Probabilities and Martingales

1MS045 10507 — 2024/2025

Inês Mesquita

Based on the notes by professor Sascha Troscheit

Contents

1 Measure Spaces

A σ-algebra is a collection of subsets of a set with specific properties that make it suitable for defining measures.

Definition 1.1. (σ-Algebra) A collection Σ of subsets of a set S is called a σ-algebra if:

1. It contains the empty set: $\emptyset \in \Sigma$
2. It is an algebra:
   - If $A \in \Sigma$, then $A^c = S \setminus A \in \Sigma$.
   - If $A, B \in \Sigma$, then $A \cup B \in \Sigma$.
3. It is closed under countable unions: If $A_i \in \Sigma$ for all $i \in \mathbb{N}$, then $\bigcup_{i=1}^{\infty} A_i \in \Sigma$.

Proposition 1.2. (Key Properties of Algebras)

- Every algebra is closed under finite unions.
- Alternative (and equivalent) formulations of σ-algebras exist. For example, replacing $\emptyset \in \Sigma$ by "S is non-empty" leads to an equivalent definition.

Example 1.3. (**Examples of σ-algebras**)

- The power set $\mathbb{P}(S) = \{A : A \subseteq S\}$ is a σ-algebra.
- $\{\emptyset, S\}$ is a σ-algebra.
- $\{\emptyset, 2\mathbb{N}, 2\mathbb{N} - 1, \mathbb{N}\}$ is a σ-algebra on $\mathbb{N}$, where $2\mathbb{N} = \{2, 4, 6, \ldots\}$, $2\mathbb{N} - 1 = \{1, 3, 5, \ldots\}$, and $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Example 1.4. (Set Algebra That is Not a σ-algebra)
Consider the set of the union of at most k intervals of the form $(a, b]$ with $a, b \in \mathbb{Z}$.

- This set is an algebra because it satisfies the properties of an algebra.
- This set is not a σ-algebra because the union of infinitely many such intervals may not be represented by a finite union.

Definition 1.5. (Measures) Let Σ_0 be a σ-algebra on S and let μ_0 be a function from $\Sigma_0 \to [0, \infty] = [0, \infty) \cup \{\infty\}$.

- μ_0 is **additive** if for all disjoint $A, B \in \Sigma_0$ we have $\mu_0(A \cup B) = \mu_0(A) + \mu_0(B)$.
- μ_0 is **σ-additive** if $\mu_0(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu_0(A_i)$ for pairwise disjoint $A_1, A_2, \dots \in \Sigma_0$ such that $\bigcup_{i=1}^{\infty} A_i \in \Sigma_0$.

Proposition 1.6. If μ_0 is additive, we also have $\mu_0(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} \mu_0(A_i)$ for all finite, pairwise disjoint $A_1, \dots, A_k$.

Example 1.7. (**Examples of Measures**)

- **Counting Measure:** Consider the σ-algebra $\Sigma = \mathbb{P}(\mathbb{N})$ on $\mathbb{N} = \{1, 2, 3, \dots\}$. Define $\mu_0(A) = \#A$ (i.e., the number of elements in A). This is σ-additive.
- **Probability Measure:** Take $\Sigma = \mathbb{P}(\{1, 2, 3, 4, 5, 6\})$ and $\mu_0(A) = \#A/6$. This represents the probability that the outcome of a fair die roll lies in A and is σ-additive.
- **Example of an Additive Measure that is Not σ-additive:** Take $\Sigma_0 = \mathbb{P}(\mathbb{N})$ and define
$$\mu_0(A) = \begin{cases} 0 & \text{if } A \text{ is finite} \\ \infty & \text{if } A \text{ is infinite} \end{cases}$$
  - This is additive.
  - This is not σ-additive because $\mu_0\left(\bigcup_{k=1}^{\infty} \{k\}\right) = \mu_0(\mathbb{N}) = \infty \neq \sum_{k=1}^{\infty} \mu_0(\{k\}) = 0$.

Definition 1.8. (Measure Spaces) A measure space consists of:

- A set S (called the "universe").
- A σ-algebra Σ on S.
- A σ-additive function $\mu : \Sigma \to [0, \infty]$, called a measure, such that $\mu(\emptyset) = 0$.

Definition 1.9. (Probability Spaces) A measure space (S, Σ, μ) is called a **probability space** if μ is a probability measure, i.e., $\mu(S) = 1$.

Example 1.10. (**Examples of Measure Spaces**)

- **(Finite probability space):** Let $S = \{s_1, \ldots, s_k\}$ be a finite set of outcomes (e.g., $S = \{1, \ldots, 6\}$ for a die, or $S = \{\text{heads}, \text{tails}\}$). Associate probabilities $p_1, \ldots, p_k$ with $s_1, \ldots, s_k$ such that $p_1 + \cdots + p_k = 1$. Define $\mu(A) = \sum\limits_{i:s_i \in A} p_i$ for $A \in \Sigma := \mathbb{P}(S)$. This defines a probability space (S, Σ, μ). $\mu(A)$ represents the probability that A occurs.
- **(Lebesgue Measure):** Let $S = \mathbb{R}$, $\Sigma = \mathcal{B}(\mathbb{R})$, the Borel σ-algebra of $\mathbb{R}$ (the smallest σ-algebra that contains all open subsets of $\mathbb{R}$). $\mathcal{B}(\mathbb{R}) \neq \mathbb{P}(\mathbb{R})$.
  For unions of open, disjoint intervals $A = (a_1, b_1] \cup \cdots \cup (a_n, b_n]$, let $\mathcal{L}(A) = (b_1 - a_1) + \cdots + (b_n - a_n)$. This can be extended to $\mathcal{B}(\mathbb{R})$ and is called the Lebesgue measure. The space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L})$ is a measure space. Restricting to $([0,1], \mathcal{B}([0,1]), \mathcal{L} \mid_{[0,1]})$ gives a probability space. $\mathcal{L} \mid_{[0,1]} = \mathcal{L}(A \cap [0,1])$ represents a uniformly random number from $[0,1]$.

Proposition 1.11. (General Properties of Measures) Let (S, Σ, μ) be a measure space. We have:

1. $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$ for all $A, B \in \Sigma$.
2. Generally, $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)$ for all $A_i \in \Sigma$.
3. More precisely,
$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$$
and
$$\mu(A \cup B \cup C) = \mu(A) + \mu(B) + \mu(C) - \mu(A \cap B) - \mu(A \cap C) - \mu(B \cap C) + \mu(A \cap B \cap C)$$
for $A, B, C \in \Sigma$. Generally,
$$\mu(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mu(A_i) - \sum_{1 \leq i < j \leq n} \mu(A_i \cap A_j) + \cdots + (-1)^{n-1} \mu(A_1 \cap \cdots \cap A_n).$$
This is known as the **inclusion-exclusion principle**.

Proof. Note that:

$\mu(A) = \mu(A \setminus B) + \mu(A \cap B)$

$\mu(B) = \mu(B \setminus A) + \mu(A \cap B)$

$\mu(A \cup B) = \mu(A \setminus B) + \mu(B \setminus A) + \mu(A \cap B) = \mu(A) - \mu(A \cap B) + \mu(B) - \mu(A \cap B) + \mu(A \cap B) = \mu(A) + \mu(B) - \mu(A \cap B)$ as required.

The general case follows by induction. □

Proposition 1.12. (Monotonicity) Let $(A_i)_{i=1}^{\infty}$ be an increasing sequence of sets in Σ, so $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \subseteq S$. Then,

$$\mu(A_i) = \mu(A_{i-1}) + \mu(A_i \setminus A_{i-1}) \geq \mu(A_{i-1})$$

and so $0 \leq \mu(A_1) \leq \mu(A_2) \leq \dots$.
Since $(\mu(A_i))$ is an increasing sequence, the limit $L = \lim_{i \to \infty} \mu(A_i)$ exists (but may be ∞). Writing $A = \bigcup_{i=1}^{\infty} A_i$, we have $\mu(A) = L$. This is because

$$A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots$$

is a disjoint union. Therefore,

$$\mu(A) = \mu(A_1) + \mu(A_2 \setminus A_1) + \mu(A_3 \setminus A_2) + \cdots = \lim_{n \to \infty} \mu(A_n) = L.$$

This also works for decreasing sequences $A_1 \supseteq A_2 \supseteq \dots$ ($A_i \in \Sigma$). Define $A = \bigcap_{i=1}^{\infty} A_i$, then $\mu(A) = \lim_{i \to \infty} \mu(A_i)$.
In particular, if $\lim_{n \to \infty} \mu(A_n) = 0$, then $\mu(A) = 0$.

Remark 1.13. Such **null sets** can be non-empty and even uncountable.

Remark 1.14. From now on, we will mostly consider **probability spaces** (i.e., $\mu(S) = 1$).

Definition 1.15. (Generated σ-algebras) Given any subset $\mathcal{A} \subseteq \mathcal{P}(S)$, the **$\sigma$-algebra generated by** $\mathcal{A}$, denoted $\sigma(\mathcal{A})$ or $\langle \mathcal{A} \rangle$, is the *smallest* σ-algebra containing $\mathcal{A}$.
Formally:

$$\sigma(\mathcal{A}) := \bigcap \{\Sigma : \text{ with } \Sigma \text{ a } \sigma\text{-algebra}, \mathcal{A} \subseteq \Sigma\}$$

- $\emptyset \in \Sigma$ for all σ-algebras, so $\emptyset \in \sigma(\mathcal{A})$.
- If $A \in \sigma(\mathcal{A})$, then $A \in \Sigma$ for all such σ-algebras, so $A^c \in \Sigma$ for all such Σ, and $A^c \in \sigma(\mathcal{A})$.
- If $A_n \in \sigma(\mathcal{A})$ for all n, then $A_n \in \Sigma$ for all such Σ, and $\bigcup A_n \in \Sigma$ for all such Σ, and $\bigcup A_n \in \sigma(\mathcal{A})$.

Definition 1.16. (Borel σ-algebra)
The σ-algebra generated by the open subsets of S is denoted $\mathcal{B}(S) = \sigma(\{A \subseteq S : A \text{ is open}\})$.
The **Borel σ-algebra** $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ is generated by:

- Open intervals: $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b) : a \leq b\})$
- Half-open intervals: $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, a] : a \in \mathbb{R}\})$ or $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, a] : a \in \mathbb{R}\})$
- Closed sets: $\mathcal{B}(\mathbb{R}) = \sigma(\{F \subseteq \mathbb{R} : F \text{ closed}\}) = \sigma(\{(q_1, q_2) : q_1 < q_2, q_1, q_2 \in \mathbb{Q}\})$

Remark 1.17. Why is the last representation of the Borel σ-algebra countable?
The set of rational numbers $\mathbb{Q}$ is countable, meaning its elements can be put into a one-to-one correspondence with the natural numbers. Therefore, any set formed by taking intervals with rational endpoints is also countable.

Lemma 1.18. The notion of generated σ-algebras is useful, as the **π-system lemma** shows. Suppose:

- $\mathcal{A}$ is a π-system (i.e., closed under finite intersections)
- μ_1, μ_2 are measures on $(S, \sigma(\mathcal{A}))$ such that $\mu_1(A) = \mu_2(A)$ for all $A \in \mathcal{A}$

Then, $\mu_1 = \mu_2$.
In other words, μ is **uniquely determined by** any π-system $\mathcal{A} \subseteq \mathcal{P}(S)$.

Example 1.19. $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, a] : a \in \mathbb{R}\})$

Hence, every probability measure $\mathbb{P}$ on $\mathbb{R}$ is determined by the values of $\mathbb{P}((-\infty, a])$, i.e., its **cumulative distribution function** F.

The following important theorem tells us that measures can be constructed from "small" collections of subsets.

Theorem 1.20. *(Carathéodory's Extension Theorem) If Σ_0 is an algebra and $\mu_0 : \Sigma_0 \to [0, \infty]$ is a σ-additive function, then there exists a* ***unique*** *measure μ on $\Sigma = \sigma(\Sigma_0)$ such that $\mu(A) = \mu_0(A)$ for all $A \in \Sigma_0$. In other words: $\mu_{|\Sigma_0} = \mu_0$.*

Proposition 1.21. (Important consequence) The Lebesgue measure is unique.

We can construct $\mathcal{L}$ on $\mathcal{B}(\mathbb{R})$ by defining

$$\mathcal{L}_0((a_1, b_1] \cup (a_2, b_2] \cup ... \cup (a_n, b_n]) = (b_1 - a_1) + ... + (b_n - a_n)$$

for all $a_1 < b_1, a_2 < b_2, \ldots, a_n < b_n$, and noting that sets of that form are an algebra.

2 Probability Spaces and Almost Sure Events

Definition 2.1. (Probability Spaces) Probability spaces are a specific type of measure space where the measure represents probability. They consist of:

- A sample space, denoted as Ω (or sometimes "universe"), representing all possible outcomes.
- A σ-algebra $\mathcal{F}$ on Ω, called "events" — subsets of Ω to which probabilities can be assigned.
- A probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$, satisfying:
  - $\mathbb{P}(\emptyset) = 0$
  - $\mathbb{P}(\Omega) = 1$
  - For pairwise disjoint events $A_i \in \mathcal{F}$,

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

Example 2.2.

- (Rolling a die): Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, and define $\mathbb{P}(E) = \frac{\#E}{6}$ for all $E \in \mathcal{F}$. This represents a formal model for rolling a die.
- (Normal Distribution): Let $\Omega = \mathbb{R}$, $\Sigma = \mathcal{B}(\mathbb{R})$, and let $\mathbb{P}$ be determined by

$$\mathbb{P}((a, b]) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx.$$

  This corresponds to the standard normal distribution.

Definition 2.3. (Almost Sure Events) An event $E \in \mathcal{F}$ happens **almost surely** if $\mathbb{P}(E) = 1$. Equivalently, $\mathbb{P}(E^c) = 0$.

Example 2.4. Consider a uniformly random number X on the interval $[0, 1]$. For every fixed $y \in [0, 1]$, we have $\mathbb{P}(X = y) = \mathbb{P}(\{y\}) = 0$. Therefore, $\mathbb{P}(X \neq y) = \mathbb{P}([0, 1] \setminus \{y\}) = 1$. In other words, $X \neq y$ almost surely.

Proposition 2.5. If $A_1, A_2, \ldots$ are almost sure events, then so is $\bigcap_{i=1}^{\infty} A_i$.
Since A_i are almost sure, $\mathbb{P}(A_i) = 1$ for all $i \in \mathbb{N}$. This implies

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = 1.$$

This only works for **countable intersections**!

Proof. By assumption $\mathbb{P}(A_i) = 1$ for all $i \in \mathbb{N}$ and so $\mathbb{P}(A_i^c) = 0$. Formally, this is because $\Omega = A_i \cup A_i^c$ is disjoint, and so $\mathbb{P}(\Omega) = \mathbb{P}(A_i) + \mathbb{P}(A_i^c) \iff 1 = \mathbb{P}(A_i) + 1$.

So $\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i^c\right) \leq \sum_{i\in\mathbb{N}} \mathbb{P}(A_i^c) = 0$.

But $\bigcup_{i\in\mathbb{N}} A_i^c = \left(\bigcap_{i\in\mathbb{N}} A_i\right)^c$ (DeMorgan's Law) and so $\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i^c\right) = \mathbb{P}\left(\left(\bigcap_{i\in\mathbb{N}} A_i\right)\right) = 0$.

Hence $\mathbb{P}\left(\bigcap_{i\in\mathbb{N}} A_i\right) = 1$ $\square$

Example 2.6. Let X be uniformly random on $[0,1]$. For any $x \in \mathbb{Q}\cap[0,1]$, $\mathbb{P}(X \neq x) = 1$. Since there are only countably many rational numbers,

$$\mathbb{P}\left(\bigcap_{x\in\mathbb{Q}\cap[0,1]} \{X \neq x\}\right) = \mathbb{P}(X \text{ is irrational}) = 1.$$

However,

$$\mathbb{P}\left(\bigcap_{x\in[0,1]} \{X \neq x\}\right) = \mathbb{P}(X \text{ takes a value outside } [0,1]) = 0 \neq 1$$

since it is an uncountable intersection.

Remark 2.7. While the intersection of countably many almost sure events is also almost sure, this property does not hold for uncountable intersections. This distinction is important in probability theory, as it highlights the difference between events with probability 1 and events that are guaranteed to occur.

Definition 2.8. (lim sup and lim inf for a Sequence of Real Numbers)

- **Recall:**

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup_{m\geq n} x_m,$$
$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf_{m\geq n} x_m$$

- The lim sup and lim inf **always exist**.
- The limit $\lim_{n\to\infty} x_n$ exists if and only if $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$.
- $\limsup_{n\to\infty} x_n \geq x$ means there exists a subsequence of x_n with a limit greater than or equal to x.

Definition 2.9. (lim sup and lim inf for Sets)

Let $E_1, E_2, \ldots$ be events (sets).

$$\liminf_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} \bigcap_{m\geq n} E_m \quad \text{(a decreasing sequence of sets)}$$
$$\limsup_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} \bigcup_{m\geq n} E_m \quad \text{(an increasing sequence of sets)}$$

- $\liminf_{n\to\infty} E_n$ contains all elements that eventually occur in **all** E_m for $m \geq n$, but only in finitely many E_m.
- $\limsup_{n\to\infty} E_n$ contains all elements that occur in **infinitely many** E_m.

Proposition 2.10. We have the relationship:

$$\liminf_{n\to\infty} E_n \subseteq \limsup_{n\to\infty} E_n$$

Lemma 2.11. (Fatou's Lemma) Given a sequence of events $(E_n)_{n\in\mathbb{N}}$, the following inequalities hold:

$$\mathbb{P}(\liminf_n E_n) \leq \liminf_n \mathbb{P}(E_n)$$
$$\mathbb{P}(\limsup_n E_n) \geq \limsup_n \mathbb{P}(E_n) \quad (\textit{Reverse Fatou's Lemma})$$

Proof. (First Inequality):

1. Define $F_n = \bigcap_{m\geq n} E_m$. Note that $(F_n)_{n\in\mathbb{N}}$ is an increasing sequence of sets.
2. $\liminf_n E_n = \bigcup_{n=1}^{\infty} F_n$
3. Since $F_n \subseteq E_m$ for all $m \geq n$, $\mathbb{P}(F_n) \leq \mathbb{P}(E_m)$ for all $m \geq n$.
4. $\mathbb{P}(F_n) \leq \inf_{m\geq n} \mathbb{P}(E_m)$
5. Since $(F_n)_{n\in\mathbb{N}}$ is increasing, $\lim_n \mathbb{P}(F_n)$ exists and equals $\mathbb{P}\left(\bigcup_{n=1}^{\infty} F_n\right) = \mathbb{P}(\liminf_n E_n)$.
6. $\lim_n \mathbb{P}(F_n) \leq \lim_n \inf_{m\geq n} \mathbb{P}(E_m) = \liminf_n \mathbb{P}(E_n)$
7. We get $\mathbb{P}(\liminf_n E_n) \leq \liminf_n \mathbb{P}(E_n)$ as required.

The proof for the second inequality (*Reverse Fatou's Lemma*) follows a similar approach. $\square$

Lemma 2.12. (Borel-Cantelli Lemma — First Lemma) Let $(E_n)_{n\in\mathbb{N}}$ be a sequence of events. If $\sum_{i=1}^{\infty} \mathbb{P}(E_i) < \infty$, then:

$$\mathbb{P}(\limsup_{n\to\infty} E_n) = \mathbb{P}(\text{infinitely many } E_n \text{ occur}) = 0.$$

Proof.

1. Recall: $\limsup_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} \bigcup_{m\geq n} E_m$, where $G_n = \bigcup_{m\geq n} E_m$ is a decreasing sequence.
2. $\limsup_{n\to\infty} E_n \subseteq G_m$ for all $m \in \mathbb{N}$.
3. Hence, $\mathbb{P}(\limsup_{n\to\infty} E_n) \leq \mathbb{P}(G_m) \leq \sum_{i=m}^{\infty} \mathbb{P}(E_i)$ for all $m \in \mathbb{N}$.
4. Since $\sum_{i=1}^{\infty} \mathbb{P}(E_i) < \infty$, we have $\lim_{m\to\infty} \sum_{i=m}^{\infty} \mathbb{P}(E_i) = 0$.
5. Therefore, $\mathbb{P}(\limsup_{n\to\infty} E_n) \leq 0$. As probabilities are non-negative, $\mathbb{P}(\limsup_{n\to\infty} E_n) = 0$.

$\square$

Example 2.13. Consider a sequence of fair coin tosses, and let E_n be the event that the first n tosses are heads.

- $\mathbb{P}(E_n) = \left(\frac{1}{2}\right)^n$, and $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = 1 < \infty$.
- Hence, the probability that E_n occurs infinitely often (we get infinitely many sequences of n heads) is 0, so the probability that we get only heads is 0.

3 Random Variables

Definition 3.1. (Measurable Functions) Let $(\Omega, \Sigma, \mathbb{P})$ be a measure space. A function $f : \Omega \to \mathbb{R}$ (or $[-\infty, \infty]$) is **measurable** if for all Borel sets $A \in \mathcal{B}(\mathbb{R})$, the pre-image satisfies

$$f^{-1}(A) = \{\omega \in \Omega : f(\omega) \in A\} \in \Sigma.$$

Notation 3.2. We write $m\Sigma$ for the measurable functions with respect to Σ, $\{m\Sigma\}^+ = m\Sigma \cap \{f \geq 0\}$ for the non-negative measurable functions, and $b\Sigma$ for the bounded measurable functions.

Proposition 3.3. (Properties of Pre-images) We have the following properties:

1. $f^{-1}(A^c) = f^{-1}(A)^c$,
2. $f^{-1}\left(\bigcup_i A_i\right) = \bigcup_i f^{-1}(A_i)$,
3. $f^{-1}\left(\bigcap_i A_i\right) = \bigcap_i f^{-1}(A_i)$.

Proof. We prove 2., the others are similar.

$$\begin{aligned} x \in f^{-1}\left(\bigcup_i A_i\right) &\iff f(x) \in \bigcup_i A_i \\ &\iff f(x) \in A_i \text{ for some } i \\ &\iff x \in f^{-1}(A_i) \text{ for some } i \\ &\iff x \in \bigcup_i f^{-1}(A_i) \end{aligned}$$

□

Proposition 3.4. (Continuous Functions are Measurable) If $f : \mathbb{R} \to \mathbb{R}$ is continuous, then it is measurable with respect to the Borel σ-algebra $\mathcal{B}(\mathbb{R})$.

Proposition 3.5. (Measurability and σ-Algebras) If $\mathcal{C} \subseteq \mathcal{P}(\mathbb{R})$ is a collection such that $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$, then $f : \Omega \to \mathbb{R}$ is measurable with respect to Σ if and only if $f^{-1}(A) \in \Sigma$ for all $A \in \mathcal{C}$.

Proof. Let $\mathcal{H}$ be the set of all $A \subseteq \mathbb{R}$ such that $f^{-1}(A) \in \Sigma$. We will show that $\mathcal{H}$ is a σ-algebra containing $\mathcal{C}$, the collection of open intervals in $\mathbb{R}$.

First, we show that $\mathcal{H}$ is a σ-algebra:

- **Closure under complements:**

  If $A \in \mathcal{H}$, then by definition $f^{-1}(A) \in \Sigma$. We need to show that $A^c \in \mathcal{H}$. Since $f^{-1}(A^c) = (f^{-1}(A))^c$ and Σ is a σ-algebra (closed under complements), it follows that $f^{-1}(A^c) \in \Sigma$. Therefore, $A^c \in \mathcal{H}$.

- **Closure under countable unions:**

  Suppose $A_i \in \mathcal{H}$ for all $i \in \mathbb{N}$. We need to show that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{H}$. By the definition of $\mathcal{H}$, we know that $f^{-1}(A_i) \in \Sigma$ for each i. Since Σ is a σ-algebra, it is closed under countable unions, so

  $$f^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(A_i) \in \Sigma.$$

  Therefore, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{H}$.

- **Contains $\mathbb{R}$:**

  We observe that $f^{-1}(\mathbb{R}) = S$, where S is the domain of f. Since $S \in \Sigma$, we have $\mathbb{R} \in \mathcal{H}$.

Thus, $\mathcal{H}$ is a σ-algebra.

Next, we show that $\mathcal{H}$ contains $\mathcal{C}$, the collection of open intervals in $\mathbb{R}$. Since f is measurable by assumption, the preimage of any open set (and hence any open interval) is in Σ. Thus, $\mathcal{C} \subseteq \mathcal{H}$.

Finally, since $\mathcal{B}(\mathbb{R})$, the Borel σ-algebra on $\mathbb{R}$, is the smallest σ-algebra containing $\mathcal{C}$, we have $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C}) \subseteq \mathcal{H}$. Therefore, for all $A \in \mathcal{B}(\mathbb{R})$, we have $f^{-1}(A) \in \Sigma$.

This shows that f is measurable with respect to the Borel σ-algebra on $\mathbb{R}$.

$\square$

Example 3.6. In order to show that f is measurable, it suffices to show:

- $f^{-1}(A) \in \Sigma$ for all open A;
- $f^{-1}(A) \in \Sigma$ for all closed A;
- $f^{-1}((-\infty, x]) \in \Sigma$ for $x \in \mathbb{R}$;
- ...

Lemma 3.7. (Measurability of Sums and Products) If $f_1, f_2 : \Omega \to \mathbb{R}$ are measurable functions, then so are $f_1 + f_2$ and $f_1 f_2$.

Proof. We need to show that $(f_1+f_2)^{-1}((x,\infty)) \in \Sigma$ for all $x \in \mathbb{R}$, given $f_1^{-1}((x,\infty)), f_2^{-1}((x,\infty)) \in \Sigma$.

1. Note that for any $s \in S$, $(f_1 + f_2)(s) > x$ if and only if there exists a rational number q such that $f_1(s) > q$ and $f_2(s) > x - q$.
2. Therefore, we can write:

$$(f_1 + f_2)^{-1}((x,\infty)) = \bigcup_{q \in \mathbb{Q}} \left(f_1^{-1}((q,\infty)) \cap f_2^{-1}((x - q,\infty)) \right).$$

3. Since f_1 and f_2 are measurable, $f_1^{-1}((q,\infty))$ and $f_2^{-1}((x - q,\infty))$ belong to Σ for any rational q. As Σ is a σ-algebra, it is closed under countable unions and intersections. Hence, $(f_1 + f_2)^{-1}((x,\infty)) \in \Sigma$, making $f_1 + f_2$ measurable.

