Probability Theory I

Robin Truax

December 2022

Contents

These notes are based on the class Math 230A taught at Stanford by Professor Sourav Chatterjee.

1 Measurability and the Lebesgue Measure

1.1 σ-Algebras and Probability Spaces

Definition 1.1 (σ-Algebra)**.** Let Ω be a set. A *σ-algebra $\mathscr{F}$ on Ω* is a collection of subsets of Ω such that

(1) $\varnothing \in \mathscr{F}$,

(2) $A \in \mathscr{F} \Rightarrow A^{\mathsf{c}} \in \mathscr{F}$.

(3) If $A_1, A_2, A_3, \ldots \in \mathscr{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$.

If the third condition is relaxed to only closure under finite unions (that is, $A_1, \ldots, A_n \in \mathscr{F}$ implies $\bigcup_{i=1}^{n} A_i \in \mathscr{F}$), then $\mathscr{F}$ is called an *algebra*.

Example 1.2. For any Ω, both the power set $\mathcal{P}(\Omega)$ and the set $\{\varnothing, \Omega\}$ are σ-algebras.

Proposition 1.3. *Let $\{\mathscr{F}_\lambda\}_{\lambda\in\Lambda}$ be a collection of σ-algebras on Ω. Then $\bigcap_{\lambda\in\Lambda} \mathscr{F}_\lambda$ is a σ-algebra on Ω.*

Proof. Left as an exercise to the reader. □

Definition 1.4 (Generating σ-Algebras)**.** Let $\mathscr{A}$ be a collection of subsets of Ω. Then, the *σ-algebra generated by $\mathscr{A}$*, denoted $\sigma(\mathscr{A})$, is defined to be the intersection of all σ-algebras containing $\mathscr{A}$. Equivalently, $\sigma(\mathscr{A})$ is the set of all subsets of Ω which can be obtained by a countable number of complements and unions.

Definition 1.5 (Borel σ-Algebra)**.** The *Borel algebra* $\mathscr{B}(\mathbb{R})$ is the σ-algebra generated by all open subsets of $\mathbb{R}$. This is equal to the σ-algebra generated by all open intervals, closed subsets, closed intervals, etc. The Borel algebra $\mathscr{B}(\mathbb{R}^n)$ is defined analogously.

Definition 1.6 (Measurable Space)**.** A *measurable space* $(\Omega, \mathscr{F})$ is a set Ω and a σ-algebra $\mathscr{F}$ on Ω.

Definition 1.7 (Measure)**.** For a measurable space $(\Omega, \mathscr{F})$, a *measure* $\mu : \mathscr{F} \to [0, \infty]$ satisfies

(1) $\mu(\varnothing) = 0$,

(2) if $A_1, A_2, \ldots$ are disjoint, then $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$.

Furthermore, if $\mu(\Omega) = 1$, then μ is called a *probability measure*.

Definition 1.8 (Measure Space)**.** A triple $(\Omega, \mathscr{F}, \mu)$ of a set Ω, a σ-algebra $\mathscr{F}$ on Ω, and a measure μ on $\mathscr{F}$ is called a *measure space*. If μ is a probability measure, then the triple is called a *probability space*.

Definition 1.9 (Events)**.** If $(\Omega, \mathscr{F}, \mu)$ is a probability space, then elements of $\mathscr{F}$ are called *events*.

The following properties of measure spaces are universally used in calculations:

Lemma 1.10. *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and $A \subseteq B$ be measurable. Then $\mu(A) \le \mu(B)$.*

Proof. Let $A_1 = A$, $A_2 = B \setminus A$, and $A_n = \varnothing$ for $n \ge 3$. Then $\mu(B) = \mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i) = \mu(A_1) + \mu(A_2) \ge \mu(A_1) = \mu(A)$. The result follows. □

Lemma 1.11. *Let $(\Omega, \mathscr{F}, \mu)$ be a finite measure space (i.e., $\mu(\Omega) < \infty$). Then, for any $A, B \in \mathscr{F}$, $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.*

Proof. Consider A and $B \setminus (A \cap B)$; they are disjoint, by definition, and have union $A \cup B$. Thus, $\mu(A \cup B) = \mu(A) + \mu(B \setminus A \cap B)$. On the other hand, $A \cap B$ and $B \setminus (A \cap B)$ are disjoint and have union B, so $\mu(B) = \mu(A \cap B) + \mu(B \setminus A \cap B) \Leftrightarrow \mu(B \setminus A \cap B) = \mu(B) - \mu(A \cap B)$. In particular, the rearrangement is valid because all measures are finite. Combining the two equations yields

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

□

More generally, we have the first and second-degree union bounds:

Lemma 1.12. *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. Then, if $A_1, A_2, \dots \in \mathscr{F}$, $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$.*

Proof. First, define the sequence $B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i$. Notice that $B_i \subseteq A_i$ and therefore $\mu(B_i) \leq \mu(A_i)$ for each i. On the other hand, $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Finally, the B_i are pairwise disjoint. Therefore, $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$. □

Lemma 1.13. *Let $(\Omega, \mathscr{F}, \mu)$ be a finite measure space. Then, if $A_1, \dots, A_n \in \mathscr{F}$, $\mu\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} \mu(A_i) - \sum_{1 \leq i < j \leq n} \mu(A_i \cap A_j)$.*

Proof. Define $B_n = A_n \setminus (A_1 \cup \dots \cup A_{n-1})$ for each n. Then, the B_i are disjoint, so $\mu\left(\bigcup_i B_i\right) = \sum_{i=1}^{n} \mu(B_i)$. On the other hand, $B_i \cup (A_i \cap A_1) \cup \dots \cup \dots \cup (A_i \cap A_{i-1}) = A_i$. Therefore, by the previous result,

$$\mu(B_i) + \sum_{j=1}^{i-1} \mu(A_i \cap A_j) \geq \mu(A_i) \Rightarrow \mu(B_i) \geq \mu(A_i) - \sum_{j=1}^{i-1} \mu(A_i \cap A_j).$$

Therefore,

$$\mu\left(\bigcup_i B_i\right) = \sum_{i=1}^{n} \mu(B_i) \geq \sum_{i=1}^{n} \mu(A_i) - \sum_{i=1}^{n} \sum_{j=1}^{i-1} \mu(A_i \cap A_j) = \sum_{i=1}^{n} \mu(A_i) - \sum_{1 \leq i < j \leq n} \mu(A_i \cap A_j).$$

□

Finally, we can measure sets using limits:

Lemma 1.14. *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. Then, if $\{A_n\}_{n \geq 1}$ is a sequence of sets in $\mathscr{F}$ that increases to a set A, $\mu(A_n)$ increases to $\mu(A)$. Similarly, if A_n decreases to a set A and $\mu(A_n) < \infty$ for some n, then $\mu(A_n)$ decreases to $\mu(A)$.*

Proof. Suppose that A_n increases to A. Then the sequence $\mu(A_n)$ is increasing. Then, notice that $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} A_n \setminus A_{n-1}$ (where $A_0 := \varnothing$), and the $A_n \setminus A_{n-1}$ are pairwise disjoint, so

$$\begin{aligned} \mu(A) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} A_n \setminus A_{n-1}\right) = \sum_{n=1}^{\infty} \mu(A_n \setminus A_{n-1}) \\ &= \lim_{N \to \infty} \sum_{n=1}^{N} \mu(A_n \setminus A_{n-1}) = \lim_{N \to \infty} \mu\left(\bigcup_{n=1}^{N} A_n \setminus A_{n-1}\right) = \lim_{N \to \infty} \mu(A_N). \end{aligned}$$

Together, these imply the desired result. The proof for decreasing sequences is similar by taking complements, with minor modifications to handle cases of infinite measure. □

1.2 Dynkin's π-λ Theorem

Definition 1.15 (π-Systems and λ-Systems)**.** Let Ω be a set. A collection $\mathscr{P}$ of subsets of Ω is called a π-system if $A, B \in \mathscr{P} \Rightarrow A \cap B \in \mathscr{P}$. Similarly, a collection $\mathscr{L}$ of subsets of Ω is called a λ-system if

(1) $\Omega \in \mathscr{L}$,

(2) $A \in \mathscr{L} \Rightarrow A^{\mathsf{c}} \in \mathscr{L}$.

(3) $A_1, \dots, A_2, \dots \in \mathscr{L}$ are disjoint implies $\bigcup_{i=1}^{\infty} A_i \in \mathscr{L}$.

Lemma 1.16. *A σ-algebra is a λ-system and a π-system.*

Lemma 1.17. *A λ-system that is also a π-system is a σ-algebra.*

Lemma 1.18. *The intersection of any family of λ-systems is again a λ-system.*

Definition 1.19 (Generating λ-System)**.** Given any system $\mathscr{A}$, the system $\lambda(\mathscr{A})$ is the intersection of all λ-systems containing $\mathscr{A}$; by Lemma 1.18, this is a λ-system, indeed the smallest λ-system containing $\mathscr{A}$.

Lemma 1.20. *If $\mathscr{P}$ is a π-system, then $\lambda(\mathscr{P})$ is a π-system.*

Proof. Take any $A \in \mathscr{P}$, and let $S_1 = \{B \in \lambda(\mathscr{P}) \mid B \cap A \in \lambda(\mathscr{P})$. It is not hard to show that S_1 is a λ-system, and it plainly contains $\mathscr{P}$ (yet is contained in $\lambda(\mathscr{P})$), so $S_1 = \lambda(\mathscr{P})$. Then, let $S_2 = \{A \in \lambda(\mathscr{P}) \mid B \in \lambda(\mathscr{P}) \Rightarrow A \cap B \in \lambda(\mathscr{P})\}$. It is not hard to show that S_2 is a λ-system, and by the previous result, $\mathscr{P} \subseteq S_2$, so $\lambda(P) = S_2$. Thus, $\lambda(\mathscr{P})$ is closed under intersection, as desired. □

Theorem 1.21 (Dynkin π-λ Theorem 1)**.** *If $\mathscr{P}$ is a π-system, then $\lambda(\mathscr{P}) = \sigma(\mathscr{P})$.*

Proof. By Lemma 1.20, $\lambda(\mathscr{P})$ is a π-system and a λ-system, so by Lemma 1.17, $\lambda(\mathscr{P})$ is a σ-algebra, and therefore contains $\lambda(\mathscr{P}) \supseteq \sigma(\mathscr{P})$. On the other hand, $\sigma(\mathscr{P})$ is a λ-system, so $\sigma(\mathscr{P}) \supseteq \lambda(\mathscr{P})$. □

Theorem 1.22 (Dynkin's π-λ Theorem 2)**.** *Let $\mathscr{P}$ be a π-system and $\mathscr{L}$ be a λ-system containing $\mathscr{P}$. Then,*

$$\mathscr{L} \supseteq \sigma(\mathscr{P}).$$

Proof. $\sigma(\mathscr{P}) = \lambda(\mathscr{P}) \subseteq \mathscr{L}$. □

Let us demonstrate a use of Dynkin's π-λ Theorem. First we need an intuitive technical lemma.

Lemma 1.23. *If μ is a measure on a σ-algebra $\mathscr{F}$ and $A_1, A_2, \dots \in \mathscr{F}$ are an increasing sequence with union A, then $\mu(A) = \lim \mu(A_i)$. Moreover, if $A_1, A_2, \dots$ are a decreasing sequence with intersection A, and $\mu(A_i) < \infty$ for some i, then $\mu(A) = \lim \mu(A) = \lim \mu(A_i)$.*

The proof is from the axioms of measures. For an example showing why the condition $\mu(A_i) < \infty$ for some i is necessary, consider the following:

Example 1.24. Let $A_i = (i, \infty)$ and μ be a length measure on $\mathbb{R}$. Then $\lim \mu(A_i) = \infty$ as $\mu(A_i) = \infty$ for each i, yet $A = \bigcap_i A_i = \varnothing$ whence $\mu(A) = 0$.

Theorem 1.25. *Let $\mathscr{P}$ be a π-system and μ_1, μ_2 be measures on $\sigma(\mathscr{P})$ that agree on $\mathscr{P}$. Suppose that there is an increasing sequence $A_1 \subseteq A_2 \subseteq A_3 \cdots$ of elements of $\mathscr{P}$ whose union is Ω such that $\mu_1(A_i) < \infty$. Then, $\mu_1 = \mu_2$ on $\sigma(\mathscr{P})$.*

Proof. Take any $A \in \mathscr{P}$ such that $\mu_1(A) < \infty$. Let $\mathscr{L} = \{B \in \sigma(\mathscr{P}) \mid \mu_1(A \cap B) = \mu_2(A \cap B)\}$. Plainly, $\mathscr{L}$ contains $\mathscr{P}$. Furthermore, notice that $\mathscr{L}$ is a λ-system. For $\Omega \in \mathscr{L}$, and if $B \in \mathscr{L}$, $\mu_1 = (A \cap B^{\mathsf{c}}) = \mu_1(A) - \mu_1(A \cap B) = \mu_2(A) - \mu_2(A \cap B) = \mu_2(A \cap B^{\mathsf{c}})$, whence $B^{\mathsf{c}} \in \mathscr{L}$ Finally, suppose $B_1, B_2, \dots \in \mathscr{L}$ are disjoint. Then $\mu_1\left(A \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right) = \mu_1\left(\bigcup_{i=1}^{\infty} A \cap B_i\right) = \sum_{i=1}^{\infty} \mu_1(A \cap B_i) = \sum_{i=1}^{\infty} \mu_2(A \cap B_i) = \mu_2\left(A \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right)$ whence $\bigcup_{i=1}^{\infty} B_i \in \mathscr{L}$.

Thus, by Dynkin's π-λ system, $\mathscr{L} \supseteq \sigma(\mathscr{P})$. Then, choose an increasing sequence $A_1 \subseteq A_2 \subseteq A_3 \cdots$ of elements of $\mathscr{P}$ whose union is Ω such that $\mu_1(A_i) < \infty$. Then, for each $B \in \sigma(\mathscr{P})$, $\bigcup_{i=1}^{\infty} A_i \cap B = B$. Thus, $\mu_1(A_i \cap B)$ converges to $\mu_1(B)$, and $\mu_2(A_i \cap B)$ converges to $\mu_2(B)$; yet $\mu_1(A_i \cap B) = \mu_2(A_i \cap B)$ for each i. Thus, $\mu_1(B) = \mu_2(B)$. The result follows. □

Example 1.26. Let $\Omega = \mathbb{R}$, and $\mathscr{P}$ be the collection of bounded open intervals. Then, $\sigma(\mathscr{P}) = \mathcal{B}(\mathbb{R})$. Suppose that μ_1 and μ_2 are measures on $\mathcal{B}(\mathbb{R})$ such that for any $a < b$, then $\mu_1((a, b)) = \mu_2((a, b)) = b - a$. Then $\mu_1 = \mu_2$ by considering the increasing sequence $A_n = (-n, n)$ and applying the previous theorem. This demonstrates that there is a unique natural "length" measure on $\mathbb{R}$.

Definition 1.27 (Monotone Class)**.** Let Ω be a set. A collection $\mathscr{C}$ of subsets of Ω is called a *monotone class* if it is closed under monotone limits, that is, if $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{C}$ then $\bigcup_i A_i \in \mathscr{C}$ and if $A_1 \supseteq A_2 \supseteq \dots \in \mathscr{C}$ then $\bigcap_i A_i \in \mathscr{C}$.

Theorem 1.28 (Monotone Class Theorem)**.** *If $\mathscr{A}$ is an algebra and $\mathscr{C}$ is a monotone class containing $\mathscr{A}$, then $\mathscr{C} \supseteq \sigma(\mathscr{A})$.*

Proof. First, notice that the intersection of any family of monotone classes is another monotone class. Therefore, given any algebra $\mathscr{A}$, there is a smallest monotone class $\mathscr{M}$ containing $\mathscr{A}$. I claim that $\mathscr{M}$ is a λ-system. Obviously, $\mathscr{M}$ is closed under increasing unions by definition and nonempty since it contains $\mathscr{A}$: it suffices to show that it is closed under complements. To see why, define fix some $S \in \mathscr{A}$. Then, define

$$\mathscr{M}_S = \{T \in \mathscr{M} \mid S \setminus T \text{ and } T \setminus S \in \mathscr{M}\}.$$

It is easy to see that $\mathscr{M}_S$ is a monotone class. Furthermore, $\mathscr{A} \subseteq \mathscr{M}_S$, so indeed $\mathscr{M} \subseteq \mathscr{M}_S$ and $\mathscr{M} = \mathscr{M}_S$. In other words, for any $S \in \mathscr{A}$ and $T \in \mathscr{M}$, $S \setminus T$ and $T \setminus S \in \mathscr{M}$. Now, suppose that $T \in \mathscr{M}$. Then, by the previous remark, $\mathscr{M}_T$ contains $\mathscr{A}$. Yet $\mathscr{M}_T$ is still a monotone class, so $\mathscr{M}_T$ contains $\mathscr{M}$ and $\mathscr{M} = \mathscr{M}_T$ for any $T \in \mathscr{M}$. In other words, for any $S, T \in \mathscr{M}$, $S \setminus T$ and $T \setminus S$ both belong to $\mathscr{M}$, so $\mathscr{M}$ is closed under complements, as desired.

Then, since $\mathscr{M}$ is a λ-system, and $\mathscr{A}$ is an algebra (and therefore a π-system), the Dynkin π-λ theorem yields that $\mathscr{M} \supseteq \sigma(\mathscr{A})$. Yet $\mathscr{C}$ contains $\mathscr{M}$, so $\mathscr{C} \supseteq \mathscr{M} \supseteq \sigma(\mathscr{A})$, as desired. □

1.3 Outer Measures

Definition 1.29 (Outer Measure)**.** Let Ω be any set. A function $\phi : 2^\Omega \to [0, \infty]$ is called an outer measure if $\phi(\varnothing) = 0$, $\phi(A) \leq \phi(B)$ when $A \subseteq B$, and for any $A_1, A_2, \dots \subseteq \Omega$, $\phi\left(\bigcup_i A_i\right) \leq \sum_i \phi(A_i)$.

Definition 1.30 (ϕ-measurable)**.** Let ϕ be an outer measure on a set Ω. A subset $A \subseteq \Omega$ is called ϕ-measurable if $\forall B \subseteq \Omega$, $\phi(B) = \phi(B \cap A) + \phi(B \cap A^\mathsf{c})$.

Theorem 1.31. *Let $\mathscr{F}$ be the collection of all ϕ-measurable subsets of Ω. Then $\mathscr{F}$ is a σ-algebra and ϕ is a measure on $\mathscr{F}$.*

Proof. The proof is a series of straightforward lemmas.

Lemma 1.32. *The collection $\mathscr{F}$ is an algebra.*

Lemma 1.33. *If $A_1, \dots, A_n \in \mathscr{F}$ are disjoint and $E \subseteq \Omega$, then*

$$\phi(E \cap (A_1 \cup \dots \cup A_n)) = \sum_{i=1}^n \phi(E \cap A_i).$$

By the previous two lemmas, we can demonstrate

Lemma 1.34. *If $A_1, A_2, \dots$ is a sequence of sets in $\mathscr{F}$ increasing to a set $A \subseteq \Omega$, then for any $E \subseteq \Omega$,*

$$\phi(E \cap A) \leq \lim_{n \to \infty} \phi(E \cap A_n).$$

From here, the conclusion follows. Indeed, let $A_1, A_2, \dots \in \mathscr{F}$ and let $A = \bigcup_i A_i$. For each n, let $B_n = \bigcup_{i=1}^n A_i$; this belongs to $\mathscr{F}$ by the first lemma. Then, for any $E \subseteq \Omega$ and any n,

$$\phi(E) = \phi(E \cap B_n) + \phi(E \cap B_n^\mathsf{c}) \geq \phi(E \cap B_n) + \phi(E \cap A^\mathsf{c})$$

Yet the third lemma demonstrates that $\lim_{n\to\infty} \phi(E \cap B_n) \geq \phi(E \cap A)$. Thus, $\phi(E) \geq \phi(E \cap A) + \phi(E \cap A^\mathsf{c})$; the other side of the inequality is immediate from subadditivity. Thus, $\phi(E) = \phi(E \cap A) + \phi(E \cap A^\mathsf{c})$, so $A \in \mathscr{F}$, as desired. This, with the first lemma, shows that $\mathscr{F}$ is a σ-algebra.
It then suffices to show that ϕ is a measure on $\mathscr{F}$. For this, take any disjoint collection $A_1, A_2, \dots \in \mathscr{F}$, and define $B_n = \bigcup_{i=1}^n A_i$ as before. Then, by Lemma 1.4.5,

$$\phi(B) \geq \phi(B_n) = \sum_{i=1}^n \phi(A_i).$$

Thus, by taking $n \to \infty$, $\phi(B) \geq \sum_i \phi(A_i)$. The opposite inequality is given by subadditivity. Thus, ϕ is a measure on $\mathscr{F}$, as desired. □

1.4 Carathéodory's Extension Theorem

Theorem 1.35 (Carathéodory's Extension Theorem)**.** *Let $\mathscr{A}$ be an algebra of subsets of a set Ω. Let μ be a measure on $\mathscr{A}$. Then, μ has an extension to $\sigma(\mathscr{A})$. Moreover, the extension is unique if μ is σ-finite on $\mathscr{A}$, meaning that $\exists A_1, A_2, \dots \in \mathscr{A}$ such that $\mu(A_i) < \infty$ for all i and $A_i \uparrow \Omega$.*

Proof. Uniqueness follows immediately from Theorem 1.25, as algebras are π-systems. For existence, define $\mu^* : 2^\Omega \to [0, \infty]$ as follows: for any $A \subseteq \Omega$, let

$$\mu^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) \,\middle|\, A_i \in \mathscr{A}, \bigcup_{i=1}^{\infty} A_i \supseteq A \right\}$$

The proof then requires two straightforward lemmas:

Lemma 1.36. *μ^* is an outer measure.*

Lemma 1.37. *For any $A \in \mathscr{A}$, $\mu^*(A) = \mu(A)$.*

Then, to conclude, let $\mathscr{A}^*$ be the set of all μ^*-measureable sets. Then, by Theorem 1.31, $\mathscr{A}^*$ is a σ-algebra and μ^* is a measure on $\mathscr{A}^*$. Thus, it suffices to show that $\mathscr{A} \subseteq \mathscr{A}^*$; that is, that any $A \in \mathscr{A}$ is μ^*-measurable.

