Notes on Probability Theory

Kai-Jyun Wang*

Fall 2025

The notes are based on the lecture taught by Prof. David Anderson and Prof. Benedek Valkó at University of Wisconsin-Madison in 2025-2026. The course assumes a certain amount of knowledges in real analysis. For some classic results in real analysis, one can refer to my notes on real analysis.

*National Taiwan University, Department of Economics.

Contents

1. Probability Space

1.1. Probability Space

Definition 1.1
*Let Ω be a set. A collection of subsets $\mathcal{F}$ forms a σ-**algebra** if*

(a) $\varnothing \in \mathcal{F}$.

(b) $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$.

(c) If $A_i \in \mathcal{F}$ *are countably many sets,* $\cup_i A_i \in \mathcal{F}$.

The dual $(\Omega, \mathcal{F})$ *is called a **measurable space** and the sets falling in* $\mathcal{F}$ *are said to be **measurable**.*

Definition 1.2
Let $(\Omega, \mathcal{F})$ *be a measurable space. A set function* $\mu : \mathcal{F} \to [0, \infty]$ *is a **measure** if*

(a) $\mu(\varnothing) = 0$.

(b) For countably many disjoint $A_i \in \mathcal{F}$, $\mu(\cup_i A_i) = \sum_i \mu(A_i)$.

The triple $(\Omega, \mathcal{F}, \mu)$ *is called a **measure space**.*

Definition 1.3
*A **probability space** is a measure space* $(\Omega, \mathcal{F}, \mathrm{P})$ *such that* $\mathrm{P}(\Omega) = 1$.

Lemma 1.4
Let $\mathcal{S}$ *be a collection of sets. Then there exists the smallest* σ*-algebra containing* $\mathcal{S}$.

Proof. Let $\mathcal{F}$ be the intersection of all σ-algebra containing $\mathcal{S}$. $\mathcal{F}$ is non-empty since the power set is a σ-algebra containing $\mathcal{S}$. Now it is clear that $\varnothing \in \mathcal{F}$ since $\varnothing \in \mathcal{A}$ for every σ-algebra $\mathcal{A}$ contiaining $\mathcal{S}$. If $A \in \mathcal{F}$, $A \in \mathcal{A}$ for all $\mathcal{A}$ containing $\mathcal{S}$ and $A^c \in \mathcal{A}$ for all $\mathcal{A}$. Thus $A^c \in \mathcal{F}$. Finally, if $A_i \in \mathcal{F}$ are countably many sets, then each A_i lies in every $\mathcal{A}$ containing $\mathcal{S}$; so does $\cup_i A_i$ and thus $\cup_i A_i \in \mathcal{F}$. The minimality follows by the construction of $\mathcal{F}$. ■

Definition 1.5
For any collection of sets $\mathcal{S}$*, the smallest* σ*-algebra is denoted as* $\sigma(\mathcal{S})$.

Theorem 1.6
Let $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space. Then*

(a) If $A, B \in \mathcal{F}$ *and* $A \subset B$*, then* $\mathrm{P}(A) \le \mathrm{P}(B)$.

(b) For countably many $A_i \in \mathcal{F}$, $\mathrm{P}(\cup_i A_i) \le \sum_i \mathrm{P}(A_i)$.

(c) If $A_i \nearrow A$, $\mathrm{P}(A_i) \to \mathrm{P}(A)$.

(d) If $A_i \searrow A$, $\mathrm{P}(A_i) \to \mathrm{P}(A)$.

Proof. (a) and (b) are clear. For (c), write $E_i = A_i - A_{i-1}$ and $A_0 = \varnothing$. Then since E_i are disjoint and $A_n = \cup_{i=1}^n E_i$,

$$\mathrm{P}(A_n) = \mathrm{P}\big(\cup_{i=1}^n E_i\big) = \sum_{i=1}^n \mathrm{P}(E_i) \to \sum_i \mathrm{P}(E_i) = \mathrm{P}(\cup_i E_i) = \mathrm{P}(A)$$

as $n \to \infty$.

For (d), note that $A_i^c \nearrow A^c$. Thus $1 - \mathrm{P}(A_i) = \mathrm{P}(A_i^c) \to \mathrm{P}(A^c) = 1 - \mathrm{P}(A)$. Thus $\mathrm{P}(A_i) \to \mathrm{P}(A)$. ■

Definition 1.7
*The **Borel σ-algebra** is the σ-algebra generated by all open sets.*

Definition 1.8
Let P *be a probability measure on* $(\mathbb{R}, \mathcal{B})$*. The **distribution function** F is defined as*

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

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

Proposition 1.9
The distribution function in $(\mathbb{R}, \mathcal{B})$ *satisfies that*

(a) $F(x) \le F(y)$ *for all* $x \le y$.

(b) $F(x) \to F(y)$ *as* $x \to y^+$.

(c) $F(-\infty) = 0$ *and* $F(\infty) = 1$.

Proof. For (a), note that $(-\infty, x] \subset (-\infty, y]$ and

$$F(x) = \mathrm{P}((-\infty, x]) \le \mathrm{P}((-\infty, y]) = F(y).$$

For (b), notice that for $x_n \to y^+$, $(-\infty, x_n] \searrow (-\infty, y]$. Hence

$$F(x_n) = \mathrm{P}((-\infty, x_n]) \to \mathrm{P}((-\infty, y]) = F(y).$$

Similarly, taking $x_n \to \pm\infty$ gives (c). ■

Definition 1.10
*A collection $\mathcal{S}$ of sets is called an **algebra** if*

(a) $\varnothing \in \mathcal{S}$.

(b) If $A \in \mathcal{S}$*, then* $A^c \in \mathcal{S}$.

(c) If $A, B \in \mathcal{S}$*, then* $A \cup B \in \mathcal{S}$.

Remark
An algebra is closed under finite unions. It is also clear that a σ-algebra is an algebra, while the converse is not true. An example is the collection of all finite unions of intervals in $\mathbb{R}$.

Definition 1.11
*A collection $\mathcal{S}$ of sets is called a **semi-algebra** if*

(a) *If $A, B \in \mathcal{S}$, then $A \cap B \in \mathcal{S}$.*

(b) *If $A \in \mathcal{S}$, then A^c can be written as a finite disjoint union of sets in $\mathcal{S}$.*

Remark
A semi-algebra must contain $\varnothing$ since for any $A \in \mathcal{S}$, $A^c = \cup_i A_i$, where $A_i \in \mathcal{S}$ are disjoint. Then $A \cap A_1 = \varnothing \in \mathcal{S}$.

Remark
An example of being a semi-algebra but not an algebra is the collection of all intervals of the form $(a_i, b_i]$ for $-\infty \leq a_i < b_i \leq \infty$ with the empty set.

Lemma 1.12
If $\mathcal{S}$ is a semi-algebra, then $\overline{\mathcal{S}} = \{$finite disjoint unions of sets in $\mathcal{S}\}$ forms an algebra.

Proof. It has been shown that $\varnothing \in \mathcal{S}$. For $A, B \in \overline{\mathcal{S}}$, write $A = \cup_{i=1}^n A_i$ and $B = \cup_{j=1}^m B_j$ for disjoint $A_i, B_j \in \mathcal{S}$, respectively. Then $A \cap B = \cup_{i,j}(A_i \cap B_j) \in \overline{\mathcal{S}}$. Thus $\overline{\mathcal{S}}$ is closed under intersection. Now if $A \in \overline{\mathcal{S}}$, $A = \cup_{i=1}^n A_i$ for disjoint $A_i \in \mathcal{S}$. Then $A^c = \cap_{i=1}^n A_i^c$. By the definition of semi-algebra, A_i^c can be written as a finite disjoint union of sets in $\mathcal{S}$ and thus $A_i^c \in \overline{\mathcal{S}}$. Since $\overline{\mathcal{S}}$ is closed under finite intersection, $A^c = \cap_{i=1}^n A_i^c \in \overline{\mathcal{S}}$. Finally, for $A, B \in \overline{\mathcal{S}}$, $A \cup B = (A^c \cap B^c)^c \in \overline{\mathcal{S}}$. We conclude that $\overline{\mathcal{S}}$ is indeed an algebra. ■

Definition 1.13
*Suppose $\mathcal{S}$ is a semi-algebra. $\overline{\mathcal{S}} = \{$finite disjoint unions of sets in $\mathcal{S}\}$ is called the **algebra generated by** $\mathcal{S}$.*

Definition 1.14
*Let $\mathcal{S}$ be an algebra. A set function $\mu_0 : \mathcal{S} \to [0, \infty]$ is called a **premeasure** if*

(a) $\mu_0(\varnothing) = 0$.

(b) *For countable disjoint $A_i \in \mathcal{S}$ such that $\cup_i A_i \in \mathcal{S}$,*

$$\mu_0(\cup_i A_i) = \sum_i \mu_0(A_i).$$

Theorem 1.15
Let ν be a set function on a semi-algebra $\mathcal{S}$ such that $\nu(\varnothing) = 0$. Suppose that

(a) *if $A \in \mathcal{S}$ and $A = \cup_{i=1}^n A_i$ for disjoint $A_i \in \mathcal{S}$, then $\nu(A) = \sum_{i=1}^n \nu(A_i)$;*

(b) *if $A_i \in \mathcal{S}$ are countably many sets and $A = \cup_i A_i \in \mathcal{S}$, then $\nu(A) \leq \sum_i \nu(A_i)$.*

Then ν can be extended to a unique premeasure μ_0 on the algebra generated by $\mathcal{S}$.

Proof. We first show the existence. From lemma 1.12 we know that $\mathcal{S}$ generates an algebra $\mathcal{A} = \{$finite disjoint union of sets in $\mathcal{S}\}$. Define our candidate μ_0 by $\mu_0(A) = \sum_i \nu(A_i)$ for

$A = \cup_i A_i$ where $A_i \in \mathcal{S}$ are disjoint. To see that μ_0 is well-defined, suppose $A = \cup_i B_i$ for disjoint $B_i \in \mathcal{S}$. Observe that

$$A_i = \cup_j (A_i \cap B_j) \quad \text{and} \quad B_j = \cup_i (A_i \cap B_j)$$

are finite disjoint unions. Then

$$\sum_i \nu(A_i) = \sum_i \sum_j \nu(A_i \cap B_j) = \sum_j \sum_i \nu(A_i \cap B_j) = \sum_j \nu(B_j)$$

by (a). Thus μ_0 is well-defined.

Now we check that μ_0 is a premeasure. Clearly $\mu_0(\varnothing) = 0$. For finitely many disjoint $A_i \in \mathcal{A}$ such that $\cup_i A_i \in \mathcal{A}$, we can write $A_i = \cup_j B_{ij}$ for disjoint $B_{ij} \in \mathcal{S}$. Then (a) implies that

$$\mu_0(\cup_i A_i) = \mu_0(\cup_{i,j} B_{ij}) = \sum_{i,j} \nu(B_{ij}) = \sum_i \sum_j \mu_0(B_{ij}) = \sum_i \mu_0(A_i).$$

Next, for countably many disjoint $A_i \in \mathcal{A}$ such that $A = \cup_i A_i \in \mathcal{A}$, write $A_i = \cup_j B_{ij}$, where $B_{ij} \in \mathcal{S}$ are finite disjoint for each i. Then $\mu_0(A_i) = \sum_j \nu(B_{ij})$ and

$$\sum_i \mu_0(A_i) = \sum_i \sum_j \nu(B_{ij}).$$

Without loss of generality, we may choose A_i to be those in $\mathcal{S}$ since otherwise we can replace A_i by B_{ij}. We assume that $A_i \in \mathcal{S}$ from now on. Since $A \in \mathcal{A}$, $A = \cup_i C_i$ for finite disjoint $C_i \in \mathcal{S}$. $C_i = \cup_j (C_i \cap A_j)$. Thus (b) gives that

$$\nu(C_i) \le \sum_j \nu(C_i \cap A_j).$$

Then

$$\mu_0(A) = \sum_i \nu(C_i) \le \sum_i \sum_j \nu(C_i \cap A_j) = \sum_j \sum_i \nu(C_i \cap A_j) = \sum_j \nu(A_j) = \sum_j \mu_0(A_j).$$

For the opposite inequality, set $B_n = \cup_{i=1}^n A_i$ and $C_n = A - B_n$. Since $\mathcal{A}$ is an algebra, $C_n \in \mathcal{A}$ and the finite additivity shows that

$$\mu_0(A) = \sum_{i=1}^n \mu_0(A_i) + \mu_0(C_n) \ge \sum_{i=1}^n \mu_0(A_i).$$

Taking $n \to \infty$ gives the desired inequality and thus μ_0 is σ-additive on $\mathcal{A}$.

Finally, if μ_1 is another premeasure on $\mathcal{A}$ extending ν, then for $A = \cup_i A_i$ for disjoint $A_i \in \mathcal{S}$,

$$\mu_1(A) = \sum_i \nu(A_i) = \mu_0(A).$$

$\blacksquare$

Definition 1.16
*A collection of sets $\mathcal{P}$ is called a π-**system** if $A, B \in \mathcal{P}$ implies $A \cap B \in \mathcal{P}$.*

Definition 1.17
*A collection of sets $\mathcal{L}$ is called a λ-**system** if*

(a) $\Omega \in \mathcal{L}$.

(b) If $A, B \in \mathcal{L}$ and $A \subset B$, then $B - A \in \mathcal{L}$.

(c) If $A_i \in \mathcal{L}$ and $A_i \nearrow A$, then $A \in \mathcal{L}$.

Theorem 1.18 (Sierpiński-Dynkin π-λ)
If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system containing $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Proof. First we show that a collection $\mathcal{S}$ is a σ-algebra if and only if it is both a π-system and a λ-system. Suppose first that $\mathcal{S}$ is a π-system and a λ-system. $\varnothing = \Omega - \Omega \in \mathcal{S}$. If $A \in \mathcal{S}$, then $A^c = \Omega - A \in \mathcal{S}$. For $A, B \in \mathcal{S}$, $A \cup B = (A^c \cap B^c)^c \in \mathcal{S}$ since we have shown that $\mathcal{S}$ is closed under complement and intersection by being a π-system. Thus $\mathcal{S}$ is also closed under finite unions. If $A_i \in \mathcal{S}$ are countably many sets, let $B_n = \cup_{i=1}^n A_i \in \mathcal{S}$. Then $B_n \nearrow \cup_i A_i$ and thus $\cup_i A_i \in \mathcal{S}$.

Conversely, if $\mathcal{S}$ is a σ-algebra, then for $A, B \in \mathcal{S}$, $A \cap B = (A^c \cup B^c)^c \in \mathcal{S}$. Thus $\mathcal{S}$ is a π-system. If $A, B \in \mathcal{S}$ and $A \subset B$, then $B - A = B \cap A^c \in \mathcal{S}$. Finally, if $A_i \in \mathcal{S}$ and $A_i \nearrow A$, then $A = \cup_i (A_i - A_{i-1}) \in \mathcal{S}$ with $A_0 = \varnothing$. Thus $\mathcal{S}$ is a λ-system.

Now set $\mathcal{L}$ to be the smallest λ-system containing $\mathcal{P}$. It suffices to show that $\mathcal{L}$ is also a π-system and thus by the above conclusion, $\mathcal{L}$ is a σ-algebra containing $\mathcal{P}$; hence $\sigma(\mathcal{P}) \subset \mathcal{L}$.

To show that $\mathcal{L}$ is a π-system, let $A, B \in \mathcal{L}$. If $A, B \in \mathcal{P}$, $A \cap B \in \mathcal{P} \subset \mathcal{L}$. To extend the result for general $A, B \in \mathcal{L}$, we first fix $B \in \mathcal{P}$ and define

$$\mathcal{L}_B = \{A \mid A \cap B \in \mathcal{L}\}.$$

We claim that $\mathcal{L}_B$ is a λ-system containing $\mathcal{P}$. For $A \in \mathcal{P}$, $A \cap B \in \mathcal{L}$. Thus $\mathcal{P} \subset \mathcal{L}_B$. Clearly $\Omega \in \mathcal{L}_B$. If $E, F \in \mathcal{L}_B$ and $E \subset F$, then

$$(F - E) \cap B = (F \cap B) - (E \cap B) \in \mathcal{L}.$$

Thus $F - E \in \mathcal{L}_B$. Finally, if $E_i \in \mathcal{L}_B$ and $E_i \nearrow E$, then

$$E \cap B = \cup_i (E_i \cap B) \in \mathcal{L}.$$

Hence $E \in \mathcal{L}_B$ and we conclude that $\mathcal{L}_B$ is a λ-system. Since it is a λ-system containing $\mathcal{P}$, it also contains the smallest λ-system $\mathcal{L}$ with the intersection property. Thus $A \cap B \in \mathcal{L}$ whenever $A \in \mathcal{L}$ and $B \in \mathcal{P}$.

Next, fix $A \in \mathcal{L}$ and define $\mathcal{L}_A = \{B \mid A \cap B \in \mathcal{L}\}$. Clearly $\mathcal{L}_A$ contains $\mathcal{L}$ and $\Omega \in \mathcal{L}_A$. If $E, F \in \mathcal{L}_A$ and $E \subset F$, then

$$(F - E) \cap A = (F \cap A) - (E \cap A) \in \mathcal{L}.$$

Thus $F - E \in \mathcal{L}_A$. Finally, if $E_i \in \mathcal{L}_A$ and $E_i \nearrow E$, then

$$E \cap A = \cup_i (E_i \cap A) \in \mathcal{L}.$$

Hence $E \in \mathcal{L}_A$ and we conclude that $\mathcal{L}_A$ is a λ-system. Since it contains $\mathcal{L}$, $A, B \in \mathcal{L}$ implies $A \cap B \in \mathcal{L}$; in other words, $\mathcal{L}$ is a π-system and the proof is complete. ∎

Corollary 1.19
Let μ and ν be two probability measures agreeing on a π-system $\mathcal{P}$, i.e., $\mu(A) = \nu(A)$ for all $A \in \mathcal{P}$. Then $\mu(A) = \nu(A)$ for all $A \in \sigma(\mathcal{P})$.

Proof. Put

$$\mathcal{L} = \{A \mid \mu(A) = \nu(A)\}.$$

We claim that $\mathcal{L}$ is a λ-system containing $\mathcal{P}$. It is clear that by our assumption, $\mathcal{P} \subset \mathcal{L}$ and $\Omega \in \mathcal{L}$. If $A, B \in \mathcal{L}$ and $A \subset B$, then

$$\mu(B - A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B - A).$$

Thus $B - A \in \mathcal{L}$. Finally, if $A_i \in \mathcal{L}$ and $A_i \nearrow A$, then

$$\mu(A) = \lim_{i\to\infty} \mu(A_i) = \lim_{i\to\infty} \nu(A_i) = \nu(A).$$

Hence $A \in \mathcal{L}$ and we conclude that $\mathcal{L}$ is a λ-system. By the Sierpiński-Dynkin π-λ theorem, $\sigma(\mathcal{P}) \subset \mathcal{L}$; in other words, μ and ν agree on $\sigma(\mathcal{P})$. ∎

Definition 1.20
*A measure μ on a measurable space $(\Omega, \mathcal{F})$ is called **σ-finite** if there exists countable $A_i \in \mathcal{F}$ such that $\cup_i A_i = \Omega$ and $\mu(A_i) < \infty$.*

Definition 1.21
*A set function $\mu^* : 2^\Omega \to [0, \infty]$ is called an **outer measure** if*

(a) $\mu^*(\varnothing) = 0$.

(b) If $A \subset B$, then $\mu^*(A) \le \mu^*(B)$.

(c) For countably many $A_i \subset \Omega$, $\mu^*(\cup_i A_i) \le \sum_i \mu^*(A_i)$.

Definition 1.22
*Let μ^* be an outer measure. A set $A \subset \Omega$ is said to be **Carathéodory measurable** or **μ^*-***

measurable *if for all $E \subset \Omega$,*

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

Lemma 1.23
Let μ^* be an outer measure on Ω. Then the collection of all μ^*-measurable sets forms a σ-algebra $\mathcal{F}$ and $\mu^*|_{\mathcal{F}}$ is a measure.

Proof. Put

$$\mathcal{F} = \{A \subset \Omega \mid \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \subset \Omega\}.$$

We first show that $\mathcal{F}$ is a σ-algebra. Clearly $\varnothing \in \mathcal{F}$ and if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$. For $A, B \in \mathcal{F}$, let $C = A \cup B$. The property of outer measure gives that $\mu^*(E) \leq \mu^*(E \cap C) + \mu^*(E \cap C^c)$. To see the opposite inequality, note that $C = A \cup (B \cap A^c)$ and

$$\begin{aligned}\mu^*(E \cap C) + \mu^*(E \cap C^c) &\leq \mu^*(E \cap A) + \mu^*(E \cap B \cap A^c) + \mu^*(E \cap A^c \cap B^c) \\ &= \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E).\end{aligned}$$

Hence $C \in \mathcal{F}$ and $\mathcal{F}$ is closed under finite unions. For countable disjoint $A_i \in \mathcal{F}$ with $A = \cup_i A_i$, let $B_n = \cup_{i=1}^n A_i \in \mathcal{F}$. Then

$$\mu^*(E \cap A) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B_n) = \sum_{i=1}^n \mu^*(E \cap A_i).$$

Taking $n \to \infty$ gives that

$$\mu^*(E \cap A) \geq \sum_i \mu^*(E \cap A_i) \geq \mu^*(E \cap A)$$

by the σ-subadditivity of outer measure. Hence $\mu^*(E \cap A) = \sum_i \mu^*(E \cap A_i)$. Note also that $E \cap A^c \subset E \cap B_n^c$ so $\mu^*(E \cap A^c) \leq \mu^*(E \cap B_n^c)$. Thus

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \sum_{i=1}^n \mu^*(E \cap A_i) + \mu^*(E \cap A^c) \to \mu^*(E \cap A) + \mu^*(E \cap A^c) \geq \mu^*(E)$$

by the σ-subadditivity of outer measure. We conclude that $\mathcal{F}$ is a σ-algebra.

Finally, denote $\mu^*|_{\mathcal{F}}$ by μ. Clearly $\mu(\varnothing) = 0$. For countably many disjoint $A_i \in \mathcal{F}$ such that $A = \cup_i A_i \in \mathcal{F}$, let $B_n = \cup_{i=1}^n A_i \in \mathcal{F}$. Then

$$\mu(A) = \mu(B_n) + \mu(A \cap B_n^c) \geq \mu(B_n) = \sum_{i=1}^n \mu(A_i) \to \sum_i \mu(A_i) \geq \mu(A).$$

Hence $\mu(A) = \sum_i \mu(A_i)$ and μ is a measure on $\mathcal{F}$. ∎

Theorem 1.24 (Carathéodory Extension)
Let ν be a finitely additive, σ-subadditive set function on a semi-algebra $\mathcal{S}$ such that $\nu(\varnothing) = 0$. Then ν can be extended to a measure on $\sigma(\mathcal{S})$.

Proof. By theorem 1.15, ν can be extended to a premeasure μ_0 on the algebra $\mathcal{A}$ generated by $\mathcal{S}$. Define the outer measure by

$$\mu^*(A) = \inf\left\{\sum_i \mu_0(E_i) \,\middle|\, A \subset \cup_i E_i, E_i \in \mathcal{A}\right\}$$

for all $A \subset \Omega$ with the convention that $\inf \varnothing = \infty$. We check that μ^* is indeed an outer measure. Clearly $\mu^*(\varnothing) = 0$. If $A \subset B$, then any cover of B by sets in $\mathcal{A}$ is also a cover of A and hence $\mu^*(A) \leq \mu^*(B)$. For countably many $A_i \subset \Omega$, we can find $\left\{E_{ij}\right\}_j$ covering A_i such that

$$\sum_j \mu_0(E_{ij}) \leq \mu^*(A_i) + 2^{-i}\epsilon$$

for some $\epsilon > 0$. Then $\cup_{i,j} E_{ij}$ covers $\cup_i A_i$ and

$$\mu^*(\cup_i A_i) \leq \sum_i \sum_j \mu_0(E_{ij}) \leq \sum_i \mu^*(A_i) + \epsilon.$$

Since ϵ is arbitrary, we conclude that $\mu^*(\cup_i A_i) \leq \sum_i \mu^*(A_i)$ and μ^* is indeed an outer measure.

It follows from lemma 1.23 that the collection of all μ^*-measurable sets forms a σ-algebra $\mathcal{F}$ and μ^* restricted on $\mathcal{F}$ is a measure. It is clear that $\mathcal{A} \subset \mathcal{F}$ and $\sigma(\mathcal{S}) \subset \sigma(\mathcal{A}) \subset \mathcal{F}$ and $\mu = \mu^*|_{\sigma(\mathcal{S})}$ is also a measure. Finally, for $A, A_i \in \mathcal{S}$ where A_i covers A,

$$\mu(A) = \mu^*(A) \leq \nu(A) \leq \sum_i \nu(A \cap A_i) \leq \sum_i \nu(A_i).$$

Taking the infimum over all such covers, we get $\nu(A) = \mu^*(A)$ and μ is indeed an extension of ν. ■

Remark
If the measures are probability measures, then we have that the extension is unique by corollary 1.19.

Theorem 1.25
If F is non-decreasing, right-continuous and satisfies that $F(-\infty) = 0$, $F(\infty) = 1$, then there is a unique probability measure such that

$$\mathrm{P}((-\infty, x]) = F(x).$$

Proof. Define

$$\mathcal{S} = \{(a, b] \mid -\infty \leq a < b \leq \infty\} \cup \{\varnothing\}.$$

It is clear that $\mathcal{S}$ is a semi-algebra. Define the set function $\mathrm{P} : \mathcal{S} \to [0,1]$ by

$$\mathrm{P}((a,b]) = F(b) - F(a)$$

and $\mathrm{P}(\varnothing) = 0$. For disjoint, at most countable $(a_i, b_i] \in \mathcal{S}$, we define

$$\mathrm{P}(\cup_i (a_i, b_i]) = \sum_i \mathrm{P}((a_i, b_i]) = \sum_i F(b_i) - F(a_i).$$

It is clear that P is finitely additive. If $(a,b] = \cup_i (a_i, b_i]$ for disjoint $(a_i, b_i] \in \mathcal{S}$, we may assume without loss of generality that $a = a_1 < b_1 < b_2 < \cdots < b_n = b$ and

$$\mathrm{P}((a,b]) = F(b) - F(a) = \sum_i F(b_i) - F(a_i) = \sum_i \mathrm{P}((a_i, b_i]).$$

Hence P is σ-additive. It now follows from the Carathéodory extension theorem that P can be extended uniquely to a probability measure on $\sigma(\mathcal{S}) = \mathcal{B}$. ■

Remark
This theorem shows that the distribution function completely characterizes the probability measure. In other words, the term "distribution function" can refer to either the CDF or the probability measure.

1.2. Random Variable

Definition 1.26

Let Ω be a probability space. A ***random variable*** *X is a measurable function $X : \Omega \to (S, \mathcal{S})$, where $(S, \mathcal{S})$ is a measurable space.*

Remark

The codomain is often taken to be $(\mathbb{R}, \mathcal{B})$ or $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, but it is also possible to define random functions, i.e., $(S, \mathcal{S})$ is a function space.

Definition 1.27

Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ be a random variable. The ***distribution*** *of X is the pushforward measure of P under X, i.e.,*

$$\mu_X(A) = \mathrm{P}(X \in A) = \mathrm{P}(X^{-1}(A)), \quad A \in \mathcal{S}.$$

Definition 1.28

Let $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B})$ be a random variable. The ***cumulative distribution function*** *of X is defined as*

$$F_X(x) = \mathrm{P}(X \le x) = \mathrm{P}(X_1 \le x_1, \ldots, X_d \le x_d), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Proposition 1.29

Let $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be a random variable and F be its cumulative distribution function. Then,

(a) F is non-decreasing, i.e., $x \le y$ implies $F(x) \le F(y)$;

(b) $F(-\infty) = 0$ and $F(\infty) = 1$;

(c) F is right-continuous, i.e., $\lim_{y \to x^+} F(y) = F(x)$;

(d) $F(x^-) = \mathrm{P}(X < x)$;

(e) $\mathrm{P}(X = x) = F(x) - F(x^-)$.

Proof. (a) comes from that $\{X \le x\} \subset \{X \le y\}$ for $x \le y$.

Take $a_n \to \infty$. Then $\{X \le a_n\} \nearrow \Omega$ and $\{X \le -a_n\} \searrow \emptyset$. By theorem 1.6, we have that

$$F(a_n) = \mathrm{P}(X \le a_n) \to \mathrm{P}(\Omega) = 1, \quad F(-a_n) = \mathrm{P}(X \le -a_n) \to \mathrm{P}(\emptyset) = 0.$$

(c) is similar to (b). Take $y_n \to x^+$, then $\{X \le y_n\} \searrow \{X \le x\}$. By theorem 1.6, we have that

$$F(y_n) = \mathrm{P}(X \le y_n) \to \mathrm{P}(X \le x) = F(x).$$

For (d), take $x_n \to x^-$, then $\{X \le x_n\} \nearrow \{X < x\}$. By theorem 1.6, we have that

$$F(x_n) = \mathrm{P}(X \le x_n) \to \mathrm{P}(X < x).$$

For (e), $P(X = x) = P(X \le x) - P(X < x) = F(x) - F(x^-)$. ■

Theorem 1.30
Let F be a non-decreasing, right-continuous function satisfying that $F(-\infty) = 0$ and $F(\infty) = 1$. Then there is a random variable X such that

$$F(x) = \mu_X((-\infty, x]).$$

Proof. Put $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}$, P be the Lebesgue measure and $X(\omega) = \sup\{x \mid F(x) < \omega\}$. Notice that

$$\begin{aligned}\{X \le x\} &= \{\omega \in \Omega \mid \sup\{y \mid F(y) < \omega\} \le x\} \\ &= \{\omega \in \Omega \mid \text{for all } y > x, F(y) \ge \omega\} \\ &= \{\omega \in \Omega \mid F(x) \ge \omega\}.\end{aligned}$$

Hence $P(X \le x) = P(\{\omega \in \Omega \mid \omega \le F(x)\}) = F(x)$. ■

Definition 1.31
If X and Y are random variables mapping to some measurable space $(S, \mathcal{S})$, then X and Y are said to be ***equal in distribution*** *if $\mu_X = \mu_Y$, denoted by $X \overset{d}{=} Y$.*

Definition 1.32
Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution F. $f : \mathbb{R} \to \mathbb{R}$ is said to be the ***density*** *of X if*

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

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

Remark
If f and g are both densities of X, then $f = g$ a.e.

Remark
If $\mu_X \ll \lambda$, where λ is the Lebesgue measure, then by Radon-Nikodym theorem, there is a density f such that

$$\mu_X(A) = \int_A f(x)d\lambda(x)$$

for all $A \in \mathcal{B}$. Or equivalently, F is absolutely continuous.

Example
Not all random variables have densities, even when its CDF is continuous. Consider the

Cantor function

$$F(x) = \begin{cases} \sum_n \frac{a_n}{2^n}, & x = \sum_n \frac{2a_n}{3^n} \in C \text{ for some } \{a_n\} \in \{0,1\}^{\mathbb{N}} \\ \sup_{y \le x, y \in C} F(y), & x \in [0,1] - C \\ 0, & x < 0 \\ 1, & x > 1 \end{cases}$$

where C is the Cantor set. Then F is a valid CDF, but has no density.

