we have determined that Y is constant, we have of course that $Y = \mathbf{E}[Y]$, but therefore:

$$\begin{aligned} Y &= \mathbf{E}[Y] \\ &= \mathbf{E}\left[\lim_{n\to\infty} \frac{S_n}{n}\right] \\ &\overset{(!)}{=} \lim_{n\to\infty} \mathbf{E}\left[\frac{S_n}{n}\right] \\ &= \mu \end{aligned}$$

Where the key step, (!), comes from the fact that the convergence $S_n/n \to 1$ is not only almost surely, but in $\mathscr{L}^1$. This finishes the claim. ♡

Example 2.45 Let us see an example related to this proof. We are going to show that if $(X_n : n \geq 1)$ is a sequence of i.i.d $\mathscr{L}^p$ for some $p > 1$ random variables with mean μ, then

$$\sup_{m\geq n} \left|\frac{S_m}{m} - \mu\right| \to 0 \qquad \text{in } \mathscr{L}^p.$$

Proof. At first glance one needs to apply some sort of Doob's $\mathscr{L}^p$ inequality, which tells us that if (Z_n) is a Martingale, with some integrability conditions, then $\left\|\sup_{0\leq t\leq n} |X_t|\right\|_p \leq \frac{p}{p-1} \|X_n\|_p$. However, in this case $\left(\frac{S_m}{m} - \mu\right)$ is not a Martingale per-se, but it is indeed a backwards Martingale with respect to the backwards filtration $\mathscr{G}_n = \sigma(S_n, S_{n+1}, \cdots)$, meaning that $\mathbf{E}[S_m/m - \mu \mid \mathscr{G}_{m+1}] = S_{m+1}/m+1-\mu$. We can however exploit this to turn $\left(\frac{S_m}{m} - \mu\right)$ into a usual Martingale by switching up the order of the filtration. Fix K and n such that $K > n$, and consider the now increasing filtration $\mathscr{G}_K \subseteq \mathscr{G}_{K-1} \subseteq \cdots \subseteq \mathscr{G}_n$, and it is clear then that

$$\left(\frac{S_{-m}}{-m} - \mu\right)_{-K\leq m\leq -n}$$

is now a usual Martingale with respect to this increasing filtration. Therefore, Doob's $\mathscr{L}^p$ inequality tells us that

$$\left\|\sup_{-K\leq t\leq -n} \left|\frac{S_{-m}}{-m} - \mu\right|\right\|_p \leq \frac{p}{p-1} \left\|\frac{S_n}{n} - \mu\right\|_p.$$

Now since the integrand on the LHS is clearly increasing in K, we can take $K \to \infty$ and by the MCT, we see that

$$\left\|\sup_{t\leq -n} \left|\frac{S_{-t}}{-t} - \mu\right|\right\|_p = \left\|\sup_{t\geq n} \left|\frac{S_t}{t} - \mu\right|\right\|_p \leq \frac{p}{p-1} \left\|\frac{S_n}{n} - \mu\right\|_p.$$

However, since we are assuming that the X_i's are all $\mathscr{L}^p$, the $\mathscr{L}^1$ convergence of the SLLN gets upgraded to $\mathscr{L}^p$, which means that the right hand side of the above expression indeed converges to zero as $n \to \infty$. ♡

Remark 2.46 The moral of the story here is that if one has a Backwards Martingale, by switching up the direction of the filtration one recovers a usual Martingale, although with a finite time horizon, and from that one can employ usual results from Martingale theory.

Theorem 2.47 (Radon-Nikodym) Let $\mathbf{P}$ and $\mathbf{Q}$ be two probability measures on $(\Omega, \mathscr{F})$. Assume $\mathscr{F}$ is countably generated, i.e: there is a countable collection $\{F_n : n \in \mathbf{N}\}$ with $\mathscr{F} = \sigma(F_n : n \in \mathbf{N})$. Then the following are equivalent

1. $\mathbf{P}[A] = 0$ implies $\mathbf{Q}[A] = 0$. (This is referred to as $\mathbf{Q}$ being absolutely continuous with respect to $\mathbf{P}$, and we write $\mathbf{Q} \ll \mathbf{P}$)
2. For all $\epsilon > 0$, there is a $\delta > 0$ so that whenever $\mathbf{P}[A] \leq \delta$, then $\mathbf{Q}[A] \leq \epsilon$.
3. There exists an almost surely unique non-negative random variable X so that

$$Q[A] = \int_\Omega X\,\mathbf{1}(A)\,d\mathbf{P} :=: \mathbf{E}_\mathbf{P}[X\,\mathbf{1}(A)]$$

Remark 2.48 The random variable X is called a version of the Radon-Nikodym derivate of $\mathbf{Q}$ with respect to $\mathbf{P}$, denoted as $X = \frac{d\mathbf{Q}}{d\mathbf{P}}$.

Main idea: By contradiction: construct a sequence $\{A_n\}$ with $\mathbf{P}[A_n] \leq 1/n^2$ yet $\mathbf{Q}[A_n] > \epsilon$. Then use Borel-Cantelli and the definition of infinitely-often.

Proof of $1 \implies 2$. Assume (2) does not hold, i.e: there exists some $\epsilon > 0$, such that for all $n \geq 1$, there exists a set A_n with $\mathbf{P}[A_n] \leq 1/n^2$ and $Q[A_n] > \epsilon$. Then by the Borel-Cantelli lemma, we have that

$$\mathbf{P}[\{A_n \text{ i.o}\}] = 0$$

Therefore, since the event $\{A_n \text{ i.o}\}$ as zero $\mathbf{P}$-measure, by hypothesis, it will also have zero $\mathbf{Q}$-measure, so $\mathbf{Q}[\{A_n \text{ i.o}\}] = 0$. However, by definition:

$$\mathbf{Q}[A_n \text{ i.o}] = \mathbf{Q}\left[\bigcap_n \bigcup_{k \geq n}\right] = \lim_{n \to \infty} \mathbf{Q}\left[\bigcup_{k \geq n} A_k\right] \geq \epsilon$$

Contradicting the fact that $\mathbf{Q}[\{A_n \text{ i.o}\}]=0$. ♡

Main idea: For $2 \implies 3$, the idea is to construct a Martingale $\{X_n\}$ as follows:

1. Use the filtration $\{\mathscr{F}_n\}$ to be $\mathscr{F}_n=\sigma(F_k : k \le n)$. Then define $\mathscr{A}_n=\{H_1\cap\cdots\cap H_n : H_i=F_i \text{ or } F_i^c\}$. Then note that $\mathscr{F}_n=\sigma(\mathscr{A}_n)$. Now define

$$X_n(\omega)=\sum_{B\in\mathscr{A}_n}\frac{\mathbf{Q}[B]}{\mathbf{P}[B]}\mathbf{1}(\omega\in B)$$

and show this is an $\mathscr{F}_n$-Martingale using the fact that sets in $\mathscr{A}_n$ are disjoint.

2. Show $\{X_n\}$ is $\mathscr{L}^1$ bounded, so we have almost sure convergence to some X.

3. Show, using the hypothesis, that $\{X_n\}$ is actually a UI family, and so we actually have $\mathscr{L}^1$ convergence.

4. Show that $\tilde{Q}[A]=\mathbf{E}[X\,\mathbf{1}(A)]$ agrees with $Q[A]$ for all sets A in the π-system of sets like $\bigcup_n\mathscr{F}_n$ that generate $\mathscr{F}$.

Proof. The proof is by UI Martingales. The filtration we will use is $\mathscr{F}_n=\sigma(F_k : k\le n)$. Then define $\mathscr{A}_n=\{H_1\cap\cdots\cap H_n : H_i=F_i \text{ or } F_i^c\}$ and set

$$X_n(\omega)=\sum_{B\in\mathscr{A}_n}\frac{\mathbf{Q}[B]}{\mathbf{P}[B]}\mathbf{1}(\omega\in B)$$

We make the following two observations: the first, and easier one, is that sets in $\mathscr{A}_n$ are pairwise disjoint. The second observation is that $\mathscr{F}_n=\sigma(\mathscr{A}_n)$. A way to see this is to note that we can recover the F_k's from $\mathscr{A}_n$ by taking unions. As a simple example, if $n=2$, then $F_1=(F_1\cap F_2)\cup(F_1\cap F_2^c)$. From this we see that in fact, any set $A\in\mathscr{F}_n$ can be written as a union of sets in $\mathscr{A}_n$, say $A=\bigcup_i S_i$ for some $S_i\in\mathscr{A}_n$. Therefore, it is now easy to see that by the fact that sets in A_n are pairwise disjoint:

$$\mathbf{E}[X_n\,\mathbf{1}(A)]=\mathbf{E}\left[\sum_{B\in\mathscr{A}_n}\sum_i\frac{\mathbf{Q}[B]}{\mathbf{P}[B]}\mathbf{1}(S_i\cap B)\right]=\mathbf{E}\left[\sum_i\frac{\mathbf{Q}[S_i]}{\mathbf{P}[S_i]}\mathbf{1}(S_i)\right]=\mathbf{Q}[A]$$

Having constructed $\{X_n\}$ and having checked this preliminary fact, we now move on to check that $\{X_n\}$ is in fact an $\mathscr{F}_n$-Martingale. And it will be quite easy. Indeed: if $A\in\mathscr{F}_n$, then by filtration property, $A\in\mathscr{F}_{n+1}$ and so

$$\mathbf{E}[X_{n+1}\,\mathbf{1}(A)]=\mathbf{Q}[A]=\mathbf{E}[X_n\,\mathbf{1}(A)]$$

Showing the Martingale property. Now we set off to show that this Martingale converges to something, and we will first extract almost sure convergence by showing $\{X_n\}$ is $\mathscr{L}^1$-bounded.

This is easy because for all n:

$$\mathbf{E}[X_n] = \mathbf{E}[X_n \mathbf{1}(\Omega)] = \mathbf{Q}[\Omega] = 1$$

and so $\{X_n\}$ is $\mathscr{L}^1$ bounded (we have also implicitly used that X_n is non-negative to skip writing the absolute value signs). We call the limiting random variable X_∞. Recall that our goal is to show that for any $A \in \mathscr{F}$, we have that $\mathbf{Q}[A] = \mathbf{E}[X_\infty \mathbf{1}(A)]$, and to do so, we note that the set of unions $\bigcup_n \mathscr{F}_n$ is a π-system generating $\mathscr{F}$, so our goal will be to show that for a fixed n, whenever $A \in \mathscr{F}_n$, we have that

$$Q[A] = \mathbf{E}[X_n \mathbf{1}(A)] = \mathbf{E}[X_\infty \mathbf{1}(A)]$$

Intuitively we see that we are going to have to pass some limits so we are now motivated to upgrade almost sure convergence to $\mathscr{L}^1$ convergence. To do so, we will finally employ the hypothesis.

To show that $\{X_n\}$ is a UI family (and hence we have $\mathscr{L}^1$ convergence), let $\epsilon > 0$ be given, then by hypothesis there exists some δ, so that whenever $\mathbf{P}[A] \leq \delta$, then $\mathbf{Q}[A] \leq \epsilon$. Set $K = 1/\delta$, then by Markov's inequality:

$$\mathbf{P}[X_n \geq K] \leq \frac{\mathbf{E}[X_n]}{K} = \frac{1}{K} = \delta$$

and so by hypothesis,

$$\mathbf{E}[X_n \mathbf{1}(X_n \geq K)] = \mathbf{Q}[X_n \geq K] \leq \epsilon$$

thus showing the UI property. Now that we have upgraded almost sure convergence to $\mathscr{L}^1$ convergence, we can finish the proof by noting that for an n fixed and $A \in \mathscr{F}_n$,

$$\mathbf{E}[X_\infty \mathbf{1}(A)] = \lim_{m \to \infty} \mathbf{E}[X_m \mathbf{1}(A)] = \mathbf{Q}[A]$$

Because eventually $m \geq n$. ♡

The proof of $3 \implies 1$ is trivial.

3 Continuous-time processes

We can generalise all the definitions we made for discrete-time processes, such as the filtration, adaptability, stopping time, etc. The only difference is that whenever we indexed, say (X_t) with $t \in \mathbf{N}$, we now index (X_t) with $t \in [0, \infty)$. This brings some potential hazards, which we really didn't have to worry about before:

- **Measurability**: In the discrete case, for a fixed t, we obviously had $X_t : \omega \mapsto X_t(\omega)$ being measurable with respect to $\mathscr{F}$, (i.e: the process really was a sequence of random variables) because of the adaptability assumption. This also holds when t is indexed in continuous-time. However, we could also inspect what happens if we keep ω fixed, and look at the map $t \mapsto X_t(\omega)$, this map corresponds to what we think of the process when we "draw it on a graph", and since the index set $\mathbf{N}$ of t, has the σ-algebra $\mathscr{P}(\mathbf{N})$, in where each subset is measurable, we have that the map $t \mapsto X_t(\omega)$ is measurable. However, when we change to continuous-time, it need not be the case that $t \mapsto X_t(\omega)$ is measurable, because now our index set for t is $\mathbf{R}$, which is endowed with the algebra $\mathscr{B}(\mathbf{R})$. We thus have the issue of measurability of the map $(\omega, t) \mapsto X_t(\omega)$ with respect to the σ-algebra $\mathscr{F} \otimes \mathscr{B}(\mathbf{R})$. A toy example is even the deterministic map $X_t = \mathbf{1}_A(t)$ where A is a Vitali set. This is not measurable with respect to $\mathscr{F} \otimes \mathscr{B}(\mathbf{R})$. Whereas if we had been working in discrete time, any subset of $\mathbf{N}$ would have been measurable so we would have not come across this issue.

- **Hitting times**: We also have something worrying, if $A \subseteq \mathbf{R}$, is an arbitrary subset, it is not necessarily the case that
$$T_A = \inf\{t \geq 0 : X_t \in A\}$$
is now a stopping time, indeed, since
$$\{T_A \leq t\} = \bigcup_{0 \leq s \leq t} \{X_s \in A\}$$
and this right hand side is an uncountable union (and A could also not be measurable), then it is not guaranteed that this belongs to $\mathscr{F}_t$.

We will now spend some time fixing these problems. For the first part, we start by introducing some more regularity on the process X:

Definition 3.1 (Continuous / Càdlàg) A continuous-time stochastic process (X_t) is continuous if the map $t \mapsto X_t(\omega)$ is continuous for a fixed ω. For a weaker notion, we say that (X_t) is càdlàg if the map $t \mapsto X_t(\omega)$ is right continuous and has a left limit.

The idea is that in both the continuous and the càdlàg case, the entire process is determined, thanks to continuity, by its values at a countable set of times.

Definition 3.2 (Measurable space of continuous/càdlàg functions) Let E be some metric space. We have the following two sets:

- $C(\mathbf{R}_+, E) = \{f : \mathbf{R}_+ \to E \text{ continuous}\}$.
- $D(\mathbf{R}_+, E) = \{f : \mathbf{R}_+ \to E \text{ càdlàg}\}$.

We endow these spaces with the σ-algebra generated by the coordinate maps $\Pi_t : f \mapsto f(t)$. Thus a continuous/càdlàg process is a random variable $\omega \mapsto C(\mathbf{R}_+, E)/D(\mathbf{R}_+, E)$.

If we impose continuous/càdlàg regularity on our random variable, it turns out that the map $(\omega, t) \mapsto X_t(\omega)$ will be measurable with respect to $\mathscr{F} \otimes \mathscr{B}(\mathbf{R})$. Let us show this for the case of continuity

Proposition 3.3 Let (X_t) be a continuous-time process, then the map $(\omega, t) \mapsto X_t(\omega)$ is be measurable with respect to $\mathscr{F} \otimes \mathscr{B}(\mathbf{R})$.

Main idea: Write $X_t(\omega)$ as a pointwise limit approximating t dyadically.

Proof. We show it for $[0,1]$ as the indexing set for simplicity. The key is to approximate t by a sequence a_n constructed as follows: partition the real line into bins of width 2^{-n}:

$$\cdots < (k-1)2^{-n} < k2^{-n} < (k+1)2^{-n} < \cdots$$

then t will fit into one of these boxes, say $k2^{-n}$ so we set $a_n = k2^{-n}$. It is clear that $a_n \to t$, so by continuity of $X_t(\omega)$ on t, we write

$$X_t(\omega) = \lim_{n\to\infty} \sum_{k=0}^{2^n-1} \underbrace{\mathbf{1}(t \in [k2^{-n}, (k+1)2^{-n}]) X_{k2^{-n}}(\omega)}_{\in \mathscr{B}(\mathbf{R}) \otimes \mathscr{F}}$$

This sum is certainly measurable in $\mathscr{F} \otimes \mathscr{B}(\mathbf{R})$, and since limits preserve measurability we are done. ♡

Now that we have clarified the situation regarding measurability, we turn to look at stopping times. First, let us present an analogous proposition to 2.19 but for continuous-time stopping times, which are defined in an essentially equal way to discrete-time stopping times, and similarly the stopped sigma algebra $\mathscr{F}_T$, is defined in an identical manner. In addition, we define the random variable $X_T(\omega) = X_{T(\omega)}(\omega)$ whenever $T < \infty$, and we also have the stopped process $(X^T)_t = X_{T \wedge t}$.

Proposition 3.4 (Stopping time properties in continuous-time) Let S and T be two stopping-times, and X be càdlàg process. Then

- $S \wedge T$ is a stopping time.
- If $S \leq T$, then $\mathscr{F}_S \subseteq \mathscr{F}_T$.
- $X_T \mathbf{1}(T < \infty)$ is $\mathscr{F}_T$-measurable.
- X^T is adapted.

Main idea: The only substance is in showing 3, and this follows from the fact that Z is $\mathscr{F}_T$ measurable if and only if $Z\mathbf{1}(T \leq t)$ is $\mathscr{F}_t$ measurable for all t which is proven by standard machine type arguments. Then perform a dyadic expansion.

Proof. Parts 1 and 2 are identical in proof to the discrete-time versions due to the definitions being identical. Now we note that 4 follows easily from 3. Indeed: the constant time t is also a stopping time, so $T \wedge t$ is a stopping time and is always finite, so by part 3, $X_{T\wedge t}$ is $\mathscr{F}_{T\wedge t}$ measurable. We now think a bit for 3. The first claim is that Z is $\mathscr{F}_T$ measurable if and only if $Z\mathbf{1}(T \leq t)$ is $\mathscr{F}_t$ measurable for all t.

- ($\implies$). Suppose Z is $\mathscr{F}_T$ measurable. Then we know that writing $Z = Z^+ - Z^-$ allows us to express Z^+ and Z^- as a pointwise limit of indicator functions of sets in $\mathscr{F}_T$. When we multiply through by $\mathbf{1}(T \leq t)$, we are effectively intersecting the sets of each of these indicator functions with the set $\{T \leq t\}$, so by definition of $\mathscr{F}_T$, we have that $Z\mathbf{1}(T \leq t)$ is indeed $\mathscr{F}_t$ measurable.
- ($\impliedby$). We apply the standard machine. First note that if $Z = c\mathbf{1}(A)$ for some $A \in \mathscr{F}$, then $Z\mathbf{1}(T \leq t) = c\mathbf{1}(A \cap \{T \leq t\})$. The hypothesis that this is $\mathscr{F}_t$ measurable means that $A \in \mathscr{F}_T$, so Z is $\mathscr{F}_T$ measurable. Now that we have shown the claim for indicator functions, we pass this into the standard machine.

With this technicality out of the way, we proceed now to prove that $X_T\mathbf{1}(T < \infty)$ is $\mathscr{F}_T$ measurable. By our detour, this is equivalent to showing that $X_T\mathbf{1}(T \leq t)$ is $\mathscr{F}_t$ measurable for all t. We go about this by expanding

$$X_T\mathbf{1}(T \leq t) = X_T\mathbf{1}(T < t) + X_t\mathbf{1}(T = t)$$

Clearly the second summand is $\mathscr{F}_t$ measurable, so we only need to show that $X_T\mathbf{1}(T < t)$ is also $\mathscr{F}_t$ measurable. We first construct a stopping time T_n taking values in the dyadic numbers

$\mathscr{D}_n = \{k2^{-n} : k \in \mathbf{N}\}$, given by

$$T_n = 2^{-n}\lceil 2^n T\rceil.$$

This is indeed a stopping time because

$$\{T_n \le t\} = \{\lceil 2^n T\rceil \le 2^n t\} = \{T \le 2^{-n}\lfloor 2^n t\rfloor\} \in \mathscr{F}_{2^{-n}\lfloor 2^n t\rfloor} \subseteq \mathscr{F}_t$$

Now using the fact that $T_n \downarrow T$, and the fact that X is càdlàg by assumption, gives that $X_{T_n} \to X_T$, but we could also write

$$X_T \mathbf{1}(T < t) = \lim_{n\to\infty} X_{T_n \wedge t}\mathbf{1}(T < t)$$

and using the fact that T_n takes values in the dyadic numbers, we see that

$$X_{T_n \wedge t}\mathbf{1}(T < t) = \sum_{q \in \mathscr{D}_n, q < t} X_q \mathbf{1}\big(T_n = q\big) + X_t \mathbf{1}(T < t < T_n)$$

which in this form is easy to see that it is $\mathscr{F}_t$ measurable, thus finishing the claim.

♡

Before we establish regularity conditions that allow us to conclude that hitting times are also stopping times in the continuous setting, let us give an explicit example of when hitting times may not be stopping times.

Example 3.5 (A hitting time that is not a stopping time) Let J take values ± 1 with equal probability, and define the process X_t to be

$$X_t = \begin{cases} t & 0 \le t \le 1 \\ 1 + J(t-1) & t > 1 \end{cases}$$

(Here J is drawn at the very start, i.e: not updated for each t). Then this process is of course adapted to the natural filtration, but if we let $A = (1,2)$, then the event $\{T_A \le 1\}$ does not belong to $\mathscr{F}_1$, (recall that $T_A = \inf\{t > 0 : X_t \in A\}$, so in the case that $J = 1$, $T_A = 1$, and in the case $J = -1$, $T_A = \infty$) because just by knowing the history of the process up to time 1, we cannot determine whether J is one or minus one. Of course here the problem is that A is open, so it is unsurprising what the first regularity condition that we will inspect is going to be.

Proposition 3.6 (Regularity condition 1 for hitting times to be stopping times) Let A be closed and let X be an adapted continuous process. Then T_A is a stopping time.

Main idea: The events $\{T_A \le t\}$ and $\{\inf_{s \in \mathbf{Q}: s \le t} d(X_s, A) = 0\}$ are the same. The right hand event is in $\mathscr{F}_t$

Proof.

- ($\subseteq$). Suppose that $T_A = s \le t$. Then there exists a sequence of rationals s_n such that $s_n \downarrow s$ and $X_{s_n} \in A$ for all n. Since X is continuous, we have that $X_{s_n} \to X_s$. Since A is closed, we also have that $X_s \in A$. Therefore $X_{T_A} \in A$. We can also find a sequence (q_n) of rationals with $q_n \uparrow T_A$, and since $d(X_{T_A}, A) = 0$, by continuity, we have that $d(X_{q_n}, A) \to 0$. Hence we have that $\inf_{s \in \mathbf{Q}: s \le t} d(X_s, A) = 0$.
- ($\supseteq$). If $\inf_{s \in \mathbf{Q}: s \le t} d(X_s, A) = 0$, then there is a sequence $s_n \in \mathbf{Q}$ with $s_n \le t$ such that $d(X_{s_n}, A) \to 0$. We can without loss of generality say that s_n converges to some s (Otherwise, since s_n takes values in the compact set $[0, t]$, we can extract a convergent subsequence and just call that s_n), and we have that $X_{s_n} \to X_s$, therefore $d(X_s, A) = 0$, which implies that $X_s \in A$, and as such $T_A \le t$.

♡

This was a regularity condition on the kind of sets that we are allowed to take in order for hitting times to be stopping times. If we want to be able to take open sets as well and still have stopping times, we can impose another kind of regularity condition. In this case, on the σ-algebra (after all recall that something is a stopping time **with respect to a sigma algebra**).

Definition 3.7 (Right-continuous sigma-algebra) Let $\{\mathscr{F}_t : t \ge 0\}$ be a filtration. Then it is easy to see that

$$\mathscr{F}_t^+ = \bigcap_{s>t} \mathscr{F}_s$$

is also a sigma-algebra. We can create a new filtration out of this and call it $\{\mathscr{F}_t^+\}$. If $\mathscr{F}_t = \mathscr{F}_t^+$, then we say $\mathscr{F}_t$ is right-continuous.

We now show how this regularity condition on our sigma-algebra does enable hitting times of open sets to be stopping times:

Proposition 3.8 Let A be an open set and X a continuous process. Then

$$T_A = \inf\{t \ge 0 : X_t \in A\}$$

is a stopping time with respect to the filtration $(\mathscr{F}_t^+)$.

Main idea: Show that $\{T_A < t\} \in \mathscr{F}_t$, by using continuity of X and openness of A, and then show how this implies the end goal using right-continuity of the filtration.

Proof. We note that if $T_A < t$, then there is some rational $q < t$ with $X_q \in A$. Since X is continuous and A is open, $\{X_q \in A\}$ is measurable, and so

$$\{T_A < t\} = \bigcup_{q \in \mathbf{Q}, q < t} \{X_q \in A\} \in \mathscr{F}_t$$

Now we use right-continuity of our filtration, to deduce that

$$\{T_A \leq t\} = \bigcap_{n \in \mathbf{N}} \underbrace{\{T_A < t + n^{-1}\}}_{\in \mathscr{F}_{t+\frac{1}{n}}} \in \mathscr{F}_t^+$$

♡

3.1 Martingale regularisation

Recall that we are treating a stochastic process X as a random variable

$$X : \Omega \to \{f : \mathbf{R}_+ \to E\}$$

Where the codomain is endowed with the σ-algebra that makes the projections $\Pi_t : f \mapsto f(t)$ measurable maps, call this algebra $\mathscr{E}$. The law of the process X is then defined as $\mu(A) = \mathbf{P}(X \in A)$. This law is a very complicated object, so we would like to simplify our attention to looking at "snapshots" of the process at specific times, say $t_1, \cdots, t_n$, and then knowing what the distribution of the process is at those times. This is the concept of a finite dimensional distribution

Definition 3.9 (Finite dimensional distribution) Let μ be a measure on the space of càdlàg functions $\mathbf{R}_+ \to E$. For each $J \subseteq \mathbf{R}+$, a finite index set of times, we let μ_J be the law of the vector $(X_t : t \in J)$. The family of measures $(\mu_J : J \subseteq \mathbf{R}_+$ finite) is called the finite dimensional distributions of μ.

The key idea is that it turns out that to check if the law of two stochastic processes X and Y is the same, it is sufficient and necessary to check whether the finite dimensional distributions agree. This is a consequence of the $\pi - \lambda$ Lemma.

Proposition 3.10 Let X and Y be two stochastic processes. Their laws coincide if and only if their finite-dimensional distributions coincide.