The proof for the product is similar. To show that $f_1 f_2$ is measurable, consider the pre-image of (x,∞) under $f_1 f_2$. □

Lemma 3.8. (Measurability of Composition) The composition of measurable functions is measurable.

Proof. Let f_1 and f_2 be measurable functions. We want to show that $f_1 \circ f_2$ is also measurable. Consider:

$$(f_1 \circ f_2)^{-1}(A) = \{x : f_1(f_2(x)) \in A\}$$

$$= \{x : f_2(x) \in f_1^{-1}(A)\} = f_2^{-1}(f_1^{-1}(A)).$$

Since f_1 is measurable, $f_1^{-1}(A)$ is a measurable set. And because f_2 is measurable, $f_2^{-1}(f_1^{-1}(A))$ is also measurable. Therefore, the composition $f_1 \circ f_2$ is measurable. □

Lemma 3.9. (Measurability of inf, sup, liminf, limsup) If $f_n : \Omega \to \mathbb{R}$ is measurable for every $n \in \mathbb{N}$, then so are $\inf_n f_n$, $\sup_n f_n$, $\liminf_{n\to\infty} f_n$, and $\limsup_{n\to\infty} f_n$.

Proof.

1. **(inf):** We want to show that $(\inf_n f_n)^{-1}((x,\infty)) \in \Sigma$. Observe that $(\inf_n f_n)(s) > x$ if and only if $f_n(s) > x$ for all n. Thus:

$$(\inf_n f_n)^{-1}((x,\infty)) = \bigcap_n f_n^{-1}((x,\infty)).$$

2. Since f_n is measurable for all n, $f_n^{-1}((x,\infty)) \in \Sigma$. Σ, being a σ-algebra, is closed under countable intersections, thus making $\inf_n f_n$ measurable.

3. **(sup, lim inf, lim sup):** The proof for $\sup_n f_n$ is similar to the proof for the $\inf_n f_n$. Note that $\liminf_{n\to\infty} f_n(s) = \sup_n \inf_{m\geq n} f_m(s)$ and so is measurable by the previous facts. Similarly $\limsup_{n\to\infty} f_n(s) = \inf_n \sup_{m\geq n} f_m(s)$ is measurable.

□

Lemma 3.10. (Measurability of a Set Defined by Limits) The set $\{s \in \Omega : \lim_{n\to\infty} f_n(s) \text{ exists and is finite}\}$ is in the σ-algebra Σ.

Definition 3.11. (Random Variables) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ that is $\mathcal{F}$-measurable is called a **random variable**.

Example 3.12. Let $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, and $\mathbb{P}(A) = \frac{\#A}{6}$ (model of a die cast) and define:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \in \{1,3,5\}, \\ 0 & \text{if } \omega \in \{2,4,6\}. \end{cases}$$

Here X is called an **indicator variable** (of odd die rolls).
$Y(\omega) = \omega$ is also a random variable.

Definition 3.13. (σ-Algebra Generated by a Random Variable) For any random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$, the **$\sigma$-algebra generated by** X, written $\sigma(X)$, is the smallest (sub) σ-algebra that makes X measurable. This is:

$$\sigma(X) = \sigma(\{X^{-1}(A) : A \text{ is a Borel set}\}).$$

Example 3.14. In the previous example:

$$\sigma(Y) = \mathcal{P}(\{1, 2, 3, 4, 5, 6\}),$$

but

$$\sigma(X) = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}.$$

Definition 3.15. (Law of a Random Variable) Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Define $\mathcal{L}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}\{\omega \in \Omega : X(\omega) \in A\}$. This is a probability measure on $\mathcal{B}(\mathbb{R})$ called the **law** of X.

Definition 3.16. (Cumulative Distribution Function) The law of X is uniquely determined by the **cumulative distribution function**:

$$F_X(t) = \mathcal{L}_X((-\infty, t]) = \mathbb{P}(X \leq t).$$

Proposition 3.17. (Properties of Distribution Functions)

- **Non-decreasing**: $F_X(t) \leq F_X(s)$ if $t \leq s$.
- $\lim_{t \to -\infty} F_X(t) = 0$, and $\lim_{t \to \infty} F_X(t) = 1$.
- **Right-continuous**: $\lim_{t \searrow a} F_X(t) = F_X(a)$.

Proof. The third property can be proven as follows:

$$\begin{aligned}
\lim_{t \searrow a} F_X(t) &= \lim_{t \searrow a} \mathbb{P}(X \in (-\infty, t]) \\
&= \mathbb{P}\left(\bigcap_{t > a} \{\omega : X(\omega) \in (-\infty, t]\} \right) \\
&= \mathbb{P}(\{\omega : X(\omega) \in (-\infty, a]\}) \\
&= F_X(a)
\end{aligned}$$

$\square$

Corollary 3.18. Given a function F satisfying all the properties above, we can define a probability measure $\mathcal{L}$ with $\mathcal{L}((-\infty, t]) = F(t)$.

4 Independence

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

Definition 4.1. (Independence of Events): Events $E_1, E_2, ..., E_n \in \mathcal{F}$ are said to be **independent** if for all choices of $i_1 < i_2 < \cdots < i_k$ where $k \leq n$, we have:

$$\mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = \mathbb{P}(E_{i_1}) \times \mathbb{P}(E_{i_2}) \times \cdots \times \mathbb{P}(E_{i_k}).$$

In simpler terms, events are independent if the occurrence of one does not affect the probability of the others occurring.

Definition 4.2. (Independence of Random Variables) Random variables $X_1, X_2, \ldots$ are said to be **independent** if for any choice of $i_1 < i_2 < \cdots < i_k$ and Borel sets $A_1, A_2, \ldots, A_k$, the events $\{X_{i_1} \in A_1\}, \{X_{i_2} \in A_2\}, \ldots, \{X_{i_k} \in A_k\}$ are independent.

Equivalently, $X_1, X_2, \ldots$ are independent if and only if:

$$\mathbb{P}(X_{i_1} \in A_1, X_{i_2} \in A_2, \ldots, X_{i_k} \in A_k) = \mathbb{P}(X_{i_1} \in A_1) \cdot \mathbb{P}(X_{i_2} \in A_2) \cdot \cdots \cdot \mathbb{P}(X_{i_k} \in A_k).$$

Definition 4.3. (Independence of Sub-σ-algebras) Sub-σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots$ of $\mathcal{F}$ are said to be **independent** if for all choices of $i_1 < i_2 < \cdots < i_k$ and events $G_1 \in \mathcal{G}_{i_1}, G_2 \in \mathcal{G}_{i_2}, \ldots, G_k \in \mathcal{G}_{i_k}$, we have:

$$\mathbb{P}(G_1 \cap G_2 \cap \cdots \cap G_k) = \mathbb{P}(G_1) \cdot \mathbb{P}(G_2) \cdot \cdots \cdot \mathbb{P}(G_k).$$

Remark 4.4. The independence of events and random variables are special cases of the independence of sub-σ-algebras.

- **Events:** For events $E_1, E_2, \ldots$, set $\mathcal{G}_i = \{\emptyset, E_i, E_i^c, \Omega\}$. Then $E_1, E_2, \ldots$ are independent if and only if $\mathcal{G}_1, \mathcal{G}_2, \ldots$ are independent.
- **Random Variables:** For random variables $X_1, X_2, \ldots$, set $\mathcal{G}_i = \sigma(X_i)$. Then, $X_1, X_2, \ldots$ are independent if and only if $\mathcal{G}_1, \mathcal{G}_2, \ldots$ are independent.

Lemma 4.5. (Independence and π-systems)
Let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$, and let $\mathcal{I}$ and $\mathcal{J}$ be π-systems that generate $\mathcal{G}$ and $\mathcal{H}$ respectively: $\sigma(\mathcal{I}) = \mathcal{G}$ and $\sigma(\mathcal{J}) = \mathcal{H}$. Then, $\mathcal{G}$ and $\mathcal{H}$ are independent if and only if:

$$\mathbb{P}(I \cap J) = \mathbb{P}(I) \cdot \mathbb{P}(J) \quad \text{for all} \quad I \in \mathcal{I}, J \in \mathcal{J}.$$

Proof.

($\Longrightarrow$) Since $\mathcal{I} \subseteq \mathcal{G}$ and $\mathcal{J} \subseteq \mathcal{H}$, the condition holds by the definition of independence of sub-σ-algebras.

($\Longleftarrow$) Using Carathéodory's Extension Theorem:

1. **Define a Measure:** Assume $\mathbb{P}(I \cap J) = \mathbb{P}(I) \cdot \mathbb{P}(J)$ holds. Define a set function $\tilde{P}$ on the collection $\{I \cap H : I \in \mathcal{I}, H \in \mathcal{H}\}$ as follows:

$$\tilde{P}(I \cap H) = \mathbb{P}(I) \cdot \mathbb{P}(H).$$

2. **Verify Measure Properties:** We need to verify that $\tilde{P}$ is a well-defined measure:
   - $\tilde{P}(\emptyset) = \mathbb{P}(\emptyset) \cdot \mathbb{P}(H) = 0$.
   - For disjoint sets $(I_i \cap H_i)$,

$$\tilde{P}\left(\bigcup_{i=1}^{\infty}(I_i \cap H_i)\right) = \sum_{i=1}^{\infty} \tilde{P}(I_i \cap H_i) = \sum_{i=1}^{\infty} \mathbb{P}(I_i) \cdot \mathbb{P}(H_i).$$

3. **Apply Extension Theorem:** By Carathéodory's Extension Theorem, $\tilde{P}$ can be extended to a unique measure on $\sigma(\{I \cap H : I \in \mathcal{I}, H \in \mathcal{H}\}) = \mathcal{G} \cap \mathcal{H}$.

4. **Uniqueness of Measures:** Since measures are uniquely determined by their values on a generating π-system, and both $\tilde{P}$ and P agree on $\mathcal{I} \times \mathcal{J}$, they must agree on $\mathcal{G} \cap \mathcal{H}$. Therefore,

$$\mathbb{P}(G \cap H) = \tilde{P}(G \cap H) = \mathbb{P}(G) \cdot \mathbb{P}(H) \quad \text{for all } G \in \mathcal{G}, H \in \mathcal{H}.$$

This proves that $\mathcal{G}$ and $\mathcal{H}$ are independent.

□

Remark 4.6. To verify the independence of random variables $X_1, X_2, \ldots$, it suffices to check:

$$\mathbb{P}(X_1 \le x_1, X_2 \le x_2, \ldots) = \mathbb{P}(X_1 \le x_1) \cdot \mathbb{P}(X_2 \le x_2) \cdot \ldots$$

for all $x_1, x_2, \ldots$. This follows because the sets of the form $\{X_i \le x_i\}$ form a π-system that generates $\sigma(X_i)$.

Example 4.7. Consider $\mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, a] : a \in \mathbb{R}\})$. Any probability measure P on $\mathbb{R}$ is determined by its cumulative distribution function $F(a) = \mathbb{P}((-\infty, a])$.

Lemma 4.8. (Borel-Cantelli Lemma — Second Lemma) Assume that $(E_n)_{n\in\mathbb{N}}$ are independent events and $\sum_{i=1}^{\infty} \mathbb{P}(E_i) = \infty$. Then:

$$\mathbb{P}(\limsup_{n\to\infty} E_n) = \mathbb{P}(E_n \text{ occurs infinitely often}) = 1.$$

Proof. Recall that:

$$\limsup_{n\to\infty} E_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n.$$

Its complement is:

$$(\limsup_{n\to\infty} E_n)^c = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} E_n^c.$$

Now,

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty} E_m^c\right) \le \mathbb{P}\left(\bigcap_{n=M}^{\infty} E_n^c\right) = \lim_{M\to\infty} \prod_{n=M}^{\infty} \mathbb{P}(E_n^c) \quad \text{(by independence)}.$$

Since $1 - x \le e^{-x}$ for $x \ge 0$,

$$\lim_{M\to\infty} \prod_{n=M}^{\infty} \mathbb{P}(E_n^c) \le \lim_{M\to\infty} \prod_{n=M}^{\infty} e^{-\mathbb{P}(E_n)} = \lim_{M\to\infty} \exp\left(-\sum_{n=M}^{\infty} \mathbb{P}(E_n)\right) = 0.$$

Hence, $\mathbb{P}\left(\bigcap_{m=1}^{\infty} E_m^c\right) = 0$ for all m, and by countable subadditivity:

$$\mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} E_n^c\right) = 0.$$

Therefore, $\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 1$. ◻

Remark 4.9. Independence is crucial! If events are not independent, the lemma might not hold. For example, if $E_1 = E_2 = \cdots = E$, then $\mathbb{P}(\limsup_{n\to\infty} E_n) = \mathbb{P}(E)$, which can take any value in $[0, 1]$.

Example 4.10. (Second Borel-Cantelli Lemma)
(Drawing Cards) Take a random card from a deck of n cards, labelled $1, 2, \ldots, n$. Let E_n be the event "card labelled 1 is drawn on the n-th draw?" Assuming all draws are independent and uniform:

$$\mathbb{P}(E_n) = \frac{1}{n}.$$

By the Second Borel-Cantelli Lemma, since $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$, $\mathbb{P}(\text{card 1 is drawn infinitely often}) = 1$.

Example 4.11. (Monkey & Typewriter) A monkey types a sequence of random characters on a keyboard. Assuming each character has a positive probability of occurring, let S be a fixed string of length s.

- **Probability of Typing 'S':** The probability of the first s characters being exactly S is some small positive number, say $\epsilon = \frac{1}{26^s}$ (assuming 26 letters on the keyboard).
- **Independence and Infinite Occurrences:** The events of typing 'S' in characters 1 to s, $s+1$ to $2s$, and so on are independent. Since we have infinitely many of these events, and their probabilities sum to infinity, the Second Borel-Cantelli Lemma tells us that $\mathbb{P}(\text{monkey types 'S' infinitely often}) = 1$.

Definition 4.12. (Tail σ-algebra)
Let $X_1, X_2, \ldots$ be a sequence of random variables. Define $\mathcal{T}_n = \sigma(X_n, X_{n+1}, \ldots)$ and $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{T}_n$. We call $\mathcal{T}$ the **tail σ-algebra**.

Intuitively, the tail σ-algebra contains events determined by the "infinite tail" of the sequence of random variables.

Example 4.13. (Tail Events)

- $\{\lim_{n\to\infty} X_n \text{ exists}\} \in \mathcal{T}$
- $\{\sum_{n=1}^{\infty} X_n \text{ converges}\} \in \mathcal{T}$

Theorem 4.14. *Let $X_1, X_2, \dots$ be independent random variables. Then, for every $T \in \mathcal{T}$, either $P(T) = 0$ or $P(T) = 1$.*

Remark 4.15. Tail events of independent random variables have a deterministic nature: they either happen with probability 1 or don't happen at all (probability 0).

Proof.

1. **Define σ-algebras:** Define $\mathcal{X}_n = \sigma(X_1, X_2, \dots, X_n)$ and $\mathcal{X}_\infty = \sigma(X_1, X_2, \dots)$. Note that $\mathcal{T}_n$ and $\mathcal{X}_n$ are independent for all n.
2. **Independence of $\mathcal{T}$ and $\mathcal{X}_n$:** Since $\mathcal{T} \subseteq \mathcal{T}_n$ for all n, $\mathcal{T}$ and $\mathcal{X}_n$ are independent for all n.
3. **Independence of $\mathcal{T}$ and $\mathcal{X}_\infty$:** Because $\mathcal{X}_\infty$ is generated by the $\mathcal{X}_n$, $\mathcal{T}$ and $\mathcal{X}_\infty$ are independent.
4. **Contradiction:** For any $F \in \mathcal{T}$, we have $P(F \cap F) = P(F) \cdot P(F)$. This implies $P(F) = P(F)^2$. The only solutions to this equation are $P(F) = 0$ or $P(F) = 1$.

□

Remark 4.16. If ξ is a $\mathcal{T}$-measurable random variable, then $P(\xi = c) = 1$ for some constant $c \in [-\infty, \infty]$. This means $\mathcal{T}$-measurable random variables are essentially constant (almost surely).

Remark 4.17. (Application of Kolmogorov's 0-1 Law) Consider a population (e.g., humans) where each member has offspring independent of other members and exactly one ancestor (e.g., considering only the paternal lineage).

- **Galton-Watson Process:** This model is called a Galton-Watson process.
- **Tail Event:** Let v_n be an arbitrary member of the population at generation n. The event "a randomly picked individual in the population will eventually be descended from v_n" is a tail event.
- **Kolmogorov's 0-1 Law:** By the Kolmogorov 0-1 Law, this tail event has probability 0 or 1. Therefore, eventually, either everyone or no one will be descended from v_n.

5 Integration

Definition 5.1. Let (S, Σ, μ) be a measure space, and let f be (Σ)-measurable. We define $\int_S f \, d\mu$ in three steps:

1. **Define the integral for indicator functions:**
   Let $I_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}$, where $A \in \Sigma$.
   We set
   $$\int_S I_A \, d\mu = \mu(A).$$
2. **Define the integral for finite linear combinations of indicator functions (step functions):**
   Let $g(s) = \sum_{k=1}^n a_k I_{A_k}(s)$, where $a_k \in \mathbb{R}^+$ and $A_k \in \Sigma$. (Here, g is called a *step function.*)
   We define
   $$\int_S g \, d\mu = \sum_{k=1}^n a_k \int_S I_{A_k} \, d\mu = \sum_{k=1}^n a_k \mu(A_k).$$
3. **Define the integral for arbitrary measurable and non-negative functions:**
   For arbitrary measurable and non-negative $f \in \mathcal{M}^+$, we define
   $$\int_S f \, d\mu = \sup \left\{ \int_S g \, d\mu : g \text{ is a non-negative step function, } g(s) \leq f(s) \right\}.$$

Remark 5.2. If f is already a step function, then the definitions in step 3 and step 2 are equivalent.

Remark 5.3. We may also write $\mu(f)$ for $\int_S f \, d\mu$.

Definition 5.4. (Extending the Integral to All Measurable Functions)
We extend step (3) to all measurable functions by defining:

$$f^+(s) = \begin{cases} f(s) & \text{if } f(s) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$f^-(s) = \begin{cases} -f(s) & \text{if } f(s) \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Both f^+ and f^- are non-negative, and $f = f^+ - f^-$. We define

$$\int_S f \, d\mu = \int_S f^+ \, d\mu - \int_S f^- \, d\mu.$$

Proposition 5.5. (Properties of the Integral)

1. **Linearity:**

$$\int_S (a \cdot f + b \cdot g) \, d\mu = a \int_S f \, d\mu + b \int_S g \, d\mu \quad \text{for } a, b \in \mathbb{R}.$$

2. **Monotonicity:** If $f \leq g$ for all $s \in S$ (or even for almost all s with respect to μ), then

$$\int_S f \, d\mu \leq \int_S g \, d\mu.$$

3. **Triangle Inequality:**

$$\left| \int_S f \, d\mu \right| = \left| \int_S f^+ \, d\mu - \int_S f^- \, d\mu \right| \leq \left| \int_S f^+ \, d\mu \right| + \left| \int_S f^- \, d\mu \right| = \int_S |f| \, d\mu$$

Example 5.6. (Integral)

- If μ is the Lebesgue measure on $\mathbb{R}$, then this is the familiar *Lebesgue integral*. If both the Lebesgue and Riemann integrals exist for f, then they are equal.
- Let μ be the counting measure on the positive integers. Then

$$\int_S f \, d\mu = \sum_{n=1}^{\infty} f(n).$$

We can restrict the domain of integration by setting

$$\int_A f\,d\mu = \int_S f \cdot I_A\,d\mu, \quad A \in \Sigma.$$

Definition 5.7. (Integrable Functions and $\mathcal{L}^1$ Space)
We say that f is *(μ)-integrable* if $\int_S f^+\,d\mu$ and $\int_S f^-\,d\mu$ are finite. If this is not the case, the integral is *undefined* (e.g., we cannot have $\infty - \infty$ in the definition).

We write $\mathcal{L}^1(S, \Sigma, \mu)$ for the space of integrable functions, i.e., all $f \in \mathcal{L}^1(S, \Sigma, \mu)$ are integrable.

Remark 5.8. If $f(s) = +\infty$ for some $s \in S$, then f can only be integrable if $\mu(\{s : f(s) = +\infty\}) = 0$.

Lemma 5.9. If f is a non-negative measurable function with $\int_S f\,d\mu = 0$, then $\mu(\{f > 0\}) = 0$; i.e., $f(s) = 0$ μ-almost everywhere.

Proof. The set $\{s : f(s) > 0\} = \bigcup_{n\in\mathbb{N}}\{s : f(s) > \frac{1}{n}\}$. Either $\mu(\{s : f(s) > \frac{1}{n}\}) = 0$ for all n, giving $\mu(\{f > 0\}) = 0$, or $\mu(\{f > \frac{1}{n}\}) > 0$ for some n. Let $A = \{f > \frac{1}{n}\}$. By monotonicity, $\int_S f\,d\mu \geq \int_S \frac{1}{n} I_A\,d\mu = \frac{1}{n}\mu(A) > 0$, a contradiction. Thus, $\mu(\{f > 0\}) = 0$.

□

Question: When is it true that

$$\lim_{n\to\infty} \int_S f_n\,d\mu = \int_S \lim_{n\to\infty} f_n\,d\mu?$$

Not always! For example, let $f_n = I_{[n,n+1]}$. Then $\int_{\mathbb{R}} f_n\,dx = 1$ for all $n \in \mathbb{N}$, but $\lim_{n\to\infty} f_n(x) = 0$ for all x. So,

$$\int_{\mathbb{R}} \lim_{n\to\infty} \mid f_n \mid \,dx = 0.$$

However, There are circumstances that allow interchanging limits & integrals.

Theorem 5.10. *(Monotone Convergence Theorem) Let $\{f_n\}$ be a sequence of non-negative measurable functions such that $f_n \uparrow f$, i.e., $f_n(x)$ is non-decreasing in n and $\lim_{n\to\infty} f_n(x) = f(x)$ for all x. Then*

$$\lim_{n\to\infty} \mu(f_n) = \mu\left(\lim_{n\to\infty} f_n\right) = \mu(f).$$

We can always approximate measurable functions by monotone sequences of *step functions*. Set:

$$\alpha^{(r)}(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ (i-1)2^{-r}x & \text{if } (i-1)2^{-r} \leq x \leq 2^{-r}, \\ 1 & \text{if } x > r. \end{cases}$$

Note that $\alpha^{(r)}(x) \uparrow x$, $\lim_{r\to\infty} \alpha^{(r)}(x) = x$, and $\alpha^{(r)}(x)$ is non-decreasing in x.

Set $f^{(r)}(x) = \alpha^{(r)}(f(x))$, $f \in m\Sigma^+$, we get a function $f^{(r)}(x)$ which:

- is a step function
- $f^{(r)} \uparrow f$

This gives a general proof strategy:

1. Prove something for indicator functions.
2. Extend to step functions by linearity.
3. Extend to arbitrary $f \in m\Sigma^+$ by approximation with step functions and monotonicity.
4. Extend to $f \in m\Sigma$ by splitting into positive and negative parts.

Lemma 5.11. Suppose f and g are integrable and $f = g$ almost everywhere. Then:

$$\int_S f\, d\mu = \int_S g\, d\mu.$$

Proof. (Sketch)

Follow the above recipe. Consider $\int_S \mid f - g \mid \, d\mu$. This is $= 0$ trivially for indicator functions. Following the steps above shows $\int_S \mid f - g \mid \, d\mu = 0$ generally. $\square$

Corollary 5.12. If $f_n \uparrow f$ almost everywhere, then $\mu(f_n) \uparrow \mu(f)$ is still true.

Lemma 5.13. (Fatou's Lemma) Suppose f_n is a sequence of non-negative measurable functions. Then:

$$\mu\left(\liminf_{n\to\infty} f_n\right) \leq \liminf_{n\to\infty} \mu(f_n).$$

Proof. Consider the sequence $g_k = \inf_{n\geq k} f_n$. Then $\lim_{k\to\infty} g_k = \liminf_{n\to\infty} f_n$. Since g_k is monotone ($g_k \uparrow \liminf_{n\to\infty} f_n$), we can apply monotone convergence and get $\mu\left(\lim_{k\to\infty} g_k\right) = \lim_{k\to\infty} \mu(g_k)$. But $g_k \leq f_n$ for all $n \geq k$, so $\mu(g_k) \leq \mu(f_n)$ for all $n \geq k$. In particular, $\mu(g_k) \leq \inf_{n\geq k} \mu(f_n)$. Hence:

$$\mu\left(\liminf_{n\to\infty} f_n\right) = \lim_{k\to\infty} \mu(g_k) \leq \lim_{k\to\infty} \inf_{n\geq k} \mu(f_n) = \liminf_{n\to\infty} \mu(f_n),$$

as required. □

Corollary 5.14. (Reverse Fatou Lemma) Suppose $f_n \leq g$ for some non-negative step function g. Then:

$$\mu\left(\limsup_{n\to\infty} f_n\right) \geq \limsup_{n\to\infty} \mu(f_n).$$

Proof. Follows from Fatou's Lemma applied to $g - f_n$. □

Theorem 5.15. *(Dominated Convergence Theorem) Let f_n be a sequence of measurable functions and assume $\mid f_n \mid\leq g$ for some integrable g; assume $f_n \to f$ pointwise; then:*

- $\mu(\mid f_n - f \mid) = \int_S \mid f_n - f \mid \, d\mu \to 0$, *and*
- $\mu(f_n) = \int_S f_n \, d\mu \to \int_S f \, d\mu = \mu(f)$.

Proof.

$$\mid f_n - f \mid\leq\mid f_n \mid + \mid f \mid\leq 2g.$$

By (Reverse) Fatou's Lemma:

$$\limsup_{n\to\infty} \mu(\mid f_n - f \mid) \leq \mu\left(\limsup_{n\to\infty} \mid f_n - f \mid\right) = 0.$$

$$\implies 0 \leq \liminf_{n\to\infty} \mu(\mid f_n - f \mid) \leq \limsup_{n\to\infty} \mu(\mid f_n - f \mid) = 0 \implies \lim_{n\to\infty} \mu(\mid f_n - f \mid) = 0.$$

It follows that $| \mu(f_n) - \mu(f) | = | \mu(f_n - f) | \leq \mu(| f_n - f |) \to 0$. So, $\mu(f_n) \to \mu(f)$.

□

Lemma 5.16. (Scheffé's Lemma) Suppose f_n, f are non-negative functions, such that $f_n \to f$ (almost) everywhere. Then, $\mu(f_n) \to \mu(f)$ if and only if $\mu(| f_n - f |) \to 0$.

