NOTES ON EXERCISE SHEETS IN PROBABILITY

CONSTANTIN KOGLER

These are my personal notes for exercise classes in probability I have taught at Oxford for third- and fourth-year undergraduates. The notes for C8.2 were written jointly with Tassilo Schwarz. I have collected these notes here for my own reference, and in case others may find them useful. I am grateful to the teaching assistants for their insights and to the students for their thoughtful questions.

Please send corrections to my personal email address shown on my website.

Contents

1. B8.1 Class 1

1.1. Motivation for Measure Theory. In integration theory, one usually first learns the Riemann integral. While the Riemann integral is intuitive, it has several disadvantages:

(1) $1_{\mathbb{Q}\cap[0,1]}$ is not Riemann integrable. But we want this function to have integral 0.
(2) It is desirable to have
$$\int \lim_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu$$
under suitable assumptions. No result like monotone convergence or dominated convergence holds for the Riemann integral.
(3) Does not generalize easily to other spaces or ways of measuring the space.
(4) The space of Riemann integrable functions is not complete, i.e. not every Cauchy sequence converges.

Integrating functions is related to measuring sets. Indeed if μ is a measure and g is a simple function , i.e. $g = \sum_i a_i 1_{A_i}$ is a finite sum of characteristic functions, then we define
$$\int f \, d\mu = \sum_i a_i \mu(A_i).$$
This allows us to define the integral of a positive function f as
$$\int f \, d\mu := \sup\left\{ \int g \, d\mu \,:\, g \text{ simple and } 0 \le g \le f \right\}.$$

In an ideal world, we would be able to measure every set. But this is not possible for $\mathbb{R}$ as the following example shows. Denote by $\mathscr{P}(\mathbb{R})$ the set of subsets of $\mathbb{R}$.

Theorem 1.1. *(Dystopia of Measure Theory) There is no function $\lambda : \mathscr{P}(\mathbb{R}) \to [0,\infty]$ satisfying the following properties:*

(1) $\lambda([a,b]) = b - a$ for all $a < b$.
(2) $\lambda(A + x) = \lambda(A)$ for all $A \in \mathscr{P}(\mathbb{R})$ and $x \in \mathbb{R}$.
(3) If $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathscr{P}(\mathbb{R})$, then it holds that
$$\lambda\left(\bigcup_{i\ge1} A_i\right) = \sum_{i\ge1} \lambda(A_i).$$

Proof. The proof is by contradiction. Assume that such a function exists and consider the quotient $\mathbb{R}/\mathbb{Q}$. Then for every equivalence class $c \in \mathbb{R}/\mathbb{Q}$, using the axiom of choice we choose an $x_c \in [0,1]$ representing that class, i.e. such that $c = x_c + \mathbb{Q}$. Denote $V = \{x_c \,:\, c \in \mathbb{R}/\mathbb{Q}\} \subset [0,1]$.

Let $q_1, q_2, \ldots$ be an enumeration of the rational numbers in $[-1,1]$ and note that the sets $V + q_i$ are all disjoint. Indeed assume that $x \in (V + q_i) \cap (V + q_j)$ for some i and j. Then there are equivalence classes $c, c' \in \mathbb{R}/\mathbb{Q}$ such that $x = x_c + q_i = x_{c'} + q_j$. It follows that $c = c'$ and thus $x_c = x_{c'}$ and therefore $q_i = q_j$.

Also it holds that $[0,1] \subset \bigcup_{i\ge1} V + q_i \subset [-1,2]$. Therefore by (1) and (2)
$$1 \le \sum_{i\ge1} \lambda(V + q_i) = \sum_{i\ge1} \lambda(V) \le 3.$$
This is a contradiction however as $\sum_{i\ge1} \lambda(V)$ is either 0 or ∞. □

1.2. **σ-algebras.** By Theorem 1.1, if we want measure subsets of $\mathbb{R}$, it only makes sense to work with a subset of $\mathscr{P}(\mathbb{R})$. This leads to the definition of a σ-algebra.

Definition 1.2. *Given a set Ω, a collection of subsets $\mathscr{A} \subset \mathscr{P}(\Omega)$ is called a σ-algebra if the following properties hold:*

(1) (Non trivial) $\emptyset, \Omega \in \mathscr{A}$
(2) (Complements) If $A \in \mathscr{A}$ then $A^c \in \mathscr{A}$.
(3) (Countable unions) If $A_1, A_2, \dots$ are in $\mathscr{A}$ then $\bigcup_{i \geq 1} A_i \in \mathscr{A}$.

We remark the following:

(1) As $(\bigcup A_i)^c = \bigcap A_i^c$, it follows that a σ-algebra is stable under countable intersections.
(2) An arbitrary intersection of σ-algebras is again a σ-algebra.
(3) A union of σ-algebras is not necessarily a σ-algebra. Indeed, if $\Omega = \{1, 2, 3\}$, Then consider the σ-algebras

$$\mathscr{A}_1 = \{\emptyset, \Omega, \{1\}, \{2, 3\}\} \quad \text{and } \mathscr{A}_2 = \{\emptyset, \Omega, \{2\}, \{1, 3\}\}.$$

Then $\mathscr{A}_1 \cup \mathscr{A}_2$ is not a σ-algebra as $\{1, 2\} = \{1\} \cup \{2\}$ is not in $\mathscr{A}_1 \cup \mathscr{A}_2$.

We give the following list of important examples:

(1) $\{\emptyset, \Omega\}$ and $\{\emptyset, \Omega, A, A^c\}$ for any $A \in \Omega$ are basic examples of σ-algebras.
(2) On a topological space X, the Borel σ-algebra $\mathscr{B}(X)$ is the smallest σ-algebra that contains all open sets of X.
(3) The set constructed in Theorem 1.1 is an example of a set not in $\mathscr{B}(\mathbb{R})$.
(4) The power set $\mathscr{P}(\Omega)$ is a σ-algebra, but it is sometimes not useful.
(5) Let $\mathcal{P} = (P_j)_{j \geq 1}$ be a partition of a set Ω. Then the collection of sets $\mathcal{U}(\mathcal{P})$ consisting of all possible unions of $\mathcal{P}$ is a σ-algebra.

Lemma 1.3. *If Ω is countable, every σ-algebra arises from a partition.*

Proof. Let $\mathscr{A}$ be a σ-algebra on Ω. For each $x \in \Omega$, we define

$$[x]_{\mathscr{A}} = \bigcap_{A \in \mathscr{A}, x \in A} A.$$

We claim that $[x]_{\mathscr{A}} \in \mathscr{A}$. Indeed for each element $y \in \Omega \backslash [x]_{\mathscr{A}}$ there is some set $A_y \in \mathscr{A}$ such that $x \in A_y$ and $y \notin A_y$. Therefore

$$[x]_{\mathscr{A}} = \bigcap_{y \in \Omega \backslash [x]_{\mathscr{A}}} A_y$$

and so $[x]_{\mathscr{A}}$ is a countable intersection of elements in $\mathscr{A}$ and therefore itself in $\mathscr{A}$.

Note that if for two elements $x, y \in X$ we have that $[x]_{\mathscr{A}} \subset [y]_{\mathscr{A}}$, then $[x]_{\mathscr{A}} = [y]_{\mathscr{A}}$ as otherwise it holds that $y \in A = [y]_{\mathscr{A}} \backslash [x]_{\mathscr{A}}$ and therefore $x \in A^c$ and $[x]_{\mathscr{A}}$ and $[y]_{\mathscr{A}}$ are disjoint. The latter implies that $[x]_{\mathscr{A}}$ is a partition of Ω. Indeed, if $[x]_{\mathscr{A}} \cap [y]_{\mathscr{A}}$ is non-empty for some $x, y \in \Omega$, then there is some $z \in [x]_{\mathscr{A}} \cap [y]_{\mathscr{A}}$. It follows that $[z]_{\mathscr{A}} \subset [x]_{\mathscr{A}}$ as well as $[z]_{\mathscr{A}} \subset [y]_{\mathscr{A}}$ and therefore $[x]_{\mathscr{A}} = [z]_{\mathscr{A}} = [y]_{\mathscr{A}}$. Moreover, every set $A \in \mathscr{F}$ is a union

$$A = \bigcup_{x \in A} [x]_{\mathscr{A}}.$$

This concludes the proof that $\mathscr{A}$ arises from a partition. $\square$

1.3. **Measures.** Recall the following definition.

Definition 1.4. *Let $\mathscr{A}$ be a collection of subsets of Ω containing the empty set $\emptyset$. A set function on A is a function $\mu : \mathscr{A}$ with $\mu(\emptyset) = 0$. We say that μ is countably additive, or σ-additive, if for all sequences $(A_n)_{n\geq 1}$ of disjoint sets in $\mathscr{A}$ with $\bigcup_{n\geq 1} A_n \in \mathscr{A}$ we have*

$$\mu\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Lemma 1.5. *Let $\mu : \mathscr{A} \to [0, \infty)$ be an additive set function on an algebra $\mathscr{A}$ taking only finite values. Show that μ is countably additive if and only if for every sequence $(A_n)_{n\geq 1}$ of sets in $\mathscr{A}$ with $A_n \downarrow \emptyset$ we have $\lim_{n\to\infty} \mu(A_n) = 0$.*

Proof. If μ is countably additive, consider $B_n = \Omega \backslash A_n$ and note $B_n \uparrow \Omega$. It follows that $\lim_{n\to\infty} \mu(B_n) = \mu(\Omega)$ (which are all finite) and therefore $\lim_{n\to\infty} \mu(A_n) = 0$.

On the other hand, if the claim holds, then let $(A_n)_{n\geq 1}$ be a sequence of disjoint sets in $\mathscr{A}$ with $A = \bigcup_{n\geq 1} A_n \in \mathscr{A}$. Then set $C_n = A \backslash \left(\bigcup_{1\leq i\leq n} A_i\right)$ and note that $C_n \downarrow \emptyset$. Thus the claim follows since μ is additive and therefore

$$\mu(A) = \mu\left(\bigcup_{1\leq i\leq n} A_i\right) + \mu(C_n) = \sum_{i=1}^{n} \mu(A_i) + \mu(C_n),$$

implying the claim by sending $n \to \infty$ and using that $\mu(C_n) \to 0$. $\square$

1.4. π-λ **systems Lemma.** We first recall the following definitions.

Definition 1.6. *A collection of sets $\mathscr{A}$ is called a π-system if it is stable under intersections, i.e. $A, B \in \mathscr{A}$ implies $A \cap B \in \mathscr{A}$.*

Definition 1.7. *A collection of sets $\mathscr{M}$ is called a λ-system if the following properties are satisfied:*

(1) $\Omega \in \mathscr{M}$,
(2) *If $A, B \in \mathscr{M}$ with $A \subset B$ then $B \backslash A \in \mathscr{M}$,*
(3) *If $A_1 \subset A_2 \subset \ldots$ is an increasing sequence of subsets of $\mathscr{M}$, then $\bigcup_{i\geq 1} A_i \in \mathscr{M}$.*

We note the following:

(1) A collection of subsets is a σ-algebra if and only if it is a π-system and a λ-system.
(2) (π-λ systems Lemma) Let $\mathscr{M}$ be a λ-system and $\mathscr{A}$ be a π-system. Then if $\mathscr{A} \subset \mathscr{M}$ it holds that $\sigma(\mathscr{A}) \subset \mathscr{M}$.

An example where the π-λ-systems lemma is useful, is the following lemma.

Lemma 1.8. *Let μ_1 and μ_2 be finite measures on a measurable space $(\Omega, \mathscr{F})$ with $\mu_1(\Omega) = \mu_2(\Omega)$. Then the collection of sets $\{A \in \mathscr{F} : \mu_1(A) = \mu_2(A)\}$ is a λ-system. In particular, if μ_1 and μ_2 agree on a π-system $\mathscr{A}$ with $\sigma(\mathscr{A}) = \mathscr{F}$ then they agree on the whole $\mathscr{F}$.*

Proof. The second claim follows from the first and the π-λ-systems lemma. To show the first, denote by $\mathscr{M} = \{A \in \mathscr{F} : \mu_1(A) = \mu_2(A)\}$. Then by assumption

$\Omega \in \mathscr{M}$. Next consider $A, B \in \mathscr{M}$ with $A \subset B$. As A and $B \backslash A$ are disjoint, it follows that

$$\mu_1(A) + \mu_1(B\backslash A) = \mu_1(B) = \mu_2(B) = \mu_2(A) + \mu_2(B\backslash A).$$

As $\mu_1(A) = \mu_2(A)$ by assumption, we conclude $\mu_1(B\backslash A) = \mu_2(B\backslash A)$ and therefore $B\backslash A \in \mathscr{M}$. Finally, if $A_1 \subset A_2 \subset \dots$ is an increasing sequence of subsets of $\mathscr{M}$, then writing $A = \bigcup_{i\geq 1} A_i$ by continuity from above,

$$\mu_1(A) = \lim_{n\to\infty} \mu_1(A_n) = \lim_{n\to\infty} \mu_2(A_n) = \mu_2(A).$$

This concludes the proof. $\square$

We recall the monotone class theorem.

Theorem 1.9. *(Monotone Class Theorem) Let $\mathscr{H}$ be a class of bounded functions from Ω to $\mathbb{R}$ satisfying the following:*

(1) $\mathscr{H}$ is a vector space.
(2) The constant function 1 is in $\mathscr{H}$.
(3) If $(f_n)_{n\geq 1}$ is a sequence in $\mathscr{H}$ such that $f_n \uparrow f$ for a bounded function f, then $f \in \mathscr{H}$.

If $\mathscr{C} \subset \mathscr{H}$ is stable under pointwise multiplication, then $\mathscr{H}$ contains all bounded $\sigma(\mathscr{C})$-measurable functions.

Theorem 1.10. *On a measurable space $(\Omega, \mathscr{F})$, let $X_1, \dots, X_k$ be random variables and let $\mathscr{G} = \sigma(X_1, \dots, X_k)$. Consider*

$$\mathscr{A} = \left\{ \bigcap_{i=1}^{k} X_i^{-1}(A_i) \,:\, A_i \in \mathscr{B}(\mathbb{R}) \right\}.$$

Then $\mathscr{A}$ is a π-system and $\sigma(\mathscr{A}) = \mathscr{G}$.

Moreover, if Y is $\mathscr{G}$-measurable, then $Y = F(X_1, \dots, X_k)$ for some measurable function $F : \mathbb{R}^k \to \mathbb{R}$.

Proof. The collection $\mathscr{A}$ is a π-system since if $B_1 = \bigcap_{i=1}^k X_i^{-1}(A_{1,i})$ and $B_2 = \bigcap_{i=1}^k X_i^{-1}(A_{2,i})$ are in $\mathscr{A}$ for $A_{1,i}, A_{2,i} \in \mathscr{B}(\mathbb{R})$ for $1 \leq i \leq k$, then

$$B_1 \cap B_2 = \bigcap_{i=1}^{k} X_i^{-1}(A_{1,i} \cap A_{2,i}).$$

Moreover, $\sigma(\mathscr{A}) = \mathscr{G}$ since $\mathscr{A}$ contains $X_i^{-1}(A)$ for every $A \in \mathscr{B}(\mathbb{R})$ and $1 \leq i \leq k$ and therefore $\sigma(\mathscr{A}) \supset \mathscr{G}$. Also $\mathscr{A} \subset \mathscr{G}$ and so $\sigma(\mathscr{A}) \subset \sigma(\mathscr{G}) = \mathscr{G}$, showing that $\sigma(\mathscr{A}) = \mathscr{G}$. We observe that we have not used the $\pi - \lambda$-systems lemma here.

Consider $\mathscr{H}$ to be the class of bounded function of the form $F(X_1, \dots, X_k)$ for some measurable $F : \mathbb{R}^k \to \mathbb{R}$. The class $\mathscr{H}$ satisfies the assumption of the Monotone Class Theorem since if $Y_n = F_n(X_1, \dots, X_k)$ with $Y_n \uparrow Y$ for Y a bounded function, then we can take $F = \limsup_{n\geq 1} F_n$ and check that $Y = F(X_1, \dots, X_k)$. We furthermore define $\mathscr{C} = \{1_C \,:\, C \in \mathscr{A}\}$. Then $\mathscr{C} \subset \mathscr{H}$ since if $C = \bigcap_{i=1}^k X_i^{-1}(A_i)$ then $1_C = \prod_{i=1}^k 1_{A_i} \circ X_i$. As $\mathscr{A}$ is a π-system and $1_A \cdot 1_B = 1_{A\cap B}$, the set $\mathscr{C}$ is stable under pointwise multiplication and since $\sigma(\mathscr{A}) = \mathscr{G}$ it holds that $\sigma(\mathscr{C}) = \mathscr{G}$. Therefore by the Monotone Class Theorem, $\mathscr{H}$ contains all bounded $\mathscr{G} = \sigma(\mathscr{C})$ measurable functions. On the other hand, since every function in $\mathscr{H}$ is $\mathscr{G}$-measurable, the class $\mathscr{H}$ is exactly the set of $\mathscr{G}$-measurable bounded functions.

It remains to deal with the case of unbounded functions. Without loss of generality by writing $Y = \max(Y, 0) - \max(-Y, 0)$, we can assume without loss of generality that Y is positive. Given a $\mathscr{G}$-measurable positive function $Y : \Omega \to \mathbb{R}$, we consider $Y_n = \max(Y, n)$. Thus there is a measurable function $F_n : \mathbb{R}^k \to \mathbb{R}$ such that $Y_n = F_n(X_1, \ldots, X_k)$. We set $F = \limsup_{n \geq 1} F_n$. Then F is measurable and we claim that $Y = F(X_1, \ldots, X_k)$. Indeed if for given $x \in \Omega$ we have that $Y(x) \leq n$ for some $n \geq 1$ then $Y_\ell(x) = Y(x)$ for all $\ell \geq n$ and hence

$$Y(x) = \sup_{\ell \geq n} Y_\ell(x) = \sup_{\ell \geq n} F_\ell(X_1(x), \ldots, X_k(x)) = F(X_1, \ldots, X_n)$$

showing the claim. □

1.5. **Product Algebras and Product Measures.** Given probability spaces $(\Omega_i, \mathscr{F}_i)$ for $1 \leq i \leq k$ be measurable spaces. Consider the space $\Omega = \Omega_1 \times \ldots \times \Omega_k$. Then the product σ-algebra $\mathscr{F} = \mathscr{F}_1 \times \ldots \times \mathscr{F}_k$ is the smallest σ-algebra on Ω containing the sets

$$A_1 \times \ldots \times A_k$$

with $A_i \in \mathscr{F}_i$ for $1 \leq i \leq k$. A few comments:

(1) Consider the projections $\pi_i : \Omega \to \Omega_i$. Then $\mathscr{F}$ is the smallest σ-algebra such that the maps π_i are measurable.
(2) Warning: The notation $\mathscr{F}_1 \times \ldots \times \mathscr{F}_k$ is slightly confusing as not all sets are of the form $A_1 \times \ldots \times A_k$. Indeed, on $\mathbb{R}^2$ every open set is Borel measurable.

I also want to simplify a Lemma from the lecture notes and the exercise sheet.

Lemma 1.11. *Let $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$ be measurable spaces and consider $(\Omega, \mathscr{F}) = (\Omega_1 \times \Omega_2, \mathscr{F}_1 \times \mathscr{F}_2)$. Let $f : \Omega \to \mathbb{R}$ be a measurable function and let $\omega_1 \in \Omega_1$. Then the map*

$$\omega_2 \mapsto f(\omega_1, \omega_2)$$

is measurable.

Proof. Consider the injection

$$\iota_1 : \Omega_2 \to \Omega, \qquad \omega_2 \mapsto (\omega_1, \omega_2).$$

Then ι_1 is measurable since

$$\iota_1^{-1}(A_1 \times A_2) = \begin{cases} A_2, & \omega_1 \in A_1, \\ \emptyset & \omega_1 \notin A_1. \end{cases}$$

Thus the map in question is the composition $\iota_1 \circ f$ and therefore measurable as the composition of measurable maps is measurable. □

Given now probability spaces $(\Omega_i, \mathscr{F}_i, \mathbb{P}_i)$ for $1 \leq i \leq k$, there is a unique probability measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$ satisfying

$$\mathbb{P}(A_1 \times \ldots \times A_k) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_k).$$

1.6. **Random Variables and Distributions.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A **random variable** is a measurable map $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Each random variable X determines a probability measure μ_X on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ defined by

$$\mu_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$$

for $A \in \mathscr{B}(\mathbb{R})$. In other words, μ_X is the push forward of $\mathbb{P}$ under X. The measure μ_X is called the **distribution** or the **law** of X. If two random variables X and Y (not necessarily defined on the same probability space) have the same distribution, then we write $X \sim Y$.

So we arrive at a map

$$\begin{aligned} D_\Omega : \{\text{random variables on } (\Omega, \mathscr{F}, \mathbb{P})\} &\to \{\text{probability measures on } (\mathbb{R}, \mathscr{B}(\mathbb{R}))\}, \\ X &\mapsto \mu_X. \end{aligned}$$

This map is highly non-injective and sometimes surjective. Indeed consider the probability space $((0,1), \mathscr{B}((0,1)), \mathrm{m}_{(0,1)})$ and the random variables

$$\begin{aligned} Y_1(\omega) &= 1_{(0,0.5)}(\omega) - 1_{[0.5,1)}(\omega), \\ Y_2(\omega) &= 1_{(0,0.25)}(\omega) + 1_{[0.75,1)}(\omega) - 1_{[0.25,0.75)}(\omega). \end{aligned}$$

Then the distribution of Y_1 and Y_2 are both $(\delta_1 + \delta_{-1})/2$. Indeed, for example,

$$\mu_{Y_1}(1) = \mathrm{m}_{(0,1)}((0,0.5)) = 0.5 \quad \text{and} \quad \mu_{Y_1}(-1) = \mathrm{m}_{(0,1)}([0.5,1)) = 0.5.$$

A further thing to notice is that to each measure μ on $\mathbb{R}$, we can define the **distribution function**

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

The function F_μ is increasing, right continuous and satisfies

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1.$$

It was proved in the lecture that

$$\begin{aligned} \{\text{probability measures on } (\mathbb{R}, \mathscr{B}(\mathbb{R}))\} &\longleftrightarrow \{\text{distribution functions}\}, \\ \mu &\longleftrightarrow F_\mu. \end{aligned}$$

is a bijection. For a random variable, the **cumulative distribution function (CDF)** is defined as the distribution function of μ_X, i.e.

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

Consider now $\Phi : \mathbb{R} \to (0,1)$ to be the cumulative distribution function (CDF) of the standard normal random variable, i.e.

$$\Phi(x) = \mathbb{P}[\mathcal{N}(0,1) \le x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx.$$

Then $\Phi : \mathbb{R} \to (0,1)$ is a bijection and consider $\Psi = \Phi^{-1}$. We define the random variable X on $((0,1), \mathscr{B}((0,1)), \mathrm{m}_{(0,1)})$ as $X(\omega) = \Psi(\omega)$. We have that $X \sim \mathcal{N}(0,1)$

since

$$\begin{aligned} F_X(x) &= m_{(0,1)}(X \leq x) \\ &= m_{(0,1)}(X^{-1}(-\infty, x)) \\ &= m_{(0,1)}(\Phi(-\infty, x)) \\ &= m_{(0,1)}((\Phi(-\infty), \Phi(x)) \\ &= \Phi(x) \\ &= \mathbb{P}[\mathcal{N}(0,1) \leq x]. \end{aligned}$$

Given two random variables X and Y that are defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$ the **joint distribution** of X and Y is defined by

$$\mu_{(X,Y)}(A \times B) = \mathbb{P}[X \in A, Y \in B] = \mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)).$$

As above, the joint distribution is determined by the joint cumulative distribution function

$$F_{(X,Y)}(x, y) = \mathbb{P}[X \leq x, Y \leq y].$$

We observe that it is a well-known fact that

$$\mu_{(X,Y)} = \mu_X \times \mu_Y \quad \text{if and only if} \quad F_{(X,Y)}(x,y) = F_X(x) \cdot F_Y(y) \ \ \forall x, y \in \mathbb{R}. \tag{1.1}$$

If the latter property holds, the random variables X and Y are called **independent**, an important topic discussed later on.

Returning to our concrete examples, we now want to study the joint distributions (X, Y_1) and (X, Y_2). We could calculate the cumulative density function, but it is a bit easier to calculate the measures directly. We note that if $x < 0$ then

$$\begin{aligned} \mu_{(X,Y_1)}((-\infty, x] \times \{1\}) &= m_{(0,1)}(X \leq x, Y_1 = 1) \\ &= m_{(0,1)}(\{\omega \in (0,1) \,:\, X(\omega) \leq x, Y_1(\omega) = 1\}) \\ &= m_{(0,1)}(\{\omega \in (0,1/2) \,:\, X(\omega) \leq x\}) \\ &= \Phi(x). \end{aligned}$$

Similarly if $x \geq 0$,

$$\mu_{(X,Y_1)}((-\infty, x] \times \{-1\}) = m_{(0,1)}(\{\omega \in (1/2, 1) \,:\, X(\omega) \leq x\}) = \Phi(x) - \frac{1}{2}.$$

Moreover, to calculate the distribution of (X, Y_2) note that for $x \leq \Psi(1/4)$ or $x \geq \Psi(3/4)$,

$$\mu_{(X,Y_2)}((-\infty, x] \times \{1\}) = \begin{cases} \Phi(x) & \text{for } x \leq \Psi(1/4), \\ \Phi(x) - \Psi(3/4) + \Psi(1/4) & \text{for } x \geq \Psi(3/4). \end{cases}$$

One similarly shows that $\mu_{(X,Y_2)}$ is supported for $y = -1$ in the range $\Psi(1/4) \leq x \leq \Psi(3/4)$. Indeed, for such an x,

$$\mu_{(X,Y_2)}((-\infty, x] \times \{-1\}) = \Phi(x) - \Psi(1/4).$$

Finally we want to consider the random variable $|X|$. We note that for $x \in \mathbb{R}$

$$\begin{aligned} F_{|X|}(x) &= \mu_{|X|}((-\infty, x]) \\ &= m_{(0,1)}(\omega \in (0,1) : |X(\omega)| \leq x) \\ &= \begin{cases} 0 & \text{if } x \leq 0, \\ 2\Phi(x) - 1 & \text{if } x \geq 0. \end{cases} \end{aligned}$$

and furthermore for $y \in \mathbb{R}$,

$$F_{Y_1}(y) = \begin{cases} 0 & \text{if } y < -1, \\ 1/2 & \text{if } -1 \leq y < 1, \\ 1 & \text{if } y \geq 1. \end{cases}$$

We claim that the joint distribution of $(|X|, Y_1)$ is $\mu_{|X|} \times \mu_{Y_1}$. To show the latter, we prove $F_{|X|,Y_1}(x,y) = F_{|X|}(x) F_{Y_1}(y)$ for all $x, y \in \mathbb{R}$ which is sufficient by (1.1). For $x \geq 0$, we note that

$$\begin{aligned} F_{(|X|,Y_1)}(x,-1) &= m_{(0,1)}(\{\omega \in [1/2, 1) : |X(\omega)| \leq x\}) \\ &= \mathbb{P}[\mathcal{N}(0,1) \in [0,x)] \\ &= \Phi(x) - \frac{1}{2} = F_{|X|}(x) F_{Y_1}(-1). \end{aligned}$$

One checks the same observation in all the other relevant ranges, which implies the claim.

2. B8.1 Class 2

2.1. Borel-Cantelli Lemmas.

First we recall the Borel-Cantelli Lemmas.

Theorem 2.1. *(First Borel-Cantelli Lemma) Let $A_1, A_2, A_3, \dots$ be events in a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and assume that*

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

Then

$$\mathbb{P}(\limsup A_i) = 0,$$

where

$$\limsup_{i\to\infty} A_i = \{\omega : \omega \in A_i \text{ for infinitely many } i\} = \bigcap_{i\geq 1} \bigcup_{j \geq i} A_j.$$

To state the second Borel-Cantelli Lemma recall that a sequence of events $A_1, A_2, A_3, \dots$ in a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ are called independent if

$$\mathbb{P}[A_{i_1} \cap \ldots \cap A_{i_j}] = \mathbb{P}[A_{i_1}] \cdots \mathbb{P}[A_{i_j}]$$

for all $i_1, \dots, i_j \in \mathbb{N}$. It is important to note that it does not suffice to check pairwise independence, meaning that $\mathbb{P}[A_i \cap A_j] = \mathbb{P}[A_i]\mathbb{P}[A_j]$ for all i and j, to conclude that $(A_n)_{n\geq 1}$ are independent.

Theorem 2.2. *(Second Borel-Cantelli Lemma) Let $A_1, A_2, A_3, \dots$ be independent events in a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and assume that*

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

Then

$$\mathbb{P}(\limsup A_i) = 1.$$

First we study the records of independent uniform random variables.

Lemma 2.3. *Let $X_1, X_2, \dots$ be independent uniform $[0,1]$ random variables. Let A_n for $n \geq 1$ be event that X_n is the record among $X_1, \dots, X_n$, i.e.*

$$A_n = \{X_n > \max(X_1, \dots, X_{n-1})\}.$$

Then almost surely infinitely many records occur, i.e. A_n happens infinitely many often.

Also denote

$$D_n = \{X_n > X_{n-1} > \max(X_1, \dots, X_{n-2})\}$$

for $n \geq 2$ the event that a double record occurs at n. Then with probability one only finitely many double records occur.

Proof. Consider the sets A_n. We first want to calculate $\mathbb{P}[A_n]$. We claim that for each permutation $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$,

$$\mathbb{P}[X_{\sigma(1)} > \ldots > X_{\sigma(n)}] = \frac{1}{n!}.$$

This follows as $\mathbb{P}[X_i = X_j] = 0$ for $i \neq j$ and as the probability density function of $(X_1, \ldots, X_n)$ is invariant by permutation since the variables are independent (see Lemma 2.8 for more details). Therefore it follows that

$$\mathbb{P}[A_n] = \frac{\#\{\sigma : \sigma(n) = n\}}{n!} = \frac{(n-1)!}{n!} = \frac{1}{n},$$

and

$$\mathbb{P}[D_n] = \frac{\#\{\sigma : \sigma(n) = n, \sigma(n-1) = n-1\}}{n!} = \frac{1}{n(n-1)}.$$

Thus it follows that $\sum_{n\geq 2} \mathbb{P}[D_n] < \infty$ and hence by the first Borel-Cantelli Lemma, with probability one, only finitely many double records occur.

On the other hand $\sum_{n\geq 1} \mathbb{P}[A_n] = \infty$. Therefore to show that A_n occurs infinitely often almost surely, by the second Borel-Cantelli Lemma, it suffices to show that the sets A_n are independent. We give two proofs of this. First, we give the following intuitive argument. Notice that given $J \subset \{1, \ldots, n-1\}$, consider the set $A = \bigcap_{j\in J} A_j$. Since a record happening at n, has no influence on records happening before, it holds that

$$\mathbb{P}[A|A_n] = \frac{\mathbb{P}[A \cap A_n]}{\mathbb{P}[A_n]} = \mathbb{P}[A].$$

Therefore the claim follows by induction on n.

For the second proof we only treat the case $A_n \cap A_m$. The general case is similar and left to the reader. We need to show that

$$\mathbb{P}[A_n \cap A_m] = \frac{1}{n \cdot m}$$

for any $n \neq m$. Without loss of generality we assume that $n > m$. By the above observations, the claim reduces to counting the permutations $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that $\sigma(n) = n$ and $\sigma(m) > \sigma(1), \ldots, \sigma(m-1)$. The condition $\sigma(n) = n$ reduces to counting permutations $\sigma : \{1, \ldots, n-1\} \to \{1, \ldots, n-1\}$ satisfying $\sigma(m) > \sigma(1), \ldots, \sigma(m-1)$. We can first choose the elements $\sigma(n-1), \ldots, \sigma(m+1)$ freely of which we have in total $(n-1)\cdots(m+1)$ many choices. Then the element $\sigma(m)$ is determined and the remaining elements $\sigma(m-1), \ldots, \sigma(1)$ can be chosen freely resulting in $(m-1)!$ more choices. Therefore there are indeed $(n-1)!/m$ many such permutations and hence

$$\mathbb{P}[A_n \cap A_m] = \frac{1}{n!} \cdot \frac{(n-1)!}{m} = \frac{1}{n \cdot m}.$$

$\square$

Next we show for the law of large numbers to hold, the variables need to have the same variance.

Lemma 2.4. *Let $(X_n)_{n\geq 2}$ be a sequence of independent random variables such that*

$$\mathbb{P}[X_n = n] = \mathbb{P}[X_n = -n] = \frac{1}{2n \log n} \quad \textit{and} \quad \mathbb{P}[X_n = 0] = 1 - \frac{1}{n \log n}.$$

Denote $S_n = \sum_{i=2}^n X_i$. Then $\frac{S_n}{n}$ converges to 0 in probability but not almost surely.

Proof. By Chebyschev's inequality, since S_n has mean zero,

$$\mathbb{P}[|S_n| \geq \varepsilon n] \leq \frac{1}{n^2 \varepsilon^2} \mathrm{Var}(S_n).$$

Moreover, by independence

$$\begin{aligned}\mathrm{Var}(S_n) &= \mathbb{E}\left[\sum_{i,j\geq 2}^{n} X_i X_j\right] = \sum_{i,j\geq 2}^{n} \mathbb{E}[X_i X_j] \\ &= \sum_{i=2}^{n} \mathbb{E}[X_i^2] = \sum_{i=2}^{n} \frac{i^2}{i\log i} = \sum_{i=2}^{n} \frac{i}{\log i} \leq \frac{n^2}{\log n}\end{aligned}$$

since $\frac{x}{\log x}$ has derivative $\frac{\log x - 1}{(\log x)^2}$ and is therefore increasing for $x > e$.

Thus,

$$\mathbb{P}[|S_n| \geq \varepsilon n] \leq \frac{1}{n^2\varepsilon^2}\mathrm{Var}(S_n) \leq \frac{1}{\varepsilon^2 \log n} \to 0$$

and so $\frac{S_n}{n}$ converges to zero in probability.

We now show that $\frac{S_n}{n}$ does not converge to 0 almost surely. Indeed notice that

$$\begin{aligned}\frac{S_{n+1}}{(n+1)} - \frac{S_n}{n} &= \frac{S_{n+1} - S_n + S_n}{(n+1)} - \frac{S_n}{n} \\ &= \frac{S_{n+1} - S_n}{(n+1)} - \frac{S_n}{n(n+1)} = \frac{X_{n+1}}{(n+1)} - \frac{S_n}{n(n+1)}.\end{aligned}$$

Thus if $\frac{S_n}{n} \to 0$, then it follows that $\frac{X_n}{n} \to 0$. However this is not the case as almost surely, by the second Borel-Cantelli Lemma, X_n is $\pm n$ infinitely many often. Indeed denote $A_n = \{X_n = \pm n\}$ and note that $(A_n)_{n\geq 2}$ are independent as the $(X_n)_{n\geq 2}$ are. Then

$$\sum_{n\geq 2} \mathbb{P}[A_n] = \sum_{n\geq 2} \frac{1}{n\log n} = \infty$$

since by the integral criterion it suffices to show

$$\int_2^\infty \frac{1}{x\log x}\,dx = \int_{\log 2}^\infty \frac{1}{y}\,dy = \infty$$

by substituting $y = \log x$. □

2.2. **Standard Random Walk on $\mathbb{Z}$.** We now give an extensive deduction that the standard simple random walk on $\mathbb{Z}$ visits every point infinitely many often. We first show that there exists a sequence of independent coinflips. This follows abstractly from the following lemma.

Lemma 2.5. *Let $\mu_1, \mu_2, \ldots$ be a sequence of probability measures on $\mathbb{R}$. Then there exists a sequence of independent random variables $X_1, X_2, \ldots$ such that X_i is distributed as μ_i for all $i \geq 1$.*

Proof. Consider $\Omega = \mathbb{R}^{\otimes\mathbb{N}}$ endowed with the product σ-algebra and the product measure $\mu = \bigotimes_{i\geq 1}\mu_i$. Setting X_i to be the i-th coordinate map, the proof is concluded. □

We can also give the following explicit construction.

Lemma 2.6. *On $([0,1], \mathscr{B}([0,1]), m_{[0,1]})$ denote*

$$A_n = \{\omega \in [0,1] \,:\, \lfloor 2^n\omega\rfloor\} \text{ is odd}\}.$$

Then the sequence of random variables $X_n = 1_{A_n}$ is independent identically distributed with

$$\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = 0] = \frac{1}{2}.$$

Proof. Note that

$$A_n = \left[\frac{1}{2^n}, \frac{2}{2^n}\right) \cup \left[\frac{3}{2^n}, \frac{4}{2^n}\right) \cup \ldots \cup \left[\frac{2^n - 1}{2^n}, \frac{2^n}{2^n}\right).$$

So these are 2^{n-1} many intervals of length 2^{-n}. This implies that $\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = 0] = 1/2$. To show that the sequence of random variables is independent, it suffices to show that the sets A_i are independent.

For simplicity, we first consider the case of $A_n \cap A_m$. We note that

$$A_n \cap A_m = \bigcup_{\substack{1 \le k \le 2^n \text{ odd} \\ k \text{ odd mod } 2^{n-m+1}}} \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right),$$

which easily implies independence. □

Let now be $(X_n)_{n \ge 1}$ be a sequence is independent identically distributed real random variables such that

$$\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = -1] = 1/2.$$

Let $S_0 = 0$ and, for all $n \ge 1$ denote $S_n = \sum_{k=1}^n X_k$.

For $x \in \mathbb{Z}$ let

$$A_x = \{S_n = x \text{ for infinitely many } n\},$$

$$B_- = \{\liminf_{n\to\infty} S_n = -\infty\} \quad \text{and} \quad B_+ = \{\limsup_{n\to\infty} S_n = \infty\}.$$

Let $\mathscr{T}_k = \sigma(X_{k+1}, X_{k+2}, \ldots)$ and $\mathscr{T} = \bigcap_{k \ge 1} \mathscr{T}_k$, which is a σ-algebra since it is the intersection of σ-algebras. We note that $B_\pm \in \mathscr{T}_k$ for all $k \ge 1$ since being in $B_\pm$ only depends on the values of $X_{k+1}, X_{k+2}, \ldots$. More formally,

$$B_\pm = \{\liminf_{n\to\infty} S_n = \pm\infty\} = \{\liminf_{n\to\infty} S_{k+n} = \pm\infty\}$$

and the map

$$\Phi_k : \Omega \to \overline{\mathbb{R}}, \qquad \omega \mapsto \liminf_{n\to\infty} S_{k+n}$$

is $\sigma(X_{k+1}, \ldots)$ measurable since each of S_{k+n} is. Thus $B_\pm = \Phi_k^{-1}(\pm 1)$ and the claim follows.