Proof. Left to finish later. See [idk, Theorem 23] ♡

We are interested in studying the sample paths of a process, and the problem is that knowing the law of a process does not give us much information about the sample path properties, indeed:

Example 3.11 (Same law, different sample paths) Let $X = (X_t)_{t\in[0,1]}$ be the process that is identical to 0 for all t. Let U be a uniform random variable on $[0,1]$, and define $X'_t = \mathbf{1}(U = t)$. Then observe that the finite dimensional distributions of X are the Dirac measures at zero, and we have the same for X'_t. As an intuitive explanation, if $0 \notin A$, and say $1 \in A$, then $\mathbf{P}(X_{t_1} \in A) = \mathbf{P}(U = t_1) = 0$, if $1 \notin A$, then obviously $\mathbf{P}(X_{t_1} \in A) = 0$. Either way, we have two processes with the same law but obviously different sample paths: one is always at zero, but the other one jumps to 1 for one value of time.

Thus we are going to talk about processes which agree almost surely on sample paths.

Definition 3.12 Let X and X' be two processes defined on the same probability space. We say that X' is a version of X if $X_t = X'_t$ a.s for each t.

We will now talk about how we can regularise Martingales: as we have seen with the previous examples of hitting times, it is generally useful to have a continuous (or at least càdlàg) Martingale. The following result will tells us that under some regularity conditions on our sigma-algebra, we can always modify a Martingale so that it is càdlàg and it agrees almost surely on sample paths.

Definition 3.13 (Usual conditions) Let $(\mathscr{F}_n)$ be a filtration, we say that it satisfies the usual conditions if it is right-continuous and it contains all $\mathbf{P}$-null sets in $\mathscr{F}$. Alternatively, we define the filtration

$$\widetilde{\mathscr{F}}_t = \sigma(\mathscr{F}_t^+, \mathscr{N})$$

Where $\mathscr{N}$ are the $\mathbf{P}$-null sets, thus "regularising" our filtration.

Theorem 3.14 (Martingale Regularisation) Let $(X_t)_{t\geq 0}$ be a Martingale with respect to the filtration $(\mathscr{F}_t)$. Then there exists a càdlàg process $\widetilde{X}$, a Martingale with respect to $(\widetilde{\mathscr{F}_t})$ that satisfies

$$X_t = \mathbf{E}\left[\widetilde{X}_t \mid \mathscr{F}_t\right]$$

Remark 3.15 Thus as an immediate consequence of this, if $(\mathscr{F}_t)$ already satisfies the usual conditions, i.e. $\mathscr{F}_t = \widetilde{\mathscr{F}_t}$, then we have

$$X_t = \mathbf{E}\left[\widetilde{X_t} \mid \mathscr{F}_t\right] = \mathbf{E}\left[\widetilde{X_t} \mid \widetilde{\mathscr{F}_t}\right] = \widetilde{X_t}$$

where this last equality is due to the fact that the Regularisation Theorem shows that $\widetilde{X_t}$ is $\widetilde{\mathscr{F}_t}$

measurable.

To prove the Regularisation Theorem we are going to need to make use of the following lemma, similar to the result we saw previously that said that a sequence (x_n) of real numbers converged to a real number (or $\pm\infty$) if and only if the number of upcrossings between every $[a,b]$ for rationals $a<b$ was finite.

Lemma 3.16 Let $f:\mathbf{Q}_+\to\mathbf{R}$ be a function such that for every bounded $I\subseteq\mathbf{Q}_+$ we have that

- $f(I)$ is bounded.
- For any $a<b$ rationals, the upcrossings over $[a,b]$ of f on I, denoted by $N([a,b],I,f)$ is finite.

Then for every $t\in\mathbf{R}_+$ we have that the limits

$$\lim_{s\downarrow t} f(s) \qquad \lim_{s\uparrow t} f(s)$$

exist and are finite.

Proof. Take $(s_n)\downarrow t$ be a sequence of rationals. Then $f(s_n)$ is also a sequence of real numbers that doesn't oscillate too much (here I mean this of course in light of the Lemma about convergence), so it converges to some limit $\lim_{s_n\downarrow t} f(s_n)$. Let us show this limit is unique. Suppose $(q_n)\downarrow t$ is a different sequence to (s_n), but $\lim_{s_n\downarrow t} f(s_n)\neq\lim_{q_n\downarrow t} f(q_n)$. Then we can create another sequence $a_n=(s_1,q_1,s_2,q_2,\cdots)$ with $a_n\downarrow t$, but $\lim_{a_n\downarrow t} f(a_n)$ doesn't exist, contradicting the first thing we have shown. By symmetry the second limit also follows. ♡

Let us now give some intuition on the proof of the Regularisation Theorem
Main idea: We use the rational convergence lemma above to show that on a set of probability 1, $\widetilde{X}_t=\lim_{s_n\downarrow t} X_{s_n}$ is a well-defined limit. Then we use the fact that (X_{s_n}) is a backwards Martingale to show that $X_t=\mathbf{E}\left[\widetilde{X}_t\mid\mathscr{F}_t\right]$ almost surely, and then finally show the Martingale property for $\widetilde{X}_t$

Proof of Theorem 3.14. The goal is to define $\widetilde{X}_t=\lim_{s\downarrow t} X_s$. To do this, we wish to employ the Lemma above. We need to tick two boxes: let $I\subseteq\mathbf{Q}_+$ be bounded:

- **$X(I)$ is almost surely bounded**: to show this, let $J=\{j_1,\cdots,j_n\}\subseteq I$ where without loss of generality $j_1<\cdots<j_n$. Then $(X_j)_{j\in J}$ is a discrete-time Martingale, so by Doob's Maximal Inequality, for any $\lambda>0$:
$$\mathbf{P}\left(\max_{j\in J}|X_j|>\lambda\right)\leq\frac{\mathbf{E}\left[|X_{j_n}|\right]}{\lambda}$$
Then taking a monotone limit over subsets $J\subseteq I$, using the fact that I is countable, we get

that

$$\mathbf{P}\left(\sup_{j\in I}|X_t|>\lambda\right)\leq\frac{\mathbf{E}\left[|X_{\sup I}|\right]}{\lambda}$$

Taking $\lambda\to\infty$ gives that $\mathbf{P}\left(\sup_{j\in I}|X_t|<\infty\right)=1$.

- **For any $a<b$ rationals,** $N([a,b],I,X)<\infty$: to show this, we take a similar approach. First note that $N([a,b],I,X)=\sup_{J\subseteq I,J\text{finite}}N([a,b],J,X)$. And by writing (in order) $J=\{a_1,\cdots,a_n\}$, we deduce that $(X_a)_{a\in J}$ is a discrete-time Martingale and so by Doob's Upcrossing Inequality, we see that

$$\mathbf{E}[N([a,b],J,X)]\leq\frac{\mathbf{E}\left[(X_{a_n}-a)^-\right]}{b-a}$$

so from this we see that $\mathbf{E}[N([a,b],I,X)]<\infty$ a.s which means $N([a,b],I,X)<\infty$ a.s. In particular, for any $M>0$

$$N([a,b],\underbrace{[0,M]\cap\mathbf{Q}_+}_{I_M},X)<\infty\text{a.s}$$

Then by setting

$$\Omega_0=\bigcap_{M\in\mathbf{N}}\bigcap_{a<b}\{N([a,b],I_M,X)<\infty\}\cap\left\{\sup_{t\in I_M}|X_t|<\infty\right\}$$

We see that $\mathbf{P}(\Omega_0)=1$, and moreover, for any $\omega\in\Omega_0$, $X(\omega)$ satisfies the two conditions of the "rational convergence lemma" (indeed any bounded $I\subseteq\mathbf{Q}_+$ falls into some I_M, and if the claims hold on all of I_M then they also hold on all of I), so we can safely set

$$\widetilde{X_t}=\lim_{s\downarrow t}X_s\,\mathbf{1}(\Omega_0)$$

Then $\widetilde{X_t}$ is $\widetilde{\mathscr{F}_t}$-adapted, because each $X_s\in m\mathscr{F}_s$, and so $\lim_{s\downarrow t}X_s\in m\bigcap_{s>t}\mathscr{F}_s$, and since $\Omega^c\in\mathscr{N}$, then it follows that $\widetilde{X_t}\in m\widetilde{\mathscr{F}_t}$. We now show the remaining properties: the expectation and Martingale properties of $\widetilde{X_t}$. Since (on Ω, hence a.s)

$$\widetilde{X_t}=\lim_{s\downarrow t}X_s=\lim_{n\to\infty}X_{s_n}=\lim_{n\to\infty}\mathbf{E}\left[X_t\mid\mathscr{F}_{s_n}\right]$$

for some $s_n\downarrow t$, we can spot two backwards Martingales in here:

- $\left(\mathbf{E}\left[X_t\mid\mathscr{F}_{s_n}\right]\right)$ is a backwards Martingale, so by the Backwards Martingale convergence Theorem, we have that almost surely:

$$\widetilde{X_s}=\mathbf{E}[X_t\mid\mathscr{F}_{s^+}]\quad(\star)$$

- $(X_{s_n})_{n\geq0}$ is also a backwards Martingale, so it converges (in particular) in $\mathscr{L}^1$, which means we can pass the convergence inside the expectation: $\mathbf{E}\left[X_{s_n} \mid \mathscr{F}_t\right] \to \mathbf{E}\left[\widetilde{X}_t \mid \mathscr{F}_t\right]$ but of course don't forget that (X_t) is a continuous time Martingale so $\mathbf{E}\left[X_{s_n} \mid \mathscr{F}_t\right] = X_t$, thus showing the expectation property

$$X_t = \mathbf{E}\left[\widetilde{X}_t \mid \mathscr{F}_t\right] \quad (\star\star)$$

Now we need to fix $\mathscr{F}_{s^+}$ into $\widetilde{\mathscr{F}}_s$, but this is an easy job because

$$\mathbf{E}[X \mid \sigma(\mathscr{G}, \mathscr{N})] = \mathbf{E}[X \mid \mathscr{G}]$$

for any σ-algebra $\mathscr{G}$ and integrable X. Now we have shown almost everything except the càdlàg property of $\widetilde{X}_t$

Let us show now the Martingale property of $\widetilde{X}_t$. Fix $s < t$, then

$$\widetilde{X}_s \overset{\star}{=} \mathbf{E}[X_t \mid \mathscr{F}_{s^+}] \overset{\star\star}{=} \mathbf{E}\left[\mathbf{E}\left[\widetilde{X}_t \mid \mathscr{F}_t\right] \mid \mathscr{F}_{s^+}\right] \overset{\text{Tower Law}}{=} \mathbf{E}\left[\widetilde{X}_t \mid \mathscr{F}_{s^+}\right]$$

All left to do now is show càdlàg property and left limits. ♡

Example 3.17 (Of when a filtration doesn't satisfy the usual conditions) Let ξ, η be independent random variables taking values in ± 1 with equal probability. Define

$$X_t = \begin{cases} 0 & t < 1 \\ \xi & t = 1 \\ \xi + \eta & t > 1 \end{cases}$$

We define $\mathscr{F}_t$ to be the natural filtration associated to the process (X_t). Then it is quite easy to check that (X_t) is a Martingale with respect to $(\mathscr{F}_t)$, as we only have a few cases to check, and the expectations of ξ and η are all zero. It is also clear by construction that X_t is not right-continuous, but by the Regularisation Theorem, we can find some $\widetilde{X}_t$, Martingale with respect to the "regularised filtration", that is right continuous. We claim that this is

$$\widetilde{X}_t = \begin{cases} 0 & t < 1 \\ \xi + \eta & t \geq 1 \end{cases}$$

It is also quite easy to check that $\mathbf{E}\left[\widetilde{X}_t \mid \mathscr{F}_t\right] = X_t$ for all t, and it is obvious that $\widetilde{X}_t$ is right-continuous. It is also easy to see that $\widetilde{X}_t$ is a Martingale with respect to $(\mathscr{F}_t^+)$, **but** since $X_1 \neq \widetilde{X}_1$, it follows that $\widetilde{X}$ is not a version of X, which means that $(\mathscr{F}_t)$ does not satisfy the usual conditions, which is quite clear to see from the way X_t is defined.

From now on, we will always consider the càdlàg version of a Martingale whenever the corresponding filtration satisfies the usual conditions. We now revisit the results we saw in discrete-time.

3.2 Doob's inequalities and convergence in continuous time

Theorem 3.18 (Martingale Convergence Theorem) Let (X_t) be an $\mathcal{L}^1$ bounded càdlàg Martingale. Then the limit $X_t \to X_\infty$ exists almost surely for some $X_\infty \in \mathcal{L}^1(\mathcal{F}_\infty)$, where $\mathcal{F}_\infty = \sigma(\mathcal{F}_t : t \geq 0)$

Theorem 3.19 (Doob's Maximal Inequality) Let $(X_t : t \geq 0)$ be a càdlàg Martingale, and let $X_t^* = \sup_{0 \leq s \leq t} |X_s|$ be the running maximum. Then, for all $\lambda \geq 0, t \geq 0$

$$\lambda \mathbf{P}(X_t^* \geq \lambda) \leq \mathbf{E}\left[|X_t| \mathbf{1}(X_t^* \geq \lambda))\right] \leq \mathbf{E}[|X_t|]$$

Theorem 3.20 (Doob's $\mathcal{L}^p$ inequality) Let $(X_t : t \geq 0)$ be a càdlàg process with $p > 1$. Then for $X_t^* = \sup_{0 \leq s \leq t} |X_s|$, we get

$$\left\| X_t^* \right\|_p \leq \frac{p}{p-1} \|X_t\|_p$$

Theorem 3.21 ($\mathcal{L}^p$ convergence) Let X be a càdlàg Martingale and $p > 1$, then the following are equivalent:

- X is bounded in $\mathcal{L}^p$.
- X converges almost surely and in $\mathcal{L}^p$ to some X_∞.
- $X_t = \mathbf{E}[Z \mid F_t]$ almost surely for some $Z \in \mathcal{L}^p$.

Theorem 3.22 (UI Martingale convergence) Let X be a càdlàg Martingale. Then X is UI if and only if X converges almost surely and in $\mathcal{L}^1$ to X_∞, and this is if and only if X is closed in $\mathcal{L}^1$

Theorem 3.23 (Optional Stopping Theorem) Let X be a càdlàg UI Martingale. Then for every stopping time $S \leq T$, we have that

$$\mathbf{E}[X_T \mid \mathcal{F}_S] = X_S$$

Main idea: We will extract this by discretising the Martingale and then using the discrete OST and the càdlàg property to return to our original Martingale.

Proof. Let $A \in \mathscr{F}_S$, the goal is to show that

$$\mathbf{E}[X_T \mathbf{1}(A)] = \mathbf{E}[X_S \mathbf{1}(A)]$$

Define sequences of stopping times

$$T_n = 2^{-n}\lceil 2^n T \rceil \qquad S_n = 2^{-n}\lceil 2^n S \rceil$$

Then it is clear that $T_n \downarrow T$ and $S_n \downarrow S$, so by the càdlàg property

$$X_{S_n} \to X_S \qquad X_{T_n} \to X_T$$

Moreover, since X is càdlàg and UI, by the UI Martingale convergence theorem for càdlàg Martingales, we have that $X_m \to X_\infty$ as $m \to \infty$ in $\mathscr{L}^1$. Then we have that since T_n is a bounded stopping time, by the OST and this convergence:

$$\mathbf{E}\left[X_\infty \mid \mathscr{F}_{T_n}\right] = \mathbf{E}\left[\lim_{m\to\infty} X_m \mid \mathscr{F}_{T_n}\right] = \lim_{m\to\infty} \mathbf{E}\left[X_m \mid \mathscr{F}_{T_n}\right] = X_{T_n}$$

(Since we are taking a limit we can without loss of generality assume that $m \geq T_n$). This has shown that X_{T_n} is UI, and as such it converges to X_T in $\mathscr{L}^1$. Moreover, since $S_n \leq T_n$, and X is UI, we have by the Optional Stopping Theorem for UI Martingales, that

$$\mathbf{E}\left[X_{T_n} \mid \mathscr{F}_{S_n}\right] = X_{S_n}$$

Of course since $S_n \geq S$, we have that any $A \in \mathscr{F}_S$ is also in $\mathscr{F}_{S_n}$, so by the property of conditional expectation we have that

$$\mathbf{E}\left[X_{T_n} \mathbf{1}(A)\right] = \mathbf{E}\left[X_{S_n} \mathbf{1}(A)\right]$$

but now we can pass the limits $n \to \infty$ inside the expectations because we argued that $X_{T_n} \to X_T$ (and hence the same for S) in $\mathscr{L}^1$. ♡

3.3 Kolmogorov's continuity criterion

We have seen in a counterexample how two processes that agree on their finite dimensional distributions need not have the same sample paths, so if there are some versions of the stochastic process that satisfy some better regularity conditions, i.e: (càdlàg) continuity as we have seen, it makes sense to work with them. The following criterion ensures that in fact a continuous version of the process exists under some regularity conditions on the moments.

Definition 3.24 (Dyadic rationals) Let $\mathscr{D}_n = \{k2^{-n} : 0 \le k \le 2^n\}$ be the set of dyadic rationals of level n, and $\mathscr{D} = \bigcup_n \mathscr{D}_n$.

Theorem 3.25 (Kolmogorov's continuity criterion) Let $(X_t)_{t\in\mathscr{D}}$ be a stochastic process with real values. Suppose that there exists $p > 0$, $\epsilon > 0$ so that

$$\mathbf{E}[|X_t - X_s|^p] \le c|t-s|^{1+\epsilon} \qquad s,t \in \mathscr{D}$$

for some $c < \infty$. Then for all $\alpha \in (0, \epsilon/p)$, we have that $(X_t)_{t\in\mathscr{D}}$ is α-Hölder continuous, i.e: there exists some random variable K_α such that

$$|X_t - X_s| \le K_\alpha |s-t|^\alpha \qquad s,t \in \mathscr{D}$$

Proof. We start by using Markov's inequality alongside the bound on the moment:

$$\mathbf{P}\left(\left|X_{k2^{-n}} - X_{(k+1)2^{-n}}\right| \ge 2^{-n\alpha}\right) \le c2^{n\alpha p}2^{-n-n\epsilon}.$$

By a union bound, we have that

$$\mathbf{P}\left(\max_{0\le k<2^n}\left|X_{k2^{-n}} - X_{(k+1)2^{-n}}\right| \ge 2^{-n\alpha}\right) \le \sum_{k=0}^{2^n-1} c2^{n\alpha p}2^{-n-n\epsilon} = c2^{-n(\epsilon-p\alpha)}.$$

Since we have assumed $\alpha \in (0, \epsilon/p)$, the exponent is negative, so the probabilities are summable in n, so by the Borel-Cantelli Lemma, we have that for sufficiently large n,

$$\max_{0\le k<2^n}\left|X_{k2^{-n}} - X_{(k+1)2^{-n}}\right| \le 2^{-n\alpha}$$

Which is equivalent to saying that there is some random variable M with

$$\sup_{n\ge 0}\max_{0\le k<2^n}\frac{\left|X_{k2^{-n}} - X_{(k+1)2^{-n}}\right|}{2^{-n\alpha}} \le M < \infty$$

We now show the existence of some random variable $M' < \infty$ almost surely so that for all $s,t \in \mathscr{D}$, we have that

$$|X_t - X_s| \le M'|t-s|^\alpha.$$

Let $s,t \in \mathscr{D}$. Pick the unique r such that

$$2^{-(r+1)} < t-s < 2^{-r}$$

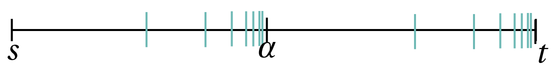
(This has nothing to do with the fact that s and t are dyadic, its just saying that there is some r for which the gap between s and t fits in a bin of width 2^{-r} but not in a bin of width $2^{-(r+1)}$). Now there exists a k such that $s < k2^{-(r+1)} < t$, and set α to be $\alpha = k2^{-(r+1)}$, then we have (since $\alpha > s$) that $0 < t - \alpha < 2^{-r}$. So in the dyadic expansion of $t - \alpha$, we get that

$$t - \alpha = \sum_{k \geq r+1} \frac{x_j}{2^j}$$

where each $x_j \in \{0, 1\}$. Similarly we have that (now because $\alpha < t$)

$$\alpha - s = \sum_{j \geq r+1} \frac{y_j}{2^j}$$

From these two sums, we see that we can partition the interval $[s, t)$ into disjoint unions of subintervals of length 2^{-n} for $n \geq r + 1$ and where at most two such intervals have the same length. See diagram:



So we can write

$$|X_s - X_t| \leq \sum_{d,n} |X_d - X_{d+2^{-n}}|,$$

where $d, d+2^{-n}$ indicate the endpoints of said subintervals in the decomposition of $[s, t)$. Recalling that

$$\sup_{n \geq 0} \max_{0 \leq k < 2^n} \frac{\left|X_{k2^{-n}} - X_{(k+1)2^{-n}}\right|}{2^{-n\alpha}} \leq M < \infty$$

We have that

$$|X_s - X_t| \leq 2 \sum_{n \geq r+1} M 2^{-n\alpha} = 2M \frac{2^{-(r+1)\alpha}}{1 - 2^{-\alpha}}$$

and so from this we see that

$$|X_s - X_t| \leq M' 2^{-(r+1)\alpha} \leq M' |t - s|^\alpha$$

♡

4 Weak convergence

4.1 Definitions

Definition 4.1 (Weak convergence) Let $(\mu_n : n \geq 0)$ be a sequence of Borel probability measures on a metric space (M, d). We say μ_n converges weakly to μ, which I write was $\mu_n \rightharpoonup \mu$, if $\mu_n(f) \to \mu(f)$ for all bounded continuous functions f on M.

Example 4.2 (Examples of weak convergence) Here are some examples:

- Let $(x_n : n \geq 0)$ be a sequence in a metric space M that converges to x. Then $\delta_{x_n} \rightharpoonup \delta_x$. Indeed, for any continuous function,

$$\delta_{x_n}(f) = f(x_n) \to f(x) = \delta_x(f).$$

- Let $M = [0,1]$ with the Euclidean metric. Let $\mu_n = n^{-1} \sum_{0 \leq k \leq n-1} \delta_{k/n}$. This corresponds to spreading lighter and lighter masses evenly spaced through the interval $[0,1]$. Then $\mu_n(f)$ is precisely a Riemann sum, and it converges to $\int_0^1 f(x)dx$, which shows that μ_n converges weakly to the Lebesgue measure on $[0,1]$. This makes sense from the intuitive picture!

Remark 4.3 Note that if $A \in \mathscr{B}(M)$, it is not necessary that $\mu_n(A) \to \mu(A)$ when $\mu_n \rightharpoonup \mu$. Indeed, let $x_n = 1/n$ and $\mu_n = \delta_{x_n}$. Then for $A = (0,1)$, we have that $\delta_{x_n}(A) \equiv 1$ but $\delta_0(A) = 0$.

We have the following equivalent characterisation of weak convergence:

Theorem 4.4 (Equivalent characterisation of weak convergence, Portmanteau's Theorem) Let (μ_n) be a sequence of probability measures. Then the following are equivalent:

1. $\mu_n \rightharpoonup \mu$.
2. $\liminf \mu_n(G) \geq \mu(G)$ for all open G. (Think of $G = (0,1)$ and $\mu_n = \delta_{1/n}$)
3. $\limsup \mu_n(A) \leq \mu(A)$ for all closed A. (Think of $A = \{0\}$ and $\mu_n = \delta_{1/n}$)
4. $\lim \mu_n(A) = \mu(A)$ for all A with $\mu(\partial A) = 0$, where $\partial A = \bar{A} \setminus A^\circ$. (Think of $A = (-\epsilon, \epsilon)$ and $\mu_n = \delta_{1/n}$)

Main idea: To show (1) $\implies$ (2), you use one of the key tricks of this section, namely that if G is open:

$$1 \wedge k d(x, G^c) \uparrow \mathbf{1}(G) \qquad k \to \infty$$

where the left hand side is a continuous and bounded function. To show (2) $\iff$ (3) is clear by taking complements. Then you show that (2)&(3) $\implies$ (4)

Proof. Let show that (1) $\implies$ (2). Let G be open with non-empty complement (otherwise there's not much to prove). For every positive k, define

$$f_k(x) = 1 \wedge [k d(x, G^c)]$$

Then it is easy to see that $f_k \uparrow \mathbf{1}(G)$, this is because G^c is closed, indeed: the only scenario where $f_k(x)$ is zero is if $d(x, G^c)$ is zero, which means that $x \in G^c$. If it were open, you could still have for some x that $f_k(x) = 0$ yet $\mathbf{1}(G)(x) = 1$. Since f_k is continuous and bounded, we have that

$$\mu_n(f_k) \to \mu(f_k)$$

Now since $f_k(x) \leq \mathbf{1}(G)$, you have that $\liminf \mu_n(G) \geq \liminf \mu_n(f_k) = \mu(f_k)$ where this last equality comes from weak convergence. Now taking $k \uparrow \infty$ we finish the claim.

It is easy to see that (2) $\iff$ (3) by taking complements. Now we show how (2) and (3) in conjunction imply (4). Let $A \in \mathscr{B}(M)$ be a Borel set with $\mu(\partial A) = \mu(\bar{A} \setminus A^\circ) = 0$. This implies that $\mu(A^\circ) = \mu(\bar{A}) = \mu(A)$. Now we have that

$$\begin{aligned}
\limsup_n \mu_n(A) &\leq \limsup_n \mu_n(\bar{A}) && \text{(Since } A \subseteq \bar{A}\text{)} \\
&\leq \mu(\bar{A}) = \mu(A) = \mu(A^\circ) && \text{(3) and } \mu(\partial A) = 0 \\
&\leq \liminf_n \mu(A^\circ) && \text{(2)} \\
&\leq \liminf_n \mu_n(A) && \text{(Since } A^\circ \subseteq A\text{)}
\end{aligned}$$

The only way that $\limsup_n \mu_n(A) \leq \mu(A) \leq \liminf_n \mu_n(A)$ can hold is if $\lim_n \mu_n(A) = \mu(A)$. Thus showing (4). Now we show that (4) implies weak convergence. For simplicity consider $f \geq 0$ continuous and bounded (then we can consider positive and negative parts). By considering the rectangle of base $f(x)$ and height 1, it is easy to see that

$$f(x) = \int_0^\infty \mathbf{1}(t \leq f(x)) dt$$

So we have that

$$\begin{aligned}
\mu_n(f) &= \int_M \left(\int_0^\infty \mathbf{1}(t \le f(x)) dt \right) d\mu_n(x) \\
&= \int_0^\infty \int_M \mathbf{1}(t \le f(x)) d\mu_n(x) dt && \text{(Fubini)} \\
&= \int_0^\infty \mu_n(\{f(x) \ge t\} dt \\
&= \int_0^{\|f\|_\infty} \mu_n(\{f(x) \ge t\}) dt
\end{aligned}$$

Our goal is to now take the limit $n \to \infty$ and to pass it inside an integral, which we can do by the Dominated Convergence Theorem, i.e:

$$\lim_n \mu_n(f) = \int_0^{\|f\|_\infty} \lim_n \mu_n(\{f(x) \ge t\}) dt$$

To conclude (1), we now will use (4), but first we need to ensure that $\partial\{f(x) \ge t\}$ has μ-measure zero (almost surely in the Lebesgue measure, since that's what we are integrating against), and then the claim will follow because $\lim_n \mu_n(f(x) \ge t) = \mu(f(x) \ge t)$ almost everywhere in t, and then we can undo the argument we did above, and reach the final conclusion. To show that $\partial\{f(x) \ge t\}$ has μ-measure zero, note that $\{f \ge t\}$ is closed by continuity of f, and since $\{f > t\}$ is open by continuity of f, we have that $\{f > t\} \subseteq \{f \ge t\}^\circ$, which means that

$$\partial\{f \ge t\} = \overline{\{f \ge t\}} \setminus \{f \ge t\}^\circ = \{f \ge t\} \setminus \{f \ge t\}^\circ \subseteq \{f \ge t\} \setminus \{f > t\} = \{f = t\}$$

Now we show that $\{f = t\}$ has μ-measure zero for Lebesgue almost all t. The argument goes like this:

$$\{t : \mu\{f = t\} > 0\} = \bigcup_{n \ge 1} \{t : \mu\{f = t\} \ge 1/n\}$$

But since μ is a probability measure, it must be that the n^{th} set on the right hand side contains at most n such values of t (otherwise the total mass would exceed 1), and so $\{t : \mu\{f = t\} > 0\}$ is countable, and any countable set has 0 Lebesgue measure. This finishes the claim, because we have shown that $\mu(\{f = t\}) = 0$ Lebesgue almost-everywhere in t, and since we showed that $\partial\{f \ge t\} \subseteq \{f = t\}$, it follows that $\mu\{\partial\{f \ge t\}\} = 0$ Lebesgue almost everywhere in t, and so the argument is done. ♡

We now talk about the relationship between weak convergence and convergence in distribution.

