A Pocket Reference to Probability and Measure Theory

Pramook Khungurn

January 14, 2023

Materials are from [Bartle, 1995], [Jacod and Protter, 2004], [Williams, 1991], and [Schilling, 2017]. Most mathematical statements are stated without proofs. (It's a reference, not a textbook!)

1 Topology

- **Definition 1.1 (topology).** *A collection $\mathcal{S}$ of subsets of S is called a* **topology on** *S if it satisfies the following conditions.*

  1. *$\emptyset \in \mathcal{S}$.*
  2. *It is closed under arbitrary union: if $\{S_\alpha : \alpha \in A\}$ is an arbitrary collection of sets, then $\bigcup_{\alpha \in A} S_\alpha \in \mathcal{S}$.*
  3. *It is closed under finite union: if $A, B \in \mathcal{S}$, then $A \cap B \in \mathcal{S}$.*

  A member of a topology is called an **open set**.

- **Definition 1.2 (topological space).** *A* **topological space** *is a tuple $(S, \mathcal{S})$ where S is a set and $\mathcal{S}$ is a topology on S.*

  When it is clear what the topology on S is, we simply say that S is a "topological space."

2 Topology of the $\mathbb{R}$ and $\mathbb{R}^d$

- **Definition 2.1 (open subset of $\mathbb{R}$).** *A subset $A \subseteq \mathbb{R}$ is called* **open** *if, for every point $a \in A$, there exists a real number $\varepsilon(a) > 0$ such that $(a - \varepsilon(a), a + \varepsilon(a)) \subseteq A$.*

- **Proposition 2.2 (charaterization of open subsets of $\mathbb{R}$).** *Any open subset of $\mathbb{R}$ is a countable union of disjoint open intervals.*

  The proof relies on the fact that $\mathbb{Q}$ is dense in $\mathbb{R}$.

- **Proposition 2.3.** *The collection of open subsets of $\mathbb{R}$ is a topology on $\mathbb{R}$.*

- **Definition 2.4 (rectangle in $\mathbb{R}^d$).** *A* **(closed) rectangle in** $\mathbb{R}^d$ *is the set of the form*

  $$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$$

  where $-\infty < a_i < b_i < \infty$ for $i = 1, 2, \ldots, d$. The **interior** *of the rectangle is the set*

  $$(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d).$$

- **Definition 2.5 (open subset of $\mathbb{R}^d$).** *A subset $A \subseteq \mathbb{R}^d$ is said to be* **open** *if, for any point $\mathbf{a} \in A$, there a rectangle R such that $\mathbf{a} \in R$ and the interior of R is contained inside A.*

- **Definition 2.6 (almost disjoint).** *A collection of rectangles are* **almost disjoint** *if the interiors of the rectangles are disjoint.*
- **Proposition 2.7 (characterization of open subsets of $\mathbb{R}^d$).** *Any open subset of $\mathbb{R}^d$ is a countable union of almost disjoint rectangles.*
- **Proposition 2.8.** *The collection of open subsets of $\mathbb{R}^d$ is a topology on $\mathbb{R}^d$.*

3 Extended Real Number System

- It is convenient to work with the **extended real number system** $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.
- We sometimes call $\overline{\mathbb{R}}$ the **extended real line**.
- For any $x \in \mathbb{R}$, we have that $-\infty < x < \infty$.
- The arithematic operations between the infiniites and real numbers are as follows:

$$\begin{aligned}(\pm\infty) + (\pm\infty) = x + (\pm\infty) = (\pm\infty) + x &= \pm\infty \\ (\pm\infty)(\pm\infty) &= +\infty \\ (\pm\infty)(\mp\infty) &= -\infty \\ (\pm\infty)x = x(\pm\infty) &= \begin{cases} \pm\infty, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ \mp\infty, & \text{if } x < 0 \end{cases}\end{aligned}$$

  for any (finite) real number x.
- Note that we do not define $(\pm\infty) - (\pm\infty)$. We also do not define quotients when the denominators are $\pm\infty$.
- We deal with supermums and infemums in the following ways.
  - If $A \neq \emptyset$, $A \subseteq \mathbb{R}$, and A has no upper bound, then $\sup A = \infty$.
  - If $B \neq \emptyset$, $B \subseteq \mathbb{R}$, and B has no lower bound, then $\inf B = -\infty$.
  - $\sup \emptyset = -\infty$.
  - $\inf \emptyset = \infty$.

  In this way, all subsets of $\overline{\mathbb{R}}$ have supremums and infemums in $\overline{\mathbb{R}}$.

4 Topology of $\overline{\mathbb{R}}$

- **Definition 4.1 (open subset of $\overline{\mathbb{R}}$).** *A subset $A \subseteq \overline{\mathbb{R}}$ is said to be* **open** *if, for all $a \in A$, a has a neighborhood $N(a) \subseteq A$ such that $a \in N(a)$. What a neighborhood is depends on what a is.*
  - *If $a \in \mathbb{R}$, then $N(a)$ must be an interval $(a - \varepsilon, a + \varepsilon)$ with $\varepsilon > 0$.*
  - *If $a = \infty$, then $N(a)$ must be a set of the form $\{x \in \overline{\mathbb{R}} : x > b\} = (b, \infty]$ for some $b \in \mathbb{R}$.*
  - *If $a = -\infty$, then $N(a)$ must be a set of the form $\{x \in \overline{\mathbb{R}} : x < b\} = [-\infty, b)$ for some $b \in \mathbb{R}$.*
- **Proposition 4.2 (open intervals in $\overline{\mathbb{R}}$).** *An open interval in $\overline{\mathbb{R}}$ has one of the following forms:*

$$(a, b), (-\infty, b), [-\infty, b), (a, \infty), (a, \infty], (-\infty, \infty), [-\infty, \infty), (-\infty, \infty], [\infty, \infty]$$

  for any $a, b \in \mathbb{R}$ such that $a < b$.

- **Proposition 4.3 (characterization of open subsets of $\overline{\mathbb{R}}$).** *Any open subset of $\overline{\mathbb{R}}$ is a countable union of disjoint open intervals in $\overline{\mathbb{R}}$.*
- **Proposition 4.4.** *The collection of open subsets of $\overline{\mathbb{R}}$ is a topology on $\overline{\mathbb{R}}$.*

5 σ-Algebras

- In this section, let S be a set.
- The power set of S is denoted by 2^S.
- **Definition 5.1 (algebra).** *A collection of sets $\mathcal{S} \subseteq 2^S$ is called an* **algebra on** *S if it satisfies the following properties.*
  1. *$\emptyset, S \in \mathcal{S}$.*
  2. *It is closed under complementation: if $A \in \mathcal{S}$, then $A^c = S - A \subseteq \mathcal{S}$.*
  3. *It is closed under finite unions: if $A, B \in \mathcal{S}$, then $A \cup B \in \mathcal{S}$.*
- It should be clear that if $\mathcal{S}$ is also closed under finite intersections. This is because of De Morgan's law: $A \cap B = (A^c \cup B^c)^c$.
- **Definition 5.2 (σ-algebra).** *A collection of sets $\mathcal{S} \subseteq 2^S$ is called a* **σ-algebra on** *S if it satisifies the following properties.*
  1. *$\emptyset, S \in \mathcal{S}$.*
  2. *It is closed under complementation: if $A \in \mathcal{S}$, then $A^c = S - A \subseteq \mathcal{S}$.*
  3. *It is closed under countable unions: if $\{A_n : n \in \mathbb{N}\}$ is a countable collection of sets in $\mathcal{S}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{S}$.*

  An element of a σ-algebra $\mathcal{S}$ is called a **$\mathcal{S}$-measurable set** *or simply a* **measurable set** *when the context is clear.*
- Let $\mathcal{S}$ be a σ-algebra on S. One can easily show that, if $\{A_n : n \in \mathbb{N}\}$ be a countable collection of sets in $\mathcal{S}$, then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{S}$ as well. So, a σ-algebra is also closed under finite intersection.
- A σ-algebra is an algebra, but an algebra is not necessarily a σ-algebra.
- **Definition 5.3 (measurable space).** *A* **measurable space** *is a tuple $(S, \mathcal{S})$ where S is a set, and $\mathcal{S}$ is σ-algebra on S.*
- **Proposition 5.4 (intersection of σ-algebras).** *Let $\{\mathcal{S}_\alpha : \alpha \in A\}$ be a collection of σ-algebra on S. Then, $\bigcap_{\alpha \in A} \mathcal{S}_\alpha$ is also a σ-algebra on S.*
  - Note that there is no restriction on the collection $\{\mathcal{S}_\alpha : \alpha \in A\}$. It can be finite, countable, or even uncountable.
- **Definition 5.5 (generated σ-algebra).** *Let $\mathcal{A}$ be a non-empty collection of subsets of S. The* **σ-algebra generated by** *$\mathcal{A}$, denoted by $\sigma(\mathcal{A})$ is the smallest σ-algebra that contains $\mathcal{A}$. In other words,*

$$\sigma(\mathcal{A}) = \bigcap \Big\{ \tilde{\mathcal{A}} \subseteq 2^S : \mathcal{A} \subseteq \tilde{\mathcal{A}} \text{ and } \tilde{\mathcal{A}} \text{ is a } \sigma\text{-algebra} \Big\}.$$

- **Definition 5.6 (Borel σ-algebra).** *Let S be a topological space. The* **Borel σ-algebra on** *S, denoted by $\mathcal{B}(S)$, is the σ-algebra generated by the canonical topology (i.e., the collection of open sets) on S. An element of $\mathcal{B}(S)$ is called a* **Borel set**.

- **Proposition 5.7.** *The following statements are true.*
  1. *The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is generated by the collection of open intervals in $\mathbb{R}$.*
  2. *The Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ is generated by the collection of rectangles in $\mathbb{R}^d$.*
  3. *The Borel σ-algebra $\mathcal{B}(\overline{\mathbb{R}})$ is generated by the collection of open intervals in $\overline{\mathbb{R}}$.*

6 Sequences of Sets and Their Limits

- In this section, we are concerned with the sequence $\{A_n : n \in \mathbb{N}\}$ of sets.
- **Definition 6.1 (monotonically increasing sequence of sets).** *We write $A_n \uparrow A$ if*
  - *$A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$, and*
  - *$\bigcup_{n=1}^{\infty} A_n = A$.*
- **Definition 6.2 (monotoically decreasing sequence of sets).** *We write $A_n \downarrow A$ if*
  - *$A_n \supseteq A_{n+1}$ for all $n \in \mathbb{N}$, and*
  - *$\bigcap_{n=1}^{\infty} A_n = A$.*
- Let $(S, \mathcal{S})$ be a measurable space. If $A_n \in \mathcal{S}$ for all n, then it follows that

$$A_n \uparrow A \implies A \in \mathcal{S},$$
$$A_n \downarrow A \implies A \in \mathcal{S}.$$

- **Definition 6.3 (limit supremum and limit infemum).** *Define*

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m,$$
$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m.$$

- **Proposition 6.4 (limit supremum = infinitely often).**

$$\begin{aligned}\liminf_{n\to\infty} A_n &= \{a : \textit{for every } m \in \mathbb{N}, \textit{ there exists } n(a,m) \geq m \textit{ such that } a \in A_{n(a,m)}\} \\ &= \{a : a \in A_n \textit{ for infinitely many } n\}.\end{aligned}$$

- **Proposition 6.5 (limit infemum = eventually).**

$$\begin{aligned}\liminf_{n\to\infty} A_n &= \{a : \textit{there exists } m(a) \in \mathbb{N} \textit{ such that } a \in A_m \textit{ for all } m \geq m(a)\} \\ &= \{a : a \in A_m \textit{ for all large } m\textit{'s}\}.\end{aligned}$$

- Again, for a measureable space $(S, \mathcal{S})$, if $A_n \in \mathcal{S}$ for all n, we have that

$$\limsup_{n\to\infty} A_n = A \implies A \in \mathcal{S},$$
$$\liminf_{n\to\infty} A_n = A \implies A \in \mathcal{S}.$$

- **Definition 6.6 (indicator function).** *Let A be a set. The* **indicator function** $\mathbb{1}_A$ *is given by*

$$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

- **Definition 6.7 (limit of sequence of sets).** *For a sequence of sets $\{A_n : n \in \mathbb{N}\}$ such that each $A_n \subseteq S$, we write*

$$\lim_{n\to\infty} A_n = A$$

or simply

$$A_n \to A$$

if

$$\lim_{n\to\infty} \mathbb{1}_{A_n}(x) = \mathbb{1}_A(x)$$

for all $x \in S$.

- **Proposition 6.8 (limit, lim sup, and lim inf).** *If $A_n \to A$, then*

$$A = \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n.$$

- The last proposition implies that, for a measurable space $(S, \mathcal{S})$ such that $A_n \in \mathcal{S}$ for all n, it holds that

$$A_n \to A \implies A \in \mathcal{S}.$$

7 Measures

- **Definition 7.1 (measure).** *Let $\mathcal{S}$ be a σ-algebra on S. A* **measure** *is a function $\mu : \mathcal{S} \to [0, \infty]$ with the following properties.*

  1. *$\mu(\emptyset) = 0$.*
  2. *μ is countably additive: for any sequence $\{E_n : n \in \mathbb{N}\}$ of disjoint measurable sets, it holds that*

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

- **Proposition 7.2 (properties of measures).** *Let μ be a measure defined on a σ-algebra $\mathcal{S}$.*

  (a) $\mu(\emptyset) = 0$.
  (b) If $E, F \in \mathcal{S}$ and $E \subseteq F$, then $\mu(E) \leq \mu(F)$. If $\mu(E) < \infty$, then $\mu(F - E) = \mu(F) - \mu(E)$.

  Moreover, let $\{E_n : n \in \mathbb{N}\}$ be a sequence of sets such that $E_n \in \mathcal{S}$ for all n.

  (c) If $E_n \uparrow E$, then $\lim_{n\to\infty} \mu(E_n) = \mu(E)$
  (d) If $E_n \downarrow E$ and $\mu(E_1) < \infty$, then $\lim_{n\to\infty} \mu(E_n) = \mu(E)$.

- **Definition 7.3 (limit supremum and limit infemum).** *Let $\{a_n : n \in \mathbb{N}\}$ be sequence of real numbers. Then,*

$$\limsup_{n\to\infty} a_n = \inf_{n\in\mathbb{N}} \sup_{m\geq n} a_m,$$
$$\liminf_{n\to\infty} a_n = \sup_{n\in\mathbb{N}} \inf_{m\geq n} a_m.$$

- **Theorem 7.4 (Fatou's lemma for sets).**

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

- **Definition 7.5 (finite measure).** *If $\mu(E) < \infty$ for all $E \in \mathcal{S}$, we say that μ is* **finite**.

- **Theorem 7.6 (reverse Fatou lemma).** *Let μ be a finite measure. For a sequence of measurable sets $\{E_n : n \in \mathbb{N}\}$, we have that*

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

  We emphasize that this also works for finite measures.

- **Corollary 7.7.** *Let μ be a finite measure. Let $\{E_n : n \in \mathbb{N}\}$ be a sequence of measurable sets such that $E_n \to E$. Then,*

$$\lim_{n\to\infty} \mu(E_n) = \mu(E).$$

  Again, this only works for finite measures.

- **Definition 7.8 (probability measure).** *A* **probability measure** *is a finite measure with $\mu(S) = 1$.*

- **Definition 7.9 (measure space).** *A* **measure space** *is a triple $(S, \mathcal{S}, \mu)$ where S is a non-empty set, $\mathcal{S}$ is a σ-algebra on S, and μ is a measure on $\mathcal{S}$.*

- **Definition 7.10 (probability space).** *A* **probability space** *is a triple $(\Omega, \mathcal{E}, P)$ where Ω is a non-empty set, $\mathcal{E}$ is a σ-algebra on Ω, and P is a probability measure on Σ. We often call Ω the* **sample space**. *An element of $\mathcal{E}$ is called an* **event**. *If E is an event, $P(E)$ is referred to as the* **probability** *of E.*

8 Measure Zero

- **Definition 8.1 (measure zero, null set, and almost everywhere).** *In a measure space $(S, \mathcal{S}, \mu)$, a set $N \in \mathcal{S}$ is set to be of* **measure zero** *or a* **null set** *if $\mu(N) = 0$. A property that holds on N^c is said to hold μ-***almost everywhere**. *In the context where μ is clear, we says that a property holds just* **almost everywhere**.

- **Definition 8.2 (completeness).** *A measure space $(S, \mathcal{S}, \mu)$ is said to be* **complete** *if every subset of a set of measure zero is also measurable.*

- **Definition 8.3 (completion).** *Let $(S, \mathcal{S}, \mu)$ be a measure space. The μ-***completion** *of $(S, \mathcal{S}, \mu)$ is the tuple $(S, \overline{S}, \overline{\mu})$ where*

$$\overline{S} = \{A \cup M : A \in \mathcal{S}, M \subseteq N \text{ where } N \in \mathcal{S} \text{ and } \mu(N) = 0\}, \text{ and } \overline{\mu}(A \cup M) = \mu(A).$$

- **Theorem 8.4.** *The μ-completion $(S, \overline{S}, \overline{\mu})$ of $(S, \mathcal{S}, \mu)$ is a complete measure space.*

- In a probability space $(\Omega, \mathcal{E}, P)$, if N is a null set and $E = N^c$, then we have that $P(E) = 1$.

- **Definition 8.5 (almost surely).** *A property that is true on a event E such that $P(E) = 1$ is said to be true* **almost surely** *or* **with probability 1**.

- **Proposition 8.6.** *If $E_n \in \mathcal{E}$ and $P(E_n) = 1$ for all n, then $P(\bigcap_{n=1}^{\infty} E_n) = 1$.*

- **Theorem 8.7 (first Borel–Cantelli lemma).** *Let $\{E_n : n \in \mathbb{N}\}$ be a sequence of events such that $\sum_{n=1}^{\infty} P(E_n) < \infty$. Then,*

$$P\left(\limsup_{n\to\infty} E_n\right) = 0.$$

9 Conditional Probability and Independence

- In this section, we work with a probability space $(\Omega, \mathcal{E}, P)$.
- **Definition 9.1 (independence).** *Two events E and F are* **independent** *if $P(E \cap F) = P(E)P(F)$. A collection of events $\{E_\alpha : \alpha \in A\}$ is an* **independent collection** *if, for every finite subset B of A, we have that*

$$P\left(\bigcap_{\beta \in B} E_\beta\right) = \prod_{\beta \in B} P(E_\beta).$$

- **Proposition 9.2.** *If E and F are independent, so are*
  1. *E and F^c,*
  2. *E^c and F, and*
  3. *E^c and F^c.*
- **Definition 9.3 (conditional probability).** *Let E and F be events such that $P(F) > 0$. The* **condition probability of E given F** *is*

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

- **Proposition 9.4.** *Suppose that $P(F) > 0$.*
  - *E and F are independent if and only if $P(E|F) = P(E)$.*
  - *The mapping $E \mapsto P(E|F)$ defines a new probability measure on $\mathcal{E}$.*
- **Proposition 9.5.** *If $E_1, E_2, \ldots, E_n \in \mathcal{E}$ and $P(E_1 \cap E_2 \cap \cdots \cap E_n) > 0$, then*

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \cdots P(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}).$$

- **Definition 9.6 (partition).** *A collection of events $\{E_\alpha : \alpha \in A\}$ is called a* **partition** *of Ω if*
  1. *the events are pairwise disjoint, and*
  2. *$\bigcup_{\alpha \in A} E_\alpha = \Omega$.*
- **Proposition 9.7 (partition equation).** *Let $\{E_n\}$ be a finite or countable partition of Ω. Then, if $E \in \mathcal{E}$, then*

$$P(E) = \sum_n P(E|E_n)P(E_n).$$

- **Theorem 9.8 (Bayes').** *Let $\{E_n\}$ be a finite or countable partition of Ω. Let E be an event such that $P(E) > 0$. Then, for any n,*

$$P(E_n|E) = \frac{P(E|E_n)P(E_n)}{\sum_m P(E|E_m)P(E_m)}.$$

10 Probabilities on a Finite or Countable Space

- Now, we assume that Ω is finite or countable.
- In this case, 2^Ω is a σ-algebra. So, we naturally take $\mathcal{E} = 2^\Omega$.
- **Definition 10.1 (atom).** *An* **atom** *is a set $\{\omega\}$ where $\omega \in \Omega$. We denote the probability of an atom $P(\{\omega\})$ by p_ω or $P(\omega)$.*
- A probability on a finite or countable set Ω is characterized by its values on the atoms.