The integral of a function f with respect to a measure μ, written as $\int f \, d\mu$, was defined in three steps:

1. **Indicator Functions:** For $I_A(s) = \begin{cases} 0, & s \notin A \\ 1, & s \in A \end{cases}$ for $A \in \Sigma$, the integral is defined as $\int I_A \, d\mu = \mu(A)$.
2. **Step Functions:** For a step function $g(s) = \sum_{k=1}^{n} a_k I_{A_k}(s)$ where $a_k \in \mathbb{R}^+$ and $A_k \in \Sigma$, the integral is defined as
$$\int g \, d\mu = \sum_{k=1}^{n} a_k \mu(A_k).$$
3. **General Measurable Functions:** For $f \in m\Sigma^+$ (non-negative measurable functions), the integral is defined as
$$\int f \, d\mu = \sup \left\{ \int g \, d\mu : g \text{ is a non-negative step function, } g \leq f \right\}.$$

For $f \in m\Sigma$, the integral is defined by splitting f into its positive and negative parts: $f = f^+ - f^-$. Then, if both are finite,
$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

An important question arises: **when can limits and integrals be exchanged?**

Two important theorems address this question:

1. **Monotone Convergence Theorem:** This theorem provides conditions for exchanging limits and integrals when the sequence of functions is monotone.
2. **Dominated Convergence Theorem:** This theorem provides alternative conditions for exchanging limits and integrals.

5.1 A New Measure and the Radon-Nikodym Theorem

Definition 5.17. (Modifying Measures) Given $f \in m\Sigma^+$, the restricted integral is defined as

$$\lambda(A) = \int_A f \, d\mu = \int f.I_A \, d\mu \quad , A \in \Sigma.$$

This defines a new measure $\lambda = \mu(f; A)$ on (S, Σ), as it satisfies:

- $\lambda(\emptyset) = 0$,
- σ-additivity: For disjoint sets $\{A_i\}$ in Σ,

$$\begin{aligned}
\lambda\left(\bigcup_{i=1}^{\infty} A_i\right) &= \int f \cdot I_{\bigcup_{i=1}^{\infty} A_i} \, d\mu = \int f \cdot \left(\sum_{i=1}^{\infty} I_{A_i}\right) d\mu \\
&= \int f \cdot \lim_{N\to\infty} \sum_{i=1}^{N} I_{A_i} \, d\mu = \int \lim_{N\to\infty} \sum_{i=1}^{N} f \cdot I_{A_i} \, d\mu \\
&= \lim_{N\to\infty} \int \sum_{i=1}^{N} f \cdot I_{A_i} \, d\mu = \lim_{N\to\infty} \sum_{i=1}^{N} \int f \cdot I_{A_i} \, d\mu \\
&= \lim_{N\to\infty} \sum_{i=1}^{N} \lambda(A_i) = \sum_{i=1}^{\infty} \lambda(A_i).
\end{aligned}$$

The function $f = \frac{d\lambda}{d\mu}$ is called the **density** of λ with respect to μ.

Important Observation: If $\mu(A) = 0$, then $\lambda(A) = \int_A f \, d\mu = 0$. This implies that all null sets of μ are also null sets of λ.

Definition 5.18. (σ-Finite Measure) A measure μ is called **σ-finite** if there exist sets $A_i \in \Sigma$, $i \in \mathbb{N}$, such that $\bigcup_{i=1}^{\infty} A_i = S$ and $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$.

Theorem 5.19. ***(Radon-Nikodym Theorem)*** *Let λ and μ be two σ-finite measures on (S, Σ). If $\lambda(A) = 0$ whenever $\mu(A) = 0$ for $A \in \Sigma$, then there exists a density function $f = \frac{d\lambda}{d\mu}$ such that $\lambda(A) = \int_A f \, d\mu$ for all $A \in \Sigma$.*

Example 5.20. (**Lebesgue Measure**) Let μ be the Lebesgue measure on $\mathbb{R}$, and consider the function $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ (the standard normal density function). Then, the measure λ defined as $\lambda(A) = \int_A f(x)\, d\mu$ represents the probability that a standard normal random variable lies in A.

6 Expectation I

Expectations, also referred to as expected values, means, or arithmetic means, are integrals defined relative to probability measures.

Definition 6.1. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$, the **expectation** of X is defined as

$$\mathbb{E}(X) = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \int X\, d\mathbb{P}.$$

Proposition 6.2. If the expectation $\mathbb{E}(X)$ exists (meaning $\int |X|\, d\mathbb{P} < \infty$), X is said to be **integrable** with respect to $\mathbb{P}$.

Example 6.3. (Finite Probability Space)
Consider a finite probability space $\Omega = \{1, 2, 3, 4, 5, 6\}$ with $\mathbb{P}(\{i\}) = \frac{1}{6}$ (representing a die roll). Every function $f : \Omega \to \mathbb{R}$ can be written as a finite sum of indicator functions:

$$X(\omega) = X(1)I_{\{1\}}(\omega) + X(2)I_{\{2\}}(\omega) + \cdots + X(6)I_{\{6\}}(\omega).$$

Therefore, the expectation of any function f on this probability space is:

$$\mathbb{E}(X) = \int X(\omega)\, d\mathbb{P}(\omega) = \frac{1}{6}\left(X(1) + X(2) + \cdots + X(6)\right).$$

Several theorems on integrals translate directly into theorems about expected values. For example:

- **Monotone Convergence:** If $0 \leq X_n \uparrow X$, then $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$.
- **Dominated Convergence:** If $\mid X_n \mid \leq Y$ with $\mathbb{E}(Y) < \infty$ and $X_n \to X$ almost surely, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.
- **Fatou's Lemma:** $\mathbb{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n)$.

Remark 6.4. We can also define the expectation of a random variable X restricted to an event E:

$$\mathbb{E}(X;A) = (\mu(X;A)) = \mathbb{E}(X \cdot I_A) = \int_A X \, d\mathbb{P}.$$

Theorem 6.5. *(Markov's Inequality)*
Let Z be a random variable taking values in $G \subseteq \mathbb{R}$ almost surely ($Z : \Omega \to G$) and let $g : G \to [0,\infty)$ be a non-decreasing measurable function. Then, for any $c \in G$:

$$\mathbb{E}(g(Z)) \geq \mathbb{E}\left(g(Z) \cdot I_{\{Z\geq c\}}\right) = \mathbb{E}\left(g(Z); \{Z \geq c\}\right) \geq \mathbb{E}\left(g(c); \{Z \geq c\}\right)$$

$$= g(c) \cdot \mathbb{E}\left(I_{\{Z\geq c\}}\right) = g(c)\mathbb{P}\left(Z \geq c\right)$$

Therefore,

$$\mathbb{P}(Z \geq c) \leq \frac{\mathbb{E}(g(Z))}{g(c)} \qquad \textit{(Markov's Inequality)}$$

Special cases:

- *For non-negative Z and $c > 0$ (for example, $g(x) = x$),*

$$\mathbb{P}(Z \geq c) \leq \frac{\mathbb{E}(Z)}{c}.$$

- *For $g(x) = e^{\theta x}$ ($\theta > 0$), for all $c > 0$:*

$$\mathbb{P}(Z \geq c) \leq e^{-\theta c}\mathbb{E}(e^{\theta Z})$$

Example 6.6. Let $Z : \Omega \to \mathbb{N} \cup \{0\}$, then

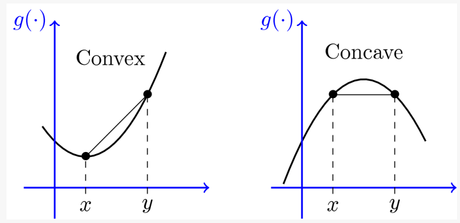
$$\mathbb{P}(Z \neq 0) = \mathbb{P}(Z \geq 1) \leq \mathbb{E}(Z)$$

Definition 6.7. A function $f : I \to \mathbb{R}$ is said to be **convex** on I if for all $x, y \in I$ and $0 \leq \lambda \leq 1$:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

Example 6.8.

- $f(x) = \mid x \mid$
- $f(x) = x^r$ for $r \geq 1$
- $f(x) = e^x$
- $f(x) = c$ for $c \in \mathbb{R}$
- $f(x) = cx$ for $c \in \mathbb{R}$



Theorem 6.9. *(Jensen's inequality)*
Let $f : I \to \mathbb{R}$ be convex and $X : \Omega \to I$ be a random variable. Then,

$$\mathbb{E}(f(X)) \geq f(\mathbb{E}(X)).$$

Proof. The proof of Jensen's inequality starts by reframing the convexity condition in terms of slopes.

A function $f(x)$ is convex on an interval I if, for any points $u < v < w$ in the interval, the slope of the line segment connecting $(u, f(u))$ and $(v, f(v))$ is less than or equal to the slope of the line segment connecting $(v, f(v))$ and $(w, f(w))$. This can be expressed mathematically as:

$$\frac{f(v) - f(u)}{v - u} \leq \frac{f(w) - f(v)}{w - v}.$$

This interpretation of convexity implies that the left and right derivatives of f exist.

A function's **left derivative** at a point is the limit of the slopes of lines drawn between that point and points to its left, as those points get arbitrarily close. The **right derivative** is defined similarly, but with points to the right.

For any x between the left and right derivatives, we have:

$$f(x) \geq f(v) + m(x - v),$$

where m is any number between the left and right derivatives of f at v.

This essentially says that the graph of a convex function lies above its tangent lines.

Substituting $v = \mathbb{E}[X]$ and $m = \frac{f(w)-f(v)}{w-v}$, and taking expectations of both sides, we get:

$$\mathbb{E}[f(X)] \geq \mathbb{E}\left[f(\mathbb{E}[X]) + m(X - \mathbb{E}[X])\right].$$

Since $\mathbb{E}[X]$ and m are constants, this simplifies to:

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]) + m[\mathbb{E}[X] - \mathbb{E}[X]] = f(\mathbb{E}[X]).$$

This concludes the proof of Jensen's inequality. □

Definition 6.10. ($\mathcal{L}^P$ Norms) For $p \geq 1$, the $\mathcal{L}^P$ norm of a random variable X is defined as:

$$\mid X \mid_p = \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p}.$$

$\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ denotes the collection of all random variables X for which $\mid X \mid_p < \infty$.

Remark 6.11. For $0 < p < 1$, modifications are needed to ensure the triangle inequality holds, but this is outside the scope of the current discussion.

Jensen's Inequality Implication

Jensen's inequality implies that if X is a non-negative random variable and $r \geq p \geq 1$, then:

$$\mid X \mid_r \geq \mid X \mid_p .$$

This means that as p increases, the space of random variables with finite $\mathcal{L}^p$ norm shrinks.

This leads to the inclusion:

$$\mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}) \subseteq \cdots \subseteq \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) \subseteq \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Theorem 6.12. *(Cauchy-Schwarz Inequality) The Cauchy-Schwarz inequality states that for any two random variables X and Y in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$:*

$$\mid \mathbb{E}[XY] \mid \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}.$$

This can be compactly written as:

$$\mid \mathbb{E}[XY] \mid \leq \mid X \mid_2 \mid Y \mid_2 .$$

Proof.

1. Define truncated random variables $X_n = \max\{-n, \min(X, n)\}$ and $Y_n = \max\{-n, \min(Y, n)\}$. These truncated variables are bounded.

2. For any real numbers a and b, the expectation $\mathbb{E}[(aX_n + bY_n)^2] \geq 0$ because the square of a real number is always non-negative.

3. Expanding the square in the expectation gives:

$$a^2\mathbb{E}[X_n^2] + 2ab\mathbb{E}[X_n Y_n] + b^2\mathbb{E}[Y_n^2] \geq 0.$$

   This is a quadratic expression in a with coefficients involving expectations.

4. For the quadratic to be non-negative for all a and b, its discriminant must be non-positive. Recall that the discriminant of a quadratic $ax^2 + bx + c$ is $b^2 - 4ac$. Applying this to our inequality gives:

$$(2\mathbb{E}[X_n Y_n])^2 - 4\mathbb{E}[X_n^2]\mathbb{E}[Y_n^2] \leq 0.$$

5. Simplifying and rearranging, we get:

$$\mathbb{E}[X_n Y_n]^2 \leq \mathbb{E}[X_n^2]\mathbb{E}[Y_n^2].$$

   Taking the square root of both sides and applying the Monotone Convergence Theorem (as X_n and Y_n converge to X and Y, respectively), we obtain the Cauchy-Schwarz inequality:

$$\mid \mathbb{E}[XY] \mid \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}.$$

□

Theorem 6.13. *(Cauchy-Schwarz Inequality) If $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$, then XY is integrable, and*

$$\mid \mathbb{E}(XY) \mid \leq \mathbb{E}(\mid XY \mid) \leq \mid |X| \mid_2 || Y ||_2 = \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$$

Cauchy-Schwarz implies that $\langle X, Y \rangle = \mathbb{E}(XY)$ becomes a well-defined inner product.

Corollary 6.14.

$$|| X + Y ||_2^2 \leq (\mid |X| \mid_2 + || Y ||_2)^2$$

Proof.

$$\mathbb{E}(\mid X + Y \mid^2) = \mathbb{E}(|X|^2 + 2\mathbb{E}(|X| \mid Y \mid) + \mathbb{E}(\mid Y \mid^2))$$

$$\leq \mathbb{E}(|X|^2) + 2 \mid |X| \mid_2 || \ Y \ ||_2 + \mathbb{E}(| \ Y \ |^2)$$

$$= \mid |X| \mid_2^2 + 2 \mid |X| \mid_2 || \ Y \ ||_2 + || \ Y \ ||_2^2$$

$$= (\mid |X| \mid_2 + || \ Y \ ||_2)^2$$

The norm $|| \cdot ||_2$ satisfies the triangle inequality. $\square$

Definition 6.15. Let X, Y be random variables with $\mu_X = \mathbb{E}(X)$ and $\mu_Y = \mathbb{E}(Y)$. We define the covariance as

$$Cov(X, Y) = \mathbb{E}((X - \mu_X)(Y - \mu_Y)) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

The variance is given by

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

Proposition 6.16. We have the following properties:

- $Var(X) \geq 0$
- If X, Y are independent, then $Cov(X, Y) = 0$
- $\mid Cov(X, Y) \mid \leq \sqrt{Var(X)Var(Y)}$ (by Cauchy-Schwarz Inequality)

Definition 6.17. (Correlation Coefficient)

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \in [-1, 1]$$

It is equal to ± 1 if and only if there is a linear relationship between X and Y.

The Cauchy-Schwarz Inequality can be extended to the Hölder Inequality:

Theorem 6.18. *(Hölder Inequality) Assume* $X \in \mathcal{L}^p$, $Y \in \mathcal{L}^q$ *for* $\frac{1}{p} + \frac{1}{q} = 1$, $p, q \geq 1$.

$$\mid \mathbb{E}(XY) \mid \leq \mathbb{E}(\mid XY \mid) \leq \mid |X| \mid_p || \ Y \ ||_q$$

Cauchy-Schwarz is the case when $p = q = 2$.

Remark 6.19. This holds for arbitrary measure spaces:

$$\int \mid fg \mid \, d\mu \leq \left(\int \mid f \mid^p \, d\mu \right)^{1/p} \left(\int \mid g \mid^q \, d\mu \right)^{1/q}$$

Corollary 6.20. (Minkowski's Inequality)

$$\mid\mid X + Y \mid\mid_p \leq \mid \mid X \mid \mid_p + \mid\mid Y \mid\mid_p \quad \text{for} \quad p \geq 1$$

Proof. We may assume $p > 1$ (otherwise it's just the triangle inequality).

$$(\mid\mid X + Y \mid\mid_p)^p = \mathbb{E}(\mid X + Y \mid^p) = \mathbb{E}(\mid X + Y \mid^{p-1} \mid X + Y \mid)$$

$$\leq \mathbb{E}(\mid X + Y \mid^{p-1} \mid X \mid) + \mathbb{E}(\mid Y \mid\mid X + Y \mid^{p-1})$$

$$\leq \mid \mid X \mid \mid_p \mid\mid\mid X + Y \mid^{p-1} \mid\mid_q + \mid\mid Y \mid\mid_p \mid\mid\mid X + Y \mid^{p-1} \mid\mid_q$$

where $\frac{1}{p} + \frac{1}{q} = 1 \Rightarrow q = \frac{p}{p-1}$.

$$= \mid \mid X \mid \mid_p \left(\mathbb{E}(\mid X + Y \mid^{q(p-1)}) \right)^{1/q} + \mid\mid Y \mid\mid_p \left(\mathbb{E}(\mid X + Y \mid^{q(p-1)}) \right)^{1/q}$$

$$= (\mid \mid X \mid \mid_p + \mid\mid Y \mid\mid_p) \left(\mathbb{E}(\mid X + Y \mid^p) \right)^{(p-1)/p}$$

$$(\mid\mid X + Y \mid\mid_p)^{p/(p-1)} \leq \mid \mid X \mid \mid_p + \mid\mid Y \mid\mid_p$$

$$\Rightarrow (\mid\mid X + Y \mid\mid_p)^p \leq (\mid \mid X \mid \mid_p + \mid\mid Y \mid\mid_p)^{p/(p-1)(p-1)}$$

But $\frac{p}{(p-1)(p-1)} = 1$, and the claim follows.

$\square$

Thus, $\mid\mid \cdot \mid\mid_p$ is a norm for $p \geq 1$.

Theorem 6.21. *$\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ is a complete space, i.e., Cauchy sequences converge.*

6.1 Densities

Proposition 6.22. Suppose $X : \Omega \to \mathbb{R}$ is a random variable with law Λ_X:

$$\Lambda_X(A) = \mathbb{P}(X \in A), \quad A \in \varepsilon$$

Then for every Borel measurable function f, we have:

$$\mathbb{E}(f(X)) = \int_\Omega f(X)\, d\mathbb{P} = \int_\mathbb{R} f(x)\, d\Lambda_X(x)$$

Proof. First for indicator functions: Let $f = \mathbb{I}_A$, where $A \in \mathcal{B}(\mathbb{R})$ then

$$\int_\Omega \mathbb{I}_A(X)\, d\mathbb{P} = \mathbb{E}(\mathbb{I}_{X \in A}) = \mathbb{P}(\{X \in A\})$$

and

$$\int_\mathbb{R} \mathbb{I}_A(x)\, d\Lambda_X = \int_A 1\, d\Lambda_X = \Lambda_X(A) = \mathbb{P}(X \in A)$$

The property then extends by linearity to step functions and to $f \in m\varepsilon^+$ by the **Monotone Convergence Theorem** (MCT). Finally, for $f \in m\varepsilon$, this is shown by splitting into f^+ and f^-.

□

If X has a density, we can also express $\mathbb{E}(f(X))$ in terms of a density.

Theorem 6.23. *If X has a density φ (and $\mathbb{P}(X \in A) = \int_A \varphi(x)\, dx$), then*

$$\mathbb{E}(f(X)) = \int_\mathbb{R} f(x)\varphi(x)\, dx$$

Proof. As before. □

Remark 6.24. φ here is the same 'density' as in the Radon-Nikodym sense.
Recall that X, Y are independent iff

$$\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \quad \text{for all} \quad A, B \in \varepsilon$$

Theorem 6.25. *If X, Y are independent and integrable, then so is XY, and*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

Proof. We can assume X, Y are non-negative (otherwise split by $X = X^+ - X^-$, etc.). We can approximate X, Y by increasing step functions $\alpha^r(X) \uparrow X$, $\alpha^r(Y) \uparrow Y$. Each is a linear combination of indicator functions:

$$\mathbb{I}_A(X) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}, \quad \mathbb{I}_B(Y) = \begin{cases} 1 & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases}$$

and for each indicator, we have

$$\mathbb{E}(\mathbb{I}_A(X)\mathbb{I}_B(Y)) = \mathbb{P}(X \in A \cap Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) = \mathbb{E}(\mathbb{I}_A(X))\mathbb{E}(\mathbb{I}_B(Y))$$

We extend this by linearity to $\mathbb{E}(\alpha^r(X)\alpha^r(Y)) = \mathbb{E}(\alpha^r(X))\mathbb{E}(\alpha^r(Y))$. Taking $r \to \infty$ and using the Monotone Convergence Theorem (MCT) gives $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, as required. ☐

Corollary 6.26. If X, Y are independent, then $Cov(X, Y) = 0$ and $Var(X + Y) = Var(X) + Var(Y)$.

Proof. The first is immediate from the definition. The second follows from:

$$\mathbb{E}((X + Y)^2) = \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2)$$

and

$$\mathbb{E}(X + Y)^2 = \mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)^2$$

Thus, subtracting gives

$$Var(X + Y) = \mathbb{E}((X + Y)^2) - \mathbb{E}(X + Y)^2 = Var(X) + Var(Y)$$

☐

Remark 6.27. $\mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY)$ does not imply X, Y are independent.

Proof. **Counterexample:**

X/Y	-1	0	1
1	λ	0	$-\lambda$
-1	0	$1-2\lambda$	0

We have $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, and

$$\mathbb{E}(XY) = \frac{1}{4}(\lambda) + \frac{1}{2}(0) + \frac{1}{4}(-\lambda) = 0$$

However, X and Y are not independent, since

$$\mathbb{P}(X = 1 \cap Y = 0) = 0 \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$\square$

7 The Strong Law of Large Numbers and Conditional Expectation

7.1 Strong Law of Large Numbers — First Version

Theorem 7.1. *(Strong Law of Large Numbers — weak version) Let $X_1, X_2, \ldots$ be a sequence of independent random variables. Suppose that $\mathbb{E}(X_i) = 0$ and $\mathbb{E}(X_i^4) \leq K < \infty$ for all i and some uniform constant $K \geq 0$. Then,*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to 0 \quad \textit{almost surely.}$$

Proof. We consider the fourth moment. Let

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Then,

$$\mathbb{E}(S_n^4) = \mathbb{E}\left((X_1 + X_2 + \cdots + X_n)^4\right) = \sum_{i=1}^{n} \mathbb{E}(X_i^4) + 6 \sum_{1\leq i<j\leq n} \mathbb{E}(X_i^2 X_j^2) + \\ + 4 \sum_{1\leq i<j\leq n} \mathbb{E}(X_i^3)\mathbb{E}(X_j) + 12 \sum_{1\leq i<j<k\leq n} \mathbb{E}(X_i^2)\mathbb{E}(X_j)\mathbb{E}(X_k) + 24 \sum_{1\leq i<j<k<l\leq n} \mathbb{E}(X_i X_j X_k X_l).$$

Remainder of form $\mathbb{E}(X_i)\mathbb{E}(\ldots)$ so:

$$\mathbb{E}(S_4^4) = \sum \mathbb{E}(X_n^4) + \sum_{i\neq j} \mathbb{E}(X_i^2 X_j^2)$$

Expanding the right-hand side gives us terms of the following forms:

1. $\mathbb{E}(X_i^4)$: These are bounded by K.
2. $\mathbb{E}(X_i^2 X_j^2) \leq \sqrt{\mathbb{E}(X_i^4)\mathbb{E}(X_j^4)} = K$ for all i, j by Cauchy-Schwarz inequality.

All other terms, such as $\mathbb{E}(X_i X_j X_k X_l)$ for distinct i, j, k, l, will be zero because of independence and $\mathbb{E}(X_i) = 0$.

Thus, there are at most $(n + 6\binom{n}{2} + 4!\binom{n}{4})$ non-zero terms in the expansion of $\mathbb{E}(S_n^4)$. Therefore,

$$\mathbb{E}(S_n^4) \leq (n + 6\binom{n}{2} + 4!\binom{n}{4})K = (n + 3n(n-1) + 3n(n-1)(n-2)(n-3))K \leq 3Kn^2.$$

So,

$$\mathbb{E}\left(\left(\frac{S_n}{n}\right)^4\right) \leq \frac{3Kn^2}{n^4} = \frac{3K}{n^2}.$$

Thus,

$$\mathbb{E}\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4\right) = \sum_{n=1}^{\infty}\mathbb{E}\left(\left(\frac{S_n}{n}\right)^4\right) \leq \sum_{n=1}^{\infty}\frac{3K}{n^2} < \infty.$$

Therefore, with probability 1, $\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4$ converges and with probability 1, $\frac{S_n}{n} \to 0$. $\square$

Remark 7.2. (**Special Case**) If $X_1, X_2, \ldots$ are independent and identically distributed (i.i.d.) with $\mathbb{E}(\mid X_i \mid) < \infty$, then:

$$Y_i = X_i - \mathbb{E}(X_i)$$

has expectation 0. $Y_1, Y_2, \ldots$ satisfy the conditions above, and

$$\frac{Y_1 + \cdots + Y_n}{n} = \frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_i) \to 0.$$

Therefore,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to m \quad \text{where } m = \mathbb{E}(X_i).$$

Theorem 7.3. *(Chebyshev's Inequality) We can derive the distance to $m = \mathbb{E}(X)$ using Chebyshev's inequality:*

$$\mathbb{P}(\mid X - \mathbb{E}(X) \mid \geq c) \leq \frac{Var(X)}{c^2}.$$

Remark 7.4. This is a special case of Markov's inequality:

$$\mathbb{P}(\mid X - m \mid^2 \geq c) \leq \frac{\mathbb{E}(\mid X - m \mid^2)}{c} = \frac{\mathrm{Var}(X)}{c^2}.$$

Remark 7.5. Applying this to

$$\frac{X_1 + \cdots + X_n}{n} - m = \frac{X_1 + X_2 + \cdots + X_n - n\mathbb{E}(X)}{n}$$

where $X_1, X_2, \dots$ are independent and identically distributed (i.i.d.) random variables with $m = \mathbb{E}(X_i)$ and $\sigma^2 = \mathrm{Var}(X_i) < \infty$, we have:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = nm,$$

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\sigma^2 \quad \text{(by independence)}.$$

Furthermore,

$$\mathrm{Var}\left(\frac{S_n}{n}\right) = \mathrm{Var}\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}\mathrm{Var}(S_n) = \frac{\sigma^2}{n}.$$

Thus,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geq c\right) \leq \frac{\sigma^2}{nc^2} \to 0 \quad \text{as} \quad n \to \infty.$$

Since c was arbitrary, we conclude that

$$\frac{S_n}{n} \to m \quad \text{in probability (not necessarily almost surely).}$$

7.2 Conditional Expectation

Example 7.6. Throw a die. All outcomes $1, 2, \ldots, 6$ are equally likely. Write $X(\omega) = \omega$ for the random variable that gives the outcome (number of eyes on the die).
We have $\mathbb{P}(X = i) = \frac{1}{6} =: p_i$.

$$\mathbb{P}(X(\omega) \leq 3) = \frac{3}{6} = \frac{1}{2}$$

Suppose we additionally know whether the outcome is even or odd. The conditional probabilities of the event $\{X \leq 3\}$ are:

$$\mathbb{P}(X \leq 3 \mid X \text{ odd}) = \frac{\mathbb{P}(\{X \leq 3\} \cap \{X \text{ odd}\})}{\mathbb{P}(X \text{ odd})} = \frac{2/6}{3/2} = \frac{2}{3},$$

$$\mathbb{P}(X \leq 3 \mid X \text{ even}) = \frac{\mathbb{P}(\{X \leq 3\} \cap \{X \text{ even}\})}{\mathbb{P}(X \text{ even})} = \frac{1/6}{2/3} = 0.$$

We can also compute conditional expectations:

$$\mathbb{E}(X \mid X \text{ even}) = \frac{2+4+6}{3} = 4 \qquad \mathbb{E}(X \mid X \text{ odd}) = \frac{1+3+5}{3} = 3.$$

In general, we define conditional expectations with respect to σ-algebras.