By Kolmogorov's 0-1 law, we conclude $\mathbb{P}[B_\pm] \in \{0, 1\}$. By symmetry it follows that $\mathbb{P}[B_+] = \mathbb{P}[B_-]$. More formally, $\{\liminf_{n\to\infty}(-S_n) = -\infty\} = \{\liminf_{n\to\infty} S_n = \infty\}$. Since (X_n) and $(-X_n)$ have the same distribution, so do $\liminf_{n\to\infty}(-S_n)$ and $\liminf_{n\to\infty} S_n$.

The event that $S_{n+k} - S_n = k$ is equivalent to $X_{n+1} = \ldots = X_{n+k} = 1$, which has probability $\frac{1}{2^k}$. Denote by $A_n = \{X_{n \cdot k+1} = \ldots = X_{n \cdot k+k} = 1\}$. Then the sets A_n are independent (since the X_i are independent) and all have probability $\frac{1}{2^k}$. Thus it follows by the second Borel-Cantelli Lemma that A_n happens infinitely many often and therefore for all $k \ge 1$,

$$\limsup_{n\to\infty} (S_{n+k} - S_n) = k$$

almost surely. This implies that $\mathbb{P}[B_-^c \cap B_+^c] = 0$ and hence $\mathbb{P}[B_+] = \mathbb{P}[B_-] = 1$. Therefore for all $x \in \mathbb{Z}$, $\mathbb{P}[A_x] = 1$.

Now we suppose that $(X_n)_{n\geq 1}$ are i.i.d. random variables but with $\mathbb{P}[X_n] = p$ and $\mathbb{P}[X_n = -1] = 1 - p$ for some $p \neq 1/2$. Then we claim that $\mathbb{P}[A_0] = 0$.

First note that $\{S_{2n+1} = 0\} = \emptyset$ for all $n \geq 0$ since one needs an even number of steps to return to 0. Observe further that

$$\mathbb{P}[S_{2n} = 0] = \binom{2n}{n} p^n (1-p)^n.$$

Note that since $4^n = (1+1)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k}$ and therefore $\binom{2n}{n} \leq 4^n$. So $\mathbb{P}[S_{2n} = 0] \leq (4p(1-p))^n$ and as $4p(1-p) < 1$ if $p \neq 1/2$ the quantity $\mathbb{P}[S_{2n} = 0]$ decays exponentially fast, showing that $\sum_{n=0}^{\infty} \mathbb{P}[S_{2n} = 0] < \infty$. Thus the claim follows by the first Borel-Cantelli Lemma.

2.3. **Hölder's Inequality.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable and denote for $p \in [1, \infty)$ by

$$||X||_p = \mathbb{E}[|X|^p]^{\frac{1}{p}} = \left(\int |X(\omega)|^p \, d\mathbb{P}(\omega) \right)^{\frac{1}{p}}.$$

Recall that the Cauchy-Schwarz inequality states that for random variables X and Y we have

$$||XY||_1 \leq ||X||_2 ||Y||_2.$$

Hölder's inequality generalises the Cauchy-Schwarz inequality and states that for $p, q \in [1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$ we have

$$||XY||_1 \leq ||X||_p ||Y||_q.$$

Using the Hölder inequality, we can prove the following lemma.

Lemma 2.7. *Let X and Y be two positive random variables such that for all $x > 0$,*

$$x\mathbb{P}[X \geq x] \leq \mathbb{E}[Y 1_{X \geq x}].$$

Then it holds that $||X||_p \leq q||Y||_p$ for any $p > 1$ and $q = p/(p-1)$.

Proof. We assume that $Y \in L^p$ as otherwise the claim is obvious. Notice that since X is positive, using Fubini since X and Y are positive,

$$\begin{aligned}
\mathbb{E}[X^p] &= \mathbb{E}\left[\int_{(0,X]} px^{p-1} \, dx \right] \\
&= \int_{\Omega} \int_0^{\infty} 1_{\{x \leq X(\omega)\}} px^{p-1} \, dx \, d\mathbb{P}(\omega) \\
&= \int_0^{\infty} \mathbb{P}[X \geq x] px^{p-1} \, dx \\
&\leq \int_0^{\infty} \mathbb{E}[Y 1_{X \geq x}] px^{p-2} \, dx \\
&= \mathbb{E}\left[Y \int_0^X px^{p-2} \, dx \right] \\
&= \frac{p}{p-1} \mathbb{E}[X^{p-1} Y] = q\mathbb{E}[X^{p-1} Y].
\end{aligned}$$

If $||X||_p < \infty$, then note that $||X^{p-1}||_q = \mathbb{E}[X^p]^{\frac{p-1}{p}} = ||X||_p^{p-1}$ is also finite and it follows by Hölder's inequality since $\frac{1}{q} + \frac{1}{p} = \frac{(p-1)}{p} + \frac{1}{p} = 1$ that

$$\mathbb{E}[X^p] \leq q\mathbb{E}[X^{p-1}Y] \leq q||X^{p-1}||_q||Y||_p.$$

The claim of the lemma follows in this case since

$$\frac{\mathbb{E}[X^p]}{||X^{p-1}||_q} = \frac{\mathbb{E}[X^p]}{\mathbb{E}[X^{q(p-1)}]^{\frac{1}{q}}} = \mathbb{E}[X^p]^{1-\frac{1}{q}} = ||X||_p.$$

For the general case, we use the standard trick of truncation. Indeed, consider $X_n = X \wedge n$. Notice that the assumed inequality also holds for X_n and Y. Then the claim follows as by monotone convergence, $\lim_{n\to\infty} ||X_n||_p = ||X||_p$. $\square$

2.4. **Interchangeability of Random Variables.** A measure μ on $\mathbb{R}$ is called absolutely continuous if there is a function $p \in L^1(\mathbb{R})$ (with respect to the Lebesgue measure) such that

$$\mu(A) = \int_A p(x)\, dm_{\mathbb{R}}(x).$$

The function p is called the density function of μ. By the Radon-Nikodym theorem this is equivalent to $\mu(N) = 0$ for every Lebesgue null set N. We say that a random variable is absolutely continuous if its distribution is. For example, uniform random variables and normal random variables are absolutely continuous.

Lemma 2.8. *Let $X_1, X_2, \ldots$ be independent identically distributed absolutely continuous random variables. Then for any permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ it holds that*

$$\mathbb{P}[X_{\sigma(1)} > \ldots > X_{\sigma(n)}] = \frac{1}{n!}$$

Proof. Let p be the density function of the X_i. Then by independence

$$p_n(x_1, \ldots, x_n) = \prod_{i=1}^n p(x_i)$$

is the density function of $X = (X_1, \ldots, X_n)$, which is permutation invariant.

Notice that the set $\{x \in \mathbb{R}^d : x_i = x_j\}$ with $i \neq j$ has Lebesgue measure zero. Therefore $\mathbb{P}[X_i = X_j] = 0$ for all $i \neq j$ and it follows that

$$\mathbb{P}\left[\bigcup_\sigma \{X_{\sigma(1)} > \ldots > X_{\sigma(n)}\}\right] = \sum_\sigma \mathbb{P}\left[X_{\sigma(1)} > \ldots > X_{\sigma(n)}\right] = 1.$$

Moreover,

$$\mathbb{P}\left[X_{\sigma(1)} > \ldots > X_{\sigma(n)}\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{x_{\sigma_1}} \cdots \int_{-\infty}^{x_{\sigma(n-1)}} p_n(x_1, \ldots, x_n)\, dx_{\sigma(1)} \cdots dx_{\sigma(n)}.$$

As p_n is permutation invariant and since the map $\Phi_\sigma : \mathbb{R}^d \to \mathbb{R}^d, x \mapsto \sigma(x)$ preserves the Lebesgue measure, the latter integral does not depend on σ. This concludes the proof. $\square$

Remark 2.9. In this proof we only used that p_n is permutation invariant and that $\mathbb{P}[X_i = X_j] = 0$. The former follows since the X_i are independent, while in the latter we used that X_i is absolutely continuous.

The following question remains: If the X_i are not absolutely continuous, is it still the case that $\mathbb{P}[X_{\sigma(1)} > \ldots > X_{\sigma(n)}]$ does not depend on σ? Indeed, does it hold that

$$\mathbb{P}[X_{\sigma(1)} > \ldots > X_{\sigma(n)}] = \frac{\mathbb{P}[(X_1, \ldots, X_n) \text{ are distinct}]}{n!}.$$

We leave it as an exercise to the reader to prove this.

3. B8.1 Class 3

3.1. Conditional Expectation.

Definition 3.1. *Let X be an integrable random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{A} \subset \mathscr{F}$ be a sub-σ-algebra. Then the conditional expectation of $\mathbb{E}[X|\mathscr{A}] \in L^1(X, \mathscr{A}, \mathbb{P})$ is uniquely characterized by*

$$\mathbb{E}\left[\mathbb{E}[X|\mathscr{A}]1_A\right] = \mathbb{E}[X1_A]$$

for all $A \in \mathscr{A}$.

The conditional expectation is the expectation (or average) of X with the information from $\mathscr{F}$. We first discuss the following important example.

3.2. Properties of conditional expectation.

Lemma 3.2. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{P} = \{A_1, A_2, \dots\}$ be a countable partition of Ω with $\mathbb{P}[A_i] > 0$ for all $i \geq 1$. Then for a random variable X and $A \in \mathscr{P}$,*

$$\mathbb{E}[X|\mathscr{A}] = \sum_{A \in \mathscr{P}} \frac{\mathbb{E}[X1_A]}{\mathbb{P}[A]} 1_A.$$

Proof. We note that for any $A \in \mathscr{A}$ it holds that

$$\mathbb{E}[\mathbb{E}[X|\mathscr{A}]1_A] = \mathbb{E}[X1_A].$$

Moreover, since $\mathbb{E}[X|\mathscr{A}]$ is $\mathscr{A}$-measurable, it must be constant on the sets $\mathscr{A}$. Therefore for each A there is c_A such that $c_A = \mathbb{E}[X|\mathscr{A}]1_A$ and hence

$$c_A \cdot \mathbb{P}[A] = \mathbb{E}[\mathbb{E}[X|\mathscr{A}]1_A] = \mathbb{E}[X1_A],$$

implying the claim. □

We can deduce the following corollary.

Corollary 3.3. *Let X and Y be discrete random variables. Then for any $y \in \mathrm{Im}(Y)$ and $\omega \in \Omega$,*

$$\mathbb{E}[X|Y] = \sum_{y \in \mathrm{Im}(Y)} \frac{\mathbb{E}[X1_{\{Y=y\}}]}{\mathbb{P}[Y=y]} 1_{\{Y=y\}}.$$

In particular, for $\omega \in \{Y = y\}$,

$$\mathbb{E}[X|Y](\omega) = \frac{\mathbb{E}[X1_{\{Y=y\}}]}{\mathbb{P}[Y=y]} = \sum_{x \in \mathrm{Im}(x)} \frac{x \cdot \mathbb{P}[X = x, Y = y]}{\mathbb{P}[Y=y]}.$$

We note that one often uses the notation $\mathbb{E}[X|Y = y]$, which means the conditional expectation $\mathbb{E}[X|Y]$ evaluated on the set $\{Y = y\}$.

The following properties of conditional expectation were discussed in the lecture.

Lemma 3.4. *Let X and Y be an integrable random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{A} \subset \mathscr{F}$ be a σ-algebra. The following properties hold:*

(i) $\mathbb{E}[\mathbb{E}[X|\mathscr{A}]] = \mathbb{E}[X]$
(ii) *The conditional expectation is linear.*
(iii) $\mathbb{E}[c \cdot 1_\Omega|\mathscr{A}] = c \cdot 1_\Omega$.
(iv) *If X is $\mathscr{A}$-measurable, then $\mathbb{E}[X|\mathscr{A}] = X$.*
(v) *If X is independent of $\mathscr{A}$, then $\mathbb{E}[X|\mathscr{A}] = \mathbb{E}[X]$.*
(vi) *If Y is $\mathscr{A}$-measurable, $\mathbb{E}[XY|\mathscr{A}] = \mathbb{E}[X|\mathscr{A}]Y$.*

(vii) If $X \leq Y$ a.s., then $\mathbb{E}[X|\mathscr{A}] \leq \mathbb{E}[Y|\mathscr{A}]$ a.s.

Lemma 3.5. *(Conditional Markov Inequality) Let X be an integrable random variable, $\mathscr{G} \subset \mathscr{F}$ be a σ-algebra and Z a strictly positive $\mathscr{G}$-measurable random variable. Then almost surely,*

$$\mathbb{E}[1_{|X|\geq Z}|\mathscr{G}] \leq \frac{1}{Z}\mathbb{E}[|X|\,|\mathscr{G}].$$

Proof. Note that $Z \cdot 1_{\{|X|\geq Z\}} \leq |X|$ and hence using Lemma 3.4 (v) and (vi),

$$Z \cdot \mathbb{E}[1_{\{|X|\geq Z\}}|\mathscr{G}] = \mathbb{E}[Z \cdot 1_{\{|X|\geq Z\}}|\mathscr{G}] \leq \mathbb{E}[|X|\,|\mathscr{G}],$$

implying the claim by dividing by Z and using that Z is positive. □

3.3. **Further explicit examples.**

Lemma 3.6. *Let X, Y be independent random variables. Then the following properties hold:*

(i) $\mathbb{E}[X|X, Y] = X$.
(ii) $\mathbb{E}[h(X, Y)|X + Y, X - Y] = h(X, Y)$.
(iii) If X and Y moreover have the same distribution,

$$\mathbb{E}[X|X + Y] = \mathbb{E}[Y|X + Y] = \frac{1}{2}(X + Y).$$

Proof. (i) follows since X is $\sigma(X, Y)$-measurable. (ii) follows since $(X + Y) + (X - Y) = 2X$ and so $\sigma(X + Y, X - Y) = \sigma(X, Y)$ and hence $\mathbb{E}[h(X, Y)|X + Y, X - Y] = \mathbb{E}[h(X, Y)|X, Y] = h(X, Y)$ as $h(X, Y)$ is $\sigma(X, Y)$-measurable.

Finally we prove (iii). We first give a heuristic deduction and then give a rigorous proof that the claim holds. We recall from Theorem 1.27 of the notes there are Borel measurable functions $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[X|X + Y] = f_1(X + Y)$ and $\mathbb{E}[Y|X + Y] = f_2(X + Y)$. By symmetry $\mathbb{E}[X|X + Y]$ and $\mathbb{E}[Y|X + Y]$ must have the same distribution. Therefore (heuristically) we have $\mathbb{E}[X|X+Y] = \mathbb{E}[Y|X+Y]$ almost surely and hence since

$$\mathbb{E}[X|X + Y] + \mathbb{E}[Y|X + Y] = \mathbb{E}[X + Y|X + Y] = X + Y$$

the claim follows.

We now give a rigorous argument that indeed $\mathbb{E}[X|X + Y] = \frac{1}{2}(X + Y)$. Since the π-system $\{\{X + Y \leq c\} : c \in \mathbb{R}\}$ generates $\sigma(X + Y)$, it suffices to show that

$$\mathbb{E}[X1_{\{X+Y\leq c\}}] = \mathbb{E}[Y1_{\{X+Y\leq c\}}]. \tag{3.1}$$

Indeed this shows that $\mathbb{E}[X1_{\{X+Y\leq c\}}] = \mathbb{E}[\frac{X+Y}{2}1_{\{X+Y\leq c\}}]$, which implies the claim. To show (3.1), we apply Lemma 3.7 below to the function $f(x, y) = y1_{\{x+y\leq c\}}$. □

Lemma 3.7. *Let X and Y be independent identically distributed random variables and let $f : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function such that $\omega \mapsto f(Y(\omega), X(\omega))$ and $\omega \mapsto f(X(\omega), Y(\omega))$ are integrable random variables. Then*

$$\mathbb{E}[f(X, Y)] = \mathbb{E}[f(Y, X)].$$

Proof. Consider $f = 1_{A\times B}$ for A, B measurable sets in $\mathbb{R}$. Then since X and Y are independent and have the same distribution it follows that

$$\begin{aligned}\mathbb{E}[f(X, Y)] &= \mathbb{P}[X \in A, Y \in B] = \mu_X(A)\mu_Y(B)\\ &= \mu_Y(A)\mu_X(B) = \mathbb{P}[Y \in A, X \in B] = \mathbb{E}[f(Y, X)].\end{aligned}$$

By the π-λ-lemma, the claim follows for all characteristic functions $f = 1_C$ with C a measurable set in $\mathbb{R}^2$. Moreover, by linearity, the claim holds for all positive simple functions and hence by taking pointwise limits for all positive measurable functions. Finally, writing $f = f^+ - f^-$ the claim follows. $\square$

3.4. **Independence and conditional expectation.**

Lemma 3.8. *Let X and Y be bounded random variables on $(\Omega, \mathscr{F}, \mathbb{P})$. Then each of the following statements implies the next:*

(i) X *and* Y *are independent.*
(ii) $\mathbb{E}[X|Y] = \mathbb{E}[X]$
(iii) $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

All of the other implications fail in general.

Proof. That (i) implies (ii) is implied by Lemma 3.4 (iii). To show that (iii) follows from (ii), we calculate

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

where we have used (v) from Lemma 3.4.

To give counterexamples for the converse directions, we consider the probability space Ω that has equal probability on three events. To give a counterexample of (iii) implying (ii), consider (X, Y) mapping to $(0, 1), (1, 0)$ and $(0, -1)$. Then $\mathbb{E}[XY] = 0$ and $\mathbb{E}[Y] = 0$ so (iii) holds but (ii) does not hold. Finally to give a counterexample to (ii) implying (i) consider (X, Y) mapping to $(1, 1), (-1, 1)$ and $(0, 0)$. Then it is easy to check that (ii) holds, yet (i) doesn't as

$$\mathbb{P}[X = 0, Y = 0] = \frac{1}{3} \neq \frac{1}{9} = \mathbb{P}[X = 0]\mathbb{P}[Y = 0].$$

To give another example that (iii) does not imply (i) consider a Gaussian X with mean zero and variance 1 and let Z be a coin flip (i.e. $\mathbb{P}[Z = 1] = \mathbb{P}[Z = -1] = 0$) independent of X. Then consider $Y = XZ$. Then note that by independence $\mathbb{E}[Y] = \mathbb{E}[X]\mathbb{E}[Z] = 0$ and $\mathbb{E}[XY] = \mathbb{E}[X^2 Z] = \mathbb{E}[X^2]\mathbb{E}[Z] = 0$. On the other hand, X and Y are not independent since Y is also distributed like a standard Gaussian and

$$\mathbb{P}[X > 1, Y > 1] = \mathbb{P}[X > 1, Z = 1] = \frac{1}{2}\mathbb{P}[X > 1] \neq \mathbb{P}[X > 1]^2 = \mathbb{P}[X > 1, Y > 1].$$

$\square$

Lemma 3.9. *Let X and Y be integrable random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ such that*

$$\mathbb{E}[X|Y] = Y \quad a.s. \quad \text{and} \quad \mathbb{E}[Y|X] = X \quad a.s.$$

Then $\mathbb{P}[X = Y] = 1$.

We first observe that if X and Y are in L^2, then by Cauchy-Schwarz XY is integrable and it holds that

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY|X]] = \mathbb{E}[X^2]$$

and by symmetry $\mathbb{E}[XY] = \mathbb{E}[Y^2]$. Thus it follows that

$$\mathbb{E}[(X - Y)^2] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] - 2\mathbb{E}[XY] = 0,$$

which implies the claim. For general X and Y we offer the following argument.

Proof. First note that

$$\mathbb{E}[Xf(Y)] = \mathbb{E}[\mathbb{E}[Xf(Y)|Y]] = \mathbb{E}[Yf(Y)]$$

and likewise $\mathbb{E}[Yf(X)] = \mathbb{E}[Xf(X)]$. Now let $f(x) = 1_{x \geq c}$, which gives

$$\mathbb{E}[(X-Y)1_{Y\geq c}] = 0 = \mathbb{E}[(X-Y)1_{X\geq c}].$$

Writing out the first term, we have

$$0 = \mathbb{E}[(X-Y)1_{Y\geq c}] = \mathbb{E}[(X-Y)1_{Y\geq c>X}] + \mathbb{E}[(X-Y)1_{Y\geq c, X\geq c}].$$

Note that the first term is non-positive the second one has to be non-negative. But we also have

$$0 = \mathbb{E}[(X-Y)1_{X\geq c}] = \mathbb{E}[(X-Y)1_{X\geq c>Y}] + \mathbb{E}[(X-Y)1_{Y\geq c, X\geq c}]$$

and now we conclude that the second term on the right has to be non-positive. This means that

$$\mathbb{E}[(X-Y)1_{Y\geq c, X\geq c}] = 0$$

and hence also $\mathbb{E}[(X-Y)1_{Y\geq c>X}]$ so that in particular $\mathbb{P}(Y \geq c > X) = 0$ for any $c \in \mathbb{R}$ and hence also

$$\mathbb{P}(Y > X) = \mathbb{P}\left(\bigcup_{c\in\mathbb{R}} \{Y \geq c > X\}\right) = 0.$$

We conclude, by symmetry, that $Y = X$ a.s. □

Lemma 3.10. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and consider three σ-algebras $\mathscr{G}_1, \mathscr{G}_2, \mathscr{G}_3 \subset \mathscr{F}$. Assume that $\sigma(\mathscr{G}_1, \mathscr{G}_3)$ is independent from $\mathscr{G}_2$ and let X be a $\mathscr{G}_3$-measurable random variable. Then*

$$\mathbb{E}[X|\sigma(\mathscr{G}_1, \mathscr{G}_2)] = \mathbb{E}[X|\mathscr{G}_1] \tag{3.2}$$

Proof. We need to show that if $A \in \sigma(\mathscr{G}_1, \mathscr{G}_2)$, then

$$\mathbb{E}[\mathbb{E}[X|\sigma(\mathscr{G}_1, \mathscr{G}_2)]1_A] = \mathbb{E}[X1_A] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]1_A].$$

Consider the π-system, $\mathscr{A} = \{A_1 \cap A_2 : A_1 \in \mathscr{G}_1 \text{ and } \mathscr{G}_2\}$. Since the collection of sets that satisfy the above are a λ-system, it suffices to check the claim on $\mathscr{A}$ (by the π-λ lemma). So consider $A_1 \in \mathscr{G}_1$ and $A_2 \in \mathscr{G}_2$. Then

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]1_{A_1\cap A_2}] &= \mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]1_{A_1}1_{A_2}] \\
&= \mathbb{E}[\mathbb{E}[X1_{A_1}|\mathscr{G}_1]1_{A_2}] \\
&= \mathbb{E}[\mathbb{E}[X1_{A_1}|\mathscr{G}_1]] \cdot \mathbb{E}[1_{A_2}] \\
&= \mathbb{E}[X1_{A_1}] \cdot \mathbb{E}[1_{A_2}] \\
&= \mathbb{E}[X1_{A_1}1_{A_2}] \\
&= \mathbb{E}[X1_{A_1\cap A_2}] \\
&= \mathbb{E}[\mathbb{E}[X|\sigma(\mathscr{G}_1, \mathscr{G}_2)]1_{A_1\cap A_2}]
\end{aligned}$$

□

We now want to give an example showing that assuming that $\mathscr{G}_2$ and $\mathscr{G}_3$ are independent is not sufficient to conclude the claim. To see this, consider two independent random variables ξ and η with exponential distribution with parameter 1, i.e. their cumulative distribution function is

$$\mathbb{P}[\xi \leq x] = \begin{cases} 1 - e^{-x} & \text{if } x \geq 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

Let $X_1 = \xi$, $X_2 = \frac{\xi}{\xi+\eta}$ and $X_3 = \xi + \eta$.

Note that

$$\mathbb{E}[X_3|X_1] = \mathbb{E}[\xi + \eta|\xi] = \mathbb{E}[\eta] + \xi = 1 + \xi$$

whereas

$$\mathbb{E}[X_3|\sigma(X_1, X_2)] = \mathbb{E}[X_3|\sigma(\xi, \eta)] = X_3 = \xi + \eta.$$

It remains to show that X_2 and X_3 are independent. Indeed to show this consider the map

$$f(x, y) = \left(\frac{x}{x+y}, x + y\right)$$

so that $(X_2, X_3) = f(\xi, \eta)$. The map $(x, y) \mapsto f(x, y) = (u, v)$ takes $(0, \infty)^2 \to (0, 1) \times (0, \infty)$. The inverse is given by $x = uv$ and $y = v(1 - u)$ and therefore the Jacobian is $J(u, v) = v$. Therefore for all $(u, v) \in (0, 1) \times (0, \infty)$ we have

$$f_{X_2,X_3}(u, v) = f_{\eta,\xi}(x, y)|J(u, v)| = e^{-(x(u,v)+y(u,v))}|J(u, v)| = ve^{-v}.$$

So the joint density factories and it follows that $X_2 \sim U[0, 1]$ and $X_3 \sim \Gamma(2, 3)$ independently of each other.

3.5. **Stopping Times.**

Definition 3.11. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $(\mathscr{F}_n)_{n\geq 1}$ be a filtration. A stopping time is a map $\tau : \Omega \to \mathbb{N}$ such that $\{\tau = n\} \in \mathscr{F}_n$.*

We remark that in the definition of a stopping time we could also equivalently require that $\{\tau \leq n\} \in \mathscr{F}_n$.

Recall that for a stopping time τ we define

$$\mathscr{F}_\tau = \{A \in \mathscr{F}_\infty \,:\, A \cap \{\tau = n\} \in \mathscr{F}_n \text{ for all } n \geq 0\}.$$

Lemma 3.12. *For a stopping time τ, the collection of sets $\mathscr{F}_\tau$ is a σ-algebra.*

Proof. It is clear that $\mathscr{F}_\tau$ contains $\emptyset$ and Ω. If $A \in \mathscr{F}_\tau$, then

$$A \cap \{\tau \leq n\} = \bigcup_{k\leq n} A \cap \{\tau = k\} \in \mathscr{F}_n$$

and therefore $A^c \cup \{\tau > n\} = (A \cap \{\tau \leq n\})^c \in \mathscr{F}_n$. Let $B_1 = A^c \cup \{\tau > n\}$ and $B_2 = A \cup \{\tau > n\}$, Then $B_1 \cup B_2 = \Omega$ and $B_1 \cap B_2 = \{\tau > n\}$, so both of these sets are in $\mathscr{F}_n$. Since B_1 is also in $\mathscr{F}_\tau$, it follows that $B_2 = \Omega\backslash(B_1\backslash(B_1 \cap B_2)) \in \mathscr{F}_n$. Therefore $A^c \in \mathscr{F}_n$ as $A^c \cap \{\tau \leq n\} = (A \cup \{\tau > n\})^c = B_2^c \in \mathscr{F}_n$.

Finally if $(A_k)_{k\geq 1}$ is a collection of events in $\mathscr{F}_\tau$, then so is $A = \bigcap_{k\geq 1} A_k$ since $A \cap \{\tau = n\} = \bigcap_{k\geq 1} A_k\{\tau = n\} \in \mathscr{F}_n$ as $\mathscr{F}_n$ is a σ-algebra. $\square$

Lemma 3.13. *If $\tau \leq \rho$ are two stopping times, then $\mathscr{F}_\tau \subset \mathscr{F}_\rho$.*

Proof. Let $A \in \mathscr{F}_\tau$. Note that $\{\rho \le n\} = \{\tau \le n, \rho \le n\} = \{\tau \le n\} \cap \{\rho \le n\}$. Therefore for $n \ge 0$,

$$A \cap \{\rho \le n\} = A \cap \{\tau \le n\} \cap \{\rho \le n\} \in \mathscr{F}_n$$

as by assumption $A \cap \{\tau \le n\} \in \mathscr{F}_n$ and $\{\rho \le n\} \in \mathscr{F}_n$. So $A \in \mathscr{F}_\rho$ and the claim follows. □

Lemma 3.14. *Let τ be a stopping time such that for some $K \ge 1$ and $\varepsilon > 0$ we have for every $n \ge 0$, almost surely*

$$\mathbb{P}[\tau \le n + K \mid \mathscr{F}_n] = \mathbb{E}[1_{\{\tau \le n+K\}} \mid \mathscr{F}_n] \ge \varepsilon.$$

Then it holds that $\mathbb{E}[\tau] < \infty$.

Proof. The assumed condition is equivalent to $\mathbb{P}[\tau > n+K \mid \mathscr{F}_n] = \mathbb{E}[1_{\{\tau > n+K\}} \mid \mathscr{F}_n] \le (1-\varepsilon)$ almost surely. For $m = 0$, the claim is obvious since $\mathbb{P}[\tau > 0] \le 1 = (1-\varepsilon)^0$. For the inductive step we calculate,

$$\begin{aligned}
\mathbb{P}[\tau > mK] &= \mathbb{E}[1_{\{\tau > mK\}}] \\
&= \mathbb{E}[1_{\{\tau > mK\}} 1_{\{\tau > (m-1)K\}}] \\
&= \mathbb{E}[\mathbb{E}[1_{\{\tau > mK\}} 1_{\{\tau > (m-1)K\}} | \mathscr{F}_{(m-1)K}]] \\
&= \mathbb{E}[1_{\{\tau > (m-1)K\}} \mathbb{E}[1_{\{\tau > (m-1)K+K\}} | \mathscr{F}_{(m-1)K}]] \\
&\le (1-\varepsilon)\mathbb{E}[1_{\{\tau > (m-1)K\}}] = (1-\varepsilon)\mathbb{P}[\tau > (m-1)K] \le (1-\varepsilon)^m.
\end{aligned}$$

We finally deduce that $\mathbb{E}[\tau] < \infty$. Indeed,

$$\begin{aligned}
\mathbb{E}[\tau] &= \sum_{\ell=0}^{\infty} \ell \cdot \mathbb{P}[\tau = \ell] \\
&= \sum_{m=0}^{\infty} \sum_{\ell=mK+1}^{(m+1)K} \ell \cdot \mathbb{P}[\tau = \ell] \\
&\le \sum_{m=0}^{\infty} \sum_{\ell=mK+1}^{(m+1)K} \ell \cdot \mathbb{P}[\tau > mK] \\
&\le \sum_{m=0}^{\infty} (m+1)K^2(1-\varepsilon)^m < \infty.
\end{aligned}$$

□

4. B8.1 Class 4

4.1. Martingales.

Definition 4.1. *An adapted and integrable stochastic process $(M_n)_{n\geq 1}$ is called a martingale if*
$$\mathbb{E}[M_{n+1}|\mathscr{F}_n] = M_n$$
for $n \geq 1$.

4.1.1. *Bellman's Optimality Principle.* Consider $(\varepsilon_n)_{n\geq 1}$ i.i.d. random variables with distribution
$$\mathbb{P}[\varepsilon_n = 1] = p \quad \text{and} \quad \mathbb{P}[\varepsilon_n = -1] = q = 1 - p$$
with $p \in (1/2, 1)$. We now want to create an investment strategy by betting on the outcome of ε_n. Denote $\mathscr{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n)$. Let $Z_0 > 1$ and consider
$$Z_n = Z_{n-1} + \varepsilon_n V_n,$$
where V_n is $\mathscr{F}_{n-1}$-measurable and strictly between 0 and Z_{n-1}.

The goal of this example is to maximise the interest rate $\mathbb{E}[\log\left(\frac{Z_n}{Z_0}\right)]$. We note that
$$\mathbb{E}\left[\log\left(\frac{Z_{n+1}}{Z_n}\right) \mid \mathscr{F}_n\right] = p\log\left(1 + \frac{V_{n+1}}{Z_n}\right) + q\log\left(1 - \frac{V_{n+1}}{Z_n}\right) = f\left(\frac{V_{n+1}}{Z_n}\right),$$
where $f(x) = p\log(1+x) + q\log(1-x)$. We note that
$$f'(x) = \frac{p}{1+x} - \frac{q}{1-x}.$$
A straightforward calculation shows that $f(x)$ is increasing on $(0, p-q]$ and is decreasing on $[p-q, 1]$. The function f has its maximum at $p-q$ and $\alpha = f(p-q) = p\log(2p) + q\log(2q)$ is its maximum. Therefore it follows that
$$\mathbb{E}\left[\log\left(\frac{Z_{n+1}}{Z_n}\right) \mid \mathscr{F}_n\right] \leq \alpha.$$
Thus $M_n = \log Z_n - n\alpha$ is a submartingale and
$$\mathbb{E}\left[\log\left(\frac{Z_n}{Z_0}\right)\right] \leq n\alpha.$$
Thus to maximise the interest rate we want M_n to be a martingale. In order for this to be the case, we need $f(\frac{V_{n+1}}{Z_n}) = \alpha$, which holds if and only if $\frac{V_{n+1}}{Z_n} = (p-q)$. Thus we require $V_{n+1} = (p-q)Z_n$, which then maximises the interest.

4.1.2. *Random Walks and Harmonic Functions.* Let $A \subset \mathbb{Z}^2$ be a finite set of points in a square lattice and let B (the boundary) be the sets of points in $\mathbb{Z}^n \backslash A$ with at least one (horizontal or vertical) neighbour in A. Denote by τ_B the hitting time of the boundary.

Given any function $g : B \to \mathbb{R}$, we consider the function
$$f(v) = \mathbb{E}_v[g(X_{\tau_B})],$$
defined on $A \cup B$, where X is a standard random walk on $\mathbb{Z}^2$ starting on v. Then it holds for every $v \in A$, that
$$f(v) = \frac{1}{4}\sum_{w\sim v} f(w).$$

Indeed, this follows since

$$f(v) = \frac{1}{4}\sum_{w\sim v} \mathbb{E}_v[g(X_{\tau_B})|X_1 = w] = \frac{1}{4}\sum_{w\sim v} \mathbb{E}_w[g(X_{\tau_B})] = \frac{1}{4}\sum_{w\sim v} f(w).$$

Now we consider a simple symmetric random walk $(X_n)_{n\geq 1}$ on $\mathbb{Z}^2$ and $X_0 \in A$. Then we claim that $M_n = f(X_{n\wedge\tau_B})$ is a martingale with respect to the natural filtration $\mathscr{F}_n = \sigma(X_1, \dots, X_n)$.

Notice that if $X_n \in A$, then

$$f(X_n) = \frac{1}{4}\sum_{w\sim X_n} f(w) = \mathbb{E}[f(X_{n+1})|X_n] = \mathbb{E}[f(X_{n+1})|\mathscr{F}_n].$$

We give a more formal argument for a similar equality in the discussion after (4.1).

Therefore it follows that,

$$\begin{aligned}
\mathbb{E}[f(X_{n+1\wedge\tau_B})|\mathscr{F}_n] &= \mathbb{E}[f(X_{n+1\wedge\tau_B})1_{\{\tau_B>n\}}|\mathscr{F}_n] + \mathbb{E}[f(X_{n+1\wedge\tau_B})1_{\{\tau_B\leq n\}}|\mathscr{F}_n] \\
&= 1_{\{\tau_B>n\}}f(X_n) + 1_{\{\tau_B\leq n\}}f(X_{\tau_B}) \\
&= f(X_{n\wedge\tau_B}).
\end{aligned}$$

4.2. **Martingale Convergence Theorems.** The following results were proved in the lecture.

Theorem 4.2. *(Doob's Forward Convergence Theorem) Let $(M_n)_{n\geq 1}$ be a sub or super-martingale that is bounded in L^1, i.e. $\sup_{n\geq 1}\mathbb{E}[|M_n|] < \infty$. Then M_n converges almost surely to a limit M_∞ and $M_\infty \in L^1$.*

Theorem 4.3. *Let $(M_n)_{n\geq 1}$ be a martingale. The following properties are equivalent:*

1. *$(M_n)_{n\geq 1}$ is uniformly integrable.*
2. *There is some $\mathscr{F}_\infty$-measurable random variable M_∞ such that $M_n \to M_\infty$ almost surely and in L^1.*
3. *There is an $\mathscr{F}_\infty$-measurable random variable M_∞ such that $M_n = \mathbb{E}[M_\infty|\mathscr{F}_n]$ almost surely for all n.*

Furthermore, under these conditions, if $M_\infty \in L^p$ for $p > 0$ then the convergence $M_n \to M_\infty$ also holds in L^p.

We can deduce the following L^p-convergence theorem.

Corollary 4.4. *(L^p-convergence theorem) Let $(M_n)_{n\geq 1}$ be a martingale that is bounded in L^p for $p > 1$, i.e. $\sup_{n\geq 1}\mathbb{E}[|M_n|^p] < \infty$. Then then M_n converges almost surely and in L^p to a random variable $M_\infty \in L^p$.*

Proof. Let $q = \frac{p}{p-1}$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then by Hölder's inequality $||M_n||_1 = \mathbb{E}[|M_n|] \leq ||M_n||_p||1||_q = ||M_n||_p = \mathbb{E}[|M_n|^p]^{1/p}$. Thus M_n is bounded in L^1. We claim that M_n is uniformly integrable. Indeed, notice that by Hölder's inequality

$$\mathbb{E}[|M_n|1_A] \leq ||M_n||_p||1_A||_q = ||M_n||_p\mathbb{P}[A]^{1/q} \to 0$$

as $\mathbb{P}[A] \to 0$. Thus by Proposition 5.22 from the lecture notes, it follows that M_n is uniformly integrable.

We furthermore notice as $M_n \to M_\infty$ almost surely, it follows by Fatou's lemma that

$$\mathbb{E}[|M_\infty|^p] = \mathbb{E}[\liminf_{n\to\infty}|M_n|^p] \leq \liminf_{n\to\infty}\mathbb{E}[|M_n|^p] < \infty.$$

Therefore $M_\infty \in L^p$ and the claim follows by Theorem 4.3. $\square$

4.3. **Consistency of Likelihood ratio.** Consider a sequence of i.i.d. tosses of a coin with X_i denoting the outcome of the ith toss. These random variables are defined on some $(\Omega, \mathscr{F})$ on which we have two probability measures $\mathbb{P}_A$ and $\mathbb{P}_B$ and we assume that the X_i are i.i.d. under both measures. Under hypothesis A, $\mathbb{P}_A$ is the true measure, and the probability of a head on any toss is $p = a$. Under hypothesis B, the measure is $\mathbb{P}_B$ and $p = b$, for some $a, b \in (0, 1)$.