  Theorem 10.2. *Let $\{p_\omega : \omega \in \Omega\}$ be a collection of real numbers indexed by members of Ω. Then, there exists a unique probability measure P such that $P(\{\omega\}) = p_\omega$ if and only if (1) $p_\omega \geq 0$ for all ω, and (2) $\sum_{\omega\in\Omega} p_\omega = 1$. In particular, we have that*

  $$P(E) = \sum_{\omega\in E} p_\omega.$$

  for any $E \in \mathcal{E}$.
- **Definition 10.3 (uniform probability measure).** *A probability measure P on a finite set Ω is called* **uniform** *if $p_\omega = P(\{\omega\})$ does not depend on ω.*
- With a uniform probability measure, we have that

  $$P(\{\omega\}) = \frac{1}{\#(\Omega)}.$$

  Moreover,

  $$P(E) = \frac{\#(E)}{\#(\Omega)}.$$

11 Random Variables on a Countable Space

- Again, we assume that Ω is countable and $\mathcal{E} = 2^\Omega$.
- **Definition 11.1 (random variable, countable case).** *A* **random variable** *is a function $\Omega \to T$ where T is a set.*
- We typically denote a random variable by uppercase letters such as X, Y, and Z.
- **Definition 11.2 (image).** *The* **image** *of X is the set $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$.*
- Because Ω is countable, the image $X(\Omega)$ is either finite or countably infinite even if T is uncountable.
- **Definition 11.3 (preimage).** *Let $A \subseteq T$. The* **preimage of A under X** *is the set*

  $$X^{-1}(A) = \{\omega : X(\omega) \in A\}.$$
- **Definition 11.4 (distribution of a random variable).** *The* **distribution of X** *is the function $P_X : 2^{X(\Omega)} \to [0, 1]$ defined by*

  $$P_X(A) = P(X^{-1}(A))$$

  for all $A \subseteq X(\Omega)$.

- We sometimes write $P_X(A)$ as $P(X \in A)$.
- **Theorem 11.5.** *Let X be a random variable. Then, P_X is a probability measure on $2^{X(\Omega)}$. It is completely determined by the collection of numbers $\{p_{X,x} : x \in X(\Omega)\}$ where*

$$p_{X,x} = P(X = x) = \sum_{\{\omega : X(\omega) = x\}} P(\omega).$$

Moreover,

$$P_X(A) = \sum_{a \in A} P(X = a).$$

12 Expectations on a Countable Space

- **Definition 12.1 (expectation).** *Let X be a real-valued random variable on a countable space Ω. (In other words, $X : \Omega \to \mathbb{R}$.) The expectation of X, denoted by $E[X]$, is defined to be*

$$E[X] = \sum_{\omega \in \Omega} X(\omega)P(\omega),$$

provided that the sum make sense: that is, when the sum evaluates to the same value even after exchanging the order of the terms. This happens when Ω is finite or when the sum is absolutely convergent: $\sum_{\omega \in \Omega} |X(\omega)|P(\omega) < \infty$.

- **Proposition 12.2 (linearity of expectation).** *The operator E is linear. In other words, if X and Y be real-valued random variables on a countable space, then we have that*
  - *$E[cX] = cE[X]$ for any $c \in \mathbb{R}$, and*
  - *$E[X + Y] = E[X] + E[Y]$.*
- **Proposition 12.3.** *Let X and Y be real-valued random variables on a countable space Ω. If (1) $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$ and (2) $E[X]$ and $E[Y]$ are finite, then $E[X] \leq E[Y]$.*
- **Theorem 12.4 (law of the unconcious statistician, aka LOTUS).** *Let $g : \mathbb{R} \to \mathbb{R}$ be an arbitrary function and X be a real-valued random variable on a countable space Ω. Then,*

$$E[g(X)] = \sum_{\omega \in \Omega} g(X(\omega))P(\omega) = \sum_{x \in X(\Omega)} g(x)P(X = x).$$

- LOTUS implies that

$$E[X] = \sum_{x \in X(\Omega)} xP(X = x).$$

- **Proposition 12.5.** *If $X = \mathbb{1}_A$ is the indicator function of an event A, then $E[X] = P(A)$.*
- **Theorem 12.6.** *Let $h : \mathbb{R} \to [0, \infty]$ be a non-negative function. Let X be a real-valued random variable. Then,*

$$P(\{\omega : h(X(\omega)) \geq a\}) \leq \frac{E[h(X)]}{a}$$

for all $a > 0$.

- **Corollary 12.7 (Markov's inequality).**

$$P(|X| \geq a) \leq \frac{E[|X|]}{a}$$

for all $a > 0$.

- **Definition 12.8 (the space $\mathcal{L}^1$).** *The space of real-valued random variables which have finite expectations is called* $\mathcal{L}^1$.

- By linearity of expectation, we have that $\mathcal{L}^1$ is a vector space.

- **Proposition 12.9.** *If* $E[X^2] < \infty$*, then so is* $E[|X|]$ *and* $E[X]$*. In other words,* $X^2 \in \mathcal{L}^1 \implies X \in \mathcal{L}^1$.

- **Definition 12.10 (variance).** *Let* X *be a random variable such that* $E[X^2] \in \mathcal{L}^1$*. The* **variance of** X*, denoted by* $\mathrm{Var}(X)$*, is defined to be*

$$\mathrm{Var}(X) = E[(X - E[X])^2].$$

- We can show that $\mathrm{Var}(X) = E[X^2] - (E[X])^2$.

- **Definition 12.11 (standard deviation).** *The* **standard deviation of** X*, denoted by* $\mathrm{Stdev}(X)$*, is defined to be the non-negative square root of* $\mathrm{Var}(X)$.

- A more common notation for the standard deviation is $\sigma(X)$. However, we will use this notation for the σ-algebra generated by X (Definition 19.2) in this note.

- **Theorem 12.12 (Chebyshev's inequality).** *If* $X^2 \in \mathcal{L}^1$*, then we have that, for all* $a > 0$*,*

$$P(|X| \geq a) \leq \frac{E[X^2]}{a^2},$$
$$P(|X - E[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

13 Construction of Measures on Uncountable Space

- In this section, we assume that the set S is uncountable.

- Given a σ-algebra $\mathcal{S}$ on S, it is very hard to define a measure on $\mathcal{S}$ from scratch.

- Instead, we follow the following strategy.
  - Come up with a collection of sets $\mathcal{S}_0$ on S such that $\mathcal{S} \subseteq \sigma(\mathcal{S}_0)$.
  - Show that S_0 is a structure called a "semi-ring."
  - Define a measure-like function μ on $\mathcal{S}_0$.
    - Since $\mathcal{S}_0$ is not a σ-algebra, we call μ a "pre-measure" instead of a measure.
  - "Extend" μ to another function μ^* so that μ^* also works on sets in $\sigma(\mathcal{S}_0) - \mathcal{S}_0$.
    - Actually, we identify a collection $\mathcal{S}_0^*$ of sets on which μ^* are defined.
    - The sets in the collection are called "μ^*-measureable sets."
    - We then show that $\sigma(\mathcal{S}_0) \subseteq \mathcal{S}_0^*$.
  - Show that μ^* is an object called an "outer measure".
  - Apply "Carathéodory extension theorem."

    - It states that, if μ^* is an outer measure defined on a collection of μ^*-measurable sets $\mathcal{S}_0^*$, then (1) $\mathcal{S}_0^*$ is a σ-algebra, and (2) μ^* is a measure on $\mathcal{S}_0^*$.
  - μ^* is the measure on $\mathcal{S}$ that we sought.

- **Definition 13.1 (semi-ring).** *Let S be a set. A collection of sets $\mathcal{S}_0$ of subsets S is called a* **semi-ring on** *S if it satisfies the following properties.*

  (a) $\emptyset \in \mathcal{S}_0$.

  (b) If $A, B \in \mathcal{S}_0$, *then* $A \cap B \in \mathcal{S}_0$ *as well.*

  (c) For any $A, B \in \mathcal{S}_0$, *there exists a finite collection of disjoint sets* $C_1, C_2, \ldots, C_n \in \mathcal{S}_0$ *such that* $A - B = \cup_{i=1}^n C_i$.

- An algebra (Definition 5.1) is always a semi-ring. This is because an algebra is closed under finite intersection. Moreover, it is also closed under set difference because $A - B = A \cup B^c$, and so property (c) is automatically satisfied.

- **Definition 13.2 (pre-measure).** *Let $\mathcal{S}_0$ a collection of subsets of S. A* **pre-measure on** *$\mathcal{S}_0$ is a function $\mu : \mathcal{S}_0 \to [0, \infty]$ that satisfies the following properties.*

  1. $\mu(\emptyset) = 0$.
  2. *It is countably additive. That is, for any sequence $\{E_n \in \mathcal{S}_0 : n \in \mathbb{N}\}$ of disjoint sets such that $\bigcup_{n=1}^\infty E_n \in \mathcal{S}_0$, we have that*

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

  Note that a pre-measure has all the properties that a measure has with the exception that $\mathcal{S}_0$ is not a σ-algebra. (This is why we have to say that $\bigcup_{n=1}^\infty E_n \in \mathcal{S}_0$.)

- **Definition 13.3 (extension).** *Let $\mathcal{S}_0$ be a semi-ring on S and μ be a pre-measure on $\mathcal{S}$. The* **extension of** *μ is the function $\mu^* : 2^S \to [0, \infty]$ such that*

$$\mu^*(E) = \inf\left\{\sum_{n=1}^\infty \mu(E_n) : \{E_n\}_{n\in\mathbb{N}} \text{ is a sequence of sets in } \mathcal{S}_0 \text{ such that } E \subseteq \bigcup_{n=1}^\infty E_n\right\}$$

  if E can be covered by a countable number of sets in $\mathcal{S}_0$. Otherwise, we set $\mu^*(E) = \infty$.

- We note that μ^* is defined on all $E \in \sigma(\mathcal{S}_0)$. This is because $\mathcal{S}_0$ is an algebra, so $S \in \mathcal{S}_0$. Hence, for every $E \subseteq S$, there is at least one cover: the one that uses S.

- **Definition 13.4 (outer measure).** *Let S be a set. An* **outer measure on** *S is a function $\mu^* : 2^S \to [0, \infty]$ that satisfies the following properties.*

  1. $\mu^*(\emptyset) = 0$.
  2. *If $E \subseteq F \subseteq S$, then $\mu^*(E) \leq \mu^*(F)$.*
  3. *μ^* is countably subadditive. That is, if $E_n : n\mathbb{N}$ is a countable collection of subsets of S, then*

$$\mu^*\left(\bigcup_{n=1}^\infty E_n\right) \leq \sum_{n=1}^\infty \mu^*(E_n).$$

- **Theorem 13.5 (outer measure from pre-measure).** *Let $\mathcal{S}_0$ be semi-ring on S. Let μ_0 be a pre-measure on $\mathcal{S}_0$. Let μ^* be the extension of μ. Then, μ^* is an outer measure on S.*

- **Definition 13.6 (μ^*-measurable).** *Let μ^* be an outer measure on S. A subset E of S is said to be* **μ^*-measurable** *if*

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$$

for every subset A of S.

- A μ^*-measurable set E splits any set A into pieces whose outer measures add up to the outer measure of A. In other words, a set is μ^*-measurable if it splits other sets in a "nice" way.

- **Theorem 13.7 (Carathédory extension theorem).** *Let $\mathcal{S}_0$ be a semi-ring on S, μ be a pre-measure on $\mathcal{S}_0$, and μ^* be the extension of μ. Then, the collection of μ^*-measurable sets is a σ-algebra that contains $\sigma(\mathcal{S}_0)$, and μ^* is a measure on this collection.*

- **Proposition 13.8 (completeness of extension).** *The μ^* measure constructed with the Carathéodory extension theorem is complete.*

14 Uniqueness of Measure

- **Definition 14.1 (σ-finite measure).** *Let $\mathcal{S}$ be a σ-algebra on S, and μ be a measure on $\mathcal{S}$. If there exists a sequence $\{E_n : n \in \mathbb{N}\}$ of sets in $\mathcal{S}$ with $\bigcup_{n=1}^{\infty} E_n = S$ and such that $\mu(E_n) < \infty$ for all n, then we say that μ is* **σ-finite**.

- We can define a **σ-finite pre-measure** in a similar way: just take $\mathcal{S}$ to be an arbitrary collection of sets instead of an algebra.

- **Definition 14.2 (Hahn extension theorem).** *Let $\mathcal{S}_0$ be a semi-ring on S and μ is a pre-measure on $\mathcal{S}_0$. If μ is σ-finite, then he extension μ^* of μ is the unique measure $\sigma(\mathcal{S}_0)$ such that $\mu^*(E) = \mu(E)$ for all $E \in \mathcal{S}_0$.*

- When the measure is finite, however, we can show that the measure is unique with much less machinery.

- **Definition 14.3 (π-system).** *Let S be a set. A collection $\mathcal{S}$ of subsets of S is called a* **π-system** *if it closed under finite intersection: if $A, B \in \mathcal{S}$, then $A \cap B \in \mathcal{S}$.*

- It should be clear that a semi-ring is a π-system.

- **Theorem 14.4 (uniqueness of finite measure).** *Let $\mathcal{S}$ be a π-system on S that contains S. Let μ_1 and μ_2 be two finite measures on $\sigma(\mathcal{S})$ such that $\mu_1(E) = \mu_2(E)$ for all $E \in \mathcal{S}$. Then, it is true that $\mu_1(E) = \mu_2(E)$ for all $E \in \sigma(\mathcal{S})$ too.*

15 Lebesgue Measures

- In this section, we constrruct a measure on $\mathbb{R}$ and another one on $\mathbb{R}^d$.

- **Definition 15.1.** *Let $\mathcal{F}$ be a collection of subsets of $\mathbb{R}$ that contains all intervals of the forms*

$$(a, b], (-\infty, b], (a, \infty), \text{ and } (-\infty, \infty)$$

and all their finite unions.

- **Proposition 15.2.** *$\mathcal{F}$ is an algebra (and so a semi-ring) on $\mathbb{R}$.*

- **Proposition 15.3.** *$\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{F})$.*

- **Definition 15.4.** *Define the* **length function** ℓ *to be a function* $\mathcal{F}$ *to* $[0,\infty]$ *with the following properties.*
  - $\ell((a,b]) = b - a$.
  - $\ell((-\infty,b]) = \ell((a,\infty)) = \ell((-\infty,\infty)) = \infty$.
  - *For* $E = E_1 \cup E_2 \cup \cdots$ *where each* E_i *is an interval of the four forms listed in* $\mathcal{F}$*'s definition and any two* E_i *and* E_j *are disjoint, we have that*

$$\ell(E) = \sum_{i=1}^{\infty} \ell(E_i).$$

- **Proposition 15.5.** ℓ *is a* σ*-finite pre-measure on* $\mathcal{F}$.
- **Definition 15.6.** *The* **Lebesgue measure on** $\mathbb{R}$ *is the extension* ℓ^* *of* ℓ.
- Using previous results, the Lebesgue measure ℓ^* the unique measure on $\mathcal{B}(\mathbb{R})$ that agrees with the natural notion of length of intervals.
- The construction above can be extended to $\mathbb{R}^d$.
- **Definition 15.7.** *Let* $\mathcal{F}_*^d = \{\mathcal{I}_1 \times \mathcal{I}_2 \times \cdots \times \mathcal{I}_d : \mathcal{I}_k$ *is one of the four forms in Definition 15.1*$\}$*. Let* $\mathcal{F}^d$ *be the smallest set that contains* $\mathcal{F}_*^d$ *and all their finite unions.*
- **Proposition 15.8.** $\mathcal{F}^d$ *is an algebra (and so a semi-ring) on* $\mathbb{R}^d$*, and* $\mathcal{B}(\mathbb{R}^d) \subseteq \sigma(\mathcal{F}^d)$.
- **Definition 15.9.** *The* **volume function v** *is a function from* $\mathcal{F}^d$ *to* $[0,\infty]$ *with the following properties.*
  - *If* $\mathcal{I}_1 \times \mathcal{I}_2 \times \cdots \times \mathcal{I}_d \in \mathcal{F}_*^d$*, then*

$$v(\mathcal{I}_1 \times \mathcal{I}_2 \times \cdots \times \mathcal{I}_d) = \ell(\mathcal{I}_1)\ell(\mathcal{I}_2)\cdots\ell(\mathcal{I}_d).$$

  - *If* $\mathcal{R} = \bigcup_{i=1}^{\infty} \mathcal{R}_i$ *where each* $\mathcal{R}_i$ *is an element of* $\mathcal{F}_*^d$ *and any two* $\mathcal{R}_i$ *and* $\mathcal{R}_j$ *are disjoint, then*

$$v(\mathcal{R}) = \sum_{i=1}^{\infty} v(\mathcal{R}_i).$$

- **Proposition 15.10.** v *is a* σ*-finite pre-measure on* $\mathcal{F}^d$.
- **Definition 15.11.** *The* **Lebesgue measure on** $\mathbb{R}^d$ *is the extension* v^* *of* v.
- **Proposition 15.12.** *The Lebesgue measure, when restricted to the Borel* σ*-algebra* $\mathcal{B}([0,1])$*, is a probability measure.*

16 Probability Measures on $\mathbb{R}$

- We just saw that there's a probability measure on $\mathcal{B}([0,1])$. In this section, we will constructor probability measures on $\mathcal{B}(\mathbb{R})$.
- **Definition 16.1 (CDF).** *A function* $F : \mathbb{R} \to [0,1]$ *is called a* **cumulative distribution function (CDF)** *if it satisfies the following properties.*
  1. F *is non-decreasing.*
  2. F *is right continuous. This means that* $\lim_{y \to x^+} F(y) = F(x)$ *for all* $x \in \mathbb{R}$.

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

- We note that such a function exists, and we will list a number of them at the end of this section.
- **Definition 16.2.** *Let $\mathcal{F}$ be defined as in Definition 15.1. Let F be a cumulative distribution function. Define $P_0 : \mathcal{F} \to [0,1]$ so that the following properties are satisifed.*

  1. $P_0((a,b]) = F(b) - F(a)$.
  2. $P_0((-\infty,b]) = F(b)$.
  3. $P_0((a,\infty)) = 1 - F(a)$.
  4. $P_0((-\infty,\infty)) = 1$.
  5. *Let E_1, E_2, ... be disjoint intervals where each E_i is of one of the forms in Definition 15.1, then* $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$.

- **Theorem 16.3 (existence of probability distribution on $\mathbb{R}$).** *P_0 is a finite pre-measure on $\mathcal{F}$. As result, its extension P to $\mathcal{B}(\mathbb{R})$ is a unique probability measure on $\mathcal{B}(\mathbb{R})$ such that $P(E) = P_0(E)$ for all $E \in \mathcal{F}$.*
- **Definition 16.4 (CDF of a probability measure).** *Let P be a probability measure on $\mathcal{B}(\mathbb{R})$. The* **cumulative distribution function (CDF) of** *P is the function $F : \mathbb{R} \to [0,1]$ such the*

$$F(x) = P((-\infty, x])$$

for all $x \in \mathbb{R}$.

- **Proposition 16.5.** *The CDF of a probability measure P is a CDF. That is, it satisfies all the properties in Definition 16.1.*
- This means that a function is a CDF if and only if it is a CDF of a probability distribution.
- **Proposition 16.6.** *A probability measure on $\mathcal{B}(\mathbb{R})$ is completely characterized by its CDF. More precisely, let P and Q be two probability measures on $\mathcal{B}(\mathbb{R})$. Let F_P and F_Q be the CDFs of P and Q, respectively. Then, $F_P = F_Q \implies P = Q$.*
- **Proposition 16.7.** *Let P be a probability measure on $\mathcal{B}(\mathbb{R})$ and let F be its CDF. Let*

$$F(x^-) = \lim_{u\to x^-} F(u).$$

Then, we have that, for any $x < y$,

$$\begin{aligned} P((x,y]) &= F(y) - F(x) \\ P([x,y]) &= F(y) - F(x^-) \\ P((x,y)) &= F(y^-) - F(x) \\ P([x,y)) &= F(y^-) - F(x^-) \\ P(\{x\}) &= F(x) - F(x^-). \end{aligned}$$

As a result, $P(\{x\}) = 0$ if and only if F is continuous at x.