Theorem 7.7. *(Conditional Expectation) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra. Let X be an integrable random variable. Then, there exists a random variable $Y : \Omega \to \mathbb{R}$, $Y = Y(\omega)$, with the following properties:*

1. *Y is $\mathcal{G}$-measurable.*
2. *Y is integrable.*
3. *For all $G \in \mathcal{G}$:*
$$\int_G Y(\omega)\, d\mathbb{P} = \int_G X(\omega)\, d\mathbb{P}.$$

Y is called the conditional expectation of X given $\mathcal{G}$, written $\mathbb{E}(X \mid \mathcal{G})$.

Proposition 7.8. (Conditional Expectation of X)
Y is almost surely unique. For any Y, Y' satisfying (1)-(3), $\mathbb{P}(Y = Y') = 1$.
This random variable is called the conditional expectation of X with respect to $\mathcal{G}$ and we write $Y(\omega) = \mathbb{E}(X \mid \mathcal{G})(\omega) = (\mathbb{E}(X \mid \mathcal{G}))(\omega)$.

Definition 7.9. If $\mathcal{G}$ is generated by random variables we write $\mathbb{E}(X \mid Z)$ instead of $\mathbb{E}(X \mid \sigma(Z))$. Similarly,

$$\mathbb{E}(X \mid Z_1, Z_2, \dots, Z_n) = \mathbb{E}(X \mid \sigma(Z_1, Z_2, \dots, Z_n)).$$

Example 7.10. In our dice throwing example:

$$\mathcal{F} = \mathcal{P}(\{1, \dots, 6\}), \quad \mathcal{G} = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \{1,2,3,4,5,6\}\} = \sigma(\{X \text{ is even}\}).$$

$\mathcal{G}$-measurability implies Y must be constant on $\{1,3,5\}$ and $\{2,4,6\}$, respectively. Say

$$Y(\omega) = \begin{cases} a & \text{for } \omega \in \{1,3,5\}, \\ b & \text{for } \omega \in \{2,4,6\}. \end{cases}$$

1.
$$\int_{\emptyset} Y(\omega)\, d\mathbb{P} = \int_{\emptyset} X(\omega)\, d\mathbb{P} \quad \text{(trivial, 0=0).}$$

2.
$$\int_{\{1,3,5\}} Y(\omega)\, d\mathbb{P} = \int_{\{1,3,5\}} X(\omega)\, d\mathbb{P} \implies a{\cdot}\mathbb{P}(\{1,3,4\}) = \frac{1}{2}a = \frac{3}{2} = \frac{1+3+5}{6} \implies a = 3.$$

with probability $\frac{1}{2}$.
3.
$$\int_{\{2,4,6\}} Y(\omega)\, d\mathbb{P} = \int_{\{2,4,6\}} X(\omega)\, d\mathbb{P} \implies b{\cdot}\mathbb{P}(\{2,4,6\}) = \frac{1}{2}b = \frac{3}{2} = \frac{2+4+6}{6} \implies b = 4.$$

with probability $\frac{1}{2}$.
4.
$$\int_{\Omega} Y(\omega)\, d\mathbb{P} = \int_{\Omega} X(\omega)\, d\mathbb{P}.$$

Sum of (2) and (3), so also satisfied.

Remark 7.11. The "ordinary" expectation $\mathbb{E}(X \mid \mathcal{G})$ is the special case $\mathcal{G} = \{\emptyset, \Omega\}$. Then Y is constant on Ω and:

$$\int_\Omega X \, d\mathbb{P} = \int_\Omega Y \, d\mathbb{P} = Y\mathbb{P}(\Omega) \quad \text{(trivial)}.$$

$$\int_\Omega X \, d\mathbb{P} = \int_\Omega Y \, d\mathbb{P} \implies Y(\omega) = \int_\Omega X \, d\mathbb{P} = \mathbb{E}(X).$$

We may interpret σ-algebras as "knowledge" of an event and will investigate sequences $X_1, X_2, \ldots$ of integrable random variables.

7.3 Martingales

Definition 7.12. (Martingale) Let $X_1, X_2, \ldots$ be integrable random variables. The sequence is a martingale if:

$$\mathbb{E}(X_{n+1} \mid \sigma(X_1, X_2, \ldots, X_n)) = X_n \quad \text{almost surely.}$$

Informally, the expectation of the $(n+1)$-th random variable conditioned on "knowing" outcomes $X_1, \ldots, X_n$ is equal to the last observed value.

Proposition 7.13. (Properties of Conditional Expectation) We have:

1. $$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X).$$
2. If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X \mid \mathcal{G}) = X$ a.s.
3. (Linearity): $$\mathbb{E}(aX + bY \mid \mathcal{G}) = a\mathbb{E}(X \mid \mathcal{G}) + b\mathbb{E}(Y \mid \mathcal{G}) \quad \text{a.s.}$$
4. (Positivity): If $X \geq 0$ a.s. then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$ a.s.

Proof. 1. Since $\Omega \in \mathcal{G}$, we have

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \int_\Omega \mathbb{E}(X \mid \mathcal{G}) \, d\mathbb{P} = \int_\Omega X \, d\mathbb{P} = \mathbb{E}(X).$$

2. X satisfies all conditions in definition/theorem.

3. Note that $a\mathbb{E}(X \mid \mathcal{G}) + b\mathbb{E}(Y \mid \mathcal{G})$ is $\mathcal{G}$-measurable. The assertion then follows by linearity of integration.

4. Suppose $Y = \mathbb{E}(X \mid \mathcal{G})$ is negative with positive probability: $\mathbb{P}(Y < 0) > 0$. Then $\exists n \in \mathbb{N}$ such that $\mathbb{P}(Y \le -\frac{1}{n}) > 0$. This is hence a $\mathcal{G}$-measurable set and

$$\int_{\{Y \le -\frac{1}{n}\}} Y \, d\mathbb{P} = \int_{\{Y \le -\frac{1}{n}\}} X \, d\mathbb{P} \le -\frac{1}{n}\mathbb{P}\left(Y \le -\frac{1}{n}\right) \le 0.$$

A contradiction. $\square$

Proposition 7.14. (Further Properties) Results on convergence carry over "na-neg." Then,

- If $X_1, X_2, \ldots$ is a sequence of random variables such that $X_n \uparrow X$, then we also have
$$\mathbb{E}(X_n \mid \mathcal{G}) \uparrow \mathbb{E}(X \mid \mathcal{G}) \quad \text{(MCT)}.$$
- If $X_1, X_2, \ldots$ is a sequence of random variables such that $\mid X_n \mid \le Y$ for some integrable Y, and $X_n \to X$, then also
$$\mathbb{E}(X_n \mid \mathcal{G}) \to \mathbb{E}(X \mid \mathcal{G}) \quad \text{(DCT)}.$$
- If $X_1, X_2, \ldots$ is any sequence of non-negative random variables, then
$$\mathbb{E}\left(\liminf_{n\to\infty} X_n \mid \mathcal{G}\right) \le \liminf_{n\to\infty} \mathbb{E}(X_n \mid \mathcal{G}) \quad \text{(Fatou)}.$$

We also get a corresponding analogue of Jensen's inequality:

Theorem 7.15. *Let $g : I \to \mathbb{R}$ be a convex function on an interval $I \subseteq \mathbb{R}$. Assume $X : \Omega \to I$ and X and $g(X)$ are integrable. Then,*

$$\mathbb{E}(g(X) \mid \mathcal{G}) \ge g(\mathbb{E}(X \mid \mathcal{G})) \quad a.s.$$

Remark 7.16. Simplification Rules:

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H}) \quad \text{for sub-}\sigma\text{-algebras } \mathcal{G}, \mathcal{H} \text{ with } \mathcal{H} \subseteq \mathcal{G}. \tag{1}$$

Remark 7.17. Simplification Rules:

1. $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H})$ for sub σ-algebras $\mathcal{G}, \mathcal{H}$ with $\mathcal{H} \subseteq \mathcal{G}$.
2. $\mathbb{E}(Z \cdot X \mid \mathcal{G}) = Z \cdot \mathbb{E}(X \mid \mathcal{G})$ if Z is $\mathcal{G}$-measurable (completely determined by $\mathcal{G}$).
3. $\mathbb{E}(X \mid \sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X \mid \mathcal{G})$ if $\mathcal{H}$ is independent of $\sigma(X, \mathcal{G})$.

   Special case:

   - If X is independent of $\mathcal{H}$, then $\mathbb{E}(X \mid \mathcal{H}) = \mathbb{E}(X)$.
   - If X is $\mathcal{H}$-measurable, then $\mathbb{E}(X \mid \mathcal{H}) = X$.
   - $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X)$

These can be proved by verifying conditions of the conditional expectation.

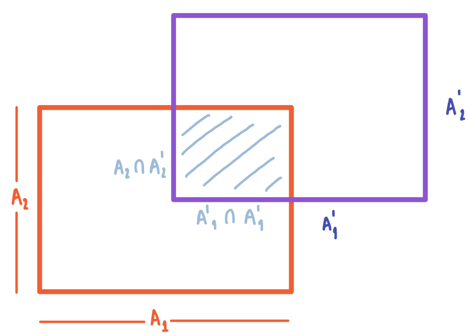
8 Product Spaces & Measures

Given two measure spaces (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2), we want to build a "canonical" measure on $S = S_1 \times S_2$.

Definition 8.1. (**Product σ-algebra**) Given two measure spaces (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2), define their product as a measure space on $S = S_1 \times S_2$:

$$\Sigma = \Sigma_1 \times \Sigma_2 = \sigma\left(\bigcup_{A \in \Sigma_1} (A \times S_2) \cup \bigcup_{B \in \Sigma_2} (S_1 \times B)\right)$$

Remark 8.2. $\{A_1 \times A_2 : A_1 \in \Sigma_1, A_2 \in \Sigma_2\}$ is a π-system that generates Σ.



Proposition 8.3. (**Measurable Projections**) If f is a bounded measurable function on (S, Σ), then the projections:

- $\pi_1 : S_1 \times S_2 \to \mathbb{R}$, $(s_1, s_2) \mapsto f(s_1, s_2)$ (for fixed s_2)
- $\pi_2 : S_1 \times S_2 \to \mathbb{R}$, $(s_1, s_2) \mapsto f(s_1, s_2)$ (for fixed s_1)

are measurable $\forall s_1, s_2$ respectively.

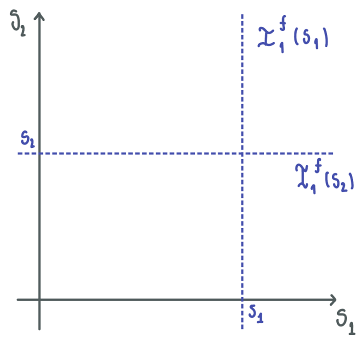
Proof. This holds for indicator functions of the form:

$$I_{A_1 \times A_2}(s_1, s_2) = \begin{cases} 1 & s_1 \in A_1, s_2 \in A_2 \\ 0 & \text{otherwise} \end{cases}$$

For arbitrary f, use approximation by step functions. $\square$

Definition 8.4. (**Product Measure**) Assume μ_1, μ_2 are finite measures on $(S_1, \Sigma_1), (S_2, \Sigma_2)$. Define two functions:

$$\mathcal{I}_1^f(s_1) = \int_{S_2} f(s_1, s_2)\, d\mu_2, \quad \mathcal{I}_2^f(s_2) = \int_{S_1} f(s_1, s_2)\, d\mu_1$$



Lemma 8.5. For bounded measurable f, $\mathcal{I}_1^f$ and $\mathcal{I}_2^f$ are bounded measurable.

Proof. For indicators $I = I_{A_1 \times A_2}$:

$$\mathcal{I}_1^f(s_1) = \int_{S_2} I_{A_1 \times A_2}(s_1, s_2)\, d\mu_2 = \int_{S_2} I_{A_1}(s_1) I_{A_2}(s_2)\, d\mu_2 = I_{A_1}(s_1) \int d\mu_2 = I_{A_1}(s_1)\mu_2(A_2)$$

(analogous for $\mathcal{I}_2^f(s_2)$). For arbitrary f_1, approximate by step functions.

Now for $F \in \Sigma$, take $f = I_F(s_1, s_2)$ and define:

$$\mu(F) = \int_{S_1} \mathcal{I}_1^f\, d\mu_1 = \int_{S_1} \left(\int_{S_2} f(s_1, s_2)\, d\mu_2 \right) d\mu_1$$

$$=^{(*)} \int_{S_2} \mathcal{I}_2^f\, d\mu_2 = \int_{S_2} \left(\int_{S_1} f(s_1, s_2)\, d\mu_1 \right) d\mu_2$$

$\square$

Theorem 8.6. ***(Fubini's Theorem)*** *This is well-defined, i.e. (*) holds. In fact:*

$$\int_{S_1}\int_{S_2} f\,d\mu_2\,d\mu_1 = \int_{S_2}\int_{S_1} f\,d\mu_1\,d\mu_2 = \int_{S_1\times S_2=S} f\,d\mu$$

for all non-negative or integrable f. Here, μ is the unique measure that satisfies:

$$\mu(A_1\times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \forall A_1\in\Sigma_1, A_2\in\Sigma_2.$$

Proof. When $f = I_{A_1\times A_2}$:

$$\int_{S_1}\int_{S_2} I_{A_1\times A_2}(s_1,s_2)\,d\mu_2\,d\mu_1 = \int_{S_1}\int_{S_2} I_{A_1}(s_1)I_{A_2}(s_2)\,d\mu_2\,d\mu_1 =$$

$$= \int_{S_1} I_{A_1}(s_1)\,d\mu_1 \int_{S_2} I_{A_2}(s_2)\,d\mu_2 = \int_{S_2}\int_{S_1} f\,d\mu_1\,d\mu_2 = \mu_1(A_1)\cdot\mu_2(A_2)$$

For general f, approximate by step functions.

Generally, $\mu(A_1\times A_2) = \mu_1(A_1)\cdot\mu_2(A_2)$. In fact, uniqueness follows since $\{A_1\times A_2 : A_1\in\Sigma_1, A_2\in\Sigma_2\}$ is a π-system, and so μ is determined uniquely by the values of $\mu(A_1\times A_2)$.

This construction defines the **product measure** μ, also written as $\mu = \mu_1\times\mu_2$. □

Remark 8.7. Extend this to products of several measure spaces/measures:

$$\mu = \mu_1\times\mu_2\times\cdots\times\mu_n$$

but in fact this also holds for countably infinite products $\mu = \mu_1\times\mu_2\times\dots$.

Example 8.8. The Lebesgue measure $\mathcal{L}^n$ on $\mathbb{R}^n$ is the same as $\mathcal{L}\times\mathcal{L}\times\cdots\times\mathcal{L}$ (n times).

Remark 8.9. Fubini's Theorem remains true for σ-finite measures, but does not necessarily hold otherwise.

Example 8.10.

- $(S_1, \mu_1) = [0,1]$, μ_1 = Lebesgue measure on $[0,1]$ (σ-finite)
- $(S_2, \mu_2) = [0,1]$, μ_2 = Counting measure on $[0,1]$ (not σ-finite)

Let

$$f(s_1, s_2) = \begin{cases} 1 & \text{if } s_1 = s_2, \\ 0 & \text{otherwise.} \end{cases}$$

Then:

$$\int_{S_1} \int_{S_2} f \, d\mu_2 \, d\mu_1 = \int_{S_1} 1 \, d\mu_1 = 1,$$

while

$$\int_{S_2} \int_{S_1} f \, d\mu_1 \, d\mu_2 = \int_{S_2} 0 \, d\mu_2 = 0.$$

This shows that the order of integration matters when the measure is not σ-finite.

An Application: Formula for $\mathbb{E}(X)$

Suppose X is a non-negative random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then:

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq x) \, dx.$$

We can prove this by expanding the integral:

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq x) \, dx = \int_0^\infty \int_\Omega I(X \geq x) \, d\mathbb{P} \, dx = \int_\Omega \int_0^\infty I(X \geq x) \, dx \, d\mathbb{P}$$

$$= \int_\Omega \int_0^{\chi(\omega)} 1 \, dx \, d\mathbb{P} = \int_\Omega \chi(\omega) \, d\mathbb{P} = \mathbb{E}(X).$$

Proof. **Proof (that $\mathbb{E}(X \mid \mathcal{G})$ is a.s. unique):**

Assume that there are two random variables Y, Y' that satisfy the conditions and that $\mathbb{P}(Y = Y') \neq 1$. Then $\mathbb{P}(Y > Y') > 0$ or $\mathbb{P}(Y < Y') > 0$. Assume, without loss of generality, that it's the former.

Note:

$$\{Y > Y'\} = \bigcup_{n \in \mathbb{N}} \{Y > Y' + \frac{1}{n}\}$$

and for some n, we have $\mathbb{P}(Y > Y' + \frac{1}{n}) > 0$.

Since Y, Y' are $\mathcal{G}$-measurable, this implies $Y - Y'$ is $\mathcal{G}$ measurable.

$$\implies \{Y - Y' > \frac{1}{n}\} \in \mathcal{G}.$$

By condition (3),

$$\int_{\{Y-Y'>\frac{1}{n}\}} Y\,d\mathbb{P} = \int_{\{Y-Y'>\frac{1}{n}\}} X\,d\mathbb{P} = \int_{\{Y-Y'>\frac{1}{n}\}} Y'\,d\mathbb{P},$$

which implies

$$\int_{\{Y-Y'>\frac{1}{n}\}} Y - Y'\,d\mathbb{P} = 0.$$

Thus, we have

$$\frac{1}{n}\mathbb{P}(Y - Y' > \frac{1}{n}) \le 0,$$

which is a contradiction because $\mathbb{P}(Y - Y' > \frac{1}{n}) > 0$. $\square$

8.1 Conditional Density and Expectation

We consider a special case where X, Y have a common density $f_{X,Y}(x, y)$ to find $\mathbb{E}(X \mid Y)$:

$$\mathbb{P}((X,Y) \in A) = \iint\limits_A f_{X,Y}(x,y)\,dx\,dy.$$

Definition 8.11. We define the conditional density:

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx$. Note that for fixed y,

$$\int_{\mathbb{R}} f_{X|Y}(x \mid y)\,dx = \int_{\mathbb{R}} \frac{f_{X,Y}(x,y)}{f_Y(y)}\,dx = \frac{1}{f_Y(y)} \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx = 1.$$

So $f_{X|Y}(x \mid y)$ is a density, provided that $f_Y(y) \neq 0$ ("the density of X given $Y = y$").

Now, set

$$g(y) = \int_X x f_{X|Y}(x \mid y)\,dx$$

("the expected value of X given $Y = y$"). We want to show that $g(y)$ satisfies the conditions of conditional expectations.

By inspection, g is $\sigma(Y)$-measurable and integrable. It remains to check condition (3). Let $A \in \Sigma_2$, then

$$\int_{\{Y\in A\}} X\, d\mathbb{P} = \int_{\{Y\in A\}} g(Y)\, d\mathbb{P}.$$

We want to show that

$$\int_{\{Y\in A\}} X\, d\mathbb{P} =$$

$$= \int_{\Omega} X\mathbb{I}_{\{Y\in A\}}\, d\mathbb{P} =$$

$$= \iint_{\mathbb{R}\times\mathbb{R}} x\mathbb{I}_{\{Y\in A\}} f_{X,Y}(x,y)\, dx\, dy =$$

$$= \int_{\mathbb{R}} \int_{A} f_{X|Y}(x \mid y)\, dy\, dx. \qquad (*)$$

On the other hand,

$$\int_{\Omega} g(Y)\mathbb{I}_{\{Y\in A\}}\, d\mathbb{P} =$$

$$= \iint_{\mathbb{R}\times\mathbb{R}} g(y)\mathbb{I}_{\{Y\in A\}} f_{X,Y}(x,y)\, dx\, dy =$$

$$= \int_{\mathbb{R}} \int_{A} g(y) f_{X,Y}(x,y)\, dy\, dx. \qquad (+)$$

By definition, $f_{X,Y}(x,y) = f_{X|Y}(x \mid y) f_Y(y)$ for Lebesgue almost every value of x. Then,

$$(*) = \int_{\mathbb{R}} \int_{A} x f_{X|Y}(x \mid y) f_Y(y)\, dy\, dx = \int_{A} f_Y(y) \int_{\mathbb{R}} x f_{X|Y}(x \mid y)\, dx\, dy =$$

$$= \int_{A} f_Y(y) g(y)\, dy = \int_{A} \int_{\mathbb{R}} f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{\mathbb{R}} \int_{A} g(y) f_{X,Y}(x,y)\, dy\, dx = (+),$$

which concludes the proof that $g(Y)$ satisfies the conditions of conditional expectations for arbitrary A.

Example 8.12. (Conditional Density)
Consider random variables X, Y on $\mathbb{R}^2$ with joint density $f_{X,Y}(x, y) = 2\mathbb{I}_{x\leq y\leq 1}$.
We get

$$f_Y(y) = \int_0^1 2\mathbb{I}_{x\leq y\leq 1}\, dx = 2\int_0^y dx = 2(1-y),$$

and

$$f_{X|Y}(x \mid y) = \frac{2\mathbb{I}_{x\leq y\leq 1}}{2(1-y)} = \frac{\mathbb{I}_{x\leq y\leq 1}}{1-y}.$$

Now, calculate:

$$g(y) = \int_0^1 x\frac{\mathbb{I}_{x\leq y\leq 1}}{1-y}\, dx = \frac{1}{1-y}\int_0^y x\, dx = \frac{1}{1-y}\frac{y^2}{2} = \frac{y^2}{2(1-y)}.$$

Thus,

$$\mathbb{E}(X \mid Y = y) = \frac{y}{2}.$$

Example 8.13. (IID Random Variables)
Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables. Let $S_n = X_1 + \cdots + X_n$. What is $\mathbb{E}(X_1 \mid S_n)$?
By symmetry, $\mathbb{E}(X_1 \mid S_n) = \mathbb{E}(X_2 \mid S_n) = \cdots = \mathbb{E}(X_n \mid S_n)$ a.s.
We must have

$$\int_A \mathbb{E}(X_1 \mid S_n)\, d\mathbb{P} + \cdots + \int_A \mathbb{E}(X_n \mid S_n)\, d\mathbb{P} = \int_A S_n\, d\mathbb{P}$$

for all $A \in \sigma(S_n)$.
Hence,

$$n\int_A \mathbb{E}(X_1 \mid S_n)\, d\mathbb{P} = \int_A S_n\, d\mathbb{P},$$

and

$$\mathbb{E}(X_1 \mid S_n) = \frac{S_n}{n} \text{ a.s.}.$$

Definition 8.14. We define conditional probabilities through conditional expectations:

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbb{E}(\mathbb{I}_A \mid \mathcal{G}) \quad \text{(Now a random variable!)}$$

Proposition 8.15. This is uniquely determined (a.s.) and satisfies:

$$\mathbb{P}(A \cup A' \mid \mathcal{G}) = \mathbb{P}(A \mid \mathcal{G}) + \mathbb{P}(A' \mid \mathcal{G}) \quad \text{a.s.}$$

if $A, A', \ldots$ are disjoint.

Proposition 8.16. If $\mathcal{G} = \sigma(B)$ is generated by an event B, then $\mathbb{P}(A \mid \mathcal{G})$ is a random variable Y with:

$$Y(\omega) = \begin{cases} a & \text{if } \omega \in B, \\ b & \text{if } \omega \in B^C, \end{cases}$$

where Y is $\mathcal{G}$-measurable.
We get:

$$a\mathbb{P}(B) = \int_B Y \, d\mathbb{P} = \int_B \mathbb{I}_A \, d\mathbb{P} = \mathbb{P}(A \cap B) \qquad \Longrightarrow \qquad a = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A \mid B)$$

and similarly,

$$b = \frac{\mathbb{P}(A \cap B^C)}{\mathbb{P}(B^C)} = \mathbb{P}(A \mid B^C).$$

8.2 Independent Random Variables

Proposition 8.17. If $X_1, X_2, \ldots, X_n$ are independent random variables, then:

$$\mathbb{E}(h(X_1, \ldots, X_n) \mid X_1) = g(X_1),$$

where $g(x) = \mathbb{E}(h(x, X_2, \ldots, X_n))$. This statement signifies that the conditional expectation of a function of independent random variables, given one of those variables, is simply the expectation of the function with the given variable fixed.