Let $P_A(x_1, \ldots, x_n)$ denote the probability of a sequence of outcomes $(x_1, \ldots, x_n)$ under hypothesis A, i.e. $P_A(x_1, \ldots, x_n) = \mathbb{P}_A[X_1 = x_1, \ldots, X_n = x_n]$ with the analogous definition for P_B.

Lemma 4.5. *With the above notation,*

$$Z_n = \frac{\mathbb{P}_A(X_1, \ldots, X_n)}{\mathbb{P}_B(X_1, \ldots, X_n)} = \prod_{i=1}^{n} \frac{\mathbb{P}_A(X_i)}{\mathbb{P}_B(X_i)}.$$

is a martingale under $\mathbb{P}_B$ under the filtration $\mathscr{F}_n = \sigma(X_1, \ldots, X_n)$. Moreover, Z_n converges almost surely to 0 if $a \neq b$.

Proof. Note that $Z_n = Y_1 \cdots Y_n$, where Y_i are independent random variables with distribution

$$\mathbb{P}[Y_i = \tfrac{a}{b}] = b \quad \text{and} \quad \mathbb{P}[Y_i = \tfrac{1-a}{1-b}] = 1 - b.$$

As $\mathbb{E}[Y_i] = 1$, it follows that Z_n is a martingale with respect to $\mathscr{F}_n$ as

$$\mathbb{E}[Z_{n+1} | \mathscr{F}_n] = Z_n \cdot \mathbb{E}[Y_{n+1} | \mathscr{F}_n] = Z_n \cdot \mathbb{E}[Y_{n+1}] = Z_n.$$

Since Z_n is positive, it is bounded in L^1. Therefore by Doob's Martingale convergence theorem (Theorem 4.2) an integrable limit Z_∞ almost surely exists. If $a = b$, then $Z_n = 1$ and thus $Z_\infty = 1$.

On the other hand, if $a \neq b$, then Y_i is bounded away from 1. We claim that $Z_\infty = 0$ almost surely. Indeed, if Z_∞ is non-zero and finite, then the tail

$$Z_\infty / Z_n = \lim_{m \to \infty} Z_m / Z_n = \lim_{m \to \infty} \prod_{i=n+1}^{m} Y_i$$

must converge to 1. Yet this can only be if the Y_i converge to 1. Indeed, if $|Z_m/Z_n - 1| < \varepsilon$ for all $m \geq m_0$, then $Z_{m+1}/Z_n = Y_{m+1} Z_m / Z_n$ and hence

$$Y_{m+1} = \frac{Z_{m+1}/Z_n}{Z_m/Z_n} \in \left[\frac{1 - \varepsilon}{1 + \varepsilon}, \frac{1 + \varepsilon}{1 - \varepsilon} \right] = 1 + O(\varepsilon).$$

This contradicts Y_i being bounded away from 1. Therefore, $Z_\infty = 0$ almost surely.

To give an alternative argument that $Z_\infty = 0$ if $a \neq b$ almost surely, we note that since log is strictly concave,

$$\mathbb{E}[\log(Y_i)] = b \log \tfrac{a}{b} + (1 - b) \log \tfrac{1-a}{1-b} < \log 1 = 0.$$

Therefore by the law of large numbers, it holds almost surely that

$$\lim_{n \to \infty} \frac{\log Z_n}{n} = \mathbb{E}[\log(Y_i)] < 0.$$

Thus $\log Z_n \to -\infty$ and therefore $Z_n \to 0$ almost surely. □

We note that Z_n if $a \neq b$ is an example of a martingale that converges almost surely but not in L^1 to $Z_\infty = 0$. Indeed, $\mathbb{E}[Z_n] = \mathbb{E}[Z_1] = 1$ for all $n \geq 1$ and therefore Z_n cannot converge to 0 in L^1.

4.4. **Polya's Urn Model.** At time 0 we have an urn with two balls, one white and one black. At each successive time, we draw at random one ball from the urn and return it back along with another ball of the same colour. This way, at time n, we have $n+2$ balls in the urn of which B_n are black and $W_n = n+2-B_n$ are white. Note that $B_n \in \{1, \ldots, n+1\}$.

Lemma 4.6. *(Polya's Urn Model) With the above notation,*

$$\mathbb{P}[B_n = k] = \frac{1}{n+1}$$

for $k \in \{1, \ldots, n+1\}$. Moreover, $M_n = \frac{B_n}{n+2}$ is a martingale with respect to $\mathscr{F}_n = \sigma(B_1, \ldots, B_n)$ that converges to M_∞, being the uniform $[0,1]$ variable.

Proof. The first claim is proved by induction. Indeed, the claim holds for $n = 0$ and $n = 1$. Assume now the claim holds for time $n-1$. Then we have for $k \in \{1, \ldots, n\}$,

$$\begin{aligned}
\mathbb{P}[B_n = k] &= \sum_{\ell=1}^{n} \mathbb{P}[B_n = k | B_{n-1} = \ell] \cdot \mathbb{P}[B_{n-1} = \ell] \\
&= \frac{1}{n}(\mathbb{P}[B_n = k | B_{n-1} = k] + \mathbb{P}[B_n = k | B_{n-1} = k-1]) \\
&= \frac{1}{n}\left(\frac{n+1-k}{n+1} + \frac{k-1}{n+1}\right) = \frac{1}{n+1}.
\end{aligned}$$

In addition,

$$\mathbb{P}[B_n = n+1] = \mathbb{P}[B_n = n+1 | B_{n-1} = n]\mathbb{P}[B_{n-1} = n] = \frac{n}{n+1} \cdot \frac{1}{n} = \frac{1}{n+1}.$$

This concludes the proof the first claim.

To show that $M_n = \frac{B_n}{n+2}$ is a martingale, we denote by

$$X_i = \begin{cases} 1 & \text{if the ith ball is black,} \\ 0 & \text{if the ith ball is white.} \end{cases}$$

Then it holds that $B_n = 1 + \sum_{i=1}^n X_i$. We note that

$$\mathbb{E}[\tfrac{B_{n+1}}{n+3} | \mathscr{F}_n] = \frac{1}{n+3}(B_n + \mathbb{E}[X_{n+1} | \mathscr{F}_n]).$$

Since X_{n+1} only depends on B_n,

$$\mathbb{E}[X_{n+1} | \mathscr{F}_n] = \mathbb{E}[X_{n+1} | B_n] = \frac{B_n}{n+2}. \tag{4.1}$$

This implies that $M_n = \frac{B_n}{n+2}$ is a martingale.

We give a more precise argument for (4.1). Indeed, we can express $X_{n+1} = \sum_{k=1}^{n+1} X_{n+1} 1_{\{B_n = k\}}$ and note that since the distribution of X_{n+1} depends only on the value of B_n, the function $\mathbb{E}[X_{n+1}|\mathscr{F}_n]$ is constant on the sets $1_{B_n=k}$. Therefore

$$\begin{aligned}\mathbb{E}[X_{n+1}1_{\{B_n=k\}}|\mathscr{F}_n] &= 1_{\{B_n=k\}}\frac{\mathbb{E}[X_{n+1}1_{\{B_n=k\}}]}{\mathbb{P}[B_n=k]}\\ &= 1_{\{B_n=k\}}\frac{\mathbb{E}[1_{\{X_{n+1}=1\}}1_{\{B_n=k\}}]}{\mathbb{P}[B_n=k]}\\ &= 1_{\{B_n=k\}}\frac{\mathbb{P}[B_{n+1}=k+1,B_n=k]}{\mathbb{P}[B_n=k]}\\ &= 1_{\{B_n=k\}}\frac{\frac{1}{n}\frac{k}{n+2}}{\frac{1}{n}}\\ &= 1_{\{B_n=k\}}\frac{k}{n+2}.\end{aligned}$$

Therefore

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] = \sum_{k=1}^{n+1}\mathbb{E}[X_{n+1}1_{\{B_n=k\}}|\mathscr{F}_n] = \sum_{k=1}^{n+1}1_{\{B_n=k\}}\frac{k}{n+2} = \frac{B_n}{n+2}.$$

The same argument shows that $\mathbb{E}[X_{n+1}|B_n]$ is the same function.

As M_n is uniformly bounded, it is uniformly integrable and hence by Theorem 4.3 converges almost surely and in L^1 to a random variable M_∞. Notice that for $x \in [0,1)$

$$\mathbb{P}[M_n \le x] = \frac{\lfloor (n+2)x\rfloor}{n+1} \to x,$$

as $n \to \infty$. So M_n converges in distribution to a uniform $[0,1]$-variable and hence also in L^1 and almost surely. □

4.5. **Galton-Watson branching process.** Let $(X_{n,r})_{n,r\ge 1}$ be an infinite array of independent identically distributed random variables, each with the same distribution as X, where

$$\mathbb{P}[X=k] = p_k$$

for $k = 0,1,\ldots$. Denote $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$ and assume that $\sigma^2 < \infty$. The sequence of random variables $(Z_n)_{n\ge 0}$ is defined by $Z_0 = 1$ and

$$Z_n = X_{n,1} + \ldots + X_{n,Z_{n-1}}.$$

We note that

$$\mathbb{E}[Z_{n+1}^2|\{Z_n=k\}] = \mathbb{E}\left[\sum_{i,j=1}^{k} X_{n+1,i}X_{n+1,j}\right] = k\sigma^2 + k^2\mu^2.$$

Therefore

$$\mathbb{E}[Z_{n+1}^2|\mathscr{F}_n] = \mathbb{E}[Z_{n+1}^2|Z_n] = \sigma^2\cdot Z_n + \mu^2\cdot Z_n^2.$$

To calculate $\mathbb{E}[Z_n^2]$, we recall that $\mathbb{E}[Z_n] = \mu^n$ and therefore

$$\begin{aligned}\mathbb{E}[Z_{n+1}^2] &= \mathbb{E}[\mathbb{E}[Z_{n+1}^2|\mathscr{F}_n]]\\ &= \sigma^2\cdot\mathbb{E}[Z_n] + \mu^2\cdot\mathbb{E}[Z_n^2]\\ &= \sigma^2\mu^n + \mu^2\cdot\mathbb{E}[Z_n^2].\end{aligned}$$

Notice that $\mathbb{E}[Z_0^2] = 1$, $\mathbb{E}[Z_1^2] = \mu^2 + \sigma^2$ and $\mathbb{E}[Z_2^2] = \mu^4 + \sigma^2(\mu + \mu^2)$. More generally we claim for $n \geq 1$ that

$$\mathbb{E}[Z_n^2] = \mu^{2n} + \sigma^2(\mu^{n-1} + \mu^n + \ldots + \mu^{2n-2}).$$

It is a straightforward calculation to check that this formula indeed satisfies the inductive relation. We conclude since $\mu > 1$,

$$\mathbb{E}[M_n^2] = \mathbb{E}\left[\frac{Z_n^2}{\mu^{2n}}\right] = 1 + \sigma^2(\mu^{-2} + \ldots + \mu^{-(n+1)}) \leq 1 + \frac{\sigma^2}{1-\mu}.$$

Therefore M_n is bounded in L^2. Hence, by Theorem 4.4, M_n converges in L^2 to a random variable M_∞ and hence also in L^1.

4.6. **Gambler's Ruin.** We are going to use the following theorem:

Theorem 4.7. *(Optional Stopping Theorem) Let $(M_n)_{n\geq 1}$ be a martingale on a filtered probability space and let τ be an almost surely finite stopping time. Assume that either*

(1) $(M_n)_{n\geq 1}$ is uniformly integrable, or
(2) $\mathbb{E}[\tau] < \infty$ and

$$\left(\sup_{n\geq 1} \mathbb{E}[|M_{n+1} - M_n| \,|\, \mathscr{F}_n]\right) < \infty$$

almost surely.

Then

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

In this section we are going to discuss the gambler's ruin problem. Let $(X_n)_{n\geq 1}$ be an i.i.d. sequence of random variables with

$$\mathbb{P}[X_i = 1] = p \quad \text{and} \quad \mathbb{P}[X_i = -1] = q = 1 - p$$

with $p \in (0, 1)$. Suppose that a and b are integers with $0 < a < b$ and consider

$$S_n = a + X_1 + \ldots + X_n \quad \text{and} \quad \tau = \inf\{n \geq 1 \,:\, S_n = 0 \text{ or } S_n = b\}.$$

Then τ is a stopping time.

We denote $p_0 = \mathbb{P}[S_\tau = 0]$ and $p_b = \mathbb{P}[S_\tau = b]$

Lemma 4.8. *It holds that $\mathbb{E}[\tau] < \infty$.*

Proof. Let A_k for $k \geq 1$ be the event that

$$A_k = \{X_{(k-1)\cdot b+1} = \ldots = X_{k\cdot b} = 1\}.$$

Then $\mathbb{P}(A_k) = p^b$ for all $k \geq 1$ since the X_i are independent. If A_k happens, then $\tau \leq k \cdot b$ since either we were below 0 some time before $k \cdot b$ or we are at b before $k \cdot b$. Therefore it follows that

$$\{\tau \geq n\} \leq (1-\gamma)^{\lfloor \frac{n}{b} \rfloor} \leq e^{-cn}$$

for suitable constants c. Therefore

$$\mathbb{E}[\tau] = \sum_{n=1}^{\infty} n\mathbb{P}[\tau = n] \leq \sum_{n=1}^{\infty} n\mathbb{P}[\tau \geq n] \leq \sum_{n=1}^{\infty} ne^{-cn} < \infty.$$

$\square$

We make one important remark that shows we need to be careful with the assumptions in the optional stopping theorem. Indeed, if $p = q$ note that S_n is a martingale. Consider now the one-sided stopping time

$$\tau' = \inf\{n \geq 1 \,:\, S_n = 0\}.$$

We note that the conclusion of the optional stopping theorem does not hold as $0 = \mathbb{E}[M_{\tau'}] \neq \mathbb{E}[M_0] = a$. The stopping time τ' is almost surely finite, as we have seen in previous sheets. So we conclude that S_n is not uniformly integrable. Also we note that $\mathbb{E}[|S_{n+1} - S_n|] \leq 1$ as we move by one step at each time. Thus it must follow that $\mathbb{E}[\tau'] = \infty$.

We note that if $\mathbb{E}[\tau']$ was finite, would could easily make arbitrage in unbiased financial markets, which I encourage the reader to ponder about.

4.6.1. *Biased Gambler's Ruin.* Assume that $p \neq q$. Throughout this section write $\gamma = p/q$. We aim to calculate $\mathbb{E}[\tau]$, p_0 and p_b.

Lemma 4.9. *The stochastic processes*

$$M_n = \gamma^{S_n} \quad \text{and} \quad N_n = S_n - n(p-q)$$

are martingales.

We note that if $\gamma = 1$, then M_n is always 1.

Proof. We first show that N_n is a martingale. Indeed,

$$\mathbb{E}[N_{n+1}|\mathscr{F}] = S_n - (n+1)(p-q) + \mathbb{E}[X_{n+1}|\mathscr{F}_n] = N_n$$

since $\mathbb{E}[X_{n+1}|\mathscr{F}_n] = \mathbb{E}[X_{n+1}] = p - q$. For the second claim, we calculate

$$\mathbb{E}[M_{n+1}|\mathscr{F}_n] = \gamma^{S_n} \cdot \mathbb{E}[\gamma^{X_{n+1}}] = M_n,$$

since $\mathbb{E}[\gamma^{X_{n+1}}] = \mathbb{E}[\gamma^{X_{n+1}}] = p\gamma + q\gamma = q + p = 1$. This concludes the proof. $\square$

We note that $\mathbb{E}[\tau] < \infty$ and that $|N_{n+1} - N_n| \leq 2$ and therefore $\mathbb{E}[|N_{n+1} - N_n| \,|\, \mathscr{F}_n] \leq 2$. So we can apply the optimal stopping theorem to N_n, which implies

$$b \cdot p_b - \mathbb{E}[\tau](p-q) = \mathbb{E}[S_\tau] - \mathbb{E}[\tau](p-q) = \mathbb{E}[N_\tau] = \mathbb{E}[N_0] = S_0 = a$$

On the other hand, none of the assumption of the optional stopping theorem are satisfies for M_n. Yet we can use the following trick. Namely, we consider stopped the stochastic processes

$$M_n^\tau = M_{n\wedge\tau},$$

which was shown in the lecture to be a martingale. Note that $0 \leq M_n^\tau \leq \max\left\{1, \gamma^b\right\}$, so it is uniformly bounded and hence uniformly integrable. Thus we can apply the optional stopping theorem to conclude that

$$p_0 + p_b\gamma^b = p_0\gamma^0 + p_b\gamma^b = \mathbb{E}[M_\tau] = \mathbb{E}[M_0] = \gamma^a.$$

Using that $p_0 + p_b = 1$, it follows that

$$p_b = \frac{\gamma^a - 1}{\gamma^b - 1} \quad \text{and} \quad p_0 = 1 - p_b = \frac{\gamma^b - \gamma^a}{\gamma^b - 1}.$$

Moreover, by the above

$$\mathbb{E}[\tau] = \frac{bp_b - a}{p-q}.$$

4.6.2. *Unbiased Gambler's Ruin.* We assume in the following that $p = q = 1/2$. Then S_n is a martingale that satisfies the assumption of the optional stopping theorem. Since $p_0 + p_b = 1$ and as $bp_b = \mathbb{E}[S_\tau] = \mathbb{E}[S_0] = a$, it follows that

$$p_b = \frac{a}{b} \quad \text{and} \quad p_a = 1 - p_b = 1 - \frac{a}{b} = \frac{b-a}{b}.$$

We observe that the above studies martingale $M_n = \gamma^{S_n}$ is not useful in the unbiased case. To deduce the equation for $\mathbb{E}[\tau]$, note that $\sigma^2 = 1$ and recall that $S_n^2 - n$ is a martingale. Thus by optional stopping,

$$ab - \mathbb{E}[\tau] = b^2 p_b - \mathbb{E}[\tau] = \mathbb{E}[S_\tau^2] - \mathbb{E}[\tau] = \mathbb{E}[S_0^2] = a^2.$$

So $\mathbb{E}[\tau] = ab - a^2 = a(b-a)$. This calculative approach does not work in the unbiased case.

4.6.3. *Alternative approach using polynomial method.* We give an example where an alternative approach to some of the above problems. We review the method of characteristic polynomials to solve recurrence relations. Indeed, assume we are given a recurrence relation with $a_0, \ldots, a_{d-1}$ to be fixed and

$$a_n = \alpha_1 a_{n-1} + \ldots + \alpha_d a_{n-d}$$

for $n \geq d$. Then we consider the equation involving the characteristic polynomial

$$\lambda^d - \alpha_1 \lambda^{d-1} - \alpha_2 \lambda^{d-2} - \ldots - \alpha_d = 0$$

Let $\lambda_1, \ldots, \lambda_k$ be distinct real roots of the above polynomial. Then any sequence

$$a_n = \sum_{i=1}^{k} c_i \lambda_i^n$$

satisfies the equation as one readily checks.

Example 4.10. *Assume that we are in a biased random walk with*

$$\mathbb{P}[X_i = 1] = \frac{2}{3} \quad \text{and} \quad \mathbb{E}[X_i = -1] = \frac{1}{3}.$$

We start at n and denote by p_n the probability that we ever hit 0. Then it holds that

$$p_n = \frac{1}{2^n}.$$

Proof. Let p_n be the probability that we hit 0. We then have $p_0 = 1$ and the recurrence relation

$$p_n = \frac{2}{3} p_{n+1} + \frac{1}{3} p_{n-1}$$

or equivalently

$$p_n = \frac{3}{2} p_{n-1} - \frac{1}{2} p_{n-2}.$$

Therefore the characteristic equation is

$$\lambda^2 - \frac{3}{2}\lambda + \frac{1}{2} = (\lambda - 1)(\lambda - \frac{1}{2}).$$

Thus it follows that

$$p_n = \frac{1}{2^n}.$$

We give an alternative solution by using the results from section 4.6.1. Indeed consider the previous example with $p = \frac{2}{3}$ and $q = \frac{1}{3}$ such that $\frac{q}{p} = \frac{1}{2}$. Let τ be the hitting time of $\{0, b\}$ for $b > n$ and assume that the random walk starts at n. Write $p_{n,0} = \mathbb{P}[S_\tau = 0]$. Then it holds that

$$p_{n,0} = \frac{\frac{1}{2^b} - \frac{1}{2^n}}{\frac{1}{2^b} - 1}.$$

The probability in question is thus

$$p_n = \lim_{b \to \infty} p_{n,0} = \frac{1}{2^n}.$$

□

5. B8.1 2023 Exam

5.1. **Question 1.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

5.1.1. *Question 1 a).* Recall that a π-system is a collection of sets that is closed under intersections. A λ-system is a collection of sets $\mathscr{M}$ such that

(1) $\Omega \in \mathscr{M}$.
(2) If $A, B \in \mathscr{M}$ and $A \subset B$ then $B \backslash A \in \mathscr{M}$.
(3) If $(A_n)_{n \geq 1} \subset \mathscr{M}$ is a collection of sets with $A_n \subset A_{n+1}$ for all $n \geq 1$, then $\bigcup_{n \geq 1} A_n \in \mathscr{M}$.

Lemma 5.1. *Let $\mathscr{A}_1$ and $\mathscr{A}_2$ be two π-systems in $\mathscr{F}$ and let $\mathscr{G}_i = \sigma(\mathscr{A}_i)$ for $i = 1, 2$. Then $\mathscr{G}_1$ and $\mathscr{G}_2$ are independent if*

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$$

for all $A \in \mathscr{A}_1$ and $B \in \mathscr{A}_2$.

Proof. Consider

$$\mathscr{M}_1 = \{A \subset \Omega : \mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B] \text{ for all } B \in \mathscr{A}_2\}.$$

We claim that $\mathscr{M}_1$ is a λ-system. Indeed, it is clear that $\Omega \in \mathscr{M}$. Assume now that $A_1, A_2 \in \mathscr{M}$ with $A_1 \subset A_2$. Then it holds that

$$\mathbb{P}[(A_2 \backslash A_1) \cap B] = \mathbb{P}[A_2 \cap B] - \mathbb{P}[A_1 \cap B] = (\mathbb{P}[A_2] - \mathbb{P}[A_1])\mathbb{P}[B] = \mathbb{P}[A_2 \backslash A_1]\mathbb{P}[B]$$

and therefore $A_2 \backslash A_1 \in \mathscr{M}$. Finally, if $(A_n)_{n \geq 1}$ is an increasing sequence of sets in $\mathscr{A}$ then write $A = \bigcup_{n \geq 1} A_n$. It holds by monotone convergence that

$$\mathbb{P}[A \cap B] = \lim_{n \to \infty} \mathbb{P}[A_n \cap B] = \lim_{n \to \infty} \mathbb{P}[A_n]\mathbb{P}[B] = \mathbb{P}[A]\mathbb{P}[AB].$$

So $A \in \mathscr{M}$ and so we have shown that $\mathscr{M}$ is a λ-system.

By our assumption $\mathscr{A}_1 \subset \mathscr{M}_1$ and therefore by the $\pi - \lambda$-systems lemma, it follows that $\sigma(\mathscr{A}_1) = \mathscr{G}_1 \subset \mathscr{M}_1$.

We next consider

$$\mathscr{M}_2 = \{B \subset \Omega : \mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B] \text{ for all } A \in \mathscr{G}_1\}.$$

Then as before $\mathscr{M}_2$ is a λ-system and by the first step $\mathscr{A}_2 \subset \mathscr{M}_2$. By the $\pi - \lambda$-systems lemma it follows that $\mathscr{G}_2 = \sigma(\mathscr{A}_2) \subset \mathscr{M}_2$, which concludes the proof. $\square$

5.1.2. *Question 1 b).* Now suppose that $Y_0, Y_1, \ldots$ are independent random variables with

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

for all $n \geq 0$. For $n \geq 1$ we define

$$X_n = Y_0 Y_1 \cdots Y_n.$$

We claim that $X_1, X_2, \ldots$ are independent.

We now consider

$$\mathcal{Y} = \sigma(Y_1, Y_2, \ldots), \quad \mathcal{T}_n = \sigma(X_k : k \geq n)$$

and

$$\mathcal{G} = \bigcap_{n \geq 1} \sigma(\mathcal{Y}, \mathcal{T}_n), \quad \mathcal{H} = \sigma\left(\mathcal{Y}, \bigcap_{n \geq 1} \mathcal{T}_n\right).$$

For every n, we have $Y_0 = Y_1^{-1} \cdots Y_n^{-1} X_n$. Therefore, since all of the random variables $Y_1, \dots, Y_n$ and X_n are $\sigma(\mathcal{Y}, \mathcal{T}_n)$-measurable and as the product of measurable random variables is again measurable, it follows that Y_0 is $\sigma(\mathcal{Y}, \mathcal{T}_n)$-measurable. As n was arbitrary, Y_0 is $\mathcal{G}$-measurable.

We now want to show that Y_0 is independent of $\mathcal{H}$ so it suffices by a) to show that

$$\mathbb{P}[\{Y_0 = i\} \cap A \cap B] = \mathbb{P}[Y_0 = i]\mathbb{P}[A]\mathbb{P}[B]$$

for $i = 1, -1$ and all $A \in \mathcal{Y}$ and $B \in \bigcap_{n\geq 1} \mathcal{T}_n$. By the Komlogorov $0-1$ law (as the X_i are independent), every element of $\bigcap_{n\geq 1} \mathcal{T}_n$ has either measure 0 or 1. So if $\mathbb{P}[B] = 0$, the above claim is obvious so we can assume that $\mathbb{P}[B] = 1$ in which case it suffices to show that

$$\mathbb{P}[\{Y_0 = i\} \cap A] = \mathbb{P}[Y_0 = i]\mathbb{P}[A],$$

which follows as Y_0 is independent from $\mathcal{Y}$ by construction.

5.1.3. *Question 1 c).* Let $(M_n)_{n\geq 1}$ be a martingale relative to a given filtration $(\mathcal{F}_n)_{n\geq 0}$ and such that $|M_{n+1} - M_n| \leq L$ for all $n \geq 0$ and some constant L. Assume that $M_0 = 0$.

(i) Let $\tau_K = \inf\{n \geq 0 : M_n \leq -K\}$ for $K > 0$. Then we claim that $\lim_{n\to\infty} M_n$ exists on $\{\tau_K = \infty\}$. Indeed, note that τ_K is a stopping time as it is the first hitting time for an adapted process. Also note that $(M_{n\wedge\tau_K} + K + L)_{n\geq 0}$ is a non-negative martingale and therefore converges almost surely. So M_n converges on $\tau_K = \infty$.

(ii) Therefore M_n converges on $\bigcup_{K=1}^{\infty}\{\tau_K = \infty\} = \{\liminf M_n > -\infty\}$. Applying the same argument for $-M$ gives the same conclusion for $\limsup M_n < \infty$. So we have shown that M_n converges on the set $\liminf M_n > -\infty$ and $\limsup M_n < \infty$, which implies the claim that

$$\begin{aligned} A &= \{\lim_{n\to\infty} M_n \text{ exists and is finite}\} \\ B &= \{\limsup M_n = \infty \quad \text{and} \quad \liminf M_n = \infty\} \end{aligned}$$

satisfies $\mathbb{P}[A \cup B] = 1$.

5.1.4. *Exercise 1 d).* Consider a sequence of events $(B_n)_{n\geq 1}$ and let $\mathcal{F}_n = \sigma(B_1, \dots, B_n)$ for $n \geq 1$ and $\mathscr{F}_0 = \{\emptyset, \Omega\}$.

(i) Now consider $(X_n)_{n\geq 0}$ the submartingale given by $X_0 = 0$ and $X_n = \sum_{k=1}^{n} 1_{B_k}$ for all $n \geq 1$. Recall that the Doob decomposition of X_n is the decomposition $X_n = M_n + A_n$, where M_n is a martingale and A_n is a predictable process. The martingale M_n is given as

$$M_n = \sum_{k=1}^{n} (X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}]) = \sum_{k=1}^{n} (1_{B_k} - \mathbb{E}[1_{B_k}|\mathcal{F}_{k-1}]).$$

(ii) Note that $|M_n - M_{n-1}| \leq 2$ and let A and B the the sets from c). From c) we conclude that if $\omega \in A$, then $\sum_{k=1}^{\infty} 1_{B_k}(\omega) = \infty$ if and only if $\sum_{k=1}^{\infty} \mathbb{E}[1_{B_k}|\mathcal{F}_{n-1}](\omega) = \infty$. When $\omega \in B$ then M_n oscillates and therefore we must have $\sum_{k=1}^{\infty} 1_{B_k}(\omega) = \infty$ then it must also hold that $\sum_{k=1}^{\infty} \mathbb{E}[1_{B_k}|\mathcal{F}_{n-1}](\omega) = \infty$. This implies

$$\mathbb{P}[B_n \text{ i.o.}] = \mathbb{P}\left[\sum_{k=1}^{\infty} \mathbb{E}[1_{B_k}|\mathcal{F}_{n-1}](\omega) = \infty\right].$$

(iii) To prove the second Borel-Cantelli lemma, we note that if $B_1, B_2, \dots$ are independent, then $\mathbb{E}[1_{B_k}|\mathcal{F}_{n-1}](\omega) = \mathbb{P}[B_k]$ and so if $\sum_{k\geq 1}\mathbb{P}[B_k] = \infty$, then $\mathbb{P}[B_n \text{ i.o.}] = 1$.

5.2. **Question 2.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

5.2.1. *Question 2 a).* (i) Let X be a random variable and let $\mathcal{B}(\mathbb{R})$ denote the Borel σ-algebra on $\mathbb{R}$.

Lemma 5.2. *It holds that*

$$\sigma(X) = \{X^{-1}(A) \,:\, A \in \mathcal{B}(\mathbb{R})\}.$$

Proof. The right hand side is a σ-algebra since it clearly contains Ω and it holds that $X^{-1}(A^c) = X^{-1}(A)^c$ and $X^{-1}(\cup_{i\geq 1}A_i) = \cup_{i\geq 1}X^{-1}(A_i)$ for any sets A and A_i in $\mathcal{B}(\mathbb{R})$. By definition, $\sigma(X) \supset \{X^{-1}(A) \,:\, A \in \mathcal{B}(\mathbb{R})\}$ and since the right hand side is a σ-algebra, $\sigma(X) \subset \{X^{-1}(A) \,:\, A \in \mathcal{B}(\mathbb{R})\}$. This concludes the proof. □

(ii) For the next exercise we recall the statement of the monotone class theorem.

Theorem 5.3. *(Monotone Class Theorem) Let $\mathscr{H}$ be a class of bounded functions from Ω to $\mathbb{R}$ satisfying the following:*

(1) $\mathscr{H}$ is a vector space.
(2) The constant function 1 is in $\mathscr{H}$.
(3) If $(f_n)_{n\geq 1}$ is a sequence in $\mathscr{H}$ such that $f_n \uparrow f$ for a bounded function f, then $f \in \mathscr{H}$.

If $\mathscr{C} \subset \mathscr{H}$ is stable under pointwise multiplication, then $\mathscr{H}$ contains all bounded $\sigma(\mathscr{C})$-measurable functions.

Lemma 5.4. *If a bounded random variable Z is $\sigma(X)$-measurable, then $Z = g(X)$ for some measurable $g : \mathbb{R} \to \mathbb{R}$.*

Proof. Consider $\mathscr{H}$ to be the class of bounded functions of the form $g(X)$ for some measurable map $g : \mathbb{R} \to \mathbb{R}$. The class $\mathscr{H}$ satisfies the assumption of the Monotone Class Theorem since if $Y_n = g_n(X)$ with $Y_n \uparrow Y$ for Y a bounded function, then we can take $g = \limsup_{n\geq 1} g_n$ and check that $Y = g(X)$. We furthermore define $\mathscr{C} = \{1_C \,:\, C \in \sigma(X)\}$. Then $\mathscr{C} \subset \mathscr{H}$ since by the previous lemma $C = X^{-1}(A)$ for some $A \in \mathcal{B}(\mathbb{R})$ and so we can set $g = 1_A$. As $\sigma(X)$ is a σ-algebra and therefore a π-system and $1_A \cdot 1_B = 1_{A\cap B}$, the set $\mathscr{C}$ is stable under pointwise multiplication. Therefore by the Monotone Class Theorem, $\mathscr{H}$ contains all bounded $\sigma(\mathscr{C}) = \sigma(X)$ measurable functions, concluding the proof. □

5.2.2. *Question 2 b).* Consider $f : [0,1] \to \mathbb{R}$ an L-Lipschitz function, that is $|f(u) - f(v)| \leq L|u - v|$ for all $u, v \in [0,1]$. Suppose X has a uniform distribution on $[0,1]$ and define

$$X_n = \frac{\lfloor 2^n X \rfloor}{2^n}, \quad \text{and} \quad Z_n = 2^n(f(X_n + 2^{-n}) - f(X_n))$$

for all $n \geq 1$, where $\lfloor x \rfloor$ denotes the integer part.

(i) We note that for all $\omega \in \Omega$ it holds that $X_n(\omega) \in [X(\omega) - 2^{-n}, X(\omega)]$ and therefore for all $\omega \in \Omega$, $|X_n(\omega) - X(\omega)| \leq 2^{-n}$ and thus the converegnce is almost surely, in probability and in L^2.

(ii) We next claim for any $n \geq 1$ that

$$\sigma(X_n, X_{n+1}, \dots) = \sigma(X).$$

It is clear that $\sigma(X_n) \subset \sigma(X)$ since X_n is $\sigma(X)$-measurable. To show the converse, it holds as $X = \lim_{n\to\infty} X_n$ that X is $\sigma(X_n, X_{n+1}, \ldots)$-measurable and therefore the claim follows.

(iii) A family of random variables $\{X_i : i \in I\}$ is uniformly integrable if

$$\lim_{K\to\infty} \sup_{i\in I} \mathbb{E}[|X_i| 1_{|X|>K}] = 0.$$

The family $(Z_n)_{n\geq 1}$ is indeed uniformly integrable as

$$|Z_n(\omega)| = 2^n |f(X_n(\omega) + 2^{-n}) - f(X_n(\omega))| \leq 2^n L |X_n(\omega) + 2^{-n} - X_n(\omega)| \leq L$$

for all $n \geq 1$ as f is L-Lipschitz. So Z_n is bounded and hence uniformly integrable.

(iv) Finally, consider $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and let $h : [0,1] \to \mathbb{R}$ be a bounded measurable function. Then by construction $\mathcal{F}_n = \{X^{-1}([\frac{i}{2^n}, \frac{i+1}{2^n})) : 0 \leq i \leq 2^n - 1\}$ and so $\mathcal{F}_n$ is a partition of Ω. For convenience denote by $A_i = X^{-1}([\frac{i}{2^n}, \frac{i+1}{2^n}))$ for $0 \leq i \leq 2^n$ and thus for $\omega \in \Omega$, almost surely

$$\begin{aligned}
\mathbb{E}[h(X)|\mathcal{F}_n](\omega) &= \sum_{i=0}^{2^n} \frac{\mathbb{E}[h(X)1_{A_i}]}{\mathbb{P}[A_i]} 1_{A_i}(\omega) \\
&= \sum_{i=0}^{2^n} 2^n \int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} h(x)\, dx \cdot 1_{X(\omega)\in[\frac{i}{2^n}, \frac{i+1}{2^n})} \\
&= 2^n \int_{X_n(\omega)}^{X_n(\omega)+2^{-n}} h(x)\, dx.
\end{aligned}$$

5.2.3. *Question 2 c).* (i) We now show that Z_n is a martingale relative to the filtration $(\mathcal{F}_n)_{n\geq 1}$. Note that

$$X_{n+1} = \frac{\lfloor 2^{n+1} X \rfloor}{2^{n+1}} \in \{X_n, X_n + \frac{1}{2^{n+1}}\}$$

and each of these cases happens with probability $1/2$. Therefore

$$\mathbb{E}[f(X_{n+1})|\mathcal{F}_n] = \frac{1}{2}(f(X_n) + f(X_n + \tfrac{1}{2^{n+1}})).$$

Thus it follows that

$$\begin{aligned}
\mathbb{E}[Z_{n+1}|\mathcal{F}_n] &= 2^{n+1} \mathbb{E}[f(X_{n+1} + \tfrac{1}{2^{n+1}}) - f(X_{n+1})|\mathcal{F}_n] \\
&= 2^{n+1} \frac{1}{2}(f(X_n + \tfrac{1}{2^n}) + f(X_n + \tfrac{1}{2^{n+1}}) - f(X_n + \tfrac{1}{2^{n+1}}) - f(X_n)) \\
&= 2^n (f(X_n + \tfrac{1}{2^n}) - f(X_n)) = Z_n.
\end{aligned}$$

Thus $(Z_n)_{n\geq 1}$ is a martingale.

(ii) The martingale convergence theorem for UI martingales states that if $(M_n)_{n\geq 1}$ is a sequence of UI martingales then there is a $\mathcal{F}_\infty$ random variable M_∞ such that $M_n \to M_\infty$ almost surely and in L^1 and $M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]$. Applied to Z_n it follows that there is a $\sigma(X)$-measurable random variable Z_∞ such that $Z_n \to Z_\infty$ almost surely and in L^1.

(iii) By a) it holds that $Z_\infty = g(X)$ for a Borel measurable function $g : [0,1] \to \mathbb{R}$. Thus by b), we have almost surely,

$$Z_n = \mathbb{E}[Z_\infty|\mathcal{F}_n] = \mathbb{E}[g(X)|\mathcal{F}_n] = 2^n \int_{X_n}^{X_n+2^{-n}} g(x)\, dx.$$

By the definition of Z_n it therefore follows that

$$f(X_n + 2^{-n}) - f(X_n) = \int_{X_n}^{X_n+2^{-n}} g(x)\,dx.$$

Thus we conclude by a telescoping sum that

$$f\left(\frac{i}{2^n}\right) - f(0) = \int_0^{\frac{i}{2^n}} g(u)\,du$$

for all $0 \leq i \leq 2^n$. As $n \to \infty$ and since f is continuous,

$$f(x) - f(0) = \int_0^x g(u)\,du$$

for almost all $x \in [0,1]$.