Definition 1.33

A probability measure P *is said to be* ***discrete*** *if there is a countable set S such that* $\mathrm{P}(S^c) = 0$. *A random variable X is said to be* ***discrete*** *if its distribution is.*

Theorem 1.34

Suppose $X : (\Omega, \mathcal{F}) \to (S, \sigma(\mathcal{A}))$ and $\mathcal{A}$ is a collection of subsets in S. If $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{A}$, then X is a random variable.

Proof. Set $\mathcal{G} = \left\{A \subset S \;\middle|\; X^{-1}(A) \in \mathcal{F}\right\}$. Clearly $\varnothing \in \mathcal{G}$ and if $A \in \mathcal{G}$, $X^{-1}(A^c) = (X^{-1}(A))^c \in \mathcal{F}$, so $A^c \in \mathcal{G}$. If $A_n \in \mathcal{G}$, then $X^{-1}(\cup_n A_n) = \cup_n X^{-1}(A_n) \in \mathcal{F}$, so $\cup_n A_n \in \mathcal{G}$. Hence $\mathcal{G}$ is a σ-algebra containing $\mathcal{A}$, so $\sigma(\mathcal{A}) \subset \mathcal{G}$. It follows that $X^{-1}(B) \in \mathcal{F}$ for all $B \in \sigma(\mathcal{A})$, so X is a random variable. ∎

Corollary 1.35

If X_i are random variables, then

$$\inf_i X_i, \quad \sup_i X_i, \quad \liminf_{i\to\infty} X_i, \quad \limsup_{i\to\infty} X_i$$

are all random variables.

Proof. Since the sets of the form $(-\infty, x]$ generate $\mathcal{B}$, it suffices to check that the inverse images of these sets are in $\mathcal{F}$. For $\inf_i X_i$,

$$\left\{\inf_i X_i \le x\right\} = \cup_i \{X_i \le x\} \in \mathcal{F}.$$

For $\sup_i X_i$, since $\sup_i X_i = -\inf_i(-X_i)$, it is also a random variable. Finally, write

$$\liminf_i X_i = \sup_n \inf_{i \ge n} X_i, \quad \limsup_i X_i = \inf_n \sup_{i \ge n} X_i.$$

The results follow from the measurability of $\inf_i X_i$ and $\sup_i X_i$. ∎

Definition 1.36

Let X be a random variable. $\sigma(X)$ is the smallest σ-algebra such that X is measurable.

Remark

If $X : \Omega \to (S, \mathcal{S})$, then $\sigma(X) = X^{-1}(\mathcal{S})$.

Definition 1.37
*Let X be a random variable. The **expectation** of X is defined as*

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

Theorem 1.38 (Jensen's Inequality)
Let $X : \Omega \to \mathbb{R}^d$ be a random variable such that $\mathrm{E}[\|X\|_1] < \infty$ and $\phi : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then

$$\phi(\mathrm{E}[X]) \le \mathrm{E}[\phi(X)].$$

Proof. For any given $y \in \mathbb{R}^d$, note that $\left\{x \in \mathbb{R}^d \,\middle|\, \phi(x) > \phi(y)\right\}$ is a open convex set. By the Hahn-Banach seperation theorem, there is a hyperplane $\{f(x) = a + \langle b, x\rangle\}$ separating $\left\{(x, \phi(x)) \in \mathbb{R}^{d+1} \,\middle|\, \phi(x) > \phi(y)\right\}$ and $\{(y, \phi(y))\}$. Note that $\phi(y) = f(y)$ and $\phi(x) \ge f(x)$ for all $x \in \mathbb{R}^d$. Take $y = \mathrm{E}[X]$, then

$$\phi(\mathrm{E}[X]) = f(\mathrm{E}[X]) = \mathrm{E}[f(X)] \le \mathrm{E}[\phi(X)].$$

■

Theorem 1.39 (Hölder's Inequality)
Let X, Y be random variables and $p, q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\mathrm{E}[|XY|] \le \mathrm{E}[|X|^p]^{1/p}\,\mathrm{E}[|Y|^q]^{1/q}.$$

Proof. If $\mathrm{E}[|X|^p]$ and $\mathrm{E}[|Y^q|]$ are zero or infinite, the result is trivial. We assume that $\mathrm{E}[|X|^p] = \mathrm{E}[|Y|^q] = 1$. For fixed $y \ge 0$, set $\phi(x) = x^p/p + y^p/p - xy$ for $x \ge 0$.

$$\phi'(x) = x^{p-1} - y, \quad \phi''(x) = (p-1)x^{p-2} \ge 0.$$

Thus ϕ is convex and minimized at $x = y^{1/(p-1)}$ with minimum $\phi(y^{1/(p-1)}) = 0$. Hence $x^p/p + y^q/q \ge xy$ for all $x, y \ge 0$.

$$\mathrm{E}[|XY|] \le \mathrm{E}\left[\frac{|X|^p}{p} + \frac{|Y|^q}{q}\right] = \frac{1}{p} + \frac{1}{q} = 1 = \mathrm{E}[|X|^p]^{1/p}\,\mathrm{E}[|Y|^q]^{1/q}.$$

■

Theorem 1.40 (Markov's Inequality)
If $X \ge 0$ is a random variable, then for any $c > 0$,

$$\mathrm{P}(X \ge c) \le \frac{1}{c}\,\mathrm{E}[X].$$

Proof.

$$\mathrm{P}(X \ge c) = \int \mathbf{1}\{X \ge c\}\, d\,\mathrm{P} \le \int \frac{X}{c} d\,\mathrm{P} = \frac{1}{c}\,\mathrm{E}[X].$$

$\blacksquare$

Example
Suppose $\phi : \mathbb{R} \to \mathbb{R}$ is a non-negative function. Put

$$I_A = \inf_{y \in A} \phi(y),$$

where A is some measurable set. Then for any random variable X,

$$I_A \mathbf{1}\{X \in A\} \leq \phi(x)\mathbf{1}\{X \in A\} \leq \phi(x).$$

Thus

$$I_A \, \mathrm{P}(X \in A) \leq \mathrm{E}\left[\phi(X)\right].$$

Corollary 1.41 (Chebyshev's Inequality)
Let X be a random variable. Then for any $c > 0$ and $\alpha \in \mathbb{R}$,

$$\mathrm{P}(|X - \alpha| \geq c) \leq \frac{1}{c^2} \mathrm{E}\left[(X-\alpha)^2\right].$$

Proof. By the Markov's inequality,

$$\mathrm{P}(|X - \alpha| \geq c) = \mathrm{P}((X-\alpha)^2 \geq c^2) \leq \frac{1}{c^2} \mathrm{E}\left[(X-\alpha)^2\right].$$

$\blacksquare$

Theorem 1.42
Suppose X is a random variable of $(S, \mathcal{S})$ with distribution μ and $f : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B})$ is measurable. If either

(a) $f \geq 0$, *or*

(b) $\mathrm{E}\left[|f(X)|\right] < \infty$,

then

$$\mathrm{E}\left[f(X)\right] = \int f(x) d\mu(x).$$

Proof. Suppose first that $f = \mathbf{1}_A$ for some $A \in \mathcal{S}$. Then

$$\mathrm{E}\left[f(X)\right] = \mathrm{P}(X \in \mathcal{A}) = \mathrm{P}(X^{-1}(A)) = \mu(A) = \int \mathbf{1}_A d\mu.$$

By linearity we can extend this result to simple functions. Now suppose first that (a) holds. For such f, there is a sequence of simple functions $s_n \nearrow f$ and $s_n \circ X \nearrow f \circ X$. By LMCT,

$$\mathrm{E}\left[f(X)\right] = \mathrm{E}\left[\lim_n s_n(X)\right] = \lim_n \mathrm{E}\left[s_n(X)\right] = \lim_n \int s_n d\mu = \int f d\mu.$$

Suppose that (b) is the case. Write $f = f^+ - f^-$ and apply the previous result.

$$\mathrm{E}[f(X)] = \mathrm{E}\left[f^+(X)\right] - \mathrm{E}[f^-(X)] = \int f^+ d\mu - \int f^- d\mu = \int f d\mu.$$

■

Definition 1.43
*The k-**th moment** of a random variable X is* $\mathrm{E}\left[X^k\right]$.

Definition 1.44
*The **variance** of a random variable X is* $\mathrm{Var}\,\mathrm{E}\left[(X - \mathrm{E}[X])^2\right]$.

Definition 1.45
*The **covariance** of two integrable random variables X, Y is*

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

Definition 1.46
For $1 \le p < \infty$, the $\mathcal{L}^p(\Omega, \mathrm{P})$ space is defined as

$$\mathcal{L}^p(\Omega, \mathrm{P}) = \{X : \Omega \to S \mid X \textit{ measurable and } \mathrm{E}[|X|^p] < \infty\}.$$

For $p = \infty$,

$$\mathcal{L}^\infty(\Omega, \mathrm{P}) = \{X : \Omega \to S \mid X \textit{ measurable and } \operatorname{ess\,sup}_{\omega\in\Omega} X(\omega) < \infty\}.$$

Proposition 1.47
Let $1 \le p < q \le \infty$. Then $\mathcal{L}^q(\mathrm{P}) \subset \mathcal{L}^p(\mathrm{P})$.

Proof. Suppose first that $q < \infty$. If $X \in \mathcal{L}^q(\mathrm{P})$, then

$$\mathrm{E}[|X|^p] \le \mathrm{E}[|X|^q\,\mathbf{1}\{|X| \ge 1\}] + \mathrm{E}[|X|^p\,\mathbf{1}\{|X| < 1\}] \le \mathrm{E}[|X|^q] + 1 < \infty.$$

Hence $X \in \mathcal{L}^p(\mathrm{P})$. If $q = \infty$, X is essentially bounded, i.e., $X \le M$ for some $M \in \mathbb{R}$ almost surely. Hence $X \in \mathcal{L}^p$. ■

1.3. Independence

Definition 1.48
*Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. Suppose $\mathcal{F}_\beta \subset \mathcal{F}$, $\beta \in B$ are a collection of sub-σ-algebras. Then $\left\{\mathcal{F}_\beta\right\}$ are **independent** if for all finite $\{\mathcal{F}_i\}_{i=1}^n \subset \left\{\mathcal{F}_\beta\right\}$,*

$$\mathrm{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathrm{P}(A_i)$$

where $A_i \in \mathcal{F}_i$.

Definition 1.49
*A collection of random variables $\left\{X_\beta \,\middle|\, \beta \in B\right\}$ on $(\Omega, \mathcal{F}, \mathrm{P})$ is **independent** if the collection of the generating σ-algebras $\left\{\sigma(X_\beta) \,\middle|\, \beta \in B\right\}$ is.*

Remark
In other words,

$$\mathrm{P}(\cap_i \left\{X_{\beta_i} \in A_i\right\}) = \prod_i \mathrm{P}(X_{\beta_i} \in A).$$

Note that these random variables can map into different measurable space.

Definition 1.50
*A collection of events $\mathcal{S}$ is **independent** if $\{\mathbf{1}_A \mid A \in \mathcal{S}\}$ is.*

Proposition 1.51
Let $X_1, \ldots, X_n$ be independent random variables and $g_1, \ldots g_n$ are measurable functions. Then $g_1(X_1), \ldots, g_n(X_n)$ are independent.

Proof. Suppose $g_i : (S_i, \mathcal{S}_i) \to (T_i, \mathcal{T}_i)$. For $A_i \in \mathcal{T}_i$, $g^{-1}(A_i) \in \mathcal{S}_i$ and

$$\mathrm{P}(\cap_i \{g_i(X_i) \in A_i\}) = \mathrm{P}(\cap_i \left\{X_i \in g^{-1}(A_i)\right\}) = \prod_i \mathrm{P}(X_i \in g^{-1}(A_i)) = \prod_i \mathrm{P}(g_i(X_i) \in A_i).$$

$g_1(X_1), \ldots, g_n(X_n)$ are independent. ∎

Theorem 1.52
Let $\mathcal{S}_1, \ldots \mathcal{S}_n$ be a collection of π-system. If $\Omega \in \mathcal{S}_i$ for all $i = 1, \ldots, n$ and for all $A_i \in \mathcal{S}_i$,

$$\mathrm{P}(\cap_i A_i) = \prod_i \mathrm{P}(A_i),$$

then $\sigma(\mathcal{S}_1), \ldots, \sigma(\mathcal{S}_n)$ are independent.

Proof. Fix $\mathcal{S}_2, \ldots, \mathcal{S}_n$. Put

$$\mathcal{L} = \left\{A \in \mathcal{F} \,\middle|\, \mathrm{P}(A \cap (\cap_{i=2}^n A_i)) = \mathrm{P}(A) \prod_{i=2}^n \mathrm{P}(A_i), A_i \in \mathcal{S}_i \text{ for } i = 2, \ldots, n\right\}.$$

We claim that $\mathcal{L}$ forms a λ-system. First, by assumption we can pick $A_i = \Omega$ for $i = 2, \ldots, n$ to see that $\Omega \in \mathcal{L}$. Suppose that $A \subset B$, $A, B \in \mathcal{L}$,

$$\begin{aligned} \mathrm{P}((B - A) \cap (\cap_{i=2}^n A_i)) &= \mathrm{P}((B \cap (\cap_{i=2}^n A_i)) - (A \cap (\cap_{i=2}^n A_i))) \\ &= \mathrm{P}(B) \prod_{i=2}^n \mathrm{P}(A_i) - \mathrm{P}(A) \prod_{i=2}^n \mathrm{P}(A_i) = \mathrm{P}(B - A) \prod_{i=2}^n \mathrm{P}(A_i). \end{aligned}$$

Hence $B - A \in \mathcal{L}$. Let $S_j \nearrow S$, $S_j \in \mathcal{L}$. Then

$$\mathrm{P}(S \cap (\cap_{i=2}^n A_i)) = \lim_{j\to\infty} \mathrm{P}(S_j \cap (\cap_{i=2}^n A_i)) = \lim_{j\to\infty} \mathrm{P}(S_j) \prod_{i=2}^n \mathrm{P}(A_i) = \mathrm{P}(S) \prod_{i=2}^n \mathrm{P}(A_i).$$

Thus $S \in \mathcal{L}$ and $\mathcal{L}$ is a λ-system. By Dynkin's π-λ, $\sigma(\mathcal{S}_1), \mathcal{S}_2, \ldots, \mathcal{S}_n$ satisfies the product property. Repeat the procedure for $\mathcal{S}_2, \ldots, \mathcal{S}_n$. We have that $\sigma(\mathcal{S}_1), \ldots, \sigma(\mathcal{S}_n)$ satisfies the product property. That is, they are independent. ■

Corollary 1.53
Let $X_1, \ldots, X_n$ be $\mathbb{R}$-valued random variables. Then they are independent if and only if

$$\mathrm{P}(X_1 \le s_1, \ldots, X_n \le s_n) = \prod_{i=1}^n \mathrm{P}(X_i \le s_i)$$

for all $s_i \in \mathbb{R}$, $1 \le i \le n$.

Proof. The sufficient part is trivial. For the converse, put $\mathcal{S}_i = \{\{X_i \le t\} \mid t \in \mathbb{R}\} \cup \{\Omega\}$. Clearly $\mathcal{S}_i$ are π-system and $\Omega \in \mathcal{S}_i$ for all i. $\sigma(\mathcal{S}_i)$ are independent and $\mathcal{S}_i$ generates $\sigma(X_i)$. Applying theorem 1.52 shows that X_i are independent. ■

Corollary 1.54
If $\mathcal{F}_{ij}$, $1 \le i \le n, 1 \le j \le m(i)$ are independent σ-algebras, then $\mathcal{G}_i = \sigma(\cup_j \mathcal{F}_{ij})$ are independent.

Proof. Put $\mathcal{H}_i = \{\cap_j A_j \mid A_j \in \mathcal{F}_{ij}\}$. We claim that $\sigma(\mathcal{H}_i) = \mathcal{G}_i$. Indeed, by choosing sets of the form

$$(\Omega, \ldots, \Omega, A_j, \Omega, \ldots, \Omega) \in \mathcal{F}_{i1} \times \cdots \times \mathcal{F}_{im(i)},$$

it is clear that $\cup_j \mathcal{F}_{ij} \subset \mathcal{H}_i$. Also, if $A \in \mathcal{H}_i$, then

$$A = \cap_j A_j = (\cup_j (A_j^c))^c \in \sigma(\cup_j \mathcal{F}_{ij}).$$

Thus $\cup_j \mathcal{F}_{ij} \subset \mathcal{H}_i \subset \sigma(\cup_j \mathcal{F}_{ij})$ and $\sigma(\mathcal{H}_i) = \sigma(\cup_j \mathcal{F}_{ij}) = \mathcal{G}_i$. Also notice that $\mathcal{H}_i$ contain Ω and form π-systems. For $A_i \in \mathcal{H}_i$, write $A_i = \cap_j A_{ij}$. Then

$$\mathrm{P}(\cap_i A_i) = \mathrm{P}(\cap_{ij} A_{ij}) = \prod_{ij} \mathrm{P}(A_{ij}) = \prod_i \mathrm{P}(\cap_j A_{ij}) = \prod_i \mathrm{P}(A_i).$$

From theorem 1.52 we know that $\mathcal{G}_i = \sigma(\mathcal{H}_i)$ are independent. ■

Corollary 1.55
If $X_{ij}, 1 \leq i \leq n, 1 \leq j \leq m(i)$ are independent random variables, then $Y_i = h_i(X_{i1}, \ldots, X_{im(i)})$ are independent provided that h_i are measurable.

Proof. Write $\mathcal{F}_{ij} = \sigma(X_{ij})$. We claim that $\sigma(Y_i) \subset \sigma(\cup_j \mathcal{F}_{ij})$. Indeed, if B_i is a measurable set, $h_i^{-1}(B_i)$ is measurable. Write $h_i^{-1}(B_i) = C_{i1} \times \cdots \times C_{im(i)}$ and since each $X_{ij}^{-1}(C_{ij}) \in \mathcal{F}_{ij}$, we see that $\sigma(Y_i) \subset \sigma(\cup_j \mathcal{F}_{ij})$. It then follows from corollary 1.54 that $\sigma(Y_i)$ are independent and Y_i are independent. ■

Theorem 1.56
If $X_1, \ldots X_n$ are independent $\mathbb{R}$-valued random variables and the distribution of X_i is μ_i. Then the joint distribution of $(X_1, \ldots, X_n)$ is $\mu_1 \times \cdots \times \mu_n$.

Proof. Let μ be the distribution of $(X_1, \ldots, X_n)$. By definition,

$$\begin{aligned}\mu((X_1, \ldots) \in A_1 \times \cdots \times A_n) &= \mu(X_1 \in A_1, \ldots, X_n \in A_n) \\ &= \prod_{i=1}^{n} \mu_i(X_i \in A_i) = (\mu_1 \times \cdots \times \mu_n)(A_1 \times \cdots \times A_n).\end{aligned}$$

Now the sets of the forms $A = A_1 \times \cdots \times A_n$ is a π-system generating the product σ-algebra. By corollary 1.19, the joint distribution is exactly $\mu_1 \times \cdots \times \mu_n$. ■

Theorem 1.57
Let X, Y be two independent random variables. If $h(x, y)$ satisfies either

(a) $\mathrm{E}\left[|h(X,Y)|\right] < \infty$*, or*

(b) *h is non-negative,*

then

$$\mathrm{E}\left[h(X,Y)\right] = \int\int h d\mu_X d\mu_Y,$$

where μ_X, μ_Y are the distributions of X and Y, respectively.

Proof. The proof follows directly from Fubini-Tonelli theorem. If one of the assumptions is true, then

$$\mathrm{E}\left[h(X,Y)\right] = \int_{\mathbb{R}^2} h d(\mu_X \times \mu_Y) = \int\int h d\mu_X d\mu_Y.$$

■

Remark
If $h(x, y) = h_1(x)h_2(y)$, then

$$\mathrm{E}\left[h_1(X)h_2(Y)\right] = \mathrm{E}\left[h(X,Y)\right] = \int\int h_1 h_2 d\mu_X d\mu_Y = \mathrm{E}\left[h_1(X)\right]\mathrm{E}\left[h_2(Y)\right].$$

Corollary 1.58

If $X_1, \dots X_n$ are independent random variables and

(a) $\mathrm{E}\left[|X_1 \cdots X_n|\right] < \infty$ *or*

(b) $X_i \geq 0$ *for all* i,

then

$$\mathrm{E}\left[X_1 \cdots X_n\right] = \prod_{i=1}^{n} \mathrm{E}\left[X_i\right].$$

Proof. Let $h(x, y) = xy$. By assumptions, we have either $\mathrm{E}\left[|h(X_1, X_2)|\right] < \infty$ or $h(X_1, X_2) \geq 0$. By theorem 1.57, $\mathrm{E}\left[X_1 X_2\right] = \mathrm{E}\left[X_1\right] \mathrm{E}\left[X_2\right]$. Substitute X_1 by $X_1 X_2$ and X_2 by X_3, we see that $\mathrm{E}\left[X_1 X_2 X_3\right] = \mathrm{E}\left[X_1\right] \mathrm{E}\left[X_2\right] \mathrm{E}\left[X_3\right]$. Repeat the procedure n times and the result follows. ∎

Definition 1.59

*Let X, Y be independent random variables with CDF F and G, respectively. The **convolution** of two CDF is defined as*

$$(F * G)(z) = \int F(z - y) dG(y).$$

Remark

If F and G are absolutely continuous with respect to the Lebesgue measure, then they have Radon-Nikodym derivatives f and g. The definition of convolution becomes

$$(F * G)(z) = \int F(z - y) dG(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f(x) g(y) dx dy.$$

Then

$$(F * G)'(z) = \int f(z - y) g(y) dy = (f * g)(z),$$

which is exactly the definition of convolution of two functions.

Proposition 1.60

Let X and Y be independent random variables. Then

$$\mathrm{P}(X + Y \leq z) = (F * G)(z).$$

Proof. By theorem 1.57,

$$\begin{aligned}\mathrm{P}(X + Y \leq z) = \mathrm{E}\left[\mathbf{1}\left\{X + Y \leq z\right\}\right] &= \int \int \mathbf{1}\left\{x + y \leq z\right\} dF(x) dG(y) \\ &= \int F(z - y) dG(y) = (F * G)(z).\end{aligned}$$

∎

Remark

Note that the convolution is commutative since

$$(F * G)(z) = \mathrm{P}(X + Y \leq z) = \mathrm{P}(Y + X \leq z) = (G * F)(z).$$

Remark
For discrete X and Y, the convolution becomes

$$\mathrm{P}(X+Y=z)=\sum_{y}\mathrm{P}(X=z-y)\,\mathrm{P}(Y=y).$$

Example
Consider $X\sim\Gamma(\alpha_1,\beta)$ and $Y\sim\Gamma(\alpha_2,\beta)$. Then the density for $X+Y$ is

$$\begin{aligned}
f_{X+Y}(z)&=\int f_X(z-y)f_Y(y)dy\\
&=\int_0^z\frac{1}{\Gamma(\alpha_1)}\beta^{\alpha_1}(z-y)^{\alpha_1-1}e^{-\beta(z-y)}\frac{1}{\Gamma(\alpha_2)}\beta^{\alpha_2}y^{\alpha_2-1}e^{-\beta y}dy\\
&=\frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\beta^{\alpha_1+\alpha_2}e^{-\beta z}\int_0^z(z-y)^{\alpha_1-1}y^{\alpha_2-1}dy\\
&=\frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\beta^{\alpha_1+\alpha_2}e^{-\beta z}z^{\alpha_1+\alpha_2-1}\int_0^1(1-t)^{\alpha_1-1}t^{\alpha_2-1}dt\\
&=\frac{B(\alpha_1,\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\beta^{\alpha_1+\alpha_2}e^{-\beta z}z^{\alpha_1+\alpha_2-1}=\frac{1}{\Gamma(\alpha_1+\alpha_2)}\beta^{\alpha_1+\alpha_2}e^{-\beta z}z^{\alpha_1+\alpha_2-1}.
\end{aligned}$$

Hence $X+Y\sim\Gamma(\alpha_1+\alpha_2,\beta)$.

1.4. Convergence of Random Variables

Definition 1.61
A sequence of probability measures P_n *on* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *are* ***consistent*** *if*

$$\mathrm{P}_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R} = \mathrm{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

for every n.

Theorem 1.62 (Kolmogorov Extension)
Suppose that a sequence of probability measures P_n *on* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *are consistent. Then there is a unique probability measure* P *on* $(\mathbb{R}^{\mathbb{N}}, \mathcal{B})$ *satisfying that*

$$\mathrm{P}(\left\{\omega \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \omega_i \in (a_i, b_i], 1 \le i \le n\right\}) = \mathrm{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n]),$$

where $\mathcal{B}$ *is generated by the collection*

$$\left\{\omega \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \omega_i \in (a_i, b_i], 1 \le i \le n, n \in \mathbb{N}\right\}.$$

Proof. Let

$$\mathcal{S} = \{(a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R} \times \cdots \mid n \in \mathbb{N}\}.$$

Define P on $\mathcal{S}$ to be

$$\mathrm{P}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R} \times \cdots) = \mathrm{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

Clearly, $\mathcal{S}$ forms a semi-algebra. From the Carathéodory extension theorem, it suffices to show that P is finitely additive, σ-additive on $\mathcal{S}$ and $\mathrm{P}(\varnothing) = 0$. Note that $\mathrm{P}(\varnothing) = \mathrm{P}(\varnothing \times \mathbb{R} \times \cdots) = \mathrm{P}_1(\varnothing) = 0$. We verify the first two conditions.

First, if $A, B \in \mathcal{S}$ are disjoint, $m \le n$,

$$A = \left\{\omega \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \omega_i \in (a_i, b_i], 1 \le i \le m\right\} \quad \text{and} \quad B = \left\{\omega \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \omega_i \in (c_i, d_i], 1 \le i \le n\right\},$$

then

$$\mathrm{P}(A \cup B) = \mathrm{P}_n((\pi_n A) \cup (\pi_n B)) = \mathrm{P}_n(\pi_n A) + \mathrm{P}_n(\pi_n B) = \mathrm{P}(A) + \mathrm{P}(B),$$

where $\pi_n : \omega \to (\omega_1, \ldots, \omega_n)$ is the projection onto the first n components. Hence P is finitely additive.

Next, suppose $A_1, \ldots \in \mathcal{S}$ are countably many disjoint measurable sets. Put $A = \cup_i A_i$. We can consider the algebra $\bar{\mathcal{S}} = \{\text{finite disjoint union of sets in } \mathcal{S}\}$ generated by $\mathcal{S}$. $B_n = \cup_{i>n} A_i \in \bar{\mathcal{S}}$. Thus

$$\mathrm{P}(A) = \mathrm{P}(B_n) + \sum_{i=1}^{n} \mathrm{P}(A_n)$$

by the previous result. It now suffices to show that $\mathrm{P}(B_n) \to 0$ for any $B_n \searrow \varnothing$. Suppose not, then there is $\delta > 0$ such that $\mathrm{P}(B_n) \to \delta$ as $B_n \to \emptyset$ by the monotonicity of P.

For such $\{B_n\}$, we claim that there is a sequence of compact set K_n such that $K_n \subset B_n$ and $\mathrm{P}(B_n - K_n) < 2^{-(n+1)}\delta$. Now since $B_1 \in \bar{S}$, there are disjoint $E_1^1, \ldots, E_{m_1}^1$ such that $B_1 = \cup_{i=1}^{m_1} E_i^1$. Now since each E_i^1 is of the product of $(\cdot, \cdot]$. We can find a compact subset K_i^1 of the product of $[\cdot, \cdot]$ such that $\mathrm{P}(E_i^1 - K_i^1) < m_1^{-1} 2^{-2}\delta$. Hence $K_1 = \cup_i K_i^1 \subset B_1$ satisfies that

$$\mathrm{P}(B_1 - K_1) = \sum_{i=1}^{m_1} \mathrm{P}(E_i^1 - K_i^1) < 2^{-2}\delta$$

as desired. Repeat the process and find K_n inductively. The claim follows.

Now, $\cap_{n=1}^m K_n \searrow K$ as $m \to \infty$. Also,

$$\mathrm{P}(B_m - (\cap_{n=1}^m K_n)) \le \sum_{n=1}^m \mathrm{P}(B_n - K_n) \le \frac{\delta}{2}.$$

Hence $\delta/2 \le \mathrm{P}(B_m) - \delta/2 \le \mathrm{P}(\cap_{n=1}^m K_n)$. We see that $\cap_{n=1}^m K_n$ is non-empty for each m. But this implies that $K \subset \cap_n B_n$ is non-empty, a contradiction. Thus $\mathrm{P}(B_n) \to 0$.

Finally, the σ-additivity follows from that we can take $n \to \infty$ so that

$$\mathrm{P}(A) = \lim_{n\to\infty} \mathrm{P}(B_n) + \sum_{i=1}^n \mathrm{P}(A_n) = \sum_i \mathrm{P}(A_n).$$

Applying Carathéodory extension theorem, such P can be extended on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B})$. ■

Remark

With Kolmogorov extension theorem, we can consider a sequence of independent variable X_i on the product probability space with $\mathcal{F} = \mathcal{B}$, $\tilde{X}_i : \omega \mapsto \omega_i$ and $\mathrm{P}(B_1 \times \cdots B_n) = \prod_{i=1}^n \mu_i(B_i)$, where μ_i is the distribution of X_i.

Definition 1.63

Let X_n be a sequence of random variable. X_n ***converges almost surely*** *to X if*

$$\mathrm{P}\left\{\lim_{n\to\infty} X_n = X\right\} = 1.$$

We denote it as $X_n \xrightarrow{a.s.} X$ or $X_n \to X$ a.s.

Definition 1.64

Let X_n be a sequence of random variable. X_n ***converges in probability*** *to X if for every $\epsilon > 0$,*

$$\mathrm{P}\{|X_n - X| > \epsilon\} \to 0$$

as $n \to \infty$. We denote it as $X_n \xrightarrow{p} X$.

Definition 1.65

A sequence of random variable $X_n \in \mathcal{L}^p$ is said to ***converge in*** *$\mathcal{L}^p$ to X if*

$$\mathrm{E}\left[|X_n - X|^p\right]^{1/p} \to 0$$

as $n \to \infty$. If $p = \infty$, the definition becomes

$$\operatorname{ess\,sup}_{\omega \in \Omega} |X_n(\omega) - X(\omega)| \to 0.$$

We denote it as $X_n \to X$ in $\mathcal{L}^p$.

Proposition 1.66

Let X_n be a sequence of independent and indentically distributed random variables. Then

(a) If $X_n \to X$ almost surely, then $X_n \xrightarrow{p} X$.

(b) If $X_n \to X$ in $\mathcal{L}^p$, then $X_n \xrightarrow{p} X$.

Proof. For (a), given $\epsilon > 0$, put

$$E_k = \cup_{n \geq k} \left\{|X_n - X| > \epsilon\right\}.$$

Note that $E_k \searrow E = \{|X_n - X| > \epsilon \text{ for infinitely many } n\} = \{\lim_{n\to\infty} X_n = X\}^c$. Hence

$$\mathrm{P}\left\{|X_k - X| > \epsilon\right\} \leq \mathrm{P}(E_k) \to \mathrm{P}\left\{\lim_{n\to\infty} X_n = X\right\}^c = 0$$

Hence $X_n \to X$ in probability.