For this, take any $A \in \mathscr{A}$ and $E \subseteq \Omega$. Then, for any sequence $A_1, A_2, \dots$ of elements of $\mathscr{A}$ that cover E, $\{A \cap A_i\}_{i=1}^{\infty}$ is a cover for $E \cap A$ and $\{A^{\mathsf{c}} \cap A_i\}_{i=1}^{\infty}$ is a cover for $E \cap A^{\mathsf{c}}$. Thus,

$$\mu^*(E \cap A) + \mu^*(E \cap A^{\mathsf{c}}) \le \sum_{i=1}^{\infty} (\mu(A \cap A_i) + \mu(A^{\mathsf{c}} \cap A_i)) = \sum_{i=1}^{\infty} \mu(A_i).$$

Taking the infimum over all choices of $\{A_i\}_{i=1}^{\infty}$, we obtain that $\mu^*(E \cap A) + \mu^*(E \cap A^{\mathsf{c}} \le \mu^*(E)$, as desired. ☐

1.5 Construction of the Lebesgue Measure

Let $\mathscr{C}$ be the collection of all sets of the form either $(a, b]$ for some $a, b \in \mathbb{R}$ or (a, ∞) for $a \in \mathbb{R}$ or $(-\infty, b]$ for $b \in \mathbb{R}$ or $\mathbb{R}$. Let $\mathscr{A}$ be the collection of all finite disjoint unions of elements of $\mathscr{C}$. Then, $\mathscr{A}$ is an algebra which generates the Borel σ-algebra of $\mathbb{R}$.

Define a functional $\lambda : \mathscr{A} \to \mathbb{R}$ by

$$\lambda \left(\bigcup_{i=1}^{n} (a_i, b_i] \cap \mathbb{R} \right) := \sum_{i=1}^{n} (b_i - a_i).$$

In other words, λ measures the length of an element of $\mathscr{A}$. Clearly, λ is finitely additive on $\mathscr{A}$ and monotone.

Lemma 1.38. *For any $A_1, \dots, A_n \in \mathscr{A}$ and any $A \subseteq A_1 \cup \dots \cup A_n$, $\lambda(A) \le \sum_{i=1}^{n} \lambda(A_i)$.*

Proof. Let $B_1 = A_1$ and $B_i = A_i \setminus (A_1 \cup \dots \cup A_{i-1})$ for $2 \le i \le n$. Then $B_1, \dots, B_n$ are disjoint and have union $A_1, \dots, A_n$. Then, $\lambda(A) = \sum_{i=1}^{n} \lambda(A \cap B_i) \le \sum_{i=1}^{n} \lambda(B_i) \le \sum_{i=1}^{n} \lambda(A_i)$. ☐

Then, we can prove the following facts about λ.

Proposition 1.39. *The functional λ defined above is a σ-finite measure on $\mathscr{A}$.*

Proof. Suppose that $A_1, A_2, \dots \in \mathscr{A}$ is a sequence of disjoint elements in $\mathscr{A}$ with union $A \in \mathscr{A}$. Since each element of $\mathscr{A}$ is a finite disjoint union of such intervals, it suffices to show the case when $A = (a, b] \cap \mathbb{R}$ and $A_i = (a_i, b_i] \cap \mathbb{R}$ for each i. Assume that $a < b$. Now, suppose that a and b are both finite.

Take any $\delta > 0$ such that $a + \delta < b$, and take any $\varepsilon > 0$. Then $[a + \delta, b] \subseteq \bigcup_{i=1}^{\infty} (a_i, b_i + 2^{-i}\varepsilon)$; since $[a + \delta, b]$ is compact, there exists some k such that $[a + \delta, b] \subseteq \bigcup_{i=1}^{k} (a_i, b_i + 2^{-i}\varepsilon)$. Therefore, by the above lemma, we have $b - a - \delta \le \sum_{i=1}^{k} (b_i + 2^{-i}\varepsilon - a_i) \le \varepsilon + \sum_{i=1}^{\infty} (b_i - a_i)$. By driving ε and δ to 0, we obtain $b - a \le \sum_{i=1}^{\infty} (b_i - a_i)$. Finite additivity and monotonicity yields the other direction. Finally, if one or both of a and b are infinite, it suffices to take finite a', b' such that $(a', b'] \subseteq (a, b]$, and take the limit. ☐

Corollary 1.39.1. *The functional λ has a unique extension to a measure on $\mathcal{B}(\mathbb{R})$.*

Definition 1.40 (Lebesgue Measure)**.** The unique extension of λ given by the above corollary is called the Lebesgue measure on the real line.

One can define the Lebesgue measure on $\mathbb{R}^n$ for general n by considering disjoint unions of products of half-open intervals and then repeating the above development. We shall do that now.

Let $\mathscr{A}$ be the set of all subsets of $\mathbb{R}^d$ that are finite disjoint unions of half-open cubes of the form $(a_1, b_1] \times \cdots \times (a_d, b_d] \cap \mathbb{R}^d$, where $-\infty \leq a \leq b \leq \infty$. Then, $\mathscr{A}$ is an algebra of subsets of $\mathbb{R}$ which generates a σ-algebra on $\mathbb{R}^d$ which we call the Borel σ-algebra on $\mathbb{R}^d$. Define $\lambda : \mathscr{A} \to \mathbb{R}$ by

$$\lambda\left(\bigcup_{i=1}^{n}(a_{i1}, b_{i1}] \times \cdots (a_{id}, b_{id}] \cap \mathbb{R}^d\right) = \sum_{i=1}^{n}(b_{i1} - a_{i1})(b_{i2} - a_{i2}) \cdots (b_{id} - a_{id}).$$

Obviously, λ satisfies finite additivity and therefore monotonicity.

Lemma 1.41. *For any $A_1, \ldots, A_n \in \mathscr{A}$ and any $A \subseteq A_1 \cup \cdots \cup A_n$, $\lambda(A) \leq \sum \lambda(A_i)$.*

Proof. Let $B_i = A_i \setminus (A_1 \cup \cdots A_{i-1})$. Then $B_1, \ldots, B_n$ are disjoint with union $\bigcup_{i=1}^{n} A_i$. Then, as desired,

$$\lambda(A) = \sum_{i=1}^{n} \lambda(A \cap B_i) \leq \sum_{i=1}^{n} \lambda(B_i) \leq \sum_{i=1}^{n} \lambda(A_i).$$

□

Lemma 1.42. *The functional λ defined above is a σ-finite measure on $\mathscr{A}$.*

Proof. σ-finitude is trivial, so it suffices to show countable additivity. Indeed, suppose that $A \in \mathscr{A}$ is a countable disjoint union of elements $A_1, A_2, \cdots \in \mathscr{A}$. Then we seek to show that $\lambda(A) = \sum_{i=1}^{\infty} \lambda(A_i)$. Of course, it suffices to show that this is true when $A = (a_1, b_1] \times \cdots \times (a_d, b_d] \cap \mathbb{R}^d \cap \mathbb{R}$ and $A_i = (a_{i1}, b_{i1}] \times \cdots (a_{id}, b_{id}] \cap \mathbb{R}^d$ for each i, since any element of $\mathscr{A}$ is a finite disjoint union of such cubes.

Now, first suppose that $-\infty < a_j < b_j < \infty$ for each j. Then, take any $\delta > 0$ such that $a_j + \delta < b_j$ for each j, and any $\varepsilon > 0$. Then $[a_1 + \delta, b_1] \times \cdots \times [a_d + \delta, b_d] \subseteq \bigcup_{i \geq 1}(a_{i1}, b_{i1} + 2^{-i}\varepsilon) \times \cdots \times (a_{id}, b_{id} + 2^{-i}\varepsilon)$. Now, since $[a_1 + \delta, b_1] \times \cdots \times [a_d + \delta, b_d]$ is compact, it is contained in the union of finitely many $(a_{i1}, b_{i1} + 2^{-i}\varepsilon) \times \cdots \times (a_{id}, b_{id} + 2^{-i}\varepsilon)$. Thus, there exists some k such that

$$[a_1 + \delta, b_1] \times \cdots \times [a_d + \delta, b_d] \subseteq \bigcup_{i=1}^{k}(a_{i1}, b_{i1} + 2^{-i}\varepsilon) \times \cdots \times (a_{id}, b_{id} + 2^{-i}\varepsilon).$$

Thus, by the preceding lemma,

$$(b_1 - a_1 - \delta) \cdots (b_d - a_d - \delta) \leq \sum_{i=1}^{k}(b_{i1} + 2^{-i}\varepsilon - a_{i1}) \cdots (b_{id} + 2^{-i}\varepsilon - a_{id}) \leq \varepsilon + \sum_{i=1}^{\infty}(b_{i1} - a_{i1}) \cdots (b_{id} - a_{id}).$$

By driving δ and ε to 0, we obtain $(b_1 - a_1) \cdots (b_d - a_d) \leq \sum_{i=1}^{\infty}(b_{i1} - a_{i1}) \cdots (b_{id} - a_{id})$. On the other hand, for any k, finite additivity and monotonicity of λ implies that $(b_1 - a_1) \cdots (b_d - a_d) = \lambda(A) \geq \sum_{i=1}^{k} = \sum_{i=1}^{k}(b_{i1} - a_{i1}) \cdots (b_{id} - a_{id})$ whence $\lambda(A) = (b_1 - a_1) \cdots (b_d - a_d) \geq \sum_{i=1}^{\infty}(b_{i1} - a_{i1}) \cdots (b_{id} - a_{id})$. Thus we have proven countable additivity when the a_j and b_j are finite. On the other hand, if either a_j or b_j is infinite, choose finite a'_j, b'_j such that $(a'_j, b'_j] \subseteq (a_j, b_j] \cap \mathbb{R}$ for each j. Repeating the above steps, we achieve

$$(b'_1 - a'_1) \cdots (b'_d - a'_d) = \sum_{i=1}^{\infty}(b_{i1} - a_{i1}) \cdots (b_{id} - a_{id})$$

for any finite $a' > a$ and $b' < b$. Since this holds for all such a'_j, b'_j, the equality $(b_1 - a_1) \cdots (b_d - a_d) = \sum_{i=1}^{\infty}(b_{i1} - a_{i1}) \cdots (b_{id} - a_{id})$ still holds. □

Corollary 1.42.1. *The function λ has a unique extension to a measure on the Borel σ-algebra on $\mathbb{R}^d$.*

Definition 1.43 (Lebesgue Measure)**.** The unique extension of λ given by the above corollary is called the Lebesgue measure on $\mathbb{R}^d$.

For an example computation of Lebesgue measure in higher dimensions, we consider the example of a line.

Example 1.44. A straight line in $\mathbb{R}^2$ has measure zero.

Proof. First, notice that by swapping the x and y-coordinates, we may assume that the line L is not vertical and therefore can be written in the form $y = f(x) = mx + b$ for some $m, b \in \mathbb{R}$. Then, for any k, let $\mathscr{F}_k$ be the following family of boxes:

$$\mathscr{F}_k = \left\{ \left[n + \frac{l}{2^{k+n}}, n + \frac{l+1}{2^{k+n}}\right] \times \left[f\left(n + \frac{l}{2^{k+n}}\right), f\left(n + \frac{l+1}{2^{k+n}}\right)\right] \;\middle|\; n \in \mathbb{Z}, 0 \leq l \leq 2^{k+n} - 1 \right\}$$

For any k, $\mathscr{F}_k$ covers the entirety of L. On the other hand, the section of $\mathscr{F}_k$ to do with a fixed $n \in \mathbb{Z}$ (that is, the section covering $[n, n+1]$) has the area $2^{k+n}\left(\frac{1}{2^{k+n}} \cdot \frac{m}{2^{k+n}}\right) = \frac{m}{2^{k+n}}$. Therefore, the area of $\mathscr{F}_k$ is $\frac{m}{2^k} \sum_{n \in \mathbb{Z}} 2^n = \frac{3m}{2^k}$. Yet then $\mu(L) \leq \mu(\mathscr{F}_k) = \frac{3m}{2^k}$ for each k, whence by taking $k \to \infty$ we obtain $\mu(L) = 0$. ☐

1.6 Completion of Measure Spaces

Definition 1.45 (Complete σ-Algebra)**.** Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. μ is said to be *complete* if whenever $A \in \mathscr{F}$, $\mu(A) = 0$, and $B \subseteq A$, then $B \subseteq \mathscr{F}$.

Proposition 1.46. *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. Then there exists a σ-algebra $\mathscr{F}' \supseteq \mathscr{F}$, and an extension of μ to $\mathscr{F}'$, such that $\mathscr{F}'$ is a complete σ-algebra.*

Proof. Define an outer measure μ^* and a σ-algebra $\mathscr{A}^*$ as in the proof of Carathéodory's extension theorem. Then, $\mathscr{A}^*$ is complete with respect to μ^*. ☐

In fact, the completion of the Borel σ-algebra of $\mathbb{R}$ is the Lebesgue σ-algebra. The Lebesgue measure is defined on this larger σ-algebra, but we work with the Borel σ-algebra most of the time. For example, when we say that a function defined on $\mathbb{R}$ is measurable, we mean Borel measurable. On the other hand, abstract probability spaces will usually be assumed to be complete.

1.7 Lebesgue vs. Borel Sets

Proposition 1.47. *A set is Lebesgue measurable if and only if it is the union of a Borel set and a null set.*

Proof. First, we begin with some straightforward lemmas.

Lemma 1.48. *Suppose that A is Lebesgue measurable. Then $m(A) = \inf\{m(U) \mid U \supseteq A \text{ open}\}$.*

Corollary 1.48.1. *Suppose that A is Lebesgue measurable. Then $m(A) = \sup\{m(V) \mid V \subseteq V \text{ closed}\}$.*

If A is the union of a Borel set and a null set, then A is clearly Lebesgue measurable. Therefore, it suffices to show that if A is Lebesgue measurable, then it is the union of a Borel set and a null set.

For this, consider first the case where $m(A) < \infty$. Then, by the above corollary, there exists a sequence of sets $V_1 \subseteq V_2 \subseteq \cdots$ such that $\lim_{i \to \infty} m(V_i) = m(A)$. Then, if V is the Borel set $\bigcup_{i=1}^{\infty} V_i$, $m(V) = m(A)$. Then $V \setminus A$ is measurable as the intersection of the Borel set V and the measurable set A^{c}, and furthermore, since $m(A) = m(V) < \infty$, $m(V \setminus A) = 0$. Thus, A is a union of the Borel set V and the null set A.

Now suppose that A is an arbitrary Lebesgue measurable set. Let A_n be the intersection of A with the open ball $B_n(0)$ of radius n around the origin. By our above work, $A_n = V_n \cap W_n$ for some Borel set V_n and null set W_n. But then $V = \bigcup_{i=1}^{\infty} V_i$ is Borel, and $W = \bigcup_{i=1}^{\infty} W_i$ is null, and $A = V \cap W$. ☐

2 Measurable Functions

Definition 2.1 (Measurable Function)**.** Let $(\Omega, \mathscr{F})$ and $(\Omega' \mathscr{F}')$ be two measurable spaces. A function $f : \Omega \to \Omega'$ is called measurable if $f^{-1}(A) \in \mathscr{F}$ for every $A \in \mathscr{F}'$. It is easy to see that the composition of measurable functions is measurable.

One way to simplify the process of computing measurability is the following:

Lemma 2.2. *Let $(\Omega, \mathscr{F})$ and $(\Omega', \mathscr{F}')$ be two measurable spaces and $f : \Omega \to \Omega'$ be a function. Suppose that there is a set $\mathscr{A} \subseteq \mathscr{F}'$ that generates $\mathscr{F}'$ and suppose that $f^{-1}(A) \in \mathscr{F}$ for all $A \in \mathscr{A}$. Then f is measurable.*

Proof. The set of all $B \subseteq \Omega'$ such that $f^{-1}(B) \in \mathscr{F}$ is a σ-algebra, and it contains $\mathscr{A}$, so it contains $\sigma(\mathscr{A}) = \mathscr{F}')$, as desired. ☐

Definition 2.3 (Borel σ-Algebra of a Topological Space)**.** The *Borel σ-algebra on Ω* is the σ-algebra generated by the open sets.

Proposition 2.4. *Suppose that Ω and Ω' are topological spaces, and $\mathscr{F}$ and $\mathscr{F}'$ are their Borel σ-algebras. Then any continuous function from Ω into Ω' is measurable.*

Proof. Apply the preceding lemma with $\mathscr{A}$ being the set of all open subsets of Ω'. ☐

Other measurable functions include:

1. Sums and products of measurable functions.
2. Right-continuous or left-continuous functions.
3. Monotone functions.
4. Lower- or upper-semicontinuous functions.
5. The infimum or supremum of a series of measurable functions.
6. The limit infimum or supremum of a series of measurable functions.
7. The pointwise limit of a series of measurable functions.
8. The sum of an infinite sequence of $[0, \infty]$-valued measurable functions.

The proof of these facts is left as an exercise to the reader.

2.1 Lebesgue Integration

We define Lebesgue integration in three steps. First, given a simple function $f = \sum_{i=1}^{n} a_i 1_{A_i}$ with $A_1, \ldots, A_n \in \mathscr{F}$ disjoint and $a_1, \ldots, a_n \geq 0$, we define $\int f d\mu = \sum_{i=1}^{n} a_i \mu(A_i)$. Next, consider any measurable function $f : \Omega \to [0, \infty)$. Let $\mathrm{SF}^+(f) = \{g \mid g \text{ non-negative simple functions} \forall \omega \in \Omega\}$. Then $\int f d\mu = \sup_{g \in \mathrm{SF}^+(f)} \int g d\mu$.

Finally, consider any measurable function $f : \Omega \to \mathbb{R}$. Let $f^+(\omega)$ be equal to $\max(f, 0)$ and $f^-(\omega) = -\min(f, 0)$. Then $f = f^+ - f^-$; if at least one of $f^+ d\mu$ and $\int f^- d\mu$ is finite, we define $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$ and say that the *integral exists.* If indeed both quantities are finite, then we say that f is *integrable.*

Lemma 2.5. *If $0 \leq f \leq g$ everywhere, then $\int f d\mu \leq \int g \mu$.*

2.2 Properties of the Lebesgue Integral

Lemma 2.6. *Let $s : \Omega \to [0, \infty)$ be a measurable simple function. For each $S \in \mathscr{F}$, let $\nu(S) = \int_S s d\mu$. Then ν is a measure on $(\Omega, \mathscr{F})$.*

Proof. Suppose that $s = \sum_{i=1}^n a_i 1_{A_i}$. Since $\nu(\varnothing) = 0$ by definition, it suffices to show that ν is countably additive. Then suppose that $S_1, S_2, \dots$ is a sequence of disjoint sets in $\mathscr{F}$ with union S. Then

$$\nu(S) = \sum_{i=1}^n a_i \mu(A_i \cap S) = \sum_{i=1}^n \left(\sum_{j=1}^\infty \mu(A_i \cap S_j) \right) = \sum_{j=1}^\infty \sum_{i=1}^n a_i \mu(A_i \cap S_j) = \sum_{j=1}^\infty \nu(S_j).$$

□

Theorem 2.7 (Monotone Convergence Theorem)**.** *Suppose that $\{f_n\}_{n\geq 1}$ is a sequence of non-negative measurable functions on Ω increasing pointwise to a limit function f. Then $\int f\mu = \lim_{n\to\infty} \int_n d\mu$.*

Proof. Since $f \geq f_n$ for every n, we have $\int f d\mu \geq \lim \int f_n d\mu$. On the other hand, consider $s \in \mathrm{SF}^+(f)$. Let ν be as in the previous lemma and fix $\alpha \in (0,1)$. Let $S_n = \{\omega \mid \alpha s(\omega) \leq f_n(\omega)\}$. These sets are measurable, increasing with n, and increase to all of Ω. Then $\int s d\mu = \nu(\Omega) = \lim_{n\to\infty} \nu(S_n) = \lim_{n\to\infty} \int_{S_n} s d\mu$. Yet $\alpha s \leq f_n$ on S_n, and since s is simple, $\int_{S_n} \alpha s d\mu = \alpha \int_{S_n} s d\mu$. Therefore,

$$\alpha \int_{S_n} s d\mu = \int_{S_n} \alpha s d\mu \leq \int_{S_n} f_n d\mu \leq \int_\Omega f_n d\mu.$$

Thus, $\alpha \int s d\mu \leq \lim \int f_n d\mu$, whence $\int s d\mu \leq \lim \int f_n d\mu$, whence $\int f d\mu \leq \lim \int f_n d\mu$. □

Proposition 2.8. *Given any measurable function $f : \Omega \to [0, \infty]$, there is a sequence of nonnegative simple functions increasing pointwise to f.*

Proof. Let $f_n(\omega) = \min\{n, \lfloor f_n 2^n \rfloor 2^{-n}\}$. The result follows. □

Proposition 2.9 (Linearity of the Integral)**.** *If f and g are two integrable functions from Ω into $\mathbb{R}^*$, then for any $\alpha, \beta \in \mathbb{R}$, the function $\alpha f + \beta g$ is integrable and $\int(\alpha f + \beta g)d\mu = \alpha \int f\mu + \beta \int g\mu$. Moreover, if f and g are measurable functions from Ω into $[0,\infty]$, then $\int(f+g)d\mu = \int f\mu + \int g d\mu$, and for any $\alpha \in \mathbb{R}$, $\int \alpha f\mu = \alpha \int f\mu$.*

Proof. First demonstrate the result for simple functions, then for non-negative measurable functions using the monotone convergence theorem, and then for all measurable functions using the traditional decomposition $f = f^+ - f^-$. Each step in this decomposition is relatively straightforward. □

Lemma 2.10 (Fatou's Lemma)**.** *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. Let $\{f_n\}_{n\geq 1}$ be a sequence of measurable functions from Ω into $[0,\infty]$. Then, $\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu$.*

Proof. Let $g_n = \inf_{f\geq n} f_k$. Then g_n is an increasing sequence of nonnegative functions converging to $f = \liminf_{n\to\infty} f_n = \lim_{n\to\infty} g_n$ By the Monotone Convergence Theorem, $\int f d\mu = \lim_{n\to\infty} \int g_n d\mu$. But $g_n \leq f_k$ everywhere for all $k \geq n$. Thus $\int g_n d\mu \leq \int f_k d\mu$ for all $k \geq n$. But this implies $\int g_n d\mu \leq \inf_{k\geq n} \int f_k d\mu$. But then $\lim_{n\to\infty} \int g_n d\mu \leq \lim_{n\to\infty} \int f_k d\mu = \liminf_{n\to\infty} \int f_n d\mu$. □

Theorem 2.11 (Dominated Convergence Theorem)**.** *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. Let $\{f_n\}_{n\geq 1}$ be a sequence of measurable functions from Ω into $\mathbb{R}$, converging pointwise to $f : \Omega \to \mathbb{R}$. Suppose that there exists a measurable function $h : \Omega \to [0,\infty)$ such that h is integrable and $|f_n(\omega)| \leq h(\omega)$ for all n, ω. Then $\int f\mu = \lim_{n\to\infty} \int f_n d\mu$.*

Proof. Let $g_n = f_n + h$. Since $|f_n| \leq h$ everywhere, $g_n \geq 0$ everywhere. Then, by Fatou's Lemma,

$$\int \liminf g_n d\mu \leq \liminf \int g_n d\mu = \int f d\mu + \int h d\mu \leq \liminf \left(\int f_n d\mu + \int h d\mu \right) = \liminf \int f_n d\mu + \int h d\mu.$$

Thus $\int f d\mu \leq \liminf \int f_n d\mu$. Next, let $g_n = h - f_n$; repeating the process, we find that $\limsup \int f_n d\mu \leq \int f d\mu$, and then combining the two results yields the desired product. □

Corollary 2.11.1. *Under the hypothesis of the DCT, we also have $\lim_{n\to\infty} |f_n - f| d\mu = 0$.*

Finally, to apply most of our familiar results about integration.