5.3. **Question 3.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

5.3.1. *Question 3 a).* Three players start with a, b and c tokens respectively. In each round, two players are selected, uniformly at random from the players in play, and then one of them, selected uniformly at random, gives the other one a token. Each choice of a player is made independently of everything that has happened so far. When a player has no tokens left, she stops playing. When one player gathers all the tokens she wins and nothing else happens in the subsequent rounds of the game.

We denote X_n, Y_n, Z_n the number of tokens owned by each of the players after the nth round. In particular, $X_0 = a$, $Y_0 = b$ and $Z_0 = c$. We let $\mathcal{F}_n = \sigma(X_k, Y_k, Z_k : 0 \leq k \leq n)$.

(i) and (ii) Let τ be the first time one of the players has no tokens left, which is clearly a stopping time. Let

$$M_n = X_n Y_n Z_n + \frac{n}{3}(a+b+c)$$

for $n \geq 0$. Then we claim that $(M_{n\wedge\tau})_{n\geq 1}$ is a martingale. Indeed, we note that while X_n, Y_n and Z_n are all non-zero it holds that

$$X_{n+1}Y_{n+1}Z_{n+1} = X_n Y_n Z_n + \frac{1}{3}(X_n + Y_n + Z_n).$$

This implies the claim.

(iii) We observe that M_n is not a martingale itself as when τ happens, $X_n Y_n Z_n$ is zero but $\frac{n}{3}(a+b+c)$ still grows. So M_n is a submartinagle since

$$\begin{aligned}
E[M_{n+1}|\mathscr{F}_n] &= E[M_{n+1}1_{n<\tau}|\mathscr{F}_n] + E[M_{n+1}1_{\tau\leq n}|\mathscr{F}_n] \\
&= M_n 1_{n<\tau} + \frac{n+1}{3}(a+b+c)1_{\tau\leq n} \\
&\geq M_n 1_{n<\tau} + \frac{n}{3}(a+b+c)1_{\tau\leq n} \\
&= M_n.
\end{aligned}$$

(iv) Finally, we compute $\mathbb{E}[\tau]$. It clearly holds that $\mathbb{E}[\tau] < \infty$ since

$$\mathbb{P}[\tau \geq (a+b+c)n] \leq e^{-cn}$$

for some constant $c > 0$. Therefore, as $|M_{n+1} - M_n| \leq L$ for some constant $L > 0$, it follows that

$$\frac{\mathbb{E}[\tau]}{3}(abc) = \mathbb{E}[M_\tau] = \mathbb{E}[M_0] = abc$$

and therefore

$$\mathbb{E}[\tau] = \frac{3abc}{a+b+c}.$$

5.3.2. *Question 3 b).* Let X be an integrable random variable and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let $Y = \mathbb{E}[X|\mathcal{G}]$ and suppose that X and Y have the same distribution.

(i) Assume for the moment that X is square integrable. Then so is Y and it holds that

$$\begin{aligned}\mathbb{E}[(Y-X)^2] &= \mathbb{E}[Y^2] + \mathbb{E}[X^2] - 2\mathbb{E}[XY] \\ &= 2\mathbb{E}[Y^2] - 2\mathbb{E}[\mathbb{E}[XY|\mathcal{G}]] \\ &= 2\mathbb{E}[Y^2] - 2\mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]] \\ &= 2\mathbb{E}[Y^2] - 2\mathbb{E}[Y^2] = 0,\end{aligned}$$

having used in the first line that X and Y have the same distribution. So it follows that $X = Y$ almost surely.

(ii) We next claim that $\mathbb{E}[X \wedge a|\mathcal{G}] = Y \wedge a$. almost surely for any $a > 0$. Indeed, by monotonicity of conditional expectation, $\mathbb{E}[X \wedge a|\mathcal{G}] \le \mathbb{E}[X|\mathcal{G}] = Y$ and $\mathbb{E}[X \wedge a|\mathcal{G}] \le \mathbb{E}[a|\mathcal{G}] = a$. So it follows that $\mathbb{E}[X \wedge a|\mathcal{G}] \le Y \wedge a$. Also, since $X \wedge a$ and $Y \wedge a$ have the same disribution, it follows that $\mathbb{E}[\mathbb{E}[X \wedge a|\mathcal{G}]] = \mathbb{E}[X \wedge a] = \mathbb{E}[Y \wedge a]$ and therefore $\mathbb{E}[X \wedge a|\mathcal{G}] = Y \wedge a$. almost surely.

(iii) Similarly one argues that that

$$\mathbb{E}[(X \wedge a) \vee (-a)|\mathcal{G}] = (Y \wedge a) \vee (-a)$$

almost surely for any $a > 0$ and therefore by the L^2-case it follows that $(X \wedge a) \vee (-a) = (Y \wedge a) \vee (-a)$ almost surely. Sending $a \to \infty$ it follows that $X = Y$ almost surely.

(iv) Now consider $T : \Omega \to \Omega$ be a $\mathcal{F}$-measurable map and assume that $\mathbb{P} \circ T = \mathbb{P}$ and that X is such that

$$\mathbb{E}[XZ] = \mathbb{E}[X(Z \circ T)]$$

for any bounded measurable random variable Z. We claim that $X = X \circ T$ $\mathbb{P}$-almost surely. Write $\mathcal{G} = T^{-1}(\mathcal{F})$. Then

$$\mathbb{E}[(X \circ T)(Z \circ T)] = \mathbb{E}[XZ] = \mathbb{E}[X(Z \circ T)] = \mathbb{E}[\mathbb{E}[X(Z \circ T)|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}](Z \circ T)],$$

having used T-invariance of $\mathbb{P}$ in the first equality and that $Z \circ T$ is $\mathcal{G}$-measurable in the last. By considering Z to be characteristic functions and since $X \circ T$ and $\mathbb{E}[X|\mathcal{G}]$ are both $\mathcal{G}$-measurable, it follows that $X \circ T = \mathbb{E}[X|\mathcal{G}]$ almost surely. Also X and $X \circ T$ have the same distribution as T preserves $\mathbb{P}$. Thus it follows from (iii) that $X = X \circ T$ almost surely with respect to $\mathbb{P}$.

6. B8.1 2024 Exam

6.1. **Question 1.** Let $(\Omega, \mathcal{F}, (F_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space.

6.1.1. *Question 1 a).* Let $(X_n)_{n\geq 0}$ be a submartingale and fix $\lambda > 0$.

(i) Recall that a stopping time is a random variable $\tau : \Omega \to \mathbb{N}$ such that for all $n \geq 1$,

$$\{\tau = n\} \in \mathcal{F}_n.$$

The random variable $\tau = \inf\{n \geq 0 : X_n \geq \lambda\}$ is a stopping time since

$$\{\tau = n\} = X_n^{-1}([\lambda, \infty)) \in \mathcal{F}_n$$

since X_n is $\mathcal{F}_n$-measurable by our assumptions.

(ii) Consider now the process $Y_n = (X_n - X_\tau)1_{\{\tau\leq n\}}$ for $n > 0$. We claim that Y_n is a submartingale. To prove this, write $\overline{X}_n = \max_{k\leq n} X_k$ and consider the predictiable process $V_n = 1_{\{\tau\leq n-1\}} = 1_{\{\overline{X}_{n-1}\geq\lambda\}}$. Then it holds that

$$(V \circ X) = \sum_{k=1}^{n} V_k(X_k - X_{k-1}) = X_{n\vee\tau} - X_\tau = (X_n - X_\tau)1_{\{\tau\leq n\}},$$

which is a submartingale as $V \circ X$ always is.

(iii) We now prove Doob's maximal inequality, i.e. that for $n \geq 1$,

$$\lambda\mathbb{P}[\max_{k\leq n} X_k > \lambda] \leq \mathbb{E}[X_n 1_{\{\max_{k\leq n} X_k>\lambda\}}] \leq \mathbb{E}[|X_n|].$$

To prove this, we further note that since $X_\tau \geq \lambda$ we have that $(X_\tau - \lambda)1_{\{\tau\leq n\}}$ is an adapted integrable and nondecreasing process and therefore a submartingale. Thut it follows that $Z_n = (X_n - \lambda)1_{\{\tau\leq n\}} = (X_n - \lambda)1_{\{\max_{k\leq n} X_k\geq\lambda\}}$ is a submartingale and therefore

$$\begin{aligned}
0 &\leq \mathbb{E}[Z_0] \\
&\leq \mathbb{E}[Z_n] \\
&= \mathbb{E}[(X_n - \lambda)1_{\{\max_{k\leq n} X_k\geq\lambda\}}] \\
&= \mathbb{E}[X_n 1_{\{\max_{k\leq n} X_k\geq\lambda\}}] - \lambda\mathbb{P}[\max_{k\leq n} X_k > \lambda],
\end{aligned}$$

showing the first inqueality. The second inequality follows as

$$\mathbb{E}[X_n 1_{\{\max_{k\leq n} X_k>\lambda\}}] \leq \mathbb{E}[|X_n|1_{\{\max_{k\leq n} X_k>\lambda\}}] \leq \mathbb{E}[|X_n|].$$

This concludes the proof of Doob's maximal inequality.

(iv) Now let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables with $\mathbb{E}[\xi_i^2] < \infty$ and $\mathbb{E}[\xi_i] = 0$ for all $i = 1, 2, \ldots$. Let

$$S_0 = 0 \quad \text{and} \quad S_n = \sum_{i=1}^{n} \xi_i, \quad n \geq 1.$$

We claim that S_n^2 is a submartingale. Indeed, by the conditional Jensen inequality since $x \mapsto x^2$ is convex,

$$\begin{aligned}
\mathbb{E}[S_{n+1}^2|\mathcal{F}_n] &= \mathbb{E}[(S_n + \xi_{n+1})^2|\mathcal{F}_n] \\
&\geq (\mathbb{E}[S_n + \xi_{n+1}|\mathcal{F}_n])^2 \\
&= (S_n + \mathbb{E}[\xi_{n+1}])^2 = S_n^2,
\end{aligned}$$

since $\mathbb{E}[\xi_{n+1}] = 0$. Thus it follows by Doob's inequality that

$$\mathbb{P}\left(\max_{0\le k\le n} |S_k| > \lambda\right) \le \lambda^{-2}\mathbb{E}[S_n^2].$$

(v) Assume further that $\sum_{i=1}^\infty \mathbb{E}[\xi_i^2] < \infty$. We claim that S_n converges almost surely to a finite limit. Indeed, we note that S_n is a martingale and is bounded in L^2 since

$$\mathbb{E}[|S_n|] \le \sqrt{\mathbb{E}[S_n^2]} \le \sqrt{\sum_{i=1}^\infty \mathbb{E}[\xi_i^2]}$$

and so by Doob's forward convergence theorem the claim follows.

6.1.2. *Question 1 b).* Let $Z_1, Z_2, \dots$ be a sequence of independent random variables. Fix $\lambda > 0$ and let $\eta_i = Z_i 1_{\{|Z_i|\le\lambda\}}, i = 1, 2, \dots$. Consider the following three conditions:

(I) $\sum_{i=1}^\infty \mathbb{P}(|Z_i| > \lambda) < \infty$,
(II) $\sum_{i=1}^\infty \mathbb{E}[\eta_i]$ converges,
(III) $\sum_{i=1}^\infty \mathrm{var}(\eta_i) < \infty$,

where $\mathrm{var}(\eta_i)$ is the variance of η_i.

(i) We first claim that if conditions (I), (II), and (III) hold, then $\sum_{i=1}^\infty Z_i$ converges a.s. Indeed, consider $\xi_i = \eta_i - \mathbb{E}[\eta_i]$. Then $\mathbb{E}[\xi_i] = 0$ and $\sum_{i=1}^\infty \mathbb{E}[\xi_i^2] < \infty$ by (III). Thus $S_n = \sum_{i=1}^n \xi_i$ converges and therefore by (II)

$$\sum_{i=1}^\infty \eta_i = S_n + \sum_{i=1}^n \mathbb{E}[\eta_i]$$

also converges almost surely. Finally, by the first Borel-Cantelli Lemma it follows from (I) that $Z_i = \eta_i$ for sufficiently large i almost surely. Thus it follows that $\sum_{i=1}^\infty Z_i$ converges from the observation that $\sum_{i=1}^\infty \eta_i$ does.

(ii) Now we suppose that $\sum_{i=1}^\infty Z_i$ converges a.s. Then we claim that (I) holds. Indeed, assume for a contradiction that (I) does not hold. Then by the second Borel-Cantelli lemma, $|Z_i| > \lambda$ infinitely often almost surely, which is not possible if $\sum_{i=1}^\infty Z_i$ converges.

(iii) Again assume that $\sum_{i=1}^\infty Z_i$ converges a.s. and that (III) holds. Then we show that (II) holds. Indeed, by (ii) we have that (I) holds and so by the argument of (i) we have that $\sum_{i=1}^n \xi_i$ converges almost surely. Also $\sum_{i=1}^n \eta_i$ converges almost surely since $\sum_{i=1}^\infty Z_i$ converges a.s. and $\eta_i = Z_i$ for sufficiently large i almost surely. Thus we conclude that

$$\sum_{i=1}^n \mathbb{E}[\eta_i] = \sum_{i=1}^n \eta_i - \sum_{i=1}^n \xi_i$$

converges.

6.2. **Question 2.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

6.2.1. *Question 2 a).* Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra.

(i) Let X be an integrable random variable. Then the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ is the unique $\mathcal{G}$-measurable integrable random variable such that

$$\mathbb{E}[X1_B] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]1_B]$$

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

(ii) For $A \in \mathcal{F}$ we write $\mathbb{P}[A|\mathcal{G}] = \mathbb{E}[1_A|\mathcal{G}]$. Then for sets $A, B \in \mathcal{F}$ we have

$$\mathbb{P}[A|\sigma(B)] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} 1_B + \frac{\mathbb{P}[A \cap B^c]}{\mathbb{P}[B^c]} 1_{B^c}.$$

Indeed, this follows as the left hand side clearly satisfies the uniqueness property written above. For example,

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} 1_B + \frac{\mathbb{P}[A \cap B^c]}{\mathbb{P}[B^c]} 1_{B^c}\right) 1_B\right] &= \mathbb{E}\left[\frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} 1_B\right] \\ &= \mathbb{P}[A \cap B] = \mathbb{E}[1_A 1_B].\end{aligned}$$

The same holds for B^c and hence the claim follows by uniqueness of conditional expectation.

(iii) It follows by monotonicity and linearity of conditional expectation that $0 \le \mathbb{P}[A|\mathcal{G}] \le 1$ and $\mathbb{P}[\emptyset|\mathcal{G}] = 0$ and $\mathbb{P}[\Omega|\mathcal{G}] = 1$.

(iv) Let $A_1, A_2, \ldots$ be disjoint events in $\mathcal{F}$. Then almost surely

$$\begin{aligned}\lim_{n\to\infty} \sum_{i=1}^n \mathbb{P}[A_i|\mathcal{G}] &= \lim_{n\to\infty} \sum_{i=1}^n \mathbb{E}[1_{A_i}|\mathcal{G}] \\ &= \lim_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^n 1_{A_i}\middle|\mathcal{G}\right] \\ &= \mathbb{E}\left[\sum_{i=1}^\infty 1_{A_i}\middle|\mathcal{G}\right] = \mathbb{P}\left[\bigcup_{i\ge 1} A_i\middle|\mathcal{G}\right],\end{aligned}$$

having used that the sets are disjoint in the second line and conditional monotone convergence in the third.

(v) We note that this does not suffice to conclude that $A \mapsto \mathbb{P}[A|\mathcal{G}](\omega)$ is a probability measure since we have no control on the null set of the various unions.

6.2.2. *Question 2 b).* Suppose that $\mathbb{Q}$ is another probability measure on $\mathcal{F}$ that is absolutely continuous with respect to $\mathbb{P}$, that is if $\mathbb{P}[A] = 0$ for any $A \in \mathcal{F}$ then $\mathbb{Q}[A] = 0$.

(i) The Radon-Nikodym theorem states that under these assumption there exists a positive function $D^{\mathcal{F}} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that for all $A \in \mathcal{F}$,

$$\mathbb{Q}(A) = \int_A D^{\mathcal{F}}\, d\mathbb{P} = \mathbb{E}_{\mathbb{P}}[D^{\mathcal{F}} 1_A].$$

(ii) Now let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Then let $D^{\mathcal{G}}$ be the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$ on $\mathcal{G}$. We claim that

$$D^{\mathcal{G}} = \mathbb{E}[D^{\mathcal{F}}|\mathcal{G}].$$

Indeed to show this we note that for $B \in \mathcal{G}$,

$$\mathbb{E}_{\mathbb{P}}[D^{\mathcal{F}} 1_B] = \mathbb{Q}(B) = \mathbb{E}_{\mathbb{P}}[D^{\mathcal{G}} 1_B].$$

Therefore the unique characterisation of conditional expectation is verified and the claim follows.

(iii) Now let $(\mathcal{F}_n)_{n\ge1}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathcal{F}_\infty = \sigma(\bigcup_{n\ge1} \mathcal{F}_n) = \mathcal{F}$. Assume that $\mathbb{Q}$ is absolutely continuous to $\mathbb{P}$ on each $\mathcal{F}_n$, but we no longer assume that this holds for $\mathcal{F}$. Denote by $D_n = D^{\mathcal{F}_n}$.

We now claim that that $\mathbb{Q}$ is absolutely continuous to $\mathbb{P}$ on $\mathcal{F}$ is equivalent to D_n being uniformly integrable. We note that D_n is a martingale since by (ii) we have that

$$\mathbb{E}[D_{n+1}|\mathcal{F}_n] = D_n.$$

We recall the following result from the lecture notes:

Theorem 6.1. *(Theorem 8.32) Let $(M_n)_{n\geq 1}$ be a martingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$. Then $(M_n)_{n\geq 1}$ is uniformly integrable if and only if there is an $\mathcal{F}_\infty$-measurable random variable M_∞ such that*

$$M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]$$

almost surely.

Indeed, if D_n is uniformly integrable, there exists a limit D such that $D_n \to D$ in L^1. So if $A \in \mathcal{F}$, then

$$\mathbb{Q}[A] = \mathbb{E}_{\mathbb{P}}[D1_A] = \lim_{n\to\infty} \mathbb{E}_{\mathbb{P}}[D_n 1_A].$$

Thus if $A \in \bigcup_{n\geq 1} \mathcal{F}_n$ and $\mathbb{P}[A] = 0$, then the same holds for $\mathbb{Q}[A]$. Since the collection of sets with $\mathbb{P}[A] = 0$ is a π-system, it follows that the same holds for all sets in $\mathcal{F}$. Converesely, if $\mathbb{Q}$ is absolutely continuous to $\mathbb{P}$ on $\mathcal{F}$, there exists a density such that by (ii) $D_n = \mathbb{E}[D|\mathcal{F}_n]$ and D_n is uniformly integrable by the above result.

(iv) Consider $\sigma(\tau)$ the σ-algebra generated by the sets $\{\tau = n\}$ and

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty \,:\, A \cap \{\tau = n\} \in \mathcal{F}_n \text{ for all } n \geq 1\}.$$

Since $\{\tau = n\}$ is a partition, it is obvious that $\sigma(\tau) \subset \mathcal{F}_\tau$. To give an easy example where the inequality is strict, we can simply consider the constant stopping time $\tau \equiv 1$. Then $\sigma(\tau) = \{\emptyset, \Omega\}$ is the trivial σ-algebra and $F_\tau = \mathcal{F}_1$. So we simply take a filtration where $\mathcal{F}_1$ is not trivial, as for example the simple random walk.

6.3. **Question 3.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

6.3.1. *Question 3 a).* (i) Let $A \subset \Omega$. Then we claim that

$$\sigma(\mathcal{F}, A) = \{(B \cap A) \cup (C \cap A^c) \,:\, B, C \in \mathcal{F}\}.$$

It is clear that $\supset$ holds. For the other direction denote the left hand side by $\mathcal{H}$. If $D \in \mathcal{F}$, then $D \in \mathcal{F}$ since $D = (D \cap A) \cup (D \cap A^c)$. Also $A \in \mathcal{H}$. So it suffices to show that $\mathcal{H}$ is a σ-algebra. We next show that $\mathcal{H}$ is closed under complements so let $E = (B \cap A) \cup (C \cap A^c)$ be a set in $\mathcal{H}$ with $B, C \in \mathcal{F}$. Then

$$E^c = (B \cap A)^c \cap (C \cap A^c)^c = (B^c \cup A^c) \cap (C^c \cup A) = (B^c \cap A) \cup (C^c \cap A^c)$$

so $\mathcal{H}$ is indeed closed under complements. Finally we show that it closed under countable untions so let $E_n = (B_n \cap A) \cup (C_n \cap A^c)$ be a sequence of sets in $\mathcal{H}$ with $B_n, C_n \in \mathcal{F}$. Then

$$\bigcup_{n\geq 1} E_n = (\cup_{n\geq 1} B_n \cap A) \cup (\cup_{n\geq 1} C_n \cap A^c)$$

so $\bigcup_{n\geq 1} \in \mathcal{H}$ since $\mathcal{F}$ is a σ-algebra. Thus we have shown that $\mathcal{H}$ is indeed a σ-algebra and hence the claim follows.

(ii) Let X, Y be two independent random variables uniformly distributed on $[0, 1]$ and write $U = \min\{X, Y\}$ and $V = \max\{X, Y\}$.

Note that for $0 \le u \le v \le 1$,

$$\begin{aligned}\mathbb{P}[U \le u, V \le v] &= 2\mathbb{P}[X \le u, X \le Y \le v] \\ &= 2\int_0^u \int_x^v dydx \\ &= 2u(v-u) = 2uv - v^2.\end{aligned}$$

So the joint density satisfies

$$f_{(U,V)}(u,v) = \frac{\partial^2}{\partial u \partial v}\mathbb{P}[U \le u, V \le v] = 2 \quad \text{on} \quad \{(u,v) \in [0,1]^2 \,:\, u \le v\}.$$

Thus it follows that

$$\mathbb{E}[U|\sigma(V)] = \frac{\int u f_{(U,V)}(u,V)\,du}{\int f_{(U,V)}(u,V)\,du} = \frac{\int_0^V 2u\,du}{\int_0^V 2\,du} = \frac{V}{2}$$

and

$$\mathbb{E}[V|\sigma(U)] = \frac{\int v f_{(U,V)}(U,v)\,dv}{\int f_{(U,V)}(U,v)\,dv} = \frac{\int_U^1 2v\,dv}{\int_U^1 2\,dv} = \frac{1-U^2}{2(1-U)} = \frac{1+U}{2}.$$

6.3.2. *Question 3 b).* Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with $\mathbb{P}(X_1 = j) = p_j > 0$, $j = 0,1,2,\ldots$, $\sum_{j=0}^\infty p_j = 1$. The sequence is revealed one at a time, so that by time n, the values of $X_1, \ldots, X_n$ are known. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ with $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

(i) Let τ be the first time that the pattern $1,1,5,7,1,1$ is observed. It is clear that τ is a $(\mathcal{F}_n)_{n\ge0}$ stopping time and we claim that $\mathbb{P}[\tau] < \infty$. To show this we prove that for $n \ge 1$,

$$\mathbb{E}[\tau > 6n] \le e^{-cn}$$

for some small constant $c > 0$. Indeed, for $k \ge 1$ consider the event

$$A_k = (X_{6k}, X_{6k+1}, X_{6k+2}, X_{6k+3}, X_{6k+4}, X_{6k+5}) = (1,1,5,7,1,1).$$

Then the events $(A_k)_{k\ge1}$ are independent since the X_i are and it holds that $\mathbb{P}[A_k] = p_1^4 p_5 p_7 > 0$. Thus we have that

$$\mathbb{P}[\tau > 6n] \le \mathbb{P}[A_1^c \cap A_2^c \cap \ldots \cap A_n^c] \le (1 - \mathbb{P}[A_1])^n \le e^{-cn}$$

for a sufficiently small constant $c > 0$. So we conclude that

$$\mathbb{E}[\tau] = \sum_{n\ge0} \mathbb{P}[\tau \ge n] \le 6 \sum_{n\ge0} \mathbb{P}[\tau \ge 6n] < \infty.$$

(ii) Consider a casino which offers fair bets according to the sequence (X_n). Specifically, if a gambler bets stake a on the outcome of the n^{th} bet being j, i.e., on $X_n = j$, then they lose their stake with probability $1 - p_j$ or, with probability p_j, they get their stake a back and win $\frac{a(1-p_j)}{p_j}$ more (so in total the player receives $\frac{a}{p_j}$). Consider a sequence of gamblers: they all start with a capital of 1 and the i^{th} gambler starts betting at time i on the pattern $(1,1,5,7,1,1)$ at subsequent rounds they bet all of their capital until they either see the sequence and retire, or they lose earlier and are out.

For $i \geq 1$, let Y_n^i denote the capital of the i^{th} gambler after round n, $n \geq 0$. We claim that $(Y_n^i)_{n\geq 0}$ is a $(\mathcal{F}_n)_{n\geq 0}$-martingale. First note that $Y_n^i = 1$ for $0 \leq n < i$. Then it holds that

$$Y_i^i = \frac{1}{p_i} 1_{X_i = i}, \qquad Y_{i+1}^i = \frac{Y_i^i}{p_1} 1_{X_{i+1}=1}, \qquad Y_{i+2}^i = \frac{Y_{i+1}^i}{p_5} 1_{X_{i+2}=5},$$

etc. and finally $Y_n^i = Y_{i+5}^i$ for all $n > i+5$.

Indeed, this follows as the game is fair. More precisely, if $i - 1 \leq n \leq i + 4$,

$$\mathbb{E}[Y_{n+1}^i | \mathcal{F}_n] = \mathbb{E}\left[\frac{Y_n^i}{p_j} 1_{X_{n+1}=j} | \mathcal{F}_n\right] = \frac{Y_n^i}{p_j} \mathbb{E}[1_{X_{n+1}=j} | \mathcal{F}_n] = Y_n^i.$$

Thus concludes the claim since for the other n we ave $Y_{n+1}^i = Y_n^i$.

(iii) Assuming the above gamblers are the only players in the casino. Denote by M_n the total winning of the casino's profit and loss by time n. Then it holds that

$$M_n = n - (Y_n^n + Y_n^{n-1} + Y_n^{n-2} + Y_n^{n-3} + Y_n^{n-4} + Y_n^{n-5}).$$

M_n is a sum of martingales and therefore a martingale it self. Then it holds by the optional stopping theorem since $|M_{n+1} - M_n|$ is uniformly bounded as at most 6 players are in the casino at the same time and as $\mathbb{E}[\tau] < \infty$ that

$$\mathbb{E}[\tau] - (Y_\tau^\tau + Y_\tau^{\tau-1} + Y_\tau^{\tau-2} + Y_\tau^{\tau-3} + Y_\tau^{\tau-4} + Y_\tau^{\tau-5}) = \mathbb{E}[M_\tau] = \mathbb{E}[M_0] = 0$$

At time τ we have that

$$Y_\tau^{\tau-5} = \frac{1}{p_1^4 p_5 p_7}, \qquad Y_\tau^{\tau-1} = \frac{1}{p_1^2}, \qquad Y_\tau^\tau = \frac{1}{p_1}$$

and $Y_\tau^{\tau-4} = 0 = Y_\tau^{\tau-3} = Y_\tau^{\tau-2}$. So it follows that

$$\mathbb{E}[\tau] = \frac{1}{p_1} + \frac{1}{p_1^2} + \frac{1}{p_1^4 p_5 p_7}.$$

7. B8.2 Class 1

7.1. **Consequences of the Monotone Class Theorem.** Recall the following result.

Definition 7.1. *Let Ω be a set. A collection of subsets $\mathscr{M} \subset \mathscr{P}(\Omega)$ is called a monotone class if the following properties hold:*

(1) $\Omega \in \mathscr{M}$.
(2) *If* $A, B \in \mathscr{M}$ *and* $A \subset B$, *then* $B \backslash A \in \mathscr{M}$.
(3) *If* $(A_n)_{n\geq 1}$ *is an increasing sequence subsets of* Ω *with* $A_n \in \mathscr{M}$ *for all* $n \geq 1$, *then*
$$\bigcup_{n\geq 1} A_n \in \mathscr{M}.$$

Theorem 7.2. *(Monotone Class Theorem) If $\mathscr{C} \subset \mathscr{P}(\Omega)$ is stable under finite intersections and $\mathscr{C} \subset \mathscr{M}$ for $\mathscr{M}$ a monotone class, then $\sigma(\mathscr{C}) \subset \mathscr{M}$.*

Lemma 7.3. *Let $(X_i)_{i\in I}$ be a collection of random variables on $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\mathscr{H} = \sigma(X_i : i \in I)$ and let $\mathscr{G} \subset \mathscr{F}$. Then $\mathscr{G}$ is independent of $\mathscr{H}$ if and only if $\mathscr{G}$ is independent of $\mathscr{H}' = \sigma(X_i : i \in I')$ for every finite $I' \subset I$.*

Proof. It is clear that if $\mathscr{G}$ and $\mathscr{H}$ are independent, then so is $\mathscr{G}$ and $\mathscr{H}'$ for every finite $I' \subset I$. So assume that $\mathscr{G}$ and $\mathscr{H}'$ are independent for every finite $I' \subset I$.

Consider $\mathscr{M} = \{A \in \mathscr{F} : A \text{ is independent from } \mathscr{G}\}$. Then one readily checks (using the monotone convergence theorem) that $\mathscr{M}$ is a monotone class. We note that by assumption $\sigma(X_i) \subset \mathscr{M}$ for all $i \in I$. Thus consider $\mathscr{C}$ to be the class of events that depend on only finitely many X_i, i.e.
$$\mathscr{C} = \bigcup_{I'\subset I \text{ finite}} \sigma(X_i : i \in I').$$
Then $\mathscr{C} \subset \mathscr{M}$ by assumption and it is closed under finite intersections. Therefore by the monotone class theorem $\sigma(X_i : i \in I) \subset \sigma(\mathscr{C}) \subset \mathscr{M}$ and the claim follows. □

Lemma 7.4. *Let $(X_i)_{i\in I}$ be a collection of random variables on $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\mathscr{H} = \sigma(X_i : i \in I)$ and let $\mathscr{G} \subset \mathscr{F}$ and let Y be a bounded random variable. Then $\mathbb{E}[Y|\sigma(X_{i_0})] = \mathbb{E}[Y|\mathscr{H}]$ if $\mathbb{E}[Y|\sigma(X_{i_0})] = \mathbb{E}[Y|\sigma(X_{i_1}, \dots, X_{i_n})]$ for every finite set of indices $\{i_1, \dots, i_k\} \subset I$.*

Proof. Clearly $\mathbb{E}[Y|\sigma(X_{i_0})]$ is $\mathscr{H}$-measurable. The collection $\mathscr{M}$ of sets $A \in \mathscr{F}$ such that $E[1_A Y] = \mathbb{E}[1_A \mathbb{E}[Y|\sigma(X_{i_0})]]$ is a monotone class. To check that it is closed under monotone limits, one uses the dominated convergence theorem. As before, we note that the π-system $\mathscr{C}$ of events which depend on finitely many X_i is contained in $\mathscr{M}$. Thus the claim follows by the monotone class theorem. □

Lemma 7.5. *Let $(X_i)_{i\in I}$ be a collection of random variables on $(\Omega, \mathscr{F}, \mathbb{P})$. Assume that $I = \bigcup_{i=1}^n I_i$ and denote by $\mathscr{H}_i = \sigma(X_s : s \in I_i)$ for $1 \leq i \leq n$. Then $\mathscr{H}_1, \dots, \mathscr{H}_n$ are independent if and only if for every finite collection of subsets $I_i' \subset I$ with $1 \leq i \leq n$ the collection if σ-algebras $(\mathscr{H}_i')_{1\leq i\leq n}$ with $\mathscr{H}_i' = \sigma(X_s : s \in I_i')$ is independent.*

Proof. We proceed by induction. For $n = 1$, there is nothing to show. Assuming the inductive hypothesis, it follows that $\mathscr{H}_2, \dots, \mathscr{H}_n$ are independent. Denote $\mathscr{H}' = \sigma(\mathscr{H}_2, \dots, \mathscr{H}_n) = \sigma(X_s : s \in I_2 \cup \ldots \cup I_n)$.

To conclude the lemma, we show that $\mathscr{H}_1$ and $\mathscr{H}'$ are independent. It follows from Lemma 7.3 that $\sigma(X_i : i \in I_1')$ for a finite subset $I_1' \subset I_1$ is independent from $\mathscr{H}'$. Let $\mathscr{M}$ be the monotone class of sets in $\mathscr{F}$ independent from $\mathscr{H}'$. Moreover, consider $\mathscr{C}$ to be the collection of events that only depend on finitely many X_i with $i \in I_1$. Then it follows that $\mathscr{C} \subset \mathscr{M}$ and it is clear that $\mathscr{C}$ is closed under finite intersections. Therefore, by the monotone class theorem, it follows that $\mathscr{H}_1 \subset \sigma(\mathscr{C}) \subset \mathscr{M}$ and therefore $\mathscr{H}_1$ and $\mathscr{H}'$ are independent. □

As another application of the monotone class theorem, we can prove the following lemma.

Lemma 7.6. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{H}_1, \dots, \mathscr{H}_n$ be independent σ-algebras. Then $\mathscr{H}_1$ is independent from $\sigma(\mathscr{H}_2, \dots, \mathscr{H}_n)$.*

Proof. Let $\mathscr{M} = \{A \in \mathscr{F} : A \text{ is independent from } \mathscr{H}_1\}$ be the sets independent of $\mathscr{H}_1$. Then as in Lemma 7.3, $\mathscr{M}$ is a monotone class. Consider

$$\mathscr{C} = \{B_2 \cap \ldots \cap B_n : B_i \in \mathscr{H}_i \text{ for } 2 \le i \le n\}.$$

We note that $\mathscr{C} \subset \mathscr{M}$. Indeed, let $B_i \in \mathscr{H}_i$ for $2 \le i \le n$. Then since $\mathscr{H}_1, \dots, \mathscr{H}_n$ are independent, of any $A \in \mathscr{H}_1$,

$$\mathbb{P}[A \cap B_2 \cap \ldots \cap B_n] = \mathbb{P}[A]\mathbb{P}[B_2] \cdots \mathbb{P}[B_n] = \mathbb{P}[A]\mathbb{P}[B_2 \cap \ldots \cap B_n] \tag{7.1}$$

Thus $B_2 \cap \ldots \cap B_n \in \mathscr{M}$. Moreover, $\mathscr{C}$ is stable under finite intersections and $\mathscr{H}_2, \dots, \mathscr{H}_n \subset \mathscr{C}$. Thus it follows by the monotone class theorem that $\sigma(\mathscr{C}) = \sigma(\mathscr{H}_2, \dots, \mathscr{H}_n) \subset \mathscr{M}$, implying the claim. □

Corollary 7.7. *Let $X_1, X_2 \dots, X_n$ be independent random variables, i.e. the σ-algebras $\sigma(X_1), \sigma(X_2), \dots, \sigma(X_n)$ are independent. Then $\sigma(X_1)$ is independent from $\sigma(X_2, \dots X_n)$.*

Proof. Follows directly from Lemma 7.6. □

7.2. **Brownian Motion and its properties.**

Definition 7.8. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A collection of σ-algebras $(\mathscr{F}_t)_{t \ge 0}$ with t ranging in $[0, \infty)$ and $\mathscr{F}_t \subset \mathscr{F}$ for all $t \in [0, \infty)$ is called a filtration if $\mathscr{F}_t \subset \mathscr{F}_s$ for all $t \le s$.*

Definition 7.9. *A (continuous-time) stochastic process $(M_t)_{t \ge 0}$ on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \ge 0}, \mathbb{P})$ is a collection of random variables $M_t : \Omega \to \mathbb{R}$ such that M_t is $\mathscr{F}_t$-measurable for all $t \ge 0$.*

Definition 7.10. *A stochastic process $(B_t)_{t \ge 0}$ is called a Brownian motion if there is some constant $\sigma > 0$, such that*

(i) (Zero at zero) $B_0 = 0$.
(ii) (Normally distributed) For each $s \ge 0$ and $t > 0$, the random variable $B_{s+t} - B_s$ is normally distributed with mean zero and variance $\sigma^2 t$.
(iii) (Independence of increments) For each $n \ge 1$ and any times $0 \le t_0 \le t_1 \le \ldots \le t_n$ the random variables $B_{t_0}, B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}$ are independent.
(iv) (Continuity) B_t is continuous in $t \ge 0$.

When $\sigma^2 = 1$, we say that we have a standard Brownian motion.

Lemma 7.11. *Suppose that $(B_t)_{t \ge 0}$ is a Brownian motion. Then $(-B_t)_{t \ge 0}$ and $(cB_{t/c^2})_{t \ge 0}$ for any $c > 0$ are Brownian motions as well.*

Proof. The first claim is obvious. For the second claim, denote $M_t = cB_{t/c^2}$. Then $M_0 = 0$ and M_t is continuous. Also $M_{s+t} - M_s = c(B_{(s+t)/c^2} - B_{s/c^2})$ is normally distributed with variance $\sigma^2 t$. Finally the independence follow similarly. □

Lemma 7.12. *It holds that* $\lim_{n\to\infty} \frac{1}{n} B_n = 0$ *almost surely and in* L^1.

Proof. Denote $X_i = B_i - B_{i-1}$ for $i \geq 1$. Then we can express $B_n = \sum_{i=1}^n X_i$ as a sum of independent random variables of mean 0. Thus by the law of large numbers, $\frac{1}{n} B_n \to \mathbb{E}[X_i] = 0$ almost surely and in L^1. □

We now aim to show that $\lim_{t\to\infty} \frac{1}{t} B_t = 0$. In order to do so, we need the following preliminary calculation.

Lemma 7.13. *Let X be a normally distributed random variable with mean* 0 *and variance* 1*. Then it holds for $x > 0$,*

$$\mathbb{P}[X \geq x] \leq \frac{e^{-x^2/2}}{\sqrt{2\pi}x}.$$

Proof. Applying integration by parts,

$$\begin{aligned}
\mathbb{P}[X \geq x] &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2}\, dy \\
&= \frac{1}{\sqrt{2\pi}} \int_x^\infty y^{-1} \cdot (ye^{-y^2/2})\, dy \\
&= \left[-\frac{e^{-y^2/2}}{\sqrt{2\pi}y} \right]_x^\infty - \frac{1}{\sqrt{2\pi}} \int_x^\infty y^{-2} \cdot e^{-y^2/2}\, dy \\
&\leq \frac{e^{-x^2/2}}{\sqrt{2\pi}x}.
\end{aligned}$$

□

Lemma 7.14. *It holds that* $\sup_{0\leq t\leq r} B_t$ *is distributed as* $|B_r|$.