- **Definition 16.8 (PDF).** *A function $f : \mathbb{R} \to \mathbb{R}$ that is non-negative and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ is called a* **probability density function (PDF)**.[1]
- Given a PDF f, it follows that $\int_{-\infty}^{x} f(u)\,\mathrm{d}u$ is a CDF. So, a PDF gives rise to a probability measure on $\mathcal{B}(\mathbb{R})$, and it makes sense to talk about the PDF of a probability measure.

[1] For now, it suffice to say that the function in Riemann integrable. However, this also apply to Lebesgue integrable functions, which we have not defined yet.

16.1 Examples of Probabilty Measures on $\mathbb{R}$

- The **step function**

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

is a CDF. The associated probability measure

$$P(E) = \begin{cases} 1, & a \in E \\ 0, & a \notin E \end{cases}$$

is called the **Dirac measure**. The PDF of this measure is the Dirac delta function $\delta(x-a)$.

- The function

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{otherwise} \end{cases}$$

is a PDF called the **uniform distribution on** $[a, b]$.

- Given $\beta > 0$, the function

$$f(x) = \begin{cases} \beta e^{-\beta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

is a PDF called the **exponential distribution with parameter** β.

- Given $\mu \in \mathbb{R}$ and $\sigma > 0$, the function

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

is a PDF called the **Gaussian distribution with mean** σ **and variance** σ^2.

17 Measurable Functions

- **Definition 17.1 (preimage).** *Let $f : S \to T$ be any function. The* **preimage of** *f is the function $f^{-1} : 2^T \to 2^S$ defined by*

$$f^{-1}(\mathtt{T}) = \{s : s \in S, f(s) \in \mathtt{T}\}$$

for any $\mathtt{T} \subseteq T$.

- **Definition 17.2 (measurable function).** *Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be two measurable spaces. A function $f : S \to T$ is* **$\mathcal{S}/\mathcal{T}$-measurable** *if $f^{-1}(\mathtt{T}) \in \mathcal{S}$ for all $\mathtt{T} \in \mathcal{T}$.*

  When the context is clear, however, we simply say f is "$\mathcal{S}$-measurable" or just "measurable."

- If f is $\mathcal{S}/\mathcal{T}$-measurable, we can summarize it with a diagram as follows:

$$S \xrightarrow{f} T$$

$$\mathcal{S} \xleftarrow{f^{-1}} \mathcal{T}$$

- **Proposition 17.3.** *Let $f : S \to T$ be an arbitrary function. The preimage function f^{-1} preserves all relevant set operations. That is, it holds that:*

  1. *$f^{-1}(\mathtt{T}^c) = (f^{-1}(\mathtt{T}))^c$, and*
  2. *For any countable sequence $\{\mathtt{T}_n : n \in \mathbb{N}\}$ such that $\mathtt{T}_n \in \mathcal{T}$ for each n, we have that*

$$f^{-1}\left(\bigcup_{n=1}^{\infty} \mathtt{T}_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(\mathtt{T}_n).$$

- **Theorem 17.4.** *Let $\mathcal{U}$ be a collection of subsets of T such that $\sigma(\mathcal{U}) = \mathcal{T}$. A function $f : S \to T$ is a $\mathcal{S}/\mathcal{T}$-measruable function if and only if $f^{-1}(\mathtt{U}) \in \mathcal{S}$ for all $\mathtt{U} \in \mathcal{U}$.*

- The above theorem tells us that, in order to show that a function f is measurable, it suffices to show that it is "measurable" on a collection $\mathcal{U}$ which generates $\mathcal{T}$.

  – For example, to show that f is $\mathcal{S}/\mathcal{B}(\mathbb{R})$-measurable, it suffices to show that it is $\mathcal{S}/\mathcal{F}$-measurable where $\mathcal{F}$ is defined as in Definition 15.1.

- **Theorem 17.5 (composition of measurable functions is measurable).** *Let $(S, \mathcal{S})$, $(T, \mathcal{T})$, and $(U, \mathcal{U})$ be measurable spaces. Let $f : S \to T$ be a $\mathcal{S}/\mathcal{T}$-measurable function and $g : T \to U$ be a $\mathcal{T}/\mathcal{U}$-measurable function. Then, the composition $g \circ f$ is a $\mathcal{S}/\mathcal{U}$-measurable function.*

- Here's the diagram of the situation in the theorem above.

$$S \xrightarrow{f} T \xrightarrow{g} U$$

$$\mathcal{S} \xleftarrow{f^{-1}} \mathcal{T} \xleftarrow{g^{-1}} \mathcal{U}$$

- **Definition 17.6 (Borel space).** *Let S be a set endowed with a topology. The measurable space $(S, \mathcal{B}(S))$ is called the* **Borel space**.

- **Definition 17.7 (Borel function).** *Let $(S, \mathcal{B}(S))$ and $(T, \mathcal{B}(T))$ be two Borel spaces. We call a $\mathcal{B}(S)/\mathcal{B}(T)$-measurable function an S/T-***Borel function*** or simply a* **Borel function**.

- **Proposition 17.8 (continuous functions are Borel).** *Let $(S, \mathcal{B}(S))$ and $(T, \mathcal{B}(T))$ be two Borel spaces. Any continuous function $f : S \to T$ is Borel.*

  This comes from the fact that, for any continuous function, the preimage of an open set is open.

18 Measurable Real-Valued Functions

- In this section, we shall fix $(T, \mathcal{T})$ to be $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ or $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. We will indicate the range by saying that the function is "real-valued" or "extended real-valued".

- **Proposition 18.1.** *A function $f : \mathcal{S} \to \mathbb{R}$ is $\mathcal{S}$-measurable if and only if, for all $a \in \mathbb{R}$, the set $f^{-1}((-\infty, a])$ is $\mathcal{S}$-measurable.*

  Note that the proposition is holds if we replace $(-\infty, a]$ with $(-\infty, a)$, (a, ∞) or $[a, \infty)$.

- Examples of $\mathcal{S}$-measurable real-valued functions:

  – The constant function $f(x) = c$ for some $c \in \mathbb{R}$.

  – The identity function $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x$. Here, the domain $(S, \mathcal{S})$ is also $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

- **Proposition 18.2.** *An indicator function $\mathbb{1}_A$ is $\mathcal{S}$-measurable if and only if $A \in \mathcal{S}$.*

- **Proposition 18.3.** *If f and g are $\mathcal{S}$-measurable real-valued functions and $c \in \mathbb{R}$, then*

$$cf, \qquad f^2, \qquad f+g, \qquad fg, \qquad |f|, \qquad 1/f, \qquad \min(f,g), \qquad \max(f,g)$$

are also $\mathcal{S}$-measurable. For the case of $1/f$, we assume that $f(x) \neq 0$ for all $x \in S$.

- It is convenient to work with extended real-valued function $f : S \to \overline{\mathbb{R}}$. This is because, if we have a sequence $\{f_n : n \in \mathbb{N}\}$ where $f_n : S \to \mathbb{R}$, we also have that $\lim f_n$, $\limsup f_n$, and $\liminf f_n$ are also functions from S to $\overline{\mathbb{R}}$ if the limits exist.

- **Definition 18.4.** *The collection of all extended real-valued $\mathcal{S}$-measurable functions is denoted by $M(S,\mathcal{S})$. The collection of non-negative functions in $M(S,\mathcal{S})$ is denoted by $M^+(S,\mathcal{S})$.*

- **Proposition 18.5.** *A function $f : S \to \overline{\mathbb{R}}$ is $\mathcal{S}$-measurable if and only if $f^{-1}([-\infty,a]) \in \mathcal{S}$ for all $a \in \mathbb{R}$.*

- Observe that

$$\{\infty\} = \bigcap_{n=1}^{\infty} [-\infty, n]^c,$$
$$\{-\infty\} = \bigcap_{n=1}^{\infty} [-\infty, -n].$$

  So,

$$f^{-1}(\{\infty\}) = f^{-1}\left(\bigcap_{n=1}^{\infty} [-\infty,n]^c\right), = \bigcap_{n=1}^{\infty} \left(f^{-1}([-\infty,n])\right)^c$$
$$f^{-1}(\{-\infty\}) = f^{-1}\left(\bigcap_{n=1}^{\infty} [-\infty,-n]\right) = \bigcap_{n=1}^{\infty} f^{-1}([-\infty,-n]).$$

- **Proposition 18.6.** *An extended real-valued function $f : S \to \overline{\mathbb{R}}$ is $\mathcal{S}$-measurable if and only if the sets $f^{-1}(\{-\infty\})$ and $f^{-1}(\{\infty\})$ belong to $\mathcal{S}$ and the real-valued function f_0 defined by*

$$f_0(x) = \begin{cases} f(x), & f(x) \notin \{-\infty,\infty\} \\ 0, & f(x) \in \{-\infty,\infty\} \end{cases}$$

is $\mathcal{S}$-measurable.

- As a consequence of the above proposition, if f and g are measurable extended-real valued functions and $c \in \mathit{Real}$, then

$$cf, \qquad f^2, \qquad fg, \qquad |f|, \qquad 1/f, \qquad \min(f,g), \qquad \max(f,g)$$

  are also measurable with the usual caveat that f should not be 0 when assessing the measurability of $1/f$.

- The measurability of $f+g$ needs greater care because we cannot say anything about it if there is an x where $f(x) = \pm\infty$ and $g(x) = \mp\infty$. Otherwise, $f+g$ is measurable given that f and g are measurable.

- **Proposition 18.7.** *Let $\{f_n : n \in \mathbb{N}\}$ be a sequence of functions in $M(S, \mathcal{S})$. Then, all of the functions*

$$\begin{aligned}
\underline{f}(x) &= \inf_{n\geq 1} f_n(x), \\
\overline{f}(x) &= \sup_{n\geq 1} f_n(x), \\
\underline{F}(x) &= \liminf_{n\in\mathbb{N}} f_n(x) = \sup_{n\geq 1}\left\{\inf_{m\geq n} f_m(x)\right\}, \\
\overline{F}(x) &= \limsup_{n\in\mathbb{N}} f_n(x) = \inf_{n\geq 1}\left\{\sup_{m\geq n} f_m(x)\right\}.
\end{aligned}$$

also belong to $M(S, \mathcal{S})$. Moreover, if $\{f_n : n \in \mathbb{N}\}$ converges to a function f, then $f \in M(S, \mathcal{S})$.

19 Random Variables on an Uncountable Spaces

- **Definition 19.1 (random variable).** *Let the domain $(S, \mathcal{S})$ be a probability space $(\Omega, \mathcal{E})$ endowed with probability measure P. An $\mathcal{E}$-measurable function $X : \Omega \to T$ is called a* **random variable**.
- Now, the diagram for a random variable X is as follows.

$$\begin{array}{ccccc}
 & & \Omega & \xrightarrow{X} & T \\
[0,1] & \xleftarrow{P} & \mathcal{E} & \xleftarrow{X^{-1}} & \mathcal{T}
\end{array}$$

- Notational conventions.
  - We denote a random variable by captical letters such as X, Y, Z.
  - We typically do not write it in functional forms such as $X(\omega)$, $Y(\omega)$, and so on.
  - We write $P(X^{-1}(\mathtt{T}))$ as $P(X \in \mathtt{T})$.
- **Definition 19.2 (σ-algebra generated by a random variable).** *Let $X : \Omega \to T$ be a random variable. The σ-algebra generated by X, denoted by $\sigma(X)$, is the smallest σ-algebra such that X is $\sigma(X)$-measurable. In other words,*

$$\sigma(X) = \bigcap \Big\{ \mathcal{S} \subseteq 2^{\Omega} : \ \mathcal{S} \text{ is a } \sigma\text{-algebra and } X \text{ is } \mathcal{S}\text{-measurable} \Big\}.$$

- **Proposition 19.3.** *Let $X : \Omega \to T$ be a random variable. Then, $\sigma(X) = \{X^{-1}(\mathtt{T}) : \mathtt{T} \in \mathcal{T}\}$.*
- **Proposition 19.4 (probability distribution measure of a random variable).** *Let $X : \Omega \to T$ be a random variable. The function $P_X : \mathcal{T} \to [0, 1]$ given by $P_X(\mathtt{T}) = P(X^{-1}(\mathtt{T})) = P(X \in \mathtt{T})$ for all $\mathtt{T} \in \mathcal{T}$ is a probability measure on $\mathcal{T}$. It is called the* **probability distribution measure of X** .
- For the rest of the section, we will discussed real-valued random variables. So, the range $(T, \mathcal{T})$ is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- **Definition 19.5 (CDF of a random variable).** *Let X be a real-valued random variable the* **cumulative distribution function (CDF) of** *X is given by $F_X(x) = P(X^{-1}((-\infty, x])) = P(X \leq x)$ for any $x \in \mathbb{R}$.*
- **Proposition 19.6.** *The CDF of a real-valued random variable X is a CDF in the sense of Definition 16.1. In other words, the following are true.*

- F_X *is non-decreasing.*
- F_X *is right continuous.*
- $\lim_{x\to\infty} F_X(x) = 1$.
- $\lim_{x\to-\infty} F_X(x) = 0$.

- **Proposition 19.7.** *Let X be a real-valued random variable with CDF F_X. Let*

$$F_X(x^-) = \lim_{u\to x^-} F_X(u).$$

Then,

$$\begin{aligned} P(a < X \le b) &= P(X^{-1}((-a,b])) = F_X(b) - F_X(a), \\ P(a < X < b) &= P(X^{-1}((a,b))) = F_X(b^-) - F_X(a), \\ P(a \le X \le b) &= P(X^{-1}([a,b])) = F_X(b) - F_X(a^-), \\ P(a \le X < b) &= P(X^{-1}([a,b))) = F_X(b^-) - F_X(a^-), \\ P(X = a) &= P(X^{-1}(\{a\})) = F_X(a) - F_X(a^-). \end{aligned}$$

So, $P(X = a) = 0$ if and only if F_X is continuous at a.

20 Lebesgue Integral of Simple Functions

- In this section, we work with the measure space $(S, \mathcal{S}, \mu)$.
- **Definition 20.1 (simple function).** *A measurable real-valued function φ is* **simple** *if it attains a finite number of values.*
- **Proposition 20.2 (standard representation).** *A simple function can be written as a linear combination of indicator functions of measurable sets.*

$$\varphi(x) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$$

where each $\alpha_i \in \mathbb{R}$ and $A_i \in \mathcal{S}$ for each i. There is a unique **standard representation** *where the α_i's are distinct, and the A_i's are disjoint from one another.*

- **Definition 20.3.** *Let $SF^+(S, \mathcal{S})$ denote the set of all non-negative simple $\mathcal{S}$-measurable function.*
- Obviously, $SF^+(S, \mathcal{S}) \subseteq M^+(S, \mathcal{S})$.
- **Proposition 20.4.** *Let $f, g \in SF^+(S, \mathcal{S})$ and $c \ge 0$. The following properties hold.*
  1. $cf \in SF^+(S, \mathcal{S})$.
  2. $f + g \in SF^+(S, \mathcal{S})$.
  3. $\{x \in S : f(x) \ne g(x)\} \in \mathcal{S}$.
  4. $\min(f, g), \max(f, g) \in SF^+(S, \mathcal{S})$.
  5. *For any $A \in \mathcal{S}$, $f\mathbb{1}_A \in SF^+(S, \mathcal{S})$.*

- **Definition 20.5 (integral of simple function).** *Let $\varphi \in SF^+(S,\mathcal{S})$. The* **(Lebesgue) integral of φ with respect to μ** *(or simply the integral) is the extended real number*

$$\int \varphi \,\mathrm{d}\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i)$$

where the α_i's and the A_i's form the standard representation of φ. Also, for any $A \in \mathcal{S}$, the **integral of φ on A with respect to μ** *is the extended real number*

$$\int_A \varphi \,\mathrm{d}\mu = \int \varphi \mathbb{1}_A \,\mathrm{d}\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i \cap A).$$

- **Proposition 20.6 (properties of integrals of simple functions).** *Let $f, g \in SF^+(S)$ and $c \geq 0$. The following properties hold:*

  1. *$\int cf \,\mathrm{d}\mu = c \int f \,\mathrm{d}\mu$.*
  2. *$\int f + g \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu + \int g \,\mathrm{d}mu$.*
  3. *Let $A = \{x \in S : f(x) \neq g(x)\}$. If $\mu(A) = 0$, then $\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu$.*
  4. *If $f \leq g$, then $\int f \,\mathrm{d}\mu \leq \int g \,\mathrm{d}\mu$.*
  5. *If $A, B \in \mathcal{S}$, and $A \subseteq B$, then $\int_A f \,\mathrm{d}\mu \leq \int_B f \,\mathrm{d}\mu$.*
  6. *Let $A_1, A_2, \ldots$ be disjoint sets in $\mathcal{S}$ such that $\bigcup_{i=1}^{\infty} A_i = S$. Then, $\int f \,\mathrm{d}\mu = \sum_{i=1}^{\infty} \int_{A_i} f \,\mathrm{d}\mu$.*
  7. *The function $\lambda_f : \mathcal{S} \to [0,\infty]$ defined as $\lambda_f(E) = \int_E f \,\mathrm{d}\mu$ is a measure on $\mathcal{S}$.*

21 Lebesgue Integral of Non-Negative Functions

- **Proposition 21.1 (approximation by non-decreasing simple functions).** *Let $f \in M^+(S,\mathcal{S})$. There exists a sequence $\{\varphi_n \in SF^+(S,\mathcal{S}) : n \in \mathcal{N}\}$ such that*

  1. *$\varphi_n \leq \varphi_{n+1} \leq f$ for all $n \in \mathbb{N}$, and*
  2. *$lim_{n\to x}\varphi_n(x) = f(x)$ for all $x \in S$.*

  Proof. For each $n \in \mathbb{N}$, define $\phi_n : [0,\infty] \to [0,n]$ as follows:

$$\phi_n(y) = \begin{cases} k/2^n, & k/2^n \leq y < (k+1)/2^n, 0 \leq k/2^n \leq n, k \in \mathbb{N} \cup \{0\} \\ n, & y > n \end{cases}..$$

  In other words, $\phi(y)$ is a discrete approximation of y. If $y > n$, then y is too high, so we approximate it with n. On the other hand, if $y \leq n$, then we approximate y with the highest multiple of $1/2^n$ that is not greater than y.

  It should be clear that $\phi_n(y) \leq \phi_{n+1}(y)$ for all $y \in [0,\infty]$. This is because (1) the range of ϕ_{n+1} is larger than ϕ_n, and (2) the division of the real line into intervals of length $1/2^n$ becomes finer as n increases, so $\phi_{n+1}(y)$ should be a more accurate approximition of y than $\phi_n(y)$.

  Take $\varphi_n = \phi_n \circ f$. The sequence $\{\varphi_n : n \in \mathbb{N}\}$ fits all the bill. □

- **Definition 21.2 (integral of non-negative function).** *Let $f \in M^+(S,\mathcal{S})$. The* **(Lebesgue) integral of f with respect to μ** *is defined to be*

$$\int f \,\mathrm{d}\mu = \sup\left\{ \int \varphi \,\mathrm{d}\mu \;\middle|\; \varphi \in SF^+(S,\mathcal{S}) \text{ and } \phi \leq f \right\}.$$

Also, for any $A \in \mathcal{S}$, define the **integral of f on A with respect to μ** *to be*

$$\int_A f \, \mathrm{d}\mu = \int f \mathbb{1}_A \, \mathrm{d}\mu.$$

In other words, the Lebesgue integral of f is the lowest upper bound of the integral of all simple functions that are not greater than f. The bound exists because of Proposition 21.1.

