Statistics 205A

Probability Theory

Notes

Druv Pai

Contents

Note: There was about a week at the end of the course on combinatorics of random walks, roughly corresponding to Chapter 4.9 and various sections of Chapter 5 from Durrett; we do not reproduce these notes here, as the results are elementary and derived almost solely by combinatorics.

1 Measure Theory Basics

1.1 Probability Spaces

The beginning motivation is to consider a random experiment. Typically this involves a **state space** Ω and some probability on it. An outcome of an experiment would be $\omega \in \Omega$.

Example 1.1. For a fair coin toss, $\Omega = \{0, 1\}$ with $\mathbb{P}(0) = \mathbb{P}(1) = \frac{1}{2}$.

Example 1.2. For a random number generator in $\Omega = [0, 1]$, if X is the outcome of the random experiment, then X is a uniformly distributed random variable with $\mathbb{P}(X \in [0, a]) = a$, $\mathbb{P}(X \in \mathbb{Q}) = 0$, and $\mathbb{P}(X \in [0, 1] \setminus \mathbb{Q}) = 1$.

If we want our probabilities to satisfy certain natural properties then one encounters issues to define probabilities of all sets.

Example 1.3. Suppose $\Omega = [0, 1]$. Let $\mathbb{P}$ be a set function which is

- translation-invariant: $\mathbb{P}(E) = \mathbb{P}((E + x) \pmod 1)$ for all $E \subseteq \Omega$ and $x \in \Omega$.
- countably additive: if $\{E_n\}$ are disjoint then $\mathbb{P}(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mathbb{P}(E_n)$.
- measures "length": $\mathbb{P}([a, b]) = b - a$.

Define the equivalence relation $x \sim y$ if and only if $x - y \in \mathbb{Q}$, for $x, y \in \mathbb{R}$. By the Axiom of Choice there is a set M which contains exactly one element from each equivalence class. Thus

$$\text{for each } x \in \mathbb{R} \text{ there exists a unique } y \in M \text{ such that } x = y + q \text{ for some } q \in \mathbb{Q}.$$

Let $M_q = \{y + q \colon y \in M\}$ for each $q \in \mathbb{Q}$. Then $\{M_q\}_{q \in \mathbb{Q}}$ are disjoint sets, and

$$\mathbb{R} = \bigcup_{q \in \mathbb{Q}} M_q.$$

Now

$$\mathbb{P}([0, 1]) = \mathbb{P}\left(\bigcup_{q \in \mathbb{Q}} M_q \cap [0, 1]\right) = \sum_{q \in \mathbb{Q}} \mathbb{P}(M_q \cap [0, 1]).$$

If $\mathbb{P}(M_q \cap [0, 1]) = 0$, then $\mathbb{P}([0, 1]) = 0$, a contradiction. If $\mathbb{P}(M_q \cap [0, 1]) > 0$ then $\mathbb{P}([0, 1]) = \infty$, a contradiction.

There are more pathological examples in high dimensions, for example, the *Banach-Tarski paradox*.

Thus our goal is to come up with a general framework with consistent definitions for probabilities on a large class of sets.

Definition 1.4 (σ-Algebra). For Ω a state space, $\Sigma \subseteq 2^{\Omega}$ is a **σ-algebra** or **σ-field** if

1. $\emptyset \in \Sigma$.
2. Closed under complements. $A \in \Sigma \implies A^c \in \Sigma$.
3. Closed under countable union. $\{A_n\} \in \Sigma \implies \bigcup_{n=1}^{\infty} A_n \in \Sigma$.

Remark. Property (3) implies $\bigcup_{n=1}^{m} A_n \in \Sigma$ for any m because the sequence $A_1, \ldots, A_m, A_m, A_m, \ldots$ has the same union as $\{A_n\}_{n=1}^{m}$.

Remark. Property (1) is redundant if Σ is nonempty. By (2), $A \in \Sigma$ implies $A^c \in \Sigma$. Taking $A \cup A^c = \Omega$ implies $\Omega \in \Sigma$ by (3). Finally $\emptyset = \Omega^c \in \Sigma$ by (2). Thus property (1) is a consequence of properties (2) and (3).

Example 1.5. $\Sigma = \{\emptyset, \Omega\}$ is the *smallest* σ-algebra for any sample space Ω. On the other hand $\Sigma = 2^{\Omega}$ is the *largest* σ-algebra for any sample space Ω.

Example 1.6. If

$$\Sigma = \{A \colon \text{either } A \text{ is countable or } A^c \text{ is countable}\}$$

then properties (1) and (2) are trivially true. To verify (3) we note that for the sequence $\{A_n\} \in \Sigma$, then:

- If all the A_n's are countable then $\bigcup_{n=1}^{\infty} A_n$ is countable, so $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.
- If any A_j is co-countable then A_j^c is countable. Then

$$A_j \subseteq \bigcup_{n=1}^{\infty} A_n \implies \left(\bigcup_{n=1}^{\infty} A_n\right)^c \subseteq A_j^c \quad \text{which is countable,}$$

  so $\bigcup_{n=1}^{\infty} A_n$ is co-countable. Thus $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

Definition 1.7 (Measure). A function $\mu \colon \Sigma \to [0, \infty]$ is a **measure** if

1. $\mu(\emptyset) = 0$.
2. Countable additivity. If $\{A_n\}$ is disjoint then $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

The measure μ is **finite** if $\mu(\Omega) < \infty$; μ is **σ-finite** if there is a sequence $\{A_n\}$ such that $\mu(A_n) < \infty$ for each n and $\bigcup_{n=1}^{\infty} A_n = \Omega$. Finally if $\mu(\Omega) = 1$ then μ is a **probability measure**.

Example 1.8. If $\Omega = [n]$ and $\Sigma = 2^{\Omega}$, then if μ is a measure on (Ω, Σ),

$$\mu(A) = \sum_{a \in A} \mu(\{a\}) \quad \text{for all } A \in \Sigma.$$

Proposition 1.9. Properties of the measure are:

1. Monotonicity. If $A \subseteq B$ then $\mu(A) \subseteq \mu(B)$.
2. Countable subadditivity. If $A \subseteq \bigcup_{n=1}^{\infty} B_n$ then $\mu(A) \leq \displaystyle\sum_{n=1}^{\infty} \mu(B_n)$.
3. Continuity from below. If $\{A_n\} \uparrow A$ (that is, $A_n \subseteq A_{n+1}$ and $A = \displaystyle\bigcup_{n=1}^{\infty} A_n$), then $\displaystyle\lim_{n\to\infty} \mu(A_n) \uparrow \mu(A)$.
4. Continuity from above. If $\{A_n\} \downarrow A$ (that is, $A_n \supseteq A_{n+1}$ and $A = \displaystyle\bigcap_{n=1}^{\infty} A_n$), and $\mu(A_1) < \infty$, then $\displaystyle\lim_{n\to\infty} \mu(A_n) \downarrow \mu(A)$.

Proof.

1. Note
$$B = A \cup (B \setminus A) = A \cup (B \cap A^c) \in \Sigma,$$
and since $A \cap B \cap A^c = \emptyset$,
$$\mu(B) = \mu(A \cup (B \cap A^c)) = \mu(A) + \mu(B \cap A^c) \geq \mu(A).$$

2. Let $C_1 = B_1$ and $C_n = B_n \setminus \left(\bigcup_{i=1}^{n-1} B_i\right)$. Then $C_n \cap C_m = \emptyset$ for all n, m, and $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} C_n$. Thus $A \subseteq \bigcup_{n=1}^{\infty} C_n$, so
$$\mu(A) \leq \mu\left(\bigcup_{n=1}^{\infty} C_n\right) = \sum_{n=1}^{\infty} \mu(C_n) \leq \sum_{n=1}^{\infty} \mu(B_n).$$

3. Let $B_n = A_1 \setminus A_n$. Then $B_1 = \emptyset$ and $\{B_n\} \uparrow A_1 \setminus A$. By (3),
$$\mu(B_n) \uparrow \mu(A_1 \setminus A) = \mu(A_1) - \mu(A),$$
and
$$\mu(B_n) = \mu(A_1 \setminus A_n) = \mu(A_1) - \mu(A_n),$$
from which it is clear that $\mu(A_n) \downarrow \mu(A)$.

□

Definition 1.10. For any $A \subseteq 2^{\Omega}$, the **σ-algebra generated by A** is
$$\sigma(A) = \bigcap_{\substack{\Sigma \text{ is a } \sigma\text{-algebra} \\ A \in \Sigma}} \Sigma.$$

Proposition 1.11. $\sigma(A)$ is a σ-algebra.

Proof. From axioms:

1. $\emptyset$ is contained in each σ-algebra that contains A, so $\emptyset \in \sigma(A)$.

2. If $B \in \sigma(A)$ then B is contained in each σ-algebra that contains A, so B^c is contained in each σ-algebra that contains A, so $B^c \in \sigma(A)$.

3. If $\{B_n\} \in \sigma(A)$ then each B_n is contained in each σ-algebra that contains A, so $\displaystyle\bigcup_{n=1}^{\infty} B_n$ is contained in each σ-algebra that contains A, so $\displaystyle\bigcup_{n=1}^{\infty} B_n \in \sigma(A)$.

□

Remark. $\sigma(A)$ is the "smallest" σ-algebra containing A.

Proposition 1.12. $\sigma(X) = \sigma(Y)$ if and only if $X \subseteq \sigma(Y)$ and $Y \subseteq \sigma(X)$.

Proof. $\sigma(X) = \sigma(Y)$ if and only if $\sigma(X) \subseteq \sigma(Y)$ and $\sigma(X) \supseteq \sigma(Y)$. And $\sigma(X) \subseteq \sigma(Y)$ if and only if $X \subseteq \sigma(Y)$, by the characterization of $\sigma(X)$ as the "smallest σ-algebra that contains X." Similarly $\sigma(Y) \subseteq \sigma(X)$ if and only if $Y \subseteq \sigma(X)$. □

Definition 1.13 (Borel σ-Algebra). Let $(\Omega, \mathcal{U})$ be a space equipped with the topology $\mathcal{U}$. Then the **Borel σ-algebra** is defined as $\mathcal{B}_\Omega = \sigma(\mathcal{U})$.

Remark. Most often we take $\Omega = \mathbb{R}$ and $\mathcal{U}$ the Euclidean topology on $\mathbb{R}$. In this case we write $\Sigma = \mathcal{B}_\mathbb{R}$.

Proposition 1.14. The Borel σ-algebra on $\mathbb{R}$ can be represented as

$$\mathcal{B}_\mathbb{R} = \sigma(\{(a,b)\colon a,b \in \mathbb{R}\}) = \sigma(\{[a,b]\colon a,b \in \mathbb{R}\}) = \sigma(\{(a,b)\colon a,b \in \mathbb{Q}\}) = \sigma(\{[a,b]\colon a,b \in \mathbb{Q}\}).$$

Proof. It follows from the fact that every open set in $\mathbb{R}$ can be written as the countable union of open intervals, so given the first definition all the open sets are contained in $\mathcal{B}_\mathbb{R}$. Taking

$$[a,b] = \bigcap_{n=1}^{\infty}\left(a - \frac{1}{n}, b + \frac{1}{n}\right) \quad \text{and} \quad (a,b) = \bigcup_{n=1}^{\infty}\left[a + \frac{1}{n}, b - \frac{1}{n}\right],$$

shows that the first two new definitions of $\mathcal{B}_\mathbb{R}$ are equivalent. These definitions are clearly equal to the last two new definitions because $\mathbb{Q}$ is dense (and can be countably approximated) by $\mathbb{R}$. $\square$

1.2 Constructing Measures

We will now try to define the uniform measure μ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. The plan of action is

1. Define μ on "intervals;" in particular define it on a "semi-algebra" containing its intervals.
2. Extend the definition of μ to an "algebra."
3. Finally extend the definition of μ to a σ-algebra.

We will henceforth omit the proofs for brevity, as they are repetitive and just use the properties of the defined structures in a predictable way. If desired, proofs can be found in Durrett.

Definition 1.15 (Semi-Algebra). A set $\mathcal{A}_{\text{semi}} \subseteq 2^\Omega$ is a **semi-algebra** if

- Closed under finite intersection. If $A_1, A_2 \in \mathcal{A}_{\text{semi}}$ then $A_1 \cap A_2 \in \mathcal{A}_{\text{semi}}$.
- If $A \in \mathcal{A}_{\text{semi}}$ then the decomposition $A^c = \bigcup_{n=1}^{m} B_n$, where $B_n \in \mathcal{A}_{\text{semi}}$, is possible.

It's straightforward to check that

$$\left\{(a,b)\colon a,b \in \overline{\mathbb{R}}\right\} \quad \text{and} \quad \left\{[a,b]\colon a,b \in \overline{\mathbb{R}}\right\} \quad \text{are not semi-algebras,}$$

but

$$\Sigma_{\text{semi}} = \left\{(a,b]\colon a,b \in \overline{\mathbb{R}}\right\} \quad \text{is a semi-algebra.}$$

On this semi-algebra we can define

$$\mu((a,b]) = b - a \quad \text{for all } (a,b] \in \Sigma_{\text{semi}} \quad .$$

Proposition 1.16. μ is countably additive on Σ_{semi}.

Now we wish to extend the definition of μ onto an algebra.

Definition 1.17 (Algebra). A set $\mathcal{A} \subseteq 2^{\Omega}$ is an **algebra** if

1. $\emptyset \in \mathcal{A}$.

2. Closed under complements. If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.

3. Closed under finite union. If $\{A_n\}_{n=1}^m \in \mathcal{A}$ then $\bigcup_{n=1}^{m} A_n \in \mathcal{A}$.

Remark. The nicest property of a semi-algebra $\mathcal{A}_{\text{semi}}$ is that there is an explicit description of the algebra generated by $\mathcal{A}_{\text{semi}}$; obtained by taking all possible disjoint finite unions.

Proposition 1.18. If $\mathcal{A}_{\text{semi}}$ is a semi-algebra then

$$\mathcal{A}_g = \left\{ \bigcup_{n=1}^{m} A_n \colon A_n \in \mathcal{A}_{\text{semi}}, \{A_n\}_{n=1}^m \text{ disjoint} \right\} \quad \text{is an algebra.}$$

Proposition 1.19. Defining μ on Σ_g by

$$\mu(A) = \sum_{n=1}^{m} \mu(A_n) \quad \text{for } A = \bigcup_{n=1}^{m} A_n, \text{ where } \{A_n\}_{n=1}^m \text{ is disjoint and } A_n \in \Sigma_{\text{semi}} \text{ for all } n$$

is well-defined and satisfies the properties of measure.

Theorem 1.20 (Caratheodory's Extension Theorem). Given a countably additive measure μ on an algebra $\mathcal{A}$, it can be extended to a measure on $\sigma(\mathcal{A})$. If μ is σ-finite on $\mathcal{A}$ then the extension is unique.

Remark. The σ-finiteness is required. Consider $\mathcal{A}_{\text{semi}} = \{(a, b] \cap \mathbb{Q} \colon a, b \in \mathbb{R}\}$ and $\mathcal{A}_\sigma = 2^{\mathbb{Q}}$. Then the **counting measure** $\mu(A) = |A|$ on this space is either 0 or ∞ for every set in $\mathcal{A}_{\text{semi}}$. One can check that the counting measure is indeed a measure. We can set $\mu' = 2\mu$, so that $\mu(A) = \mu'(A)$ for every $A \in \mathcal{A}_{\text{semi}}$, and $\mu'(A) = 2\mu(A)$ for every $A \in 2^{\mathbb{Q}}$. In particular μ and μ' disagree on finite sets and are thus not equal. And μ is not a σ-finite measure.

To prove the existence theorem, we introduce a general approach to show a property is true for a σ-algebra.

Remark. To show that some property is true for a σ-algebra, a good method is to show that the sets satisfying these properties must be closed under some natural operations (complement, intersection, union, etc.). Any such family of sets must contain a σ-algebra. The next result will formalize one such set of properties.

Definition 1.21. A class of sets $\mathcal{P}$ is said to be a **π-system** if it is closed under finite intersection, that is, if $A, B \in \mathcal{P}$ then $A \cap B \in \mathcal{P}$.

A class of sets $\mathcal{G}$ is said to be a **λ-system** if

1. $\Omega \in \mathcal{G}$.

2. Closed under differences. If $A \subseteq B$ and $A, B \in \mathcal{G}$, then $B \setminus A \in \mathcal{G}$.

3. Closed under increasing. If $\{A_n\} \in \mathcal{G}$ and $\{A_n\} \uparrow A$ then $A \in \mathcal{G}$.

Theorem 1.22 (π-λ Theorem). If $\mathcal{P}$ a π-system is contained in $\mathcal{G}$ a λ-system, then $\sigma(\mathcal{P}) \subseteq \mathcal{G}$.

Proof of Uniqueness in Theorem 1.20. We first deal with the case that $\mu_1(\Omega) = \mu_2(\Omega) < \infty$.

Let $\mathcal{C} = \{A \in \sigma(A) \colon \mu_1(A) = \mu_2(A)\}$. We know by definition $\mathcal{A} \subseteq \mathcal{C}$.

A σ-algebra is also a λ-system. And given any π-system $\mathcal{P}$, $\sigma(\mathcal{P})$ is the smallest λ-system containing $\mathcal{P}$. Thus to show $\sigma(\mathcal{A}) \subseteq \mathcal{C}$ it's sufficient to show that $\mathcal{C}$ is a λ-system. We verify the axioms:

1. Since $\Omega \in \mathcal{A}$, $\Omega \in \mathcal{C}$.
2. Suppose $A, B \in \mathcal{C}$ with $A \subseteq B$. Then
$$\mu_1(A) = \mu_2(A) \quad \text{and} \quad \mu_1(B) = \mu_2(B).$$
Since μ_1, μ_2 are finite measures,
$$\mu_1(B \setminus A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B \setminus A).$$
Thus $B \setminus A \in \mathcal{C}$.
3. If $\{A_n\} \in \mathcal{C}$ with $\{A_n\} \uparrow A$ then from continuity from below,
$$\mu_1(A_n) \uparrow \mu_1(A) \quad \text{and} \quad \mu_2(A_n) \uparrow \mu_2(A).$$
Then
$$\mu_1(A_n) = \mu_2(A_n) \quad \text{for every } n \qquad \Longrightarrow \qquad \mu_1(A) = \mu_2(A),$$
so $A \in \mathcal{C}$.

Thus $\sigma(\mathcal{A}) \subseteq \mathcal{C}$, so $\sigma(\mathcal{A}) = \mathcal{C}$.

Finally, if μ_1 and μ_2 are σ-finite, then there is a sequence $\{A_n\}$ for which $\mu_1(A_n) < \infty$, $\mu_2(A_n) < \infty$ for all n, and $\bigcup_{n=1}^{\infty} A_n = \Omega$. Then $\mu_1 = \mu_2$ on $\{A \cap A_n \colon A \in \sigma(\mathcal{A})\}$ for each n, so $\mu_1 = \mu_2$ in general. □

Proof Sketch of Uniqueness in Theorem 1.20. We will construct the **outer measure**, a function from 2^{Ω} to $[0, \infty]$:
$$\mu_*(B) = \inf_{\substack{\{A_n\} \in \mathcal{A} \\ B \subseteq \bigcup_{n=1}^{\infty} A_i}} \sum_{n=1}^{\infty} \mu(A_n).$$
One can show that the outer measure possesses the following properties:

- $\mu_*(\emptyset) = 0$.
- Monotonicity. If $B_1, B_2 \in 2^{\Omega}$ with $B_1 \subseteq B_2$ then $\mu_*(B_1) \subseteq \mu_*(B_2)$.
- Countable subadditivity. If $\{C_n\} \in 2^{\Omega}$ then $\mu_*(\bigcup_{n=1}^{\infty} C_n) \leq \sum_{n=1}^{\infty} \mu_*(C_n)$.

Finally, $\mu_* = \mu$ on $\mathcal{A}$.

To get a measure we need to upgrade from countable subadditivity to countable additivity. Defining
$$\mathcal{A}_{\sigma} = \{A \colon \mu_*(E) = \mu_*(E \cap A) + \mu_*(E \cap A^c) \text{ for all } E \subseteq \Omega\}$$
we see that $\mathcal{A}_{\sigma}$ is a σ-algebra, $\sigma(\mathcal{A}) \subseteq \mathcal{A}_{\sigma}$, and μ_* is a measure when restricted to $\mathcal{A}_{\sigma}$. □

Applying this theorem to Σ_g defines the uniform measure on Borel sets.

Remark. We use the notation $\mathcal{A}$ for the underlying algebra, $\mathcal{A}^{\sigma}$ as the elements formed by taking countable unions of elements of $\mathcal{A}$, and $\mathcal{A}^{\sigma\delta}$ as the elements formed by taking countable intersections of elements of $\mathcal{A}^{\sigma}$. Then

$$\text{for any set } B \subseteq \Omega \text{ there is a set } B' \in \mathcal{A}^{\sigma\delta} \text{ such that } B \subseteq B' \text{ and } \mu_*(B) = \mu_*(B'), \text{ or equivalently } \mu_*(B' \setminus B) = 0.$$

Moreover, for any B such that $\mu_*(B) = 0$, $B \in \mathcal{A}_{\sigma}$.

Remark. The construction defines the measure μ_* on sets of the form $A \cup B$ where A is a Borel set and $\mu_*(B) = 0$. It is easy to check that if $\mu_*(B) = 0$ then there is a Borel set $B' \supseteq B$ such that $\mu_*(B') = 0$.

Remark. The set of Lebesgue measurable subsets Σ_{σ} is a strictly larger σ-algebra than $\mathcal{B}_{\mathbb{R}}$. In particular $|\mathcal{B}_{\mathbb{R}}| = 2^{\aleph_0}$ while $|\Sigma_{\sigma}| \geq 2^{|\text{Cantor set}|} = 2^{\aleph_1}$.

1.3 Distributions

Definition 1.23 (Single-Variable Distribution Function). A function $F : \mathbb{R} \to \mathbb{R}$ is a **distribution function** if it is:

1. Bounded. $0 \leq F(x) \leq 1$ for all x.
2. Non-decreasing. If $x \leq y$ then $F(x) \leq F(y)$.
3. Right-continuous. If $\{x_n\} \downarrow x$ then $\lim_{n\to\infty} F(x_n) \downarrow F(x)$.
4. $\lim_{x\to-\infty} F(x) = 0$, and $\lim_{x\to\infty} F(x) = 1$.

Given a distribution function F, one can define $\mu((a, b]) = F(b) - F(a)$ and construct the probability measure μ on all Borel sets. And given a probability measure μ, one can define $F(a) = \mu((-\infty, a])$; it is easy to show F is a distribution function.

We now want to extend this idea to higher-dimension distribution functions. The Borel sets in $\mathbb{R}^d$, $\mathcal{B}_{\mathbb{R}^d}$, can be represented as

$$\mathcal{B}_{\mathbb{R}^d} = \sigma\left(\left\{ \bigtimes_{n=1}^{d} (a_i, b_i) : a_i, b_i \in \overline{\mathbb{R}} \right\}\right) = \sigma\left(\left\{ \bigtimes_{n=1}^{d} B_n : B_n \in \mathcal{B}_{\mathbb{R}} \right\}\right).$$

We introduce the partial ordering on $\mathbb{R}^d$, that is, $x \preceq y$ if and only if $x_n \leq y_n$ for $n \in [d]$.

Definition 1.24 (Distribution Function). A function $F : \mathbb{R}^d \to \mathbb{R}$ is a **distribution function** if it is:

1. Non-decreasing. If $x \preceq y$ then $F(x) \leq F(y)$.
2. Right-continuous. If $\{x_n\} \downarrow x$ then $\lim_{n\to\infty} F(x_n) \downarrow F(x)$.
3. $\lim_{x_n\to-\infty} F(x) = 0$ for all $n \in [d]$, and $\lim_{x\to\infty} F(x) = 1$.
4. For all boxes B, the mass F puts on B is ≥ 0.

The points (3) and (4) imply that $0 \leq F(x) \leq 1$. We will talk about the formula in (4). Given a distribution function F, one can define $\mu((a, b])$ via the Inclusion-Exclusion Principle to be

$$\mu((a, b]) = \sum_{n=0}^{d} (-1)^n \sum_{\substack{S \subseteq [d] \\ |S| = n}} F(a_S, b_{[d]\setminus S}).$$

For example, for $d = 2$,

$$\mu(((a_1, a_2), (b_1, b_2)]) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

And (4) says that $\mu((a, b]) \geq 0$. In this case, a and b are diametrically opposite coordinates of the box $(a, b]$.

We can use the same machine as before to develop a probability measure μ on $\mathcal{B}_{\mathbb{R}^d}$. Similarly to before, given a probability measure μ on $\mathcal{B}_{\mathbb{R}^d}$, we may define a distribution function F by $F(a) = \mu((-\infty, a])$.

Thus we have discussed measure spaces and how to construct measures on $\mathcal{B}_{\mathbb{R}^d}$ for $d \geq 1$.

1.4 Random Variables

Suppose (Ω_1, Σ_1) and (Ω_2, Σ_2) are two measure spaces.

Definition 1.25 (Measurable Function, Random Variable). The function $f\colon \Omega_1 \to \Omega_2$ is **measurable** with respect to Σ_1 and Σ_2 if

$$f^{-1}(A) \in \Sigma_1 \quad \text{for all } A \in \Sigma_2.$$

If $(\Omega, \Sigma_2) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ then f is a **random variable**.

Remark. We use the notation $\{f \in B\}$ as shorthand for $\{x \in \Omega_1 \colon f(x) \in B\} = f^{-1}(B)$, for $B \in \Sigma_2$.

Proposition 1.26. If $\Sigma_2 = \sigma(\mathcal{A})$ then f is measurable with respect to Σ_1 and Σ_2 if $f^{-1}(B) \in \Sigma_1$ for all $B \in \mathcal{A}$.

Proof. Let $\Sigma' = \left\{B \in \mathcal{A} \colon f^{-1}(B) \in \Sigma_1\right\}$. Then Σ' is a σ-algebra:

1. $f^{-1}(\emptyset) = \emptyset$, so $\emptyset \in \Sigma'$.
2. If $B \in \Sigma'$ then $f^{-1}(B) \in \Sigma_1$. But
$$f^{-1}(B^c) = (f^{-1}(B))^c \in \Sigma_1 \quad \text{so} \quad B^c \in \Sigma_1.$$
3. If $\{B_n\} \in \Sigma'$ then $f^{-1}(\bigcup_{n=1}^{\infty} B_n) \in \Sigma_1$. But
$$f^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(B_n) \in \Sigma_1 \quad \text{so} \quad \bigcup_{n=1}^{\infty} B_n \in \Sigma'.$$

Thus Σ' is a σ-algebra, and $\mathcal{A} \subseteq \Sigma'$. Thus $\sigma(\mathcal{A}) = \Sigma_2 \subseteq \Sigma'$, so $\Sigma_2 = \Sigma'$. □

Definition 1.27 (σ-Algebra Generated by Random Variable). If f is a random variable, then $\sigma(f) = \{\{f \in B\} \colon B \in \mathcal{B}_{\mathbb{R}}\}$ is the **σ-algebra generated by f**. If $\{f_\lambda\}$ are random variables then $\sigma(f_\lambda \colon \lambda \in \Lambda) = \sigma\left(\bigcup_{\lambda \in \Lambda} \sigma(f_\lambda)\right)$.

By the previous proposition, $\sigma(f)$ is indeed a σ-algebra.

Proposition 1.28. If $f\colon (\Omega_1, \Sigma_1, \mu_1) \to (\Omega_2, \Sigma_2)$ is measurable with respect to Σ_1 and Σ_2, then f induces a measure μ_2 on (Ω_2, Σ_2) by

$$\mu_2(B) = \mu_1(f^{-1}(B)) \quad \text{for all } B \in \Sigma_2.$$

This is called the **push-forward measure** of μ_1.

Proposition 1.29. If $\{(\Omega_n, \Sigma_n)\}_{n=1}^m$ are measure spaces and $\{f_n : \Omega_n \to \Omega_{n+1}\}_{n=1}^{m-1}$ are measurable functions with respect to Σ_n and Σ_{n+1}, then $g = f_{m-1} \circ \cdots \circ f_1$ is measurable.

Proof. It suffices to prove it for $m = 3$ and induct. Indeed,

$$g^{-1}(B) = \underbrace{f_1^{-1}(\underbrace{f_2^{-1}(\underbrace{B}_{\in \Sigma_3})}_{\in \Sigma_2})}_{\in \Sigma_1} \quad \text{for all } B \in \Sigma_3,$$

and the claim is proved by induction. $\square$

Proposition 1.30. If (Ω, Σ) is a measure space and X_1, X_2 are two random variables, then $(X_1, X_2) : \Omega \to \mathbb{R}^2$ is a measurable function with respect to Σ and $\mathcal{B}_{\mathbb{R}^2}$.

Proof. Recall that

$$\mathcal{B}_{\mathbb{R}^2} = \sigma(\{B_1 \times B_2 : B_1, B_2 \in \mathcal{B}_{\mathbb{R}}\}).$$

Thus it suffices to check $(X_1, X_2)^{-1}(B_1, B_2) \in \Sigma$, for $\mathcal{B}_1, \mathcal{B}_2 \in \mathcal{B}_{\mathbb{R}}$. But

$$(X_1, X_2)^{-1}(B_1 \times B_2) = \underbrace{X_1^{-1}(B_1)}_{\in \Sigma} \cap \underbrace{X_2^{-1}(B_2)}_{\in \Sigma} \in \Sigma \quad \text{for all } B_1, B_2 \in \mathcal{B}_{\mathbb{R}}$$

as desired. $\square$

Proposition 1.31. If $(\Omega_1, \mathcal{B}_{\Omega_1})$ and $(\Omega_2, \mathcal{B}_{\Omega_2})$ are topological measure spaces, and $f : \Omega_1 \to \Omega_2$ is continuous, then f is measurable.

Proof. Note that

$$\mathcal{B}_{\Omega_1} = \sigma(\{U \subseteq \Omega_1 : U \text{ is open}\}) \quad \text{and} \quad \mathcal{B}_{\Omega_2} = \sigma(\{U \subseteq \Omega_2 : U \text{ is open}\}).$$

Thus it suffices to check $f^{-1}(U) \in \mathcal{B}_{\Omega_2}$ for all U open in Ω_2. Indeed, if f is continuous then $f^{-1}(U)$ is open in Ω_1, for all U open in Ω_2. Then $f^{-1}(U) \in \mathcal{B}_{\Omega_1}$. $\square$

Proposition 1.32. If (Ω, Σ) is a measure space and $\{X_n\}_{n=1}^m$ are random variables, then $\displaystyle\sum_{n=1}^m X_n$ is a random variable.

Proof. Both the mappings $f_1 : \Omega \to \mathbb{R}^m$ and $f_2 : \mathbb{R}^m \to \mathbb{R}$ given by

$$f_1(\omega) = \begin{bmatrix} X_1(\omega) \\ \vdots \\ X_m(\omega) \end{bmatrix} \quad \text{and} \quad f_2(z_1, \ldots, z_m) = \sum_{n=1}^m z_n,$$

are continuous and thus measurable, so their composition

$$(f_2 \circ f_1)(\omega) = f_2(f_1(\omega)) = \sum_{n=1}^m X_n(\omega),$$

is measurable. And in fact $f_2 \circ f_1 = \sum_{n=1}^m X_n$, so $\sum_{n=1}^m X_n$ is measurable. $\square$

Proposition 1.33. If $\{X_n\}$ is a sequence of random variables and $\{X_n\} \to X$ pointwise, then X is measurable.

Proof. Recall that
$$\mathcal{B}_{\mathbb{R}} = \sigma\big(\{(a, b) \colon a, b \in \overline{\mathbb{R}}\}\big) = \sigma\big(\{(a, \infty) \colon a \in \overline{\mathbb{R}}\}\big).$$

If $X(\omega) > z$ then $X_n(\omega) > z$ for all large n. Thus
$$\{X > z\} = \limsup_{n \to \infty} \{X_n > z\} = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{X_m > z\} = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \underbrace{X_m^{-1}((z, \infty))}_{\in \Sigma} \in \Sigma.$$

Thus $\{X > z\} = X^{-1}((z, \infty)) \in \Sigma$ for all $z \in \mathbb{R}$, so X is measurable. $\square$

Recall that if $X \colon \Omega \to \mathbb{R}$ induces the push-forward measure $\mu \colon \mathcal{B}_{\mathbb{R}} \to [0, \infty]$ by
$$\mu_X(B) = \mu(X^{-1}(B)) \quad \text{for all } B \in \Sigma_2.$$
It also introduces a distribution function.

We now wish to do the reverse, that is, given a distribution function F we wish to recover a random variable X.

In the case $X \colon ([0, 1], \mathcal{B}_{[0,1]}, \text{uniform measure}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, for X to induce the distribution F, we have
$$F(y) = \mu_X((-\infty, y]) = \mu(X^{-1}((-\infty, y])) \implies X^{-1}((-\infty, y]) = (0, F(y)].$$
Indeed, X "acts like an inverse" to F; the above is true whenever $X(F(y)) = y$ for all y. Sadly, the range of F may not cover $[0, 1]$; if F has a jump discontinuity at y, then $X(F(y))$ is not well-defined.

For general distributions one can come up with various definitions of "inverses" which induce the desired properties. For $\omega \in (0, 1)$ we may write
$$X(\omega) = \sup\{y \colon F(y) < \omega\}.$$
Clearly X is measurable.

1.5 Integration of Measurable Functions

Assume that the underlying measure space is (Ω, Σ, μ), where μ is a σ-finite measure. Assume $f \colon \Omega \to \mathbb{R}$ is measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$. The broad goal is to define the integral of f, $\int_{\Omega} f \, \mathrm{d}\mu$, and the strategy is:

1. First define the integral $\int_{\Omega} f \, \mathrm{d}\mu$ for f a simple function.
2. Extend the definition of $\int_{\Omega} f \, \mathrm{d}\mu$ for f a bounded measurable function with bounded support.
3. Extend the definition of $\int_{\Omega} f \, \mathrm{d}\mu$ for f a non-negative measurable function.
4. Finally, extend the definition of $\int_{\Omega} f \, \mathrm{d}\mu$ for f a general measurable function.

As before, we will omit the proofs, since they are repetitive and will follow from applying definitions straightforwardly. If desired, proofs can be found in Durrett.

Remark. We use the notation
$$\int_\Omega f \mathbb{1}_A \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu \,.$$

Definition 1.34 (Lebesgue Integral for Characteristic Functions). If f is a characteristic function, that is, if $f = \mathbb{1}_A$ for $A \in \Sigma$, then
$$\int_\Omega f \, \mathrm{d}\mu = \int_A \mathrm{d}\mu = \mu(A).$$

Definition 1.35 (Lebesgue Integral for Simple Functions). If f is a simple function, that is, if $f = \sum_{n=1}^{m} c_n \mathbb{1}_{A_n}$ for $\{A_n\} \in \Sigma$ disjoint, then
$$\int_\Omega f \, \mathrm{d}\mu = \sum_{n=1}^{m} c_n \mu(A_n).$$

Proposition 1.36. The Lebesgue integral for simple functions is well defined; if $\{A_n\}_{n=1}^{m} \in \Sigma$ and $\{B_n\}_{n=1}^{m'} \in \Sigma$, then
$$f = \sum_{n=1}^{m} c_n \mathbb{1}_{A_n} = \sum_{n=1}^{m'} d_n \mathbb{1}_{B_n} \implies \int_\Omega f \, \mathrm{d}\mu = \sum_{n=1}^{m} c_n \mu(A_n) = \sum_{n=1}^{m'} d_n \mu(B_n).$$

Proof. For n, m such that $A_n \cap B_m \neq \emptyset$, $c_n = d_m$. And $\{A_i \cap B_j\}_{i \in [m], j \in [m']}$ are disjoint. Then immediately we have
$$\sum_{n=1}^{m} c_i \mu(A_i) = \sum_{i=1}^{m} \sum_{j=1}^{m'} c_i \mu(A_i \cap B_j) = \sum_{i=1}^{m} \sum_{j=1}^{m'} d_j \mu(A_i \cap B_j) = \sum_{j=1}^{m'} d_j \mu(B_j).$$
□

Proposition 1.37. Some properties of $\int_\Omega f \, \mathrm{d}\mu$ are

1. If $f \geq 0$ then $\int_\Omega f \, \mathrm{d}\mu \geq 0$.
2. $\int_\Omega a f \, \mathrm{d}\mu = a \int_\Omega f \, \mathrm{d}\mu$.
3. $\int_\Omega (f + g) \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu + \int_\Omega g \, \mathrm{d}\mu$.
4. If $g \leq f$ then $\int_\Omega g \, \mathrm{d}\mu \leq \int_\Omega f \, \mathrm{d}\mu$.
5. If $g = f$ then $\int_\Omega g \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu$.
6. $\left| \int_\Omega f \, \mathrm{d}\mu \right| \leq \int_\Omega |f| \, \mathrm{d}\mu$.

Definition 1.38 (Lebesgue Integral for Bounded Measurable Functions with Bounded Support). Suppose f is measurable and $|f| \le M$. Further suppose f vanishes outside E with $\mu(E) < \infty$. Then

$$\int_\Omega f \, \mathrm{d}\mu = \sup\left\{\int_\Omega g \, \mathrm{d}\mu : g \text{ is simple}, g \le f\right\}.$$

Proposition 1.39. We can write the approximation in a different way:

$$\sup\left\{\int_\Omega g \, \mathrm{d}\mu : g \text{ is simple}, g \le f\right\} = \inf\left\{\int_\Omega h \, \mathrm{d}\mu : h \text{ is simple}, h \ge f\right\}.$$

Proof. It suffices to prove that there exists simple functions g and h such that

$$\int_\Omega h \, \mathrm{d}\mu - \int_\Omega g \, \mathrm{d}\mu \le \varepsilon.$$

It suffices to consider g, h such that

$$h - f < \delta \text{ pointwise} \quad \text{and} \quad f - g < \delta \text{ pointwise} \implies h - g < 2\delta \text{ pointwise} \implies \int_\Omega h \, \mathrm{d}\mu - \int_\Omega g \, \mathrm{d}\mu < 2\delta\mu(E),$$

and since ε, δ are arbitrary we can fix $\delta = \frac{\varepsilon}{2\mu(E)}$.

So our goal is to find g, h simple functions such that $h - f < \delta$ pointwise and $f - g < \delta$ pointwise. Since f is bounded, $f(\Omega) \subseteq [-M, M]$. Let

$$K = \frac{2M}{\delta} \quad \text{and} \quad I_n = [-M + (n-1)\delta, -M + n\delta] \quad \text{for} \quad n \in [K].$$

Define

$$A_n = f^{-1}(I_n) \cap E \quad \text{and} \quad g = \sum_{n=1}^{K} (n\delta - M)\mathbb{1}_{A_n}, \quad h = \sum_{n=1}^{K} ((n+1)\delta - M)\mathbb{1}_{A_n}.$$

Then $h - f < \delta$ and $f - g < \delta$ pointwise as desired. □

Proposition 1.40. Proposition 1.37 holds for integrals of bounded measurable functions with bounded support. In addition the integral is well-defined on bounded measurable functions with bounded support.

Remark. In case f is simple then the integral definitions Definition 1.35 and Definition 1.38 are the same.

Definition 1.41 (Lebesgue Integral for Non-Negative Measurable Functions). Suppose $f \ge 0$ is measurable. Then

$$\int_\Omega f \, \mathrm{d}\mu = \sup\left\{\int_\Omega h \, \mathrm{d}\mu : h \text{ is bounded with bounded support}, h \le f\right\}.$$

Proposition 1.42. Proposition 1.37 holds for integrals of non-negative measurable functions. In addition the integral is well-defined on non-negative measurable functions.

Remark. In case f is bounded with bounded support then the integral definitions Definition 1.38 and Definition 1.41 are the same.

Lemma 1.43. Suppose $\{E_n\} \in \Sigma$ with $E_n \uparrow \Omega$. Further suppose $\mu(E_n) < \infty$ for all n. Then

$$\lim_{n\to\infty} \int_{E_n} f \wedge n \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu .$$

Remark. We use the notation $x \wedge y = \min\{x, y\}$. Correspondingly $x \vee y = \max\{x, y\}$.