Definition 4.5 (Distribution function) For μ a finite measure on $\mathbf{R}$, the distribution function of μ, F_μ is given by

$$F_\mu(x) = \mu(-\infty, x]$$

Proposition 4.6 (Relation between weak convergence and convergence in distribution) Let $\{\mu_n\}$ be a sequence of probability measures on $\mathbf{R}$. Then we have that $\mu_n \to \mu$ weakly (1) if and only if $F_{\mu_n}(x) \to F_\mu(x)$ for all x that are points of continuity of F_μ (2).

Main idea: To show (1) $\implies$ (2), you just show that $\mu\{x\} = 0$ by expressing this quantity as a difference of F_μ's and using its continuity at x.

To prove that (2) $\implies$ (1) we want to use Portmanteau's Theorem with the case of open sets, i.e: we want to show that for any open set U, $\liminf_n \mu_n(U) \geq \mu(U)$. To do this, we express U as a countable union of intervals and use the distribution functions to deal with those, as well as the fact that the points of continuity will be dense in $\mathbf{R}$.

Proof. (1) $\implies$ (2): fix a continuity point x of F_μ. Our goal is to show that $\mu_n(-\infty, x] \to \mu(\infty, x]$. To do so we wish to use Portmanteau's Theorem, and to apply the Theorem we need to show that $\partial(-\infty, x] = \{x\}$ has μ measure zero. This is easy to see because

$$\mu(\{x\}) = \mu(-\infty, x] - \lim_n \mu(-\infty, x - 1/n] = \lim_n (F_\mu(x) - F_\mu(x - 1/n)) = 0$$

Where this last equality follows from continuity of F_μ at x.

To show that (2) $\implies$ (1), we will consider open sets of the real line and use Portmanteau's Theorem. First note that for any open set $U \subseteq \mathbf{R}$, we can write

$$U = \bigcup_k (a_k, b_k)$$

for some disjoint collection of intervals (a_k, b_k), so that $\mu_n(U) = \sum_k \mu_n(a_k, b_k)$. Since F_{μ_n} is always increasing, it can have at most countably many discontinuities, which means that the set of points of contuinity is dense (since the complement of a countable set is dense), so for any interval (a, b) we may choose continuity points a' and b' such that $a < a' < b' < b$.

Now note that

$$\mu_n(a,b) = \lim_{\beta \uparrow b} F_{\mu_n}(\beta) - F_{\mu_n}(a) \geq F_{\mu_n}(b') - F_{\mu_n}(a')$$

So taking limits, we get that

$$\liminf_n \mu_n(a,b) \geq F_\mu(a',b')$$

Now we can take $a' \downarrow a$ and $b' \uparrow b$ along continuity points of F_μ, and this means that

$$\liminf_n \mu_n(a,b) \geq \mu(a,b)$$

Now back in the case of a generic open set $U = \bigcup_k (a_k, b_k)$ we have that

$$\liminf_n \mu_n(U) \geq \sum_k \liminf_n \mu_n(a_k, b_k) \geq \sum_k \mu(a_k, b_k) = \mu(U)$$

Where in the first inequality we used Fatou's Lemma (interpreting $\sum_k \mu_n(a_k, b_k)$ as an integral against the counting measure). ♡

This proposition motivates the following definition, that generalises the concept of convergence of distribution for general random variables taking values in a metric space (M,d).

Definition 4.7 (Convergence in distribution) Let (X_n) be a collection of random variables taking values in a metric space (M,d), with the random variables possibly being defined on different probability spaces. Then we say that $X_n \to X$ in distribution, if the laws $\mathscr{L}(X_n)$ have $\mathscr{L}(X_n) \rightharpoonup \mathscr{L}(X)$.

Remark 4.8 If say X_n is defined on $(\Omega, \mathscr{F}_n, \mathbf{P}_n)$ and X is defined on $(\Omega, \mathscr{F}, \mathbf{P})$, then it is clear that convergence in distribution occurs if and only if

$$\mathbf{E}_{\mathbf{P}_n}[f(X_n)] \to \mathbf{E}_{\mathbf{P}}[f(X)]$$

for all continuous and bounded $f : M \to \mathbf{R}$

We have also the following fact (which we already saw in usual case of $\mathbf{R}$ in a first course in Probability Theory)

Proposition 4.9 Let (X_n) be a sequence of random variables taking values in (M,d) with $X_n \to X$ in probability. Then $X_n \to X$ in distribution. Moreover, there is the partial converse that if $X_n \to c$ for some constant c in distribution, then the convergence $X_n \to c$ also occurs in probability.

Example 4.10 A classic example of convergence in distribution is that of the Central Limit Theorem: if X_n are all iid $\mathscr{L}^2$ integrable random variables with mean μ and variance σ^2, then

$$\frac{\sum_i X_i - n\mu}{\sigma\sqrt{n}} \to N(0,1)$$

in distribution.

We now present a tool that allows us to check for weak convergence of measures, Proharov's Theorem.

Definition 4.11 (Tightness) A sequence $\{\mu_n\}$ of probability measures on a metric space (M,d) is said to be tight if given any $\epsilon > 0$, there exists a compact set $K \subseteq M$ such that

$$\sup_n \mu_n(M \setminus K) < \epsilon$$

In other words, the sequence is tight if there is a large enough compact set, such that the mass of all measures is mostly concentrated around that set.

Example 4.12 (Example and non-example) The following is an example of a tight family of probability measures: on $\mathbf{R}$, $\mu_n = \delta_{1/n}$, then if we set $K = (-2,2)$, $\mu_n(\mathbf{R} \setminus K) = 0$ for all n, so the sequence is tight.

The following is an example of a family of measures that is not tight: on $\mathbf{R}$, set $\mu_n = \delta_n$, then any compact set K, must in particular be bounded, and so eventually some μ_n will have its mass "escape" K, i.e: for some n large enough, we will have $\mu_n(\mathbf{R} \setminus K) = 1$.

Theorem 4.13 (Prohorov's Theorem) If $\{\mu_n\}$ is a tight sequence, then there exists a subsequence $\{\mu_{n_k}\}$ and a probability measure μ such that

$$\mu_{n_k} \rightharpoonup \mu$$

Main idea: Enumerate the rationals by $\{x_j\}$. Since $F_n(x_1) \in [0,1]$ is abounded subsequence ($F_n := \mu_n(-\infty, n]$), we have that $F_n(x_1)$ has a convergent subsequence. Iterating this diagonal argument, we get a sequence n_k, such that $F_{n_k}(x_j)$ converges for all j, and we set $F(q) = \lim_{k\to\infty} F_{n_k}(q)$ for $q \in \mathbf{Q}$. This convergence holds along the rationals, so we extend along the real numbers by taking monotone limits along rationals, i.e: $F(x) := \lim_{q\downarrow x : q\in\mathbf{Q}} F(q)$. Now construct a measure μ out of this distribution function and by Relation between weak convergence and convergence in distribution, we are done.

Proof. We begin by letting $F_n := F_{\mu_n}$ be the distribution functions of the measures μ_n. The goal is to construct the "alleged limiting measure" μ, and show there exists a subsequence of $\{\mu_n\}$ on which the corresponding subsequence of $\{F_n\}$ converges to some F_μ on the rationals, then we will extend to the reals, and construct μ from there. Then Proposition 4.6 will finish off the proof.

Start by fixing an enumeration $\{x_j\}$ of the rationals. Note that $F_n(x_1) \in [0,1]$ for all n so the sequence $\{F_n(x_1)\}$ has a convergent subsequence, $\{F_{n_k^{(1)}}(x_1)\}$, note that now $\{F_{n_k^{(1)}}(x_2)\}$ is a bounded sequence, hence is has a convergent subsequence $\{F_{n_k^{(2)}}(x_2)\}$. Naturally this subsequence still converges when we replace x_2 by x_1 because its a subsequence of the original subsequence which converged. Iterating this argument, we find a subsequence $\{F_{n_k^j}(x_i)\}$ that converges for all $1 \le i \le j$. In light of this, call $(n_k^{(m+1)})$ a sequence of rationals such that $\{F_{n_k^{(m+1)}}(x_{m+1})\}$ converges. Then the sequence $n_k = n_k^{(k)}$ has that $F_{n_k}(x_j)$ converges for all j (eventually k will "catch up" with j, and then convergence will be guaranteed). We now set

$$F(x) = \lim_{k\to\infty} F_{n_k}(x) \qquad x \in \mathbf{Q}$$

Now note that since each F_n is a distribution function, it is non-decreasing, so we still have that F is non-decreasing, so it has right-hand limits. Therefore we can safely construct

$$F(x) := \lim_{q\downarrow x, q\in\mathbf{Q}} F(q) \qquad x \in \mathbf{R}$$

This is our candidate function. Now by construction, this function is right continuous, indeed:

$$\lim_{t\downarrow x} F(t) = \lim_{t\downarrow x} \lim_{q\downarrow t} F(q) = \lim_{q\downarrow x} F(q) = F(x)$$

And this limit business held because $F(q)$ is right continuous along rationals. Since it is also monotone, it implies it has left limits as well, so F is càdlàg. We now go on to show convergence of F_n to F on continuity points of F along the convergent subsequence, i.e, we want to show that

$$\lim_{k\to\infty} F_{n_k}(t) = F(t)$$

whenever t is a point of continuity of F. By assumption of continuity at t, for any $\epsilon > 0$ we can find sufficiently close points $s_1 < t < s_2$ where $s_1, s_2 \in \mathbf{Q}$ such that

$$\max_i |F(t) - F(s_i)| < \frac{\epsilon}{2}$$

In other words, for k large enough, we have that

$$F(t) - \epsilon \overset{(1)}{<} F(s_1) - \frac{\epsilon}{2} \overset{(2)}{<} F_{n_k}(s_1) \overset{(3)}{<} F_{n_k}(t) < F_{n_k}(s_2) < F(s_2) + \frac{\epsilon}{2} < F(t) + \epsilon$$

Where inequality (1) comes from the construction of s_1, inequality (2) comes from the fact that at rational points, like s_1, we have that $F(s_1) \equiv \lim_{k\to\infty} F_{n_k}(s_1)$, in inequality (3) we use the monotonicity of F_{n_k} and the rest of the inequalities follow in a symmetric manner. This shows that $F_{n_k}(t) \to F(t)$ for continuity points of F. We still haven't really shown that F is a distribution function, i.e: that $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. This will be shown using tightness. Recall that tightness guarantees that for any $\epsilon > 0$ there is some $N \in \mathbf{N}$ large enough such that for all $n > N$, we have that

$$\sup_k \mu_k([-n, n]^c) < \epsilon$$

So with the ϵ above, choose continuity points of F, $a < -N$ and $b > N$, then

$$F(a) < \epsilon \qquad F(b) > 1 - \epsilon$$

as needed (here we used the fact that a is a continuity point as well as the fact that we have established convergence of $F_{n_k}(a) \to F(a)$). We are now ready to construct the measure μ. First we set

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

and now having established that F is a càdlàg distribution function, it follows that μ is a well defined measure. Then we extend by Carethedory's Extension Theorem to a probability measure on $\mathscr{B}(\mathbf{R})$. ♡

4.2 Characteristic functions

Definition 4.14 (Characteristic function) Let X be a random variable taking values in $\mathbf{R}^d$, with law $\mu = \mathscr{L}(X)$. Then the characteristic function, φ_X is defined as

$$\varphi(u) = \mathbf{E}[\exp(i\langle u, X\rangle)] = \int_{\mathbf{R}^d} \exp[i\langle u, x\rangle] d\mu(x)$$

Example 4.15 Let us compute the characteristic function of a Poisson law. Let $X \sim \text{Poi}(\lambda)$, then

$$\begin{aligned}
\varphi_X(u) &= \mathbf{E}[\exp(iuX)] \\
&= \sum_{n=0}^{\infty} \exp(iun) \frac{e^{-\lambda}\lambda^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{\left(\lambda e^{iu}\right)^n}{n!} e^{-\lambda} \\
&= \exp\left\{\lambda\left(e^{iu} - 1\right)\right\}
\end{aligned}$$

Remark 4.16 (Properties of characteristic functions) We have the following

- $\varphi(0) = 1$. (Duh)
- φ is continuous, indeed:

$$\begin{aligned}
\lim_{\mathbf{h}\to\mathbf{0}} \varphi(\mathbf{u}+\mathbf{h}) &= \lim_{\mathbf{h}\to\mathbf{0}} \mathbf{E}[\exp\{i\langle \mathbf{u}+\mathbf{h}, X\rangle\}] \\
&= \mathbf{E}\left[\lim_{\mathbf{h}\to\mathbf{0}} \exp\{i\langle \mathbf{u}+\mathbf{h}, X\rangle\}\right] && \text{(DCT)} \\
&= \varphi(\mathbf{u})
\end{aligned}$$

- φ determines the law of X (Fourier Inversion).

Now we have a Theorem by Lévy that relates convergence of characteristic functions to weak convergence. This Theorem can be used to prove the Central Limit Theorem among other things.

Theorem 4.17 (Lévy's Continuity Theorem) Let (X_n) be a sequence of random variables with values in $\mathbf{R}^d$.

- If $\mathscr{L}(X_n) \to \mathscr{L}(X)$ weakly for some random variable X, then $\varphi_{X_n}(\xi) \to \varphi_X(\xi)$ for all $\xi \in \mathbf{R}^d$.

- If there exists some $\psi : \mathbf{R}^d \to \mathbb{C}$ with $\psi(0) = 1$, ψ is continuous at zero, and for all $\xi \in \mathbf{R}^d$

$$\varphi_{X_n}(\xi) \to \psi(\xi)$$

  then ψ is the characteristic function for some random variable X and $\mathscr{L}(X_n) \to \mathscr{L}(X)$ weakly.

To prove this we first need a technical Lemma:

Lemma 4.18 Let X be a random variable with values in $\mathbf{R}^d$, then

$$\mathbf{P}(\|X\|_\infty \geq k) \leq C \cdot \left(\frac{k}{2}\right)^d \int_{[-k^{-1},k^{-1}]^d} (1 - \varphi_X(\xi)) d\xi$$

Main idea: The proof is not hard, but just tedious: We first show that

$$\left(\frac{k}{2}\right)^d \int_{[-k^{-1},k^{-1}]^d} (1 - \varphi_X(\xi)) d\xi = \mathbf{E}\left[1 - \prod_{j=1}^d \frac{\sin(k^{-1}X_j)}{k^{-1}X_j}\right]$$

This is just done via slightly unpleasant integration. Once this is done, we note that whenever $x \geq 1$, $|\sin(x)| \leq x\sin(1)$. From this we extend and see that whenever $\|u\|_\infty \geq 1$, we have that for $f(u) = \prod_{j=1}^d \frac{\sin(u_j)}{u_j}$: $|f(u)| \leq \sin(1)$. Rearranging this gives that for $C = (1 - \sin(1))^{-1}$,

$$\mathbf{1}\left(\left\|\frac{X}{k}\right\|_\infty \geq 1\right) \leq C\left[1 - \prod_{j=1}^d \frac{\sin(k^{-1}X_j)}{k^{-1}X_j}\right]$$

Proof. Let us show the first part of the intuition section:

$$\begin{aligned}
\int_{[-\lambda,\lambda]^d} \varphi_X(u)du &:= \int_{[-\lambda,\lambda]^d} \int_{\mathbf{R}^d} \exp(i\langle u, x\rangle) d\mu(x) du \\
&= \int_{[-\lambda,\lambda]^d} \int_{\mathbf{R}^d} \prod_{j=1}^{d} \exp(i u_j x_j) d\mu(x) du_1 du_2 \cdots du_d \\
&\overset{(!)}{=} \int_{\mathbf{R}^d} \prod_{j=1}^{d} \int_{[-\lambda,\lambda]} \exp(i u_j x_j) du_j d\mu(x) \\
&= \int_{\mathbf{R}^d} \prod_{j=1}^{d} \left\{ \frac{1}{i x_j} \left(e^{i\lambda x_j} - e^{-i\lambda x_j} \right) \right\} d\mu(x) \\
&= \int_{\mathbf{R}^d} \prod_{j=1}^{d} \left(\frac{2\sin(\lambda x_j)}{x_j} \right) d\mu(x) \\
&= \mathbf{E}\left[\prod_{j=1}^{d} \frac{\sin(\lambda X_j)}{\lambda X_j} \right]
\end{aligned}$$

And step (!) is just Fubini's Theorem, and hence the first part of the intuition section is proven. Now we prove the remaining: Note that when $|x| \geq 1$, we have that $\left|\frac{\sin(x)}{x}\right| \leq \sin(1)$. Hence, if we call $f(u) = \prod_{j=1}^{d} \frac{\sin(u_j)}{u_j}$, we observe the following: whenever $\|u\|_\infty \geq 1$, we must have that for some index $j \in [d]$, $|u_j| \geq 1$. This means that $\left|\frac{\sin(u_j)}{u_j}\right| \leq \sin(1)$. Generally, we always have that $\left|\frac{\sin(x)}{x}\right| \leq 1$, and so whenever $\|u\|_\infty \geq 1$, we conclude that

$$f(u) \leq |f(u)| = \left| \prod_{j=1}^{d} \frac{\sin(u_j)}{u_j} \right| \leq \sin(1)$$

Rearranging, we have that whenever $\|u\|_\infty \geq 1$,

$$\frac{1 - f(u)}{1 - \sin(1)} \geq 1$$

and so from this we conclude that

$$\mathbf{1}\left(\left\| \frac{X}{k} \right\|_\infty \geq 1 \right) \leq C \left[1 - \prod_{j=1}^{d} \frac{\sin(k^{-1} X_j)}{k^{-1} X_j} \right]$$

Where $C+(1-\sin(1))^{-1}$. Now the proof follows immediately:

$$\begin{aligned}\mathbf{P}[\|X\|_\infty \geq k] &= \mathbf{P}\left[\left\|\frac{X}{k}\right\|_\infty \geq 1\right] \\ &= \mathbf{E}\left[\mathbf{1}\left(\left\|\frac{X}{k}\right\|_\infty \geq 1\right)\right] \leq C\mathbf{E}\left[1-\prod_{j=1}^{d}\frac{\sin(k^{-1}X_j)}{k^{-1}X_j}\right] \\ &= C\cdot\left(\frac{k}{2}\right)^d \int_{[-k^{-1},k^{-1}]^d}(1-\varphi_X(\xi))d\xi\end{aligned}$$

as required. ♡

Now we are ready to prove Levy's Theorem:

Main idea: The direct part is trivial, by simply noting that $f(x)=\exp(i\langle\xi,x\rangle)$ is a bounded and continuous function, now use weak convergence. For the converse we have a few steps:

- Show that that $\{\mu_n\}$ is a tight family, for this we will use the Lemma and the hypothesis.
- Extract a subsequence $\{n_k\}$ such that $\mu_{n_k}\rightharpoonup\mu$. I.e: $\mathscr{L}(X_{n_k})\rightharpoonup\mathscr{L}(X)$ (every probability measure arises as the law of a random variable).
- By the direct part, $\varphi_{X_{n_k}}\to\varphi_X$ pointwise, and so $\varphi_X=\psi$.
- Extend the weak convergence along the subsequence to convergence along the whole sequence. To do this, assume $\mathscr{L}(X_n)$ does not converge weakly to $\mathscr{L}(X)$, then for any $\epsilon>0$, there is some bounded continuous function f, and some subsequence n_k so that $|\mathbf{E}[f(X_{n_k})]-\mathbf{E}[f(X)]|>\epsilon$. Now use tightness to extract a further subsequence of $\{n_k\}$ where we do have convergence in distribution and contradict the above statement.

Proof of Lévy's Theorem. The first part is easy: If $\underbrace{\mathscr{L}(X_n)}_{\mu_n}\to\underbrace{\mathscr{L}(X)}_{\mu}$ weakly, then for any continuous and bounded function, we have that $\mu_n(f)\to\mu(f)$, but applying this to $f(x)=\exp(i\langle\xi,x\rangle)$ gives the result. The converse will be more technical. Step 1: show that μ_n are a tight family of measures. For this we will use the previous lemma.

We first note that $|1-\Re(\varphi_{X_n}(u))|\leq 2$ for all n, we can apply the DCT and see that

$$\lim_{n\to\infty}\int_{[-K^{-1},K^{-1}]}1-\varphi_{X_n}(\xi)d\xi=\int_{[-K^{-1},K^{-1}]}1-\psi(\xi)d\xi$$

And since ψ is continuous at zero and 1 at zero, we have that for large enough k, this right hand side is $\lesssim \epsilon/2$ by the Lemma, for n large enough,

$$\mathbf{P}(\|X_n\|_\infty \geq k) \leq \epsilon$$

We can make k even larger to take care of the finite amount of n's that came before the "for n large enough", and then this expression will holds for all n, so we get tightness. From Proharov's Theorem, we now get a subsequence X_{n_k} such that its laws converge in distribution to the law $\mathscr{L}(X)$ of some random variable X. (This is because every distribution function gives rise to a random variable). Then by the previous part we have that $\varphi_{X_{n_k}} \to \varphi_X$ pointwise, therefore $\varphi_X = \phi$ and so ϕ is the characteristic function of a random variable as we claimed. To show that X_n converges in distribution to X, we suppose that $\mathscr{L}(X_n)$ did not converge weakly to $\mathscr{L}(X)$, then there would exists a subsequence (n_k) and a continuous bounded function f, for which $\mathbf{E}[f(X_{n_k})] \not\to \mathbf{E}[f(X)]$ In particular, for some $\epsilon > 0$

$$|\mathbf{E}[f(X_{n_k})] - \mathbf{E}[f(X)]| > \epsilon$$

for all k. However, by tightness of the $\mathscr{L}(X_{n_k})$ there exists a further subsequence $(X_{n'_k})$ with

$$\mathscr{L}(X_{n'_k}) \to \mathscr{L}(Y)$$

weakly for some random variable Y. By the direct part of this theorem, we have that for all $\xi \in \mathbf{R}^d$,

$$\lim_{k' \to \infty} \varphi_{X_{n_{k'}}}(\xi) \to \varphi_Y(\xi)$$

which means that $\varphi_Y = \varphi_X$, because we had that $\varphi_{X_{n_k}} \to \varphi_X$ pointwise. Since characteristic functions determine distributions, we have that $Y \overset{(d)}{=} X$, and in particular, $\mathbf{E}[f(X)] = \mathbf{E}[f(Y)]$. However, since we also have that $\mathscr{L}(X_{n'_k}) \to \mathscr{L}(Y)$, it means that for k' large enough,

$$|\mathbf{E}[f(X_{n_{k'}}) - \mathbf{E}[f(X)]| = |\mathbf{E}[f(X_{n_{k'}}) - \mathbf{E}[f(Y)]| < \epsilon$$

contradicting our assumption that $\mathscr{L}(X_n)$ did not convrege weakly to $\mathscr{L}(X)$.

♡

4.3 Large Deviation Theory

We now present a self-contained section where we talk about Large Deviation Theory. Recall from the Central Limit Theorem, that if (X_i) is a sequence of iid random variables with finite mean and variance, say μ and σ^2 respectively, then

$$\mathbf{P}(S_n \geq n\mu + a\sigma\sqrt{n}) \to \mathbf{P}(Z \geq a)$$

where $Z \sim N(0,1)$. In this section we want to nail these asymptotics precisely. There are some cases where we can explicitly compute these:

Example 4.19 (Gaussian case) Let $(X_n) \overset{iid}{\sim} N(0,1)$. Then $S_n \sim N(0,n)$, and as such $n^{-1}S_n \sim N(0,n^{-1})$, and since $\frac{1}{\sqrt{n}}X_1 \sim N(0,n^{-1})$, we have that

$$\mathbf{P}\left(\frac{1}{n}S_n \geq a\right) = \mathbf{P}(X_1 \geq a\sqrt{n})$$

And using the fact that the asymptotic behavior of the tail probability of a standard normal is given by

$$\mathbf{P}(X_1 \geq x) \sim \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

We have that

$$\mathbf{P}\left(\frac{1}{n}S_n \geq a\right) \sim \frac{1}{2\sqrt{2\pi n}} \exp\left(\frac{-a^2 n}{2}\right)$$

In other words,

$$-\frac{1}{n}\log \mathbf{P}(S_n \geq an) \to \frac{a^2}{2} =: I(a)$$

The function $I(a)$, sometimes called the rate function, captures (as implied by the name) the rate at which the exponential decay of the probability occurs.

We now have a Theorem by Cramer that generalises this last observation we made in the Gaussian example.

4.4 Cramer's Theorem

Let's recall some definitions:

Definition 4.20 (Moment Generating Function) Let X_1 be a random variable. For $\lambda \geq 0$, the moment generating function of X, $M(\lambda)$ is defined as

$$M(\lambda) = \mathbf{E}[e^{\lambda X_1}]$$

which could be infinite. The log moment generating function $\psi(\lambda)$ is simply defined to be $\psi(\lambda) = \log M(\lambda)$.

Remark 4.21 (Motivation to Cramer's Theorem) Note that by a Chernoff bound:

$$\begin{aligned}
\mathbf{P}(S_n \geq na) &\leq \inf_{\lambda \geq 0} e^{-\lambda n a} \mathbf{E}[\exp(\lambda S_n)] \\
&= \inf_{\lambda \geq 0} (\exp(-\lambda a) M(\lambda))^n \\
&= \inf_{\lambda \geq 0} \exp\big(-n(\lambda a - \psi(\lambda)\big) \\
&= \exp\left(-n \underbrace{\sup_{\lambda \geq 0} (\lambda a - \psi(\lambda))}_{\psi^*(a)}\right)
\end{aligned}$$

Where $\psi^*(a)$ is called the Legendre transform of ψ, and since $\psi(0) = 0$, we have that $\psi^*(a) \geq -\psi(0) = 0$. Rearranging the above expression gives that

$$\liminf_{n \to \infty} -\frac{1}{n} \log \mathbf{P}(S_n \geq an) \geq \psi^*(a)$$

Cramer's Theorem tells us that the limit actually exists, and is equal to $\psi^*(a)$. In other words, $\psi^*(a)$ captures the rate of exponential decay.