Proposition 2.12. *Let $[a,b]$ be a closed interval in $\mathbb{R}$, and let $f : [a,b] \to \mathbb{R}$ be a continuous function. Let λ be a Lebesgue measure on $[a,b]$. Show that $\int f d\lambda$ is equal to the Riemann integral $\int_a^b f(x)dx$.*

Proof. First, notice that the minimum and maximum of continuous functions are continuous, so f^+ and f^- are both continuous. Then, notice that $\int f d\lambda = \int f^+ d\lambda - \int f^- d\lambda$ *and* $\int_a^b f(x)dx = \int_a^b f^+(x)dx - \int_a^b f^-(x)dx$. Therefore, if we can establish the result in the case that f is nonnegative, then the result follows in the general case. Thus, we may assume that f is non-negative.

Suppose that f is continuous. Then,

$$\int_a^b f(x) = \lim_{\max_k(x_k - x_{k-1}) \to 0} \sum_{k=1}^n f(x_k^*)(x_k - x_{k-1})$$

where $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$ is a sequence of points in $[a,b]$ and $x_k^* \in [x_k, x_{k-1}]$ for each k. Now, fix $\varepsilon > 0$. Then, since continuous functions are uniformly continuous on closed intervals, there exists some δ such that whenever $x, y \in [a,b]$, $|x-y| < \delta$ implies $|f(x) - f(y)| < \frac{\varepsilon}{2(a-b)}$. Furthermore, there exists a pair of sequences $x_1 \leq \cdots \leq x_n$ and $x_1^*, \ldots, x_n^*$ such that $\max_k(x_k - x_{k-1}) < \delta$ and

$$\left| \int_a^b f(x) - \sum_{k=1}^n f(x_k^*)(x_k - x_{k-1}) \right| < \frac{\varepsilon}{2}.$$

Then, it follows that

$$\begin{aligned} \int_a^b f(x) - \varepsilon &\leq \sum_{k=1}^n f(x_k^*)(x_k - x_{k-1}) - \frac{\varepsilon}{2} = \sum_{k=1}^n f(x_k^*)(x_k - x_{k-1}) - \sum_{k=1}^n \frac{\varepsilon}{2(a-b)}(x_k - x_{k-1}) \\ &= \sum_{k=1}^n \left(f(x_k^*) - \frac{\varepsilon}{2(a-b)} \right)(x_k - x_{k-1}) \leq \sum_{k=1}^n \inf_{x \in [x_{k-1}, x_k]} f(x)(x_k - x_{k-1}). \end{aligned}$$

But $g = \inf_{x \in [x_{k-1}, x_k]} f(x)$ is a simple function with integral $\sum_{k=1}^n \inf_{x \in [x_{k-1}, x_k]} f(x)(x_k - x_{k-1})$. Thus,

$$\int_a^b f(x)dx - \varepsilon \leq \sum_{k=1}^n \inf_{x \in [x_{k-1}, x_k]} f(x)(x_k - x_{k-1}) = \int g d\lambda \leq \sup_{g \in \mathrm{SF}^+} \int g d\lambda \leq \int f d\lambda$$

given that $g \leq f$ by definition. Therefore, for any $\varepsilon > 0$, we have $\int_a^b f(x)dx - \varepsilon \leq \int f d\lambda$ whence by driving $\varepsilon \to 0$, we obtain $\int_a^b f(x) \leq \int f d\lambda$.

On the other hand, fix any simple function $g \leq f$ defined on $[a,b]$. Let $y_0, y_1, \ldots, y_n$ be the points at which g changes value. Then, for any j, define the sequence x^j to be given by subdividing each interval in the sequence y into j parts, and let x^{j*} be the sequence given by defining $x_k^{j*} = \frac{x_k^j - x_{k-1}^j}{2}$. Then as $j \to \infty$,$\max_k(x_k^j - x_{k-1}^j) \to 0$. Thus,

$$\int g d\lambda \leq \sum f(x^{j*}k)(x_k^j - x_{k-1}^j) = \lim_{\max_k(x_k - x_{k-1}) \to 0} \sum_{k=1}^n f(x_k^*)(x_k - x_{k-1}) = \int_a^b f(x).$$

But then, by taking the supremum over all simple functions $g \leq f$, we obtain $\int f d\lambda \leq \int_a^b f(x)$, as desired. Combining this with the result of the preceding paragraph yields the desired result $\int f d\lambda \leq \int_a^b f(x)$. ☐

2.3 "Almost Everywhere"

Definition 2.13 (Almost Everywhere)**.** Given a measure space $(\Omega, \mathscr{F}, \mu)$, an event $A \in \mathscr{F}$ is said to "happen almost everywhere (a.e.)" (or, in probability theory, almost surely) if $\mu(A^c) = 0$.

For example, we say that $f = g$ almost everywhere if the set of points at which they are different is null.

Proposition 2.14. *Let $f : \Omega \to [0, \infty]$ be a measurable function. Then $\int f d\mu = 0$ if and only if $f = 0$ almost everywhere.*

Proof. If $f = 0$ almost everywhere, then it is clear the integral of any simple function $g \leq f$ is 0, so $\int f d\mu = 0$. On the other hand, suppose that $\mu(f^{-1}((0, \infty])) > 0$. Then,

$$\mu(f^{-1}((0,\infty])) = \mu\left(\bigcup_{n=1}^{\infty}\{f^{-1}((1/n,\infty])\right) = \lim_{n\to\infty}\mu(f^{-1}((1/n,\infty]) > 0.$$

But then, for some n, $\mu(f^{-1}((1/n, \infty])) > 0$. Yet then we obtain the desired result:

$$\int f\mu \geq \int f 1_{A_n} d\mu \geq \int n^{-1} 1_{A_n} d\mu = n^{-1}\mu(A_n) > 0.$$

□

Any result about integration can usually have its hypotheses replaced with almost-everywhere versions of these hypotheses to get maximally general results, as the above theorem shows precisely that null sets are the largest sets on which functions can be modified without changing their integrals.

2.4 Finite-Dimensional Product Spaces

Definition 2.15 (Product σ-Algebra)**.** Let $(\Omega_1, \mathscr{F}_1), \ldots, (\Omega_n, \mathscr{F}_n)$ be measurable spaces. Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$. Then, the product σ-algebra $\mathscr{F}$ (often denoted $\mathscr{F}_1 \times \cdots \times \mathscr{F}_n$ by abuse of notation) on Ω is defined to be the σ-algebra generated by sets of the form $A_1 \times \cdots \times A_n$, where $A_i \in \mathscr{F}_i$ for each i.

Proposition 2.16 (Product Measure)**.** *Let Ω and $\mathscr{F}$ be as above. Then, if Ω_i is endowed with a σ-finite measure μ_i for each i, there is a unique measure μ on Ω which satisfies, for any $A_i \in \mathscr{F}_i$,*

$$\mu(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mu_i(A_i).$$

Proof. Now, the collection of finite disjoint unions of sets of the form $A_1 \times \cdots \times A_n$ form an algebra. Therefore, by Carathédory's Theorem, it suffices to show that μ, as defined above, is a measure on this algebra. We prove this by induction on n; the base case $n = 1$ is obvious, so assume that the result holds for $n-1$.

Therefore, take any rectangular set $A_1 \times \cdots \times A_n$. Suppose that this set is a disjoint union of $A_{i,1} \times \cdots \times A_{i,n}$ for $i = 1, 2, \ldots$ where $A_{i,j} \in \mathscr{F}_j$ for each i, j. Then, it suffices to show that

$$\mu(A_1 \times \cdots \times) = \sum_{i=1}^{\infty}(A_{i,1} \times \cdots \times A_{i,n}).$$

Now, take $x \in A_1 \times \cdots \times A_{n-1}$. Let $I = \{i \mid x \in A_{i,1} \times \cdots \times A_{i,n-1}\}$. Then $\mu_n(A_n) = \sum_{i\in I} \mu_n(A_{i,n})$. On the other hand, if $x \notin A_1 \times \cdots \times A_{n-1}$ and $x \in A_{i,1} \times \cdots \times A_{i,n-1}$ for some i, then $A_{i,n}$ must be empty. Thus,

$$1_{A_1\times\cdots\times A_{n-1}}(x)\mu_n(A_n) = \sum_{i=1}^{\infty} 1_{A_{i,1}\times\cdots\times A_{i,n-1}}(x)\mu_n(A_{i,n}).$$

Then, let $\mu' = \mu_1 \times \cdots \times \mu_{n-1}$ be the measure given by the induction hypothesis. Integrating both sides with respect to μ' on $\Omega_1 \times \cdots \times \Omega_{n-1}$, we find that

$$\mu'(A_1 \times \cdots \times A_{n-1})\mu_n(A_n) = \sum_{i=1}^{\infty} \mu'(A_{i,1} \times \cdots \times A_{i,n-1})\mu_n(A_{i,n})$$

which is the desired result. □

2.5 Fubini's Theorem

Lemma 2.17. *Let $(\Omega_i, \mathscr{F}_i)$, $i = 1,2,3$ be measurable spaces. Let $f : \Omega_1 \times \Omega_2 \to \Omega_2$ be a measurable function. Then for all $x \in \Omega_1$, the map $y \mapsto f(x,y)$ is measurable on Ω_2.*

Proof. Take any $A \in \mathscr{F}_3$ and $x \in \Omega_1$. Let $B = f^{-1}(A)$ and $B_x = \{y \in \Omega \mid f(x,y) \in A\}$. Our goal is to demonstrate that $B_x \in \mathscr{F}_2$. Fixing x, let $\mathscr{G} = \{E \in \mathscr{F}_1 \times \mathscr{F}_2 \mid E_x \in \mathscr{F}\}$ where $E_x = \{y \in \Omega_2 \mid (x,y) \in E\}$. Then $\mathscr{G}$ is a σ-algebra which contains every rectangular set. Thus, $\mathscr{G}$ contains $\mathscr{F}_1 \times \mathscr{F}_2$, so $B_x \in \mathscr{F}_2$ for every $x \in \Omega_1$, as desired. □

Theorem 2.18 (Fubini's Theorem)**.** *Let $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$ be two σ-finite measure spaces. Let $\mu = \mu_1 \times \mu_2$ and let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}^*$ be a measurable function. If f is either nonnegative or integrable, then the map $x \mapsto \int_{\Omega_2} f(x,y) d\mu_2(y)$ on Ω_1 and the map $y \mapsto \int_{\Omega_1} f(x,y) d\mu_1(x)$ on Ω_2 are well-defined and measurable (when set equal to zero if the integral is undefined). Moreover,*

$$\int_{\Omega_1 \times \Omega_2} f(x,y) d\mu(x,y) = \int_{\Omega_1} \int_{\Omega_2} f(x,y) d\mu_2(y) d\mu_1(x) = \int_{\Omega_2} \int_{\Omega_1} f(x,y) d\mu_1(x) d\mu_2(y).$$

Finally, if either of

$$\int_{\Omega_1} \int_{\Omega_2} |f(x,y)| d\mu_2(y) d\mu_1(x) = \int_{\Omega_2} \int_{\Omega_1} |f(x,y)| d\mu_1(x) d\mu_2(y)$$

is finite, then f is integrable.

Proof. First, suppose that $f = 1_A$ for some $A \in \mathscr{F}_1 \times \mathscr{F}_2$. Then, for any $x \in \Omega_1$, $\int_{\Omega_2} f(x,y) d\mu_2(y) = \mu_2(A_x)$, where $A_x = \{y \in \Omega_2 \mid (x,y) \in A\}$. Now, our goal is to show that $x \mapsto \mu_2(A_x)$ is a measurable map.

Let $\mathscr{L}$ be the set of all $E \in \mathscr{F}_1 \times \mathscr{F}_2$ such that $x \mapsto \mu_2(E_x)$ is a measurable map on Ω_1 whose integral is $\mu(E)$. We demonstrate that $\mathscr{L}$ is a λ-system, first under the assumption that μ_1 and μ_2 are both finite measures. Now, clearly $\Omega_1 \times \Omega_2 \in \mathcal{L}$. Suppose $E_1, E_2, \dots \in \mathscr{L}$ are disjoint with union E, then E_x is the disjoint union of $(E_1)_x, (E_2)_x, \dots$ whence $\mu_2(E_x) = \sum_{i=1}^{\infty} \mu((E_i)_x)$. Thus, $x \mapsto \mu_2(E_x)$ is measurable. By the monotone convergence theorem,

$$\int_{\Omega_1} \mu_2(E_x) d\mu_1(x) = \sum_{i=1}^{\infty} \int_{\Omega_1} \mu_2((E_i)_x) d\mu_1(x) = \sum_{i=1}^{\infty} \mu(E_i) = \mu(E).$$

Thus $E \in \mathscr{L}$ and $\mathscr{L}$ is closed under countable disjoint unions. Finally, take $E \in \mathscr{L}$. Since μ_1 and μ_2 are finite, $\mu_2((E^{\mathsf{c}})_x) = \mu_2((E_x)^{\mathsf{c}}) = \mu_2(\Omega_2) - \mu_2(E_x)$ whence $x \mapsto \mu_2((E^{\mathsf{c}})_x)$ is measurable. Then,

$$\int_{\Omega_1} \mu_2((E^{\mathsf{c}})_x) d\mu_1(x) = \mu_1(\Omega_1)\mu_2(\Omega_2) - \int_{\Omega_1} \mu_2(E_x) d\mu_1(x) = \mu(\Omega) - \mu(E) = \mu(E^{\mathsf{c}}).$$

whence $\mathscr{L}$ is a λ-system. Furthermore, it contains the π-system of all rectangles, which generates $\mathscr{F}_1 \times \mathscr{F}_2$, so by the Dynkin π-λ theorem it contains $\mathscr{F}_1 \times \mathscr{F}_2$, as desired.

Now, let μ_1 and μ_2 be σ-finite measures. Then let $\{E_{n,1}\}_{n \geq 1}$ and $\{E_{n,2}\}_{n \geq 1}$ be sequences of measure sets of finite measure increasing to Ω_1 and Ω_2. For each n, let $E_n = E_{n,1} \times E_{n,2}$, and define the functionals $\mu_{n,i}(A) = \mu_i(A \cap E_{n,i})$ for each n and $i = 1,2$. Also define $\mu_n(E) = \mu(E \cap E_n)$. These are finite measures increasing up to μ_i and μ respectively. Then, if $f : \Omega_1 \to [0, \infty]$ is a measurable function,

$$\int_{\Omega_1} f(x) d\mu_{n,1}(x) = \int_{\Omega_1} f(x) 1_{E_{n,1}}(x) d\mu_1(x),$$

where we use the convention $\infty \cdot 0 = 0$ on the right (this follows first for indicator functions, then for simple functions by linearity, and then for nonnegative measurable functions by the monotone convergence theorem).

Then, for any $E \in \mathscr{F}_1 \times \mathscr{F}_2$ and any $x \in \Omega_1$, $\mu_2(E_x)$ is the increasing limit of $\mu_{n,2}(E_x)1_{E_{n,1}}(x)$. This demonstrates that $x \mapsto \mu_2(E_x)$ is measurable. Furthermore, $\mu_n = \mu_{n,1} \times \mu_{n,2}$ because they agree on the generating set of all rectangles, so the monotone convergence theorem yields that

$$\int_{\Omega_1} \mu_2(E_x) d\mu_1 = \lim_{n\to\infty} \int_{\Omega_1} \mu_{n,2}(E_x) 1_{E_{n,1}}(x) d\mu_1(x) = \lim_{n\to\infty} \int_{\Omega_1} \mu_{n,2}(E_x) d\mu_{n,1}(x) = \lim_{n\to\infty} \mu_n(E) = \mu(E).$$

This shows that Fubini's theorem holds for all indicator functions. By linearity, it holds for all simple functions, and by the monotone convergence theorem, it holds for all nonnegative measurable functions. Then we can conclude the result for any integrable f using the case of Fubini's Theorem for nonnegative measurable functions separately for f^+ and f^-. □

As an application of Fubini's Theorem, we have the following:

Theorem 2.19. *If $f_1, f_2, \dots$ are measurable functions from Ω into $\mathbb{R}$ such that*

$$\sum_{i=1}^{\infty} \int |f_i| d\mu < \infty,$$

then show that the set of ω where $\sum f_i(\omega)$ does not exist is a measurable set of measure zero, and if we define $\sum f_i$ arbitrarily on this set (e.g., equal to zero), then $\int \sum f_i d\mu = \sum \int f_i d\mu$.

Proof. For this, we apply Fubini's theorem. Indeed, let $\Omega_1 = \mathbb{R}$, $\mathscr{F}_1$ be the Borel σ-algebra, and μ_1 be the Lebesgue measure. On the other hand, let $\Omega_2 = \mathbb{Z}^+$, $\mathscr{F}_2 = \mathcal{P}(\mathbb{Z}^+)$, and define $\mu_2(S) = |S|$. Now, define $f(x, n) = f_n(x)$. First, let us demonstrate that f is measurable. Indeed, consider a measurable subset $M \subseteq \mathbb{R}$. Then $f^{-1}(M)$ is the countable union of the measurable sets $\bigcup_{n=1}^{\infty} f_n^{-1}(M) \times \{n\}$ and therefore measurable, so f is indeed measurable. Finally, notice that integration with respect to the described measure μ_2 is simply summation. That is, if $g : \mathbb{Z}^+ \to \mathbb{R}$ is a function, then $\int_{\mathbb{Z}^+} g d\mu_2 = \sum_{n=1}^{\infty} g(n)$. Then, it suffices to use Fubini's theorem.

Indeed, notice that $\int_{\Omega_2} \int_{\Omega_1} |f(x, y)| d\mu_1(x) d\mu_2(y) = \sum_{i=1}^{\infty} \int |f_i| d\mu < \infty$, so f is integrable. Then, by Fubini's Theorem, the map $x \mapsto \Omega_2 f(x, y) d\mu_2(y) = \sum_{n=1}^{\infty} f(x, n) = \sum f_i(x)$ is defined almost everywhere. Furthermore, if we set this map equal to zero where the integral is undefined, Fubini's Theorem also yields

$$\int \sum f_i d\mu = \int_{\Omega_1} \int_{\Omega_2} f(x, y) d\mu_2(y) d\mu_1(x) = \int_{\Omega_2} \int_{\Omega_1} f(x, y) d\mu_1(x) d\mu_2(y) = \sum \int f_i d\mu$$

which is the desired result. □

2.6 Infinite-Dimensional Product Spaces

Definition 2.20 (Infinite Product of Probability Spaces)**.** Let $\{(\Omega_i, \mathscr{F}_i, \mu_i)\}_{i \geq 1}$ be a countable collection of probability spaces. Then the *product σ-algebra* $\mathscr{F} = \prod_{i \geq 1} \mathscr{F}_i$ on $\Omega = \Omega_1 \times \Omega_2 \times \cdots$ by sets of the form $A_1 \times A_2 \times \cdots$ where at most finitely many A_i are not equal to Ω_i.

Theorem 2.21. *In the above case, there exists a unique probability measure μ on $(\Omega, \mathscr{F})$ such that $\mu(A_1 \times A_2 \times \cdots) = \prod_{i=1}^{\infty} \mu(A_i)$ whenever all but finitely many A_i are equal to Ω_i.*

Proof. For each n, let $\nu_n = \mu_1 \times \cdots \times \mu_n$. Let $\Omega^{(n)} = \Omega_{n+1} \times \Omega_{n+2} \times \cdots$. A set $A \in \mathscr{F}$ is called a *cylinder set* if it is of the form $B \times \Omega^{(n)}$ for some n and $B \in \mathscr{F}_1 \times \mathscr{F}_n$. Then let $\mathscr{A}$ be the collection of all cylinder sets. Then $\mathscr{A}$ is an algebra and $\sigma(\mathscr{A}) = \mathscr{F}$. Then define μ on $\mathscr{A}$ as follows: if $A \in \mathscr{A}$ is $B \times \Omega^{(n)}$, let $\mu(A) = \nu_n(B)$. This can be easily verified to be well-defined.

Now, to show that μ is a measure, it suffices to show that μ is countably additive on $\mathscr{A}$ by Carathéodory's Theorem. Let $A_1, A_2, \dots \in \mathscr{A}$ be disjoint such that $A = \bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$. Then, for each n, let $B_n = A \setminus (\bigcup_{i=1}^{n} A_i)$. Then, since $\mathscr{A}$ is an algebra, $B_n \in \mathscr{A}$, and A is the disjoint union of $A_1, \dots, A_n, B_n$. But μ is clearly finitely additive, so $\mu(A) = \mu(B_n) + \mu(A_1) + \cdots + \mu(A_n)$ for each n. Therefore, it suffices to show that $\lim \mu(B_n) = 0$.