Proof. This result can be proven using Fubini's Theorem. Since $\sigma(X_1)$ (the sigma-algebra generated by X_1) is generated by sets of the form $\{X_1 \in A\}$, where A is a Borel set in $\mathbb{R}$, to prove the statement, it suffices to show that:

$$\int_{\{X_1 \in A\}} h(X_1, \ldots, X_n) \, d\mathbb{P} = \int_{\{X_1 \in A\}} g(X_1) \, d\mathbb{P}.$$

Let's break down the proof step by step:

1. **Expressing the Integrals in Terms of Densities:**
   - The left-hand side integral can be expressed using the joint density function $f(x_1, \ldots, x_n)$ of the random variables $X_1, \ldots, X_n$ and the Lebesgue measure Λ on $\mathbb{R}$:

$$\int_{\{X_1 \in A\}} h(X_1, \ldots, X_n)\, d\mathbb{P} = \int \int \cdots \int h(x_1, x_2, \ldots, x_n)\, d\mathbb{P}.$$

   - The right-hand side integral can be expressed similarly using the density function of X_1:

$$\int_{\{X_1 \in A\}} g(X_1)\, d\mathbb{P} = \int_{\{X_1 \in A\}} g(x_1)\, d\Lambda_1.$$

2. **Applying Independence and Fubini's Theorem:**
   - Using the independence of $X_1, \ldots, X_n$, we can rewrite the left-hand side integral as:

$$\int_{\{X_1 \in A\}} \int_{\mathbb{R}^{n-1}} f(h(x_1, x_2, \ldots, x_n))\, d(\Lambda_2 \times \cdots \times \Lambda_n)\, d\Lambda_1.$$

   - Now, applying Fubini's Theorem, we can change the order of integration:

$$\int_{\mathbb{R}} \mathbb{I}_{\{X_1 \in A\}} h(x_1, \ldots, x_n)\, d(\Lambda_2 \times \cdots \times \Lambda_n)\, d\Lambda_1 =$$

$$= \int_{\mathbb{R}} \mathbb{I}_{\{X_1 \in A\}} \left[\int_{\mathbb{R}^{n-1}} f(h(x_1, \ldots, x_n))\, d(\Lambda_2 \times \cdots \times \Lambda_n) \right] d\Lambda_1.$$

3. **Relating to $g(x)$:**
   - Notice that the inner integral is precisely the definition of $g(x_1)$:

$$\int_{\mathbb{R}^{n-1}} f(h(x_1, \ldots, x_n))\, d(\Lambda_2 \times \cdots \times \Lambda_n) = g(x_1).$$

4. **Conclusion:**
   - Substituting this back into the integral, we obtain:

$$\int_{\mathbb{R}} \mathbb{I}_{\{X_1 \in A\}} \left[\int_{\mathbb{R}^{n-1}} f(h(x_1, \ldots, x_n))\, d(\Lambda_2 \times \cdots \times \Lambda_n) \right] d\Lambda_1 = \int_{\{X_1 \in A\}} g(x_1)\, d\Lambda_1,$$

     which is exactly the right-hand side integral we started with, confirming that $g(X_1)$ satisfies the conditions for being the conditional expectation $\mathbb{E}(h(X_1, \ldots, X_n) \mid X_1)$.

$\square$

Example 8.18. (Conditional Expectation of the Average)
Let's illustrate this concept with an example. Consider independent random variables $X_1, \ldots, X_n$ and define X as their average:

$$X = \frac{X_1 + \cdots + X_n}{n}.$$

What is $\mathbb{E}(X_1 \mid X)$?

1. **Identify the Function:** In this case, we have $h(x_1, \ldots, x_n) = \frac{x_1 + \cdots + x_n}{n}$.
2. **Calculate $g(x)$:** We need to compute:

$$g(x_1) = \mathbb{E}\left(\frac{x_1 + X_2 + \cdots + X_n}{n}\right).$$

   Using the linearity of expectation and the fact that $X_2, \ldots, X_n$ are independent of X_1, we get:

$$g(x_1) = \frac{x_1 + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)}{n}.$$

3. **The Result:** Therefore, the conditional expectation of X_1 given X is:

$$\mathbb{E}(X_1 \mid X) = \frac{X_1 + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)}{n}.$$

This example demonstrates how the concept of independent random variables simplifies the calculation of conditional expectations. The result intuitively means that given the average of independent random variables, the expected value of one of the variables is adjusted based on the expected values of the remaining variables.

9 Martingales

9.1 Stochastic Processes & Filtrations

Definition 9.1. A (discrete) **stochastic process** is a sequence $X_0, X_1, X_2, \dots$ of random variables.

Definition 9.2. A **filtration** is a sequence of σ-algebras

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \cdots \subseteq \mathcal{F}.$$

We write $\mathcal{F}_n = \sigma(X_0, \dots, X_n) \subseteq \mathcal{F}$.

Definition 9.3. The process $X_0, X_1, \dots$ is said to be **adapted** to the filtration $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$-measurable. This means that at time n, the value of X_n is known based on the information contained in $\mathcal{F}_n$.

Definition 9.4. A martingale with respect to $(\mathcal{F}_n)$ is a stochastic process (X_n) adapted to $(\mathcal{F}_n)$ such that

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = X_{n-1}.$$

Equivalently expressed in increments,

$$\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_N \mid \mathcal{F}_{n-1}) - X_{n-1} = 0.$$

Definition 9.5. A sub-martingale with respect to $(\mathcal{F}_n)$ is a stochastic process (X_n) adapted to $(\mathcal{F}_n)$ such that

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \geq X_{n-1}.$$

Equivalently expressed in increments,

$$\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_N \mid \mathcal{F}_{n-1}) - X_{n-1} \geq 0.$$

Definition 9.6. A **super-martingale** with respect to $(\mathcal{F}_n)$ is a stochastic process (X_n) adapted to $(\mathcal{F}_n)$ such that

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \leq X_{n-1}.$$

This condition implies that, given the information up to time $n-1$, the expected value of the process at time n is less than or equal to its value at time $n-1$.
Equivalently, in terms of increments:

$$\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1} \leq 0.$$

Example 9.7. (Standard Random Walk)
Let $Y_1, Y_2, \ldots$ be independent random variables with

$$\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = \frac{1}{2},$$

and set $X_n = Y_1 + Y_2 + \cdots + Y_n$ with $X_0 = 0$.
With $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$, the sequence $X_0, X_1, \ldots$ is adapted, and

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_{n-1} + Y_n \mid \mathcal{F}_{n-1}) = X_{n-1} + \mathbb{E}(Y_n) = X_{n-1}.$$

This also works for any other Y_i with $\mathbb{E}(Y_i) = 0$. In this case, the standard random walk is a martingale, meaning that the expected value of the process at any time, given the past information, is equal to its current value.

Example 9.8. Let $Y_1, Y_2, \ldots$ be independent random variables with $\mathbb{E}(Y_i) = 1$. Set

$$X_n = X_0 \prod_{i=1}^{n} Y_i.$$

This is a martingale since

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_{n-1} \cdot Y_n \mid \mathcal{F}_{n-1}) = X_{n-1}\mathbb{E}(Y_n) = X_{n-1}.$$

This example illustrates a martingale, where the process is formed by taking the product of independent random variables with an expected value of 1.

Example 9.9. Let X be a fixed random variable. Fix a filtration $(\mathcal{F}_n)$ and set

$$X_n = \mathbb{E}(X \mid \mathcal{F}_n).$$

This is a martingale:

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_n) \mid \mathcal{F}_{n-1}) = \mathbb{E}(X \mid \mathcal{F}_{n-1}) = X_{n-1}.$$

In this example, the martingale is constructed by taking the conditional expectation of a fixed random variable with respect to a given filtration.

Remark 9.10. Let $m < n$. For every martingale, we have $\mathbb{E}(X_n \mid \mathcal{F}_m) = X_m$:

$$\mathbb{E}(X_n \mid \mathcal{F}_m) = \mathbb{E}(\mathbb{E}(\ldots(\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \mid \mathcal{F}_{n-2}) \cdots \mid \mathcal{F}_m)) = X_m.$$

This result generalizes the martingale property to any two time points $m < n$, indicating that the expected value of the process at a future time n, given the information up to time m, is equal to its value at time m.

Definition 9.11. A **previsible process** is a sequence $C_1, C_2, \ldots$ of random variables such that C_n is $\mathcal{F}_{n-1}$-measurable for all n.
Intuitively, a previsible process is one where its value at any time n is determined by the information available up to the previous time $n - 1$.

Proposition 9.12. Let C_n be a previsible process. The **martingale transform** of X by C is

$$(C \cdot X)_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1}).$$

In particular, if $C_k = 1$ for all k, then

$$(C \cdot X)_n = \sum_{k=1}^{n} (X_k - X_0).$$

The martingale transform can be interpreted as a betting strategy where C_k represents the bet placed at time $k-1$ based on the information available then, and $X_k - X_{k-1}$ is the change in the process between times $k-1$ and k.

Proposition 9.13. If C is a bounded previsible process with $\mid C_n(\omega) \mid \leq K$ for all n and $\omega \in \Omega$, then:

- $(C \cdot X)_n$ is a martingale, if X_n is.
- If C is also non-negative, then $(C \cdot X)_n$ is a sub/super-martingale whenever X_n is.

This proposition states that under certain conditions on the previsible process C, the martingale transform preserves the martingale property or transforms a martingale into a sub/supermartingale.

Proof.

$$\mathbb{E}((C \cdot X)_n - (C \cdot X)_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(C_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}) = C_n(\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1})$$

$$= \begin{cases} 0 & \text{if } X_n \text{ is a martingale} \\ \geq 0 & \text{if } C_n \geq 0 \text{ and } X_n \text{ is a submartingale} \\ \leq 0 & \text{if } C_n \geq 0 \text{ and } X_n \text{ is a supermartingale} \end{cases}$$

This proof uses the properties of conditional expectations and the previsibility of C to show the desired results. □

10 Martingales and Processes

Remark 10.1. **Usually** we take $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, but **not always**!

Remark 10.2. Let $m < n$. For every martingale, we have:

$$\begin{aligned}\mathbb{E}(X_n \mid \mathcal{F}_m) &= \mathbb{E}(\mathbb{E}(\ldots \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \mid \mathcal{F}_{n-2}) \cdots \mid \mathcal{F}_m) \\ &= \mathbb{E}(\mathbb{E}(\ldots \mathbb{E}(X_{n-1} \mid \mathcal{F}_{n-2}) \cdots \mid \mathcal{F}_m)) \\ &\vdots \\ &= \mathbb{E}(X_{m+1} \mid \mathcal{F}_m) = X_m.\end{aligned}$$

10.1 Previsible Processes

Definition 10.3. A **previsible process** is a sequence $C_1, C_2, \ldots$ of random variables such that C_n is $\mathcal{F}_{n-1}$-measurable for all n.

Definition 10.4. Let C_n be a previsible process. The **martingale transform** of X by C is:

$$(C \cdot X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}).$$

In particular, if $C_k = 1$ for all k, then $(C \cdot X)_n = X_n - X_0$.

Proposition 10.5. If C is a bounded previsible process with $\mid C_n(\omega) \mid \leq K$ for all n and $\omega \in \Omega$, then $(C \cdot X)_n$ is a:

- **martingale** if X_n is.
- **submartingale** if C is also non-negative and X_n is.
- **supermartingale** if C is also non-negative and X_n is.

Proof. We have:

$$\begin{aligned}
\mathbb{E}((C\cdot X)_n - (C\cdot X)_{n-1} \mid \mathcal{F}_{n-1}) &= \mathbb{E}(C_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}) \\
&= C_n\left(\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - \mathbb{E}(X_{n-1} \mid \mathcal{F}_{n-1})\right) \\
&= C_n\left(\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1}\right) \\
&\begin{cases} = 0 & \text{if } X_n \text{ is a martingale,} \\ \geq 0 & \text{if } C_n \geq 0 \text{ and } X_n \text{ is a submartingale,} \\ \leq 0 & \text{if } C_n \geq 0 \text{ and } X_n \text{ is a supermartingale.} \end{cases}
\end{aligned}$$

□

10.2 Stopping Times

Definition 10.6. A **stopping time** is a random variable T with values in $\{0,1,2,\ldots,\infty\}$ and the property that $\{T \leq n\} \in \mathcal{F}_n$ for all n. Equivalently, $\{T = n\} \in \mathcal{F}_n$ for all n. This follows from $\{T \leq n\} = \{T \leq n-1\} \cup \{T = n\}$.

Example 10.7.

- **All constants** are stopping times.
- **"First occurrence":** For example: $T = \min\{n : X_n \in B\}$ for an adapted process X_n.
- If S, T are stopping times, then so are:
  - $\min\{S,T\} = S \wedge T$ (*either stopped*)
  - $\max\{S,T\} = S \vee T$ (*both stopped*).
- **"Counting":** For example, set:
  - N_n = number of indices $k \leq n$ with $X_k = 0$
  - $T = \min\{n : N_n = 10\}$.

Example 10.8. The following are **(generally) not** stopping times:

- $T = \max\{n \leq 5 : N_n = 0\}$ (we cannot determine whether $N_5 = 0$ or not).
- $T = \min\{n : X_n = \sup_{k\leq n} X_k\}$ ($\sup_{k\leq n} X_k$ is not measurable w.r.t. $\mathcal{F}_n$. It could be for larger k!).

11 Stopped Processes and Stopping Times

Definition 11.1. Let X_n be an adapted process and T a stopping time with respect to a given filtration. The **stopped process** X^T is:

$$X_n^T(\omega) = X_{n \wedge T(\omega)}(\omega) = \begin{cases} X_{T(\omega)}(\omega) & \text{if } n \geq T(\omega), \\ X_n(\omega) & \text{if } n < T(\omega). \end{cases}$$

Proposition 11.2. If X_n is a **martingale**/**supermartingale**/**submartingale**, then so is X_n^T. In particular, for every n, $\mathbb{E}(X_n^T) \leq \mathbb{E}(X_0)$ (submartingale).

Proof. Note that:

$$C_n = \mathbb{I}_{\{n \leq T\}} = 1 - \mathbb{I}_{\{T \leq n-1\}}$$

is previsible (*not yet stopped at time* $n-1$).

We have:

$$\begin{aligned}(C \cdot X)_n &= \sum_{k=1}^{n} C_k (X_k - X_{k-1}) \\ &= \sum_{k=1}^{n} \mathbb{I}_{\{k \leq T\}} (X_k - X_{k-1}) = \sum_{k=1}^{T \wedge n} (X_k - X_{k-1}) \\ &= X_{T \wedge n} - X_0 = X_n^T - X_0.\end{aligned}$$

Since $\mathbb{E}((C \cdot X)_n) = \mathbb{E}(X_0)$, the second conclusion follows. □

Remark 11.3. *So for every fixed* n, $\mathbb{E}(X_n^T) = \mathbb{E}(X_0)$. *Is it true that* $\mathbb{E}(X_T) = \mathbb{E}(X_0)$*?* In general, **no!**

Example 11.4. Consider the martingale:

$$X_0 = 1, \quad X_n = \begin{cases} 2X_{n-1} & \text{with prob. } \frac{1}{2}, \\ 0 & \text{with prob. } \frac{1}{2}. \end{cases}$$

Let $T = \min\{n : X_n = 0\}$. Clearly, $\mathbb{E}(X_T) = 0 \neq \mathbb{E}(X_0)$.

Example 11.5. Consider the simple random walk:

$$X_0 = 0, \quad X_n = \begin{cases} X_{n-1} + 1 & \text{with prob. } \frac{1}{2}, \\ X_{n-1} - 1 & \text{with prob. } \frac{1}{2}. \end{cases}$$

Define $T = \min\{n : X_n = 1\}$. This is a stopping time and one can show that $T < \infty$ a.s. Hence, $\mathbb{E}(X_T) = 1 \neq \mathbb{E}(X_0)$.

However, under simple conditions, $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

Theorem 11.6. ***(Doob's Optional Stopping Theorem)***
Let T be a stopping time and let X be a super/sub-martingale. Suppose one of the following holds:

i.) T is bounded (almost surely)
ii.) X_n is bounded and $T < \infty$ for almost every $\omega \in \Omega$
iii.) $\mathbb{E}(T) < \infty$ and $\mid X_n(\omega) - X_{n-1}(\omega) \mid \leq K$ for all n and almost every $\omega \in \Omega$.
Then,

$$\mathbb{E}(X_T) \begin{cases} \leq & & \textit{super-} \\ = & \mathbb{E}(X_0) & \textit{martingale} \\ \geq & & \textit{sub-}. \end{cases}$$

Proof.

i.) If T is bounded by N (a.s.), we have $T \wedge N = T$ (a.s.) and so $\mathbb{E}(X_T) = \mathbb{E}(X_{T\wedge N}) \leq \mathbb{E}(X_0)$.

ii.) We have $\mathbb{E}(X_{T\wedge N}) \leq \mathbb{E}(X_0)$ for fixed n. Since X is bounded, we can use dominated convergence:

$$\mathbb{E}(X_0) \geq \lim_{n\to\infty} \mathbb{E}(X_{T\wedge n}) = \mathbb{E}(\lim_{n\to\infty} X_{T\wedge n}) = \mathbb{E}(X_T).$$

iii.) We have:

$$\mid X_{T\wedge n} - X_0 \mid = \left| \sum_{k=1}^{T\wedge n} (X_k - X_{k-1}) \right| \leq \sum_{k=1}^{T\wedge n} \mid X_k - X_{k-1} \mid \leq K \cdot (T \wedge n) \leq KT.$$

So $\mathbb{E}(KT) = K\mathbb{E}(T) < \infty$ and we can apply DCT above.

$\square$

Corollary 11.7. If X_n is a non-negative supermartingale and T an (almost surely) finite stopping time, then $\mathbb{E}(X_T) \le \mathbb{E}(X_0)$.

Proof. By Fatou's Lemma:

$$\mathbb{E}(X_T) = \mathbb{E}(\lim_{n\to\infty} X_{T\wedge n}) \text{ (why does it exist?) } = \mathbb{E}(\liminf_{n\to\infty} X_{T\wedge n}) \le \liminf_{n\to\infty} \mathbb{E}(X_{T\wedge n}) \le \mathbb{E}(X_0).$$

$\square$

Remark 11.8.

- $X^T = \lim_{n\to\infty} X^{T\wedge n}$, if it exists.
- **Doob's Optional Stopping Theorem**: $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ under nice X_n and T.

The following lemma is useful to show that $\mathbb{E}(T) < \infty$ for specific stopping times.

Lemma 11.9. Suppose there exists $\epsilon > 0$ and a positive integer N such that

$$\mathbb{P}(T \le n + N \mid \mathcal{F}_n) \ge \epsilon \text{ for all } n.$$

Then $\mathbb{E}(T) < \infty$.

Proof. The probability of stopping within the next N steps is at least ϵ.

- $\mathbb{P}(T > N) \le 1 - \epsilon$ (first N steps),
- $\mathbb{P}(T > 2N \mid T > N) \le 1 - \epsilon$ (steps $N + 1, \ldots, 2N$),
- $\mathbb{P}(T > 3N \mid T > 2N) \le 1 - \epsilon$ (steps $2N + 1, \ldots, 3N$),
- ...

So,

$$\begin{aligned}\mathbb{E}(T) &\le N\epsilon + 2N\epsilon(1-\epsilon) + 3N\epsilon(1-\epsilon)^2 + \ldots \\ &= N\epsilon\left(1 + 2(1-\epsilon) + 3(1-\epsilon)^2 + \ldots\right) \\ &= N\epsilon \cdot \frac{1}{\epsilon^2} = \frac{N}{\epsilon} < \infty.\end{aligned}$$

$\square$

Example 11.10. Consider the simple random walk with $X_0 = 0$.

$$X_n = \begin{cases} X_{n-1} + 1 & \text{with probability } \frac{1}{2}, \\ X_{n-1} - 1 & \text{with probability } \frac{1}{2}. \end{cases}$$

Take $T = \min\{n :| X_n |\geq a\}$ for $a \in \mathbb{N}_0$. Then $\mathbb{E}(T) < \infty$. This follows by taking $N = a$, $\epsilon = 2^{-a}$. More generally, we can consider:

$$T = \min\{n : X_n = a \text{ or } X_n = -b\}.$$

For $a, b \in \mathbb{N}_0$. Assume identical with $N = \max\{a, b\}, \epsilon = 2^{-N}$. So $\mathbb{E} < \infty$. Since $| X_{n+1} - X_n |= 1$, the "hitting the second line" of Doob's optional stopping theorem applies.

This allows us to answer questions such as:

- What is the probability that we reach a before $-b$?
- What is the expected time for one of the two to happen?

We get $\mathbb{E}(X_T) = \mathbb{E}(X_0) = 0$ (from **DOST**).

$$\begin{aligned}
&\implies a \cdot \mathbb{P}(X_T = a) + (-b) \cdot \mathbb{P}(X_T = -b) = 0 \quad \text{and} \quad \mathbb{P}(X_T = a) + \mathbb{P}(X_T = -b) = 1 \implies \\
&\implies a \cdot \mathbb{P}(X_T = a) + (-b) \cdot (1 - \mathbb{P}(X_T = a)) = 0 \iff \\
&\iff \mathbb{P}(X_T = a) = \frac{b}{a+b} \iff \\
&\iff \mathbb{P}(X_T = -b) = \frac{a}{a+b}.
\end{aligned}$$

Now look at X_n^2:

$$\begin{aligned}
\mathbb{E}(X_n^2 \mid \mathcal{F}_{n-1}) &= \frac{1}{2}(X_{n-1} + 1)^2 + \frac{1}{2}(X_{n-1} - 1)^2 \\
&= X_{n-1}^2 + 1.
\end{aligned}$$

It follows that:

$$\mathbb{E}(X_n^2 \mid \mathcal{F}_{n-1}) = X_{n-1}^2 + 1 = n - X_{n-1}^2 - (n-1).$$

Hence, $Y_n = X_n^2 - n$ is a **martingale**! X_n isn't.

The second and third item of **DOST** apply, and $\mathbb{E}(Y_T) = \mathbb{E}(Y_0) = 0$.

$$Y_T = X_T^2 - T \text{ is either } a^2 - T \text{ or } b^2 - T \implies \mathbb{E}(X_T^2) = \mathbb{E}(T) \implies$$
$$\implies a^2 \cdot \frac{b}{a+b} + b^2 \cdot \frac{a}{a+b} = \mathbb{E}(T) \implies \mathbb{E}(T) = ab.$$

11.1 Convergence Theorem

Are there conditions under which a martingale converges to a limit $X_\infty(\omega) = \lim n \to \infty X_n(\omega)$***?***
(Limit may still be random!)

Example 11.11.

$$X_0 = 0, \quad X_n = X_{n-1} + \begin{cases} 2^n & \text{with probability } \frac{1}{2}, \\ -2^n & \text{with probability } \frac{1}{2}. \end{cases}$$

We can express X_n as:

$$X_n = \sum_{k=1}^{n} \frac{Y_k}{2^k}, \quad \text{where } Y_k = \pm 1.$$

$X_\infty = \sum_{k=1}^{\infty} \frac{Y_k}{2^k}$ **always exists**, because the sum is absolutely convergent. In fact, X_∞ is **uniformly distributed** on $[-1, 1]$. (Bernoulli convolutions proof omitted.)

We want to establish conditions under which martingales **converge almost surely**.

11.2 Upcrossings

Definition 11.12. Fix $a < b$. An **upcrossing** starts from a value below a and ends with a value above b.

Formally, let X_n be an adapted process, and let $U_N[a, b](\omega)$ be the largest k such that there exist times

$$0 \leq t_1 < s_1 < \cdots < s_k < t_k \leq N$$

with $X_{t_i}(\omega) \leq a$ and $X_{s_i}(\omega) \geq b$ for all i.

Consider the previsible process that is equal to 1 within an upcrossing and 0 otherwise:

$$C_n = \mathbf{1}_{\{X_{n-1} \leq a < b \leq X_n\}}.$$

Then

$$(C \cdot X)_N = \sum_{k=1}^{N} C_k(X_k - X_{k-1}) \geq (b-a)U_N[a,b]$$

Set

$$Y_n = (C \cdot X)_n - (b-a)U_n[a,b]$$

Apply expectation to both sides to get:

Lemma 11.13. (**Doob's Upcrossing Lemma**) X is a supermartingale, then

$$(b-a)\mathbb{E}(U_N[a,b]) \leq \mathbb{E}((X_N - a)^{-1})$$

Proof. This follows since the transform of a supermartingale by a non-negative previsible process is still a supermartingale:

- X_n is a supermartingale.
- C_n is non-negative and previsible.
- Therefore, $(C \cdot X)_n$ is a supermartingale.

So, Y is a supermartingale; thus, $\mathbb{E}(Y_N) \leq \mathbb{E}(Y_0) = 0$. □

Corollary 11.14. If X is a supermartingale with $\sup_n \mathbb{E}(\mid X_n \mid) < \infty$, we have

$$\sup_n ((b-a)\mathbb{E}(U_n[a,b])) \leq \sup_n \mathbb{E}(\mid X_n \mid) + \mid a \mid < \infty$$

where $U_\infty[a,b] = \lim_{N\to\infty} U_N[a,b]$. In particular, $U_\infty[a,b]$ is almost surely finite.

Proof. We have:

$$\begin{aligned}(b-a)\mathbb{E}(U_N[a,b]) &\leq \mathbb{E}((X_N - a)^-) \\ &\leq \mathbb{E}(\mid X_N - a \mid) \leq \mathbb{E}(\mid X_N \mid) + \mid a \mid \\ &\leq \sup_n \mathbb{E}(\mid X_n \mid) + \mid a \mid .\end{aligned}$$

We take $N \to \infty$ and apply the Monotone Convergence Theorem (MCT). □

Theorem 11.15. ***(Doob's Convergence Theorem)*** *Let X_n be a supermartingale with $\sup_n \mathbb{E}(\mid X_n \mid) < \infty$. Then, $X_\infty = \lim_{n\to\infty} X_n$ exists almost surely and is finite.*

Remark 11.16. To make X_∞ well-defined when the limit does not exist, one can define it as $X_\infty = \lim_{n\to\infty} \sup X_n$. The statement above becomes $\lim_{n\to\infty} X_n = X_\infty$ almost surely and $X_\infty < +\infty$ almost surely ...)

12 L^2-Martingales

In the following, consider martingales X_n with finite second moment:

$$\text{Var}(X_n) < \infty \implies \mathbb{E}[X_n^2] < \infty.$$

Define the **inner product**

Definition 12.1.

$$\langle U, V\rangle = \mathbb{E}[UV],$$

and have the following **orthogonality property**.

Proposition 12.2. For $s \le t \le u \le v$ and an L^2-martingale M_n,

$$\langle M_t - M_s, M_v - M_u\rangle = 0.$$

"Increments at different times are orthogonal".