Theorem 4.22 (Cramer's Theorem) Let (X_i) be iid random variables with finite mean $\bar{x}$. Then

$$\lim_{n \to \infty} \log \mathbf{P}(S_n \geq na) = \psi^*(a) \qquad a \geq \bar{x}$$

We need the following technical Lemma:

Lemma 4.23 (Continuity and differentiability of $M(\lambda), \psi(\lambda)$) The functions $M(\lambda)$ and $\psi(\lambda)$ are continuous in $D = \{\lambda : M(\lambda) < \infty\}$ and differentiable in D° with

$$M'(\lambda) = \mathbf{E}[X_1 e^{\lambda X_1}] \qquad \psi'(\lambda) = \frac{M'(\lambda)}{M(\lambda)}$$

Proof. For continuity, we take a sequence $\lambda_n \to \lambda$ in D. Since $\lambda \in D$, we have that $\sup_n \lambda_n \in D$ (convergent sequence is bounded, which means that either the supremum is attained by some λ_n, or the supremum is λ, in each case the supremum is in D). Therefore we have that

- $|\exp(\lambda_n X_1)| \to |\exp(\lambda X_1)|$.
- $|\exp(\lambda_n X_1)| \leq |\exp(\sup_n \lambda_n X_1)|$, and since $\sup_n \lambda_n \in D$, we have that the integral of the latter function is bounded

We may therefore apply the DCT, and see that

$$\lim_{n\to\infty} M(\lambda_n) = M(\lambda)$$

Thus establishing continuity of $M(\lambda)$, and hence continuity of $\log M(\lambda)$. To check differentiability, choosing $\eta \in D^\circ$, we will also use the DCT to swap limit and expectations in

$$\lim_{\epsilon\to 0} \frac{M(\eta+\epsilon)-M(\eta)}{\epsilon} = \lim_{\epsilon\to 0} \mathbf{E}\left[\frac{\exp((\eta+\epsilon)X_1)-\exp(\eta X_1)}{\epsilon}\right]$$

[Justify why we have that]

$$\left|\frac{e^{(\eta+\epsilon)X_1}-e^{\eta X_1}}{\epsilon}\right| \leq e^{\eta X_1}\left(\frac{e^{\delta|X_1|}-1}{\delta}\right)$$

And since $\eta \in D^\circ$, for a sufficiently small δ, we have that $\eta+\delta \in D$
Makes no sense. Come back later ♡

Proof of Cramer's Theorem. Recall from our initial discussion that we have already shown that

$$\liminf_n -\frac{1}{n}\log \mathbf{P}(S_n \geq na) \geq \psi^*(a)$$

Replacing X_i by $\widetilde{X_i} = X_i - a$, we get that

$$\mathbf{P}(S_n \geq na) = \mathbf{P}(\widetilde{S_n} \geq 0)$$

and that

$$\widetilde{M}(\lambda) = e^{-\lambda a} M(\lambda)$$

so that

$$\widetilde{\psi}(\lambda) = \psi(\lambda) - \lambda a$$

Thus we need to show that, dropping tildes,

$$\liminf_n \frac{1}{n}\mathbf{P}(S_n \geq 0) \geq \inf_{\lambda\geq 0}\psi(\lambda)$$

whenever $\bar{x} < 0$. We distinguish two cases:

- $\mathbf{P}(X_1 \leq 0) = 1$, in which case

$$\begin{aligned}\inf_{\lambda \geq 0} \psi(\lambda) &\leq \lim_{\lambda \to \infty} \psi(\lambda) \\ &= \log \mathbf{E}\Big[\lim_{\lambda \to \infty} e^{\lambda X_1}(\mathbf{1}(X_1 = 0) + \mathbf{1}(X_1 < 0))\Big] \\ &= \log \mathbf{P}[X_1 = 0]\end{aligned}$$

  And This somehow shows that

$$\liminf \frac{1}{n} \log \mathbf{P}(S_m \geq 0) \geq \inf_{\lambda \geq 0} \psi(\lambda)$$

  (Yes, look at Chua's notes)

- $\mathbf{P}(X_1 > 0) > 0$:

Assume however that $M(\lambda)$ exists for all λ. The idea is to modify X_1 so that it has mean zero. We do so by introducing a distribution μ^θ with density relative to $\mu = \mu_{X_1}$ given by

$$\frac{d\mu^\theta}{d\mu}(x) = \frac{e^{\theta x}}{M(\theta)}$$

And now we define the function

$$g(\theta) = \mathbf{E}_\theta[X_1] = \int x \, d\mu^\theta(x)$$

The claim is that $g(\theta)$ is continuous. Indeed:

$$g(\theta) = \frac{\int x e^{\theta x} d\mu(x)}{M(\theta)}$$

We showed that $M(\theta)$ was continuous in θ and One can also show that the numerator is continuous in θ using DCT. Of course $g(0) = \mathbf{E}[X_1] = \bar{x} < 0$. Now note that

$$\lim_{\theta \uparrow \infty} g(\theta) > 0$$

Because the denominator is always non-negative, and for the integrand of the numerator, we can split the integral into the integral over the set where x takes negative values, and in this case the limit $\theta \to \infty$ kills this part, and then the other part which will be strictly greater than zero. Therefore by the intermediate value theorem, there exists some θ such that $g(\theta) = \mathbf{E}_\theta[X_1] = 0$.

Now we have that

$$\mathbf{P}(S_n \geq 0) \geq \mathbf{P}(S_n \in [0, \epsilon n]) \geq \mathbf{E}\left[e^{\theta(S_n - n\epsilon)} \mathbf{1}(S_n \in [0, \epsilon n])\right]$$

Now since we defined

$$\frac{d\mu^\theta}{d\mu}(x) = \frac{e^{\theta x}}{M(\theta)}$$

we have that

$$\mathbf{E}\left[e^{\theta X_1} \mathbf{1}(A)\right] = M(\theta)\mathbf{P}_\theta(A)$$

so

$$\mathbf{E}\left[e^{\theta(S_n - n\epsilon)} \mathbf{1}(S_n \in [0, \epsilon n])\right] = M(\theta)^n \mathbf{P}_\theta(S_n \in [0, \epsilon n]) e^{-\theta \epsilon n}$$

Under θ, each X_i has mean zero (this is the way we have constructed θ), and as such $\mathbf{P}_\theta(S_n \in [0, \epsilon n]) \to 1/2$ by the CLT. Therefore

$$\liminf_n \frac{1}{n} \log \mathbf{P}(S_n \geq 0) \geq \phi(\theta) - \theta\epsilon$$

and so taking $\epsilon \downarrow 0$ finishes the claim for the case when $M(\lambda)$ exists for all λ.

Let $\mu_n = \mathscr{L}(S_n)$ and let ν be the law of X_1 conditioned on the event $\{|X_1| \leq K\}$, i.e:

$$\nu(A) = \mathbf{P}(A \mid \{|X_1| \leq K\})$$

Define also ν_n to be the law of S_n conditioned on the event $\bigcap_i \{|X_i| \leq K\}$. Then we have that

$$\mu_n[0, \infty) \geq \nu_n[0, \infty)\mu[-K, K]^n$$

Indeed: $\mathbf{P}(S_n \in [0, \infty)) \geq \mathbf{P}(S_n \in [0, \infty)$ and $|X_i| \leq K \ \forall i) = RHS$. We can now write

$$\psi_K(\lambda) = \log \int_{-K}^{K} e^{\lambda x} d\mu(x)$$

Now note that

$$\log \int_{-\infty}^{\infty} e^{\lambda x} d\nu(x) = \log\left(\frac{\int_{-\infty}^{\infty} e^{\lambda x} \mathbf{1}(|x| \leq K) d\mu(x)}{\mathbf{P}(\{|X_1| \leq K\})}\right) = \psi_K(\lambda) - \log \mu[-K, K]$$

Then we have that

$$\liminf_n \frac{1}{n} \log \mu_n([0, \infty)) \geq \liminf \frac{1}{n} \log \nu_n[0, \infty) + \log \mu[-K, K] \qquad (\star)$$

But we note that $\frac{1}{n}\log \nu_n[0,\infty) \geq \inf_{\lambda\geq 0}\log\int_{-\infty}^{\infty} e^{\lambda x}\,d\,\nu(x)$. Therefore

$$\liminf_n \frac{1}{n}\log \mu_n[0,\infty) \geq \inf_{\lambda\geq 0}\psi_K(\lambda) =: J_K$$

Taking $K\to\infty$, we have that $\psi_K(\lambda)\to\psi(\lambda)$, and so $J_K\to J$ for some J. Since each ψ_K is continuous, then $\{\lambda : \psi_K \leq J\}$ are non-empty, compact and nested in K, so by Cantor's Theorem there is some

$$\lambda_0 \in \bigcap_K \{\lambda : \psi_K \leq J\}$$

So

$$J \geq \sup_K \psi_K(\lambda_0) = \psi(\lambda_0) \geq \inf_{\lambda\geq 0}\psi(\lambda)$$

as required. ♡

5 Brownian motion

I should write something nice here one day

Definition 5.1 (Brownian motion) Let $B = (B_t : t \geq 0)$ be a continuous process in $\mathbf{R}^d$. We say that B is Brownian Motion started from $x \in \mathbf{R}^d$ if:

- $B_0 = x$ a.s,
- $B_t - B_s \sim N(0, (t-s)I)$ for all $t > s$.
- B has independent increments, independent of B_0. I.e: for all $t_1 < t_2 < \cdots < t_n$, the random variables

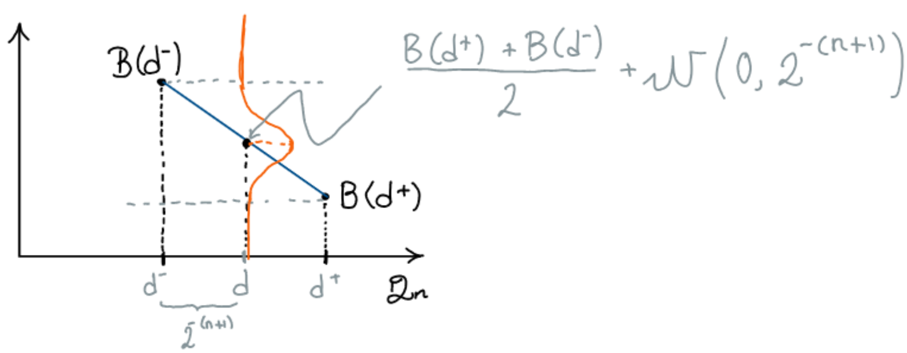
$$B_{t_1} - B_0, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}}$$

are all independent.

5.1 Wiener's Theorem

Theorem 5.2 (Wiener's Theorem) Brownian motion exists on some probability space.

Main idea: There are three steps, the first one is to construct inductively Brownian motion on the dyadics as follows: once B is defined on $\mathscr{D}_n$, define $B(d)$ for $d \in \mathscr{D}_{n+1}$ by grabbing the halfway value between $B(d^-)$ and $B(d^+)$ and adding a Gaussian fluctuation of the corresponding width.



Once this is done, one can check without too much trouble that this satisfies the properties of Brownian Motion on $\mathscr{D} = \bigcup_n \mathscr{D}_n$. Then one employs Kolmogorov's Criterion to show that Brownian Motion on the dyadics is indeed continuous, and so we use this to extend it to the whole interval $[0, 1]$. After this, we must check, via characteristic functions that this extension is indeed Brownian Motion too.

Proof. We first construct Brownian motion on the interval $[0,1]$. Then we extend to the whole real line and then to higher dimensions.

Let $\mathscr{D}_0 = \{0,1\}$, and $\mathscr{D}_n = \{k2^{-n} : 0 \le k \le 2^n\}$ for $n \ge 1$. Finally let $\mathscr{D} = \bigcup_n \mathscr{D}_n$ be the set of dyadic rational numbers in $[0,1]$. Now consider a sequence of independent $\mathscr{N}(0,1)$ random variables indexed in the dyadics: $(Z_d : d \in \mathscr{D})$ on some probability space. We will first construct Brownian motion on the dyadics by induction, and then extend to the Real line.

Set $B_0 = 0$, and let $B_1 = Z_1$. Now suppose by induction that we have constructed $(B_d : d \in \mathscr{D}_{n-1})$. Let us construct $(B_d : d \in \mathscr{D}_n)$ as follows. Take $d \in \mathscr{D}_n \setminus \mathscr{D}_{n-1}$ and let $d^- = d - 2^{-n}$ and $d^+ = d + 2^{-n}$. Therefore d^- and d^+ are consecutive dyadics in $\mathscr{D}_{n-1}$. We construct now:

$$B_d = \frac{B_{d^-} + B_{d^+}}{2} + \frac{Z_d}{2^{(n+1)/2}}$$

And show that this satisfies the Brownian properties:

- **Independence of increments**: We first note that $B_d - B_{d^-}$ and $B_{d^+} - B_d$ are independent. Indeed, letting $N_d = \frac{B(d^+) - B(d^-)}{2}$ and $N_{d'} = \frac{Z_d}{2^{(n+1)/2}}$:

$$\begin{aligned}\mathrm{Cov}(B_d - B_{d^-}, B_{d^+} - B_d) &=: \mathrm{Cov}\left(N_d + N_d', N_d - N_d'\right) \\ &\overset{(1)}{=} \mathrm{Var}(N_d) - \mathrm{Var}\left(N_d'\right) \\ &\overset{(2)}{=} 0\end{aligned}$$

  Indeed: By using the induction hypothesis on $B_{d^+} - B_{d^-}$, we can easily check that $N_d \sim \mathscr{N}(0, 2^{-n-1})$ and similarly, by construction of Z_d, $N_d' \sim \mathscr{N}(0, 2^{-n-1})$ and moreover, these are independent, so from this comes (1) and then (2) comes from the calculations on the variances we have just done. (Recall that for Gaussians, zero covariance implies independence). Then, once we have established this, we have that all increments $(B_d - B_{d-2^{-n}} : d \in \mathscr{D}_n)$ are independent. If they are consecutive, we have just proven it. Otherwise, we simply consider the consecutive increments in between and then note that since they are all pairwise independent, and these random variables are Gaussian, the whole collection is independent. his now easily extends to increments that are not just of 2^{-n} width.

- **Gaussian distribution of increments**: This is also quite easy, consider $t_1, t_2 \in \mathscr{D}_n$, then $B_{t_1} - B_{t_2}$ can be written as a big telescoping sum of increments of width 2^{-n}, and since each has variance 2^{-n-1} and there are $(t_2 - t_1)2^n$ of them we get the result.

Thus $(B_t : t \in \mathscr{D})$ satisfies the Brownian properties. Let $s \le t \in \mathscr{D}$, and notice that for each $p > 0$, since $B_t - B_s \sim \mathcal{N}(0, t-s) = (t-s)^{1/2}\mathcal{N}(0,1)$:

$$\mathbf{E}[|B_t - B_s|^p] = |t-s|^{p/2}\mathbf{E}[|N|^p]$$

where $N \sim \mathcal{N}(0,1)$. Since N has moments of all orders, by Kolmogorov's continuity criterion, it follows that $(B_d : d \in \mathcal{D})$ is α-Hölder continuous for all $\alpha < 1/2$ almost surely. Hence to extend to the interval $[0,1]$, we set

$$B_t = \lim_{i\to\infty} B_{d_i}$$

where $d_i \to t$ is a sequence of dyadics. From this definition, $(B_t, t \in [0,1])$ is also α-Hölder continuous for all $\alpha < 1/2$. Now one needs to check that $(B_t : t \in [0,1])$ is Brownian motion. For this, we let $0 = t_0 < t_1 < \cdots < t_k$ and let $0 = t_0^n \le t_1^n \le \cdots \le t_k^n$ be dyadic rationals converging to their respective t_i. Let $\mathbf{B} = (B_{t_1} - B_{t_0}, \cdots, B_{t_k} - B_{t_{k-1}})$ and similarly define $\mathbf{B}_n$. Then by considering the characteristic functions, we have that

$$\begin{aligned}
\mathbf{E}[\exp(i\langle \mathbf{u}, \mathbf{B}\rangle] &\overset{(1)}{=} \mathbf{E}[\lim_{n\to\infty} \exp(i\langle \mathbf{u}, \mathbf{B}_n\rangle]\\
&\overset{(2)}{=} \lim_{n\to\infty} \mathbf{E}[\exp(i\langle \mathbf{u}, \mathbf{B}_n\rangle]\\
&= \lim_{n\to\infty} \prod_{j=1}^k \exp(-(t_j^n - t_{j-1}^n)u_j^2/2)\\
&= \prod_{j=1}^k \exp(-(t_j - t_{j-1})u_j^2/2)
\end{aligned}$$

Where (1) comes from the continuity Brownian motion, 2 comes from the DCT, and so we get that the Characteristic function of the vector of differences is that of a vector of independent Gaussian random variables with respective variances $t_j - t_{j-1}$, which means that the distributions are in fact equal. This shows the Brownian properties in $[0,1]$. Now to extend to $\mathbf{R}_+$ and higher dimensions, you glue a lot of BM together but I have spent already enough on this proof so that's that. ♡

Remark 5.3 (α-Hölder continuity) Wiener's Theorem also gives that Brownian paths are a.s α-Hölder continuous for all $\alpha < 1/2$. However, a.s there exists no interval $[a,b]$ for which Brownian motion if Hölder continuous with $\alpha \ge 1/2$.

5.2 Invariance Properties

We have the following invariance properties of Brownian motion which follow straight from the definition using very simple properties:

Proposition 5.4 Let B be standard Brownian motion in $\mathbf{R}^d$.

- **Rotation:** If $U^T U = I$, i.e: U is an orthonormal matrix, then $UB = (UB_t, t \geq 0)$ is again Brownian motion. In particular, $-B$ is also standard Brownian motion.
- **Rescaling** If $\lambda > 0$, then $\left(\frac{1}{\sqrt{\lambda}} B_{\lambda t} : t \geq 0\right)$ is standard Brownian motion.
- **Simple Markov Property** For every $s \geq 0$, the shifted process $(B_{t+s} - B_s : t \geq 0)$ is standard Brownian motion independent of $\mathscr{F}_s^B = \sigma(B_u : u \leq s)$.

Main idea: The rotation property is a simple fact of Gaussian processes: that if $W \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, then $UW \sim \mathcal{N}(\mathbf{0}, U\mathbf{\Sigma}U^T)$ In particular, when $\mathbf{\Sigma} = \mathbf{I}$, we get the claim. Rescaling property follows form the way variance scales. The simple Markov Property follows by the property of independence of increments, $B_{t+h} - B_t$ is independent of all B_s for $s \leq t$.

A slightly more tricky property is the following:

Theorem 5.5 (Time inversion) Let $(B_t : t \geq 0)$ be standard Brownian motion, then the process defined by $(X_t : t \geq 0)$ defined by

$$X_t = \begin{cases} 0 & t = 0 \\ t B_{1/t} & t > 0 \end{cases}$$

is standard Brownian motion.

Let us first note a useful fact:

Lemma 5.6 Brownian motion is a Gaussian process, i.e: the vector $(B_{t_1}, \cdots, B_{t_n})$ is a Gaussian vector for all $t_1 < \cdots < t_n$, with covariances:

$$\mathrm{Cov}(B_t, B_s) = t \wedge s$$

Proof. We know that

$$(B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}})$$

is a Gaussian vector, and moreover, it is the image under a linear isomorphism of the vector

$(B_{t_1}, \cdots, B_{t_n})$. Therefore it is Gaussian. To show the covariances:

$$\text{Cov}(B_t, B_s) = \text{Cov}(B_t - B_s + B_s, B_s) = \underbrace{\text{Cov}(B_t - B_s, B_s)}_{0 \quad (\perp)} + \text{Cov}(B_s, B_s) = s$$

♡

Main idea: The key idea here is that if you want to show that some process (X_t) is Brownian motion, all you have to do is show that the collections $(X_{t_1}, \cdots, X_{t_n})$ are Gaussian random vectors with the same mean and covariance structure as that of the corresponding random vector of Brownian motion $(B_{t_1}, \cdots, B_{t_n})$. This is because from ES2 Q2.1 We know that the law of a Gaussian random vector is uniquely characterised by its mean and covariance structure, and if two processes have the same finite dimensional distributions, then they have the same law.

Proof of time inversion. A useful property to remember is that the law of a process is uniquely determined by its finite dimensional distributions. In our case, the finite dimensional distributions

$$(B_{t_1}, \cdots, B_{t_n})$$

of Brownian motion are Gaussian random vectors, and therefore characterized by their means $\mathbf{E}[B_{t_i}] = 0$ and covariances $\text{Cov}\big(B_{t_i}, B_{t_j}\big) = t_i \wedge t_j$. Thus if we can show that $(X_{t_1}, \cdots, X_{t_n})$ is a Gaussian random vector with the same means and covariances we are done. To show that it is a random vector, we simply note that

$$(t_1 B_{1/t_1}, \cdots, t_n B_{1/t_n})$$

is the image under a linear isomorphism of the random vector

$$(B_{1/t_1}, \cdots, B_{1/t_n})$$

which is Gaussian. It is clear that $\mathbf{E}[t_i B_{1/t_i}] = t_i \mathbf{E}[B_{1/t_i}] = 0$ so all we need to understand now is the covariances. Say $s \le t$

$$\text{Cov}\big(s B_{1/s}, t B_{1/t}\big) = s\, t\, \text{Cov}\big(B_{1/s}, B_{1/t}\big) = s\, t \cdot \frac{1}{t} = s$$

as required. All left to show is continuity. For $t > 0$ there is no question, so it remains to show that (X_t) is continuous at $t = 0$. We have just determined that (X_t) and (B_t) have the same laws as processes, so let (q_n) be a sequence of rationals with $q_n \downarrow 0$. Since the laws agree, we have that $X_{q_n} \stackrel{d}{=} B_{q_n}$. By continuity of (B_t), we have that $B_{q_n} \downarrow 0$ a.s. Now we use a fact that says that if two sequences of random variables (X_n) and (Y_n) are equal in distribution for each n and $Y_n \to c$ a.s, then $X_n \to c$ a.s as well. ♡

Corollary 5.7 (Law of Large numbers) Almost surely, $\lim_{t\to\infty} \frac{B_t}{t} = 0$

Main idea: Immediate application of time inversion along with the fact that by continuity of Brownian motion $\lim_{t\downarrow 0} B_t = 0$ almost surely.

Proof. From time inversion, we know that $tB_{1/t} \overset{(d)}{=} B_t$, which means that

$$\mathbf{P}\left(\lim_{t\to\infty} \frac{B_t}{t} = 0\right) = \mathbf{P}\left(\lim_{t\downarrow 0} tB_{1/t} = 0\right) = \mathbf{P}\left(\lim_{t\downarrow 0} B_t = 0\right) = 1$$

♡

Remark 5.8 An alternative proof to this, which gives more intuition as to why this is called "Law of Large numbers", is that we can also phrase:

$$\lim_{n\to\infty} \frac{B_n}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \underbrace{B_i - B_{i-1}}_{\overset{\text{iid}}{\sim} \mathcal{N}(0,1)}$$

We now consider the following σ-algebra, which lets you look *slightly into the future*.

Definition 5.9 (Slightly augmented σ-algebra) Let $(\mathscr{F}_t^B : t \geq 0)$ be the natural filtration of $(B_t : t \geq 0)$ and $\mathscr{F}_s^+$ be the σ-algebra given by

$$\mathscr{F}_s^+ = \bigcap_{t>s} \mathscr{F}_t^B$$

Remark 5.10 The simple Markov property of Brownian motion tells us that $B_{t+s} - B_s$ is independent of $\mathscr{F}_s^B$. Clearly $\mathscr{F}_s^B \subseteq \mathscr{F}_s^+$ for all s, since in $\mathscr{F}_s^+$ we allow an infinitesimal glance into the future. The next theorem says that $(B_{t+s} - B_s)$ is still independent of $\mathscr{F}_s^+$

Theorem 5.11 For every $s \geq 0$, the process $(B_{t+s} - B_s : t \geq 0)$ is independent of $\mathscr{F}_s^+$.

Main idea: Recall the characterisation of a process being independent of a sigma-algebra that is discussed in the Appendix. Then, let $(s_n) \downarrow s$. It follows that for any $A \in \mathscr{F}_s^+$, A will also be in $\mathscr{F}_{s_n}$ for all n, and since $(B_{t+s_n} - B_{s_n})$ is a process independent of $\mathscr{F}_{s_n}$ (this is the Simple MP), we will be able to separate expectations. Then some fiddling with DCT gives it.

Proof. We start by choosing any $A \in \mathscr{F}_s^+$, and in light of the discussion in the Appendix about independence, we want to show that for any $f \in C_b\left(\left(\mathbf{R}^d\right)^m\right)$ and any $t_1, t_2, \cdots, t_m > 0$,

$$\mathbf{E}[f(B_{t_1+s} - B_s, B_{t_2+s} - B_s, \cdots, B_{t_m+s} - B_s)\mathbf{1}(A)] = \mathbf{P}[A]\mathbf{E}[f(B_{t_1+s} - B_s, B_{t_2+s} - B_s, \cdots, B_{t_m+s} - B_s)\mathbf{1}(A)]$$

To do this, we will choose a sequence $(s_n) \downarrow s$, then by continuity of Brownian motion we have that

$$\lim_{n\to\infty} B_{t+s_n} - B_{s_n} = B_{t+s} - B_s$$

Now everything follows nicely.

$$\begin{aligned}
\mathbf{E}[f(B_{t_1+s} - B_s, B_{t_2+s} - B_s, \cdots, B_{t_m+s} - B_s)\mathbf{1}(A)] &\overset{(1)}{=} \lim_{n\to\infty} \mathbf{E}[f(B_{t_1+s_n} - B_{s_n}, B_{t_2+s_n} - B_{s_n}, \cdots, B_{t_m+s_n} - B_{s_n})\mathbf{1}(A)] \\
&\overset{(2)}{=} \lim_{n\to\infty} \mathbf{P}[A]\mathbf{E}[f(B_{t_1+s_n} - B_{s_n}, B_{t_2+s_n} - B_{s_n}, \cdots, B_{t_m+s_n} - B_{s_n})] \\
&\overset{(3)}{=} \mathbf{P}[A]\mathbf{E}[f(B_{t_1+s} - B_s, B_{t_2+s} - B_s, \cdots, B_{t_m+s} - B_s)\mathbf{1}(A)]
\end{aligned}$$

Where in (1) we have used the DCT to pull the limit outside of the expectation, in (2) we have noted that since $\mathscr{F}_s^+ \subseteq \mathscr{F}_{s_n}$ (this is because we assumed $s_n \downarrow s$), it follows that $A \in \mathscr{F}_{s_n}$ and so by the Markov property, $(B_{t+s_n} - B_{s_n} : t \geq 0)$ is a process independent of $\mathscr{F}_{s_n}$ which means that we can pull $\mathbf{1}(A)$ out of the expectation. Step (3) follows by using DCT. ♡

Theorem 5.12 (Blumenthal's $0-1$ Law) The σ-algebra $\mathscr{F}_0^+$ is trivial, i.e: if $A \in \mathscr{F}_0^+$, then $\mathbf{P}(A) \in \{0, 1\}$

Main idea: Any set $A \in \mathscr{F}_0^+$ is $\sigma(B_t : t \geq 0)$ measurable. However, by a previous result $(B_t : t \geq 0)$ is independent of $\mathscr{F}_0^+$. Therefore A is independent of itself.