Proof of Lemma 1.43. Since $h_n = (f \wedge n)\mathbb{1}_{E_n}$ is bounded with bounded support, and $h_n \le f$, then

$$\int_\Omega h_n \,\mathrm{d}\mu \le \int_\Omega f \,\mathrm{d}\mu \quad \text{for all } n \qquad \Longrightarrow \lim_{n\to\infty} \int_\Omega h_n \,\mathrm{d}\mu \le \int_\Omega f \,\mathrm{d}\mu\,.$$

Now suppose $0 \le h \le f$ and $h \le M$ and $\operatorname{supp}(h) = \{x : h(x) > 0\}$ obeys $\mu(\operatorname{supp}(h)) < \infty$. Then for $n \ge M$,

$$\int_{E_n} f \wedge n \,\mathrm{d}\mu \ge \int_{E_n} h \,\mathrm{d}\mu = \int_\Omega h \,\mathrm{d}\mu - \int_{E_n^c} h \,\mathrm{d}\mu \ge \int_\Omega h \,\mathrm{d}\mu - \underbrace{M\mu(E_n^c \cap \operatorname{supp}(h))}_{\to 0}$$

so that

$$\liminf_{n\to\infty} \int_{E_n} f \wedge n \,\mathrm{d}\mu \ge \int_\Omega h \,\mathrm{d}\mu\,,$$

which proves the result. □

Definition 1.44 (Positive and Negative Parts). Suppose f is measurable. Define $f^+ = f \vee 0$ and $f^- = (-f) \vee 0$.

Remark. Clearly

$$f = f^+ - f^- \quad \text{and} \quad |f| = f^+ + f^-.$$

Definition 1.45 (Lebesgue Integral for General Measurable Functions). Let f be measurable. Then

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

The function f is **integrable** if $\int_\Omega |f| \,\mathrm{d}\mu < \infty$.

Proposition 1.46. Proposition 1.37 holds for integrals of general measurable functions. In addition the integral is well-defined on general measurable functions.

Remark. In case f is non-negative then the integral definitions Definition 1.41 and Definition 1.45 are the same.

We have been talking about measurable functions, but we can easily translate this theory into the language of random variables.

Definition 1.47 (Expectation). If X is a random variable and $\int_\Omega |X| \,\mathrm{d}\mu < \infty$ then we call

$$\mathbb{E}[X] = \int_\Omega X \,\mathrm{d}\mu$$

the **expectation** of X.

Definition 1.48 (Variance). If X is a random variable and $\mathbb{E}\left[X^2\right] < \infty$ then we call

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

the **variance** of X.

Definition 1.49 (Covariance). If X, Y are random variables, $\mathbb{E}\left[X^2\right] < \infty$, and $\mathbb{E}\left[Y^2\right] < \infty$, then we call

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

the **covariance** of X and Y.

1.6 Limits and Integrals

Suppose for this section that μ is a finite measure. Everything may be applied to σ-finite measures with a bit more work.

Definition 1.50 (Almost Surely). An event $E \in \Sigma$ happens **almost surely** if $\mu(E^c) = 0$.

Definition 1.51 (Almost Sure Convergence). A sequence of functions $\{f_n\}$ which are measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$ **converges almost surely** to f if there is an event $E \in \Sigma$ such that $\mu(E) = 0$ and $\lim_{n\to\infty} f_n(x) = f(x)$ for $x \in E^c$. We write

$$\overset{\text{a.s.}}{\lim_{n\to\infty}} f_n = f.$$

Remark. When we write $f_n \to f$, $f_n \uparrow f$, or $f_n \downarrow f$, this convergence is almost sure convergence.

Proposition 1.52. The following related functions are measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$:

$$\left(\lim_{n\to\infty} f_n\right)\mathbb{1}_{E^c} \quad \text{and} \quad \liminf_{n\to\infty} f_n \quad \text{and} \quad \limsup_{n\to\infty} f_n.$$

Definition 1.53 (Convergence in Measure). A sequence of functions $\{f_n\}$ which are measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$ **converges in measure** to f if

$$\lim_{n\to\infty} \mu(|f_n - f| > \varepsilon) = 0 \quad \text{for all } \varepsilon > 0.$$

We write

$$\overset{\mu}{\lim_{n\to\infty}} f_n = f.$$

Proposition 1.54. If $\mu(\Omega) < \infty$ then $\overset{\text{a.s.}}{\lim_{n\to\infty}} f_n = f$ implies $\overset{\mu}{\lim_{n\to\infty}} f_n = f$.

Example 1.55. We need $\mu(\Omega) < \infty$. For a counterexample take μ the uniform measure on $\mathbb{R}$, and let $f_n = \mathbb{1}_{[-n,n]}$, with $\overset{\text{a.s.}}{\lim_{n\to\infty}} f_n = 1$. But

$$\mu(\{|f_n - f| > \varepsilon\}) = \mu((-\infty, -n)) + \mu((n, \infty)) = \infty \quad \text{for every } n \text{ and } \varepsilon \in (0, 1).$$

Thus $\overset{\mu}{\lim_{n\to\infty}} f_n \neq 1$.

Theorem 1.56 (Bounded Convergence Theorem (BCT)). Let $E \in \Sigma$ have $\mu(E) < \infty$. Suppose $\{f_n\}$ are measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$, vanishes on E^c, and $|f_n| \leq M$, and $\overset{\mu}{\lim_{n\to\infty}} f_n = f$. Then

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

Proof. Let $\varepsilon > 0$ and $F_n = \{|f_n - f| < \varepsilon\}$. Then $\lim^{\mu}_{n\to\infty} f_n = f$ implies $\lim_{n\to\infty} \mu(F_n^c) = 0$. So

$$\begin{aligned}
\lim_{n\to\infty} \left| \int_\Omega f \, \mathrm{d}\mu - \int_\Omega f_n \, \mathrm{d}\mu \right| = \lim_{n\to\infty} \left| \int_E f \, \mathrm{d}\mu - \int_E f_n \, \mathrm{d}\mu \right| &= \lim_{n\to\infty} \int_E |f_n - f| \, \mathrm{d}\mu \\
&= \lim_{n\to\infty} \int_{E\cap F_n} |f_n - f| \, \mathrm{d}\mu + \lim_{n\to\infty} \int_{E\cap F_n^c} |f_n - f| \, \mathrm{d}\mu \\
&= \lim_{n\to\infty} (\varepsilon\mu(F_n) + 2M\mu(F_n^c)) = \lim_{n\to\infty} \varepsilon\mu(F_n) \leq \varepsilon\mu(\Omega)
\end{aligned}$$

and since ε is arbitrary and $\mu(\Omega) < \infty$,

$$\lim_{n\to\infty} \left| \int_\Omega f \, \mathrm{d}\mu - \int_\Omega f_n \, \mathrm{d}\mu \right| = 0 \implies \lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu \,,$$

which completes the proof. □

Theorem 1.57 (Fatou's Lemma). If $\{f_n\}$ are measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$, we have $f_n \geq 0$ then

$$\liminf_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \geq \int_\Omega \left(\liminf_{n\to\infty} f_n \right) \mathrm{d}\mu \,.$$

Remark. The inequality in Theorem 1.57 can be tight. Take for example μ the Lebesgue measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, and let $f_n = n\mathbb{1}_{(0, \frac{1}{n}]}$. Then

$$\liminf_{n\to\infty} f_n = 0 \implies \int_\Omega \left(\liminf_{n\to\infty} f_n \right) \mathrm{d}\mu = 0 \quad \text{but} \quad \int_\Omega f_n \, \mathrm{d}\mu = 1 \implies \liminf_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu = 1.$$

Proof of Theorem 1.57. Let $g_n = \inf_{m \geq n} f_m$. Then $f_n \geq g_n$ and

$$\lim_{n\to\infty} g_n = g = \liminf_{n\to\infty} f_n.$$

Then

$$f_n \geq g_n \implies \int_\Omega f_n \, \mathrm{d}\mu \geq \int_\Omega g_n \, \mathrm{d}\mu \,.$$

Let $\{E_n\} \uparrow \Omega$. Since $g_n \geq 0$, for fixed m,

$$\overset{\text{a.s.}}{\lim_{n\to\infty}} (g_n \wedge m)\mathbb{1}_{E_m} = (g \wedge m)\mathbb{1}_{E_m} \,,$$

and the BCT Theorem 1.56 implies

$$\liminf_{n\to\infty} \int_\Omega g_n \, \mathrm{d}\mu \geq \int_{E_m} g_n \wedge m \, \mathrm{d}\mu \quad \text{and} \quad \lim_{n\to\infty} \int_{E_m} g_n \wedge m \, \mathrm{d}\mu = \int_{E_m} g \wedge m \, \mathrm{d}\mu \,.$$

Taking the supremum over m and using Lemma 1.43 gives the result. □

Theorem 1.58 (Monotone Convergence Theorem (MCT)). If $\{f_n\}$ are measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$, we have $f_n \geq 0$, and $\overset{\text{a.s.}}{\lim_{n\to\infty}} f_n = f$ then

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

Proof. Directly

$$f \geq f_n \implies \int_\Omega f \, \mathrm{d}\mu \geq \int_\Omega f_n \, \mathrm{d}\mu \implies \int_\Omega f \, \mathrm{d}\mu \geq \limsup_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \,,$$

and by Fatou's Lemma, Theorem 1.57,

$$\int_\Omega f \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \,.$$

Thus

$$\limsup_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \leq \int_\Omega f \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \implies \int_\Omega f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \,.$$

□

Theorem 1.59 (Dominated Convergence Theorem (DCT)). If $\{f_n\}$ are measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$ has $\overset{\text{a.s.}}{\lim_{n\to\infty}} f_n = f$ with f measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$, $|f_n| \leq g$ for all n, and g is integrable, then

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

Proof. We know $f_n + g \geq 0$ and $\lim^{\text{a.s.}}_{n\to\infty} f_n + g = f + g$. Thus Fatou's Lemma implies

$$\liminf_{n\to\infty} \int_\Omega (f+g) \, \mathrm{d}\mu \geq \int_\Omega (f+g) \, \mathrm{d}\mu \implies \liminf_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \geq \int_\Omega f \, \mathrm{d}\mu \, .$$

Applying the same reasoning to $\{-f_n\}$ and $-f$ obtains

$$\limsup_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \leq \int_\Omega f \, \mathrm{d}\mu \, .$$

Thus

$$\limsup_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \leq \int_\Omega f \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \implies \lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu \, ,$$

and the theorem is proved. □

In the language of random variables, we have

- (Theorem 1.58) If $X_n \geq 0$ and $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$ then $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.
- (Theorem 1.59) If $|X_n| \leq Y$, $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$, and $\mathbb{E}[Y] < \infty$, then $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

1.7 Function Spaces

We will take this interlude to discuss classes of functions which are defined by their integrability. This will give us the language to introduce a new form of convergence and discuss different types of random variables.

Definition 1.60 ($\mathcal{L}^p$ Space). For $1 \leq p \leq \infty$, the space $\mathcal{L}^p(\Omega, \Sigma, \mu)$ is the set of functions f which are measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$ for which $\|f\|_{\mathcal{L}^p(\Omega,\Sigma,\mu)} < \infty$. Here

$$\|f\|_{\mathcal{L}^p(\Omega,\Sigma,\mu)} = \begin{cases} \left(\int_\Omega |f|^p \, \mathrm{d}\mu\right)^{1/p} & 1 \leq p < \infty \\ \inf\{M \colon \mu(\{|f| > M\}) = 0\} & p = \infty \end{cases}$$

This norm $\|\cdot\|_{\mathcal{L}^p(\Omega,\Sigma,\mu)} = \|\cdot\|_{\mathcal{L}^p}$ is the $\mathcal{L}^p(\Omega, \Sigma, \mu)$ **norm**, or just the $\mathcal{L}^p$ norm.

Remark. If a function $f \in \Sigma$ is integrable then $f \in \mathcal{L}^1(\Omega, \Sigma, \mu)$; if f is square-integrable then $f \in \mathcal{L}^2(\Omega, \Sigma, \mu)$; and so on.

Definition 1.61 (Convergence in $\mathcal{L}^p$). The sequence $\{f_n\} \in \mathcal{L}^p(\Omega, \Sigma, \mu)$ converges to $f \in \mathcal{L}^p(\Omega, \Sigma, \mu)$ in $\mathcal{L}^p(\Omega, \Sigma, \mu)$ (or just $\mathcal{L}^p$ if the measure space is obvious) if

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

We write $\overset{\mathcal{L}^p(\Omega,\Sigma,\mu)}{\lim_{n\to\infty}} f_n = f$ or $\overset{\mathcal{L}^p}{\lim_{n\to\infty}} f_n = f$.

Remark. Since two functions which are equal almost surely have the same integral, if $\{f_n\}$ converges to g_1 in $\mathcal{L}^p(\Omega, \Sigma, \mu)$, and $g_1 = g_2$ almost surely, then $\{f_n\}$ converges to g_2 in $\mathcal{L}^p(\Omega, \Sigma, \mu)$. This is obviously undesirable.

Definition 1.62 (L^p Space). The space $L^p(\Omega, \Sigma, \mu)$ is defined by

$$L^p(\Omega, \Sigma, \mu) = \mathcal{L}^p(\Omega, \Sigma, \mu)/\ker\left(\|\cdot\|_p\right).$$

It is equpped with the norm

$$\|[f]\|_{L^p(\Omega,\Sigma,\mu)} = \|f\|_{\mathcal{L}^p(\Omega,\Sigma,\mu)}.$$

This norm $\|\cdot\|_{L^p(\Omega,\Sigma,\mu)} = \|\cdot\|_{L^p} = \|\cdot\|_p$ is called the $L^p(\Omega, \Sigma, \mu)$ norm, or just the L^p norm.

Remark. $L^p(\Omega, \Sigma, \mu)$ is the set of *equivalence classes* of functions f for which $\|f\|_p < \infty$, where $f \sim g$ if $f = g$ almost surely. Thus every convergent sequence in $L^p(\Omega, \Sigma, \mu)$ has a unique limit.

Proposition 1.63 (Minkowski's Inequality). The $L^p(\Omega, \Sigma, \mu)$ norm is indeed a norm, and obeys the triangle inequality (also called **Minkowski's Inequality**). If $[f], [g] \in L^p(\Omega, \Sigma, \mu)$ then

$$\|[f] + [g]\|_p \leq \|[f]\|_p + \|[g]\|_p.$$

Definition 1.64 (Convergence in L^p). The sequence $\{[f_n]\} \in \mathcal{L}^p(\Omega, \Sigma, \mu)$ converges to $[f] \in L^p(\Omega, \Sigma, \mu)$ in $L^p(\Omega, \Sigma, \mu)$ (or just L^p if the measure space is obvious) if $\overset{\mathcal{L}^p}{\lim_{n\to\infty}} f_n = f$. We write

$$\overset{L^p(\Omega,\Sigma,\mu)}{\lim_{n\to\infty}} [f_n] \to [f] \quad \text{or} \quad \overset{L^p}{\lim_{n\to\infty}} [f_n] = [f].$$

Proposition 1.65. For $1 \leq p \leq \infty$, $L^p(\Omega, \Sigma, \mu)$ is a Banach space; all Cauchy sequences converge, so it is complete.

Proposition 1.66. For $p \leq r$,

$$\mathcal{L}^p(\Omega, \Sigma, \mu) \supseteq \mathcal{L}^r(\Omega, \Sigma, \mu) \quad \text{and} \quad L^p(\Omega, \Sigma, \mu) \supseteq L^r(\Omega, \Sigma, \mu).$$

As a corollary, if $\overset{\mathcal{L}^r}{\lim_{n\to\infty}} f_n = f$ then $\overset{\mathcal{L}^p}{\lim_{n\to\infty}} f_n = f$, and if $\overset{L^r}{\lim_{n\to\infty}} [f_n] = [f]$ then $\overset{L^p}{\lim_{n\to\infty}} [f_n] = [f]$.

Remark. For the rest of the notes, and throughout other resources, the notion of $\mathcal{L}^p$ vs L^p is usually dropped, and equivalence classes $[f] \in L^p(\Omega, \Sigma, \mu)$ are identified with their representatives $f \in \mathcal{L}^p(\Omega, \Sigma, \mu)$.

Remark. We will use the notation $f \in \mathcal{L}^0(\Omega, \Sigma, \mu)$ or $[f] \in L^0(\Omega, \Sigma, \mu)$ to denote f is measurable with respect to Σ and $\mathcal{B}_{\mathbb{R}}$.

1.8 Change of Measure

We discuss the change of measure formulas for integrals. Suppose $(\Omega_1, \Sigma_1, \mu_1)$ is a measure space, (Ω_2, Σ_2) is a measurable space, $X\colon \Omega_1 \to \Omega_2$ is measurable with respect to Σ_1 and Σ_2, and $f\colon \Omega_2 \to \mathbb{R}$ is measurable with respect to Σ_2 and $\mathcal{B}_{\mathbb{R}}$.

Then $f \circ X\colon \Omega_1 \to \mathbb{R}$ is measurable with respect to Σ_1 and $\mathcal{B}_{\mathbb{R}}$. If $\displaystyle\int_{\Omega_1} |f \circ X| \, \mathrm{d}\mu_1 < \infty$, then X induces the push-forward measure μ_2 on (Ω_2, Σ_2), by

$$\mu_2(A) = \mu_1(X^{-1}(A)) \quad \text{for } A \in \Sigma_2.$$

Theorem 1.67 (Change of Measure). With the above conditions,

$$\int_{\Omega_1} (f \circ X) \, \mathrm{d}\mu_1 = \int_{\Omega_2} f \, \mathrm{d}\mu_2 \, .$$

Proof. We use the "integral" machine, that is, prove the statement for simple functions, then bounded measurable functions, then general functions.

Suppose $f = \mathbb{1}_E$ for $E \in \Sigma_2$. Then

$$\int_{\Omega_2} f \, \mathrm{d}\mu_2 = \mu_2(E) = \mu_1(X^{-1}(E)) = \int_{\Omega_1} \mathbb{1}_{X^{-1}(E)} \, \mathrm{d}\mu_1 = \int_{\Omega_1} f \circ X \, \mathrm{d}\mu_1 \, .$$

To extend to f a simple function, it's straightforward by the linearity of the integrals.

To extend to f a non-negative function, we take a sequence of simple functions that increase to f. Let $f_n = \frac{\lfloor 2^n f \rfloor}{2^n}$ and $\widehat{f}_n = f_n \wedge n$. Then $\widehat{f}_n \geq 0$ and $\lim_{n \to \infty}^{\text{a.s.}} \widehat{f}_n \uparrow f$ so $\lim_{n \to \infty}^{\text{a.s.}} \widehat{f}_n \circ X \uparrow f \circ X$. By Theorem 1.58,

$$\int_{\Omega_2} f \, \mathrm{d}\mu_2 = \lim_{n \to \infty} \int_{\Omega_2} \widehat{f}_n \, \mathrm{d}\mu_2 = \lim_{n \to \infty} \int_{\Omega_1} X \circ \widehat{f}_n \, \mathrm{d}\mu_1 = \int_{\Omega_1} X \circ f \, \mathrm{d}\mu_1 \, .$$

To extend to f a general function, linearity is once again required. □

1.9 Product Measures

Suppose $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are measure spaces, with μ_1, μ_2 σ-finite measures. Then we obtain as a reasonable "product" the measurable space $(\Omega_1 \times \Omega_2, \Sigma_p)$, where $\mathcal{A}_p = \{A \times B \colon A \in \Sigma_1, B \in \Sigma_2\}$ and $\Sigma_p = \sigma(\mathcal{A}_p)$.

Proposition 1.68. $\mathcal{A}_p$ is a semi-algebra.

Proof. We verify from axioms.

1. $\emptyset \in \Sigma_1$ and $\emptyset \in \Sigma_2$, so $\emptyset \times \emptyset \in \mathcal{A}_p$. But $\emptyset \times \emptyset = \emptyset$, so $\emptyset \in \mathcal{A}_p$.
2. If $(A_1 \times A_2), (B_1 \times B_2) \in \mathcal{A}_p$, then
$$(A_1 \times A_2) \cap (B_1 \times B_2) = \underbrace{(A_1 \cap B_1)}_{\in \Sigma_1} \times \underbrace{(A_2 \cap B_2)}_{\in \Sigma_2} \in \mathcal{A}_p.$$
The closure under finite intersection property follows from induction.
3. If $(A_1 \times A_2) \in \mathcal{A}_p$, then
$$(A_1 \times A_2)^c = (A_1^c \times A_2) \cup (A_1 \times A_2^c) \in \mathcal{A}_p.$$

□

The goal is to construct the product measure, or the measure on the product space.

Theorem 1.69 (Existence of Product Measures). There exists a unique measure μ on $(\Omega_1 \times \Omega_2, \Sigma_p)$ such that

$$\mu(A \times B) = \mu_1(A)\mu_2(B) \quad \text{for all } A \times B \in \mathcal{A}_p.$$

This measure μ is the **product measure**, and we sometimes write $\mu = \mu_1 \otimes \mu_2$.

Proof. Given a countably additive σ-finite measure μ on a semi-algebra $\mathcal{A}_p$, Theorem 1.20 shows that there is a unique extension to the generated σ-algebra Σ_p. So to prove existence and uniqueness of the product measure, it suffices to show that μ is countably additive on $\mathcal{A}_p$. Suppose $\{A_n \times B_n\} \in \mathcal{A}_p$ is disjoint and $\bigcup_{n=1}^{\infty} A_n \times B_n = A \times B$.

The strategy is to project onto the lower-dimensional space A. For a given $x \in A$, the set $\{B_n : x \in A_n\}$ is a disjoint partition of B. Thus

$$\mu_2(B) = \sum_{n\colon x \in A_n} \mu_2(B_n) \quad \text{for all } x \in A.$$

Thus

$$\mathbb{1}_A \mu_2(B) = \sum_{n=1}^{\infty} \mathbb{1}_{A_n} \mu_2(B_n).$$

Integrating both sides and using Theorem 1.58, we obtain

$$\mu_1(A)\mu_2(B) = \int_{\Omega_1} \mathbb{1}_A \mu_2(B) \, \mathrm{d}\mu_1 = \int_{\Omega_1} \sum_{n=1}^{\infty} \mathbb{1}_{A_n} \mu_2(B_n) = \sum_{n=1}^{\infty} \mu_1(A_n)\mu_2(B_n).$$

Thus we have

$$\mu(A \times B) = \mu_1(A)\mu_1(B) = \sum_{n=1}^{\infty} \mu_1(A_n)\mu_2(B_n) = \sum_{n=1}^{\infty} \mu(A_n \times B_n)$$

and so μ is countably additive, and so the theorem is proved. $\square$

We now want to integrate over product spaces.

Lemma 1.70. Let $E \in \Sigma_p$. For any $x \in \Omega_1$, $E_x = \{y \in \Omega_2 \colon (x, y) \in \Sigma_p\}$ has $E_x \in \Sigma_2$.

Remark. If $E \in \mathcal{A}_p$ then $E = E_1 \times E_2$ where $E_1 \in \Sigma_1$ and $E_2 \in \Sigma_2$. Then $E_x = \emptyset$ or $E_x = E_2$, so $E_x \in \Sigma_2$.

Proof of Lemma 1.70. The strategy is to show that the sets E with this property form a σ-algebra $\mathcal{A}_\sigma$. We do this by axioms.

1. If $E = \emptyset$ then $E_x = \emptyset$ and thus $E_x \in \Sigma_2$ for all x. Thus $\emptyset \in \mathcal{A}_\sigma$.

2. If $E \in \mathcal{A}_\sigma$ then, for any $x \in \Omega_2$, $E_x \in \Sigma_2$. Then $(E_x)^c \in \Sigma_2$. But $(E_x)^c = (E^c)_x$, so $(E^c)_x \in \Sigma_2$. Thus $E^c \in \mathcal{A}_\sigma$.

3. If $\{E_n\} \in \mathcal{A}_\sigma$ and $\bigcup_{n=1}^{\infty} E_n = E$, then for a given $x \in \Omega_2$,

$$E_x = \left(\bigcup_{n=1}^{\infty} E_n\right)_x = \bigcup_{n=1}^{\infty} \underbrace{(E_n)_x}_{\in \Sigma_2} \in \Sigma_2.$$

   Thus $E_x \in \Sigma_2$ for all x, so $E \in \mathcal{A}_\sigma$.

Since $\mathcal{A}_\sigma$ is a σ-algebra which contains $\mathcal{A}_p$, $\mathcal{A}_\sigma \supseteq \Sigma_p$. This implies the claim. $\square$

Theorem 1.71 (Integral of Product Measure). For any $E \in \Sigma_p$,

$$\mu(E) = \int_{\Omega_1} \mu_2(E_x) \, \mathrm{d}\mu_1(x)\,.$$

Proof. The result is obvious for rectangles. In particular let $E \in \mathcal{A}_p$. Then we can write $E = E_1 \times E_2$ for $E_1 \in \Sigma_1, E_2 \in \Sigma_2$. Then

$$\int_{\Omega_1} \mu_2(E_x) \, \mathrm{d}\mu_1(x) = \int_{E_1} \mu_2(E_2) \, \mathrm{d}\mu_1 = \mu_1(E_1)\mu_2(E_2) = \mu(E).$$

Let $\mathcal{A}$ be the set of $E \in \Sigma_p$ for which the result holds. Now $\mathcal{A}_p \subseteq \mathcal{A}$ by the above. We claim $\mathcal{A}$ is a λ-system. We verify the axioms.

1. We have
$$\int_{\Omega_1} \mu_2((\Omega_1 \times \Omega_2)_x)\,\mathrm{d}\mu_1(x) = \int_{\Omega_1} \mu_2(\Omega_2)\,\mathrm{d}\mu_1 = \mu_1(\Omega_1)\mu_2(\Omega_2) = \mu(\Omega_1 \times \Omega_2).$$
Thus $\Omega_1 \times \Omega_2 \in \mathcal{A}$.

2. Suppose $E_1, E_2 \in \mathcal{A}$ and $E_1 \subseteq E_2$. Then
$$\begin{aligned} (E_1 \setminus E_2)_x &= (E_1)_x \setminus (E_2)_x \\ \implies \mu_2((E_1 \setminus E_2)_x) &= \mu_2((E_1)_x) - \mu_2((E_2)_x) \\ \implies \int_{\Omega_1} \mu_2((E_1 \setminus E_2)_x)\,\mathrm{d}\mu_1(x) &= \int_{\Omega} \mu_2((E_1)_x)\,\mathrm{d}\mu_1(x) - \int_{\Omega} \mu_2((E_2)_x)\,\mathrm{d}\mu_1(x) = \mu(E_1) - \mu(E_2) = \mu(E_1 \setminus E_2). \end{aligned}$$
Thus $E_1 \setminus E_2 \in \mathcal{A}$.

3. Suppose $\{E_n\} \in \mathcal{A}$ and $\{E_n\} \uparrow E$. Then $\{(E_n)_x\} \uparrow E_x$ for all x. Thus $\mu_2((E_n)_x) \uparrow \mu_2(E_x)$. By MCT Theorem 1.58,
$$\mu(E) = \lim_{n\to\infty} \mu(E_n) = \lim_{n\to\infty} \int_{\Omega_1} \mu_2((E_n)_x)\,\mathrm{d}\mu_1(x) = \int_{\Omega_1} \mu_2(E_x)\,\mathrm{d}\mu_1(x)\,.$$
Thus $E \in \mathcal{A}$.

Thus $\mathcal{A}$ is a λ-system with $\mathcal{A}_p \subseteq \mathcal{A}$. By the π-λ theorem Theorem 1.22, $\Sigma_p = \sigma(\mathcal{A}_p) \subseteq \mathcal{A}$, so $\Sigma_p = \mathcal{A}$. And so the property holds for all $E \in \Sigma_p$. □

Theorem 1.72 (Fubini's Theorem). For any $f \geq 0$ or $\int_{\Omega_1 \times \Omega_2} |f|\,\mathrm{d}\mu < \infty$, the following holds:

1. For all $x \in \Omega_1$, $f(x, \cdot) \in \Sigma_2$.
2. $\displaystyle\int_{\Omega_2} f(\cdot, y)\,\mathrm{d}\mu_2(y) \in \Sigma_1$.
3. $\displaystyle\int_{\Omega_1} \left(\int_{\Omega_2} f(x, y)\,\mathrm{d}\mu_2(y)\right)\mathrm{d}\mu_1(x) = \int_{\Omega_1 \times \Omega_2} f\,\mathrm{d}\mu.$

Proof. We have already shown it for $f = \mathbb{1}_E$. By linearity of the integral and the fact that the sum of measurable functions is measurable, the above holds for simple functions. Suppose f is non-negative measurable. Taking $f_n = \frac{\lfloor 2^n f \rfloor}{2^n} \wedge n$, we note $\{f_n\} \uparrow f$ and each f_n is simple, so
$$\int_{\Omega_1} \left(\int_{\Omega_2} f_n(x, y)\,\mathrm{d}\mu_2(y)\right)\mathrm{d}\mu_1(x) = \int_{\Omega_1 \times \Omega_2} f_n\,\mathrm{d}\mu\,.$$
Then by MCT Theorem 1.58,
$$\lim_{n\to\infty} \int_{\Omega_1} \left(\int_{\Omega_2} f_n(x, y)\,\mathrm{d}\mu_2(y)\right)\mathrm{d}\mu_1(x) = \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y)\,\mathrm{d}\mu_2(y)\right)\mathrm{d}\mu_1(x) \quad \text{and} \quad \lim_{n\to\infty} \int_{\Omega_1 \times \Omega_2} f_n\,\mathrm{d}\mu = \int_{\Omega_1 \times \Omega_2} f\,\mathrm{d}\mu\,.$$
Thus
$$\int_{\Omega_1} \left(\int_{\Omega_2} f(x, y)\,\mathrm{d}\mu_2(y)\right)\mathrm{d}\mu_1(x) = \lim_{n\to\infty} \int_{\Omega_1} \left(\int_{\Omega_2} f_n(x, y)\,\mathrm{d}\mu_2(y)\right)\mathrm{d}\mu_1(x) = \lim_{n\to\infty} \int_{\Omega_1 G \times \Omega_2} f_n\,\mathrm{d}\mu = \int_{\Omega_1 \times \Omega_2} f\,\mathrm{d}\mu\,.$$
To extend to general f with $\int_{\Omega_1 \times \Omega_2} |f|\,\mathrm{d}\mu < \infty$, write $f = f^+ - f^-$ and use linearity of the integrals. □

Corollary 1.73 (Undergraduate-Level Fubini's Theorem). By symmetry, Fubini's Theorem implies

1. For all $y \in \Omega_2$, $f(\cdot, y) \in \Sigma_1$.

2. $\displaystyle\int_{\Omega_1} f(x, \cdot)\,\mathrm{d}\mu_1(x) \in \Sigma_1$.

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

This implies the undergraduate-level Fubini's theorem: if $\int_{\Omega_1 \times \Omega_2} |f|\,\mathrm{d}\mu < \infty$ or $f \geq 0$,

$$\int_{\Omega_1} \int_{\Omega_2} f\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1 = \int_{\Omega_2} \int_{\Omega_1} f\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2\,.$$

Remark. Things can go very wrong when the assumptions are not met. The following examples violate some conditions of Fubini's Theorem Theorem 1.72.

Example 1.74. Take $\Omega_1, \Omega_2 = \mathbb{N}$ and $\mu_1 = \mu_2$ the counting measure. Then if

$$f(m, n) = \begin{cases} 1 & m = n \\ -1 & m = n + 1 \\ 0 & \text{otherwise} \end{cases},$$

$\int_{\Omega_1 \times \Omega_2} |f|\,\mathrm{d}\mu = \sum_{n=1}^{\infty} 2 = \infty$, and

$$\int_{\Omega_1} \int_{\Omega_2} f\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1 = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} f(m, n) = 1 \quad \text{and} \quad \int_{\Omega_2} \int_{\Omega_1} f\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2 = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} f(m, n) = 0.$$

Thus in this case Fubini's Theorem Theorem 1.72 does not hold.

Example 1.75. Take $\Omega_1, \Omega_2 = (0, 1)$, μ_1 the uniform measure, μ_2 the counting measure. Let $E = \{(x, x) \colon x \in (0, 1)\}$. Then

$$\int_{\Omega_1} \mu_2(E_x)\,\mathrm{d}\mu_1(x) = 1 \quad \text{and} \quad \int_{\Omega_2} \mu_1(E_y)\,\mathrm{d}\mu_2(y) = 0.$$

Remark. We can iteratively define such integrals for higher dimensional product spaces.

2 Laws of Large Numbers

Henceforth we will work in the measure space $(\Omega, \Sigma, \mathbb{P})$, where $\mathbb{P}$ is a probability measure, so that $\mathbb{P}(\Omega) = 1$. In this space convergence in measure becomes **convergence in probability** and is denoted by $\overset{p}{\lim_{n\to\infty}} X_n = X$.

Let $\{X_n\} \in \Sigma$ be a sequence of independent random variables. Define $S_n = \sum_{m=1}^{n} X_n$. The prevailing question is the asymptotic behavior of $\frac{S_n}{n}$. These are called **Laws Of Large Numbers** (LLN). The Weak LLN (or WLLN) are results about the convergence in probability of $\frac{S_n}{n}$, while the Strong LLN (or SLLN) are results about the convergence almost surely of $\frac{S_n}{n}$.

2.1 Independence

Definition 2.1 (Mutual Independence of Collections of Sets). Collections $\{\mathcal{A}_\lambda\}$ are mutually independent if for any finite subset $I \subseteq \Lambda$ and sets $A_\lambda \in \mathcal{A}_\lambda$,

$$\mathbb{P}\left(\bigcap_{\lambda\in I} A_\lambda\right) = \prod_{\lambda\in I} \mathbb{P}(A_\lambda).$$

Remark. To check mutual independence of a finite subset $\{\mathcal{A}_\lambda\}_{\lambda\in\Lambda_F}$ it's equivalent to check the mutual independence condition for $I = \Lambda_F$, assuming that $\Omega \in \mathcal{A}_\lambda$ for each $\lambda \in \Lambda_F$. This is because, if the condition has been shown for $I = \Lambda_F$, then for any $I \subseteq \Lambda_F$ setting $A_\lambda = \Omega$ for $\lambda \notin I$ implies that the condition is true for I as well. And if the condition is shown for all $I \subseteq \Lambda_F$ then it's also been shown for $I = \Lambda_F$.

Definition 2.2 (Mutual Independence of σ-Algebras). The σ-algebras $\{\Sigma_\lambda\} \subseteq \Sigma$ are said to be **mutually independent** if and only if they are independent in the sense of Definition 2.1.

Remark. Since $\Omega \in \Sigma_\lambda$ for each λ, to check mutual independence of a finite subset Λ_F it suffices to check the independence condition for $I = \Lambda_F$.

Recall the definition of $\sigma(X)$ from Definition 1.27 as $\sigma(X) = \left\{X^{-1}(B) \colon B \in \mathcal{B}_{\mathbb{R}}\right\}$, and that $\sigma(X)$ is indeed a σ-algebra.

Definition 2.3 (Mutual Independence of Random Variables). Random variables $\{X_\lambda\}$ are independent if and only if $\{\sigma(X_\lambda)\}$ are independent.

We now want to find a good way to check independence of σ-algebras.

Proposition 2.4. Suppose $\{\mathcal{A}_\lambda\} \subseteq \Sigma$ are independent π systems in the sense of Definition 2.1. Then $\{\sigma(\mathcal{A}_\lambda)\}$ are independent σ-algebras.

Proof. Let $\Lambda_F = \{\lambda_1, \dots, \lambda_m\} \subseteq \Lambda$ be finite. Fix $B_n \in \mathcal{A}_{\lambda_n}$ for $n \in [m-1]$. We induct on m, for the hypothesis that

$\{\sigma(\mathcal{A}_{\lambda_n})\}_{n=1}^m$ are independent. The base case $m = 1$ is trivial. Now consider arbitrary m. Define two measures on $\sigma(\mathcal{A}_{\lambda_m})$:

$$\mathbb{P}'(B) = \mathbb{P}\left(\left(\bigcap_{n=1}^{m-1} B_n\right) \cap B\right) \quad \text{and} \quad \mathbb{P}''(B) = \mathbb{P}(B)\prod_{n=1}^{m-1}\mathbb{P}(B_n).$$

By hypothesis $\mathbb{P}' = \mathbb{P}''$ on $\mathcal{A}_{\lambda_m}$. By the proof of uniqueness of Caratheodory's Extension Theorem Theorem 1.20, $\mathbb{P}' = \mathbb{P}''$ on $\sigma(\mathcal{A}_{\lambda_m})$. Thus

$$\mathbb{P}\left(\bigcap_{n=1}^{m} B_n\right) = \mathbb{P}'(B_m) = \mathbb{P}''(B_m) = \prod_{n=1}^{m}\mathbb{P}(B_m),$$

for any sets $\{B_n\}_{n=1}^m$ where $B_n \in \mathcal{A}_{\lambda_n}$. Thus $\{\sigma(\mathcal{A}_{\lambda_n})\}_{n=1}^m = \{\sigma(\mathcal{A}_\lambda)\}_{\lambda\in\Lambda_F}$ are independent. This holds for every Λ_F, so $\{\sigma(\mathcal{A}_\lambda)\}$ are independent. □

Example 2.5 (Pairwise Independence Doesn't Imply Mutual Independence). Let $X_1, X_2, X_3 \in \{0, 1\}$ and let the distribution of (X_1, X_2, X_3) be uniform over all triples for which $X_1 + X_2 + X_3 = 0 \pmod 2$. Then by symmetry each $X_n \sim \mathrm{Bern}\left(\frac{1}{2}\right)$ and each pair (X_n, X_m) are independent. However, (X_1, X_2, X_3) is not mutually independent, since

$$\mathbb{P}(X_1 = X_2 = X_3 = 1) = 0 \neq \frac{1}{8} = \prod_{n=1}^{3}\mathbb{P}(\{X_n = 1\}).$$

Definition 2.6 (Tail σ-Algebra). For an ordered index set Λ and a set of random variables $\{X_\lambda\}$, we define $\mathcal{T}_\lambda^X = \sigma(X_\eta \colon \eta \geq \lambda, \eta \in \Lambda)$. Then the **tail σ-algebra** is defined as $\mathcal{T}^X = \bigcap_{\lambda\in\Lambda}\mathcal{T}_\lambda^X$.

Theorem 2.7 (Kolmogorov 0-1 Law). If $\{X_\lambda\}$ are independent then $\mathcal{T}^X$ is μ-trivial:

$$\mathbb{P}(E) = 0 \quad \text{or} \quad \mathbb{P}(E) = 1 \quad \text{for every } E \in \mathcal{T}^X.$$

Proof. Fix $E \in \mathcal{T}^X$. Since

$$E \in \mathcal{T}^X \implies E \in \sigma(X_\eta \colon \eta \geq \lambda, \eta \in \Lambda) \implies E \perp\!\!\!\perp \{X_\eta\}_{\eta\in\Lambda\colon \eta<\lambda} \implies E \perp\!\!\!\perp \bigcup_{\substack{\eta\in\Lambda \\ \eta<\lambda}}\sigma(X_\eta)$$

for all $\lambda \in \Lambda$,

$$E \perp\!\!\!\perp \bigcup_{\lambda\in\Lambda}\sigma(X_\lambda),$$

and $\bigcup_{\lambda\in\Lambda}\sigma(X_\lambda)$ is a π-system, by the π-λ Theorem Theorem 1.22,

$$E \perp\!\!\!\perp \sigma(X_\lambda \colon \lambda \in \Lambda).$$

And

$$E \in \mathcal{T}^X \subseteq \sigma(X_\lambda \colon \lambda \in \Lambda).$$

Thus

$$E \perp\!\!\!\perp E \quad\implies\quad \mathbb{P}(E) = \mathbb{P}(E \cap E) = \mathbb{P}(E)^2 \quad\implies\quad \mathbb{P}(E) = 0 \quad \text{or} \quad \mathbb{P}(E) = 1,$$

which proves the theorem. □

Proposition 2.8. If $\{\mathcal{A}_{ij}\}_{i\in[n], j\in[m_i]}$ are π-systems containing Ω which are mutually independent, then $\left\{\sigma(\bigcup_{j=1}^{m_i}\mathcal{A}_{ij})\right\}_{i=1}^n$ are mutually independent.

Proof. Write $\mathcal{A}_i = \left\{\bigcap_{j=1}^k A_{ij} \colon k \in \mathbb{N}, A_{ij} \in \mathcal{A}_{ij}\right\}$. Then $\mathcal{A}_i$ is a π-system such that $\Omega \in \mathcal{A}_i$ and $\bigcup_{j=1}^{m_i}\mathcal{A}_{i,j} \in \mathcal{A}_i$. Thus by Proposition 2.4, $\{\sigma(\mathcal{A}_i)\}_{i=1}^n = \left\{\sigma(\bigcup_{j=1}^{m_i}\mathcal{A}_{ij})\right\}_{i=1}^n$ are independent. □

Finally we want to discuss the existence of infinitely many independent random variables – for if such a thing does not exist then our theory goes down the drain. Fortunately they do exist and are constructible.