- **Proposition 21.3.** *Let $f \in M^+(S, \mathcal{S})$. If $\int f \, \mathrm{d}\mu = 0$, then $f = 0$ almost everywhere. (In other words, the set $\{x \in S : f(x) > 0\}$ has measure zero. Symbolically, $\mu(\{x \in S : f(x) > 0\}) = 0$.)*

- **Theorem 21.4 (Monotone Convergence Theorem).** *Let $\{f_n : n \in \mathbb{N}\}$ be a sequence of functions in $M^+(S, \mathcal{S})$ such that $f_n \leq f_{n+1}$ for all n and $\lim_{n\to\infty} f_n(x) = f(x)$ for all x. Then,*

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu.$$

- **Proposition 21.5.** *Let $f, g \in M^+(S, \mathcal{S})$. If $f = g$ almost everywhere, then $\int f \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu$.*

- With the above proposition, the condition of the Monotone Convergence Theorem can be relaxed.

  Corollary 21.6 (Monotone Convergence Theorem 2.0). *Let $\{f_n : n \in \mathbb{N}\}$ be a sequence of functions in $M^+(S, \mathcal{S})$ such that $f_n \leq f_{n+1}$ for all n and $\lim_{n\to\infty} f_n(x) = f(x)$ almost everywhere. Then,*

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu.$$

- **Proposition 21.7 (properites of integral of non-negative function).** *Let $f, g \in M^+(S, \mathcal{S})$ and $c \geq 0$. The following properties are true.*

  1. *$\int cf \, \mathrm{d}\mu = c \int f \, \mathrm{d}\mu$.*
  2. *$\int f + g \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$.*
  3. *If $f \leq g$, then $\int f \, \mathrm{d}\mu \leq \int g \, \mathrm{d}\mu$.*
  4. *If $A, B \in \mathcal{S}$, and $A \subseteq B$, then $\int_A f \, \mathrm{d}\mu \leq \int_B f \, \mathrm{d}\mu$.*
  5. *Let $A_1, A_2, \ldots$ be disjoint sets in $\mathcal{S}$ such that $\bigcup_{i=1}^{\infty} A_i = S$. Then, $\int f \, \mathrm{d}\mu = \sum_{i=1}^{\infty} \int_{A_i} f \, \mathrm{d}\mu$.*
  6. *The function $\lambda_f : \mathcal{S} \to [0, \infty]$ define as $\lambda_f(E) = \int_E f \, \mathrm{d}\mu$ is a measure on $\mathcal{S}$.*

  This proposition is almost the same as Proposition 20.6. The difference is that the functions now belong to $M^+(S, \mathcal{S})$ instead of $SF^+(S, \mathcal{S})$. Most properties can be proven by applying the Monotone Convergence Theorem to the properties in Proposition 20.6.

- **Definition 21.8 (absolute continuity).** *Consider two measures λ and μ on $(S, \mathcal{S})$. We say that λ is* **absolutely continuous with respect to μ** *if $\mu(E) = 0$ implies that $\lambda(E) = 0$. If this is the case, we write $\lambda \ll \mu$.*

- **Proposition 21.9.** *Let $f \in M^+(S, \mathcal{S})$. The measure $\lambda_f(E) = \int_E f \, \mathrm{d}\mu$ is absolutely continuous with respect to μ.*

- **Theorem 21.10 (Fatou's lemma).** *For any sequence $\{f_n : n \in \mathbb{N}\}$ of functions in $M^+(S, \mathcal{S})$, we have that*

$$\liminf_{n\to\infty} f_n \in M^+(S, \mathcal{S}),$$

and

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

- **Theorem 21.11 (reverse Fatou lemma).** *Let $g \in M^+(S, \mathcal{S})$ be a function such that $\int g \, \mathrm{d}\mu < \infty$. Let $\{f_n : n \in \mathbb{N}\}$ be a sequence of functions in $M^+(S, \mathcal{S})$ such that $f_n \leq g$ for all n. Then,*

$$\limsup_{n\to\infty} f_n \in M^+(S, \mathcal{S}),$$

and

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

22 Lebesgue Integration

- In this section, we work with measurable extended real-valued functions. Not all of these functions can be integrated, however.
- **Definition 22.1 (positive and negative parts).** *Let $f \in M(S, \mathcal{S})$. Define*

$$f^+ = \max(0, f),$$
$$f^- = \max(0, -f).$$

It follows that $f^+, f^- \in M^+(S, \mathcal{S})$, $f = f^+ - f^-$, and $|f| = f^+ + f^-$.

- **Definition 22.2 (integrable function).** *Let $f \in M(S, \mathcal{S})$. We say that f is* **(Lebesgue) integrable with respect to μ** *if*

$$\int |f| \, \mathrm{d}\mu = \int f^+ \, \mathrm{d}\mu + \int f^- \, \mathrm{d}\mu < \infty.$$

Let $\mathcal{L}^1(S, \mathcal{S}, \mu)$ denote the set of all integrable functions in $M(S, \mathcal{S})$ with respect to μ.

- **Definition 22.3 (Lebesgue integral).** *Let $f \in \mathcal{L}^1(S, \mathcal{S}, \mu)$. The* **(Lebesgue) integral of f with respect to μ** *is the real number*

$$\int f \, \mathrm{d}\mu = \int f^+ \, \mathrm{d}\mu - \int f^- \, \mathrm{d}\mu.$$

Also, for any $A \in \mathcal{S}$, the **(Lebesgue) integral of f on A with respect to μ** *is the real number*

$$\int_A f \, \mathrm{d}\mu = \int_A f^+ \, \mathrm{d}\mu - \int_A f^- \, \mathrm{d}\mu.$$

- **Proposition 22.4 (linearity of Lebesgue integral).** *Let $f, g \in \mathcal{L}^1(S, \mathcal{S}, \mu)$ and $c \in \mathbb{R}$. The following properties are true.*
  1. $\int cf \, \mathrm{d}\mu = c \int f \, \mathrm{d}\mu$.
  2. $\int f + g \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$.
- In literature, there are several different equivalent notations for the integral:

$$\int f \, \mathrm{d}\mu = \int f(x) \, \mathrm{d}\mu = \int f(x) \, \mathrm{d}\mu(x) = \int f(x) \, \mu(\mathrm{d}x).$$

and

$$\int_A f \, \mathrm{d}\mu = \int_A f(x) \, \mathrm{d}\mu = \int_A f(x) \, \mathrm{d}\mu(x) = \int_A f(x) \, \mu(\mathrm{d}x).$$

Some texts such as [Williams, 1991] even propose using

$$\mu(f) \text{ to denote } \int f \,\mathrm{d}\mu, \text{ and}$$
$$\mu(f; A) \text{ to denote } \int_A f \,\mathrm{d}\mu.$$

These notations are too terse for my taste, but they give the flavor that the integral is, in a sense, the measure of a function. One may also use

$$\mu[f] \text{ to denote } \int f \,\mathrm{d}\mu, \text{ and}$$
$$\mu_A[f] \text{ to denote } \int_A f \,\mathrm{d}\mu.$$

These notations emphasize that the fact that integration is a (linear) operator on functions, much like the expection $E[\cdot]$.

- **Definition 22.5 (charge).** *Let $(S, \mathcal{S})$ be a measurable space. A function $\mu : \mathcal{S} \to \mathbb{R}$ is said to be a* **charge** *on $\mathcal{S}$ if the following properties are satisfied.*

  1. $\mu(\emptyset) = 0$.
  2. *μ is countably additive. This is, for a sequence $\{E_n \in \mathcal{S} : n \in \mathbb{N}\}$ of disjoint sets, it holds that*

$$\mu\Big(\bigcup_{n=1}^{\infty} E_n\Big) = \sum_{n=1}^{\infty} \mu(E_n).$$

  The difference between a charge and a measure is that a measure is always non-negative, but a charge can be negative. Moreover, a charge cannot take infinite values.

- **Proposition 22.6.** *If $f \in \mathcal{L}^1(S, \mathcal{S}, \mu)$, then the function $\lambda_f : \mathcal{S} \to \mathbb{R}$ defined by $\lambda_f(E) = \int_E f \,\mathrm{d}\mu$ is a charge. If $f \in M^+(S, \mathcal{S})$, then λ_f is a measure.*

- **Theorem 22.7 (Dominated Convergence Theorem).** *Let $\{f_n : n \in \mathbb{N}\}$ be a sequence of functions in $M(S, \mathcal{S})$ such that the following properties hold.*

  - *There exists $f \in M(S, \mathcal{S})$ such that f_n convergers to f almost everywhere. In other words, $\lim_{n\to\infty} f_n(x) = f(x)$ almost everywhere.*
  - *There exists $g \in \mathcal{L}^1(S, \mathcal{S}, \mu)$ such that f_n is dominated by g. In other words, $|f_n(x)| \leq g(x)$ for all $x \in S, n \in \mathbb{N}$.*

  Then, the following are true.

  1. *$f_n \in \mathcal{L}^1(S, \mathcal{S}, \mu)$ for all n.*
  2. *$f \in \mathcal{L}^1(S, \mathcal{S}, \mu)$.*
  3. *f_n converges to f in $\mathcal{L}^1(S, \mathcal{S}, \mu)$, which means that*

$$\lim_{n\to\infty} \int |f_n - f| \,\mathrm{d}\mu = 0.$$

  4. *The integral of f_n converges to the integral of f. That is,*

$$\lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu.$$

- **Theorem 22.8 (Scheffé's lemma).** *Let $f, f_1, f_2, \ldots \in \mathcal{L}^1(S, \mathcal{S}, \mu)$. Suppose that $f_n \to f$ almost everywhere. Then, $\int |f_n - f| \,\mathrm{d}\mu \to 0$ if and only if $\int |f_n| \,\mathrm{d}\mu \to \int |f| \,\mathrm{d}\mu$.*

23 Radon–Nikodym Thoerem

- From Proposition 21.7, we can create a measure from a non-negative extended real-valued measurable function by integrating with respect to an existing measure.
- The Radon–Nikodym theorem is the converse of the above property. It indicates when a measure λ can be expressed as an integration of a function f with respect to an existing measure μ. Hence, it is useful in deriving probability density functions.
- A necessary and sufficient condition for the theorem to hold is given below.

  Definition 23.1 (absolute continuity). *Let λ and μ be measures on $\mathcal{S}$. We say that λ is* **absolutely continuous with respect to** *μ if $\mu(E) = 0$ implies $\lambda(E) = 0$ for all $E \in \mathcal{S}$. We write $\lambda \ll \mu$.*

- **Proposition 23.2.** *Let λ and μ be finite measures on $\mathcal{S}$. Then, $\lambda \ll \mu$ if and only if, for every $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that $\lambda(E) < \varepsilon$ for all E such that $\mu(E) < \delta(\varepsilon)$.*
- **Theorem 23.3 (Radon–Nikodym theorem).** *Let λ and μ be σ-finite measures defined on $\mathcal{S}$, and suppose that λ is absolutely continous with respect to μ. Then, there exists a function $f \in M^+(S, \mathcal{S})$ such that*

  $$\lambda(E) = \int_E f \, \mathrm{d}\mu.$$

  Moreover, the function f is uniquely determined μ-almost everywhere. The function is called the **Random–Nikodym derivative** *of λ with respective to μ, and it is denoted by*

  $$\frac{\mathrm{d}\lambda}{\mathrm{d}\mu}.$$

24 Random Variables and Their Expectations

- In this section, the measure space $(S, \mathcal{S}, \mu)$ becomes the probability space $(\Omega, \mathcal{E}, P)$.
- **Definition 24.1 (PDF of a random variable).** *Let X be a real-valued random variable. A measurable function $f_X(x) : \mathbb{R} \to [0, \infty]$ is called the* **probability density function (PDF) of** *X if, for any $B \in \mathcal{B}(\mathbb{R})$,*

  $$P(X \in B) = P_X(B) = \int_B f_X(x) \, \mathrm{d}x = \int_B f_X(x) \, \ell^*(\mathrm{d}x) = \int_B f_X \, \mathrm{d}\ell^*$$

  where ℓ^* is the Lebesgue measure on $\mathbb{R}$ (Definition 15.6).
- **Proposition 24.2.** *If P_X is absolutely continuous with respect to the Lebesgue measure ℓ^*, then f_X exists and is unique ℓ^*-almost everywhere. In particular,*

  $$f_X = \frac{\mathrm{d}P_X}{\mathrm{d}\ell^*}$$

  where the RHS is the Radon–Nikodym derivative.
- **Proposition 24.3.** *Let X be a real-valued random variable with PDF f_X. Then, its CDF, F_X, is given by*

  $$F_X(x) = P_X((-\infty, x]) = \int_{(-\infty, x]} f_X(u) \, \ell^*(\mathrm{d}u) = \int_{-\infty}^{x} f_X(u) \, \mathrm{d}u.$$

Hence,

$$f_X = \frac{\mathrm{d}F_X}{\mathrm{d}x}.$$

- **Definition 24.4.** *For a random variable $X \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$, define the* **expectation of** X *to be*

$$E[X] = \int X \,\mathrm{d}P = \int X(\omega)\, P(\mathrm{d}\omega).$$

- A lot of properties of integrals of random variables carry over to expectations.
- **Proposition 24.5 (linearity of expectation).** *Let $X, Y \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ and $c \in \mathbb{R}$. Then,*

$$E[cX] = cE[X],$$
$$E[X + Y] = E[X] + E[Y].$$

- **Theorem 24.6 (convergence theorems for expectation).** *Let $\{X_n : n \in \mathbb{N}\}$ be a sequence of random variables. Suppose there exists a random variable X such that $X_n \to X$ almost surely. The following statements hold.*

  1. *(Monotone Convergence Theorem) If $0 \leq X_n \leq X_{n+1} \leq X$ for all n, then $E[X_n] \to E[X]$.*
  2. *(Fatou's lemma) If $0 \leq X_n$ for all n, then $E[X] \leq \liminf_{n\to\infty} E[X_n]$.*
  3. *(Dominated Convergence Theorem) Suppose there is a random variable $Y \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ such that $|X_n| \leq Y$ for all n. Then, $X_n \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ for all n, $X \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$, $E[|X_n - X|] \to 0$, and $E[X_n] \to E[X]$.*
  4. *(Scheffé's lemma) $E[|X_n - X|] \to 0 \iff E[|X_n|] \to E[|X|]$.*
  5. *(Bounded Convergence Theorem) Suppose there is constant $K \geq 0$ such that $X_n < K$ for all n. Then, $X_n \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ for all n, $X \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$, $E[|X_n - X|] \to 0$, and $E[X_n] \to E[X]$.*

  The added Bounded Convergence Theorem is a special case of the Dominated Convergence theorem when applied to expectation. Here, we choose $Y(\omega) = K$ for all ω, which gives $E[Y] = K < \infty$.

- **Theorem 24.7 (LOTUS 2.0).** *Let X be a real-valued random variable in $(\Omega, \mathcal{E}, P)$. Let $g : \mathbb{R} \to \mathbb{R}$ be $\mathcal{B}(\mathbb{R})$-measurable. Then, the following statements are true.*

  – *$g(X) \in \mathcal{L}^1(\Omega, \mathcal{E}, P) \iff g \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.*
  – *If $g \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$, then*

$$E[g(X)] = \int_\Omega g(X) \,\mathrm{d}P = \int_\Omega g(X(\omega))\, P(\mathrm{d}\omega) = \int_{\mathbb{R}} g(x)\, P_X(\mathrm{d}x) = \int_{\mathbb{R}} g \,\mathrm{d}P_X.$$

  The diagram of the situation in in the theorem is as follows.

$$\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$$

$$[0,1] \xleftarrow{P} \mathcal{E} \xleftarrow{X^{-1}} \mathcal{B}(\mathbb{R}) \xleftarrow{g^{-1}} \mathcal{B}(\mathbb{R})$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \overset{P_X}{\curvearrowleft}$$

- For the analogue of Theorem 12.6, we now require that $h(X)$ must be integrable.

Theorem 24.8. *Let $h : \mathbb{R} \to [0, \infty)$ and X be a random variable. If $h(X) \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$, then, for any $a > 0$,*

$$P(h(X) \geq a) \leq \frac{E[h(X)]}{a}.$$

- **Corollary 24.9 (Markov's inequality).** *If $X \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$, then, for any $a > 0$,*

$$P(|X| \geq a) \leq \frac{E[|X|]}{a}.$$

- **Proposition 24.10.** *If $X \in M^+(\Omega, \mathcal{E})$ and $E[X] < \infty$, then $X < \infty$ almost surely.*
- **Proposition 24.11 (results on sums random variables).** *The following statements are true.*
  - *If $\{X_n : n \in \mathbb{N}\}$ be a sequence of random variables in $M^+(\Omega, \mathcal{E})$, then*

$$E\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty} E[X_n]. \tag{1}$$

  - *Let $\{X_n : n \in \mathbb{N}\}$ be a sequence of random variables in $\mathcal{L}^1(\Omega, \mathcal{E}, P)$. If $\sum_{n=1}^{\infty} E[|X_n|] < \infty$, then $Y = \sum_{n=1}^{\infty} X_n$ converges almost surely. (In other words, Y is finite almost surely.) Moreover, $Y \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$, Equation (1) holds, and $X_n \to 0$.*
- **Theorem 24.12 (the first Borel–Cantelli lemma).** *Let $\{E_n : n \in \mathbb{N}\}$ be a sequence of events such that $\sum_{n=1}^{\infty} P(E_n) < \infty$. Take $X_n = \mathbb{1}_{E_n}$. Then,*

$$\sum_{n=1}^{\infty} \mathbb{1}_{E_n} = \text{number of events } E_n \text{ that occur}$$

is finite almost surely.

- **Definition 24.13 (convex function).** *Let I be an open subinterval of $\mathbb{R}$. A function $f : I \to \mathbb{R}$ is called* **convex** *if, for every $x, y \in I$ and $\alpha \in [0, 1]$, it holds that*

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

In other words, the curve of f on any interval $[x, y] \subseteq I$ lies below the straight line that connects the point $(x, f(x))$ to the point $(y, f(y))$.

- **Theorem 24.14 (Jensen's inequality).** *Let $f : I \to \mathbb{R}$ be a convex function. Let $X \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ be a random variable such that $P(X \in I) = 1$, and $f(X) \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$. Then,*

$$f(E[X]) \leq E[f(X)].$$

25 Lebesgue spaces L^p

- **Definition 25.1.** *Let $(S, \mathcal{S}, \mu)$ be a measure space. For $1 \leq p < \infty$, define $\mathcal{L}^p(S, \mathcal{S}, \mu)$ to be the set of functions $f \in M(S, \mathcal{S})$ such that $|f|^p$ is integrable. In other words,*

$$\int |f|^p \, \mathrm{d}\mu = \int (f^+)^p \, \mathrm{d}\mu + \int |f^-|^p \, \mathrm{d}\mu < \infty.$$

When the measure space $(S, \mathcal{S}, \mu)$ is clear from the context, we will simply write $\mathcal{L}^p$ instead of $\mathcal{L}^p(S, \mathcal{S}, \mu)$.

- **Definition 25.2 ($\mathcal{L}^p$-norm).** *For any $f \in \mathcal{L}^p(S, \mathcal{S}, \mu)$, define the* **$\mathcal{L}^p$-norm of** *$f$ to be*

$$\|f\|_p = \left(\int |f|^p \, \mathrm{d}\mu \right)^{\frac{1}{p}}$$

- **Proposition 25.3 (monotonicity of $\mathcal{L}^p$-norm).** *Let $1 \leq p \leq q < \infty$. If $f \in \mathcal{L}^q$, then $f \in \mathcal{L}^p$, and $\|f\|_p \leq \|f\|_q$.*

- **Proposition 25.4 (vector space properties of $\mathcal{L}^p$).** *Let $f, g \in \mathcal{L}^p$. The following statements are true.*

  1. *For any $c \in \mathbb{R}$, $cf \in \mathcal{L}^p$.*
  2. *$f + g \in \mathcal{L}^p$.*

- **Proposition 25.5 (properties of the $\mathcal{L}^p$-norm).** *Let $f, g \in \mathcal{L}^p$. The following statements are true.*

  (a) $\|f\|_p \geq 0$.
  (b) If $\|f\|_p = 0$, then $f = 0$ almost everywhere.
  (c) $\|cf\|_p = |c|\|f\|_p$ *for any $c \in \mathcal{R}$.*
  (d) (Minkowsky's inequality) $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

- **Theorem 25.6 (Hölder's inequality).** *Let $p, q > 1$ be such that $1/p + 1/q = 1$. If $f \in \mathcal{L}^p$, $g \in \mathcal{L}^q$, then $fg \in \mathcal{L}^1$ and $\|fg\|_1 \leq \|f\|_p \|g\|_q$.*

- **Theorem 25.7 (Cauchy–Schwarz inequality).** *If $f, g \in \mathcal{L}^2(S, \mathcal{S}, \mu)$, then $fg \in \mathcal{L}^1$, and*

$$\left| \int fg \, \mathrm{d}\mu \right| \leq \int |fg| \, \mathrm{d}\mu = \|fg\|_1 \leq \|f\|_2 \|g\|_2.$$

- **Definition 25.8 (normed vector space).** *A* **normed vector space** *is a vector space V endowed with a norm function $\|\cdot\| : V \to [0, \infty]$ such that, for every $\mathbf{u}, \mathbf{v} \in V$ and $c \in \mathbb{R}$, it is true that*

  (a) $\|\mathbf{u}\| \geq 0$,
  (b) $\|\mathbf{u}\| = 0 \iff \mathbf{u} = \mathbf{0}$,
  (c) $\|c\mathbf{u}\| = |c|\|\mathbf{u}\|$, *and*
  (d) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

- $\mathcal{L}^p$ is a vector space with the zero function $f(x) = 0$ serving as the zero vector. However, it is not a normed vector space because Property (b) is not satisifed. Instead, if $\|f\|_p = 0$, then we can only say that $f = 0$ almost everywhere instead of everywhere.