Since $\{B_n\}_{n\geq 1}$ is a decreasing sequence of sets, there is some $\varepsilon > 0$ such that $\mu(B_n) \geq \varepsilon$ for all n. We will use this fact to yield a contradiction with the fact $\bigcap_{n=1}^{\infty} B_n = \varnothing$.

For each n, let $\mathscr{A}^{(n)}$ be the algebra of all cylinder sets in $\Omega^{(n)}$, and let $\mu^{(n)}$ be the analogue of μ for $A^{(n)}$. Then, for any n, m and $(x_1, \dots, x_m) \in \Omega_1 \times \cdots \times \Omega_m$, define $B_n(x_1, \dots, x_m) = \{(x_{m+1}, x_{m+2}, \dots) \mid (x_1, \dots, x_m, x_{m+1}, x_{m+2}, \dots) \in B_n\}$. By a previous lemma, $B_n(x_1) \in \mathscr{A}^{(1)}$ and by Fubini's Theorem, the map $x_1 \mapsto \mu^{(1)}(B_n(x_1))$ is measurable ($\mu^{(1)}$ is evidently a measure on the σ-algebra of all sets of the form $D \times \Omega^{(m)} \subseteq \Omega^{(1)}$). Thus, the set $F_n = \{x_1 \in \Omega_1 \mid \mu^{(1)}(B_n(x_1)) \geq \frac{\varepsilon}{2}\} \in \mathscr{F}_1$.

Then, by Fubini's Theorem,

$$\mu(B_n) = \int \mu^{(1)}(B_n(x_1)) d\mu_1(x_1) = \int_{F_n} \mu^{(1)}(B_n(x_1)) d\mu_1(x_1) + \int_{F_n^{\mathsf{c}}} \mu^{(1)}(B_n(x_1)) d\mu_1(x) \leq \mu_1(F_n) + \frac{\varepsilon}{2}.$$

Therefore, $\mu_1(F_n) \geq \varepsilon/2$. Since $\{F_n\}_{n\geq 1}$ is a decreasing sequence of sets, $\bigcap F_n \neq \varnothing$. Choose $x_1^* \in \bigcap F_n$. Repeating the above argument for the product space $\Omega^{(1)}$ and the sequence $\{B_n(x_1^*)\}_{n\geq 1}$, we find $x_2^* \in \Omega_2$ such that $\mu^{(2)}(B_n(x_1^*, x_2^*)) \geq \varepsilon/4$ for every n.

Then, we get a point $x = (x_1^*, x_2^*, \dots) \in \Omega$ such that for any m, n, $\mu^{(m)}(B_n(x_1^*, \dots, x_m^*)) \geq \frac{\varepsilon}{2^m}$. Then, for any n, notice that since B_n is a cylinder set, it is of the form $C_n \times \Omega^{(m_n)}$ for some m_n and some $C_n \in \mathscr{F}_1 \times \cdots \times \mathscr{F}_{m_n}$. Since $\mu^{(m_n)}(B_n(x_1^*, \dots, x_{m_n}^*) > 0$, there is some $(x_{m_n+1}, x_{m+2}, \dots) \in \Omega^{(m_n)}$ such that $(x_1^*, \dots, x_{m_n}^*, x_{m_n+1}, \dots) \in B_n$. But then $x \in B_n$, so $x \in \bigcap_n B_n$, yielding the desired contradiction. $\square$

3 Random Variables

Definition 3.1 (Random Variable)**.** A random variable X is a measurable map from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to $\mathbb{R}$. The interpretation of this definition is that each element $\omega \in \Omega$ is the outcome of some randomized experiment, and that $X(\omega)$ is a value attached to this outcome.

Definition 3.2 (Law of a Random Variable)**.** The *law of a random variable* X is a measure μ_X defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ as $\mu_X(A) = \mathbb{P}(X \in A) := \mathbb{P}(\{\omega \mid X(\omega) = A\})$

Given any probability measure μ on $\mathbb{R}$, there exists a random variable X with $\mu_X = \mu$. We construct this random variable by letting $\Omega \in \mathbb{R}$, $\mathscr{F} = \mathcal{B}(\mathbb{R})$, $\mathbb{P} = \mu$, and then letting $X : \Omega \to \mathbb{R}$ be $X(\omega) = \omega$.

Definition 3.3 (Cumulative Distribution Function)**.** The *cumulative distribution function* (c.d.f.) for F_X of a random variable X is defined as $F_X(t) = P(X \leq t) = \mu_X((-\infty, t])$. Notice that since half-open intervals generate the Borel sets as a σ-algebra, the c.d.f. uniquely determines the law.

Given any non-decreasing, right continuous function $F : \mathbb{R} \to [0, 1]$ such that $\lim_{t\to\infty} F(t) = 1$ and $\lim_{t\to-\infty} F(t) = 0$, then there exists a random-variable with c.d.f. F. To construct this, we simply define $\mu((a, b]) = F(b) - F(a)$ and then use Caratheodory's theorem to construct μ everywhere.

Definition 3.4 (Probability Density Function)**.** A measurable function $f : \mathbb{R} \to [0, \infty)$ is called a *probability density function* (p.d.f.) if $\int_{-\infty}^{\infty} f(x)dx = 1$. A p.d.f. defines a probability measure on $\mathbb{R}$ (precisely, on the set of Lebesgue-measurable subsets of $\mathbb{R}$) given by $\mu(A) = \int_A f dx$.

Definition 3.5 (Random Variables and P.D.F.s)**.** A random variable X is said to *have a p.d.f.* f if the probability measure defined by f is the law of X.

Example 3.6. Not all random variables have a p.d.f. Indeed, any random variable X such that $\mathbb{P}(X = a) > 0$ for any fixed a has no probably density function.

Notice that if f and g are two densities of X, they must be equal almost everywhere. Therefore, up to almost-everywhere equality, it makes sense to define $X \sim f$.

3.1 Expected Value and Variance

Definition 3.7 (Expected Value of a Random Variable)**.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable. We say that the expected value $\mathbb{E}[X]$ of X exists if $\int X dP$ is defined; in this case, we define $\mathbb{E}[X] = \int X d\mathbb{P}$.

Immediately, we notice from this definition that expectation is linear: i.e., for any integrable random variables X and Y, $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

Proposition 3.8. *Given any random variable X and any measurable function $g : \mathbb{R} \to \mathbb{R}$. Then*

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) d\mu_X(x)$$

in the sense that the left-hand side exists iff the right-hand side exists.

Proof. Take an arbitrary random variable $X : \Omega \to \mathbb{R}$, and let $g : \mathbb{R} \to \mathbb{R}$ be a simple function; i.e., $g = \sum_{i=1}^{n} a_i 1_{A_i}$ for $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$. Then, $\int g \circ X d\mathbb{P} = \sum_{i=1}^{n} g(a_i)\mathbb{P}(\{\omega \mid X(\omega) \in A_i\}) = \sum_{i=1}^{n} g(a_i)\mu_X(A_i) = \int g d\mu_X$. Now, let $g : \mathbb{R} \to [0, \infty)$ be a measurable function and $\{g_n\}$ be a sequence of nonnegative simple functions increasing to g. Then $0 \leq g_n \circ X$ and $g_n \circ X$ increases to $g \circ X$. Then by the Monotone Convergence Theorem, $\int g_n \circ X d\mathbb{P} \to \int g \circ X d\mathbb{P}$. But $\int g_n \circ X d\mathbb{P} = \int g_n d\mu_X$ and the latter increases to $\int g d\mu_X$. It is then straightforward to generalize to any measurable function $g : \mathbb{R} \to \mathbb{R}$ by splitting g into its positive and negative parts. ☐

Corollary 3.8.1. *$\mathbb{E}[X]$ exists if and only if $\int_{\mathbb{R}} x d\mu_X(x)$ exists, and then the two are equal.*

Proof. Apply the above proposition with the identity function $\mathrm{id} : \mathbb{R} \to \mathbb{R}$. ☐

Proposition 3.9. *Suppose $X \sim f$. Then, for any measurable g such that $g(X)$ is integrable*

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x) dx.$$

Proposition 3.10. *If X is a nonnegative random variable, prove that*

$$\sum_{n=1}^{\infty} \mathbb{P}(X \geq n) \leq \mathbb{E}(X) \leq \sum_{n=0}^{\infty} \mathbb{P}(X \geq n)$$

with equality on the left if X is integer-valued.

Proof. Notice that $\mathbb{P}(X \geq n) = \sum_{k=n}^{\infty} \mathbb{P}(k \leq X \leq k+1)$. Furthermore, since both sums have non-negative terms, rearrangement is valid. Therefore,

$$\sum_{n=1}^{\infty} \mathbb{P}(X \geq n) = \sum_{n=1}^{\infty}\sum_{k=n}^{\infty} \mathbb{P}(k \leq X < k+1) = \sum_{n=0}^{\infty} n\mathbb{P}(n \leq X < n+1)$$

$$\sum_{n=0}^{\infty} \mathbb{P}(X \geq n) = \sum_{n=0}^{\infty}\sum_{k=n}^{\infty} \mathbb{P}(k \leq X < k+1) = \sum_{n=0}^{\infty} (n+1)\mathbb{P}(n \leq X < n+1)$$

Yet, notice that $\Omega = \bigcup_{n=0}^{\infty}\{\omega \mid n \leq X(\omega) < n+1\}$ and furthermore this is a disjoint union. Thus,

$$\mathbb{E}(X) = \int_{\Omega} X d\mathbb{P} = \sum_{n=0}^{\infty} \int_{\{\omega \mid n \leq X(\omega) < n+1\}} X d\mathbb{P}.$$

But, of course, $n\mathbb{P}(n \leq X < n+1) \leq \int_{\{\omega \mid n \leq X(\omega) < n+1\}} X d\mathbb{P} \leq (n+1)\mathbb{P}(n \leq X < n+1)$ with equality on the left when X is integer-valued (for then $X = n$ on $\{\omega \mid n \leq X(\omega) < n+1\}$). This yields that, as desired,

$$\sum_{n=1}^{\infty} \mathbb{P}(X \geq n) = \sum_{n=0}^{\infty} n\mathbb{P}(n \leq X < n+1) \leq \mathbb{E}(X) \leq \sum_{n=0}^{\infty}(n+1)\mathbb{P}(n \leq X < n+1) = \sum_{n=0}^{\infty} \mathbb{P}(X \geq n)$$

with equality on the left when X is integer-valued, as desired. ☐

Theorem 3.11. *If X is a nonnegative random variable,*

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq t)dt = \int_0^\infty \mathbb{P}(X > t)dt$$

interpreting both integrals on the right as Lebesgue integrals with respect to Lebesgue measure.

Proof. Let $f(t,\omega) = 1$ if $X(\omega) \geq t$ and 0 otherwise. Then, $\int_0^\infty \mathbb{P}(X \geq t) = \int_0^\infty \int_{X \geq t} d\mathbb{P}dt = \int_{[0,\infty)} \int_\Omega f d\mathbb{P}dt = \int_\Omega \int_{[0,\infty)} f dt d\mathbb{P} = \int_\Omega X d\mathbb{P} = \mathbb{E}(X)$, where we may apply Fubini's Theorem on the basis that f is non-negative. Similarly, define $g(t,\omega) = 1$ if $X(\omega) > t$ and 0 otherwise. Then, again, $\int_0^\infty \mathbb{P}(X > t) = \int_0^\infty \int_{X>t} d\mathbb{P}dt = \int_{[0,\infty)} \int_\Omega g d\mathbb{P}dt = \int_\Omega \int_{[0,\infty)} g dt d\mathbb{P} = \int_\Omega X d\mathbb{P} = \mathbb{E}(X)$. ☐

Definition 3.12 (Variance)**.** The *variance* of a random variable X is defined to be

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

These two quantities can be seen to be equal by computing

$$\mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Proposition 3.13. *For any random variable X, $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$.*

Definition 3.14 (Covariance)**.** The *covariance* of random variables is defined to be

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

3.2 Standard Distributions

Following are a series of possible distributions for random variables:

Definition 3.15 (Normal Distribution)**.** A random variable X has the *normal* or *Gaussian* distribution with mean parameter $\mu \in \mathbb{R}$ and standard deviation parameter $\sigma > 0$ (denoted by $X \sim \mathcal{N}(\mu, \sigma)$) if it has p.d.f.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Definition 3.16 (Exponential Distribution)**.** A random variable X has the *exponential* distribution with rate parameter λ if it has p.d.f.

$$f(x) = \lambda e^{-\lambda x}$$

when $x \geq 0$ and 0 otherwise.

Definition 3.17 (Bernoulli Distribution)**.** A random variable X has the *Bernoulli distribution* with parameter p if $\mathbb{P}(X = 0) = 1 - p$ and $\mathbb{P}(X = 1) = p$.

Definition 3.18 (Binomial Distribution)**.** A random variable X has the *binomial distribution* with parameters n and p if $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for all $0 \leq k \neq n$ and 0 otherwise.

Definition 3.19 (Geometric Distribution)**.** A random variable X has the *geometric distribution* with parameter p if $\mathbb{P}(X = k) = (1-p)^{k-1}p$.

Definition 3.20 (Poisson Distribution)**.** A random variable X has the *Poisson distribution* with parameter λ if $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

3.3 Characteristic Functions

Definition 3.21 (Characteristic Function)**.** The *characteristic function* of a random variable X is defined by $\phi_X(t) = \mathbb{E}[e^{iXt}] = \mathbb{E}[\cos tX] + i\mathbb{E}[\sin tX]$.

The characteristic function of any random variable exists because $\mathbb{E}|e^{iXt}| = \mathbb{E}[1] = 1$; thus, e^{iXt} is integrable.

Proposition 3.22 (Boundedness of ϕ_X)**.** $|\phi_X(t)| \leq 1$.

Proof. This follows immediately from the identity $\left|\int f d\mu\right| \leq \int |f| d\mu$. This can be deduced by letting $re^{i\theta} = \int f d\mu$, and then working out

$$\left|\int f d\mu\right| = r = re^{-i\theta}e^{i\theta} = \int e^{-i\theta} f d\mu = \mathrm{Re}\left(\int e^{-i\theta} f d\mu\right) = \int \mathrm{Re}\left(e^{-i\theta} f\right) d\mu = \int |e^{-i\theta} f| d\mu = \int |f| d\mu.$$

□

Proposition 3.23 (Uniform Continuity of ϕ_X)**.** *For any random variable X, the characteristic function ϕ_X is continuous.*

Proof. Fix s, t. Then $|\phi_X(t) - \phi_X(s)| = |\mathbb{E}[e^{itX}(1 - e^{i(s-t)X})]| \leq \mathbb{E}|e^{itX}(1 - e^{i(s-t)X})| \leq \mathbb{E}|1 - e^{i(s-t)X}|$. Yet $1 - e^{i(s-t)X}$ converges to 0 for each $\omega \in \Omega$ and is dominated by the constant 2, so by the dominated convergence theorem $\mathbb{E}|1 - e^{i(s-t)X}| \to 0$ as $s - t \to 0$. In other words, for any $\varepsilon > 0$, there exists δ such that if $|s - t| < \delta$, then $|\phi_X(t) - \phi_X(s)| \leq \mathbb{E}|1 - e^{i(s-t)X}| < \varepsilon$, as desired. □

Proposition 3.24 (Symmetry and Real ϕ_X)**.** *A random variable X is symmetric around* 0 *(i.e. $\mathbb{P}(X \geq k) = \mathbb{P}(X \leq -k)$ for any $k \geq 0$) if and only if ϕ_X is real.*

Proof. This follows from the identity $\phi_{-X} = \overline{\phi_X}$. □

Proposition 3.25 (Convolution of ϕ_X)**.** *For any random variables X and Y, $\phi_{X*Y}(t) = \phi_X(t)\phi_Y(t)$.*

Proposition 3.26. *Suppose that $X \sim \mathcal{N}(0,1)$. Then, $\phi_X(t) = e^{-t^2}/2$.*

Proof. Now, $\phi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx = \frac{e^{-t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-it)^2/2} dx$.

Now, fix $R > 0$. Let C be the contour integral from $-R$ to R to $R - it$ to $-R - it$. Since the map $z \mapsto e^{-z^2/2}$ is entire, the integral of $e^{-z^2/2}$ along C is 0. Now, it can be easily shown that the vertical sections R to $R - it$ and $-R$ to $-R - it$ go to 0 as $R \to \infty$. Therefore, as $R \to \infty$, if C_1 denotes the contour $-R$ to R and C_2 denotes the contour $-R - it$ to $R - it$, then as $R \to \infty$

$$\int_{C_1} e^{-z^2/2} dz - \int_{C_2} e^{-z^2/2} dz \to 0.$$

But as $R \to \infty$, $\int_{C_1} e^{-z^2/2} dz \to \sqrt{2\pi}$ and $\int_{C_2} e^{-z^2/2} dz \to \int_{-\infty}^{\infty} e^{-(x-it)^2/2} dx$. The result follows. □

3.4 Independence

For the remainder of this section, assume that $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space.

Definition 3.27 (Independent Events)**.** Two events $A, B \in \mathscr{F}$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. More generally, $\{A_i\}_{i \in I}$ are independent if $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ whenever $i \neq j$.

Definition 3.28 (Independent σ-Algebras)**.** Let $\{G_i\}_{i \in I}$ be a collection of sub-σ-algebras of $\mathscr{F}$. Then these σ-algebras are independent if for any distinct $i_1, \ldots, i_k \in I$ and any $A_1 \in G_{i_1}, \ldots, A_k \in G_{i_k}$, we have $\mathbb{P}(\bigcap_{j=1}^k A_j) = \prod_{j=1}^k \mathbb{P}(A_j)$.

Proposition 3.29. *A collection $\{A_i\}_{i \in I}$ of events is independent if and only if the collection $\{\sigma(\{A_i\})\}_{i \in I} = \{\{\varnothing, A_i, A_i^{\mathsf{c}}, \Omega\}\}$ of σ-algebras is independent.*

Definition 3.30 (σ-Algebra Generated by Random Variables)**.** Let $\{X_i\}_{i\in I}$ be a collection of random variables defined on Ω. Then, the σ-algebra generated by $\{X_i\}_{i\in I}$, denoted $\sigma(\{X_i\}_{i\in I})$ is the σ-algebra of all sets of the form $X_i^{-1}(A)$ for $i \in I$, $A \in \mathcal{B}(\mathbb{R})$. This is the smallest σ-algebra such that all of the X_i are measurable.

Definition 3.31 (Independent Collections)**.** A set of collections of random variables $\{\{X_i\}_{i\in I_\alpha}\}_{\alpha\in\mathscr{A}}$ is independent if the σ-algebras $\{\sigma(\{X_i\}_{i\in I_\alpha})\}_{\alpha\in\mathscr{A}}$ are independent. In particular, $\{X_i\}_{i\in I}$ are independent if $\{\sigma(X_i)\}_{i\in I}$ are independent. This is equivalent to the statement that for all $i_1, \dots, i_k \in I$ and all $A_1, \dots, A_k \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}(X_{i_1} \in A_1, \dots, X_{i_k} \in A_k) = \mathbb{P}(X_{i_1} \in A_1) \cdots \mathbb{P}(X_{i_k} \in A_k)$.

Proposition 3.32. *Let $\mu_1, \mu_2, \dots$ be a sequence of probability measures on $\mathbb{R}$. Then there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and independent random variables $X_1, X_2, \dots$ on Ω such that μ_i is the law of X_i for each i.*

Proof. This is done by taking $\mathbb{P} = \mu_1 \times \mu_2 \times \cdots$, $\Omega = \mathbb{R}^{\mathbb{N}}$, and X_i defined in the obvious way. ☐

Example 3.33. There exist three three random variables X_1, X_2, X_3 that are pairwise independent but not independent.

Proof. Consider the uniform distribution on $\Omega = \{a, b, c, d\}$ (i.e. $\mathscr{F} = 2^\Omega$ and $\mu(S) = \frac{|S|}{4}$). Then let $X_1(\omega) = 1_{\{a,b\}}$, $X_2(\omega) = 1_{\{a,c\}}$, and $X_3(\omega) = 1_{\{b,c\}}$. To show that X_i and X_j are independent for $i \neq j$, it suffices to show that

$$\begin{aligned} \mathbb{P}(X_i = 1, X_j = 1) &= \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1) \qquad & \mathbb{P}(X_i = 0, X_j = 1) &= \mathbb{P}(X_i = 0)\mathbb{P}(X_j = 1) \\ \mathbb{P}(X_i = 1, X_j = 0) &= \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 0) \qquad & \mathbb{P}(X_i = 0, X_j = 0) &= \mathbb{P}(X_i = 0)\mathbb{P}(X_j = 0) \end{aligned}$$

Yet notice that $\mathbb{P}(X_i = n, X_j = m) = \frac{1}{4}$ and $\mathbb{P}(X_i = n) = \frac{1}{2}$ and $\mathbb{P}(X_j = m) = \frac{1}{2}$ for any $n, m \in \{0, 1\}$. Thus, all the above equalities hold and indeed X_i and X_j are independent whenever $i \neq j$. Yet X_1, X_2, and X_3 are not independent as $\mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1) = 0$ yet $\mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)\mathbb{P}(X_3 = 1) = \frac{1}{8}$. ☐

Theorem 3.34 (Multiplicativity of Expectation)**.** *Suppose $X_1, X_2, \dots, X_n$ are independent and integrable random variables defined on $(\Omega, \mathscr{F}, \mathbb{P})$. Then, $X_1 \cdots X_n$ is also integrable and $\mathbb{E}[X_1 \dots X_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n]$.*

Proof. By induction, it suffices to demonstrate the result for $n = 2$. First, suppose X and Y are nonnegative independent simple random variables, i.e., $X = \sum_i a_i 1_{A_i}$ and $Y = \sum_j b_j 1_{B_j}$. Then,

$$\mathbb{E}[XY] = \mathbb{E}[\sum_i \sum_j a_i b_j 1_{A_i} 1_{B_j}] = \sum_i \sum_j a_i b_j \mathbb{E}[1_{A_i \cap B_j}] = \sum_i \sum_j a_i b_j \mathbb{P}(A_i \cap B_j) = \sum i \sum_j a_i b_j \mathbb{P}(A_i)\mathbb{P}(B_j).$$

But the last expression is precisely $\mathbb{E}[X]\mathbb{E}[Y]$.