Proof.

$$\mathbb{E}[M_v - M_u \mid \mathcal{F}_k] = \mathbb{E}[M_v \mid \mathcal{F}_k] - \mathbb{E}[M_u \mid \mathcal{F}_k] = M_v - M_u = 0 \quad \text{a.s}$$

$$\implies \mathbb{E}[(M_t - M_s)(M_v - M_u) \mid \mathcal{F}_t)] = Y = (M_t - M_s)\mathbb{E}(M_v - M_u \mid \mathcal{F}_t) = 0 \quad \text{a.s}$$

$$\implies \langle M_t - M_s, M_v - M_u\rangle = \mathbb{E}[(\ldots)(\ldots)] = \mathbb{E}[Y] = \mathbb{E}[0] = 0$$

□

Proposition 12.3. We can write $M_n = M_0 + (M_1 - M_0) + (M_2 - M_1) + \ldots + (M_n - M_{n-1})$ and by Pithagora's Theorem:

$$\mathbb{E}[M_n^2] = \langle M_n, M_n\rangle = \mathbb{E}[M_0^2] + \mathbb{E}[(M_1 - M_0)^2] + \mathbb{E}[(M_2 - M_1)^2] + \ldots + \mathbb{E}[(M_n - M_{n-1})^2]$$

$$\sup_n \mathbb{E}[M_n^2] < \infty \iff \sum_{k=1}^{\infty} \mathbb{E}[(M_k - M_{k-1})^2]$$

Proposition 12.4. Hence, also

$$\mathbb{E}[| M_t |] = \sqrt{\mathbb{E}[M_t^2]} < \infty,$$

$$\mathbb{E}[(| M_n |) \leq \sqrt{\mathbb{E}[M_n^2)}$$

so the convergence theorem applies:

$$M_t \to M_\infty \quad \text{a.s.}$$

Proposition 12.5. It also holds that

$$\mathbb{E}\left[(X_\infty - X_u)^2\right] = midX_\infty - X_u mid_2^2$$

tends to 0. That is,

$$M_u \to M_\infty$$

with respect to the norm
$mid \cdot$
mid_2.

Proposition 12.6. We say

$$X_n \longrightarrow_{L^2} X_\infty$$

Proof. One can verify this as follows:

$$\mathbb{E}\left[(M_{n+r} - M_r)^2\right] = \sum_{k=r+1}^{n+r} \mathbb{E}\left[(M_k - M_{k-1})^2\right]$$

by orthogonality. Now, let $n \to \infty$:

$$\mathbb{E}\left[(M_\infty - M_r)^2\right] = \sum_{k=r+1}^{\infty} \mathbb{E}\left[(M_k - M_{k-1})^2\right] \leq \lim_{n\to\infty} \inf \sum_{k=r+1}^{n} \mathbb{E}\left[(M_k - M_{k-1})^2\right]$$

$$= \sum_{k=r+1}^{\infty} \mathbb{E}\left[(M_k - M_{k-1})^2\right] < \infty$$

by Fatou's Lemma. Now, as $r \to \infty$, it follows that

$$\mathbb{E}((M_\infty - M_r)^2) \to 0.$$

□

12.1 Special Case: M_n as a Sum of Independent Random Variables

Now consider the special case where M_n is a sum of independent random variables $X_1, X_2, \ldots, X_n$.

Proposition 12.7.

- $M_0 = 0$
- $M_n = X_1 + X_2 + \ldots + X_n$
- $\sigma_k^2 = \mathrm{Var}(X_k) < \infty$

If $\mathbb{E}(X_k) = 0$ for all k, then M_n is a martingale.

Theorem 12.8. *If $\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$, then $\sum_{k=1}^{\infty} \frac{X_k}{k} = \lim_{n\to\infty} M_n = M_\infty$ exists and is almost surely finite.*

Proof.

$$\sum_{k=1}^{\infty} \mathbb{E}((M_k - M_{k-1})^2) = \sum_{k=1}^{\infty} \mathbb{E}(X_k^2) = \sum_{k=1}^{\infty} \frac{\sigma_k^2 \,\mathrm{Var}(X_k)}{\mathrm{Var}(X_k) k^2} = \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2}.$$

So convergence follows. (Why? Work out details).

□

Remark 12.9. If the X_k are also uniformly bounded, the converse also holds: If the sum $\sum_k X_k$ converges a.s., then $\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$. (Why? Exercise).

Example 12.10. (Sum of Random Variables)
Let $X_1, X_2, \dots$ be i.i.d. random variables with

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2} \qquad \text{(symmetric random walk)},$$

and consider the random sum

$$\sum_{k=1}^{\infty} \frac{a_k X_k}{k}, \quad \sup_k \mid a_k \mid < \infty.$$

Note that

$$\text{Var}(a_k X_k) = \mathbb{E}((a_k X_k)^2) = a_k^2.$$

So the theorem above shows that the random sum converges (a.s.) if and only if $\sum_k a_k^2 < \infty$.

12.2 Strong Law of Large Numbers for L^2 Random Variables

Combine L^2 martingale results with results from real analysis.

Lemma 12.11. (Cesaro's Lemma)
If b_n is a sequence of non-negative reals with $b_n \uparrow \infty$ as $n \to \infty$ and u_n is a convergent sequence of reals with $u_n \to u_\infty$, then

$$\frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) u_k \to u_\infty.$$

Remark 12.12. Without loss of generality, let $b_0 = 0$ and then

$$\frac{\sum_{k=1}^{n} b_k - b_{k-1}}{b_n} = 1.$$

So the left-hand side on Cesaro's Lemma is a weighted average of u_k.

Lemma 12.13. (Kronecker's Lemma)
If b_n is a sequence of non-negative reals with $b_n \uparrow \infty$ as $n \to \infty$ and x_n is a arbitrary sequence of reals.

$$\sum_{n=1}^{\infty} \frac{x_n}{b_n} \text{ converges } \implies \frac{\sum_{k=1}^{n} X_k}{b_n} \to 0.$$

Remark 12.14. If X_n is an i.i.d. sequence of random variables with mean μ and variance $\sigma^2 = \mathrm{Var}(X_i)$, then $Y_n = X_n - \mu$ satisfies:

- $\mathbb{E}(Y_n) = 0$
- $\mathrm{Var}(Y_n) = \sigma^2$

$$\implies \sum_{n=1}^{\infty} \frac{\mathrm{Var}(Y_n)}{n^2} = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Hence,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{Y_1 + Y_2 + \cdots + Y_n}{n} + \mu \to \mu$$

almost surely.
Why? $\mathrm{Var}\left(\frac{Y_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(Y_n)$ and apply convergence theorem and Kronecker's lemma with $b_n = n$ for any event $\omega \in \Omega$, $\frac{\sum_{k=1}^{n} X_n}{n} \to 0$

Remark 12.15. The Law of Large numbers holds for all X_n such that $\sum \frac{\mathrm{Var}(X_n)}{n^2} < \infty \implies$ if X_n is i.i.d. with Var $< \infty$ the LLN applies.

12.3 Kolmogorov's Truncation Lemma

We will slightly tweak this method with a truncation approach.

Lemma 12.16. (Kolmogorov's Truncation Lemma):
Let (X_n) be a sequence of i.i.d. random variables. Assume X_n is integrable and $\mathbb{E}(X) = \mu$.
Write

$$Y_n = \begin{cases} X_n & \text{if } |X_n| \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

Then, the following hold:

1. $\mathbb{E}(Y_n) \to \mu$ as $n \to \infty$
2. $\mathbb{P}(Y_n = X_n \text{ for all but finitely many } n) = 1$
3. $\sum_{n=1}^{N} \frac{\mathrm{Var}(Y_n)}{n^2} \to \infty$.

Proof. 1. $\mid Y_n \mid \leq |X_n|$ and hence

$$\mathbb{E}(\mid Y_n \mid) \leq \mathbb{E}(|X|) = \mathbb{E}(|X_n|) < \infty$$

and using the DCT, we get the conclusion.

2. $\mathbb{P}(Y_n \neq X_n) = \mathbb{P}(|X| > n)$

$$\sum_{n\in\mathbb{N}} \mathbb{P}(|X_n| > n) = \sum_{n\in\mathbb{N}} \mathbb{P}(|X_n| > n) = \sum_{n\in\mathbb{N}} \mathbb{E}(I_{\{|X|>n\}}) = \mathbb{E}\left[\sum_{n\in\mathbb{N}} I_{\{|X|>n\}}\right] \leq \mathbb{E}(|X_n|) \leq \infty$$

So, using Borel-Cantelli, we get 2.

3. We have $\mathrm{Var}(Y_n) = \mathbb{E}(Y_n^2) - \mathbb{E}(Y_n)^2 \leq \mathbb{E}(Y_n^2)$

$$\sum_{n\in\mathbb{N}} \frac{\mathrm{Var}(Y_n)}{n^2} \leq \sum_{n\in\mathbb{N}} \frac{\mathbb{E}(Y_n^2)}{n^2} = \sum_{n\in\mathbb{N}} \frac{\mathbb{E}\left[\mid X_n^2 \mid - I_{\{|X|\leq n\}}\right]}{n^2} = \mathbb{E}\left[|X_n|^2 \sum_{n\in\mathbb{N}} I_{\{|X|\leq n\}} \frac{1}{n^2}\right]$$

$$= \mathbb{E}\left[|X_n|^2 \sum_{n\geq|X_n|} \frac{1}{n^2}\right] \leq (*)\mathbb{E}\left[|X_n|^2 \frac{2}{\max\{1, |X_n|\}}\right] - \mathbb{E}\left[2|X_n|\right] < \infty$$

$\square$

Where $(*)$ follows from:

$$\frac{1}{n^2}\sigma_n^2 \leq \frac{2}{n(n+1)} \sum_{k=1}^{n} \frac{k^2}{n+1}$$

and

$$\sum_{n\geq k}\frac{1}{n^2}\leq\sum_{n\geq k}\left(\frac{2}{n}-\frac{2}{n+1}\right)=\left(\frac{2}{k}-\frac{2}{k+1}\right)+\left(\frac{2}{k+1}-\frac{2}{k+2}\right)+\ldots=\frac{2}{k}.$$

Kolmogorov's Strong Law of Large Numbers (LLN)

Proposition 12.17. (Kolmogorov's Strong Law of Large Numbers)
Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with $\mathbb{E}(\mid X_i \mid) = \mu$. Then,

$$\frac{1}{n}(X_1+X_2+\ldots+X_n)\to\mu \quad \text{a.s.} \quad \text{as } n\to\infty$$

Proof. Define Y_n as above and note that $\frac{1}{n}(X_1+\ldots+X_n)$ and $\frac{1}{n}(Y_1+\ldots+Y_n)$ have the same limit a.s., as they differ only finitely many times. So, the limits of both are identical (if they exist).

Now,

$$\frac{1}{n}(Y_1+\ldots+Y_n)=\frac{(1-\mathbb{E}(Y_1))Y_1+\ldots+(1-\mathbb{E}(Y_n))Y_n}{n}+\frac{1}{n}(\mathbb{E}(Y_1)+\ldots+\mathbb{E}(Y_n)).$$

The first summoned satisfies criteria from previous convergence theory,

$$\mathbb{E}\left[Y_i-\mathbb{E}\left[Y_i\right]\right]=0 \qquad \sum_i\frac{\operatorname{Var}\left(Y_i-\mathbb{E}\left[Y_i\right]\right)}{i^2}=\sum\frac{\operatorname{Var}(Y_i)}{i^2}$$

which is finite by 3. Hence,

$$\lim_{n\to\infty}\frac{1}{n}\left(Y_1+\ldots+Y_n\right)=\lim_{n\to\infty}\frac{1}{n}\left(\mathbb{E}(Y_1)+\ldots+\mathbb{E}(Y_n)\right)$$

which equal μ by Cesarios's lemma and 1. □

12.4 Doob Decomposition

Remark 12.18. Recall:

$$\mathbb{E}(X_n\mid\mathcal{F}_{n-1})=\begin{cases}\leq X_{n-1} & \text{sub-martingale}\\ = X_{n-1} & \text{martingale}\\ \geq X_{n-1} & \text{super-martingale}\end{cases}$$

Proposition 12.19. Let X_n be an adapted process w.r.t $(\mathcal{F}_n)$. Then we can always find a predictable process A_n and a martingale M_n such that:

- $X_n = X_0 + M_n + A_n$
- $M_0 = A_0 = 0$

This decomposition is unique (up to a null set).

From this, we also get:

$$X_n \text{ is a super/sub-martingale } \iff A_n \text{ is decreasing/increasing a.s.}$$

Proof. Suppose we are given the decomposition:

$$X_n - X_{n-1} = M_n - M_{n-1} + A_n - A_{n-1}$$

This gives

$$\mathbb{E}\left(X_n - X_{n-1} \mid \mathcal{F}_{n-1}\right) = \mathbb{E}\left(M_n - M_{n-1} \mid \mathcal{F}_{n-1}\right) + \mathbb{E}\left(A_n - A_{n-1} \mid \mathcal{F}_{n-1}\right) =$$

$$= \text{ martingale } + \text{ previsible process } = 0 + A_n - A_{n-1}.$$

Then $A_n = \sum_{k=1}^n A_k - A_{k-1} = \sum_{k=1}^n \mathbb{E}\left(X_n - X_{n-1} \mid \mathcal{F}_{n-1}\right)$ is uniquely determined and so is $M_n = X_n - X_0 - A_n$ (a.s.). Conversely, one can check that this choice of M_n, A_n works. $\square$

13 Uniform Integrability

Problem: Given $X_n \to X_\infty$, when can we say that $\mathbb{E}(X_n) \to \mathbb{E}(X_\infty)$?

Example 13.1.

$$X_n = \begin{cases} n^2 & \text{with probability } \frac{1}{n^2} \\ 0 & \text{otherwise} \end{cases}$$

Then $\mathbb{E}(X_n) = 1$. Since $\sum_n \mathbb{P}(X_n \neq 0) = \sum_n \frac{1}{n^2} < \infty$, we have $X_n \to X_\infty = 0$ a.s. But $\mathbb{E}(X_\infty) = 0 \neq 1 = \lim_n \mathbb{E}(X_n)$!

Uniform integrability is a key condition that allows exchange of $\mathbb{E}$ and lim.

Lemma 13.2. (Special case of **Egorov's Theorem**)
Let X be an integrable random variable. For every $\epsilon > 0$, there exists $\delta > 0$ such that for all events E with $\mathbb{P}(E) < \delta$, we have

$$\mathbb{E}(|X|; E) = \mathbb{E}(|X|\mathbb{I}_E) < \epsilon.$$

.

Proof. Suppose this was not the case: For some $\epsilon_0 > 0$, there exists a sequence of events E_n such that $\mathbb{P}(E_n) < 2^{-n}$ but $\mathbb{E}(|X|; I_{E_n}) > \epsilon_0$. Since $\sum_n \mathbb{P}(E_n) < \infty$, the Borel-Cantelli lemma implies that only finitely many E_n occur.

Let $F = \limsup_{n\to\infty} E_n$.

Then $\mathbb{P}(F) = 0$.

Hence $\mathbb{E}(|X|\mathbb{I}_F) = 0$.

But by the reverse Fatou lemma:

$$\limsup_{n\to\infty} \mathbb{E}(|X|\mathbb{I}_{E_n}) \leq \mathbb{E}(|X| \limsup_{n\to\infty} \mathbb{I}_{E_n}) = \mathbb{E}(|X|\mathbb{I}_F) = 0.$$

But the LHS is bounded below by $\epsilon_0 > 0$, a contradiction. □

Proposition 13.3. In particular, there exists $K > 0$ such that

$$\mathbb{E}(|X|; |X| > K) < \epsilon.$$

Proof. This holds because:

- $\mathbb{P}(|X| > K) \leq \frac{\mathbb{E}(|X|)}{K}$ by Markov's inequality.
- So we can take $K > \frac{\mathbb{E}(|X|)}{\epsilon}$.

□

Remark 13.4. K generally depends on ϵ and X.

Definition 13.5. Let $\mathcal{C}$ be a family of random variables. We say $\mathcal{C}$ is **uniformly integrable** if, for every $\epsilon > 0$, there exists $K > 0$ such that

$$\mathbb{E}(|X|; |X| > K) < \epsilon \quad \text{for all } X \in \mathcal{C}.$$

Remark 13.6. K does not depend on X (only on ϵ and $\mathcal{C}$).

Example 13.7.

$$X_n = \begin{cases} n^2 & \text{with probability } \frac{1}{n^2}, \\ 0 & \text{otherwise.} \end{cases}$$

X_n is **not uniformly integrable**: No matter the choice of $K > 0$, for large enough n,

$$\mathbb{E}(|X_n|; |X_n| > K) = n^2 \cdot \frac{1}{n^2} = 1.$$

Uniform integrability and whether

$$\lim_n \mathbb{E}(\lim_n X_n)$$

are closely connected.

Proposition 13.8. (Sufficient Condition for UI) Assume there exists $p > 1$ and $\mathcal{C} > 0$ s.t. $\mathbb{E}(|X|^p \leq C$ for all $X \in \mathcal{C}$. Then $(X)_{X \in \mathcal{C}}$ is uniformly integrable.

Proof. We have, for all $K > 0$,

$$\mathbb{E}(|X|; |X| \geq K) \leq \mathcal{E}\left(|X| \cdot \left(\frac{|X|}{k}\right)^{p-1}; |X| \geq K\right) = \frac{1}{K^{p-1}} \mathbb{E}\left(|X|^p; |X| \geq K\right) \leq$$

$$\leq K^{1-p}\mathbb{E}(|X|)^p \leq CK^{1-p}$$

Hence, choosing $K = \left(\frac{\epsilon}{C}\right)^{\frac{1}{1-p}} = \left(\frac{C}{\epsilon}\right)^{\frac{1}{p-1}}$ suffices. □

Proposition 13.9. If $|X| \leq Y$ for all $X \in \mathcal{C}$ where Y is an integrable random variable, then $\mathcal{C}$ is uniformly integrable.

Theorem 13.10. *Let X be an integrable random variable. The family*

$$\mathcal{C} = \{\mathbb{E}(X|\mathcal{G}) : \mathcal{G} \textit{ is a sub-}\sigma\textit{-algrebra of } \mathcal{F}\}$$

is uniformly integrable.

Proof. For given $\epsilon > 0$ choose δ such that $\mathbb{P}(F) < \delta \implies \mathbb{E}(X; F) < \epsilon$ for all $F \in \mathcal{F}$. Now take $K > \frac{\mathbb{E}(|X|}{\delta}$. For $Y = \mathbb{E}(X|\mathcal{G})$ we get

$$|Y| =| \mathbb{E}(X \mid \mathcal{G}) |\leq \mathbb{E}(|X| \mid \mathcal{G}) \qquad \text{Jensen}$$

and so

$$\mathbb{E}(|Y|) \leq \mathbb{E}(\mathbb{E}(|X| \mid \mathcal{G})) = \mathbb{E}(X)$$

and

$$K\mathbb{P}(|Y| > K) \leq_{\text{Markov}} \mathbb{E}(|y|) \leq \mathbb{E}(|X|) < K\delta$$

and so $\mathbb{P}(|Y| > K) < \delta$. And so we get

$$\mathbb{E}(|Y|; |Y| > K) \leq \mathbb{E}(|X|; |Y| > K) < \epsilon \quad (\text{as } Y = X \text{ with probability } \leq \epsilon)$$

□

Definition 13.11. (Convergence in Probability) A sequence X_n of random variables is said to **converge in probability** to X ($X_n \xrightarrow{\mathbb{P}} X$) if, for all $\epsilon > 0$,

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0 \quad \text{as } n \to \infty.$$

Lemma 13.12.

- If $X_n \xrightarrow{a.s.} X$, then also $X_n \xrightarrow{\mathbb{P}} X$.
- If $X_n \xrightarrow{L^P} X$ (i.e. $\|X_n - X\|_p \to 0$) for some $p > 1$, then also $X_n \xrightarrow{\mathbb{P}} X$.

Proof. For the first part, assume $X_n \xrightarrow{a.s.} X$ and apply the reverse Fatou lemma:

$$\limsup_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}\left(\limsup_{n\to\infty}\{|X_n - X| > \epsilon\}\right)$$

$$= \mathbb{P}(\{|X_n - X| > \epsilon \text{ infinitely often}\}) \leq \mathbb{P}(\{X_n \nrightarrow X\}) = 0 \quad \text{by a.s. convergence.}$$

Thus, $X_n \xrightarrow{\mathbb{P}} X$.

For the second part, we have:

$$\mathbb{E}(|X_n - X|) = \mathbb{E}(|X_n - X|; |X_n - X| \leq \frac{\epsilon}{3}) + \mathbb{E}(|X_n - X|; |X_n - X| > \frac{\epsilon}{3})$$

$$\leq \frac{\epsilon}{3} + \mathbb{E}(|X_n| + |X|; |X_n - X| > \frac{\epsilon}{3})$$

$$\leq \frac{\epsilon}{3} + \mathbb{E}(|X_n|; |X_n - X| > \frac{\epsilon}{3}) + 2K \cdot \mathbb{P}(|X_n - X| > \frac{\epsilon}{3})$$

$$< \frac{\epsilon}{3} + \frac{\epsilon}{3K} \cdot 2K \cdot \frac{\epsilon}{3K} = \epsilon.$$

Since $\epsilon > 0$ was arbitrary, $\mathbb{E}(|X_n - X|) \to 0$ and $X_n \xrightarrow{L^1} X$. $\square$

Theorem 13.13. *Suppose that $X_n \xrightarrow{\mathbb{P}} X_\infty$ and $|X_n| \leq K$ for some $K \geq 0$ for all $n \in \mathbb{N}$. Then we have $\mathbb{E}(|X_n - X|) \to 0$ and thus $X_n \xrightarrow{L^1} X$.*

Proposition 13.14. For any $k \in \mathbb{N}$, we have

$$\mathbb{P}(X_\infty > K + \frac{1}{k}) \leq \mathbb{P}(|X_n - X_\infty > \frac{1}{K}) \to 0$$

so $|X_\infty| \leq K$ a.s.
Let $\epsilon > 0$ and pick n_0 large enough such that $\mathbb{P}(|X_n - X| > \frac{\epsilon}{3K}$ for all $n \geq n_0$. Then

$$\mathbb{E}(|X_n - X|) = \mathbb{E}(|X_n - X|; |X_n - X| \leq \frac{\epsilon}{3}) + \mathbb{E}(|X_n - X|) =$$

$$= \mathbb{E}(|X_n - X|; |X_n - X| > \frac{\epsilon}{3}) \leq \frac{\epsilon}{3} + \mathbb{P}(|X_n - X| > \frac{\epsilon}{3})2K < \frac{\epsilon}{2} + 2K\frac{\epsilon}{3K} = \epsilon$$

Since $\epsilon > 0$ was arbitrary, $\mathbb{E}(|X_n - X|) \to 0$ and $X_n \xrightarrow{L^1} X$

Theorem 13.15. *(Equivalent Conditions for Convergence in L^1)*
Suppose X_n is a sequence of integrable random variables. The following are equivalent:

1. $\mathbb{E}(|X_n - X_\infty|) \to 0 \qquad X_n \xrightarrow{L_1} X_\infty$.
2. $X_n \xrightarrow{\mathbb{P}} X_\infty$ *and* $\{X_n\}$ *is uniformly integrable.*

Proof. Exercise (maybe). ☐

13.1 Uniformly Integrable Martingales

Definition 13.16. Uniform Martingale is a sequence M_n martingale such that $\{M_n\}$ is uniformly integrable.

Let M_n be a uniformly integrable martingale $M_n \to M_\infty$ a.s. by Doob's Convergence theorem. But by uniform integrability, $M_n \xrightarrow{L_1} M\infty$.

For any fixed n, we have $\mathbb{E}(M_r \mid \mathcal{F}_n) = M_n$ a.s. for $r \geq n$. This implies

$$\mathbb{E}(M_r \mid F) = \mathbb{E}(M_n \mid F) \quad \text{ for all } F \in \mathcal{F}_n$$

We get $|\mathbb{E}(M_n \mid F) - \mathbb{E}(M_\infty \mid F)|$

$$= |\mathbb{E}(M_r \mid F) - \mathbb{E}(M_\infty \mid F)| = |\mathbb{E}(M_n - M_\infty \mid F)| \leq \mathbb{E}(|M_r - M_\infty; F) \quad \forall r \geq n$$

which tends to 0 as $r \to \infty$.

So we must have $\mathbb{E}(M_n; F) = \mathbb{E}(M_\infty; F)$ for all $F \in \mathcal{F}$. So

$$M_n = \mathbb{E}(M_\infty \mid \mathcal{F}_n) \quad \text{a.s.}$$

We have shown:

Theorem 13.17. *(Martingale Convergence for Uniformly Integrable Martingales)*
If M_n is a uniformly integrable martingale with respect to filtration $\mathcal{F}_n$, then:

- $M_\infty = \lim_{n\to\infty} M_n$ *exists a.s. and is finite.*
- $M_n = \mathbb{E}(M_\infty | \mathcal{F}_n)$ *a.s. for all* $n \in \mathbb{N}$.

Remark 13.18. This also holds for super/sub-martingales with appropriate inequalities.

Theorem 13.19. *Consider a non-negative sub-martingale Z_n. For every $c > 0$, we have:*

$$c \cdot \mathbb{P}\left(\sup_{k\le n} Z_k \ge c\right) \quad \le \quad \mathbb{E}(Z_n; \sup_{k\le n} Z_k \ge c) \quad \le \quad \mathbb{E}(Z_n).$$

Remark 13.20. This result is similar to Markov's inequality but for all $k \le n$.

Proof. The event $\{\sup_{k\le n} Z_k \ge c\}$ can be decomposed into disjoint events:

- $F_0 = \{Z_0 \ge c\}$,
- $F_1 = \{Z_0 < c\} \cap \{Z_1 \ge c\}$,
- $F_2 = \{Z_0 < c\} \cap \{Z_1 < c\} \cap \{Z_2 \ge c\}, \ldots$

Note that $F_k \in \mathcal{F}_k = \sigma(Z_0, \ldots, Z_k)$.

Then,

$$\mathbb{E}(Z_n; F_k) = \int_{F_k} Z_n \, d\mathbb{P} = \int_{F_k} \mathbb{E}(Z_n|\mathcal{F}_k) \, d\mathbb{P} \ge \int_{F_k} Z_k \, d\mathbb{P} = \mathbb{E}(Z_k; F_k).$$

Since $Z_k \ge c$ on F_k, it follows that

$$\mathbb{E}(Z_n; F_k) \ge \int_{F_k} c \, d\mathbb{P} = c \cdot \mathbb{P}(F_k).$$

Summing gives

$$\sum_{k=0}^{n} \mathbb{E}(Z_n; F_k) \geq c \cdot \sum_{k=0}^{n} \mathbb{P}(F_k) = c \cdot \mathbb{P}\left(\bigcup_{k=0}^{n} F_k\right) = c \cdot \mathbb{P}\left(\sup_{k \leq n} Z_k \geq c\right).$$

The left-hand side gives

$$\sum_{k=0}^{n} \mathbb{E}(Z_n; F_k) = \sum_{k=0}^{n} \mathbb{E}(Z_n \cdot I_{F_k}) = \mathbb{E}\left(Z_n \cdot \sum_{k=0}^{n} I_{F_k}\right) = \mathbb{E}(Z_n \cdot I_{\bigcup_{k=0}^{n} F_k}) = \mathbb{E}(Z_n; \sup_{k \leq n} Z_k \geq c) \leq \mathbb{E}(Z_n).$$

Thus, we conclude that $\mathbb{E}(Z_n) \geq c \cdot \mathbb{P}\left(\sup_{k \leq n} Z_k \geq c\right)$ as required.