- To make a normed vector space out of $\mathcal{L}^p$, we instead view the set of functions that are equal almost every as a unit.

  Definition 25.9 (equivalent class w.r.t. $\mathcal{L}^p$-norm). *Let $f \in \mathcal{L}^p$. The* **equivalent class with respect to $\mathcal{L}^p$-norm of** *f is the set*

$$[f]_p = \{g \in \mathcal{L}^p : f = g \text{ almost everywhere}\}.$$

- **Proposition 25.10.** $[f]_p = \{f + g : g \in [0]_p\}$.

- **Definition 25.11 (operations on equivelent classes).** *Let $f, g \in \mathcal{L}^p$, and $c \in \mathbb{R}$. Define*

$$\begin{aligned} c[f]_p &:= \{cg : g \in [f]_p\}, \\ [f]_p + [g]_p &:= \{f + g : f \in [f]_p + g \in [g]_p\}, \\ \|[f]_p\|_p &:= \|f\|_p. \end{aligned}$$

  The operations are well defined despite the non-determinism in the definition.

- **Definition 25.12 (Lebesgue space L^p).** *The* **Lebesgue space** *$L^p(S, \mathcal{S}, \mu)$ is defined to be the set $\{[f]_p : f \in \mathcal{L}^p(S, \mathcal{S}, \mu)\}$.*

  When the measure space is clear from the context, we will write L^p instead of $L^p(S, \mathcal{S}, \mu)$.

- **Theorem 25.13 (L^p is a normed vector space).** *The space L^p together with the operations in Definition 25.11 is a normed vector space with $[0]_p$ serving as the zero vector.*

- **Definition 25.14 (Cauchy sequence).** *Let V be a vector space with some norm function $\|\cdot\|$. A* **Cauchy sequence in** *V is a sequence $\{\mathbf{u}_n \in V : n \in \mathbb{N}\}$ such that*

$$\lim_{n\to\infty} \sup_{i,j\geq n} \|\mathbf{u}_i - \mathbf{u}_j\| = 0.$$

- **Theorem 25.15 (Cauchy sequence converged in $\mathcal{L}^p$).** *Let $f_n \in \mathcal{L}^p : n \in \mathbb{N}$ be a Cauchy sequence. Then, there exists a function $f \in \mathcal{L}^p$ such that $f_n \to f$ in $\mathcal{L}^p$; that is,*

$$\lim_{n\to\infty} \|f_n - f\|_p = 0.$$

- **Definition 25.16 (completeness).** *A normed vector space V is said to be* **complete** *if every Cauchy sequence in V has a limit in V. In other words, if $\{u_n : n \in \mathbb{N}\}$ is a Cauchy sequence in V, then there exists $\mathbf{u} \in V$ such that $\lim_{n\to\infty} \|\mathbf{u}_n - \mathbf{u}\| = 0$.*

- **Definition 25.17 (Banach space).** *A Banach space is a complete normed vector space.*

- **Theorem 25.18 (L^p is Banach).** *The space L^p is complete, and so is a Banach space.*

26 Variance on an Uncountable Space

- Let $X \in \mathcal{L}^2(\Omega, \mathcal{E}, P)$. By the monotonicity of $\mathcal{L}^p$-norm (Proposition 25.3), we have that

$$\left|E[X]\right| = \left|\int X \,\mathrm{d}P\right| \leq \int |X| \,\mathrm{d}P = \|X\|_1 \leq \|X\|_2 < \infty.$$

  This means that $E[X]$ is finite. As a result,

$$E[(X - E[X])^2] = E[X^2] - 2(E[X])^2 + (E[X])^2 = E[X^2] + (E[X])^2 = \|X\|_2^2 + (E[X])^2$$

  is also finite. So, $(X - E[X])^2 \in \mathcal{L}^2(\Omega, \mathcal{E}, P)$ as well.

- **Definition 26.1 (variance and standard deviation).** *Let $X \in \mathcal{L}^2(\Omega, \mathcal{E}, P)$. The* **variance of X**, *denoted by* $\mathrm{Var}(X)$, *is defined to be*

$$\mathrm{Var}(X) = \|X - E[X]\|_2^2 = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

  The **standard deviation of** *X, denoted by* $\mathrm{Stdev}(X)$, *is the non-negative square root of the variance.*

$$\mathrm{Stdev}(X) = \sqrt{\mathrm{Var}(X)} = \|X - E[X]\|_2.$$

- **Theorem 26.2 (Chebyshev's inequality).** *Let $X \in \mathcal{L}^2(\Omega, \mathcal{E}, P)$. We have that, for any $a > 0$,*

$$P(|X| \geq a) = \frac{E[X^2]}{a^2},$$

so

$$P(|X - E[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

One can prove the theorem by applying Markov's inequality to the random variable X^2 and $(X - E[X])^2$.

27 Product Measures and Double Integrals

- In this section, two measurable spaces $(X, \mathcal{X}, \mu)$, $(Y, \mathcal{Y}, \nu)$, we want to construct a new measure space whose underlying set is $X \times Y$.
- The natural candidate for the σ-algebra to define a measure on is $\mathcal{X} \times \mathcal{Y}$. The problem, however, is that $\mathcal{X} \times \mathcal{Y}$ is not a σ-algebra in general. Still...

  Proposition 27.1. *Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces. Then, $\mathcal{X} \times \mathcal{Y}$ is a semi-ring on $X \times Y$.*

- As a result, to construct a measure, we must work with $\sigma(\mathcal{X} \times \mathcal{Y})$ instead. Since this construction is very common, we will give it a special notation.

  Definition 27.2 (product of σ-algebras). *Let $\mathcal{X}$ and $\mathcal{Y}$ be σ-algebras. Let $\mathcal{X} \otimes \mathcal{Y}$ denote the σ-algebra generated by $\mathcal{X} \times \mathcal{Y}$.*

$$\mathcal{X} \otimes \mathcal{Y} = \sigma(\mathcal{X} \times \mathcal{Y}).$$

- **Definition 27.3 (product measure theorem).** *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure spaces. Then, there exists a measure ρ on $\mathcal{X} \otimes \mathcal{Y}$ such that*

$$\rho(A \times B) = \mu(A)\nu(B)$$

for all $A \in \mathcal{X}$ and $B \in \mathcal{Y}$.

- When working a σ-algebra, however, we generally start from a simpler collection of sets that generates that σ-algebra. The following definitions and statements allows us to conveniently work with the generators in the production measure settings.

  Definition 27.4 (exhausting sequence). *Let X be a set and $\mathcal{A} \subseteq 2^X$ be a collection of sets. An* **exhausting sequence of X in $\mathcal{A}$** *is a sequence of sets $\{A_n : n \in \mathbb{N}\}$ such that $A_n \uparrow X$. In other words, $A_n \subseteq A_{n+1}$ for all n and $\bigcup_{n=1}^{\infty} A_n = X$. If there is such a sequence, we say that $\mathcal{A}$* **exhausts X.**

  Proposition 27.5. *Let X and Y be sets. Let $\mathcal{A} \subseteq 2^X$ and $\mathcal{B} \subseteq 2^Y$ be collections of sets. Let $\mathcal{X} = \sigma(\mathcal{A})$, and $\mathcal{Y} = \sigma(\mathcal{B})$. If $\mathcal{A}$ exhausts X and $\mathcal{B}$ exhausts Y, then*

$$\sigma(\mathcal{A} \times \mathcal{B}) = \sigma(\mathcal{X} \times \mathcal{Y}) = \mathcal{X} \otimes \mathcal{Y}.$$

- When the generators are π-systems and the measures are σ-finite, we have that the product measure is unique.

  Theorem 27.6 (uniqueness of product measure). *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure spaces. Let $\mathcal{A} \subseteq 2^X$ and $\mathcal{B} \subseteq 2^Y$ be such that the following properties hold.*

1. *$\sigma(\mathcal{A}) = \mathcal{X}$, and $\sigma(\mathcal{B}) = \mathcal{Y}$.*
2. *$\mathcal{A}$ and $\mathcal{B}$ are π-systems.*
3. *There exist exhausting sequences $A_n \uparrow X$ in $\mathcal{A}$ and $B_n \uparrow Y$ in $\mathcal{B}$ such that $\mu(A_n) < \infty$ and $\nu(B_n) < \infty$ for all n.*

Then, there exists one and at most one measure ρ on $\mathcal{X} \otimes \mathcal{Y}$ such that $\rho(A \times B) = \mu(A)\nu(B)$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Moreover, ρ is also σ-finite.

- **Definition 27.7 (product measure space).** *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure spaces such that μ and ν are σ-finite. The unique measure on $\mathcal{X} \otimes \mathcal{Y}$ is called the* **product of** *mu* **and** *ν and is denoted by $\mu \times \nu$, and we call $(X \times Y, \mathcal{X} \otimes \mathcal{Y}, \mu \times \nu)$ a* **producte measure space**.
- For $d \geq 1$, let v^d denote the Lebesgue meausure on $\mathbb{R}^d$ (Definition 15.11).

  Corollary 27.8. *For any $n > d \geq 1$, we have*

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), v^n) = (\mathbb{R}^d \times \mathbb{R}^{n-d}, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^{n-d}), v^d \times v^{n-d}).$$

- **Definition 27.9 (sections of a set).** *Let $A \subseteq X$. For each $x \in X$, the x-***section of** *A is the set*

$$A_{x,\square} = \{y \in Y : (x, y) \in A\}.$$

  *For each $y \in Y$, the y-***section of** *A is the set*

$$A_{\square,y} = \{x \in X : (x, y) \in A\}.$$

- **Definition 27.10 (sections of two-argument function).** *Let $f : X \times Y \to T$. For each $x \in X$, the x-section of f is the function $f_{x,\square} : Y \to T$ such that*

$$f_{x,\square}(y) = f(x, y).$$

  For each $y \in Y$, the y-section of f is the function $f_{\square,y} : X \to T$ such that

$$f_{\square,y}(x) = f(x, y).$$

- **Proposition 27.11 (sections are measurable).** *Let $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ be a measurable space. Let $A \in \mathcal{X} \otimes \mathcal{Y}$. Then,*
  - *$A_{x,\square} \in \mathcal{Y}$ for all $x \in X$, and*
  - *$A_{\square,y} \in \mathcal{X}$ for all $y \in Y$.*

  Moreover, let $(T, \mathcal{T})$ be a measurable space and $f : X \times Y \to T$ be a $(\mathcal{X} \otimes \mathcal{T})/\mathcal{T}$-measurable function.
  - *$f_{x,\square}$ is $\mathcal{Y}/\mathcal{T}$-measurable for all $x \in X$, and*
  - *$f_{\square,y}$ is $\mathcal{X}/\mathcal{T}$-measurable for all $y \in Y$.*
- **Proposition 27.12.** *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure spaces with μ and ν being σ-finite. If $A \in \mathcal{X} \otimes \mathcal{Y}$, the functions defined by*

$$f(x) = \nu(A_{x,\square}),$$
$$g(y) = \mu(A_{\square,y})$$

  are $\mathcal{X}$-measurable and $\mathcal{Y}$-measurable, respectively. Moreover, if $\phi = \mu \times \nu$, we have that

$$\mu(A) = \int_X f(x)\,\mathrm{d}\mu(x) = \int_Y g(y)\,\mathrm{d}\mu(y).$$

- **Theorem 27.13 (Tonelli).** *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure spaces with μ and ν being σ-finite. Let $\rho = \mu \times \nu$. Let $f \in M^+(X \times Y, \mathcal{X} \otimes \mathcal{Y})$. Define*

$$g(x) = \int_Y f_{x,\square}(y)\,\mathrm{d}\nu(y),$$
$$h(y) = \int_X f_{\square,y}(x)\,\mathrm{d}\mu(x).$$

Then,

$$\int_{X\times Y} f\,\mathrm{d}\rho = \int_X g(x)\,\mathrm{d}\mu(x) = \int_Y h(y)\,\mathrm{d}\nu(y).$$

Tonelli's theorem states that an integral on a product measure space of a non-negative function can be evaluated as a double integral, and the order of the integral can be exchanged.

- The notation in Theorem 27.13 is a bit opaque. We commonly write the integrals as:

$$\int_Y f(x,y)\,\mathrm{d}\mu(y) := \int_Y f_{x,\square}(y)\,\mathrm{d}\nu(y),$$
$$\int_X f(x,y)\,\mathrm{d}\nu(x) := \int_X f_{\square,y}(x)\,\mathrm{d}\mu(x),$$
$$\int_{X\times Y} f(x,y)\,\mathrm{d}\rho(x,y) := \int_{X\times Y} f\,\mathrm{d}\rho.$$

With the more familiar notation, we have that

$$\int_{X\times Y} f(x,y)\,\mathrm{d}\rho(x,y) = \int_X \left(\int_Y f(x,y)\,\mathrm{d}\nu(y)\right)\mathrm{d}\mu(x) = \int_Y \left(\int_X f(x,y)\,\mathrm{d}\mu(x)\right)\mathrm{d}\nu(y),$$

or, simply,

$$\int_{X\times Y} f(x,y)\,\mathrm{d}\rho(x,y) = \int_X \int_Y f(x,y)\,\mathrm{d}\nu(y)\mathrm{d}\mu(x) = \int_Y \int_X f(x,y)\,\mathrm{d}\mu(x)\mathrm{d}\nu(y),$$

or just

$$\int_{X\times Y} f\,\mathrm{d}(\mu\times\nu) = \int_X \int_Y f\,\mathrm{d}\nu\mathrm{d}\mu = \int_Y \int_X f\,\mathrm{d}\mu\mathrm{d}\nu.$$

- **Theorem 27.14 (Fubini).** *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure spaces with μ and ν being σ-finite. Let $f \in M(X \times Y, \mathcal{X} \otimes \mathcal{Y})$. If at least one of the three integrals,*

$$\int_{X\times Y} |f|\,\mathrm{d}(\mu\times\nu), \qquad \int_X\int_Y |f|\,\mathrm{d}\nu\mathrm{d}\mu, \qquad \int_Y\int_X |f|\,\mathrm{d}\mu\mathrm{d}\nu,$$

is finite, then all of them are finite, and $f \in \mathcal{L}^1(X \times Y, \mathcal{X} \otimes \mathcal{Y}, \mu \times \nu)$. Moreover, we have that

1. *$f_{x,\square} \in \mathcal{L}^1(Y, \mathcal{Y}, \nu)$ for x almost everywhere in X with respect to μ.*
2. *$f_{\square,y} \in \mathcal{L}^1(X, \mathcal{X}, \mu)$ for y almost everywhere in Y with respect to ν.*
3. *$g(x) = \int_Y f_{x,\square}\,\mathrm{d}\nu = \int_Y f(x,y)\,\mathrm{d}\nu(y) \in \mathcal{L}^1(X, \mathcal{X}, \mu)$.*
4. *$h(y) = \int_X f_{\square,y}\,\mathrm{d}\mu = \int_X f(x,y)\,\mathrm{d}\mu(x) \in \mathcal{L}^1(Y, \mathcal{Y}, \nu)$.*
5. *Lastly,*

$$\int_{X\times Y} f\,\mathrm{d}(\mu\times\nu) = \int_X\int_Y f\,\mathrm{d}\nu\mathrm{d}\mu = \int_Y\int_X f\,\mathrm{d}\mu\mathrm{d}\nu.$$

Fubini's theorem is a corollary to Tonelli's theorem after applying the latter to a general extended real-valued function. We use it every time we evaluate a double integral by slicing.

28 Two Random Variables Considered Together

- **Definition 28.1 (joint probability distribution measure).** *Let $X : \Omega \to U$ and $Y : \Omega \to V$ be random variables (with associated σ-algebras $\mathcal{U}$ and $\mathcal{V}$, respectively). Then, $Z = (X, Y)$ is a random variable that maps Ω to $\mathcal{U} \times \mathcal{V}$. The* **joint probability distribution measure of X and Y** *is the function $P_Z : \mathcal{U} \otimes \mathcal{V} \to [0, 1]$ such that*

$$P_Z(W) = P(Z^{-1}(W)) = P(\{\omega : (X(\omega), Y(\Omega)) \in W\})$$

for all $W \in \mathcal{U} \times \mathcal{V}$. P_Z is a measure on $\mathcal{U} \otimes \mathcal{V}$. We also denote P_Z with $P_{X,Y}$.

- Consider the special case where X and Y are real-valued random variables. That is, $(U, \mathcal{U})$ and $(V, \mathcal{V})$ are $(\mathbb{R} \times \mathbb{R}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})) = (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. Because the collection $\mathcal{I} = \{(-\infty, x] : x \in Real\}$ is a π-system that generates $\mathbb{R}$, we have that P_Z is completely determined by its values on $\mathcal{I} \times \mathcal{I}$.

- **Definition 28.2 (joint CDF).** *Let X and Y be two real-valued random variables. Let $Z = (X, Y)$. The* **joint cumulative distribution function of Z,** *denoted by fF_Z and $F_{X,Y}$, is the function*

$$F_Z(x, y) = F_{X,Y}(x, y) = P(X \leq x \wedge Y \leq y) = P(\{\omega : X(\omega) \leq x \wedge Y(\omega) \leq y\}).$$

- **Definition 28.3 (joint PDF).** *Let X and Y be two real-valued random variables. A measurable function $f_{X,Y} : \mathbb{R}^2 \to [0, \infty]$ is called a* **joint probability distribution of X and Y** *if, for every $W \in \mathcal{B}(\mathbb{R}^2)$,*

$$P((X, Y) \in W) = \int_W f_{X,Y}(x, y) \, \mathrm{d}v^2(x, y)$$

where v^2 is the Lebesgue measure in $\mathbb{R}^2$ (i.e., the area measure).

- **Proposition 28.4.** *If the joint probability distribution measure $P_{X,Y}$ is absolutely continuous with respect to the area measure v^2, then the PDF $f_{X,Y}$ exists and is unique v^2-almost everywhere. In particular,*

$$f_{X,Y} = \frac{\mathrm{d}P_{X,Y}}{\mathrm{d}v^2}$$

where the RHS is the Radon–Nikodym derivative.

- **Proposition 28.5.** *Let X and Y be two real-valued random variables with joint PDF $f_{X,Y}$ and joint CDF $F_{X,Y}$. We have that*

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{(-\infty, x] \times (-\infty, y]} f_{X,Y}(u, v) \, \mathrm{d}v^2(u, v) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) \, \mathrm{d}v^1(v) \, \mathrm{d}v^1(u) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) \, \mathrm{d}u \, \mathrm{d}v. \end{aligned}$$

So, we may say that

$$f_{X,Y} = \frac{\mathrm{d}F_{X,Y}}{\mathrm{d}v^2} = \frac{\partial F_{X,Y}}{\partial x \, \partial y}.$$

29 Independent Random Variables

- We discussed the notion of independent events in Section 9. In this section, we focus on formulating independence through the σ-algebras. This formulation will allow us to more elegantly define independent random variables.