Now, suppose that X and Y are arbitrary nonnegative independent random variables. Then, there exist nonnegative simple random variables X_n increasing to X and Y_n increasing to Y. Yet then, looking at the construction, X_n is $\sigma(X)$-measurable and Y_n is $\sigma(Y)$-measurable, so that X_n and $Y - n$ are independent and $\mathbb{E}[X_n Y_n] = \mathbb{E}[X_n]\mathbb{E}[Y_n]$. But then $X_n Y_n$ increases to XY, so by the monotone convergence theorem we have $\mathbb{E}[XY] = \lim_{n\to\infty} \mathbb{E}[X_n Y_n] = \lim_{n\to\infty} \mathbb{E}[X_n]\mathbb{E}[Y_n] = \mathbb{E}[X]\mathbb{E}[Y]$.

Finally, let X and Y be arbitrary. Then X^+ and X^- are $\sigma(X)$-measurable and Y^+ and Y^- are $\sigma(Y)$-measurable, so that X^+ and X^- are independent with Y^+ and Y^-. Thus,

$$\begin{aligned} \mathbb{E}|XY| &= \mathbb{E}[(X^+ - X^-)(Y^+ - Y^-)] = \mathbb{E}[X^+Y^+] - \mathbb{E}[X^+Y^-] - \mathbb{E}[X^-Y^+] + \mathbb{E}[X^-Y^-] \\ &= \mathbb{E}[X^+]\mathbb{E}[Y^+] - \mathbb{E}[X^+]\mathbb{E}[Y^-] - \mathbb{E}[X^-]\mathbb{E}[Y^+] + \mathbb{E}[X^-]\mathbb{E}[Y^-]. \end{aligned}$$

This shows that XY is integrable, as $\mathbb{E}|X|$ and $\mathbb{E}|Y|$ are both finite. Then, repeating the processes with $\mathbb{E}[XY]$ instead of $\mathbb{E}|XY|$, we also obtain $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ as desired. ☐

Definition 3.35 (Uncorrelated)**.** Two random variables X and Y are said to be *uncorrelated* if $\mathrm{Cov}(X, Y) = 0$.

Proposition 3.36. *If X and Y are independent and integrable, then they are uncorrelated.*

Proof. First, suppose that X and Y are simple; that is, $X = \sum_{i=1}^k a_i 1_{A_i}$ and $Y = \sum_{j=1}^m b_j 1_{B_j}$ for distinct non-negative a_i and b_j and measurable A_i and B_j. Then, $A_i = X^{-1}(\{a_i\}) \in \sigma(X)$ and $B_j = Y^{-1}(\{b_j\}) \in \sigma(Y)$ whence A_i and B_j are independent for each i, j. Yet then

$$\mathbb{E}[XY] = \sum_{i=1}^k \sum_{j=1}^m a_i b_j \mathbb{E}(1_{A_i} 1_{B_j}) = \sum_{i=1}^k \sum_{j=1}^m a_i b_j \mathbb{P}(A_i \cap B_j) = \sum_{i=1}^k \sum_{j=1}^m a_i b_j \mathbb{P}(A_i)\mathbb{P}(B_j) = \mathbb{E}[X]\mathbb{E}[Y].$$

Then, suppose that X and Y are non-negative and independent. Then, there exist simple random variables X_n increasing to X and Y_n increasing to Y by Proposition 2.8. Then, X_n is $\sigma(X)$-measurable and Y_n is $\sigma(Y)$-measurable by the construction in Proposition 2.8. Thus X_n and Y_n are independent, so that $\mathbb{E}[X_n Y_n] = \mathbb{E}[X_n]\mathbb{E}[Y_n]$ by our work above. Then, $X_n \uparrow X$ and $Y_n \uparrow Y$ implies $X_n Y_n \uparrow XY$, so that by the monotone convergence theorem $\mathbb{E}[XY] = \lim_{n\to\infty} \mathbb{E}[X_n Y_n] = \lim_{n\to\infty} \mathbb{E}[X_n]\mathbb{E}[Y_n] = \mathbb{E}[X]\mathbb{E}[Y]$.

Then, suppose X and Y are independent. Then X^+ and X^- are $\sigma(X)$-measurable, and similarly Y^+ and Y^- are $\sigma(Y)$-measurable. Thus,

$$\begin{aligned}\mathbb{E}|XY| &= \mathbb{E}[(X^+ - X^-)(Y^+ - Y^-)] = \mathbb{E}[X^+Y^+] - \mathbb{E}[X^+Y^-] - \mathbb{E}[X^-Y^+] + \mathbb{E}[X^-Y^-] \\ &= \mathbb{E}[X^+]\mathbb{E}[Y^+] - \mathbb{E}[X^+]\mathbb{E}[Y^-] - \mathbb{E}[X^-]\mathbb{E}[Y^+] + \mathbb{E}[X^-]\mathbb{E}[Y^-].\end{aligned}$$

This shows that XY is integrable, as $\mathbb{E}[X^+]$, $\mathbb{E}[X^-]$, $\mathbb{E}[Y^+]$, and $\mathbb{E}[Y^-]$ are all finite. Then, repeating the processes with $\mathbb{E}[XY]$ instead of $\mathbb{E}|XY|$, we also obtain $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ as desired. □

4 Inequalities, L^p Spaces, and Lemmas

This section dives into the details of random variables and provides tools for analyzing them.

4.1 Concentration Inequalities

In this section, we develop some useful tools for determining when a random variable is close to a fixed value.

Proposition 4.1 (Markov's Inequality)**.** *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. Let $f : \Omega \to [0, \infty]$ be a measurable function. Then, for any $t > 0$, $\mu\{\omega \mid f(\omega) \geq t\}) \leq \frac{\int f d\mu}{t}$.*

Proof. Let $A = \{\omega \mid f(\omega) \leq t\}$. Then let $g = 1_A$ and $h = \frac{f}{t}$. Then $g \leq h$, so $\int g d\mu \leq \int h d\mu$. But $\int g d\mu = \mu\{\omega \mid f(\omega) \geq t\}$ and $\int h d\mu = \frac{\int f d\mu}{t}$, so we are done. □

Proposition 4.2 (Chebyshev's Inequality)**.** *Let X be any random varible with $\mathbb{E}[X^2] < \infty$. Then, for any $t > 0$,*

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

Proof. By Markov's inequality, $\mathbb{P}(|X - \mathbb{E}[X]| \geq t) = \mathbb{P}((X - \mathbb{E}[X])^2 \geq t^2) \leq \frac{\mathbb{E}[X-\mathbb{E}[X]]^2}{t^2} = \frac{\text{Var}(X)}{t^2}$. □

Proposition 4.3 (Cantelli's Inequality)**.** *Let X be a random variable with $\mathbb{E}[X^2] < \infty$. Then, for $t > 0$,*

$$\mathbb{P}(X - \mathbb{E}[X] \geq \lambda) \leq \frac{\sigma^2}{\sigma^2 + \lambda^2}.$$

Proof. Fix $u \geq 0$. Define $Y = X - \mathbb{E}[X]$. Then,

$$\mathbb{P}(X - \mathbb{E}[X] \geq \lambda) = \mathbb{P}(Y \geq \lambda) = \mathbb{P}(Y + u \geq \lambda + u) \leq \mathbb{P}((Y + u)^2 \geq (\lambda + u)^2) \leq \frac{\mathbb{E}[(Y + u)^2]}{(\lambda + u)^2} = \frac{\sigma^2 + u^2}{(\lambda + u)^2}$$

where the last inequality is an application of Markov's inequality. Then, notice that $\frac{\sigma^2+u^2}{(\lambda+u)^2}$ can be minimized by letting $u = \frac{\sigma^2}{\lambda}$, from which the desired inequality follows. □

Corollary 4.3.1. *Let X be a real-valued random variable with $\mathbb{E}[X^2] < \infty$. Then, for $t > 0$,*

$$\mathbb{P}(X - \mathbb{E}[X] \leq -\lambda) \leq \frac{\sigma^2}{\sigma^2 + \lambda^2}.$$

Proof. Apply the Cantelli inequality to $-X$. □

This is superior to Chebyshev's inequality for one-sided bounds, and inferior for two-sided bounds.

Following is an exploration of the Chernoff bound for independent random variables, which is useful for applying the probabilistic method. For these, we use the *moment generating function.*

Definition 4.4 (Moment Generating Function)**.** Let X be a random variable. Then the *moment generating function* $M_X(s)$ is defined to be $\mathbb{E}[e^{sX}]$.

Lemma 4.5. *Suppose that $X = \sum_{i=1}^n X_i$ where the X_i are independent random variables. Then,*

$$M_X(s) = \prod_{i=1}^n M_{X_i}(s).$$

Proposition 4.6 (Multiplicative Chernoff Bound)**.** *Let $X = \sum_{i=1}^n X_i$, where $X_i = 1$ with probability p_i, and $X_i = 0$ with probability $1 - p_i$, and all the X_i are independent. Let $\mu = \mathbb{E}[X]$. Then,*

(i) $\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2}{2+\delta}\mu}$ *for all* $\delta > 0$.

(ii) $\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2}$ *for all* $0 < \delta < 1$.

(iii) $\mathbb{P}(|X - \mu| \geq \delta\mu) \leq 2e^{-\mu\delta^2/3}$ *for all* $0 < \delta < 1$.

Proof.
(i): First, notice that

$$M_X(s) = \prod_{i=1}^n M_{X_i}(s) = \prod_{i=1}^n (p_i \cdot e^s + (1-p_i)) = \prod_{i=1}^n 1 + p_i(e^s - 1) \leq \prod_{i=1}^n e^{p_i(e^s-1)} = e^{(e^s-1)\mu}$$

Then, it follows that for any $s > 0$,

$$\mathbb{P}(X \geq (1+\delta)\mu) = \mathbb{P}(e^{sX} \geq e^{s(1+\delta)\mu}) \leq \frac{M_X(s)}{e^{s(1+\delta)\mu}} \leq e^{(e^s-1)\mu}e^{-s(1+\delta)\mu}.$$

Defining $s = \log(1+\delta)$, we obtain that

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{(e^{\log(1+\delta)}-1)\mu}e^{-\log(1+\delta)(1+\delta)\mu} = \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.$$

Now, $\log(1+x) \geq \frac{x}{1+\frac{x}{2}}$ for all $x > 0$. Thus,

$$\left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu = e^{\log\left(\left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu\right)} = e^{\mu(\delta-(1+\delta)\log(1+\delta))} \leq e^{\mu\delta\left(1-\frac{1+\delta}{1+\frac{\delta}{2}}\right)} = e^{-\mu\delta\frac{\frac{\delta}{2}}{1+\frac{\delta}{2}}} = e^{-\mu\frac{\delta^2}{2+\delta}}.$$

(ii): Now, it follows that for any $s < 0$

$$\mathbb{P}(X \leq (1-\delta)\mu) = \mathbb{P}(e^{sX} \geq e^{s(1-\delta)\mu}) \leq \frac{M_X(s)}{e^{s(1-\delta)\mu}} = e^{(e^s-1)\mu}e^{-s(1-\delta)\mu}.$$

Defining $s = \log(1-\delta)$, we obtain that

$$\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{(e^{\log(1-\delta)}-1)\mu}e^{-\log(1-\delta)(1-\delta)\mu} = \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu.$$

Now, $(1-x)\log(1-x) \geq -x + \frac{x^2}{2}$. Thus,

$$\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} = e^{\log\left(\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}\right)} = e^{\mu(-\delta-(1-\delta)\log(1-\delta))} \leq e^{\mu(-\delta+\delta-\delta^2/2)} = e^{-\mu\delta^2/2}.$$

(iii): This follows immediately from the prior two bounds. ☐

A similar additive result can be shown using the same methods:

Proposition 4.7 (Additive Chernoff Bound). *Let $X = \sum_{i=1}^{n} X_i$, where $X_i = 1$ with probability p_i and $X_i = 0$ with probability $1-p_i$, and all the X_i are independent. Let $\mu = \mathbb{E}[X]$. Then,*

(i) $\mathbb{P}(X \geq \mu + \delta n) \leq e^{-2n\delta^2}$.

(ii) $\mathbb{P}(X \leq \mu - \delta n) \leq e^{-2n\delta^2}$.

(iii) $\mathbb{P}(|X-\mu| \geq \delta n) \leq 2e^{-2n\delta^2}$.

4.2 The Borel-Cantelli Lemmas

Lemma 4.8 (First Borel-Cantelli Lemma). *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. Let $A_1, A_2, \dots \in \mathscr{F}$. Suppose that $\sum_{i=1}^{\infty} \mu(A_i) < \infty$. Then $\mu(\{\omega \mid \omega \in A_i \text{ i.o.}\}) = 0$.*

Proof. Notice that $\{\omega \mid \omega \in A_i \text{ i.o.}\} \subseteq \bigcap_{i=1}^{\infty} \bigcup_{j=i}^{\infty} A_j$. Thus let $B_i = \bigcup_{j=i}^{\infty} A_i$. Then $\mu(\{\omega \mid \omega \in A_i \text{ i.o.}\}) = \lim_{i\to\infty} \mu(B_i)$ because $\mu(B_1) < \infty$. But then $\mu(\{\omega \mid \omega \in A_i \text{ i.o.}\}) = \lim_{i\to\infty} \mu(B_i) \leq \lim_{i\to\infty} \sum_{n=i}^{\infty} A_n = 0$, because the tails of convergent series go to zero. ☐

Lemma 4.9 (Second Borel-Cantelli Lemma). *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of independent events. Then, if $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ diverges to $+\infty$, then $\mathbb{P}(A_n$ happens i.o.$) = 1$.*

Proof. Let B be the event of A_n happening infinitely often. Then $B = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$. Thus, $B^{\mathsf{c}} = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^{\mathsf{c}}$. Then, since $\bigcap_{k=n}^{\infty} A_k^{\mathsf{c}}$ is increasing in n, $\mathbb{P}(B^{\mathsf{c}}) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^{\mathsf{c}}\right)$. Yet then for any n and any $m > n$,

$$\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^{\mathsf{c}}\right) \leq \mathbb{P}\left(\bigcap_{k=n}^{m} A_k^{\mathsf{c}}\right) = \prod_{k=n}^{m} \mathbb{P}(A_k^{\mathsf{c}}) = \prod_{k=n}^{m} (1 - \mathbb{P}(A_k))$$

and since $1 - x \leq e^{-x}$ for all $x \geq 0$,

$$\prod_{k=n}^{m} (1 - \mathbb{P}(A_k)) \leq \prod_{k=n}^{m} e^{-\mathbb{P}(A_k)} = e^{-\sum_{k=n}^{m} \mathbb{P}(A_k)}$$

which goes to 0 as $m \to \infty$ by the assumption that $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ diverges to $+\infty$. Thus, $\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^{\mathsf{c}}\right) = 0$ for any n, whence $\mathbb{P}(B^{\mathsf{c}}) = \lim_{n\to\infty} 0 = 0$, yielding the desired result. ☐

Example 4.10. Suppose you have a random infinite string of the 26 letters of the English alphabet, where each letter is drawn independently and uniformly at random. Then, the probability that every word appears infinitely often is 1. To prove this, let $\mathcal{W}$ be the set of words; since each word has finite length, $\mathcal{W}$ is countable. Now, notice that it suffices to show that any single word appears infinitely often with probability 1, because then by using countable additivity on the complement, we obtain that every word appears infinitely often with probability 1. Yet this is immediate from the Borel-Cantelli Lemma, by letting A_i be the event that the $(i-1)|w|$th to $i|w|-1$th characters form w (so that $\mathbb{P}(A_i) = 26^{-|w|}$ and the sum diverges).

4.3 L^p Spaces

Definition 4.11 (L^p Spaces)**.** Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. For $p \in [1, \infty)$, let $L^p(\Omega, \mathscr{F}, \mu)$ denote the set of all measurable functions $f : \Omega \to \mathbb{R}$ such that $\int |f|^p d\mu < \infty$. Let $\|f\|_{L^p}$ denote $\left(\int |f|^p d\mu\right)^{1/p}$.

Proposition 4.12 (Jensen's Inequality)**.** *Let $(\Omega, \mathscr{F}, \mu)$ be a probability space. Let $f : \Omega \to I$ be a measurable function where $I \subseteq \mathbb{R}$ is an interval. Let $\phi : I \to \mathbb{R}$ be a convex function (i.e. $\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y)$ for all x, y). Informally, the line between any two points of the graph of ϕ is above the graph itself. Then, if $\phi \circ f$ is measurable,*

$$\int \phi \circ f d\mu \geq \phi\left(\int f d\mu\right)$$

Proof. Let $x = \int f d\mu$; then $x \in I$. Then, by choosing a to be any number in the interval

$$\left[\lim_{y \uparrow x} \frac{\phi(x) - \phi(y)}{x - y}, \lim_{y \downarrow x} \frac{\phi(y) - \phi(x)}{y - x}\right].$$

and defining $b = \phi(x) - ax$, a and b satisfy $ax + b = \phi(x)$ and $ay + b \leq \phi(y)$ for all $y \in I$. Then $\phi\left(\int f d\mu\right) = \phi(x) = ax + b = a \int f d\mu + b = \int (af + b)\, d\mu \leq \int \phi \circ f d\mu$. ☐

Proposition 4.13 (Young's Inequality)**.** *Suppose $p, q \in [1, \infty)$ are such that $\frac{1}{p} + \frac{1}{q} = 1$. Then $\forall x, y > 0$, $xy \leq \frac{x^p}{p} + \frac{y^q}{q}$.*

Proof. Let $\Omega = \{0, 1\}$ and $f : \Omega \to \mathbb{R}$ be $f(0) = x^p$ and $f(1) = y^q$. Let $\phi(z) = -\log z$. Then ϕ is convex, so Jensen's gives $\phi(\int f d\mu) = -\log\left(\frac{x^p}{p} + \frac{y^q}{q}\right) \leq \int \phi \circ f d\mu = \frac{1}{p}(-\log x^p) + \frac{1}{q}(-\log y^q) = -\log(xy)$. ☐

Proposition 4.14 (Hölder's Inequality)**.** *Suppose that $p, q \in [1, \infty)$ are such that $\frac{1}{p} + \frac{1}{q} = 1$. Take any $f \in L^p(\mu)$, $g \in L^q(\mu)$. Then $fg \in L^1(\mu)$ and $\|fg\|_{L^1} \leq \|f\|_{L^p} \|g\|_{L^q}$.*

Proof. Suppose that $\|f\|_{L^p} = \|g\|_{L^q} = 1$. Then, by Young's inequality, $|fg| \leq \frac{|f|^p}{p} + \frac{|g|^q}{q}$. Then $\int |fg| d\mu \leq \frac{\int |f|^p d\mu}{p} + \frac{\int |g|^q d\mu}{q} = \frac{1}{p} + \frac{1}{q} = 1$. Therefore, the result holds when f and g have norm 1, and we can obtain the general result by replacing f and g with $\frac{f}{\|f\|_{L^p}}$ and $\frac{g}{\|g\|_{L^q}}$. ☐

Lemma 4.15. *$f + g \in L^p$ implies that $f + g \in L^p$.*

Proof. Suppose that $p \geq 1$. Then $x \mapsto x^p$ is convex on $[0, \infty)$, s, $\left|\frac{a+b}{2}\right|^p \leq \left|\frac{|a|+|b|}{2}\right|^p \leq \frac{|a|^p}{2} + \frac{|b|^p}{2}$ by Jensen's. Then $\int \left|\frac{f+g}{2}\right|^p d\mu \leq \int \frac{|f|^p + |g|^p}{2} d\mu < \infty$ whence $\int |f + g|^p\, d\mu$ is finite. ☐

Corollary 4.15.1. *Any L^p space is a vector space.*

Theorem 4.16 (Minkowski's Inequality)**.** *For all $f, g \in L^p(\mu)$ and all $p \in [1, \infty)$, $\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}$.*

Proof. Notice that $p = 1$ is the trivial case of the triangle inequality. Thus, assume that $p \in (1, \infty)$. Furthermore, for now we will assume that $f + g \in L^p(\mu)$. Then

$$\int |f + g|^p d\mu = \int |f + g||f + g|^{p-1} d\mu \leq \int |f| \cdot |f + g|^{p-1} d\mu + \int |g||f + g|^{p-1} d\mu$$

Then, let q be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then $(p-1)q = p$. On the other hand, by Hölder's Inequality,

$$\int |f| \cdot |f+g|^{p-1} d\mu + \int |g||f+g|^{p-1} d\mu \leq \left(\int |f|^p\right)^{1/p} \left(\int |f + g|^{p-1} q d\mu\right)^{1/q} + \left(\int |g|^p\right)^{1/p} \left(\int |f + g|^{p-1} q d\mu\right)^{1/q}$$

But this just equals $(\|f\|_{L^p} + \|g\|_{L^p})\left(\int |f + g|^p d\mu\right)^{1/q}$, so indeed $\int |f+g|^p d\mu \leq (\|f\|_{L^p} + \|g\|_{L^p})\left(\int |f + g|^p d\mu\right)^{1/q}$ which can be rearranged to give the desired result. ☐

Theorem 4.17 (Riesz-Fischer Theorem)**.** *For any measure space* $(\Omega, \mathscr{F}, \mu)$ *and any* $p \in [1, \infty)$*,* $L^p(\Omega, \mathscr{F}, \mu)$ *is a complete normed space (i.e. any Cauchy sequence convergences).*

Proof. Let $\{f_n\}_{n\geq 1}$ be a Cauchy sequence in $L^p(\mu)$. Then, we can find a subsequence $\{f_{n_k}\}_{k\geq 1}$ such that $\|f_{n_k} - f_n\|_{L^p} < \frac{1}{2^k}$ for any $n > n_k$. Then, the sequence $\{f_{n_k}\}_{k=1}^{\infty}$ converges pointwise almost everywhere. To see why, define $A_k = \{\omega \mid |f_{n_k}(\omega) - f_{n_{k+1}}(\omega)| \geq 2^{-k/2}\}$, and notice that by Markov's inequality, $\mu(A_k) \leq 2^{-kp/2}$. Thus, $\sum_{k=1}^{\infty} \mu(A_k) \leq \sum_{k=1}^{\infty} 2^{-kp/2} < \infty$. Then, by the Borel-Cantelli Lemma, $\mu(\{\omega \mid \omega \in A_k \text{ infinitely often}\}) = 0$. But if ω is not in A_k infinitely often, then ω is in only finitely many A_k. Then $|f_{n_k}(\omega) - f_{n_{k+1}}(\omega)| < 2^{-k/2}$ for all but finitely many k. Thus $\sum_{k=1}^{\infty} |f_{n_{k+1}}(\omega) - f_{n_k}(\omega)| < \infty$ which implies that $\lim_{k\to\infty} f_{n_k}(\omega)$ exists. Define $f(\omega) = \lim_{k\to\infty} f_{n_k}(\omega)$ if the limit exists and 0 otherwise. By our work above, the latter case happens with measure 0.