Theorem 2.9 (Kolmogorov's Extension Theorem). Suppose we are given a measure space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \mu_n)$. Then there exists a unique measure μ_{n+1} on $(\mathbb{R}^{n+1}, \mathcal{B}_{\mathbb{R}^{n+1}})$ such that

$$\mu_{n+1}(A \times \mathbb{R}) = \mu_n(A) \quad \text{for all } A \in \mathcal{B}_{\mathbb{R}^n}.$$

We can use this to inductively define a sequence of independent random variables, where the first n variables have joint measure μ_n, that is,

$$\mu_n(A) = \mathbb{P}((X_1, \ldots, X_n) \in A) \quad \text{for } A \in \mathcal{B}_{\mathbb{R}^n}.$$

Remark. If a sequence of random variables $\{X_n\}$ is independent but their distribution is a function of n, we write $X_n \overset{\text{ind}}{\sim} \mathrm{Distribution}(n)$. If furthermore they are identically distributed, we write $X_n \overset{\text{i.i.d.}}{\sim} \mathrm{Distribution}$, and the random variables $\{X_n\}$ are i.i.d. or independent and identically distributed.

2.2 L^2 WLLN

Theorem 2.10 (Markov's Inequality). If $X \in L^1$ is a non-negative integrable random variable, then for any positive t,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

Corollary 2.11 (Chebyshev's Inequality). If $X \in L^2$ then for any positive t,

$$\mathbb{P}(|X| > t) \leq \frac{\mathbb{E}\left[X^2\right]}{t^2}.$$

Proof. Apply Markov's inequality to X^2. ☐

Proposition 2.12. For $X, Y \in L^2$, $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Proof. We have

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

☐

Proposition 2.13. If $\{X_n\} \in L^2$ are independent then they are **uncorrelated**, that is, $\mathrm{Cov}(X_m, X_n) = 0$ for all $m \neq n$.

Proof. We use the standard integral machine. Let $X_n = \mathbb{1}_{E_n}$, so $\mathbb{E}[X_n] = \mathbb{P}(E_n)$, for all n. Then

$$\begin{aligned} \mathrm{Cov}(X_m, X_n) &= \mathbb{E}[X_m X_n] - \mathbb{E}[X_m]\mathbb{E}[X_n] = \mathbb{P}(E_m E_n) - \mathbb{P}(E_m)\mathbb{P}(E_n) \\ &= \mathbb{P}(E_m)\mathbb{P}(E_n) - \mathbb{P}(E_m)\mathbb{P}(E_n) = 0 \end{aligned}$$

since $\sigma(X_m) \perp\!\!\!\perp \sigma(X_n)$. Now if $X_n = \sum_{k=1}^{\ell_n} c_{nk}\mathbb{1}_{E_{nk}}$ for $\{E_{nk}\}_{k=1}^{\ell_n} \in \sigma(X_n)$ disjoint. Then

$$\mathrm{Cov}(X_m, X_n) = \mathrm{Cov}\left(\sum_{k=1}^{\ell_m} c_{mk}\mathbb{1}_{E_{mk}}, \sum_{k=1}^{\ell_n} c_{nk}\mathbb{1}_{E_{nk}}\right) = \sum_{k=1}^{\ell_m}\sum_{k'=1}^{\ell_n} c_{mk}c_{nk'}\,\mathrm{Cov}\left(\mathbb{1}_{E_{mk}}, \mathbb{1}_{E_{nk'}}\right) = 0,$$

again since $\sigma(X_m) \perp\!\!\!\perp \sigma(X_n)$. Then if $X_n \geq 0$ is measurable then taking $\overline{X}_{nk} = \frac{\lfloor 2^k X_n \rfloor}{2^k} \wedge k$ and using MCT Theorem 1.58,

$$\begin{aligned}
\operatorname{Cov}(X_m, X_n) &= \mathbb{E}[X_m X_n] - \mathbb{E}[X_m]\mathbb{E}[X_n] = \int_\Omega X_m X_n \, \mathrm{d}\mathbb{P} - \mathbb{E}[X_m]\mathbb{E}[X_n] \\
&= \lim_{k\to\infty} \int_\Omega \overline{X}_{mk} \overline{X}_{nk} \, \mathrm{d}\mathbb{P} - \mathbb{E}[X_m]\mathbb{E}[X_n] = \lim_{k\to\infty} \mathbb{E}\left[\overline{X}_{mk}\overline{X}_{nk}\right] - \mathbb{E}[X_m]\mathbb{E}[X_n] \\
&= \lim_{k\to\infty} \mathbb{E}\left[\overline{X}_{mk}\right]\mathbb{E}\left[\overline{X}_{nk}\right] - \mathbb{E}[X_m]\mathbb{E}[X_n] = \mathbb{E}[X_m]\mathbb{E}[X_n] - \mathbb{E}[X_m]\mathbb{E}[X_n] = 0.
\end{aligned}$$

For general X_n it follows by linearity. □

So if $\{X_n\}$ are uncorrelated then $\mathbb{E}[X_m X_n] = \mathbb{E}[X_m]\mathbb{E}[X_n]$ for all $m \neq n$.

Proposition 2.14. If $\{X_n\} \in L^2$ are uncorrelated, then $\operatorname{Var}(S_n) = \sum_{m=1}^{n} \operatorname{Var}(X_m)$.

Proof. We have

$$\begin{aligned}
\operatorname{Var}(S_n) = \mathbb{E}\left[(S_n - \mathbb{E}[S_n])^2\right] &= \mathbb{E}\left[\left(\sum_{m=1}^{n}(X_m - \mathbb{E}[X_m])\right)^2\right] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right] \\
&= \sum_{m=1}^{n} \mathbb{E}\left[(X_m - \mathbb{E}[X_m])^2\right] + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \\
&= \sum_{m=1}^{n} \operatorname{Var}(X_m) + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathbb{E}[X_i - \mathbb{E}[X_i]]\mathbb{E}[X_j - \mathbb{E}[X_j]] = \sum_{m=1}^{n}\operatorname{Var}(X_m).
\end{aligned}$$

□

Proposition 2.15 (Convergence in L^p Implies Convergence In Probability). If $p \geq 1$ then

$$\overset{L^p}{\lim_{n\to\infty}} Z_n = 0 \quad \Longrightarrow \quad \overset{p}{\lim_{n\to\infty}} Z_n = 0.$$

Proof. Let $\varepsilon > 0$. By Markov's inequality Theorem 2.10 on the random variables $\{|Z_n|\}$ shows that

$$\lim_{n\to\infty} \mathbb{P}(|Z_n| > \varepsilon) = \lim_{n\to\infty} \mathbb{P}(|Z_n|^p > \varepsilon^p) \leq \lim_{n\to\infty} \frac{\mathbb{E}[|Z_n|^p]}{\varepsilon^p} = 0.$$

Thus $\lim^p_{n\to\infty} Z_n = 0$. □

Theorem 2.16 (L^2 WLLN). Suppose $\{X_n\} \in L^2$ are independent with $\mathbb{E}[X_n] = \mu$ and $\mathbb{E}\left[X_n^2\right] \leq C < \infty$ for all n. Then $\frac{S_n}{n} \xrightarrow{L^2} \mu$ and $\overset{p}{\lim_{n\to\infty}} \frac{S_n}{n} = \mu$.

Proof. By linearity $\mathbb{E}\left[\frac{S_n}{n}\right] = \mu$. We have

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] &= \operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\operatorname{Var}(S_n) = \frac{1}{n^2}\operatorname{Var}\left(\sum_{m=1}^{n} X_m\right) = \frac{1}{n^2}\sum_{m=1}^{n}\operatorname{Var}(X_m) \\
&\leq \frac{nC}{n^2} = \frac{C}{n}
\end{aligned}$$

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] = \lim_{n\to\infty} \frac{C}{n} = 0.$$

Thus $\frac{S_n}{n} \xrightarrow{L^2} \mu$, so by Proposition 2.15, $\lim^p_{n\to\infty} \frac{S_n}{n} = \mu$. □

Example 2.17 (Polynomial Approximation). Given continuous $f\colon [0,1] \to \mathbb{R}$ and $\varepsilon > 0$, we want to find a sequence of polynomials $\{f_n\}$ which converge to f uniformly on $[0,1]$, that is,

$$\lim_{n\to\infty} \sup_{x\in[0,1]} |f_n(x) - f(x)| = 0.$$

In particular we pick

$$f_n(x) = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m} f\left(\frac{m}{n}\right) = \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] \quad \text{where } X_n \overset{\text{i.i.d.}}{\sim} \operatorname{Bern}(x).$$

By the L^2 weak law Theorem 2.16, $\lim^p_{n\to\infty} \frac{S_n}{n} = x$. Since f is continuous on a compact interval $[0,1]$ it is uniformly continuous and bounded on $[0,1]$. Let $M = \sup_{x\in[0,1]} |f(x)|$. Let $\varepsilon > 0$ and pick δ such that if $|x-y| < \delta$ then $|f(x) - f(y)| < \varepsilon$. Using Jensen's inequality and that $\lim^p_{n\to\infty} \frac{S_n}{n} = x$,

$$\begin{aligned}
\lim_{n\to\infty} |f_n(x) - f(x)| &= \lim_{n\to\infty} \left|\mathbb{E}\left[f\left(\frac{S_n}{n}\right) - f(x)\right]\right| \leq \lim_{n\to\infty} \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|; \left|\frac{S_n}{n} - x\right| < \delta\right] + \lim_{n\to\infty} \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|; \left|\frac{S_n}{n} - x\right| \geq \delta\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\varepsilon; \left|\frac{S_n}{n} - x\right| < \delta\right] + \lim_{n\to\infty} \mathbb{E}\left[2M; \left|\frac{S_n}{n} - x\right| \geq \delta\right] \\
&\leq \lim_{n\to\infty} \mathbb{E}[\varepsilon] + 2M \lim_{n\to\infty} \mathbb{E}\left[\mathbb{1}_{\left|\frac{S_n}{n} - x\right| \geq \delta}\right] \\
&\leq \varepsilon + 2M \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - x\right| \geq \delta\right) \\
&\leq \varepsilon.
\end{aligned}$$

Since ε is arbitrary $\lim_{n\to\infty} |f_n(x) - f(x)| = 0$ for each $x \in [0,1]$, so since $[0,1]$ is compact,

$$\sup_{x\in[0,1]} \lim_{n\to\infty} |f_n(x) - f(x)| = 0 \implies \lim_{n\to\infty} \sup_{x\in[0,1]} |f_n(x) - f(x)| = 0.$$

This proves the claim.

2.3 WLLN for Triangular Arrays

The goal is to now prove the weak law without the L^2 assumption. Our goal is to *truncate*, that is, consider random variables of the form $X_M = X\mathbb{1}_{|X|\leq M}$. Note that $\mathbb{E}\left[X_M^2\right] \leq M^2 < \infty$ for all M even if $\mathbb{E}\left[X^2\right] < \infty$. That is, $X_M \in L^2$ even if $X \notin L^2$.

Definition 2.18 (Triangular Array). A **triangular array** of random variables is a collection $\{X_{mn}\}_{n\in[m], m\in\mathbb{N}} \in L^0$ for which $\{X_{mn}\}_{n\in[m]}$ are independent for each m.

Theorem 2.19 (WLLN for Triangular Arrays). Suppose $\{X_{mn}\}_{n\in[m], m\in\mathbb{N}} \in L^0$ is a triangular array. Further suppose for each m there is a constant b_m such that, if $\overline{X}_{mn} = X_{mn}\mathbb{1}_{|X_{mn}|<b_m}$ for which

(i) $\displaystyle\lim_{m\to\infty} \sum_{n=1}^{m} \mathbb{P}(|X_{mn}| > b_m) = 0,$

(ii) $\lim_{m\to\infty} \frac{1}{b_m^2} \sum_{n=1}^{m} \mathbb{E}\left[\overline{X}_{mn}^2\right] = 0.$

Then

$$\lim^p_{m\to\infty} \frac{1}{b_m} \sum_{n=1}^{m} \left(X_{mn} - \mathbb{E}\left[\overline{X}_{mn}\right]\right) = 0.$$

Proof. Let $\varepsilon > 0$. Then for any m,

$$\mathbb{P}\left(\left|\frac{1}{b_m}\sum_{n=1}^{m}\left(X_{mn} - \mathbb{E}\left[\overline{X}_{mn}\right]\right)\right| > \varepsilon\right) \leq \mathbb{P}\left(\sum_{n=1}^{m} X_{mn} \neq \sum_{n=1}^{m} \overline{X}_{mn}\right) + \mathbb{P}\left(\left|\frac{1}{b_m}\sum_{n=1}^{m}\left(\overline{X}_{mn} - \mathbb{E}\left[\overline{X}_{mn}\right]\right)\right| > \varepsilon\right).$$

To estimate the first term, we have

$$\mathbb{P}\left(\sum_{n=1}^{m} X_{mn} \neq \sum_{n=1}^{m} \overline{X}_{mn}\right) \leq \mathbb{P}\left(\bigcup_{n=1}^{m} \left\{X_{mn} \neq \overline{X}_{mn}\right\}\right) \leq \sum_{n=1}^{m} \mathbb{P}\left(X_{mn} \neq \overline{X}_{mn}\right) = \sum_{n=1}^{m} \mathbb{P}(|X_{mn}| > b_m).$$

To estimate the second term, we first note that for any $X \in L^2$,

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

Then by Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{b_m}\sum_{n=1}^{m}\left(\overline{X}_{mn} - \mathbb{E}\left[\overline{X}_{mn}\right]\right)\right| > \varepsilon\right) &\leq \frac{1}{\varepsilon^2}\operatorname{Var}\left(\frac{1}{b_m}\sum_{n=1}^{m}\left(\overline{X}_{mn} - \mathbb{E}\left[\overline{X}_{mn}\right]\right)\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[\left(\frac{1}{b_m}\sum_{n=1}^{m}\left(\overline{X}_{mn} - \mathbb{E}\left[\overline{X}_{mn}\right]\right)\right)^2\right] \\
&= \frac{1}{\varepsilon^2 b_m^2}\mathbb{E}\left[\left(\sum_{n=1}^{m}\overline{X}_{mn} - \mathbb{E}\left[\sum_{n=1}^{m}\overline{X}_{mn}\right]\right)^2\right] = \frac{1}{\varepsilon^2 b_m^2}\operatorname{Var}\left(\sum_{n=1}^{m}\overline{X}_{mn}\right) \\
&= \frac{1}{\varepsilon^2 b_m^2}\sum_{n=1}^{m}\operatorname{Var}\left(\overline{X}_{mn}\right) \leq \frac{1}{\varepsilon^2 b_m^2}\sum_{n=1}^{m}\mathbb{E}\left[\overline{X}_{mn}^2\right].
\end{aligned}$$

Thus by our assumptions

$$\lim_{m\to\infty}\mathbb{P}\left(\left|\frac{1}{b_m}\sum_{n=1}^{m}\left(X_{mn} - \mathbb{E}\left[\overline{X}_{mn}\right]\right)\right| > \varepsilon\right) \leq \lim_{m\to\infty}\sum_{n=1}^{m}\mathbb{P}(|X_{mn}| > b_m) + \lim_{m\to\infty}\frac{1}{\varepsilon^2 b_m^2}\sum_{n=1}^{m}\mathbb{E}\left[\overline{X}_{mn}^2\right] = 0.$$

Thus $\lim^p_{m\to\infty}\frac{1}{b_m}\sum_{n=1}^{m}\left(X_{mn} - \mathbb{E}\left[\overline{X}_{mn}\right]\right) = 0$. □

2.4 General WLLN

The following estimate will be continually of use to us.

Lemma 2.20 (Power Estimate). If $p > 0$ and $Y \in L^p$ has $Y \geq 0$, then $\mathbb{E}[Y^p] = \int_0^\infty p y^{p-1}\mathbb{P}(Y > y)\,\mathrm{d}y$.

Proof. Using Fubini's Theorem Theorem 1.72, and the definition of expectation Definition 1.47,

$$\begin{aligned}
\int_0^\infty p y^{p-1}\mathbb{P}(Y > y)\,\mathrm{d}y &= \int_0^\infty\int_\Omega p y^{p-1}\mathbb{1}_{Y>y}\,\mathrm{d}\mathbb{P}\,\mathrm{d}y = \int_\Omega\int_0^\infty p y^{p-1}\mathbb{1}_{Y>y}\,\mathrm{d}y\,\mathrm{d}\mathbb{P} = \int_\Omega\int_0^Y p y^{p-1}\,\mathrm{d}y\,\mathrm{d}\mathbb{P} \\
&= \int_\Omega Y^p\,\mathrm{d}\mathbb{P} = \mathbb{E}[Y^p].
\end{aligned}$$

□

Lemma 2.21 (General WLLN). Let $\{X_n\} \in L^0$ be i.i.d. with

$$\lim_{x\to\infty} x\mathbb{P}(|X_1| > x) = 0.$$

Let $\mu_n = \mathbb{E}[X_1; |X_1| \leq n]$. Then $\overset{p}{\lim_{n\to\infty}} \left(\frac{S_n}{n} - \mu_n\right) = 0$.

Proof. We use the WLLN for Triangular Arrays Theorem 2.19, letting $X_{mn} = X_n$ and $b_m = m$. To check (i) of the assumptions of Theorem 2.19,

$$\lim_{m\to\infty} \sum_{n=1}^{m} \mathbb{P}(|X_{mn}| > b_m) = \lim_{m\to\infty} m\mathbb{P}(|X_{mn}| > m) = \lim_{m\to\infty} m\mathbb{P}(|X_1| > m) = 0.$$

To check (ii) of the assumptions of Theorem 2.19,

$$\begin{aligned}
\lim_{m\to\infty} \frac{1}{b_m^2} \sum_{n=1}^{m} \mathbb{E}\left[\overline{X}_{mn}^2\right] &= \lim_{m\to\infty} \frac{2}{m^2} \sum_{n=1}^{m} \int_0^\infty y\mathbb{P}\left(\left|\overline{X}_{mn}\right| > y\right) \mathrm{d}y \\
&= \lim_{m\to\infty} \frac{2}{b_m^2} \sum_{n=1}^{m} \int_0^\infty y\mathbb{P}\left(\left|X_{mn}\mathbb{1}_{|X_{mn}|\leq b_m}\right| > y\right) \mathrm{d}y \\
&= \lim_{m\to\infty} \frac{2}{b_m^2} \sum_{n=1}^{m} \int_0^\infty y\mathbb{P}\left(|X_{mn}|\mathbb{1}_{|X_{mn}|\leq b_m} > y\right) \mathrm{d}y \\
&= \lim_{m\to\infty} \frac{2}{m^2} \sum_{n=1}^{m} \int_0^\infty y\mathbb{P}\left(|X_n|\mathbb{1}_{|X_n|\leq m} > y\right) \mathrm{d}y \\
&= \lim_{m\to\infty} \frac{2}{m^2} \sum_{n=1}^{m} \int_0^\infty y\mathbb{P}\left(|X_n|\mathbb{1}_{|X_n|\leq m} > y\right) \mathrm{d}y \\
&= \lim_{m\to\infty} \frac{2}{m^2} \sum_{n=1}^{m} \int_0^m y\mathbb{P}(|X_n| > y) \,\mathrm{d}y \\
&= \lim_{m\to\infty} \frac{2}{m} \int_0^m y\mathbb{P}(|X_1| > y) \,\mathrm{d}y \\
&= 0
\end{aligned}$$

where the last equality is derived from DCT Theorem 1.59 and the assumption of the General WLLN.

Thus the WLLN for Triangular Arrays says that

$$\overset{p}{\lim_{m\to\infty}} \frac{1}{b_m} \sum_{n=1}^{m} (X_{mn} - \mathbb{E}\left[\overline{X}_{mn}\right]) = \overset{p}{\lim_{m\to\infty}} \frac{1}{m} \sum_{n=1}^{m} (X_n - \mu_m) = \frac{S_m}{m} - \mu_m = 0,$$

as desired. ∎

Corollary 2.22 (Traditional WLLN). Suppose $\{X_n\} \in L^1$ are i.i.d. and $\mu = \mathbb{E}[X_1]$. Then $\overset{p}{\lim_{n\to\infty}} \frac{S_n}{n} = \mu$.

Proof. By DCT Theorem 1.59,

$$\lim_{x\to\infty} x\mathbb{P}(|X_1| > x) = \lim_{x\to\infty} \mathbb{E}[x; |X_1| > x] \leq \lim_{x\to\infty} \mathbb{E}[|X_1|; |X_1| > x] = \mathbb{E}\left[|X_1|\left(\lim_{x\to\infty} \mathbb{1}_{|X_1|>x}\right)\right] = 0.$$

And again by DCT Theorem 1.59,

$$\lim_{n\to\infty} \mu_n = \lim_{n\to\infty} \mathbb{E}[X_1; |X_1| \leq n] = \mathbb{E}\left[X_1\left(\lim_{n\to\infty} \mathbb{1}_{|X_1|\leq n}\right)\right] = \mathbb{E}[X_1] = \mu.$$

By Lemma 2.21, if $\varepsilon > 0$ then
$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) &\leq \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \frac{\varepsilon}{2}\right) + \lim_{n\to\infty} \mathbb{P}\left(|\mu_n - \mu| > \frac{\varepsilon}{2}\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \frac{\varepsilon}{2}\right) + \lim_{n\to\infty} \mathbb{1}_{|\mu_n - \mu| > \frac{\varepsilon}{2}} = 0.\end{aligned}$$
Thus $\lim^p_{n\to\infty} \frac{S_n}{n} = \mu$ as desired. □

2.5 Growth Rates

Suppose $\{X_n\} \in L^0$ are i.i.d. with $X_n \geq 0, \mathbb{E}[X_n] = \infty$. Fix $M > 0$ and let $X_n' = X_n \mathbb{1}_{|X_n| < M}$ and $S_n' = \sum_{m=1}^{n} X_m'$. Then by Traditional WLLN Corollary 2.22, $\lim^p_{n\to\infty} \frac{S_n'}{n} = \mathbb{E}[X_1']$, and
$$\lim^p_{n\to\infty} \frac{S_n}{n} \geq \lim^p_{n\to\infty} \frac{S_n'}{n} = \mathbb{E}[X_1'].$$
By making M arbitrarily high $\mathbb{E}[X_1']$ can be made arbitrarily large. By MCT Theorem 1.58,
$$\lim^p_{n\to\infty} \frac{S_n}{n} = \infty,$$
or it can be said that $\dfrac{S_n}{n}$ does not converge in probability.

The salient question now is the rate of growth of S_n.

Example 2.23 (St. Petersburg Paradox). Suppose $X \in L^0$ with
$$\mathbb{P}(X = 2^n) = 2^{-n} \quad \text{for all } n \geq 1.$$
Then
$$\mathbb{E}[X] = \sum_{n=1}^{\infty} 2^n \mathbb{P}(X = 2^n) = \sum_{n=1}^{\infty} 2^n \cdot 2^{-n} = \sum_{n=1}^{\infty} 1 = \infty.$$
Let $\{X_n\} \overset{\text{i.i.d.}}{\sim} X$. Then $\lim^p_{n\to\infty} \frac{S_n}{n} = \infty$, so S_n grows superlinearly in probability. To apply Theorem 2.19 with $X_{mn} = X_n$, we have to pick b_n. We are guided by the principle that in checking condition (ii) of Theorem 2.19 we want to take b_n as small as possible while allowing condition (i) to hold. We observe that
$$\mathbb{P}(X \geq 2^m) = \sum_{n=m}^{\infty} 2^{-n} = 2^{-m+1}.$$
Let
$$k(n) = \log_2(n) + K(n)$$
where $\lim_{n\to\infty} K(n) = \infty$ and is chosen so that $k(n)$ is an integer. Let
$$b_n = 2^{k(n)} = n2^{K(n)}.$$
Then
$$\begin{aligned}\lim_{m\to\infty} \sum_{n=1}^{m} \mathbb{P}(|X_{mn}| > b_m) &= \lim_{m\to\infty} \sum_{n=1}^{m} \mathbb{P}(|X_n| \geq b_m) = \lim_{m\to\infty} m\mathbb{P}(|X| \geq b_m) = \lim_{m\to\infty} m\mathbb{P}(X \geq b_m) \\ &= \lim_{m\to\infty} m\mathbb{P}\left(X \geq 2^{k(m)}\right) = \lim_{m\to\infty} m2^{-k(m)+1} = \lim_{m\to\infty} \frac{2m}{2^{k(m)}} = \lim_{m\to\infty} \frac{2m}{m2^{K(m)}} \\ &= \lim_{m\to\infty} 2^{1-K(m)} = 0.\end{aligned}$$

Thus condition (i) of Theorem 2.19 is satisfied. To check condition (ii) of Theorem 2.19, let

$$\overline{X}_{mn} = X_{mn}\mathbb{1}_{|X_{mn}| \leq b_m} = X_n \mathbb{1}_{|X_n| \leq b_n}$$

then

$$\begin{aligned}
\mathbb{E}\left[\overline{X}_{mn}^2\right] &= \mathbb{E}\left[X_n^2; |X_n| \leq b_m\right] = \mathbb{E}\left[X^2; |X| \leq b_m\right] = \sum_{n=1}^{\infty} 2^{2n}\mathbb{P}(X = 2^n)\mathbb{1}_{2^n \leq b_m} = \sum_{n=1}^{\infty} 2^{2n} \cdot 2^{-n}\mathbb{1}_{2^n \leq b_m} \\
&= \sum_{n=1}^{\infty} 2^n \mathbb{1}_{2^n \leq 2^{k(m)}} = \sum_{n=1}^{k(m)} 2^n \leq 2^{k(m)} \sum_{n=0}^{\infty} 2^{-n} = 2b_m
\end{aligned}$$

$$\lim_{m\to\infty} \frac{1}{b_m^2} \sum_{n=1}^{m} \mathbb{E}\left[\overline{X}_{mn}^2\right] \leq \lim_{m\to\infty} \frac{1}{b_m^2} \sum_{n=1}^{m} (2b_m) = \lim_{m\to\infty} \frac{2m}{b_m} = \lim_{m\to\infty} \frac{2m}{m2^{K(m)}} = \lim_{m\to\infty} 2^{1-K(m)} = 0.$$

Thus condition (2) of Theorem 2.19 is satisfied. It remains to evaluate $\mathbb{E}\left[\overline{X}_{mn}\right]$ and b_m. Indeed,

$$\begin{aligned}
\mathbb{E}\left[\overline{X}_{mn}\right] &= \mathbb{E}[X_{mn}; |X_{mn}| \leq b_m] = \mathbb{E}[X_n; |X_n| \leq b_m] = \mathbb{E}[X; |X| \leq b_m] = \mathbb{E}[X; X \leq b_m] = \sum_{n=1}^{\infty} 2^n \mathbb{P}(X = 2^n)\mathbb{1}_{2^n \leq b_m} \\
&= \sum_{n=1}^{\infty} 2^n \cdot 2^{-n}\mathbb{1}_{2^n \leq 2^{k(m)}} = \sum_{n=1}^{\infty} \mathbb{1}_{n \leq k(m)} = \sum_{n=1}^{k(m)} 1 = k(m).
\end{aligned}$$

$$\sum_{n=1}^{m} \mathbb{E}\left[\overline{X}_{mn}\right] = \sum_{n=1}^{m} k(m) = mk(m) = m(\log_2(m) + K(m)).$$

Thus

$$\lim_{m\to\infty} \frac{K(m)}{\log_2(m)} = 0 \implies \lim_{m\to\infty} \frac{1}{m\log_2(m)} \sum_{n=1}^{m} \mathbb{E}\left[\overline{X}_{mn}\right] = 1.$$

Therefore by Theorem 2.19,

$$\overset{p}{\lim_{m\to\infty}} \frac{\left(\sum_{n=1}^{m} X_n\right) - m(\log_2(m) + K(m))}{m2^{K(m)}} = 0.$$

If $K(m) \approx \log_2(\log_2(m))$ then

$$\overset{p}{\lim_{m\to\infty}} \frac{\left(\sum_{n=1}^{m} X_n\right) - m(\log_2(m) + K(m))}{m2^{K(m)}} = \overset{p}{\lim_{m\to\infty}} \frac{\sum_{n=1}^{m} X_n}{m\log_2(m)} - \underbrace{\frac{m(\log_2(m) + \log_2(\log_2(m)))}{m\log_2(m)}}_{\to 1} = 0$$

and so

$$\overset{p}{\lim_{m\to\infty}} \frac{1}{m\log_2(m)} \sum_{n=1}^{m} X_n = 1.$$

2.6 Borel-Cantelli Lemmas

Definition 2.24 (Limits of Sets). Let $\{A_n\} \subseteq \Omega$. Then

$$\limsup_{n\to\infty} A_n = \lim_{m\to\infty} \bigcup_{n=m}^{\infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega \text{ that are in infinitely many } A_n\} = \{\omega \colon \omega \in A_n \text{ i.o.}\} = \{A_n \text{ i.o.}\}.$$

Similarly

$$\liminf_{n\to\infty} A_n = \lim_{m\to\infty} \bigcap_{n=m}^{\infty} A_n = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n = \{\omega \text{ that are in all but finitely many } A_n\} = \{\omega \colon \omega \in A_n \text{ ev.}\} = \{A_n \text{ ev.}\}.$$

Proposition 2.25. We have the following identities:

$$\limsup_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\limsup_{n\to\infty} A_n} \quad \text{and} \quad \liminf_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\liminf_{n\to\infty} A_n}.$$

Theorem 2.26. $\lim^{\text{a.s.}}_{n\to\infty} X_n = X$ if and only if for every $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.}) = 0$.

Proof.

$$\mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.}) = 0\ \forall\varepsilon \iff \mathbb{P}(|X_n - X| \leq \varepsilon \text{ ev.}) = 1\ \forall\varepsilon \iff \overset{\text{a.s.}}{\lim_{n\to\infty}} |X_n - X| \leq \varepsilon\ \forall\varepsilon \iff \overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X.$$

□

Theorem 2.27 (First Borel-Cantelli Lemma). If $\{A_n\} \in \Sigma$ then

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}(A_n \text{ i.o.}) = 0.$$

Proof. Let $N = \sum_{n=1}^{\infty} \mathbb{1}_{A_n}$. Then $N(\omega)$ is the number of events ω belongs to. By Fubini's Theorem Theorem 1.72,

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_{n=1}^{\infty} \mathbb{1}_{A_n}\right] = \sum_{n=1}^{\infty} \mathbb{E}[\mathbb{1}_{A_n}] = \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty.$$

Thus $N < \infty$ a.s.. □

Theorem 2.28 (Second Borel-Cantelli Lemma). If $\{A_n\} \in \Sigma$ are independent then

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \implies \mathbb{P}(A_n \text{ i.o.}) = 1.$$

Proof. Let $M < N < \infty$. Since $1 - x \leq \mathrm{e}^{-x}$ and $\{A_n\}$ are independent then

$$\mathbb{P}\left(\bigcap_{n=M}^{N} A_n^c\right) = \prod_{n=M}^{N} \mathbb{P}(A_n^c) = \prod_{n=M}^{N} (1 - \mathbb{P}(A_n^c)) \leq \prod_{n=M}^{N} \exp(-\mathbb{P}(A_n)) = \exp\left(-\sum_{n=M}^{N} \mathbb{P}(A_n)\right),$$

so

$$\lim_{N\to\infty} \mathbb{P}\left(\bigcap_{n=M}^{N} A_n^c\right) \leq \lim_{N\to\infty} \exp\left(-\sum_{n=M}^{N} \mathbb{P}(A_n)\right) = \exp\left(-\sum_{n=M}^{\infty} \mathbb{P}(A_n)\right) = 0 \quad \text{for all } M \in (0, \infty) \quad .$$

Thus

$$\mathbb{P}\left(\bigcup_{n=M}^{\infty} A_n\right) = 1 - \mathbb{P}\left(\left(\bigcup_{n=M}^{\infty} A_n\right)^c\right) = 1 - \mathbb{P}\left(\bigcap_{n=M}^{\infty} A_n^c\right) = 1 \quad \text{for all } M.$$

Since $\{A_n \text{ i.o.}\} = \lim_{M\to\infty} \bigcup_{n=M}^{\infty} A_n$, by continuity of measure,

$$\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}\left(\lim_{M\to\infty} \bigcup_{n=M}^{\infty} A_n\right) = \lim_{M\to\infty} \mathbb{P}\left(\bigcup_{n=M}^{\infty} A_n\right) = 1.$$

□

An application of the Borel-Cantelli Lemmas is to prove the SLLN provided $\{X_n\} \in L^4$.

Theorem 2.29 (Fourth-Moment SLLN). Let $\{X_n\} \in L^4$ be i.i.d. with $\mathbb{E}[X_n] = \mu$. Then $\overset{\text{a.s.}}{\lim_{n\to\infty}} \dfrac{S_n}{n} = \mu$.

Proof. Let $X'_n = X_n - \mu$. Then $\{X'_n\} \in L^4$ are independent. Then

$$\mathbb{E}\left[S_n'^4\right] = \mathbb{E}\left[\left(\sum_{m=1}^{n} X'_m\right)^4\right] = \sum_{1\leq i,j,k,\ell\leq n} X'_i X'_j X'_k X'_\ell = \sum_{i=1}^{n}\mathbb{E}\left[X_i'^4\right] + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathbb{E}\left[X_i'^2 X_j'^2\right]$$

$$= \sum_{i=1}^{n}\mathbb{E}\left[X_i'^4\right] + 2\sum_{i=1}^{n}(i-1)\mathbb{E}\left[X_i'^2\right]^2 = n\mathbb{E}\left[X_1'^4\right] + \mathbb{E}\left[X_1'^2\right]\sum_{i=1}^{n} 2(i-1)$$

$$= n\mathbb{E}\left[X_1'^4\right] + 3n(n-1)\mathbb{E}\left[X_1'^2\right]^2 \leq Cn^2,$$

where $C < \infty$. Then by Chebyshev's inequality, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S'_n}{n}\right| > \varepsilon\right) = \mathbb{P}(|S'_n| > n\varepsilon) \leq \frac{\mathbb{E}\left[S_n'^4\right]}{\varepsilon^4 n^4} \leq \frac{C}{\varepsilon^4 n^2}.$$

Thus

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{S'_n}{n}\right| > \varepsilon\right) \leq \sum_{n=1}^{\infty}\frac{C}{\varepsilon^4 n^2} = \frac{C\pi^2}{6\varepsilon^4} < \infty.$$

Thus by Borel-Cantelli Lemma Theorem 2.27 and Theorem 2.26,

$$\mathbb{P}\left(\left|\frac{S'_n}{n}\right| > \varepsilon \text{ i.o.}\right) = 0 \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S'_n}{n} = 0.$$

Finally,

$$\overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S'_n}{n} = 0 \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S_n - n\mu}{n} = 0 \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S_n}{n} - \mu = 0 \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S_n}{n} = \mu.$$

□

2.7 Etemadi's SLLN

The goal is to now prove the SLLN without fourth moment assumptions, and in fact all we require is $\{X_n\} \in L^1$.

Theorem 2.30 (SLLN). Let $X \in L^1$, and let $\mathbb{E}[X] = \mu$. Let $\{X_n\} \sim X$ be pairwise independent and identically distributed. Then $\overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S_n}{n} = \mu$.

Proof. The proof follows Etemadi's idea. Without loss of generality assume $X_n \geq 0$, since at the end we can apply the theorem to $\{X_n^-\}, \{X_n^+\} \in L^1$. Let $Y_n = X_n \mathbb{1}_{|X_n|\leq n}$ and $T_n = \sum_{m=1}^{n} Y_m$.

Lemma 1. It is sufficient to prove that $\overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{T_n}{n} = \mu$.

Proof. By Lemma 2.20,

$$\sum_{n=1}^{\infty}\mathbb{P}(X_n \neq Y_n) = \sum_{n=1}^{\infty}\mathbb{P}(|X_n| > n) = \sum_{n=1}^{\infty}\mathbb{P}(|X| > n) \leq \int_0^{\infty}\mathbb{P}(|X| > t)\,\mathrm{d}t = \mathbb{E}[|X|] < \infty.$$

Then by Theorem 2.27, there exists a function $R \in L^1$ such that

$$\mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0 \implies \mathbb{P}(X_n = Y_n \text{ ev.}) = 1 \implies \mathbb{P}(|S_n - T_n| \leq R) = 1 \quad \text{for all } n.$$

Thus $\lim_{n\to\infty}^{\text{a.s.}} \frac{S_n}{n} = \mu$. ■

Lemma 2. If $y \geq 0$ then

$$2y \sum_{n>y}^{\infty} \frac{1}{n^2} \leq 4.$$

Proof. We begin with the observation that if $m \geq 2$ then

$$\sum_{n=m}^{\infty} \frac{1}{n^2} \leq \int_{m-1}^{\infty} \frac{\mathrm{d}x}{x^2} = \frac{1}{m-1}.$$

When $y \geq 1$,

$$2y \sum_{n>y} \frac{1}{n^2} = 2y \sum_{n=\lfloor y \rfloor+1} \frac{1}{n^2} \leq \frac{2y}{\lfloor y \rfloor} \leq 4,$$

since $\frac{y}{\lfloor y \rfloor} \leq 2$ for $y \geq 1$. For $0 \leq y < 1$, we note

$$2y \sum_{n>y} \frac{1}{n^2} \leq 2\left(1 + \sum_{n=2}^{\infty} \frac{1}{n^2}\right) \leq 4.$$

This establishes the lemma. ∎

Lemma 3. We have the inequality,

$$\sum_{n=1}^{\infty} \frac{\mathrm{Var}(Y_n)}{n^2} \leq 4\mathbb{E}[|X|] < \infty.$$

Proof. To bound the sum, we first note that for any $Y \in L^2$,

$$\mathrm{Var}(Y) = \mathbb{E}\left[Y^2\right] - \mathbb{E}[Y]^2 \leq \mathbb{E}\left[Y^2\right].$$

Then by Lemma 2.20, Fubini's theorem Theorem 1.72, and the previous lemma,

$$\mathrm{Var}(Y_n) \leq \mathbb{E}\left[Y_n^2\right] = \int_0^{\infty} 2y\mathbb{P}(|Y_n| > y) \leq \int_0^{n} 2y\mathbb{P}(|X_n| > y)\,\mathrm{d}y = \int_0^{n} 2y\mathbb{P}(|X| > y)\,\mathrm{d}y\,.$$

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{\mathrm{Var}(Y_n)}{n^2} &\leq \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[Y_n^2\right]}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \int_0^{n} 2y\mathbb{P}(|X| > y)\,\mathrm{d}y = \int_0^{\infty} \left(\sum_{n=1}^{\infty} \frac{\mathbb{1}_{y<n}}{n^2}\right) 2y\mathbb{P}(|X| > y)\,\mathrm{d}y \\ &\leq \int_0^{\infty} 4\mathbb{P}(|X| > y) = 4\mathbb{E}[|X|] < \infty.\end{aligned}$$

This establishes the lemma. ∎

The general proof strategy from here is to prove the result first for a subsequence and then use monotonicity to control the values in between. Let $\alpha > 1$ and let $k(n) = \lfloor \alpha^n \rfloor$. Let $\varepsilon > 0$. Then by Fubini's Theorem Theorem 1.72 and the previous lemma,

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{T_{k(n)} - \mathbb{E}\left[T_{k(n)}\right]}{k(n)}\right| > \varepsilon\right) &\leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{\mathrm{Var}\left(T_{k(n)}\right)}{k(n)^2} = \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k(n)^2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m) \\ &= \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{n\colon k(n) \geq m} \frac{1}{k(n)^2} = \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{n\colon k(n) \geq m} \frac{1}{\lfloor \alpha^n \rfloor^2} \\ &\leq \frac{4}{\varepsilon^2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{n\colon k(n) \geq m} \frac{1}{\alpha^{2n}} \leq \frac{4}{(1-\alpha^2)\varepsilon^2} \sum_{m=1}^{\infty} \frac{\mathrm{Var}(Y_m)}{m^2} \\ &\leq \frac{4}{(1-\alpha^2)\varepsilon^2} \sum_{m=1}^{\infty} \frac{\mathbb{E}\left[Y_m^2\right]}{m^2} < \infty.\end{aligned}$$

Since ε is arbitrary, by Theorem 2.27 and then Theorem 2.26,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{T_{k(n)} - \mathbb{E}\left[T_{k(n)}\right]}{k(n)}\right| > \varepsilon\right) < \infty \implies \mathbb{P}\left(\left|\frac{T_{k(n)} - \mathbb{E}\left[T_{k(n)}\right]}{k(n)}\right| > \varepsilon \text{ i.o.}\right) = 0 \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{T_{k(n)} - \mathbb{E}\left[T_{k(n)}\right]}{k(n)} = 0.$$

By DCT Theorem 1.59,

$$\lim_{n\to\infty} \mathbb{E}[Y_n] = \mathbb{E}\left[\lim_{n\to\infty} Y_n\right] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right] = \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X] \implies \lim_{n\to\infty} \frac{\mathbb{E}\left[T_{k(n)}\right]}{k(n)} = \mathbb{E}[X].$$

Thus

$$\overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{T_{k(n)} - \mathbb{E}\left[T_{k(n)}\right]}{k(n)} = 0 \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{T_{k(n)}}{k(n)} - \frac{\mathbb{E}\left[T_{k(n)}\right]}{k(n)} = 0 \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{T_{k(n)}}{k(n)} - \mathbb{E}[X] = 0 \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{T_{k(n)}}{k(n)} = \mathbb{E}[X].$$

Since $X_n \geq 0$, $Y_n \geq 0$, so T_n is monotone in n. Now

$$\frac{T_{k(n)}}{k(n+1)} \leq \frac{T_m}{m} \leq \frac{T_{k(n+1)}}{k(n)},$$

and $\lim_{n\to\infty} \frac{k(n+1)}{k(n)} = \lim_{n\to\infty} \frac{\lfloor \alpha^{n+1} \rfloor}{\lfloor \alpha^n \rfloor} = \alpha$. Then

$$\frac{\mathbb{E}[X]}{\alpha} \leq \liminf_{n\to\infty} \frac{T_n}{n} \leq \limsup_{n\to\infty} \frac{T_n}{n} \leq \alpha \mathbb{E}[X].$$

Taking $\alpha \downarrow 1$ shows that $\overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{T_n}{n} = \mu$ so by the first lemma $\overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S_n}{n} = \mu$ as desired. ☐

Corollary 2.31. If $\{X_n\}$ are i.i.d. with $\mathbb{E}[X_n^+] = \infty$ and $\mathbb{E}[X_n^-] < \infty$, then $\overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S_n}{n} = \infty$.