Proof. This is proved in section 5 of the lecture notes. □

Lemma 7.15. *It holds that* $\lim_{t\to\infty} \frac{1}{t} B_t = 0$ *almost surely.*

Proof. We first show that

$$\lim_{n\to\infty} \sup_{t\in(0,1)} \frac{B_{n+t} - B_n}{n} = 0$$

almost surely. Indeed, denote $M_n = \sup_{t\in(0,1)} B_{n+t} - B_n$. Consider the event $A_n = \{M_n > \sqrt{n}\}$. Moreover, by Lemma 7.14, M_n is distributed like $|B_1|$. Thus by Lemma 7.13,

$$\mathbb{P}[A_n] = \mathbb{P}[|B_1| \geq \sqrt{n}] \leq \frac{\sqrt{2}}{\sqrt{\pi n}} e^{-n/2}$$

and so

$$\sum_{n\geq 1} \mathbb{P}[A_n] < \infty.$$

Therefore by the first Borel-Cantelli lemma (which does not require independence), almost surely, A_n only occurs finitely many often and thus $\frac{M_n}{n} \leq \frac{1}{\sqrt{n}}$ for large enough n almost surely. This implies the claim.

To show the main claim, we calculate,

$$\begin{aligned}\lim_{t\to\infty}\frac{1}{t}B_t &= \lim_{t\to\infty}\frac{\lfloor t\rfloor}{t}\left(\frac{B_t - B_{\lfloor t\rfloor}}{\lfloor t\rfloor} + \frac{B_{\lfloor t\rfloor}}{\lfloor t\rfloor}\right)\\ &\leq \lim_{t\to\infty}\frac{\lfloor t\rfloor}{t}\left(\sup_{t\in[\lfloor t\rfloor,\lceil t\rceil)}\frac{B_t - B_{\lfloor t\rfloor}}{\lfloor t\rfloor} + \frac{B_{\lfloor t\rfloor}}{\lfloor t\rfloor}\right) = 0\end{aligned}$$

almost surely. This implies that $\lim_{t\to\infty}\frac{B_t}{0} \leq 0$ almost surely. Applying the same conclusion to $-B_t$, concludes the proof. □

By arguments from the appendix, we can further show that the M_n are independent, for which we need to show that $M_0,\dots,M_n$ are independent. The claim follows from Lemma 7.5. Indeed, denote $\mathscr{H}_k = \sigma(B_{k+s} - B_k \,:\, s\in[0,1))$ and observe that it suffices to show that $\mathscr{H}_0,\dots,\mathscr{H}_k$ are independent. By Lemma 7.5 it suffices to show that for a finite collection of times $0\leq t_{k,1}\leq\ldots\leq t_{k,i_k} < 1$ for $0\leq k\leq n$ the σ-algebras $\mathscr{H}_k' = \sigma(B_{k+t_{k,j}} - B_k \,:\, 1\leq j\leq i_k)$ are independent. To see this we note that

$$\mathscr{H}_k' = \sigma(B_{k+t_{k,1}} - B_k, B_{k+t_{k,2}} - B_{k+t_{k,1}},\dots, B_{k+t_{k,i_k}} - B_{k+t_{k,i_k-1}})$$

and the claim follows form the independence of increments.

7.3. **Modifications and Indistinguishable Processes.**

Definition 7.16. *Let $(X)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be two stochastic processes defined on a common probability space. Then we say that:*

*(i) X is a **modification** of Y if for all $t\geq 0$ it holds that $X_t = Y_t$ almost surely.*
*(ii) X is **indistinguishable** from Y if*

$$\mathbb{P}[X_t = Y_t \textit{ for all } t\in\mathbb{R}_{\geq 0}] = 1.$$

Lemma 7.17. *Let $(X)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be two stochastic processes. Assume that X and Y have almost surely right continuous paths and are modifications of each other. Then X and Y are indistinguishable.*

Proof. Since X and Y are indistinguishable and $\mathbb{Q}$ is countable, it follows that

$$\mathbb{P}\left[X_t = Y_t \,:\, t\in\mathbb{Q}_{\geq 0}\right] = 1.$$

The claim then follows by right continuity. □

Lemma 7.18. *Brownian motion is indistinguishable from a γ-locally Hölder continuous process for order γ for every $\gamma\in(0,1/2)$.*

Proof. By the above lemma and since Brownian motion is continuous, it suffices to show that Brownian motion is a modification of a γ-Hölder continuous process. Recall that Kolmogorov's continuity criterion states the following. Assume that a stochastic process $(X_t)_{t\geq 0}$ satisfies

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq C|t-s|^{1+\beta}$$

for all $s,t\geq 0$ and some strictly positive constants $\alpha,\beta,C > 0$. Then X_t has a γ-locally Hölder continuous process for every $\gamma\in(0,\beta/\alpha)$.

To show the claim, we note that for $s, t \geq 0$ the random variable $B_t - B_s$ is normally distributed with mean zero and variance $|t-s|$. Thus $|t-s|^{-1/2}(B_t - B_s)$ is a normally distributed standard Gaussian. Therefore it follows that

$$\mathbb{E}[|B_t - B_s|^n] = \mathbb{E}\left[\frac{|t-s|^{-n/2}}{|t-s|^{-n/2}}|B_t - B_s|^n\right] \leq C_n |t-s|^{n/2},$$

where $C_n = \mathbb{E}\left[|t-s|^{-n/2}|B_t - B_s|^n\right]$, which is finite and strictly positive since a standard Gaussian has moments of arbitrary degree. Thus B_t admits a modification with $\gamma \in (0, \frac{n/2-1}{n})$. The claim follows as $\lim_{n\to\infty} \frac{n/2-1}{n} = 1/2$. □

7.4. **Gaussian Processes.**

Definition 7.19. *A multivariate Gaussian is a multivariate random variable* $\mathbf{X} = (X_1, \ldots, X_n)$ *such that for every* $u \in \mathbb{R}^\ell$ *it holds that*

$$\langle u, X \rangle = \sum_{i=1}^{\ell} u_i X_i$$

is a Gaussian random variable. We say that $\mathbf{X}$ *is centred if* $(\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n]) = 0$.

Multivariate Gaussians have the following very useful property.

Lemma 7.20. *Let* $\mathbf{X} = (X_1, \ldots, X_n)$ *be a multivariate Gaussian. Then the random variables* $X_1, \ldots, X_n$ *are independent if and only if the covariance matrix* $(\Gamma_X)_{ij} = \operatorname{cov}(X_i, X_j)$ *is diagonal.*

Proof. If the random variables are independent, then for $i \neq j$ it holds that $\operatorname{cov}(X_i, X_j) = 0$. For the other direction, we assume without loss of generality that $\mathbf{X}$ is centred. So $\langle u, X \rangle = \sum_{i=1}^n u_i X_i$ is a centred random variable with variance $u^T \Gamma_X u$.

To prove the claim, we consider the characteristic function $u \mapsto \mathbb{E}[e^{i\langle u, X\rangle}]$ and recall that $\mathbf{X}$ is independent whenever $\mathbb{E}[e^{i\langle u,X\rangle}] = \prod_{k=1}^n \mathbb{E}[e^{iu_k X_k}]$. We note that if $Y \sim \mathcal{N}(0, \sigma^2)$, $\mathbb{E}[e^{itY}] = e^{-\sigma^2 t^2/2}$. By assumption, write $\Gamma_X = \operatorname{diag}(\lambda_1^2, \ldots, \lambda_n^2)$ with $\lambda_i^2 = \operatorname{Var}(X_i)$ and therefore $\langle u, X \rangle$ is a centred Gaussian with variance $\sum_{k=1}^n \lambda_k^2 u_k^2$. Thus it follows that

$$\mathbb{E}[e^{i\langle u,X\rangle}] = e^{-\frac{1}{2}\sum_{k=1}^n \lambda_k^{-2} u_k^2} = \prod_{k=1}^n e^{-\lambda_k^2 u_k^2/2} = \prod_{k=1}^n \mathbb{E}[e^{iu_k X_k}].$$

Thus the claim follows. □

It is important to note that it is necessary to assume that the Gaussian is multivariate. It is a classical fallacy in probability theory to assume only that the marginals are Gaussian, as the following example shows.

Example 7.21. *We give an example of two uncorrelated random variables* X *and* Y *that are both distributed as* $\mathcal{N}(0, 1)$*, yet that are not independent.*

Indeed consider X be a $\mathcal{N}(0,1)$-distributed random variable and let Z be an independent uniform $\{\pm 1\}$ variable. Denote $Y = X \cdot Z$ and note $Y \sim \mathcal{N}(0,1)$ since

$$\begin{aligned}\mathbb{P}[Y \le x] &= \frac{1}{2}\mathbb{P}[Y \le x \mid Z = 1] + \frac{1}{2}\mathbb{P}[Y \le x \mid Z = -1] \\ &= \frac{1}{2}\mathbb{P}[X \le x] + \frac{1}{2}\mathbb{P}[-X \le x] \\ &= \frac{1}{2}\left(\Phi(x) + \Phi(-x)\right) = \Phi(x),\end{aligned}$$

where $\Phi(x) = \mathbb{P}[X \le x]$ is the CDF of $\mathcal{N}(0,1)$ and we use that $\Phi(x) = \Phi(-x)$. In addition X and Y are both uncorrelated since

$$\operatorname{cov}(X,Y) = \mathbb{E}[XY] = \mathbb{E}[X^2 Z] = \mathbb{E}[X^2]\mathbb{E}[Z] = 0.$$

However, X and Y are not independent as for $x \le 0$,

$$\mathbb{P}[X \le x, Y \le x] = \mathbb{P}[X \le x, Z = 1] = \frac{1}{2}\Phi(x),$$

which is not equal to $\Phi(x)^2$ for $x < 0$.

Definition 7.22. *A continuous stochastic process $(X_t)_{t \ge 0}$ is called a (centred) Gaussian process if for every finite set $\{t_1, \dots, t_n\} \subset \mathbb{R}_{>0}$ it holds that the random variable*

$$\mathbf{X} = (X_{t_1}, X_{t_2}, \dots, X_{t_n})$$

is a multivariate (centred) Gaussian.

Lemma 7.23. *Let $(X_t)_{t \ge 0}$ be a Gaussian process, let $\{t_1, \dots, t_n\} \subset \mathbb{R}_{>0}$ be a finite set and write $\mathbf{X} = (X_{t_1}, X_{t_2}, \dots, X_{t_n})$. Let $B \in M_{k,n}(\mathbb{R})$ be a matrix. Then the random vector $B\mathbf{X}$ is a multivariate Gaussian.*

Proof. Let $u \in \mathbb{R}^k$. Then $\langle u, B\mathbf{X}\rangle = \langle B^T u, \mathbf{X}\rangle$ and therefore it follows that $B\mathbf{X}$ is a multivariate Gaussian. □

Lemma 7.24. *Brownian motion is a Gaussian process.*

Proof. Let $\{t_1, \dots, t_n\} \subset \mathbb{R}_{>0}$ be a finite set, write $\mathbf{X} = (B_{t_1}, B_{t_2}, \dots, B_{t_n})$ and let $u \in \mathbb{R}^n$. We want to show that

$$\langle u, \mathbf{X}\rangle = \sum_{i=1}^{n} u_i B_{t_i}$$

is a Gaussian random variable. Without loss of generality we assume that $t_1 \le \dots \le t_n$. Notice that

$$\begin{aligned}\langle u, \mathbf{X}\rangle = u_n(B_{t_n} - B_{t_{n-1}}) + (u_{n-1} - u_n)(B_{t_{n-1}} - B_{t_{n-2}}) + \dots \\ \dots + c_{n-1}(B_{t_2} - B_{t_1}) + c_n B_{t_1}\end{aligned}$$

for suitable constants c_k. Thus $\langle u, \mathbf{X}\rangle$ is a sum of independent Gaussians and therefore a Gaussian itself. □

Lemma 7.25. *(Characterization of Brownian motion) Let $(X_t)_{t \ge 0}$ be a continuous centred Gaussian process with $X_0 = 0$. Assume that $\operatorname{cov}(X_t, X_s) = \min\{t, s\}$ for $t, s \in \mathbb{R}_{\ge 0}$. Then X_t is a standard Brownian motion.*

Proof. We first show that for each $n \geq 1$ and any times $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$ the random variables $B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent. Indeed, write $\mathbf{X} = (B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}})$ and we calculate the covariance matrix. Indeed, notice that for $1 \leq i < j$,

$$\begin{aligned}\operatorname{cov}(B_{t_i} - B_{t_{i-1}}, B_{t_j} - B_{t_{j-1}}) &= \operatorname{cov}(B_{t_i}, B_{t_j}) - \operatorname{cov}(B_{t_{i-1}}, B_{t_j}) \\ &\quad - \operatorname{cov}(B_{t_i}, B_{t_{j-1}}) + \operatorname{cov}(B_{t_{i-1}}, B_{t_{j-1}}) \\ &= t_i - t_{i-1} - t_i + t_{i-1} = 0.\end{aligned}$$

Moreover if $j \geq 1$,

$$\operatorname{cov}(B_{t_0}, B_{t_j} - B_{t_{j-1}}) = \operatorname{cov}(B_{t_0}, B_{t_j}) - \operatorname{cov}(B_{t_0}, B_{t_{j-1}}) = t_0 - t_0 = 0.$$

Thus we have shown that the covariance matrix of $\mathbf{X}$ is zero and therefore by Lemma 7.20 it follows that $B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent.

Moreover, by Lemma 7.23, $X_{s+t} - X_s$ is a centred Gaussian and

$$\begin{aligned}\operatorname{var}(X_{s+t} - X_s) &= \operatorname{cov}(X_{s+t} - X_s, X_{s+t} - X_s) \\ &= \operatorname{cov}(X_{s+t}, X_{s+t}) - 2\operatorname{cov}(X_{s+t}, X_s) + \operatorname{cov}(X_s, X_s) \\ &= s + t - 2s + s = t.\end{aligned}$$

This implies the claim. □

7.5. **Applications of Gaussian Processes.**

Proposition 7.26. *The stochastic process $(tB_{1/t})_{t\geq 0}$ is a Brownian motion.*

Proof. Write $X_t = tB_{1/t}$. Since $(0, \infty) \ni t \mapsto \frac{1}{t}$ is continuous, it follows that $t \mapsto X_t$ is almost surely continuous on $(0, \infty)$. By the Lemma 7.15, it holds that $\lim_{t\to 0} X_t = 0$, so the process is also continuous at zero and zero at zero.

It is clear that X_t is a centred Gaussian process and note that

$$\operatorname{cov}(X_t, X_s) = ts \min\{\tfrac{1}{s}, \tfrac{1}{t}\} = \min\{s, t\}.$$

Therefore by Lemma 7.25 X_t is a Brownian motion. □

For the purposes of the next lemma, we define the conditional variance of a random variable X defined on $(\Omega, \mathscr{F}, \mathbb{P})$ with respect to a σ-algebra $\mathscr{A} \subset \mathscr{F}$ as

$$\operatorname{Var}(X|\mathscr{A}) = \mathbb{E}[X^2 \mid \mathscr{A}] - \mathbb{E}[X|\mathscr{A}]^2.$$

Lemma 7.27. *Let $(B_t)_{t\geq 0}$ be a Brownian motion. Fix $0 \leq s < t < \infty$. Then conditionally on B_s and B_t it holds that $B_{\frac{s+t}{2}}$ is normally distributed with mean $\frac{1}{2}(B_s + B_t)$ and variance $\frac{1}{2}(t - s)$.*

Proof. Consider the random variable

$$Y = B_{\frac{s+t}{2}} - \frac{1}{2}(B_s + B_t) = \frac{1}{2}(B_{\frac{s+t}{2}} - B_s) + \frac{1}{2}(B_{\frac{s+t}{2}} - B_t).$$

Note that Y is a sum of independent mean-zero Gaussian. Therefore it is a Gaussian itself, $\mathbb{E}[Y] = 0$ and

$$\begin{aligned}\operatorname{Var}(Y) &= \frac{1}{4}\operatorname{Var}(B_{\frac{s+t}{2}} - B_s) + \frac{1}{4}\operatorname{Var}(B_{\frac{s+t}{2}} - B_t) \\ &= \frac{1}{4}\left(\frac{s+t}{2} - s + t - \frac{s+t}{2}\right) = \frac{t-s}{4}.\end{aligned}$$

We show below that Y is independent from B_s and B_t. Assume for the moment, that this is the case. Then the distribution of Y conditioned on B_s and B_t is the same as the distribution of Y, which is a Gaussian with mean zero and variance $\frac{t-s}{4}$. Then $\mathbb{E}[Y|B_s, B_t] = \mathbb{E}[Y] = 0$ and thus

$$\mathbb{E}[B_{\frac{s+t}{2}}|B_s, B_t] = \mathbb{E}\left[\frac{1}{2}(B_s + B_t) \,|\, B_s, B_t\right] = \frac{B_s + B_t}{2}.$$

Similarly, $\operatorname{Var}(Y|B_s, B_t) = \operatorname{Var}(Y) = \frac{t-s}{2}$ and therefore

$$\begin{aligned}\operatorname{Var}(B_{\frac{s+t}{2}}|B_s, B_t) &= \operatorname{Var}(B_{\frac{s+t}{2}} - \mathbb{E}[B_{\frac{s+t}{2}}|B_s, B_t]|B_s, B_t)\\ &= \operatorname{Var}(B_{\frac{s+t}{2}} - \frac{1}{2}(B_s + B_t)|B_s, B_t)\\ &= \operatorname{Var}(Y|B_s, B_t)\\ &= \frac{t-s}{2}\end{aligned}$$

showing the claim.

It remains to show that Y is independent of B_s and B_t. We first show that Y and B_s are independent. Indeed by Lemma 7.23, (Y, B_s) is a multivariate Gaussian. Thus by Lemma 7.20 it suffices to show that $\operatorname{cov}(Y, B_s) = 0$. Note first that for $r_1 \neq r_2$ it holds that $\operatorname{cov}(B_{r_1}, B_{r_2}) = \min\{r_1, r_2\}$. Indeed, assuming without loss of generality that $r_1 \leq r_2$,

$$\begin{aligned}\operatorname{cov}(B_{r_1}, B_{r_2}) &= \operatorname{cov}(B_{r_1}, B_{r_2} - B_{r_1} + B_{r_1})\\ &= \operatorname{cov}(B_{r_1}, B_{r_2} - B_{r_1}) + \operatorname{cov}(B_{r_1}, B_{r_1}) = 0 + \operatorname{Var}(B_{r_1}) = r_1.\end{aligned}$$

We calculate

$$\operatorname{cov}(Y, B_s) = \operatorname{cov}(B_{\frac{s+t}{2}}, B_s) - \frac{1}{2}\left(\operatorname{cov}(B_s, B_s) + \operatorname{cov}(B_t, B_s)\right) \quad = s - \frac{1}{2}(s+s) = 0.$$

Similarly,

$$\begin{aligned}\operatorname{cov}(Y, B_t) &= \operatorname{cov}(B_{\frac{s+t}{2}}, B_t) - \frac{1}{2}\left(\operatorname{cov}(B_s, B_t) + \operatorname{cov}(B_t, B_t)\right)\\ &= \frac{s+t}{2} - \frac{1}{2}(s+t) = 0.\end{aligned}$$

In particular, $Y, B_s - B_t$ and B_t is a collection of independent random variables as their covariance matrix is zero. Thus by Corollary 7.7, it follows that Y is independent of $\sigma(B_s - B_t, B_t) = \sigma(B_s, B_t)$. This concludes the proof. $\square$

8. B8.2 Class 2

8.1. **Stopping Times.**

Definition 8.1. *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space. A stopping time τ is a measurable map $\tau : \Omega \to [0, \infty]$ such that*

$$\{\tau \leq t\} \in \mathscr{F}_t$$

for all $t \geq 0$.

Lemma 8.2. *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space and let τ be a stopping time. Then for any $t \geq 0$ it holds that*

$$\{\tau > t\}, \{\tau < t\}, \{\tau \geq t\} \in \mathscr{F}_t.$$

Proof. As $\{\tau > t\} = \{\tau \leq t\}^c$, the first claim follows. Similarly since $\{\tau < t\}^c = \{\tau \geq t\}$ it suffices to show that $\{\tau < t\} \in \mathscr{F}_t$, which follows since

$$\{\tau < t\} = \bigcup_{n\geq 1} \{\tau \leq t - \tfrac{1}{n}\} \in \mathscr{F}_t.$$

□

Lemma 8.3. *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space and let τ and ρ be stopping times. The following properties hold:*

(i) $\tau \wedge \rho = \min\{\tau, \rho\}$, $\tau \vee \rho = \max\{\tau, \rho\}$ and $\tau + \rho$ are all stopping times.
(ii) The collection of sets

$$\mathscr{F}_\tau = \{A \in \mathscr{F} \,:\, A \cap \{\tau \leq t\} \in \mathscr{F}_t \text{ for all } t \geq 0\}$$

is a σ-algebra.
(iii) If $\tau \leq \rho$ then $\mathscr{F}_\tau \leq \mathscr{F}_\rho$.
(iv) $\mathscr{F}_{\tau\wedge\rho} = \mathscr{F}_\tau \cap \mathscr{F}_\rho$ and $\{\tau \leq \rho\}$ is $\mathscr{F}_{\tau\wedge\rho}$-measurable.

Proof. To show (i), note that $\{\tau \wedge \rho \leq t\} = \{\tau \leq t\} \cup \{\rho \leq t\}$ and $\{\tau \vee \rho \leq t\} = \{\tau \leq t\} \cap \{\rho \leq t\}$ implying that $\tau \wedge \rho$ and $\tau \vee \rho$ are stopping times. It remains to show that $\tau + \rho$ is a stopping times. To see this, observe

$$\{\tau + \rho > t\} = \{\tau = 0, \rho > t\} \cup \{\rho = 0, \tau > t\} \cup \bigcup_{\substack{s\in\mathbb{Q} \\ s\in[0,t]}} \big(\{\tau > s\} \cap \{\rho > t - s\}\big) \in \mathscr{F}_t.$$

Indeed this inequality of sets holds since if $\tau + \rho > t$, then, as $\mathbb{Q}$ is dense in $\mathbb{R}$, there is $\varepsilon > 0$ such that $\tau + \rho - \varepsilon > t$ and $s = \tau - \varepsilon$ is rational. The claim follows since $\{t + \rho \leq t\} = \{t + \rho > t\}^c$.

Next, to show (ii) observe that it is clear that $\mathscr{F}_\tau$ contains $\emptyset$ and Ω and that it is closed under countable unions. To show that it is closed under complements, notice that for any set A and $t \geq 0$ we have a disjoint union.

$$\{\tau \leq t\} = (A \cap \{\tau \leq t\}) \cup (A^c \cap \{\tau \leq t\}).$$

Therefore if $A \in \mathscr{F}_\tau$, we conclude

$$A^c \cap \{\tau \leq t\} = \{\tau \leq t\} \backslash (A \cap \{\tau \leq t\}) = \{\tau \leq t\} \cap (A \cap \{\tau \leq t\})^c \in \mathscr{F}_\tau$$

and the claim follows.

To show (iii), let $A \in \mathscr{F}_\tau$. Then since $\{\rho \leq t\} \subset \{\tau \leq t\}$,

$$A \cap \{\rho \leq t\} = A \cap \{\rho \leq t\} \cap \{\tau \leq t\} = (A \cap \{\tau \leq t\}) \cap \{\rho \leq t\} \in \mathscr{F}_t$$

as $A \cap \{\tau \leq t\} \in \mathscr{F}_t$.

For (iv), observe it follows by (iii) that $\mathscr{F}_{\tau\wedge\rho} \subset \mathscr{F}_\tau \cap \mathscr{F}_\rho$. For the other direction let $A \in \mathscr{F}_\tau \cap \mathscr{F}_\rho$. Then

$$A \cap \{\tau \wedge \rho \leq t\} = A \cap (\{\tau \leq t\} \cup \{\rho \leq t\}) = (A \cap \{\tau \leq t\}) \cup (A \cap \{\rho \leq t\}) \in \mathscr{F}_t.$$

Therefore $A \in \mathscr{F}_{\tau\wedge\rho}$. Finally, note that

$$\{\tau \leq \rho\} \cap \{\tau \leq t\} = \{\tau \wedge t \leq \rho \wedge t\} \cap \{\tau \leq t\} \in \mathscr{F}_t,$$

where it follows that $\{\tau\wedge t \leq \rho\wedge t\} \in \mathscr{F}_t$ since both $\tau\wedge t$ and $\rho\wedge t$ are $\mathscr{F}_t$-measurable functions. Similarly,

$$\{\tau \leq \rho\} \cap \{\rho \leq t\} = \{\tau \wedge t \leq \rho \wedge t\} \cap \{\tau \leq t\} \cap \{\rho \leq t\}.$$

Therefore $\{\tau \leq \rho\} \in \mathscr{F}_{\tau\wedge\rho}$. □

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space. Denote

$$\mathscr{F}_{t+} = \bigcap_{\varepsilon>0} \mathscr{F}_{t+\varepsilon}.$$

Note that $\mathscr{F}_{t+}$ is a σ-algebra as it is an intersection of σ-algebras. Recall that we say that the filtration $(\mathscr{F}_t)_{t\geq 0}$ is **right-continuous** if $\mathscr{F}_t = \mathscr{F}_{t+}$ for all $t \geq 0$

Given an adapted stochastic process $(X_t)_{t\geq 0}$, for a Borel-measurable subset $\Gamma \subset \mathbb{R}$ denote

$$H_\Gamma = \inf\{t \geq 0, X_t \in \Gamma\}.$$

Lemma 8.4. *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space and let $(X_t)_{t\geq 0}$ be an adapted stochastic process. Then the following holds:*

(i) If $(X_t)_{t\geq 0}$ has right-continuous paths, then for an open set Γ, $H_\Gamma = \inf\{t \geq 0 : X_t \in \Gamma\}$ is a stopping time relative to $(\mathscr{F}_{t+})_{t\geq 0}$.

(ii) If $(X_t)_{t\geq 0}$ has continuous paths, then for a closed set Γ, $H_\Gamma = \inf\{t \geq 0 : X_t \in \Gamma\}$ is a stopping time relative to $(\mathscr{F}_t)_{t\geq 0}$.

Proof. For (i) we want to show that $\{H_\Gamma \leq t\} \in \mathscr{F}_{t+}$. Note that it is enough to show that $\{H_\Gamma < t\} \in \mathscr{F}_t$ since then for any $k > 0$,

$$\{H_\Gamma \leq t\} = \bigcap_{n=k}^{\infty} \left\{H_\Gamma < t + \frac{1}{n}\right\} \in \mathscr{F}_{t+\frac{1}{k}}$$

and hence $\{H_\Gamma \leq t\} \in \mathscr{F}_{t+}$.

Now since $(X_t)_{t\geq 0}$ is right continuous and Γ is open, if $X_r \in \Gamma$ then necessarily $X_s \in \Gamma$ for some $s > r$ and $s \in \mathbb{Q}$. Thus it follows that

$$\{H_\Gamma < t\} = \bigcup_{s<t} \{X_s \in \Gamma\} = \bigcup_{s<t, s\in\mathbb{Q}} \{X_s \in \Gamma\},$$

which is in $\mathscr{F}_t$ as required.

For (ii) we first introduce the distance between a point $x \in \mathbb{R}$ and a subset $\Gamma \subset \mathbb{R}$. We define

$$d(x, \Gamma) = \inf_{\gamma\in\Gamma} d(x, \gamma).$$

If Γ is closed then it holds that

$$d(x, \Gamma) = \min_{\gamma\in\Gamma} d(x, \Gamma). \tag{8.1}$$

Indeed, if $\gamma_n \in \Gamma$ is a sequence with $d(x,\gamma_n) < d(x,\Gamma) + \frac{1}{n}$ then γ_n ranges within a compact set. Thus we can pass to a converging subsequence, showing that there is $\gamma \in \Gamma$ with $d(x,\Gamma) = d(x,\gamma)$.

Returning to (ii), we will first use that Γ is closed, to show that

$$\{H_\Gamma \le t\} = \left\{ \inf_{s\in[0,t]} d(X_s,\Gamma) = 0 \right\}. \tag{8.2}$$

Indeed, we note that $\subset$ in (8.2) is clear and for the other direction let $\omega \in \Omega$ be an event such that $\inf_{s\in[0,t]} d(X_s(\omega),\Gamma) = 0$. Then there is a sequence $s_n \in [0,t]$ such that $\lim_{n\to\infty} d(X_{s_n}(\omega),\Gamma) = 0$. Using that $[0,t]$ is compact, we can assume without loss of generality that s_n converges to a limit point $s \in [0,t]$. Thus, as X has continuous paths it follows that $d(X_s(\omega),\Gamma) = 0$. Finally, by (8.1), there is $\gamma \in \Gamma$ with $d(X_s(\omega),\gamma) = 0$ and hence $X_s(\omega) = \gamma \in \Gamma$.

We finally deduce (ii) from (8.2). Indeed, since X_s has continuous paths it holds that

$$\begin{aligned} \{H_\Gamma \le t\} &= \left\{ \inf_{s\in[0,t], s\in\mathbb{Q}} d(X_s,\Gamma) = 0 \right\} \\ &= \bigcap_{n\ge1} \bigcup_{\substack{s\le t,\\ s\in\mathbb{Q}}} \{X_s \in B_{\frac{1}{n}}(\Gamma)\}, \end{aligned}$$

where $B_{\frac{1}{n}}(\Gamma) = \{x \in X : d(x,\Gamma) < \frac{1}{n}\}$. As $\{X_s \in B_{\frac{1}{n}}(\Gamma)\} \in \mathscr{F}_s$, the claim follows. □

Assuming that X_t is continuous and $(\mathscr{F}_t)_{t\ge0}$ is right-continuous, we can use Lemma 8.4 to conclude that the hitting time of a Borel-measurable sets is also a stopping time.

Theorem 8.5. *Let $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\ge0},\mathbb{P})$ be a filtered probability space with right-continuous filtration and let $(X_t)_{t\ge0}$ be a continuous adapted stochastic process. Then for every Borel measurable set A, the hitting time H_A is a stopping time.*

Before proving the result, we establish the following lemma:

Lemma 8.6. *Let $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\ge0},\mathbb{P})$ be a filtered probability space with right-continuous filtration and let $(\tau_n)_{n\ge1}$ be a sequence of stopping times. Then*

$$\inf_{n\ge1} \tau_n \qquad \textit{and} \qquad \sup_{n\ge1} \tau_n$$

are stopping times.

Proof. As in the proof of proposition 8.4, it suffices to show that $\{\liminf_{n\to\infty} \tau_n < t\} \in \mathscr{F}_t$ for all $t \ge 0$. Note that $\inf_{n\to\infty} \tau_n < t$ whenever $\tau_n \le s$ infinitely often for some $s < t$ enabling us to write

$$\{\inf_{n\ge1} \tau < t\} = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{\tau_m \le t - \tfrac{1}{k}\},$$

which is in $\mathscr{F}_t$. A similar argument applies to $\sup_{n\ge1} \tau_n$, concluding the proof. □

Proof. (of Theorem 8.5) It suffices to show the claim for Borel measurable sets. Consider the collection of sets

$$\mathscr{A} = \{A \subset \mathbb{R} : H_A \text{ and } H_{A^c} \text{ are stopping times}\}.$$

By Lemma 8.4, $\mathscr{A}$ contains all open subsets of $\mathbb{R}$. It therefore suffices to show that $\mathscr{A}$ is a σ-algebra.

It is clear that $\emptyset, \mathbb{R} \in \mathscr{A}$ and that $\mathscr{A}$ is closed under complements. To show that $\mathscr{A}$ is closed under countable unions, let $(A_n)_{n\geq 1}$ be a collection of sets in $\mathscr{A}$ and write $A = \bigcup_{n\geq 1} A_n$ and $U_k = \bigcup_{n=1}^k A_n$. Then by Lemma 8.3 (i), H_{U_k} is a stopping time and observe that $H_A = \lim_{k\to\infty} H_{U_k} = \sup_{k\geq 1} H_{U_k}$. Thus by Lemma 8.6 it follows that H_A is also a stopping time. A similar argument applies to H_{A^c}, concluding the proof. $\square$

8.2. **Optional Stopping Theorem.**

Theorem 8.7. *(Optional Stopping Theorem) Let $(M_t)_{t\geq 0}$ be a uniformly integrable martingale and let τ be an almost surely finite stopping time. Then*

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

We note that a bounded martingale is uniformly integrable. To apply the optional stopping theorem, the following proposition is useful.

Proposition 8.8. *Let $(M_t)_{t\geq 0}$ be a martingale with right continuous paths and let τ be an almost surely finite stopping time. Then the stopped process $M^\tau = M_{\tau\wedge t}$ is also a martingale.*

Lemma 8.9. *Let $(B_t)_{t\geq 0}$ be a Brownian motion. Then*

$$\limsup_{t\to\infty} B_t = \infty \quad \text{and} \quad \liminf_{t\to\infty} B_t = -\infty$$

almost surely.

Proof. By symmetry it suffices to show that for every $M > 0$,

$$\mathbb{P}[\sup_{s\geq 0} B_s > M] = 1.$$

Let $\delta > 0$ and recall that $B_t^\delta = B_{\delta^2 t}/\delta$ is a standard Brownian motion. Thus

$$\mathbb{P}\left[\sup_{0\leq s\leq 1} B_s > M\delta\right] = \mathbb{P}\left[\sup_{0\leq s\leq 1/\delta^2} B_{\delta^2 s}/\delta > M\right] = \mathbb{P}\left[\sup_{0\leq s\leq 1/\delta^2} B_t^\delta > M\right].$$

Letting $\delta \to 0$, the right-hand sice converges to $\mathbb{P}[\sup_{s\geq 0} B_s > M]$ whereas the left-hand side converges to $\mathbb{P}\left[\sup_{0\leq s\leq 1} B_s > 0\right] = 1$, as seen in the lecture.

Alternatively the claim follows from the reflection principle, or the fact that if $S_n = \sum_{i=1}^n X_i$ with $(X_i)_{i\geq 1}$ independent $\mathcal{N}(0,1)$-random variables then $\limsup_{t\to\infty} S_n = \infty$ almost surely. $\square$

Let B_t be the standard Brownian motion. Denote by $H_a = \inf\{t \geq 0 : B_t = a\}$.

Lemma 8.10. *Let $a < 0 < b$. Then $B_{H_a\wedge H_b}$ is distributed as*

$$\frac{b}{b-a}\delta_a + \frac{-a}{b-a}\delta_b.$$

Proof. Note $H_a \wedge H_b < \infty$ almost surely by Lemma 8.9. Therefore $B_{H_a\wedge H_b}$ is distributed like $p\delta_a + (1-p)\delta_b$ for some $p \in [0,1]$. Moreover, by the optional stopping theorem, since $B^{H_a\wedge H_b}$ is bounded and therefore uniformly integrable,

$$p \cdot a + (1-p) \cdot b = \mathbb{E}[B_{H_a\wedge H_b}] = \mathbb{E}[B^{H_a\wedge H_b}_{H_a\wedge H_b}] = \mathbb{E}[B_0] = 0.$$

This implies $p = \frac{b}{b-a}$ and $(1-p) = \frac{-a}{b-a}$. $\square$

Lemma 8.11. *For $a, \lambda > 0$ it holds that*

$$\mathbb{E}[e^{-\lambda H_a}] = e^{-a\sqrt{2\lambda}}.$$

Proof. Since $M_t = e^{\alpha B_t - \alpha^2 t/2}$ is a martingale and M^{H_a} is bounded it follows from the optional stopping theorem,

$$1 = \mathbb{E}[e^{\alpha B_{H_a} - \alpha^2 H_a/2}] = e^{\alpha a}\mathbb{E}[e^{-\alpha^2 H_a/2}].$$

Therefore $\mathbb{E}[e^{-\alpha^2 H_a/2}] = e^{-\alpha a}$. Setting $\alpha = \sqrt{2\lambda}$ implies the claim. $\square$

Lemma 8.12. *For $a, \lambda > 0$ it holds that*

$$\mathbb{E}[e^{-\lambda H_a \wedge H_{-a}}] = \frac{1}{\cosh(a\sqrt{2\lambda})}.$$

Proof. Denote by $A = \{H_a \wedge H_{-a} = H_a\} = \{B_{H_a \wedge H_{-a}} = a\}$. Then it holds that $A^c = \{H_a \wedge H_{-a} = H_{-a}\} = \{B_{H_a \wedge H_{-a}} = -a\}$ and by Lemma 8.10 we have that $\mathbb{P}[A] = \mathbb{P}[A^c] = 1/2$. Also by symmetry we have that, $\mathbb{E}[e^{-\lambda H_a \wedge H_{-a}}|A] = \mathbb{E}[e^{-\lambda H_a \wedge H_{-a}}|A^c] = \mathbb{E}[e^{-\lambda H_a \wedge H_{-a}}]$. Indeed this follows since

$$\mathbb{E}[e^{-\lambda H_a \wedge H_{-a}}] = \mathbb{E}[e^{-\lambda H_a \wedge H_{-a}}|A]\mathbb{P}[A] + \mathbb{E}[e^{-\lambda H_a \wedge H_{-a}}|A^c]\mathbb{P}[A^c].$$

Recall that for any $\alpha > 0$ it holds that $M_t = e^{\alpha B_t - \alpha^2 t/2}$ is a martingale. Since $M^{H_a \wedge H_b}$ is bounded, it follows by the optional stopping theorem,

$$\begin{aligned}
1 = \mathbb{E}[e^{\alpha B_0}] &= \mathbb{E}[e^{\alpha B_{H_a \wedge H_{-a}} - \alpha^2 (H_a \wedge H_{-a})/2}] \\
&= \mathbb{E}[e^{\alpha B_{H_a \wedge H_{-a}} - \alpha^2 (H_a \wedge H_{-a})/2}|A]\mathbb{P}[A] \\
&+ \mathbb{E}[e^{\alpha B_{H_a \wedge H_{-a}} - \alpha^2 (H_a \wedge H_{-a})/2}|A^c]\mathbb{P}[A^c] \\
&= \frac{e^{\alpha a}}{2}\mathbb{E}[e^{-\alpha^2 (H_a \wedge H_{-a})/2}|A] + \frac{e^{-\alpha a}}{2}\mathbb{E}[e^{-\alpha^2 (H_a \wedge H_{-a})/2}|A^c] \\
&= \cosh(\alpha a)\mathbb{E}[e^{-\alpha^2 H_a \wedge H_{-a}/2}].
\end{aligned}$$

Therefore $\mathbb{E}[e^{-\alpha^2 H_a \wedge H_{-a}/2}] = \frac{1}{\cosh(a\alpha)}$ and setting $\alpha = \sqrt{2\lambda}$ implies the claim. $\square$

8.3. **Reflection Principle.**

Theorem 8.13. *(The reflection principle) Let $(B_t)_{t \geq 0}$ be a standard Brownian motion and let τ be a stopping time. Then the process*

$$B'_t = \begin{cases} B_t & \text{if } t < \tau, \\ 2B_\tau - B_t & \text{if } t \geq \tau \end{cases}$$

is a standard Brownian motion.