Furthermore, by applying Fatou's Lemma to f_{n_k}, we find that f is in L^p. Indeed, $f_{n_k} \to f$ in L^p. To complete the proof, recall that if a Cauchy sequence in a metric space has a convergent subsequence, then the full Cauchy sequence converges to the same limit. □

Theorem 4.18 (Monotonicity in p)**.** *Let* $(\Omega, \mathscr{F}, \mu)$ *be a probability space. Then, for all* $1 \leq p \leq q$*,* $L^q(\mu) \subseteq L^p(\mu)$*. Moreover* $\|f\|_{L^p} \leq \|f\|_{L^q}$ *for all* f *in* L^q*.*

Proof. Now, since $x \mapsto x^{q/p}$ is convex, $\int |f|^q d\mu = \int (|f|^p)^{q/p} d\mu \geq \left(\int |f|^p d\mu\right)^{q/p}$, where the final step is by Jensen's inequality. Thus, $\|f\|_{L^q} \geq \|f\|_{L^p}$ □

On the other hand, monotonicity does not necessarily hold when Ω has infinite measure:

Proposition 4.19. *Let* λ *be the Lebesgue measure on* $\mathbb{R}$*. Then neither of* $L^1(\lambda)$ *and* $L^2(\lambda)$ *is a subset of the other.*

Proof. Let f be the function $x \mapsto \frac{1_{[1,\infty)}(x)}{x}$. Then,

$$\int_{\mathbb{R}} |f| d\lambda = \int_1^{\infty} \frac{1}{x} dx = \infty.$$

Thus, $f \notin L^1$. On the other hand,

$$\int_{\mathbb{R}} |f|^2 d\lambda = \int_1^{\infty} \frac{1}{x^2} dx = -\frac{1}{x}\Big|_1^{\infty} = 1.$$

Therefore, $f \in L^2(\lambda)$ but $f \notin L^1(\lambda)$, so indeed $L^2(\lambda) \not\subseteq L^1(\lambda)$.

On the other hand, let g be the function $x \mapsto \frac{1_{(0,1]}}{\sqrt{x}}$. Then,

$$\int_{\mathbb{R}} |g|^2 d\lambda = \lim_{t\to 0^+} \int_t^1 \frac{1}{x} dx = \lim_{t\to 0} (\ln(1) - \ln(t)) = \lim_{t\to 0} \ln\left(\frac{1}{t}\right) = \lim_{u\to\infty} \ln(u) = \infty.$$

Therefore, $g \notin L^2$. On the other hand,

$$\int_{\mathbb{R}} |g| d\lambda = \lim_{t\to 0^+} \int_t^1 \frac{1}{\sqrt{x}} dx = \lim_{t\to 0^+} (2\sqrt{1} - 2\sqrt{t}) = 2.$$

Therefore, $f \in L^1(\lambda)$ but $f \notin L^2(\lambda)$, so indeed $L^1(\lambda) \not\subseteq L^2(\lambda)$. □

4.4 The Kolmogorov Zero-One Law

Definition 4.20 (Tail σ-Algebra)**.** Let $X_1, X_2, \ldots$ be a sequence of random variables on a given probability space. Then the tail σ-algebra generated by this family is

$$\mathcal{T}(X_1, X_2, \ldots) := \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \ldots).$$

Theorem 4.21 (Kolmogorov Zero-One Law)**.** *If* $\{X_n\}_{n=1}^{\infty}$ *is a sequence of independent random variables defined on a probability space* $(\Omega, \mathscr{F}, \mathbb{P})$ *and* $\mathcal{T}$ *is the tail* σ*-algebra of this sequence, then for any* $A \in \mathcal{T}$*,* $\mathbb{P}(A)$ *is either* 0 *or* 1*.*

Proof. Take any n. Then, since $A \in \sigma(X_{n+1}, X_{n+2}, \dots)$ and the X_i's are independent, it follows that A is independent of the σ-algebra $\sigma(X_1, \dots, X_n)$. Then, let $\mathcal{A} = \bigcup_{n=1}^{\infty} \sigma(X_1, \dots, X_n)$. Then, $\sigma(\mathcal{A}) = \sigma(X_1, X_2, \dots)$. Then, A is independent of B for every $B \in \mathcal{A}$, so A is independent of $\sigma(X_1, X_2, \dots)$. But then A is independent of itself, so $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ whence $\mathbb{P}(A)$ is either 0 or 1. □

Example 4.22. Consider independent random variables $X_1, X_2, \dots$. Let $S_n = \sum_{i=1}^{n} X_i$ and $\{a_n\}$ be a sequence of positive real numbers increasing to ∞. Then, let $L = \limsup_{n\to\infty} \frac{S_n}{a_n}$. Then, for any $t \in \mathbb{R}$, the event $\{L \leq t\}$ is a tail event. Thus, for all t, $\mathbb{P}(L \leq t)$ is either 0 or 1. Therefore, there exists some $c \in [-\infty, \infty]$ such that $\mathbb{P}(L = c) = 1$. In summary, it follows that for any $a_n \uparrow \infty$, there exists some c such that $\mathbb{P}(\limsup \frac{S_n}{a_n} = c) = 1$.

5 Convergence Results

This section covers the laws of large numbers and the central limit theorem, which are the two main results which are used to show the convergence of sums of random variables.

5.1 Types of Convergence

First, we begin with a recap of various types of convergence. Then, we discuss equivalent formulations of these notions as well as various relationships between them:

Definition 5.1 (Convergence Almost Everywhere)**.** A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ on a probability space Ω *converges almost everywhere* to a random variable X on Ω if for almost all $\omega \in \Omega$, $\lim_{n\to\infty} X_n(\omega) \to \lim_{n\to\infty} X(\omega)$. This is denoted "$X_n \to X$ almost everywhere" or "$X_n \to X$ a.s.".

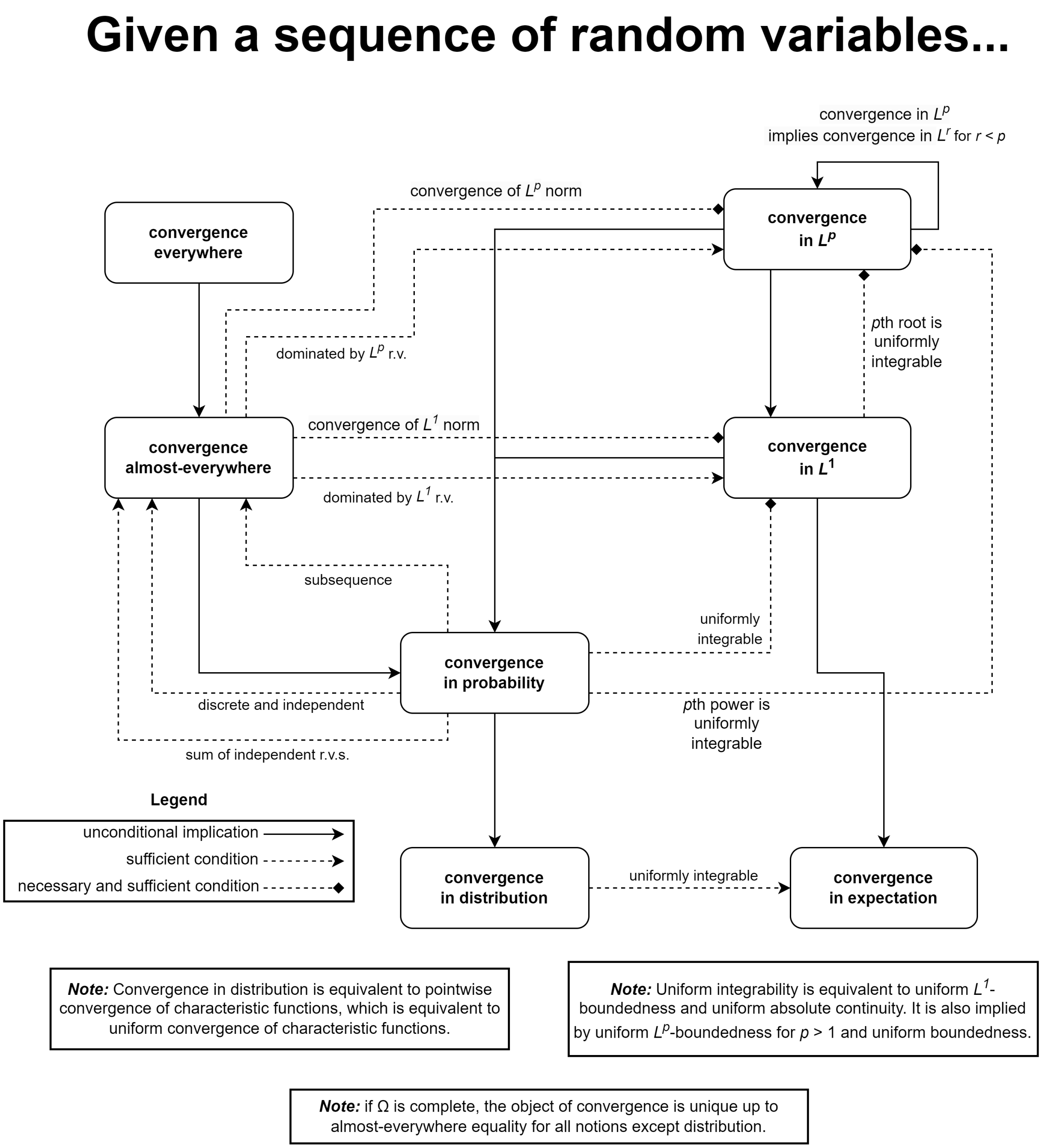
Definition 5.2 (Convergence in Probability)**.** A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ on a probability space Ω *converges in probability* to a random variable X on Ω if for all $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. This is denoted "$X_n \to X$ in probability" or "$X_n \xrightarrow{p} X$".

Definition 5.3 (Convergence in Distribution)**.** A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ with respective c.d.f. F_n *converges in distribution* to a random variable X with c.d.f. F if for any $t \in \mathbb{R}$ which is a continuity point of F, $\lim_{n\to\infty} F_n(t) = F(t)$. This is denoted "$X_n \to X$ in distribution" or "$X_n \xrightarrow{d} X$".

Definition 5.4 (Convergence in L^p)**.** A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ *converges in* L^p to a random variable X if $\|X_n - X\|_{L^p} \to 0$ as $n \to 0$. This is denoted "$X_n \to X$ in L^p" or "$X_n \xrightarrow{L^p} X$". The special case of $p = 1$, convergence in L^1, is particularly important.

Definition 5.5 (Convergence in Expectation)**.** A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ *converges in expectation* to a random variable X if $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Following is a diagram of the relations between the various types of convergence.



The remainder of this subsection is dedicated to proving this result.

5.1.1 Unconditional Relationships

Proposition 5.6. *$X_n \to X$ everywhere implies $X_n \to X$ a.e.*

Proof. Trivial. □

Proposition 5.7. *$X_n \to X$ a.e. implies that $X_n \to X$ in probability.*

Proof. Fix $\varepsilon > 0$. Let $A = \{\omega \mid \exists N_\omega \ \forall n > N_\omega \ |X_n(\omega) - X(\omega)| < \varepsilon\}$. By definition of a.e. convergence, $\mathbb{P}(A) = 1$. Now, for all N, let $A_N = \{\omega \mid \forall n > N \ |X_n(\omega) - X(\omega)| < \varepsilon\}$. Since $A_N \uparrow A$, $\mathbb{P}(A_N) \uparrow \mathbb{P}(A) = 1$. Thus, there exists N' such that $\mathbb{P}(A_{N'}) > 1 - \varepsilon$; then, for all $n > N'$, $\mathbb{P}(|X_n - X| \geq \varepsilon) < 1 - \varepsilon$. □

Proposition 5.8. *For $1 \leq p \leq \infty$, $X_n \to X$ in L^p implies $X_n \to X$ in probability.*

Proof. Fix $\varepsilon > 0$. If $p < \infty$, then by Markov's inequality,

$$\lim_{n\to\infty} \mu(|f_n - f| \geq \varepsilon) = \lim_{n\to\infty} \mu(|f_n - f|^p \geq \varepsilon^p) \leq \lim_{n\to\infty} \frac{1}{\varepsilon^p} \int |f_n - f|^p d\mu = \lim_{n\to\infty} \frac{1}{\varepsilon^p} \|f_n - f\|_{L^p}^p = 0.$$

On the other hand, if $p = \infty$, there exists N such that $\|f_n - f\|_\infty < \varepsilon$ for all $n > N$. But then $\|f_n - f\|_\infty < \varepsilon$ implies $\mu(|f_n - f| \geq \varepsilon) = 0$, so the result also follows in this case. ☐

Corollary 5.8.1. *$X_n \to X$ in L^1 implies $X_n \to X$ in probability.*

Proposition 5.9. *$X_n \to X$ in probability implies $X_n \to X$ in distribution.*

Proof. Let t be a continuity point of F_X. Fix $\varepsilon > 0$. Then

$$F_{X_n}(t) = \mathbb{P}(X_n \leq t) \leq \mathbb{P}(X \leq t + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon).$$

Then, $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$, so $\limsup F_{X_n}(t) \leq F_X(t+\varepsilon)$. Then, since F_X is continuous at t, this implies that $\limsup F_{X_n}(t) \leq F_X(t)$. A similar argument shows that $\liminf F_{X_n}(t) \geq F_X(t)$, so $\lim F_{X_n}(t) = F_X(t)$. The result follows. ☐

Proposition 5.10. *$X_n \to X$ in L^p implies $X_n \to X$ in L^r whenever $p > r$.*

Proof. First, notice that $f(x) = x^{p/r}$ is convex. Thus, by Jensen's inequality, $\mathbb{E}[|X - X_n|^r]^{p/r} \leq \mathbb{E}[|X - X_n|^p] \to 0$ as $n \to \infty$, so $\mathbb{E}[|X - X_n|^r]^{p/r} \to 0$ as $n \to \infty$ whence $\mathbb{E}[|X - X_n|^r] \to 0$ as $n \to \infty$. ☐

Corollary 5.10.1. *For any $p \geq 1$, $X_n \to X$ in L^p implies $X_n \to X$ in L^1.*

5.1.2 Necessary and Sufficient Conditions

Definition 5.11 (Uniformly Integrable)**.** A sequence of random variables $\{X_n\}_{n\geq 1}$ is *uniformly integrable* if for any $\varepsilon > 0$, there is some $K > 0$ such that for all n,

$$\int_{|X_n|>K} |X_n| d\mu \leq \varepsilon.$$

Proposition 5.12 (Alternate Definition of Uniform Integral)**.** *A sequence of random variables $\{X_n\}_{n\geq 1}$ is uniformly integrable if and only if $\sup_n \mathbb{E}|X_n| < \infty$ and, for all $\varepsilon > 0$, there exists $\delta > 0$ such that $\mu(F) < \delta$ implies $\int_F |X_n| d\mu < \infty$ for all n.*

Proof. Suppose $\{X_n\}$ is uniformly integrable. Then, there exists K such that $\mathbb{E}(|X_n| \mid |X_n| > K) \leq 1$. Then, for all n, $\mathbb{E}|X_n| = \int_{|X_n|\leq K} |X_n| d\mu + \int_{|X_n|>K} |X_n| d\mu = K + 1$, so $\sup_n \mathbb{E}|X_n| < \infty$. Then, fix $\varepsilon > 0$. By definition, there exists K such that $\int_{|X_n|>K} |X_n| d\mu < \frac{\varepsilon}{2}$. Then, let $\delta = \frac{\varepsilon}{2K}$. If $\mu(F) < \delta$, then for any n,

$$\int_F |X_n| d\mu = \int_{F\cap\{|X_n|\leq K\}} |X_n| d\mu + \int_{F\cap\{|X_n|>K\}} |X_n| d\mu \leq \frac{K\varepsilon}{2K} + \frac{\varepsilon}{2} = \varepsilon.$$

Suppose $\sup_n \int |f_n| d\mu < \infty$ and, for all ε, there exists δ such that $\mu(F) < \delta$ implies $\int_F |X_n| < \varepsilon$ for all n. Then, fix X_n is uniformly integrable. Then, for all $K > \frac{\sup_n \int |f_n| d\mu}{\delta}$, Markov's inequality implies that $\mu\{|f_n| > K\} \leq K^{-1} \int |f_n| d\mu \leq K^{-1} \sup_n \int |f_n| d\mu < \delta$, so that $\int_{|f_n|>\alpha} |f_n| d\mu < \varepsilon$, as desired. ☐

Lemma 5.13 (Absolute Continuity of the Integral)**.** *Let f be an L^1 function. Then, for any $\varepsilon > 0$, there exists $\delta > 0$ such that $\mu(A) < \delta$ implies $\int_A |f| d\mu < \varepsilon$.*

Proof. First, notice that we may replace f by $|f|$; that is, it suffices to show the result for f nonnegative. Suppose, for the sake of contradiction, there exists $\varepsilon > 0$ and a sequence of sets A_n so that $\mu(A_n) < 2^{-n}$ but $\int_{A_n} f d\mu \geq \varepsilon$. Consider $g_n(x) = f(x)\chi_{A_n}(x)$. Then $g_n(x) \to 0$ as $n \to \infty$ except for points x which lie in infinite many A_ns. But the collection of such points has measure 0, so $g_n(x) \to 0$ almost everywhere. Then, set $f_n = f - g_n$, so $f_n \geq 0$ and $f_n \to f$ almost everywhere. Then Fatou's Lemma yields the contradiction

$$\int_E f d\mu \leq \liminf \int_E f_n d\mu \leq \int_E f d\mu - \limsup \int_E g_n d\mu \leq \int_E f d\mu - \limsup \int_{A_n} f_n d\mu \leq \int_E f d\mu - \varepsilon.$$

□

Corollary 5.13.1. *Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables which is dominated by an L^1 random variable X. Then $\{X_n\}$ is uniformly integrable.*

Lemma 5.14. *Suppose that $1 < p < \infty$. Then, if $\sup_n \|X_n\|_{L^p}$ is finite, $\{X_n\}_{n\geq 1}$ is uniformly integrable.*

Proof. Fix R. Then, $\chi_{\{|f|>R\}}|f(x)|R^{p-1} \leq |f(x)|^p$. Then, integrating,

$$\int_{|f_n|>R} |f_n| d\mu \leq R^{1-p} \sup_n \int |f_n|^p d\mu$$

which tends to 0 as $R \to \infty$. The result follows. □

Proposition 5.15 (Vitali Covergence Theorem)**.** *Suppose $\{X_n\}_{n\geq 1}$ is a sequence of L^p random variables and X is a random variable. For any $1 \leq p < \infty$, $X_n \to X$ converges in probability and $|X_n|^p$ is uniformly integrable if and only if $X_n \to X$ in L^p.*

Proof. Assume that $X_n \to X$ in L^p. Then, Proposition 5.8 implies that $X_n \to X$ in probability. Similarly, fix $\varepsilon > 0$. Then, select N such that $\int |X_n - X_N|^p d\mu < \frac{\varepsilon}{2}$ when $n \geq N$. Now, there exists $\delta > 0$ such that $\mu(E) < \delta$ implies $\int_E |X_n|^p d\mu < \frac{\varepsilon}{2}$ for $n \leq N$. On the other hand, for $n > N$, if $\mu(E) < \delta$, $\int_E |X_n|^p d\mu \leq \int_E |X_n - X_N|^p d\mu + \int_E |X_N|^p d\mu < \varepsilon$. Thus $\{X_n^p\}$ is uniformly integrable.