For (b), suppose first that $p < \infty$. By Markov inequality,

$$\mathrm{P}\left\{|X_n - X| > \epsilon\right\} = \mathrm{P}\left\{|X_n - X|^p > \epsilon^p\right\} \leq \frac{1}{\epsilon^p}\,\mathrm{E}\left[|X_n - X|^p\right] \to 0.$$

Let $p = \infty$. Note that $\operatorname{ess\,sup}|X_n - X| = \inf\left\{c \mid \mathrm{P}\left\{|X_n - X| > c\right\} = 0\right\}$. Convergence in $\mathcal{L}^\infty$ implies that for $\epsilon > 0$, there is N such that if $n \geq N$, $\inf\left\{c \mid \mathrm{P}\left\{|X_n - X| > c\right\} = 0\right\} < \epsilon$. That is, $\mathrm{P}\left\{|X_n - X| > \epsilon\right\} = 0$ for $n \geq N$. Hence $X_n \xrightarrow{p} X$. ∎

2. Asymptotic Theory

2.1. Law of Large Number

Definition 2.1
*Let X_i be random variables with $\mathrm{E}\left[X_i^2\right] < \infty$. They are called **uncorrelated** if*

$$\mathrm{E}\left[X_i X_j\right] = \mathrm{E}\left[X_i\right]\mathrm{E}\left[X_j\right].$$

Theorem 2.2 (Weak Law of Large Number I)
Suppose that X_n are uncorrelated random variables with $\mathrm{Var}\left[X_n\right] \leq C\infty$ and $\mathrm{E}\left[X_n\right] = \mu$ for all n. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{1}{n}S_n \to \mu$$

in $\mathcal{L}^2$ and hence in probability.

Proof. Compute that

$$\mathrm{E}\left[\left(\frac{1}{n}S_n - \mu\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) \leq \frac{C}{n} \to 0.$$

Hence $\frac{1}{n}S_n \to \mu$ in $\mathcal{L}^2$ and thus in probability. ∎

Theorem 2.3 (Weak Law of Large Number II, Khinchin)
Suppose that X_i is a sequence of independent and identically distributed random variables with $\mathrm{E}\left[|X_1|\right] < \infty$. Let $S_n = \sum_{i=1}^n X_i$ and $\mu = \mathrm{E}\left[X_1\right]$. Then

$$\frac{1}{n}S_n \to \mu$$

in $\mathcal{L}^1$ and hence in probability.

Proof. By replacing X_i with $X_i - \mu$, we may assume without loss of generality that $\mu = 0$. Now, for $C > 0$,

$$0 = \mathrm{E}\left[X_i\right] = \mathrm{E}\left[X_i \mathbf{1}\left\{|X_i| > C\right\}\right] + \mathrm{E}\left[X_i \mathbf{1}\left\{|X_i| \leq C\right\}\right].$$

Also,

$$\begin{aligned}
\frac{1}{n}S_n &= \frac{1}{n}\sum_{i=1}^n X_i \mathbf{1}\left\{|X_i| > C\right\} + \frac{1}{n}\sum_{i=1}^n X_i \mathbf{1}\left\{|X_i| \leq C\right\} \\
&= \frac{1}{n}\sum_{i=1}^n (X_i \mathbf{1}\left\{|X_i| > C\right\} - \mathrm{E}\left[X_i \mathbf{1}\left\{|X_i| > C\right\}\right]) + \frac{1}{n}\sum_{i=1}^n (X_i \mathbf{1}\left\{|X_i| \leq C\right\} - \mathrm{E}\left[X_i \mathbf{1}\left\{|X_i| \leq C\right\}\right]).
\end{aligned}$$

Notice that by LDCT,

$$\mathrm{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i\mathbf{1}\{|X_i|>C\}-\mathrm{E}[X_i\mathbf{1}\{|X_i|>C\}])\right|\right]\le 2\,\mathrm{E}[|X_1|\mathbf{1}\{|X_1|>C\}]\to 0$$

as $C\to\infty$ since $|X_1|\mathbf{1}\{|X_1|>C\}\le|X_1|$ and $\mathrm{E}[|X_1|]<\infty$. Also, by Hölder inequality and the independence,

$$\mathrm{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i\mathbf{1}\{|X_i|\le C\}-\mathrm{E}[X_i\mathbf{1}\{|X_i|\le C\}])\right|\right]\le\sqrt{\frac{1}{n}\mathrm{Var}(X_i\mathbf{1}\{|X_i|\le C\})}\le\frac{C}{\sqrt{n}}$$

For any given $\epsilon>0$, there is C such that $2\,\mathrm{E}[|X_1|\mathbf{1}\{|X_1|>C\}]<\epsilon$ and

$$\begin{aligned}\mathrm{E}\left[\left|\frac{1}{n}S_n\right|\right]&\le\mathrm{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i\mathbf{1}\{|X_i|>C\}-\mathrm{E}[X_i\mathbf{1}\{|X_i|>C\}])\right|\right]\\&\quad+\mathrm{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i\mathbf{1}\{|X_i|\le C\}-\mathrm{E}[X_i\mathbf{1}\{|X_i|\le C\}])\right|\right]\\&\le\epsilon+\frac{C}{\sqrt{n}}\to\epsilon\end{aligned}$$

as $n\to\infty$. Since ϵ can be arbitrarily small, we conclude that $\frac{1}{n}S_n\to 0$ in $\mathcal{L}^1$ and hence in probability. ∎

Definition 2.4
Let A_n be a sequence of events.

$$\limsup_{n\to\infty}A_n=\cap_{m=1}^{\infty}\cup_{n=m}^{\infty}A_n$$

and

$$\liminf_{n\to\infty}A_n=\cup_{m=1}^{\infty}\cap_{n=m}^{\infty}A_n.$$

Remark
Observe that

$$\limsup_{n\to\infty}A_n=\{\omega\in\Omega\mid\omega\in A_n\textit{ for infinitely many }n\}$$

and

$$\liminf_{n\to\infty}A_n=\{\omega\in\Omega\mid\omega\in A_n\textit{ for all but finitely many }n\}.$$

Theorem 2.5 (Borel-Cantelli I)
Let A_n be a sequence of events. If $\sum_n\mathrm{P}(A_n)<\infty$, then

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

Proof. Let $\epsilon > 0$ be given. By assumption, there is n_0 such that $\sum_{n \geq n_0} \mathrm{P}(A_n) < \epsilon$. Then

$$\mathrm{P}\left(\limsup_{n\to\infty} A_n\right) = \mathrm{P}(\cap_{m=1}^{\infty} \cup_{n=m}^{\infty} A_n) \leq \mathrm{P}(\cup_{n=n_0}^{\infty} A_n) \leq \sum_{n=n_0}^{\infty} \mathrm{P}(A_n) < \epsilon.$$

Since ϵ can be arbitrarily small, $\mathrm{P}(\limsup_{n\to\infty} A_n) = 0$. ∎

Corollary 2.6
Suppose for $\epsilon > 0$, $\sum_n \mathrm{P}(|X_n - X| > \epsilon) < \infty$. Then $X_n \to X$ almost surely.

Proof. Let $E_k = \left\{|X_n - X| > k^{-1} \text{ for finitely many } n\right\}$. Note that $E_{k+1} \subset E_k$ and $E_k \searrow E = \{X_n \to X\}$. Now we claim that $\mathrm{P}(E_k) = 1$. Consider $E_k^n = \left\{|X_n - X| > k^{-1}\right\}$. For fixed k, by assumption we have $\sum_n \mathrm{P}(E_k^n) < \infty$. By Borel-Cantelli, $\mathrm{P}(\limsup_{n\to\infty} E_k^n) = 0$. Hence

$$\mathrm{P}(E_k) = \mathrm{P}(\left\{|X_n - X| > k^{-1} \text{ for infinitely many } n\right\}^c) = 1 - \mathrm{P}(\limsup_{n\to\infty} E_k^n) = 1.$$

It now follows by the monotone convergence of measures that $\mathrm{P}(E) = 1$. ∎

Remark
Intuitively, if the convergence is sufficiently fast, the convergence in probability may recover almost sure convergence.

Theorem 2.7 (Strong Law of Large Number I)
Let X_i be independent and identically distributed with $\mu = \mathrm{E}[X_1]$ and $\mathrm{E}\left[X_1^4\right] < \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{1}{n} S_n \to \mu$$

almost surely.

Proof. Note that

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{1}{n}S_n - \mu\right)^4\right] &= \frac{1}{n^4}\left(\sum_i \mathrm{E}\left[(X_i-\mu)^4\right] + \sum_{i\neq j} \mathrm{E}\left[(X_i-\mu)^2(X_j-\mu)^2\right]\right) \\
&\leq \frac{1}{n^3}\mathrm{E}\left[(X_1-\mu)^4\right] + \frac{1}{n^4}\binom{n}{2}\binom{4}{2}\mathrm{E}\left[(X_1-\mu)^2\right]^2 \leq \frac{C}{n^2}
\end{aligned}$$

for some constant C. By Checyshev's inequality, for $\epsilon > 0$,

$$\mathrm{P}\left\{\left|\frac{1}{n}S_n - \mu\right| > \epsilon\right\} \leq \frac{1}{\epsilon^4}\mathrm{E}\left[\left(\frac{1}{n}S_n - \mu\right)^4\right] \leq \frac{C}{\epsilon^2 n^2}$$

is absolute summable. Hence by corollary 2.6,

$$\frac{1}{n} S_n \to \mu$$

almost surely. ∎

Theorem 2.8
$X_n \xrightarrow{p} X$ if and only if every subsequence of X_n has a further subsequence converging almost surely.

Proof. Suppose first that $X_n \xrightarrow{p} X$. Given a subsequence $X_{n(k)}$, we can choose $n(k_1) < n(k_2) < \cdots$ such that

$$P(|X_{n(k_i)} - X| > 2^{-i}) < 2^{-i}.$$

Since 2^{-i} is summable, by Borel-Cantelli we have

$$P(|X_{n(k_i)} - X| > 2^{-i} \text{ for infinitely many } i) = 0.$$

In other words,

$$P\left\{X_{n(k_i)} \to X\right\} = P\left\{|X_{n(k_i)} - X| > 2^{-i} \text{ for infinitely many } i\right\}^c = 1.$$

For the converse, suppose that $X_n \not\to X$ in probability. Then there exist $\epsilon, \delta > 0$ and

$$P\left\{|X_{n(k)} - X| > \epsilon\right\} \geq \delta.$$

By assumption there is a further subsequence converging almost surely and thus in probability, i.e.,

$$P\left\{|X_{n(k_j)} - X| > \epsilon\right\} \to 0.$$

This is a contradiction. Hence $X_n \to X$ in probability. ■

Corollary 2.9
Suppose $X_n \xrightarrow{p} X$. Then the followings are true:

(a) *If f is continuous, then $f(X_n) \xrightarrow{p} f(X)$.*

(b) *If $|X_n| \leq Y$ for some $Y \in \mathcal{L}^1$, then* $E[X_n] \to E[X]$.

(c) *If $X_n \in \mathcal{L}^1$ with $X_n \geq 0$, then* $E[X] \leq \liminf_{n\to\infty} E[X_n]$.

Proof. For (a), by theorem 2.8, every subsequence has a further subsequence $X_{n(k_j)} \to X$ almost surely and hence $f(X_{n(k_j)}) \to f(X)$ almost surely. Then by theorem 2.8 again we see that $f(X_n) \xrightarrow{p} f(X)$.

For (b), by theorem 2.8, every subsequence has a furhter subsequence $X_{n(k_j)} \to X$ almost surely and LDCT gives $E\left[X_{n(k_j)}\right] \to E[X]$. This implies that $E[X_n] \to E[X]$ as well.

For (c), note that there is a subsequence X_{n_k} such that $E\left[X_{n_k}\right] \to \liminf_{n\to\infty} E[X_n]$. Now, by theorem 2.8, there is a further subsequence of X_{n_k} such that $X_{n_{k(j)}} \to X$ almost surely. By the Fatou's lemma,

$$E[X] = E\left[\liminf_{j\to\infty} X_{n_{k(j)}}\right] \leq \liminf_{j\to\infty} E\left[X_{n_{k(j)}}\right] = \lim_{k\to\infty} E\left[X_{n_k}\right] = \liminf_{n\to\infty} E[X_n].$$

■

Definition 2.10
Let $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space.*

$$\mathcal{L}^0(\Omega) = \{X : \Omega \to \mathbb{R} \mid X \text{ is } \mathcal{F}\text{-measurable}\}.$$

Remark
In general, the almost convergence notion on $\mathcal{L}^0$ *is not metrizable, i.e., there is no metric* d *on* $\mathcal{L}^0$ *such that*

$$d(X_n, X) \to 0 \quad \Leftrightarrow \quad X_n \to X \quad a.s.$$

To see this, suppose that the almost sure convergence is metrizable. If $X_n \xrightarrow{p} X$*, any subsequence* $X_{n(k)}$ *converges to* X *in probability as well. By theorem 2.8, we can find a further subsequence converging almost surely and hence in metric* d*, but this implies that* $d(X_n, X) \to 0$*. Then* $X_n \to X$ *almost surely, which is absurd since convergence in probability does not imply almost sure convergence in general.*

However, convergence in probability on $\mathcal{L}^0$ *can be metrized. For instance,*

$$d(X, Y) = \mathrm{E}\left[\max\{|X - Y|, 1\}\right].$$

Theorem 2.11 (Borel-Cantelli II)
Let A_n *be independent events and* $\sum_{n=1}^{\infty} \mathrm{P}(A_n) = \infty$*. Then*

$$\mathrm{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

Proof. By assumption we have that $\sum_{n\geq m} \mathrm{P}(A_n) = \infty$ for every $m \in \mathbb{N}$. Notice that $1 + x \leq e^x$. Then

$$\begin{aligned}
\mathrm{P}(\limsup_{n\to\infty} A_n) &= \lim_{m\to\infty} \mathrm{P}(\cup_{n\geq m} A_n) = 1 - \lim_{m\to\infty} \mathrm{P}(\cap_{n\geq m} A_n^c) \\
&= 1 - \lim_{m\to\infty} \lim_{N\to\infty} \mathrm{P}(\cap_{n=m}^{N} A_n^c) = 1 - \lim_{m\to\infty} \lim_{N\to\infty} \prod_{n=m}^{N} \mathrm{P}(A_n^c) \\
&= 1 - \lim_{m\to\infty} \prod_{n=m}^{\infty} (1 - \mathrm{P}(A_n)) \geq 1 - \lim_{m\to\infty} \exp\left(-\sum_{n=m}^{\infty} \mathrm{P}(A_n)\right) = 1.
\end{aligned}$$

Hence $\mathrm{P}(\limsup_{n\to\infty} A_n) = 1$. ■

Lemma 2.12
Let X *be a non-negative random variable and* $h : \mathbb{R} \to \mathbb{R}$ *be a differentiable function with* $h(0) = 0$ *and* $h' \geq 0$*. Then*

$$\mathrm{E}[h(X)] = \int_0^{\infty} h'(t)\, \mathrm{P}(X > t)\, dt.$$

Proof. By Fubini-Tonelli theorem,

$$\begin{aligned}\mathrm{E}\left[h(X)\right] &= \mathrm{E}\left[\int_0^X h'(t)dt\right] = \mathrm{E}\left[\int_0^\infty \mathbf{1}\left\{t < X\right\} h'(t)dt\right] \\ &= \int_0^\infty h'(t)\,\mathrm{E}\left[\mathbf{1}\left\{t < X\right\}\right] dt = \int_0^\infty h'(t)\,\mathrm{P}(X > t)dt.\end{aligned}$$

■

Proposition 2.13

Suppose that X_i are independent and identically distributed random variables with $\mathrm{E}\left[|X_i|\right] = \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then

(a) $\mathrm{P}\left\{|X_n| > n \text{ for infinitely many } n\right\} = 1$.

(b) $\mathrm{P}\left\{\frac{1}{n}S_n \text{ has finite limit}\right\} = 0$.

Proof. For (a), using lemma 2.12 with h being identity,

$$\begin{aligned}\infty = \mathrm{E}\left[|X_1|\right] &= \int_0^\infty \mathrm{P}(|X_1| > t)dt \le \sum_{n=0}^\infty \int_n^{n+1} \mathrm{P}(|X_1| > t)dt \\ &\le \sum_{n=0}^\infty \int_n^{n+1} \mathrm{P}(|X_1| > n)dt = \sum_{n=0}^\infty \mathrm{P}(|X_n| > n).\end{aligned}$$

Now by the second Borel-Cantelli, $\mathrm{P}\left\{|X_n| > n \text{ for infinitely many } n\right\} = 1$.

For (b), consider ω with $\frac{S_n(\omega)}{n} \longrightarrow Y(\omega) \in \mathbb{R}$. Then for such ω,

$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \longrightarrow 0$$

as $n \longrightarrow \infty$. Thus

$$\begin{aligned}\mathrm{P}\left\{\frac{1}{n}S_n \text{ has finite limit}\right\} &\le \mathrm{P}\left\{|X_n| > n \text{ for finitely many } n\right\} \\ &= 1 - \mathrm{P}\left\{|X_n| > n \text{ for infinitely many } n\right\} = 0.\end{aligned}$$

(b) follows. ■

Definition 2.14

*A collection of σ-algebra $\{\mathcal{H}_k\}$ is **pairwise independent** if for any $\mathcal{H}_1, \mathcal{H}_2 \in \{\mathcal{H}_k\}$,*

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$$

for any $A \in \mathcal{H}_1$ and $B \in \mathcal{H}_2$.

Remark

As before, a sequence of random variables $\{X_k\}$ is pairwise independent if $\{\sigma(X_k)\}$ is.

Theorem 2.15 (Strong Law of Large Number II, Kolmogorov)
Let X_i be pairwise independent, identically distributed random variables with $\mathrm{E}\left[|X_1|\right] < \infty$ *and* $S_n = \sum_{i=1}^n X_i$. *Then*

$$\frac{1}{n}S_n \to \mathrm{E}\left[X_1\right] = \mu$$

almost surely.

Proof. Since we can always decompose $X_i = X_i^+ - X_i^-$ and X_i^+, X_i^- satisfy the assumption of the theorem, we may assume without loss of generality that $X_i \geq 0$. Let $Y_i = X_i \mathbf{1}\left\{X_i \leq i\right\}$ and $T_n = \sum_{i=1}^n Y_i$. Let $\alpha > 1$ and put $k_n = \lfloor \alpha^n \rfloor$. By Chebyshev inequality, for any given $\epsilon > 0$ we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathrm{P}\left(\left|\frac{T_{k_n} - \mathrm{E}\left[T_{k_n}\right]}{k_n}\right| > \epsilon\right) &\leq \frac{1}{\epsilon^2}\sum_{n=1}^{\infty}\frac{1}{k_n^2}\operatorname{Var}(T_{k_n}) \\
&= \frac{1}{\epsilon^2}\sum_{n=1}^{\infty}\frac{1}{k_n^2}\sum_{i=1}^{k_n}\operatorname{Var}(Y_i) = \frac{1}{\epsilon^2}\sum_{i=1}^{\infty}\operatorname{Var}(Y_i)\sum_{n:k_n\geq i}\frac{1}{k_n^2}.
\end{aligned}$$

Since $1/k^2$ is summable and k_n repeat at most m_α times, where m_α is an integer such that $\alpha^{m_\alpha+1} \geq \alpha^{m_\alpha} + 1$, we can find a constant $c_\alpha > 0$ such that

$$\sum_{n:k_n\geq i}\frac{1}{k_n^2} \leq \frac{c_\alpha}{i^2}.$$

Let F be the distribution of X. We have

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathrm{P}\left(\left|\frac{T_{k_n} - \mathrm{E}\left[T_{k_n}\right]}{k_n}\right| > \epsilon\right) &\leq \frac{c_\alpha}{\epsilon^2}\sum_{i=1}^{\infty}\frac{\operatorname{Var}(Y_i)}{i^2} \leq \frac{c_\alpha}{\epsilon^2}\sum_{i=1}^{\infty}\frac{\mathrm{E}\left[Y_i^2\right]}{i^2} \\
&= \frac{c_\alpha}{\epsilon^2}\sum_{i=1}^{\infty}\frac{1}{i^2}\int_0^i x^2 dF(x) = \frac{c_\alpha}{\epsilon^2}\sum_{i=1}^{\infty}\sum_{k=0}^{i-1}\frac{1}{i^2}\int_k^{k+1} x^2 dF(x) \\
&= \frac{c_\alpha}{\epsilon^2}\sum_{k=0}^{\infty}\left(\sum_{i=k+1}^{\infty}\frac{1}{i^2}\right)\int_k^{k+1} x^2 dF(x)
\end{aligned}$$

Also, notice that there is a constant C such that

$$\sum_{i=k+1}^{\infty}\frac{1}{i^2} \leq \frac{C}{k+1}.$$

Hence,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathrm{P}\left(\left|\frac{T_{k_n} - \mathrm{E}\left[T_{k_n}\right]}{k_n}\right| > \epsilon\right) &\leq \frac{c_\alpha C}{\epsilon^2}\sum_{k=0}^{\infty}\frac{1}{k+1}\int_k^{k+1} x^2 dF(x) \\
&\leq \frac{c_\alpha C}{\epsilon^2}\sum_{k=0}^{\infty}\int_k^{k+1} x dF(x) = \frac{c_\alpha C}{\epsilon^2}\mathrm{E}\left[X_1\right] < \infty.
\end{aligned}$$

Note that for $\delta > 0$ there is an integer M such that $\mathrm{E}\left[X_1 \mathbf{1}\left\{X_1 > M\right\}\right] \leq \delta \leq \mathrm{E}\left[X_1\right]$.

$$\frac{\mathrm{E}\left[T_{k_n}\right]}{k_n} = \frac{1}{k_n}\sum_{i=1}^{k_n} \mathrm{E}\left[Y_i\right] \geq \frac{1}{k_n}\sum_{i=1}^{M} \mathrm{E}\left[Y_i\right] + \frac{1}{k_n}\sum_{i=M+1}^{k_n} \mathrm{E}\left[X_1\right] - \delta.$$

Also,

$$\frac{1}{k_n}\sum_{i=1}^{k_n} \mathrm{E}\left[Y_i\right] \leq \frac{1}{k_n}\sum_{i=1}^{k_n} \mathrm{E}\left[X_1\right] = \mathrm{E}\left[X_1\right].$$

Taking $n \longrightarrow \infty$ and since δ is arbitrary, we conclude that

$$\frac{\mathrm{E}\left[T_{k_n}\right]}{k_n} \longrightarrow \mathrm{E}\left[X_1\right].$$

Thus, by the Borel-Cantelli lemma,

$$\mathrm{P}\left\{\frac{T_{k_n}}{k_n} \not\rightarrow \mathrm{E}\left[X_1\right]\right\} = \mathrm{P}\left\{\left|\frac{T_{k_n} - \mathrm{E}\left[T_{k_n}\right]}{k_n}\right| > \epsilon \text{ for infinitely many } n\right\} = 0.$$

In other words, $T_{k_n}/k_n \longrightarrow \mathrm{E}\left[X_1\right]$ almost surely. Also,

$$\begin{aligned}\sum_{k=1}^{\infty} \mathrm{P}\left\{X_k \neq Y_k\right\} = \sum_{k=1}^{\infty} \mathrm{P}\left\{X_k > k\right\} &\leq \sum_{k=1}^{\infty} \int_{k-1}^{k} \mathrm{P}(X_1 > t)dt \\ &= \int_0^{\infty} \mathrm{P}(X_1 > t)dt = \mathrm{E}\left[X_1\right] < \infty\end{aligned}$$

by lemma 2.12. Hence by Borel-Cantelli lemma, $X_k \neq Y_k$ for finitely many k almost surely. This implies that

$$\lim_{n\to\infty} \frac{1}{k_n} S_{k_n} = \lim_{n\to\infty} \frac{T_{k_n}}{k_n} = \mathrm{E}\left[X_1\right]$$

almost surely. Note that S_m is monotone and for each m, we may find $k(n_m) \leq m \leq k(n_{m+1})$ so that

$$\frac{S_{k(n_m)}}{k(n_{m+1})} \leq \frac{S_m}{m} \leq \frac{S_{k(n_{m+1})}}{k(n_m)}.$$

Take $m \longrightarrow \infty$, we conclude that

$$\frac{1}{\alpha}\mu \leq \liminf_{m\to\infty} \frac{S_m}{m} \leq \limsup_{m\to\infty} \frac{S_m}{m} \leq \alpha\mu$$

almost surely. Taking $\alpha \longrightarrow 1^+$ gives the desired result. ■

Theorem 2.16

Let X_i be independent and identically distributed with $\mathrm{E}\left[X_1^+\right] = \infty$ and $\mathrm{E}\left[X_1^-\right] < \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{1}{n} S_n \longrightarrow \infty$$

almost surely.

Proof. Write $X_i = X_i^+ - X_i^-$. For X_i^+, consider $Y_i^M = \min\left\{X_i^+, M\right\}$ for some $M > 0$. Note that Y_i^M is independent and identically distributed with finite mean. By the strong law of large number,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^M \to \mathrm{E}\left[Y_1^M\right]$$

almost surely. Hence,

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^+ \geq \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i^M = \mathrm{E}\left[Y_1^M\right].$$

Notice that $Y_1^M \nearrow X_1^+$ as $M \to \infty$. By LMCT,

$$\lim_{M\to\infty} \mathrm{E}\left[Y_1^M\right] = \mathrm{E}\left[X_1^+\right] = \infty.$$

We conclude that $\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^+ = \infty$. On the other hand, by the strong law of large number,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^- \to \mathrm{E}\left[X_1^-\right]$$

almost surely. We end up with

$$\lim_{n\to\infty} \frac{1}{n} S_n = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^+ - \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^- = \infty$$

almost surely. ■

Example
Let Y_i be independent and identically distributed with density

$$f(y) = \mathbf{1}\{y \geq 1\}\frac{1}{c}\frac{1}{y^2},$$

where c is some normalizing constant. Let $H_i \sim Ber(2^{-i})$. Put $X_i = Y_i H_i$. Then $\mathrm{E}[X_i] = \infty$ for all i, but since

$$\sum_i \mathrm{P}(X_i > 0) = \sum_i 2^{-i} < \infty,$$

by the Borel-Cantelli lemma, $X_i \to 0$ almost surely and

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to 0$$

almost surely.

Example

$Y \geq 0$ is a random variable with $\mathrm{E}[Y] = \infty$. Put $X_i = Y$ for all i. Then X_i are identically distributed with $\mathrm{E}[X_i] = \infty$. But

$$\frac{1}{n}\sum_{i=1}^{n} X_i = Y \not\to \infty$$

almost surely.

Example (Event Streaks)
$X_i \overset{iid}{\sim} Ber(2^{-1})$. Let L_n be the longest streaks of 1 in the first n trials. We have the following:

$$\lim_{n\to\infty} \frac{L_n}{\log_2(n)} = 1.$$

To see this, let ℓ_n be the length of the current streaks. For instance, the following sequence

$$1, 0, 1, 1, 1, 1, 0, \ldots$$

generates $\ell_1 = 1, \ell_2 = 0, \ell_6 = 4$. Observe that $L_n = \max_{m\leq n} \ell_m$. Now,

$$\mathrm{P}(\ell_n \geq k) = \sum_{m=k}^{n} \mathrm{P}(\ell_n = k) = \sum_{m=k}^{n} 2^{-k-1} \leq 2^{-k}$$

as $n \to \infty$. For $\epsilon > 0$,

$$\mathrm{P}(\ell_n \geq (1+\epsilon)\log_2(n)) = \mathrm{P}(\ell_n \geq \lceil (1+\epsilon)\log_2(n) \rceil) \leq 2^{-\lceil (1+\epsilon)\log_2(n)\rceil} \leq 2^{-(1+\epsilon)\log_2(n)} = \frac{1}{n^{1+\epsilon}}$$

is summable. By the Borel-Cantelli lemma,

$$\mathrm{P}\left\{\ell_n \geq (1+\epsilon)\log_2(n) \text{ for infinitely many } n\right\} = 0.$$

Hence

$$\mathrm{P}\left\{\ell_n < (1+\epsilon)\log_2(n) \text{ for all but finitely many } n\right\} = 1.$$

That is, for almost every ω, there is $N(\omega)$ such that $\ell_n < (1+\epsilon)\log_2(n)$ for $n \geq N(\omega)$. For such ω, we have

$$L_n(\omega) = \max_{m\leq n} \ell_n(\omega) \leq \max_{m\leq n}(1+\epsilon)\log_2(n) = (1+\epsilon)\log_2(n)$$

as $n > N(\omega)$. Thus

$$\limsup_{n\to\infty} \frac{L_n}{\log_2(n)} \leq 1+\epsilon$$

almost surely. Note that

$$\left\{\limsup_{n\to\infty} \frac{L_n}{\log_2(n)} \leq 1+\epsilon\right\} \searrow \left\{\limsup_{n\to\infty} \frac{L_n}{\log_2(n)} \leq 1\right\}$$

as $\epsilon \to 0^+$ and by the monotone convergence of the measures,

$$\limsup_{n\to\infty} \frac{L_n}{\log_2(n)} \le 1$$

almost surely.

For the other side, note that for large n, we may split the sequence into blocks of size $\lceil (1-\epsilon)\log_2(n) \rceil$ and

$$\frac{n}{\lceil (1-\epsilon)\log_2(n) \rceil} \ge \frac{n}{\log_2(n)}$$

for large n.

$$\begin{aligned}
\mathrm{P}(L_n \le (1-\epsilon)\log_2(n)) &\le \mathrm{P}(\textit{each block did not have all 1s}) \\
&\le (1 - 2^{-\lceil (1-\epsilon)\log_2(n)/2 \rceil})^{n/\lceil (1-\epsilon)\log_2(n)/2 \rceil} \\
&\le \left(1 - \frac{1}{n^{1-\epsilon}}\right)^{n^{1-\epsilon}\frac{n^\epsilon}{\log_2(n)}} \le \exp\left(-\frac{n^\epsilon}{\log_2(n)}\right),
\end{aligned}$$

whcih is summable, so by the Borel Cantelli lemma,

$$\mathrm{P}\left\{L_n \le (1-\epsilon)\log_2(n) \textit{ for infinitely many } n\right\} = 0.$$

By a similar argument as above,

$$\liminf_{n\to\infty} \frac{L_n}{\log_2(n)} \ge 1-\epsilon$$

almost surely and by the monotone convergence of the measures

$$\liminf_{n\to\infty} \frac{L_n}{\log_2(n)} \ge 1$$

almost surely. We conclude that

$$1 \le \liminf_{n\to\infty} \frac{L_n}{\log_2(n)} \le \limsup_{n\to\infty} \frac{L_n}{\log_2(n)} \le 1$$

and the claim follows.