Proof. This follows rather immediately from the strong Markov property. $\square$

Corollary 8.14. *Let $S_t = \sup_{0 \leq s \leq t} B_s$. Then for $a \geq 0$ and $b \leq a$ we have for all $t \geq 0$*

$$\mathbb{P}[S_t \geq a, B_t \leq b] = \mathbb{P}[B_t \geq 2a - b].$$

Moreover, S_t and $|B_t|$ have the same distribution.

Proof. We apply the reflection principle with the stopping time H_a and B' the at τ reflected process:

$$\begin{aligned}\mathbb{P}[S_t \geq a, B_t \leq b] &= \mathbb{P}[T_a \leq t, B_t \leq b] \\ &= \mathbb{P}[T_a \leq t, B'_t \leq b] \\ &= \mathbb{P}[T_a \leq t, 2a - B_t \leq b] \\ &= \mathbb{P}[B_t \geq 2a - b],\end{aligned}$$

where we used that as $2a - b \geq > a$ we have that $\{B_t \geq 2a - b\} \subset \{T_a \leq t\}$.

To show the claim about the distribution of S_t we calculate

$$\begin{aligned}\mathbb{P}[S_t \geq a] &= \mathbb{P}[S_t \geq a, B_t \geq a] + \mathbb{P}[S_t \geq a, B_t \leq a] \\ &= 2\mathbb{P}[B_t \geq a] \\ &= \mathbb{P}[B_t \geq a] + \mathbb{P}[B_t \leq a] = \mathbb{P}[|B_t| \geq a].\end{aligned}$$

□

Lemma 8.15. *Let $a \neq 0$. Then the probability density function of H_a is*

$$f_{H_a}(t) = \frac{|a|}{\sqrt{2\pi t^3}} e^{-a^2/2t}.$$

Proof. By symmetry, $H_a \sim H_{-a}$ and therefore it suffices to assume $a > 0$. Denote $S_t = \sup_{0 \leq s \leq t} B_s$ and observe that by continuity $\{H_a \leq t\} = \{S_t \geq a\}$. Write

$$\Phi(y) = \mathbb{P}[\mathcal{N}(0,1) \leq y] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Then upon applying a substitution

$$\mathbb{P}[\mathcal{N}(0,\sigma^2) \leq y] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}\, dx = \Phi(y/\sigma)$$

Therefore it holds that $\mathbb{P}[H_a \leq t] = \mathbb{P}[S_t \geq a] = \mathbb{P}[|B_t| \geq a] = 1 - 2\Phi(a/\sqrt{t})$. Thus the probability density function of H_a is

$$\frac{d}{dt}\mathbb{P}[H_a \leq t] = \frac{d}{dt}(1 - 2\Phi(a/\sqrt{t})) = -2\Phi'(a/\sqrt{t})(-1/2)at^{-3/2} = \frac{a}{\sqrt{2\pi t^3}} e^{-a^2/2t}.$$

□

We observe that Lemma 8.15 is a rather elegant result. The hitting times of the standard random walk on $\mathbb{Z}$ don't have as explicit a closed form.

Lemma 8.16. *Let $a \neq 0$ and $U_a = \sup\{t \geq 0 : B_t = at\}$ be the last time that Brownian motion hits the line at. Then $U_a = 1/H_a$ in distribution.*

Proof. We assume $a > 0$. Denote $W_s = sB_{1/s}$. Then for $r \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}[U_a \le r] &= \mathbb{P}[\sup\{t \ge 0 \,:\, B_t = at\} \le r] \\ &= \mathbb{P}[\sup\{t \ge 0 \,:\, B_t = at\} < r] \\ &= \mathbb{P}[\tfrac{1}{t} B_t \neq a \text{ for all } t \ge r] \\ &= \mathbb{P}[sB_{s^{-1}} \neq a \text{ for all } s^{-1} \ge r] \\ &= \mathbb{P}[W_s \neq a \text{ for all } r^{-1} \ge s] \\ &= \mathbb{P}[\max_{0 \le s \le r^{-1}} W_s < a] \\ &= 1 - \mathbb{P}[H_a < r^{-1}] = \mathbb{P}[H_a^{-1} \le r],\end{aligned}$$

having used in the second line that $\mathbb{P}[B_r = ar] = 0$. □

We note that U_a is not a stopping time, since the event $\{U_a \le t\}$ does not only depend on the values of $(B_s)_{0 \le s \le t}$.

Lemma 8.17. *It holds that* $\mathbb{E}[U_a] = \frac{1}{a^2}$ *and therefore* $\mathbb{E}[B_{U_a}] = \mathbb{E}[aU_a] = \frac{1}{|a|}$.

Proof. Again we assume $a > 0$. Note that $\frac{d}{dt} e^{-a^2/2t} = \frac{a^2}{2t^2} e^{-a^2/2t}$. Therefore by Lemma 8.15 and partial integration,

$$\begin{aligned}\mathbb{E}[U_a] = \mathbb{E}[\tfrac{1}{H_a}] &= \int_0^\infty \frac{1}{t} f_{H_a}(t)\, dt \\ &= \int_0^\infty \frac{a}{\sqrt{2\pi t^5}} e^{-a^2/2t}\, dt \\ &= \frac{1}{a\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{t}} \cdot \frac{a^2}{2t^2} e^{-a^2/2t}\, dt \\ &= \frac{1}{a\sqrt{2\pi}} \int_0^\infty \frac{a}{\sqrt{t^3}} e^{-a^2/2t}\, dt \\ &= \frac{1}{a^2} \int_0^\infty \frac{a}{\sqrt{2\pi t^3}} e^{-a^2/2t}\, dt = \frac{1}{a^2}.\end{aligned}$$

□

8.4. **An Exercise on Gaussian Processes.**

Lemma 8.18. *Let* $(B_t)_{t \ge 0}$ *be a Brownian motion. Fix* $0 \le s < t < \infty$. *Then conditionally on* B_s *and* B_t *for* $\alpha \in [0,1]$ *it holds that* $B_{\alpha s + (1-\alpha)t}$ *is normally distributed with mean* $\alpha B_s + (1-\alpha)B_t$ *and variance* $\alpha(1-\alpha)(t-s)$.

Proof. Consider the random variable

$$Y = B_{\alpha s+(1-\alpha)t} - \alpha B_s - (1-\alpha)B_t = \alpha(B_{\alpha s+(1-\alpha)t} - B_s) + (1-\alpha)(B_{\alpha s+(1-\alpha)t} - B_t).$$

Note that Y is a sum of independent mean-zero Gaussian. Therefore it is a Gaussian itself, $\mathbb{E}[Y] = 0$ and

$$\mathrm{Var}(Y) = \alpha^2((\alpha-1)s + (1-\alpha)t) + (1-\alpha)^2(\alpha t - \alpha s) = \alpha(1-\alpha)(t-s).$$

We show below that Y is independent from B_s and B_t. Assume for the moment, that this is the case. Then the distribution of Y conditioned on B_s and B_t is the same as the distribution of Y, which is a Gaussian with mean zero and variance $\alpha(1-\alpha)(t-s)$. Then $\mathbb{E}[Y|B_s, B_t] = \mathbb{E}[Y] = 0$ and thus

$$\mathbb{E}[B_{\alpha s+(1-\alpha)t} | B_s, B_t] = \mathbb{E}\left[\alpha B_s + (1-\alpha)B_t \,|\, B_s, B_t\right] = \alpha B_s + (1-\alpha)B_t.$$

Similarly, $\mathrm{Var}(Y|B_s, B_t) = \mathrm{Var}(Y) = \alpha(1-\alpha)(t-s)$ and therefore

$$\begin{aligned}\mathrm{Var}(B_{\alpha s+(1-\alpha)t}|B_s, B_t) &= \mathrm{Var}(B_{\alpha s+(1-\alpha)t} - \mathbb{E}[B_{\alpha s+(1-\alpha)t}|B_s, B_t]|B_s, B_t)\\ &= \mathrm{Var}(B_{\alpha s+(1-\alpha)t} - \alpha B_s - (1-\alpha)B_t)|B_s, B_t)\\ &= \mathrm{Var}(Y|B_s, B_t)\\ &= \alpha(1-\alpha)(t-s)\end{aligned}$$

showing the claim.

It remains to show that Y is independent of B_s and B_t. We first show that Y and B_s are independent. Indeed (Y, B_s) is a multivariate Gaussian and so it suffices to show that $\mathrm{cov}(Y, B_s) = 0$. We calculate

$$\begin{aligned}\mathrm{cov}(Y, B_s) &= \mathrm{cov}(B_{\alpha s+(1-\alpha)t}, B_s) - \alpha \cdot \mathrm{cov}(B_s, B_s) - (1-\alpha) \cdot \mathrm{cov}(B_t, B_s)\\ &= s - \alpha s + (1-\alpha)s = 0.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathrm{cov}(Y, B_t) &= \mathrm{cov}(B_{\alpha s+(1-\alpha)t}, B_t) - \alpha\mathrm{cov}(B_s, B_t) - (1-\alpha)\mathrm{cov}(B_t, B_t)\\ &= \alpha s + (1-\alpha)t - (\alpha s + (1-\alpha)t) = 0.\end{aligned}$$

In particular, $Y, B_s - B_t$ and B_t is a collection of independent random variables as their covariance matrix is zero. Thus it follows that Y is independent of $\sigma(B_s - B_t, B_t) = \sigma(B_s, B_t)$. This concludes the proof. □

Lemma 8.19. *Let B_t be a standard Brownian motion on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$. Consider for $0 \leq t \leq 1$,*

$$X_t = x(1-t) + yt + (B_t - tB_1).$$

Then X_t is a continuous Gaussian process with mean $x(1-t) + yt$ and variance $t(1-t)$. Moreover, X_t has the same law as $(B_t|B_0 = x, B_1 = y)$.

We note that X_t is not adapted to the filtration $\mathscr{F}_t$, which would imply that B_1 is $\mathscr{F}_t$ measurable. However this is not possible since if this was the case then $B_1 = \mathbb{E}[B_1|\mathscr{F}_t] = B_t$ for $0 \leq t \leq 1$, which is a contradiction. Also B_t is not a $\mathscr{F}_t^X$-Brownian motion. If it were, then B_1 would be $\mathscr{F}_t^X$-measurable since B_t would be adapted. Therefore again $B_1 = \mathbb{E}[B_1|\mathscr{F}_t^X]$, which contradicts B_t being a $\mathscr{F}_t^X$-Brownian motion.

Proof. We note that $X_0 = x$ and $X_1 = y$. Moreover, since Brownian motion is a Gaussian process, it follows that X_t is a Gaussian process. Note that

$$\mathbb{E}[X_t] = x(1-t) + yt$$

and

$$\begin{aligned}\mathrm{Var}(X_t) &= \mathrm{Var}(B_t - tB_1)\\ &= \mathrm{Var}((1-t)B_t - t(B_1 - B_t))\\ &= (1-t)^2 t + t^2(1-t) = t(1-t).\end{aligned}$$

By Lemma 8.18 it follows that X_t has the same law as $(B_t|B_0 = x, X_1 = y)$. □

9. B8.2 Class 3

9.1. **An exercise on stopping times.**

Lemma 9.1. *Let M be a positive continuous martingale converging almost surely to zero as $t \to \infty$. Let $M^* = \sup_{t\geq 0} M_t$. Then for $x > 0$*

$$P[M^* \geq x \,|\, \mathscr{F}_0] = \min\left(1, \frac{M_0}{x}\right).$$

Moreover, the distribution of M^* is distributed as M_0/U for U an uniform $[0,1]$ variable independent of M_0.

Proof. Denote $\tau_x = \inf\{t \geq 0 : M_t \geq x\}$. Then we note that the process $Y_t = M_{t\wedge\tau_x}$ is bounded by $\max(x, M_0)$ and hence is a uniformly integrable martingale. We note that since $M_t \to 0$ almost surely,

$$\begin{aligned} Y_\infty &= M_0 \cdot 1_{\{M_0 \geq x\}} + x \cdot 1_{\{\tau_x < \infty\}} \cdot 1_{\{M_0 < x\}} \\ &= M_0 \cdot 1_{\{M_0 \geq x\}} + x \cdot 1_{\{M^* \geq x\}} \cdot 1_{\{M_0 < x\}} \end{aligned}$$

Therefore it follows that

$$1_{\{M^* \geq x\}} \cdot 1_{\{M_0 < x\}} = \frac{Y_\infty - M_0 \cdot 1_{\{M_0 \geq x\}}}{x}.$$

Finally using that $E[Y_\infty \,|\, \mathscr{F}_0] = M_0$, we conclude that

$$\begin{aligned} \mathbb{P}[M^* \geq x \,|\, \mathscr{F}_0] &= \mathbb{E}[1_{\{M^* \geq x\}} \,|\, \mathscr{F}_0] \\ &= \mathbb{E}[1_{\{M^* \geq x\}}(1_{\{M_0 < x\}} + 1_{\{M_0 \geq x\}}) \,|\, \mathscr{F}_0] \\ &= \mathbb{E}\left[\frac{Y_\infty - M_0 \cdot 1_{\{M_0 \geq x\}}}{x} \middle| \mathscr{F}_0\right] + 1_{\{M_0 \geq x\}} \\ &= \frac{M_0}{x} 1_{\{M_0 < x\}} + 1_{\{M_0 \geq x\}} \\ &= \min\left(1, \frac{M_0}{x}\right), \end{aligned}$$

showing the first equation.

To show the final claim, we notice that

$$\begin{aligned} \mathbb{P}[M^* \geq x] &= \mathbb{E}[1_{\{M^* \geq x\}}] \\ &= \mathbb{E}[\mathbb{E}[1_{\{M^* \geq x\}} \,|\, \mathscr{F}_0]] \\ &= \mathbb{E}\left[\min\left(1, \frac{M_0}{x}\right)\right] \end{aligned}$$

and furthermore,

$$\begin{aligned} \mathbb{P}\left[\frac{M_0}{U} \geq x\right] &= \mathbb{E}[1_{\{\frac{M_0}{x} \geq U\}}] \\ &= \mathbb{E}[\mathbb{E}[1_{\{\frac{M_0}{x} \geq U\}} \,|\, \mathscr{F}_0]] \\ &= \mathbb{E}\left[\min\left(1, \frac{M_0}{x}\right)\right] \end{aligned}$$

showing the final claim. $\square$

Lemma 9.2. *Let $a > 0$ and $B_t^a = a + B_t$ be a Brownian motion starting at a. Let $\tau = H_0(B^a) = \inf\{t \geq 0 : B_t^a = 0\}$. Then the distribution of $Y = \sup_{t \leq \tau} B_t^a$ is a/U with U a uniform $[0,1]$ random variable.*

Proof. Consider the positive continuous martingale $M_t = B_{t\wedge\tau}^a$. Since M_t is positive, it is bounded in L^1 and therefore there is $M_\infty \in L^1$ such that $M_t \to M_\infty$ almost surely. We note that τ is almost surely finite and therefore $M_\infty = 0$ almost surely. The distribution of $\sup_{t\geq 0} M_t$ is the same as the one of $Y = \sup_{t\leq\tau} B_t^a$ and thus by the previous lemma it holds that Y is distributed as a/U for U a uniform $[0,1]$ variable. □

9.2. **Continuous Local Martingales and Quadratic Variation.**

Definition 9.3. *An adapted process $(M_t)_{t\geq 0}$ is called a **continuous local martingale** if $M_0 = 0$, it has continuous trajectories a.s. and if there exists a non-decreasing sequence of stopping times $(\tau_n)_{n\geq 1}$ such that $\tau_n \uparrow \infty$ a.s. and for each n, $M^{\tau_n} = (M_{\tau_n \wedge t})_{t\geq 0}$ is a martingale. We say $(\tau_n)_{n\geq 1}$ reduces or localizes M.*

More generally, when we do not assume that $M_0 = 0$, we say that M is a continuous local martingale if $N_t = M_t - M_0$ is a continuous local martingale.

The most important property of continuous local martingales is that they have quadratic variation processes.

Theorem 9.4. *Let M be a continuous local martingale. There exists a unique (up to indistinguishability) non-decreasing, continuous adapted finite variation process $\langle M\rangle = (\langle M, M\rangle_t)_{t\geq 0}$, starting in zero, such that $(M_t^2 - \langle M, M\rangle_t)_{t\geq 0}$ is a continuous local martingale. The process $\langle M\rangle$ is called the **quadratic variation** of M.*

9.2.1. *Characterisation of the quadratic variation being zero.* It is important to note that local martingales and finite variation processes are orthogonal to each other.

Theorem 9.5. *Let M be a continuous local martingale with $M_0 = 0$. Then the following properties are equivalent:*

(i) *M is indistinguistable from zero.*
(ii) *$\langle M\rangle_t = 0$ for all $t \geq 0$.*
(iii) *M is a process of finite variation, i.e. $t \mapsto M_t$ has finite variation almost surely.*

We draw the following corollary of the previous theorem.

Corollary 9.6. *Let M be a continuous local martingale and let $\tau_1 \leq \tau_2$ be two stopping times. Then the following are equivalent:*

(i) *M is a.s. constant on $[\tau_1, \tau_2]$.*
(ii) *$\langle M\rangle_t = 0$ is a.s. constant on $[\tau_1, \tau_2]$.*

Proof. Consider $M_s' = M^{\tau_2} - M^{\tau_1} = M_{s\wedge\tau_2} - M_{s\wedge\tau_1}$. Then by Proposition 7.27 from the lecture notes,

$$\begin{aligned}
\langle M'\rangle_s &= \langle M^{\tau_2}\rangle_s + \langle M^{\tau_1}\rangle_s - 2\langle M^{\tau_2}, M^{\tau_1}\rangle_s \\
&= \langle M\rangle_{s\wedge\tau_2} + \langle M\rangle_{s\wedge\tau_1} - 2\langle M, M\rangle_{s\wedge\tau_1\wedge\tau_2} \\
&= \langle M\rangle_{s\wedge\tau_2} - \langle M\rangle_{s\wedge\tau_1}.
\end{aligned}$$

Since $M_0' = 0$, the claim follows from the previous proposition. □

Lemma 9.7. *Assume that the filtration $(\mathscr{F}_t)_{t\geq 0}$ is right continuous. Let Y be a continuous stochastic process and let $t \in \mathbb{R}_{\geq 0}$. Then*

$$T_t = \inf\{s > t \,:\, Y_s \neq Y_t\}$$

is a stopping time.

Proof. We give two proofs, first we consider the continuous process $X_s = 1_{s>t}(Y_s - Y_t)$. Then T_t is the first hitting time of the open set $\mathbb{R}\backslash\{0\}$ and hence since the filtration is right continuous it follows that T_t is a stopping time.

Let $T \geq 0$. If $T < t$, then $\{T_t \leq T\} = \emptyset \in \mathscr{F}_T$ so we assume that $T \geq t$. Then it holds that

$$\begin{aligned}\{T_t < T\}^c &= \{T_t \geq T\} \\ &= \bigcap_{s\in[t,T]} \{Y_s = Y_t\} \\ &= \bigcap_{\substack{s\in[t,T]\\ s\in\mathbb{Q}}} \{Y_s = Y_t\},\end{aligned}$$

using in the last line that Y_s is continuous. Since $\{Y_s = Y_t\} \in \mathscr{F}_s$, it thus follows that $\{T_t < T\} \in \mathscr{F}_T$. To conclude the proof we notice that for any $k > 0$,

$$\{T_t \leq T\} = \bigcap_{n=k}^{\infty} \{T_t < T + \frac{1}{n}\} \in \mathscr{F}_{T+\frac{1}{k}}.$$

Therefore the claim follows by right continuity of the filtration. $\square$

Proposition 9.8. *Let M be a continuous L^2-bounded martingale. Then the intervals of constancy for M and $\langle M\rangle$ coincide. More precisely, if $S \leq S'$ are two random times then for almost every $\omega \in \Omega$ it holds that*

$$M_s(\omega) = M_{S(\omega)}(\omega) \text{ for all } s \in [S(\omega), S'(\omega)]$$

if and only if for almost all $\omega \in \Omega$ it holds that

$$\langle M\rangle_s(\omega) = \langle M\rangle_{S(\omega)}(\omega) \text{ for all } s \in [S(\omega), S'(\omega)].$$

Proof. By the previous corollary and lemma, the claim holds for (t, T_t) for any t. Assume that the first claim holds. Let $\Omega' \subset \Omega$ be the set of full measure such that every $\omega \in \Omega'$ satisfies $M_s(\omega) = M_{S(\omega)}(\omega)$ and $\langle M\rangle_{T_t}(\omega) = \langle M\rangle_t(\omega)$ for all $t \in \mathbb{Q}$.

Fix any $\omega \in \Omega'$. Either the interval $[S(\omega), S'(\omega)]$ consists of a single point, in which case there is nothing to show, or it contains a rational t and hence $S'(\omega) \leq T_t(\omega)$. However, since $\omega \in \Omega'$ it holds that $s \mapsto \langle M\rangle_s(\omega)$ is constant on $[t, T_t(\omega)]$. As t is an arbitrary rational, this implies that $s \mapsto \langle M\rangle_s(\omega)$ is constant on $(S(\omega), S'(\omega)]$. Finally since $s \mapsto \langle M\rangle_s(\omega)$ is continuous, it cannot vary at a single point, so it is constant on $[S(\omega), S'(\omega)]$. A similar argument applies for the converse direction. $\square$

9.2.2. *Quadratic Covariation.* Another useful property is the following.

Theorem 9.9. *Let M be a martingale with $M_0 \in L^2$. Then $\mathbb{E}[\langle M\rangle_\infty] < \infty$ and $M^2 - \langle M\rangle$ is a uniformly integrable martingale.*

Definition 9.10. *Let M and N be continuous local martingales. Then we define the* ***quadratic covariation*** *of M and N as*

$$\langle M, N\rangle = \frac{1}{2}\left(\langle M+N, M+N\rangle - \langle M, M\rangle - \langle N, N\rangle\right)$$

Proposition 9.11. *Let M and N be continuous local martingales. Then the following properties hold:*

(1) $(M_tN_t - \langle M, N\rangle_t)$ is the unique finite variation process that is a continuous local martingale and that is zero at zero.
(2) The mapping $M, N \mapsto \langle M, N\rangle$ is bilinear and symmetric.
(3) For any stopping time τ and $t \geq 0$,

$$\langle M^\tau, N^\tau\rangle_t = \langle M^\tau, N\rangle_t = \langle M, N^\tau\rangle_t = \langle M, N\rangle_{t\wedge\tau}.$$

Lemma 9.12. *Let M and N be continuous local martingales and let τ be a stopping time. Then*

$$M^\tau(N - N^\tau) = M^\tau N - M^\tau N^\tau$$

is a continuous local martingale.

Proof. By Proposition 9.11 (i) it holds that $M^\tau N - \langle M^\tau, N\rangle$ and $M^\tau N^\tau - \langle M^\tau, N^\tau\rangle$ are continuous local martingales. Also by Proposition 9.11 (ii) $\langle M^\tau, N\rangle = \langle M^\tau, N^\tau\rangle$ and therefore the difference of these two continuous local martingales is $M^\tau N - M^\tau N^\tau$ and so this is again a continuous local martingale. □

Lemma 9.13. *Let $(\Omega, \mathscr{G}, (\mathscr{G}_t)_{t\geq 0}, \mathbb{P})$ and let B be a standard Brownian motion with respect to $(\mathscr{G}_t)_{t\geq 0}$. Let X be a positive $\mathscr{G}_0$-measurable random variable that is independent of B_t for every $t \geq 0$ and write $M_t = B_{tX}$. Then $\rho_t = tX$ can be viewed as a $(\mathscr{G}_t)_{t\geq 0}$-stopping time and consider the stopping time σ-algebra $\mathscr{F}_t = \mathscr{G}_{\rho_t}$. Then the following properties hold*

(i) M_t is a continuous local martingale with respect to $(\mathscr{F}_t)_{t\geq 0}$.
(ii) If $\mathbb{E}_X[X^{1/2}] < \infty$, then $(M_t)_{t\geq 0}$ is a martingale.
(iii) $\langle M\rangle_t = tX$.

Proof. Consider the stopping times $\tau_n = \inf\{t \geq 0 : |M_t| \geq n\}$ and $\tau_n^B = \inf\{t \geq 0 : |B_t| \geq n\}$. Then $B^{\tau_n^B}$ is a uniformly integrable martingale and it holds that $M_t^{\tau_n} = B_{\rho_t}^{\tau_n^B}$. Thus it holds by the optional stopping theorem that

$$\mathbb{E}[M_t^{\tau_n}|\mathscr{F}_s] = \mathbb{E}[B_{\rho_t}^{\tau_n^B}|\mathscr{G}_{\rho_s}] = B_{\rho_s}^{\tau_n^B} = M_s^{\tau_n}. \tag{9.1}$$

Thus it follows that $M_t^{\tau_n}$ is a martingale and the claim follows.

To show (ii), it holds by using independence,

$$\mathbb{E}[|M_t|] = \mathbb{E}[\mathbb{E}[|B_{tX}|\,|\,\sigma(X)]] = \mathbb{E}\left[\sqrt{\tfrac{2tX}{\pi}}\right] < \infty$$

if $\mathbb{E}_X[X^{1/2}] < \infty$. Thus M_t is integrable. To show the martingale condition, we want to apply dominated convergence to (9.1). To do so, note that by the maximum principle, $\sup_{t\in[0,s]} M_t = \sup_{t\in[0,s]} B_{tX} \sim |B_{sX}| = M_s$ and therefore $\mathbb{E}[\sup_{t\in[0,s]} M_t] = \mathbb{E}[|M_s|] < \infty$. Similarly $\mathbb{E}[\sup_{t\in[0,s]} M_t^-] \leq \mathbb{E}[|M_s|]$ and therefore $\mathbb{E}[\sup_{t\in[0,s]} |M_t|] \leq 2\mathbb{E}[|M_s|] < \infty$. Finally notice that $|M_s^{\tau_n}| \leq \sup_{t\in[0,s]} |M_t| \in L^1$ and hence

$$\mathbb{E}[M_t|\mathscr{F}_s] = \lim_{n\to\infty} \mathbb{E}[M_t^{\tau_n}\,|\,\mathscr{F}_s] = \lim_{n\to\infty} M_s^{\tau_n} = M_s,$$

having used conditional dominated convergence in the first equality and (9.1) in the second. This shows that M_t is indeed a martingale.

Finally, to show (iii), we want to show that $N_t = M_t^2 - tX$ is a continuous local martingale. Denote $L_t = B_t^2 - t$ and recall that L_t is a martingale. We proceed as in (i). Let $\tau_n = \inf\{t \geq 0 : |N_t| \geq n\}$ and $\tau_n^L = \inf\{t \geq 0 : |L_t| \geq n\}$. Then it again holds that $N_t^{\tau_n} = L_{\rho_t}^{\tau_n^L}$ and

$$\mathbb{E}[N_t^{\tau_n}|\mathscr{F}_s] = \mathbb{E}[L_{\rho_t}^{\tau_n^L}|\mathscr{G}_{\rho_s}] = L_{\rho_s}^{\tau_n^L} = N_t^{\tau_n}.$$

Therefore N_t is a continuous local martingale and since the quadratic variation is the unique non-decreasing adapted finite variation process $\langle M\rangle$ such that $M_t^2 - \langle M\rangle$ is a continuous local martingale, it follows that $\langle M\rangle_t = tX$, concluding the proof of the lemma. □

9.3. Square Integrable Continuous Martingales.

Definition 9.14. *We define*

$$\mathscr{H}^{2,c} = \left\{ \text{continuous martingales } (M_t)_{t\geq 0} \text{ with } \sup_{t\geq 0} \mathbb{E}[M_t^2] < \infty \right\}.$$

We recall the following results from the lecture notes:

Proposition 9.15. *Let $M, N \in \mathscr{H}^{2,c}$. Then the following properties hold:*

(i) M is uniformly integrable and therefore M_∞ exists and as $t \to \infty$, $M_t \to M_\infty$ almost surely and in L^2. Indeed, $\sup_{t\geq 0}|M_t|$ is a square integrable random variable.

(ii) The inner product

$$\langle M, N\rangle_{\mathscr{H}^{2,c}} := \mathbb{E}[M_\infty N_\infty] \leq \mathbb{E}[M_\infty^2]^{1/2}\mathbb{E}[N_\infty^2]^{1/2} < \infty$$

defines an inner product on $\mathscr{H}^{2,c}$. Therefore $\mathscr{H}^{2,c}$ is a Hilbert space.

(iii) It holds for all $t \in [0,\infty]$ that $|\langle M, N\rangle_t| \leq \sqrt{\langle M\rangle_t}\cdot\sqrt{\langle N\rangle_t} \leq \sqrt{\langle M\rangle_\infty}\cdot\sqrt{\langle N\rangle_\infty}$

Lemma 9.16. *A continuous local martingale M such that there exists a random variable $Z \in L^1$ with $|M_t| \leq Z$ for every $t \geq 0$ is a uniformly integrable martingale.*

We use these results to deduce the following proposition.

Proposition 9.17. *Let $M, N \in \mathscr{H}^{2,c}$. Then $\mathbb{E}[\langle M, N\rangle_\infty] < \infty$ and $MN - \langle M, N\rangle$ is a uniformly integrable martinagle and therefore $\mathbb{E}[M_t N_t] = \mathbb{E}[M_0 N_0] + \mathbb{E}[\langle M, N\rangle_t]$.*

Proof. We give two proofs. First we recall that by Theorem 7.24 (i), if $X \in \mathscr{H}^{2,c}$ it follows that $X^2 - \langle X\rangle$ is a uniformly integrable martingale. Note that

$$2(MN - \langle M, N\rangle) = (M+N)^2 - \langle M+N\rangle - (M^2 - \langle M\rangle) - (N^2 - \langle N\rangle)$$

and therefore since $M + N \in \mathscr{H}^{2,c}$ it holds that $MN - \langle M, N\rangle$ is a sum of three uniformly integrable martingales and hence uniformly integrable itself.

For a second proof, we adapt the proof from Theorem 7.24 of the lecture notes. By Doob's L^2-inequality,

$$\mathbb{E}[\sup_{0\leq t\leq T} M_t^2] \leq 4\mathbb{E}[M_T^2]$$

and so letting $T \to \infty$,

$$\mathbb{E}[\sup_{t\geq 0} M_t^2] \leq 4\sup_{t\geq 0}\mathbb{E}[M_t^2] \leq C$$

for some $C > 0$. The same holds for N by suitably adjusting the constant C. Therefore by Cauchy-Schwarz,

$$\begin{aligned}\mathbb{E}\left[\left(\sup_{t\geq 0}|M_t|\right)\cdot\left(\sup_{t\geq 0}|N_t|\right)\right] &\leq \mathbb{E}\left[\left(\sup_{t\geq 0}|M_t|\right)^2\right]^{1/2}\cdot\mathbb{E}\left[\left(\sup_{t\geq 0}|N_t|\right)^2\right]^{1/2}\\ &= \mathbb{E}\left[\sup_{t\geq 0}M_t^2\right]^{1/2}\cdot\mathbb{E}\left[\sup_{t\geq 0}N_t^2\right]^{1/2}\leq C.\end{aligned}$$

Furthermore we note that by Proposition 9.15, Cauchy-Schwarz and Theorem 9.9,

$$\begin{aligned}\mathbb{E}[\sup_{t\geq 0}|\langle M,N\rangle_t|] &\leq \mathbb{E}[\sqrt{\langle M\rangle_\infty}\sqrt{\langle N\rangle_\infty}]\\ &\leq \mathbb{E}[\langle M\rangle_\infty]^{1/2}\mathbb{E}[\langle N\rangle_\infty]^{1/2}<\infty.\end{aligned}$$

To conclude the proof, we note that for all $M_tN_t-\langle M,N\rangle_t$ is bounded from above by $\left(\sup_{t\geq 0}|M_t|\right)\cdot\left(\sup_{t\geq 0}|N_t|\right)+\sup_{t\geq 0}|\langle M,N\rangle_t|$, which is therefore integrable. This implies the claim by Lemma 9.16. ☐

9.4. **Concrete Examples of Stochastic Integrals.** Suppose that $(B_t)_{t\geq 0}$ is Brownian motion. For a partition π of $[0,T]$, write $||\pi||$ for the mesh of the partition and $0=t_0<t_1<\ldots<t_{N(\pi)}=T$ for the endpoints of the intervals of the partition.

Lemma 9.18. *It holds that*

$$\lim_{||\pi||\to 0}\sum_{j=0}^{N(\pi)-1}B_{t_{j+1}}(B_{t_{j+1}}-B_{t_j})=\frac{1}{2}(B_T^2-B_0^2+T)$$

in probability.

Proof. Note that

$$\begin{aligned}S_{\pi,1}:=\sum_{j=0}^{N(\pi)-1}B_{t_{j+1}}(B_{t_{j+1}}-B_{t_j}) &= B_T^2-B_TB_{t_{N(\pi)-1}}\\ &+B_{t_{N(\pi)-1}}^2-B_{t_{N(\pi)-1}}B_{t_{N(\pi)-2}}+\ldots\\ &+B_{t_1}^2+B_{t_1}B_0.\end{aligned}$$

Similarly,

$$\begin{aligned}S_{\pi,2}:=\sum_{j=0}^{N(\pi)-1}B_{t_j}(B_{t_{j+1}}-B_{t_j}) &= B_TB_{t_{N(\pi)-1}}-B_{t_{N(\pi)-1}}^2\\ &+B_{t_{N(\pi)-1}}B_{t_{N(\pi)-2}}-B_{t_{N(\pi)-2}}^2+\ldots\\ &+B_0B_{t_1}-B_0^2.\end{aligned}$$

Therefore follows that

$$S_{\pi,1}=B_T^2-B_0^2-S_{\pi,2}.$$

Thus we conclude that

$$\begin{aligned}2S_{\pi,1} &= B_T^2-B_0^2+\sum_{j=0}^{N(\pi)-1}(B_{t_{j+1}}-B_{t_j})(B_{t_{j+1}}-B_{t_j})\\ &\to B_T^2-B_0^2+T\end{aligned}$$

as $||\pi|| \to 0$ in probability since $\langle B\rangle_T = T$ and by the definition of quadratic variation. $\square$

The Stratonovich integral is defined as

$$\int_0^T B_s \circ dB_s = \lim_{||\pi||\to 0} \sum_{j=0}^{N(\pi)-1} \frac{1}{2}(B_{t_{j+1}} + B_{t_j})(B_{t_{j+1}} - B_{t_j}).$$

Lemma 9.19. *It holds that*

$$\int_0^T B_s \circ dB_s = \frac{1}{2}(B_T^2 - B_0^2).$$

Proof. It holds that $\frac{1}{2}(B_{t_{j+1}} + B_{t_j})(B_{t_{j+1}} - B_{t_j}) = \frac{1}{2}(B_{t_{j+1}}^2 - B_{t_j}^2)$ and therefore the claim follows by noticing that the above sum is telescoping. $\square$

10. B8.2 Class 4

10.1. **Stochastic Integrals.** We recall some definitions.

Definition 10.1. *We define*

$$\mathscr{H}^{2,c} = \left\{ \textit{continuous martingales } (M_t)_{t\geq 0} \textit{ with } \sup_{t\geq 0} \mathbb{E}[M_t^2] < \infty \right\}.$$

This is a Hilbert space when endowed with the norm

$$||M||^2_{\mathscr{H}^{2,c}} = \mathbb{E}[M_\infty^2].$$

Definition 10.2. *An adapted right-continuous process $A = (A_t)_{t\geq 0}$ is called a finite variation process if $A_0 = 0$ and $t \to A_t$ is of finite variation a.s.*

Definition 10.3. *Let A be a finite variation process. Then we define $L^1(|dA|)$ as the space of progressively measurable processes such that for all $t \geq 0$,*

$$\int_0^t |K_t|\,|dA_t| < \infty$$

almost surely.

A few remarks:

(i) (Proposition 7.13) If $K \in L^1(|dA|)$, then
$$(K \cdot A)_t = \int_0^t K_s\,dA_s$$
is a finite variation process.

(ii) (Proposition 7.6) If $F \in L^1(|dA|)$ and $KF \in L^1(|dA|)$, then
$$\int_0^t K_s F_s\,dA_s = ((KF)\cdot A)_t = (K\cdot(F\cdot A))_t = \int_0^t K_s\,d(F\cdot A)_s.$$

(iii) $A_t = t$ for all $t \geq 0$ is a finite variation process. Thus the above definition generalizes the Lebesgue integral on $\mathbb{R}$.