□

Remark 13.21. In particular, the theorem holds if $|X_n| \leq K$ for all n (almost surely).

Remark 13.22. If X_n is a non-negative supermartingale, then $\mathbb{E}(|X_1|) = \mathbb{E}(X_1) \geq \mathbb{E}(X_n)$ for all n, and the convergence holds, provided $\mathbb{E}(|X_1|) < \infty$.

Jensen inequality also implies:

Lemma 13.23. If M_n is a martingale and f is a convex function such that $f(M_n)$ is integrable for all n, then $f(M_n)$ is a sub-martingale.

Theorem 13.24. *(Kolmogorov's Inequality))*
Let X_n be a sequence of independent random variables with $\mathbb{E}(X_n) = 0$ and $Var(X_n) = \sigma_n^2 < \infty$. Define $S_n = X_1 + \cdots + X_n$.
Then, for every $c > 0$,

$$c^2 \mathbb{P}\left(\sup_{k \leq n} |S_k| \geq c\right) \leq V_n = Var(S_n) = \sum_{k=1}^{n} \sigma_k^2.$$

Proof. S_n is a martingale, and S_n^2 is a sub-martingale because $x \mapsto x^2$ is convex.

By Doob's sub-martingale inequality, we have:

$$c^2 \mathbb{P}\left(\sup_{k \leq n} |S_k| \geq c\right) = c^2 \mathbb{P}\left(\sup_{k \leq n} S_k^2 \geq c^2\right)$$

$$\leq \mathbb{E}(S_n^2) = \mathrm{Var}(S_n).$$

$\square$

14 Pricing & Arbitrage

Definition 14.1. Stock Market: Assets are modelled on random processes.

Definition 14.2. Derivatives: Assets determined by other assets.

- **Futures/Forward contracts**: Buy/Sell an asset at time T at a fixed price K. If the value at time T is S_T, the win/loss ("payoff") is $S_T - K$ (buyer) / $K - S_T$ (seller).
- **Swaps**: Exchange of future cash flow, e.g., currency swaps.
- **Options**: Right, but not obligation, to buy/sell an asset at a future time T for a price K.
  - **Call option** (buying): payoff $(S_T - K)^+$
  - **Put option** (selling): payoff $(K - S_T)^+$
  - **European option**: Right to buy/sell can only be exercised at time T.
  - **American option**: Right to buy/sell can be exercised at any time up to T.

Proposition 14.3. Options are "safe" (payoff non-negative), so we must have a cost.

Pricing: We want to know the **fair price** for an option. We need some concepts to make this precise, and well-defined.

The key concept is **arbitrage** (risk-free gains).

Example 14.4. Suppose Sweden vs Canada are playing the final of Curling World Championships. Two betting sites offer bests on the winner:

Site A offers:

- Sweden wins: $1.5\times$ bet
- Canada wins: $2\times$ bet

Site B offers:

- Sweden wins: $3\times$ bet
- Canada wins: $1.2\times$ bet

We can use the following betting strategy: bet 3 units on site A for Canada, bet 2 on site B for Sweden. No matter who wins, we get 6 units, having paid only 5. We gain one unit, free off risk.

Definition 14.5. More precisely, one assumes that there is a certain **risk-free rate r** at which money can be invested. A unit becomes $(1+r)^T$ at time T (with compound interest) / e^{rT} at time T.
$r = 0 \iff$ 'no risk-free interest'. **Absence of arbitrage** means that we cannot do better than **r** without risks.

Definition 14.6. Another tool we will use are **hedging portfolios** $\longrightarrow$ combining assets by a 'replicating strategy'.

Example 14.7. Call-Put Parity It relates the prices of (European) call and put options on the same asset with the same parameters K,T.
Note that the difference in payoffs is:

$$(S_T - K)^+ - (K - S_T)^+ = \begin{cases} (S_T - K) & 0 \leq S_T \leq K \\ (K - S_T) & \text{otherwise} \end{cases} = S_T - K.$$

Compare the following strategies:

1. Buy a call option, sell a put option. $\longrightarrow$ payoff at time T is $S_T - K$.
2. Buy the asset at its current price S_0 and borrow $e^{-rT}K$ at risk-free interest. $\longrightarrow$ at time T portfolio is worth S_T - risk-free interest.

Since both strategies are equal, their cost at time 0 must coincide. Otherwise, there is the possibility for arbitrage. Hence, $C_0 - P_0 = S_0 - e^{-rT}K$ (**call-put parity**). Where,

- C_0 = call price at time 0.
- P_0 = put price at time 0.

We do not know C_0 and P_0 (yet), but one determines the other.

Example 14.8. (Fair Price of a Call Option) Consider a model where there is only one time period $T = 1$ and only two outcomes: $S_1 = \begin{cases} 13 & \text{with } p \\ 8 & \text{with } (1-p) \end{cases}$ and $S_0 = 10$, we have a call option with $K = 11$, and assume $r = 0$.

What is the fair price of a call option? If we knew probabilities p, $1-p$ of outcomes, we'd have

$$\mathbb{E}((S_1 - K)^+) = p(13-10)^+ + (1-p)(8-10)^+ = 2p.$$

But how to choose p? We try to replicate the option with a portfolio of:

- η units cash
- θ units asset

At time $T = 1$ this is worth

$$\begin{cases} \eta + 13\theta = 2 \\ \eta + 8\theta = 0 \end{cases} \implies \eta = -3.2, \theta = 0.4.$$

So, the fair price of the option has to be the value of the portfolio at time 0 (otherwise, there would be arbitrage): $C_0 = \eta + 10\theta = 0.8$.

Note that this price corresponds to the expected payoff, with probability 0.4 that the price goes up. For such p, we get

$$S_1 = \begin{cases} 13 & \text{with prob } p = 0.4 \\ 8 & \text{with prob } 1-p \end{cases}, \qquad S_0 = 10.$$

$$\text{And} \qquad \mathbb{E}(S_1|S_0) = p \cdot 13 + (1-p)8 = 10 = S_0 \qquad \text{so a } \textbf{martingale}!!$$

This is not a coincidence; we will see that it holds in much greater generality.

Fair option price = expected payoff assuming that the asset price is a martingale.

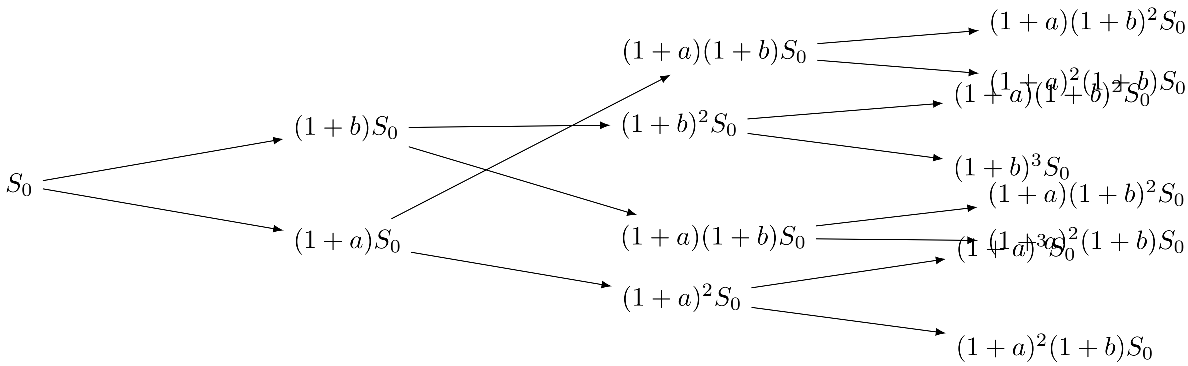
Remark 14.9. Finally, a replicating strategy was possible because there were only two possible outcomes. This might not be true in general. Models where every contingent claim (options) can be obtained by a hedging strategy are called **complete**.

14.1 Binomial Model

At time periods T_i, the asset price can change by a factor of $(1+a)$ or $(1+b)$, $a < b$.

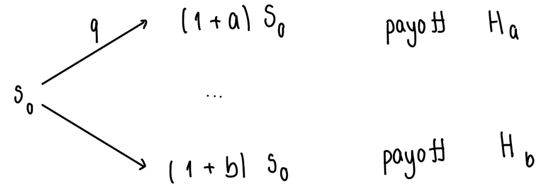
$$S_i = \begin{cases} (1+a)S_{i-1} \\ (1+b)S_{i-1} \end{cases}$$

for all time steps i. The risk-free rate satisfies $a < r < b$.



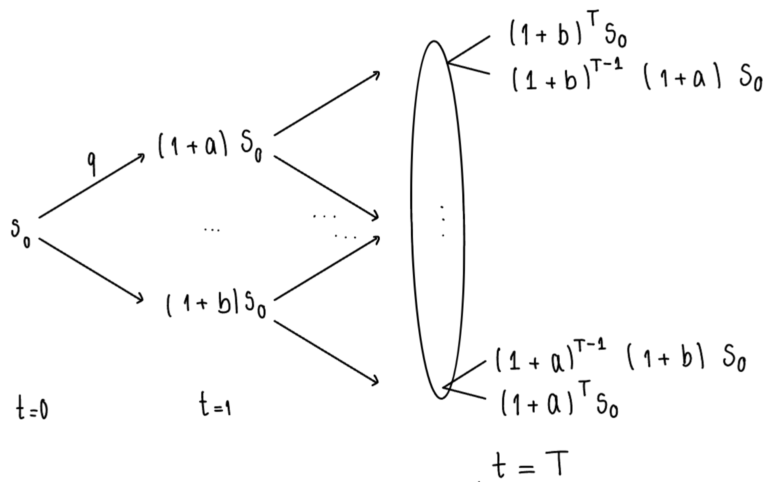
.... i haven't written the rest YET

15 Discrete Markets



The value of a portfolio that replicates the payoff is $q\beta H_a + (1-q)\beta H_b$ at time 0, where $\beta = \frac{1}{1+r}$ is the discounting factor and q is such that $\mathbb{E}(\beta S_1|S_0) = S_0$.

For arbitrary time steps, use payoffs at time T to calculate the replicating portfolio at time



$T-1$.

Repeating the argument gives

$$\mathbb{E}(\beta^T \cdot \textbf{ payoff })$$

with binomial distribution and probabilities $q = \frac{b-r}{b-a}$ and $1-q = \frac{r-a}{b-a}$. We get a martingale for the asset price. The probabilities are such that the discounted asset price is a martingale

$$\mathbb{E}(\beta^n S_n|\mathcal{F}_{n-1}) = \beta^{n-1} S_{n-1}.$$

The probability of asset price at time T to be $(1+b)^{T-k}(1+a)^k S_0$ for $0 \leq k \leq T$ is

$$\binom{T}{k} q^k (1-q)^{T-k}.$$

Let $H(x)$ be the payoff if $S_T = x$. Then

$$\mathbb{E}(\beta^T \cdot \text{ payoff }) = \beta^T \sum_{k=0}^{T} \binom{T}{k} q^k (1-q)^{T-k} H\left((1+a)^k (1+b)^{T-k} S_0\right).$$

The expected payoff (discounted) with binomial model. For a call option (European) $H(x) = (x-k)^T$ [**Car-Ross-Rubinstein Formula**]:

$$\mathbb{E}(\beta^T \cdot \text{ payoff }) = \beta^T \sum_{k=0}^{T} \binom{T}{k} q^k (1-q)^{T-k} \left((1+a)^k (1+b)^{T-k} S_0 - k\right).$$

15.1 Some General Bounds

1. **European/American Options**: Let $C_0(A)$, $C_0(E)$ be the cost of the options, where we assume same S, T, K. We have

   $$0 \leq C_0(E) \leq C_0(A)$$

2. **Call-Put Parity**: If C_0 and P_0 are the prices on the call and on the put, respectively:

   $$C_0(E) - P_0(E) = S_0 - \beta K$$

   So we get:

   $$C_0(A) \geq C_0(E) \geq S_0 - \beta^T K \geq S_0 - K,$$

   assuming $r \geq 0 \iff \beta \leq 1$. By the same argument,

   $$C_t(A) \geq S_t - K$$

   Hence the current pay-off of an American option at time t is

   $$(S_t - K)^+.$$

   So, it is "always better" to wait until maturity T. A "perfect trading strategy" keeps the option until time T. Thus $C_0(E) = C_0(A)$, making American options equivalent to European options.

15.2 General Discrete Models

- **Probability Space** $(\Omega, \mathcal{F}, \mathbb{P})$ to model the market;

- **Filtration** $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{F}$ to model time;
- **Price Process**: vector $S = \left(S^0, S^1, \ldots, S^d\right)$ where:
  - S_t^0 is the risk-free investment ("cash in bank") and a discrete process;
  - S_t^i is the value of asset i at time t and is adapted to filtration $(\mathcal{F}_\backslash)$;
  - At least one of S_t^i is strictly positive;
  - The discounting factor is $\beta_t = \frac{1}{S_t^0}$, normalizing all the vectors.

15.3 Trading Strategies

- **Portfolio** at time t is $\Theta_t = \left(\theta_t^0, \theta_t^1, \ldots, \theta_t^d\right)$ describes the amount of each asset. We want $\Theta = (\Theta_t)$ to be previsible/predictable and $\mathcal{F}_{t-1}$ measurable. The value at time t is

$$V_t(\Theta) = \Theta_t \cdot S_t = \sum_{i=0}^{d} \theta_t^i S_t^i.$$

- A strategy is **self-financing** if we have no access to outside funds

$$\Theta_{t-1} \cdot S_t = \Theta_t \cdot S_t.$$

  Equivalently,

$$\Delta V_t(\Theta) = V_t(\Theta) - V_{t-1}(\Theta) = \Theta_t \cdot S_t - \Theta_{t-1} \cdot S_{t-1} = \Theta_t \cdot S_t - \Theta_t \cdot S_{t-1} = \Theta_t \cdot (S_t - S_{t-1}) = \Theta_t \cdot \Delta S_t$$

- The **Gains Process** is

$$G_0(\Theta) = 0, \qquad G_t(\Theta) = V_t(\Theta) - V_{t-1}(\Theta)$$

- To account for the risk-free investment, we define **discounted** versions:
  - If X_t is a random variable (possible vector) the discounted version

$$\overline{X_t} = \beta_t X_t = \frac{X_t}{S_t^0}.$$

  - The portfolio above is self-financing if (also)

$$\Delta\Theta_t \cdot \overline{S_{t-1}} = (\Theta_t - \Theta_{t-1}) \cdot \overline{S_{t-1}} = (\Theta_t - \Theta_{t-1}) \cdot \beta_{t-1} S_{t-1} = 0.$$

– It is always possible to turn Θ into a self-financing strategy by changing Θ_t^0 (a linear equation to solve).

Proposition 15.1. Arbitrage
The existence of arbitrage, i.e., the existence of a strategy θ with $V_0(\theta) = 0$, $V_t(\theta) \geq 0$ for all t, and $\mathbb{E}(V_t(\theta)) > 0$, is equivalent to weak arbitrage, i.e., the existence of a strategy with $V_0(\theta) = 0$, $V_t(\theta) \geq 0$ at time T, and $\mathbb{E}(V_t(\theta)) > 0$.

Proof. Arbitrage implies weak arbitrage, which is trivial.

To prove the converse, suppose θ is a portfolio that offers weak arbitrage but $V_t(\theta)$ is not almost surely non-negative for all t. Then there exists a time $t < T$ and an event $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, such that

$$V_t(\theta)(\omega) = (\theta \cdot S_t)(\omega) < 0 \text{ for } \omega \in A.$$

It can be assumed that t is the last such time (discrete model!) and $V_u(\theta)(\omega) \geq 0$ a.s. for all $t < u \leq T$. Now construct a portfolio that offers (strong) arbitrage:

$$\begin{cases} \psi_u(\omega) = 0 & \text{for } \omega \in A \text{ and all } u \\ \psi_u(\omega) = 0 & \text{for } \omega \notin A \text{ and } u \leq t \\ \phi_u^0(\omega) = \theta_u^0(\omega) - \frac{\theta_t(\omega) \cdot S_t(\omega)}{S_t^0(\omega)} & \text{for } \omega \in A,\ u > t \\ \phi_u^i(\omega) = \theta_u^i(\omega) & \end{cases}$$

This is previsible/predictable self-financing strategy. We also have $V_t(\phi) \geq 0$ for all t, $V_0(\phi) = 0$ and $\mathbb{E}(V_T(\psi)) > 0$. Thus we have arbitrage. ☐

15.4 Arbitrage Prices

Definition 15.2. If H is a claim with maturity T which has a generating/replicating strategy θ (i.e. self-financing strategy with $V_T(\theta) = H$) then we say H is **attainable**.
If H is an attainable claim with maturity T, $V_0(\theta)$ can be taken as a **fair price** for H.

We want to know if this price is **unique**.

Lemma 15.3. For all generating strategies of H, the associated value process is the same all the time (a.s.), provided the model is free of arbitrage.

Proof. Assume to the contrary and we have two strategies *theta*, ϕ such that $V_T(\theta) = H =$

$V_T(\phi)$ a.s. but $V_t(\theta) \neq V_t(\phi)$ for some t with positive probability. WLOG we may assume $A = \{V_t(\theta) > V_t(\phi)\} \in \mathcal{F}_t$ has positive probability. Now consider the strategy ψ with $\psi_u(\omega) = \theta_u(\omega) - \phi_u(\omega)$ for all u if $\omega \notin A$ and for $\omega \in A$ let

$$\begin{cases} \psi_u(\omega) = \theta_u(\omega) - \phi_u(\omega) & \text{for } u \leq t \\ \psi_u^0(\omega) = \frac{V_t(\omega) - V_t(\phi)}{S_t^0} & \text{for } u > t \\ \phi_u^i(\omega) = 0 & \text{for } u > t,\ i > 0 \end{cases}$$

- If A does not occur, $V_T(\psi) = V_T(\theta) - V_T(\psi) = 0$.
- If A occurs, positive difference is converted into cash and $V_T(\psi) > 0$.

$\Longrightarrow$ Hence there is weak arbitrage, contradicting the arbitrage-free assumption.

$\Longrightarrow$ Thus the fair price can be uniquely defined by $V_0(\theta)$ a.s.

$\square$

16 Martingales and Pricing

Remark 16.1. Recall that if M_t is a martingale, then

$$\mathbb{E}\left(M_t \mid \mathcal{F}_{t-1}\right) = M_{t-1} \iff \mathbb{E}\left(M_t - M_{t-1} \mid \mathcal{F}_{t-1}\right) = 0 \iff \mathbb{E}\left(\Delta M_t \mid \mathcal{F}_{t-1}\right) = 0.$$

For predictable/previsible φ, define the martingale transform $X = \varphi \cdot M$:

$$X_t = \varphi_1 \Delta M_1 + \varphi_2 \Delta M_2 + \ldots + \varphi_t \Delta M_t = \sum_{k=1}^{t} \varphi_k \left(M_k - M_{k-1}\right).$$

Remark 16.2. Note that

$$\mathbb{E}\left(\varphi_k \Delta M_k \mid \mathcal{F}_{k-1}\right) = \varphi_k \mathbb{E}\left(\Delta M_k \mid \mathcal{F}_{k-1}\right) = 0$$

since M_k is a martingale and φ_k is predictable.

Thus

$$\mathbb{E}\left(\varphi_k \Delta M_k\right) = 0 \implies \mathbb{E}\left(X_k\right) = 0 \text{ for all } t \leq k.$$

Now suppose we have a probability measure $\mathcal{Q}$ such that the discounted price process $\overline{S_t}$ becomes a martingale:

$$\mathbb{E}_{\mathcal{Q}}\left(\Delta \overline{S_t^i} \mid \mathcal{F}_{t-1}\right) = 0 \text{ for all } i,\, t.$$

Equivalently,

$$\mathbb{E}_{\mathcal{Q}}\left(\overline{S_t^i} \mid \mathcal{F}_{t-1}\right) = \overline{S^i}_{t-1} \text{ for all } i,\, t.$$

Let θ be an admissible strategy. The discounted value process can be expressed as a martingale transform:

$$\overline{V_t}\left(\theta\right) = V_0\left(\theta\right) + \sum_{u=1}^{t} \theta_u \cdot \Delta \overline{S_u} = \sum_{i=0}^{d} \theta_0^i S_0^i + \sum_{i=1}^{d} \left(\sum_{u=1}^{t} \theta_u^i \Delta \overline{S_u^i}\right).$$

$i = 0$ is the cash term, which when discounted is a constant:

$$\Delta \overline{S_u^0} = 0.$$

Since

$$\mathbb{E}_{\mathcal{Q}}\left(\sum_{u=1}^{t}\theta_u^i \Delta\overline{S_u^i}\right) = 0$$

by our observations and the assumption that $\overline{S_u}S$ is a martingale, we get:

$$\mathbb{E}_{\mathcal{Q}}\left(\overline{V_t}\left(\theta\right)\right) = \mathbb{E}_{\mathcal{Q}}\left(V_0\left(\theta\right)\right) = V_0\left(\theta\right).$$

Let $\mathbb{P}$ be the probability measure in our model. If $\mathbb{P}$ and $\mathcal{Q}$ are equivalent (written $\mathbb{P} \sim \mathcal{Q}$, i.e. $\mathbb{P}$ and $\mathcal{Q}$ have the same null sets), then this rules out arbitrage.

Assume there is a portfolio θ such that $V_0\left(\theta\right) = 0$ but $V_T\left(\theta\right) \geq 0$ $\mathbb{P}$ almost surely. Then also

$$V_T\left(\theta\right) \geq 0$$

$\mathcal{Q}$ almost surely since $\mathcal{Q}$ and $\mathbb{P}$ are equivalent. Note also that

$$\mathbb{E}_{\mathcal{Q}}\left(\overline{V}_t\left(\theta\right)\right) = \mathbb{E}_{\mathcal{Q}}\left(V_0\left(\theta\right)\right) = V_0\left(\theta\right) = 0.$$

From this, it follows that

$$V_T\left(\theta\right) = 0$$

$\mathcal{Q}$ a.s. and so $\mathbb{P}$ a.s. In other words, there is no weak (and hence full) arbitrage.

$\mathcal{Q}$ is called the equivalent martingale measure.

Proposition 16.3. If H is an attainable claim (i.e., H has a replicating strategy), then for any replicating strategy θ we have:

$$\overline{V}_t\left(\theta\right) = \mathbb{E}_{\mathcal{Q}}\left(\beta H \mid \mathcal{F}_t\right)$$

a.s. (with respect to $\mathbb{P}$ and $\mathcal{Q}$).

Proof. This follows by taking conditional expectations in

$$\overline{V}_t\left(\theta\right) = V_0\left(\theta\right) + \sum_{u=1}^{t}\theta_u \cdot \Delta\overline{S_u}$$

and using the martingale property of $\overline{S}$. □

Definition 16.4. We can define the fair price of H by $\overline{V}_0(\theta)$:

$$\pi(H) = \overline{V}_0(\theta) = \mathbb{E}_{\mathcal{Q}}(\beta_T H \mid \mathcal{F}_0) = \mathbb{E}_{\mathcal{Q}}(\beta_T H).$$

Example 16.5. In the binomial model, the measure $\mathcal{Q}$ was determined by the probability q that turned $\frac{S_t}{S_t^0}$:

$$S_{t-1} \longrightarrow \begin{cases} S_t(1+a) & \text{prob } 1-q \\ S_t(1+b) & \text{prob } q \end{cases}$$

into a martingale, up to the discounting factor β.

$$\mathbb{E}\left(\overline{S_t} \mid \mathcal{F}_{t-1}\right) = q\beta\overline{S}_{t-1}(1+a) + (1-q)\beta(1+b)\overline{S}_{t-1}$$

q is then determined by the equation:

$$1 = q\beta(1+a) + (1-q)\beta(1+b) = \frac{q(1+a)}{(1+r)} + \frac{(1-q)(1+b)}{1+r}$$

$$\iff 1+r = q(1+a) + (1+b) - q(1+b) = q(a-b) + 1 + b \iff q = \frac{r-b}{a-b} = \frac{b-r}{b-a}.$$

Proposition 16.6. The price of a European call can then be expressed as:

$$\mathbb{E}_{\mathcal{Q}}\left(\beta_T (S_T - K)^+\right).$$

Evaluating this expectation gives the **Cox-Ross-Rubinstein formula**.

Note that the formula requires the existence of a replicating strategy, even if not explicitly referenced. In a **complete** market model, all European contingent claims have a replicating strategy and can be priced this way.

16.1 Uniqueness of Equivalent Martingale Measures

Proposition 16.7. Uniqueness of Equivalent Martingale Measures
In principle, the martingale measure may not be unique. Suppose $\mathcal{Q}$ and $\mathbb{R}$ are two equivalent measures in a **complete model**. Then,

$$\mathbb{E}_{\mathcal{Q}}\left(\beta_T H\right) = \mathbb{E}_{\mathbb{R}}\left(\beta_T H\right)$$

for all H,

$$\implies \mathbb{E}_{\mathcal{Q}}\left(H\right) = \mathbb{E}_{\mathbb{R}}\left(H\right).$$

The fair price is unique in the absence of arbitrage.
So,

$$\mathbb{E}_{\mathcal{Q}}\left(\mathbb{I}_A\right) = \mathbb{E}_{\mathbb{R}}\left(\mathbb{I}_A\right)$$

for all indicators $\mathbb{I}_A$, and

$$\mathcal{Q}\left(A\right) = \mathbb{R}\left(A\right)$$

for all A.

17 The Black-Scholes Formula

Background:

- If there exists an equivalent martingale measure $\mathcal{Q}$ ($\mathcal{Q} \sim P$, $\overline{S}_t = \beta_t S_t$ is $\mathcal{Q}$-martingales;
- If H is an attainable claim, then $\overline{V}_t(\theta) = \mathbb{E}_{\mathcal{Q}}[\beta_T h | \mathcal{F}_t]$;
- If the model is complete, then $\mathcal{Q}$ is unique.

17.1 Super hedging

Definition 17.1. Given a claim $H = f(S_T)$ (function of an asset price at time N), an (z, H)-hedge is an admissible strategy θ with $V_0(\theta) = z$ and $V_N(\theta) = H$ a.s.