Proof. Let $A \in \mathscr{F}_0^+$. Then $A \in \sigma(B_t : t \geq 0)$ (Yes: if $A \in \mathscr{F}_0^+$, then $A \in \sigma(B_s : 0 \leq s \leq t$ for all $t > 0$, of course, these sigma algebras are contained in $\sigma(B_s : 0 \leq s)$) . But by the previous theorem we know that the process $(B_t : t \geq 0)$ is independent of $\mathscr{F}_0^+$, i.e: A is independent of $\mathscr{F}_0^+$ and in particular A is independent of itself. ♡

Now we have some remarkable property of Brownian motion in one dimension, it showcases how erratically it oscillates immediately after it starts running:

Theorem 5.13 Let (B_t) be standard Brownian motion in 1-d. Let $\tau = \inf\{t > 0 : B_t > 0\}$ and $\sigma = \inf\{t > 0 : B_t = 0\}$. Then

$$\mathbf{P}(\tau = 0) = \mathbf{P}(\sigma = 0) = 1.$$

Main idea: The key idea is that the event $\{\tau = 0\}$ is something that depends on the immediate future of Brownian Motion after it starts, it will be an $\mathscr{F}_0^+$-measurable event and so it will have probability zero or one. Then by using the fact that $B_t \sim \mathcal{N}(0,t)$ you can show that $\mathbf{P}[\tau = 0] > 0$. To show $\mathbf{P}[\sigma = 0] = 1$, use the intermediate value Theorem.

Proof. The strategy to show these type of claims is to show that the event $\{\tau = 0\}$ is in $\mathscr{F}_0^+$, and then show that it has positive probability. We start by noting that

$$\{\tau = 0\} = \bigcap_{n \geq 1} \underbrace{\left\{ \text{for some } 0 < \epsilon < \frac{1}{n}, B_\epsilon > 0 \right\}}_{\in \mathscr{F}^B_{\frac{1}{n}}}$$

Therefore $\{\tau = 0\} \in \mathscr{F}^B_{1/n}$ for all n, and as such $\{\tau = 0\} \in \mathscr{F}_0^+$. Now we show that its probability is not zero. To do this, we note that

$$\begin{aligned} \mathbf{P}(\tau = 0) &= \mathbf{P}\left(\bigcap_n \{\tau \leq 1/n\} \right) \\ &= \lim_{n \to \infty} \mathbf{P}(\tau \leq 1/n) \\ &\geq \lim_{n \to \infty} \mathbf{P}(B_{1/n} > 0) = 1/2 \end{aligned}$$

Where the inequality comes from the fact that if $B_{1/n} > 0$, then the first time that Brownian motion goes above zero must be $\leq 1/n$. The final equality comes from the fact that $B_{1/n} \sim \mathcal{N}(0, 1/n)$, which is symmetric about zero. The one for σ goes as follows. By a symmetric argument, we have that $\inf\{t > 0 : B_t < 0\} = 0$ almost surely. Therefore, by the intermediate value theorem and the fact that B is continuous, we get that $\mathbf{P}[\sigma = 0] = 1$. ♡

Proposition 5.14 (Brownian motion oscillates wildly near the origin) Let (B_t) be Standard Brownian Motion in one-dimension. Define S_t and I_t to be the running supremum and infimum up to time t. Then for any $\epsilon > 0$ one has that almost surely:

$$S_\epsilon > 0 \qquad \text{and} \qquad I_\epsilon < 0$$

In particular, due to continuity of Brownian motion, there exists a zero in any interval $(0, \epsilon)$.

Main idea: The key is to let $(t_n) \downarrow 0$ be a sequence, and then show that $\mathbf{P}[B_{t_n} > 0 \text{ i.o }] = 1$. To do this use the fact that $\{B_{t_n} > 0 \text{ i.o }\} = \{\limsup_{n\to\infty} B_{t_n} > 0\}$ and by the Reverse Fatou Lemma bound this latter quantity by $1/2$. Then show that $\{B_{t_n} > 0 \text{ i.o }\}$ is $\mathscr{F}_0^+$ measurable, which makes sense at it depends only on the immediate future after the start of BM.

Proof. Let $\epsilon > 0$ be given, we will show that with probability one, there will be some $t_n < \epsilon$ for which $B_{t_n} > 0$, thus establishing the first claim. Indeed: let (t_n) be a sequence with $t_n \downarrow 0$. Then

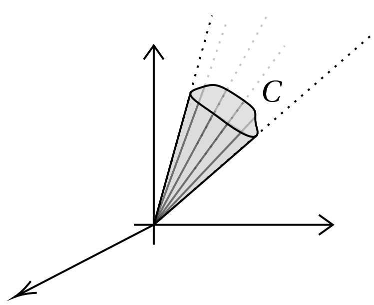
$$\begin{aligned}\mathbf{P}(B_{t_n} > 0 \text{ infinitely often}) &= \mathbf{P}\left(\limsup_{n\to\infty} B_{t_n} > 0\right) \\ &\geq \limsup_{n\to\infty} \mathbf{P}(B_{t_n} > 0) = \frac{1}{2}\end{aligned}$$

Moreover, since

$$\{B_{t_n} > 0 \text{ infinitely often}\} = \bigcap_{n\geq 0} \underbrace{\bigcup_{m\geq n} \{B_{t_m} > 0\}}_{\in \mathscr{F}^B_{t_n} \quad (\star)}$$

Where $(\star)$ follows because the event "for some $m \geq n$, $B_{t_m} > 0$" can be determined by knowing the process up to time t_n (since (t_n) is a decreasing sequence). Therefore it follows that $\{B_{t_n} > 0 \text{ infinitely often}\} \in \mathscr{F}_0^+$ and so by Blumenthal's Law, we have that $S_\epsilon > 0$ with probability one. The second claim follows similarly by plugging in minus signs, which will change a sup into an inf, but then using that $-B_t \overset{(d)}{=} B_t$ finishes the claim. ♡

Now we have one last property of the behavior of Brownian motion near the origin, which has to do with Brownian motion hitting cones



Proposition 5.15 (Brownian motion hits cones immediately) Let (B_t) be Standard Brownian motion in $\mathbf{R}^d$ and let C be a cone in $\mathbf{R}^d$, i.e: $C = \{ut : t > 0, u \in A\}$ where A is any non-empty open subset of the unit sphere in d-dimensions, then letting $T_C = \inf\{t \geq 0 : B_t \in C\}$, we have that $T_C = 0$ almost surely.

Proof. First we note that since the cone expands to infinity, it is scaling invariant in the sense that for any $\lambda > 0$, $\lambda C = C$, so for any $t > 0$,

$$\mathbf{P}(B_t \in C) = \mathbf{P}\left(\frac{B_t}{\sqrt{t}} \in C\right) = \mathbf{P}(B_1 \in C)$$

But since C has non-empty interior, this last probability is just $\mu(C)$ where μ is the standard Gaussian measure on $\mathbf{R}^d$, and hence $\mu(C) > 0$. Therefore for any t, $\mathbf{P}(B_t \in C) > 0$. We will not show this again explicitly, but since the event $\{T_C = 0\}$ depends on the immediate behavior of the Brownian motion after $t = 0$, it is in $\mathscr{F}_0^+$ (Indeed the argument is verbatim the same we used to show that the event $\{\tau_0 = 0\} \in \mathscr{F}_0^+$, where τ_0 is the return time of zero). To show that $\mathbf{P}(T_C = 0) = 1$ it suffices to show (once again by Blumenthal) that $\mathbf{P}(T_C = 0) > 0$, this is once again a standard argument:

$$\begin{aligned}\mathbf{P}(T_C = 0) &= \mathbf{P}\left(\bigcap_{n \geq 0}\left\{T_C \leq \frac{1}{n}\right\}\right)\\ &= \lim_{n\to\infty} \mathbf{P}\left\{T_C \leq \frac{1}{n}\right\}\\ &\geq \lim_{n\to\infty} \mathbf{P}\left(B_{\frac{1}{n}} \in C\right) > 0\end{aligned}$$

♡

The moral of the property above is that when you start Brownian motion in $\mathbf{R}^d$ it must hit all of its neighbouring regions, rather than move away from the origin immediately in a particular direction, hence avoiding a particular cone. Now we have explored some properties of Brownian motion near the origin, let us give one property of Brownian motion as $t \to \infty$.

Proposition 5.16 (Brownian motion oscillates between $\pm\infty$ for $t \to \infty$) Let (B_t) be Standard Brownian motion in one dimension. Then almost surely

$$\sup_{t\geq 0} B_t = \infty \qquad \text{and} \qquad \inf_{t \geq 0} B_t = -\infty$$

Proof. For this we will utilise the scaling invariance of Brownian motion. In particular, we note that for any $\lambda > 0$

$$\sup_{t\geq 0} B_t = \sup_{t\geq 0} B_{\lambda t} \overset{(d)}{=} \sqrt{\lambda} \sup_{t\geq 0} B_t$$

Therefore, for any $x > 0$,

$$\mathbf{P}\left(\sup_{t\geq 0} B_t > x\right) = \mathbf{P}\left(\sup_{t\geq 0} B_t > \frac{1}{\sqrt{\lambda}} x\right)$$

which means that for any $x>0$, the probability that $\sup_{t\geq 0} B_t > x$ is a constant. We can now use this as follows:

$$\begin{aligned}\mathbf{P}\left(\sup_t B_t = \infty\right) &= \lim_{n\to\infty} \mathbf{P}\left(\sup_t B_t > n\right)\\ &\overset{(!)}{=} \lim_{n\to\infty} \mathbf{P}\left(\sup_t B_t > \frac{1}{n}\right)\\ &= \mathbf{P}\left(\sup_t B_t > 0\right) = 1.\end{aligned}$$

Where the step (!) came from the fact that for any $x>0$, $\mathbf{P}\big(\sup_t B_t > x\big)$ is constant. The second claim follows immediately now. ♡

5.3 Strong Markov Property

Definition 5.17 (($\mathscr{F}_t$)-adapted Brownian motion) Let $B=(B_t)$ be Brownian motion on some probability space. We say that B is ($\mathscr{F}_t$)-adapted if B_t is $\mathscr{F}_t$-measurable and if for all $s\geq 0$, the process:

$$(B_{t+s}-B_s : t\geq 0)$$

is independent of $\mathscr{F}_s$. (In some textbooks this is referred to as $\{\mathscr{F}_t\}$ being admissible)

Theorem 5.18 (Strong Markov Property ☕) Let (B_t) be Brownian motion and T an almost surely finite stopping time. Then

$$(B_{T+t}-B_T : t\geq 0)$$

is standard Brownian motion independent of $\mathscr{F}_T^+$.

Main idea: Discretize T into the usual $T_n = 2^{-n}[2^n T]$, and so T_n takes values of the form $k2^{-n}$. It is easy to see that $B_t^{(k)} = B_{t+k2^{-n}} - B_{k2^{-n}}$ is Brownian motion (Simple Markov Property) and so one can show that $B_{t+T_n}-B_{T_n}$ is Brownian motion independent of $\mathscr{F}_{T_n}^+$. One then passes limits.

Proof. We proceed as usual by approximating the stopping time from above by a stopping time that takes on discrete values. We shall show the Strong Markov Property holds for these stopping times and then use continuity to finish the claim.

We start by defining the Brownian motion $B_t^{(k)} = B_{t+k2^{-n}} - B_{k2^{-n}}$ (this is indeed Brownian motion by the Simple Markov Property). Now we discretize the stopping time T into the sequence $T_n = 2^{-n}[2^n T]$ which has the property $T_n \downarrow T$ (Since $T<\infty$ almost surely). Now define the object $B_t^* = B_{t+T_n} - B_{T_n}$ (it is not yet clear whether this is Brownian motion, the only thing that is clear is that it is continuous). First we show that this process, whatever it may be, is independent of $\mathscr{F}_{T_n}^+$, and then we show it is Brownian motion. Consider any event of the form $\{B^* \in A\}$ where A is in the sigma algebra given to $C\left(\mathbf{R}_+, \mathbf{R}^d\right)$ and let $E \in \mathscr{F}_{T_n}^+$. We can write the following:

$$\begin{aligned}
\mathbf{P}(\{B^* \in A\} \cap E) &\overset{(1)}{=} \sum_{k=0}^{\infty} \mathbf{P}\left(\{B^* \in A\} \cap E \cap \{T_n = k2^{-n}\}\right) \\
&\overset{(2)}{=} \sum_{k=0}^{\infty} \mathbf{P}(\{B^{(k)} \in A\} \cap E \cap \{T_n = k2^{-n}\}) \\
&\overset{(3)}{=} \sum_{k=0}^{\infty} \mathbf{P}(B^{(k)} \in A)\mathbf{P}(E \cap \{T_n = k2^{-n}\}) \\
&\overset{(4)}{=} \sum_{k=0}^{\infty} \mathbf{P}(B \in A)\mathbf{P}(E \cap \{T_n = k2^{-n}\}) \\
&\overset{(5)}{=} \mathbf{P}(B \in A)\mathbf{P}(E)
\end{aligned}$$

Where (1) comes from the law of Total Probability, (2) comes from the fact that when working on the event $\{T_n = k2^{-n}\}$, $B^* = B^{(k)}$, (3) is slightly non-trivial: if $E \in \mathscr{F}_{T_n}^+$, then $E \in \mathscr{F}_{T_n+s}$ for all $s > 0$, which means that by definition of the stopped σ-algebra,

$$E \cap \{T_n = k2^{-n}\} = E \cap \{T_n + s = k2^{-n} + s\} \in \mathscr{F}_{k2^{-n}+s} \qquad \text{for all } s$$

and so $E \cap \{T_n = k2^{-n}\} \in \mathscr{F}_{k2^{-n}}^+$, then you use the simple Markov Property to establish that $B^{(k)}$ is independent to $\mathscr{F}_{k2^{-n}}^+$ (see definition of $B^{(k)}$ to convince yourself), and hence since we have just established $E \cap \{T_n = k2^{-n}\} \in \mathscr{F}_{k2^{-n}}^+$ we can split the probability as a product. (4) comes from the fact that $B^{(k)}$ is Standard Brownian Motion hence has the same distribution as B, and (5) comes again from the law of Total Probability. We have just established that

$$\mathbf{P}(\{B^* \in A\} \cap E) = \mathbf{P}(\{B \in A\})\mathbf{P}(E)$$

so we may take $E = \Omega$ i.e: the whole probability space, and we get that **the processes** B^* and B have the same law. Therefore since B^* is continuous, we get that B^* is Brownian motion, and in particular, we get that

$$\mathbf{P}(\{B^* \in A\} \cap E) = \mathbf{P}(B^* \in A)\mathbf{P}(E)$$

which establishes that B^* is Standard Brownian Motion independent of $\mathscr{F}_{T_n}^+$, now we just need to establish that we can pass limits and thus finish the result.

First we note that due to continuity of Brownian motion, we have that almost surely,

$$\underbrace{B_{t+T} - B_T}_{\tilde{B}(t)} = \lim_{n\to\infty} \underbrace{(B_{t+T_n} - B_{T_n})}_{B_n^*(t)}$$

and so the marginal distributions of the process $(\tilde{B})$ are described by

$$\mathbf{P}(\tilde{B}(t_1) \in A_1, \cdots, \tilde{B}(t_k) \in A_k) = \mathbf{P}\Big(\lim_{n\to\infty} B_n^*(t_1) \in A, \cdots, \lim_{n\to\infty} B_n^*(t_k) \in A_k)\Big)$$

But since for all n we have shown that each $B_n^*(t_i)$ is equal in distribution to Brownian motion $B(t_i)$, we have that $\tilde{B}$ has the same marginal laws as Brownian motion and hence its law is also the same as of Brownian motion. All left to determine now is whether $\tilde{B}$ is independent of $\mathscr{F}_T^+$. This is not too bad, let $A \in \mathscr{F}_T^+$, $t_1, \cdots, t_k$ be given, and $f \in C_b\big((\mathbf{R}^d)^k\big)$ be given as well. Then by the DCT we have that

$$\mathbf{E}[\mathbf{1}(A) f(\tilde{B}_{t_1}, \cdots, \tilde{B}_{t_k})] = \lim_{n\to\infty} \mathbf{E}[\mathbf{1}(A) f([B_{t_1+T_n} - B_{T_n}], \cdots, [B_{t_k+T_n} - B_{T_n}])]$$

And since $T_n > T$ for all n and $A \in \mathscr{F}_T^+$, we have that for all $s > 0$, $A \in \mathscr{F}_{T+s} \subseteq \mathscr{F}_{T_n+s}$ (indeed $T + s < T_n + s$ after all) so $A \in \mathscr{F}_{T_n}^+$, so using what we have proven earlier in this proof, we have that

$$\mathbf{E}[\mathbf{1}(A) f(\tilde{B}_{t_1}, \cdots, \tilde{B}_{t_k})] = \mathbf{P}(A) \lim_{n\to\infty} \mathbf{E}[f([B_{t_1+T_n} - B_{T_n}], \cdots, [B_{t_k+T_n} - B_{T_n}])]$$

and now passing the limit inside with the DCT finishes the proof. ♡

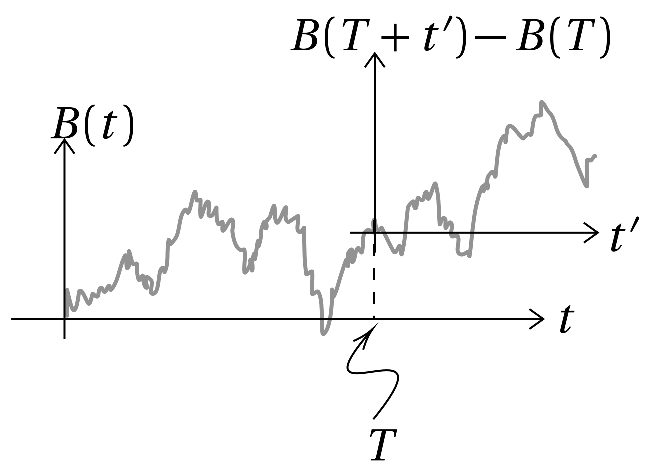
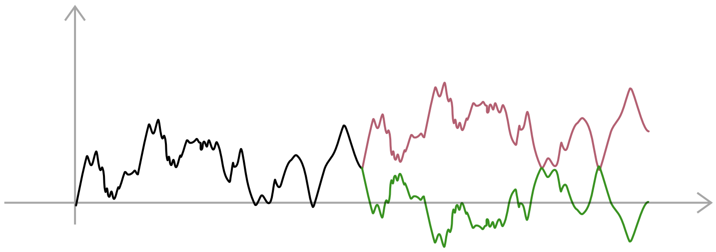


Figure 6: The Strong Markov Property: if you stop Brownian Motion at a random time, and put on a new set of axes, the picture you see has the same distribution as usual Brownian Motion.

5.4 The Reflection Principle

Here we have an amazing property of Brownian Motion that is a very straightforward corollary to the Strong Markov Property: the Reflection Principle. This is not only beautiful but also useful in calculations.



Theorem 5.19 (The Reflection Principle) Let (B_t) be Standard Brownian motion in $\mathbf{R}$, and T an almost surely finite stopping time. Then the process defined by $(\tilde{B}_t)$ defined by

$$\tilde{B}(t) = B(t)\mathbf{1}(t \leq T) + [2B(T) - B(t)]\mathbf{1}(t > T)$$

is also Standard Brownian Motion, called Standard Brownian Motion reflected at T.

Remark 5.20 (Proof by waffle) You run your Brownian motion until time T, then if you put on a new set of axes, the picture you get after time T is Standard Brownian Motion, which means that if you reflect it, by invariance property, its also Standard Brownian Motion. Now you can concatenate the processes and see that the laws stay the same.

Proof. We start by noting that the process

$$B^{(T)} = (B_{t+T} - B_T : t \geq 0)$$

is by the Strong Markov Property, a Standard Brownian Motion independent of $(B_t : 0 \leq t \leq T)$, and similarly, the process $-B^{(T)}$ is also SBM independent of the history of B up to the random time T. Therefore, given that $-B^{(T)} \overset{(d)}{=} B^{(T)}$, we have that the processes, on $C(\mathbf{R}_+, \mathbf{R}) \times C(\mathbf{R}_+, \mathbf{R})$ given by:

$$((B_t : 0 \leq t \leq T), B^{(T)}) \qquad \text{and} \qquad ((B_t : 0 \leq t \leq T), -B^{(T)})$$

Have the same laws. Now we just need to somehow concatenate these maps and preserve the equality of laws. Let us define the concatenation map that takes two processes $X, Y \in C(\mathbf{R}_+, \mathbf{R})$

$$\Psi_T(X, Y) := X_t \mathbf{1}(t \leq T) + (X_T + Y_{t-T})\mathbf{1}(t > T)$$

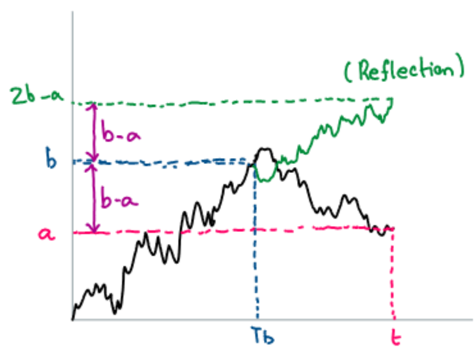
if Y is started at zero, then it is easy to see that $\Psi_T : C(\mathbf{R}_+, \mathbf{R}) \times C(\mathbf{R}_+, \mathbf{R}) \to C(\mathbf{R}_+, \mathbf{R})$, and moreover, it is also easy to see, by approximating T by discrete stopping times, that it is $\mathscr{A} \times \mathscr{A} - \mathscr{A}$ measurable, where $\mathscr{A}$ is the sigma algebra of $C(\mathbf{R}_+, \mathbf{R})$. This finishes the proof, because

$$\begin{aligned}
\mathbf{P}(B \in A) &= \mathbf{P}\left(\psi_T[((B_t : 0 \le t \le T), B^{(T)})] \in A\right) \\
&= \mathbf{P}\left(((B_t : 0 \le t \le T), B^{(T)}) \in (\psi_T^{-1}(A))\right) \\
&\stackrel{!}{=} \mathbf{P}\left(((B_t : 0 \le t \le T), -B^{(T)}) \in (\psi_T^{-1}(A))\right) \\
&= \mathbf{P}(\tilde{B} \in A)
\end{aligned}$$

Where the only tricky step, marked by an exclamation mark is due to the fact that we showed that the two processes $((B_t : 0 \le t \le T), B^{(T)})$ and $((B_t : 0 \le t \le T), -B^{(T)})$ had the same law. This shows that B and $\tilde{B}$ have the same law, and since $\tilde{B}$ is continuous, then it is also SBM.

♡

The reflection principle sometimes takes on a different fashion, given by the following result, which tells us that the joint law of the running supremum of Brownian motion and the Brownian motion itself can be understood simply in terms of the law of Brownian motion itself.



Corollary 5.21 (Reflection Principle, ♩) Let (B_t) be Standard Brownian Motion in one dimension, and let (S_t) be the running supremum up to time t, then $\mathbf{P}(S_t \ge b, B_t \le a) = \mathbf{P}(B_t \ge 2b - a)$

Main idea: The proof by English here is that on the event that $S_t \ge b$, it means we have hit it the level b at some stopping time T_b. Then we can reflect the motion at that point, and the event $B_t \le a$, means that the original BM travelled a distance of $b-a$ downwards, so the reflected Brownian motion travelled a distance of $b-a$ upwards, so that the reflected BM, $\tilde{B}$ ended up at $\tilde{B}_t \ge 2b-a$. Then the reflection principle ensures that all these probabilities match up.

Proof. The idea is to reflect B at the hitting time T_b. Note that $\{S_t \ge b\} = \{T_b \le t\}$, so we have the following

$$\begin{aligned}\mathbf{P}(S_t \geq b, B_t \leq a) &= \mathbf{P}(T_b \leq t, 2b - B_t \geq 2b - a)\\ &\overset{(1)}{=} \mathbf{P}(T_b \leq t, \tilde{B}_t \geq 2b - a)\\ &\overset{(2)}{=} \mathbf{P}(\tilde{B}_t \geq 2b - a)\\ &\overset{(3)}{=} \mathbf{P}(B_t \geq 2b - a)\end{aligned}$$

Where (1) comes from the fact that since B is continuous, $B_{T_b} = b$, so by definition of $\tilde{B}$ and the fact that we are working on the event that $t \geq T_b$, we get that $2b - B_t = \tilde{B}_t$. (The reason why we need the fact that $B(T_b) = b$ is that $\widetilde{B}(t) = B(t)\mathbf{1}(t < T) + (2B(T_b) - B(t))\mathbf{1}(t \geq T)$) Then (2) comes from the fact that since $a \leq b$, then if $\tilde{B}_t \geq 2b - a \geq b$, it must be that $t \geq T_b$, indeed, the only way $\tilde{B}_t$ can go above b is if B_t has gone above b, therefore $\{\tilde{B}_t \geq 2b - a\} \subseteq \{T_b \leq t\}$. Finally (3) comes from the Reflection Principle, which tells us that $\tilde{B}$ has the same law as B. ♡

Finally we have the following connection between the running maximum and the Brownian motion:

Corollary 5.22 For each t, $S_t \overset{(d)}{=} |B_t|$.

Proof. We have the following simple calculation

$$\begin{aligned}\mathbf{P}(S_t \geq a) &\overset{(1)}{=} \mathbf{P}(S_t \geq a, B_t \leq a) + \mathbf{P}(S_t \geq a, B_t \geq a)\\ &\overset{(2)}{=} \mathbf{P}(S_t \geq a, B_t \leq a) + \mathbf{P}(B_t \geq a)\\ &\overset{(3)}{=} 2\mathbf{P}(B_t \geq a)\\ &\overset{(4)}{=} \mathbf{P}(|B_t| \geq a)\end{aligned}$$

Where (1) comes from the Law of Total Probability and the fact that since B_t is a Normal random variable, the probability that it is equal to a is zero, (2) comes from the fact that if $B_t \geq a$, then of course $S_t \geq a$, and as such $\{B_t \geq a\} \subseteq \{S_t \geq a\}$, (3) comes from the Reflection Principle, and (4) comes from the fact that B_t is normally distributed and hence symmetrically distributed. ♡

Remark 5.23 (Proof by waffle) A nice intuitive way to see what this has to do with the reflection principle at a moral level, consider the following. If a Brownian path went above level a by time t, then by continuity it touched a at some stopping time T_a. Then you have to possibilities, either by time t the path finished above a, which happens with probability $\mathbf{P}[B_t \geq a]$, or it finished below a. If it finished below a, we can consider the reflection $\widetilde{B}(t)$ at time T_a, and now the probability that our original path finished below a has become the probability that the reflection finishes above a, but by the reflection principle this probability is simply the probability that the original path finished above a.