Example (Counting Process)
Let $X_i \in (0,\infty)$ be independent and identically distributed random variable. Put $\mu = \mathrm{E}[X_1]$, $T_n = \sum_{i=1}^n X_i$ and $N_t = \sup\{n \mid T_n \le t\}$. Then we have the following claim:

$$\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{\mu}$$

almost surely. To see this, note that since $X_i < \infty$ for all i,

$$\lim_{t\to\infty} N_t = \lim_{t\to\infty} \sup\{n \mid T_n \le t\} = \infty.$$

Now, observe that $T_{N_t} \le t \le T_{N_t+1}$ and hence

$$\frac{T_{N_t}}{N_t} \le \frac{t}{N_t} \le \frac{T_{N_t+1}}{N_t+1}\frac{N_t+1}{N_t}.$$

By the strong law of large number, $T_{N_t}/N_t \to \mu$ almost surely. Thus

$$\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{\mu}.$$

Theorem 2.17 (Glivenko-Cantelli)
Suppose that $X_i \overset{iid}{\sim} F$ with $X_i \in (-\infty, \infty)$ and

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \le x\}$$

is the empirical CDF. Then

$$\|F_n - F\|_\infty \to 0$$

almost surely when $n \to \infty$.

Proof. We first claim that for $\epsilon > 0$, we may find a finite partition $\{t_j\}$ such that $-\infty = t_0 < \cdots < t_j = \infty$ and

$$F(t_{j+1}^-) - F(t_j) \le \epsilon$$

for all j. To see the existence of such partition, put $t_0 = -\infty$ and let

$$t_{j+1} = \sup\left\{t \in \mathbb{R} \,\middle|\, F(t) \le F(t_j) + \epsilon\right\}.$$

Observe that $F(t_{j+1}) \ge F(t_j) + \epsilon$. If not, then $F(t_{j+1}) < F(t_j) + \epsilon$. By the right-continuity of F, there is $\delta > 0$ such that $F(t_{j+1} + \delta) \le F(t_j) + \epsilon$, contradicting to the definition of t_{j+1}. It now also follows from the definition that

$$F(t_{j+1}^-) \le F(t_j) + \epsilon.$$

Finally, since F is of finite total variation, the jumps of sizes greater than ϵ can occur only finitely many times and we conclude the existence of such partition.

Next, by the strong law of large number, for almost every ω there is $N(\omega)$ uniform in j such that

$$\left|F_n(t_j) - F(t_j)\right| \le \epsilon$$

for all $n > N(\omega)$. For any $t \in [t_j, t_{j+1})$, we have

$$F(t) - F(t_j) \le F(t_{j+1}^{-1}) - F(t_j) \le \epsilon.$$

Again, by the strong law of the large number,

$$F_n(t_{j+1}^-) - F_n(t_j) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\left\{t_j < X_i < t_{j+1}\right\} \to \mathrm{E}\left[\mathbf{1}\left\{t_j < X_i < t_{j+1}\right\}\right] = F(t_{j+1}^-) - F(t_j)$$

almost surely. That is, for almost every ω, there is $N'(\omega) > N(\omega)$ such that for all j,

$$F_n(t_{j+1}^-) - F_n(t_j) \le F(t_{j+1}^-) - F(t_j) + \epsilon$$

if $n \ge N'(\omega)$. Combining the above estimates, if $n \ge N'(\omega)$,

$$\begin{aligned}
|F_n(t) - F(t)| &\le \left|F_n(t) - F_n(t_j)\right| + \left|F_n(t_j) - F(t_j)\right| + \left|F(t_j) - F(t)\right| \\
&\le \left|F_n(t_{j+1}^-) - F_n(t_j)\right| + 2\epsilon \\
&\le F(t_{j+1}^-) - F(t_j) + 3\epsilon \le 4\epsilon.
\end{aligned}$$

Since ϵ is arbitrary, we conclude that $F_n \to F$ uniformly for almost every ω and the proof is complete. ■

Theorem 2.18 (Kolmogorov Maximal Inequality)
Suppose that X_i are independent with $\mathrm{E}[X_i] = 0$ and $\mathrm{Var}[X_i] < \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\mathrm{P}\left(\max_{1\le k\le n} |S_k| \ge x\right) \le \frac{1}{x^2}\mathrm{Var}[S_n].$$

Proof. Let $A_k = \left\{|S_k| \ge x \text{ and } \left|S_j\right| < x \text{ for } 1 \le j \le k-1\right\}$. Note that

$$\sum_{k=1}^n \mathbf{1}_{A_k} = \mathbf{1}\left\{\max_{1\le k\le n} |S_k| \ge x\right\}.$$

$$\begin{aligned}
\mathrm{E}\left[S_n^2\right] \ge \mathrm{E}\left[S_n^2 \sum_{k=1}^n \mathbf{1}_{A_k}\right] &= \sum_{k=1}^n \mathrm{E}\left[S_n^2 \mathbf{1}_{A_k}\right] = \sum_{k=1}^n \mathrm{E}\left[(S_n - S_k + S_k)^2 \mathbf{1}_{A_k}\right] \\
&= \sum_{k=1}^n \mathrm{E}\left[(S_n - S_k)^2\mathbf{1}_{A_k}\right] + 2\,\mathrm{E}\left[(S_n - S_k)S_k\mathbf{1}_{A_k}\right] + \mathrm{E}\left[S_k^2\mathbf{1}_{A_k}\right] \\
&\ge \sum_{k=1}^n \mathrm{E}\left[S_k^2\mathbf{1}_{A_k}\right] + 2\,\mathrm{E}\left[(S_n - S_k)S_k\mathbf{1}_{A_k}\right].
\end{aligned}$$

Notice that $S_n - S_k \in \sigma(X_{k+1}, \ldots, X_n)$ and $S_k \mathbf{1}_{A_k} \in \sigma(X_1, \ldots, X_k)$ are independent. Thus

$$\begin{aligned} \mathrm{E}\left[S_n^2\right] &\geq \sum_{k=1}^n \mathrm{E}\left[S_k^2 \mathbf{1}_{A_k}\right] + 2\,\mathrm{E}\left[(S_n - S_k) S_k \mathbf{1}_{A_k}\right] \\ &= \sum_{k=1}^n \mathrm{E}\left[S_k^2 \mathbf{1}_{A_k}\right] \geq x^2 \sum_{k=1}^n \mathrm{E}\left[\mathbf{1}_{A_k}\right] = x^2\, \mathrm{P}\left(\max_{1\leq k\leq n} |S_k| \geq x\right). \end{aligned}$$

Hence

$$\mathrm{P}\left(\max_{1\leq k\leq n} |S_k| \geq x\right) \leq \frac{1}{x^2}\, \mathrm{E}\left[S_n^2\right].$$

■

Definition 2.19

*Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. A sub-σ-algebra $\mathcal{G}$ is P-**trivial** if for all $A \in \mathcal{G}$, $\mathrm{P}(A) \in \{0, 1\}$.*

Theorem 2.20 (Kolmogorov Zero-One Law)

Let $\mathcal{F}_i$ be independent σ-algebras, $\mathcal{G}_n = \sigma(\mathcal{F}_n, \ldots)$ and $\mathcal{G}_\infty = \cap_{n=1}^\infty \mathcal{G}_n$. Then $\mathcal{G}_\infty$ is P-trivial.

Proof. Observe that a σ-algebra $\mathcal{G}$ satisfies that for $A \in \mathcal{G}$, $\mathrm{P}(A) \in \{0, 1\}$ if $\mathcal{G}$ is independent of itself. Indeed, if $\mathcal{G}$ is independent of itself, then for any $A \in \mathcal{G}$,

$$\mathrm{P}(A) = \mathrm{P}(A \cap A) = \mathrm{P}(A)^2$$

implies that $\mathrm{P}(A) = 0$ or 1. Now, for any given n, $\sigma(\mathcal{F}_1, \ldots, \mathcal{F}_{n-1})$ is independent with $\mathcal{G}_n$ and $\mathcal{G}_\infty \subset \mathcal{G}_n$. Hence $\mathcal{G}_\infty$ is independent of $\sigma(\mathcal{F}_1, \ldots, \mathcal{F}_{n-1})$ for all n.

In particular, $\mathcal{G}_\infty$ is independent of $\sigma(\cup_n \mathcal{F}_n)$. To see this, note that $\cup_n \sigma(\cup_{k=1}^n \mathcal{F}_k)$ is a π system that generates $\sigma(\cup_n \mathcal{F}_n)$ and for $A \in \cup_n \sigma(\cup_{k=1}^n \mathcal{F}_k)$, $A \in \sigma(\cup_{k=1}^n \mathcal{F}_k)$ for some n. For $B \in \mathcal{G}_\infty$,

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$$

since $\sigma(\cup_{k=1}^n \mathcal{F}_k)$ and $\mathcal{G}_\infty$ is independent. Now it follows from theorem 1.52 that $\mathcal{G}_\infty$ and $\sigma(\cup_n \mathcal{F}_n)$ are independent. Notice that $\mathcal{G}_\infty \subset \sigma(\cup_n \mathcal{F}_n)$ and hence $\mathcal{G}_\infty$ is independent of itself. The proof is complete. ■

Corollary 2.21

Let X_i be independent and identically distributed and put $S_n = \sum_{i=1}^n X_i$. Then

(a) S_n is either almost surely convergent or almost surely divergent.

(b) If $\frac{1}{n} S_n$ converges almost surely, its limit is almost surely a constant.

Proof. Define $\mathcal{F}_i = \sigma(X_i)$ and note that $\{S_n \text{ converges}\} \in \mathcal{G}_\infty = \cap_n \sigma(\cup_{i\geq n} \mathcal{F}_i)$ since

$$\{S_n \text{ converges}\} = \left\{\lim_{n\to\infty} \sum_{i\geq n} X_i = 0\right\}$$

is $\mathcal{G}_\infty$-measurable. By the Kolmogorov zero-one law,

$$\mathrm{P}\left\{S_n \text{ converges}\right\} \in \{0, 1\}.$$

This proves (a).

For (b), note that by a similar argument, we have

$$\left\{\frac{1}{n}S_n \text{ converges}\right\} \in \mathcal{G}_\infty,$$

where $\mathcal{G}_\infty$ is P-trivial. Also, $\lim_{n\to\infty} \frac{1}{n}S_n$ is $\mathcal{G}_\infty$-measurable. Since $\mathcal{G}_\infty$ is P-trivial,

$$F(t) := \mathrm{P}\left\{\lim_{n\to\infty} \frac{1}{n}S_n \leq t\right\} \in \{0, 1\}.$$

Thus

$$\lim_{n\to\infty} \frac{1}{n}S_n = \sup\left\{t \mid F(t) = 0\right\}$$

almost surely, proving that the limit is almost surely a constant. ∎

2.2. Convergence in Distribution

Definition 2.22

Let μ_n and μ be probability measures on (S, d). We say that $\mu_n \to \mu$ ***in distribution*** *or* ***weakly*** *if for all $f \in C_b(S, \mathbb{R})$,*

$$\int f d\mu_n \to \int f d\mu.$$

We denote the convergence as $\mu_n \xrightarrow{d} \mu$.

Remark

The notion of convergence is the smallest topology such that the linear functional of the form $\ell_f : \mu \to \int f d\mu$ where $f \in C_b(S)$ is continuous.

Definition 2.23

Let X_n and X be random variables. $X_n \xrightarrow{d} X$ if the corresponding distributions $\mu_n \xrightarrow{d} \mu$.

Remark

The convergence in distribution of the random variables does not depend on the space where the random variables are defined; in fact, they can be defined on different spaces. The definition can also be written as

$$\mathrm{E}[f(X_n)] \to \mathrm{E}[f(X)]$$

for all $f \in C_b(S)$.

Theorem 2.24 (Scheffé)

If $f_n : S \to \mathbb{R}$ are density functions such that $f_n \to f$ almost everywhere, where f is a density function, then

$$\sup_{B \in \mathcal{B}} \left| \int_B f_n dx - \int_B f dx \right| \to 0$$

as $n \to 0$. In particular, when $S = \mathbb{R}$, taking $B = [-\infty, x]$ gives the uniform convergence of the CDFs.

Proof. Since

$$\sup_{B \in \mathcal{B}} \left| \int_B f_n dx - \int_B f dx \right| \leq \sup_{B \in \mathcal{B}} \int_B |f_n - f| \, dx \leq \int |f_n - f| \, dx,$$

the theorem follows once we prove that $f_n \to f$ in $\mathcal{L}^1$. Now, since $|f_n - f| \to 0$ almost everywhere and

$$|f_n - f| \leq |f_n| + |f| \quad \Rightarrow \quad 0 \leq |f_n| + |f| - |f_n - f|.$$

By the assumptions that f_n and f are density functions,

$$\int f_n dx = 1 = \int f dx.$$

By the Fatou's lemma,

$$\begin{aligned}2\int |f|\,dx &= \int \liminf_{n\to\infty} |f_n| + |f| - |f_n - f| \le \liminf_{n\to\infty} \int f_n dx + \int f\,dx - \int |f_n - f|\,dx \\ &= 2\int f\,dx - \limsup_{n\to\infty} \int |f_n - f|\,dx.\end{aligned}$$

Hence

$$\limsup_{n\to\infty} \int |f_n - f|\,dx \le 0 \quad \Rightarrow \quad \int |f_n - f|\,dx \to 0.$$

Hence $f_n \to f$ in $\mathcal{L}^1$ and the proof is complete. ∎

Proposition 2.25
Let X_n, X be random variables on a separable metric space (S, d). Then $X_n \xrightarrow{p} X$ implies $X_n \xrightarrow{d} X$.

Proof. Let $X_n \xrightarrow{p} X$. To show that $X_n \xrightarrow{d} X$, we need to show that $\mathrm{E}[f(X_n)] \to \mathrm{E}[f(X)]$. Suppose that this does not hold. There is a subsequence X_{n_k} such that $X_{n_k} \to X$ almost surely by theorem 2.8. By continuity we have $f(X_{n_k}) \to f(X)$ almost surely. Since f is bounded, the bounded convergence theorem implies that $\mathrm{E}\left[f(X_{n_k})\right] \to \mathrm{E}[f(X)]$, contradicting to our hypothesis. Hence $\mathrm{E}[f(X_n)] \to \mathrm{E}[f(X)]$ and $X_n \xrightarrow{d} X$. ∎

Theorem 2.26 (Skorokhod Representation)
Suppose $\mu_n \xrightarrow{d} \mu$ on $\mathbb{R}$. Then there are corresponding random variables X_n, X for μ_n and μ such that $X_n \sim \mu_n$, $X \sim \mu$ and $X_n \to X$ almost surely.

Proof. Let F_n and F be the CDFs for μ_n and μ, respectively. Take $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}$ and P be the Lebesgue measure on $[0,1]$. Put

$$X_n(\omega) = \sup\{x \in \mathbb{R} \mid F_n(x) < \omega\} \quad \text{and} \quad X(\omega) = \sup\{x \in \mathbb{R} \mid F(x) < \omega\}$$

with the convention that $\sup\varnothing = -\infty$. Then

$$\mathrm{P}\{X_n \le x\} = \mathrm{P}\{\omega \mid \omega \le F_n(x)\} = F_n(x) \quad \text{and} \quad \mathrm{P}\{X \le x\} = \mathrm{P}\{\omega \mid \omega \le F(x)\} = F(x).$$

It now suffices to show that $X_n \to X$ almost surely. Indeed, since F_n, F are CDFs, there are only at most countable discontinuities. Let ω be a point of continuity of X. We may find another continuity point such that $F(y) < \omega$. The convergence in distribution implies that $F_n(y) \to F(y)$. Hence for n large enough, we have $F_n(y) < \omega$ and hence $X_n(\omega) > y$. Thus

$$\liminf_{n\to\infty} X_n(\omega) \ge y$$

for all $y \le X(\omega)$. Thus

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

Similarly, pick a continuity point y such that $F(y) \geq \omega$ would give

$$\limsup_{n\to\infty} X_n(\omega) \leq y$$

for all $y \geq X(\omega)$ and thus

$$\limsup_{n\to\infty} X_n(\omega) \leq X(\omega).$$

Combining the above results gives that $X_n \to X$ almost surely, since X is continuous almost surely. ■

Corollary 2.27
Let $g \geq 0$ be a continuous measurable function on $\mathbb{R}$ and $X_n \xrightarrow{d} X$. Then

$$\mathrm{E}\left[g(X)\right] \leq \liminf_{n\to\infty} \mathrm{E}\left[g(X_n)\right].$$

Proof. Let Y_n and Y be the Skorokhod representations for X_n and X, respectively. Since now $g(Y_n) \to g(Y)$ almost surely, the Fatou's lemma shows that

$$\mathrm{E}\left[g(X)\right] = \mathrm{E}\left[g(Y)\right] \leq \liminf_{n\to\infty} \mathrm{E}\left[g(Y_n)\right] = \liminf_{n\to\infty} \mathrm{E}\left[g(X_n)\right].$$

■

Theorem 2.28 (Helly-Bray)
Suppose that X_n and X are $\mathbb{R}$-valued random variables. Then $X_n \xrightarrow{d} X$ if and only if

$$\mathrm{E}\left[g(X_n)\right] \to \mathrm{E}\left[g(X)\right]$$

for all $g \in C_b(\mathbb{R})$.

Proof. Assume first that $X_n \xrightarrow{d} X$. By the Skorokhod representation theorem, we may assume that X_n and X are defined on the same probability space and $X_n \to X$ almost surely. Now, since for all $g \in C_b(\mathbb{R})$, $g(X_n) \to g(X)$ almost surely and are uniformly bounded, the bounded convergence theorem implies that

$$\mathrm{E}\left[g(X_n)\right] \to \mathrm{E}\left[g(X)\right].$$

Conversely, suppose that $\mathrm{E}\left[g(X_n)\right] \to \mathrm{E}\left[g(X)\right]$ for all $g \in C_b(\mathbb{R})$. Let F_n and F be the distribution functions for X_n and X respectively and x be a continuity point of F. For $\epsilon > 0$, consider

$$g_\epsilon(y) = \begin{cases} 1 & y \leq x \\ 1 - \frac{y-x}{\epsilon} & x < y \leq x+\epsilon \\ 0 & y \geq x+\epsilon. \end{cases}$$

Clearly $g_\epsilon \in C_b(\mathbb{R})$. Let $g(y) = \mathbf{1}\{y \le x\}$.

$$\limsup_{n\to\infty} F_n(x) = \limsup_{n\to\infty} \mathrm{E}[g(X_n)] \le \limsup_{n\to\infty} \mathrm{E}[g_\epsilon(X_n)] = \mathrm{E}[g_\epsilon(X)] \le F(x+\epsilon).$$

Since ϵ is arbitrary and F is continuous at x, we have

$$\limsup_{n\to\infty} F_n(x) \le F(x).$$

On the other hand,

$$\liminf_{n\to\infty} F_n(x) = \liminf_{n\to\infty} \mathrm{E}[g(X_n)] \ge \liminf_{n\to\infty} \mathrm{E}[g_\epsilon(X_n+\epsilon)] = \mathrm{E}[g_\epsilon(X+\epsilon)] \ge F(x+\epsilon) \ge F(x).$$

Hence $F_n(x) \to F(x)$ and the proof is complete. ∎

Remark
The theorem gives an equivalent definition for the convergence in distribution in $\mathbb{R}$-valued case.

Theorem 2.29 (Continuous Mapping Theorem)
Let $X_n \xrightarrow{d} X$ in $\mathbb{R}$ and g be a measurable function continuous μ_X-almost surely. Then $g(X_n) \xrightarrow{d} g(X)$.

Proof. By the Skorokhod representation theorem, we may assume that X_n and X are on the same space and $X_n \to X$ almost surely. By the continuity of g, we have $g(X_n) \to g(X)$ almost surely. For all $f \in C_b(\mathbb{R})$, $f(g(X_n)) \to f(g(X))$ almost surely as well. Since $f \circ g$ is bounded, the bounded convergence theorem gives $\mathrm{E}[f(g(X_n))] \to \mathrm{E}[f(g(X))]$. By the Helly-Bray theorem, this implies that $g(X_n) \xrightarrow{d} g(X)$. ∎

Remark
If g is bounded, then $\mathrm{E}[g(X_n)] \to \mathrm{E}[g(X)]$ directly by applying bounded convergence theorem on $g(X_n)$ and $g(X)$.

Example
$X_n \sim U[-\frac{1}{n}, \frac{1}{n}] \xrightarrow{d} \delta_0$. Let $g(x) = \mathbf{1}\{x \ge 0\}$. Then $g(X_n) \sim Ber(\frac{1}{2}) \stackrel{d}{\not\to} g(X) \sim \delta_1$.

Definition 2.30
*A metric space (S, d) is **Polish** if it is complete and separable.*

Theorem 2.31 (Portmanteau)
Suppose that (S, d) is a Polish space. Let $X_n, X : \Omega \to S$ be random variables. Then the followings are equivalent:

(a) $X_n \xrightarrow{d} X$.

(b) *If $G \subset S$ is an open set, then $\liminf_{n\to\infty} \mathrm{P}(X_n \in G) \ge \mathrm{P}(X \in G)$.*

(c) *If $F \subset S$ is a closed set, then* $\limsup_{n\to\infty} \mathrm{P}(X_n \in F) \le \mathrm{P}(X \in F)$.

(d) *If $A \subset S$ satisfies* $\mathrm{P}(X \in \partial A) = 0$, *then* $\lim_{n\to\infty} \mathrm{P}(X_n \in A) = \mathrm{P}(X \in A)$.

Proof. We start from proving (a) implies (b). Let $G \subset S$ be an open set. For any continuous function with $0 \le f \le \mathbf{1}_G$, we have $\mathrm{E}[f(X_n)] \to \mathrm{E}[f(X)]$ and $\mathrm{E}[f(X_n)] \le \mathrm{P}(X_n \in G)$. Hence

$$\mathrm{E}[f(X)] \le \liminf_{n\to\infty} \mathrm{P}(X_n \in G).$$

Take $f \nearrow \mathbf{1}_G$ and by LMCT, since $f(X) \nearrow \mathbf{1}_G(X)$, $\mathrm{E}[\mathbf{1}_G(X)] \le \liminf_{n\to\infty} \mathrm{P}(X_n \in G)$.

To see that (b) and (c) are equivalent, note that if $\liminf_{n\to\infty} \mathrm{P}(X_n \in G) \ge \mathrm{P}(X \in G)$ for all open G, for every closed F,

$$\begin{aligned}\limsup_{n\to\infty} \mathrm{P}(X_n \in F) &= -\liminf_{n\to\infty} -\mathrm{P}(X_n \in F) = 1 - \liminf_{n\to\infty} \mathrm{P}(X_n \in F^c)\\ &\le 1 - \mathrm{P}(X \in F^c) = \mathrm{P}(X \in F).\end{aligned}$$

Conversely, if $\limsup_{n\to\infty} \mathrm{P}(X_n \in F) \le \mathrm{P}(X \in F)$ for every closed F, then

$$\liminf_{n\to\infty} \mathrm{P}(X_n \in G) = 1 - \limsup_{n\to\infty} \mathrm{P}(X_n \in G^c) \ge 1 - \mathrm{P}(X \in G^c) = \mathrm{P}(X \in G).$$

Now, assume that (b) and (c) holds. If $A \subset S$ satisfies that $\mathrm{P}(X \in \partial A) = 0$, then we have $\mathrm{P}(X \in \overline{A}) = \mathrm{P}(X \in A^\circ)$. Now,

$$\begin{aligned}\mathrm{P}(X \in A^\circ) &\le \liminf_{n\to\infty} \mathrm{P}(X_n \in A^\circ) \le \liminf_{n\to\infty} \mathrm{P}(X_n \in A)\\ &\le \limsup_{n\to\infty} \mathrm{P}(X_n \in A) \le \limsup_{n\to\infty} \mathrm{P}(X_n \in \overline{A}) \le \mathrm{P}(X \in \overline{A}) = \mathrm{P}(X \in A^\circ).\end{aligned}$$

Hence $\lim_{n\to\infty} \mathrm{P}(X_n \in A) = \mathrm{P}(X \in A)$.

Finally, assume (d) holds. We prove that (d) implies (c) and (b) implies (a). For any closed F, consider the closed ϵ-neighborhood

$$F_\epsilon = \{s \in S \mid d(F, s) \le \epsilon\}.$$

Now $\partial F_\epsilon = \{s \in S \mid d(F, s) = \epsilon\}$ are disjoint for distinct $\epsilon > 0$. Since X can accumulate mass on only countably many such sets, $\mathrm{P}(X \in \partial F_\epsilon) = 0$ for almost every ϵ. Picking $\epsilon_k \to 0$ such that $\mathrm{P}(X \in \partial F_{\epsilon_k}) = 0$, we have $\mathrm{P}(X_n \in F) \le \mathrm{P}(X_n \in F_\epsilon)$. Taking $n \to \infty$,

$$\limsup_{n\to\infty} \mathrm{P}(X_n \in F) \le \limsup_{n\to\infty} \mathrm{P}(X_n \in F_{\epsilon_k}) = \mathrm{P}(X \in F_{\epsilon_k}).$$

Now taking $\epsilon_k \to 0$,

$$\limsup_{n\to\infty} \mathrm{P}(X_n \in F) \le \mathrm{P}(X \in F).$$

To see that (b) implies (a), let $f \in C_b(S)$ with $f \geq 0$. By lemma 2.12, (b), and Fatou's lemma,

$$\begin{aligned} \mathrm{E}\left[f(X)\right] &= \int_0^\infty \mathrm{P}(f(X) > t)dt \leq \int_0^\infty \liminf_{n\to\infty} \mathrm{P}(f(X_n) > t)dt \\ &\leq \liminf_{n\to\infty} \int_0^\infty \mathrm{P}(f(X_n) > t)dt = \liminf_{n\to\infty} \mathrm{E}\left[f(X_n)\right]. \end{aligned}$$

Suppose $f \leq C$, replacing f with $C - f$ gives

$$C - \mathrm{E}\left[f(X)\right] \leq \liminf_{n\to\infty} C - \mathrm{E}\left[f(X_n)\right] = C - \limsup_{n\to\infty} \mathrm{E}\left[f(X_n)\right].$$

Hence $\mathrm{E}\left[f(X)\right] \geq \limsup_{n\to\infty} \mathrm{E}\left[f(X_n)\right]$. We see that $\mathrm{E}\left[f(X)\right] = \lim_{n\to\infty} \mathrm{E}\left[f(X_n)\right]$ for $f \geq 0$, $f \in C_b(S)$. For general f, write $f = f^+ - f^-$ and the conclusion follows by applying above arguments to f^+ and f^-. ■

Example
For a random variable X on $\mathbb{R}^d$, consider the perturbated version $X_n = X + \sigma_n Z$ with $\sigma_n \searrow 0$ and Z being independent of X, $\mu_Z \ll \lambda$ where λ is the Lebesgue measure. We claim that $X_n \xrightarrow{d} X$. For any open G and $\epsilon > 0$, define

$$G_{-\epsilon} = \left\{ x \in \mathbb{R}^d \,\middle|\, \overline{B_\epsilon(x)} \subset G \right\}.$$

Then

$$\mathrm{P}(X_n \in G) \geq \mathrm{P}(X \in G_{-\epsilon}, \|\sigma_n Z\| \leq \epsilon) = \mathrm{P}(X \in G_{-\epsilon})\,\mathrm{P}(\|\sigma_n Z\| \leq \epsilon).$$

Taking $n \to \infty$, $\liminf_{n\to\infty} \mathrm{P}(X_n \in G) \geq \mathrm{P}(X \in G_{-\epsilon})$. Then taking $\epsilon \to 0$ gives

$$\liminf_{n\to\infty} \mathrm{P}(X_n \in G) \geq \mathrm{P}(X \in G).$$

By the portmanteau theorem, $X_n \xrightarrow{d} X$.

Theorem 2.32 (Helly's Selection)
For every sequence of distribution F_n on $\mathbb{R}$, there is a subsequence $F_{n_k} \to F$ pointwise at continuities where F is a non-decreasing, right-continuous function.

Proof. Since $\mathbb{R}$ is separable, there is a countable dense subset D. For $x_1 \in D$, pick a subsequence F_{n_1} such that F_{n_1} converges at x_1. This is possible due to the Bolzano-Weierstrass theorem. Proceed with $x_2 \in D$ and extract subsequence from F_{n_1}. Continue this process and take the diagonal. We obtain a subsequence F_{n_k} such that for every $x_m \in D$, $F_{n_k}(x_m)$ converges. For general $x \in \mathbb{R}$, define

$$F(x) = \inf_{y \in D, y > x} \lim_{k\to\infty} F_{n_k}(y)$$

where $x_m \in D$ is such that $x_m \nearrow x$. It is clear that F is non-decreasing.

Next, for each $x \in \mathbb{R}$ and $\epsilon > 0$, there is $y \in D$ with $y > x$ such that

$$F(y) = \lim_{k\to\infty} F_{n_k}(y) \leq F(x) + \epsilon.$$

Hence F is right-continuous.

If x is a continuity of F, then for $\epsilon > 0$ we can also choose $z \in D$ with $z < x$ such that $F(z) \geq F(x) - \epsilon$. Now

$$F_{n_k}(z) \leq F_{n_k}(x) \leq F_{n_k}(y).$$

Taking $k \to \infty$,

$$F(x) - \epsilon \leq F(z) \leq \liminf_{k\to\infty} F_{n_k}(x) \leq \limsup_{k\to\infty} F_{n_k}(x) \leq F(y) \leq F(x) + \epsilon.$$

Since ϵ is arbitrary, we conclude that $F_{n_k}(x) \to F(x)$. Hence F is the desired function. ■

Remark
A sequence of distribution function F_n converges pointwise at continuities to F does not imply that F is a distribution function. One may consider $F_n(x) = \mathbf{1}\{x \geq n\}$. As $n \to \infty$, F_n converges pointwise at continuities to $F = 0$.

Definition 2.33
*A collection of probability measures $\{\mu_\alpha\}$ on S is **uniformly tight** or simply **tight** if for every $\epsilon > 0$, there is a compact set $K \subset S$ such that $\mu_\alpha(K^c) < \epsilon$ for every α.*

Remark
An alternative definition is that

$$\liminf_{n\to\infty} \mu_n(B_r) \to 1$$

as $r \to \infty$ for $\{\mu_n\}$ with μ_n being defined on normed space.

Definition 2.34
*A sequence of probability measures $\{\mu_n\}$ on $\mathbb{R}$ is said to converge **vaguely** if*

$$\int f d\mu_n \to \int f d\mu$$

for all $f \in C_c(\mathbb{R})$.

Remark
Since $C_c(\mathbb{R}) \subset C_b(\mathbb{R})$, converging weakly implies converging vaguely. Also, the μ in the definition of the vague convergence is not necessarily a probability distribution. This is exactly what we see in the remark of Helly's selecetion theroem.

Theorem 2.35
Let μ_n be probability measures on $\mathbb{R}^d$ such that $\mu_n \xrightarrow{v} \mu$. Then the followings are equivalent:

(a) $\{\mu_n\}$ *is tight.*

(b) μ *is a probability measure, i.e.,* $\mu(\mathbb{R}^d) = 1$.

(c) $\mu_n \xrightarrow{d} \mu$.

Proof. Assuming (c), then

$$\int f d\mu_n \to \int f d\mu$$

for all $f \in C_b(\mathbb{R}^d)$. Take $f = 1$ shows (b).