Definition 10.4. *Given $M \in \mathscr{H}^{2,c}$ we denote by $L^2(M)$ the space of progressively measurable processes K such that*

$$\mathbb{E}\left[\int_0^\infty K_t^2\,d\langle M\rangle_t\right] < \infty.$$

Definition 10.5. *For a continuous local martingale M, denote by $L^2_{\text{loc}}(M)$ the space of progressively measurable processes K such that for all $t \geq 0$,*

$$\int_0^t K_s^2\,d\langle M\rangle_s < \infty \qquad \textit{a.s.}$$

We observe the following:

(i) (Theorem 8.5) The space $L^2(M)$ is a Hilbert space when endowed with the inner product
$$\langle H, K\rangle_{L^2(M)} = \mathbb{E}[(HK\cdot\langle M\rangle)_\infty] = \mathbb{E}\left[\int_0^\infty H_t K_t\,d\langle M\rangle_t\right].$$

Moreover, the map $L^2(M) \to \mathscr{H}^{2,c}_0, K \mapsto K \bullet M$ is an isometry, i.e.
$$\mathbb{E}\left[\int_0^\infty K_t^2\,d\langle M\rangle_t\right] = ||K||^2_{L^2(M)} = ||K\bullet M||^2_{\mathscr{H}^{2,c}} = \mathbb{E}\left[\left(\int_0^\infty K_t\,dM_t\right)^2\right].$$

(ii) (Theorem 8.12) If $K \in L^2_{\text{loc}}(M)$, there exists a unique continuous local martingale, zero in zero, denoted $K \bullet M$ such that for any continuous local martingale N,

$$\langle K \bullet M, N\rangle = K \cdot \langle M, N\rangle.$$

(iii) (Follows from (i)) if H is a further progressively measurable process and if $KH \in L^2_{\text{loc}}(M)$,

$$(KH) \bullet M = K \bullet (H \bullet M).$$

(iv) (Follows from (ii)) If τ is a stopping time

$$(K \bullet M)^\tau = K \bullet M^\tau.$$

Lemma 10.6. *(Generalized Ito Isometry) Let $M, N \in \mathscr{H}^{2,c}$ and let $K \in L^2(M)$ and $F \in L^2(N)$. Then for each $t \in [0, \infty]$,*

$$\mathbb{E}\left[\left(\int_0^t K_s\, dM_s\right)\left(\int_0^t F_s\, dN_s\right)\right] = \mathbb{E}\left[\int_0^t K_s F_s\, d\langle M, N\rangle_s\right]$$

Proof. Recall (Remark 7.29) that if M' and N' are in $\mathscr{H}_0^{2,c}$, then $\mathbb{E}[M'_t N'_t] = \mathbb{E}[\langle M', N'\rangle_t]$. Also, $K \bullet M$ and $F \bullet N$ are in $\mathscr{H}_0^{2,c}$. Note furthermore that

$$\langle K \bullet M, F \bullet N\rangle = K \cdot \langle M, F \bullet N\rangle = K \cdot (F \cdot \langle M, N\rangle) = KF \cdot \langle M, N\rangle.$$

Therefore, it follows that

$$\mathbb{E}[(K \bullet M)_t (F \bullet N)_t] = \mathbb{E}[\langle K \bullet M, F \bullet N\rangle_t] = \mathbb{E}[(KF \cdot \langle M, N\rangle)_t]$$

concluding the proof. □

Lemma 10.7. *Let M be a continuous local martingale and let $K \in L^2_{\text{loc}}(M)$. Fix $t > 0$. Then if $\mathbb{E}[\int_0^t K_s^2 d\langle M\rangle_s] < \infty$, the stopped process $(K \bullet M)^t$ is a martingale,*

$$\mathbb{E}\left[\int_0^t K_s\, dM_s\right] = 0 \quad \textit{and} \quad \mathbb{E}\left[\left(\int_0^t K_s dM_s\right)^2\right] = \mathbb{E}\left[\int_0^t K_s^2 d\langle M\rangle_s\right].$$

Proof. By Theorem 7.24 from the notes, to show that $(K \bullet M)^t$ is a martingale (bounded in L^2), it suffices to show that $\mathbb{E}[\langle K \bullet M^t\rangle_\infty] < \infty$. The latter quantity is equal to $\mathbb{E}[\int_0^t K_s^2 d\langle M\rangle_s]$ and therefore the claim follows. Since $(K \bullet M)^t$ is a martingale bounded in L^2 and therefore uniformly integrable,

$$\mathbb{E}\left[\int_0^t K_s\, dM_s\right] = \mathbb{E}[(K \bullet M)^t_\infty] = \mathbb{E}[(K \bullet M)^t_0] = 0.$$

Finally, by Theorem 7.24 $((K \bullet M)^t)^2 - \langle (K \bullet M)^t\rangle$ is uniformly integrable and therefore, since it is zero at zero the final claim follows. □

Lemma 10.8. *Let f be a continuous function on $[0, \infty)$ and let B be a standard Brownian motion. Then for $t \geq 0$,*

$$X_t = \int_0^t f(s)\, dB_s$$

is Gaussian and $\operatorname{cov}(X_t, X_r) = \int_0^{t \wedge r} f(s)^2\, ds$ for $t, s \geq 0$. Moreover, X_t is a martingale and a Gaussian process.

Proof. Fix $t \geq 0$. If f is a step function, then X_t is a sum of independent Gaussian and therefore Gaussian itself. Let $(f_n)_{n\geq 0}$ be a sequence of step functions such that $f_n \to f$ uniformly on $[0,t]$. Recall that $B^t \in \mathscr{H}^{2,c}$ and $f, f_n \in L^2(B^t)$. Then it holds that

$$\begin{aligned} ||f_n - f||^2_{L^2(B^t)} &= \mathbb{E}\left[\int_0^t (f(s) - f_n(s))^2 \, d\langle B\rangle_s\right] \\ &= \int_0^t (f(s) - f_n(s))^2 \, ds \leq t \cdot ||f - f_n||_\infty. \end{aligned}$$

Thus $f_n \to f$ in $L^2(B^t)$ and therefore $f_n \bullet B^t \to f \bullet B^t$ in $\mathscr{H}^{2,c}_0$. Thus

$$||(f\bullet B)_t - (f_n \bullet B)_t||_2 = ||(f\bullet B)^t_\infty - (f_n\bullet B)^t_\infty||_2 = ||(f\bullet B)^t - (f_n\bullet B)^t||_{\mathscr{H}^{2,c}} \to 0,$$

showing that $(f_n \bullet B)_t$ converges to $(f \bullet B)_t$ in L^2. Since the space of Gaussian's is L^2 closed, it follows that $(f \bullet B)_t$ is a Gaussian. It holds that $\mathbb{E}[X_t] = 0$ since $\mathbb{E}[(f_n \bullet X)_t] = 0$ and by Ito's isometry,

$$\begin{aligned} \operatorname{cov}(X_t, X_r) &= \mathbb{E}[X_t X_r] \\ &= \mathbb{E}[(f \bullet B)^t_\infty (f \bullet B)^r_\infty] \\ &= \mathbb{E}[\langle f \bullet B, f \bullet B\rangle^{t\wedge r}_\infty] = \int_0^{t\wedge r} f(s)^2 \, ds. \end{aligned}$$

Next we show that $X_t = \int_0^t f(s)\, dB_s$ is a martingale. Indeed $X_t = (f \bullet B)_t$ is a local continuous martingale and $\langle X\rangle_t = (f^2 \cdot \langle B\rangle)_t$. Therefore

$$\mathbb{E}[\langle X\rangle_t] = \mathbb{E}\left[\int_0^t f^2(s)\, ds\right] = \int_0^t f^2(s)\, ds < \infty.$$

It follows that X is a martingale by Theorem 7.24.

Finally we show that X_t is a Gaussian process. Indeed if $f_n \to f$ is again a approximating sequence of step functions converging uniformly to f, then for each n the random vector $((f_n \bullet B)_{t_1}, \dots, (f_n \bullet B)_{t_n})$ is a multivariate Gaussian since Brownian motion is a Gaussian process. As $((f_n \bullet B)_{t_1}, \dots, (f_n \bullet B)_{t_n})$ converges to $(X_{t_1}, \dots, X_{t_n})$ in L^2, it follows that X is a Gaussian process. ☐

10.2. **Continuous Semi-Martingales and Ito's Theorem.**

Definition 10.9. *A stochastic process $X = (X_t)_{t\geq 0}$ is called a continuous semi-martingale if it can be written as*

$$X_t = X_0 + M_t + A_t,$$

where M is a continuous local martingale, A is a continuous process of finite variation and $M_0 = A_0 = 0$.

Definition 10.10. *Let $X = X_0 + M + A$ be a continuous semimartingale. Then*

$$L(X) = L^2_{\text{loc}}(M) \cap L^1(|dA|)$$

and for $K \in L(X)$ we define

$$K \bullet X = K \bullet M + K \cdot A.$$

We recall that if $X = X_0 + M + A$ is a continuous semimartingale, then

$$\langle X \rangle = \langle M \rangle.$$

The main theorem of Stochastic Calculus is Ito's Theorem, which generalizes the Fundamental Theorem of Stochastic Calculus.

Theorem 10.11. *Let $X^1, \ldots, X^d$ be continuous semiartingales and $F : \mathbb{R}^d \to \mathbb{R}$ a C^2-function. Then $(F(X_t^1, \ldots, X_t^d))_{t\geq 0}$ is a continuous semimartingale and up to indistinguishability,*

$$\begin{aligned} F(X_t^1, \ldots, X_t^d) = F(X_0^1, \ldots, X_0^d) &+ \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x_i}(X_s^1, \ldots, X_s^d)\, dX_s^i \\ &+ \frac{1}{2} \sum_{1\leq i,j\leq d} \int_0^t \frac{\partial^2 F}{\partial x_i \partial x_j}(X_s^1, \ldots, X_s^d)\, d\langle X^i, X^j \rangle_s. \end{aligned}$$

In particular, for $d = 1$, we have that

$$F(X_t) = F(X_0) + \int_0^t F'(X_s)\, dX_s + \frac{1}{2} \int_0^t F''(X_s)\, d\langle X \rangle_s.$$

We note that in Ito's theorem, the integrability of all of the processes involved follows from continuity of F.

Lemma 10.12. *Suppose that $(B_t)_{t\geq 0}$ is a standard Brownian motion and let f and g be C^2-functions. Then*

$$Y_t = \exp\left(f(B_t) - \int_0^t g(B_s)\, ds \right)$$

is a local martingale whenever $g(x) = \frac{1}{2}(f''(x) + f'(x)^2)$.

Proof. Consider $F(x, y) = \exp(f(x) - y)$ and consider $X_t^1 = B_t$ and $X_t^2 = \int_0^t g(B_s)\, ds$. Since we can view $ds = \langle B \rangle_s$, X^2 is a finite variation process and therefore $\langle X^2 \rangle = 0 = \langle X^1, X^2 \rangle$. Note that

$$\int_0^t F(X_t^1, X_t^2)\, dX_s^2 = \int_0^t F(X_t^1, X_t^2) g(B_s)\, ds$$

is a process of finite variation.

Thus by Ito's Theorem and since $\frac{\partial F}{\partial x}(x, y) = F(x, y) f'(x)$, $\frac{\partial F}{\partial y}(x, y) = -F(x, y)$ and $\frac{\partial^2 F}{\partial^2 x}(x, y) = F(x, y) f''(x) + F(x, y)(f'(x))^2$,

$$\begin{aligned} Y_t = F(X_t^1, X_t^2) = \exp(f(B_0)) &+ \int_0^t \frac{\partial F}{\partial x}(X_s^1, X_s^2)\, dB_s \\ &- \int_0^t F(X_s^1, X_s^2) g(B_s)\, ds + \frac{1}{2} \int_0^t F(X_s^1, X_s^2)(f''(B_s) + f'(B_s)^2)\, ds, \end{aligned}$$

which is a decomposition of Y into a local martingale part and processes of finite variation. The finite variation part is zero if and only if $g(x) = \frac{1}{2}(f''(x) + f'(x)^2)$ for all $x \in \mathbb{R}$ and therefore the claim follows. ☐

Lemma 10.13. *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Then*

$$M_t = e^{t/2} \cos(B_t)$$

is a martingale.

Proof. Let $F(x,y) = e^{x/2}\cos(y)$ and $X_t^1 = t$ and $X_t^2 = B_t$. Then it holds by Ito's Theorem that

$$\begin{aligned} M_t = F(X_t^1, X_2^t) &= \cos(B_0) + \int_0^t e^{s/2}\cos(B_s)\,ds \\ &- \int_0^t e^{s/2}\sin(B_s)\,dB_s - \int_0^t e^{s/2}\cos(B_s)\,ds \\ &= \cos(B_0) - \int_0^t e^{s/2}\sin(B_s)\,dB_s. \end{aligned}$$

Therefore M_t is a continuous local martingale. Thus to show that it is a martingale, it suffices by Theorem 7.24 to check that $\mathbb{E}[\langle M\rangle_t] < \infty$. It follows that $\langle M\rangle_t = \int_0^t e^s \sin^2(B_s)\,ds$ and therefore

$$\big|\,\mathbb{E}[\langle M\rangle_t]\,\big| \leq \int_0^t e^s\,ds = e^t - 1 < \infty.$$

□

10.3. **A primer on stochastic differential equations.** A stochastic differential equation determines a process by

$$dX_t = f(t, X_t)dt + g(t, X_t)dB_t,$$

for two functions $f, g : \mathbb{R}^2 \to \mathbb{R}$. More precisley, X_t is a stochastic process that satisfies

$$X_t = X_0 + \int_0^t f(t, X_t)\,dt + \int_0^t g(t, X_t)\,dB_t.$$

It can be shown that under weak assumptions on f and g, the process X_t exists, is a continuous semimartingale and is uniquely determined by f and g.

For simplicity we study in this exposition the stochastic differential equation

$$dX_t = \sigma X_t dB_t \qquad \text{or equivalently} \qquad X_t = X_0 + \sigma\cdot(X \bullet B)_t \tag{10.1}$$

for $\sigma > 0$. This is a model for the price of a stock with volatility σ.

We now apply Ito's Lemma to understand (10.1). Indeed let $F(x) = \log(x)$, then by associativity of the stochastic integral,

$$\begin{aligned} \log(X_t) &= \log(X_0) + \int_0^t \frac{1}{X_s}\,dX_s - \frac{1}{2}\int_0^t \frac{1}{X_s^2}\,d\langle X\rangle_s \\ &= \log(X_0) + \left(\frac{1}{X}\bullet X\right)_t - \frac{1}{2}\left(\frac{1}{X^2}\cdot\langle X\rangle\right)_t \\ &= \log(X_0) + \sigma\left(\frac{1}{X}\bullet(X\bullet B)\right)_t - \frac{\sigma^2}{2}\left(\frac{1}{X^2}\cdot X^2\langle B\rangle\right)_t \\ &= \log(X_0) + \sigma B_t - \frac{\sigma^2 t}{2}. \end{aligned}$$

Therefore it follows that

$$X_t = X_0 \exp\left(\sigma B_t - \frac{\sigma^2 t}{2}\right).$$

Recall that if $Y \sim \mathcal{N}(\mu, \sigma^2)$, then

$$E[\exp(cY)] = \exp\left(\frac{c^2\sigma^2}{2} + c\mu\right).$$

We now assume that X_0 is constant. Then it follows that $E[X_t] = X_0$ and

$$\begin{aligned}\operatorname{Var}(X_t) &= E[(X_t - X_0)^2] \\ &= E[X_t^2] - X_0^2 \\ &= X_0^2 \cdot \big(E\left[\exp(2\sigma B_t)\right]\exp(-\sigma^2 t) - 1\big) \\ &= X_0^2 \cdot (\exp(\sigma^2 t) - 1).\end{aligned}$$

10.4. **Harmonic Functions, the Heat Equation and Brownian Motion.** Let $\Omega \subset \mathbb{R}^d$ be an open bounded subset.

Definition 10.14. *A continuous function $h : \Omega \to \mathbb{R}$ is said to be **harmonic** if for each $x \in \Omega$ and $r < d(x, \partial\Omega)$ the mean-value property is satisfied*

$$h(x) = \int_{\partial B_r(x)} h(z)\, \mathrm{dvol}_{\partial B_r(x)}(z),$$

where $\partial B_r(x) = \{z \in \mathbb{R}^d : d_{\mathbb{R}^d}(x, z) = r\}$.

We recall that a harmonic function is an eigenfunction of the Laplacian and smooth.

Definition 10.15. *(Dirichlet problem) Given a continuous function f on $\partial\Omega$, does there exists a harmonic function h with $h|_{\partial\Omega} = f$.*

The Dirichlet problem can be solved very elegantly by using Brownian motion. Indeed, we consider the function

$$h(x) = E_x[f(B_T)],$$

where T is the first hitting time of the boundary $\partial\Omega$. We won't show that h defined as above is continuous, yet we prove that it satisfies the mean-value property.

Indeed, let $x \in \Omega$ and $r < d(x, \partial\Omega)$. We consider a Brownian motion starting at x. Let τ be the first hitting time of $\partial B_r(x)$. Then it holds that

$$h(x) = E_x[f(B_T)] = E_x[E[f(B_T)|\mathscr{F}_\tau]],$$

as conditioning doesn't alter the expected value. On the other hand, by the strong Markov property, conditionally on $\mathscr{F}_\tau$ the process $(B_{\tau+t})_{t\geq 0}$ is a Brownian motion starting from B_τ, thus

$$E_x[E[f(B_T)|\mathscr{F}_\tau]] = E_x[E[f(B_T)|\mathscr{F}_\tau]] = E_x[E_{B_\tau}[f(B_T)]] = E_x[h(B_\tau)].$$

The claim follows since when starting at x the point B_τ is by symmetry uniformly distributed on $\partial B_r(x)$.

In similar vein, we can use Brownian motion to solve the Heat equation. Let $f : \mathbb{R} \to \Omega$ be a continuous function and assume that $f|_{\partial\Omega} \equiv 0$. We now want to find a continuous solution $u : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ satisfying for all $x \in \Omega$,

$$u(0, x) = f(x), \qquad \partial_t u = \tfrac{1}{2}\partial_{xx} u \quad \text{for all } t > 0$$

and

$$u(t, x) = 0$$

for all $t > 0$ and $x \in \partial\Omega$. Then the solution to the Heat equation is given as

$$u(t, x) = E_x[f(B_t)1_{\{t<T\}}],$$

where again T is the first hitting time of the boundary.

11. C8.2: Class 1

11.1. **Markov Processes.**

11.1.1. *Measures and Operators.* We first make some general remarks between measures and operators. Indeed, we state the following theorem.

Theorem 11.1. *(Riesz-Markov-Kakutani representation theorem) Let E be a compact space and let $C(E)$ be the space of real-valued continuous functions on E. Let $\Phi : C(E) \to \mathbb{R}$ be a positive linear map, i.e. if $f \geq 0$ then $\Phi(f) \geq 0$.*

Then there exists a unique measure ν on E that represents Φ in the sense that for every $f \in C(E)$ it holds that

$$\Phi(f) = \int f \, d\nu.$$

In particular we have a bijection

$$\{\text{pos. linear maps on } C(E) \to \mathbb{R}\} \overset{1:1}{\longleftrightarrow} \{\text{measures on } E\}.$$

Moreover, we can easily restrict to probability measures

$$\{\text{pos. linear maps } C(E) \to \mathbb{R} \text{ with } \Phi(1_E) = 1\} \overset{1:1}{\longleftrightarrow} \{\text{prob. measures on } E\},$$

where 1_E is the function being constant $\equiv 1$ on E.

11.1.2. *Definition and general remarks.*

Definition 11.2. *Let $(E, \mathscr{E})$ be a measurable space. A Markovian transition kernel (or a Markov kernel) from E into E is a mapping $T : E \times \mathscr{E} \to [0, 1]$ such that:*

1. *For every $x \in E$, the mapping $\mathscr{E} \ni A \mapsto T(x, A)$ is a probability measure on $(E, \mathscr{E})$.*
2. *For every $A \in \mathscr{E}$, the mapping $E \ni x \mapsto T(x, A)$ is $\mathscr{E}$-measurable.*

Given a Markov kernel T and a bounded measurable function $f : E \to \mathbb{R}$ we define

$$(Tf)(x) = \int_E f(y) T(x, dy).$$

Definition 11.3. *A collection $(T_t)_{t \geq 0}$ of transition kernels on E is called a transition semigroup if the following three properties hold:*

1. *For every $x \in E$ and $A \in \mathscr{E}$, $T_0(x, A) = \delta_x(A)$.*
2. *(Chapman-Kolmogorov identity) For every $s, t \geq 0$ and $A \in \mathscr{E}$ we have*
$$T_{t+s}(x, A) = \int_E T_t(x, dy) T_s(y, A).$$
3. *For every $A \in \mathscr{E}$, the function $(t, x) \mapsto T_t(x, A)$ is measurable with respect to the σ-field $\mathscr{B}(\mathbb{R}_+) \otimes \mathscr{E}$.*

Let $B(E)$ be the vector space of all bounded measurable real functions on E, which is equipped with the norm $||f|| = \sup\{|f(x)| : x \in E\}$. Then the linear mapping $B(E) \ni f \mapsto T_t f$ is a contraction of $B(E)$. From this point of view, the Chapman-Kolmogorov identity is equivalent to the relation for every $s, t \geq 0$,

$$T_{t+s} = T_t T_s.$$

Definition 11.4. *Let $(T_t)_{t\geq 0}$ be a transition semigroup on E. A Markov process (with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$) with transition semigroup $(T_t)_{t\geq 0}$ is an $(\mathscr{F}_t)$-adapted process $(X_t)_{t\geq 0}$ with values in E such that, for every $s, t \geq 0$ and $f \in B(E)$ we have*

$$\mathbb{E}[f(X_{s+t})|\mathscr{F}_s] = (T_t f)(X_s).$$

We note that a Markov process is a martingale if and only if for all $t \geq 0$,

$$T_t 1 = 1.$$

More abstractly, we can also study contraction semi-groups defined as follows. We use some strong continuity assumptions to make our life easier.

Definition 11.5. *A family of bounded operators $(T_t)_{t\geq 0}$ on a Banach space B is called a strongly continuous contraction semigroup if the following properties hold:*

(1) $T_0 = I$,
(2) $T_{s+t} = T_s T_t$ *for all* $s, t \in \mathbb{R}_{\geq 0}$,
(3) $||T(t)|| \leq 1$ *for all* $t \geq 0$,
(4) for any $z \in B$ *the map* $t \mapsto T_t z$ *is continuous.*

Proposition 11.6. *Let T_t be a strongly continuous semigroup on a Banach space B and define for $t > 0$,*

$$A_t = \frac{1}{t}(T_t - I).$$

Let $\mathcal{D}(A) = \{z \in B \,:\, \lim_{t\to 0} A_t z \text{ exists}\}$. Then for $z \in \mathcal{D}(A)$ we define

$$Az = \lim_{t\to 0} A_t z = \left.\frac{dT_t}{dt}\right|_{t=0} z.$$

Then A is a densely defined closed operator and it is called the infinitesimal generator of T_t. Moreover, for any $z \in \mathcal{D}(A)$ and $t \geq 0$ we have that $T_t z \in \mathcal{D}(A)$ and

$$\frac{dT_t z}{dt} = AT_t z = T_t A z.$$

Definition 11.7. *Let $\lambda > 0$. The λ-resolvent of the transition semigroup $(T_t)_{t\geq 0}$ is the linear operator $R_\lambda : B(E) \to B(E)$ defined by*

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda t} T_t f(x)\, dt$$

for $f \in B(E)$ and $x \in E$.

Similarly we can define the operator R_λ for a strongly continuous contraction semigroup. The following then holds:

Lemma 11.8. *Let $\lambda > 0$ and let $(T_t)_{t\geq 0}$ be a strongly continuous contraction semigroup on a Banach space B. The operator $(\lambda - A) : \mathcal{D}(A) \to C$ is invertible with inverse*

$$(\lambda - A)^{-1} = R_\lambda.$$

Proof. This is a little exercise with partial integration. ☐

We can now prove an interesting property the resolvent of strongly continuous contraction semigroups by knowing they have an inverse.

Lemma 11.9. *Let* $(T_t)_{t\geq 0}$ *be a strongly continuous contraction semigroup and let* R_λ *be the associated resolvents. Then for* $\lambda, \mu > 0$ *it holds that*

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu.$$

Proof. It is clear that R_λ and R_μ commute as T_t commutes and since

$$R_\lambda R_\mu = \int_0^\infty \int_0^\infty e^{-\lambda t_1 - \mu t_2} T_{t_1+t_2}\, dt_1 dt_2.$$

So it follows on $\mathcal{D}(A)$ by the previous lemma that

$$\begin{aligned} R_\lambda - R_\mu &= R_\lambda R_\mu R_\mu^{-1} - R_\mu R_\lambda R_\lambda^{-1} \\ &= R_\lambda R_\mu (R_\mu^{-1} - R_\lambda^{-1}) \\ &= R_\lambda R_\mu (\mu - \lambda). \end{aligned}$$

Since $\mathcal{D}(A)$ is dense the claim follows. □

Lemma 11.10. *Let* $(T_t)_{t\geq 0}$ *be a strongly continuous contraction semigroup and let* R_λ *be the associated resolvents. Then if* $|\lambda - \mu| < ||R_\lambda||^{-1}$, *we have that*

$$R_\mu = \sum_{n=0}^\infty (\lambda - \mu)^n R_\lambda^{n+1}.$$

In particular, on $\mathcal{D}(A)$ *it holds that*

$$(\mu - A)^{-1} = R_\mu = \sum_{n=0}^\infty (\lambda - \mu)^n R_\lambda^{n+1}.$$

Proof. Note that since by assumption $||(\lambda - \mu) R_\lambda|| \leq |\lambda - \mu| \cdot ||R_\lambda|| < 1$, it follows by standard properties of Banach spaces that

$$(I - (\lambda - \mu) R_\lambda)^{-1} = \sum_{n=0}^\infty (\lambda - \mu)^n R_\lambda^n.$$

By the resolvent equation it follows that

$$R_\lambda = R_\mu - (\lambda - \mu) R_\lambda R_\mu = R_\mu (I - (\lambda - \mu) R_\lambda).$$

By the above formula for $(I - (\lambda - \mu) R_\lambda)^{-1}$ it therefore follows that

$$R_\mu = R_\lambda (I - (\lambda - \mu) R_\lambda)^{-1} = \sum_{n=0}^\infty (\lambda - \mu)^n R_\lambda^{n+1}.$$

□

Note that by (iii) of the definition of a transition semigroup it holds that the mapping $t \mapsto T_t f(x)$ is measurable and clearly bounded, so the resolvent always exist in the case of transition semigroups. We now give an alternative proof using abstractly only properties of the transition semigroup.

Lemma 11.11. *Let* $(T_t)_{t\geq 0}$ *be a transition semigroup. If* $\lambda, \mu > 0$, *it holds that*

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu.$$

Proof. We calculate

$$\begin{aligned}
R_\lambda R_\mu &= \int_0^\infty e^{-\lambda t_1} T_{t_1} R_\mu \, dt_1 \\
&= \int_0^\infty e^{-\lambda t_1} T_{t_1} \left(\int_0^\infty e^{-\mu t_2} T_{t_2} \, dt_2 \right) dt_1 \\
&= \int_0^\infty e^{-\lambda t_1} \left(\int_0^\infty e^{-\mu t_2} T_{t_1+t_2} \, dt_2 \right) dt_1 \\
&= \int_0^\infty e^{-\lambda t_1 + \mu t_2} \left(\int_0^\infty e^{-\mu(t_1+t_2)} T_{t_1+t_2} \, dt_2 \right) dt_1 \\
&= \int_0^\infty e^{-\lambda t_1 + \mu t_1} \left(\int_{t_1}^\infty e^{-\mu r} T_r \, dr \right) dt_1 \\
&= \int_0^\infty e^{-(\lambda-\mu) t_1} \left(\int_{t_1}^\infty e^{-\mu r} T_r \, dr \right) dt_1 \\
&= \int_0^\infty e^{-\mu r} T_r \left(\int_0^r e^{-(\lambda-\mu) t_1} \, dt_1 \right) dr \\
&= \int_0^\infty T_r \left(\frac{e^{-\mu r} - e^{-\lambda r}}{\lambda - \mu} \right) dr,
\end{aligned}$$

concluding the proof. □

We give a probabilistic interpretation of the resolvent equation. Denote by $(Z_\lambda)_{\lambda \geq 0}$ a collection of independent exponential distributions of parameter λ and denote their densities by f_λ. Then a direct calculation shows that the density of $Z_\lambda + Z_\mu$ is for $x \geq 0$ given by

$$(f_\lambda * f_\mu)(x) = \frac{\lambda\mu}{(\mu - \lambda)}(e^{-\lambda x} - e^{-\mu x}) = \frac{\mu f_\lambda(x) - \lambda f_\mu(x)}{(\mu - \lambda)}. \tag{11.1}$$

To connect this to transition functions we note that

$$\lambda R_\lambda = \int_0^\infty f_\lambda(t) T_t \, dt = \mathbb{E}[T_{Z_\lambda}].$$

Also it holds that

$$(\lambda R_\lambda)(\mu R_\mu) = \mathbb{E}[T_{Z_\lambda}]\mathbb{E}[T_{Z_\mu}] = \mathbb{E}[T_{Z_\lambda} T_{Z_\mu}] = \mathbb{E}[T_{Z_\lambda + Z_\mu}].$$

We can therefore easily deduce the resolvent equation by using (11.1). Indeed,

$$\begin{aligned}
\lambda\mu R_\lambda R_\mu &= \mathbb{E}[T_{Z_\lambda + Z_\mu}] \\
&= \int_0^\infty (f_\lambda * f_\mu)(t) T_t \, dt \\
&= \int_0^\infty \frac{\mu f_\lambda(t) - \lambda f_\mu(t)}{(\mu - \lambda)} T_t \, dt \\
&= \frac{\mu}{(\mu - \lambda)} R_\lambda - \frac{\lambda}{(\mu - \lambda)} R_\mu,
\end{aligned}$$

which easily implies the resolvent equation. Thus from the probabilistic interpretation of the resolvent, we observe that the resolvent equation is nothing more than a statement about the product of densities of exponential random variables.

11.1.3. *Cauchy Process.* As example, let us discuss the Cauchy process. Indeed it is given as the process for which $X_{st} - X_s$ is distributed as a Cauchy random variable with density

$$\frac{1}{\pi}\frac{t}{t^2+x^2}$$

and increments to disjoint time intervals are independent.

Denote by T_t the expectation semigroup of X that is $T_t f(x) = \mathbb{E}[f(X_t) \mid X_0 = x]$. We note that for $t > 0$ we have

$$\frac{T_t f(x) - f(x)}{t} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(x+y) - f(x)}{t^2+y^2}\, dy.$$

Now assume that f is $C_c^2(\mathbb{R})$. Then by Taylor's theorem for every $x, y \in \mathbb{R}$ there exists $\xi_{x,y} \in [x, x+y]$ such that

$$f(x+y) - f(x) = yf'(x) + \frac{y^2}{2} f''(\xi_{x,y}).$$

Plugging the latter into the first equation it follows that

$$\frac{T_t f(x) - f(x)}{t} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{yf'(x) + \frac{y^2}{2} f''(\xi_{x,y})}{t^2+y^2}\, dy$$

We now observe that $\frac{yf'(x)}{t^2+y^2}$ is an odd function and therefore integrates to zero. Thus we conclude that

$$\frac{T_t f(x) - f(x)}{t} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y^2 f''(\xi_{x,y})}{2(t^2+y^2)} dy.$$

Letting t tend to zero it therefore follows that

$$\lim_{t\to 0} \frac{T_t f(x) - f(x)}{t} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f''(\xi_{x,y})}{2}\, dy,$$

which is a description of the infinitesimal generator.

11.1.4. *Finite State Spaces.* We now consider the case that E is finite and $\mathscr{E}$ is the corresponding power set. We can view the transition probabilities T_t as matrices and the infitesimal operator Q is a matrix as well and it follows that

$$T_t = \exp(tQ).$$

Lemma 11.12. *In the above setting, the $Q = (q_{ij}))$-matrix satisfies the following properties:*

(1) $q_{ij} \geq 0$ *for all* $i \neq j$.
(2) $\sum_{k\in E} q_{ik} = 0$ *for all* $i \in E$.

Proof. As $Q = \lim_{t\to 0} \frac{1}{\varepsilon}(T_t - I)$ it holds that all non-diagonal entries of the right-hand side are ≥ 0 and therefore (1) follows. For (2) we note that $T_t 1 = 1$ and it holds that $T_t 1 - I1 = 0$ and so the same holds for Q. □

Assume for a moment that we have a basis $v_1, \ldots, v_n$ of eigenvectors of A. Then we note that if $Qv_i = \lambda_i v_i$ it holds that

$$T_t v_i = \exp(tQ)v_i = \sum_{n=0}^{\infty} \frac{t^n Q^n}{n!} v_i = \sum_{n=0}^{\infty} \frac{(t\lambda_i)^n}{n!} v_i = \exp(t\lambda_i)v_i.$$

So an eigenvalue decomposition describes the dynamics of T_t very well.

For a concrete example suppose now that we have three states $\{i, j, k\}$ and the infinitesimal generator Q is given by

$$Q = \begin{pmatrix} -3 & 1 & 2 \\ 0 & -2 & 2 \\ 2 & 1 & -3 \end{pmatrix}.$$

Let's note that

$$v_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_2 = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \qquad \text{and} \qquad v_3 = \begin{pmatrix} 2 \\ 2 \\ -3 \end{pmatrix}$$

are eigenvectors of A with eigenvalues $0, -3, -5$. So it follows that if $v = c_1 v_1 + c_2 v_2 + c_3 v_3$ it holds that

$$T_t v = c v_1 + \exp(-3t) c_2 v_2 + \exp(-5t) c_3 v_3,$$

which shows that T_t converges to the uniform distribution. Moreover it holds that

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \frac{4}{15} v_1 - \frac{1}{3} v_2 + \frac{1}{5} v_3$$

and therefore it follows that

$$\mathbb{P}_i[X_t = i] = \frac{4}{15} + \frac{1}{3} e^{-3t} + \frac{2}{5} e^{-5t}.$$

11.1.5. *Feller Semigroups.* In this section we work with compact spaces E. We note that we can always compactify a metric space (with a finite metric) by adding a point at infinity.

Definition 11.13. *Let E be a compact space and let $C(E)$ be the Banach space of continuous functions on E. A strongly continuous contraction semigroup on $C(E)$ with the additional properties:*

(1) $T_t 1 = 1$ and
(2) $T_t f \geq 0$ for all non-negative $f \in C(E)$

is called a Feller semigroup.

Lemma 11.14. *If X is a Feller process and f is a non-negative function, then for $\lambda > 0$*

$$Y_t^\lambda = e^{-\lambda t} R_\lambda f(X_t)$$

is a supermartingale as $t \geq 0$.

Proof. We first note that by the definition of a Markov process

$$Y_t^\lambda = e^{-\lambda t} \int_0^\infty e^{-\lambda r} (T_r f)(X_t)\, dr = \int_0^\infty e^{-\lambda(t+r)} \mathbb{E}[f(X_{t+r}) \,|\, \mathscr{F}_t]\, dr.$$

Therefore by the tower property of conditional expectation and again using the definition of a Markov process for $s \leq t$,

$$\begin{aligned}
\mathbb{E}[Y_t^\lambda|\mathscr{F}_s] &= \int_0^\infty e^{-\lambda(t+r)}\mathbb{E}\big[\mathbb{E}[f(X_{t+r}) \mid \mathscr{F}_t] \mid \mathscr{F}_s\big]\,dr \\
&= \int_0^\infty e^{-\lambda(t+r)}\mathbb{E}[f(X_{t+r}) \mid \mathscr{F}_s]\,dr \\
&= \int_0^\infty e^{-\lambda(t+r)}T_{(t-s)+r}f(X_s)\,dr \\
&= e^{-\lambda s}\int_0^\infty e^{-\lambda((t-s)+r)}T_{(t-s)+r}f(X_s)\,dr \\
&= e^{-\lambda s}\int_{(t-s)}^\infty e^{-\lambda r}T_r f(X_s)\,dr,
\end{aligned}$$

where we have substituted $(t-s)+r$ with r in the last line. Now using that X is Feller and f is non-negative it follows that

$$\begin{aligned}
\mathbb{E}[Y_t^\lambda|\mathscr{F}_s] &= e^{-\lambda s}\int_{(t-s)}^\infty e^{-\lambda r}T_r f(X_s)\,dr \\
&\leq e^{-\lambda s}\int_0^\infty e^{-\lambda r}T_r f(X_s)\,dr \\
&= e^{-\lambda s}R_\lambda f(X_s) = Y_s^\lambda,
\end{aligned}$$

completing the proof. □

11.1.6. *Markov generators and Hille-Yosida Theorem.*

Definition 11.15. *A (usually unbounded) linear operator A on $C(E)$ with domain $\mathcal{D}(A)$ is said to be a Markov pregenerator if it satisfies the following conditions:*

(1) $1 \in \mathcal{D}(A)$ and $A1 = 0$,
(2) $\mathcal{D}(A)$ is dense in $C(E)$,
(3) If $f \in \mathcal{D}(A), \lambda \geq 0$ and $f - \lambda Af = g$, then

$$\min_{\zeta\in E} f(\zeta) \geq \min_{\zeta\in E} g(\zeta).$$

Lemma 11.16. *Let A be a Markov propagator and let $f \in \mathcal{D}(A), \lambda \geq 0$ and $g = f - \lambda Af$. Then $||f|| \leq ||g||$ and moreover, g determines f uniquely.*

Proof. By applying a minus sign to the defining equation of g we conclude that $(-g) = (-f) - \lambda A(-f)$ and therefore by (iii) of the definition of a Markov operator we conclude that

$$\min_{\zeta\in E} -f(\zeta) \geq \min_{\zeta\in E} -g(\zeta),$$

which implies that $||f|| \leq ||g||$. To shows that g uniquely determines f, let $f_1, f_2 \in \mathcal{D}(A)$ be such that $f_1 - \lambda Af_1 = f_2 - \lambda Af_2$. Then it follows that $(f_1 - f_2) - \lambda A(f_1 - f_2) = 0$ and therefore by the established properties that $f_1 = f_2$. □

Lemma 11.17. *Let A be a linear operator on $C(E)$ and assume that that if $f \in \mathcal{D}(A)$ and $f(\eta) = \min_{\zeta\in E} f(\zeta)$ for $\eta \in E$, then $Af(\eta) \geq 0$. Then A satisfies property (3) of the definition of a Markov pregenerator.*

Proof. Let $f \in \mathcal{D}(A)$, $\lambda \geq 0$ and $g = f - \lambda Af$. Then $g(\eta) = f(\eta) - \lambda Af(\eta) \leq f(\eta)$ by assumption. Therefore

$$\min_{\zeta \in E} f(\zeta) = f(\eta) \geq g(\eta) \geq \min_{\zeta \in E} g(\zeta).$$

□

Let's discuss some examples of Markov pregenerators:

(1) $A = G - I$ where G is a positive operator defined on all of $C(E)$ such that $G1 = 1$.