5.5 Martingales and Brownian Motion

Proposition 5.24 Let $(B_t : t \geq 0)$ be Standard Brownian Motion on $\mathbf{R}$. Then the processes (B_t) and $(B_t^2 - t)$ are both $(\mathscr{F}_t^+)$ martingales.

Proof. To show that Brownian motion itself is a Martingale, we simply note that by the Markov Property, the increment $B_t - B_s$, where $t > s$, is independent of the sigma algebra $\mathscr{F}_s^+$ (SMP with the constant stopping time s). Then by the independence property of conditional expectation:

$$\mathbf{E}[B_t - B_s \mid \mathscr{F}_s^+] = \mathbf{E}[B_t - B_s] = 0$$

To show that the second process is a Martingale, we write

$$\begin{aligned}
\mathbf{E}[B_t^2 - t \mid \mathscr{F}_s^+] &= \mathbf{E}[(B_t - B_s + B_s)^2 - t \mid \mathscr{F}_s^+] \\
&= \mathbf{E}[(B_t - B_s)^2 \mid \mathscr{F}_s^+] + 2\mathbf{E}[B_t B_s \mid \mathscr{F}_s^+] - B_s^2 - t \\
&\stackrel{!}{=} (t - s) + B_s^2 - t = B_s^2 - s.
\end{aligned}$$

Where in the only tricky step, denoted by (!), we used the fact that $B_t - B_s \sim \mathcal{N}(0, t - s)$ and is independent of $\mathscr{F}_S^+$, which means that $\mathbf{E}[(B_t - B_s)^2 \mid \mathscr{F}_s^+] = \mathbf{E}[(B_t - B_s)^2] = \mathrm{Var}(B_t - B_s) = t - s$. We also used the fact that $B_s \in m\mathscr{F}_s^+$ and that (B_t) is an $\mathscr{F}_t^+$ Martingale to show that $\mathbf{E}[B_t B_s \mid \mathscr{F}_s^+] = B_s^2$. ♡

Remark 5.25 In the above proposition, we saw that if we consider the function $f(x) = x^2$, the correct thing to subtract from $f(B_t)$ to turn this into a Martingale was t. Let's observe in a discrete setting, say on a random walk on the integers, whether we can see a more general behavior. Let (S_n) be simple symmetric random walk on $\mathbf{Z}$ and consider a function $f : \mathbf{Z} \to \mathbf{R}$. Then

$$\begin{aligned}
\mathbf{E}[f(S_{n+1}) \mid \mathscr{F}_n] &= f(S_n) + \frac{1}{2}\big[f(S_n + 1) - 2f(S_n) + f(S_n - 1)\big] \\
&:= f(S_n) + \frac{1}{2}\tilde{\Delta} f(S_n)
\end{aligned}$$

Therefore we see that

$$f(S_n) - \frac{1}{2}\sum_{t=0}^{n-1} \tilde{\Delta} f(S_t)$$

is a Martingale, indeed:

$$\begin{aligned}\mathbf{E}\left[f(S_n)-\frac{1}{2}\sum_{t=0}^{n-1}\tilde{\Delta}f(S_t)\middle|\mathscr{F}_{n-1}\right] &= f(S_{n-1})+\frac{1}{2}\tilde{\Delta}f(S_{n-1})-\frac{1}{2}\sum_{t=0}^{n-1}\tilde{\Delta}f(S_t)\\ &= f(S_{n-1})-\frac{1}{2}\sum_{t=0}^{n-2}\tilde{\Delta}f(S_t)\end{aligned}$$

This gives some motivation, and in the continuous case, instead of having $\tilde{\Delta}$, we will have the Laplacian:

$$\Delta f = \sum_{i=1}^{d}\frac{\partial^2 f}{\partial x_i^2}$$

Note that the discrete-Laplacian we wrote before does indeed give rise to this new when going from the discrete case to the continuous. Because in a very brutal approximation, we may say

$$f''(x)\approx f(x+1)-2f(x)+f(x-1)$$

We now see how this generalises to Brownian motion:

Theorem 5.26 Let B be Brownian Motion in $\mathbf{R}^d$ and let $f(t;x)\in C_b^{1,2}\left(\mathbf{R}_+\times\mathbf{R}^d\right)$. Then the process defined by

$$M_t = f(t,B_t)-f(0,B_0)-\int_0^t\left(\frac{\partial}{\partial t}+\frac{1}{2}\Delta\right)f(s,B_s)ds$$

is an $\mathscr{F}_t^+$-Martingale.

Proof. Integrability is clear due to the boundedness of f and its derivatives. The meaty substance will be to show the Martingale property. For this we must show that $\mathbf{E}[M_{t+s}-M_s\,|\,\mathscr{F}_s^+]=0$. First note that

$$\begin{aligned}M_{t+s}-M_s &= f(t+s,B_{t+s})-f(s,B_s)-\int_s^{t+s}\left(\frac{\partial}{\partial r}+\frac{1}{2}\Delta\right)f(r,B_r)dr\\ &= f(t+s,B_{t+s})-f(s,B_s)-\int_0^{t}\left(\frac{\partial}{\partial r}+\frac{1}{2}\Delta\right)f(r+s,B_{r+s})dr\end{aligned}$$

Let's first take a look at $f(t+s,B_{t+s})$. We have the following:

$$\begin{aligned}\mathbf{E}[f(t+s,B_{t+s})\,|\,\mathscr{F}_s^+] &= \mathbf{E}[f(t+s,B_{t+s}-B_s+B_s)\,|\,\mathscr{F}_s^+]\\ &\overset{(☕)}{=}\int_{\mathbf{R}^d}f(t+s,x+B_s)p_t(0,x)dx\end{aligned}$$

Where $p_t(0,x)$ is the standard d-dimensional Gaussian distribution, which we refer to as the transition density. The important step, ☕, is intuitively due to the following facts: since $B_{t+s} - B_s \sim \mathcal{N}(\mathbf{0}, t\mathbf{I})$, we can condition f on the value of $B_t - B_s$ and then integrate against the density. Moreover, once we have conditioned on the value of $B_t - B_s$, f becomes $\mathscr{F}_s^+$ measurable, so we can forget about the expectation. The actual result being used here can be found in the appendix 8.10.

Now we look at the big integral term in the definition of $M_{t+s} - M_s$, and compute its conditional expectation. We quickly refer the reader to the Fubini Trick on the Appendix. Using this trick and the boundedness of f, we have that

$$\mathbf{E}\left[\int_0^t \left(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\right) f(r+s, B_{r+s}) dr \middle| \mathscr{F}_s^+\right] = \int_0^t \mathbf{E}\left[\left(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\right) f(r+s, B_{r+s}) \middle| \mathscr{F}_s^+\right] dr$$

Now we can compute the expectation inside the integral on the right hand side using an analogous trick to what we did before. We note that writing $B_{r+s} = B_{r+s} - B_s + B_s$, and that $B_{r+s} - B_s$ is independent of $\mathscr{F}_s^+$ and that once we condition on its value, f and its derivatives will be $\mathscr{F}_s^+$-measurable, we get that

$$\begin{aligned}
&\mathbf{E}\left[\int_0^t \left(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\right) f(r+s, B_{r+s}) dr \middle| \mathscr{F}_s^+\right] = \int_0^t \int_{\mathbf{R}^d} \left(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\right) f(r+s, x+B_s) p_r(0,x)\, dx\, dr \\
&= \lim_{\epsilon \downarrow 0} \int_\epsilon^t \int_{\mathbf{R}^d} \left(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\right) f(r+s, x+B_s) p_r(0,x)\, dx\, dr \\
&= \lim_{\epsilon \downarrow 0} \int_{\mathbf{R}^d} \int_\epsilon^t \left(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\right) f(r+s, x+B_s) p_r(0,x)\, dr\, dx \\
&= \lim_{\epsilon \downarrow 0} \int_{\mathbf{R}^d} \left(\left\{ p_t(0,x) f(t+s, x+B_s) - p_\epsilon(0,x) f(\epsilon+s, x+B_s) \right\} - \int_\epsilon^t f(r+s, x+B_s) \underbrace{\left(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\right) p_r(0,x)}_{=0} \right) dx \\
&= \lim_{\epsilon \downarrow 0} \int_{\mathbf{R}^d} \left\{ p_t(0,x) f(t+s, x+B_s) - p_\epsilon(0,x) f(\epsilon+s, x+B_s) \right\} dx \\
&= \mathbf{E}\left[f(t+s, B_{t+s}) \mid \mathscr{F}_s^+\right] + \lim_{\epsilon \downarrow 0} \int_{\mathbf{R}^d} p_\epsilon(0,x) f(\epsilon+s, x+B_s) = \mathbf{E}\left[f(t+s, B_{t+s}) \mid \mathscr{F}_s^+\right] + f(s, B_s)
\end{aligned}$$

Thus showing the Martingale property. ♡

Let us see one more example of a Martingale arising from Brownian motion:

Proposition 5.27 Let B be Standard Brownian Motion in d dimensions, and let $u \in \mathbf{R}^d$. Then the process

$$M_t^u = \exp\left(\langle u, B_t\rangle - \frac{|u|^2 t}{2}\right)$$

is a $(\mathscr{F}_t^+)$-Martingale.

Proof. We first recall from elementary probability, that if $X \sim \mathcal{N}(\mathbf{a}, \mathbf{v})$ is a multidimensional normal distribution, then the MGF

$$M_X(\mathbf{t}) = \exp\left(\mathbf{t} \cdot \mathbf{a} + \frac{1}{2}\mathbf{t}^T \mathbf{v}\mathbf{t}\right)$$

From this we get that

- M_t^u is integrable, because:

$$\mathbf{E}[M_t^u] = \exp\left(\frac{t}{2}\sum_{i=1}^{d} u_i^2 - \frac{|u|^2 t}{2}\right) < \infty$$

- M_t^u satisfies the Martingale property, because

$$\begin{aligned}
\mathbf{E}[M_t^u \mid \mathscr{F}_s^+] &= \exp\left(-\frac{|u|^2 t}{2}\right)\mathbf{E}\left[\langle u, B_t - B_s + B_s\rangle \mid \mathscr{F}_s^+\right] \\
&= \exp\left(-\frac{|u|^2 t}{2}\right)\exp(\langle u, B_s\rangle)\mathbf{E}\left[\exp\langle u, B_t - B_s\rangle \mid \mathscr{F}_s^+\right] \\
&\overset{(1)}{=} \exp\left(-\frac{|u|^2 t}{2}\right)\exp(\langle u, B_s\rangle)\mathbf{E}[\exp\langle u, B_t - B_s\rangle] \\
&\overset{(2)}{=} \exp(\langle u, B_s\rangle)\exp\left(-\frac{|u|^2 t}{2}\right)\exp\left(\frac{|u|^2(t-s)}{2}\right) \\
&= M_s^t
\end{aligned}$$

Where (1) is due to the fact that $\langle u, B_t - B_s\rangle$ is independent of $\mathscr{F}_s^+$ as shown by the Markov property, and (2) is due to the fact that $B_t - B_s \sim \mathcal{N}(\mathbf{0}, (t-s)\mathbf{I})$, so the expectation is nothing but the MGF and so we can use the formula at the start of the proof.

♡

5.6 Recurrence and Transience

Armed with these Martingale results, we are ready to explore some more properties of Brownian motion. In particular, we will see how the dependence on the dimension of Brownian motion can tell us information about how often (if at all) it revisits certain points.

Definition 5.28 We denote with $\mathbf{P}_x$ the probability measure $\mathbf{P}$ conditioned to Brownian motion starting at x, i.e:

$$\mathbf{P}_x[B_t \in A] = \mathbf{P}[B_t \in A \mid B_0 = x]$$

For example, under $\mathbf{P}_x$, the process $(B_t - x : t \geq 0)$ is Standard Brownian Motion

Theorem 5.29 (Recurrence and Transience of Brownian Motion) Let B be Brownian motion in d dimensions:

- If $d = 1$, then for any $x \in \mathbf{R}$, we have that $\mathbf{P}_0$-almost surely:

$$\{t \geq 0 : B_t = x\}$$

  is unbounded. Meaning that Brownian motion is point recurrent.

- If $d = 2$, then B is neighbourhood recurrent, meaning that for any $x, z \in \mathbf{R}^2$ and any fixed $\epsilon > 0$, we have that $\mathbf{P}_x$-almost surely:

$$\{t \geq 0 : |B_t - z| \leq \epsilon\}$$

  is unbounded. However, B_t does not hit points, i.e for any $x \in \mathbf{R}^2$: $\mathbf{P}_0(\exists t > 0 : B_t = x) = 0$. (This is wild)

- If $d = 3$, then Brownian motion is transient, i.e: $\mathbf{P}_0$-almost surely,

$$|B_t| \to \infty$$

Proof. The proof for $d = 1$ follows from the previous result where we showed that

$$\mathbf{P}_0\left(\limsup_{t\to\infty} B_t = \infty\right) = \mathbf{P}_0\left(\liminf_{t\to\infty} B_t = -\infty\right) = 1$$

and so by continuity of Brownian motion, it must be that each point $x \in \mathbf{R}$ is visited by Brownian motion infinitely many times with probability 1.

For $d = 2$, we show the claim for $z = 0$ and BM started from any $x \in \mathbf{R}^2$. We first show that for any $\epsilon > 0$, almost surely B will hit the ϵ-neighbourhood of zero. To do so, we fix $\epsilon < |x| < R$ and select a function $\varphi \in C_b^2(\mathbf{R}^2)$ that agrees with $\log|y|$ on $\epsilon < |y| < R$ (Although we haven't shown this function exists, intuitively, we know it should). Then one can verify that on $\epsilon \leq |y| \leq R$, one has that $\Delta\varphi = 0$, this just amounts to computing a bunch of derivatives. Therefore, using the Theorem we showed about Martingales for Brownian motion, we know that

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

is a $\mathscr{F}_t^+$-martingale. (Notice that we have dropped the $\varphi(B_0)$ that would be missing if we wrote everything according to the Theorem, but if M_t is a Martingale, then so is $M_t + C$ for any constant). Notice that to write this integral we noted that φ is only spatially dependent, not time dependent. Now we can construct two stopping times, $S = \inf\{t > 0 : |B_t| = \epsilon\}$ and $T_R = \inf\{t > 0 : |B_t| = R\}$. Then $H = S \wedge T_R$ is an almost surely finite stopping time (to see why its a.s finite, just think of 2d BM as two 1d BM that run together, each of these BM will oscillate in their axis between $\pm\infty$). And since R was chosen large enough (and x can be assumed to not start in the epsilon ball), we know not only that the Martingale $M_{t\wedge H}$ is bounded, but also that it will take values in the annulus of radii ϵ and R, which means that on this range, $\Delta\varphi = 0$, and so

$$M_{t\wedge H} = \log|B_{t\wedge H}| \text{ is a bounded Martingale.}$$

We may therefore apply the Optional Stopping Theorem, and see that

$$\mathbf{E}_x[\log|B_H|] = \log|x|$$

But on the other hand:

$$\mathbf{E}_x[\log|B_H|] = \log\epsilon\mathbf{P}_x(S < T_R) + \log R\mathbf{P}_x(S > T_R)$$

Solving these equations yields that

$$\mathbf{P}_x(S < T_R) = \frac{\log R - \log|x|}{\log R - \log\epsilon} \to 1 \qquad (R \to \infty)$$

Therefore

$$\mathbf{P}_x(S < \infty) = \mathbf{P}_x\left(\bigcup_{R>0}\{S < T_R\}\right) = \lim_{R\to\infty}\mathbf{P}_x(S < T_R) = 1$$

since the events $\{S < T_R\}$ form an increasing sequence. In conclusion, we have that $\mathbf{P}_x$ almost surely, there is some finite time $t > 0$ for which B_t reaches the ϵ-neighbourhood around zero. To extend this result to show that actually B_t visits this neighbourhood infinitely often, we will first let B_t reach the neighbourhood, then let it run for a bit, and use the Markov property to restart the motion at this new point, then we can reuse the result, and keep going forever.

Now we finally have that for any $n > 0$, we have that

$$\begin{aligned}
\mathbf{P}_x\left\{|B_t| = \epsilon \text{ for some } t \geq n\right\} &= \mathbf{P}_x\left\{|B_{n+t} - B_t + B_t| = \epsilon \text{ for some } t \geq 0\right\} \\
&\overset{(1)}{=} \int_{\mathbf{R}^2} \mathbf{P}_0\left\{|B_t + y| = \epsilon \text{ for some } t \geq 0\right\} p_n(x,y)dy \\
&\overset{(2)}{=} \int_{\mathbf{R}^2} \mathbf{P}_y\left\{|B_t| = \epsilon \text{ for some } t \geq 0\right\} p_n(x,y)dy \\
&\overset{(3)}{=} 1
\end{aligned}$$

Where (1) comes from the fact that $B_{n+t} - B_n$ has the same distribution as a Brownian Motion started at zero, and in this same step, B_n is the position at time n of Brownian motion started at x, hence why we condition on its position being y and then integrate against $p_n(x,y)$. Step (2) comes from the fact that if we add a constant to standard Brownian Motion, then the result is Brownian Motion started at said constant. Step (3) is due to the fact that we have shown that for all $y \in \mathbf{R}^2$, $\mathbf{P}_y\left\{|B_t| = \epsilon \text{ for some } t \geq 0\right\} = 1$, and then $p_n(x,y)$ is a probability distribution so it integrates to one (the x is just a fixed constant, don't get confused). Finally:

$$\begin{aligned}
\mathbf{P}_x(\#\{t > 0 : |B_t| = \epsilon\} = \infty) &= \mathbf{P}_x\left(\bigcap_{n\geq 0}\left\{|B_t| = \epsilon \text{ for some } t \geq n\right\}\right) \\
&= \lim_{n\to\infty} \mathbf{P}_x\left\{|B_t| = \epsilon \text{ for some } t \geq n\right\} = 1
\end{aligned}$$

Where the limit was extracted as $\left\{|B_t| = \epsilon \text{ for some } t \geq n\right\}$ is a decreasing sequence of events.

To now show that the probability of hitting any actual point is zero, we will show that the probability to hit zero from any starting point is zero, and then by translational invariance of Brownian motion, the general claim will follow. Recall that S was defined as the least time for which Brownian motion reached a distance of ϵ away from the origin, and we saw that

$$\mathbf{P}_x(S < T_R) = \frac{\log R - \log|x|}{\log R - \log \epsilon} \to 0 \qquad (\epsilon \to 0)$$

And if we also send $R \to \infty$, we have that the probability of reaching 0 from $x \neq 0$ in any finite time is of zero. To show the case for $x = 0$, i.e: to show that

$$\mathbf{P}_0(B_t = 0 \text{ for some } t > 0) = 0$$

We use once again a similar technique to what we have for $a>0$:

$$\begin{aligned}\mathbf{P}_0(B_t=0:t\geq a)&=\mathbf{P}_0(B_{t+a}-B_a+B_a=0:t\geq 0)\\&=\int_{\mathbf{R}^2}\mathbf{P}_0(B_{t+a}-B_a+y=0:t\geq 0)p_a(0,y)dy\\&=\int_{\mathbf{R}^2}\underbrace{\mathbf{P}_y(B_t=0:t\geq 0)}_{=0}p_a(0,y)y=0\end{aligned}$$

Now sending $a\to 0$ finishes the claim.

For $d\geq 3$: since the first three components of BM in d dimensions form a Brownian Motion in 3 dimensions, it suffices to treat the case for $d=3$. The strategy is very similar to the case $d=2$, with the difference that we will get a different result. Let $B_0=x$, and choose ϵ and R small and large enough so that $\epsilon\leq|x|\leq R$. If we define $S(\epsilon)$ to be the first hitting time of the ϵ-ball, and $T(R)$ to be the first exit time of the R-ball, we can now choose a function $f\in C_b^2(\mathbf{R}^3)$ such that

$$f(y)=\frac{1}{|y|}$$

for $\epsilon\leq|y|\leq R$. Then one can compute that $\Delta f=0$ on the annulus $\epsilon\leq|y|\leq R$. Therefore, we have that $\{f(B_t)\}_t$ is a $\mathscr{F}_t^+$ Martingale. Moreover, if we let $H=S(\epsilon)\wedge T(R)$, we have that $\{f(B_{t\wedge H})\}$ is a bounded Martingale, so by the OST:

$$\mathbf{E}_x\left[\frac{1}{|B_H|}\right]=\frac{1}{|x|}$$

but likewise:

$$\mathbf{E}_x\left[\frac{1}{|B_H|}\right]=\mathbf{P}_x[S(\epsilon)<T(R)]\cdot\frac{1}{\epsilon}+\mathbf{P}_x[S(\epsilon)>T(R)]\cdot\frac{1}{R}$$

and using the fact that $\mathbf{P}_x[S(\epsilon)<T(R)]+\mathbf{P}_x[S(\epsilon)>T(R)]=1$, we can solve and see that

$$\mathbf{P}_x[S(\epsilon)<T(R)]=\frac{|x|^{-1}-R^{-1}}{\epsilon^{-1}-R^{-1}}$$

And so if we take $R\to\infty$, we have that the probability of ever visiting the ball centred at 0 with radius ϵ when starting from $|x|\geq\epsilon$ is $\epsilon/|x|$. We now use this to show that

$$\mathbf{P}_0[|B_t|\to\infty\text{ as }\to\infty]=1.$$

Let $\tau(R)$ be the return time to distance R: $\inf\{t>0:|B_t|=R\}$ (Observe the strictly greater than sign). Then we define the events

$$A_n = \{|B_t| > n \text{ for all } t \geq \tau(n^3)\}$$

i.e: the events that once distance n^3 from the origin is reached, the distance to the origin is always above n. Of course, we also have by unboundedness that $\mathbf{P}_0[\tau(n^3) < \infty] = 1$. We will show that A_n occurs for all n large enough, and so the claim will follow. First note that

$$\begin{aligned}
\mathbf{P}_0[A_n^c] &= \mathbf{P}_0\left[\left|B_{t+\tau(n^3)} - B_{\tau(n^3)} + B_{\tau(n^3)}\right| \leq n \text{ for some } t \geq 0\right] \\
&= \mathbf{E}_0\left[\mathbf{P}_0\left(|W_t + B_{\tau(n^3)}| \leq n \text{ for some } t \geq 0 \mid B_{\tau(n^3)}\right)\right] \\
&= \mathbf{E}_0[\mathbf{P}_{B(\tau_{n^3})}(\tau(n) < \infty)]
\end{aligned}$$

Where $W_t = B_{t+\tau(n^3)} - B_{\tau(n^3)}$ is Standard Brownian Motion by the Strong Markov Property, and we have conditioned on the value of $B(\tau_{n^3})$. Now we can use the result we obtained above, namely that if you start at some $|x| \geq \epsilon$, the probability of ever hitting the ϵ ball is $\epsilon/|x|$, and so we conclude that

$$\mathbf{P}_{B(\tau_{n^3})}(\tau(n) < \infty) = \frac{n}{n^3} = \frac{1}{n^2}$$

Where we have of course used the continuity of Brownian Motion to use that $|B(\tau_{n^3})| = n^3$. There we have that $\mathbf{P}_0[A_n^c]$ is summable and so it can only occur for a finite amount of times by Borel-Cantelli. This implies that eventually, A_n holds, and so $|B_t|$ gets further and further away from the origin. ♡

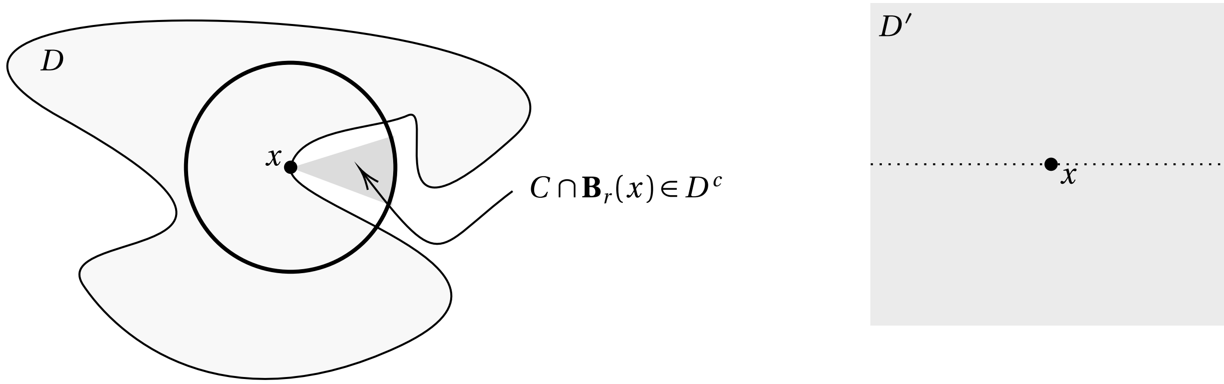


Figure 7: On the left an example that satisfies the cone condition at x, on the right, we have the plane $\mathbf{R}^2$ with the x-axis cut, i.e: $\mathbf{R}^2 \setminus \{(x,0) : x \in \mathbf{R}\}$. The boundary of this set is the x-axis but it does not satisfy the cone condition at any point in the boundary.

5.7 Dirichlet Problem

Definition 5.30 (Domain) An open connected subset $D \subseteq \mathbf{R}^d$ is called a domain.

Let's recall the Dirichlet Problem:

Dirichlet Problem: Let D be a domain and $f : \partial D \to \mathbf{R}$ be a continuous function on the boundary. Find a function $u : \bar{D} \to \mathbf{R}$ that satisfies:

- u is harmonic in D, i.e: $\Delta u = 0$ in D.
- u has the boundary condition $u = f$ on ∂D.

It turns out that one can solve the Dirichlet problem on certain "good" subsets of $\mathbf{R}^d$ by using Brownian motion. This regularity condition is given by the Poincaré cone condition:

Definition 5.31 (Poincaré cone condition) A domain D is said to satisfy the Poincaré Cone Condition at a point $x \in \partial D$ if there exists a nonempty open cone C with origin at x such that $C \cap \mathbf{B}(x,r) \subseteq D^c$ for some $r > 0$.

Theorem 5.32 (Dirichlet Problem) Let D be a bounded domain in $\mathbf{R}^d$ such that each point on the boundary satisfies the cone condition. Suppose φ is a continuous function on ∂D. Let $T_{\partial D}$ be the hitting time of the Brownian motion of the boundary of the domain, which is an almost surely finite stopping time when starting in D. Then the function

$$u(x) = \mathbf{E}_x[\varphi(B_{T_{\partial D}})] \qquad x \in \bar{D}$$

is the unique continuous function that satisfies

- $\Delta u = 0$ on D, i.e: u is harmonic.
- $u(x) = \varphi(x)$ on ∂D

We first provide a useful characterisation of harmonic functions, which in plain English tell you that harmonic functions are those whose value at a point x is the average of the values of the function in its neighbouring regions.