Assume $X_n \to X$ in probability and $|X_n|^p$ is uniformly integrable. Fix $\varepsilon > 0$. Then let $E_n = \{|X_n - X| \geq \left(\frac{\varepsilon}{3}\right)^{1/p}\}$. Choose $\delta > 0$ such that $\int_E X_n^p d\mu < \frac{\varepsilon}{3}$ and $\int_E X^p d\mu < \frac{\varepsilon}{3}$ whenever $\mu(E) < \delta$. Then, take N such that if $n > N$ then $\mu(E_n) < \delta$. It follows that for $n > N$, $\int_{E_n} |X_n - X|^p d\mu < \frac{2\varepsilon}{3}$. On the other hand, $\int_{E_n^c} |X_n - X|^p d\mu < \frac{\varepsilon}{3}$. Thus, $\int |X_n - X|^p d\mu < \varepsilon$, as desired. □

Corollary 5.15.1. *Suppose that $\{X_n\}$ is a sequence of L^1 random variables and X is a random variable. Then $X_n \to X$ converges in probability and $|X_n|$ is uniformly integrable if and only if $X_n \to X$ in L^1.*

Corollary 5.15.2. *Suppose $X_n \to X$ in L^1. Then $X_n \to X$ in L^p if and only if $\{X_n^p\}$ is uniformly integrable.*

Proof. Trivial. □

Proposition 5.16. *For any $1 \leq p < \infty$, if $X_n \to X$ a.e. and $\|X_n\|_{L^p} \to \|X\|_{L^p}$, $X_n \to X$ in L^p.*

Proof. Let $Y_n = |X|^p + |X_n|^p - |X - X_n|^p$. Then Y_n is non-negative for each n and $Y_n \to 2|X|^p$ pointwise almost everywhere. Thus, by the almost-everywhere version of Fatou's Lemma,

$$\begin{aligned}\int_E 2|X|^p d\mu \leq \liminf_{n\to\infty} \int_E (|X|^p + |X_n|^p - |X - X_n|^p) d\mu &= \int_E |X|^p + \liminf_{n\to\infty} \int_E |X_n|^p + \liminf_{n\to\infty} \int_E (-|X - X_n|^p) d\mu \\ &= \int_E |X|^p + \liminf_{n\to\infty} \int_E |X_n|^p - \limsup_{n\to\infty} \int_E |X - X_n|^p d\mu \\ &= \int_E 2|X|^p - \limsup_{n\to\infty} \int_E |X - X_n|^p d\mu\end{aligned}$$

where the final equality follows from the assumption $\lim_{n\to\infty} \int_E |X_n|^p \to \int_E |X|^p$. Now, if we rearrange the inequality given by the above calculation, we obtain $\limsup_{n\to\infty} \int_E |X - X_n|^p d\mu \leq 0$. Of course, $\liminf_{n\to\infty} \int_E |X - X_n|^p d\mu \geq 0$, so indeed $\lim_{n\to\infty} \int_E |X - X_n|^p d\mu = 0$ and $X_n \to X$ in L^p. □

Corollary 5.16.1. *If $X_n \to X$ a.e. and $\mathbb{E}[X_n] \to \mathbb{E}[X]$, then $X_n \to X$ in L^1.*

5.1.3 Sufficient Conditions

Proposition 5.17. *$X_n \xrightarrow{p} X$ implies that there exists a subsequence $\{n_k\}$ such that $X_{n_k} \to X$ a.e.*

Proof. By convergence in probability, there exists a subsequence $\{X_{n_k}\}_{k\geq 1}$ such that for all k, $\mathbb{P}(|X_{n_k} - X_{n_{k+1}}| > 2^{-k}) \leq 2^{-k}$. Thus, by the Borel-Cantelli Lemma, $\mathbb{P}(|X_{n_k} - X_{n_{k+1}}| > 2^{-k} \text{ i.o.}) = 0$. Thus, $\{X_{n_k}(\omega)\}_{k\geq 1}$ is a Cauchy sequence with probability 1. Then, define $Y(\omega)$ to be $\lim_k X_{n_k}(\omega)$ if $X_{n_k}(\omega)$ is a Cauchy sequence and 0 otherwise. Then $X_{n_k} \to Y$ a.e. But then, $X_{n_k} \to Y$ in probability by Proposition 5.7. But then, by Proposition 5.23, $X = Y$ a.e., so that $X_{n_k} \to X$ a.e. □

Corollary 5.17.1. *Suppose that $\{X_n\}_{n\geq 1}$ is a non-decreasing sequence which converges to X in probability. Then $X_n \to X$ a.e.*

Proof. By the above proposition, there is a subsequence $(X_{n_k})_{k\geq 1}$ converging to X a.e. But then (X_n) converges to X a.e. by monotonicity. □

Proposition 5.18. *$X_n \xrightarrow{L^1} X$ implies that $\mathbb{E}[X_n] \to \mathbb{E}[X]$.*

Proof. This is immediate:

$$0 = \lim_{n\to\infty} \int |X_n - X| d\mu \geq \lim_{n\to\infty} \left| \int (X_n - X) d\mu \right| = \lim_{n\to\infty} \left| \int X_n d\mu - \int X d\mu \right| = \lim_{n\to\infty} |\mathbb{E}[X_n] - \mathbb{E}[X]|.$$

□

Proposition 5.19. *Suppose $\{X_n\}$ is a sequence of discrete and independent random variables. Then $X_n \to X$ in probability implies that $X_n \to X$ a.e.*

Proof. This follows from the Second Borel-Cantelli Lemma. □

Lemma 5.20 (Ottoviani's Inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables. Let $S_{k,n} = \sum_{i=k+1}^{n} X_i$ and $S_n = S_{0,n}$. Then, for all $\varepsilon > 0$,*

$$\min_{1\leq k\leq n} \mathbb{P}(|S_{k,n}| \leq \varepsilon)\mathbb{P}(\max_{1\leq i\leq n} |S_i| > 2\varepsilon) \leq \mathbb{P}(|S_n| > \varepsilon).$$

Proof. Let A_k be the event that $|S_k|$ is the first $|S_j|$ strictly greater than 2ε. Then the event $\max_{1\leq i\leq n} |S_i| > 2\varepsilon$ is the disjoint union $\bigcup_{i=1}^{n} A_i$. Then, since $|S_{k,n}|$ is independent of $|S_1|, \ldots, |S_k|$,

$$\mathbb{P}(A_k) \min_{1\leq j\leq n} \mathbb{P}(|S_{j,n}| \leq \varepsilon) \leq \mathbb{P}(A_k \text{ and } |S_{k,n}| \leq \varepsilon) = \mathbb{P}(A_k \text{ and } |S_{k,n}| \leq \varepsilon) \leq \mathbb{P}(A_k \text{ and } |S_n| > \varepsilon)$$

where the final step is because A_k and $|S_{k,n}| \leq \varepsilon$ implies $|S_n| > \varepsilon$. Then, sum over k to conclude. □

Proposition 5.21. *Suppose that $\{X_n\}$ is a sequence of independent random variables. Then, if $\sum_n X_n$ converges in probability (i.e., if for any $\varepsilon > 0$, there exists N such that $\mathbb{P}(|S_{m,n}| > \varepsilon) \leq \varepsilon$ when $n > m > N$), $\sum_n X_n$ converges almost surely.*

Proof. First, notice that S_n doesn't converge if and only if $I = \inf_{m\in\mathbb{Z}^+} \sup_{k\in\mathbb{Z}^+} |S_{m,m+k}|$ doesn't equal 0. Thus, it suffices to show that $I = 0$ with probability 1. Let $\varepsilon > 0$. Then, by Ottoviani,

$$\min_{1\leq k\leq j} \mathbb{P}\left(|S_{(m+k,m+j)}| \leq \frac{\varepsilon}{2}\right) \mathbb{P}\left(\max_{1\leq k\leq j} |S_{m,m+k}| > \varepsilon\right) \leq \mathbb{P}\left(|S_{m,j+m}| > \frac{\varepsilon}{2}\right).$$

For any $\delta > 0$, by convergence in probability, there exists N_δ such that $\mathbb{P}(|S_{N_\delta+k,N_\delta+j}| > \frac{\varepsilon}{2}) \leq \delta$ for $0 \leq k \leq j$. Then $\mathbb{P}\left(|S_{(N_\delta+k,N_\delta+j)}|\right) \geq 1 - \delta$ and $\mathbb{P}\left(|S_{N_\delta,N_\delta+j}| > \frac{\varepsilon}{2}\right) \leq \delta$, so $\mathbb{P}\left(\max_{1\leq k\leq j} |S_{N_\delta,N_\delta+k}| > \varepsilon\right) \leq \frac{\delta}{1-\delta}$. Then,

$$\mathbb{P}\left(\inf_{m\in\mathbb{Z}^+} \sup_{k\in\mathbb{Z}^+} |S_{m,m+k}| > \varepsilon\right) \leq \mathbb{P}\left(\max_{1\leq k\leq j} |S_{N_\delta,N_\delta+k}| > \varepsilon\right) \leq \frac{\delta}{1-\delta}$$

Then, since δ is arbitrary, by driving $\delta \to 0$ we find that $\mathbb{P}(\inf_{m\in\mathbb{Z}^+} \sup_{k\in\mathbb{Z}^+} |S_{m,m+k}| > \varepsilon) = 0$ for any $\varepsilon > 0$. Therefore, taking $\varepsilon \to 0$, $\mathbb{P}(\inf_{m\in\mathbb{Z}^+} \sup_{k\in\mathbb{Z}^+} |S_{m,m+k}| > 0) = 0$ and the result follows. □

Proposition 5.22. *Suppose that $X_n \to X$ a.e. Also suppose that there exists an L^p random variable Y such that $X_n \leq Y$ for all n for some $1 \leq p \leq \infty$. Then $X_n \to X$ in L^p.*

Proof. This follows immediately from the dominated convergence theorem. □

Corollary 5.22.1. *Suppose that $X_n \to X$ a.e. Also suppose that there exists an L^1 random variable Y such that $X_n \leq Y$ for all n. Then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.*

5.1.4 Additional Notes

Proposition 5.23. *The following hold:*

1. *Suppose that $X_n \to X$ a.e. and $X_n \to Y$ a.e. Then $X = Y$ a.e.*
2. *Suppose that $X_n \to X$ in probability and $X_n \to Y$ in probability. Then $X = Y$ a.e.*
3. *Suppose that $X_n \to X$ in L^p and $X_n \to Y$ in L^p. Then $X = Y$ a.e.*
4. *Suppose that $X_n \to X$ in L^1 and $X_n \to Y$ in L^1. Then $X = Y$ a.e.*

Proof. (1) is trivial, and (4) follows from (3). Furthermore, (3) follows from the fact that $X_n \to X$ and $X_n \to Y$ in L^p implies that $\int |X-Y|^p d\mu = 0$ whence $|X-Y|^p = 0$ a.e. whence $X = Y$ a.e. Finally, (2) follows from the fact that there exists a subsequence $\{X_{n_k}\}$ of $\{X_n\}$ converging to X a.e. and converging to Y in probability, and then a subsequence $\{X_{n_{k_m}}\}$ of $\{X_{n_k}\}$ converging to X and Y a.e., so $X = Y$ a.e. □

Finally, the statements about convergence in distribution shall be proven in the following subsections.

5.2 The Weak Law of Large Numbers

Theorem 5.24 (Quantitative Weak Law of Large Numbers)**.** *If $X_1, X_2, \ldots, X_n$ be L^2 random variables defined on the same probability space. Let $\mu_i = \mathbb{E}[X_i]$ and $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2 n^2}\sum_{i,j=1}^n \sigma_{ij}.$$

Proof. Apply Chebychev's inequality to the variance of a sum of random variables. □

Corollary 5.24.1 (L^2 Weak Law of Large Numbers)**.** *If $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. with common mean μ and uniformly bounded finite second moment, then $\frac{\sum_{i=1}^n X_i}{n}$ converges in probability to μ as $n \to \infty$.*

5.3 The Strong Law of Large Numbers

Theorem 5.25 (Strong Law of Large Numbers)**.** *Let $\{X_n\}_{n\geq 1}$ be a sequence of pairwise independent and identically distributed random variables with $\mathbb{E}[X_1] < \infty$. Then $\frac{1}{n}\sum_{i=1}^n X_i$ tends to $\mathbb{E}[X_1]$ almost surely as $n \to \infty$.*

Proof. First, notice that by splitting into positive and negative parts, we may assume that the X_i are non-negative. Then, define $Y_i = X_i 1_{\{X_i < i\}}$. Then $\sum_{i=1}^\infty \mathbb{P}(X_i \neq Y_i) = \sum_{i=1}^\infty \mathbb{P}(X \leq i) = \sum_{i=1}^\infty \mathbb{P}(X_i \geq i) \leq \mathbb{E}[X_1] < \infty$. Therefore, by the first Borel-Cantelli Lemma, $\mathbb{P}(X_i \neq Y_i \text{ i.o.}) = 0$. Yet if $X_i \neq Y_i$ finitely often, $\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n Y_i \to 0$ as $n \to \infty$. Next, notice that $|\mathbb{E}[Y_i] - \mathbb{E}[X_1]| = |\mathbb{E}[Y_i - X_1]| \leq \mathbb{E}[|Y_i - X_1|] = \mathbb{E}[|Y_i - X_i|] \leq \mathbb{E}[X_i 1_{\{X_i > i\}}] = \mathbb{E}[X_1 1_{\{X_1 \geq i\}}]$. Then, notice that as $i \to \infty$, then $X_1 1_{\{X_i > i\}} \to 0$ by Markov's inequality as $\mathbb{E}[X_1] < \infty$. Thus, by the Dominated Convergence Theorem, $\mathbb{E}[X_1 1_{\{X_i \geq i\}}) \to 0$ as $i \to \infty$. Thus, $\mathbb{E}[Y_i] \to \mathbb{E}[X_1]$ as $i \to \infty$, so that $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i] \to \mathbb{E}[X_1]$ as $n \to \infty$.

Thus, since $\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n Y_i \to 0$ as $n \to \infty$ almost surely, and $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i] \to \mathbb{E}[X_1]$ as $n \to \infty$, it suffices to show that $\frac{1}{n}\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) \to 0$ almost surely. Let $Z_n = \frac{1}{n}\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i])$. For any $n > 1$, let $k_n = [\alpha^n]$. We show that for any $\alpha > 1$, $Z_{k_n} \to 0$ almost surely.

Fix $\varepsilon > 0$. Then, by the Weak Law of Large Numbers, $\mathbb{P}(|Z_{k_n}| > \varepsilon) \leq \frac{1}{\varepsilon^2 k_n^2} \sum_{i=1}^{n} \sum_{i=1}^{k_n} \operatorname{Var}(Y_i)$. Thus, $\sum_{n=1}^{\infty} \mathbb{P}(|Z_{k_n}| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \operatorname{Var}(Y_i) \sum_{n|k_n \geq i} \frac{1}{k_n^2}$. Yet, there exists some β with $k_{n+1}/k_n \geq \beta$ for all sufficiently large n, so that $\sum_{n|k_n \geq i} \frac{1}{k_n^2} \leq \frac{1}{i^2} \sum_{n=0}^{\infty} \beta^{-n} \leq \frac{C}{i^2}$. Then, by increasing C if necessary, this inequality holds for all n.

Thus, by the monotone convergence theorem,

$$\sum_{n=1}^{\infty} \mathbb{P}(|Z_{k_n}| > \varepsilon) \leq \frac{C}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{\operatorname{Var}(Y_i)}{i^2} \leq \frac{C}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{\mathbb{E}[Y_i]^2}{i^2} = \frac{C}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{\mathbb{E}[X_1^2 \mid X_1 < i]}{i^2} \leq \frac{C}{\varepsilon^2} \mathbb{E}\left[X_1^2 \sum_{i > X_1} \frac{1}{i^2}\right] \leq \frac{C'}{\varepsilon^2} \mathbb{E}[X_1] < \infty.$$

Hence by the first Borel-Cantelli Lemma, $\mathbb{P}(|Z_{k_n}|) > \varepsilon \text{ i.o.}) = 0$. Thus, $Z_{k_n} \to 0$ a.s. as $n \to \infty$.

Now, our goal is to show that $Z_n \to 0$ a.s. Let $T_n = Y_1 + \cdots + Y_n$, and take $k_n < m \leq k_{n+1}$. Then,

$$\frac{k_n}{k_{n+1}} \frac{T_{k_n}}{k_n} = \frac{T_{k_n}}{k_{n+1}} \leq \frac{T_m}{m} \leq \frac{T_{k_{n+1}}}{k_n} = \frac{T_{k_{n+1}}}{k_{n+1}} \frac{k_{n+1}}{k_n}.$$

But then, taking $m \to \infty$, since $k_{n+1}/k_n \to \alpha$ and $T_{k_n}/k_n \to \mu$ almost surely, the above imply that

$$\frac{\mu}{\alpha} \leq \liminf_{m \to \infty} \frac{T_m}{m} \leq \limsup_{m \to \infty} \frac{T_m}{m} \leq \alpha\mu$$

for any $\alpha > 1$, which is sufficient to establish the desired result. ☐

5.4 Prerequisites for the Central Limit Theorem

This section prepares us to prove the Central Limit Theorem, which is the following result:

Theorem 5.26 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean 0 and variance 1. Let $S_n = X_1 + \cdots + X_n$. Then $\frac{S_n}{\sqrt{n}}$ converges in distribution to $\mathcal{N}(0, 1)$.*

For, this we need some preliminary material.

Definition 5.27 (Tight Family)**.** Let $\{X_i\}_{i \in I}$ be any collection of random variables. Then, $\{X_i\}_{i \in I}$ is a *tight family* if for all $\varepsilon > 0$, there exists $K > 0$ such that $\mathbb{P}(|X_i| > K) \leq \varepsilon$ for all $i \in I$.

Proposition 5.28. *If $X_n \xrightarrow{d} X$ in distribution, then $\{X_n\}_{n \geq 1}$ is tight.*

Theorem 5.29 (Helly's Selection Theorem)**.** *If $\{X_n\}_{n \geq 1}$ is a tight family, then there is a subsequence $\{X_{n_k}\}_{k \geq 1}$ that converges in distribution.*

Proof. Let F_n be the c.d.f. of X_n. By the standard diagonal argument, there is a subsequence $\{n_k\}_{k \geq 1}$ such that $F_*(q) = \lim_{k \to \infty} F_{n_k}(q)$ for every rational q. Then, for every $x \in \mathbb{R}$, define $F(x) = \inf_{q \in \mathbb{Q}, q > x} F_*(q)$. Then, F is non-decreasing and it can be straightforwardly shown that it satisfies both $\limsup_{k \to \infty} F_{n_k}(x) \leq F(x)$ and $\liminf_{k \to \infty} F_{n_k}(x) \geq F(x)$, so $\lim_{k \to \infty} F_{n_k}(x) = F(x)$ whenever x is a continuity point of F. ☐

Theorem 5.30. *X_n converges to X in distribution if and only if $\mathbb{E}[f(X_n)]$ converges to $\mathbb{E}[f(X)]$ for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$.*

Proof. Suppose that $X_n \xrightarrow{d} X$. Then, take any $f : \mathbb{R} \to \mathbb{R}$ bounded and continuous. Then $\{X_n\}_{n \geq 1}$ is tight, so that there exists $K > 0$ such that for all n, $\mathbb{P}(|X_n| > K) < \varepsilon$ and $\mathbb{P}(|X| > K) < \varepsilon$. Since f is bounded, there exists $M > 0$ such that $|f(x)| \leq M$ for all x. Since f is continuous, it is uniformly continuous in $[-K, K]$. Thus, there exists some $\delta > 0$ such that $x, y \in [-K, K]$ with $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$. Now, we may choose $-K = x_1 \leq x_2 \leq \cdots \leq x_m = K$ such that each x_i is a continuity point of F_X and $x_{i+1} - x_i \leq \delta$ for each i.

Let $g(x) = 0$ if $|x| > K$ and $f(x_i)$ if $x \in (x_{i-1}, x_i]$. Then $|\mathbb{E}f(X_n) - \mathbb{E}g(X_n)| = |\mathbb{E}[f(X_n) - g(X_n)]| \leq \mathbb{E}|f(X_n) - g(X_n)|$. Now, if $X_n \in (-K, K]$, this quantity is at most ε, and if $X_n \notin (-K, K)$, this is bounded above by M. Thus, $\mathbb{E}|f(X_n) - g(X_n)| \leq M\mathbb{P}(X_n \notin (-K, K]) + \varepsilon\mathbb{P}(X_n \in (-K, K]) = M\varepsilon + \varepsilon = (M+1)\varepsilon$. By the same argument, $|\mathbb{E}f(X) - \mathbb{E}g(X)| \leq (M+1)\varepsilon$. Yet, $\mathbb{E}g(X_n) = \sum_{i=1}^m f(y_i)\mathbb{P}(y_i < X_n \leq y_{i+1}) = \sum_{i=1}^m f(y_i)(F_n(y_i) - F_n(y_{i-1})) \to \sum_{i=1}^m f(y_i)(F(y_i) - F(y_{i-1})) = \mathbb{E}[g(x)]$. Thus, $\limsup_{n\to\infty} |\mathbb{E}f(X_n) - \mathbb{E}f(X)| \leq 2(M+1)\varepsilon$, and since this holds for any ε, we obtain $\limsup_{n\to\infty} |\mathbb{E}f(X_n) - \mathbb{E}f(X)| = 0$ and the result follows.

Then, suppose that $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for any bounded continuous function f, and take a continuous point t of F_X. Then take $\varepsilon > 0$. Let f_ε be the function that is 1 below t, 0 above $t+\varepsilon$, and goes down linearly from 1 to 0 in the interval $[t, t+\varepsilon]$. Then $\limsup_{n\to\infty} F_{X_n}(t) \leq \limsup_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}f(X) \leq F_X(t+\varepsilon)$. Since F_X is right-continuous, taking $\varepsilon \to 0$ yields $\limsup_{n\to\infty} F_{X_n}(t) \leq F_X(t)$. Similarly, $\liminf_{n\to\infty} F_{X_n}(t) \geq F_X(t)$. Thus, $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$ whenever t is a continuity point of F_X, so that $X_n \to X$ in distribution. ☐

Corollary 5.30.1. *Two random variables X and Y have the same law if and only if $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$ for all bounded continuous f.*

Corollary 5.30.2. *If $\{X_n\}_{n=1}^\infty$ is a sequence of random variables converging in distribution to a random variable X. Then for any continuous $f : \mathbb{R} \to \mathbb{R}$, $f(X_n) \xrightarrow{d} f(X)$.*

Theorem 5.31 (Slutsky's Theorem). *If $X_n \to c \in \mathbb{R}$ in probability and $Y_n \to Y$ in distribution, show that $X_nY_n \to cY$ and $X_n + Y_n \to c + Y$ in distribution.*

Proof. Let F be the c.d.f. of $Y + c$ and t be a continuity point of F. Fix $\varepsilon > 0$. Then, if $X_n + Y_n \leq t$, either $Y_n + c \leq t + \varepsilon$ or $X_n - c < -\varepsilon$. Thus, the union bound yields $\mathbb{P}(X_n + Y_n \leq t) \leq \mathbb{P}(Y_n + c \leq t+\varepsilon) + \mathbb{P}(X_n - c < -\varepsilon)$. Then, if $t+\varepsilon$ is also a continuity point of F, $\limsup_{n\to\infty} \mathbb{P}(Y_n + c \leq t+\varepsilon) = F(t+\varepsilon)$, and $\limsup_{n\to\infty} \mathbb{P}(X_n - c < -\varepsilon) = 0$. Thus, $\limsup_{n\to\infty} \mathbb{P}(X_n + Y_n \leq t) \leq F(t+\varepsilon)$.