Proof. Let $M > 0$ and $X_n^M = X_n \wedge M$. Then the $\{X_n^M\}$ are i.i.d. and in L^1, so if we let $S_n^M = \sum_{m=1}^n X_m^M$, the general SLLN Theorem 2.30 implies that $\overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S_n^M}{n} = \mathbb{E}\left[X_1^M\right]$. Since $X_n \geq X_n^M$, it follows that

$$\liminf_{n\to\infty} \frac{S_n}{n} \geq \lim_{n\to\infty} \frac{S_n^M}{n} = \mathbb{E}\left[X_1^M\right].$$

The MCT Theorem 1.58 implies

$$\lim_{M\to\infty} \mathbb{E}\left[(X_n^M)^+\right] = \mathbb{E}\left[X_n^+\right] = \infty \implies \lim_{M\to\infty} \mathbb{E}\left[X_n^M\right] = \lim_{M\to\infty} \left(\mathbb{E}\left[(X_n^M)^+\right] - \mathbb{E}\left[(X_n^M)^-\right]\right) = \infty \implies \liminf_{n\to\infty} \frac{S_n}{n} \geq \infty$$

which implies the result. ☐

2.8 Quantitative SLLN

We will chase a more quantitative proof of the SLLN. The first tool is Kolmogorov's Maximal Inequality.

Theorem 2.32 (Kolmogorov's Maximal Inequality). Let $\{X_n\} \in L^2$ be independent with $\mathbb{E}[X_n] = 0$ for all n. Then

$$\mathbb{P}\left(\sup_{m\in[n]} |S_m| \geq \varepsilon\right) \leq \frac{\operatorname{Var}(S_n)}{\varepsilon^2}.$$

Proof. Let

$$A_n = \{n = \inf\{m \in \mathbb{N} : |S_m| \geq \varepsilon\}\}.$$

That is, A_n is the event that n is the first step for which $|S_n| \geq \varepsilon$. Then the $\{A_n\}$ are disjoint. Then

$$\begin{aligned}
\mathbb{E}\left[S_n^2\right] &\geq \mathbb{E}\left[S_n^2 \sum_{m=1}^{\infty} \mathbb{1}_{A_m}\right] = \sum_{m=1}^{\infty} \mathbb{E}\left[S_n^2; A_m\right] \geq \sum_{m=1}^{n} \mathbb{E}\left[S_n^2; A_m\right] = \sum_{m=1}^{n} \mathbb{E}\left[\left(S_m^2 + 2S_m(S_n - S_m) + (S_n - S_m)^2\right); A_m\right] \\
&= \sum_{m=1}^{n} \mathbb{E}\left[S_m^2; A_m\right] + 2\sum_{m=1}^{n} \mathbb{E}[S_m(S_n - S_m); A_m] + \sum_{m=1}^{n} \mathbb{E}\left[(S_n - S_m)^2; A_m\right] \\
&\geq \sum_{m=1}^{n} \mathbb{E}\left[S_m^2; A_m\right] + 2\sum_{m=1}^{n} \mathbb{E}[S_m(S_n - S_m); A_m]
\end{aligned}$$

Since

$$\mathbb{E}[S_m] = \mathbb{E}[S_n] \implies \mathbb{E}[S_n - S_m] = 0 \quad \text{and} \quad S_n - S_m \in \sigma(X_{m+1}, \ldots, X_n) \perp\!\!\!\perp \sigma(X_1, \ldots, X_m) \ni S_m, \mathbb{1}_{A_m},$$

we have

$$\begin{aligned}
\mathbb{E}\left[S_n^2\right] &\geq \sum_{m=1}^{n} \mathbb{E}\left[S_m^2; A_m\right] + 2\sum_{m=1}^{n} \mathbb{E}[S_m(S_n - S_m); A_m] = \sum_{m=1}^{n} \mathbb{E}\left[S_m^2; A_m\right] + 2\sum_{m=1}^{n} \mathbb{E}[S_m; A_m]\underbrace{\mathbb{E}[S_n - S_m]}_{0} \\
&= \sum_{m=1}^{n} \mathbb{E}\left[S_m^2; A_m\right] \geq \sum_{m=1}^{n} \mathbb{E}\left[\varepsilon^2; A_m\right] = \sum_{m=1}^{n} \varepsilon^2 \mathbb{E}[\mathbb{1}_{A_m}] = \sum_{m=1}^{n} \varepsilon^2 \mathbb{P}(A_m) = \varepsilon^2 \sum_{m=1}^{n} \mathbb{P}(A_m) \geq \varepsilon^2 \mathbb{P}\left(\bigcup_{m=1}^{n} A_m\right) \\
&= \varepsilon^2 \mathbb{P}\left(\sup_{m \in [n]} |S_m| \geq \varepsilon\right)
\end{aligned}$$

Since $\mathbb{E}[S_n] = 0$, $\operatorname{Var}(S_n) = \mathbb{E}\left[S_n^2\right]$. Then

$$\varepsilon^2 \mathbb{P}\left(\sup_{m \in [n]} |S_m| \geq \varepsilon\right) \leq \mathbb{E}\left[S_n^2\right] = \operatorname{Var}(S_n) \implies \mathbb{P}\left(\sup_{m \in [n]} |S_k| \geq \varepsilon\right) \leq \frac{\operatorname{Var}(S_n)}{\varepsilon^2}.$$

This proves the claim. $\square$

Theorem 2.33. Suppose $\{X_n\} \in L^2$ have $\mathbb{E}[X_n] = 0$ for all n. If $\sum_{n=1}^{\infty} \operatorname{Var}(X_n) < \infty$ then S_n converges almost surely.

Proof. From Kolmogorov's Maximal Inequality Theorem 2.32,

$$\mathbb{P}\left(\sup_{m \in [M,N]} |S_m - S_M| > \varepsilon\right) \leq \frac{\operatorname{Var}(S_N - S_M)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \sum_{m=M+1}^{N} \operatorname{Var}(X_m).$$

Taking $N \to \infty$,

$$\mathbb{P}\left(\sup_{m \geq M} |S_m - S_M| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{m=M+1}^{\infty} \operatorname{Var}(X_m) \to \infty \quad \text{as } M \to \infty.$$

Let $w_M = \sup_{m,n \geq M} |S_m - S_n|$. Then $w_m \to 0$ as $M \to \infty$ and

$$\mathbb{P}(w_M > 2\varepsilon) \leq \mathbb{P}\left(\sup_{m \geq M} |S_m - S_M| > \varepsilon\right) \to 0 \quad \text{as } M \to \infty$$

so $\lim_{n\to\infty}^{\text{a.s.}} w_m = 0$. If $w_M(\omega) \to 0$ then $S_n(\omega)$ is a Cauchy sequence and thus converges, so $\lim_{n\to\infty} S_n(\omega)$ exists. Thus $\lim_{n\to\infty} S_n$ exists almost surely. $\square$

Proposition 2.34. If $\{X_n\} \in L^1$ are i.i.d. with $\mathbb{E}[X_n] = 0$ for all n, then $\displaystyle\sum_{m=1}^{n} \frac{X_m}{m}$ converges almost surely as $n \to \infty$.

Proof. Let $Y_n = X_n \mathbb{1}_{|X_n| \leq n}$. Then by the third lemma in the proof of Theorem 2.30,

$$\sum_{n=1}^{\infty} \mathrm{Var}\left(\frac{Y_n}{n}\right) = \sum_{n=1}^{\infty} \frac{\mathrm{Var}(Y_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[Y_n^2\right]}{n^2} < 4\mathbb{E}[|X_1|] < \infty.$$

Thus by the previous theorem, $\sum_{m=1}^{n} \frac{Y_m}{m}$ converges almost surely. Since $\{X_n\} \in L^1$,

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > n) \leq \int_0^{\infty} \mathbb{P}(|X_1| > t)\, \mathrm{d}t = \mathbb{E}[|X|] < \infty.$$

By Borel-Cantelli Theorem 2.27,

$$\mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0 \implies \mathbb{P}(X_n = Y_n \text{ ev.}) = 1 \implies \sum_{n=1}^{\infty} \frac{X_n}{n} \text{ converges a.s.}$$

□

The last ingredient in the proof of the quantitative SLLN is a particular real-analytic statement about convergence of series.

Theorem 2.35 (Kronecker's Lemma). If $\{a_n\}$ and $\{x_n\}$ are such that $\lim_{n\to\infty} a_n = \infty$ and $\displaystyle\sum_{n=1}^{\infty} \frac{x_n}{a_n}$ converges then

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{m=1}^{n} x_m = 0.$$

Proof. Let $a_0 = b_0 = 0$, and for $m \geq 1$, let $b_m = \sum_{k=1}^{m} \frac{x_k}{a_k}$, and let $b_\infty = \lim_{m\to\infty} b_m$. Then

$$\begin{aligned} x_m &= a_m(b_m - b_{m-1}) \\ \frac{1}{a_n}\sum_{m=1}^{n} x_m &= \frac{1}{a_n}\sum_{m=1}^{n} a_m(b_m - b_{m-1}) = \frac{1}{a_n}\left(\sum_{m=1}^{n} a_m b_m - \sum_{m=1}^{n} a_m b_{m-1}\right) \\ &= \frac{1}{a_n}\left(a_n b_n + \sum_{m=2}^{n} a_{m-1} b_{m-1} - \sum_{m=1}^{n} a_m b_{m-1}\right) = b_n - \sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} b_{m-1}. \end{aligned}$$

We claim that the latter term converges to b_∞. Indeed, by hypothesis, $\lim_{n\to\infty} b_n = b_\infty < \infty$. Then $B = \sup_n |b_n| < \infty$. Fix $\varepsilon > 0$. Pick M such that $|b_m - b_\infty| < \frac{\varepsilon}{2}$ for $m \geq M$. Pick N such that $\frac{a_M}{a_n} < \frac{\varepsilon}{4B}$ for $n \geq N$. Then for $n \geq N$ we have

$$\left|\sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} b_{m-1} - b_\infty\right| \leq \sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} |b_{m-1} - b_\infty| \leq 2\frac{a_M}{a_n} B + \frac{\varepsilon(a_n - a_M)}{2a_n} < \varepsilon.$$

Since ε is arbitrary,

$$\lim_{n\to\infty} \sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} b_{m-1} = b_\infty.$$

And so

$$\lim_{n\to\infty} \frac{1}{a_n}\sum_{m=1}^{n} x_m = \lim_{n\to\infty}\left(b_n - \sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} b_{m-1}\right) = \lim_{n\to\infty} b_n - \lim_{n\to\infty} \sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} b_{m-1} = b_\infty - b_\infty = 0.$$

This proves the claim. □

Now the proof of the SLLN is quite straightforward.

Theorem 2.36 (SLLN). Let $\{X_n\} \in L^1$ be i.i.d. with $\mathbb{E}[X_1] = \mu$. Then $\displaystyle\overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{S_n}{n} = \mu$.

Proof. Let

$$Y_n = X_n \mathbb{1}_{|X_n| \le n} \quad \text{and} \quad T_n = \sum_{m=1}^{n} Y_m.$$

By the first lemma of Theorem 2.30 it suffices to show that $\lim^{\text{a.s.}}_{n\to\infty} \frac{T_n}{n} = \mu$. Let $Z_n = Y_n - \mathbb{E}[Y_n]$, so that $\mathbb{E}[Z_n] = 0$. By Proposition 2.34, $\sum_{n=1}^{\infty} \frac{Z_n}{n}$ converges almost surely. By Kronecker's Lemma Theorem 2.35,

$$0 = \lim^{\text{a.s.}}_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} Z_m = \lim^{\text{a.s.}}_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n}(Y_m - \mathbb{E}[Y_m]) = \lim^{\text{a.s.}}_{n\to\infty}\left(\frac{T_n}{n} - \frac{1}{n}\sum_{m=1}^{n}\mathbb{E}[Y_m]\right).$$

By DCT Theorem 1.59,

$$\lim_{n\to\infty}\mathbb{E}[Y_n] = \lim_{n\to\infty}\mathbb{E}[X_n; |X_n| \le n] = \lim_{n\to\infty}\mathbb{E}[X; |X| \le n] = \mathbb{E}\left[\lim_{n\to\infty} X; |X| \le n\right] = \mathbb{E}[X] = \mu.$$

Thus

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\mathbb{E}[Y_m] = \lim_{n\to\infty}\mathbb{E}[Y_n] = \mu \implies \lim^{\text{a.s.}}_{n\to\infty}\frac{T_n}{n} - \mu = 0 \implies \lim^{\text{a.s.}}_{n\to\infty}\frac{T_n}{n} = \mu.$$

Thus $\lim^{\text{a.s.}}_{n\to\infty}\frac{S_n}{n} = \mu$. □

2.9 Growth Rates Revisited

We will now approach the problem of growth rates using the more quantitative SLLN Theorem 2.36.

Theorem 2.37. Let $\{X_n\} \in L^2$ be i.i.d. with $\mathbb{E}[X_n] = 0$ and $\mathbb{E}\left[X_n^2\right] = \sigma^2$ for all n. Then for every $\varepsilon > 0$,

$$\lim^{\text{a.s.}}_{n\to\infty}\frac{S_n}{n^{1/2}\log(n)^{1/2+\varepsilon}} = 0.$$

Proof. Let $a_n = n^{1/2}\log(n)^{1/2+\varepsilon}$ for $n \ge 2$ and $a_1 > 0$. Then

$$\sum_{n=1}^{\infty}\operatorname{Var}\left(\frac{X_n}{a_n}\right) = \sum_{n=1}^{\infty}\frac{\operatorname{Var}(X_n)}{a_n^2} = \sigma^2\sum_{n=1}^{\infty}\frac{1}{a_n^2} = \sigma^2\left(\frac{1}{a_1^2} + \sum_{n=2}^{\infty}\frac{1}{a_n^2}\right) = \sigma^2\left(\frac{1}{a_1^2} + \sum_{n=2}^{\infty}\frac{1}{n\log(n)^{1+2\varepsilon}}\right) < \infty.$$

Then by Theorem 2.33, $\sum_{n=1}^{\infty}\frac{X_n}{a_n}$ converges almost surely, so by Kronecker's lemma Theorem 2.35, $\lim^{\text{a.s.}}_{n\to\infty}\frac{S_n}{a_n} = 0$. □

Theorem 2.38. If $\{X_n\} \in L^p$ for $p \in (1,2)$, with $\mathbb{E}[X_n] = 0$ for all n, then $\lim\limits^{\text{a.s.}}_{n\to\infty}\frac{S_n}{n^{1/p}} = 0$.

Proof. Let $Y_n = X_n\mathbb{1}_{|X_n|\le n^{1/p}}$ and $T_n = \sum_{m=1}^{n} Y_m$. Then

$$\sum_{n=1}^{\infty}\mathbb{P}(X_n \ne Y_n) = \sum_{n=1}^{\infty}\mathbb{P}(|X_n|^p > n) = \sum_{n=1}^{\infty}\mathbb{P}(|X_1|^p > n) \le \mathbb{E}[|X_1|^p] < \infty.$$

By the Borel-Cantelli lemma Theorem 2.27,

$$\mathbb{P}(X_n \ne Y_n \text{ i.o.}) = 0 \implies \mathbb{P}(X_n = Y_n \text{ ev.}) = 1 \implies \frac{T_n}{n^{1/p}} \text{ converges} \iff \frac{S_n}{n^{1/p}} \text{ converges}.$$

Using

$$\operatorname{Var}(Y_m) = \mathbb{E}\left[Y_m^2\right] - \mathbb{E}[Y_m]^2 \le \mathbb{E}\left[Y_m^2\right],$$

the power estimate Lemma 2.20, and Fubini's theorem Theorem 1.72, we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \operatorname{Var}\left(\frac{Y_n}{n^{1/p}}\right) &\leq \sum_{n=1}^{\infty} \mathbb{E}\left[\left(\frac{Y_n}{n^{1/p}}\right)^2\right] = \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[Y_n^2\right]}{n^{2/p}} = \sum_{n=1}^{\infty} \frac{1}{n^{2/p}} \int_0^{\infty} 2y\mathbb{P}(|Y_n| > y)\,\mathrm{d}y \\
&= \sum_{n=1}^{\infty} \frac{2}{n^{2/p}} \int_0^{\infty} y\mathbb{P}(|Y_n| > y)\,\mathrm{d}y = \sum_{n=1}^{\infty} \frac{2}{n^{2/p}} \int_0^{\infty} y\mathbb{P}\left(|X_n|\mathbb{1}_{|X_n| \leq n^{1/p}} > y\right)\mathrm{d}y \\
&= \sum_{n=1}^{\infty} \frac{2}{n^{2/p}} \int_0^{n^{1/p}} y\mathbb{P}(|X_n| > y)\,\mathrm{d}y = \sum_{n=1}^{\infty} \frac{2}{n^{2/p}} \int_0^{n^{1/p}} y\mathbb{P}(|X_1| > y)\,\mathrm{d}y \\
&= \sum_{n=1}^{\infty} \frac{2}{n^{2/p}} \sum_{m=1}^{n} \int_{(m-1)^{1/p}}^{m^{1/p}} y\mathbb{P}(|X_1| > y)\,\mathrm{d}y = \sum_{m=1}^{\infty} \int_{(m-1)^{1/p}}^{m^{1/p}} 2y\mathbb{P}(|X_1| > y) \sum_{n=m}^{\infty} \frac{1}{n^{2/p}}\,\mathrm{d}y \\
&\leq \sum_{m=1}^{\infty} \int_{(m-1)^{1/p}}^{m^{1/p}} 2y\mathbb{P}(|X_1| > y)\frac{p}{2-p}(m-1)^{(p-2)/p}\,\mathrm{d}y \leq \sum_{m=1}^{\infty} \int_{(m-1)^{1/p}}^{m^{1/p}} 2y\mathbb{P}(|X_1| > y)(Cy^{p-2})\,\mathrm{d}y \\
&= C\sum_{m=1}^{\infty} \int_{(m-1)^{1/p}}^{m^{1/p}} 2y^{p-1}\mathbb{P}(|X_1| > y)\,\mathrm{d}y = \frac{2C}{p} \sum_{m=1}^{\infty} \int_{(m-1)^{1/p}}^{m^{1/p}} py^{p-1}\mathbb{P}(|X_1| > y)\,\mathrm{d}y \\
&= \frac{2C}{p} \int_0^{\infty} py^{p-1}\mathbb{P}(|X_1| > y)\,\mathrm{d}y = \frac{2C}{p}\mathbb{E}[|X_1|] < \infty.
\end{aligned}$$

Let $\mu_n = \mathbb{E}[Y_n]$. Thus by Theorem 2.33 and Kronecker's lemma Theorem 2.35,

$$\overset{\text{a.s.}}{\lim_{n\to\infty}} \sum_{m=1}^{n} \frac{Y_m - \mu_m}{m^{1/p}} = 0 \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} n^{-1/p} \sum_{m=1}^{n} (Y_m - \mu_m) = 0.$$

To estimate μ_m, note that

$$\mathbb{E}[X_m] = \mathbb{E}\left[X_m\left(\mathbb{1}_{|X_m| \leq m^{1/p}} + \mathbb{1}_{|X_m| > m^{1/p}}\right)\right] = \mathbb{E}\left[X_m; |X_m| \leq m^{1/p}\right] + \mathbb{E}\left[X_m; |X_m| > m^{1/p}\right] = \mu_m + \mathbb{E}\left[X_m; |X_m| > m^{1/p}\right]$$

and so

$$\mathbb{E}[X_m] = 0 \implies \mu_m = -\mathbb{E}\left[X_m; |X_m| > m^{1/p}\right].$$

Thus

$$\begin{aligned}
|\mu_m| &= \left|-\mathbb{E}\left[X_m; |X_m| > m^{1/p}\right]\right| = \left|\mathbb{E}\left[X_m; |X_m| > m^{1/p}\right]\right| \leq \mathbb{E}\left[|X_m|; |X_m| > m^{1/p}\right] = m^{1/p}\mathbb{E}\left[\frac{|X_m|}{m^{1/p}}; |X_m| > m^{1/p}\right] \\
&\leq m^{1/p}\mathbb{E}\left[\left(\frac{|X_m|}{m^{1/p}}\right)^p; |X_m| > m^{1/p}\right] = m^{(1/p)-1}\mathbb{E}\left[|X_m|^p; |X_m| > m^{1/p}\right] = m^{(1/p)-1}\mathbb{E}\left[|X_1|^p; |X_1| > m^{1/p}\right].
\end{aligned}$$

Then

$$\sum_{m=1}^{n} m^{(1/p)-1} \leq Cn^{1/p} \quad \text{and} \quad \lim_{m\to\infty} \mathbb{E}\left[|X_1|^p; |X_1| > m^{1/p}\right] = 0 \implies \lim_{n\to\infty} n^{-1/p} \sum_{m=1}^{n} \mu_m = 0.$$

Thus

$$0 = \overset{\text{a.s.}}{\lim_{n\to\infty}} n^{-1/p} \sum_{m=1}^{n} (Y_m - \mu_m) = \overset{\text{a.s.}}{\lim_{n\to\infty}} \left(n^{-1/p} \sum_{m=1}^{n} Y_m - n^{-1/p} \sum_{m=1}^{n} \mu_m\right)$$

which implies

$$\overset{\text{a.s.}}{\lim_{n\to\infty}} n^{-1/p} \sum_{m=1}^{n} Y_m = \lim_{n\to\infty} n^{-1/p} \sum_{m=1}^{n} \mu_m = 0.$$

Thus $\lim^{\text{a.s.}}_{n\to\infty} \frac{S_n}{n^{1/p}} = 0$ as desired. $\square$

Theorem 2.39 (Glivenko-Cantelli Theorem). Let $\{X_n\}$ be i.i.d. with distribution F and let

$$F_n(x) = \frac{1}{n} \sum_{m=1}^{n} \mathbb{1}_{X_m \leq x}.$$

Then
$$\lim_{n\to\infty} \overset{\text{a.s.}}{\sup_{x\in\mathbb{R}}} |F_n(x) - F(x)| = 0.$$

Proof. Since F is a distribution function, F is non-decreasing, right-continuous, $\lim_{x\to\infty} F(x) = 1$, and $\lim_{x\to-\infty} F(x) = 0$.

Fix x and let $Y_n = \mathbb{1}_{X_n \leq x}$. Then $\{Y_n\}$ are i.i.d. with
$$\mathbb{E}[Y_n] = \mathbb{P}(X_n \leq x) = F(x).$$

By Theorem 2.30,
$$\overset{\text{a.s.}}{\lim_{n\to\infty}} F_n(x) = \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{1}{n}\sum_{m=1}^{n} Y_m = F(x).$$

Fix again x and let $Z_n = \mathbb{1}_{X_n < x}$. Then $\{Z_n\}$ are i.i.d. with
$$\mathbb{E}[Z_n] = \mathbb{P}(X_n < x) = F(x-) = \lim_{y\uparrow x} F(y).$$

Thus Theorem 2.30 implies that
$$\overset{\text{a.s.}}{\lim_{n\to\infty}} F_n(x-) = \overset{\text{a.s.}}{\lim_{n\to\infty}} \frac{1}{n}\sum_{m=1}^{n} Z_m = F(x-).$$

We divide up the real line and look at pointwise a.s. convergence on the partition points. For a fixed j, let $i \in [j-1]$ and $x_{ij} = \inf\left\{y\colon F(y) \geq \frac{i}{j}\right\}$. For notation, set $x_{0j} = -\infty$ and $x_{jj} = \infty$. Then the pointwise a.s. convergence of $F_n(x)$ and $F_n(x-)$ imply that there is an $N_j(\omega)$ such that, for all $i \in \{0, \ldots, j\}$,
$$|F_n(x_{ij}) - F(x_{ij})| < \frac{1}{j} \quad \text{and} \quad |F_n(x_{ij}-) - F(x_{ij}-)| < \frac{1}{j} \quad \text{for all } n \geq N_j(\omega).$$

If $x \in \left(x_{(i-1)j}, x_{ij}\right)$ for $1 \leq i \leq j$ and $n \geq N_k(\omega)$, then using the monotonicity of F_n and F and that $F(x_{ij}-) - F(x_{(i-1)j}) \leq \frac{1}{j}$, we have
$$\begin{aligned} F_n(x) &\leq F_n(x_{ij}-) \leq F(x_{ij}-) + \frac{1}{j} \leq F(x_{(i-1)j}) + \frac{2}{j} \leq F(x) + \frac{2}{j} \\ F_n(x) &\geq F_n(x_{(i-1)j}) \geq F(x_{(i-1)j}) - \frac{1}{j} \geq F(x_{ij}-) - \frac{2}{j} \geq F(x) - \frac{2}{j} \end{aligned}$$

so that
$$\begin{aligned} |F_n(x) - F(x)| \leq \frac{2}{j} \quad \text{for a.s. } x \text{ and } n \geq N_j(\omega) \quad &\implies \overset{\text{a.s.}}{\sup_{x\in\mathbb{R}}} |F_n(x) - F(x)| \leq \frac{2}{j} \quad \text{for all } n \geq N_j(\omega) \\ &\implies \lim_{n\to\infty} \overset{\text{a.s.}}{\sup_{x\in\mathbb{R}}} |F_n(x) - F(x)| \leq \frac{2}{j}. \end{aligned}$$

Thus setting j arbitrarily high has
$$\lim_{n\to\infty} \overset{\text{a.s.}}{\sup_{x\in\mathbb{R}}} |F_n(x) - F(x)| = 0,$$

so the theorem is proved. $\square$

3 Central Limit Theorems

3.1 Convergence in Distribution

We will start by discussing the type of convergence that appears in the central limit theorem and explore some of its properties.

Some simple notation will help us first.

Remark. If $f\colon X \to Y$ is continuous then define

$$\mathcal{D}_f = \{x \in X \colon f \text{ is discontinuous at } x\}$$

as the discontinuity set of f. And $\mathcal{C}_f = X \setminus \mathcal{D}_f$ is the continuity set of f.

Proposition 3.1. If $X \in L^0$ is a random variable and $F(x) = \mathbb{P}(X \leq x)$ is its distribution function then $\mathcal{D}_F$ is countable.

Definition 3.2 (Convergence in Distribution). A sequence of distribution functions $\{F_n\}$ is said to **converge in distribution** to a limit F (written $\overset{d}{\lim\limits_{n\to\infty}} F_n = F$, although it is also written $\overset{w}{\lim\limits_{n\to\infty}} F_n = F$) if

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{for all } x \in \mathcal{C}_F.$$

A sequence of random variables $\{X_n\} \in L^0$ converges in distribution to a random variable X if the distribution functions of X_n converge in distribution to the distribution function of X (written $\overset{d}{\lim\limits_{n\to\infty}} X_n = X$).

Remark. The distribution function of a random variable X is $F\colon \mathbb{R} \to \mathbb{R}$ regardless of the space on which X is defined. Thus $\{X_n\}$ and X can be defined on different spaces and still have $\overset{d}{\lim\limits_{n\to\infty}} X_n = X$.

Example 3.3. Take any random variable $X \in L^0$ with distribution F. Let $X_n = X + \frac{1}{n}$ have distribution F_n. Then for $x \in \mathcal{C}_F$,

$$F_n(x) = F\left(x - \frac{1}{n}\right) \quad \text{and} \quad \lim_{y\to x} F(y) = F(x) \implies \lim_{n\to\infty} F\left(x - \frac{1}{n}\right) = \lim_{n\to\infty} F_n(x) = F(x).$$

Thus $\overset{d}{\lim\limits_{n\to\infty}} X_n = X$.

Theorem 3.4. If $\{X_n\}, X \in L^0$ are random variables then

$$\overset{p}{\lim_{n\to\infty}} X_n = X \implies \overset{d}{\lim_{n\to\infty}} X_n = X.$$

Proof. Suppose F_n is the distribution function of X_n and F is the distribution function of X.

If $X, Y \in L^0$ and $x \in \mathbb{R}$, then for any $\varepsilon > 0$,

$$F_Y(x) = \mathbb{P}(Y \leq x) \leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|Y - X| > \varepsilon) = F_X(x + \varepsilon) + \mathbb{P}(|Y - X| > \varepsilon).$$

Let $x \in \mathcal{C}_F$. Then by the previous statement

$$\varlimsup_{n\to\infty} F_n(x) \leq \varlimsup_{n\to\infty} \left(F(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)\right) = F(x + \varepsilon).$$

And

$$F(x-\varepsilon) = \lim_{n\to\infty} F(x-\varepsilon) \le \lim_{n\to\infty} (F_n(x) + \mathbb{P}(|X_n - X| > \varepsilon)) = \lim_{n\to\infty} F_n(x).$$

Thus

$$F(x-\varepsilon) \le \lim_{n\to\infty} F_n(x) \le F(x+\varepsilon) \quad \text{for all } \varepsilon > 0 \text{ and } x \in \mathcal{C}_F \quad \implies F(x) = \lim_{n\to\infty} F_n(x) \quad \text{for all } x \in \mathcal{C}_F.$$

Thus $\lim^d_{n\to\infty} X_n = X$. □

Example 3.5. Suppose $X \sim \mathrm{Bern}(\pm 1)$ and $X_n = -X$. Then $\overset{d}{\lim_{n\to\infty}} X_n = X$ but $\overset{p}{\lim_{n\to\infty}} X_n \neq X$.

Theorem 3.6 (Skorokhod Embedding Theorem). Suppose $\{F_n\}$ are distribution functions such that $\overset{d}{\lim_{n\to\infty}} F_n = F$. Then there exists a probability space $(\Omega, \Sigma, \mathbb{P})$ and random variables $\{X_n\}, X \in L^0(\Omega, \Sigma, \mathbb{P})$ such that $X_n \sim F_n$ and $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$.

Remark. Recall we did a similar thing; when given a distribution F we constructed a random variable with distribution F.

Proof of Theorem 3.6. Let $(\Omega, \Sigma, \mathbb{P}) = ([0,1], \mathcal{B}_{[0,1]}, \text{uniform measure})$, and let $X_n(\omega) = \sup\{y \colon F_n(y) < \omega\}$. Then X_n has distribution F_n. We will now show that $X_n(\omega) \to X(\omega)$ for a co-countable set of ω.

Let $a_\omega = \sup\{y \colon F(y) < \omega\}$, $b_\omega = \inf\{y \colon F(y) > \omega\}$, and $\Omega_0 = \{\omega \colon a_\omega = b_\omega\}$. Then $\Omega \setminus \Omega_0$ is countable since the (a_ω, b_ω) are disjoint and each nonempty interval contains a different rational number. If $\omega \in \Omega_0$ then $F(y) < \omega$ for $y < X_n(\omega)$ and $F(z) > \omega$ for $z > X_n(\omega)$. To prove that $X_n(\omega) \to X(\omega)$ for $\omega \in \Omega_0$ there are two things to show:

Claim 1. $\liminf_{n\to\infty} X_n(\omega) \ge X(\omega)$.

> *Proof.* Let $y \in \mathcal{C}_F$ have $y < X(\omega)$. Since $\omega \in \Omega_0$, $F(y) < \omega$ and if n is sufficiently large $F_n(y) < \omega$, i.e., $X_n(\omega) \ge y$. This holds for all such y, so the result holds. ■

Claim 2. $\limsup_{n\to\infty} X_n(\omega) \le X(\omega)$.

> *Proof.* Let $y \in \mathcal{C}_F$ have $y > X(\omega)$. Since $\omega \in \Omega_0$, $F(y) > \omega$ and if n is sufficiently large $F_n(y) > \omega$, i.e., $X_n(\omega) \le y$. This holds for all such y, so the result holds. □

Thus $X_n(\omega) \to X(\omega)$ for $\omega \in \Omega_0$, and Ω_0 is co-countable, which completes the proof. □

Proposition 3.7. Random variables $\{X_n\}, X \in L^0$ have $\overset{d}{\lim_{n\to\infty}} X_n = X$ if and only if for any bounded continuous function g, $\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$.

Proof. By Theorem 3.6, let Y_n have the same distribution as X_n and converge a.s. to Y which has the same distribution as X. Since g is continuous, $\lim^{\text{a.s.}}_{n\to\infty} g(Y_n) = g(Y)$. The BCT Theorem 1.56 implies

$$\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \lim_{n\to\infty} \mathbb{E}[g(Y_n)] = \mathbb{E}[g(Y)] = \mathbb{E}[g(X)].$$

To prove the converse let

$$g_{x,\varepsilon}9y) = \begin{cases} 1 & y \le x \\ 0 & y \ge x + \varepsilon \\ \text{linear} & x \le y \le x + \varepsilon \end{cases}.$$

Since $g_{x,\varepsilon}(y) = 1$ for $y \leq x$, $g_{x,\varepsilon}$ is continuous, and $g_{x,\varepsilon}(y) = 0$ for $y > x + \varepsilon$,

$$\limsup_{n\to\infty} \mathbb{P}(X_n \leq x) \leq \limsup_{n\to\infty} \mathbb{E}[g_{x,\varepsilon}(X_n)] = \mathbb{E}[g_{x,\varepsilon}(X)] \leq \mathbb{P}(X \leq x + \varepsilon).$$

Letting $\varepsilon \to 0$ gives

$$\limsup_{n\to\infty} \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x).$$

The last conclusion is valid for any x. To get the other direction, we observe

$$\liminf_{n\to\infty} \mathbb{P}(X_n \leq x) \geq \liminf_{n\to\infty} \mathbb{E}[g_{x-\varepsilon,\varepsilon}(X_n)] = \mathbb{E}[g_{x-\varepsilon,\varepsilon}(X)] \geq \mathbb{P}(X \leq x - \varepsilon).$$

Lettting $\varepsilon \to 0$ gives that for $x \in \mathcal{C}_F$,

$$\liminf_{n\to\infty} \mathbb{P}(X_n \leq x) \geq \mathbb{P}(X < x) = \mathbb{P}(X \leq x).$$

The results combine to give the desired result. $\square$

Theorem 3.8 (Continuous Mapping Theorem). Let $\{X_n\}, X \in L^0$ be random variables such that $\overset{d}{\lim_{n\to\infty}} X_n = X$ and g be a measurable function such that $\mathbb{P}(X \in \mathcal{D}_g) = 0$. Then $\overset{d}{\lim_{n\to\infty}} g(X_n) = g(X)$.

Proof. First we note that $\mathcal{C}_g$ is a Borel set. We use Theorem 3.6 to find $Y_n \overset{d}{=} X_n$ and $Y \overset{d}{=} Y$ such that $\lim^{\text{a.s.}}_{n\to\infty} Y_n = Y$. If f is continuous then $\mathcal{D}_{f\circ g} \subseteq \mathcal{D}_g$, so $\mathbb{P}(Y \in \mathcal{D}_{f\circ g}) = 0$. Thus $\lim^{\text{a.s.}}_{n\to\infty} f(g(Y_n)) = f(g(Y))$. If f is bounded then $f \circ g$ is bounded, so the BCT Theorem 1.56 implies

$$\lim_{n\to\infty} \mathbb{E}[f(g(X_n))] = \lim_{n\to\infty} \mathbb{E}[f(g(Y_n))] = \lim_{n\to\infty} \mathbb{E}[(f\circ g)(Y_n)] = \mathbb{E}[(f\circ g)(Y)] = \mathbb{E}[f(g(Y))] = \mathbb{E}[f(g(X))].$$

Thus by the previous proposition $\lim^{d}_{n\to\infty} g(X_n) = g(X)$. $\square$

Remark 3.9. Let $\{X_n\}, X \in L^0$ have $\overset{d}{\lim_{n\to\infty}} X_n = X$. Further suppose $X_n \sim F_n$ and $X \sim F$. Define *push-forward* measures μ_n and μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ by $\mu_n = \mathbb{P} \circ X_n^{-1}$ and $\mu = \mathbb{P} \circ X^{-1}$. Then we write

$$\overset{d}{\lim_{n\to\infty}} \mu_n = \mu \quad \text{if and only if} \quad \overset{d}{\lim_{n\to\infty}} X_n = X \quad \text{if and only if} \quad \overset{d}{\lim_{n\to\infty}} F_n = F.$$

Theorem 3.10 (Portmanteau's Lemma). The following are equivalent:

1. $\overset{d}{\lim_{n\to\infty}} \mu_n = \mu$.
2. For any open set U, $\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$.
3. For any closed set V, $\limsup_{n\to\infty} \mu_n(V) \leq \mu(V)$.
4. Let $\partial A = \overline{A} \setminus A^\circ$. For any set A such that $\mu(\partial A) = 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.

Proof. The proof follows by a series of equivalences.

Claim 1. (1) implies (2).

Proof. By Theorem 3.6, let $Y_n \overset{d}{=} X_n$, $Y \overset{d}{=} X$, and $\lim_{n\to\infty}^{\text{a.s.}} Y_n = Y$. Since U is open, Fatou's Lemma Theorem 1.57 gives

$$\liminf_{n\to\infty} \mathbb{1}_{Y_n \in U} \geq \mathbb{1}_{Y \in U} \implies \liminf_{n\to\infty} \mu_n(U) = \liminf_{n\to\infty} \mathbb{E}[\mathbb{1}_{Y_n \in U}] \geq \mathbb{E}\left[\liminf_{n\to\infty} \mathbb{1}_{Y_n \in U}\right] = \mathbb{E}[\mathbb{1}_{Y \in U}] = \mu(U),$$

which is (2). ∎

Claim 2. (2) is equivalent to (3).

Proof. This follows easily from the fact that A is open if and only if A^c is closed, and $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$. ∎

Claim 3. (2) and (3) imply (4).

Proof. Let $U = A^\circ$ and $V = \overline{A}$ be the interior and closure of A respectively. Then $\mu(\partial A) = 0$, so

$$\mu(U) = \mu(A) = \mu(V).$$

Then using (2) and (3),

$$\limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(V) \leq \mu(V) = \mu(A) \quad \text{and} \quad \liminf_{n\to\infty} \mu_n(A) \geq \liminf_{n\to\infty} \mu_n(U) \geq \mu(U) = \mu(A).$$

Thus

$$\lim_{n\to\infty} \mu_n(A) = \mu(A)$$

for all A for which $\partial A = 0$. ∎

Claim 4. (4) implies (1).