Theorem 5.33 (Characterisation of Harmonic functions) Let $D \subseteq \mathbf{R}^d$ be a domain and $u : D \to \mathbf{R}$ be measurable and locally bounded. Then the following are equivalent:

- $u \in C^2(\mathbf{R}^d)$ and $\Delta u = 0$.
- For any ball $\mathbf{B}(x,r) \subseteq D$ we have that

$$u(x) = \frac{1}{\mathscr{L}(\mathbf{B}(x,r))} \int_{\mathbf{B}(x,r)} u(y)dy.$$

- For any ball $\mathbf{B}(x,r) \subseteq D$ we have that

$$u(x) = \frac{1}{\sigma_{xy}(\partial \mathbf{B}(x,r))} \int_{\partial \mathbf{B}(x,r)} u(y) d\sigma_{xy}(y)$$

  Where σ_{xy} is the surface area measure on $\mathbf{B}(x,r)$.

We now need some further tools that will aid in the proof of the Dirichlet Problem. If you stare at the statement of the Maximum Principle, the result will not seem surprising. Indeed, if we have established that harmonic functions are those whose values at a point are determined by the average around the point, if the function attains a maximum at a point in the domain where it is harmonic, then the average of the function around that point must also be the maximal, but since the other points can't be any "hotter", it must be that all the neighbouring points have the same temperature.

Theorem 5.34 (Maximum principle) Suppose that $u: \mathbf{R}^d \to \mathbf{R}$ is a harmonic function on $D \subseteq \mathbf{R}^d$. Then

- If u attains its maximum in D, then u is a constant in D.
- If u is continuous on $\bar{D}$ and D is bounded, then

$$\max_{x \in \partial D} u(x) = \max_{x \in \bar{D}} u(x)$$

Main idea: The proof is just a formalisation of the discussion above. The idea is to show that the set of points that attain the maximum inside D is clopen. Since D is connected then the proof follows. Showing its closed follows from continuity of u, openness follows from formalising the argument above.

Proof. For the first part, let M be its maximum and let $V = \{x \in D : u(x) = M\}$. Then notice two things:

- If (x_n) is a converging sequence in V, with limit $x \in D$, we have that by continuity of u, $u(x) = \lim u(x_n) = M$ so V is closed in D.
- Since D is open, for any $x \in V$, there is some radius r such that $\mathbf{B}(x,r) \subseteq D$, which means, by the characterisation of harmonic functions, that

$$M = u(x) = \frac{1}{\mathscr{L}(\mathbf{B}(x,r))} \int_{\mathbf{B}(x,r)} u(y) dy \leq M$$

Which can only be true if for almost all $y \in \mathbf{B}(x,r)$ we have that $u(y) = M$, but by continuity of u it must actually be that $u(y) = M$ for all $y \in \mathbf{B}(x,r)$ which means that $\mathbf{B}(x,r) \subseteq V$ and so V is open.

It follows V is clopen, but by assumption V is not empty, and since D is connected, it only admits trivial clopen sets, which means that $V = D$.

For the second part, since u is continuous on $\bar{D}$, it must attain its maximum on $\bar{D}$. If u is constant then the claim follows trivially. If u is not constant, then by part one, it can't attain its maximum on D, so it must attain its maximum on ∂D.

♡

The following corollary of the Maximum Principle will be crucial for proving unicity of the solution to the Dirichlet Problem.

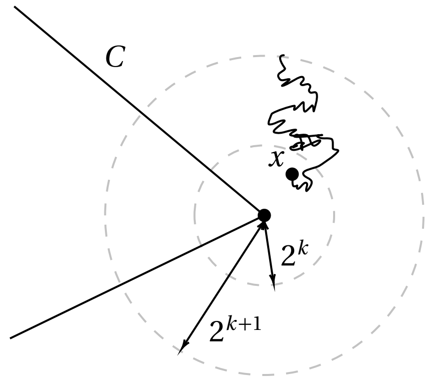


Figure 8: Proof by picture, cone hitting Lemma

Corollary 5.35 Suppose that two functions $u_1, u_2 : \mathbf{R}^d \to \mathbf{R}$ are harmonic on a bounded domain D and continuous on $\bar{D}$. If they agree on ∂D, then they are the same function.

Main idea: This follows immediately from the fact that harmonic functions form a linear space, so in particular $u_1 - u_2$ is harmonic, and the fact that harmonic functions attain their maximum in the boundary of the domains.

Proof. It is clear that harmonic functions form a linear space, so $u_1 - u_2$ is also harmonic on D. Since $u_1 - u_2$ is also continuous on $\bar{D}$ and D is bounded by assumption, the maximum principle says that

$$\max_{\bar{D}}(u_1 - u_2) = \max_{\partial D}(u_1 - u_2) = 0,$$

where the last equality follows by hypothesis of agreement on the boundary. Therefore $u_1 \leq u_2$. Repeating the argument switching u_1 and u_2 gives the desired equality. ♡

We are now ready to solve the Dirichlet problem. Just kidding, we have one more Lemma:

Lemma 5.36 (Cone hitting Lemma) Let $x \in \mathbf{R}^d$ be at a distance at most 2^{-k} from the origin, and let C be a cone centered at the origin. Let $T_{\partial \mathbf{B}(0,1)}$ and T_C denote the hitting times of the radius one ball about the origin and the cone respectively of Brownian Motion started at x. Then there is some $a < 1$ such that

$$\mathbf{P}_x[T_{\partial \mathbf{B}(0,1)} < T_C] \leq a^k$$

Proof. The proof is by the Strong Markov Property and scale invariance. First of all, let's define

$$a = \sup_{|x| < \frac{1}{2}} \mathbf{P}_x[T_{\partial \mathbf{B}(0,1) < T_C}] < 1$$

Now, if we have Brownian Motion started at x where $x \in \mathbf{B}(0, 2^{-k})$. Then we note that the probability of reaching $\partial\mathbf{B}(0, 2^{-k+1})$ before hitting the cone is at most some constant a, this is by scale invariance. If we hit the boundary of the ball of radius 2^{-k+1} before the cone, we can restart Brownian Motion here, and since this is at half the distance to the boundary of the next ball, i.e: the ball of radius 2^{-k+2}, once again by scale invariance we have that the probability of reaching the boundary of the ball before the cone is a. Iterating this until we reach the ball of radius 1, we get the desired conclusion. (Where the fact that we multiply the a's comes from the fact that SMP guarantees that all these Brownian motions are independent) ♡

Now we are finally ready.

Proof of Dirichlet Problem, Theorem 5.32. By hypothesis of D being bounded, we automatically get that u is bounded (and hence locally bounded). We now show that u as defined, i.e:

$$u(x) = \mathbf{E}_x[\varphi(B_{T(\partial D)})]$$

is a harmonic function, we do so by checking the third equivalent condition found in Theorem Characterisation of Harmonic Functions (Theorem 5.33). Let $\mathbf{B}(x, r) \subseteq D$ be any ball centered at x that is contained in D. We will run the motion until we reach the boundary of the ball, almost surely in finite time, and then restart the motion. It will be clear how the averaging is done. Let H be the hitting time of the boundary of the ball $\mathbf{B}(x, r)$, then:

$$\begin{aligned}
u(x) &= \mathbf{E}_x[\varphi \circ B(T_{\partial D})] \\
&= \mathbf{E}_x[\mathbf{E}_x[\varphi \circ B(T_{\partial D}) \mid \mathscr{F}_H]] && \text{(Tower Law)} \\
&= \mathbf{E}_x[\mathbf{E}_{B(H)}[\varphi \circ B(T_{\partial D})]] && \text{(Strong Markov Property)} \\
&= \mathbf{E}_x[u(B_H)] \\
&= \frac{1}{\sigma_{x,r}(\mathbf{B}(x, r))} \int_{\partial\mathbf{B}(x,r)} u(y) \sigma_{x,r}(dy)
\end{aligned}$$

Where the last equality simply comes from the fact that $B(H)$ lives on the boundary of the ball $\mathbf{B}(x, r)$, and since Brownian Motion was started at the centre of the ball, intuitively, it will hit the boundary of the ball uniformly at random. Here of course $\sigma_{x,r}$ simply denotes the area measure of the ball centered at x of radius r. Since the value of u at a point $x \in D$ is simply the average of the value of its neighbours, its intuitively clear that it is continuous at x. The only thing we have to show is that u is continuous in ∂D. For this we are going to need the Cone hitting Lemma. Note the following facts:

- If $y \in \partial D$, then $u(y) = \varphi(y)$, which is assumed to be continuous on the boundary, meaning that if z and y are both points in ∂D, if $\epsilon > 0$ is given, then for some tolerance $|z - y| < \delta$, we have that $|\varphi(z) - \varphi(y)| < \epsilon$.

- If $z \in \partial D$ and C is the cone centered at z that is guaranteed to exist by the cone condition, by the Cone hitting Lemma, we have that if $x \in \bar{D}$ is such that $x \in \mathbf{B}(z, 2^{-k}\delta)$, then $\mathbf{P}[T_{\partial \mathbf{B}(z,\delta)} \leq T_C] < a^k$ for some $a < 1$.

Finally we are armed to show continuity of u at ∂D and finish the proof. If $z \in \partial D$ and $x \in \bar{D}$ with $|z - x| \leq 2^{-k}\delta$, then:

$$\begin{aligned}
|u(x) - u(z)| &= \left| \mathbf{E}_x[\varphi \circ B(T_{\partial D}) - \varphi(z)] \right| \\
&= \left| \mathbf{E}_x\left[\left(\varphi \circ B(T_{\partial D}) - \varphi(z) \right) \left(\mathbf{1}(T_C \leq T_{\partial \mathbf{B}(z,\delta)}) + \mathbf{1}(T_C > T_{\partial \mathbf{B}(z,\delta)}) \right) \right] \right| \\
&\leq \epsilon \mathbf{P}_x[T_C \leq T_{\partial B(z,\delta)}] + 2 \left\| \varphi \right\|_\infty \mathbf{P}_x[\mathbf{1}(T_C > T_{\partial \mathbf{B}(z,\delta)})] \\
&\leq \epsilon + a^k
\end{aligned}$$

Now by taking k large enough, i.e: imposing the original x to be just closer to z if needed, this can be made as small as you want. ♡

5.8 Donsker's Invariance Principle

We know very well that if $\{X_n\}$ is an i.i.d sequence of finite mean and variance, then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$$

converges in distribution to a $\mathcal{N}(0,\sigma^2)$ random variable. Thus in some sense, the Normal random variable is the universal object for random variables. It would not seem a surprise if a similar result held for Brownian Motion, as we would expect in some sense Brownian Motion to be the universal object for random walks.

Definition 5.37 (Norm of a function) For a function $f \in C([0,1],\mathbf{R})$, we define uniform its norm $\|f\|$ as $\sup_{x\in[0,1]}|f(x)|$.

Theorem 5.38 (Donsker's Invariance) Let $\{X_n\}$ be a sequence of IID random variables with common law μ, such that $\mathbf{E}(X_n) = 0$ for all n and $0 < \mathbf{E}(X_n^2) = \sigma < \infty$. Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$. Define the linear interpolation

$$S_t = (1-\{t\})S_{[t]} + \{t\}S_{[t]+1}$$

Where $\{t\} = t - [t]$. Then we can speed up time by a factor of N and rescale space by a factor of $\sqrt{N\sigma^2}$, and we will obtain that

$$S^{[N]} := \left(\frac{S_{Nt}}{\sqrt{N\sigma^2}} : 0 \le t \le 1\right) \implies (B_t)_{t\in[0,1]}$$

We will first show another result which we will use in proving Donsker's Invariance Principle.

Theorem 5.39 (Skorokhod Embedding) Let μ be a measure on $\mathbf{R}$ of mean zero and finite variance σ^2. There exists a Probability Space $(\Omega, \mathscr{F}, \mathbf{P})$ equipped with a filtration $\{\mathscr{F}_t\}$, an adapted Brownian Motion B and a sequence of stopping times $\{T_n\}$, so that

- The process (T_n) is a random walk (has independent increments) of mean σ^2.
- The process S_n defined by $B(T_n)$ is a random walk with step distributed according to μ.

Remark 5.40 What this Theorem is telling us is that if we have any random walk S_n, we can construct a sequence of stopping times such that when we look at the Brownian Motion at those times we see exactly S_n.

Proof of Skorokhod's Theorem. The idea is to choose our stopping times $\{T_n\}$ to be the exit times of our Brownian Motion out of a sequence of randomly chosen intervals. The structure of μ will determine how we choose these intervals. Let us first define the following two objects:

$$\mu_{+/-}(A) = \mu(+/-A)$$

For a Borel set $A \in \mathscr{B}([0,\infty))$. These objects are measures, but not probability measures. Construct a probability space with a Brownian Motion (B_t) and a sequence (X_n, Y_n) of Random Variables that are independent of the Brownian Motion, that are IID and chosen from the probability measure $\nu(dx, dy)$ where

$$\nu(dx, dy) = C(x+y)\mu_-(dx)\mu_+(dy)$$

As for the filtration on which we are going to define our stopping times, we set $\mathscr{F}_0 = \sigma((X_n, Y_n) : n \in \mathbf{N})$ and $\mathscr{F}_t = \sigma(\mathscr{F}_0, \mathscr{F}_t^B)$. Then define $T_0 = 0$ and inductively:

$$T_n = \inf\{t > T_{n-1} : B_{t+T_{n-1}} - B_{T_{n-1}} \in \{-X_n, Y_n\}\}$$

It is clear that T_n is an $\mathscr{F}_n$ stopping time. Now we just need to show that this choice of stopping times satisfies the claim.

- Calculation: Since μ was assumed to have zero mean, we conclude that

$$0 = \int_{-\infty}^{0} x\mu(dx) + \int_0^\infty y\mu(dy) \qquad (\star)$$

Moreover, since ν was normalised by C to be a probability measure, we have that

$$\begin{aligned}1 = \int \nu(dx, dy) &= C\int\int(x+y)\mu_-(dx)\mu_+(dy)\\ &= C\int_0^\infty\left(\int_0^\infty x\mu_-(dx) + y\mu_-(\mathbf{R}^+)\right)\mu_+(dy)\\ &= C\left(\int_0^\infty x\mu_-(dx)\mu_+(\mathbf{R}^+) + \int_0^\infty y\mu_+(dy)\mu_-(\mathbf{R}^+)\right)\end{aligned}$$

Now we note from $(\star)$ that

$$\int_0^\infty y\mu(dy) = -\int_{-\infty}^0 x\mu(dx) = \int_0^\infty x\mu_-(dx)$$

And so combining we have that

$$1 = C(\mu_+(\mathbf{R}^+) + \mu_-(\mathbf{R}^+)) \int_0^\infty x\mu_-(dx)$$

But since $\mu_-(\mathbf{R}^+) = \mu(\mathbf{R}^-)$ we have that

$$1 = C \int_0^\infty x\mu_-(dx)$$

Which once again, rearranging and whatnot gives that

$$C \int_{-\infty}^0 (-x)\mu(dx) = C \int_0^\infty y\mu(dy) = 1$$

We're ready to roll. First note that our objective number one is to show that the steps $T_{n+1} - T_n$ are independent and have mean σ^2. However, to simplify our calculations, we make the following observation. Recall how

$$T_n = \inf\{t > T_{n-1} : B_{t+T_{n-1}} - B_{T_n} \in \{-X_n, Y_n\}\}$$

This is of course equivalent to saying that $T_n - T_{n-1} = \inf\{t > 0 : B_{t+T_{n-1}} - B_{T_{n-1}} \in \{-X_n, Y_n\}\}$. Of course by the Strong Markov Property, $\{B(t + T_{n-1}) - B(T_{n-1})\}$ is Standard Brownian Motion independent of $\mathscr{F}^+_{T_{n-1}}$. Since the pair (X_n, Y_n) is independent of everything, it follows that $T_n - T_{n-1}$ is independent of $\mathscr{F}^+_{T_{n-1}}$ and as such the increments are independent. Since the pairs (X_n, Y_n) are identically distributed, it also follows that the increments $T_n - T_{n-1}$ are identically distributed, so we only need to really show that $\mathbf{E}[T_1 - T_0]$ which by definition is just $\mathbf{E}[T_1]$ equals σ^2. Now recall how T_1 is just the hitting time of the set $\{X_1, Y_1\}$, so by conditioning on the value of (X_1, Y_1) we can just employ Gambler's Ruin:

$$\begin{aligned}
\mathbf{E}[T_1] &= \mathbf{E}[\mathbf{E}[T_1 \mid X_1, Y_1]] \\
&\overset{(!)}{=} \int_0^\infty \int_0^\infty xy\,\nu(dx, dy) \\
&= C \int_0^\infty \int_0^\infty xy(x+y)\mu_-(dx)\mu_+(dy) \\
&= \int_0^\infty x^2 \left(C \int_0^\infty y\mu_+(dy) \right) \mu_-(dx) + \int_0^\infty y^2 \left(C \int_0^\infty x\mu_-(dx) \right) \mu_+(dy) \\
&\overset{(!!)}{=} \int_{-\infty}^0 x^2 \mu(dx) + \int_0^\infty y^2 \mu(dy) = \sigma^2
\end{aligned}$$

Where step (!) is our beloved Independence Lemma, and step (!!) comes from the computations we did in the "Calculation" step above. A similar argument, which I'm too tired to reproduce, shows the claim for S_n. ♡

Remark 5.41 A much much easier version of this Theorem is for the case when X takes only two values a and b. In this case, we define T as the first exit time of the interval (a,b), and by some easy Martingale computations using the Martingale B and the Martingale B^2-t, we see that $B(T)\sim X$ and that $\mathbf{E}[T]=\mathbf{E}[X^2]$.

With this Theorem we can now prove finally Donsker's Theorem.

Proof of Donsker's Theorem. I will reproduce the proof of Donsker's Invariance from the "simpler" assumption that for a random variable X with mean zero and variance 1 (this is without loss of generality since we can just rescale the random walk later on), we have a stopping time T such that $B(T)\sim X$.

- Let us first construct the sequence of stopping times that embed the random walk into the Brownian Motion: By our assumption, we have a T_1 such that $B(T_1)\sim X$. Then by starting the Brownian Motion afresh at T_1, we can repeat the use of the assumption, and for this restarted Brownian Motion, find a stopping time T_2' such that at T_2', the restarted walk has law of X. Hence if w let $T_2=T_1+T_2$, we see that for our original Brownian Motion, $B(T_2)\sim X$. Inductively we find all the stopping times $T_1<T_2<\cdots$.

Now we start proving the Invariance principle.

- Recall that if $(S_n)_n$ is the random walk with step distribution given by X, we may extend linearly, and then rescale, so that we have

$$S_n^*(t):=\frac{1}{\sqrt{n}}S_n(t).$$

- Define $W_n(t)=\frac{B(nt)}{\sqrt{n}}$ and let $\epsilon>0$ be given. We are going to prove that

$$\mathbf{P}\left(\sup_{t\le 1}\left|W_n(t)-S_n^*(t)\right|>\epsilon\right)\to 0. \tag{1}$$

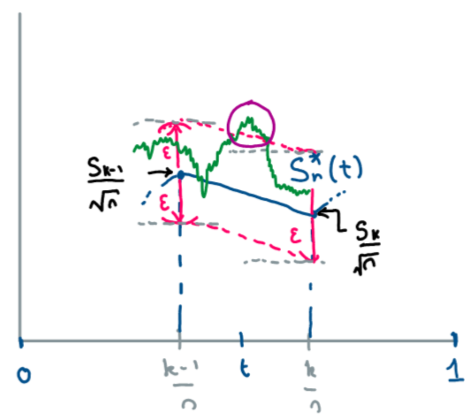
  Suppose momentarily that we have shown this to be true. Let us quickly show how this actually finishes the claim. Let $K\subseteq\mathscr{C}([0,1])$ be a closed subset of the space of continuous functions on the interval $[0,1]$ (closed with respect to the supremum norm). By Portmanteau's Theorem condition 2, it suffices to show that $\limsup_{n\to\infty}\mathbf{P}\left(S_n^*\in K\right)\le\mathbf{P}(B\in K)$, but observe that if we define $K[\epsilon]$ to be the ϵ-fattening of K with respect to the supremum norm,

$$\mathbf{P}\left(S_n^* \in K[\epsilon]\right) \leq \mathbf{P}(W_n \in K[\epsilon]) + \mathbf{P}\left(\sup_{t\leq 1}\left|W_n(t) - S_n^*(t)\right| > \epsilon\right).$$

(To verify this inequality, notice that if $W_n \notin K[\epsilon]$ and S_n^* is ϵ-close to W_n, then S_n^* cannot be in $K[\epsilon]$ either). Now we notice that since W_n has the distribution of a Brownian Motion by scaling invariance, $\mathbf{P}(W_n \in K[\epsilon]) = \mathbf{P}(B \in K[\epsilon])$, and the second summand vanishes as $n \to \infty$ by our assumption that 1 holds. Therefore sending $\epsilon \downarrow 0$ gives the claim.

Now all left to do is show that 1 holds.

- Define $k = k(t)$ to be the unique integer such that $\frac{k-1}{n} \leq t < \frac{k}{n}$. Then notice that on this interval, $S_n^*(t)$ is linear, so by looking at the *diagram that says it all:*



We see that if the event $A_n = \left\{\sup_{t\leq 1}|W_n(t) - S_n^*(t)| > \epsilon\right\}$ holds, then

$$A_n^* = \left\{\left|W_n(t) - \frac{S_{k-1}}{\sqrt{n}}\right| > \epsilon \text{ for some } t \in [0,1)\right\} \cup \left\{\left|W_n(t) - \frac{S_k}{\sqrt{n}}\right| > \epsilon \text{ for some } t \in [0,1)\right\}$$

holds. But since $S_k = B(T_k)$, this event can be rewritten as

$$A_n^* = \left\{|W_n(t) - W_n(T_{k-1}/n)| > \epsilon \text{ for some } t \in [0,1)\right\} \cup \left\{|W_n(t) - W_n(T_k/n)| > \epsilon \text{ for some } t \in [0,1)\right\}.$$

Now we notice that if A_n^* holds, then for a fixed $\delta \in (0,1)$,

$$\left\{\text{there are } t,s \in [0,2] \text{ with } |W_n(t) - W_n(s)| > \epsilon\right\} \cup \{|T_k/n - t| \wedge |T_{k-1}/n - t| > \delta \text{ for some } t \in [0,1)\}. \tag{2}$$

Indeed, if both of this don't hold, meaning that $|T_{k-1}/n - t| < \delta$ and $|T_k/n - t| \leq \delta$ for all $t \in [0,1)$ and whenever two values $t,s \in [0,2]$ are closer than δ, $|W_n(t) - W_n(s)| < \epsilon$, it can't be that A_n^* holds. (The reason for the $[0,2]$ interval is that since δ could be almost 1 and we are allowing the distance to be δ, then since $t \in [0,1)$ it could be that T_k/n is almost 2). However, the first term of 2 vanishes as $n \to \infty$ by uniform continuity of Brownian Motion on the interval $[0,2]$, and the second term vanishes because since the family $\{T_k - T_{k-1}\}$ are i.i.d with mean one (this is from the construction of the stopping time having expectation equal to the variance of X), we have that $\lim_{n\to\infty} T_n/n = 1$. Then we have the deterministic fact that if $(a_n)_n$ is a sequence with $a_n/n \to 1$, then $\sup_{k\le n} |a_k - k|/n \to 0$. Finally use the fact that $t \le k/n$ ♡

5.9 Zeroes of Brownian Motion

Let us recall a topological definition

Definition 5.42 (Isolated Point) Let X be a topological space and S be a subset. An element $x \in S$ is called an isolated point if there exists some neighbourhood U of x such that $U \cap S = \{x\}$, i.e: U contains no other points of S. For example, the set $S = \{1\} \cup (2,3)$ has 1 as an isolated point, as the neighbourhood $(1-\epsilon, 1+\epsilon)$ contains no other points of S other than $\{1\}$.

Theorem 5.43 (Zeros of Brownian Motion) Let B be Standard Brownian Motion and let $\mathscr{Z}$ be the zero set, i.e:

$$\mathscr{Z} = \{t; B(t) = 0\}$$

Then $\mathscr{Z}$ is closed and contains no isolated points.

Proof. To show that $\mathscr{Z}$ is almost surely closed, we make the following observation:

- Brownian motion is almost surely continuous. The singleton $\{0\}$ is closed in the topology we are working with, and since the preimage of a closet set under a continuous function is closed, we get that $\mathscr{Z} = B^{-1}\{0\}$ is almost surely closed. In other words:

$$1 = \mathbf{P}(B \text{ is continuous}) \leq \mathbf{P}(\mathscr{Z} \text{ is a closed set}).$$

To show that it contains no isolated points, we make the following definition: for a rational number $q \in \mathbf{Q}$, we define

$$\tau_q = \inf\{t \geq q : B_t = 0\}.$$

Then

- If $t \in \mathscr{Z}$ is a zero of the form τ_q for some $q \in \mathbf{Q}$, we can apply the Strong Markov Property at τ_q, which is an almost surely finite stopping time, and since the Brownian Motion starts afresh, we can apply the fact that for a Standard Brownian Motion, for any interval $[0, \epsilon)$, there will be a zero different from $t = 0$. Therefore these kind of zeroes are not isolated from the right, so not isolated.

- If $t \in \mathscr{Z}$ is not of the form τ_q for some $q \in \mathbf{Q}$, we can take a sequence of rational numbers $q_n \uparrow t$, and note that by definition of τ_{q_n}, $q_n \leq \tau_{q_n}$. Moreover, since t is a zero of Brownian motion and $q_n \leq t$, it follows by definition of τ_{q_n} that $\tau_{q_n} \leq t$. Therefore, $q_n \leq \tau_{q_n} \uparrow t$, and so $\tau_{q_n} \uparrow t$. This means in particular that for any $\epsilon > 0$, there is some τ_{q_n} (a zero of Brownian

Motion) with $|\tau_{q_n} - t| < \epsilon$, and since $\tau_{q_n} \leq t$, it follows that t is not isolated form the left, and so t is not isolated either.