Similarly, if $Y_n + c \leq t - \varepsilon$, either $X_n + Y_n \leq t$ or $X_n - c > \varepsilon$. Thus, the union bound yields $\mathbb{P}(Y_n + c \leq t-\varepsilon) \leq \mathbb{P}(X_n + Y_n \leq t) + \mathbb{P}(X_n - c > \varepsilon)$. Then, if $t-\varepsilon$ is a continuity point of F, $\liminf_{n\to\infty} \mathbb{P}(Y_n + c \leq t-\varepsilon) = F(t-\varepsilon)$, and $\liminf_{n\to\infty} \mathbb{P}(X_n - c > \varepsilon) = 0$. Thus, $\liminf_{n\to\infty} \mathbb{P}(X_n + Y_n \leq t) \geq F(t-\varepsilon)$.

Now, since F has only countably many points of discontinuity, there exists a sequence $\{\varepsilon_j\} \to 0$ such that $t + \varepsilon_i$ and $t - \varepsilon_i$ are continuity points of F for each i. Furthermore, since F is continuous at t,

$$F(t) = \lim_{j\to\infty} F(t+\varepsilon_j) \geq \limsup_{n\to\infty} \mathbb{P}(X_n + Y_n \leq t) \geq \liminf_{n\to\infty} \mathbb{P}(X_n + Y_n \leq t) \geq \lim_{j\to\infty} F(t-\varepsilon_j) = F(t)$$

Thus, $\lim_{n\to\infty} \mathbb{P}(X_n + Y_n \leq t) = F(t)$, and the result follows.

Suppose $c > 0$. Let F be the c.d.f of cY. Fix $\varepsilon > 1$. Then, if $X_nY_n \leq t$, either $cY_n \leq \varepsilon t$ or $\frac{X_n}{c} < \frac{1}{\varepsilon}$. Thus, $\mathbb{P}(X_nY_n \leq t) \leq \mathbb{P}(cY_n \leq \varepsilon t) + \mathbb{P}(\frac{X_n}{c} < \frac{1}{\varepsilon})$. Then, if $t\varepsilon$ is a continuity point of F, $\limsup_{n\to\infty} \mathbb{P}(cY_n \leq t\varepsilon) = F(t\varepsilon)$, and $\limsup_{n\to\infty} \mathbb{P}(\frac{X_n}{c} < \frac{1}{\varepsilon}) = 0$. Thus, $\limsup_{n\to\infty} \mathbb{P}(X_nY_n \leq t) \leq F(t+\varepsilon)$. Similarly, if $cY_n \leq \frac{t}{\varepsilon}$, either $X_nY_n \leq t$ or $\frac{X_n}{c} > \varepsilon$. Thus, $\mathbb{P}(cY_n \leq \frac{t}{\varepsilon}) \leq \mathbb{P}(X_nY_n \leq t) + \mathbb{P}(\frac{X_n}{c} > \varepsilon)$. Then, if $\frac{t}{\varepsilon}$ is a continuity point of F, $\liminf_{n\to\infty} \mathbb{P}(cY_n \leq \frac{t}{\varepsilon}) = F(\frac{t}{\varepsilon})$, and $\liminf_{n\to\infty} \mathbb{P}(\frac{X_n}{c} > \varepsilon) = 0$. Thus, $\liminf_{n\to\infty} \mathbb{P}(X_nY_n \leq t) \geq F(\frac{t}{\varepsilon})$.

Now, since F has only countably many points of discontinuity, there exists a sequence $\{\varepsilon_j\} \to 1$ such that $t\varepsilon_i$ and $\frac{t}{\varepsilon_i}$ are continuity points of F for each i. Furthermore, since F is continuous at t,

$$F(t) = \lim_{j\to\infty} F(t\varepsilon_j) \geq \limsup_{n\to\infty} \mathbb{P}(X_nY_n \leq t) \geq \liminf_{n\to\infty} \mathbb{P}(X_nY_n \leq t) \geq \lim_{j\to\infty} F\left(\frac{t}{\varepsilon_j}\right) = F(t)$$

Thus, $\lim_{n\to\infty} \mathbb{P}(X_nY_n \leq t) = F(t)$, and the result follows. Similarly, for the case $c < 0$, notice that $-X_n \to -c$ in probability, so that $-(X_nY_n) \to -cY$ in distribution by the above work, which immediately implies that $X_nY_n \to cY$, as desired.

All that remains is to show that if $X_n \to 0$ in probability and $Y_n \to Y$ in distribution, then $X_nY_n \to 0$ in distribution. Fix $t > 0$. Then, $|X_nY_n| > t$ implies that either $|Y_n| > \frac{t}{\varepsilon}$ or $|X_n| > \varepsilon$. Thus, $\mathbb{P}(|X_nY_n| > t) \leq \mathbb{P}(|Y_n| > \frac{t}{\varepsilon}) + \mathbb{P}(|X_n| > \varepsilon)$. Now, if $\frac{t}{\varepsilon}$ and $-\frac{t}{\varepsilon}$ are both continuity points of the c.d.f. of Y, $\lim_{n\to\infty} \mathbb{P}(|Y_n| > \frac{t}{\varepsilon}) = \mathbb{P}(|Y| > \frac{t}{\varepsilon})$. Yet as $\varepsilon \to 0$, $\mathbb{P}(|Y| > \frac{t}{\varepsilon}) \to 0$. Since F has only countably many points of discontinuity, there exists a decreasing sequence $\{\varepsilon_j\} \to 0$ such that $\frac{t}{\varepsilon_i}$ and $t\varepsilon_i$ are continuity points of F for each i. Thus,

$$0 \leq \lim_{n\to\infty} \mathbb{P}(|X_nY_n| > t) \leq \lim_{j\to\infty} \lim_{n\to\infty} \left(\mathbb{P}\left(|Y_n| > \frac{t}{\varepsilon_j}\right) + \mathbb{P}(|X_n| > \varepsilon_j) \right) \leq \lim_{j\to\infty} \mathbb{P}\left(|Y| > \frac{t}{\varepsilon_j}\right) = 0.$$

Thus, $\mathbb{P}(|X_nY_n| > t) = 0$ for any $t > 0$. Therefore, $\mathbb{P}(|X_nY_n| \leq t) = \mathbb{P}(X_nY_n \leq t) - \mathbb{P}(X_nY_n < -t) = 1$, whence $\mathbb{P}(X_nY_n \leq t) = 1$ and $\mathbb{P}(X_nY_n < -t)$ for any $t > 0$. That is, $\mathbb{P}(X_nY_n \leq t)$ is equal to 0 if $t < 0$ and 1 if $t > 0$, and therefore X_nY_n converges in distribution to the random variable 0. □

Theorem 5.32. *Let X be a random variable with characteristic function ϕ. Then, for each $\theta > 0$, define $f_\theta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx-\theta t^2}\phi(t)dt$. Then, for any bounded continuous $g : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[g(X)) = \lim_{\theta\to 0}\int_{-\infty}^{\infty} g(x)f_\theta(x)dx$.*

Proof. Let μ be the law of X, so $\phi(t) = \int_{-\infty}^{\infty} e^{ity}d\mu(y)$. Then, applying Fubini's theorem, $f_\theta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(y-x)t-\theta t^2}dtd\mu(y)$. Yet $\int_{-\infty}^{\infty} e^{i(y-x)t-\theta t^2}dt = \sqrt{\frac{\pi}{\theta}}\int_{-\infty}^{\infty} e^{i(2\theta)^{-1/2}(y-x)s}\frac{e^{-s^2/2}}{\sqrt{2\pi}}ds = \sqrt{\frac{\pi}{\theta}}e^{-(y-x)^2/4\theta}$. Thus $f_\theta(x) = \int_{-\infty}^{\infty}\frac{e^{-(y-x)^2/4\theta}}{\sqrt{4\pi\theta}}d\mu(y)$. Then $f_\theta(x)$ is the p.d.f. of $X + Z_\theta$, where $Z_\theta = N(0, 2\theta)$, so that $\int_{-\infty}^{\infty} g(x)f_\theta(x)dx = \mathbb{E}[g(X + Z_\theta)]$. But $Z_\theta \to 0$ in probability as θ to 0, so $X + Z_\theta \to X$ in distribution by Slutksy's theorem, and $\mathbb{E}[g(X + Z_\theta)] \to \mathbb{E}[g(X)]$ as $\theta \to 0$ by the previous theorem. □

Corollary 5.32.1. *Two random variables X and Y have the same law if and only if they have the same characteristic function.*

Corollary 5.32.2. *Let X be a random variable with characteristic function ϕ. Suppose that*

$$\int_{-\infty}^{\infty} |\phi(t)|dt < \infty.$$

Then X has a probability density function f given by $f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)dt$.

Proof. Now, recall that f_θ is the p.d.f. of $X + Z_\theta$, where $Z_\theta \sim N(0, 2\theta)$. Then, if ϕ is integrable, the dominated convergence theorem shows that $f(x) = \lim_{\theta\to 0} f_\theta(x)$. Furthermore, by integrability of ϕ, $|f_\theta(x)| \leq \frac{1}{2\pi}\int_{-\infty}^{\infty}|\phi(t)|dt$. Thus, by the dominated convergence theorem, for $-\infty < a \leq b < \infty$, $\int_a^b f(x)dx = \lim_{\theta\to 0}\int_a^b f_\theta(x)dx$. Therefore, by Slutsky's Theorem, if a and b are continuity points of the c.d.f. of X, $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$. □

Theorem 5.33. *Let X be an integer-valued random variable with characteristic function ϕ. Then for any $x \in \mathbb{Z}$, $\mathbb{P}(X = x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itx}\phi(t)dt$.*

Proof. If μ is the law of X, then by Fubini's Theorem,

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itx}dt &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\infty}^{\infty} e^{-itx}e^{ity}d\mu(y)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\pi}^{\pi} e^{it(y-x)}dtd\mu(y) \\ &= \sum_{y\in\mathbb{Z}}\mathbb{P}(X = y)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{it(y-x)}dt\right) = \mathbb{P}(X = x).\end{aligned}$$

□

Theorem 5.34 (Levy's Continuity Theorem)**.** *A sequence of random variables $\{X_n\}_{n\geq 1}$ converges in distribution to a random variable X if and only if the sequence of characteristic functions $\{\phi_{X_n}\}_{n\geq 1}$ converges to the characteristic function ϕ_X pointwise.*

Proof. One direction follows immediately from our work earlier in this section. For the other, suppose $\phi_{X_n}(t) \to \phi_X(t)$ for every t. Take any $\varepsilon > 0$. Then, there exists a such that $|\phi_X(s) - 1| \leq \varepsilon/2$ whenever $|s| \leq a$. Thus, $\frac{1}{a}\int_{-a}^{a}(1 - \phi_X(s))ds \leq \varepsilon$. Thus, by the dominated convergence theorem, $\lim_{n\to\infty} \frac{1}{a}\int_{-a}^{a}(1 - \phi_{X_n}(s))ds = \frac{1}{a}\int_{-a}^{a}(1 - \phi_X(s))ds \leq \varepsilon$. Let $t = 2/a$. Then, $\limsup_{n\to\infty} \mathbb{P}(|X_n| \geq t) \leq \varepsilon$, so that $\mathbb{P}(|X_n| \geq t) \leq 2\varepsilon$ for all large enough n. Then, increasing t to T as necessary, we may assume that there exists T such that $\mathbb{P}(|X_n| \geq T) \leq 2\varepsilon$, so that $\{X_n\}$ is tight.

Then, suppose that $X_n \overset{d}{\not\to} X$; then there exists a bounded continuous function f with $\mathbb{E}[f(X_n)] \not\to \mathbb{E}[f(X)]$. But then, passing to a subsequence if necessary, we find that there exists $\varepsilon > 0$ such that $|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \geq \varepsilon$ for all n. Then, by tightness, there is a subsequence $\{X_{n_k}\}$ that converges in distribution to a limit Y. But then $\mathbb{E}f(X_{n_k}) \to \mathbb{E}f(Y)$ whence $|\mathbb{E}f(Y) - \mathbb{E}f(X)| \geq \varepsilon$. But by the first direction and the fact that $\phi_{X_n} \to \phi_X$ pointwise, $\phi_Y = \phi_X$. But then Y and X have the same law, yielding a contradiction with the above work. ☐

5.5 The Central Limit Theorem

In this section, we will be proving various forms of the Central Limit Theorem. We begin with the classical Central Limit Theorem, which is for i.i.d. random sums.

Theorem 5.35. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean μ and variance σ^2.Then, the random variable*

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

converges weakly to the standard Gaussian distribution as $n \to \infty$.

Proof. First, we need the following two lemmas:

Lemma 5.36. *For any $x \in \mathbb{R}$,*

$$\left| e^{ix} - 1 - ix + \frac{x^2}{2} \right| \leq \min\left\{ x^2, \frac{|x|^3}{6} \right\}.$$

Proof. Now, by Taylor expansion, $\left| e^{ix} - \sum_{j=0}^k \frac{(ix)^j}{j!} \right| \leq \frac{|x|^{k+1}}{(k+1)!}$. Thus, $\left| e^{ix} - 1 - ix + \frac{x^2}{2} \right| \leq \frac{|x|^3}{6}$. Yet also $\left| e^{ix} - 1 - ix + \frac{x^2}{2} \right| \leq \left| e^{ix} - 1 - ix \right| + \frac{x^2}{2} \leq \frac{x^2}{2} + \frac{x^2}{2} = x^2$. The result follows. ☐

Lemma 5.37. *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be complex numbers such that $|a_i| \leq 1$ and $|b_i| \leq 1$. Then, $|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i| \leq \sum_{i=1}^n |a_i - b_i|$.*

Proof.

$$\begin{aligned}
\left| \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \right| &= \left| \sum_{i=1}^n a_1 \cdots a_{i-1} b_i \cdots b_n - a_1 \cdots a_i b_{i+1} \cdots b_n \right| \leq \sum_{i=1}^n |a_1 \cdots a_{i-1} b_i \cdots b_n - a_1 \cdots a_i b_{i+1} \cdots b_n| \\
&= \sum_{i=1}^n |a_1 \cdots a_{i-1}(b_i - a_i) b_{i+1} \cdots b_n| \leq \sum_{i=1}^n |b_i - a_i|.
\end{aligned}$$

☐

First, notice that by replacing X_i by $(X_i - \mu)/\sigma$, we may assume that $\mu = 0$ and $\sigma = 1$. Then, let $S_n = \sqrt{n}\sum_{i=1}^n X_i$. Then, take any $t \in \mathbb{R}$; it suffices, by Levy's continuity theorem, to show that $\phi_{S_n}(t) \to e^{-t^2/2}$ as $n \to \infty$. Yet because the X_i are i.i.d., $\phi_{S_n}(t) = \prod_{i=1}^n \phi_{X_i}(t/\sqrt{n}) = (\phi_{X_1}(t/\sqrt{n}))^n$. Thus, by Lemma 8.10.4, when n is large enough that $t^2 \leq 2n$,

$$\left| \phi_{S_n}(t) - \left(1 - \frac{t^2}{2n}\right)^n \right| \leq n \left| \phi_{X_1}(t/\sqrt{n}) - \left(1 - \frac{t^2}{2n}\right) \right|.$$

Now, it suffices to show that the right-hand side tends to zero as $n \to \infty$. Yet

$$n \left| \phi_{X_1}(t/\sqrt{n}) - \left(1 - \frac{t^2}{2n}\right) \right| = n \left| \mathbb{E}\left(e^{itX_1/\sqrt{n}} - 1 - \frac{itX_1}{\sqrt{n}} + \frac{t^2 X_1^2}{2n}\right) \right| \leq \mathbb{E} \min\left\{t^2 X_1^2, \frac{|t|^3 |X_1|^3}{6\sqrt{n}}\right\} \to 0$$

by the finiteness of $\mathbb{E}_1^2$ and the dominated convergence theorem. □

Following is a result which can be applied in combination with the Central Limit Theorem to yield useful results:

Theorem 5.38. *Suppose that $X_1, X_2, \ldots$ is a sequence of random variables (not necessarily independent), and $\mu \in \mathbb{R}$ and $\sigma > 0$ are constants such that $\sqrt{n}(X_n - \mu)$ converges in distribution to $N(0, \sigma^2)$. Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function such that f' is continuous at μ. Then, $\sqrt{n}(f(X_n) - f(\mu))$ converges in distribution to $N(0, f'(\mu)^2\sigma^2)$.*

Proof. First, let $Y_n = \frac{X_n - \mu}{\sigma}$, so that $\sqrt{n}Y_n$ converges in distribution to $N(0,1)$ by Slutsky's Theorem. Secondly, let $g(x) = \frac{f(\sigma x + \mu) - f(\mu)}{\sigma}$, so that

$$\sqrt{n}g(Y_n) = \sqrt{n} \cdot \frac{f(\sigma Y_n + \mu) - f(\mu)}{\sigma} = \sqrt{n} \cdot \frac{f(X_n) - f(\mu)}{\sigma}$$

converges in distribution to $N(0, f'(0)^2)$ if and only if $\sqrt{n}(f(X_n) - f(\mu))$ converges in distribution to $N(0, f'(\mu)^2\sigma^2)$ by Slutsky's Theorem. Thus, we may assume that $\mu = f(\mu) = 0$ and $\sigma = 1$.

Now, since f is differentiable and such that $f(0) = 0$, by Taylor's Theorem, there exists $h(x)$ such that $\lim_{x\to 0} h(x) = 0$ and $f(x) = f'(0)x + h(x)x^2$. Then, $f(X_n) = f'(0)X_n + h(X_n)X_n^2$, so that $\sqrt{n}f(X_n) = f'(0)(\sqrt{n}X_n) + (\sqrt{n}X_n)h(X_n)X_n$. Now, by assumption, $\sqrt{n}X_n$ converges in distribution to $N(0,1)$. Yet since $\frac{1}{\sqrt{n}}$ converges in probability to 0, by Slutsky's theorem this implies that X_n converges in distribution to 0. Yet then, X_n converges in probability to 0.

Then, since $\lim_{x\to 0} h(x) = 0$, $h(X_n)$ converges in probability to 0. Then, since $h(X_n) \xrightarrow{p} 0$ and $X_n \xrightarrow{p} 0$, clearly $h(X_n)X_n \xrightarrow{p} 0$. But then, by Slutsky's Theorem, $(\sqrt{n}X_n)h(X_n)X_n \xrightarrow{d} 0$, whence $(\sqrt{n}X_n)h(X_n)X_n \xrightarrow{p} 0$. Furthermore, $f'(0)\sqrt{n}X_n$ converges in distribution to $f'(0)N(0,1) = N(0, f'(0)^2)$ by Slutsky's Theorem, so by a final application of Slutsky's Theorem, $\sqrt{n}f(X_n) = f'(0)(\sqrt{n}X_n) + (\sqrt{n}X_n)h(X_n)X_n$ converges in distribution to $N(0, f'(0)^2)$, which is the desired result. □

Following is an example calculation using this result:

Example 5.39. A p-coin is a coin that has probability p of turning up heads. Let S_n be the number of heads in n tosses of a p-coin. Then $\sqrt{S_n} - \sqrt{np}$ converges in distribution as $n \to \infty$ to $N(0, \frac{1-p}{4})$.

Proof. Let $X_1, X_2, \ldots$ be i.i.d. random variables taking the value 0 with probability $1-p$ and the value 1 with probability p, so that $S_n = \sum_{i=1}^n X_i$. Then X_i has mean p and variance $p(1-p)$. Thus, by CLT for i.i.d. sums, $\frac{S_n - np}{\sqrt{np(1-p)}}$ converges in distribution to $N(0,1)$. But then $\sqrt{n}\left(\frac{\frac{S_n}{n} - p}{\sqrt{p(1-p)}}\right)$ converges in distribution to $N(0,1)$, so that $\sqrt{n}\left(\frac{S_n}{n} - p\right) \xrightarrow{d} N(0, p(1-p))$.

Now, p is the mean of $\frac{S_n}{n}$. Thus, defining $f(x) = \sqrt{x}$, and noticing that f is differentiable with continuous derivative if $p = (0,1)$, we can apply the previous problem to see that $\sqrt{n}\left(\sqrt{\frac{S_n}{n}} - \sqrt{p}\right)$ converges in distribution to $N(0, p(1-p)f'(p)^2)$. Now, $f'(p) = \frac{1}{2\sqrt{p}}$, so $f'(p)^2 = \frac{1}{4p}$. Thus,

$$\sqrt{n}\left(\sqrt{\frac{S_n}{n}} - \sqrt{p}\right) = \sqrt{S_n} - \sqrt{np} \xrightarrow{d} N\left(0, \frac{1-p}{4}\right).$$

□

We also mention two other forms of the Central Limit Theorem.

Theorem 5.40 (Lindeberg-Feller CLT)**.** *Let $\{k_n\}_{n\geq 1}$ be a sequence of positive integers increasing to infinity. For each n, let $\{X_{n,i}\}_{1\leq i\leq k_n}$ is a collection of independent random variables. Let $\mu_{n,i} = \mathbb{E}[X_{n,i}]$, $\sigma_{n,i}^2 = \mathrm{Var}(X_{n,i})$, and*

$$s_n^2 = \sum_{i=1}^{k_n} \sigma_{n,i}^2.$$

Suppose that for any $\varepsilon > 0$, $\lim_{n\to\infty} \frac{1}{s_n^2}\mathbb{E}[(X_{n,i} - \mu_{n,i})^2 \mid |X_{n,i} - \mu_{n,i}| \geq \varepsilon s_n] = 0$. Then, the random variable $\frac{\sum_{i=1}^{k_n}(X_{n,i}-\mu_{n,i})}{s_n}$ converges in distribution to the standard Gaussian law as $n \to \infty$.

Proof. Similar to the above proof with only a few details changed. □

Theorem 5.41 (Lyapunov CLT)**.** *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent random variables. Let $\mu_i = \mathbb{E}[X_i]$, $\sigma_i^2 = \mathrm{Var}(X_i)$, and $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$. If, for some $\delta > 0$,*

$$\lim_{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}|X_i - \mu_i|^{2+\delta} = 0$$

then the random variable $\frac{\sum_{i=1}^{n}(X_i-\mu_i)}{s_n}$ converges weakly to the standard Gaussian distribution as $n \to \infty$.

Proof. This follows from letting $k_n = n$ and $X_{n,i} = X_i$. Then, the Lyapunov condition will yield the desired result. □