Proof. Let x be such that $\mu(\{x\}) = 0$, so $x \in \mathcal{C}_F$, and let $A = (-\infty, x]$. ∎

Thus (1) through (4) are equivalent. □

Theorem 3.11 (Helly's Selection Theorem). Given a sequence of distribution functions $\{F_n\}$ there exists a subsequence $\{m(n)\}$ and a right-continuous non-decreasing function F such that $\lim\limits_{n\to\infty} F_{m(n)}(x) = F(x)$ for all $x \in \mathcal{C}_F$.

Remark. The reason F is not a distribution function is because of cases like the following. Suppose $F_n(x) = 0$ for all $n \geq x$. Any subsequential limit is 0. A more sophisticated example is that F_n puts $\frac{1}{2}$ probability at 0 and $\frac{1}{2}$ probability at n. Then all the subsequential limits are the same F which is 0 for $x < 0$ and $\frac{1}{2}$ for $x \geq 0$. We will learn how to patch this to get distribution functions as limits.

Proof of Theorem 3.11. The first step is a diagonal argument. Let $\{q_n\}$ be an enumeration of $\mathbb{Q}$. Since $\{F_n\}$ are distribution functions, $F_m(q_k) \in [0, 1]$ for each k. Then there is an inductively defined subsequence $\{m_k(n)\} \subseteq \{m_{k-1}(n)\}$, with $m_0(n) = n$, such that $\lim_{n\to\infty} F_{m_k(n)}(q_k)$ converges, say to $G(q_k)$. Define $m(n) = m_n(n)$. By construction $\lim_{n\to\infty} F_{m(n)}(q) = G(q)$ for all $q \in \mathbb{Q}$. And G may not be right-continuous but $F(x) = \inf\{G(q) : q \in \mathbb{Q}, q > x\}$ is right-continuous since

$$\begin{aligned}\lim_{x_n \downarrow x} F(x_n) &= \inf\{G(q) : q \in \mathbb{Q}, q > x_n \text{ for all } n \text{ large enough}\} = \inf\{G(q) : q \in \mathbb{Q}, q > x\} \\ &= F(x).\end{aligned}$$

To complete the proof, let $\varepsilon > 0$ and $x \in \mathcal{C}_F$. Then there are $r_1, r_2, s \in \mathbb{Q}$ with $r_1 < r_2 < x < s$ such that

$$F(x) - \varepsilon < F(r_1) \leq F(r_2) \leq F(x) \leq F(s) < F(x) + \varepsilon.$$

Since

$$\lim_{n\to\infty} F_{m(n)}(r_2) = G(r_2) \geq F(r_1) \quad \text{and} \quad \lim_{n\to\infty} F_{m(n)}(s) = G(s) \leq F(s)$$

it follows that if n is large,

$$F(x) - \varepsilon < F_{m(n)}(r_2) \leq F_{m(n)}(x) \leq F_{m(n)}(s) < F(x) + \varepsilon$$

and since ε is arbitrary the result follows. $\square$

So now we want to find when F is a distribution function.

Definition 3.12 (Tightness). A set of measures $\{\mu_\lambda\}$ are **tight** if given $\varepsilon > 0$ there exists a compact set K_ε such that

$$\mu_\lambda(K_\varepsilon) \geq 1 - \varepsilon \quad \text{for all } \lambda \in \Lambda.$$

A set of distribution functions $\{F_\lambda\}$ are tight if, supposing $\{X_\lambda\} \in L^0$ has $X_\lambda \sim F_\lambda$ and $\mu_\lambda = \mathbb{P} \circ X_\lambda^{-1}$, that $\{\mu_\lambda\}$ are tight.

Remark. It's equivalent to say the same for all *but finitely many* $\lambda \in \Lambda$ since supremums or infimums over that set will be finite.

Theorem 3.13. Under the conditions of Theorem 3.11, $\{F_n\}$ are tight if and only if F is a distribution function.

Proof. Suppose $\{F_n\}$ are tight and there is a subsequence $\{m(n)\}$ such that $\lim_{n\to\infty} F_{m(n)}(x) = F(x)$ for all $x \in \mathcal{C}_F$. Let $\varepsilon > 0$. Then there is a compact set $K_\varepsilon \subseteq \mathbb{R}$ such that $\mu_n(K_\varepsilon) \geq 1 - \varepsilon$ for all n. Since K_ε is compact, K_ε is closed and bounded, so there exists $M_\varepsilon > 0$ such that $K_\varepsilon \subseteq [-M_\varepsilon, M_\varepsilon]$, so

$$\mu_n([-M_\varepsilon, M_\varepsilon]) = F(M_\varepsilon) - F(-M_\varepsilon) \geq 1 - \varepsilon \quad \text{for all } n.$$

Let $r, s \in \mathcal{C}_F$ with $r < -M_\varepsilon$ and $s > M_\varepsilon$. Then

$$\lim_{n\to\infty} F_n(r) = F(r) \quad \text{and} \quad \lim_{n\to\infty} F_n(s) = F(s),$$

so

$$\begin{aligned} F(s) - F(r) &= \lim_{n\to\infty} (F_n(s) - F_n(r)) = \lim_{n\to\infty} \left(F_{m(n)}(s) - F_{m(n)}(r)\right) \geq \lim_{n\to\infty} \left(F_{m(n)}(M_\varepsilon) - F_{m(n)}(-M_\varepsilon)\right) \\ &\geq 1 - \varepsilon. \end{aligned}$$

The last result implies

$$\liminf_{x\to\infty} (F(x) - F(-x)) \geq 1 - \varepsilon \quad \text{for all } \varepsilon > 0 \qquad \implies \qquad F(\infty-) - F(-\infty+) = 1.$$

Thus F is a distribution function.

To prove the converse, suppose $\{F_n\}$ are not tight. Then there is an $\varepsilon > 0$ and a subsequence $\{m(n)\} \uparrow \infty$ such that

$$\mu_{m(n)}([-n, n]) = F_{m(n)}(n) - F_{m(n)}(-n) < 1 - \varepsilon \quad \text{for all } n.$$

By passing to a further subsequence $\{m(k(n))\} \subseteq \{m(n)\}$, we can suppose $\lim_{n\to\infty} F_{m(k(n))}(x) = F(x)$ for all $x \in \mathcal{C}_F$. Let $r, s \in \mathcal{C}_F$ with $r < 0 < s$. Then

$$\begin{aligned} F(s) - F(r) &= \lim_{n\to\infty} (F_n(s) - F_n(r)) = \lim_{n\to\infty} \left(F_{m(k(n))}(s) - F_{m(k(n))}(r)\right) < \lim_{n\to\infty} \left(F_{m(k(n))}(k(n)) - F_{m(k(n))}(k(n))\right) \\ &< 1 - \varepsilon. \end{aligned}$$

Letting $s \to \infty$ and $r \to -\infty$, we see that F is not a distribution function. $\square$

3.2 Characteristic Functions

Definition 3.14. Let $X \in L^0$ have distribution function F. Then the **characteristic function** of X is defined as

$$\phi_X(t) = \mathbb{E}\left[e^{itX}\right].$$

Remark. Since $e^{itX} = \cos(tX) + i\sin(tX)$ which is bounded, $\phi_X(t)$ exists for each t, and for $X \notin L^1$.

Proposition 3.15. Let $X \in L^0$. Then

1. $\phi_X(0) = 1$.
2. $\overline{\phi_X(t)} = \phi_X(-t)$.
3. $|\phi_X(t)| \leq 1$.
4. ϕ_X is uniformly continuous.
5. For $a, b \in \mathbb{R}$, $\phi_{aX+b}(t) = \mathrm{e}^{\mathrm{i}bt}\phi_X(at)$.

Proof.

1. We have
$$\phi_X(0) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\cdot 0\cdot X}\right] = \mathbb{E}\left[\mathrm{e}^0\right] = \mathrm{e}^0 = 1.$$
2. We have
$$\overline{\phi_X(t)} = \overline{\mathbb{E}[\mathrm{e}^{\mathrm{i}tX}]} = \mathbb{E}\left[\overline{\mathrm{e}^{\mathrm{i}tX}}\right] = \mathbb{E}\left[\mathrm{e}^{-\mathrm{i}tX}\right] = \phi_X(-t).$$
3. We have by Jensen's inequality on the convex function $(x, y) \mapsto \sqrt{x^2 + y^2}$,
$$|\phi_X(t)| = \left|\mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\right]\right| \leq \mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}tX}\right|\right] = 1.$$
4. We have
$$\begin{aligned}|\phi_X(t+h) - \phi_X(t)| &= \left|\mathbb{E}\left[\mathrm{e}^{\mathrm{i}(t+h)X} - \mathrm{e}^{\mathrm{i}tX}\right]\right| = \left|\mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\left(\mathrm{e}^{\mathrm{i}hX} - 1\right)\right]\right| \\ &\leq \mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}tX}\left(\mathrm{e}^{\mathrm{i}hX} - 1\right)\right|\right] \leq \mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}tX}\right|\left|\mathrm{e}^{\mathrm{i}hX} - 1\right|\right] \\ &= \mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}hX} - 1\right|\right]\end{aligned}$$
so uniform continuity follows by the BCT Theorem 1.56.
5. We write
$$\phi_{aX+b}(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}t(aX+b)}\right] = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}atX}\mathrm{e}^{\mathrm{i}tb}\right] = \mathrm{e}^{\mathrm{i}tb}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}(at)X}\right] = \mathrm{e}^{\mathrm{i}bt}\phi_X(at).$$

□

Theorem 3.16. If $X, Y \in L^0$ are independent then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

Proof. Since $X \perp\!\!\!\perp Y$, $\mathrm{e}^{\mathrm{i}tX} \perp\!\!\!\perp \mathrm{e}^{\mathrm{i}tY}$. Then
$$\phi_{X+Y}(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}t(X+Y)}\right] = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\mathrm{e}^{\mathrm{i}tY}\right] = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\right]\mathbb{E}\left[\mathrm{e}^{\mathrm{i}tY}\right] = \phi_X(t)\phi_Y(t).$$

□

Theorem 3.17. If $X \in L^n$ for n an integer then ϕ_X has a continuous derivative of order n given by
$$\phi_X^{(n)}(t) = \mathbb{E}\left[(\mathrm{i}X)^n\mathrm{e}^{\mathrm{i}tX}\right].$$

Proof. We use induction, with base case $n = 0$ clearly true. Then

$$\begin{aligned}
\left|\phi_X^{(n)}(t) - \mathbb{E}\left[(\mathrm{i}X)^n \mathrm{e}^{\mathrm{i}tX}\right]\right| &= \lim_{h \to 0}\left|\frac{\phi_X^{(n-1)}(t+h) - \phi_X^{(n-1)}(t)}{h} - \mathbb{E}\left[(\mathrm{i}X)^n \mathrm{e}^{\mathrm{i}tX}\right]\right| \\
&= \lim_{h \to 0}\left|\mathbb{E}\left[\frac{(\mathrm{i}X)^{n-1}\mathrm{e}^{\mathrm{i}(t+h)X} - (\mathrm{i}X)^{n-1}\mathrm{e}^{\mathrm{i}tX}}{h} - (\mathrm{i}X)^n \mathrm{e}^{\mathrm{i}tX}\right]\right| \\
&\leq \lim_{h \to 0}\mathbb{E}\left[\left|\frac{(\mathrm{i}X)^{n-1}\mathrm{e}^{\mathrm{i}(t+h)X} - (\mathrm{i}X)^{n-1}\mathrm{e}^{\mathrm{i}tX}}{h} - (\mathrm{i}X)^n \mathrm{e}^{\mathrm{i}tX}\right|\right] \\
&= \lim_{h \to 0}\mathbb{E}\left[\left|(\mathrm{i}X)^n \mathrm{e}^{\mathrm{i}tX}\left(\frac{\mathrm{e}^{\mathrm{i}hX} - 1}{\mathrm{i}hX} - 1\right)\right|\right] \\
&\leq \lim_{h \to 0}\mathbb{E}\left[\left|(\mathrm{i}X)^n \mathrm{e}^{\mathrm{i}tX}\right|\left|\frac{\mathrm{e}^{\mathrm{i}hX} - 1}{\mathrm{i}hX} - 1\right|\right] \\
&\leq \lim_{h \to 0}\mathbb{E}\left[\left|(\mathrm{i}X)^n \mathrm{e}^{\mathrm{i}tX}\right|\sup_{0 \leq u \leq h}\left|\mathrm{e}^{\mathrm{i}uX} - 1\right|\right] = 0
\end{aligned}$$

by the DCT Theorem 1.59. And with this formula $\phi_X^{(n)}$ is clearly continuous. □

Theorem 3.18. For $n \geq 0$ and $x \in \mathbb{R}$,

$$\left|\mathrm{e}^{\mathrm{i}x} - \sum_{m=0}^{n}\frac{(\mathrm{i}x)^m}{m!}\right| \leq \min\left\{\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right\}.$$

Proof. Tedious real-analytic argument. □

Corollary 3.19. For positive integer n and $X \in L^n$,

$$\left|\phi_X(t) - \sum_{m=0}^{n}\mathbb{E}\left[\frac{(\mathrm{i}tX)^m}{m!}\right]\right| \leq \mathbb{E}\left[\min\left\{|tX|^{n+1}, 2|tX|^n\right\}\right].$$

Proof. Using Jensen's inequality and Theorem 3.18, we have

$$\begin{aligned}
\left|\phi_X(t) - \sum_{m=0}^{n}\mathbb{E}\left[\frac{(\mathrm{i}tX)^m}{m!}\right]\right| &= \left|\mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\right] - \sum_{m=0}^{n}\mathbb{E}\left[\frac{(\mathrm{i}tX)^m}{m!}\right]\right| = \left|\mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX} - \sum_{m=0}^{n}\frac{(\mathrm{i}tX)^m}{m!}\right]\right| \\
&\leq \mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}tX} - \sum_{m=0}^{n}\frac{(\mathrm{i}tX)^m}{m!}\right|\right] \leq \mathbb{E}\left[\min\left\{\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\right\}\right] \\
&\leq \mathbb{E}\left[\min\left\{|tX|^{n+1}, 2|tX|^n\right\}\right]
\end{aligned}$$

as desired. □

Theorem 3.20. If $X \in L^2$ then

$$\phi_X(t) = 1 + \mathrm{i}t\mathbb{E}[X] - \frac{t^2\mathbb{E}\left[X^2\right]}{2} + o\left(t^2\right).$$

Proof. By Corollary 3.19, the error term is

$$\left|\phi_X(t) - 1 - \mathrm{i}t\mathbb{E}[X] - \frac{t^2\mathbb{E}\left[X^2\right]}{2}\right| \leq t^2\mathbb{E}\left[\min\left\{|t||X|^3, 2|X|^2\right\}\right].$$

For small t, and in particular as $t \to 0$, we have by the DCT Theorem 1.59,

$$\lim_{t\to 0}\left|\phi_X(t) - 1 - \mathrm{i}t\mathbb{E}[X] - \frac{t^2\mathbb{E}\left[X^2\right]}{2}\right| \leq \lim_{t\to 0} t^2\mathbb{E}\left[\min\left\{|t||X|^3, 2|X|^2\right\}\right] \leq \lim_{t\to 0} t^2\mathbb{E}\left[2|X|^2\right] = 0.$$

□

Corollary 3.21. If $\phi_X(t)$ is twice-differentiable at 0 then $X \in L^2$.

Proof. We have the identity

$$\phi''(0) = \lim_{h\to 0}\frac{\phi'(h) - \phi'(0)}{h} = \lim_{h\to 0}\frac{1}{h}\left(\lim_{h'\to h}\frac{\phi(h') - \phi(h)}{h' - h} - \lim_{h'\to 0}\frac{\phi(h') - \phi(0)}{h'}\right) = \lim_{h\to 0}\frac{\phi_X(h) + \phi_X(-h) - 2\phi_X(0)}{h^2}$$
$$= \lim_{h\to 0}\frac{\phi_X(h) + \phi_X(-h) - 2}{h^2}.$$

We have

$$\frac{\mathrm{e}^{\mathrm{i}hx} + \mathrm{e}^{-\mathrm{i}hx} - 2}{h^2} = -\frac{2(1 - \cos(hx))}{h^2} \leq 0 \quad \text{and} \quad \lim_{h\to 0}\frac{2(1 - \cos(hx))}{h^2} = x^2.$$

Thus by Fatou's lemma Theorem 1.57 and Fubini's theorem Theorem 1.72,

$$\mathbb{E}\left[X^2\right] \leq 2\liminf_{h\to 0}\mathbb{E}\left[\frac{1 - \cos(hX)}{h^2}\right] = -\limsup_{h\to 0}\frac{\phi_X(h) + \phi_X(-h) - 2}{h^2} < \infty.$$

Therefore $X \in L^2$ as desired. □

We now explore a Fourier inversion formula. which obtains a measure from a characteristic function.

Theorem 3.22 (Characteristic Function Inversion Formula). Let $X \in L^0$ with distribution function F and $\mu = \mathbb{P}\circ X^{-1}$ the push-forward of X. Let $a, b \in \mathbb{R}$ with $a < b$. Then

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\phi_X(t)\,\mathrm{d}t = \mu((a,b)) + \frac{1}{2}\mu(\{a,b\}).$$

Proof. Let

$$I_T = \int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\phi_X(t)\,\mathrm{d}t = \int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\right]\mathrm{d}t = \int_{-T}^{T}\mathbb{E}\left[\frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\mathrm{e}^{\mathrm{i}tX}\right]\mathrm{d}t\,.$$

The integrand may look bad near $t = 0$, but observing that

$$\frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} = \int_a^b \mathrm{e}^{-\mathrm{i}ty}\,\mathrm{d}y \leq b - a,$$

we observe that the magnitude of the expectation is bounded by $b - a$. By Fubini's theorem Theorem 1.72, we have

$$I_T = \int_{-T}^{T}\mathbb{E}\left[\frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\mathrm{e}^{\mathrm{i}tX}\right]\mathrm{d}t = \mathbb{E}\left[\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\mathrm{e}^{\mathrm{i}tX}\,\mathrm{d}t\right] = \mathbb{E}\left[\int_{-T}^{T}\frac{\sin(t(X-a))}{t}\,\mathrm{d}t - \int_{-T}^{T}\frac{\sin(t(X-b))}{t}\,\mathrm{d}t\right].$$

Making a substitution,

$$R(\theta, T) = \int_{-T}^{T}\frac{\sin(\theta t)}{t}\,\mathrm{d}t \implies I_T = \mathbb{E}[R(X-a,T) - R(X-b,T)].$$

Letting

$$S(T) = \int_0^T\frac{\sin(x)}{x}\,\mathrm{d}x$$

then for $\theta > 0$ changing variables $t = \frac{x}{\theta}$ gives

$$R(\theta, T) = 2\int_0^{T\theta} \frac{\sin(x)}{x}\,\mathrm{d}x = 2S(T\theta),$$

while for $\theta < 0$, $R(\theta, T) = -R(|\theta|, T)$. Thus we can write the formulas together as

$$R(\theta, T) = 2\,\mathrm{sign}(\theta) S(T|\theta|).$$

It's a simple fact from complex analysis that

$$\lim_{T\to\infty} S(T) = \frac{\pi}{2} \implies \lim_{T\to\infty} R(\theta, T) = \pi\,\mathrm{sign}(\theta) \implies \lim_{T\to\infty}(R(x-a, T) - R(x-b, T)) = \begin{cases} 2\pi & x \in (a,b) \\ \pi & x \in \{a,b\} \\ 0 & x \notin [a,b] \end{cases}.$$

And

$$|R(\theta, T)| \leq 2\sup_y S(y) < \infty,$$

so using the BCT Theorem 1.56 implies

$$\lim_{T\to\infty} \frac{I_T}{2\pi} = \lim_{T\to\infty} \mathbb{E}\left[\frac{R(X-a,T) - R(X-b,T)}{2\pi}\right] = \mathbb{E}\left[\lim_{T\to\infty} \frac{R(X-a,T) - R(X-b,T)}{2\pi}\right] = \mathbb{E}\left[\begin{cases} 1 & X \in (a,b) \\ \frac{1}{2} & X \in \{a,b\} \\ 0 & X \notin [a,b] \end{cases}\right]$$

$$= \mathbb{P}(X \in (a,b)) + \frac{1}{2}\mathbb{P}(X \in \{a,b\}) = \mu((a,b)) + \frac{1}{2}\mu(\{a,b\}).$$

$\square$

Lemma 3.23. Let $X, Y \in L^0$. If $\phi_X(t) = \phi_Y(t)$ then $X \overset{d}{=} Y$.

Proof. We want to show that $\mu_X(B) = \mu_Y(B)$ for $B \in \mathcal{B}_{\mathbb{R}}$; it suffices to show it for $B = (a,b)$ an interval in $\mathbb{R}$. By Theorem 3.22, since $\phi_X = \phi_Y$,

$$\mu_X((a,b)) + \frac{1}{2}(\mu_X(\{a\}) + \mu_X(\{b\})) = \mu_Y((a,b)) + \frac{1}{2}(\mu_Y(\{a\}) + \mu_Y(\{b\})) \quad \text{for all } a, b \in \mathbb{R} \text{ with } a < b \quad .$$

If $\mu_X(\{x\}) = \mu_Y(\{x\}) = 0$ for all $x \in \mathbb{R}$ then $\mu_X((a,b)) = \mu_Y((a,b))$ for $a, b \in \mathbb{R}$. Sending $a \to -\infty$ and $b \to \infty$ shows that $\mu_X(B) = \mu_Y(B)$ for all open intervals B and thus for $B \in \mathcal{B}_{\mathbb{R}}$. Otherwise, the set $A = \{x \in \mathbb{R} \colon \mu_X(\{x\}) > 0 \text{ or } \mu_Y(\{x\}) > 0\}$ is countable (in fact finite), or else μ_X or μ_Y are not σ-finite measures, a contradiction. Then for all $a, b \in A^c$ with $a < b$, $\mu_X((a,b)) = \mu_Y((a,b))$; since A^c is dense and $x \mapsto \mu_X((-\infty, x])$ and $x \mapsto \mu_Y((-\infty, x])$ are right-continuous, $\mu_X(B) = \mu_Y(B)$ for $B \in \mathcal{B}_{\mathbb{R}}$. $\square$

Theorem 3.24. If $\phi_X(t) \in L^1(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \text{Lebesgue measure})$ then μ_X has bounded continuous density

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}tx}\phi_X(t)\,\mathrm{d}t\,.$$

Proof. We observed in the proof of Theorem 3.22 that

$$\left|\frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\right| = \left|\int_a^b \mathrm{e}^{-\mathrm{i}ty}\,\mathrm{d}y\right| \leq |b-a|,$$

so the inversion formula integral converges absolutely and

$$\mu_X((a,b)) + \frac{1}{2}\mu_X(\{a,b\}) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\phi_X(t)\,\mathrm{d}t \leq \frac{b-a}{2\pi}\int_{\mathbb{R}} |\phi_X(t)|\,\mathrm{d}t < \infty.$$

The last result implies μ_X has no point masses, and by Fubini's theorem Theorem 1.72,

$$\begin{aligned}\mu_X((a,b)) &= \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\mathrm{e}^{-\mathrm{i}ta}-\mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\phi_X(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{\mathbb{R}}\left(\int_a^b \mathrm{e}^{-\mathrm{i}ty}\,\mathrm{d}y\right)\phi_X(t)\,\mathrm{d}t \\ &= \int_a^b\left(\frac{1}{2\pi}\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}ty}\phi_X(t)\,\mathrm{d}t\right)\mathrm{d}y\end{aligned}$$

so that μ_X has density

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}tx}\phi_X(t)\,\mathrm{d}t\,.$$

And DCT Theorem 1.59 implies f_X is continuous. □

Theorem 3.25 (Continuity Theorem). Let $\{\mu_n\}$ be probability measures with associated characteristic functions $\{\phi_n\}$.

1. If μ is a probability measure with associated characteristic function ϕ, and $\overset{d}{\lim\limits_{n\to\infty}}\,\mu_n = \mu$, then $\lim\limits_{n\to\infty}\phi_n(t) = \phi(t)$ for all t.

2. If $\phi_n(t)$ converges pointwise to a limit $\phi(t)$ that is continuous at 0, then $\{\mu_n\}$ is tight and converges weakly to the measure μ with characteristic function ϕ.

Remark. To see why continuity of the limit at 0 is needed in (2), let μ_n have a normal distribution with mean 0 and variance n. Then $\phi_n(t) = \exp\left(-nt^2/2\right) \to 0$ for $t \neq 0$, and $\phi_n(0) = 1$ for all n, but the measures do not converge weakly since $\mu_n((-\infty, x]) \to \frac{1}{2}$ for all x.

Proof of Theorem 3.25.

1. The function $f_t(x) = \mathrm{e}^{\mathrm{i}tx}$ is bounded and continuous, so if $X_n \sim \mu_n$ and $X \sim \mu$, then $\lim^d_{n\to\infty} X_n = X$, so
$$\lim_{n\to\infty}\phi_n(t) = \lim_{n\to\infty}\mathbb{E}[f_t(X_n)] = \mathbb{E}[f_t(X)] = \phi(t).$$

2. Our first goal is to prove tightness. Suppose $X_n \sim \mu_n$ and X have characteristic function ϕ. We compute
$$\begin{aligned}\int_{-u}^{u}\left(1-\mathrm{e}^{\mathrm{i}tx}\right)\mathrm{d}t &= 2u - \int_{-u}^{u}(\cos(tx)+\mathrm{i}\sin(tx))\,\mathrm{d}t = 2u - \frac{2\sin(ux)}{x} \\ \frac{1}{u}\int_{-u}^{u}\left(1-\mathrm{e}^{\mathrm{i}tx}\right)\mathrm{d}t &= 2 - \frac{2\sin(ux)}{ux} \\ \mathbb{E}\left[\frac{1}{u}\int_{-u}^{u}\left(1-\mathrm{e}^{\mathrm{i}tX_n}\right)\mathrm{d}t\right] &= \mathbb{E}\left[2-\frac{2\sin(uX_n)}{uX_n}\right] \\ \frac{1}{u}\int_{-u}^{u}\mathbb{E}\left[1-\mathrm{e}^{\mathrm{i}tX_n}\right]\mathrm{d}t &= 2\mathbb{E}\left[1-\frac{\sin(uX_n)}{uX_n}\right] \\ \frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))\,\mathrm{d}t &= 2\mathbb{E}\left[1-\frac{\sin(uX_n)}{uX_n}\right]\end{aligned}$$
To bound the right-hand side, we note that $|\sin(x)| \le |x|$ for all x, so
$$\begin{aligned}2\mathbb{E}\left[1-\frac{\sin(uX_n)}{uX_n}\right] &= 2\int_{\mathbb{R}}\left(1-\frac{\sin(ux)}{ux}\right)\mathrm{d}\mu_n(x) \ge 2\int_{|x|>2/u}\left(1-\frac{\sin(ux)}{ux}\right)\mathrm{d}\mu_n(x) \\ &\ge 2\int_{|x|>2/u}\left(1-\frac{1}{|ux|}\right)\mathrm{d}\mu_n(x) \ge \mu_n\left(\left\{x\colon |x| > \frac{2}{u}\right\}\right).\end{aligned}$$

Since $\lim_{t\to 0}\phi(t) = 1$,

$$\lim_{u\to 0}\frac{1}{u}\int_{-u}^{u}(1-\phi(t))\,\mathrm{d}t = 0.$$

Pick u such that

$$\frac{1}{u}\int_{-u}^{u}(1-\phi(t))\,\mathrm{d}t < \varepsilon.$$

Since $\lim_{n\to\infty}\phi_n(t) = \phi(t)$, it follows from BCT Theorem 1.56 that for large enough n,

$$2\varepsilon \geq \frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))\,\mathrm{d}t \geq \mu_n\left(\left\{x\colon |x| > \frac{2}{u}\right\}\right).$$

Since ε is arbitrary, $\{\mu_n\}$ are tight.

To complete the proof we observe that if $\{m(n)\}$ is a subsequence for which $\lim^d_{n\to\infty}\mu_{m(n)} = \mu$, then it follows from (1) that μ has characteristic function ϕ. Thus, from this observation and tightness, *every* subsequence $\{m(n)\}$ has a further subsequence $\{m(k(n))\}$ such that $\lim^d_{n\to\infty}\mu_{m(k(n))} = \mu$. We have shown that if f is bounded and continuous then every subsequence of $\{\mathbb{E}[f(X_n)]\}$ has a further subsequence that converges to $\mathbb{E}[f(X)]$. Thus the whole sequence $\{\mathbb{E}[f(X_n)]\}$ converges to $\mathbb{E}[f(X)]$. Thus $\lim^d_{n\to\infty}X_n = X$, or $\lim^d_{n\to\infty}\mu_n = \mu$.

□

3.3 Central Limit Theorems

We will now prove a series of central limit theorems (or CLTs). Let $\{X_n\} \in L^0$ and $S_n = \sum_{m=1}^{n}X_m$. We have shown, under various circumstances, convergence in probability and pointwise a.s. of $\frac{S_n}{n}$. We will show, under various circumstances,

$$\lim^d_{n\to\infty}\frac{S_n}{\sqrt{n}} = \mathcal{N}(0,\sigma^2).$$

Here's a teaser.

Theorem 3.26 (Bernoulli CLT). Suppose $\{X_n\} \overset{\text{i.i.d.}}{\sim} \mathrm{Bern}(\pm 1)$. Then

$$\lim^d_{n\to\infty}\frac{S_n}{\sqrt{n}} = \mathcal{N}(0,1).$$

Proof. Note that $\{X_n\}$ are i.i.d. with $\mathbb{E}[X_1] = 0$ and $\mathbb{E}\left[X_1^2\right] = 1$. Let $Z \sim \mathcal{N}(0,1)$. Then

$$\phi_Z(t) = \mathrm{e}^{-t^2/2}.$$

Then by Theorem 3.20 estimate we have

$$\phi_{\frac{S_n}{\sqrt{n}}}(t) = \phi_{S_n}\left(\frac{t}{\sqrt{n}}\right) = \left(\phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right)^n$$

and so

$$\lim_{n\to\infty}\phi_{\frac{S_n}{\sqrt{n}}}(t) = \lim_{n\to\infty}\left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right)^n = \mathrm{e}^{-t^2/2} = \phi_Z(t).$$

Therefore by Theorem 3.25, $\lim^d_{n\to\infty}\frac{S_n}{\sqrt{n}} = \mathcal{N}(0,1)$. □

To progress we need basic results from complex analysis.

Proposition 3.27. If $\{c_n\}, c \in \mathbb{C}$ and $\lim_{n\to\infty} c_n = c$, then
$$\lim_{n\to\infty}\left(1+\frac{c_n}{n}\right)^n = \mathrm{e}^c.$$

Proposition 3.28. Let $\{z_n\}_{n\in[m]}$ and $\{w_n\}_{n\in[m]}$ be complex numbers with $|z_n| \le \theta$ and $|w_n| \le \theta$ for $n \in [m]$. Then
$$\left|\prod_{n=1}^m z_n - \prod_{n=1}^m w_n\right| \le \theta^{n-1}\sum_{n=1}^m |z_n - w_n|.$$

The next step is to generalize CLT to triangular array of random variables.

Theorem 3.29 (Lindeberg-Feller CLT). Let $\{X_{mn}\}_{n\in[m],m\in\mathbb{N}} \in L^0$ be a triangular array, for which $\{X_{mn}\}_{n\in[m]}$ are independent and $\mathbb{E}[X_{mn}] = 0$. Suppose

(i) $\displaystyle\lim_{m\to\infty}\sum_{n=1}^m \mathbb{E}\left[X_{mn}^2\right] = \sigma^2 > 0,$

(ii) For all $\varepsilon > 0$, $\displaystyle\lim_{m\to\infty}\sum_{n=1}^m \mathbb{E}\left[|X_{mn}|^2; |X_{mn}| > \varepsilon\right] = 0.$

Then $\displaystyle\overset{d}{\lim_{m\to\infty}} S_n = \sum_{n=1}^m X_{mn} = \mathcal{N}(0,\sigma^2).$

Proof. Let $\phi_{mn}(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX_{mn}}\right]$ and $\sigma_{mn}^2 = \mathbb{E}\left[X_{mn}^2\right]$. By Theorem 3.25, it suffices to show that
$$\lim_{m\to\infty}\prod_{n=1}^m \phi_{mn}(t) = \mathrm{e}^{-\sigma^2t^2/2}.$$
Fix t; let $z_{mn} = \phi_{mn}(t)$ and $w_{mn} = (1-\sigma_{mn}^2t^2/2)$. By Corollary 3.19, (i) and (ii),
$$\begin{aligned}
|z_{mn} - w_{mn}| = \left|\phi_{mn}(t) - \left(1 - \frac{\sigma_{mn}^2t^2}{2}\right)\right| &\le \mathbb{E}\left[|tX_{mn}|^3 \wedge 2|tX_{mn}|^2\right] \\
&\le \mathbb{E}\left[|tX_{mn}|^3; |X_{mn}| \le \varepsilon\right] + \mathbb{E}\left[2|tX_{mn}|^2; |X_{mn}| > \varepsilon\right] \\
&\le \varepsilon t^3\mathbb{E}\left[|X_{mn}|^2; |X_{mn}| \le \varepsilon\right] + 2t^2\mathbb{E}\left[|X_{mn}|^2; |X_{mn}| > \varepsilon\right]. \\
\sum_{n=1}^m |z_{mn} - w_{mn}| &\le \sum_{n=1}^m\left(\varepsilon t^3\mathbb{E}\left[|X_{mn}|^2; |X_{mn}| \le \varepsilon\right] + 2t^2\mathbb{E}\left[|X_{mn}|^2; |X_{mn}| > \varepsilon\right]\right) \\
\lim_{m\to\infty}\sum_{n=1}^m |z_{mn} - w_{mn}| &\le \lim_{m\to\infty}\sum_{n=1}^m\left(\varepsilon t^3\mathbb{E}\left[|X_{mn}|^2; |X_{mn}| \le \varepsilon\right] + 2t^2\mathbb{E}\left[|X_{mn}|^2; |X_{mn}| > \varepsilon\right]\right) \\
&\le \varepsilon t^3\left(\lim_{m\to\infty}\mathbb{E}\left[|X_{mn}|^2; |X_{mn}| \le \varepsilon\right]\right) + 2t^2\left(\lim_{m\to\infty}\sum_{n=1}^m\mathbb{E}\left[|X_{mn}|^2; |X_{mn}| > \varepsilon\right]\right) \\
&\le \varepsilon t^3\left(\lim_{m\to\infty}\mathbb{E}\left[|X_{mn}|^2; |X_{mn}| \le \varepsilon\right]\right) \le \varepsilon t^3\left(\lim_{m\to\infty}\mathbb{E}\left[|X_{mn}|^2\right]\right) \le \varepsilon\sigma^2t^3.
\end{aligned}$$
Since ε is arbitrary,
$$\lim_{m\to\infty}\sum_{n=1}^m |z_{mn} - w_{mn}| \le \lim_{\varepsilon\to 0}\varepsilon\sigma^2t^3 = 0.$$

We want to use the Proposition 3.28 with $\theta = 1$. In particular

$$|z_{mn}| = |\phi_{mn}(t)| = \left|\mathbb{E}\left[e^{itX_{mn}}\right]\right| \leq \mathbb{E}\left[\left|e^{itX_{mn}}\right|\right] \leq 1$$

and

$$\sigma_{mn}^2 = \mathbb{E}\left[X_{mn}^2\right] = \mathbb{E}\left[X_{mn}^2; |X_{mn}| \leq \varepsilon\right] + \mathbb{E}\left[X_{mn}^2; |X_{mn}| > \varepsilon\right] \leq \varepsilon^2 + \mathbb{E}\left[X_{mn}^2; |X_{mn}| > \varepsilon\right],$$

and ε is arbitrary so (ii) implies $\lim_{m\to\infty} \sup_{n\in[m]} \sigma_{mn}^2 = 0$. Thus if n is large, $w_{mn} = 1 - \sigma_{mn}^2 t^2/2 \in [-1, 1]$. Thus by Proposition 3.28, taking limits,

$$\lim_{m\to\infty}\left|\prod_{n=1}^m z_{mn} - \prod_{n=1}^m w_{mn}\right| \leq \lim_{m\to\infty}\sum_{n=1}^m |z_{mn} - w_{mn}| = 0.$$

Finally, by (i),

$$\lim_{m\to\infty}\sum_{n=1}^m -\frac{\sigma_{mn}^2 t^2}{2} = -\frac{\sigma^2 t^2}{2} \implies \lim_{m\to\infty}\prod_{n=1}^m\left(1 - \frac{\sigma_{mn}^2 t^2}{2}\right) = \lim_{m\to\infty}\prod_{n=1}^m w_{mn} = \exp\left(-\frac{\sigma^2 t^2}{2}\right).$$

One notices that

$$\begin{aligned} 0 &= \lim_{m\to\infty}\left|\prod_{n=1}^m z_{mn} - \prod_{n=1}^m w_{mn}\right| = \lim_{m\to\infty}\left|\prod_{n=1}^m \phi_{mn}(t) - \prod_{n=1}^m w_{mn}\right| = \lim_{m\to\infty}\left|\phi_{\sum_{n=1}^m X_n}(t) - \exp\left(-\frac{\sigma^2 t^2}{2}\right)\right| \\ &= \lim_{m\to\infty}\left|\phi_{S_m}(t) - \exp\left(-\frac{\sigma^2 t^2}{n}\right)\right| = \lim_{m\to\infty}\left|\phi_{S_m}(t) - \phi_{\mathcal{N}(0,\sigma^2)}(t)\right| \end{aligned}$$

Thus $\lim_{n\to\infty}^d S_n = \mathcal{N}\left(0, \sigma^2\right)$ as desired. □

Remark. The Bernoulli CLT is a special case of the Lindeberg-Feller CLT.

Theorem 3.30 (Kolmogorov's Three-Series Theorem). Let $\{X_n\} \in L^0$ be independent. Let $A > 0$ and let $Y_n = X_n \mathbb{1}_{|X_n| \leq A}$. Then

$$\sum_{n=1}^\infty X_n \quad \text{converges a.s.}$$

if and only if

$$\text{(i)} \sum_{n=1}^\infty \mathbb{P}(|X_n| > A) < \infty, \quad \text{(ii)} \sum_{n=1}^\infty \mathbb{E}[Y_n] \text{ converges}, \quad \text{(iii)} \sum_{n=1}^\infty \mathrm{Var}(Y_n) < \infty.$$

Proof. The "if" direction is as follows. Let $\mu_n = \mathbb{E}[Y_n]$. Then (iii) and Theorem 2.33 imply that $\sum_{n=1}^\infty (Y_n - \mu_n)$ converges a.s.. Using (ii) gives that $\sum_{n=1}^\infty Y_n$ converges a.s.. Finally (i) and the Borel-Cantelli lemma Theorem 2.27 imply $\mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0$, so $\sum_{n=1}^\infty X_n$ converges a.s..

The "only if" direction is as follows. If (i) is false, then $\sum_{n=1}^\infty \mathbb{P}(|X_n| > A) = \infty$. By the Borel-Cantelli converse Theorem 2.28, $\mathbb{P}(|X_n| > A \text{ i.o.}) > 0$, so that $\sum_{n=1}^\infty X_n$ does not converge a.s.. Suppose now that the sum in (i) is finite but the sum in (iii) is infinite. Let

$$c_n = \sum_{m=1}^n \mathrm{Var}(Y_m) \quad \text{and} \quad X_{mn} = \frac{Y_n - \mu_n}{\sqrt{c_m}}.$$

Then

$$\mathbb{E}[X_{mn}] = 0, \quad \sum_{n=1}^m \mathbb{E}\left[X_{mn}^2\right] = 1, \qquad \text{for every } \varepsilon > 0, \lim_{m\to\infty}\sum_{n=1}^m \mathbb{E}[|X_{mn}|; |X_{mn}| > \varepsilon] = 0$$

since $\sum_{n=1}^m \mathbb{E}[|X_{mn}|; |X_{mn}| > \varepsilon] = 0$ for m large enough that $2A/\sqrt{c_m} < \varepsilon$. Then applying the Lindeberg-Feller CLT Theorem 3.29, if $S_m = \sum_{n=1}^m X_{mn}$, then $\lim_{n\to\infty}^d S_n = \mathcal{N}(0, 1)$. Now it's easy to see that

1. if $\sum_{n=1}^\infty X_n$ exists then $\sum_{n=1}^\infty Y_n$ exists.

2. if we let $T_n = (\sum_{m=1}^{n} Y_m)/\sqrt{c_n}$ then $\lim_{n\to\infty}^{d} T_n = 0$.