Suppose that (b) holds. For any open ball $B_r \subset \mathbb{R}^d$, consider $f \in C_c(\mathbb{R}^d)$ with $0 \le f \le \mathbf{1}_{B_r}$. For every n, we have

$$\int f d\mu_n \le \mu_n(B_r).$$

Also,

$$\int f d\mu_n \to \int f d\mu \le \mu(B_r)$$

by the vague convergence. Thus

$$\int f d\mu = \liminf_{n\to\infty} \int f d\mu \le \liminf_{n\to\infty} \mu_n(B_r).$$

Taking $f \nearrow \mathbf{1}_{B_r}$ gives $\mu(B_r) \le \liminf_{n\to\infty} \mu_n(B_r)$. Now taking $r \to \infty$ gives $\liminf_{n\to\infty} \mu_n(B_r) \to 1$ as $r \to \infty$ and hence $\{\mu_n\}$ is tight.

Suppose (a) is true. Fix any $f \in C_b(\mathbb{R}^d)$. For every open ball B_r, consider $g_r \in C_c(\mathbb{R}^d)$ with $f|_{B_r} \le g_r \le f$. Then

$$\left|\int f d\mu_n - \int f d\mu\right| \le \left|\int g_r d\mu_n - \int g_r d\mu\right| + \int |f - g_r|\, d\mu_n + \int |f - g_r|\, d\mu.$$

Since $\{\mu_n\}$ is tight, for every $\epsilon > 0$ there is some M such that $r > M$ implies $\mu_n(B_r^c) < \epsilon$ and

$$\left|\int f d\mu_n - \int f d\mu\right| \le \left|\int g_r d\mu_n - \int g_r d\mu\right| + \|f\|_\infty\, \epsilon + \int |f - g_r|\, d\mu.$$

Taking $n \to \infty$ yields

$$\limsup_{n\to\infty} \left|\int f d\mu_n - \int f d\mu\right| \le \|f\|_\infty\, \epsilon + \int |f - g_r|\, d\mu.$$

Taking $r \to \infty$ gives $g_r \to f$ and

$$\limsup_{n\to\infty} \left|\int f d\mu_n - \int f d\mu\right| \le \|f\|_\infty\, \epsilon.$$

Since ϵ is arbitrary, we conclude that $\int f d\mu_n \to \int f d\mu$ and $\mu_n \xrightarrow{d} \mu$. ■

Theorem 2.36 (Prokhorov)

Suppose μ_n is a sequence of probability measures on $\mathbb{R}$. $\{\mu_n\}$ is tight if and only if every subsequence of μ_n has a further subsequence converging weakly.

Proof. Suppose that $\{\mu_n\}$ is tight. By Helly's selection theorem, for every subsequence of μ_n we may extract a further subsequence converging vaguely to a measure μ. By the theorem 2.35, μ is a probability measure.

Conversely, suppose that $\{\mu_n\}$ is not tight. For every $\epsilon > 0$ and $k \in \mathbb{N}$, we can pick μ_{n_k} such that $\mu_{n_k}(B_k^c) \geq \epsilon$. For this sequence, we may extract a further subsequence $\mu_{n(k_j)}$ such that it converges weakly to μ and it follows from theorem 2.35 that $\left\{\mu_{n(k_j)}\right\}$ is tight. However $\mu_{n(k_j)}(B_{k_j}^c) \geq \epsilon$ for every k_j, posing a contradiction. Hence $\{\mu_n\}$ is tight. ∎

Proposition 2.37
Let X_n be a sequence of random variables. If there is $\phi(x) \geq 0$ such that $\phi(x) \to \infty$ as $\|x\| \to \infty$ and $\mathrm{E}\left[\phi(X_n)\right] \leq C$ for all n, then $\{X_n\}$ is tight.

Proof. Notice that

$$C \geq \mathrm{E}\left[\phi(X_n)\right] \geq \mathrm{E}\left[\phi(X_n)\mathbf{1}\left\{\|X_n\| \geq r\right\}\right] \geq \left(\inf_{\|x\|\geq r} \phi(x)\right) \mathrm{P}(\|X_n\| \geq r).$$

Hence

$$\mathrm{P}(\|X_n\| \geq r) \leq \frac{C}{\inf_{\|x\|\geq r} \phi(x)} \to 0$$

as $r \to \infty$. Thus $\{X_n\}$ tight. ∎

2.3. Characteristic Functions

Definition 2.38
Let X be an $\mathbb{R}$-valued random variable. The characteristic function of X is defined as

$$\varphi_X(t) = \mathrm{E}\left[e^{itX}\right] = \mathrm{E}\left[\cos(tX)\right] + i\,\mathrm{E}\left[\sin(tX)\right].$$

Remark
The characteristic function always exists for every $t \in \mathbb{R}$ since $x \mapsto \cos(tx)$ and $x \mapsto \sin(tx)$ are bounded functions.

Remark
*If X has distribution μ, we also write $\hat{\mu} = \varphi_X$. It is somtimes called the **Fourier transform** of the probability measure μ.*

Example
If $X \sim U[a, b]$, then

$$\varphi_X(t) = \mathrm{E}\left[\cos(tX)\right] + i\,\mathrm{E}\left[\sin(tX)\right] = \int_a^b \frac{\cos(tx)}{b-a}dx + i\int_a^b \frac{\sin(tx)}{b-a}dx = \frac{e^{ibt} - e^{iat}}{(b-a)it}.$$

Example
If

$$X = \begin{cases} n & \text{with prob. } \frac{1}{2} \\ -n & \text{with prob. } \frac{1}{2}, \end{cases}$$

then

$$\varphi_X(t) = \mathrm{E}\left[e^{itX}\right] = \frac{1}{2}e^{int} + \frac{1}{2}e^{-int} = \cos(nt).$$

Example
If $X \sim \text{Poisson}(\lambda)$, then

$$\varphi_X(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^{it}\lambda)^k}{k!} = \exp\left(\lambda\left(e^{it} - 1\right)\right).$$

Proposition 2.39
Let X be an $\mathbb{R}$-valued random variable. Then the followings are true:

(a) $\varphi_X(0) = 1$.

(b) $\varphi_X(-t) = \overline{\varphi_X(t)}$.

(c) $|\varphi_X(t)| \leq 1$.

(d) $\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at)$.

(e) If X and Y are independent, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

Proof. (a) is trivial. For (b), $\varphi_X(-t) = \mathrm{E}\left[e^{-itX}\right] = \mathrm{E}\left[\cos(tX)\right] - i\,\mathrm{E}\left[\sin(tX)\right] = \overline{\varphi_X(t)}$.

For (c), $|\varphi_X(t)| \leq \mathrm{E}\left[\left|e^{itX}\right|\right] \leq \mathrm{E}[1] = 1$.
For (d), $\varphi_{aX+b}(t) = \mathrm{E}\left[e^{iatX+ibt}\right] = e^{ibt}\,\mathrm{E}\left[e^{iatX}\right] = e^{ibt}\varphi_X(at)$.
For (e), $\varphi_{X+Y}(t) = \mathrm{E}\left[e^{itX+itY}\right] = \mathrm{E}\left[e^{itX}\right]\mathrm{E}\left[e^{itY}\right] = \varphi_X(t)\varphi_Y(t)$. ■

Remark

For (e), inductively we have that for independent variables X_n,

$$\varphi_{X_1+\cdots+X_n}(t) = \prod_{i=1}^{n} \varphi_{X_i}(t).$$

Example

$Z \sim N(0,1)$. *Then*

$$\begin{aligned}\varphi_Z(t) &= \mathrm{E}[\cos(tZ)] + i\,\mathrm{E}[\sin(tZ)]\\ &= \int \cos(tx)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx + i\int \sin(tx)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx\\ &= \int \cos(tx)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx.\end{aligned}$$

Now

$$\frac{\varphi_Z(t)-\varphi_Z(s)}{t-s} = \int \frac{\cos(tx)-\cos(sx)}{t-s}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx.$$

By the mean value theorem, $|\cos(tx)-\cos(sx)| \leq |\sin(c)|\,|tx-sx| \leq |x|\,|t-s|$ for some constant lying between tx and sx. Hence

$$\left|\frac{\cos(tx)-\cos(sx)}{t-s}\right|\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \leq |x|\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

which is integrable. It follows by LDCT that taking $t \to s$

$$\begin{aligned}\varphi_Z'(s) &= -\int \sin(sx)x\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx = \int \sin(sx)\,d\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\right)\\ &= -s\int \cos(sx)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx = -s\varphi_Z(s).\end{aligned}$$

Solving the ODE with initial condition $\varphi_Z(0) = 1$ gives

$$\varphi_Z(t) = \exp\left(-\frac{t^2}{2}\right).$$

In general, if $X \sim N(\mu,\sigma^2)$, $X = \mu + \sigma Z$ and

$$\varphi_X(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right).$$

Theorem 2.40 (Inversion Formula)
Let μ be a probability measure on $\mathbb{R}$ with characteristic function φ. Then

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)dt = \mu(a,b) + \frac{1}{2}\mu\{a,b\}.$$

Proof. Put

$$g(T,\lambda) = \int_{-T}^{T} \frac{\sin(\lambda t)}{t} dt = \int_{0}^{T} \frac{\sin(\lambda t)}{t} dt.$$

Note that $g(T,\lambda) \to 2\,\mathrm{sgn}(\lambda)\frac{\pi}{2} = \pi\,\mathrm{sgn}(\lambda)$. Now

$$\begin{aligned}
f(T) &= \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)dt \\
&= \frac{1}{2\pi} \int_{-T}^{T} \int_{a}^{b} e^{-ity} \varphi(t) dy dt \\
&= \frac{1}{2\pi} \int_{-T}^{T} \int_{a}^{b} \int e^{-it(y-x)} d\mu(x) dy dt \\
&= \frac{1}{2\pi} \int \int_{-T}^{T} \int_{a}^{b} e^{-it(y-x)} dy dt d\mu(x) \\
&= \frac{1}{2\pi} \int \int_{-T}^{T} \frac{\sin(t(b-x)) - \sin(t(a-x))}{t} dt d\mu(x) \\
&\quad + \frac{1}{2\pi} \int \int_{-T}^{T} \frac{[\cos(t(b-x)) - \cos(t(a-x))]}{t} dt d\mu(x),
\end{aligned}$$

where the fourth equality uses Fubini's theorem, which is valid since $[-T,T]\times[a,b]\times\mathbb{R}$ is of finite measure ($\mu(\mathbb{R}) = 1$) and $\left|e^{-it(y-x)}\right| \le 1$. Notice that

$$t \mapsto \frac{[\cos(t(b-x)) - \cos(t(a-x))]}{t}$$

is a bounded odd function. Hence

$$\begin{aligned}
f(T) &= \frac{1}{2\pi} \int \int_{-T}^{T} \frac{\sin(t(b-x)) - \sin(t(a-x))}{t} dt d\mu(x) \\
&= \frac{1}{2\pi} \int g(T, b-x) - g(T, a-x) d\mu(x).
\end{aligned}$$

Since $g(T,\lambda) \to \pi\,\mathrm{sgn}(\lambda)$,

$$g(T,b-x) - g(T,a-x) \to \begin{cases} 2\pi & \text{if } x \in (a,b) \\ \pi & \text{if } x \in \{a,b\} \\ 0 & \text{otherwise} \end{cases}$$

as $T \to \infty$. By the bounded convergence theorem,

$$f(T) \to \frac{1}{2\pi}(2\pi\mu(a,b) + \pi\mu\{a,b\}) = \mu(a,b) + \frac{1}{2}\mu\{a,b\}$$

as $T \to \infty$. ■

Remark

The theorem establishes that the characteristic function is unique, i.e., if $\varphi_X = \varphi_Y$, then $X \overset{d}{=} Y$.

Corollary 2.41

Let X be a random variable on $\mathbb{R}$ with characteristic function φ_X. Then X is symmetric if and only if φ_X is real.

Proof. Suppose that X is symmetric. Then $\varphi_X(t) = \varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$. Hence $\varphi_X(t)$ is real. Conversely, if φ_X is real, $\varphi_X(t) = \overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t)$. By the uniqueness of characteristic functions, $X \overset{d}{=} -X$ and X is symmetric. ■

Lemma 2.42

Let μ be a probability measure on $\mathbb{R}$ with characteristic function φ. Then for $r > 0$, we have

$$\mu\{x \mid |x| \geq r\} \leq \frac{r}{2}\int_{-2/r}^{2/r} 1 - \varphi(t)dt$$

Proof. By Fubini's theorem, since $(t,x) \mapsto \left|1 - e^{itx}\right|$ is integrable on $[-2/r, 2/r] \times \mathbb{R}$,

$$\begin{aligned}\int_{-2/r}^{2/r} 1 - \varphi(t)dt &= \int_{-2/r}^{2/r}\int 1 - e^{itx}d\mu(x)dt = \int\int_{-2/r}^{2/r} 1 - e^{itx}dtd\mu(x) \\ &= \frac{4}{r}\int 1 - \frac{\sin(2x/r)}{2x/r}d\mu(x) \geq \frac{2}{r}\mu\left\{x \,\middle|\, \left|\frac{2x}{r}\right| \geq 2\right\}\end{aligned}$$

since $\sin(y) \leq y/2$ for all $y \geq 2$. Rearrange the inequality

$$\mu\{x \mid |x| \geq r\} \leq \frac{r}{2}\int_{-2/r}^{2/r} 1 - \varphi(t)dt.$$

■

Proposition 2.43

Let $\{\mu_\alpha\}$ be a set of probability measures on $\mathbb{R}$ and $\{\varphi_\alpha\}$ be their characteristic function. Then $\{\mu_\alpha\}$ is tight if and only if $\{\varphi_\alpha\}$ is equicontinuous at zero.

Proof. Suppose that $\{\varphi_\alpha\}$ is equicontinuous at zero. Note that $\varphi_\alpha(0) = 1$ for all α. For every $\epsilon > 0$, there is $r > 0$ such that $1 - \varphi_\alpha(t) < \epsilon$ for all $t \in [-2/r, 2/r]$. Then lemma 2.42 gives

$$\mu_\alpha\{x \mid |x| \geq r\} \leq \frac{r}{2}\int_{-2/r}^{2/r} 1 - \varphi_\alpha(t)dt \leq \frac{r}{2} \cdot 2 \cdot \frac{2}{r}\epsilon = 2\epsilon.$$

Since ϵ is arbitrary, $\{\mu_\alpha\}$ is tight.

Conversely, assume that $\{\mu_\alpha\}$ is tight. Let $X_\alpha \sim \mu_\alpha$ be the corresponding random variables. For any $t \in \mathbb{R}$,

$$|\mu_\alpha(t) - \mu_\alpha(0)| = \left|\mathrm{E}\left[e^{itX_\alpha} - 1\right]\right| \leq \mathrm{E}\left[\left|1 - e^{itX_\alpha}\right|\right] \leq \mathrm{E}\left[\min\{2, |tX_\alpha|\}\right],$$

where the last inequality is due to $\left|1 - e^{itX_\alpha}\right| \leq 2$ and

$$\left|1 - e^{itX_\alpha}\right| = \left|-i\int_0^{tX_\alpha} e^{is}ds\right| \leq \int_0^{|tX_\alpha|} \left|e^{is}\right| ds = |tX_\alpha|.$$

Since $\{X_\alpha\}$ is tight, for every $\epsilon > 0$ there is $M > 0$ such that $\mathrm{P}(|X_\alpha| \geq M) < \epsilon$. Hence

$$|\mu_\alpha(t) - \mu_\alpha(0)| \leq \mathrm{E}\left[\min\{2, |tX_\alpha|\}\right] \leq 2\,\mathrm{P}(|X_\alpha| \geq M) + 2M\,|t| \leq 2(\epsilon + M\,|t|).$$

Take $|t| \to 0$ and since ϵ is arbitrary, the equicontinuity follows. ■

Theorem 2.44 (Levy's Continuity Theorem)
Suppose μ_n are random variables on $\mathbb{R}$ with characteristic function $\varphi_n(t)$. Then

(a) *If $\mu_n \xrightarrow{d} \mu$, then $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$ where φ is the characteristic function of μ.*

(b) *If $\varphi_n(t) \to \varphi(t)$ and $\varphi(t)$ is continuous at 0, then $\{\mu_n\}$ is tight and $\mu_n \xrightarrow{d} \mu$ where μ has characteristic function φ.*

Proof. For (a), suppose that $\mu_n \xrightarrow{d} \mu$. Then for every $f \in C_b(\mathbb{R})$,

$$\int f d\mu_n \to \int f d\mu.$$

In particular, taking $f : x \mapsto \cos(tx)$ and $f : x \mapsto \sin(tx)$ shows that

$$\varphi_n(t) = \int \cos(tx)d\mu_n + i\int \sin(tx)d\mu_n \to \int \cos(tx)d\mu + i\int \sin(tx)d\mu = \varphi(t).$$

For (b), by lemma 2.42 and bounded convergence theorem,

$$\limsup_{n\to\infty} \mu_n\{x \mid |x| \geq r\} \leq \lim_{n\to\infty} \frac{r}{2}\int_{-2/r}^{2/r} 1 - \varphi_n(t)dt = \frac{r}{2}\int_{-2/r}^{2/r} 1 - \varphi(t)dt.$$

Since $\varphi(t)$ is continuous at 0,

$$\frac{r}{2}\int_{-2/r}^{2/r} 1 - \varphi(t)dt \to 0$$

as $r \to \infty$. This implies that $\{\mu_n\}$ is tight.

Now, suppose that μ_n does not converge weakly. Then we can find $\epsilon \geq 0$, $f \in C_b$ and a subsequence μ_{n_k} such that

$$\left|\int f d\mu_{n_k} - \int f d\mu\right| \geq \epsilon$$

for every k. By Prokhorov's theorem, we can extract a further subsequence converging weakly, which is a contradiction. Hence $\mu_n \xrightarrow{d} \mu$. ■

Theorem 2.45

Suppose that X is a random variable on $\mathbb{R}$ with $\mathrm{E}\left[|X|^n\right] < \infty$ and has characteristic function $\varphi(t)$. Then $\varphi \in C^n(\mathbb{R})$ and

$$\varphi^{(n)}(t) = \mathrm{E}\left[(iX)^n e^{itX}\right].$$

Proof. We prove the result for $n = 1$. The other orders follow inductively. For $t \in \mathbb{R}$,

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \mathrm{E}\left[e^{itX}\frac{e^{ihX} - 1}{h}\right].$$

Note that

$$\left|e^{itx}\frac{e^{ihx} - 1}{h}\right| \leq \left|\frac{\cos(hx) - 1}{h}\right| + \left|\frac{\sin(hx)}{h}\right| \leq 2\,|x|\,.$$

Since $\mathrm{E}\left[2\,|X|\right] < \infty$, LDCT gives

$$\varphi'(t) = \lim_{h \to 0} \mathrm{E}\left[e^{itX}\frac{e^{ihX} - 1}{h}\right] = \mathrm{E}\left[e^{itX}(iX)\right].$$

■

Remark

The converse does not hold in general. We provide an example where $\varphi(x)$ is differentiable at 0 but $\mathrm{E}\left[|X|\right] = \infty$. Consider the random variable defined by $\mathrm{P}(X = \pm k) = \frac{c}{k^2 \log(k)}$ for $k \geq 2$ where c is some normalizing constant. Then

$$\mathrm{E}\left[|X|\right] = 2c \sum_{k=2}^{\infty} \frac{1}{k \log(k)}.$$

Since the integral

$$\int_2^{\infty} \frac{1}{x \log(x)} dx = \log(\log(x))|_2^{\infty} = \infty,$$

the first moment is not finite. However,

$$\varphi(t) = 2c \sum_{k=2}^{\infty} \frac{\cos(kt)}{k^2 \log(k)}, \quad \text{and} \quad \varphi'(0) = 2c \sum_{k=2}^{\infty} \frac{0}{k \log(k)} = 0.$$

Corollary 2.46 (Taylor Expansion)

Let φ be a characteristic function of a random variable X with $\mathrm{E}\left[|X|^n\right] < \infty$. Then as $t \to 0$,

$$\varphi(t) = \sum_{k=0}^{n} \frac{(it)^k \,\mathrm{E}\left[X^k\right]}{k!} + o(t^n).$$

Proof. By theorem 2.45, we can consider the Taylor expansion of φ up to order n at $t = 0$.

$$\varphi(t) = \sum_{k=0}^{n} \frac{\varphi^{(k)}(0)t^k}{k!} + o(t^n) = \sum_{k=0}^{n} \frac{\mathrm{E}\left[(iX)^k\right] t^k}{k!} + o(t^n) = \sum_{k=0}^{n} \frac{(it)^k \,\mathrm{E}\left[X^k\right]}{k!} + o(t^n)$$

as $t \to 0$. ∎

2.4. Central Limit Theorems

Theorem 2.47 (Central Limit Theorem I, Lindeberg-Levy)
Let X_i be independent and identically distributed random variables with $\mathrm{E}[X_i] = 0$ *and* $\mathrm{E}\left[X_i^2\right] = 1$*. Let* $X \sim N(0,1)$ *and* $S_n = \sum_{i \le n} X_i$*. Then*

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{d} X$$

as $n \to \infty$*.*

Proof. Notice that $\varphi_X(t) = \exp(-t^2/2)$ and by corollary 2.46,

$$\varphi_{\frac{1}{\sqrt{n}}S_n}(t) = \prod_{i=1}^{n} \varphi_{X_i}\left(\frac{t}{\sqrt{n}}\right) = \left(1 - \frac{t^2}{2n} + o(1/n)\right)^n \to \exp(-t^2/2)$$

as $n \to \infty$ for fixed t. Hence $\frac{1}{\sqrt{n}} S_n \xrightarrow{d} X$ by the Levy's continuity theorem. ∎

Theorem 2.48 (Central Limit Theorem II, Lindeberg-Feller)
For each n, let $X_{n,i}$, $1 \le i \le n$ *be random variables with mean* 0 *and* $S_n = \sum_{i \le n} X_{n,i}$*. Suppose*

(a) For given n, $X_{n,i}$ are independent.

(b) $\mathrm{Var}\left[\sum_{i=1}^n X_{n,i}\right] = \sum_{i=1}^n \mathrm{E}\left[X_{n,i}^2\right] \to \sigma^2$*. as* $n \to \infty$*.*

(c) For all $\epsilon > 0$, $\lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}\left[X_{n,i}^2 \mathbf{1}\left\{\left|X_{n,i}\right| > \epsilon\right\}\right] = 0$.

Then $S_n \xrightarrow{d} N(0, \sigma^2)$*.*

Proof. We begin the proof of the theorem by the following claim. For $t \in \mathbb{R}$, we have

$$\left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| \le \min\left\{ \frac{2|t|^n}{n!}, \frac{|t|^{n+1}}{(n+1)!} \right\}.$$

To see this, let $g_n(t)$ be the difference on the left hand side. For $n = 0$, the inequality holds trivially. Now suppose that the inequality holds for n.

$$g_{n+1}(t) = e^{it} - \sum_{k=0}^{n+1} \frac{(it)^k}{k!} = i \int_0^t e^{is} - \sum_{k=0}^{n} \frac{(is)^k}{k!} ds = i \int_0^t g_n(s) ds.$$

Now

$$|g_{n+1}(t)| \le \int_0^t \frac{2|s|^n}{n!} ds = \frac{2|t|^{n+1}}{(n+1)!} \quad \text{and} \quad |g_{n+1}(t)| \le \int_0^t \frac{|s|^{n+1}}{(n+1)!} ds = \frac{|t|^{n+2}}{(n+2)!}.$$

The claim follows by induction.

Now notice that for given n, if $Z_{n,i} \sim N(0, \sigma_{n,i}^2)$ are independent with $\sigma_{n,i}^2 = \mathrm{E}\left[X_{n,i}^2\right]$, then $\sum_{i \le n} Z_{n,i} \xrightarrow{d} Z \sim N(0, \sigma^2)$. Let the characteristic functions for $Z_{n,i}$ be $\psi_{n,i}$. Note that for complex numbers z_i and w_i with $|z_i|, |w_i| \le 1$, we have

$$\left|\prod_{i\le n} z_i - \prod_{i\le n} w_i\right| \le \sum_{i\le n} |z_i - w_i| .$$

If $n = 2$, $|z_1z_2 - w_1w_2| \le |z_1z_2 - z_1w_2| + |z_1w_2 - w_1w_2| \le |z_2 - w_2| + |z_1 - w_1|$. The general case follows by induction. It follows that if $\varphi_{n,i}$ are characteristic functions for $X_{n,i}$,

$$\begin{aligned}
\left|\prod_{i\le n} \varphi_{n,i}(t) - \prod_{i\le n} \psi_{n,i}(t)\right| &\le \sum_{i\le n} \left|\varphi_{n,i}(t) - \psi_{n,i}(t)\right| \\
&\le \sum_{i\le n} \left|\varphi_{n,i}(t) - 1 + \frac{\sigma_{n,i}^2 t^2}{2}\right| + \sum_{i\le n} \left|\psi_{n,i}(t) - 1 + \frac{\sigma_{n,i}^2 t^2}{2}\right| \\
&\le \sum_{i\le n} \mathrm{E}\left[\left|e^{itX_{n,i}} - 1 + \frac{X_{n,i}^2 t^2}{2}\right|\right] + \sum_{i\le n} \mathrm{E}\left[\left|e^{itZ_{n,i}} - 1 + \frac{Z_{n,i}^2 t^2}{2}\right|\right].
\end{aligned}$$

Now by the claim we have for all $\epsilon > 0$,

$$\begin{aligned}
\mathrm{E}\left[\left|e^{itX_{n,i}} - 1 + \frac{X_{n,i}^2 t^2}{2}\right|\right] &\le \mathrm{E}\left[\min\left\{\frac{2\left|tX_{n,i}\right|^2}{2!}, \frac{\left|tX_{n,i}\right|^3}{3!}\right\}\right] \\
&\le |t|^2\, \mathrm{E}\left[\left|X_{n,i}\right|^2 \mathbf{1}\left\{\left|X_{n,i}\right| > \epsilon\right\}\right] + \frac{|t|^3}{6}\mathrm{E}\left[\left|X_{n,i}\right|^3 \mathbf{1}\left\{\left|X_{n,i}\right| \le \epsilon\right\}\right] \\
&\le |t|^2\, \mathrm{E}\left[\left|X_{n,i}\right|^2 \mathbf{1}\left\{\left|X_{n,i}\right| > \epsilon\right\}\right] + \frac{\epsilon\,|t|^3}{6}\mathrm{E}\left[\left|X_{n,i}\right|^2 \mathbf{1}\left\{\left|X_{n,i}\right| \le \epsilon\right\}\right] \\
&\le |t|^2\, \mathrm{E}\left[\left|X_{n,i}\right|^2 \mathbf{1}\left\{\left|X_{n,i}\right| > \epsilon\right\}\right] + \frac{\epsilon\,|t|^3\,\sigma_{n,i}^2}{6}
\end{aligned}$$

Let $\sigma_n^2 = \sum_{i\le n} \sigma_{n,i}^2$. The claim applies to $Z_{n,i}$ gives

$$\begin{aligned}
\sum_{i\le n} \mathrm{E}\left[\left|e^{itZ_{n,i}} - 1 + \frac{Z_{n,i}^2 t^2}{2}\right|\right] &\le \frac{|t|^3}{6}\sum_{i\le n} \mathrm{E}\left[\left|Z_{n,i}\right|^3\right] = \frac{|t|^3}{6}\sum_{i\le n} \sigma_{n,i}^3\, \mathrm{E}\left[|Y|^3\right] \\
&\le \frac{|t|^3}{6}\mathrm{E}\left[|Y|^3\right]\sigma_n^2 \sup_{i\le n} \sigma_{n,i}
\end{aligned}$$

Hence

$$\begin{aligned}
\left|\prod_{i\le n} \varphi_{n,i}(t) - \prod_{i\le n} \psi_{n,i}(t)\right| &\le \sum_{i\le n} |t|^2\, \mathrm{E}\left[\left|X_{n,i}\right|^2 \mathbf{1}\left\{\left|X_{n,i}\right| > \epsilon\right\}\right] + \frac{\epsilon\,|t|^3\,\sigma_{n,i}^2}{6} + \frac{|t|^3}{6}\mathrm{E}\left[|Y|^3\right]\sigma_n^2 \sup_{i\le n}\sigma_{n,i} \\
&\to \frac{\epsilon\,|t|^3}{6}\sigma^2 + \frac{|t|^3}{6}\mathrm{E}\left[|Y|^3\right]\sigma^2\epsilon
\end{aligned}$$

as $n \to \infty$ since

$$\begin{aligned}\sup_{i\le n}\sigma_{n,i} = \left(\sup_{i\le n}\sigma_{n,i}^2\right)^{1/2} &= \left(\sup_{i\le n}\mathrm{E}\left[X_{n,i}^2\mathbf{1}\left\{\left|X_{n,i}\right|>\epsilon\right\}\right]+\mathrm{E}\left[X_{n,i}^2\mathbf{1}\left\{\left|X_{n,i}\right|\le\epsilon\right\}\right]\right)^{1/2}\\ &\le\left(\epsilon^2+\sum_{i\le n}\mathrm{E}\left[X_{n,i}^2\mathbf{1}\left\{\left|X_{n,i}\right|>\epsilon\right\}\right]\right)^{1/2} \to \epsilon\end{aligned}$$

as $n \to \infty$. Since ϵ is arbitrary, we conclude that $\left|\prod_{i\le n}\varphi_{n,i}(t)-\prod_{i\le n}\psi_{n,i}(t)\right| \to 0$ and $S_n \xrightarrow{d} N(0,\sigma^2)$. ■

Corollary 2.49 (Central Limit Theorem III, Lyapunov)
Let X_j be independent with $\mathrm{E}\left[X_j\right]=0$ and $\alpha_n^2=\sum_{j\le n}\mathrm{Var}(X_j)$. Suppose that there is $\delta>0$ such that $\mathrm{E}\left[\left|X_j\right|^{2+\delta}\right]<\infty$ and

$$\lim_{n\to\infty}\frac{1}{\alpha_n^{2+\delta}}\sum_{j\le n}\mathrm{E}\left[\left|X_j\right|^{2+\delta}\right]=0.$$

Then

$$\frac{1}{\alpha_n}\sum_{j\le n}X_j\xrightarrow{d}N(0,1).$$

Proof. Put $Y_{n,j}=\alpha_n^{-1}X_j$. Note that

$$\mathrm{Var}\left[\sum_{j\le n}Y_{n,j}\right]=\frac{1}{\alpha_n^2}\cdot\alpha_n^2=1$$

and for $\epsilon>0$,

$$\sum_{j\le n}\mathrm{E}\left[Y_{n,j}^2\mathbf{1}\left\{\left|Y_{n,j}\right|>\epsilon\right\}\right]=\frac{1}{\alpha_n^2}\sum_{j\le n}\mathrm{E}\left[X_{n,j}^2\mathbf{1}\left\{\left|X_{n,j}\right|^\delta>(\alpha_n\epsilon)^\delta\right\}\right]\le\frac{1}{\alpha_n^{2+\delta}\epsilon^\delta}\sum_{j\le n}\mathrm{E}\left[\left|X_{n,j}\right|^{2+\delta}\right]\to 0$$

as $n\to\infty$. Hence by the Lindeberg-Feller theorem, $\alpha_n^{-1}\sum_{j\le n}X_j\xrightarrow{d}N(0,1)$. ■

Example
Let $X_j \sim Ber(p_j)$ be independent random variables and $S_n=\sum_{j\le n}X_j$. Then

$$\frac{S_n-\mathrm{E}\left[S_n\right]}{\sqrt{\mathrm{Var}(S_n)}}\xrightarrow{d}N(0,1)$$

if $\mathrm{Var}(S_n)\to\infty$. *To see this, write $Y_j=X_j-p_j$ and $\sigma_j^2=\mathrm{Var}(Y_j)=\mathrm{E}\left[Y_j^2\right]=p_j(1-p_j)$. Let $\delta>0$ be given. Since $\left|Y_j\right|\le 1$, $\mathrm{E}\left[\left|Y_j\right|^{2+\delta}\right]\le\mathrm{E}\left[Y_j^2\right]=\sigma_j^2$ and*

$$\frac{1}{\mathrm{Var}(S_n)^{2+\delta}}\sum_{j\le n}\mathrm{E}\left[\left|Y_j\right|^{2+\delta}\right]\le\frac{1}{\mathrm{Var}(S_n)^{2+\delta}}\sum_{j\le n}\sigma_j^2=\mathrm{Var}(S_n)^{-\delta}\to 0.$$

By the Lyapunov theorem, we conclude that

$$\frac{S_n - \mathrm{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow{d} N(0,1).$$

Theorem 2.50

For given n, $X_{n,j} \sim Ber(p_{n,j})$ with $1 \le j \le n$ are independent. Let $S_n = \sum_{j\le n} X_{n,j}$. If

(a) $\sum_{j\le n} p_{n,j} \to \lambda \in (0,\infty)$

(b) $\max_{j\le n} p_{n,j} \to 0$,

then $S_n \xrightarrow{d} Poisson(\lambda)$.