It is clear that (1) and (2) holds. To check (3) we apply Lemma 11.17. Indeed let $f \in C(E)$ and $\eta \in E$ such that $f(\eta) = \min_{\zeta \in E} f(\zeta)$. Consider then $f' = f - f(\eta)1_E$ and note that by construction $f' \geq 0$. Thus $Gf' \geq 0$ and it follows by linearity that $Gf - f(\eta)1_E \geq 0$. In particular $Af(\eta) = Gf(\eta) - f(\eta) \geq 0$, implying the claim.

(2) $E = [0, 1]$ and $Af(\eta) = \frac{1}{2} f''(\eta)$ with

$$\mathcal{D}(A) = \{f \in C(E) \,:\, f'' \in C(E), f'(0) = 0 = f'(1)\}.$$

(1) is clear and that $\mathcal{D}(A)$ is dense in $C(E)$ easily follows from the Stone-Weierstrass theorem. To check (3) we apply Lemma 11.17. Indeed let $f \in C(E)$ and $\eta \in E$ such that $f(\eta) = \min_{\zeta \in E} f(\zeta)$. If $\eta \in (0, 1)$, then $f'(\eta) = 0$ and we have by assumption that $f'(0) = f'(1) = 0$. So by Taylor's theorem it holds that

$$f(x) = f(\eta) + \frac{(x - \eta)^2}{2} f''(\xi_{\eta,x})$$

for some $\xi_{\eta,x} \in [\eta, x]$. As $x \to \eta$ and $x \in E$ and since f'' is continuous it follows that $f''(\eta) \geq 0$. Thus $Af(\eta) \geq 0$.

(3) $E = [0, 1]$ and $Af(\eta) = \frac{1}{2} f''(\eta)$ with

$$\mathcal{D}(A) = \{f \in C(E) \,:\, f'' \in C(E), f''(0) = 0 = f''(1)\}.$$

(1) and (2) holds as before. For (2) we again apply Lemma 11.17. Indeed let $f \in C(E)$ and $\eta \in E$ such that $f(\eta) = \min_{\zeta \in E} f(\zeta)$. If $\eta = 0$ or $\eta = 1$, then by assumption $Af(\eta) \geq 0$. On the other hand, if $\eta \in (0, 1)$, then $f'(\eta) = 0$ and the same argument as in example (2) applies. Thus the claim follows.

Definition 11.18. *A linear operator A on a Banach space B with domain $\mathcal{D}(A)$ is closed if its graph*

$$\operatorname{graph}(A) = \{(f, Af) \,:\, f \in \mathcal{D}(A)\}$$

is closed. In other words, if $f_n \in \mathcal{D}(A)$ is a sequence such that $(f_n, Af_n) \to (f, h)$ for some $f, h \in B$, then it holds that $(f, h) \in \operatorname{graph}(A)$ that is $h = Af$.

Definition 11.19. *Let A be a linear operator on a Banach space B. We say that A admits a closure if there exists a linear operator $\overline{A}$ such that $\mathcal{D}(A) \subset \mathcal{D}(\overline{A})$, $\overline{A}|_{D(A)} = A$ and $\operatorname{graph}(\overline{A}) = \overline{\operatorname{graph}(A)}$.*

Lemma 11.20. *The following properties hold:*

(1) A linear operator is closed if and only if it is its own closure.

(2) A linear operator A admits a closure if and only if for every sequence $f_n \in \mathcal{D}(A)$ such that $(f_n, Af_n) \to (0, y)$ for some $y \in B$ satisfies that $y = 0$.

Proof. (1) is obvious and for (2) we note that if $\overline{A}$ is a closure, then by linearity $y = \overline{A}0 = 0$ and the one direction follows. For the other direction, assume that the assumption holds. Then assume that for a sequence $f_n \in \mathcal{D}(A)$ we have that $(f_n, Af_n) \to (f, h)$ for some $f, h \in B$. Then by the assumption h is uniquely determined by f and so we set $\overline{A}f = h$. It is straightforward to check that $\overline{A}$ is a closure of A. □

With this lemma at hand, it is quite easy to show that certain operators do not admit a closure.

Lemma 11.21. *Let $E = [0,1]$ and consider the operator $Af(x) = f'(0)$ with*

$$\mathcal{D}(A) = \{f \in C([0,1]) : f'(0) \text{ exists}\}.$$

Then A does not admit a closure.

Proof. By the above lemma, it suffices to construct a sequence of functions such that $f_n \to 0$ in $C([0,1])$, yet $f_n'(0)$ does not converge to 0, which is obviously possible, as for example we can take $f_n = \frac{1}{n}(1-x)^n$. □

Definition 11.22. *A Markov generator is a closed Markov pregenerator A for which* $\mathrm{Im}(I - \lambda A) = C(E)$ *for all $\lambda \geq 0$.*

Proposition 11.23. *The following properties holds:*

1. *For a closed Markov pregenerator A, if* $\mathrm{Im}(I - \lambda A) = C(E)$ *for all sufficiently small positive λ, then A is a Markov generator.*
2. *If a Markov generator is everywhere defined and is a bounded operator, then it is a Markov generator.*

Proof. (2) follows from (1). Indeed, for $|\lambda| < ||A||^{-1}$ sufficiently small we have that $(I - \lambda A)^{-1} = \sum_{i=0}^{\infty} \lambda^n A^n$ is a bounded operator and therefore for every $g \in C(E)$ we set $f = (I - \lambda A)^{-1} g$, showing that $\mathrm{Im}(I - \lambda A) = C(E)$ for sufficiently small λ. This implies the claim by (1).

To show (1) consider the set

$$\rho(A) = \{\lambda > 0 : \mathrm{Im}(I - \lambda A) = C(E)\}.$$

We claim that $\rho(A)$ is open and closed, which implies the claim as $\rho(A)$ is non-empty by our assumption.

We first observe the following. If $\lambda \in \rho(A)$, then by Lemma 11.16 the operator $(I - \lambda A)^{-1}$ exists and its operator norm is ≤ 1. Denote by

$$R_\rho = (\rho I - A)^{-1}$$

the resolvent of A whenever it exists. Since $(I - \lambda A)^{-1} = \lambda^{-1} R_{\lambda^{-1}}$ it therefore follows by our assumption that for sufficiently large λ the resolvent R_λ exists, is bijective and satisfies

$$||R_\lambda|| \leq \lambda. \tag{11.2}$$

We first show that $\rho(A)$ is open. Indeed, if $\lambda \in \rho(A)$, then by Lemma 11.10 it follows that if $\mu \in \mathbb{R}_{>0}$ with $|\mu - \lambda| < ||R_\lambda||^{-1}$ then

$$R_\mu = \sum_{n=0}^{\infty} (\lambda - \mu)^n R_\lambda^{n+1}$$

and so $\mu \in \rho(A)$.

It remains to show that $\rho(A)$ is closed. Indeed, let $\lambda_n \in \rho(A)$ with $\lambda_n \to \lambda$. We recall that by resolvent equation,

$$R_{\lambda_n^{-1}} - R_{\lambda_m^{-1}} = (\lambda_m^{-1} - \lambda_n^{-1}) R_{\lambda_n^{-1}} R_{\lambda_m^{-1}}.$$

Since by (11.2) and as $\lambda_n \to \lambda$, it holds that $\sup_{n\geq 1} ||R_{\lambda_n^{-1}}|| < \infty$ and therefore the family of operators R_{λ_n} is Cauchy with respect to the operator norm and therefore converges to a bounded operator R. It remains to show that $R = R_{\lambda^{-1}}$, implying that $\lambda \in \rho(A)$. Indeed, let $g \in C(E)$ and consider $f_n = R_{\lambda_n^{-1}} g$. Then it follows that f_n converges to $f = Rg$ as $f_n = \lim_{n\to\infty} R_{\lambda_n^{-1}} g = Rg$. Moreover, since by construction $Af_n = \lambda_n^{-1} f_n - g$ it follows that Af_n is a Cauchy sequence as well and converges to a some element $h \in C(E)$. Since A is closed we conclude that $h = Af$ and so Af_n converges to Af. Thus it follows that

$$(\lambda_n^{-1} I - A) f = \lim_{n\to\infty} (\lambda_n^{-1} I - A) f_n = g. \tag{11.3}$$

By Lemma 11.16 it follows that given g, the solution f is unique and the resolvent $R_{\lambda^{-1}}$ exists as a bounded operator. This concludes the proof. □

Actually we have established the following:

Corollary 11.24. *For a closed Markov pregenerator A, if* $\mathrm{Im}(I - \lambda A) = C(E)$ *for a single positive λ, then A is a Markov generator.*

We now return to the previously discusses Markov pregenerators and aim to show that the closure of all of them are actually Markov generators:

(1) $A = G - I$ where G is a positive operator defined on all of $C(E)$ such that $G1 = 1$. This is a Markov generator as $||G|| \leq 1$ and so the claim follows by Lemma 11.23 (2).

(2) $E = [0, 1]$ and $Af(\eta) = \frac{1}{2} f''(\eta)$ with

$$\mathcal{D}(A) = \{f \in C(E) : f'' \in C(E), f'(0) = 0 = f'(1)\}.$$

We solve the problem with a standard ODE approach. Indeed, we want to find a solution to

$$f - \frac{\lambda}{2} f'' = g$$

with appropriate boundary conditions. To do so set $\alpha^2 = 2/\lambda$. Then the associated homogeneous equation has solution $u(x) = e^{\alpha x}$ and $v(x) = e^{-\alpha x}$. As appears to be standard in the theory of ODE's, we guess that a solution has the form $f = \phi u + \psi v$ and make the Ansatz $\phi' u + \psi' v = 0$. Substituting to the original equation, we find that ϕ' and ψ' shall satisfy

$$\phi' u + \psi' v = 0 \qquad \text{and} \qquad \phi' u' + \psi' u' = -\alpha^2 g.$$

It thus follows that by reformulating the terms that

$$\phi' = \frac{-\alpha^2 g v}{u'v - uv'} = -\frac{\alpha g v}{2} \qquad \text{and} \qquad \psi' = \frac{-\alpha^2 g v}{uv' - u'v} = \frac{\alpha g u}{2}.$$

Thus it follows that the general solution to our equation is the of the form

$$f(x) = e^{\alpha x} \int_x^1 \frac{\alpha}{2} g(y) e^{-\alpha y}\, dy + e^{-\alpha x} \int_0^x \frac{\alpha}{2} g(y) e^{\alpha y}\, dy + Ae^{\alpha x} + Be^{-\alpha x}$$

and we simply choose A and B in such a way that the boundary conditions are satisfied.

(3) $E = [0,1]$ and $Af(\eta) = \frac{1}{2}f''(\eta)$ with

$$\mathcal{D}(A) = \{f \in C(E) : f'' \in C(E), f''(0) = 0 = f''(1)\}.$$

One applies a similar method as in the previous case.

Theorem 11.25. *(Hille-Yoshida) There is a bijection*

$$\{\textit{Feller semigroups on } C(E)\} \xleftrightarrow{1:1} \{\textit{Markov generators on } C(E)\}.$$

Indeed, a Feller semigroup T_t is mapped to its derivative $A = \lim_{t\to 0} \frac{1}{t}(T_t - I)$.

11.2. **Martingale Problems.**

11.2.1. *Definition.* We use the same notation as in the previous section and denote by $\mathcal{P}(E)$ the set of probability measures on E. We denote by $D[0,\infty)$ for the space of cadlag functions from $[0,\infty)$ to E, that is the space of functions that are right continuous and has left limits. For $s \in [0,\infty)$, the evaluation mapping $\pi_s : D[0,\infty) \to E$ is defined by $\pi_s(\eta) = \eta_s$. Let $\mathscr{F}$ be the smallest σ-algebra with respect to which all the mappings π_s are measurable and for $t \in [0,\infty)$ let $\mathscr{F}_t$ be the smallest σ-algebra on $D[0,\infty)$ relative to which all the mappings π_s for $0 \le s \le t$ are measurable.

Definition 11.26. *Suppose that A is a Markov pregenerator on $C(E)$ and $\mu \in \mathcal{P}(E)$. A probability measure $\mathbb{P}$ on $D[0,\infty)$ is said to solve the martingale problem (A,μ) if*

(1) $\mathbb{P}[\zeta \in D\, \zeta_0 \in A] = \mu(A)$ *for all $A \in \mathscr{E}$ and*
(2)

$$f(\eta_t) - \int_0^t Af(\eta_s)\, ds$$

is a local martingale relative to $\mathbb{P}$ and the σ-algebras $\{\mathscr{F}_t : t \ge 0\}$ for all $f \in \mathcal{D}(A)$.

Theorem 11.27. *Suppose that E is compact and separable and that A is a Markov pregenerator for which the closure $\overline{A}$ is a Markov generator. Let $\{\mathbb{P}^x, x \in E\}$ be the unique Feller process that corresponds to $\overline{A}$. Then for each $x \in E$, $\mathbb{P}^x$ is the unique solution to the martingale problem for A with initial point x.*

11.2.2. *Discrete Time Martingales.* In the following two example we assume that the state space is discrete.

Lemma 11.28 (Sheet 2, Exercise 5a)**.** *Let E be a compact space and denote by*

- *$B(E)$ bounded Borel measurable functions on E.*
- *$\mu(x,\Gamma)$: transition function on $E \times B(E)$*
- *$\{X_n\}_{n\in\mathbb{N}}$ sequence of E-valued random variables*
- *Operator $A : B(E) \to B(E)$ given by*

$$Af(x) = \int_E f(y)\mu(x,dy) - f(x)$$

,

suppose that for each $f \in B(E)$, the following is a martingale with respect to the natural filtration generated by $(X_n)_n$:

$$f(X_n) - \sum_{k=0}^{n-1} Af(X_k)$$

Then X is a Markov Chain with transition function $\mu(x, \Gamma)$.

Proof. Let $\mathcal{F}_n$ be the natural filtration. By the martingale property, we have

$$\mathbb{E}[f(X_{n+1}) - \sum_{k=0}^{n} Af(X_k) \mid F_n] = f(X_n) - \sum_{k=0}^{n-1} Af(X_k)$$

Note that we have on the right hand side a function depending on all X_k for $k \leq n-1$ – highlighting that we don't have Markovianity here.

We rearrange the above to

$$\begin{aligned}\mathbb{E}[f(X_{n+1}) \mid F_n] &= f(X_n) - \sum_{k=0}^{n-1} Af(X_k) + \mathbb{E}[\sum_{k=0}^{n} Af(X_k) \mid F_n] &(11.4)\\ &= f(X_n) + Af(X_n) &(11.5)\\ &= \int_E f(y)\mu(X_n, dy) &(11.6)\end{aligned}$$

Note that we now have only X_n on the right hand side, so it is Markov. □

Lemma 11.29 (Sheet 2, Exercise 5b)**.** *Let $X(n)$, $n = 0, 1, \ldots,$ be a sequence of $\mathbb{Z}$-valued random variables such that for each $n \geq 0$,*

$$\left|X(n+1) - X(n)\right| \ = \ 1.$$

Let $g : \mathbb{Z} \to [-1, 1]$ be a function, and suppose that

$$X(n) \ - \ \sum_{k=0}^{n-1} g\big(X(k)\big)$$

is a martingale with respect to the natural filtration generated by X.

Then X is a Markov chain with

$$\mathbb{P}[X(n+1) - X(n) = 1 \mid F_n] = \frac{g(X(n)) + 1}{2}.$$

Proof. Note that we have the two equations

$$\begin{aligned}&\mathbb{P}[X(n+1) - X(n) = 1 \mid F_n] + \mathbb{P}[X(n+1) - X(n) = -1 \mid F_n] = 1 &(11.7)\\ &\mathbb{P}[X(n+1) - X(n) = 1 \mid F_n] - \mathbb{P}[X(n+1) - X(n) = -1 \mid F_n] = \\ &\qquad\qquad = \mathbb{E}[X(n+1) - X(n) = 1 \mid F_n] = g(X_n) &(11.8)\end{aligned}$$

where we used the Martingale property in the last equality, and used throughout that the absolute difference between $X(n+1)$ and $X(n)$ must be 1.

Solving this yields

$$\mathbb{P}[X(n+1) - X(n) = 1 \mid F_n] = \frac{g(X(n)) + 1}{2}, \qquad (11.9)$$

which is all we need to answer the exercise (it's Markov – the RHS depends only on $X(n)$ – and the above yields the transition probability right away). □

11.2.3. *Wright-Fisher diffusion.* The Wright-Fischer diffusion, which takes values in $[0,1]$ has generator

$$Af(x) = \frac{1}{2}x(1-x)f''(x),$$

when restricted to the subset of twice continuously differentiable functions on $[0,1]$.

By the previous section, the martingale problem

$$f(X_t) - \int_0^t Af(X_s)\,ds$$

will be a $\mathbb{P}$-local martingale, with suitable functions f.

We note that by setting $f(x) = x$, the integral term in the matringale inequality is zero and therefore it follows that X_t is a martingale itself. Moreover, it is positive and bounded and therefore uniformly integrable. So $X_\infty = \lim_{t\to\infty} X_t$ exists and satisfies $\mathbb{E}[X_\infty] = \mathbb{E}[X_0]$.

We claim that $\mathbb{P}[X_\infty \in \{0,1\}] = 1$ and therefore $\mathbb{P}[X_\infty = 1] = \mathbb{E}[X_\infty] = \mathbb{E}[X_0]$. To show the claim, we consider $f(x) = x(1-x)$ and so

$$X_t(1-X_t) + \int_0^t X_s(1-X_s)\,ds$$

is a positive martingale and therefore converges to a bounded limit. However, this is only possible if $\mathbb{P}[X_\infty \in \{0,1\}]$.

Finally we calculate $\mathbb{E}[\int_0^\infty X_s(1-X_s)\,dx]$. Indeed, we apply the previous martingale to conclude that $\mathbb{E}[X_\infty(1-X_\infty)] + \mathbb{E}[\int_0^\infty X_s(1-X_s)\,dx] = \mathbb{E}[X_0(1-X_0)]$. Note that $\mathbb{E}[X_\infty(1-X_\infty)] = 0$ as $\mathbb{P}[X_\infty \in \{0,1\}]$ and therefore $\mathbb{E}[\int_0^\infty X_s(1-X_s)\,dx] = \mathbb{E}[X_0(1-X_0)]$.

Now take $f(x) = 2x\log x + 2(1-x)\log(1-x)$ so that $Af(x) \equiv 1$ and so

$$f(X_t) - \int_0^t 1\,ds = f(X_t) - t$$

is a martingale. We note that $f(X_t) - t$ is negative and therefore unifomly integrable. Thus by the optional stopping theorem, for τ_ε the hitting time of $\{\varepsilon, 1-\varepsilon\}$ it holds that

$$\mathbb{E}[f(X_{\tau_\varepsilon}) - \tau_\varepsilon] = \mathbb{E}[f(X_0)].$$

As $f(x) = f(1-x)$ it therefore follows that

$$\mathbb{E}[\tau_\varepsilon] = f(\varepsilon) - \mathbb{E}[f(X_0)]$$

and so as $\varepsilon \to 0$ we conclude that the hitting time of $\{0,1\}$ is $-\mathbb{E}[f(X_0)]$.

12. C8.2: Class 2

12.1. **Stroock-Varadhan theory of diffusion approximation.** Sometimes, it is easier to deal with continuous space processes rather than discrete space processes – the latter often resulting in tedious recurrence equations. In this section, our goal is to show how to do such an approximation.

Take a sequence of (discrete space) Markov chains Z^h, indexed by $h > 0$ with the idea that we will take $h \downarrow 0$. Let Π^h be its transition kernel Π^h:

$$\Pr Z^h_{n+1} \in A \mid Z^h_n = z =: \Pi^h(z, A). \tag{12.1}$$

We rescale time (only time! space stays the same) and define for $t \in [0, 1]$

$$X^h_t := Z^h_{\lfloor t/h \rfloor} \quad t \in [0, 1] \tag{12.2}$$

with rescaled transition kernel

$$K_h(x, dy) := \frac{1}{h}\Pi_h(x, dy). \tag{12.3}$$

Theorem 12.1. *If we have $\forall R > 0, \epsilon > 0$:*

(1) $\lim_{h\downarrow 0} \sup_{|x|\le R} |b^h(x) - b(x)| = 0$ *with* $b^h(x) := \int_{|y-x|\le 1}(y - x)K^h(x, dy)$
(2) $\lim_{h\downarrow 0} \sup_{|x|\le R} |a^h(x) - a(x)| = 0$ *with* $a^h(x) := \int_{|y-x|\le 1}(y - x)^2 K^h(x, dy)$
(3) $\lim_{h\downarrow 0} \sup_{|x|\le R} K^h(x, B_\epsilon(x)^C) = 0$ i.e. for any fixed $\epsilon > 0$, probability of jumping further than ϵ away goes to 0.

(and some well-posedness conditions + initial condition, for details see lecture notes) then the sequence of Markov chains X^h_t on $[0, 1]$ converges weakly to a process solving the martingale problem M(a,b).

Writing ΔZ^h for the space increment at a single jump of the chain, we observe

$$K^h(x, B_\epsilon(x)^C) = \frac{1}{h}\Pr|\Delta Z^h| > \epsilon = \frac{1}{h}\Pr|\Delta Z^h|^4 > \epsilon^4 \le \frac{1}{h}\frac{1}{\epsilon^4}|\Delta Z^h|^4 \tag{12.4}$$

by Markov, so

$$\frac{1}{h}|\Delta Z^h|^4 \longrightarrow 0 \qquad \text{as } h \downarrow 0 \tag{12.5}$$

is sufficient for (3) of thm:stroock-varadhan-approximation to hold. (This solves Question 2 on sheet 3).

We now show that two very distinct discrete chains can converge to the same limiting diffusion as $h \downarrow 0$.

Lemma 12.2 (Sheet 3, Exercise 7)**.** *In the neutral Wright–Fisher model a population of N genes evolves in discrete generations. Generation $(t + 1)$ is formed from generation t by choosing N genes uniformly at random* with replacement*. In other words, each gene in generation $(t + 1)$ chooses its parent independently at random from among those present in generation t. Let us write $X^{(N)}_t$ for the proportion of type a genes in the population at time t under this model.*

In the neutral Moran model, generations overlap. At exponential rate $\binom{N}{2}$ a pair of genes is sampled uniformly at random from the population. One of the pair is selected at random to die, and the other splits into two copies. Let us write $Y^{(N)}_t$ for the proportion of type a genes in the population at time t under this model.

Show that the processes $\{X^{(N)}_{\lfloor Nt \rfloor}\}_{t\ge 0}$ and $\{Y^{(N)}_t\}_{t\ge 0}$ both converge as $N \to \infty$, and identify the limiting diffusion.

Proof. We first consider the wright fisher model. Let $Z_t^{(N)}$ be the #a-alleles in generation $t \in \mathbb{N}$. Then

$$Z_{t+1}^{(N)} \sim N\frac{Z_t^{(N)}}{N}. \tag{12.6}$$

To get sequence of MCs getting finer and finer, we define

$$X_t^{(N)} := \frac{Z_t^{(N)}}{N} \tag{12.7}$$

which gives rise to the rescaled transition kernel for $y \in E^N := \{\frac{k}{N} : 0 \leq k \leq N\}$

$$K_N(x, y) = N\Pi^N(x, y) = N \Pr Nx = yN = N\binom{N}{yN}x^{yN}(1-x)^{N-yN} \tag{12.8}$$

Now note that for condition (1): $b^h(x)$ is just the expected $\Delta X^{(h)}/h$ (started a point x) [here $h = 1/N$]. Since we have

$$\Delta X^{(N)} = \frac{Nx}{N} - x \tag{12.9}$$

so

$$\Delta X^{(N)}/h = N\Delta X^{(N)} = Nx - Nx \tag{12.10}$$

So

$$b^{(N)}(x) = \int (y - s)K^{(N)}(x, dy) = Nx - Nx = 0. \tag{12.11}$$

Similarly, for $a^{(N)}(x)$ note that

$$a^{(N)}(x) = N\cdot(\Delta X^{(N)})^2 = \frac{1}{N}Nx = x(1-x) \tag{12.12}$$

For condition (3), we use the observation above and note that it suffices to check that $\Delta X^{(N)} = O(1/N^2)$: The 4-th centered moment of a Bin(n,p) RV is:

$$Np(1-p)(3p(1-p)(N-2)+1) = O(N^2) \tag{12.13}$$

so indeed

$$(\Delta X^{(N)})^4 = \frac{1}{N^4}O(N^2) = O(1/N^2), \tag{12.14}$$

so condition (3) holds.

Thus, in the limit, the process:

$$X_{Nt} \tag{12.15}$$

converges to a process solving the MG problem $M(a, b)$ with

$$b(x) \equiv 0 \tag{12.16}$$

$$a(x) = x(1-x) \tag{12.17}$$

so generator is

$$Af = \frac{1}{2}af'' + b' = \frac{1}{2}x(1-x)f'' \tag{12.18}$$

For the neutral Moran model, note that this continuous time Markov chain $Y_t^{(N)}$ has state space $E^{(N)} = \{k/N : 0 \leq k \leq N\}$. While the chain is absorbing at 0 and 1 (you should verify that and treat that separately) we focus on the case $k \in \{1, \ldots, N-1\}$. Note that from a state k/N we can either go to state $(k-1)/N$, stay in state k/N (the killing and birth evened out) or go to state $(k+1)/N$. The first occurs at rate $\binom{N}{2}k/N(1-k/N)$, the second at rate $2\binom{N}{2}k/N(1-k/N)$, and the last at rate $\binom{N}{2}k/N(1-k/N)$. We can thus write for the generator of the N-th refined process:

$$A_N f(y) = \binom{N}{2} y(1-y)\left(f(y-1/N) + f(y+1/N) - 2f(y)\right).$$

Doing a Taylor expansion thereof, we observe that for fixed $y \in [0,1]$ we have

$$A_N f(y) \longrightarrow \frac{1}{2} y(1-y) f''(y) \qquad \text{as } N \to \infty$$

where we note that the conditions of convergence are satisfied (exercise: check that).

We thus observe that both discrete processes above converge to the same continuous process. $\square$

12.2. **Duality.** We recall the following theorem from the lecture notes.

Theorem 12.3. *(The method of duality) Let E_1 and E_2 be metric spaces and suppose that $\mathbb{P}$ and $\mathbb{Q}$ are probability distributions on the space of cadlag functions from $[0,\infty)$ to E_1 and E_2 repsectively. Let f and g be two bounded functions for which the following are true:*

(1) For each $y \in E_2$, $f(\cdot, y)$ and $g(\cdot, y)$ are continuous functions on E_1.
(2) For each $x \in E_1$, $f(x, \cdot)$ and $g(x, \cdot)$ are continuous functions on E_2.
(3) For $y \in E_2$,

$$f(X(t), y) - \int_0^t g(X(s), y)\, ds$$

is a $\mathbb{P}$-martingale.

(4) For $x \in E_1$,

$$f(x, Y(t)) - \int_0^t g(x, Y(s))\, ds$$

is a $\mathbb{Q}$-martingale.

Then

$$\mathbb{E}^{\mathbb{P}}_{X(0)}[f(X(t), Y(0))] = \mathbb{E}^{\mathbb{Q}}_{Y(0)}[f(X(0), Y(t))].$$

We can also give the following variant of the duality results. If we denote by A_X the generator of X on E_1 and by A_Y the generator of Y on E_2. If instead of the setting of Theorem 12.3 we have

$$A_X f(x,y) + \alpha(x) f(x,y) = A_Y f(x,y) + \beta(y) f(x,y),$$

then if we assume that $\int_0^t |\alpha(X(s))|\, ds < \infty$ and $\int_0^t |\beta(Y(s))|\, ds < \infty$ and we have the additional integrability conditions

$$\mathbb{E}\left[\left| f(X(t), Y(0)) \exp\left(\int_0^t \alpha(X(s))\, ds\right)\right|\right] < \infty$$

and

$$\mathbb{E}\left[\left| f(X(0), Y(t)) \exp\left(\int_0^t \beta(Y(s))\, ds\right)\right|\right] < \infty,$$

then the duality formula can be modified to

$$\begin{aligned}&\mathbb{E}\left[f(X(t),Y(0))\exp\left(\int_0^t \alpha(X(s))\,ds\right)\right]\\=&\mathbb{E}\left[f(X(0),Y(t))\exp\left(\int_0^t \beta(Y(s))\,ds\right)\right].\end{aligned} \tag{12.19}$$

We discuss the following examples as applications.

Wright-Fisher diffusion. We consider the Wright-Fisher diffusion model with generator

$$A_X f(x) = \frac{1}{2}x(1-x)f''(x).$$

We also consider the pure death process $N_t \mapsto N_t - 1$ at rate $\binom{N_t}{2}$. Indeed, the generator of the pure death process A_Y is

$$A_Y f(n) = \binom{n}{2}(f(n-1)-f(n)).$$

Therefore if we set $f(x,n) = x^n$, then we can take $g(x,n) = \binom{n}{2}(x^{n-1}-x^n)$, and we conclude that

$$\mathbb{E}[X_t^{N_0}] = \mathbb{E}[X_0^n].$$

Wright-Fisher diffusion with mutation. The Wright-Fisher diffusion model with mutation is discussed next. The model is uniquely characterized by the generator

$$A_X f(x) = \frac{1}{2}x(1-x)f''(x) + (a-bx)f'(x)$$

with $a < b$. Denote by $X(t)$ the associated Markov process. Now let's consider the function $f(x,n) = x^n$ for a positive integer n. Then we have

$$\begin{aligned}g(x,n) = A_X f(\cdot,n) &= \binom{n}{2}(x^{n-1}-x^n) + anx^{n-1} - bnx^n\\ &= \binom{n}{2}(x^{n-1}-x^n) + an(x^{n-1}-x^n) - (b-a)nx^n.\end{aligned}$$

We can write this in the form of (12.19). Indeed, we set N_t to be the pure death process N_t at rate $\binom{N_t}{2} + aN_t$, so the corresponding operator is

$$A_Y f(n) = \left(\binom{n}{2} + an\right)(f(n-1)-f(n)).$$

Then $\alpha(x) = 0$ and $\beta(y) = (b-a)y$. Since the integrebility conditions are clear, it follows that

$$\mathbb{E}[X_t^{N_0}] = \mathbb{E}\left[X_0^{N_t}\exp\left(-\int_0^t (b-a)N_s\,ds\right)\right].$$

Setting $N_0 = k$, it follows that all of the moments of X_t are determined and therefore, as X_t is compactly supported, it is uniquely determined. Moreover, since N is a pure death process with strictly non-zero rates, it will hit zero in finite time with probability one, and so the right hand side of the previous equation is independent of X_0 as $t \to \infty$. Therefore X_t converges in distribution to a limiting distribution which does not depend on X_0.

Absorbing and Reflecting Brownian Motion. Denote by A_1 the generator of Brownian motion on $[0,\infty)$ with absorbing boundary condition, that is $A_1 f = \frac{1}{2}f''$ and with

$$\mathcal{D}(A_1) = \{f \in C^2 \,:\, f''(0) = 0\}.$$

We take A_2 to be the generator of reflecting Brownian motion on $[0,\infty)$ so $A_2 f = \frac{1}{2}f''$ and with

$$\mathcal{D}(A_1) = \{f \in C^2 \,:\, f'(0) = 0\}.$$

We denote by X_t absorbing Brownian motion and by Y_t reflecting Brownian motion.

Let $h : \mathbb{R} \to \mathbb{R}$ be C^2 such that $h(z) = -h(-z)$ for $z \in \mathbb{R}$. Consider

$$F(x,y) = h(x+y) + h(x-y).$$

Note that since h is odd, $h''(0) = 0$ and therefore

$$\left.\frac{\partial F}{\partial x^2}\right|_{x=0} = (h''(x+y) + h''(x-y))|_{x=0} = 0.$$

Thus for a fixed y, $F(\cdot,y) \in \mathcal{D}(A_1)$. Similarly, for each fixed x,

$$\left.\frac{\partial F}{\partial y}\right|_{y=0} = (h'(x+y) - h'(x-y))|_{y=0} = 0$$

and so $F(x,\cdot) \in \mathcal{D}(A_2)$. Moreover $A_1F(x,y)$ (as a function of x) equals $A_2F(x,y)$ (as a function of y) and they are both equal to $h''(x+y) + h''(x-y)$. Thus (3) and (4) is satisfied by Theorem 11.27 and therefore it holds that

$$\mathbb{E}_x[h(X_t+y) + h(X_t-y)] = \mathbb{E}_y[h(x+Y_t) + h(x-Y_t)].$$

Now take

$$g(x) = \begin{cases} -\frac{1}{2} & x < 0, \\ 0 & x = 0, \\ \frac{1}{2} & x > 0, \end{cases}$$

or rather a twice continuously differentiable approximation to this. If $X_t > y$, then $g(X_t+y) + g(X_t-y) = 1$ and if $X_t \le y$ then $g(X_t+y) + g(X_t-y) = 0$. Similarly if $x > Y_t$ we have that $g(x+Y_t) + g(x-Y_t) = 1$, whereas if $x < Y_t$ we have $g(x+Y_t) + g(x-Y_t) = 0$. The result follows.

12.3. **Theory of Speed and Scale.** Assume that a one-dimensional Markov process $(X_t)_{t\ge 0}$ is governed by the infinitesimal generator

$$Af(x) = \frac{1}{2}\sigma^2(x)f''(x) + \mu(x)f'(x)$$

for f a twice continuously differentiable function on (a,b). We assume that μ and σ are bounded and locally Lipschitz on (a,b) with $\sigma^2(x) > 0$ on (a,b).

In this setting the scale function is defined as

$$S(x) = \int_{x_0}^{x} \exp\left(-\int_{\eta}^{y} \frac{2\mu(z)}{\sigma^2(z)}\right) dy,$$

where x_0, η are arbitrary chosen points in (a,b) The density of the speed function is given by

$$m(\xi) = \frac{1}{\sigma^2(\xi)S'(\xi)}$$

and we write
$$M(x) = \int_{x_0}^{x} m(\xi)\,d\xi.$$

We then consider
$$u(x) = \int_{x}^{x_0} M\,dS, \qquad v(x) = \int_{x}^{x_0} S\,dM.$$
We are interested in the behaviour of u and v at a. We recall that we call a to be a regular boundary if $u(a) < \infty$ and $v(a) < \infty$.

12.3.1. *Wright-Fisher diffusion with mutation.* We are now interested in the Wright-Fisher diffusion model with mutation, with generator
$$Af(x) = \frac{1}{2}x(1-x)f''(x) + (\nu_2 - (\nu_1+\nu_2)x)f'(x).$$

We calculate
$$\begin{aligned} S'(p) &= \exp\left(-\int_{p_0}^{p} \frac{2\mu(z)}{\sigma^2(z)}\,dz\right) \\ &= \exp\left(-\int_{p_0}^{p} \frac{2\nu_2(1-z) - 2\nu_1 z}{z(1-z)}\,dz\right) \\ &= C_{p_0}\exp(-2\nu_2\log p - 2\nu_2\log(1-p)) \\ &= C_{p_0}p^{-2\nu_2}(1-p)^{-2\nu_1}, \end{aligned}$$
where the constant C depends on p_0. We then have
$$m(p) = \frac{1}{\sigma^2(p)S'(p)} = C_{p_0}^{-1}p^{2\nu_2-1}(1-p)^{2\nu_1-1}.$$
Thus
$$\int_0^{1/2} M\,dS = \int_0^{1/2}\int_x^{1/2} \xi^{2\nu_2-1}(1-\xi)^{2\nu_1-1}\,d\xi\,x^{-2\nu_2}(1-x)^{-2\nu_1}\,dx$$
which is of the same order as
$$\int_0^{1/2}(c_1x^{2\nu_2}+c_1)x^{-2\nu_2}$$
which is finite if and only if $2\nu_2 < 1$. In the other order,
$$\int_0^{1/2}\int_\xi^{1/2} dS\,dM$$
one checks that the resulting term is finite for $\nu_2 > 0$ and infinite for $\nu_2 = 0$. Thus the boundary is regular for $0 < \nu_2 < 1/2$.

12.3.2. *Bessel process.* We now consider the Bessel process with parameter $\alpha \geq 0$, which is determines as the one-dimensional diffusion process on $[0,\infty)$ with generator
$$Af(x) = \frac{1}{2}f''(x) + \frac{\alpha-1}{2x}f'(x).$$
When α is an integer, this is the norm of a Brownian motion in $\mathbb{R}^\alpha$. We now find expressions for the speed and scale.

Note that
$$\exp(-\int_1^{\eta}\frac{\alpha-1}{z}\,dz) = \eta^{1-\alpha}$$

and so the scale function is

$$S(\xi) = \int_1^\xi \eta^{1-\alpha} d\eta = \begin{cases} \frac{1}{2-\alpha}(\xi^{2-\alpha} - 1) & \text{if } \alpha \neq 2, \\ \log \xi & \text{if } \alpha = 2. \end{cases}$$

and therefore

$$m(\eta) = \eta^{\alpha-1}.$$

Substituting these expressions, we obtain, by doing the calculation for $\alpha \neq 0$,

$$\begin{aligned} u(0) &= \int_0^1 \int_\xi^1 m(\eta)\, d\eta\, S'(\xi)\, d\xi \\ &= \int_0^1 \frac{1}{\alpha}(1 - \xi^\alpha)\xi^{1-\alpha}\, d\xi \\ &= \frac{1}{\alpha} \int_0^1 \xi^{1-\alpha} - \xi\, d\xi \\ &\begin{cases} \int -\xi \log \xi & \text{if } \alpha = 0, \\ \frac{1}{\alpha}(\frac{1}{2-\alpha} - \frac{1}{2}) & \text{if } \alpha \in (0,2), \\ \infty & \text{if } \alpha \geq 2. \end{cases} \end{aligned}$$

One similarly checks that $v(0)$ is finite if and only if $v > 0$. Combining these observations:

$$\text{The boundary at 0 is } \begin{cases} \text{an entrance boundary} & \text{for } \alpha \geq 2, \\ \text{a regular boundary} & \text{if } \alpha \in (0,2), \\ \text{an exit boundary} & \text{if } \alpha = 0. \end{cases}$$