These results imply $\lim_{n\to\infty}^{d}(S_n - T_n) = \mathcal{N}(0,1)$. Since

$$S_n - T_n = -\frac{1}{\sqrt{c_n}}\sum_{m=1}^{n}\mu_m$$

is nonrandom, this is absurd.

Finally, assume the series in (i) and (iii) are finite. Theorem 2.33 implies that $\sum_{n=1}^{\infty}(Y_n - \mu_n)$ converges, so if $\sum_{n=1}^{\infty} X_n$ and hence $\sum_{n=1}^{\infty} Y_n$ converges, taking differences shows that (ii) holds. $\square$

Example 3.31 (Infinite Second Moment CLT). Suppose $\{X_n\} \in L^1$ are i.i.d. symmetric, that is, $\mathbb{P}(X_1 > x) = \mathbb{P}(X_1 < -x)$, and $\mathbb{P}(|X_1| > x) = x^{-2}$, for $x \geq 1$. Then by Lemma 2.20,

$$\mathbb{E}\left[X_1^2\right] = \int_0^{\infty} 2x\mathbb{P}(|X_1| > x)\,\mathrm{d}x = \int_0^{\infty}\frac{2}{x}\,\mathrm{d}x = \infty.$$

However, when S_n is normalized it converges to a normal distribution. Let

$$Y_{mn} = X_n \mathbb{1}_{|X_n| \leq c_m}.$$

The truncation level c_m is chosen large enough to make

$$\lim_{m\to\infty}\sum_{n=1}^{m}\mathbb{P}(Y_{mn} \neq X_n) \leq \lim_{m\to\infty} m\mathbb{P}(|X_1| > c_m) = 0.$$

And we want $\mathrm{Var}(Y_{mn})$ to be minimal, so we keep the truncation at the lowest possible level. It turns out that $c_n = n^{1/2}\log(\log(n))$ works.

Our next step is to show $\mathbb{E}\left[Y_{mn}^2\right] = O(\log(m))$. For this we need upper and lower bounds. Since $\mathbb{P}(|Y_{mn}| > x) \leq \mathbb{P}(|X_1| > x)$ and is 0 for $x > c_0$, we have

$$\begin{aligned}
\mathbb{E}\left[Y_{mn}^2\right] &= \int_0^{\infty} 2y\mathbb{P}(|Y_{mn}| > y)\,\mathrm{d}y = \int_0^{c_m} 2y\mathbb{P}(|Y_{mn}| > y)\,\mathrm{d}y \\
&\leq \int_0^{c_m} 2y\mathbb{P}(|X_1| > y)\,\mathrm{d}y \leq 1 + \int_1^{c_m} 2y\mathbb{P}(|X_1| > y)\,\mathrm{d}y = 1 + \int_1^{c_m}\frac{2}{y}\,\mathrm{d}y = 1 + 2\log(c_m) \\
&= 1 + \log(m) + 2\log(\log(m)) = O(\log(m)).
\end{aligned}$$

In the other direction, we observe

$$\mathbb{P}(|Y_{mn}| > x) = \mathbb{P}(|X_1| > x) - \mathbb{P}(|X_1| > c_m) \geq \left(1 - \log(\log(m))^{-2}\right)\mathbb{P}(|X_1| > x) \quad \text{when } x \leq \sqrt{m}$$

so

$$\mathbb{E}\left[Y_{mn}^2\right] \geq \left(1 - \log(\log(m))^{-2}\right)\int_1^{\sqrt{m}}\frac{2}{y}\,\mathrm{d}y = O(\log(m)).$$

Thus $\mathbb{E}\left[Y_{mn}^2\right] = O(\log(m))$. If $S_m' = \sum_{n=1}^{m} Y_{mn}$ then $\mathrm{Var}(S_n') = O(n\log(m))$, so we apply Theorem 3.29 to $X_{mn} = Y_{mn}/\sqrt{m\log(m)}$. The condition (i) is satisfied. And for m large enough the sum in (ii) is 0. It follows that $\lim_{n\to\infty}^{d} S_n'/\sqrt{n\log(n)} = \mathcal{N}(0,1)$. The choice of c_n guarantees $\lim_{n\to\infty}\mathbb{P}(S_n \neq S_n') = 0$ so $\lim_{n\to\infty}^{d} S_n/\sqrt{n\log(n)} = \mathcal{N}(0,1)$.

3.4 Limit Theorems in $\mathbb{R}^d$

We begin by considering weak convergence in a general metric space (S,ρ).

Definition 3.32 (General Convergence in Distribution). A sequence of random variables $\{X_n\} \in L^0(\Omega, S)$ converges to

$X \in L^0(\Omega, S)$ if

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \quad \text{for all bounded continuous functions } f\colon S \to \mathbb{R}.$$

Definition 3.33 (Lipschitz Continuous). Let $f\colon (M_1, d_1) \to (M_2, d_2)$ be a function between metric spaces. Then, for $C \in \mathbb{R}$, f is **C-Lipschitz continuous** if

$$d_2(f(x), f(y)) \leq C d_1(x, y) \quad \text{for all } x, y \in M_1.$$

The corresponding definition in the case of functions $f\colon S \to \mathbb{R}$ is that

$$|f(x) - f(y)| \leq C\rho(x, y) \quad \text{for all } x, y \in S.$$

Theorem 3.34 (General Portmanteau Theorem). The following statements are equivalent.

1. $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ for all bounded continouus f.
2. $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ for all bounded Lipschitz continouus f.
3. For all closed sets V, $\limsup_{n\to\infty} \mathbb{P}(X_n \in V) \leq \mathbb{P}(X \in V)$.
4. For all open sets U, $\liminf_{n\to\infty} \mathbb{P}(X_n \in U) \geq \mathbb{P}(X \in U)$.
5. For all sets A with $\mathbb{P}(X \in \partial A) = 0$, $\lim_{n\to\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$.
6. For all bounded functions f with $\mathbb{P}(X \in \mathcal{D}_f) = 0$, $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$.

Proof.

Claim 1. (1) implies (2).

Proof. Lipschitz continuous functions are continuous. ∎

Claim 2. (2) implies (3).

Proof. Let $\rho(x, K) = \inf\{\rho(x, y)\colon y \in V\}$, and $f_j(x) = (1 - j\rho(x, K))$. Then f_j is Lipschitz-continuous, has values in $[0, 1]$, and decreases almost everywhere to $\mathbb{1}_V(x)$ as $j \to \infty$. So

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in V) \leq \lim_{j\to\infty} \lim_{n\to\infty} \mathbb{E}[f_j(X_n)] = \lim_{j\to\infty} \mathbb{E}[f_j(X)] = \mathbb{P}(X \in V).$$

∎

Claim 3. (3) and (4) are equivalent.

Proof. It follows easily from the fact that A is open if and only if A^c is closed; and $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$. ∎

Claim 4. (3) and (4) imply (5).

Proof. Let $U = A^\circ$ and $V = \overline{A}$. Then $\mu(\partial A) = 0$, so

$$\mu(U) = \mu(A) = \mu(V).$$

Then using (3) and (4),

$$\limsup_{n\to\infty} \mu_n(A) \le \limsup_{n\to\infty} \mu_n(V) \le \mu(V) = \mu(A) \quad \text{and} \quad \liminf_{n\to\infty} \mu_n(A) \ge \liminf_{n\to\infty} \mu_n(U) \ge \mu(U) = \mu(A).$$

Thus

$$\lim_{n\to\infty} \mu_n(A) = \mu(A)$$

for all A for which $\partial A = 0$. ■

Claim 5. (5) implies (6).

Proof. Suppose f is bounded with $|f(x)| \le M$. Pick $\{\alpha_n\}_{n=0}^{\ell}$ such that $\mathbb{P}(f(X) = \alpha_n) = 0$ for $0 \le n \le \ell$ and $\alpha_0 < -M < M < \alpha_\ell$, and $\alpha_n - \alpha_{n-1} < \varepsilon$. This is always possible since $\{\alpha \colon \mathbb{P}(f(X) = \alpha) > 0\}$ is at most countable. Let $A_n = \{x \colon \alpha_{n-1} < f(x) \le \alpha_n\}$. Then $\partial A_n \subseteq \{x \colon f(x) \in \{\alpha_{n-1}, \alpha_n\}\} \cup \mathcal{D}_f$. Thus $\mathbb{P}(X \in \partial A_n) = 0$. It follows from (5) that

$$\lim_{n\to\infty} \sum_{m=1}^{\ell} \alpha_m \mathbb{P}(X_n \in A_m) = \sum_{m=1}^{\ell} \alpha_m \mathbb{P}(X \in A_m).$$

The definition of the α_n implies

$$0 \le \sum_{m=1}^{\ell} \alpha_m \mathbb{P}(X_n \in A_m) - \mathbb{E}[f(X_n)] \le \varepsilon \quad \text{for all } n \qquad \text{and} \quad 0 \le \sum_{m=1}^{\ell} \alpha_m \mathbb{P}(X \in A_m) - \mathbb{E}[f(X)] \le \varepsilon.$$

Since ε is arbitrary, $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$. ■

Claim 6. (6) implies (1).

Proof. Continuous functions f have $\mathcal{D}_f = \emptyset$, so $\mathbb{P}(X \in \mathcal{D}_f) = \mathbb{P}(X \in \emptyset) = 0$. ■

□

Now we will specialize to $S = \mathbb{R}^d$ and ρ the Euclidean metric. Let $X : \Omega \to \mathbb{R}^d$ be a random vector. We have already discussed that its distribution function is defined by $F(x) = \mathbb{P}(X \preceq x)$, and it has four specific properties:

1. F is non-decreasing. If $x \preceq y$ then $F(x) \le F(y)$.
2. F is right-continuous. If $\{x_n\} \downarrow x$ componentwise, then $\lim_{n\to\infty} F(x_n) = F(x)$.
3. $\lim_{x\to\infty} F(x) = 1$, and for all $n \in [d]$, $\lim_{x_n \to -\infty} F(x) = 0$.
4. For all boxes B, $\mathbb{P}(X \in B) \ge 0$.

We have already discussed that if $B = \bigtimes_{n=1}^{d} (a_n, b_n]$ then

$$\mathbb{P}(X \in B) = \sum_{n=0}^{d} (-1)^n \sum_{\substack{S \subseteq [d] \\ |S| = n}} F(a_S, b_{[d]\setminus S}).$$

Definition 3.35 (General Convergence of Distribution Functions). If $\{F_n\}$ and F are distribution functions on $\mathbb{R}^d$, then we define convergence in distribution in the usual way:

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{for all } x \in \mathcal{C}_F \qquad \Longrightarrow \overset{d}{\lim_{n\to\infty}} F_n = F.$$

Example 3.36. Let $Y \sim \mathrm{Uni}([0,1])$. Then the distribution function of $(0, Y)$ is

$$F(x,y) = \begin{cases} 1 & x \geq 0, y \geq 1 \\ y & x \geq 0, 0 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Then this distribution has no atoms, but F is discontinuous at $(0, y)$ when $y > 0$.

But in spite of this the definition still makes sense. Observe that if $x_n \prec x$ and $x_n \uparrow x$, then

$$\lim_{n\to\infty} (F(x) - F(x_n)) = \lim_{n\to\infty} (\mathbb{P}(X \preceq x) - \mathbb{P}(X \preceq x_n)) = \mathbb{P}(X \preceq x) - \mathbb{P}(X \prec x).$$

Let $H_c^n = \left\{x \in \mathbb{R}^d : x_n = c\right\}$ be the hyperplane where the n^{th} coordinate is c. Then for each n, the H_c^n are disjoint, so $\mathcal{D}^n = \{c : \mathbb{P}(X \in H_c^n) > 0\}$ is at most countable. It's easy to see that if $x \notin \mathcal{D}^n$ for all $n \in [d]$ then $x \in \mathcal{C}_F$, so $\left(\bigcup_{n=1}^d \mathcal{D}^n\right)^c \subseteq \mathcal{C}_F$. This gives us more than enough points to reconstruct F.

Theorem 3.37. If $\{X_n\} \in L^0(\Omega, \mathbb{R}^n)$ are random vectors such that $X_n \sim F_n$, then

$$\overset{d}{\lim_{n\to\infty}} X_n = X \quad \text{if and only if} \quad \overset{d}{\lim_{n\to\infty}} F_n = F.$$

Proof.

Claim 1. $\lim_{n\to\infty}^d X_n = X$ implies $\lim_{n\to\infty}^d F_n = F$.

Proof. We show that (5) of Portmanteau Theorem Theorem 3.34 implies $\lim_{n\to\infty}^d F_n = F$. If $x \in \mathcal{C}_F$, then $A = (-\infty, x]$ has $\mu(\partial A) = 0$, so

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A) = F(x).$$

Thus $\lim_{n\to\infty}^d F_n = F$. ■

Claim 2. $\lim_{n\to\infty}^d F_n = F$ implies $\lim_{n\to\infty}^d X_n = X$.

Proof. We show that $\lim_{n\to\infty}^d F_n = F$ implies (4) of Portmanteau Theorem Theorem 3.34. Let $\mathcal{D}^n = \{c : \mathbb{P}(X \in H_c^n)\}$. We say a box $A = (a, b]$ is good if $a_n, b_n \notin \mathcal{D}^n$ for all n. Since $\lim_{n\to\infty}^d F_n = F$, $\lim_{n\to\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$ for all good boxes A. This is true for B that are a finite disjoint union of good boxes. For any open set U, there is a sequence of sets $\{B_n\}$ which are themselves a finite disjoint union of good boxes, so

$$\liminf_{n\to\infty} \mathbb{P}(X_n \in U) \geq \lim_{k\to\infty} \liminf_{n\to\infty} \mathbb{P}(X_n \in B_k) = \lim_{k\to\infty} \mathbb{P}(X \in B_k) = \mathbb{P}(X \in G).$$

Therefore (4) of Portmanteau Theorem Theorem 3.34 holds so $\lim_{n\to\infty}^d X_n = X$. ■

□

Definition 3.38 (General Characteristic Function). The characteristic function of X is

$$\phi_X(t) = \mathbb{E}\left[e^{i\langle t, X\rangle}\right]$$

where $\langle t, X\rangle = \sum_{n=1}^d t_n X_n$ is the standard inner product.

Theorem 3.39 (General Inversion Formula). Let $A = \bigtimes_{n=1}^{d} [a_n, b_n]$ with $\mu(\partial A) = 0$, where $\mu = \mathbb{P} \cdot X^{-1}$. Then

$$\mu(A) = \lim_{T\to\infty} \frac{1}{(2\pi)^d} \int_{[-T,T]^d} \phi_X(t) \prod_{n=1}^{d} \psi_n(t_n) \, \mathrm{d}t \, .$$

where $\psi_j(t) = \dfrac{\mathrm{e}^{-\mathrm{i}ta_j} - \mathrm{e}^{-\mathrm{i}tb_j}}{\mathrm{i}t}$.

Proof. We use Fubini's theorem Theorem 1.72 to write

$$\begin{aligned}
\int_{[-T,T]^d} \phi_X(t) \prod_{n=1}^{d} \phi_n(t_n) \, \mathrm{d}t &= \int_{[-T,T]^d} \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\langle t, X\rangle}\right] \prod_{n=1}^{d} \psi_n(t_n) \, \mathrm{d}t \\
&= \int_{[-T,T]^d} \mathbb{E}\left[\prod_{n=1}^{d} \mathrm{e}^{\mathrm{i}tX_n}\right] \prod_{n=1}^{d} \psi_n(t_n) \, \mathrm{d}t \\
&= \int_{[-T,T]^d} \prod_{n=1}^{d} \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX_n}\right] \psi_n(t_n) \, \mathrm{d}t \\
&= \int_{[-T,T]^d} \prod_{n=1}^{d} \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX_n} \psi_n(t_n)\right] \mathrm{d}t \\
&= \prod_{n=1}^{d} \int_{-T}^{T} \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX_n} \psi_n(t_n)\right] \mathrm{d}t_n \\
&= \prod_{n=1}^{d} \mathbb{E}\left[\int_{-T}^{T} \mathrm{e}^{\mathrm{i}tX_n} \psi_n(t_n) \, \mathrm{d}t_n\right]
\end{aligned}$$

From the proof of Theorem 3.22, we know

$$\lim_{T\to\infty} \int_{-T}^{T} \psi_n(t_n) \mathrm{e}^{\mathrm{i}t_n x_n} \, \mathrm{d}t_n = \pi\big(\mathbb{1}_{(a_n,b_n)}(x) + \mathbb{1}_{[a_n,b_n]}(x)\big).$$

An application of BCT Theorem 1.56 allows us to pass the limit inside:

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{(2\pi)^d} \int_{[-T,T]^d} \phi_X(t) \prod_{n=1}^{d} \phi_n(t_n) \, \mathrm{d}t &= \lim_{T\to\infty} \frac{1}{(2\pi)^d} \prod_{n=1}^{d} \mathbb{E}\left[\int_{-T}^{T} \mathrm{e}^{\mathrm{i}tX_n} \psi_n(t_n) \, \mathrm{d}t_n\right] \\
&= \frac{1}{(2\pi)^d} \prod_{n=1}^{d} \mathbb{E}\left[\lim_{T\to\infty} \int_{-T}^{T} \mathrm{e}^{\mathrm{i}tX_n} \psi_n(t_n) \, \mathrm{d}t_n\right] \\
&= \frac{1}{(2\pi)^d} \prod_{n=1}^{d} \mathbb{E}\left[\pi\big(\mathbb{1}_{(a_n,b_n)}(X_n) + \mathbb{1}_{[a_n,b_n]}(X_n)\big)\right] \\
&= \prod_{n=1}^{d} \frac{1}{2\pi} \mathbb{E}\left[\pi\big(\mathbb{1}_{(a_n,b_n)}(X_n) + \mathbb{1}_{[a_n,b_n]}(X_n)\big)\right] \\
&= \prod_{n=1}^{d} \frac{1}{2} \mathbb{E}\left[\mathbb{1}_{(a_n,b_n)}(X_n) + \mathbb{1}_{[a_n,b_n]}(X_n)\right] \\
&= \prod_{n=1}^{d} \frac{1}{2} (\mathbb{P}(X_n \in (a_n, b_n)) + \mathbb{P}(X_n \in [a_n, b_n])) \\
&= \prod_{n=1}^{d} \left(\mathbb{P}(X_n \in [a_n, b_n]) + \frac{1}{2} \mathbb{P}(X_n \in \{a_n, b_n\})\right)
\end{aligned}$$

$$= \prod_{n=1}^{d} \mathbb{P}(X_n \in [a_n, b_n]) = \mathbb{P}(X \in A) = \mu(A)$$

where the penultimate equality is due to the fact that $\mu(\partial A) = \mathbb{P}(X \in \partial A) = 0$. □

Theorem 3.40 (General Continuity Theorem). Let $\{X_n\} \in L^0(\Omega, \mathbb{R}^d)$ be random vectors with characteristic functions $\{\phi_n\}$.

1. If $X \in L^0(\Omega, \mathbb{R}^n)$ has characteristic function ϕ, and $\overset{d}{\lim_{n\to\infty}} X_n = X$, then $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ for all t.

2. If $\phi_n(t)$ converges pointwise to a limit $\phi(t)$ which is continuous at 0, then $\{X_n\}$ have tight distributions and $\overset{d}{\lim_{n\to\infty}} X_n = X$ with characteristic function ϕ.

Proof.

1. The function $f_t(x) = \mathrm{e}^{\mathrm{i}\langle t, x\rangle}$ is bounded and continuous, so if $\lim_{n\to\infty}^{d} X_n = X$, then $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ for all t.

2. By the proof of Theorem 3.25, it suffices to prove that $\{X_n\}$ has a tight sequence of distributions; the rest of the proof goes exactly the same. To do this, we observe that if we fix $\theta \in \mathbb{R}^d$, then $\lim_{n\to\infty} \phi_n(s\theta) = \phi(s\theta)$. It follows from Theorem 3.25 that the distributions of $\langle \theta, X_n\rangle$ are tight. Letting $\theta = e_k$ for $k \in [d]$ shows that the distributions of $\{X_{n,k}\}$ are tight. From the definition of tightness it is clear that $\{X_n\}$ have distributions that are tight.

□

Theorem 3.41 (Cramér-Wold Device). Let $\{X_n\}, X \in L^0(\Omega, \mathbb{R}^d)$ are random vectors. If $\overset{d}{\lim_{n\to\infty}} \langle \theta, X_n\rangle = \langle \theta, X\rangle$ for all $\theta \in \mathbb{R}^d$, then $\overset{d}{\lim_{n\to\infty}} X_n = X$.

Proof. Since $\lim_{n\to\infty}^{d} \langle \theta, X_n\rangle = \langle \theta, X\rangle$ for all $\theta \in \mathbb{R}^d$, then using the bounded continuous function $f(x) = \mathrm{e}^{\mathrm{i}x}$ gives

$$\lim_{n\to\infty} \phi_{X_n}(\theta) = \lim_{n\to\infty} \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\langle \theta, X_n\rangle}\right] = \lim_{n\to\infty} \mathbb{E}[f(\langle \theta, X_n\rangle)] = \mathbb{E}[f(\langle \theta, X\rangle)] = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\langle \theta, X\rangle}\right] = \phi_X(\theta).$$

Thus by Theorem 3.40, $\lim_{n\to\infty}^{d} X_n = X$. □

Theorem 3.42 (Multivariate CLT). Let $\{X_n\} \in L^0(\Omega, \mathbb{R}^d)$ be i.i.d. with $\mathbb{E}[X_1] = \mu$ and finite covariance matrix $\Gamma = \mathbb{E}\left[(X_1 - \mu)(X_1 - \mu)^\mathsf{T}\right]$. Then $\overset{d}{\lim_{n\to\infty}} \dfrac{S_n - n\mu}{\sqrt{n}} = \mathcal{N}(0, \Gamma)$.

Proof. Consider $X_n' = X_n - \mu$. Then $\mathbb{E}[X_1'] = 0$ and $\Gamma = \mathbb{E}\left[X_1' X_1'^\mathsf{T}\right]$. Let $\theta \in \mathbb{R}^d$. Then $\langle \theta, X_n\rangle$ is a random variable, with expectation

$$\mathbb{E}[\langle \theta, X_n'\rangle] = \langle \theta, \mathbb{E}[X_n']\rangle = 0$$

and variance

$$\mathbb{E}\left[\langle \theta, X_n'\rangle^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{d} \theta_i X_{n,i}'\right)^2\right] = \sum_{i=1}^{d}\sum_{j=1}^{d} \theta_i\theta_j \mathbb{E}\left[X_{n,i}' X_{n,j}'\right] = \sum_{i=1}^{d}\sum_{j=1}^{d} \theta_i\theta_j \Gamma_{i,j}.$$

The $\{\langle \theta, X_n\rangle\}$ are i.i.d., so from the one-dimensional CLT Theorem 3.29,

$$\overset{d}{\lim_{n\to\infty}} \frac{1}{\sqrt{n}} \sum_{m=1}^{n} \langle X_m', \theta\rangle = \mathcal{N}\left(0, \theta^\mathsf{T}\Gamma\theta\right) \quad \text{for all } \theta.$$

Therefore by Theorem 3.41,

$$\mathcal{N}(0,\Gamma) = \lim_{n\to\infty}^{d} \frac{1}{\sqrt{n}} \sum_{m=1}^{n} X'_m = \lim_{n\to\infty}^{d} \frac{1}{\sqrt{n}} \sum_{m=1}^{n} (X_m - \mu) = \lim_{n\to\infty}^{d} \frac{S_n - n\mu}{\sqrt{n}}.$$

□

3.5 Poisson Convergence

Theorem 3.43 (Weak Law of Small Numbers). Let $\{X_{mn}\}_{n\in[m],m\in\mathbb{N}} \in L^0$ be a triangular array where $\{X_{mn}\}_{n\in[m]}$ are independent, and $X_{mn} \sim \mathrm{Bern}(p_{mn})$. Suppose

(i) $\displaystyle\lim_{m\to\infty} \sum_{n=1}^{m} p_{mn} = \lambda \in (0,\infty)$.

(ii) $\displaystyle\lim_{m\to\infty} \sup_{n\in[m]} p_{mn} = 0$.

Suppose further that $S_m = \displaystyle\sum_{n=1}^{m} X_{mn}$ and $X \sim \mathrm{Pois}(\lambda)$. Then $\displaystyle\lim_{n\to\infty}^{d} S_n = X$.

Proof. Let

$$\phi_{mn}(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX_{mn}}\right] = 1 - p_{mn} + p_{mn}\mathrm{e}^{\mathrm{i}t} = 1 + p_{mn}\left(\mathrm{e}^{\mathrm{i}t} - 1\right).$$

Then

$$\phi_{S_m}(t) = \phi_{\sum_{n=1}^{m} X_{mn}}(t) = \prod_{n=1}^{m} \phi_{mn}(t) = \prod_{n=1}^{m} \left(1 + p_{mn}\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right).$$

For $0 \le p \le 1$,

$$\left|\exp\left(p\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right)\right| = \exp\left(p\,\mathrm{Re}\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right) \le 1 \quad \text{and} \quad \left|1 + p\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right| \le 1.$$

The latter inequality is because $1 + p\left(\mathrm{e}^{\mathrm{i}t} - 1\right)$ is on the line segment connecting $\mathrm{e}^{\mathrm{i}t}$ to 1 in $\mathbb{C}$. Using Proposition 3.28 with $\theta = 1$ and then the inequality $\left|\mathrm{e}^{b} - (1+b)\right| \le b^2$ for $|b| \le 1$, which is valid when $\sup_{n\in[m]} p_{mn} \le \frac{1}{2}$ since $\left|\mathrm{e}^{\mathrm{i}t} - 1\right| \le 2$,

$$\begin{aligned}
\left|\exp\left(\sum_{n=1}^{m} p_{mn}\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right) - \prod_{n=1}^{m}\left(1 + p_{mn}\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right)\right| &\le \sum_{n=1}^{m} \left|\exp\left(p_{mn}\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right) - \left(1 + p_{mn}\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right)\right| \\
&\le \sum_{n=1}^{m} p_{mn}^2 \left|\mathrm{e}^{\mathrm{i}t} - 1\right|^2 \\
&\le 4\left(\sup_{n\in[m]} p_{mn}\right) \sum_{n=1}^{m} p_{mn} \\
\lim_{m\to\infty} \left|\exp\left(\sum_{n=1}^{m} p_{mn}\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right) - \prod_{n=1}^{m}\left(1 + p_{mn}\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right)\right| &\le \lim_{m\to\infty} 4\left(\sup_{n\in[m]} p_{mn}\right) \sum_{n=1}^{m} p_{mn} = 0.
\end{aligned}$$

The final equality uses both assumptions (i) and (ii). Since $\lim_{m\to\infty} \sum_{n=1}^{m} p_{mn} = \lambda$, then

$$\lim_{n\to\infty} \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tS_n}\right] = \exp\left(\lambda\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right) = \phi_X(t).$$

The conclusion follows by Theorem 3.25. □

4 Conditional Expectations

4.1 Radon-Nikodym Theorem

In this section we will prove the Radon-Nikodym theorem. This will allow us to show the existence of conditional expectations which are needed in the future.

Throughout this section, let (Ω, Σ) be a measurable space. Further, we use the shorthand $L^p(\Sigma_0) = L^p(\Omega, \Sigma_0, \mathbb{P})$ for $p \geq 0$.

Definition 4.1 (Signed Measure). A function $\alpha\colon \Sigma \to (-\infty, \infty]$ is a **signed measure** if

1. $\alpha(\emptyset) = \emptyset$.

2. If $\{E_n\} \in \Sigma$ are disjoint and $E = \bigcup_{n=1}^{\infty} E_n$, then $\alpha(E) = \sum_{n=1}^{\infty} \alpha(E_n)$ in the following sense:

   - If $\alpha(E) < \infty$, then $\sum_{n=1}^{\infty} \alpha(E_n)$ converges absolutely and is equal to $\alpha(E)$.

   - If $\alpha(E) = \infty$, then $\sum_{n=1}^{\infty} \alpha(E_n)^- < \infty$ and $\sum_{n=1}^{\infty} \alpha(E_n)^+ = \infty$.

As such, a signed measure cannot be allowed to take both the values ∞ and $-\infty$. In most formulations, a singed measure is allowed to take values in either $(-\infty, \infty]$ or $[-\infty, \infty)$; we ignore the second formulation in order to simplify statements later.

Example 4.2. Let μ be a regular measure; let f be a function with $\int_\Omega f^- \,\mathrm{d}\mu < \infty$. Let

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

Then α is a signed measure.

Example 4.3. Let μ_1 and μ_2 be measures with $\mu_2(\Omega) < \infty$, and let $\alpha(A) = \mu_1(A) - \mu_2(A)$.

Indeed, the above example is the general case; this is called the *Jordan decomposition*, which will be defined later.

Definition 4.4 (Positive, Negative, Null). Let α be a signed measure. A set A is **positive** with respect to α if every measurable $B \subseteq A$ has $\alpha(B) \geq 0$. A set A is **negative** with respect to α if every measurable $B \subseteq A$ has $\alpha(B) \leq 0$. A set A is **null** with respect to α if every measurable $B \subseteq A$ has $\alpha(B) = 0$.

Proposition 4.5.

1. Every measurable subset of a positive set is positive.

2. If the sets $\{A_n\}$ are positive then $A = \bigcup_{n=1}^{\infty} A_n$ is also positive.

Proof.

1. Let B be a measurable subset of A which is a positive measurable set. Then every measurable subset C of B is also a measurable subset of A, so $\mu(C) \geq 0$. Therefore B is positive.

2. Observe that
$$B_n = A_n \cap \left(\bigcap_{m=1}^{n-1} A_m^c\right) \subseteq A_n$$
are positive, disjoint, and $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. Let $E \subseteq A$ be measurable, and let $E_n = E \cap B_n$. Since B_n is positive, $\alpha(E_n) \geq 0$, so $\alpha(E) = \sum_{n=1}^{\infty} \alpha(E_n) \geq 0$.

□

The conclusions remain valid if "positive" is replaced by "negative".

Lemma 4.6. Let E be a measurable set with $\alpha(E) < 0$. Then there is a negative set $F \subseteq E$ with $\alpha(F) < 0$.

Proof. If E is negative, this is true. If not, we do an inductive process. let n_1 be the smallest positive integer so that there is an $E_1 \subset E$ with $\alpha(E_1) > n_1^{-1}$. For the induction, let $k \geq 2$. If $F_k = E \setminus \left(\bigcup_{n=1}^{k-1} E_n\right)$ is negative, we are done. If not, let n_k be the smallest positive integer so that there is an $E_k \subset F_k$ with $\alpha(E_k) \geq n_k^{-1}$. If the construction does not stop for any $k < \infty$, let
$$F = \bigcap_{k=1}^{\infty} F_k = E \setminus \left(\bigcap_{k=1}^{\infty} E_k\right).$$
Since $0 > \alpha(E) > -\infty$ and $\alpha(E_k) \geq 0$, it follows from the definition of signed measure that
$$\alpha(E) = \alpha(F) + \sum_{k=1}^{\infty} \alpha(E_k),$$
so that
$$\alpha(F) \leq \alpha(E) < 0$$
and the sum is finite. From the last observation and the construction, it follows that F can have no subset G with $\alpha(G) > 0$, for then $\alpha(G) \geq N^{-1}$ for some N and we would have a contradiction. □

Theorem 4.7 (Hahn Decomposition). Let α be a signed measure. Then there is a positive set A and a negative set B so that $\Omega = A \cup B$ and $A \cap B = \emptyset$.

Proof. Let $c = \inf\{\alpha(B) \colon B \in \Sigma \text{ is negative}\} \leq 0$. Let $\{B_n\}$ be neagtive measurable sets with $\lim_{n\to\infty} \alpha(B_n) \downarrow c$. By Proposition 4.5, B is negative, so $\alpha(B) \geq c$. To prove $\alpha(B) \leq c$, we observe that, since B is negative,
$$\alpha(B) = \lim_{n\to\infty} (\alpha(B_n) + \alpha(B \setminus B_n)) \leq \lim_{n\to\infty} \alpha(B_n) = c.$$
Thus $\alpha(B) = c$, it follows from Definition 4.1 that $c > -\infty$.

Let $A = B^c$. To show A is positive, if $E \subset A$ with $\alpha(E) < 0$, then by Lemma 4.6, $F \subseteq E$ with $\alpha(F) < 0$. But then $B \bigcup F$ is negative and
$$\alpha(B \cup F) = \alpha(B) + \alpha(F) = c + \alpha(F) < c,$$

a contradiction. □

The Hahn decomposition is not unique, but rather unique up to null sets. In particular if $\Omega = A_1 \cup B_1 = A_2 \cup B_2$ are two Hahn decompositions, $A_1 \cap B_2$ and $A_2 \cap B_1$ are null sets.

Definition 4.8 (Mutually Singular). Two measures μ_1 and μ_2 are **mutually singular** if there is a set A with $\mu_1(A) = 0$ and $\mu_2(A^c) = 0$. In this case we also say μ_1 is **singular with respect to** μ_2 and write $\mu_1 \perp \mu_2$.

Theorem 4.9 (Jordan Decomposition). Let α be a signed measure. There are mutually singular measures α_+ and α_- so that $\alpha = \alpha_+ - \alpha_-$. Moreover there is only one such pair.

Proof. Let $\Omega = A \cup B$ be a Hahn decomposition. Let

$$\alpha_+(E) = \alpha(E \cap A) \quad \text{and} \quad \alpha_-(E) = -\alpha(E \cap B).$$

Since A is positive and B is negative, α_+ and α_- are measures. Indeed, $\alpha_+(A^c) = 0$ and $\alpha_-(A) = 0$, so they are mutually singular. To prove uniqueness, suppose $\alpha = \nu_1 - \nu_2$ and D is a set with $\nu_1(D) = 0$ and $\nu_2(D^c) = 0$. If we set $C = D^c$, then $\Omega = C \cup D$ is a Hahn decomposition, and it follows from the choice of D that

$$\nu_1(E) = \alpha(C \cap E) \quad \text{and} \quad \nu_2(E) = -\alpha(D \cap E).$$

Our uniqueness result for the Hahn decomposition shows that $A \cap D = A \cap C^c$ and $B \cap C = A^c \cap C$ are null sets, so

$$\alpha(E \cap C) = \alpha(E \cap (A \cup C)) = \alpha(E \cap A)$$

and $\nu_1 = \alpha_+$. □

Our final decomposition is

Theorem 4.10 (Lebesgue Decomposition). Let μ, ν be σ-finite measures. Then ν can be written as $\nu_r + \nu_s$, where $\nu_s \perp \mu$, and

$$\mu_r(E) = \int_E g \, \mathrm{d}\mu\,.$$

Proof. By decomposing $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$, we can suppose without loss of generality that μ and ν are finite measures.

Let

$$\mathcal{G} = \left\{g \in L^0(\Omega, \Sigma, \mu) \colon g \geq 0, \int_E g \, \mathrm{d}\mu \leq \nu(E) \; \forall E \in \Sigma\right\}.$$

Claim 1. If $g, h \in \mathcal{G}$ then $g \vee h \in \mathcal{G}$.

Proof. Let $A = \{g > h\}$ and $B = A^c$. Then

$$\int_E g \vee h \, \mathrm{d}\mu = \int_{E \cap A} g \, \mathrm{d}\mu + \int_{E \cap B} h \, \mathrm{d}\mu \leq \nu(E \cap A) + \nu(E \cap B) = \nu(E).$$

Thus $g \vee h \in \mathcal{G}$. ■

Let

$$\kappa = \sup\left\{\int_\Omega g \, \mathrm{d}\mu : g \in \mathcal{G}\right\} \leq \nu(\Omega) < \infty.$$

Pick g_n so that $\int_\Omega g_n \, \mathrm{d}\mu > \kappa - n^{-1}$ and let $h_n = \min\{g_1, \ldots, g_n\}$. By Claim 1, $h_n \in \mathcal{G}$. And $\lim_{n\to\infty} h_n = h$. Then the defintiion of κ, the MCT Theorem 1.58, and the choice of g_n imply that

$$\kappa \geq \int_\Omega h \, \mathrm{d}\mu = \lim_{n\to\infty} \int_\Omega h_n \, \mathrm{d}\mu \geq \lim_{n\to\infty} \int_\Omega g_n \, \mathrm{d}\mu = \kappa.$$

Let

$$\nu_r(E) = \int_E h \, \mathrm{d}\mu \quad \text{and} \quad \nu_s(E) = \nu(E) - \nu_r(E).$$

Claim 2. ν_s is singular with respect to μ.

Proof. Let $\varepsilon > 0$ and let $\Omega = A_\varepsilon \cup B_\varepsilon$ be a Hahn decomposition for $\nu_s - \varepsilon\mu$. Since A_ε is positive for $\nu_s - \varepsilon\mu$, $\varepsilon\mu(A_\varepsilon \cap E) \leq \nu_s(A_\varepsilon \cap E)$. Using the definition of ν_r and this fact,

$$\int_E (h + \varepsilon \mathbb{1}_{A_\varepsilon}) \, \mathrm{d}\mu = \nu_r(E) + \varepsilon\mu(A_\varepsilon \cap E) \leq \nu_r(E) + \nu_s(A_\varepsilon \cap E) \leq \nu_r(E) + \nu_s(E) = \nu(E).$$

This holds for all E, so $k = h + \varepsilon \mathbb{1}_{A_\varepsilon} \in \mathcal{G}$. It follows that $\mu(A_\varepsilon) = 0$, for if not, then $\int_\Omega k \, \mathrm{d}\mu > \kappa$, a contradiction. Letting $A = \bigcup_{n=1}^\infty A_{1/n}$, we have $\mu(A) = 0$. Now observe that if $\nu_s(A^c) > 0$, then $(\nu_s - \varepsilon\mu)(A^c) > 0$ for small ε, a contradiction since $A^c \subset B_\varepsilon$, a negative set for $\nu_s - \varepsilon\mu$. Thus $\nu_s(A^c) = 0$, so $\mu \perp \nu_s$. ■

□

Definition 4.11 (Absolute Continuity). A measure ν is **absolutely continuous** with respect to μ (and write $\nu \ll \mu$) if $\mu(A) = 0$ implies $\nu(A) = 0$.

Theorem 4.12 (Radon-Nikodym Theorem). If μ, ν are σ-finite measures and ν is absolutely continuous with respect to μ, then there is a $g \geq 0$ so that $\nu(E) = \int_E g \, \mathrm{d}\mu$. If h is another such function then $g = h$ μ-a.e..

Proof. Let $\nu = \nu_r + \nu_s$ be a Lebesgue decomposition of ν. Let A be chosen so that $\nu_s(A^c) = 0$ and $\mu(A) = 0$. Since $\nu \ll \mu$, $0 = \nu(A) \geq \nu_s(A)$. So $\nu_s(A) = \nu_s(A^c) = 0$, hence $\nu_s \equiv 0$. Therefore $\nu(A) = \nu_r(A)$ which has the desired representation.

To prove uniqueness a.e., observe that if $\int_E g \, \mathrm{d}\mu = \int_E h \, \mathrm{d}\mu$ for all E, then letting $E \subset \{g > h, g \leq n\}$ be any subset of finite measure, we conclude $\mu(\{g > h, g \leq n\}) = 0$ for all n, so $\mu(\{g > h\}) = 0$. Similarly $\mu(\{g < h\}) = 0$, so $\mu(\{g \neq h\}) = 0$ as desired. □

Remark. The function g is usually written as $g = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. This suggests a host of derivative properties, which we do not prove here.

1. If $\nu_1, \nu_2 \ll \mu$ then $\nu_1 + \nu_2 \ll \mu$, and
$$\frac{\mathrm{d}(\nu_1 + \nu_2)}{\mathrm{d}\mu} = \frac{\mathrm{d}\nu_1}{\mathrm{d}\mu} + \frac{\mathrm{d}\nu_2}{\mathrm{d}\mu}.$$
2. If $\nu \ll \mu$ and $f \geq 0$ then
$$\int_\Omega f \, \mathrm{d}\nu = \int_\Omega f \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \, \mathrm{d}\mu\,.$$
3. If $\pi \ll \nu \ll \mu$ then
$$\frac{\mathrm{d}\pi}{\mathrm{d}\mu} = \frac{\mathrm{d}\pi}{\mathrm{d}\nu} \frac{\mathrm{d}\nu}{\mathrm{d}\mu}.$$
4. If $\nu \ll \mu$ and $\mu \ll \nu$ then
$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu} = \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)^{-1}.$$

4.2 Conditional Expectation

Definition 4.13 (Conditional Expectation). Let $X \in L^1(\Sigma)$. Let $\Sigma_{\text{small}} \subseteq \Sigma$ be a σ-algebra. Then the **conditional expectation of X given Σ_{small}**, $\mathbb{E}[X \mid \Sigma_{\text{small}}]$, as any random variable Y that has

(i) $Y \in L^0(\Sigma_{\text{small}})$.