Proof. Let $\varphi_{n,j}(t) = 1 + p_{n,j}(e^{it} - 1)$ be the characteristic functions for $X_{n,j}$. Then

$$\begin{aligned}\log(\varphi_{S_n}(t)) &= \sum_{j=1}^{n} \log(\varphi_{n,j}(t)) = \sum_{j=1}^{n} \log(1 + p_{n,j}(e^{it} - 1)) \\ &= \sum_{j=1}^{n} p_{n,j}(e^{it} - 1) + o(p_{n,j}) \to \lambda(e^{it} - 1),\end{aligned}$$

where we use the fact that $\log(1+x) = x + o(x)$ as $x \to 0$ and condition (b) implies that the Taylor expansion is valid. Hence $\varphi_{S_n}(t) \to \exp(\lambda(e^{it} - 1))$ and $S_n \xrightarrow{d} \mathrm{Poisson}(\lambda)$. ■

Example

The independence assumption of the random variables can sometimes be relaxed. Consider the random permutations of $\{1,\dots,n\}$. We are interested in finding the distribution of the number of fixed points for the random permutations. Define $X_{n,j} = \mathbf{1}\{\pi(j) = j\}$ where π is a random permutation and $S_n = \sum_{j\le n} X_{n,j}$. Put $A_{n,j} = \{X_{n,j} = 1\}$. Then

$$\mathrm{P}(S_n > 0) = \mathrm{P}(\cup_{j=1}^{n} A_{n,j}) = \sum_{j} \mathrm{P}(A_{n,j}) - \sum_{k\le j} \mathrm{P}(A_{n,j} \cap \mathrm{P}(A_{n,k})) + \cdots + (-1)^{n+1}\mathrm{P}(A_{n,1} \cap \cdots \cap A_{n,n}).$$

Note that

$$\mathrm{P}(A_{n,i_1} \cap \cdots \cap A_{n,i_r}) = \frac{(n-r)!}{n!}.$$

Hence

$$\mathrm{P}(S_n > 0) = n \cdot \frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} + \cdots + (-1)^{n+1}\frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1}\frac{1}{n!}.$$

Thus $\mathrm{P}(S_n = 0) = 1 - 1 + \frac{1}{2!} \cdots + (-1)^n \frac{1}{n!} \to e^{-1}$ as $n \to \infty$. Fix k entries. The remaining $n-k$ entries with no fixed points has combinatorics a_{n-k}. Thus

$$\mathrm{P}(S_n = k) = \binom{n}{k}\frac{a_{n-k}}{n!}, \quad \textit{and} \quad \mathrm{P}(S_{n-k} = 0) = \frac{a_{n-k}}{(n-k)!}.$$

So

$$\mathrm{P}(S_n = k) = \binom{n}{k} \frac{(n-k)!\,\mathrm{P}(S_{n-k} = 0)}{n!} \to \frac{1}{k!} e^{-1}$$

for given k*. We see that* $S_n \xrightarrow{d}$ *Poisson*(1).

3. Martingale

3.1. Conditional Expectation

Definition 3.1
Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $X \in \mathcal{L}^1(\Omega)$. The ***conditional expectation*** *of X with respect to a σ-algebra $\mathcal{G} \subset \mathcal{F}$ is a $\mathcal{G}$-measurable random variable Z such that*

$$\int_A Z d\,\mathrm{P} = \int_A X d\,\mathrm{P}$$

for every $A \in \mathcal{G}$.

Remark
Define

$$\nu(A) = \int_A X d\,\mathrm{P}\,.$$

It is clear that $\nu \ll \mathrm{P}$ — in particular on $\mathcal{G}$ — and hence the Radon-Nikodym theorem implies that there is a almost surely unique random variable $Z \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathrm{P})$ such that $\nu(A) = \int_A Z d\,\mathrm{P}$. Hence the conditional expectation is well-defined. We may denote the conditional expectation of X with respect to $\mathcal{G}$ as $\mathrm{E}[X|\mathcal{G}]$. For random variable Y, we write $\mathrm{E}[X|Y] = \mathrm{E}[X|\sigma(Y)]$.

Proposition 3.2
Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra. The followings are true:

(a) $\mathrm{E}[X|\mathcal{F}] = X$.

(b) If $\sigma(X)$ is independent of $\mathcal{G} \subset \mathcal{F}$, then $\mathrm{E}[X|\mathcal{G}] = \mathrm{E}[X]$.

(c) For $c \in \mathbb{R}$, $\mathrm{E}[cX + Y|\mathcal{G}] = c\,\mathrm{E}[X|\mathcal{G}] + \mathrm{E}[Y|\mathcal{G}]$.

(d) If $X \leq Y$ almost surely, then $\mathrm{E}[X|\mathcal{G}] \leq \mathrm{E}[Y|\mathcal{G}]$ almot surely.

(e) $\mathrm{E}[|\mathrm{E}[X|\mathcal{G}]|] \leq \mathrm{E}[|X|]$.

(f) If f is convex, $f(\mathrm{E}[X|\mathcal{G}]) \leq \mathrm{E}[f(X)|\mathcal{G}]$.

Proof. For (a), note that X is $\mathcal{F}$-measurable and for every $A \in \mathcal{F}$, $\mathrm{E}[X\mathbf{1}_A] = \mathrm{E}[X\mathbf{1}_A]$. Hence X satisfies the required condition for it to be the conditional expectation. $\mathrm{E}[X|\mathcal{F}] = X$ by the uniqueness of the conditional expectation.

For (b), since the constant is $\mathcal{G}$-measurable and $\mathrm{E}[X\mathbf{1}_A] = \mathrm{E}[X]\,\mathrm{E}[\mathbf{1}_A] = \mathrm{E}[\mathrm{E}[X]\,\mathbf{1}_A]$ for every $A \in \mathcal{G}$, we have $\mathrm{E}[X|\mathcal{G}] = \mathrm{E}[X]$.

For (c), it is clear that $c\,\mathrm{E}[X|\mathcal{G}] + \mathrm{E}[Y|\mathcal{G}]$ is measurable. Also, for every $A \in \mathcal{G}$,

$$\begin{aligned}\mathrm{E}[(cX+Y)\mathbf{1}_A] &= c\,\mathrm{E}[X\mathbf{1}_A] + \mathrm{E}[Y\mathbf{1}_A] = c\,\mathrm{E}[\mathrm{E}[X|\mathcal{G}]\,\mathbf{1}_A] + \mathrm{E}[\mathrm{E}[Y|\mathcal{G}]\,\mathbf{1}_A] \\ &= \mathrm{E}[(c\,\mathrm{E}[X|\mathcal{G}] + \mathrm{E}[Y|\mathcal{G}])\mathbf{1}_A]\,.\end{aligned}$$

Hence $\mathrm{E}[cX + Y|\mathcal{G}] = c\,\mathrm{E}[X|\mathcal{G}] + \mathrm{E}[Y|\mathcal{G}]$.

For (d), let $A = \{\mathrm{E}[Y|\mathcal{G}] - \mathrm{E}[X|\mathcal{G}] > 0\} = \{\mathrm{E}[Y - X|\mathcal{G}] > 0\}$. By the definition of conditional expectation, since $A \in \mathcal{G}$,

$$\mathrm{E}\left[(\mathrm{E}[Y|\mathcal{G}] - \mathrm{E}[X|\mathcal{G}])\mathbf{1}_A\right] = \mathrm{E}\left[\mathrm{E}[Y - X|\mathcal{G}]\,\mathbf{1}_A\right] = \mathrm{E}\left[(Y - X)\mathbf{1}_A\right] \geq 0$$

due to the assumption that $X \leq Y$ almost surely. If $\mathrm{P}(A) > 0$, we also have

$$\mathrm{E}\left[(\mathrm{E}[Y|\mathcal{G}] - \mathrm{E}[X|\mathcal{G}])\mathbf{1}_A\right] < 0,$$

which is a contradiction. Thus $\mathrm{P}(A) = 0$ and we conclude that $\mathrm{E}[X|\mathcal{G}] \leq \mathrm{E}[Y|\mathcal{G}]$ almost surely.

For (e), we may take $A = \{\mathrm{E}[X|\mathcal{G}] \geq 0\}$. Then

$$\mathrm{E}\left[|\mathrm{E}[X|\mathcal{G}]|\right] = \mathrm{E}\left[\mathrm{E}[X|\mathcal{G}]\,\mathbf{1}_A\right] - \mathrm{E}\left[\mathrm{E}[X|\mathcal{G}]\,\mathbf{1}_{A^c}\right] = \mathrm{E}\left[X\mathbf{1}_A\right] - \mathrm{E}\left[X\mathbf{1}_{A^c}\right] \leq \mathrm{E}\left[|X|\right],$$

proving (e).

For (f), note that by the convexity, there are a, b such that $f(x) \geq ax + b$ and $f(x_0) = x_0$ for some x_0. Taking $x_0 = \mathrm{E}[X|\mathcal{G}]$,

$$f(\mathrm{E}[X|\mathcal{G}]) = a\,\mathrm{E}[X|\mathcal{G}] + b = \mathrm{E}[aX + b|\mathcal{G}] \leq \mathrm{E}[f(X)|\mathcal{G}]$$

by (c) and (d). ■

Proposition 3.3

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, $X_n, X \in \mathcal{L}^1(\Omega)$ and $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. Then the followings are true:

(a) If $X_n \nearrow X$ and $X_n \geq 0$, then $\mathrm{E}[X_n|\mathcal{G}] \nearrow \mathrm{E}[X|\mathcal{G}]$ almost surely.

(b) If $X_n \to X$ almost surely and $|X_n| \leq Y \in \mathcal{L}^1(\Omega)$, then $\mathrm{E}[X_n|\mathcal{G}] \to \mathrm{E}[X|\mathcal{G}]$ almost surely.

(c) $\mathrm{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leq \liminf_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}]$ almost surely.

Proof. For (b), by assumption and LDCT we have $X_n \to X$ in $\mathcal{L}^1$. The proposition 3.2 (e) implies that $\mathrm{E}\,|\mathrm{E}[X_n - X|\mathcal{G}]| \leq \mathrm{E}[|X_n - X|] \to 0$. Hence (b) follows.

Now for (a), proposition 3.2 (d) implies that $\mathrm{E}[X_n|\mathcal{G}]$ is incresing. Since $X_n \leq |X| \in \mathcal{L}^1$, (b) shows that $\mathrm{E}[X_n|\mathcal{G}] \to \mathrm{E}[X|\mathcal{G}]$.

Again, by proposition 3.2 (d), $\mathrm{E}[\inf_{k\geq n} X_k|\mathcal{G}] \leq \mathrm{E}[X_m|\mathcal{G}]$ for every $m \geq n$. Thus

$$\mathrm{E}\left[\inf_{k\geq n} X_k|\mathcal{G}\right] \leq \inf_{m\geq n} \mathrm{E}[X_m|\mathcal{G}].$$

Since $\inf_{k\geq n} X_k \nearrow \liminf_{n\to\infty} X$, we may take $n\to\infty$ and the result from (a) gives

$$\mathrm{E}\left[\liminf_{n\to\infty} X_n|\mathcal{G}\right] \leq \liminf_{n\to\infty} \mathrm{E}\left[X_n|\mathcal{G}\right]$$

as desired. ∎

Proposition 3.4
If $\sigma(X)\subset\mathcal{G}$ and $\mathrm{E}\left[|Y|\right], \mathrm{E}\left[|XY|\right]<\infty$, then $\mathrm{E}\left[XY|\mathcal{G}\right]=X\,\mathrm{E}\left[Y|\mathcal{G}\right]$.

Proof. Suppose first that $X=\mathbf{1}_G$ where $G\in\mathcal{G}$. Then for any $A\in\mathcal{G}$,

$$\mathrm{E}\left[X\,\mathrm{E}\left[Y|\mathcal{G}\right]\mathbf{1}_A\right]=\mathrm{E}\left[\mathbf{1}_G\mathbf{1}_A\,\mathrm{E}\left[Y|\mathcal{G}\right]\right]=\mathrm{E}\left[\mathbf{1}_G Y\mathbf{1}_A\right]=\mathrm{E}\left[\mathrm{E}\left[\mathbf{1}_G Y|\mathcal{G}\right]\mathbf{1}_A\right]=\mathrm{E}\left[\mathrm{E}\left[XY|\mathcal{G}\right]\mathbf{1}_A\right]$$

since $G\cap A\in\mathcal{G}$. By the linearity this extends to the case where X is simple. For $X\geq 0$, there are simple functions $X_n\nearrow X$ almost surely and hence $X_nY\to XY$ with $|X_nY|\leq|XY|\in\mathcal{L}^1$. By proposition 3.3 (b), $\mathrm{E}\left[X_nY|\mathcal{G}\right]\to\mathrm{E}\left[XY|\mathcal{G}\right]$ almost surely. Also, $|\mathrm{E}\left[X_nY|\mathcal{G}\right]\mathbf{1}_A|\leq \mathrm{E}\left[|X_nY|\,|\mathcal{G}\right]\mathbf{1}_A\leq\mathrm{E}\left[|XY|\right]\mathbf{1}_A\in\mathcal{L}^1$.

$$\mathrm{E}\left[X\,\mathrm{E}\left[Y|\mathcal{G}\right]\mathbf{1}_A\right]=\lim_{n\to\infty}\mathrm{E}\left[X_n\,\mathrm{E}\left[Y|\mathcal{G}\right]\mathbf{1}_A\right]=\lim_{n\to\infty}\mathrm{E}\left[\mathrm{E}\left[X_nY|\mathcal{G}\right]\mathbf{1}_A\right]=\mathrm{E}\left[\mathrm{E}\left[XY|\mathcal{G}\right]\mathbf{1}_A\right]$$

where the first equality follows by $|X_n\,\mathrm{E}\left[Y|\mathcal{G}\right]\mathbf{1}_A|\leq|XY\mathbf{1}_A|\in\mathcal{L}^1$ and then applying proposition 3.2 (f) with $f(x)=|x|$, $X_n\,\mathrm{E}\left[Y|\mathcal{G}\right]\mathbf{1}_A\to X\,\mathrm{E}\left[Y|\mathcal{G}\right]\mathbf{1}_A$ and LDCT. The general case follows from the decomposition $X=X^+-X^-$ and $X_n=X_n^+-X_n^-$. ∎

Proposition 3.5 (Law of Iterated Expectation, Tower Property)
Let $(\Omega,\mathcal{F},\mathrm{P})$ be a probability space and $\mathcal{H}\subset\mathcal{G}\subset\mathcal{F}$ are σ-algebras. Then

(a) $\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{H}\right]|\mathcal{G}\right]=\mathrm{E}\left[X|\mathcal{H}\right]$.

(b) $\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{G}\right]|\mathcal{H}\right]=\mathrm{E}\left[X|\mathcal{H}\right]$.

Proof. For (a), since $\mathrm{E}\left[X|\mathcal{H}\right]$ is $\mathcal{G}$-measurable, proposition 3.4 implies that $\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{H}\right]|\mathcal{G}\right]=\mathrm{E}\left[X|\mathcal{H}\right]\mathrm{E}\left[1|\mathcal{G}\right]=\mathrm{E}\left[X|\mathcal{H}\right]$.

For (b), first not that both sides are $\mathcal{H}$-measurable. For $A\in\mathcal{H}\subset\mathcal{G}$,

$$\mathrm{E}\left[\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{G}\right]|\mathcal{H}\right]\mathbf{1}_A\right]=\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{G}\right]\mathbf{1}_A\right]=\mathrm{E}\left[X\mathbf{1}_A\right]=\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{H}\right]\mathbf{1}_A\right].$$

The conclusion follows. ∎

Corollary 3.6
For random variables $X,Y\in\mathcal{L}^1$, we have

$$\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{G}\right]Y\right]=\mathrm{E}\left[X\,\mathrm{E}\left[Y|\mathcal{G}\right]\right]=\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{G}\right]\mathrm{E}\left[Y|\mathcal{G}\right]\right].$$

Proof. By the tower property,

$$\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{G}\right]Y\right]=\mathrm{E}\left[\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{G}\right]Y|\mathcal{G}\right]\right]=\mathrm{E}\left[\mathrm{E}\left[X|\mathcal{G}\right]\mathrm{E}\left[Y|\mathcal{G}\right]\right].$$

Similarly,

$$\mathrm{E}\left[X\,\mathrm{E}\left[Y|\mathcal{G}\right]\right] = \mathrm{E}\left[\mathrm{E}\left[X\,\mathrm{E}\left[Y|\mathcal{G}\right]|\mathcal{G}\right]\right] = \mathrm{E}\left[\mathrm{E}\left[X|\mathcal{G}\right]\mathrm{E}\left[Y|\mathcal{G}\right]\right].$$

■

Remark

If we view $\mathrm{E}[XY] = \langle X, Y\rangle$, *then corollary 3.6 is effectively saying that the conditional expectation operator is self-adjoint.*

Proposition 3.7

Let $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space and* $\mathcal{G} \subset \mathcal{F}$ *be a* σ*-algebra. Consider the minimization problem*

$$\inf_{Z\in\mathcal{L}^2(\mathcal{G})} \mathrm{E}\left[(X-Z)^2\right]$$

where $X \in \mathcal{L}^2(\mathcal{F})$. *Then the minimum is attained in* $\mathcal{L}^2(\mathcal{G})$ *and the minimizer is given by* $Z = \mathrm{E}[X|\mathcal{G}]$.

Proof. First notice that $\mathcal{L}^2(\mathcal{G})$ is a closed convex subset of $\mathcal{L}^2(\mathcal{F})$. Hence the minimum is attained by a unique minimizer. Now rewrite the expression as

$$\mathrm{E}\left[(X-Z)^2\right] = \mathrm{E}\left[(X-\mathrm{E}[X|\mathcal{G}])^2\right] + \mathrm{E}\left[(Z-\mathrm{E}[X|\mathcal{G}])^2\right] - 2\,\mathrm{E}\left[(X-\mathrm{E}[X|\mathcal{G}])(Z-\mathrm{E}[X|\mathcal{G}])\right].$$

By the tower property, the third term becomes

$$\begin{aligned}\mathrm{E}\left[(X-\mathrm{E}[X|\mathcal{G}])(Z-\mathrm{E}[X|\mathcal{G}])\right] &= \mathrm{E}\left[\mathrm{E}\left[(X-\mathrm{E}[X|\mathcal{G}])(Z-\mathrm{E}[X|\mathcal{G}])|\mathcal{G}\right]\right] \\ &= \mathrm{E}\left[(\mathrm{E}[X|\mathcal{G}]-\mathrm{E}[X|\mathcal{G}])(Z-\mathrm{E}[X|\mathcal{G}])\right] = 0.\end{aligned}$$

To minimize the expression, it suffices to minimize the second term, which gives the solution $Z = \mathrm{E}[X|\mathcal{G}]$. ■

Remark

The proposition gives a characterization of conditional expectations. In fact, we can conversely view the conditional expectation as an projection operator $\mathrm{E}[\cdot|\mathcal{G}] : \mathcal{L}^2(\mathcal{F}) \to \mathcal{L}^2(\mathcal{G})$. *Since* $\mathcal{L}^2$ *is dense in* $\mathcal{L}^1$ *under* $\|\cdot\|_1$, *the projection operater extends to* $\mathrm{E}[\cdot|\mathcal{G}] : \mathcal{L}^1(\mathcal{F}) \to \mathcal{L}^1(\mathcal{G})$. *For* $X \in \mathcal{L}^1(\mathcal{F})$, *we may choose* $X_n \to X$ *in* $\mathcal{L}^1$ *where* $X_n \in \mathcal{L}^2$ *and define* $\mathrm{E}[X|\mathcal{G}] = \lim_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}]$.

Definition 3.8

Let $X, Y \in \mathcal{L}^2$ *be random variables. For any* σ*-algebra* $\mathcal{G} \subset \mathcal{F}$, *the* ***conditional covariance*** *is defined as*

$$\mathrm{Cov}[X, Y|\mathcal{G}] = \mathrm{E}\left[(X-\mathrm{E}[X|\mathcal{G}])(Y-\mathrm{E}[Y|\mathcal{G}])|\mathcal{G}\right] = \mathrm{E}[XY|\mathcal{G}] - \mathrm{E}[X|\mathcal{G}]\,\mathrm{E}[Y|\mathcal{G}].$$

The ***conditional variance*** *is defined as* $\mathrm{Var}[X|\mathcal{G}] = \mathrm{Cov}[X, X|\mathcal{G}] = \mathrm{E}\left[X^2|\mathcal{G}\right] - \mathrm{E}[X|\mathcal{G}]^2$.

Lemma 3.9 (Law of Total Variance)
Let $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space and* $\mathcal{G} \subset \mathcal{F}$ *be a* σ*-algebra. For any random variables* $X, Y \in \mathcal{L}^2$*,*

(a) $\mathrm{Var}[X] = \mathrm{E}[\mathrm{Var}[X|\mathcal{G}]] + \mathrm{Var}[\mathrm{E}[X|\mathcal{G}]]$.

(b) $\mathrm{Cov}[X,Y] = \mathrm{E}[\mathrm{Cov}[X,Y|\mathcal{G}]] + \mathrm{Cov}[\mathrm{E}[X|\mathcal{G}], \mathrm{E}[Y|\mathcal{G}]]$.

Proof. We prove only (b). (a) follows immediately by replacing Y with X. By the tower property,

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

■

3.2. Martingale

Definition 3.10
*Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. A **stochastic process** is a collection of random variables $X_t : \Omega \to (S, \mathcal{S})$ where $t \in T$. T is a totally ordered index set.*

Remark
*T is often taken to be $\mathbb{Z}_+$ or $\mathbb{R}_+$, which represents the time. If T is countable, we say that X_t is a **discrete time** process and **continuous time** if T is some uncountable subset of $\mathbb{R}$.*

Definition 3.11
*A **filtration** $\{\mathcal{F}_t\}_{t \in T}$ is a collection of σ-algebras such that $\mathcal{F}_t \subset \mathcal{F}_s$ for every $t < s$, $t, s \in T$.*

Definition 3.12
*Let X_t be a stochastic process and $\mathcal{F}_t$ be a filtration. We say that X_t is **adapted** to $\mathcal{F}_t$ or $\mathcal{F}_t$-**adapted** if $\sigma(X_t) \subset \mathcal{F}_t$ for all t.*

Remark
In many cases, the filtration is not mentioned since we may consider the natural filtration generated by X_t through the definition $\mathcal{F}_t = \sigma(\{X_s | s \le t\})$.

Definition 3.13
*Let X_t be a $\mathbb{R}$-valued stochastic process. We say that X_t is an $\mathcal{F}_t$-**martingale** if*

(a) $\mathrm{E}[|X_t|] < \infty$.

(b) *X_t is $\mathcal{F}_t$-adapted.*

(c) *$X_t = \mathrm{E}[X_s | \mathcal{F}_t]$ for all $s > t$.*

*We say that X_t is a **supermartingale** if $X_t \ge \mathrm{E}[X_s|\mathcal{F}_t]$ and **submartingale** if $X_t \le \mathrm{E}[X_s|\mathcal{F}_t]$ for all $s > t$.*

Remark
X_t is a martingale if and only if X_t is both a submartingale and a supermartingale.

Proposition 3.14
Let X_n be a discrete time stochastic process.

(a) *If $X_n \ge \mathrm{E}[X_{n+1}|\mathcal{F}_n]$ for all n, then X_n is a supermartingale.*

(b) *If $X_n \le \mathrm{E}[X_{n+1}|\mathcal{F}_n]$ for all n, then X_n is a submartingale.*

Proof. The proof for (b) is similar to (a). We prove only the case (a). Let $m > n$.

$$\mathrm{E}[X_m|\mathcal{F}_n] = \mathrm{E}[\mathrm{E}[X_m|\mathcal{F}_{m-1}]|\mathcal{F}_n] \ge \mathrm{E}[X_{m-1}|\mathcal{F}_n] = \cdots \ge \mathrm{E}[X_n|\mathcal{F}_n] = X_n.$$

Hence X_n is a supermartingale. ■

Example (Random Walk)
Let $\xi_i \in \mathcal{L}^1$ be independent and identically distributed with $\mathrm{E}[\xi_i] = 0$. *Set* $X_0 = 0$ *and* $X_n = X_{n-1} + \xi_n$ *for* $n \in \mathbb{N}$. *Clearly* X_n *is adapted to the natural filtration* $\mathcal{F}_n$ *generated by* X_n *and* $\mathrm{E}[|X_n|] \le n\,\mathrm{E}[|\xi_i|] < \infty$ *for given* n. *Also,*

$$\mathrm{E}[X_{n+1}|\mathcal{F}_n] = \mathrm{E}[X_n + \xi_{n+1}|\mathcal{F}_n] = X_n.$$

Hence X_n *is a martingale.*

Example (Quadratic Martingale)
Let ξ_i *be independent and identically distributed with* $\mathrm{E}[\xi_i] = 0$ *and* $\mathrm{E}\left[\xi_i^2\right] = \sigma^2 < \infty$. *Put* $X_n = \sum_{i \le n} \xi_i$. *Then* $M_n = X_n^2 - n\sigma^2$ *is a martingale. It is clear that* M_n *is adapted to the natural filtration* $\mathcal{F}_n$ *generated by* X_n *and*

$$\mathrm{E}[|M_n|] \le \mathrm{E}\left[X_n^2\right] + n\sigma^2 = 2n\sigma^2 < \infty.$$

Also,

$$\begin{aligned}\mathrm{E}[M_{n+1}|\mathcal{F}_n] &= \mathrm{E}\left[X_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n\right] = \mathrm{E}\left[X_n^2 + 2\xi_{n+1}X_n + \xi_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n\right]\\ &= X_n^2 + \sigma^2 - (n+1)\sigma^2 = M_n.\end{aligned}$$

Hence M_n *is a martingale.*

Example (Exponential Martingale)
Let ξ_i *be independent and identically distributed with* $M(t) = \mathrm{E}[\exp(t\xi_i)] < \infty$ *for given* t. *Put* $X_i = \frac{\exp(t\xi_i)}{M(t)}$ *and thus* $\mathrm{E}[X_i] = 1$. *Let* $M_n = \prod_{i \le n} X_i$. M_n *is adapted to the natural filtration* $\mathcal{F}_n$ *generated by* X_n *and*

$$\mathrm{E}[|M_n|] = \mathrm{E}\left[\prod_{i \le n} X_i\right] = \prod_{i \le n} \mathrm{E}[X_i] = 1.$$

Also,

$$\mathrm{E}[M_{n+1}|\mathcal{F}_n] = M_n\,\mathrm{E}[X_{n+1}|\mathcal{F}_n] = M_n\,\mathrm{E}[X_{n+1}] = M_n.$$

Hence M_n *is a martingale.*

Lemma 3.15
Let X_t *be a* $\mathcal{F}_t$*-martingale and* φ *is a function such that* $\mathrm{E}[|\varphi(X_t)|] < \infty$.

(a) If φ *is convex, then* $\varphi(X_t)$ *is a submartingale.*

(b) If φ *is concave, then* $\varphi(X_t)$ *is a supermartingale*

Proof. The proof for (b) is similar to (a). We only prove the case (a). Let $s > t$.

$$\mathrm{E}[\varphi(X_s)|\mathcal{F}_t] \ge \varphi(\mathrm{E}[X_s|\mathcal{F}_t]) = \varphi(X_t).$$

Hence $\varphi(X_t)$ is a submartingale. ■

Lemma 3.16
Let X_t be a stochastic process and φ is an increasing function such that $\mathrm{E}\left[|\varphi(X_t)|\right] < \infty$.

(a) If X_t is a $\mathcal{F}_t$-submartingale and φ is convex, then $\varphi(X_t)$ is a $\mathcal{F}_t$-submartingale.

(b) If X_t is a $\mathcal{F}_t$-supermartingale and φ is concave, then $\varphi(X_t)$ is a $\mathcal{F}_t$-supermartingale.

Proof. For (a), let $s > t$.

$$\mathrm{E}\left[\varphi(X_s)|\mathcal{F}_t\right] \geq \varphi(\mathrm{E}\left[X_s|\mathcal{F}_t\right]) \geq \varphi(X_t).$$

Hence $\varphi(X_t)$ is a submartingale.

For (b),

$$\mathrm{E}\left[\varphi(X_s)|\mathcal{F}_t\right] \leq \varphi(\mathrm{E}\left[X_s|\mathcal{F}_t\right]) \leq \varphi(X_t).$$

Hence $\varphi(X_t)$ is a supermartingale. ■

Definition 3.17
Let X_n be a stochastic process and $\mathcal{F}_n$ be the filtration generated by X_n. A stochastic process H_n is said to be ***predictable*** *if H_{n+1} is $\mathcal{F}_n$-adapted.*

Definition 3.18
Let X_n, Y_n be two discrete-time stochastic processes. The ***discrete-time stochastic integral*** *is defined as*

$$(X \cdot Y)_n = \sum_{i \leq n} X_i (Y_i - Y_{i-1}).$$

Theorem 3.19
If X_n is a $\mathcal{F}_n$-supermartingale and $H_n \geq 0$ is predictable and is bounded for each n. Then $(H \cdot X)_n$ is a supermartingale.