- **Definition 29.1 (independent σ-algebras).** *Sub σ-algebras $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, ... are* **independent** *if, for every set of finite indices $\{i_1, i_2, \ldots, i_n\} \subset \mathbb{N}$, we have that*

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = \prod_{k=1}^{n} P(E_{i_k})$$

  given that $E_{i_k} \in \mathcal{E}_{i_k}$ for all k.

  The definition can be extended to other collections of sets, including algebra, semi-ring, and π-system.

- **Proposition 29.2.** *Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be sub-σ-algebras such that $\mathcal{E}_1 = \sigma(\Pi_1)$ and $\mathcal{E}_2 = \sigma(\Pi_2)$ where Π_1 and Π_2 are π-systems. Then, $\mathcal{E}_1$ and $\mathcal{E}_2$ are independent if and only if Π_1 and Π_2 are indepenent.*

- **Proposition 29.3 (independent events in terms of σ-algebras).** *Events E_1, E_2, ..., are* **independent** *if and only if their generated σ-algebras, given by*

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

  are indepenent.

- **Definition 29.4 (independent random variables).** *Random variables X_1, X_2, X_3, ... are* **independent** *if their generated σ-algebras $\sigma(X_1)$, $\sigma(X_2)$, $\sigma(X_3)$, ... are independent.*

  The above definition only requires that the random variables have the same probability space $(\Omega, \mathcal{E}, P)$ as the domain. The ranges can be different. That is, X_1 maps to $(T_1, \mathcal{T}_1)$, X_2 maps to $(T_2, \mathcal{T}_2)$, and so on.

- By the above definition, to check if random variables $X : \Omega \to U$ and $Y : \Omega \to V$ are independent, we would need to check that $P(X \in A \wedge Y \in B)$ for all $A \in \mathcal{U}$ and $B \in \mathcal{V}$. This can be cumbersome. However, the following proposition allows us to work with π-systems that generate $\mathcal{U}$ and $\mathcal{V}$.

  Proposition 29.5. *Let $X : \Omega \to U$ and $Y : \Omega \to V$ be random variables (with associated σ-algebras $\mathcal{U}$ and $\mathcal{V}$, respectively). The random variables are independent if and only if*

$$P(X \in A \wedge Y \in B) = P(X \in A)P(Y \in B)$$

  for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$ where $\mathcal{A}$ and $\mathcal{B}$ are π-systems such that $\mathcal{U} = \sigma(\mathcal{A})$ and $\mathcal{V} = \sigma(\mathcal{B})$.

- As an example, consider two real-valued random variables X and Y. We have that $\mathcal{U} = \mathcal{V} = \mathcal{B}(\mathbb{R})$. Because the collecction $\{(-\infty, x] : x \in \mathbb{R}\}$ is a π-system that generates $\mathcal{B}(\mathbb{R})$, it only suffices to check whether

$$P(X \leq x \wedge P \leq y) = P(X \leq x)P(Y \leq y)$$

  for all $x, y \in \mathbb{R}$ to check if the random variables are independent.

- Independent random variables can also be viewed through the lens of joint distributions and product measures.

  Proposition 29.6. *Let $X : \Omega \to X$ and $Y : \Omega \to Y$ be random variables (with associated σ-algebras $\mathcal{X}$ and $\mathcal{Y}$, respectively). The random variables are independent if and only if $P_{X,Y} = P_X \times P_Y$.*

- **Theorem 29.7 (second Borel–Cantelli lemma).** *If $\{E_n : n \in \mathbb{N}\}$ be a sequence of independent events, then*

$$\sum_{n=1}^{\infty} P(E_n) = \infty \implies P(\limsup_{n\to\infty} E_n) = 1.$$

Note that $\limsup_{n\to\infty} E_n$ is the set of elements in Ω that occurs in infinitely many events in the sequence. Hence, if $\sum P(E_n) > \infty$, then "occurs in infinitely many events" is a property that is true almost everywhere in Ω.

- **Definition 29.8 (tail σ-algebra).** *Let X_1, X_2, X_3, ... be a sequence of random variables. For each n, let $\mathcal{E}_n = \sigma(X_n)$ be the σ-algebra generated by X_n. The* **tail σ-algebra of the sequence** *is defined to be*

$$\mathcal{E}_\infty = \bigcap_{n=1}^{\infty} \sigma\Big(\bigcup_{m=n}^{\infty} \mathcal{E}_m \Big).$$

The tail σ-algebra contains **tail events**. A tail event is an event that do not appear in only a finite number of σ-algebras in the sequence.

- **Theorem 29.9 (Kolmogorov's zero-one law).** *Let X_1, X_2, X_3, ... be a sequence of independent random variables. Let $\mathcal{E}_\infty$ be the associated tail σ-algebra. Then, if $E \in \mathcal{E}_\infty$, then $P(E) = 0$ or $P(E) = 1$.*

- **Theorem 29.10.** *Let X and Y be random variables in $\mathcal{L}^2(\Omega, \mathcal{E}, P)$. Then, if X and Y are independent, then $E[XY] = E[X]E[Y]$.*

Proof. We include the proof because it is instructive. The strategy is to proceed like how we define Lebesque integral.

(Step 1) First, if X and Y are independent, then their generated σ-algebras, $\sigma(X)$ and $\sigma(Y)$, are independent. As a result, for any event $A \in \sigma(X)$ and $B \in \sigma(Y)$, we have that $P(A \cap B) = P(A)P(B)$. Now, consider the indicator function $\mathbb{1}_A$ of A. We have that $\sigma(\mathbb{1}A) = \sigma(A) \subseteq \sigma(X)$, and the same can be said for $\mathbb{1}_B$. We also have that $\mathbb{1}_A$ is $\sigma(X)$-measurable and is in $\mathcal{L}^2(\Omega, \mathcal{E}, P)$ because $E[\mathbb{1}_A] = P(A)$, which is finite. Again, the same can be said for $\mathbb{1}_B$. Thus,

$$E[\mathbb{1}_A \mathbb{1}_B] = P(A \cap B) = P(A)P(B) = E[\mathbb{1}_A]E[\mathbb{1}_B].$$

As a result, we can say that the theorem is true for pairs of indicator functions in $\sigma(X)$ and $\sigma(Y)$.

(Step 2) Second, by algebraic manipulation, we have that

$$E[(a_1X_1 + a_2X_2)(b_1Y_1 + b_2Y_2)] = a_1b_1E[X_1Y_1] + a_1b_2E[X_1Y_2] + a_2b_1E[X_2Y_1] + a_2b_2E[X_2Y_2].$$

If $X_1, X_2 \in \mathcal{L}^2(\Omega, \sigma(X), P)$ and $Y_1, Y_2 \in \mathcal{L}^2(\Omega, \sigma(Y), P)$, we have that

$$\begin{aligned}
&E[(a_1X_1 + a_2X_2)(b_1Y_1 + b_2Y_2)] \\
&= a_1b_1E[X_1Y_1] + a_1b_2E[X_1Y_2] + a_2b_1E[X_2Y_1] + a_2b_2E[X_2Y_2] \\
&= a_1b_1E[X_1]E[Y_1] + a_1b_2E[X_1]E[Y_2] + a_2b_1E[X_2]E[Y_1] + a_2b_2E[X_2]E[Y_2] \\
&= (a_1E[X_1] + a_2E[X_2])(b_1E[Y_1] + b_2EY_2).
\end{aligned}$$

This means that, if X_1, X_2 are $\sigma(X)$-measurable with finite variance, it means that their linear combinations would also satisfy the theorem.

(Step 3) Using Step 1 and Step 2, we have that all pairs of simple functions in $SF^+(\Omega, \sigma(X))$ and $SF^+(\Omega, \sigma(Y))$ satsify the theorem.

(Step 4) Let $f \in M^+(\Omega, \sigma(X))$ and $g \in M^+(\Omega, \sigma(Y))$ such that $E[f^2] < \infty$ and $E[g^2] < \infty$. We have that $fg \in M^+(\Omega, \Sigma)$ and $E[fg] < \infty$. Using Proposition 21.1, there exists a sequence of non-decreasing simple functions $\{f_n \in SF^+(\Omega, \sigma(X)) : n \in \mathbb{N}\}$ and $\{g_n : SF^+(\omega, \sigma(Y)) : n \in \mathbb{N}\}$ such that $f_n \to f$ and $g_n \to fg$. We have that $\{f_n g_n : n \in \mathbb{N}\}$ is also a sequence of simple functions such that $f_n g_n \to g$. It follows from the monotone convergence theorem that

$$\lim_{n\to\infty} E[f_n] = E[f],$$
$$\lim_{n\to\infty} E[g_n] = E[g],$$
$$\lim_{n\to\infty} E[f_n g_n] = E[fg].$$

Because $E[f_n g_n] = E[f_n]E[g_n]$ by Step 3, we may conclude that $E[fg] = E[f]E[g]$ as well. So, the theorem is true for any pair of non-negative functions in $\mathcal{L}^2(\Omega, \sigma(X), P)$ and $\mathcal{L}^2(\Omega, \sigma(Y), P)$.

(Step 5) Because we can write $X = X^+ - X^-$ and $Y = Y^+ - Y^-$, we can conclude that $E[XY] = E[X]E[Y]$ using Step 4 and Step 2 together. □

30 Covariance

- **Definition 30.1 (covariance).** *Let $X, Y \in \mathcal{L}^2(\Omega, \mathcal{E}, P)$. The* **covariance of X and Y,** *denoted by $\mathrm{Cov}(X, Y)$, is defined as:*

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

- **Proposition 30.2.** *If $X, Y \in \mathcal{L}^2(\Omega, \mathcal{E}, P)$ are independent random variables, then $\mathrm{Cov}(X, Y) = 0$.*

- **Proposition 30.3.** *Let $X, Y \in \mathcal{L}^2(\Omega, \mathcal{E}, P)$. We have that*

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y).$$

If X and Y are independent, then

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

31 L^2 and Hilbert Space

- Among all the Lebesgue spaces, L^2 is special because we can define an inner product on it.

- **Definition 31.1 (inner product on $\mathcal{L}^2$).** *The* **inner product** *is a function $\langle \cdot, \cdot \rangle : \mathcal{L}^2 \times \mathcal{L}^2 \to \mathbb{R}$ defined by*

$$\langle f, g \rangle = \int fg \, \mathrm{d}\mu.$$

- We note that the inner product is always defined. This is because the Cauchy–Schwarz inequality (Theorem 25.7) tells us that $|\int fg \, \mathrm{d}\mu| \leq \|f\|_2 \|g\|_2$. So, it follows that $\int fg \, \mathrm{d}\mu$ must be finite.

- **Proposition 31.2 (properties of inner product on $\mathcal{L}^2$).** *Let $f, g, h \in \mathcal{L}^2$, and $a, b \in \mathbb{R}$. The following statements are true.*

  1. *(Symmetry) $\langle f, g \rangle = \langle g, f \rangle$.*
  2. *(Linearity in the first argument) $\langle af + bg, h \rangle = a\langle f, h \rangle + b\langle g, h \rangle$.*
  3. *(Positivity) $\langle f, f \rangle \geq 0$, and $\langle f, f \rangle = 0$ implies $f = 0$ almost everywhere.*

- **Definition 31.3 (inner product on L^2).** *Going from $\mathcal{L}^2$ to L^2, we define the inner product on L^2 according to the following equation:*

$$\langle [f]_2, [g]_2 \rangle = \langle f, g \rangle = \int f g \, \mathrm{d}\mu.$$

  The inner product on L^2 is well defined despect the non-determinism in the definition.

- Since writing $[f]_2$, $[g]_2$, and so on is quite handful, we shall denote an element of L^2 by just f, g, and so on.

- **Proposition 31.4 (properties of inner product on L^2).** *Let $f, g, h \in L^2$, and $a, b \in \mathbb{R}$. The following statements are true.*

  1. *(Symmetry) $\langle f, g \rangle = \langle g, f \rangle$.*
  2. *(Linearity in the first argument) $\langle af + bg, h \rangle = a\langle f, h \rangle + b\langle g, h \rangle$.*
  3. *(Positive definiteness) $\langle f, f \rangle \geq 0$, and $\langle f, f \rangle > 0$ if and only if $f \neq 0$.*

- **Definition 31.5 (inner product).** *Let V be a vector space. A function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is called an* **inner product** *if it satisfies the following conditions for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a, b \in \mathbb{R}$.*

  1. *(Symmetry) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.*
  2. *(Linearity in the first argument) $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$.*
  3. *(Positive definiteness) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, and $\langle \mathbf{u}, \mathbf{u} \rangle > 0$ if and only if $\mathbf{u} \neq \mathbf{0}$.*

- **Definition 31.6 (inner product space).** *A vector space endowed with an inner product is called an* **inner product space**.

- **Theorem 31.7.** *L^2 is an inner product space.*

- **Theorem 31.8 (norm from inner product).** *An inner product space is also a normed vector space. The norm function $\| \cdot \| : V \to \mathbb{R}$ is given by $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$.*

- The norm of L^2 is just the $\mathcal{L}^2$-norm.

- **Definition 31.9 (Hilbert space).** *A* **Hilbert space** *is an inner product space that is also a complete normed vector space.*

- **Theorem 31.10.** *L^2 is a Hilbert space.*

32 Hilbert Space Theory

- The definitions and statements in this section applies to any Hilbert spaces, so they apply to any (equivalent classes of functions) in L^2.

- In this section, we let H be a Hilbert space. We denote members of H with boldfaced lowercase letters such as $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$. The inner product is denoted by $\langle \cdot, \cdot \rangle$, and the induced norm is denoted by $\| \cdot \|$ (no subscription).

- **Proposition 32.1 (Cauchy–Schwarz indequality).** $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$.

- **Proposition 32.2 (triangle inequality).** $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

- **Proposition 32.3 (parallelogram law).** $\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$.

- **Proposition 32.4 (polarization identity).** $4\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2$.

- **Definition 32.5 (metric).** *A metric on a set M is a function $d : M \times M \to [0, \infty)$ that satisfies the following properties for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in M$.*

  1. *$d(\mathbf{u}, \mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$.*
  2. *$d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$.*
  3. *$d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$.*

- **Definition 32.6.** *The function $d : H \times H \to [0, \infty)$ defined by $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$ is a metric on H.*

- Once we have a metric, we can define the notation of (1) open ball, (2) open set, (3) limit, (4) limit point, (5) closed set, (6) continuity, etc. that we should be familiar with from point set theory employed.

- **Definition 32.7.** *The inner product is a continous function in both of its arguments. In particular, if $\mathbf{u}_n \to \mathbf{u}$ and $\mathbf{v}_n \to \mathbf{v}$ (in terms of the induced distance d in Definition 32.6), then $\langle \mathbf{u}_n, \mathbf{v}_n \rangle \to \langle \mathbf{u}, \mathbf{v} \rangle$.*

- **Definition 32.8 (perpendicularity).** *Let $\mathbf{u}, \mathbf{v} \in H$ and $W \subset H$. We say that* **$\mathbf{u}$ is perpendicular to $\mathbf{v}$** *if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. We say that* **$\mathbf{u}$ is perpendicular to W** *if $\mathbf{u}$ is perpendicular to every $\mathbf{w} \in W$. Notationally, we write $\mathbf{u} \perp \mathbf{v}$, and $\mathbf{u} \perp W$.*

- **Definition 32.9 (Hilbert subspace).** *Let G be a subset of H. We say that G* **is a subspace of H** *if G is a vector space and G is closed (i.e., G contains all of its limit points).*

- **Proposition 32.10 (perpendicular subspace).** *Let U be a non-empty subset of H. Let $U^\perp$ denote the set of all elements in H that are perpendicular to U. Then, $U^\perp$ is a subspace of H.*

- **Proposition 32.11 (distance from a subspace).** *Let $\mathbf{u} \in H$ and V be a non-empty subspace of H. Let $d(\mathbf{u}, V)$ be defined as*

$$d(\mathbf{u}, V) = \inf\{d(\mathbf{u}, \mathbf{v}) : \mathbf{v} \in V\}.$$

- **Proposition 32.12 (closest point property).** *Let $\mathbf{u} \in H$ and V be a non-empty subspace of H. Then, there is a unique point $\mathbf{v}^* \in V$ such that $d(\mathbf{u}, \mathbf{v}^*) = d(\mathbf{u}, V)$. We say that $\mathbf{v}^*$ is the* **closest point in V to $\mathbf{u}$**.

- **Definition 32.13.** *Let $\mathbf{u} \in H$ and V be a non-empty subspace of H. The* **projecction of $\mathbf{u}$ onto V** *is the closest point in V to $\mathbf{u}$. The* **projection operator**, Π, *maps $\mathbf{u}$ to the projection of $\mathbf{u}$ onto V.*

  For the reason that should be almost immediately, we write the projection of $\mathbf{u}$ as $\Pi\mathbf{u}$.

- **Proposition 32.14.** *Let Π be the projection operator that sends elements of H to a Hilbert subspace V. It has the following properties.*

  1. *Π is linear. For any $\mathbf{u}$, $\mathbf{v} \in H$ and $a, b \in \mathbb{R}$, we have that $\Pi(a\mathbf{u} + b\mathbf{v}) = a\Pi\mathbf{u} + b\Pi\mathbf{b}$. (In other words, Π is like a matrix, so we write it without parentheses.)*
  2. *$\langle \Pi\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \Pi\mathbf{v} \rangle$ for all $\mathbf{u}, \mathbf{v} \in H$.*
  3. *Π is idempotent: $\Pi^2 = \Pi$.*
  4. *$\Pi\mathbf{u} = \mathbf{u}$ for all $\sqcap \in V$, and $\Pi\mathbf{u} = \mathbf{0}$ for all $\mathbf{u} \in V^\perp$.*
  5. *For every $\mathbf{u} \in H$, $\mathbf{u} - \Pi\mathbf{u}$ is orthogonal to V.*

- **Proposition 32.15.** *Let Π be a projection operation of H onto a subspace V. Then, $\mathbf{u} = \Pi\mathbf{u} + (\mathbf{u} - \Pi\mathbf{u})$ is the unique representation of $\mathbf{u}$ as the sum of a vector in V and a vector in $V^\perp$. Moreover, $\mathbf{u} - \Pi\mathbf{u}$ is the projection of $\mathbf{u}$ into $V^\perp$.*

- **Proposition 32.16.** *For any subspace V of H, $(V^\perp)^\perp = V$.*

33 Conditional Expectation

- **Definition 33.1.** *Let X be a random variable that can only take a finitely or countably many infinite values: $X(\omega) \in \{x_1, x_2, \dots\}$ for all $\omega \in \Omega$. Let Y be another random variable. If $P(X = x_j) > 0$, the* **conditional expectation of Y given the event $\{X = x_j\}$** *is defined to be*

$$E[Y|X = x_j] = \int Y \, \mathrm{d}Q$$

where Q is the probability measure defined by $Q(E) = P(Y \in E|X = x_j)$ for all $E \in \mathcal{E}$, provided that $Y \in \mathcal{L}^1(\Omega, \mathcal{E}, Q)$.

- **Definition 33.2.** *Let X be a random variable that can only take a finite or countably many infinite values: $X(\omega) \in \{x_1, x_2, \dots\}$ for all $\omega \in \Omega$. Let*

$$f(x) = \begin{cases} E[Y|X = x], & P(X = x) > 0 \\ \textit{arbitrary value}, & P(X = x) = 0 \end{cases}.$$

Then, the **conditional expectation of Y given X** *is defined to be*

$$E[Y|X] = f(X).$$

It is defined only if $Y \in \mathcal{L}^2(\Omega, \mathcal{E}, Q_j)$ where Q_j is the probability measure defined by $Q_j(E) = P(Y \in E|X = x_j)$ for all j such that $P(X = x_j) \neq 0$.

- **Proposition 33.3.** *Let X and Y be random variables that can only take a finite or countably many infinite values. That is, $X(\omega) \in \{x_1, x_2, \dots\}$ and $Y(\omega) \in \{y_1, y_2, \dots\}$ for all $\omega \in \Omega$. We have that*

$$E[Y|X] = \sum_{i=1}^{\infty} \sum_{\{x_j : P(X=x_j) \neq 0\}} y_i P(Y = y_i | X = x_j),$$

provided that the sum converges absolutely.

- Continuing from the above proposition, we can look at $E[Y|X]$ in a new perspective. First, it is a random variable, so let us denote it with $\hat{Y}$.