(ii) For all $A \in \Sigma_{\text{small}}$, $\mathbb{E}[X; A] = \mathbb{E}[Y; A]$.

Any Y satisfying (i) and (ii) is said to be a **version of $\mathbb{E}[X \mid \Sigma_{\text{small}}]$**.

Proposition 4.14. The conditional expectation $\mathbb{E}[X \mid \Sigma_{\text{small}}]$ exists.

Proof. Suppose first that $X \geq 0$ and define the measure

$$\nu(A) = \mathbb{E}[X; A] \quad \text{for } A \in \Sigma_{\text{small}}.$$

Then the DCT Theorem 1.59 implies ν is a measure. And the definition of the integral implies $\nu \ll \mu$. By the Radon-Nikodym theorem Theorem 4.12, $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^0(\Sigma_{\text{small}})$, and for any $A \in \Sigma_{\text{small}}$,

$$\mathbb{E}[X; A] = \nu(A) = \mathbb{E}\left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu}; A\right].$$

Taking $A = \Omega$, we see that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \geq 0$ is integrable (that is, $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^1(\Sigma_{\text{small}})$). So $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ is a version of $\mathbb{E}[X \mid \Sigma_{\text{small}}]$.

To treat the general case now, write $X = X^+ - X^-$, let $Y_1 = \mathbb{E}[X^+ \mid \Sigma_{\text{small}}]$ and $Y_2 = \mathbb{E}[X^- \mid \Sigma_{\text{small}}]$. Now $Y_1 - Y_2 \in L^1(\Sigma_{\text{small}})$, and for all $A \in \Sigma_{\text{small}}$ we have

$$\begin{aligned}\mathbb{E}[X; A] &= \mathbb{E}\left[X^+; A\right] - \mathbb{E}\left[X^-; A\right] = \mathbb{E}[Y_1; A] - \mathbb{E}[Y_2; A] \\ &= \mathbb{E}[Y_1 - Y_2; A].\end{aligned}$$

This shows that $Y_1 - Y_2$ is a version of $\mathbb{E}[X \mid \Sigma_{\text{small}}]$ and completes the proof. □

Proposition 4.15. If $X_1 = X_2$ on $B \in \Sigma_{\text{small}}$ then $\mathbb{E}[X_1 \mid \Sigma_{\text{small}}] \overset{\text{a.s.}}{=} \mathbb{E}[X_2 \mid \Sigma_{\text{small}}]$ on B.

Proof. We start with a technical point.

Lemma 1. If Y satisfies (i) and (ii) of the conditional expectation conditions, then $Y \in L^1(\Sigma_{\text{small}})$.

Proof. Let $A = \{Y > 0\} \in \Sigma_{\text{small}}$. Using (ii) twice, we get

$$\begin{aligned}\mathbb{E}[Y; A] &= \mathbb{E}[X; A] \leq \mathbb{E}[|X|; A] \\ \mathbb{E}[-Y; A^c] &= \mathbb{E}[-X; A^c] \leq \mathbb{E}[|X|; A^c]\end{aligned}$$

and summing gets

$$\begin{aligned}\mathbb{E}[Y; A] - \mathbb{E}[Y; A^c] &= \mathbb{E}[|X|; A] + \mathbb{E}[|X|; A^c] \\ \mathbb{E}[|Y|] &= \mathbb{E}[|X|].\end{aligned}$$

Thus $\mathbb{E}[|Y|] \leq \mathbb{E}[|X|] < \infty$, so $Y \in L^1(\Sigma_{\text{small}})$. ■

Now let $Y_1 = \mathbb{E}[X_1 \mid \Sigma_{\text{small}}]$ and $Y_2 = \mathbb{E}[X_2 \mid \Sigma_{\text{small}}]$. Take $A = \{Y_1 - Y_2 \geq \varepsilon > 0\}$, we see

$$0 = \mathbb{E}[X_1 - X_2; A \cap B] = \mathbb{E}[Y_1 - Y_2; A \cap B] \geq \varepsilon \mathbb{P}(A).$$

Thus $\mathbb{P}(A) = 0$ independently of ε. Thus $Y_1 = Y_2$ a.s. on B as desired. □

Remark. Taking $B = \Omega$ is the more traditional independence result; if $X_1 \overset{\text{a.s.}}{=} X_2$ then $\mathbb{E}[X_1 \mid \Sigma_{\text{small}}] \overset{\text{a.s.}}{=} \mathbb{E}[X_2 \mid \Sigma_{\text{small}}]$.

Remark. Technically, all equalities such as $Y = \mathbb{E}[X \mid \Sigma_{\text{small}}]$ should have been written as $Y \overset{\text{a.s.}}{=} \mathbb{E}[X \mid \Sigma_{\text{small}}]$ but we have ignored this point in previous chapters and will continue to do so.

(After all, who cares what happens on measure 0 sets?)

Remark. If $X \in L^0(\Sigma_{\text{small}})$, then $\mathbb{E}[X \mid \Sigma_{\text{small}}] = X$; X always satisfies condition (ii) of Definition 4.13, and it's left to confirm condition (i), which is true by assumption.

We will state and prove analogies of regular expectation properties for conditional expectation.

Theorem 4.16. Let $X, Y \in L^0(\Sigma)$ and $\Sigma_{\text{small}} \subseteq \Sigma$ be a σ-algebra.

1. If $X, Y \in L^1(\Sigma)$ then conditional expectation is linear:
$$\mathbb{E}[aX + Y \mid \Sigma_{\text{small}}] = a\mathbb{E}[X \mid \Sigma_{\text{small}}] + \mathbb{E}[Y \mid \Sigma_{\text{small}}].$$

2. If $X, Y \in L^1(\Sigma)$ then conditional expectation is monotone:
$$X \leq Y \implies \mathbb{E}[X \mid \Sigma_{\text{small}}] \leq \mathbb{E}[Y \mid \Sigma_{\text{small}}].$$

3. If $\{X_n\} \geq 0$ and $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n \uparrow X$ with $X \in L^1(\Sigma)$ then
$$\overset{\text{a.s.}}{\lim_{n\to\infty}} \mathbb{E}[X_n \mid \Sigma_{\text{small}}] \uparrow \mathbb{E}[X \mid \Sigma_{\text{small}}].$$

Proof.

1. To prove (a), we need to check that the right-hand side is a version of the left. Clearly $a\mathbb{E}[X \mid \Sigma_{\text{small}}] + \mathbb{E}[Y \mid \Sigma_{\text{small}}] \in L^0(\Sigma_{\text{small}})$, confirming (i) of the conditional expectation condition. To check (ii), if $A \in \Sigma_{\text{small}}$ then by linearity of the integral,
$$\begin{aligned}\mathbb{E}[a\mathbb{E}[X \mid \Sigma_{\text{small}}] + \mathbb{E}[Y \mid \Sigma_{\text{small}}]; A] &= a\mathbb{E}[\mathbb{E}[X \mid \Sigma_{\text{small}}]; A] + \mathbb{E}[\mathbb{E}[Y \mid \Sigma_{\text{small}}]; A] \\ &= a\mathbb{E}[X; A] + \mathbb{E}[Y; A] = \mathbb{E}[aX + Y; A] \quad \text{for } A \in \Sigma_{\text{small}}.\end{aligned}$$
Thus (1) is proven.

2. To prove (b), we note that the integral is monotone, so for $A \in \Sigma_{\text{small}}$,
$$\mathbb{E}[\mathbb{E}[X \mid \Sigma_{\text{small}}]; A] = \mathbb{E}[X; A] \leq \mathbb{E}[Y; A] = \mathbb{E}[\mathbb{E}[Y \mid \Sigma_{\text{small}}]; A].$$
Letting $A = \{\mathbb{E}[X \mid \Sigma_{\text{small}}] - \mathbb{E}[Y \mid \Sigma_{\text{small}}] \geq \varepsilon > 0\}$, we see that $\mathbb{P}(A) = 0$ for all $\varepsilon > 0$. Thus (2) is proven.

3. Let $Y_n = X - X_n$ and $Z_n = \mathbb{E}[Y_n \mid \Sigma_{\text{small}}]$. It suffices to show that $\lim_{n\to\infty} Z_n = 0$. Since $\lim^{\text{a.s.}}_{n\to\infty} Y_n \downarrow 0$, (2) implies $\lim^{\text{a.s.}}_{n\to\infty} Z_n \downarrow Z$, where Z is defined as this limit. If $A \in \Sigma_{\text{small}}$ then by DCT Theorem 1.59,
$$\mathbb{E}[Z_n; A] = \mathbb{E}[\mathbb{E}[Y_n \mid \Sigma_{\text{small}}]; A] = \mathbb{E}[Y_n; A] \implies \mathbb{E}[Z; A] = \lim_{n\to\infty} \mathbb{E}[Z_n; A] = \lim_{n\to\infty} \mathbb{E}[Y_n; A] = 0.$$
Thus $Z = 0$ and (3) is proven.

$\square$

Theorem 4.17 (Conditional Jensen's Inequality). If ϕ is convex and $X \in L^1(\Sigma)$ and $\phi \circ X \in L^1(\Sigma)$, then for $\Sigma_{\text{small}} \subseteq \Sigma$ a

σ-algebra,

$$\phi(\mathbb{E}[X \mid \Sigma_{\text{small}}]) \leq \mathbb{E}[\phi(X) \mid \Sigma_{\text{small}}].$$

Proof. If ϕ is linear, say $\phi(x) = ax + b$. Then

$$a\mathbb{E}[X \mid \Sigma_{\text{small}}] + b = \mathbb{E}[aX + b \mid \Sigma_{\text{small}}] = \mathbb{E}[\phi(X) \mid \Sigma_{\text{small}}]$$

which proves the result. Now suppose ϕ is non-linear and let

$$S = \left\{(a, b) \in \mathbb{Q}^2 \colon ax + b \leq \phi(x) \text{ for all } x \in \mathbb{R}\right\}.$$

Then

$$\phi(x) = \sup\{ax + b \colon (a, b) \in S\}.$$

If $\phi(x) \geq ax + b$ then

$$\mathbb{E}[\phi(X) \mid \Sigma_{\text{small}}] \geq \mathbb{E}[aX + b \mid \Sigma_{\text{small}}] = a\mathbb{E}[X \mid \Sigma_{\text{small}}] + b \quad \text{a.s.}$$

and taking the supremum over $(a, b) \in S$ gives

$$\mathbb{E}[\phi(X) \mid \Sigma_{\text{small}}] \geq \sup_{(a,b) \in S} a\mathbb{E}[X \mid \Sigma_{\text{small}}] + b = \phi(\mathbb{E}[X \mid \Sigma_{\text{small}}])\text{a.s.},$$

as desired. □

Remark. In this case there is a set where convergence does not hold for each $(a, b) \in S$; to maintain a.s. convergence, we need to take the supremum over a countable set, as then the set where convergence does not hold for the final inequality will be (a subset of) a countable union of countable sets, and thus countable and measure 0.

Theorem 4.18. Conditional expectation is a contraction in L^p for $p \geq 1$.

Proof. Let $\phi(x) = |x|^p$. Then $\phi(x)$ is convex. Let $X \in L^1(\Sigma)$ and let $\Sigma_{\text{small}} \subseteq \Sigma$ be a σ-algebra. Then Theorem 4.17 implies

$$|\mathbb{E}[X \mid \Sigma_{\text{small}}]|^p \leq \mathbb{E}[|X|^p \mid \Sigma_{\text{small}}] \implies \mathbb{E}[|\mathbb{E}[X \mid \Sigma_{\text{small}}]|^p] \leq \mathbb{E}[\mathbb{E}[|X|^p \mid \Sigma_{\text{small}}]] = \mathbb{E}[|X|^p]$$

where the last equality is a consequence of Definition 4.13 with $A = \Omega$. □

Theorem 4.19. Suppose $\Sigma_1 \subseteq \Sigma_2 \subseteq \Sigma$ are σ-algebras and $\mathbb{E}[X \mid \Sigma_2] \in L^0(\Sigma_1)$ then $\mathbb{E}[X \mid \Sigma_1] = \mathbb{E}[X \mid \Sigma_2]$.

Proof. We show that $\mathbb{E}[X \mid \Sigma_2]$ is a version of $\mathbb{E}[X \mid \Sigma_1]$. By assumption $\mathbb{E}[X \mid \Sigma_2] \in L^0(\Sigma_1)$. To check the other part of the definition we note that if $A \in \Sigma_1 \subseteq \Sigma_2$ then

$$\mathbb{E}[\mathbb{E}[X \mid \Sigma_1]; A] = \mathbb{E}[X; A] = \mathbb{E}[\mathbb{E}[X \mid \Sigma_2]; A].$$

□

Theorem 4.20 (Tower Property). If $\Sigma_1 \subseteq \Sigma_2 \subseteq \Sigma$ are σ-algebras then

1. $\mathbb{E}[\mathbb{E}[X \mid \Sigma_1] \mid \Sigma_2] = \mathbb{E}[X \mid \Sigma_1]$.
2. $\mathbb{E}[\mathbb{E}[X \mid \Sigma_2] \mid \Sigma_1] = \mathbb{E}[X \mid \Sigma_1]$.

In other words, the smaller σ-field always wins.

Proof of Theorem 4.20.

1. Since $\Sigma_1 \subseteq \Sigma_2$, $\mathbb{E}[X \mid \Sigma_1] \in L^0(\Sigma_2)$. Then by Definition 4.13,
$$\mathbb{E}[\mathbb{E}[X \mid \Sigma_1] \mid \Sigma_2] = \mathbb{E}[X \mid \Sigma_1]$$
as desired.

2. Notice that $\mathbb{E}[X \mid \Sigma_1] \in L^0(\Sigma_1)$. And if $A \in \Sigma_1 \subseteq \Sigma_2$ then
$$\mathbb{E}[\mathbb{E}[X \mid \Sigma_1]; A] = \mathbb{E}[X; A] = \mathbb{E}[\mathbb{E}[X \mid \Sigma_2]; A].$$
Thus it follows that $\mathbb{E}[X \mid \Sigma_1]$ is a version of $\mathbb{E}[\mathbb{E}[X \mid \Sigma_2] \mid \Sigma_1]$.

□

The next result shows that for conditional expectations with respect to Σ_{small}, random variables $X \in L^0(\Sigma_{\text{small}})$ are like constants.

Theorem 4.21. If $X \in L^0(\Sigma_{\text{small}})$ and $Y, XY \in L^1(\Sigma)$, then
$$\mathbb{E}[XY \mid \Sigma_{\text{small}}] = X\mathbb{E}[Y \mid \Sigma_{\text{small}}].$$

Proof. We have $X\mathbb{E}[Y \mid \Sigma_{\text{small}}] \in L^0(\Sigma_{\text{small}})$, so we have to check (ii) of Definition 4.13. To do this, we use the usual four-step procedure. First, suppose $X = \mathbb{1}_B$ with $B \in \Sigma_{\text{small}}$. If $A \in \Sigma_{\text{small}}$ then
$$\mathbb{E}[X\mathbb{E}[Y \mid \Sigma_{\text{small}}]; A] = \mathbb{E}[\mathbb{E}[Y \mid \Sigma_{\text{small}}]; A \cap B] = \mathbb{E}[Y; A \cap B] = \mathbb{E}[XY; A],$$
so (ii) holds. The last result extends to simple X by linearity. If $X, Y \geq 0$, let X_n be simple random variables that increase to X a.s., and use the MCT Theorem 1.58 to conclude that
$$\mathbb{E}[X\mathbb{E}[Y \mid \Sigma_{\text{small}}]; A] = \mathbb{E}[XY; A].$$
To prove the result in general, split X and Y into their positive and negative parts. □

Theorem 4.22. Let $X \in L^2(\Sigma)$. Then for $\Sigma_{\text{small}} \subseteq \Sigma$ a σ-algebra, $\mathbb{E}[X \mid \Sigma_{\text{small}}]$ is the variable $Y \in L^0(\Sigma_{\text{small}})$ that minimizes the mean square error $\mathbb{E}\left[(X - Y)^2\right]$.

Proof. We begin by observing that if $Z \in L^2(\Sigma_{\text{small}})$ then Theorem 4.21 implies
$$Z\mathbb{E}[X \mid \Sigma_{\text{small}}] = \mathbb{E}[ZX \mid \Sigma_{\text{small}}] < \infty.$$
Taking expectations gives
$$\mathbb{E}[Z\mathbb{E}[X \mid \Sigma_{\text{small}}]] = \mathbb{E}[\mathbb{E}[ZX \mid \Sigma_{\text{small}}]] = \mathbb{E}[ZX],$$
or rearranging,
$$\mathbb{E}[Z(X - \mathbb{E}[X \mid \Sigma_{\text{small}}])] = 0 \quad \text{for } Z \in L^2(\Sigma_{\text{small}}).$$
If $Y \in L^2(\Sigma_{\text{small}})$ and $Z = \mathbb{E}[X \mid \Sigma_{\text{small}}] - Y$ then
$$\mathbb{E}\left[(X - Y)^2\right] = \mathbb{E}\left[(X - \mathbb{E}[X \mid \Sigma_{\text{small}}] + Z)^2\right] = \mathbb{E}\left[(X - \mathbb{E}[X \mid \Sigma_{\text{small}}])^2\right] + \mathbb{E}\left[Z^2\right].$$
It is easy to see $\mathbb{E}\left[(X - Y)^2\right]$ is minimized when $Z = 0$. □

4.3 Regular Conditional Probability

Let $(S, \mathcal{S})$ be a measurable space and $X \colon (\Omega, \Sigma) \to (S, \mathcal{S})$ a measurable map, and $\Sigma_{\text{small}} \subseteq \Sigma$ a σ-algebra.

Definition 4.23 (Regular Conditional Distribution). $\mu\colon \Omega \times S \to [0,1]$ is a **regular conditional distribution** for X given Σ_{small} if

(i) For each $A \in \Sigma$, $\omega \mapsto \mu(\omega, A)$ is a version of $\mathbb{P}(X \in A \mid \Sigma_{\text{small}})$.

(ii) For a.s. ω, $A \mapsto \mu(\omega, A)$ is a probability measure on $(S, \mathcal{S})$.

When $S = \Omega$ and X is the identity map, μ is called a **regular conditional probability**.

Regular conditional distributions are useful because they allow us to simultaneously compute the conditional expectation of all functions of X and to generalize properties of ordinary expectation in a more straightforward way.

Theorem 4.24. Let $\mu(\Omega, A)$ be a r.c.d. for X given Σ_{small}. If $f\colon (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ has $\mathbb{E}[|f(X)|] < \infty$ then

$$\mathbb{E}[f(X) \mid \Sigma_{\text{small}}](\omega) = \int_{\Omega} f(x)\, \mathrm{d}\mu(\omega, x) \quad \text{a.s.}$$

Proof. If $f = \mathbb{1}_A$ this follows from the definition. Linearity extends the result to simple f and monotone convergence to nonnegative f. Finally we get the result in general by writing $f = f^+ - f^-$. $\square$

Unfortunately, r.c.d.'s don't always exist. If $\{A_n\}$ are disjoint, then

$$\mathbb{P}\left(X \in \bigcup_{n=1}^{\infty} A_n \,\middle|\, \Sigma_{\text{small}}\right) = \sum_{n=1}^{\infty} \mathbb{P}(X \in A_n \mid \Sigma_{\text{small}}) \quad \text{a.s.}$$

but if $\mathcal{S}$ contains enough countable collections of disjoint sets, the exceptional sets may pile up. We will state but not prove results regarding when r.c.d.'s exist.

Theorem 4.25. If $(S, \mathcal{S})$ is *nice*, in that there is an injective map $\phi\colon S \to \mathbb{R}$ so that ϕ and ϕ^{-1} are measurable, then r.c.d.'s exist.

Theorem 4.26. Suppose X and Y take values in a nice space $(S, \mathcal{S})$. Then there is a function $\mu\colon S \times \mathcal{S} \to [0,1]$ so that

(i) For each A, $\omega \mapsto \mu(Y(\omega), A)$ is a version of $\mathbb{P}(X \in A \mid \sigma(Y))$.

(ii) For a.s. ω, $A \mapsto \mu(\omega, A)$ is a probability measure on $(S, \mathcal{S})$.

4.4 Martingales

Definition 4.27 (Stochastic Process). A sequence of random variables $\mathcal{X} = \{X_t\}_{t \in I} \in L^0(\Sigma)$ indexed by a *time variable* $t \in I \subseteq \mathbb{R}$ is a **stochastic process**. If $|I| \leq |\mathbb{N}|$ then it's a **discrete stochastic process**, otherwise a **continuous stochastic process**. A **trajectory** is a sequence $\mathcal{X}(\omega) = \{X_t(\omega)\}_{t \in I}$ for a fixed $\omega \in \Omega$.

Remark. In the future we will assume $t \in I$ is implied.

Remark. If we write $\mathcal{X} \in L^p(\Sigma_0)$ then that means $X_t \in L^p(\Sigma_0)$ for all t. And most things we do are componentwise, that is, $\mathcal{X} \geq 0$ if and only if $X_t \geq 0$ for each t, by $\phi(\mathcal{X})$ we mean $\{\phi(X_t)\}$, etc..

Definition 4.28 (Filtration). A **filtration** $\mathcal{F} = \{\Sigma_t\}$ is an increasing sequence of σ-algebras, that is, if $t_1 \leq t_2$ then $\Sigma_{t_1} \subseteq \Sigma_{t_2}$. A sequence $\mathcal{X}$ is **adapted** to $\mathcal{F}$ if $X_t \in L^0(\Sigma_t)$ for all t.

Now we let $I = \mathbb{N} \cup \{0\}$, and study the case of discrete martingales. Note, our sequences start at 1; we usually define X_0 separately.

Definition 4.29 (Martingale). Let $\mathcal{X} \in L^0(\Sigma)$ be a random sequence and $\mathcal{F}$ a filtration. Then if

(i) $\mathcal{X} \in L^1(\Sigma)$,

(ii) $\mathcal{X}$ is adapted to $\mathcal{F}$,

(iii) $\mathbb{E}[X_{n+1} \,|\, \Sigma_n] = X_n$ for all n,

then X is said to be a **martingale** (with respect to $\mathcal{F}$). If $\mathbb{E}[X_{n+1} \,|\, \Sigma_n] \leq X_n$ then $\{X_n\}$ is a **supermartingale**; if $\mathbb{E}[X_{n+1} \,|\, \Sigma_n] \geq X_n$ then $\{X_n\}$ is a **submartingale**.

Definition 4.30 (Predictable Process). The sequence $\mathcal{H} \in L^1(\Sigma)$ is a **predictable process** if $H_n \in L^0(\Sigma_{n-1})$ for all n.

Remark. Let $m \geq n$. Since $X_n \in L^1(\Sigma_n)$, and $\mathcal{F} = \{\Sigma_n\}$ is a filtration and hence increasing, $X_n \in L^1(\Sigma_m)$. Thus

$$\mathbb{E}[X_n \,|\, \Sigma_m] = X_n \quad \text{a.s. for } m > n.$$

Thus it's only interesting to talk about $m < n$.

We will talk about some examples related to random walks.

Example 4.31 (Linear Martingale). Let X_0 be a constant and $\mathcal{X} \in L^1(\Sigma)$ be an i.i.d. random sequence. Let $S_n = \sum_{m=0}^{n} X_n$. Let $\mu = \mathbb{E}[X_1]$. Then $\{S_n - n\mu\}$ is a martingale with respect to $\mathcal{F}$. Indeed,

$$\begin{aligned}
\mathbb{E}[S_{n+1} - (n+1)\mu \,|\, \Sigma_n] &= \mathbb{E}[S_{n+1} \,|\, \Sigma_n] - (n+1)\mu = \mathbb{E}[S_n + X_{n+1} \,|\, \Sigma_n] - (n+1)\mu = S_n + \mathbb{E}[X_{n+1} \,|\, \Sigma_n] - (n+1)\mu \\
&= (S_n - n\mu) + (\mathbb{E}[X_{n+1} \,|\, \Sigma_n] - \mu) = (S_n - n\mu) + (\mathbb{E}[X_{n+1}] - \mu) = (S_n - n\mu) + (\mu - \mu) \\
&= S_n - n\mu.
\end{aligned}$$

It follows that if $\mu = 0$ then $\mathcal{S}$ is a martingale, if $\mu < 0$ then $\mathcal{S}$ is a supermartingale, and if $\mu > 0$ then $\mathcal{S}$ is a submartingale.

Example 4.32 (Quadratic Martingale). Let X_0 be a constant and $\mathcal{X} \in L^1(\Sigma)$ be an i.i.d. random sequence with $\mathbb{E}[X_1] = 0$. Let $S_n = \sum_{m=0}^{n} X_n$. Let $\sigma^2 = \operatorname{Var}(X_1)$. Then $\left\{S_n^2 - n\sigma^2\right\}$ is a martingale. Indeed,

$$\begin{aligned}
\mathbb{E}\left[S_{n+1}^2 - (n+1)\sigma^2 \,\middle|\, \Sigma_n\right] &= \mathbb{E}\left[(S_n + X_{n+1})^2 - (n+1)\sigma^2 \,\middle|\, \Sigma_n\right] = \mathbb{E}\left[S_n^2 + 2X_{n+1}S_n + X_{n+1}^2 - (n+1)\sigma^2 \,\middle|\, \Sigma_n\right] \\
&= S_n^2 - (n+1)\sigma^2 + 2\mathbb{E}[X_{n+1}S_n \,|\, \Sigma_n] + \mathbb{E}\left[X_{n+1}^2 \,\middle|\, \Sigma_n\right] \\
&= S_n^2 - (n+1)\sigma^2 + 2S_n\mathbb{E}[X_{n+1} \,|\, \Sigma_n] + \mathbb{E}\left[X_{n+1}^2 \,\middle|\, \Sigma_n\right] \\
&= S_n^2 - (n+1)\sigma^2 + 2S_n\mathbb{E}[X_{n+1}] + \mathbb{E}\left[X_{n+1}^2\right] \\
&= S_n^2 - (n+1)\sigma^2 + \operatorname{Var}(X_{n+1}) = S_n^2 - n\sigma^2.
\end{aligned}$$

Example 4.33 (Exponential Martingale). Let $\mathcal{Y} \in L^1(\Sigma)$ be a non-negative i.i.d. random sequence with $\mathbb{E}[Y_1] = 1$, and let $\Sigma_n = \sigma(Y_m \colon m \in [n])$. Then $M_n = \prod_{m=1}^{n} Y_m$ defines a martingale. Indeed,

$$\mathbb{E}[M_{n+1} \,|\, \Sigma_n] = \mathbb{E}[M_n Y_{n+1} \,|\, \Sigma_n] = M_n\mathbb{E}[Y_{n+1} \,|\, \Sigma_n] = M_n\mathbb{E}[Y_{n+1}] = M_n.$$

Now suppose that $\mathcal{X} \in L^\infty(\Sigma)$ is an i.i.d. random sequence and $S_n = \sum_{m=1}^{n} X_m$. Let $\psi(\theta) = \mathbb{E}\left[e^{\theta X_1}\right]$ and $Y_n = \frac{\exp(\theta X_n)}{\psi(\theta)}$. Then $Y_n \geq 0$ and $\mathbb{E}[Y_n] = 1$, so if

$$M_n = \prod_{m=1}^{n} Y_m = \frac{\exp(\theta S_n)}{\psi(\theta)^n}$$

then $\mathcal{M}$ is a martingale.

From here onwards, we fix a filtration $\mathcal{F}$; all (sub- or super-) martingales are with respect to this filtration.

Theorem 4.34.

1. If $\mathcal{X}$ is a supermartingale then for $n > m$, $\mathbb{E}[X_n \mid \Sigma_m] \leq X_m$.

2. If $\mathcal{X}$ is a submartingale then for $n > m$, $\mathbb{E}[X_n \mid \Sigma_m] \geq X_m$.

3. If $\mathcal{X}$ is a martingale then for $n > m$, $\mathbb{E}[X_n \mid \Sigma_m] = X_m$.

Proof.

1. For $n = m + 1$, the definition Definition 4.29 gives
$$\mathbb{E}[X_n \mid \Sigma_m] = \mathbb{E}[X_{m+1} \mid \Sigma_m] \leq X_m.$$
Suppose $n = m + k$ with $k \geq 2$. Then by Theorem 4.20,
$$\mathbb{E}[X_n \mid \Sigma_m] = \mathbb{E}[X_{m+k} \mid \Sigma_m] = \mathbb{E}[\mathbb{E}[X_{m+k} \mid \Sigma_{m+k-1}] \mid \Sigma_m] \leq \mathbb{E}[X_{m+k-1} \mid \Sigma_m].$$
The result follows from induction.

2. If $\mathcal{X}$ is a submartingale it's easy to see that $-\mathcal{X}$ is a supermartingale, and (1) gives the result.

3. If $\mathcal{X}$ is a martingale then $\mathcal{X}$ is both a submartingale and a supermartingale, so (1) and (2) give the result.

$\square$

The ideas in the proofs of (2) and (3) are very common, so we will state the result for submartingales or supermartingales and the remaining pair of results can follow easily.

We will discuss the creation of new martingales out of old martingales. We introduce the discrete convolution operator, between a predictable process $\mathcal{H}$ and a martingale $\mathcal{X}$:

$$(\mathcal{H} \cdot \mathcal{X})_n = \sum_{m=1}^{n} H_m (X_m - X_{m-1}).$$

Theorem 4.35. Let $\mathcal{X} \in L^1(\Sigma)$ be a supermartingale. If $\mathcal{H} \in L^\infty(\Sigma)$ is a predictable process and $\mathcal{H} \geq 0$, then $\mathcal{H} \cdot \mathcal{X}$ is a supermartingale.

Proof. Clearly $(\mathcal{H} \cdot \mathcal{X})_n \in L^1(\Sigma_n)$, so now we use the fact that $H_{n+1} \in L^\infty(\Sigma_n)$ to obtain

$$\begin{aligned}
\mathbb{E}\left[(\mathcal{H} \cdot \mathcal{X})_{n+1} \mid \Sigma_n\right] &= \mathbb{E}[(\mathcal{H} \cdot \mathcal{X})_n + H_{n+1}(X_{n+1} - X_n) \mid \Sigma_n] = (\mathcal{H} \cdot \mathcal{X})_n + \mathbb{E}[H_{n+1}(X_{n+1} - X_n) \mid \Sigma_n] \\
&= (\mathcal{H} \cdot \mathcal{X})_n + H_{n+1}\mathbb{E}[X_{n+1} - X_n \mid \Sigma_n] = (\mathcal{H} \cdot \mathcal{X})_n + H_{n+1}(\mathbb{E}[X_{n+1} \mid \Sigma_n] - X_n) \\
&\leq (\mathcal{H} \cdot \mathcal{X})_n
\end{aligned}$$

since $H_{n+1} \geq 0$ and by the supermartingale property $\mathbb{E}[X_{n+1} \,|\, \Sigma_n] \leq X_n$. Thus $\mathcal{H} \cdot \mathcal{X}$ is a supermartingale. $\square$

Remark. The same result is true for submartingales; for martingales, the same result holds with one difference. In particular $\mathbb{E}[X_{n+1} \,|\, \Sigma_n] = X_n$, so the coefficient of H_{n+1} is 0, and hence there is no restriction $H \geq 0$.

Definition 4.36 (Stopping Time). A random variable $\tau \in L^0(\Sigma)$ which takes values in $\mathbb{N} \cup \{\infty\}$ is a **stopping time** if $\{\tau = n\} \in \Sigma_n$ for all n.

Example 4.37. The variable τ_1, the "first time to hit a", is a stopping time, since the event that $\{\tau_1 = n\}$ depends only on the first n timesteps, so $\{\tau_1 = n\} \in L^0(\Sigma_n)$. The variable τ_2, the "last time to hit a", is *not* a stopping time, since the event that $\{\tau_2 = n\}$ depends on the whole sequence X. Thus $\{\tau_2 = n\} \notin L^0(\Sigma_n)$

Theorem 4.38. If τ is a stopping time and $\mathcal{X}$ is a supermartingale, then $\mathcal{X}_{\wedge\tau} = \{X_{n\wedge\tau}\}$ is a supermartingale.

Proof. Let $H_n = \mathbb{1}_{\{\tau \geq n\}}$. In particular $\{\tau \geq n\} = \{\tau \leq n-1\}^c \in \Sigma_{n-1}$, so H is predictable. Thus $\mathcal{H} \cdot \mathcal{X}$ is a supermartingale. But $(\mathcal{H} \cdot \mathcal{X})_n = X_{n\wedge\tau} - X_0$. The constant sequence $\mathcal{X}_0 = \{X_0\}$ is a supermartingale, and the sum of two supermartingales $\mathcal{H} \cdot \mathcal{X} + \mathcal{X}_0 = \mathcal{X}_{\wedge\tau}$ is a supermartingale. $\square$

Remark. The correct generalization is that if τ is a stopping time and $\mathcal{X}$ is a (sub)martingale, then $\mathcal{X}_{\wedge\tau}$ is a (sub)martingale.

Theorem 4.39. If $\mathcal{X}$ is a submartingale and τ is a stopping time with $\mathbb{P}(\tau \leq n) = 1$ then

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_\tau] \leq \mathbb{E}[X_n].$$

Proof. Theorem 4.38 implies $\mathcal{X}_{\wedge\tau}$ is a submartingale. It follows that

$$\mathbb{E}[X_0] = \mathbb{E}[X_{\tau\wedge 0}] \leq \mathbb{E}[X_{n\wedge\tau}] = \mathbb{E}[X_\tau].$$

To prove the other inequalikty, let $H_n = \mathbb{1}_{\tau \leq n-1}$. Then $\mathcal{H}$ is predictable, so Theorem 4.35 implies $\mathcal{H} \cdot \mathcal{X}$ is a submartingale. But $(\mathcal{H} \cdot \mathcal{X})_n = X_n - X_{n\wedge\tau}$, so it follows that

$$\mathbb{E}[X_n] - \mathbb{E}[X_\tau] = \mathbb{E}[(H \cdot X)_n] \geq \mathbb{E}[(H \cdot X)_0] = 0.$$

Thus $\mathbb{E}[X_0] \leq \mathbb{E}[X_\tau] \leq \mathbb{E}[X_n]$. $\square$

Remark. The correct generalization is that if $\mathcal{X}$ is a supermartingale and τ is a stopping time with $\mathbb{P}(\tau \leq n) = 1$ then

$$\mathbb{E}[X_0] \geq \mathbb{E}[X_\tau] \geq \mathbb{E}[X_n],$$

and the case for martingales has the conclusion

$$\mathbb{E}[X_0] = \mathbb{E}[X_\tau] = \mathbb{E}[X_n].$$

Remark. If $\mathcal{S}$ is a simple random walk with $S_0 = 1$, and $\tau = \inf\{n : S_n = 0\}$, then $\mathbb{E}[S_0] = 1$ but $\mathbb{E}[S_\tau] = 0$. The first inequality thus need not hold for unbounded stopping times.

Theorem 4.40. If $\mathcal{X}$ is a martingale and ϕ is a convex function with $\phi(\mathcal{X}) \in L^1(\Sigma)$, then $\phi(\mathcal{X})$ is a sub-martingale.

Proof. By conditional Jensen's inequality Theorem 4.17 and the definition,

$$\mathbb{E}[\phi(X_{n+1}) \,|\, \Sigma_n] \geq \phi(\mathbb{E}[X_{n+1} \,|\, \Sigma_n]) = \phi(X_n).$$

$\square$

Theorem 4.41. If $\mathcal{X}$ is a submartingale and ϕ is an increasing convex function with $\phi(\mathcal{X}) \in L^1(\Sigma)$, then $\phi(\mathcal{X})$ is a submartingale.

Proof. By conditional Jensen's inequality Theorem 4.17 and the assumptions,

$$\mathbb{E}[\phi(X_{n+1}) \,|\, \Sigma_n] \geq \phi(\mathbb{E}[X_{n+1} \,|\, \Sigma_n]) \geq \phi(X_n).$$

□

Remark. The correct generalization is that if $\mathcal{X}$ is a supermartingale and ϕ is a decreasing concave function with $\phi(\mathcal{X}) \in L^1(\Sigma)$, then $\phi(\mathcal{X})$ is a supermartingale.

Remark. Consequently,

- If $\mathcal{X}$ is a submartingale then $(\mathcal{X} - a)^+$ is a submartingale.
- If $\mathcal{X}$ is a supermartingale then $\mathcal{X} \wedge a$ is a supermartingale.

4.5 Inequalities and Convergence

We will develop a theory about which types of martingales converge and the modes of convergence. Define $\overline{X}_n = \sup_{0 \leq m \leq n} X_m^+$.

Theorem 4.42 (Doob's Maximal Inequality). Let $\mathcal{X}$ be a submartingale, let $a > 0$, and $A = \left\{\overline{X}_n \geq a\right\}$. Then

$$a\mathbb{P}(A) \leq \mathbb{E}[X_n; A] \leq \mathbb{E}\left[X_n^+\right].$$

Proof. Let $A = \left\{\overline{X}_n \geq a\right\}$. Let $\tau = \inf\{m \colon X_m \geq a\} \wedge n$. Since $X_\tau \geq a$ on A,

$$a\mathbb{P}(A) \leq \mathbb{E}[X_\tau; A].$$

Now Theorem 4.39 implies $\mathbb{E}[X_\tau] \leq \mathbb{E}[X_n]$, and we have $X_\tau = X_n$ on A^c. Thus

$$\mathbb{E}[X_\tau; A] + \mathbb{E}[X_\tau; A^c] = \mathbb{E}[X_\tau] \leq \mathbb{E}[X_n] = \mathbb{E}[X_n; A] + \mathbb{E}[X_n; A^c] \implies \mathbb{E}[X_\tau; A] \leq \mathbb{E}[X_n; A].$$

Now

$$\mathbb{E}[X_n; A] = \mathbb{E}\left[X_n \mathbb{1}\left(\overline{X}_n \geq a\right)\right] \leq \mathbb{E}[X_n; X_n \geq 0] = \mathbb{E}\left[X_n^+\right],$$

as desired. □

Remark. Let $\{X_n\} \in L^2(\Sigma)$ be independent with $\mathbb{E}[X_n] = 0$, $\sigma_n^2 = \mathbb{E}\left[X_n^2\right]$, and $S_n = \sum_{m=1}^{n} X_m$. Then $\mathcal{S}$ is a martingale, so Theorem 4.40 implies $\mathcal{S}^2$ is a submartingale. Letting $a = x^2$ and applying Doob's maximal inequality Theorem 4.42, we obtain

$$\begin{aligned}\mathbb{P}\left(\sup_{1 \leq m \leq n} |S_m| \geq x\right) &= \mathbb{P}\left(\sup_{1 \leq m \leq n} |S_m|^2 \geq x^2\right) \\ &\leq \frac{\mathbb{E}\left[S_n^2\right]}{x^2} = \frac{\mathrm{Var}(S_n)}{x^2}\end{aligned}$$

which is KMI Theorem 2.32.

Now we will discuss convergence of martingales almost surely. Suppose $\mathcal{X}$ is a submartingale. Let $a < b$, let $\tau_0 = -1$, and for $k \geq 1$ let

$$\tau_{2k-1} = \inf\{m > \tau_{2k-2} \colon X_m \leq a\}, \quad \tau_{2k} = \inf\{m > \tau_{2k-1} \colon X_m \geq b\}.$$

Then the τ_n are stopping times and

$$\{\tau_{2k-1} < m \leq \tau_{2k}\} = \{\tau_{2k-1} \leq m-1\} \cap \{\tau_{2k} \leq m-1\}^c \in \Sigma_{m-1},$$

so

$$H_m = \begin{cases} 1 & \tau_{2k-1} < m \leq \tau_{2k} \text{ for some } k \\ 0 & \text{otherwise} \end{cases}$$

defines a predictable sequence $\mathcal{H}$. Finally, let $U_n = \sup\{k \colon \tau_{2k} \leq n\}$ be the number of upcrossings competed by time n.