Proof. First, it is clear that $(H \cdot X)_n$ is $\mathcal{F}_n$-adapted since the discrete-time stochastic integral is a function of $(H_1, \ldots H_n, X_0, \ldots, X_n)$. Second, since H_n is bounded, say by c_i,

$$\mathrm{E}\left[|(H \cdot X)_n|\right] \leq \sum_{i \leq n} \mathrm{E}\left[|H_i|\,|X_i - X_{i-1}|\right] \leq \sum_{i \leq n} c_i \max_{1 \leq i \leq n} 2\,\mathrm{E}\left[|X_i|\right] < \infty$$

for given n since X_n is a supermartingale and satisfies that $X_n \in \mathcal{L}^1$. Finally,

$$\begin{aligned}
\mathrm{E}\left[(H \cdot X)_{n+1}|\mathcal{F}_n\right] &= (H \cdot X)_n + \mathrm{E}\left[H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n\right] \\
&= (H \cdot X)_n + H_{n+1}\,\mathrm{E}\left[X_{n+1}|\mathcal{F}_n\right] - H_{n+1}X_n \\
&\leq (H \cdot X)_n + H_{n+1}X_n - H_{n+1}X_n = (H \cdot X)_n.
\end{aligned}$$

Hence $(H \cdot X)_n$ is a supermartingale. ■

Remark
If we instead assume that X_n is a $\mathcal{F}_n$-submartingale, then $(H \cdot X)_n$ is a submartingale. Furthermore, if X_n is a $\mathcal{F}_n$-martingale, then for every predictable H_n, $(H \cdot X)_n$ is a martingale.

Definition 3.20
A random variable $\tau \in T$ is a ***stopping time*** *if $\{\tau \le t\} \in \mathcal{F}_t$.*

Example (Hitting Time)
Let X_t be a stochastic process and $\mathcal{F}_t$ be the natural filtration generated by X_t. The hitting time of A is $\tau = \inf\{s \in T \mid X_s \in A\}$. Then τ is a stopping time.

Proposition 3.21
Suppose that τ_1 and τ_2 are stopping times with respect to $\mathcal{F}_t$. Then $\tau_1 \wedge \tau_2$ and $\tau_1 \vee \tau_2$ are stopping times.

Proof. Notice that

$$\{\tau_1 \wedge \tau_2 \le t\} = \{\tau_1 \le t\} \cup \{\tau_2 \le t\} \in \mathcal{F}_t, \quad \text{and} \quad \{\tau_1 \vee \tau_2 \le t\} = \{\tau_1 \le t\} \cap \{\tau_2 \le t\} \in \mathcal{F}_t.$$

Hence $\tau_1 \wedge \tau_2$ and $\tau_1 \vee \tau_2$ are stopping times. ■

Remark
Since any constant n is also a stopping time, $n \wedge \tau$ is a stopping time provided that τ is. The conclusion from theorem 3.19 also applies to the strategy of the form $H'_n = H_{n\wedge\tau}$. In particular, taking the predictable process

$$H_n = \begin{cases} 1 & \tau \le n \\ 0 & \tau > n \end{cases}$$

shows that $X_{n\wedge\tau}$ is a martingale provided that X_n is.

Definition 3.22
Let X_t be a stochastic process with index set T and $a < b$ be constants. Given ω, the ***$[a,b]$-crossing time*** *$N_a^b(\omega)$ is defined as the supremum over $n \in \mathbb{Z}_+$ such that there is $s_1 < t_1 < s_2 < t_2 < \cdots < s_n < t_n$, $s_i, t_i \in T$ with $X_{s_1}(\omega) \le a$ and $X_{t_1}(\omega) \ge b$.*

Lemma 3.23 (Doob's Upcrossing Lemma)
Let X_n be a submartingale on $\mathbb{Z}_+$ and $N_a^b(n)$ be the crossing time up to $n \in \mathbb{Z}_+$. Then

$$\mathrm{E}\left[N_a^b(n)\right] \le \frac{1}{b-a}\,\mathrm{E}\left[(X_n - a)^+\right]$$

for all $n \in \mathbb{Z}_+$.

Proof. Let $Y_n = a + (X_n - a)^+$. By lemma 3.16, Y_n is also a submartingale since $x \mapsto a + (x-a)^+$ is a convex increasing function. Notice that the crossing time up to time n for Y_n is also $N_a^b(n)$. Now define the stopping times $0 = \tau_0 \le \sigma_1 < \tau_1 < \cdots$ by

$$\sigma_k = \inf\{n \ge \tau_{k-1} \mid Y_n = a\}, \quad \text{and} \quad \tau_k = \inf\{n \ge \sigma_k \mid Y_n \ge b\}.$$

Next, we consider the predictable process

$$H_n = \sum_{k \geq 1} \mathbf{1} \{\sigma_k < n \leq \tau_k\}.$$

By theorem 3.19, $((1 - H) \cdot Y)_n$ is a submartingale and thus

$$\mathrm{E}\left[((1 - H) \cdot Y)_n\right] \geq \mathrm{E}\left[((1 - H) \cdot Y)_0\right] = 0.$$

Notice that $(H \cdot Y)_n \geq (b - a) N_a^b(n)$. We have

$$\mathrm{E}\left[(X_n - a)^+\right] = \mathrm{E}\left[(H \cdot Y)_n\right] + \mathrm{E}\left[((1 - H) \cdot Y)_n\right] \geq (b - a)\, \mathrm{E}\left[N_a^b(n)\right].$$

Rearranging the inequality gives the desired result. ∎

Theorem 3.24 (Martingale Convergence Theorem I)
Let X_n be a $\mathcal{F}_n$-submartingale. Suppose that $\sup_n \mathrm{E}\left[X_n^+\right] < \infty$. Then there is $X \in \mathcal{L}^1$ such that $X_n \to X$ almost surely.

Proof. For $a, b \in \mathbb{Q}$, consider the event $A_{a,b} = \{\liminf_n X_n < a < b < \limsup_n X_n\}$. Put S be the union of all $A_{a,b}$ over $a, b \in \mathbb{Q}$. For given $a < b$, consider the crossing time N_a^b on $\mathbb{Z}_+$ and $N_a^b(n)$ which is up to time n. Clearly $N_a^b(n) \nearrow N_a^b$. By LMCT and Doob's upcrossing lemma,

$$\mathrm{E}\left[N_a^b\right] = \lim_{n \to \infty} \mathrm{E}\left[N_a^b(n)\right] \leq \limsup_{n \to \infty} \frac{1}{b - a} \mathrm{E}\left[(X_n - a)^+\right] \leq C \sup_n \mathrm{E}\left[X_n^+\right]$$

for some constant C. Hence $\mathrm{P}(N_a^b < \infty) = 0$. This implies that X_n crosses $[a, b]$ only finite times almost surely. Thus $\mathrm{P}(A_{a,b}) = 0$ and $\mathrm{P}(S) = 0$. We see that

$$\mathrm{P}(\liminf_n X_n < \limsup_n X_n) = 0.$$

Thus the limit of X_n exists, possibly in extended sense, almost surely. Denote the limit of X_n as X. It now remains to show that $X \in \mathcal{L}^1$.

$$\mathrm{E}\left[X^+\right] = \mathrm{E}\left[\lim_{n \to \infty} X_n^+\right] \leq \liminf_{n \to \infty} \mathrm{E}\left[X_n^+\right] \leq \sup_n \mathrm{E}\left[X_n^+\right] < \infty$$

by Fatou's lemma. Also, $X_n^- = X_n^+ - X_n$.

$$\mathrm{E}\left[X_n^-\right] = \mathrm{E}\left[X_n^+\right] - \mathrm{E}\left[X_n\right] \leq \mathrm{E}\left[X_n^+\right] - \mathrm{E}\left[X_0\right] < \infty.$$

Hence $X \in \mathcal{L}^1$. The proof is complete. ∎

Remark
The theorem also holds for supermartingale. If X_n is a $\mathcal{F}_n$-supermartingale and $\sup_n \mathrm{E}\left[X_n^-\right] < \infty$. Then there is $X \in \mathcal{L}^1$ such that $X_n \to X$ almost surely.

Corollary 3.25
If $X_n \geq 0$ is a supermartingale, then there is $X \in \mathcal{L}^1$ such that $X_n \to X$ almost surely and $\mathrm{E}[X] \leq \mathrm{E}[X_0]$.

Proof. Since $\sup_n \mathrm{E}\left[X_n^-\right] = 0$, the martingale convergence theorem implies that there is $X \in \mathcal{L}^1$ such that $X_n \to X$ almost surely. By Fatou's lemma,

$$\mathrm{E}[X] = \mathrm{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathrm{E}[X_n] \leq \liminf_{n\to\infty} \mathrm{E}[X_0] = \mathrm{E}[X_0].$$

The proof is complete. ∎

Remark
The hard bound on X_n is neccessary. Consider ξ_i with $\mathrm{P}(\xi_i = \pm 1) = 1/2$. Let $X_n = \sum_{i\leq n} \xi_i$ and $X_0 = 1$. Consider the stopping time $\tau = \inf\{n \mid X_n = 0\}$. Then $\mathrm{P}(\tau < \infty) = 1$. Since $X_{n\wedge\tau}$ is a martingale, $\mathrm{E}[X_{n\wedge\tau}] = X_0 = 1$. Also, the martingale convergence theorem shows that there is X where $X_{n\wedge\tau} \to X$ as $n \to \infty$. Since $\tau < \infty$ almost surely, we have $X = 0$ almost surely and $\mathrm{E}[X] = 0$.

Example (Pólya's Urn)
Let $X_0, Y_0 \in \mathbb{N}$. Define

$$(X_n, Y_n) = \begin{cases} (X_{n-1}+1, Y_{n-1}), & \textit{prob.} = \frac{X_{n-1}}{X_{n-1}+Y_{n-1}}, \\ (X_{n-1}, Y_{n-1}+1), & \textit{prob.} = \frac{Y_{n-1}}{X_{n-1}+Y_{n-1}}. \end{cases}$$

Let $Z_n = X_n/(X_n+Y_n)$. Notice that

$$\begin{aligned} \mathrm{E}[Z_{n+1}|\mathcal{F}_n] &= Z_n \cdot \frac{X_n+1}{X_n+1+Y_n} + (1-Z_n)\cdot\frac{X_n}{X_n+Y_n+1} = \frac{X_n}{X_n+Y_n}\cdot\frac{1}{X_n+1+Y_n} + \frac{X_n}{X_n+Y_n+1} \\ &= \frac{X_n}{X_n+Y_n} = Z_n. \end{aligned}$$

Hence Z_n is a martingale. Since Z_n is clearly bounded, the martingale convergence theorem implies that there is some $Z \in \mathcal{L}^1$ such that $Z_n \to Z$ almost surely. In fact, one can show that $Z \sim Beta(X_0, Y_0)$.

Theorem 3.26 (Optional Stopping Theorem)
Let τ be a stopping time and X_n be a martingale. Then $\mathrm{E}[|X_\tau|] < \infty$ and $\mathrm{E}[X_\tau] = \mathrm{E}[X_0]$ if one of the followings holds:

(a) There is some N such that $\tau \leq N$ almost surely.

(b) There is some M such that $|X_{\tau\wedge n}| \leq M$ and $\mathrm{P}(\tau < \infty) = 1$.

(c) $\mathrm{E}[\tau] < \infty$ and $|X_{n+1} - X_n| \leq M$ almost surely for some M.

Proof. For (a), we have that

$$|X_{n\wedge\tau} - X_0| \le \sum_{k=1}^{n\wedge\tau} |X_k - X_{k-1}| \le \sum_{k=1}^{N} |X_k - X_{k-1}| \in \mathcal{L}^1$$

for every n. Hence by the LDCT, since $\tau < \infty$ almost surely and $X_{n\wedge\tau} \to X_\tau$,

$$\mathrm{E}\left[X_\tau\right] = \mathrm{E}\left[\lim_{n\to\infty} X_{n\wedge\tau}\right] = \lim_{n\to\infty} \mathrm{E}\left[X_{n\wedge\tau}\right] = \mathrm{E}\left[X_0\right]$$

by the fact that $X_{n\wedge\tau}$ is a martingale by theorem 3.19.

For (b), by the bounded convergence theorem and that $\tau < \infty$ almost surely,

$$\mathrm{E}\left[X_\tau\right] = \mathrm{E}\left[\lim_{n\to\infty} X_{n\wedge\tau}\right] = \lim_{n\to\infty} \mathrm{E}\left[X_{n\wedge\tau}\right] = \mathrm{E}\left[X_0\right].$$

For (c), notice that

$$|X_{n\wedge\tau} - X_0| \le \sum_{k=1}^{n\wedge\tau} |X_k - X_{k-1}| \le M\tau \in \mathcal{L}^1$$

since $\tau \in \mathcal{L}^1$. Thus $|X_{n\wedge\tau}| \le M\tau \in \mathcal{L}^1$. LDCT implies that

$$\mathrm{E}\left[X_\tau\right] = \mathrm{E}\left[\lim_{n\to\infty} X_{n\wedge\tau}\right] = \lim_{n\to\infty} \mathrm{E}\left[X_{n\wedge\tau}\right] = \mathrm{E}\left[X_0\right].$$

■

Corollary 3.27 (Wald's Identity)
Let X_i be independent and identically distributed with $\mathrm{E}\left[X_i\right] = \mu < \infty$. $|X_i| \le K$ almost surely. Let $S_n = \sum_{i\le n} X_i$, $S_0 = 0$ and τ be a stopping time with $\mathrm{E}\left[\tau\right] < \infty$. Then $\mathrm{E}\left[S_\tau\right] = \mu\,\mathrm{E}\left[\tau\right]$.

Proof. Consider $M_n = S_n - n\mu$. Notice that

$$\mathrm{E}\left[M_{n+1} | \mathcal{F}_n\right] = \mathrm{E}\left[X_{n+1} - \mu | \mathcal{F}_n\right] + S_n - n\mu = M_n.$$

Thus M_n is a martingale. Also,

$$|M_{n+1} - M_n| = |X_{n+1} - \mu| \le K + |\mu|$$

almost surely. By the optional stopping theorem (c),

$$\mathrm{E}\left[S_\tau - \tau\mu\right] = \mathrm{E}\left[M_\tau\right] = \mathrm{E}\left[M_0\right] = 0.$$

Hence $\mathrm{E}\left[S_\tau\right] = \mu\,\mathrm{E}\left[\tau\right]$. ■

Example
Consider a random walk $X_n = \sum_{i\le n} \xi_i$ where $\mathrm{P}(\xi_i = \pm 1) = 1/2$ are independent and $X_0 = x \in$

$\mathbb{Z}$. Define the hitting time $\tau = \inf\{k \mid X_k = a, b\}$ where $a < x < b$, $a, b \in \mathbb{Z}$. We first show that

$$\mathrm{P}(X_\tau = a) = \frac{b-x}{b-a}.$$

Since X_n is a martingale, $X_{n\wedge\tau}$ is again a martingale by theorem 3.19. Thus

$$\begin{aligned} x = \mathrm{E}\left[X_{n\wedge\tau}\right] &= \mathrm{E}\left[X_\tau \mathbf{1}\{n \geq \tau\}\right] + \mathrm{E}\left[X_n \mathbf{1}\{n < \tau\}\right] \\ &= a\,\mathrm{P}(X_\tau = a, \tau \leq n) + b\,\mathrm{P}(X_\tau = b, \tau \leq n) + \mathrm{E}\left[X_n \mathbf{1}\{n < \tau\}\right]. \end{aligned}$$

Notice that $|X_n|\,\mathbf{1}\{n < \tau\} \leq |a| \vee |b|$ and $|X_n|\,\mathbf{1}\{n < \tau\} \to 0$ since $\tau < \infty$ almost surely. Letting $n \to \infty$ gives

$$x = a\,\mathrm{P}(X_\tau = a) + b\,\mathrm{P}(X_\tau = b) = a\,\mathrm{P}(X_\tau = a) + b(1 - \mathrm{P}(X_\tau = a)), \quad \Rightarrow \quad \mathrm{P}(X_\tau = a) = \frac{b-x}{b-a}.$$

Next, we show that

$$\mathrm{E}\left[\tau\right] = (b-x)(x-a).$$

Consider $M_n = X_n^2 - n$. Then clearly the integrability and measurability holds. Also,

$$\mathrm{E}\left[M_{n+1}|\mathcal{F}_n\right] = \mathrm{E}\left[X_n^2 + 2X_n\xi_{n+1} + \xi_{n+1}^2 - (n+1)|\mathcal{F}_n\right] = M_n.$$

Hence M_n is a martingale. Again, by theorem 3.19, $M_{n\wedge\tau}$ is still a martingale.

$$\begin{aligned} x^2 = \mathrm{E}\left[M_{n\wedge\tau}\right] &= \mathrm{E}\left[X_\tau^2 \mathbf{1}\{\tau \leq n\}\right] + \mathrm{E}\left[X_n^2 \mathbf{1}\{n < \tau\}\right] - \mathrm{E}\left[n \wedge \tau\right] \\ &= a^2\,\mathrm{P}(X_\tau = a, \tau \leq n) + b^2\,\mathrm{P}(X_\tau = b, \tau \leq n) + \mathrm{E}\left[X_n^2 \mathbf{1}\{n < \tau\}\right] - \mathrm{E}\left[n \wedge \tau\right]. \end{aligned}$$

Notice that $\tau < \infty$ almost surely and thus $n \wedge \tau \nearrow \tau$. Also, $X_n^2 \mathbf{1}\{n < \tau\} < (a^2 \vee b^2)$ and $X_n^2 \mathbf{1}\{n < \tau\} \to 0$ as $n \to \infty$. By the LMCT and LDCT, taking $n \to \infty$,

$$x^2 = a^2\,\mathrm{P}(X_\tau = a) + b^2\,\mathrm{P}(X_\tau = b) - \mathrm{E}\left[\tau\right] = a^2 \frac{b-x}{b-a} + b^2 \frac{x-a}{b-a} - \mathrm{E}\left[\tau\right].$$

Hence $\mathrm{E}\left[\tau\right] = (b-x)(x-a)$.

Theorem 3.28 (Doob's Decomposition)
Let X_n be an integrable and $\mathcal{F}_n$-adapted process. Then there is an almost surely unique decomposition $X_n = M_n + A_n$ where M_n is an $\mathcal{F}_n$-martingale and A_n is $\mathcal{F}_n$-predictable with $A_0 = 0$. In particular, if X_n is a submartingale, then A_n is non-decreasing almost surely.

Proof. Put $A_n = \sum_{k \leq n} \mathrm{E}\left[X_k|\mathcal{F}_{k-1}\right] - X_{k-1}$ with $A_0 = 0$ and $M_n = X_n - A_n$. Notice that A_n is $\mathcal{F}_{n-1}$ measurable and hence $\mathcal{F}_n$ predictable. Also, M_n is integrable and $\mathcal{F}_n$-adapted. We verify the martingale condition:

$$\begin{aligned} \mathrm{E}\left[X_n|\mathcal{F}_{n-1}\right] = \mathrm{E}\left[M_n|\mathcal{F}_{n-1}\right] + A_n &= \mathrm{E}\left[M_n|\mathcal{F}_{n-1}\right] + A_{n-1} + \mathrm{E}\left[X_n|\mathcal{F}_{n-1}\right] - X_{n-1} \\ &= \mathrm{E}\left[M_n|\mathcal{F}_{n-1}\right] - M_{n-1} + \mathrm{E}\left[X_n|\mathcal{F}_{n-1}\right]. \end{aligned}$$

Rearranging the equation shows that $M_{n-1} = \mathrm{E}[M_n|\mathcal{F}_{n-1}]$. Thus M_n is indeed a martingale.

To see the uniqueness, suppose that $X_n = M'_n + A'_n$ is another decomposition. Then

$$\mathrm{E}[X_n|\mathcal{F}_{n-1}] = M_{n-1} + A_n = X_{n-1} + A_n - A_{n-1} = X_{n-1} + A'_n - A'_{n-1}.$$

Since $A_0 = A'_0 = 0$ and $A_n - A_{n-1} = A'_n - A'_{n-1}$ for all n, it follows that A_n are unique and also the M_n.

Now suppose that X_n is a submartingale,

$$A_{n-1} + M_{n-1} = X_{n-1} \le \mathrm{E}[X_n|\mathcal{F}_{n-1}] = M_{n-1} + A_n.$$

Hence $A_{n-1} \le A_n$ almost surely for all n. ■

Example

Let X_n be independent with $\mathrm{E}[X_i] = 0$ and $\sigma_i^2 = \mathrm{E}\left[X_i^2\right] < \infty$. Let $S_n = \sum_{i\le n} X_i$ be a martingale. Then S_n^2 is a submartingale. Consider the Doob's decomposition $S_n^2 = M_n + A_n$. We claim that $M_n = S_n^2 - \sum_{i\le n}\sigma_i^2$ and $A_n = \sum_{i\le n}\sigma_i^2$ is the Doob's decomposition for S_n^2. First, the integrability and adaptedness of M_n are clear.

$$\mathrm{E}[M_{n+1}|\mathcal{F}_n] = S_n^2 + 2S_n\,\mathrm{E}[X_{n+1}|\mathcal{F}_n] + \mathrm{E}\left[X_{n+1}^2|\mathcal{F}_n\right] - \sum_{i\le n+1}\sigma_i^2 = S_n^2 - \sum_{i\le n}\sigma_i^2 = M_n.$$

Thus M_n is a martingale. Also, A_n is increasing almost surely and predictable. We conclude that M_n and A_n are the desired decomposition.

Theorem 3.29

X_n is a martingale with $|X_{n+1} - X_n| \le M$. Let

$$C = \left\{\lim_{n\to\infty} X_n \text{ exists and is finite}\right\}, \quad D = \left\{\liminf_{n\to\infty} X_n = -\infty, \text{ and } \limsup_{n\to\infty} X_n = \infty\right\}.$$

Then $\mathrm{P}(C \cup D) = 1$.

Proof. Without loss of generality, we assume that $X_0 = 0$. For fixed $k > 0$, define the stopping time $\tau = \inf\{n \mid X_n \le k\}$. Since the increments of X_n are bounded, we have $X_{n\wedge\tau} \ge -k - M$. Since $X_{n\wedge\tau}$ is a martingale, by the martingale convergence theorem $X_{n\wedge\tau} \to X$ for some $X \in \mathcal{L}^1$ almost surely. On $\{\tau = \infty\}$, $\lim_{n\to\infty} X_n = \lim_{n\to\infty} X_{n\wedge\tau} = X$ is finite. Now taking $k \to \infty$, then on the event $\{\liminf_{n\to\infty} X_n > -\infty\}$, $\lim_{n\to\infty} X_n$ exists and is finite almost surely. The same argument for $-X_n$ shows that on $\{\limsup_{n\to\infty} < \infty\}$, $\lim_{n\to\infty} X_n$ exists and is finite. Combining the results,

$$D^c = \left\{\liminf_{n\to\infty} X_n > -\infty\right\} \cup \left\{\limsup_{n\to\infty} < \infty\right\} \subset C$$

with some exception of measure zero. Hence $\mathrm{P}(C \cup D) = 1$. ■

Theorem 3.30 (Borel-Cantelli III)
Let $\mathcal{F}_n$ be a filtration with $\mathcal{F}_0 = \{\varnothing, \Omega\}$ and $A_n \in \mathcal{F}_n$. Then

$$\mathrm{P}\left\{A_n \text{ } i.o.\right\} = \mathrm{P}\left\{\sum_n \mathrm{P}(A_n|\mathcal{F}_{n-1}) = \infty\right\}.$$

Proof. Let $X_0 = 0$ and $X_n = \sum_{k\leq n} \mathbf{1}_{A_k} - \mathrm{P}(A_k|\mathcal{F}_{k-1})$. Notice that X_n is bounded and $\mathcal{F}_n$-adapted. Also,

$$\mathrm{E}\left[X_{n+1}|\mathcal{F}_n\right] = X_n + \mathrm{E}\left[\mathbf{1}_{A_{n+1}}|\mathcal{F}_n\right] - \mathrm{P}(A_{n+1}|\mathcal{F}_n) = X_n.$$

Thus X_n is a martingale with bounded increments. By theorem 3.29,

$$C = \left\{\lim_{n\to\infty} X_n \text{ exists and is finite}\right\}, \quad \text{and} \quad D = \left\{\liminf_{n\to\infty} X_n = -\infty, \text{ and } \limsup_{n\to\infty} X_n = \infty\right\}$$

satisfy that $\mathrm{P}(C \cup D) = 1$. Notice that $\left\{\sum_n \mathbf{1}_{A_n} = \infty\right\} = \{A_n \text{ i.o.}\}$. On C, $\sum_n \mathbf{1}_{A_n} = \infty$ if and only if $\sum_n \mathrm{P}(A_n|\mathcal{F}_{n-1}) = \infty$. On D, we have $\sum_n \mathbf{1}_{A_n} = \infty$ and $\sum_n \mathrm{P}(A_n|\mathcal{F}_{n-1}) = \infty$. Since $\mathrm{P}(C \cup D) = 1$, we conclude that $\{A_n \text{ i.o.}\} = \{\sum_n \mathrm{P}(A_n|\mathcal{F}_{n-1}) = \infty\}$ up to some measure zero sets and the result follows. ∎

Lemma 3.31
Let X_n be a submartingale and τ be a stopping time with $\tau \leq T$. Then

$$\mathrm{E}\left[X_0\right] \leq \mathrm{E}\left[X_\tau\right] \leq \mathrm{E}\left[X_T\right].$$

Proof. Note that $X_{n\wedge\tau}$ is again a submartingale. Then

$$\mathrm{E}\left[X_0\right] = \mathrm{E}\left[X_{0\wedge\tau}\right] \leq \mathrm{E}\left[\mathrm{E}\left[X_{T\wedge\tau}|\mathcal{F}_0\right]\right] = \mathrm{E}\left[X_\tau\right].$$

Also, for fixed $0 \leq k \leq T$,

$$\mathrm{E}\left[X_\tau \mathbf{1}\left\{\tau = k\right\}\right] = \mathrm{E}\left[X_k \mathbf{1}\left\{\tau = k\right\}\right] \leq \mathrm{E}\left[\mathrm{E}\left[X_T|\mathcal{F}_k\right]\mathbf{1}\left\{\tau = k\right\}\right] = \mathrm{E}\left[X_T \mathbf{1}\left\{\tau = k\right\}\right].$$

Summing over $0 \leq k \leq T$ gives $\mathrm{E}\left[X_\tau\right] \leq \mathrm{E}\left[X_T\right]$. ∎

Theorem 3.32 (Doob's Submatingale Inequality)
Let X_n be a submartingale and $\lambda > 0$. Define $X_n^* = \max_{0\leq k\leq n} X_k^+$. Then

$$\mathrm{P}\left\{X_n^* \geq \lambda\right\} \leq \frac{1}{\lambda}\mathrm{E}\left[X_n \mathbf{1}\left\{X_n^* \geq \lambda\right\}\right] \leq \frac{1}{\lambda}\mathrm{E}\left[X_n^+\right].$$

Proof. Let $A = \left\{X_n^* \geq \lambda\right\}$ and $\tau = \inf\left\{k \mid X_k \geq \lambda\right\} \wedge n$. On A, $X_\tau \geq \lambda$ and hence

$$\mathrm{P}\left\{X_n^* \geq \lambda\right\} = \mathrm{E}\left[\mathbf{1}_A\right] \leq \frac{1}{\lambda}\mathrm{E}\left[X_\tau \mathbf{1}_A\right].$$

Notice that by lemma 3.31, $\mathrm{E}\left[X_\tau\right] \leq \mathrm{E}\left[X_n\right]$. Also, $X_\tau \mathbf{1}_{A^c} = X_n \mathbf{1}\tau_{A^c}$. Thus $\mathrm{E}\left[X_\tau \mathbf{1}_A\right] \leq$

$\mathrm{E}[X_n \mathbf{1}_A]$. The first inequality holds. The second inequality is trivial since $X_n \mathbf{1}_A \le X_n^+ \mathbf{1}_A \le X_n^+$. The proof is complete. ■

Corollary 3.33
Let M_n be a martingale and $\mathrm{E}[|M_n|^p] < \infty$ for some $p \ge 1$. For any $\lambda > 0$,

$$\mathrm{P}\left\{\max_{0\le k\le n} |M_k| \ge \lambda\right\} \le \frac{1}{\lambda^p} \mathrm{E}[|M_n|^p].$$

Proof. For $p \ge 1$, $x \mapsto |x|^p$ is convex and thus $|M_n|^p$ is a submartingale. By the Doob's submartingale inequality,

$$\mathrm{P}\left\{\max_{0\le k\le n} |M_k| \ge \lambda\right\} = \mathrm{P}\left\{\max_{0\le k\le n} |M_k|^p \ge \lambda^p\right\} \le \frac{1}{\lambda^p} \mathrm{E}[|M_k|^p].$$

■

Theorem 3.34 ($\mathcal{L}^p$ Maximal Inequality)
Let X_n be a submartingale and $X_n^* = \max_{0\le k\le n} X_k^+$. For $1 < p < \infty$, we have

$$\mathrm{E}\left[|X_n^*|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathrm{E}\left[X_n^+\right]^p.$$

Furthermore, if X_n is a martingale, then

$$\mathrm{E}\left[\left(\max_{0\le k\le n} |X_k|\right)^p\right] \le \left(\frac{p}{p-1}\right)^p \mathrm{E}[|X_n|^p].$$

Proof. Suppose first that $X_n^* \in \mathcal{L}^p$. By lemma 2.12, Doob's submartingale inequality, Fubini theorem and Hölder's inequality,

$$\begin{aligned}
\mathrm{E}\left[(X_n^*)^p\right] &= \int_0^\infty p\lambda^{p-1} \mathrm{P}(X_n^* \ge \lambda) d\lambda \le \int_0^\infty p\lambda^{p-1} \frac{1}{\lambda} \mathrm{E}\left[X_n \mathbf{1}\left\{X_n^* \ge \lambda\right\}\right] d\lambda \\
&\le \int_0^\infty p\lambda^{p-2} \mathrm{E}\left[X_n^+ \mathbf{1}\left\{X_n^* \ge \lambda\right\}\right] d\lambda = p\,\mathrm{E}\left[X_n^+ \int_0^{X_n^*} \lambda^{p-2} d\lambda\right] \\
&= \frac{p}{p-1} \mathrm{E}\left[X_n^+ (X_n^*)^{p-1}\right] \le \frac{p}{p-1} \mathrm{E}\left[|X_n^+|^p\right]^{1/p} \mathrm{E}\left[|X_n^*|^p\right]^{(p-1)/p}.
\end{aligned}$$

Rearranging the inequality gives

$$\mathrm{E}\left[|X_n^*|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathrm{E}\left[(X_n^+)^p\right].$$

For general X_n, consider $Y_n \in \mathcal{L}^p$ such that $Y_n \nearrow X_n$. Since $Y_n^* \in \mathcal{L}^p$ as well, the result extends to X_n by the monotone convergence theorem. For the second part, apply the previous result to the submartingale $|X_n|$. Then

$$\mathrm{E}\left[\left(\max_{0\le k\le n} X_k\right)^p\right] \le \left(\frac{p}{p-1}\right)^p \mathrm{E}\left[|X_n|^p\right].$$

This finishes the proof. ■

Example

Let S_n be a symmetric simple random walk on $\mathbb{Z}$ with $S_0 = 1$. Let $\tau = \inf\{k \mid S_k = 0\}$ be the hitting time. Put $X_n = S_{n\wedge\tau}$. Then

$$\mathrm{P}\left\{\sup_{k\ge 0} X_k \ge M\right\} = \frac{1}{M}.$$

To see this, note that by the martingale convergence theorem, X_k converges to some $X \in \mathcal{L}^1$ almost surely since X_k is bounded. By the optional stopping theorem, for the hitting time $N = \inf\{k \mid X_k = M \text{ or } X_k = 0\}$,

$$1 = \mathrm{E}[S_0] = \mathrm{E}[X_0] = \mathrm{E}[X_N] = M\,\mathrm{P}(X_N = M) = M\,\mathrm{P}\left\{\sup_{k\ge 0} X_k \ge M\right\}.$$

Rearranging the equality shows the desired equation. Now, by the LMCT,

$$\lim_{n\to\infty} \mathrm{E}\left[\max_{k\le n} |X_k|\right] \ge \lim_{n\to\infty} \mathrm{E}\left[\max_{k\le n} X_k\right] = \mathrm{E}\left[\sup_{k\ge 0} X_k\right] = \sum_{M\ge 1} \mathrm{P}\left\{\sup_{k\ge 0} X_k \ge M\right\} = \sum_{M\ge 1} \frac{1}{M} = \infty.$$

However, $\mathrm{E}[X_n] = \mathrm{E}[S_{n\wedge\tau}] = \mathrm{E}[S_0] = 1$. *We see that the $\mathcal{L}^p$ inequality fails for $p = 1$.*

Theorem 3.35 (Martingale Convergence Theorem II)
Let X_n be a martingale and $\sup_n \mathrm{E}[|X_n|^p] < \infty$ for some $p > 1$. Then $X_n \to X$ almost surely and in $\mathcal{L}^p$.