  Consider the sample space Ω. We have that we may think of Ω as being partitioned into a collection of disjoint sets $\{X = x_1\}$, $\{X = x_2\}$, and so on. We have that

$$\begin{aligned} \int_{\{X=x_j\}} \hat{Y} \, \mathrm{d}P &= \Big(\sum_{i=1}^{\infty} y_i P(Y = y_i | X = x_j) \Big) P(X = x_j) \\ &= \sum_{i=1}^{\infty} y_i P(Y = y_i \wedge X = x_j) \\ &= \int_{\{X=x_j\}} Y \, \mathrm{d}P. \end{aligned}$$

  Now, consider the σ-algebra generated by X. We have that it contains sets of the form $\{\omega : X(\omega) \in E\}$ where $E \subseteq \{x_1, x_2, \dots\}$. Hence, such a set is a disjoint union of the $\{X \in x_j\}$'s. As a result, we have that

$$\int_E \hat{Y} \, \mathrm{d}P = \int_E Y \, \mathrm{d}P.$$

  This is the identity we will be using to define conditinal expectation in the general case.

- **Theorem 33.4 (Kolmogorov, 1933).** *Let $X \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{E}$. Then, there exists a random variable Y such that the following properties hold.*

  1. *Y is $\mathcal{L}^1(\Omega, \mathcal{F}, P)$.*
  2. *For every set $F \in \mathcal{F}$, we have that $\int_F Y \, \mathrm{d}P = \int_F X \, \mathrm{d}P$.*

  Moreover, if Y' is another random variable with these properties, then $Y' = Y$ almost surely (i.e., $P(Y' = Y) = 1$). A random variable with the above properties is called a version of **the conditional expectation $E[X|\mathcal{F}]$ of X given $\mathcal{F}$.** *We write $Y = E[X|\mathcal{F}]$.*

- One way to prove a special case of the theorem is to use the Radon–Nikodym theorem. If X is a non-negative random variable in $\mathcal{L}^1(\Omega, \mathcal{E}, P)$, then we can define a measure $Q : \mathcal{F} \to [0, \infty]$ by

$$Q(F) = \int_F X \, \mathrm{d}P$$

  for all $F \in \mathcal{F}$. Because $E[|X|] = \int |X| \, \mathrm{d}P < \infty$, we have that $Q(F) \leq E[|X|] < \infty$ for all F, so it is a finite measure. If we assume further that $Q \ll P$ (i.e., no point masses and other pathologies), then the Radon–Nikodym derivative exists and is unique P-almost everywhere. We can now define $Y := \mathrm{d}Q/\mathrm{d}P$, and Y would have all the properties that we want.

- Note, however, there are proofs of Theorem 33.4 that does not require the Radon–Nikodym theorem. They are presented in [Williams, 1991] and [Jacod and Protter, 2004].

- **Definition 33.5.** *Let $X, Y \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$. The* **conditional expectation** *$E[Y|X]$ is defined to be $E[Y|\sigma(X)]$. Also, for $X_1, X_2, \ldots \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$, we denote $E[Y|\sigma(X_1, x_2, \ldots)]$ with $E[Y|X_1, X_2, \ldots]$.*

- **Definition 33.6.** *Let $X : \Omega \to \mathbb{R}^d$ and $Y : \Omega \to \mathbb{R}$ be random variables. Then, Y is $\sigma(X)$-measurable if and only if there exists a Borel function $f : \mathbb{R}^d \to \mathbb{R}$ such that $Y = f(X)$.*

- **Corollary 33.7.** *Let $X \in \mathcal{L}^2(\Omega, \mathcal{E}, P)$, and $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{E}$. Then, the conditional expectation $Y = E[X|\mathcal{F}]$ is a version of the projection of X onto the Hilbert subspace $L^2(\Omega, \mathcal{F}, P)$. As a result, Y (or any other function in the same equivalent class) is a $\mathcal{F}$-measurable function that minimizes $\|Y - X\|_2$ or, equivalently, $E[(Y - X)^2]$.*

- **Corollary 33.8.** *$Y = E[X|\mathcal{F}]$ (as an equivalent class) is the unique element in $L^2(\Omega, \mathcal{F}, P)$ such that*

$$E[XZ] = E[YZ]$$

  for all $Z \in L^2(\Omega, \mathcal{F}, P)$.

- **Theorem 33.9 (properties of conditional expectation).** *Let $X \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$. Let $\mathcal{F}$ and $\mathcal{G}$ be sub-σ-algebras of $\mathcal{E}$. The following properties are true.*

  1. *If X is $\mathcal{F}$-measurable, then $E[X|\mathcal{F}] = X$.*
  2. *If Y is any version of $E[X|\mathcal{F}]$, then $E[Y] = E[X]$. (In other words, $E[E[X|\mathcal{F}]] = E[X]$.)*
  3. *(Linearity) Let X_1 and X_2 be random variables and $a_1, a_2 \in G$. Then,*

$$E[a_1X_1 + a_2X_2|\mathcal{F}] = a_1E[X_1|\mathcal{F}] + a_2E[X_2|\mathcal{F}].$$

  4. *(Positivity) If $X \geq 0$, then $E[X|\mathcal{F}] \geq 0$.*
  5. *(Tower property) If $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, then $E[E[X|\mathcal{F}]|\mathcal{G}] = E[X|\mathcal{G}]$.*
  6. *(Monotone convergence theorem) If $0 \leq X_n$ for all n and $X_n \uparrow X$, then, $E[X_n|\mathcal{F}] \uparrow E[X|\mathcal{F}]$.*
  7. *(Fatou's lemma) If $X_n \geq 0$, then $E[\liminf_{n\to} X_n|\mathcal{F}] = \liminf_{n\to\infty} E[X_n|\mathcal{F}]$.*

8. *(Dominated convergence theorem) Let $Y \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$. If $|X_n(\omega) \leq Y(\omega)$ for all n and $X_n \to X$ almost surely, then $E[X_n|\mathcal{F}] \to E[X|\mathcal{F}]$.*
9. *(Jensen's inequality) If $c : \mathbb{R} \to \mathbb{R}$ is convex and $|c(X)| \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$, then $E[c(X)|\mathcal{F}] \geq c(E[X|\mathcal{F}])$ almost surely.*
10. *$\|E[X|\mathcal{F}]\|_p \leq \|X\|_p$ for all $p \geq 1$.*
11. *If Y is $\mathcal{F}$-measurable and bounded, then $E[YX|\mathcal{F}] = YE[X|\mathcal{F}]$ almost surely.*
12. *If $\mathcal{G}$ is independent of $\sigma(\sigma(X), \mathcal{F})$, then $E[X|\sigma(\mathcal{F}, \mathcal{G})] = E[X|\mathcal{F}]$ almost surely. In particular, if X is independent of $\mathcal{F}$, then $E[X|\mathcal{F}] = E[X]$.*

- Let X and Z be random variables with joint PDF $f_{X,Z}(x, z)$. The marginal PDF of X and Z are given by

$$f_X(x) = \int_{\mathbb{R}} f_{X,Z}(x, z)\,\mathrm{d}z,$$
$$f_Z(z) = \int_{\mathbb{R}} f_{X,Z}(x, z)\,\mathrm{d}x.$$

We can define **elementary conditional pdf $f_{X|Z}$ of X given Z** as follow:

$$f_{X|Z}(x|z) = \begin{cases} f_{X,Z}(x, z)/f_Z(z), & \text{if } f_Z(z) \neq 0 \\ 0, & \text{otherwise} \end{cases}.$$

Proposition 33.10. *Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function such that*

$$E[|h(X)|] = \int_{\mathbb{R}} |h(x)| f_X(x)\,\mathrm{d}x < \infty.$$

Set

$$g(z) = \int_{\mathbb{R}} h(x) f_{X|Z}(x|z)\,\mathrm{d}x.$$

Then, $Y = g(Z)$ is a version of the conditional expection of $h(X)$ given $\sigma(Z)$.

34 Martingales

- **Definition 34.1 (filtration).** *Let $\mathcal{E}$ be a σ-algebra. A sequence of sub-σ-algebras $\{\mathcal{E}_n \subseteq \mathcal{E} : n \geq 0\}$ is called a* **filtration in** *$\mathcal{E}$ if $\mathcal{E}_0 \subseteq \mathcal{E}_1 \subseteq \mathcal{E}_2 \subseteq \cdots \subseteq \mathcal{E}$. We define*

$$\mathcal{E}_\infty = \sigma\Big(\bigcup_{n=0}^{\infty} \mathcal{E}_n\Big).$$

Note that $\mathcal{E}_\infty \subseteq \mathcal{E}$.

- Here's a way to think about filtrations. At the start of our random experiment, an element of the sample space ω is picked according to P. The sample ω determines which events would occur. At each time step, we observe more and more events. Each $\mathcal{E}_n$ as representing all observable events up to time n. Typically, $\mathcal{E}_0 = \{\emptyset, \Omega\}$, which means that there's no information available at all at the start of our observation.

- **Definition 34.2 (filtered space).** *A* **filtered space** *is a quadruple $(\Omega, \mathcal{E}, \{\mathcal{E}_n\}_{n\geq 0}, P)$ where $(\Omega, \mathcal{E}, P)$ is a probability space and $\{\mathcal{E}_n\}_{n\geq 0}$ is a filteration in $\mathcal{E}$.*

- **Definition 34.3 (discrete-time stochastic process).** *A* **(discrete-time) stochastic process** *is a sequence* $\mathbf{X} = \{X_n\}_{n\geq 0}$ *of random variables. It induces the* **natural filtration** $\mathcal{E}_n = \sigma(X_0, X_1, \ldots, X_n)$.

- **Definition 34.4 (adapted process).** *A stochastic process* $\{X_n\}_{n\geq 0}$ *is said to be adapted to a filtration* $\{\mathcal{E}_n\}_{n\geq 0}$ *if* X_n *is* $\mathcal{E}_n$*-measurable for all* $n \geq 0$.

- Here's one way to think about adapted process. Again, recall that ω is picked at the start to the random experiment. Then, from time $n = 0, 1, 2, \ldots$, information about ω is revealed to us through the value of X_n. When the process is adapted to $\{\mathcal{E}\}_{n\geq 0}$, the σ-algebra $\mathcal{E}_n$ is one where all of X_0, X_1, $\ldots$, X_n are measurable, so it would allow us to evaluate $P(X_i \in E)$ for any $\mathcal{E} \in \mathcal{E}$. So, $\mathcal{E}_n$ contains all relevant information about X_0, X_1, $\ldots$, X_n.

- It should be clear that a stochastic process is adapted to its natural filtration.

- **Definition 34.5 (martingle).** *A stochastic process* $\mathbf{X}$ *is called a* **martingle relative to** $(\{\mathcal{E}_n\}_{n\geq 0}, P)$ *if the followign conditions are satisfied.*

  1. $\mathbf{X}$ *is adapted to* $\{\mathcal{E}_n\}_{n\geq 0}$.
  2. $X_n \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ *for all* n.
  3. $E[X_n|\mathcal{E}_{n-1}] = X_{n-1}$ *almost surely for all* $n \geq 1$.

  A **supermartingle** *is defined similarly with Condition 3 replaced by "*$E[X_n|\mathcal{E}_{n-1}] \leq X_{n-1}$ *almost surely." A* **submatingle** *is defined with Condition 3 replaced by "*$E[X_n|\mathcal{E}_{n-1}] \geq X_{n-1}$ *almost surely."*

- The word martingle comes from a gambling context. Let's see an example. We can imagine that we can play a gambling games in rounds. In each round, you flip a coin. If it turns up head, you win \$1. If it turns up tail, you lose \$1. You may say X_n is the money you have won at the end of round n. So, $X_0 = 0$. Moreover,

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

  So,

$$E[X_n|\mathcal{E}_{n-1}] = E[X_n|X_{n-1}] = X_{n-1} + 2p - 1.$$

  If $p = 0.5$, then $\mathbf{X}$ is a martingle. The game is fair because

$$E[X_n - X_{n-1}|\mathcal{E}_{n-1}] = 0.$$

  On the other hand, if $p < 0.5$, then $\mathbf{X}$ is a supermartingale, and the game is stacked against you. If $p > 0.5$, then $\mathbf{X}$ is a submartingale, and the game is rigged in your favor.

- **Proposition 34.6.** *If* $\mathbf{X}$ *is a martingle relative to* $\{\mathcal{E}_n\}_{n\geq 0}$*, then* $E[X_n|\mathcal{E}_m] = X_m$ *for all* $m \leq n$*. The equation changes to* $E[X_n|\mathcal{E}_m] \leq X_m$ *for a supermartingale and* $E[X_n|\mathcal{E}_m] \geq X_m$ *for a submartingale.*

- **Proposition 34.7.** *If* $\mathbf{X}$ *is a martingle, then* $E[X_n] = E[X_0]$ *for all* n.

  Proof. $E[X_n] = E[E[X_n|\mathcal{F}_0]] = E[X_0]$. □

- **Proposition 34.8.** *If* $\mathbf{X}$ *is a martingale, and if* f *is a convex function and* $f(X_n) \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ *for each* n*, then* $\{f(X_n)\}_{n\geq 0}$ *is a submartingale.*

  This means that if $\mathbf{X}$ is a martingale, then $|\mathbf{X}| = \{|X_n|\}_{n\geq 0}$ is a submartingale.

- **Theorem 34.9 (Doob decomposition).** *Let* $\mathbf{X}$ *be a submartingale adapted to* $\{\mathcal{E}_n\}_{n\geq 0}$*. There exists a martingale* $\mathcal{M}$ *and a stochastic process* $\mathbf{A}$ *with* $A_{n+1} \geq A_n$ *almost surely and* A_{n+1} *being* $\mathcal{E}_n$ *measurable for all* $n \geq 0$ *such that*

$$X_n = X_0 + M_n + A_n$$

with $M_0 = A_0 = 0$*. Moreover, such a decomposition in unique. If* $\mathbf{X}$ *is a supermartingale, the decomposition is*

$$X_n = X_0 + M_n - A_n.$$

- Let $\{\mathcal{E}_n\}_{n\geq 0}$ be a filtration. Let $Y \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$. Define $X_n = E[Y|\mathcal{E}_n]$. We have that

$$E[X_n|\mathcal{E}_{n-1}] = E[E[Y|\mathcal{E}_n]|\mathcal{E}_{n-1}] = E[Y|\mathcal{E}_{n-1}] = X_{n-1}.$$

  So, $\mathbf{X}$ is a martingle. A way to think about this martingle is as follows. With each $\mathcal{E}_n$, we get more information about ω. $X_n = E[Y|\mathcal{E}_n]$ is the closest random variable to Y given the information we have so far. So, X_∞ is the best approximation to Y given all the information.

- **Definition 34.10.** *Let* $Y \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$*. A martingle* $\mathbf{X}$ *adapted to* $\{\mathcal{E}_n\}_{n\geq 0}$ *is said to be* **closed by** Y *if* $X_n = E[Y|\mathcal{E}_n]$.

35 Stopping Times

- **Definition 35.1 (stopping time).** *A map* $T : \Omega \to \{0, 1, 2, \ldots, \infty\}$ *is called a* **stopping time** *if*

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

  for all $n \leq \infty$ *or, equivalently, if*

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

  for all $n \leq \infty$*.*

  Intuitively, with the information up to the n-th time step, we have all the information to decide whether $T \leq n$.

- For example, let $\mathbf{X}$ be adapted to $\{\mathcal{E}_n\}_{n\geq 0}$. Let $B \in \mathcal{B}(\mathbb{R})$. Let $T = \inf\{n \geq 0 : X_n \in B\}$. In other words, T is the first time X_n enters B. Obviously, T is a stopping time because we can decide whether $T \leq n$ by the history of X_0, X_1, $\ldots$, X_n only.

- **Definition 35.2 (bounded stopping time).** *We say that a stopping time* T *is* **bounded** *if there exists a constant* c *such that* $P(T \leq c) = 1$*.*

- **Definition 35.3.** *If* T *is a finite stopping time, we denote by* X_T *the random variable* $X_{T(\omega)}(\omega)$*. This random variable takes the value* X_n *whenever* $T = n$*.*

- **Proposition 35.4.** *Let* T *be a stopping time bounded by* c*, and let* $\mathbf{X}$ *be a martingle. Then,*

$$E[X_T] = E[X_0].$$

- **Definition 35.5 (stopping time σ-algebra).** *Let* T *be a stopping time. The* **stopping time σ-algebra** $\mathcal{E}_T$ *is defined to be*

$$\mathcal{E}_T = \{E \in \mathcal{E} : E \cap \{T \leq n\} \in \mathcal{E}_n \text{ for all } n\}$$

- **Proposition 35.6.** *The stopping time σ-algebra is a σ-algebra.*
- **Proposition 35.7.** *Let S, T be stopping times with $S \leq T$. Then, $\mathcal{E}_S \subseteq \mathcal{E}_T$.*
- **Proposition 35.8.** *X_T is $\mathcal{F}_T$-measurable.*
- **Theorem 35.9 (Doob's optinal sampling theorem).** *Let $\mathbf{X}$ be a martingle. Let S, T be stopping times bounded by a constant c with $S \leq T$. Then, $E[X_T|\mathcal{E}_S] = X_S$ almost surely.*
- **Theorem 35.10.** *Let $\mathbf{X}$ be a stochastic process adapted to $\{\mathcal{E}_n\}_{n\geq 0}$. Suppose $X_n \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ for all n, and $E[X_T] = E[X_0]$ for all bounded stopping time T. Then, $\mathbf{X}$ is a martingale.*
- **Theorem 35.11.** *Let T be a stopping time bounded by $c \in \mathbb{N}$. Let $\mathbf{X}$ be a submartingale. Then, $E[X_T] \leq E_c$.*

  In particular, $E[X_{\min(T,n)}] \leq E[X_n]$.

36 Martingale Inequalities

- We will now discuss various inequalities concerning martintgales. We shall deal with a filtered space $(\Omega, \mathcal{E}, \{\mathcal{E}_n\}_{n\geq 0}, P)$. We let $\mathbf{X} = \{X_n\}_{n\geq 0}$ be a sequence of random variables adapted to $\{\mathcal{E}_n\}_{n\geq 0}$ that is integrable (i.e., $X_n \in \mathcal{L}^1(\Omega, \mathcal{E}_n, P)$). Let $X_n^* = \sup_{i\leq n} |X_i|$. We have that $\{X_n^*\}_{n\geq 0}$ is a non-negative and non-decreasing stochastic process and a submartingale.
- **Theorem 36.1 (Doob's first martingale inequality).** *Let $\mathbf{X}$ be a martingale or a non-negative submartingle. Then*

$$P(X_n^* \geq a) \leq \frac{E[|X_n|]}{a}.$$

- **Theorem 36.2 (Doob's $\mathcal{L}^p$ martingle inequalities #1).** *Let $\mathbf{X}$ be a martingale or a non-negative submartingle. Let $1 < p < \infty$. There exists a constant c depending on p such that*

$$E[(X_n^*)^p] \leq cE[|X_n|^p].$$

- **Theorem 36.3 (Doob's $\mathcal{L}^p$ martingle inequalities #2).** *Let $\mathbf{X}$ be a martingale or a non-negative submartingle. Let $1 < p < \infty$. There exists a constant c depending on p such that*

$$E[(X_n^*)^p]^{\frac{1}{p}} \leq \frac{p}{p-1} E[|X_n|^p]^{\frac{1}{p}}.$$

  Equivalently,

$$\|X_n^*\|_p \leq \frac{p}{p-1} \|M_n\|_p.$$

- **Definition 36.4 (upcrossing).** *Let $\mathbf{X}$ be a submartingale. Let $a < b$. The* **number of upcrossings of an inteval** *$[a, b]$ is the nujmber of times a process crosses from below a to above b at a later time.*
- The number of upcrossing can be defined as follows. We define stopping times $T_0, T_1, \ldots$ and $S_1, S_2, \ldots$ recursively. We start with

$$T_0 = 0.$$

  Then,

$$S_{j+1} = \min\{k > T_j : X_k \leq a\},$$
$$T_{j+1} = \min\{k > S_{j-1} : X_k \geq b\}.$$

So,

$$S_{j+1} = \text{earliest time when } \mathbf{X} \text{ first dropped down to } a \text{ after the } j\text{th time it goes above } b,$$
$$T_{j+1} = \text{earliest time when } \mathbf{X} \text{ goes above } b \text{ after the } (j+1)\text{th time it drops down to } a.$$

We can now define

$$U_n = \max\{j : T_j \leq n\},$$

and this equals to the number of upcrossings of $[a, b]$ before time n.