Theorem 4.43 (Uprcrossing Inequality). If $\mathcal{X}$ is a submartingale then

$$(b-a)\mathbb{E}[U_n] \leq \mathbb{E}\left[(X_n - a)^+\right] - \mathbb{E}\left[(X_0 - a)^+\right].$$

Proof. Let $Y_m = a + (X_m - a)^+$. By Theorem 4.41, $\mathcal{Y}$ is a submartingale. And $\mathcal{Y}$ upcrosses $[a,b]$ the same number of times that $\mathcal{X}$ does, and we have

$$(b-a)U_n \leq (\mathcal{H} \cdot \mathcal{Y})_n,$$

since each upcrossing results in a gain of $\geq b-a$ to the right hand side, and a final incomplete upcrossing adds a non-negative amount to the right side as well.

We have

$$Y_n - Y_0 = (\mathcal{H} \cdot \mathcal{Y})_n + ((1-\mathcal{H}) \cdot \mathcal{Y})_n$$

and it follows from Theorem 4.35 that

$$\mathbb{E}[((1-\mathcal{H}) \cdot \mathcal{Y})_n] \geq \mathbb{E}[((1-\mathcal{H}) \cdot \mathcal{Y})_0] = 0 \implies (b-a)\mathbb{E}[U_n] \leq \mathbb{E}[(\mathcal{H} \cdot \mathcal{Y})_n] \leq \mathbb{E}[Y_n] - \mathbb{E}[Y_0].$$

Therefore

$$(b-a)\mathbb{E}[U_n] \leq \mathbb{E}\left[(X_n - a)^+\right] - \mathbb{E}\left[(X_0 - a)^+\right]$$

as desired. $\square$

Theorem 4.44 (Submartingale Convergence Theorem). If $\mathcal{X} \in L^1(\Sigma)$ is a submartingale with $\sup_n \mathbb{E}\left[X_n^+\right] < \infty$ then there is a finite limit X for which $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$.

Proof. Since $(X_n - a)^+ \leq X_n^+ + |a|$, Theorem 4.43 implies that

$$\mathbb{E}[U_n] \leq \frac{\mathbb{E}[(X_n - a)^+] - \mathbb{E}[(X_0 - a)^+]}{b-a} \leq \frac{|a| + \mathbb{E}[X_n^+]}{b-a}.$$

Then $\lim_{n\to\infty} U_n = U$ the limit of upcrossings of $[a,b]$ by the whole sequence. So if $\sup_n \mathbb{E}[X_n^+] < \infty$ then $\mathbb{E}[U] < \infty$ and hence $U < \infty$ a.s.. Since the last conclusion holds for all rational a and b,

$$\mathbb{P}\left(\bigcup_{\substack{a,b\in\mathbb{Q} \\ a<b}} \left\{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right\}\right) = 0.$$

Hence

$$\limsup_{n\to\infty} X_n \overset{\text{a.s.}}{=} \liminf_{n\to\infty} X_n \implies \lim_{n\to\infty} X_n \text{ exists a.s..} \implies \overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X.$$

Fatou's lemma Theorem 1.57 guarantees

$$\mathbb{E}\left[X_\infty^+\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[X_n^+\right] < \infty \implies X \overset{\text{a.s.}}{<} \infty.$$

To see that $X > -\infty$, we observe that

$$\mathbb{E}\left[X_n^-\right] = \mathbb{E}\left[X_n^+\right] - \mathbb{E}[X_n] \leq \mathbb{E}\left[X_n^+\right] - \mathbb{E}[X_0]$$

since $\mathcal{X}$ is a submartingale. Another application of Fatou's lemma Theorem 1.57 shows

$$\mathbb{E}\left[X_\infty^-\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[X_n^-\right] \leq \sup_n \mathbb{E}\left[X_n^+\right] - \mathbb{E}[X_0] < \infty,$$

and completes the proof. $\square$

Remark. The correct generalization is that if $\mathcal{X} \in L^1(\Sigma)$ is a supermartingale with $\sup_n \mathbb{E}\left[X_n^-\right] < \infty$ then there is a limit X for which $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$. And for martingales we can pick either result.

Corollary 4.45. If $\mathcal{X} \in L^1(\Sigma)$ has $\mathcal{X} \geq 0$ and is a supermartingale then $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$ and $\mathbb{E}[X] \leq \mathbb{E}[X_0]$.

Proof. Let $Y_n = -X_n \leq 0$. Then $\mathcal{Y}$ is a submartingale with $\mathbb{E}[Y_n^+] = 0$. Convergence follows from Theorem 4.44. Since $\mathbb{E}[X_0] \geq \mathbb{E}[X_n]$, the inequality follows from Fatou's lemma Theorem 1.57. $\square$

Remark. The correct generalization is that if $\mathcal{X} \in L^1(\Sigma)$ has $\mathcal{X} \leq 0$ and is a submartingale then $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$ and $\mathbb{E}[X] \geq \mathbb{E}[X_0]$. And for martingales we can pick either result.

Martingales with bounded increments either converge or oscillate between $+\infty$ and $-\infty$.

Theorem 4.46. Let $\mathcal{X} \in L^1(\Sigma)$ be a martingale with $|X_{n+1} - X_n| \leq M < \infty$. Let

$$A = \left\{\lim_{n\to\infty} X_n \text{ exists and is finite}\right\} \quad \text{and} \quad B = \left\{\limsup_{n\to\infty} X_n = \infty \text{ and } \liminf_{n\to\infty} X_n = -\infty\right\}.$$

Then $\mathbb{P}(A \cup B) = 1$.

Proof. Let $0 < K < \infty$ and let $\tau = \inf\{n \colon X_n \leq -K\}$. Then $\mathcal{X}_{\wedge\tau}$ is a martingale with $\mathcal{X}_{\wedge\tau} \geq -K - M$ a.s. so applying Corollary 4.45 to $\mathcal{X}_{\wedge\tau} + K + M$ shows $\lim_{n\to\infty} X_n$ exists on $\{\tau = \infty\}$. Letting $K \to \infty$, we see that the limit exists on $\{\liminf_{n\to\infty} X_n > -\infty\}$. Applying the last conclusion to $-\mathcal{X}$, we see that $\lim_{n\to\infty} X_n$ exists on $\{\limsup_{n\to\infty} X_n < \infty\}$, and the proof is complete. $\square$

Example 4.47 (Counterexample in Absence of Bounded Increments). Let

$$X_n = \begin{cases} 2^n & \text{w.p. } 2^{-n} \\ -(1-2^{-n})^{-1} & \text{w.p. } 1 - 2^{-n} \end{cases}$$

independently. Then $\mathcal{S}$ is a martingale. By Theorem 2.27, it's easy to see that $\overset{\text{a.s.}}{\lim_{n\to\infty}} S_n = -\infty$.

4.6 L^p Theory, $1 < p < \infty$

Theorem 4.48 (L^p Maximal Inequality). If $\mathcal{X}$ is a submartingale then for $1 < p < \infty$,

$$\mathbb{E}\left[\overline{X}_n^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[(X_n^+)^p\right].$$

Consequently, if $\mathcal{Y}$ is a martingale and $Y_n^* = \sup_{0 \le m \le n} |Y_m|$,

$$\mathbb{E}[|Y_n^*|^p] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|Y_n|^p].$$

Proof. The second inequality follows by applying the first to $\mathcal{X} = |\mathcal{Y}|$. To prove the first we will work with $\overline{X}_n \wedge M$ rather than $\overline{X}_n$. Since $\{\overline{X}_n \wedge M \ge a\}$ is always $\{\overline{X}_n \ge a\}$ or $\emptyset$, this does not change the application of Theorem 4.42. Using this theorem, Lemma 2.20, Fubini's theorem Theorem 1.72, and finally Holder's theorem,

$$\begin{aligned}
\mathbb{E}\left[\left(\overline{X}_n \wedge M\right)^p\right] &= \int_0^\infty px^{p-1}\mathbb{P}\left(\overline{X}_n \wedge M \ge x\right) \mathrm{d}x \\
&\le \int_0^\infty px^{p-1}\left(\frac{\mathbb{E}\left[X_n^+; \overline{X}_n \wedge M \ge x\right]}{x}\right) \\
&= \mathbb{E}\left[X_n^+ \int_0^{\overline{X}_n \wedge M} px^{p-2} \,\mathrm{d}x\right] \\
&= \frac{p}{p-1}\mathbb{E}\left[X_n^+\left(\overline{X}_n \wedge M\right)^{p-1}\right] \\
&\le \left(\frac{p}{1-p}\right)\mathbb{E}\left[|X_n^+|^p\right]^{1/p}\mathbb{E}\left[\left|\overline{X}_n \wedge M\right|^p\right]^{1/(1-p)} \\
\mathbb{E}\left[\left|\overline{X}_n \wedge M\right|^p\right] &\le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[\left(X_n^+\right)^p\right].
\end{aligned}$$

Letting $M \to \infty$ and using the MCT Theorem 1.58 gives

$$\mathbb{E}\left[\overline{X}_n^p\right] = \lim_{M\to\infty} \mathbb{E}\left[\left(\overline{X}_n \wedge M\right)^p\right] \le \lim_{M\to\infty}\left(\frac{p}{p-1}\right)^p \mathbb{E}\left[\left(X_n^+\right)^p\right] = \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[\left(X_n^+\right)^p\right]$$

as desired. $\square$

Theorem 4.49 (**L^p Convergence Theorem**)**.** If $\mathcal{X}$ is a martingale with $\sup_n \mathbb{E}[|X_n|^p] < \infty$ where $p > 1$, then there is a limit X with $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$ and $\overset{L^p}{\lim_{n\to\infty}} X_n = X$.

Proof. We have

$$\mathbb{E}\left[X_n^+\right]^p \le \mathbb{E}[|X_n|]^p \le \mathbb{E}[|X_n|^p],$$

so it follows from the Martingale Convergence Theorem Theorem 4.44 that $\lim_{n\to\infty}^{\text{a.s.}} X_n = X$. The second conclusion in Theorem 4.48 implies

$$\mathbb{E}\left[\sup_{0\le m\le n}|X_m|\right]^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p].$$

Thus by MCT Theorem 1.58,

$$\mathbb{E}\left[\sup_n |X_n|^p\right] = \lim_{n\to\infty}\mathbb{E}\left[\sup_{0\le m\le n}|X_m|^p\right] \le \lim_{n\to\infty}\left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p] < \infty.$$

Thus $\sup_n |X_n| \in L^p(\Sigma)$. Finally

$$|X_n - X|^p \le \left(2\sup_n |X_n|\right)^p < \infty$$

so it follows from DCT Theorem 1.59 that

$$\lim_{n\to\infty}\mathbb{E}[|X_n - X|^p] = \mathbb{E}\left[\lim_{n\to\infty}|X_n - X|^p\right] = 0.$$

$\square$

4.7 L^1 Theory

Before getting into L^1 theory, we will see a decomposition result.

Theorem 4.50 (Doob's Decomposition Theorem). Any submartingale $\mathcal{X}$ can be written in a unique way as $\mathcal{X} = \mathcal{M} + \mathcal{A}$, where $\mathcal{M}$ is a martingale and $\mathcal{A}$ is a predictable increasing sequence with $A_0 = 0$.

Proof. We want $X_n = M_n + A_n$, $\mathbb{E}[M_{n+1} \,|\, \Sigma_n] = M_n$, and $A_{n+1} \in L^0(\Sigma_n)$. So we must have

$$\mathbb{E}[X_{n+1} \,|\, \Sigma_n] = \mathbb{E}[M_{n+1} + A_{n+1} \,|\, \Sigma_n] = \mathbb{E}[M_{n+1} \,|\, \Sigma_n] + \mathbb{E}[A_{n+1} \,|\, \Sigma_n] = M_n + A_{n+1} = X_n - A_n + A_{n+1}.$$

It follows that

$$A_{n+1} - A_n = \mathbb{E}[X_{n+1} \,|\, \Sigma_n] - X_n.$$

Since $A_0 = 0$, we have

$$A_n = \sum_{m=1}^{n} \mathbb{E}[X_m - X_{m-1} \,|\, \Sigma_{m-1}] \quad \text{and} \quad M_n = X_n - A_n.$$

To check that our recipe works, note $A_n - A_{n-1} \geq 0$ since $\mathcal{X}$ is a submartingale and $\mathcal{A}$ is predictable. Now to show $\mathcal{M}$ is a martingale, we note that using $\mathcal{A}$ is predictable,

$$\mathbb{E}[M_{n+1} \,|\, \Sigma_n] = \mathbb{E}[X_{n+1} - A_{n+1} \,|\, \Sigma_n] = \mathbb{E}[X_{n+1} \,|\, \Sigma_n] - A_{n+1} = X_n - A_n = M_n$$

which completes the proof. □

Now we will start on L^1 theory.

Definition 4.51 (Uniform Integrability). A sequence of random variables $\{X_\lambda\} \in L^1(\Sigma)$ is **uniformly integrable** if

$$\lim_{M \to \infty} \sup_{\lambda \in \Lambda} \mathbb{E}[|X_\lambda|; |X_\lambda| > M] = 0.$$

If we pick M large enough so that the supremum is < 1, it follows that

$$\sup_{\lambda \in \Lambda} \mathbb{E}[|X_\lambda|] \leq M + 1 < \infty.$$

Example 4.52. Given $X \in L^1(\Sigma)$, then the family

$$\{\mathbb{E}[X \,|\, \Sigma_{\text{small}}] \colon \Sigma_{\text{small}} \subseteq \Sigma, \Sigma_{\text{small}} \text{ is a } \sigma\text{-algebra}\}$$

is uniformly integrable. Indeed, let $\varepsilon > 0$. By DCT Theorem 1.59 we can pick $\delta > 0$ such that if $\mathbb{P}(A) \leq \delta$ then $\mathbb{E}[|X|; A] \leq \varepsilon$. Pick M large enough so that $\mathbb{E}[|X|] \leq M\delta$. Jensen's inequality and the definition of conditional expectation give

$$\begin{aligned}
\mathbb{E}[|\mathbb{E}[X \,|\, \Sigma_{\text{small}}]|; |\mathbb{E}[X \,|\, \Sigma_{\text{small}}]| > M] &\leq \mathbb{E}[\mathbb{E}[|X| \,|\, \Sigma_{\text{small}}]; |\mathbb{E}[X \,|\, \Sigma_{\text{small}}]| > M] \\
&\leq \mathbb{E}[\mathbb{E}[|X| \,|\, \Sigma_{\text{small}}]; \mathbb{E}[|X| \,|\, \Sigma_{\text{small}}] > M] \\
&= \mathbb{E}[|X|; \mathbb{E}[|X| \,|\, \Sigma_{\text{small}}] > M].
\end{aligned}$$

Using Chebyshev's inequality Corollary 2.11 and recalling the definition of M, we have

$$\mathbb{P}(\mathbb{E}[|X| \,|\, \Sigma_{\text{small}}] > M) \leq \frac{\mathbb{E}[\mathbb{E}[|X| \,|\, \Sigma_{\text{small}}]]}{M} = \frac{\mathbb{E}[|X|]}{M} \leq \delta.$$

Thus by the choice of δ we have

$$\mathbb{E}[|\mathbb{E}[X \,|\, \Sigma_{\text{small}}]|; |\mathbb{E}[X \,|\, \Sigma_{\text{small}}]| > M] \leq \mathbb{E}[|X|; \mathbb{E}[|X| \,|\, \Sigma_{\text{small}}] > M] \leq \varepsilon.$$

Since ε is arbitrary, the family is uniformly integrable.

Theorem 4.53. If $\mathcal{X}$ is a martingale then the following are equivalent.

1. $\{X_n\}$ are uniformly integrable.
2. There is a limit $X \in L^1(\Sigma)$ such that $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$ and $\overset{L^1}{\lim_{n\to\infty}} X_n = X$.
3. There is a limit $X \in L^1(\Sigma)$ such that $X_n = \mathbb{E}[X \mid \Sigma_n]$.

Proof.

Claim 1. (1) implies (2).

Proof. Since $\{X_n\}$ are uniformly integrable, $\sup_n \mathbb{E}[|X_n|] < \infty$. Thus the martingale convergence theorem Theorem 4.44 says that there is a finite limit $X \in L^1(\Sigma)$ such that $\lim^{\text{a.s.}}_{n\to\infty} X_n = X$. To show it converges in L^1, let $\phi_M(x) = x \wedge M \vee -M$. The triangle inequality implies

$$|X_n - X| \leq |X_n - \phi_M(X_n)| + |\phi_M(X_n) - \phi_M(X)| + |\phi_M(X) - X|.$$

We have the inequality

$$|\phi_M(Y) - Y| = (|Y| - M)^+ \leq |Y|\mathbb{1}_{|Y|>M},$$

taking expectations gives

$$\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|\phi_M(X_n) - \phi_M(X)|] + \mathbb{E}[|X_n|; |X_n| > M] + \mathbb{E}[|X|; |X| > M].$$

We know that $\lim^{\text{a.s.}}_{n\to\infty} \phi_M(X_n) = \phi_M(X)$, so by BCT Theorem 1.56,

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

If $\varepsilon > 0$ and M is large, uniform integrability implies that

$$\mathbb{E}[|X_n|; |X_n| > M] \leq \varepsilon.$$

To bound the last term, we observe that $X \in L^1(\Sigma)$, so for M large enough,

$$\mathbb{E}[|X|; |X| > M] \leq \varepsilon.$$

This implies

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|] \leq \lim_{n\to\infty} (\mathbb{E}[|\phi_M(X_n) - \phi_M(X)|] + \mathbb{E}[|X_n|; |X_n| > M] + \mathbb{E}[|X|; |X| > M]) \leq 2\varepsilon.$$

Since ε is arbitrary, (2) is shown. ■

Claim 2. (2) implies (3).

Proof. The martingale property implies if $m > n$ then $\mathbb{E}[X_m \mid \Sigma_n] = X_n$. Thus if $A \in \Sigma_n$ then $\mathbb{E}[X_n; A] = \mathbb{E}[X_m; A]$. And

$$\lim_{m\to\infty} |\mathbb{E}[X_m; A] - \mathbb{E}[X; A]| \leq \lim_{m\to\infty} \mathbb{E}[|X_m - X|; A] \leq \lim_{n\to\infty} \mathbb{E}[|X_n - X|] = 0$$

so sending $m \to \infty$, we have $\mathbb{E}[X_n; A] = \mathbb{E}[X; A]$ for all $A \in \Sigma_n$. Thus $X_n = \mathbb{E}[X \mid \Sigma_n]$. ■

Claim 3. (3) implies (1).

Proof. Follows from Example 4.52. ■

□

Corollary 4.54. Suppose $X \in L^1(\Sigma)$ and define $X_n = \mathbb{E}[X \mid \Sigma_n]$. If $\overset{\text{a.s.}}{\lim\limits_{n\to\infty}} X_n = Y$ and $\overset{L^1}{\lim\limits_{n\to\infty}} X_n = Y$, then

$$Y = \mathbb{E}\left[X \,\middle|\, \sigma\left(\bigcup_{n=1}^{\infty} \Sigma_n\right)\right].$$

Proof. To show this it suffices to show

$$\mathbb{E}[X; A] = \mathbb{E}[Y; A] \quad \text{for all } A \in \sigma\left(\bigcup_{n=1}^{\infty} \Sigma_n\right).$$

By the π-λ theorem Theorem 1.22, it suffices to show

$$\mathbb{E}[X; A] = \mathbb{E}[Y; A] \quad \text{for all } A \in \bigcup_{n=1}^{\infty} \Sigma_n.$$

Suppose $A \in \bigcup_{n=1}^{\infty} \Sigma_n$. Then $A \in \Sigma_n$ for some n. Then since $X_n = \mathbb{E}[X \mid \Sigma_n]$,

$$\mathbb{E}[X; A] = \mathbb{E}[X_n; A] = \mathbb{E}[X_m; A] \quad \text{for } m > n \qquad \Longrightarrow \quad \mathbb{E}[X; A] = \lim_{m\to\infty} \mathbb{E}[X_m; A] = \mathbb{E}[Y; A]$$

since $\lim_{m\to\infty}^{L^1} X_m = Y$. Thus $\mathbb{E}[X; A] = \mathbb{E}[Y; A]$ for all $A \in \bigcup_{n=1}^{\infty} \Sigma_n$, so $Y = \mathbb{E}[X \mid \sigma(\bigcup_{n=1}^{\infty} \Sigma_n)]$ as desired. □

Example 4.55 (There is no L^1 Maximal Inequality). Let $\mathcal{S}$ be a simple random walk starting from $S_0 = 1$, $\tau = \inf\{n : S_n = 0\}$, and $\mathcal{X} = \mathcal{S}_{\wedge\tau}$. Then Theorem 4.39 implies

$$\mathbb{E}[X_n] = \mathbb{E}[S_{\tau\wedge n}] = \mathbb{E}[S_0] = 1 \quad \text{for all } n.$$

Using hitting probabilities for the simple random walk, we have

$$\mathbb{P}\left(\sup_m X_m \geq M\right) = \frac{1}{M}$$

so by Lemma 2.20

$$\mathbb{E}\left[\sup_m X_m\right] = \sum_{M=1}^{\infty} \mathbb{P}\left(\sup_m X_m \geq M\right) = \sum_{M=1}^{\infty} \frac{1}{M} = \infty.$$

The MCT Theorem 1.58 implies that

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{0\leq m\leq n} X_m\right] \uparrow \infty.$$

4.8 Optional Stopping

Let $\tau \in L^0(\Sigma)$ be a stopping time and $\mathcal{X}$ be a submartingale. We want to find conditions under which

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_\tau].$$

This is not always true. Theorem 4.39 shows that this is true if $\tau \in L^\infty(\Sigma)$; our attention will be focused on the case of unbounded τ.

Like in the proof of Theorem 4.34, we present results for submartingales or supermartingales; for the other type, the inequalities of the conclusion are reversed.

Theorem 4.56. If $\mathcal{X} \in L^1(\Sigma)$ is a uniformly integrable submartingale and $\tau \in L^0(\Sigma)$ is a stopping time, then $\mathcal{X}_{\wedge\tau}$ is uniformly integrable.

Proof. Since $\mathcal{X}$ is a submartingale, $\mathcal{X}^+$ is a submartingale. Thus Theorem 4.39 implies

$$\mathbb{E}\left[X_{n\wedge\tau}^+\right] \leq \mathbb{E}\left[X_n^+\right].$$

Since $\mathcal{X}^+$ is uniformly integrable, it follows that

$$\sup_n \mathbb{E}\left[X_{n\wedge\tau}^+\right] \leq \sup_n \mathbb{E}\left[X_n^+\right] < \infty.$$

Using the martingale convergence theorem Theorem 4.44 gives

$$\overset{\text{a.s.}}{\lim_{n\to\infty}} X_{n\wedge\tau} = X_\tau \quad \text{and} \quad X_\tau \in L^1(\Sigma).$$

With this established, we write

$$\begin{aligned}
\mathbb{E}[|X_{n\wedge\tau}|; |X_{n\wedge\tau}| > M] &= \mathbb{E}[|X_{n\wedge\tau}|; |X_{n\wedge\tau}| > M, \tau \leq n] + \mathbb{E}[|X_{n\wedge\tau}|; |X_{n\wedge\tau}| > M, \tau > n] \\
&= \mathbb{E}[|X_\tau|; |X_\tau| > M, \tau \leq n] + \mathbb{E}[|X_n|; |X_n| > M, \tau > n] \\
&\leq \mathbb{E}[|X_\tau|; |X_\tau| > M] + \mathbb{E}[|X_n|; |X_n| > M].
\end{aligned}$$

Since $X_\tau \in L^1(\Sigma)$ and $\mathcal{X}$ is uniformly integrable, if M is large enough then both terms are $\leq \frac{\varepsilon}{2}$. Since ε is arbitrary, $\mathcal{X}_{\wedge\tau}$ is uniformly integrable. $\square$

Theorem 4.57. If $\mathcal{X} \in L^1(\Sigma)$ is a submartingale, $\tau \in L^0(\Sigma)$ is a stopping time, $X_\tau \in L^1(\Sigma)$, and $\{X_n \mathbb{1}_{\tau > n}\}$ is uniformly integrable, then $\mathcal{X}_{\wedge\tau}$ is uniformly integrable and hence $\mathbb{E}[X_0] \leq \mathbb{E}[X_\tau]$.

Proof. The last computation in the proof of Theorem 4.56. $\square$

Theorem 4.58. If $\mathcal{X} \in L^1(\Sigma)$ is a uniformly integrable submartingale then for any stopping time $\tau \in L^0(\Sigma)$, we have $\mathbb{E}[X_0] \leq \mathbb{E}[X_\tau] \leq \mathbb{E}[X]$, where $\overset{\text{a.s.}}{\lim_{n\to\infty}} X_n = X$.

Proof. Theorem 4.39 implies

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_{n\wedge\tau}] \leq \mathbb{E}[X_n].$$

Taking $n \to \infty$ and noting Theorem 4.56 implies $\lim_{n\to\infty}^{L^1} X_{n\wedge\tau} = X_\tau$ and $\lim_{n\to\infty}^{L^1} X_n = X$ gives

$$\mathbb{E}[X_0] \leq \lim_{n\to\infty} \mathbb{E}[X_{n\wedge\tau}] = \mathbb{E}[X_\tau] \leq \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

$\square$

The next result does not require uniform integrability.

Theorem 4.59. If $\mathcal{X} \in L^1(\Sigma)$ has $\mathcal{X} \geq 0$ and $\tau \in L^0(\Sigma)$ is a stopping time, then $\mathbb{E}[X_0] \geq \mathbb{E}[X_\tau]$.

Proof. Theorem 4.39 gives that $\liminf_{n\to\infty} X_{n\wedge\tau} = X_\tau$. By Fatou's lemma Theorem 1.57,

$$\mathbb{E}[X_0] \geq \liminf_{n\to\infty} \mathbb{E}[X_{n\wedge\tau}] \geq \mathbb{E}\left[\liminf_{n\to\infty} X_{n\wedge\tau}\right] = \mathbb{E}[X_\tau]$$

as desired. $\square$

The next result is useful in some situations.

Theorem 4.60. Suppose $\mathcal{X}$ is a submartingale and $\mathbb{E}[|X_{n+1} - X_n| \,|\, \Sigma_n] \leq B$. If $\tau \in L^1(\Sigma)$ is a stopping time then $\mathcal{X}_{\wedge\tau}$ is integrable and hence $\mathbb{E}[X_\tau] \geq \mathbb{E}[X_0]$.

Proof. We observe that

$$|X_{n\wedge\tau}| \leq |X_0| + \sum_{m=0}^{\infty} |X_{m+1} - X_m| \mathbb{1}_{\tau > m}.$$

To prove uniform integrability it usffices to show that the right-hand side has finite expectation, for then $|\mathcal{X}_{\wedge\tau}|$ is dominated by an integrable random variable. Now,

$$\{\tau > m\} = \{\tau \leq m - 1\}^c \in \Sigma_{m-1} \subseteq \Sigma_m$$

so

$$\begin{aligned}\mathbb{E}[|X_{m+1} - X_m|; \tau > m] &= \mathbb{E}[\mathbb{E}[|X_{m+1} - X_m|; \tau > m \mid \Sigma_m]] = \mathbb{E}[\mathbb{E}[|X_{m+1} - X_m| \mid \Sigma_m]; \tau > m] \\ &\leq \mathbb{E}[B; \tau > m] = B\mathbb{P}(\tau > m).\end{aligned}$$

Then

$$\begin{aligned}\mathbb{E}\left[|X_0| + \sum_{m=0}^{\infty} |X_{m+1} - X_m| \mathbb{1}_{\tau > m}\right] &= \mathbb{E}[|X_0|] + \sum_{m=0}^{\infty} \mathbb{E}[|X_{m+1} - X_m|; \tau > m] \leq \mathbb{E}[|X_0|] + B \sum_{m=0}^{\infty} \mathbb{P}(\tau > m) \\ &= \mathbb{E}[|X_0|] + B\mathbb{E}[\tau] < \infty.\end{aligned}$$

Thus $\mathcal{X}_{\wedge\tau}$ is uniformly integrable. ☐

4.9 Reverse Martingale

Definition 4.61. A **reverse martingale** is a martingale indexed by the negative integers, i.e., $\mathcal{X} = \{X_n\}_{n \leq 0} \in L^1(\Sigma)$, adapted to an increasing sequence of σ-algebras $\mathcal{F} = \{\Sigma_n\}_{n \leq 0}$ with

$$\mathbb{E}[X_{n+1} \mid \Sigma_n] = X_n \quad \text{for } n \leq -1.$$

Because the σ-algebras decrease as $n \downarrow -\infty$, the convergence theory for backwards martingales is particularly simple.

Theorem 4.62. Let $\mathcal{X}$ be a backwards martingale. Then there is a finite limit X such that $\overset{\text{a.s.}}{\lim_{n\to-\infty}} X_n = X$ and $\overset{L^1}{\lim_{n\to-\infty}} X_n = X$.

Proof. Let U_n be the number of upcrossings of $[a, b]$ by $\{X_n\}_{-n \leq m \leq 0}$, and let U be the number of upcrossings of $[a, b]$ by $\mathcal{X}$. The upcrossing inequality, Theorem 4.43, implies

$$(b - a)\mathbb{E}[U_n] \leq \mathbb{E}\left[(X_0 - a)^+\right].$$

Letting $n \to \infty$ and using the MCT Theorem 1.58, we have

$$\mathbb{E}[U] \leq \frac{\mathbb{E}\left[(X_0 - a)^+\right]}{b - a} < \infty.$$

By Theorem 4.44 the limit exists a.s., so write $X = \lim^{\text{a.s.}}_{n\to\infty} X_{-n}$. The martingale property implies $X_n = \mathbb{E}[X_0 \mid \Sigma_n]$, so Example 4.52 implies $\{X_n\}_{n \leq 0}$ is uniformly integrable, and Theorem 4.53 tells us that $\lim^{L^1}_{n\to-\infty} X_n = X$ as well. ☐

We now want to construct this limit.

Theorem 4.63. If $\mathcal{X}$ is a backwards martingale, $X = \lim_{n\to-\infty} X_n$, and $\Sigma_{\text{small}} = \bigcap_{n \leq 0} \Sigma_n$, then $X = \mathbb{E}[X_0 \mid \Sigma_{\text{small}}]$.

Proof. Clearly, $X \in \Sigma_{\text{small}}$. The martingale property says $X_n = \mathbb{E}[X_0 \mid \Sigma_n]$, so if $A \in \Sigma_{\text{small}} \subseteq \Sigma_n$ then $\mathbb{E}[X_n; A] = \mathbb{E}[X_0; A]$.

Then Theorem 4.62 and parts of the proof of Theorem 4.53 imply $\lim_{n\to-\infty}\mathbb{E}[X_n; A] = \mathbb{E}[X; A]$, so $\mathbb{E}[X; A] = \mathbb{E}[X_0; A]$ for all $A \in \Sigma_{\text{small}}$, proving the desired conclusion. ☐

Theorem 4.64. If $\Sigma_{\text{small}} = \bigcap_{n\leq 0} \Sigma_n$, then

$$\lim_{n\to-\infty}^{\text{a.s.}} \mathbb{E}[Y \mid \Sigma_n] = \mathbb{E}[Y \mid \Sigma_{\text{small}}] \quad \text{and} \quad \lim_{n\to-\infty}^{L^1} \mathbb{E}[Y \mid \Sigma_n] = \mathbb{E}[Y \mid \Sigma_{\text{small}}].$$

Proof. If $X_n = \mathbb{E}[Y \mid \Sigma_n]$ then $\mathcal{X}$ is a backwards martingale, so Theorem 4.62 implies that, if

$$X = \mathbb{E}[X_0 \mid \Sigma_{\text{small}}] = \mathbb{E}[\mathbb{E}[Y \mid \Sigma_0] \mid \Sigma_{\text{small}}] = \mathbb{E}[Y \mid \Sigma_{\text{small}}]$$

then $\lim_{n\to-\infty}^{\text{a.s.}} X_n = X$ and $\lim_{n\to-\infty}^{L^1} X_n = X$. ☐

Even though the convergence theory for backwards martingales is easy, there are some nice applications. Suppose random variables take values in the measurable space $(S, \mathcal{S})$, and that

$$\Omega = \{\{\omega_n\} : \omega_n \in S\}, \quad \Sigma = \sigma(\mathcal{S}^\infty), \quad X_n(\omega) = \omega_n.$$

Let $\mathcal{E}_n$ be the σ-algebra generated by events that are invariant under permutations of $\mathbb{N}$ that leave m unchanged for all $m \geq n$. Finally let $\mathcal{E} = \bigcap_{n=1}^{\infty} \mathcal{E}_n$ be the exchangeable σ-algebra.

Theorem 4.65 (Hewitt-Savage 0-1 Law). If $\{X_n\}_{n\in\mathbb{N}}$ are i.i.d. and $A \in \mathcal{E}$ then $\mathbb{P}(A) \in \{0, 1\}$.

Proof.

Lemma 1. If $X \in L^2(\Sigma)$ and $\mathbb{E}[X \mid \Sigma_1] \in L^0(\Sigma_2)$ with X independent of Σ_2 then $\mathbb{E}[X \mid \Sigma_1] = \mathbb{E}[X]$.

Proof. Let $Y = \mathbb{E}[X \mid \Sigma_1]$ and note that Theorem 4.18 implies

$$\mathbb{E}\left[Y^2\right] \leq \mathbb{E}\left[X^2\right] < \infty.$$

By independence,

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

From Theorem 4.22, $\mathbb{E}[(X - Y)Y] = 0$, so

$$\mathbb{E}\left[Y^2\right] = \mathbb{E}[XY] = \mathbb{E}[Y]^2 \implies \operatorname{Var}(Y) = \mathbb{E}\left[Y^2\right] - \mathbb{E}[Y]^2 = 0$$

so Y is almost surely constant. ■

Lemma 2. Suppose $\{X_n\}$ are i.i.d. and let

$$A_n(\phi) = \frac{1}{(n)_k} \sum_i \phi(X_{i_1}, \ldots, X_{i_k})$$

where the sum is over all sequences of distinct integers $1 \leq i_1, \ldots, i_k \leq n$ and $(n)_k = \dfrac{n!}{(n-k)!}$ is the number of such sequences. If ϕ is bounded, $\lim_{n\to\infty}^{\text{a.s.}} A_n(\phi) = \mathbb{E}[\phi(X_1, \ldots, X_k)]$.

Proof. It's obvious that $A_n(\phi) \in L^\infty(\mathcal{E}_n)$, so

$$A_n(\phi) = \mathbb{E}[A_n(\phi) \mid \mathcal{E}_n] = \frac{1}{(n)_k} \sum_i \mathbb{E}[\phi(X_{i_1}, \ldots, X_{i_k}) \mid \mathcal{E}_n] = \mathbb{E}[\phi(X_1, \ldots, X_k) \mid \mathcal{E}_n].$$

Then Theorem 4.63 with $\Sigma_{-m} = \mathcal{E}_m$ for $m \geq 1$ implies
$$\lim_{n\to\infty} \mathbb{E}[\phi(X_1, \dots, X_k) \,|\, \mathcal{E}_n] = \mathbb{E}[\phi(X_1, \dots, X_k) \,|\, \mathcal{E}].$$
We want to show that the limit is $\mathbb{E}[\phi(X_1, \dots, X_k)]$. The first step is to observe that there are $k(n-1)_{k-1}$ terms in $A_n(\phi)$ involving X_1 and $\phi \in L^\infty(\mathcal{E}_n)$, so if we let $1 \in i$ denote the sum over sentences that contain 1,
$$\lim_{n\to\infty} \frac{1}{(n)_k} \sum_{1\in i} \phi(X_{i_1}, \dots, X_{i_k}) \leq \lim_{n\to\infty} \frac{k(n-1)_{k-1}}{(n)_k} \sup_{x\in S^k} \phi(x) = 0.$$
This shows that
$$\mathbb{E}[\phi(X_1, \dots, X_k) \,|\, \mathcal{E}] \in \sigma(X_m \colon m \geq 2).$$
Repeating the same argument for $2, 3, \dots, k$ shows
$$\mathbb{E}[\phi(X_1, \dots, X_k) \,|\, \mathcal{E}] \in \sigma(X_m \colon m > k).$$
Since $\{X_n\}$ are i.i.d., $\sigma(X_m \colon m \leq k) \perp\!\!\!\perp \sigma(X_m \colon m > k)$. By Lemma 1,
$$\mathbb{E}[\phi(X_1, \dots, X_k) \,|\, \mathcal{E}] = \mathbb{E}[\phi(X_1, \dots, X_k)].$$
■

Lemma 2 holds for all bounded ϕ, so $\mathcal{E}$ is independent of $\sigma(X_m \colon m \in [k])$. Since this holds for all k, and $\bigcup_{k=1}^\infty \sigma(X_m \colon m \in [k])$ is a π-system that contains Ω, Theorem 1.22 implies $\mathcal{E} \perp\!\!\!\perp \sigma\left(\bigcup_{k=1}^\infty \sigma(X_m \colon m \in [k])\right) \supseteq \mathcal{E}$. Thus $\mathcal{E} \perp\!\!\!\perp \mathcal{E}$, so if $A \in \mathcal{E}$ then
$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2.$$
□

Definition 4.66 (Exchangeability). A sequence $\{X_n\}$ is **exchangeable** if for each n and permutation π of $[n]$, $(X_1, \dots, X_n) \overset{d}{=} (X_{\pi(1)}, \dots, X_{\pi(n)})$.

Theorem 4.67 (De Finetti's Theorem). If $\{X_n\}$ are exchangeable then conditional on $\mathcal{E}$, $\{X_n\}$ are i.i.d..

Proof. For any exchangeable sequence X,
$$\begin{aligned} A_n(\phi) = \mathbb{E}[A_n(\phi) \,|\, \mathcal{E}_n] &= \frac{1}{(n)_k} \sum_i \mathbb{E}[\phi(X_{i_1}, \dots, X_{i_k}) \,|\, \mathcal{E}_n] \\ &= \frac{1}{(n)_k} \sum_i \mathbb{E}[\phi(X_1, \dots, X_k) \,|\, \mathcal{E}_n] = \mathbb{E}[\phi(X_1, \dots, X_k) \,|\, \mathcal{E}_n]. \end{aligned}$$
Again, Theorem 4.63 implies
$$\lim_{n\to\infty} A_n(\phi) = \mathbb{E}[\phi(X_1, \dots, X_k) \,|\, \mathcal{E}].$$
This time, $\mathcal{E}$ may be nontrivial, so we cannot hope to show that the limit is $\mathbb{E}[\phi(X_1, \dots, X_k)]$.

Let $f \colon \mathbb{R}^{k-1} \to \mathbb{R}$ and $g \colon \mathbb{R} \to \mathbb{R}$ be bounded. If $I_{n,k}$ is the set of all sequences of distinct integers $1 \leq i_1, \dots, i_k \leq n$, then
$$\begin{aligned} (n)_{k-1} A_n(f) n A_n(g) &= \sum_{i \in I_{n,k-1}} f(X_{i_1}, \dots, X_{i_{k-1}}) \sum_{m=1}^{k} g(X_m) \\ &= \sum_{i\in I_{n,k}} f(X_{i_1}, \dots, X_{i_{k-1}}) g(X_{i_k}) + \sum_{i \in I_{n,k-1}} \sum_{j=1}^{k-1} f(X_{i_1}, \dots, X_{i_{k-1}}) g(X_{i_j}). \end{aligned}$$

If we let $\phi(x_1, \dots, x_k) = f(x_1, \dots, x_{k-1})g(x_k)$, then

$$\frac{(n)_{k-1}n}{(n)_k} = \frac{n}{n-k+1} \quad \text{and} \quad \frac{(n)_{k-1}}{(n)_k} = \frac{1}{n-k+1}$$

then rearrange, we have

$$A_n(\phi) = \frac{n}{n-k+1}A_n(f)A_n(g) - \frac{1}{n-k+1}\sum_{j=1}^{k-1} A_n(\phi_j)$$

where $\phi_j(x_1, \dots, x_{k-1}) = f(x_1, \dots, x_{k-1})g(x_j)$. Applying the limit to ϕ, f, g, and all the ϕ_j gives

$$\mathbb{E}[f(X_1, \dots, X_{k-1})g(X_k) \,|\, \mathcal{E}] = \mathbb{E}[f(X_1, \dots, X_{k-1}) \,|\, \mathcal{E}]\mathbb{E}[g(X_k) \,|\, \mathcal{E}].$$

It follows by induction that

$$\mathbb{E}\left[\prod_{j=1}^{k} f_j(X_j) \,\middle|\, \mathcal{E}\right] = \prod_{j=1}^{k} \mathbb{E}[f_j(X_j) \,|\, \mathcal{E}]$$

which demonstrates equality a.s. as desired. $\square$