Definition 17.2. The seller's price $\Pi_S(H)$ can now be defined as:

$$\Pi_S = \inf\{z \in \mathbb{R} : \text{ there exists a } (z, H) - \text{hedge } \theta\}$$

$$V_T(\theta) = z + G_T(\theta) \geq H$$

where G_T is the gains process.

This guarantees the seller of H not to incur losses.

Definition 17.3. The buyer's price $\Pi_B(H)$ is analogously:

$$\Pi_B = \sup\{z \in \mathbb{R} : \text{ there exists a } (-z, -H) - \text{hedge } \theta\}$$

$$V_T(\theta) = -z + G_T(\theta) \geq -H$$

where G_T is the gains process.

This guarantees the buyer not to incur losses.

Proposition 17.4. If there is a replicating strategy θ, then:

$$V_t(\theta) = H \implies \begin{cases} \theta \text{ is a } (z, H) - \text{hedge} \\ -\theta \text{ is a } (-z, -H) - \text{hedge} \end{cases}$$

for $z = V_0(\theta)$. In this case, $\Pi_S = \Pi_B = \Pi(H)$.
In general, we only have $\Pi_B \leq \Pi_S$.

17.2 Equivalent Martingale Measure

Proposition 17.5. If we have an equivalent martingale measure $\mathcal{Q}$, then for a seller's strategy θ we have:

$$\mathbb{E}_{\mathcal{Q}}\left[\overline{H}\right] = \mathbb{E}_{\mathcal{Q}}\left[\overline{V}_N(\theta)\right] = \mathbb{E}_{\mathcal{Q}}\left[V_0(\theta)\right] = V_0(\theta)$$

$$\implies \mathbb{E}_{\mathcal{Q}}\left[\overline{H}\right] \leq \Pi_S \text{ upon taking } \inf.$$

In the same way,

$$\mathbb{E}_{\mathcal{Q}}\left[\overline{H}\right] \geq \Pi_B \implies \Pi_B \leq \mathbb{E}_{\mathcal{Q}}\left[\overline{H}\right] \leq \Pi_S.$$

If the claim is attainable (i.e. exists a replicating strategy), we have equalities throughout.

$$\Pi_B = \mathbb{E}_{\mathcal{Q}}\left[\overline{H}\right] = \Pi_S$$

17.3 Strategies Involving Contingent Claims

We now expand our standard model with asset prices $S^0, S^1, \ldots, S^d$ by adding some attainable European claims $Z^1, Z^2, \ldots, Z^n$.

Definition 17.6. **A trading strategy** is now a pair $\Phi = (\theta, \varphi)$ with initial value

$$V_0(\Phi) = \theta_0 \cdot S_0 + \varphi \cdot Z_0.$$

- "standard" trading strategy θ
- trading strategy φ with attainable claims $Z^1, Z^2, \ldots, Z^n$

Proposition 17.7. It is **self-financing** if:

$$\theta_t \cdot S_t + \varphi_t \cdot Z_t = \theta_{t-1} \cdot S_t + \varphi_{t-1} \cdot Z_t$$

17.4 Arbitrage-free Model

Theorem 17.8. *The model is arbitrage-free if and only if every attainable European claim with payoff Z has value process:*

$$Z_t = S_t^0 \mathbb{E}_{\mathcal{Q}} \left(\frac{Z_T}{S_T^0} \mid \mathcal{F}_t \right)$$

where $\mathcal{Q}$ is an equivalent martingale measure for the price process S.

Proof.

$\Longrightarrow$ By assumption, there is a replicating strategy θ for Z. Its value is $V_0(\theta) = S_0^0 \mathbb{E}_{\mathcal{Q}} \left(\frac{Z_T}{S_T^0} \mid \mathcal{F}_t \right)$ by the martingale property.

Now suppose $V_t(\theta) \neq z_t$ on a set of positive measure. Without loss of generality, we may assume $\mathbb{P}(Z_u > V_u(\theta)) > 0$ for some time u. There now exists an arbitrage strategy:

- Do nothing until time u.
- If the event $\{Z_u > V_u(\theta)\}$ does not occur, keep doing nothing.
- If $Z_u > V_u(\theta)$:
  - Sell Z for price Z_u.
  - Invest in θ (at price $V_u(\theta)$).
  - Put positive difference in bank.

At time T, $V_T(\theta) = Z$, so θ and Z cancel. We are left with the difference. This happens with positive probability and is hence (local) arbitrage, a contradiction in an arbitrage-free market.

Hence

$$Z_t = S_t^0 \mathbb{E}_{\mathcal{Q}} \left(\frac{Z_T}{S_T^0} \mid \mathcal{F}_t \right) \text{ a.s.}$$

for all t. This proves the first direction.

$\Longleftarrow$ Assume now that all contingent claims follow this rule:

$$Z_t = S_t^0 \mathbb{E}_{\mathcal{Q}} \left(\frac{Z_T}{S_T^0} \mid \mathcal{F}_t \right).$$

Consider a self-financing strategy $\Phi = (\theta, \varphi)$ with $V_0(\Phi) = 0$, $V_T(\Phi) = 0$ (a.s.).

$$\mathbb{E}_{\mathcal{Q}}\left(\overline{V_t}(\Phi) \mid \mathcal{F}_{t-1}\right) = \mathbb{E}_{\mathcal{Q}}\left(\sum_i \theta_t^i \frac{\overline{S_t^i}}{S_t^0} + \sum_i \varphi_t^i \frac{\overline{Z_t^i}}{S_t^0} \mid \mathcal{F}_{t-1}\right)$$

$$= \mathbb{E}_{\mathcal{Q}}\left(\sum_i \theta_t^i \frac{\overline{S_t^i}}{S_t^0} \mid \mathcal{F}_{t-1}\right) + \mathbb{E}_{\mathcal{Q}}\left(\sum_i \varphi_t^i \frac{\overline{Z_t^i}}{S_t^0} \mid \mathcal{F}_{t-1}\right) = \sum_i \theta_{t-1}^i \mathbb{E}_{\mathcal{Q}}\left(\frac{\overline{S_t^i}}{S_t^0} \mid \mathcal{F}_{t-1}\right) + \sum_i \varphi_{t-1}^i \mathbb{E}_{\mathcal{Q}}\left(\frac{\overline{Z_t^i}}{S_t^0} \mid \mathcal{F}_{t-1}\right)$$

$$= \mathbb{E}_{\mathcal{Q}}\left(\mathbb{E}_{\mathcal{Q}}\left(\sum_i \theta_t^i \frac{\overline{S_t^i}}{S_t^0} + \sum_i \varphi_t^i \frac{\overline{Z_t^i}}{S_t^0} \mid \mathcal{F}_t\right) \mid \mathcal{F}_{t-1}\right) = \mathbb{E}_{\mathcal{Q}}\left(\frac{\overline{Z_t^i}}{S_t^0} \mid \mathcal{F}_{t-1}\right) = \frac{\overline{Z_{t-1}^i}}{S_{t-1}^0}$$

Combining gives:

$$V_{t-1}(\Phi) = 0 \text{ and } V_t(\Phi) \text{ is a martingale.}$$

$$\implies \mathbb{E}\left[\overline{V_T}(\Phi)\right] = \mathbb{E}\left[\overline{V_0}(\Phi)\right] = 0 \implies V_T(\Phi) = 0 \text{ a.s.}$$

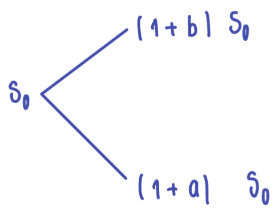
since $V_t(\Phi) \geq 0$ a.s. by assumption.

It follows that there is no arbitrage. $\square$

17.5 The Black-Scholes Formula

Random walk $\to$ Brownian motion

We define a, b, r in such a way that the process converges.



with probability q s.t. $q(1+b) + (1-q)(1+q) = 1+r$

$q = \dfrac{r-a}{b-a}$

Let $h_N = \frac{T}{N}$ (length of one time step), and $\rho_N = r h_N$ (risk-free rate). Observe that

$$(1+\rho_N)^N = \left(1 + r\frac{T}{N}\right)^N \to e^{rT}.$$

Let a_N, b_N satisfy:

$$\frac{\log(1+a_N)}{(1+\rho_N)} = \sigma\sqrt{h_N} = \sigma\sqrt{\frac{T}{N}} \frac{\log(1+b_N)}{(1+\rho_N)} = -\sigma\sqrt{h_N} = -\sigma\sqrt{\frac{T}{N}}$$

chosen s.t. the process converges turns multiplication into addition where σ is a constant that measures the **volatility** of an asset. We get:

$$1 + a_N = (1 + \rho_N)\, e^{\sigma\sqrt{h_N}} \qquad 1 + b_N = (1 + \rho_N)\, e^{-\sigma\sqrt{h_N}}$$

and we find:

$$q_N = \frac{b_N - \rho_N}{b_N - a_N} = (1 + \rho_N)\, e^{-\sigma\sqrt{h_N}} - \frac{1 + \rho_N}{1 + \rho_N} e^{-\sigma\sqrt{h_N}} - (1 + \rho_N)\, e^{\sigma\sqrt{h_N}} \sim \frac{1}{2}$$

and $1 - q_N \sim \frac{1}{2}$.

The discounted price process $\frac{S_N}{S_0^0}$ becomes:

$$\frac{S_N}{S_0^0} = \frac{S_N}{S_0} (1 + \rho_N)^{-N} = \prod_{k=1}^{N} \frac{\frac{S_k}{S_{k-1}}}{1 + \rho_N} = \prod_{k=1}^{N} \left(\text{either } \ (1 + a_N) \ \text{ or } \ \frac{1 + b_N}{1 + \rho_N} \right)$$

$\uparrow$ dependence on final time

$$= S_0 \prod_{k=1}^{N} \left(\frac{1 + \rho_N}{1 + \rho_N} \right) \ (\text{either } e^{\sigma\sqrt{h}} \text{ or } e^{-\sigma\sqrt{h}}) = S_0 \exp \mathbb{P} \left(\sum_{k=1}^{N} Y_k(n) \right)$$

where $Y_k(n)$ is $\pm\sigma\sqrt{h_k}$. The sum has mean:

$$\mathbb{E}_{\mathcal{Q}} \left(\sum_{k=1}^{N} Y_k(n) \right) = N \left[q_N \sigma \sqrt{\frac{T}{N}} + (1 - q_N) \left(-\sigma \sqrt{\frac{T}{N}} \right) \right] \sim \sigma^2 \frac{T}{2}$$

and variance:

$$\mathrm{Var}_Q \left(\sum_{k=1}^{N} Y_k(n) \right) = N \mathrm{Var}_Q (Y_k(n)) \sim \sigma^2 T$$

The *Central Limit Theorem* gives us that $\sum_{k=1}^{N} Y_k$ converges in distribution to a normal distribution:

$$\sum_{k=1}^{N} Y_k(n) \sim \left(-\sigma^2 \frac{T}{2}, \sigma^2 T \right).$$

This justifies our choice of $\pm\sigma\sqrt{h_k}$. If the factors were too large there would be no convergence, if too small there would be a fixed deterministic limit.

The discounted final price $\frac{S_N}{S_N^0}$ under the martingale measure $\mathcal{Q}$ is obtained as:

$$\exp \mathbb{P}\left(-\sigma^2 \frac{T}{2}, \sigma^2 T\right) \sim \exp \mathbb{P}\left(-\sigma^2 \frac{T}{2} + \sigma\sqrt{T} \cdot (0,1)\right).$$

We can now plug this into the general formula:

$$\mathbb{E}_{\mathcal{Q}}\left(\beta^T \cdot \text{payoff at time } T\right)$$

and we obtain the **Black-Scholes Formula**.

18 The Separating Hyperplane Theorem

Theorem 18.1. ***(Separating Hyperplane Theorem):*** *Let $\mathcal{L}$ be a linear subspace of $\mathbb{R}^n$ and K a convex compact subset of $\mathbb{R}^n$ disjoint from $\mathcal{L}$. Then there exists a linear functional $\Phi : \mathbb{R}^n \to \mathbb{R}$ such that:*

- *$\Phi(x) = 0$ for all $x \in \mathcal{L}$,*
- *$\Phi(x) \geq c$ for all $x \in K$, where $c > 0$.*

For the proof, we will first prove the following lemma:

Lemma 18.2. Let $C \subseteq \mathbb{R}^d$ be closed and convex such that $0 \notin C$. Then there exists a linear functional $\Phi : \mathbb{R}^d \to \mathbb{R}$ and $c > 0$ such that $\Phi(x) \geq c$ for all $x \in C$.

Proof. If C is empty, the statement is trivial, so we assume C is non-empty.

Take $r > 0$ large enough such that the closed ball

$$B(0, r) = \{x \in \mathbb{R}^d : \|x\| \leq r\}$$

intersects C.

The intersection $C \cap B(0, r)$ is non-empty and closed, hence compact. Since the norm $\|\cdot\|$ is continuous, it has a minimum in $C \cap B(0, r)$. That is, there exists $x_0 \in C \cap B(0, r)$ such that $\|z\| \geq \|x_0\|$ for all $z \in C \cap B(0, r)$. Further, $\|x_0\| \leq r$ and $\|x_0\| > 0$ (since $0 \notin C$).

For all $z \in C$, by convexity, if $z \in C$, then $\lambda x_0 + (1 - \lambda)z \in C$ for all $\lambda \in [0, 1]$. So,

$$\|\lambda x_0 + (1 - \lambda)z\|^2 \leq \|x_0\|^2.$$

Expanding the squared norm, we have:

$$\|\lambda x_0 + (1 - \lambda)z\|^2 = \langle \lambda x_0 + (1 - \lambda)z, \lambda x_0 + (1 - \lambda)z \rangle.$$

Expanding further,

$$\|\lambda x_0 + (1 - \lambda)z\|^2 = \lambda^2 \|x_0\|^2 + (1 - \lambda)^2 \|z\|^2 + 2\lambda(1 - \lambda)\langle x_0, z \rangle.$$

Since $\|\lambda x_0 + (1 - \lambda)z\|^2 \leq \|x_0\|^2$, we have:

$$\lambda^2 \|x_0\|^2 + (1 - \lambda)^2 \|z\|^2 + 2\lambda(1 - \lambda)\langle x_0, z \rangle \leq \|x_0\|^2.$$

Simplify this inequality to deduce properties about $\langle x_0, z\rangle$.

Taking $\lambda \to 1$, we deduce that

$$2\langle x_0, x_0\rangle \leq 2\langle x_0, z\rangle \quad \Rightarrow \quad \langle x_0, x_0\rangle \leq \langle x_0, z\rangle.$$

Hence the linear functional $\Phi(z) = \langle x_0, z\rangle$ satisfies $\Phi(z) \geq \|x_0\|^2 > 0$ for all $z \in C$. $\square$

Proof. (of the Separating Hyperplane Theorem): Let

$$C = K - \mathcal{L} = \{k - \ell : k \in K, \ell \in \mathcal{L}\}.$$

This set is convex: for $x_1 = k_1 - \ell_1 \in C$ and $x_2 = k_2 - \ell_2 \in C$, we have

$$\lambda x_1 + (1-\lambda)x_2 = \lambda(k_1 - \ell_1) + (1-\lambda)(k_2 - \ell_2) = (\lambda k_1 + (1-\lambda)k_2) - (\lambda \ell_1 + (1-\lambda)\ell_2) \in K - \mathcal{L} = C.$$

C is also closed. Let $C_n = k_n - \ell_n$ be a sequence that converges (in $\mathbb{R}^n$) as $n \to \infty$. By compactness of K, there exists a subsequence $k_{n_r} \to k_\infty$ for some $k_\infty \in K$. $\square$

18.1 Construction of Martingale Measures

Definition 18.3. A probability space is called *finitely generated* if every measurable function takes at most finitely many distinct values.

Proposition 18.4. Ω can be partitioned into n disjoint sets

$$\Omega = \Omega_1 \cup \Omega_2 \cup \ldots \cup \Omega_n \quad \text{such that every measurable function is constant on } \Omega_i.$$

We now assume our model is finitely generated. Then, the condition of being viable (arbitrage-free) is equivalent to gains $G_T(\theta)$ not belonging to

$$C = \{X : X(\omega) \geq 0 \text{ for all } \omega,\ X(\omega) > 0 \text{ for at least one } \omega\}.$$

18.2 First Fundamental Theorem of Asset Pricing

Theorem 18.5. ***First Fundamental Theorem of Asset Pricing*** *Assume that the model is finitely generated. The following are equivalent:*

(i) The model is viable.

(ii) There exists an equivalent martingale measure $\mathcal{Q}$.

Proof. (ii) $\implies$ (i)

If S is a martingale under $\mathcal{Q}$, then so is $G_T(\theta)$ for any attainable θ, as it is a martingale transform. Thus,

$$\mathbb{E}_{\mathcal{Q}}[G_T(\theta)] = \mathbb{E}_{\mathcal{Q}}[G_T(0)] = 0.$$

If $G_T(\theta) \geq 0$ a.s., then $G_T(\theta) = 0$. So there is no arbitrage, and the model is viable.

(i) $\implies$ (ii) First, let

$$\mathcal{L} = \{\overline{G_T(\theta)} : \theta \text{ attainable strategies}\}$$

be a linear subspace of the space of all random variables:

- $0 \in \mathcal{L}$.
- θ_1, θ_2 attainable strategies $\implies$ $\theta_1 + \theta_2$ attainable strategy.
- $-\theta$ attainable strategy, $c \in \mathbb{R}$ $\implies$ $c\theta$ attainable strategy.

Since $\overline{G_T(\theta)}$ is linear in θ, linearity of $\mathcal{L}$ follows.

Further, $K = \{X \text{ non-neg. random var. with } \mathbb{E}(X) = 1\}$ is compact w.r.t. $||\cdot||_1$ since any open cover $(U_i)_i$ of K must contain a U_i which contains an element $X \in K$. But then

$$B(X, \epsilon) = \{Y : ||Y||_1 \leq \epsilon\} \supseteq \{Y : \mathbb{E}|Y| \leq \epsilon\} \supseteq \{Y : \mathbb{E}|Y - X| \leq \epsilon - K\}.$$

Hence, U_i is a cover of K and K is compact.

K is also convex, since

$$\mathbb{E}(\lambda X_1 + (1-\lambda)X_2) = \lambda\mathbb{E}(X_1) + (1-\lambda)\mathbb{E}(X_2) = \lambda + (1-\lambda) = 1$$

and

$$\lambda X_1 + (1-\lambda)X_2 \in K.$$

Note that $K \cap \mathcal{L} = \varnothing$ since any attainable strategy with $\mathbb{E}(\widetilde{G}_T(\theta)) = 1$ would be arbitrage.

We can apply the separating hyperplane theorem and thus exists a linear functional φ that is zero on $\mathcal{L}$ and greater than $c > 0$ on K.

By linearity, $\varphi(X) = \sum_{i=1}^{n} q_i X(\omega_i)$ for some constants q_i. Consider the random var. $\xi_i = I_{\Omega_i}$. Then φ_i is non-negative and

$$\mathbb{E}(\xi_i) = 1 = \mathbb{E}(I_{\omega_i}) = \mathbb{P}(\{\omega_i\}) = p_i.$$

So $\xi_i \in K$ and $\varphi(\xi_i) = q_i = \varphi_i > c > 0$.

Hence $q_i > 0$ for all i. Define $\mathcal{Q}$ by

$$\mathcal{Q}(\{\omega_i\}) = \frac{q_i}{\sum q_j} > 0.$$

This is a probability measure as

$$\sum \frac{q_i}{\sum q_j} = \frac{\sum q_i}{\sum q_j} = 1.$$

We have $\mathbb{E}_{\mathcal{Q}}(\widetilde{G}_T(\theta)) = \frac{1}{\sum q_j} \sum q_i \widetilde{G}_T(\theta)(\omega_i) = \frac{1}{\sum q_j} \varphi(\widetilde{G}_T(\theta)) = 0$ for all attainable θ. ($\widetilde{G}_T(\theta) \in \mathcal{L}$)

Now $\mathcal{Q}$ is an equivalent martingale measure as:

- Equivalent to $\mathbb{P}$ since $\mathcal{Q}(\{\omega_i\}) > 0$.
- $\mathbb{E}_{\mathcal{Q}}(\widetilde{G}_T(\theta)) = 0$ for all predictable processes, so $\mathbb{E}_{\mathcal{Q}}(\Delta \widetilde{S}_t | \mathcal{F}_{t-1}) = 0$ for all t.

$\square$

Remark 18.6. This holds in greater generality for non-finite models.

18.3 Completeness of Market Models

Definition 18.7. A market model is **complete** if every contingent claim X has a replicating (generating) strategy θ: a strategy with $V_t(\theta) = X$.

Remark 18.8. We are only considering finite models.

But then,

$$M_t = \overline{V}_t(\theta) = V_0(\theta) + \sum_{u=1}^{t} \theta_u \Delta \overline{S}_u = M_0 + \sum_{u=1}^{t} \lambda_u \Delta \overline{S}_u$$

so we have a representation of M_t in the desired form.

For the course, consider a claim C. Define a martingale by $M_t = \mathbb{E}(\beta_T C | \mathcal{F}_t)$. There must be a representation of the form

$$M_t = M_0 + \sum_{u=1}^{t} \lambda_u \Delta \overline{S}_u$$

for some

$$\lambda_u = (\lambda_u^0, \ldots, \lambda_u^d).$$

Definition 18.9. Define a strategy by $\theta_u^i = \lambda_u^i$ which uniquely determines θ_u^0 as well by the self-financing property.
The precise choice is $\theta_u^0 = M_u - \lambda_u \cdot S_u$ which gives the required replicating strategy:

$$V_t(\theta) = \theta_t^0 S_t^0 + \sum_{(} i = 1)^d \theta_t^i S_t^i = S_t^0 (\theta_t^0 + \sum_{(} j = 1)^d \lambda_t^i \overline{S}_t^i) = S_t^0 M_t$$

In particular, $C = S_t^0 M_T = V_T(\theta)$.

18.4 Second Fundamental Theorem of Asset Pricing

Theorem 18.10. *(**Second Fundamental Theorem of Asset Pricing**) A finite market model is complete if and only if it has a **unique** equivalent martingale measure.*

Proof. Suppose first that the model is complete and assume that $\mathcal{Q}$, $\mathcal{Q}'$ are equivalent martingale measures. Let X be a contingent claim with generating strategy θ. We have

$$\beta_T X = \overline{V}_T(\theta) = V_0(\theta) + \sum_{t=1}^{T} \theta_t \Delta \overline{S}_t$$

By the martingale property, we have

$$E_{\mathcal{Q}}(\beta_T X) = V_0(\theta) = E_{\mathcal{Q}'}(\beta_T X).$$

This holds for all claims and in particular

$$X = I_A$$

for all events A. Hence,

$$\mathcal{Q}(A) = E_{\mathcal{Q}}(I_A) = E_{\mathcal{Q}'}(I_A) = \mathcal{Q}'(A)$$

and $\mathcal{Q} = \mathcal{Q}'$. Hence the equivalent martingale measure is unique. $\square$

For the course assume there exists a claim X that does not have a replicating strategy and let $\mathcal{Q}$ be an equivalent martingale measure.

Definition 18.11. Define

$$\mathcal{L} = \{c + \sum_{t=1}^{T} \theta_k \cdot \Delta\overline{S}_t | c \in \mathbb{R}, \theta_t \text{ predictable}\}.$$

This is a linear subspace of the vector space of all random variables in $(\Omega, \mathcal{F})$.

It is a proper subspace as $X \notin \mathcal{L}$. We assumed Ω is finite, so $\mathcal{L} \neq \Omega$. There exists a random variable Z that is orthogonal to $\mathcal{L}$, i.e. non-zero Z s.t. $E_{\mathcal{Q}}(YZ) = 0$ for all $Y \in \mathcal{L}$.

We define a new measure $\mathcal{Q}' \neq \mathcal{Q}$: Set

$$\mathcal{Q}'(\{\omega\}) = \mathcal{Q}(\{\omega\}) \left(1 + Z\frac{\omega}{2||Z||_\infty}\right),$$

where $||Z||_\infty = \max|Z(\omega)|$, and $\mathcal{Q}'$ is positive whenever $\mathcal{Q}$ is positive.

$$\sum_{\omega\in\Omega} \mathcal{Q}'(\omega) = \sum_{\omega\in\Omega} \mathcal{Q}(\omega) + \frac{1}{2||Z||_\infty} \sum_{\omega\in\Omega} \mathcal{Q}(\omega)Z(\omega) = 1$$

because $\mathcal{Q}$ is a prob. measure and $= E_{\mathcal{Q}}(Z) = 0$ as $Y = 1$ is in $\mathcal{L}$

So $\mathcal{Q}'$ is a probability measure with $\mathcal{Q} \sim \mathcal{Q}'$. We have

$$E_{\mathcal{Q}'}(\sum_{t=1}^{T} \theta_t \cdot \Delta\overline{S}_t) = \sum_{\omega\in\Omega} \mathcal{Q}'(\omega) \sum_{t=1}^{T} \theta_t(\omega) \cdot \Delta\overline{S}_t(\omega) =$$

$$\sum_{\omega\in\Omega} \mathcal{Q}(\omega) \sum_{t=1}^{T} \theta_t(\omega) \cdot \Delta\overline{S}_t(\omega) + 1/(2||Z||_\infty) \sum_{\omega\in\Omega} \mathcal{Q}(\omega)Z(\omega) \sum_{t=1}^{T} \theta_t(\omega) \cdot \Delta\overline{S}_t(\omega) = E_{\mathcal{Q}}(\sum_{t=1}^{T} \theta_t \cdot \Delta\overline{S}_t) = 0$$

by orthogonality.

$$= E_{\mathcal{Q}}(\sum_{t=1}^{T} \theta_t \cdot \Delta \overline{S}_t) = 0.$$

So

$$E_{\mathcal{Q}'}(\sum_{t=1}^{T} \theta_t \cdot \Delta \overline{S}_t) = 0$$

for all choices of θ, which is only possible if $\overline{S}_t$ is a martingale wrt $\mathcal{Q}'$.

Now both $\mathcal{Q}$, $\mathcal{Q}'$ are equivalent martingale measures and since Z is non-zero, there exists $\omega \in \Omega$ with $\mathcal{Q}(\omega), \mathcal{Q}'(\omega) > 0$ and $\mathcal{Q}(\omega) \neq \mathcal{Q}'(\omega)$. Hence $\mathcal{Q}$ is not unique, a contradiction. This proves the claim.