- **Theorem 36.5 (Doob's upcrossing inquality).** *Let $\mathbf{X}$ be a submartingale. Let $a < b$, and let U_n be the number of upcrossings of $[a, b]$ before time n. Then,*

$$E[U_n] \leq \frac{1}{b-a} E[\max(X_n - a, 0)].$$

37 Convergence of Random Variables

- In this section, we deal with a sequence of random variable $\{X_n\}_{n\in\mathbb{N}}$.
- **Definition 37.1 (pointwise convergence).** *We say that $\{X_n\}_{n\in\mathbb{N}}$* **converges (pointwise)** *to a random variable X if*

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

for all ω. We write "$X_n \to X$."

This notion of convergence is often too strong. We introduce three more notions of convergence that are weaker.

- **Definition 37.2 (almost sure convergence).** *We say that $\{X_n\}_{n\in\mathbb{N}}$* **converges almost surely** *to a random variable X if*

$$P(N) = 0 \qquad \textit{where} \qquad N = \left\{\omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\right\}.$$

We write "$X_n \to X$ almost surely" or "$X_n \to X$ a.s." or "$X_n \xrightarrow{a.s.} X$."

- **Definition 37.3 (convergence in $\mathcal{L}^p$).** *Let $1 \leq p < \infty$. We say that $\{X_n\}_{n\in\mathbb{N}}$* **converges in $\mathcal{L}^p$** *to a random variable X if X_n, X are in $\mathcal{L}^p(\Omega, \mathcal{E}, P)$ and*

$$\lim_{n\to\infty} E\big[|X_n - X|^p\big] = \lim_{n\to\infty} \big(\|X_n - X\|_p\big)^p = 0.$$

We write "$X_n \xrightarrow{\mathcal{L}^p} X$."

- **Definition 37.4 (convergence in probability).** *We say that $\{X_n\}_{n\in\mathbb{N}}$* **converges in probability** *to a random variable X if, for any $\varepsilon > 0$, we have*

$$\lim_{n\to\infty} P\big(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}\big) = 0.$$

Equivalently, for any $\varepsilon > 0$, $\delta > 0$, there exists $N = N(\varepsilon, \delta)$ such that $P(|X_n - X| > \varepsilon) < \delta$ for all $n \geq N$. We write "$X_n \xrightarrow{P} X$."

- **Theorem 37.5.** *$X_n \xrightarrow{P} X$ if and only if*

$$\lim_{n\to\infty} E\left[\frac{|X_n - X|}{1 + |X_n - X|}\right] = 0.$$

- Convergence in probability is the weakest notion of convergence that has been introduced so far.

  Proposition 37.6. *The following are true.*

  1. $X_n \xrightarrow{\mathcal{L}^p} X \implies X_n \xrightarrow{P} X$.
  2. $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X$.

- **Proposition 37.7.** *If $X_n \xrightarrow{P} X$, then there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ such that $X_{n_k} \to X$ almost surely.*

- **Proposition 37.8.** *Let $Y \in \mathcal{L}^p(\Omega, \mathcal{E}, P)$. Let $X_n \xrightarrow{P} X$ and $|X_n| \leq Y$ for all n. Then, we have that $|X| \in \mathcal{L}^p(\Omega, \mathcal{E}, P)$ and $X_n \xrightarrow{\mathcal{L}^p} X$.*

- **Proposition 37.9.** *Let f be a continuous function.*

  1. $X_n \xrightarrow{a.s.} X \implies f(X_n) \xrightarrow{a.s.} f(X)$.
  2. $X_n \xrightarrow{P} X \implies f(X_n) \xrightarrow{P} f(X)$.

38 Martingale Convergence Theorems

- **Theorem 38.1 (martingale convergence theorem #1).** *Let $\{X_n\}_{n\geq 0}$ be a submartingale such that $\sup_{n\geq 0} E[\max(X_n, 0)] < \infty$. Then, $X = \lim_{n\to\infty} X_n$ exists almost surely. Moreover, $X \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$.*

  However, it is not true in general that $X_n \xrightarrow{\mathcal{L}^1} X$.

- **Corollary 38.2.** *Let $\{X_n\}_{n\geq 0}$ is a non-negative supermartingale or a martingle bounded above or bounded below. Then, $X = \lim_{n\to\infty} X_n$ exists almost surely. Moreover, $X \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$.*

- **Definition 38.3 (uniformly integrable collection of random variables).** *A subset H of L^1 is said to be* **uniformly integrable collection of random variables** *if*

$$\lim_{c\to\infty} \sup_{X\in H} E[\mathbb{1}_{\{|X|\geq c\}}|X|] = 0.$$

- **Proposition 38.4.** *Let H be a class of random variables.*

  1. *If $\sup_{X\in H} E[|X|^p] < \infty$ for some $p > 1$, then H is uniformly integrable.*
  2. *If there exists a random variable $Y \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ such that $|X| < Y$ almost surely for all $X \in H$, then H is uniformly integrable.*

- **Theorem 38.5 (martingale convergence theorem #2).** *Let $\mathbf{X}$ be a martingale that is a uniformly integrable collection of random variables. Then, $\lim_{n\to\infty} X_n = X_\infty$ exists almost surely, $X_\infty \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$, and $X_n \xrightarrow{\mathcal{L}^1} X_\infty$. Moreover, $X_n = E[X|\mathcal{F}_\infty]$.*

  Conversely, let $Y \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$. Consider the martingle $X_n = E[Y|\mathcal{E}_n]$ that is closed by Y. Then, $\{X_n\}_{n\geq 0}$ is a uniformly integrable collection of random variables.

- **Corollary 38.6.** *Let $\{\mathcal{E}_n\}_{n\geq 0}$ be a filtration. Let $\mathcal{E}_\infty = \sigma(\bigcup_{n=0}^\infty \mathcal{E}_n)$. If $Y \in \mathcal{L}^1(\Omega, \mathcal{E}_\infty, P)$, then $E[Y|\mathcal{E}_n] \xrightarrow{\mathcal{L}^1} Y$.*

- When we define martingales, the indices of the random variables are non-negative. However, we can also consider non-positive indices.

  Definition 38.7 (backwards martingale). *Let $\{\mathcal{E}_{-n}\}_{n\geq 0}$ be an increasing sequence of sub-σ-algebras of $\mathcal{E}$. (That is, $\mathcal{E}_{-n-1} \subseteq \mathcal{E}_{-n}$ for all $n \geq 0$. In other words, a filtration with non-positive indices.) A* **backwards martingale** *is a sequence $\{X_{-n}\}_{n\geq 0}$ of random variables such that the following properties are satisfied.*

  1. *X_{-n} is $\mathcal{E}_{-n}$-measurable for all $n \geq 0$.*
  2. *$X_{-n} \in \mathcal{L}^1(\Omega, \mathcal{E}, P)$ for all $n \geq 0$.*
  3. *$E[X_{-n}|\mathcal{E}_{-n-1}] = X_{-n-1}$ for all $n \geq 0$.*

- **Theorem 38.8 (backwards martingale convergence theorem).** *Let $\{X_{-n}\}_{n\geq 0}$ be a backward martingale adapted to a filtration $\{\mathcal{E}_{-n}\}_{n\geq 0}$. Let $\mathcal{E}_{-\infty} = \bigcap_{n=0}^\infty \mathcal{E}_{-n}$. Then, the sequence $\{X_{-n}\}_{n\geq 0}$ converges almost surely and in $\mathcal{L}^1$ to a limit X as $n \to \infty$. This X is a member of $\mathcal{L}^1(\Omega, \mathcal{E}, P)$.*

- **Theorem 38.9 (strong law of large numbers).** *Let $\{X_n : n \in \mathbb{N}\}$ be a sequence of i.i.d. random variables in $\mathcal{L}^1(\Omega, \mathcal{E}, P)$. Then,*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{a.s.} E[X_1].$$

39 Weak Convergence

- The notion of weak convergence deals with convergence of probability measures. When applied to the probabiliy measures induced by random variables, it gives another notion of convergence or random variables. However, the values of the random variable in question is not of any concern at all. As a result, weak convergence is very different from other types of convergence discussed in the last section.

- **Definition 39.1 (weak convergence).** *Let $\{\mu_n\}_{n\geq 0}$ and μ be probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. We say that $\{\mu_n\}_{n\geq 0}$* **converges weakly to** *μ if*

$$\lim_{n\to\infty} \int f \, \mathrm{d}\mu_n = \int f \, \mathrm{d}\mu$$

  for each $f : \mathbb{R}^d \to \mathbb{R}$ that is continuous and bounded.

- **Definition 39.2 (convergence in distribution).** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}^d$-valued random variables. We say that X_n* **converges in distribution** *to X if the distribution measures $\{P_{X_n}\}_{n\in\mathbb{N}}$ converges weakly to P_X. In other words,*

$$\lim_{n\to\infty} E[f(X_n)] = E[f(X)]$$

  for all $f : \mathbb{R}^d \to \mathbb{R}$ that is continuous and bounded. We write "$X_n \xrightarrow{\mathcal{D}} X$."

- One thing to note is that the limit X might not be in the same probability space as the X_n's. So, convergence in distribution is really different from other notions of convergence.

- However, if the limit is in the same space as the sequence, we can say nice things about it.

Proposition 39.3. *Let $\{X_n\}_{n\in\mathbb{N}}$ and X be random variables defined on a given probability space $(\Omega, \mathcal{E}, P)$. Then,*

$$X_n \xrightarrow{P} X \implies X_n \xrightarrow{\mathcal{D}} X.$$

- **Proposition 39.4.** *Let $\{X_n\}_{n\in\mathbb{N}}$ and X be random variables defined on a given probability space $(\Omega, \mathcal{E}, P)$. If $X_n \xrightarrow{\mathcal{D}} X.$ and X is equal to a constant almost surely, then $X_n \xrightarrow{P} X$.*
- **Proposition 39.5.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real-valued random variables.*

  1. *If $X_n \xrightarrow{\mathcal{D}} X$, then $F_{X_n}(x) \to F_X(x)$ for all x in the set D of continuity points of F_X. That is,*

  $$D = \{x : F_X(x^-) = F_X(x)\}.$$

  We also have that D is a dense subset of $\mathbb{R}$.

  2. *If $F_{X_n}(x) \to F_X(x)$ for all x in a dense subset of $\mathbb{R}$, then $X_n \xrightarrow{\mathcal{D}} X$.*

- **Proposition 39.6.** *Let $\{X_n\}_{n\in\mathbb{N}}$ and X be $\mathbb{R}^d$-valued random variables with probability densitie functions $\{f_{X_n}\}_{n\in\mathbb{N}}$ and f_X, respetively. If $f_n \to f$ pointwisely, then $X_n \xrightarrow{\mathcal{D}} X$.*
- **Proposition 39.7.** *Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}(\mathbb{R})$. Suppose*

$$\lim_{m\to\infty} \sup_{n\in\mathbb{N}} \mu_n([-m, m]^c) = 0.$$

  Then, there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ such that $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ converges weakly.

- **Definition 39.8 (Lipschitz continuity).** *A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be* **Lipschitz continous** *if $|f(\mathbf{x}) - f(\mathbf{y})| \le k\|\mathbf{x} - \mathbf{y}\|$ for some constant $k > 0$ and for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.*
- **Proposition 39.9.** *Let $\{X_n\}_{n\in\mathbb{N}}$ and X be $\mathbb{R}^d$-valued random variables. Then, $X_n \xrightarrow{\mathcal{D}} X$ if and only if $E[g(X_n)] \to E[g(X)]$ for all bounded Lipschitz continuous function $g : \mathbb{R}^d \to \mathbb{R}$.*
- **Definition 39.10 (uniform continuity).** *A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be* **uniformly continuous** *if, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have that $\|\mathbf{x} - \mathbf{y}\| < \delta \implies |f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$.*
- **Corollary 39.11.** *Let $\{X_n\}_{n\in\mathbb{N}}$ and X be $\mathbb{R}^d$-valued random variables. Then, $X_n \xrightarrow{\mathcal{D}} X$ if and only if $E[g(X_n)] \to E[g(X)]$ for all bounded uniformly continuous function $g : \mathbb{R}^d \to \mathbb{R}$.*

  This follows from the fact that any Lipschitz continuous function is also uniformly continuous.
- **Theorem 39.12 (Slutsky's theorem).** *Let X, $\{X_n\}_{n\in\mathbb{N}}$, and $\{Y_n\}_{n\in\mathbb{N}}$ be $\mathbb{R}^d$-valued random variables. If $X_n \xrightarrow{\mathcal{D}} X$ and $\|X_n - Y_n\| \xrightarrow{P} 0$, then $Y_n \xrightarrow{\mathcal{D}} X$.*
- **Proposition 39.13.** *Let $\{X_n\}_{n\in\mathbb{N}}$ and X be random variables that take at most countably infinitely many values. Then, $X_n \xrightarrow{\mathcal{D}} X$ if and only if $P(X_n = x) = P(X = x)$ for all possible values x that these varialbles can take.*

40 Characteristic Functions

- **Definition 40.1 (Fourier transform of a measure).** *Let μ be a measure on $\mathcal{B}(\mathbb{R}^d)$. Its* **Fourier transform** *is a complex-valued function $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\hat{\mu}(\boldsymbol{\xi}) = \int e^{i(\boldsymbol{\xi}\cdot\mathbf{x})} \, \mathrm{d}\mu(\mathbf{x}) = \int \cos(\boldsymbol{\xi}\cdot\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) + i \int \sin(\boldsymbol{\xi}\cdot\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x})$$

for any $\boldsymbol{\xi} \in \mathbb{R}^d$.

- **Definition 40.2 (characteristic function).** *Let X be a $\mathbb{R}^d$-valued random variable. Its* **characteristic function** *is the function $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\varphi_X(\boldsymbol{\xi}) = E[e^{i(\boldsymbol{\xi}\cdot X)}] = \int e^{i(\boldsymbol{\xi}\cdot\mathbf{x})}\,\mathrm{d}P_X(\mathbf{x}) = \widehat{P_X}(\boldsymbol{\xi})$$

where P_X is the probability distribution measure (Proposition 19.4) of X.

- Why do we study characteristic functions? Well, it can do several things.
  1. It can be used to compute its random variable's moments: $E[X]$, $E[X^2]$, $E[X^3]$, ...
  2. It can be used to prove the Central Limit Theorem.

- **Proposition 40.3 (properties of characteristic functions).** *Let X be $\mathbb{R}^d$-valued random variable.*
  1. $\varphi_X(\mathbf{0}) = 1$.
  2. $|\varphi_X(\boldsymbol{\xi})| \leq 1$ *for all* $\boldsymbol{\xi} \in \mathbb{R}^d$.
  3. φ_X *is a continous function.*
  4. $\varphi_{(-X)}(\boldsymbol{\xi}) = \overline{\varphi_X(\boldsymbol{\xi})}$ *for all* $\boldsymbol{\xi} \in \mathbb{R}$.
  5. $\varphi_{(AX+\boldsymbol{b})}(\boldsymbol{\xi}) = e^{i(\mathbf{b}\cdot\boldsymbol{\xi})}\varphi_X(A^T\boldsymbol{\xi})$ *for all* $\boldsymbol{\xi}, \boldsymbol{b} \in \mathbb{R}^d$ *and* $A \in \mathbb{R}^{d\times d}$.

- **Theorem 40.4.** *Let X be a real-valued random variable. If $E[|X|^m] < \infty$, for some integer mgeq1. Then, the characteristic function φ_X has continuous partial derivative up to order m, and*

$$\varphi_X^{(m)}(\xi) = i^m E[X^m e^{i\xi X}].$$

As a result,

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

- **Theorem 40.5 (uniqueness theorem).** *If two real-valued random variables have the same characteristic functions, then they are equal.*

- **Corollary 40.6.** *Let $X = (X_1, X_2, \ldots, X_d)$ be a $\mathbb{R}^d$-valued random variable. Then, the components of X are independent if and only if*

$$\varphi_X(\boldsymbol{\xi}) = \prod_{i=1}^{n} \varphi_{X_i}(\xi_i)$$

for all $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_d) \in \mathbb{R}^d$.

- **Theorem 40.7.** *Two real-valued random variables X and Y are independent if and only if*

$$\varphi_{X+Y}(\xi) = \varphi_X(\xi)\varphi_Y(\xi)$$

for all $\xi \in \mathbb{R}$.

- **Theorem 40.8 (Lévy's inversion formula).** *Let X be a real-valued random variable. Then, for $a < b$,*

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-i\xi a} - e^{-i\xi b}}{i\xi}\phi_X(\xi)\,\mathrm{d}\xi &= \frac{1}{2}P_X(\{a\}) + P_X((a,b)) + P_X(\{b\}) \\ &= \frac{1}{2}\Big(F_X(b) + F_X(b^-) - F_X(a) - F_X(a-)\Big).\end{aligned}$$

- **Theorem 40.9 (Lévy's continuity theorem #1).** *Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of measure on $\mathcal{B}(\mathbb{R}^d)$. Let $\hat{\mu}_n$ denote μ_n's Fourier transform.*

  1. *If μ_n converges weakly to a probability measure μ, then $\hat{\mu}_n(\boldsymbol{\xi}) \to \hat{\mu}(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \mathbb{R}^d$ where $\hat{\mu}$ is the Fourier transform of μ.*
  2. *Suppose $\hat{\mu}_n(\boldsymbol{\xi})$ converges to a function $f(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \mathbb{R}^d$, and suppose that f is continuous at $\mathbf{0}$. Then, there exists a measure μ on $\mathcal{B}(\mathbb{R}^d)$ such that $f(\boldsymbol{\xi}) = \hat{\mu}(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \mathbb{R}^d$. Moreover, $\{\mu_n\}_{n\in\mathbb{N}}$ converges weakly to μ.*

- The following is Lévy's continuity theorem written in the language of "random variables" and "convergence in distribution."

  Theorem 40.10 (Lévy's continuity theorem #2). *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}^d$-valued random variables.*

  1. *If there exists a random variable X such that $X_n \xrightarrow{\mathcal{D}} X$, then $\varphi_{X_n}(\boldsymbol{\xi}) \to \varphi_X(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \mathbb{R}^d$.*
  2. *Suppose φ_{X_n} converges to a function $f(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \mathbb{R}^d$, and suppose that f is continuous at $\mathbf{0}$. Then, there exists a probability measure Q on $\mathcal{B}(\mathbb{R}^d)$ such that $f(\boldsymbol{\xi}) = \hat{Q}(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \mathbb{R}^d$. Morever, $\{P_{X_n}\}_{n\in\mathbb{N}}$ converges weakly to Q.*

- One of the most well-known consequence of Lévy's continuity theorem is the Central Limit Theorem.

  Theorem 40.11 (Central Limit Theorem). *Let $\{X_n : n \in \mathbb{N}\}$ be a sequence of i.i.d. real-valued random variable such that $E[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ for all i with $0 < \sigma^2 < \infty$. Let*

$$S_n = \sum_{i=1}^{n} X_i,$$
$$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

  Then, Y_n converges in distribution to the standard Gaussian random variable. In other words, $Y_n \xrightarrow{\mathcal{D}} Y$ where

$$f_Y(y) = \mathcal{N}(y; 0, 1) = \frac{e^{-y^2/2}}{\sqrt{2\pi}}.$$

  Consequently,

$$\lim_{n\to\infty} P(Y_n \le y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2}\,\mathrm{d}u$$

  for any $y \in \mathbb{R}$.

References

[Bartle, 1995] Bartle, R. G. (1995). *The Elements of Integration and Lebesgue Measure*. John Wiley & Sons, New York.

[Jacod and Protter, 2004] Jacod, J. and Protter, P. (2004). *Probability Essentials*. Springer-Verlag, Berlin Heidelberg.

[Schilling, 2017] Schilling, R. (2017). *Measures, Integrals, and Martingales*. Cambridge University Press, Cambridge, 2nd edition.

[Williams, 1991] Williams, D. (1991). *Probability with Martingales*. Cambridge University Press, Cambridge.