♡

6 Poisson Random Measures

We begin by recalling a basic definition

Definition 6.1 (Poisson Random Variable) A random variable X that takes values in $\mathbf{Z}^{+}$ is said to have a Poisson distribution with parameter λ if

$$\mathbf{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}$$

An elementary property of the Poisson distribution is the following:

Proposition 6.2 (Addition Property) Let $N_i \sim \text{Poi}(\lambda_i)$ for $i \in \{1, \cdots, k\}$ be independent. Then

$$\sum_{i=1}^{k} N_i \sim \text{Poi}\left(\sum_{i=1}^{k} \lambda_i\right)$$

Proof. We do the argument by characteristic functions.

$$\begin{aligned}
\varphi\left(\sum_{i=1}^{k} N_i\right)(u) &= \mathbf{E}\left[\exp\left(iu\sum_{i=1}^{k} N_i\right)\right] \\
&\stackrel{(1)}{=} \prod_{i=1}^{k} \mathbf{E}[iuN_i] \\
&\stackrel{(2)}{=} \prod_{j=1}^{k} \exp\left(\lambda_j\{e^{iu} - 1\}\right) \\
&\stackrel{(3)}{=} \exp\left(\left(\sum_{j=1}^{k} \lambda_j\right)\{e^{iu} - 1\}\right)
\end{aligned}$$

Where (1) comes from the fact that the N_i are assumed to be independent, step (2) comes from direct computation of the characteristic function of the Poisson Law (See example 4.15. Step (3) is obvious and this finishes the proof since the last result is the characteristic function of a Poisson random variable with the desired parameter. The proof is finished because characteristic functions uniquely determine the law of a random variable. ♡

Proposition 6.3 (Splitting property) Let N and $\{Y_n\}$ be independent random variables, with $N \sim \text{Poi}(\lambda)$ and $\mathbf{P}[Y_n = j] = p_j$ for $j \in \{1, \cdots, k\}$ for some j. Then the random variables $N_1, \cdots, N_k$ defined by

$$N_i = \sum_{t=1}^{N} \mathbf{1}(\{Y_t = i\})$$

have that $N_i \sim \text{Poi}(\lambda p_i)$

Proof. We have the following

$$\mathbf{E}\left[\exp(iuN_j)\right] = \sum_{t=0}^{\infty} \mathbf{E}[\exp(iuN_j) \mid N = t] \frac{\lambda^t e^{-\lambda}}{t!}$$

Now we compute

$$\begin{aligned}
\mathbf{E}[\exp(iuN_j) \mid N = t] &= \mathbf{E}\left[\exp\left(iu \sum_{n=1}^{t} \mathbf{1}(\{Y_t = j\})\right)\right] \\
&= \prod_{n=1}^{t} \mathbf{E}[\exp(iu\,\mathbf{1}(\{Y_t = j\})] \\
&= \left(p_j\left(e^{iu} - 1\right) + 1\right)^t
\end{aligned}$$

So putting it all together we have that

$$\begin{aligned}
\mathbf{E}\left[\exp(iuN_j)\right] &= \sum_{t=0}^{\infty} \left(p_j\left(e^{iu} - 1\right) + 1\right)^t \frac{\lambda^t}{t!} e^{-\lambda} \\
&= \exp\left(\lambda p_j \left\{e^{iu} - 1\right\}\right)
\end{aligned}$$

as required. ♡

Definition 6.4 (Poisson random measure) Let $(\Omega, \mathscr{F}, \mathbf{P})$ be our underlying probability space, and let $(E, \mathscr{E}, \mu)$ be a σ-finite measure space. A Poisson random measure M with intensity μ is a map

$$M : \Omega \times \mathscr{E} \to \mathbf{Z}^+ \cup \{\infty\}$$

satisfying the following properties: if $\{A_k\}$ is a disjoint sequence of $\mathscr{E}$-measurable sets:

1. $M\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} M(A_k)$.
2. $\{M(A_k)\}_k$ is a collection of independent random variables.

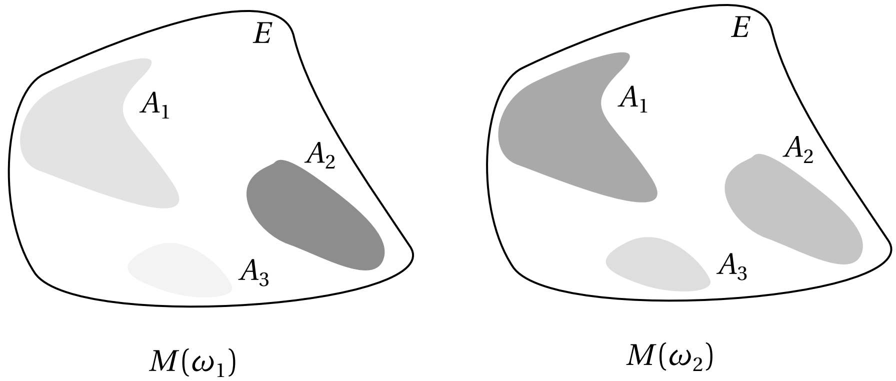


Figure 9: Illustration of point (3) of the definition of Poisson Random Measure.

3. $M(A_k) \sim \text{Poi}(\mu(A_k))$

It is not clear from the definition whether such objects exist. We now give a positive answer with the following Theorem:

Theorem 6.5 (Poisson Random Measures exist) Let $(E, \mathcal{E})$ be a measurable space, and denote by E^* the set of all $\mathbf{Z}^+ \cup \{\infty\}$-valued measures on $(E, \mathcal{E})$. For a set $A \in \mathcal{E}$ define the maps

$$X : E^* \times \mathcal{E} \to \mathbf{Z}^+ \cup \{\infty\} \qquad X_A : E^* \to \mathbf{Z}^+ \cup \{\infty\}$$

by

$$X(m, A) = X_A(m) = m(A).$$

Let $\mathcal{E}^*$ be the σ-algebra on E^* given by $\mathcal{E}^* = \sigma(X_A : A \in \mathcal{E})$. Then there exists a unique probability measure μ^* on $(E^*, \mathcal{E}^*)$, such that using $(E^*, \mathcal{E}^*, \mu^*)$ as our underlying probability space, we have that X is a Poisson Random Measure with intensity μ (where μ is a measure on $(E, \mathcal{E})$).

Proof. We first show uniqueness. We will do this through a π-system argument. Let $A_1, \cdots, A_k$ be disjoint sets of $\mathcal{E}$. Then define

$$\begin{aligned} A^* &= \{m \in E^* : m(A_1) = n_1, \cdots, m(A_k) = n_k\} \\ &=: \left\{m \in E^* : X_{A_1}(m) = n_1, \cdots, X_{A_k}(m) = n_k\right\} \\ &\equiv \left\{X_{A_1} = n_1, \cdots, X_{A_k} = n_k\right\} \end{aligned}$$

But by definition of μ^* making X a Poisson Random measure with intensity μ, we have, noting that $\{X(A_i)\} \equiv \{X_{A_i}\}$ are independent random variables, that

$$\begin{aligned}\mu^*(A^*) &= \mu^*\left\{X_{A_1} = n_1, \cdots, X_{A_k} = n_k\right\} \\ &= \prod_{i=1}^{n} \mu^*\left(X_{A_i} = n_i\right) \\ &\overset{(!)}{=} \prod_{i=1}^{n} \frac{\mu(A_i)^{n_i}}{n_i!} e^{-\mu(A_i)}\end{aligned}$$

Where step (!) comes from the fact that under μ^*, the random variable X_{A_i} is Poisson distributed with parameter $\mu(A_i)$. We have shown that any measure μ^* that turns X into a Poisson Random Measure with intensity μ, must have the same value on any A^* as defined above. Now we claim that the set S of all such A^* is a π-system. Indeed:

- $\varnothing \in S$. Indeed: such a set A^* could be formed by taking only one set $A_1 \in \mathscr{E}$, namely the empty set, then by setting

  $$A^* = \{m \in E^* : m(\varnothing) = 3\}$$

  we see that $A^* \in S$ and $A^* = \varnothing$ because any measure must assign 0 to the empty set.

- S is closed under finite intersections. Indeed: if A^*, B^* are two sets in S, then writing

  $$A^* = \{m : m(A_1) = n_1, \cdots, m(A_k) = n_k\} \qquad B^* = \{m : m(B_1) = m_1, \cdots, m(B_l) = m_l\}$$

  Where $\{A_i\}$ are disjoint and $\{B_j\}$ are disjoint. We now distinguish between the following cases:

  1. If **all** sets $A_1, \cdots, A_k, B_1, \cdots, B_l$ are disjoint then,

     $$A^* \cap B^* = \{m : m(A_1) = n_1, \cdots, m(A_k) = n_k, m(B_1) = m_1, \cdots, m(B_l) = m_l\}$$

     which is obviously an element of S.

  2. If not all of them are disjoint, then without loss of generality we have that for some i and j, $A_i \subseteq B_j$. We now further distinguish the following cases:

     – If in A^* A_i is prescribed to have measure n_i, and B_j is prescribed to have measure m_j, then if $n_i > m_j$, then $A^* \cap B^* = \varnothing$ which we have already established to be a set in S.

     – if $n_i \leq m_j$, then we may express:

     $$A^* \cap B^* = \{m : m(A_1) = n_1, \cdots, m(A_i) = n_i, \cdots, m(B_j \setminus A_i) = m_j - n_1, \cdots\}$$

     and so $A^* \cap B^*$ can be written as a valid set in S.

Therefore we have established that S is a π-system. It is easy to see that S generates $\mathcal{E}^*$. Therefore by the uniqueness Lemma, we have that μ^* is unique. We now show existence:

We first suppose that $\lambda := \mu(E) < \infty$. We now invoke $N \sim$ Poi(λ) and $Y_n \sim \mu/\lambda$ to be independent random variables. We now set

$$M(A) = \sum_{n=1}^{N} \mathbf{1}(Y_n \in A)$$

and we check that this is indeed a Poisson Random measure. We first note that if $\{A_k\}$ is a collection of disjoint sets,

$$\begin{aligned} M\left(\bigcup_{k=1}^{\infty} A_k\right) &= \sum_{n=1}^{N} \mathbf{1}\left(Y_n \in \bigcup_{k=1}^{\infty} A_k\right) \\ &\overset{(!)}{=} \sum_{k=1}^{\infty} \sum_{n=1}^{N} \mathbf{1}(Y_n \in A_k) = \sum_{k=1}^{N} M(A_k) \end{aligned}$$

Where step (!) comes from the fact that $\{A_k\}$ are disjoint. ♡

7 Homework Problems

Question 7.1 (1, ES1) Let X and Y be integrable random variables and suppose that

$$\mathbf{E}[X \mid Y] = Y \quad \mathbf{E}[Y \mid X] = X$$

almost surely. Show then that $X = Y$ almost surely.

Proof. Fix a $c \in \mathbf{R}$. Then we note that since $\{X > c\} \in \sigma(X)$, we have $\mathbf{E}[(X - Y)\mathbf{1}_{\{X>c\}}] = 0$. But we can write this set as the union of two disjoint sets, namely

$$\{X > c\} = \{X > c, Y \geq c\} \cup \{X > c, Y < c\}$$

and plugging in and using the fact that the indicator function will split as a sum,

$$\mathbf{E}\big[(X - Y)\mathbf{1}_{\{X>c,Y\geq c\}}\big] + \mathbf{E}\big[(X - Y)\mathbf{1}_{\{X>c,Y<c\}}\big] = 0$$

Since the right hand summand is greater than or equal to zero, we must have that the left summand is less than or equal to zero, i.e:

$$\mathbf{E}\big[(X - Y)\mathbf{1}_{\{X>c,Y\geq c\}}\big] \leq 0$$

We can repeat this argument switching X and Y and using the second assumption to obtain

$$\mathbf{E}\big[(X - Y)\mathbf{1}_{\{X>c,Y\geq c}\}\big] \geq 0$$

i.e: $\mathbf{E}\big[(X - Y)\mathbf{1}_{\{X>c,Y\geq c\}}\big] = 0$. Recall the equation

$$\mathbf{E}\big[(X - Y)\mathbf{1}_{\{X>c,Y\geq c\}}\big] + \mathbf{E}\big[(X - Y)\mathbf{1}_{\{X>c,Y<c\}}\big] = 0$$

Using what we have just shown, we must have that

$$\mathbf{E}\big[(X - Y)\mathbf{1}_{\{X>c,Y<c\}}\big] = 0$$

But the integrand is strictly positive, so it must be that the set $\{X > c > Y\}$ has zero probability. Since the event

$$\{X > Y\} = \bigcup_{q\in\mathbf{Q}} \{X > q > Y\}$$

we have that $\mathbf{P}\{X > Y\} = 0$. Which means that $X \leq Y$ almost surely. By symmetry, we obtain that $X \geq Y$ almost surely and as such the claim follows.

♡

Question 7.2 (2, ES1) Let X, Y be iid Bernoulli p RVs. Let $Z = \mathbf{1}_{\{X+Y=0\}}$. Compute $\mathbf{E}[X \mid Z]$ and $\mathbf{E}[Y \mid Z]$.

Proof. We start by noting that $\sigma(\mathbf{1}_A) = \{\varnothing, A, A^c, \Omega\}$. Therefore $\sigma(Z) = \{\varnothing, \{Z = 0\}, \{Z = 1\}, \Omega\}$. Which means we are in the case of a countable sigma-algebra, and as such we know that

$$\mathbf{E}[X \mid Z] = \mathbf{E}[X \mid Z = 0]\,\mathbf{1}_{\{Z=0\}} + \mathbf{E}[X \mid Z = 1]\,\mathbf{1}_{\{Z=1\}}$$

We now only need to compute $\mathbf{E}[X \mid Z = 0]$, which is just

$$\mathbf{P}(X = 1 \mid X + Y = 0) = 0$$

and also $\mathbf{E}[X \mid Z = 1]$, which is just

$$\mathbf{P}(X = 1 \mid X + Y = 1) = \mathbf{P}(Y = 0) = p$$

which gives

$$\mathbf{E}[X \mid Z] = p\,\mathbf{1}_{\{Z=1\}}$$

By symmetry that's the same for Y. ♡

Question 7.3 (3, ES1) Let X and Y be two independent Exp(θ) RVs and let $Z = X + Y$. Show that Z is $\Gamma(2, \theta)$, i.e: its density is $\theta^2 x \exp(-\theta x)\mathbf{1}(x \geq 0)$. Show that for any non-negative Borel h:

$$\mathbf{E}[h(X) \mid Z] = \frac{1}{Z}\int_0^Z h(u)\,du$$

Proof. We first begin by determining the density of $Z = X + Y$. But it is a standard result that this is given by a convolution:

$$f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\,dx = \theta^2 \int_{-\infty}^{\infty} e^{-\theta x} e^{-\theta(z-x)}\,\mathbf{1}(0 < x < z)\,dx = \theta^2 z e^{-\theta z}.$$

Now that we have a density for Z, we can compute the elementary conditional density

$$f_{X|Z}(x \mid z) = \frac{f_{X,Z}(x,z)}{f_Z(z)}.$$

To find $f_{X,Z}(x,z)$ given that we know $f_{X,Y}$ we are going to have to use transformation of random variables (the dreaded Jacobean), and we see that

$$f_{X,Z}(x,z) = f_X(x) f_Y(z-x)$$

So combining we have that

$$f_{X|Z}(x \mid z) = \frac{f_X(x) f_Y(z-x)}{f_Z(z)} \mathbf{1}(0 \le x \le z)$$

and as such we have that $\mathbf{E}[X \mid Z = z] = \frac{1}{z}\int_0^z h(x) dx$ so our desired result is that

$$\mathbf{E}[h(X) \mid Z] = \frac{1}{Z}\int_0^Z h(x) dx$$

♡

Question 7.4 Conversely, let Z be a random variable with a $\Gamma(2,\theta)$ distribution and suppose X is a random variable whose conditional distribution given Z is uniform on $[0,Z]$. Namely that

$$\mathbf{E}[h(X) \mid Z] = \frac{1}{Z}\int_0^Z h(x) dx$$

Show that X and $X - Z$ are independent exponentials.

Proof. We have that by simple transformations of random variables and assumption of conditional density being uniform:

$$f_{X,X-Z}(x,z) = f_{X,Z}(x, z-x) = \frac{f_Z(z)}{z} = \theta^2 e^{-\theta z} = \theta^2 e^{-\theta(x+z-x)} = \theta e^{-\theta x} \theta e^{-\theta(z-x)}$$

as required. ♡

Question 7.5 (4, ES1) Let $X \ge 0$ be a random variable on $(\Omega, \mathscr{F}, \mathbf{P})$ and $\mathscr{G} \subseteq \mathscr{F}$ be a sub-sigma algebra. Show that $X > 0$ implies $\mathbf{E}[X \mid \mathscr{G}] > 0$ almost surely. Furthermore, show that $\{\mathbf{E}[X \mid \mathscr{G}] > 0\}$ is the smallest $\mathscr{G}$-measurable event that contains $\{X > 0\}$ up to zero probability events.

Proof. For convenience let $X' = \mathbf{E}[X \mid \mathscr{G}]$. It is clear that the event $A = \{X' \le 0\}$ is $\mathscr{G}$-measurable. Therefore

$$\mathbf{E}[X \mathbf{1}_A] = \mathbf{E}\left[X' \mathbf{1}_A\right] \le 0$$

But since $X > 0$ by assumption, then the only way it can be the case that $\mathbf{E}[X\,\mathbf{1}_A]$ is if $\mathbf{P}(A) = 0$. That is, $X' > 0$ almost surely. To show the second part, we suppose that there exists an $A \in \mathscr{G}$ with

$$\{X' > 0\} \supseteq A \supseteq \{X > 0\}$$

such that A is not just $\{X > 0\}$ union a collection of events of zero probability. That is to say, there is some $S \in A \setminus \{X > 0\}$ of non-zero probability. Then

$$\mathbf{E}[X \mid \mathscr{G}]\,\mathbf{1}_S > 0$$

because $S \subseteq \{X' > 0\}$, but also, $X\,\mathbf{1}_S \leq 0$, which by monotonicity of expectation implies $\mathbf{E}[X\,\mathbf{1}_S] \leq 0$. Since $0 \geq \mathbf{E}[X\,\mathbf{1}_S] = \mathbf{E}[X'\,\mathbf{1}_S] > 0$ we reach a contradiction. ♡

Question 7.6 (2.1, ES1) Let (X_n) be an integrable process with values in a countable subset $E \subseteq \mathbf{R}$. Show that (X_n) is a Martingale with respect to the natural filtration if and only if for every n, every $i_0, \cdots, i_{n-1} \in E$, we have that

$$\mathbf{E}[X_n \mid X_0 = i_0, \cdots, X_{n-1} = i_{n-1}] = i_{n-1}.$$

Proof. Suppose (X_n) is a Martingale, then we know that for every $A \in \mathscr{F}_{n-1}$, we have that $\mathbf{E}[X_n\,\mathbf{1}_A] = \mathbf{E}[X_{n-1}\,\mathbf{1}_A]$. Taking $A = \{X_0 = i_0, \cdots, X_{n-1} = i_{n-1}\} \in \mathscr{F}_{n-1}$, we determine that

$$\mathbf{E}[X_n \mid X_0 = i_0, \cdots, X_{n-1} = i_{n-1}] = \frac{\mathbf{E}[X_n\,\mathbf{1}_A]}{\mathbf{P}(A)} = \frac{\mathbf{E}[X_{n-1}\,\mathbf{1}_A]}{\mathbf{P}(A)} = \frac{\mathbf{E}[i_{n-1}\,\mathbf{1}_A]}{\mathbf{P}(A)} = \frac{i_{n-1}\mathbf{E}[\mathbf{1}\,A]}{\mathbf{P}(A)} = i_{n-1}$$

For the converse, we simply note that since (X_n) takes values in a countable set, then $\mathscr{F}_{n-1}$ is generated by sets of the form $\{X_0 = i_0, \cdots, X_{n-1} = i_{n-1}\}$, which establishes the property that $\mathbf{E}[X_n\,\mathbf{1}_A] = \mathbf{E}[X_{n-1}\,\mathbf{1}_A]$. ♡

8 Appendix

Here are all the proofs, Lemmas, or technicalities that for the sake of clarity were skipped from the main manuscript, but I still deemed them worthy of a discussion.

8.1 Measure Theory

8.1.1 π-systems, extensions and uniqueness

In general, σ-algebras are not "nice-enough" to obtain an explicit description of its typical element, hence defining a measure directly by specifying its values on every set of the σ-algebra is impossible.

The way around it is to find a more manageable set that generates the σ-algebra, and define the measure on that set. Of course, two questions arise naturally: can this function be extended to a measure on the whole σ-algebra, and is this extension unique? These two questions are dealt by Carethedory's Theorem and Dynkin's π-system lemma.

Definition 8.1 (π-system, d-system) A collection of sets E that is closed under finite intersections is called a π-system.

A collection of sets E that contains the entire space and has the property that for each $B \subseteq A \in E$, one has that

$$A \setminus B \in E$$

and for each increasing sequence of sets (A_n), one has

$$\bigcup_n A_n \in E$$

is called a d-system.

Lemma 8.2 (Dynkin's π-system Lemma) Let $\mathscr{A}$ be a π-system. Then if $\mathscr{B}$ is a d-system with $\mathscr{A} \subseteq \mathscr{B}$, then $\sigma(\mathscr{A}) \subseteq \mathscr{B}$.

Theorem 8.3 (Carethedory's Extension Theorem) Let $\mathscr{R}$ be a ring, i.e: a collection of sets that contains $\emptyset$, and for all $A, B \in \mathscr{R}$:

$$A \setminus B \in \mathscr{R} \qquad A \cup B \in \mathscr{R}$$

If $f : \mathscr{R} \to [0, \infty]$ is a countably additive set function, then f extends to a measure on $\sigma(\mathscr{R})$.

Theorem 8.4 (Uniqueness of extension) Let μ_1, μ_2 be two finite measures on a measurable space $(\Omega, \mathscr{F})$ that agree on a π-system that generates $\mathscr{F}$. Then $\mu_1 = \mu_2$.

Theorem 8.5 (Monotone Class Theorem) Suppose that $(\Omega, \mathscr{F})$ is a measurable space, and $\mathscr{F}$ is generated by some π-system $\mathscr{A}$. Suppose V is a linear function space such that

1. $\mathbf{1}_\Omega \in V$.
2. For any $A \in \mathscr{A}$, $\mathbf{1}_A \in V$.

3. If (f_n) is a bounded, non-negative sequence in V such that $f_n \uparrow f$ where f is also bounded and non-negative, then $f \in V$.

Then $V = \mathrm{b}\mathscr{F}$

Theorem 8.6 (Characterisation of measurability) Let X and Y be two random variables. If X is $\sigma(Y)$-measurable, then $X = f(Y)$ for some Borel f.

8.2 Conditional Expectation

Lemma 8.7 Given a sequence of events (F_n), then

$$\limsup \mathbf{1}_{F_n} = \mathbf{1}_{\limsup F_n}$$

Proof. Recall the definition

$$\limsup F_n = \{\omega \in F_n \text{ infinitely often}\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} F_n$$

Thus suppose $\omega \in \limsup F_n$, then $\mathbf{1}_{F_n}(\omega) = 1$ for infinitely many n, and as such, $\limsup \mathbf{1}_{F_n}(\omega) = 1$. On the contrary, suppose that $\omega \notin \limsup F_n$, then ω ceases to be in F_n eventually, that is to say, $\mathbf{1}_{F_n}(\omega) = 0$ for all n large enough, and as such, $\limsup \mathbf{1}_{F_n}(\omega) = 0$. ♡

Lemma 8.8 (Fubini-Trick) Let X_t be a process and $\mathscr{G}$ a sub-sigma algebra, for which we have $\int_a^b \mathbf{E}[|X_u|]du = \mathbf{E}\int_a^b |Xu|du$ are equal and finite. Then

$$\mathbf{E}\left[\int_a^b X_u du \middle| \mathscr{G}\right] = \int_a^b \mathbf{E}[X_u \mid \mathscr{G}]du$$

Proof. Let $A \in \mathscr{G}$ be any $\mathscr{G}$-measurable set. For convenience label $Y_r = \mathbf{E}[X_r \mid \mathscr{G}]$ Then by definition of conditional expectation, $\mathbf{E}[Y_r \mathbf{1} A] = \mathbf{E}[X_r \mathbf{1} A]$, and so we have that

$$\int_a^b \mathbf{E}[Y_r \mathbf{1} A]dr = \int_a^b \mathbf{E}[X_r \mathbf{1} A]dr$$

Then using the regularity conditions (the integrals of the absolute value are finite) and Fubini's

Theorem:

$$\mathbf{E}\left[\int_a^b Y_r \mathbf{1} A d r\right] = \mathbf{E}\left[\int_a^b X_r \mathbf{1} A d r\right]$$

Taking $\mathbf{1} A$ out of the integrals, for the function $\mathbf{1} A$ does not depend on r finishes the claim. ♡

8.3 Independence

Definition 8.9 (Independence of a process and a σ-algebra) Let (X_t) be a stochastic process on a filtered space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, \mathbf{P})$ and let $\mathscr{G} \subseteq \mathscr{F}$ be a sub-sigma algebra. Recall that a process (X_t) is a random variable $\Omega \to D := \{\text{functions } f : \mathbf{R}^+ \to E\}$ where $(E, \mathscr{E})$ is some other measurable space. The space D is endowed with the sigma algebra $\mathscr{D}$ that makes the projection maps measurable, and so we say that (X_t) is independent of $\mathscr{G}$ if for any $A \in \mathscr{D}$, and any $B \in \mathscr{G}$,

$$\mathbf{P}(\{(X_t) \in A\}, B) = \mathbf{P}(\{X_t \in A\})\mathbf{P}(B)$$

Since the law of the process (X_t) is determined by its finite dimensional distributions, it is enough to check that for all $t_1, t_2, \cdots, t_k \in \mathbf{R}$, one has that the random vector $(X_{t_1}, X_{t_2}, \cdots, X_{t_k})$ is independent of $\mathscr{G}$. This is equivalent to checking that for all $B \in \mathscr{G}$, and all $A \in \mathscr{E}^k$,

$$\mathbf{P}[(X_{t_1}, \cdots, X_{t_k}) \in A, B] = \mathbf{P}[(X_{t_1}, \cdots, X_{t_k}) \in A]\mathbf{P}[B]$$

Of course we can express this by checking that

$$\mathbf{E}[\mathbf{1}(\{(X_{t_1}, \cdots, X_{t_k}) \in A\})\mathbf{1}(B)] = \mathbf{P}[B]\mathbf{E}[\mathbf{1}(\{(X_{t_1}, \cdots, X_{t_k}) \in A\})]$$

and the reason we've written it like this is that if we can show that for any bounded and measurable $f : E^k \to \mathbf{R}$ we can show that

$$\mathbf{E}[f(X_{t_1}, \cdots, X_{t_k})\mathbf{1}(B)] = \mathbf{P}[B]\mathbf{E}[f(X_{t_1}, \cdots, X_{t_k})] \qquad (\star)$$

we can just take $f(\cdot) = \mathbf{1}(\cdot \in A)$ and we are done. Even better, if we can show that $(\star)$ holds for bounded and continuous functions, then by using DCT we can pass a limit and approximate any measurable function by the continuous functions. This is the heart of the argument of say the Simple Markov Property and similar results.

Lemma 8.10 (Independence Lemma) Let X and Y be two random variables, and let $\mathscr{G}$ be a sigma algebra for which Y is $\mathscr{G}$-measurable and $\sigma(X)$ is independent of $\mathscr{G}$. Then for all bounded and measurable Φ:

$$\mathbf{E}[\Phi(X,Y) \mid \mathscr{G}] = \mathbf{E}[\Phi(X,Y) \mid X]$$

Which means that if X has a density f with respect to the Lebesgue measure:

$$\mathbf{E}[\Phi(X,Y) \mid \mathscr{G}] = \int f(x)\Phi(x,Y)\,dx$$

The proof of this goes by the standard machine and it can be found in [SP12, Lemma A.3]

Index

$\mathscr{L}^p$ convergence of Martingales, 38

Doob's $\mathscr{L}^p$ inequality, 37
Doob's maximal inequality, 36

References

[idk] idk. *notes*. `https://www.stat.cmu.edu/~cshalizi/754/notes/lecture-02.pdf`.

[SP12] René L. Schilling and Lothar Partzsch. *Brownian Motion: An Introduction to Stochastic Processes*. 2nd. Vol. 51. De Gruyter Studies in Mathematics. Berlin: De Gruyter, 2012. isbn: 978-3-11-025382-3. doi: `10.1515/9783110253830`.

[Wil14] David Williams. *Probability with Martingales*. Cambridge University Press, 2014.