Proof. First, note that $|X_n|^p$ is a submartingale. The almost sure convergence follows immediately from the martingale convergence theorem I. Next, since $|X_n - X|^p \le 2\sup_n |X_n|^p$ and by LMCT and the $\mathcal{L}^p$ martingale inequality,

$$\mathrm{E}\left[\sup_n |X_n|^p\right] = \lim_{n\to\infty} \mathrm{E}\left[\max_{0\le k\le n} |X_k|^p\right] \le \lim_{n\to\infty}\left(\frac{p}{p-1}\right)^p \mathrm{E}[|X_n|^p] \le \left(\frac{p}{p-1}\right)^p \sup_n \mathrm{E}[|X_n|^p] < \infty.$$

We see that $2\sup_n |X_n|^p$ is integrable and hence by LDCT, $X_n \to X$ in $\mathcal{L}^p$. ■

3.3. Uniform Integrability

Definition 3.36
A collection of random variable $\{X_i\}_{i\in\mathcal{I}}$ is uniformly integrable if

$$\lim_{M\to\infty}\sup_{i\in\mathcal{I}} \mathrm{E}\left[|X_i|\,\mathbf{1}\{|X_i|>M\}\right]=0.$$

Remark
The uniform integrability implies that $\sup_i \mathrm{E}[|X_i|]<\infty$. To see this, note that we can find M such that

$$\sup_i \mathrm{E}\left[|X_i|\,\mathbf{1}\{|X_i|>M\}\right]<\epsilon.$$

Then

$$\sup_i \mathrm{E}[|X_i|]=\sup_i \mathrm{E}\left[|X_i|\,\mathbf{1}\{|X_i|>M\}\right]+\mathrm{E}\left[|X_i|\,\mathbf{1}\{|X_i|\le M\}\right]\le \epsilon+M<\infty.$$

Proposition 3.37
If $\{X_i\}$ and $\{Y_i\}$ are two collections of random variables, then $\{X_i+Y_i\}$ is also uniformly integrable.

Proof. Note that

$$\begin{aligned}|X_i+Y_i|\,\mathbf{1}\{|X_i+Y_i|>M\}&\le 2\max\{|X_i|,|Y_i|\}\,\mathbf{1}\{2\max\{|X_i|,|Y_i|\}>M\}\\&\le 2(|X_i|\,\mathbf{1}\{|X_i|>M/2\}+|Y_i|\,\mathbf{1}\{|Y_i|>M/2\}).\end{aligned}$$

Then

$$\begin{aligned}&\lim_{M\to\infty}\sup_i \mathrm{E}\left[|X_i+Y_i|\,\mathbf{1}\{|X_i+Y_i|>M\}\right]\\&\quad\le \lim_{M\to\infty}2\left(\sup_i \mathrm{E}\left[|X_i|\,\mathbf{1}\{|X_i|>M/2\}\right]+\sup_i \mathrm{E}\left[|Y_i|\,\mathbf{1}\{|Y_i|>M/2\}\right]\right)=0.\end{aligned}$$

■

Proposition 3.38
If $\sup_i \mathrm{E}[|X_i|^p]<\infty$ for $p>1$, then $\{X_i\}$ is uniformly integrable.

Proof. Let $C=\sup_i \mathrm{E}[|X_i|^p]^{1/p}<\infty$. Notice that by the Hölder's inequality,

$$\mathrm{E}\left[|X_i|\,\mathbf{1}\{|X_i|>M\}\right]\le \mathrm{E}[|X_i|^p]^{1/p}\,\mathrm{E}\left[\mathbf{1}\{|X_i|>M\}\right]^{1/q}\le C\,\mathrm{P}\{|X_i|>M\}^{1/q}.$$

Suppose that for every M we can find a subsequence such that $\mathrm{P}\left\{\left|X_{i_k}\right|>M\right\}\ge\delta>0$. Then $C\ge \mathrm{E}\left[\left|X_{i_k}\right|^p\right]\ge \mathrm{E}\left[\left|X_{i_k}\right|^p\mathbf{1}\left\{\left|X_{i_k}\right|>M\right\}\right]\ge M^p\delta\to\infty$ as $M\to\infty$, which is a contradiction. Hence $\sup_i \mathrm{P}\{|X_i|>M\}\to 0$ as $M\to\infty$ and

$$\sup_i \mathrm{E}\left[|X_i|\,\mathbf{1}\{|X_i|>M\}\right]\le C\,\mathrm{P}\{|X_i|>M\}^{1/q}\to 0$$

as $M \to \infty$. We conclude that $\{X_i\}$ is uniformly integrable. ■

Example

If we only have $\sup_i \mathrm{E}[|X_i|] < \infty$*, then the result does not hold in general. For instance, consider* X_n *such that* $\mathrm{P}(X_n = 0) = 1 - 1/n$ *and* $\mathrm{P}(X_n = n) = 1/n$*. Then* $\mathrm{E}[X_n] = 1 < \infty$ *but*

$$\lim_{M\to\infty} \sup_n \mathrm{E}[|X_n| \mathbf{1}\{|X_n| > M\}] = 1 \neq 0.$$

Theorem 3.39

$\{X_i\}$ *is uniformly integrable if and only if* $\sup_i \mathrm{E}[|X_i|] < \infty$ *and for all* $\epsilon > 0$*, there is* $\delta > 0$ *such that* $\mathrm{P}(A) \le \delta$ *implies that* $\mathrm{E}[|X_i| \mathbf{1}_A] < \epsilon$*.*

Proof. Assume that $\{X_i\}$ is uniformly integrable. Then $\sup_i \mathrm{E}[|X_i|] < \infty$. For every $\epsilon > 0$, we can find M such that $\mathrm{E}[|X_i| \mathbf{1}\{|X_i| > M\}] < \epsilon/2$. Pick $\delta = \epsilon/(2M)$. We have

$$\begin{aligned}\mathrm{E}[|X_i| \mathbf{1}_A] &= \mathrm{E}[|X_i| \mathbf{1}\{|X_i| > M\} \mathbf{1}_A] + \mathrm{E}[|X_i| \mathbf{1}\{|X_i| \le M\} \mathbf{1}_A] \\ &\le \mathrm{E}[|X_i| \mathbf{1}\{|X_i| > M\}] + M\,\mathrm{P}(A) \le \frac{\epsilon}{2} + M\frac{\epsilon}{2M} = \epsilon.\end{aligned}$$

For the converse, notice that by the Markov inequality,

$$\mathrm{P}\{|X_i| > M\} \le \frac{1}{M}\mathrm{E}[|X_i|] \le \frac{1}{M}\sup_i \mathrm{E}[|X_i|].$$

For any $\epsilon > 0$, there is some $\delta > 0$ such that if $\mathrm{P}(A) \le \delta$, $\mathrm{E}[|X_i| \mathbf{1}_A] < \epsilon$. Put $A_M = \{|X_i| > M\}$. For any $M > M_0 = \delta^{-1}\sup_i \mathrm{E}[|X_i|]$, $\mathrm{P}(A_M) \le \delta$ and hence $\mathrm{E}\left[|X_i| \mathbf{1}_{A_M}\right] < \epsilon$. Since the choice of M_0 is uniform in i, $\{X_i\}$ is uniformly integrable. ■

Lemma 3.40

$X_n \to 0$ *in* $\mathcal{L}^1$ *if and only if* $X_n \xrightarrow{p} 0$ *and* $\{X_n\}$ *is uniformly integrable.*

Proof. Assume that $X_n \to 0$ in $\mathcal{L}^1$. Then $X_n \xrightarrow{p} 0$. By theorem 3.39, it suffices to check that $\sup_n \mathrm{E}[|X_n|] < \infty$ and for all $\epsilon > 0$, there is $\delta > 0$ such that $\mathrm{P}(A) \le \delta$ implies $\mathrm{E}[|X_n| \mathbf{1}_A] < \epsilon$. Note that $\{X_n\}$ is convergent in $\mathcal{L}^1$ and hence must be $\mathcal{L}^1$ bounded. We have $\sup_n \mathrm{E}[|X_n|] < \infty$. For every $\epsilon > 0$, ■

Theorem 3.41

Let $X_n \xrightarrow{p} X$ *and* $\mathrm{E}[|X_n|] < \infty$*. Then the followings are equivalent:*

(a) $\{X_n\}$ *is uniformly integrable.*

(b) $X_n \to X$ *in* $\mathcal{L}^1$*.*

(c) $\mathrm{E}[|X_n|] \to \mathrm{E}[|X|]$.

Proof. We first prove that (a) implies (b). By the unifom integrability, $\sup_n \mathrm{E}[|X_n|] < \infty$. By the Fatou's lemma (corollary 2.9),

$$\mathrm{E}[|X|] \le \liminf_{n\to\infty} \mathrm{E}[|X_n|] \le \sup_n \mathrm{E}[|X_n|] < \infty.$$

Hence X itself is uniformly integrable. By proposition 3.37, $\{X_n - X\}$ is uniformly integrable. Let $Y_n = |X_n - X| \xrightarrow{p} 0$. It now suffices to prove that $\mathrm{E}[Y_n] \to 0$. Notice that for $M > 0$,

$$\begin{aligned}\mathrm{E}[|Y_n|] &= \mathrm{E}[|Y_n| \mathbf{1}\{|Y_n| \le M\}] + \mathrm{E}[|Y_n| \mathbf{1}\{|Y_n| > M\}] \\ &\le \mathrm{E}[|Y_n| \mathbf{1}\{|Y_n| \le M\}] + \sup_n \mathrm{E}[|Y_n| \mathbf{1}\{|Y_n| > M\}]\end{aligned}$$

Taking $n \to \infty$, the first term converges to 0 by corollary 2.9 since $|Y_n| \mathbf{1}\{|Y_n| \le M\} \le M \in \mathcal{L}^1$. We see that

$$\limsup_{n\to\infty} \mathrm{E}[|Y_n|] \le \sup_n \mathrm{E}[|Y_n| \mathbf{1}\{|Y_n| > M\}].$$

Taking $M \to \infty$, the uniform integrability implies that the right hand side converges to 0. We conclude that $\mathrm{E}[|Y_n|] \to$ and $X_n \to X$ in $\mathcal{L}^1$.

Now assume (b) holds.

$$|\mathrm{E}[|X_n|] - \mathrm{E}[|X|]| \le \mathrm{E}[|X_n - X|] \to 0$$

by the assumption.

Suppose that (c) holds. Note that for $M > 0$,

$$\mathrm{E}[|X_n| \mathbf{1}\{|X_n| > M\}] = \mathrm{E}[|X_n|] - \mathrm{E}[|X_n| \mathbf{1}\{|X_n| \le M\}] \to \mathrm{E}[|X|] - \mathrm{E}[|X| \mathbf{1}\{|X| \le M\}]$$

as $n \to \infty$ by corollary 2.9 with $|X_n| \mathbf{1}\{|X_n| \le M\} \to |X| \mathbf{1}\{|X| \le M\}$ and that $|X_n| \mathbf{1}\{|X_n| \le M\} \le M \in \mathcal{L}^1$. Notice that the right hand side convegres to 0 as $M \to \infty$ by LMCT. Hence, for $\epsilon > 0$, we can find M_0 so that if $M > M_0$, there is some $n_0(M_0)$ such that $\mathrm{E}[|X_n| \mathbf{1}\{|X_n| > M\}] < \epsilon$ for all $n > n_0$. Note that $\{X_n\}_{n \le n_0}$ is uniformly integrable. We conclude that $\{X_n\}$ is uniformly integrable. ∎

Theorem 3.42
If $X \in \mathcal{L}^1$, then $\{\mathrm{E}[X|\mathcal{G}] \mid \mathcal{G} \subset \mathcal{F}$ is a σ-algebra$\}$ is uniformly integrable.

Proof. To prove the theorem, we use the characterization theorem 3.39. Note that for $\mathcal{G} \subset \mathcal{F}$,

$$\mathrm{E}[|\mathrm{E}[X|\mathcal{G}]|] \le \mathrm{E}[\mathrm{E}[|X|\,|\mathcal{G}]] = \mathrm{E}[|X|] < \infty.$$

Hence $\sup_{\mathcal{G}\subset\mathcal{F}} \mathrm{E}[|\mathrm{E}[X|\mathcal{G}]|] < \infty$. Also, for $M > 0$,

$$\begin{aligned}\mathrm{E}[|\mathrm{E}[X|\mathcal{G}]| \mathbf{1}\{|\mathrm{E}[X|\mathcal{G}]| > M\}] &\le \mathrm{E}[\mathrm{E}[|X|\,|\mathcal{G}] \mathbf{1}\{\mathrm{E}[|X|\,|\mathcal{G}] > M\}] \\ &= \mathrm{E}[|X| \mathbf{1}\{\mathrm{E}[|X|\,|\mathcal{G}] > M\}].\end{aligned}$$

Since $\{X\}$ is uniformly integrable, for any $\epsilon > 0$, there is $\delta > 0$ such that $\mathrm{P}(A) \le \delta$ implies that $\mathrm{E}[|X| \mathbf{1}_A] < \epsilon$. For M sufficiently large such that

$$\mathrm{P}\{\mathrm{E}[|X|\,|\mathcal{G}] > M\} \le \frac{1}{M} \mathrm{E}[\mathrm{E}[|X|\,|\mathcal{G}]] = \frac{1}{M} \mathrm{E}[|X|] \le \delta,$$

we have

$$\mathrm{E}\left[|\mathrm{E}\left[X|\mathcal{G}\right]|\,\mathbf{1}\left\{|\mathrm{E}\left[X|\mathcal{G}\right]| > M\right\}\right] \le \mathrm{E}\left[|X|\,\mathbf{1}\left\{\mathrm{E}\left[|X|\,|\mathcal{G}\right] > M\right\}\right] < \epsilon.$$

Since the chocie of M does not depend on $\mathcal{G}$, we see that $\{\mathrm{E}\left[X|\mathcal{G}\right] \mid \mathcal{G} \subset \mathcal{F} \text{ is a } \sigma\text{-algebra}\}$ is indeed uniformly integrable. ■

Theorem 3.43
Let X_n be a submartingale. Then the followings are equivalent.

(a) X_n is uniformly integrable.

(b) X_n converges almost surely and in $\mathcal{L}^1$.

(c) X_n converges in $\mathcal{L}^1$.

If, furthermore, X_n is a martingale, then (a), (b) and (c) are also equivalent to that there is an $X \in \mathcal{L}^1$ such that $X_n = \mathrm{E}\left[X|\mathcal{F}_n\right]$.

Proof. Assume (a). By the uniform integrability we have $\sup_n \mathrm{E}\left[|X_n|\right] < \infty$. By the martingale convergence theorem, X_n converges almost surely. By theorem 3.41, X_n converges in $\mathcal{L}^1$. Hence (b) holds.

(b) to (c) is trivial.

Assume (c). Since X_n converges in $\mathcal{L}^1$, it also converges in probability. Also, $\mathrm{E}\left[|X_n|\right] < \infty$ by the submartingale property. By theorem 3.41, X_n is uniform integrable. (a) holds.

Now suppose that X_n is a martingale. If there is an X such that $X_n = \mathrm{E}\left[X|\mathcal{F}_n\right]$. Then by theorem 3.42, $\{X_n\}$ is uniformly integrable. Conversely, assume (c). Write $X_n \to X$ in $\mathcal{L}^1$. Let $A \in \mathcal{F}_n$. By the martingale property, $\mathrm{E}\left[X_m|\mathcal{F}_n\right] = X_n$ for all $m > n$ and $\mathrm{E}\left[X_m\mathbf{1}_A\right] = \mathrm{E}\left[X_n\mathbf{1}_A\right]$. Also, since

$$\mathrm{E}\left[|X_m\mathbf{1}_A - X\mathbf{1}_A|\right] \le \mathrm{E}\left[|X_m - X|\right] \to 0,$$

$\mathrm{E}\left[X_m\mathbf{1}_A\right] \to \mathrm{E}\left[X\mathbf{1}_A\right]$. We see that $\mathrm{E}\left[X_n\mathbf{1}_A\right] = \mathrm{E}\left[X\mathbf{1}_A\right]$ and $X_n = \mathrm{E}\left[X|\mathcal{F}_n\right]$. ■

Theorem 3.44 (Levy's Upward Theorem)
Let $X \in \mathcal{L}^1$ and $\mathcal{F}_n \nearrow \mathcal{F}$. Then $\mathrm{E}\left[X|\mathcal{F}_n\right] \to \mathrm{E}\left[X|\mathcal{F}\right]$ almost surely and in $\mathcal{L}^1$.

Proof. Let $Y_n = \mathrm{E}\left[X|\mathcal{F}_n\right]$. By theorem 3.43, Y_n converges almost surely and in $\mathcal{L}^1$, say to Y, which we can pick to be $\mathcal{F}$-measurable. It now suffices to verify that $Y = \mathrm{E}\left[X|\mathcal{F}\right]$. To see this, let $\mathcal{P}$ be the union of $\mathcal{F}_n$. Since $\mathcal{F}_n$ form a filtration, $\mathcal{P}$ is a π-system. Consider

$$\mathcal{L} = \left\{A \in \mathcal{F} \mid \mathrm{E}\left[Y\mathbf{1}_A\right] = \mathrm{E}\left[X\mathbf{1}_A\right]\right\}.$$

We claim that $\mathcal{L}$ is a λ-system. First,

$$\mathrm{E}\left[Y\mathbf{1}_\Omega\right] = \mathrm{E}\left[Y\right] = \mathrm{E}\left[\lim_{n\to\infty} \mathrm{E}\left[X|\mathcal{F}_n\right]\right] = \lim_{n\to\infty} \mathrm{E}\left[\mathrm{E}\left[X|\mathcal{F}_n\right]\right] = \mathrm{E}\left[X\right] = \mathrm{E}\left[X\mathbf{1}_\Omega\right].$$

Hence $\Omega \in \mathcal{L}$. Next, if $A, B \in \mathcal{L}$ and $A \subset B$, then

$$\mathrm{E}\left[Y\mathbf{1}_{B-A}\right] = \mathrm{E}\left[Y\mathbf{1}_B\right] - \mathrm{E}\left[Y\mathbf{1}_A\right] = \mathrm{E}\left[X\mathbf{1}_B\right] - \mathrm{E}\left[X\mathbf{1}_A\right] = \mathrm{E}\left[X\mathbf{1}_{B-A}\right].$$

Hence $B - A \in \mathcal{L}$. Finally, if $A_n \in \mathcal{L}$ and $A_n \nearrow A$, then by the LMCT,

$$\begin{aligned}\mathrm{E}\left[Y\mathbf{1}_A\right] &= \mathrm{E}\left[Y^+\mathbf{1}_A\right] - \mathrm{E}\left[Y^-\mathbf{1}_A\right] = \lim_{n\to\infty}\mathrm{E}\left[Y^+\mathbf{1}_{A_n}\right] - \lim_{n\to\infty}\mathrm{E}\left[Y^-\mathbf{1}_{A_n}\right] \\ &= \lim_{n\to\infty}\mathrm{E}\left[X^+\mathbf{1}_{A_n}\right] - \lim_{n\to\infty}\mathrm{E}\left[X^-\mathbf{1}_{A_n}\right] = \mathrm{E}\left[X^+\mathbf{1}_A\right] - \mathrm{E}\left[X^-\mathbf{1}_A\right] = \mathrm{E}\left[X\mathbf{1}_A\right]\end{aligned}$$

We conclude that $A \in \mathcal{L}$ and hence $\mathcal{L}$ is a λ-system. For any $A \in \mathcal{P}$, $A \in \mathcal{F}_n$ for some n.

$$\mathrm{E}\left[Y\mathbf{1}_A\right] = \mathrm{E}\left[\lim_{n\to\infty} Y_n\mathbf{1}_A\right] = \mathrm{E}\left[\lim_{n\to\infty}\mathrm{E}\left[X\mathbf{1}_A|\mathcal{F}_n\right]\right] = \lim_{n\to\infty}\mathrm{E}\left[\mathrm{E}\left[X\mathbf{1}_A|\mathcal{F}_n\right]\right] = \mathrm{E}\left[X\mathbf{1}_A\right].$$

Thus $A \in \mathcal{L}$ and $\mathcal{P} \subset \mathcal{L}$. By the π-λ theorem, for all $A \in \mathcal{F}$, $\mathrm{E}\left[X\mathbf{1}_A\right] = \mathrm{E}\left[Y\mathbf{1}_A\right]$. This shows that $Y = \mathrm{E}\left[X|\mathcal{F}\right]$. ■

Corollary 3.45 (Levy's Zero-One Law)
If $\mathcal{F}_n \nearrow \mathcal{F}$ and $A \in \mathcal{F}$, then $\mathrm{P}(A|\mathcal{F}_n) \to \mathbf{1}_A$ almost surely.

Proof. By Levy's upward theorem, $\mathrm{P}(A|\mathcal{F}_n) = \mathrm{E}\left[\mathbf{1}_A|\mathcal{F}_n\right] \to \mathrm{E}\left[\mathbf{1}_A|\mathcal{F}\right] = \mathbf{1}_A$ almost surely. ■

Corollary 3.46
Suppose that $X_n \to X$ almost surely with $X_n \in \mathcal{L}^1$. If $|X_n| \le Y$ for some $Y \in \mathcal{L}^1$ and $\mathcal{F}_n \nearrow \mathcal{F}$, $\mathrm{E}\left[X_n|\mathcal{F}_n\right] \to \mathrm{E}\left[X|\mathcal{F}\right]$ almost surely.

Proof. Notice that

$$\left|\mathrm{E}\left[X_n|\mathcal{F}_n\right] - \mathrm{E}\left[X|\mathcal{F}\right]\right| \le \left|\mathrm{E}\left[X_n|\mathcal{F}_n\right] - \mathrm{E}\left[X|\mathcal{F}_n\right]\right| + \left|\mathrm{E}\left[X|\mathcal{F}_n\right] - \mathrm{E}\left[X|\mathcal{F}\right]\right|.$$

It now sufices to prove that the two terms on the right hand side converges almost surely to 0. For the first term, let $Z_N = \sup_{m,n\ge N}|X_m - X_n| \le 2Y$. Note that $Z_N \to 0$ almost surely. For fixed N,

$$\limsup_{n\to\infty}\mathrm{E}\left[|X_n - X|\,|\mathcal{F}_n\right] \le \lim_{n\to\infty}\mathrm{E}\left[Z_N|\mathcal{F}_n\right] = \mathrm{E}\left[Z_N|\mathcal{F}\right]$$

almost surely by Levy's upward theorem. Taking $N \to \infty$ shows that

$$\left|\mathrm{E}\left[X_n|\mathcal{F}_n\right] - \mathrm{E}\left[X|\mathcal{F}_n\right]\right| \le \mathrm{E}\left[|X_n - X|\,|\mathcal{F}_n\right] \to 0.$$

almost surely since $\mathrm{E}\left[Z_N|\mathcal{F}\right] \to 0$ by LDCT. For the second term, directly applying Levy's upward theorem gives

$$\left|\mathrm{E}\left[X|\mathcal{F}_n\right] - \mathrm{E}\left[X|\mathcal{F}\right]\right| \to 0$$

almost surely. Hence the conclusion follows. ■

Theorem 3.47 (Optional Stopping Theorem II)
Let X_n be a uniformly integrable martingale and $\sigma \le \tau$ almost surely are stopping times. Then $\mathrm{E}\left[X_\tau\right] < \infty$ and

$$X_\sigma = \mathrm{E}\left[X_\tau|\mathcal{F}_\sigma\right].$$

In particular, τ can be taken as ∞.

Proof. Since X_n is a uniformly integrable martingale, there is $X_\infty \in \mathcal{L}^1$ such that $X_n = \mathrm{E}\left[X_\infty | \mathcal{F}_n\right]$ by theorem 3.43. We may take X_∞ to be $\mathcal{F}_\infty$-measurable. We first show that

$$X_\tau = \mathrm{E}\left[X_\infty | \mathcal{F}_\tau\right].$$

Define $Y_n = \mathrm{E}\left[X_\infty^+ | \mathcal{F}_n\right]$ and $Z_n = \mathrm{E}\left[X_\infty^- | \mathcal{F}_n\right]$ and $Y_\infty = X_\infty^+$, $Z_\infty = X_\infty^-$. Clearly, $X_n = Y_n - Z_n$. For $A \in \mathcal{F}_\tau$,

$$\mathrm{E}\left[Y_\infty \mathbf{1}_{A \cap \{\tau = \infty\}}\right] = \mathrm{E}\left[Y_\tau \mathbf{1}_{A \cap \{\tau = \infty\}}\right]$$

since $Y_n = \mathrm{E}\left[Y_\infty | \mathcal{F}_n\right] \to Y_\infty$ almost surely by Levy's upward theorem. Now, since $A \cap \{\tau = k\} \in \mathcal{F}_\tau$,

$$\begin{aligned}
\mathrm{E}\left[Y_\infty \mathbf{1}_{A \cap \{\tau < \infty\}}\right] &= \sum_{k \geq 0} \mathrm{E}\left[Y_\infty \mathbf{1}_{A \cap \{\tau = k\}}\right] = \sum_{k \geq 0} \mathrm{E}\left[Y_k \mathbf{1}_{A \cap \{\tau = k\}}\right] \\
&= \sum_{k \geq 0} \mathrm{E}\left[Y_\tau \mathbf{1}_{A \cap \{\tau = k\}}\right] = \mathrm{E}\left[Y_\tau \mathbf{1}_{A \cap \{\tau < \infty\}}\right].
\end{aligned}$$

Thus, $\mathrm{E}\left[Y_\tau \mathbf{1}_A\right] = \mathrm{E}\left[Y_\infty \mathbf{1}_A\right]$ and $\mathrm{E}\left[Y_\infty | \mathcal{F}_\tau\right] = Y_\tau$ almost surely. Similarly, $\mathrm{E}\left[Z_\infty | \mathcal{F}_\tau\right] = Z_\tau$ and hence $\mathrm{E}\left[X_\infty | \mathcal{F}_\tau\right] = X_\tau$. Now for $\sigma \leq \tau$,

$$\mathrm{E}\left[X_\tau | \mathcal{F}_\sigma\right] = \mathrm{E}\left[\mathrm{E}\left[X_\infty | \mathcal{F}_\tau\right] | \mathcal{F}_\sigma\right] = \mathrm{E}\left[X_\infty | \mathcal{F}_\sigma\right] = X_\sigma.$$

The proof is complete. ■

3.4. Backward Martingale

Definition 3.48
Consider the filtration $\cdots \subset \mathcal{F}_{-1} \subset \mathcal{F}_0$. *A process* X_n *is called an* $\mathcal{F}_n$***-backward martingale*** *if it is* $\mathcal{F}_n$*-adapted and* $\mathrm{E}\left[X_{n+1}|\mathcal{F}_n\right] = X_n$ *for all* $n < 0$.

Remark
Alternatively, we can define $Y_n = X_{-n}$ *and* $\mathcal{G}_n = \mathcal{F}_{-n}$. *Then the underlying process* X_n *is a* $\mathcal{F}_n$*-backward martingale if and only if* Y_n *is* $\mathcal{G}_n$*-adapted and* $\mathrm{E}\left[Y_n|\mathcal{G}_{n+1}\right] = Y_{n+1}$.

Theorem 3.49
Let X_n *be a backward martingale. Then* $X_n \to X_{-\infty}$ *as* $n \to -\infty$ *for some* $X_{-\infty} \in \mathcal{L}^1$ *almost surely and in* $\mathcal{L}^1$. *Furthermore,* $X_{-\infty} = \mathrm{E}\left[X_0| \cap_{n\leq 0} \mathcal{F}_n\right]$.

4. Continuous-Time Stochastic Processes

4.1. Poisson Process

4.2. Semigroup Theory

Definition 4.1
Let $(S, \mathcal{S})$ be a measurable space. A ***probability kernel*** *is a function $\mu : (s, A) \mapsto [0, 1]$ such that*

(a) For every fixed $s \in S$, $A \mapsto \mu(s, A)$ is a probability measure on S.

(b) For every fixed $A \in \mathcal{S}$, $s \mapsto \mu(s, A)$ is measurable.

Definition 4.2
Let $(S, \mathcal{S})$ be a measurable space and μ be a probability kernel on S. The ***transition operator*** *T associated with μ is an operator acting on*

$$M = \{ f : S \to \mathbb{R} \mid f \text{ is measurable and non-negative or bounded} \},$$

and defined by

$$Tf(x) = \int \mu(x, dy) f(y).$$

Remark
For every $f \in M$, Tf is again measurable. Indeed, for any characteristic function $\mathbf{1}_A$, $A \in \mathcal{S}$, $(T\mathbf{1}_A)^{-1}(B) = \{x \in S \mid \mu(x, A) = B\} \in \mathcal{S}$ for all $B \in \mathcal{S}$ by the measurability of $x \mapsto \mu(x, A)$. The result extends by linearity and approximation to all functions in M.

Definition 4.3
Let $\{T_t\}_{t \geq 0}$ be a family of parametrized operator. T_t is called a ***semigroup*** *if*

(a) For $s, t \geq 0$, $T_s T_t = T_{s+t}$.

(b) $T_0 = I$.

Proposition 4.4
For each $t \geq 0$, consider the probability kernel μ_t and the associated transition operator T_t. Then $T_t T_s = T_{t+s}$ if and only if $\mu_t \mu_s = \mu_{t+s}$.

Proof. Let $B \in \mathcal{S}$. Then

$$T_t T_s \mathbf{1}_B(x) = \int \mu_t(x, dy)(T_s \mathbf{1}_B)(y) = \int \mu_t(x, dy) \mu_s(y, B) = (\mu_t \mu_s)(x, B).$$

Now the result holds for every measurable characteristic functions. The result extends by the linearity and the monotone convergence. ■

Definition 4.5
Let S be a locally compact and separable metric space and $(C_0, \|\cdot\|_\infty)$ collect all the functions on S that are continuous and vanishes at the infinities. A semigroup on C_0 is called a ***Feller semigroup*** *if*

(a) $T_t C_0 \subset C_0$, *and*

(b) $T_t f(x) \to f(x)$ *as* $t \to 0$ *for* $f \in C_0$ *and* $x \in S$.