MATH 587: Advanced Probability Theory
Final Exam: 14 December 2021 18:30-21:30

Contents

1 Measure Space. Probability Space

1.1 Review of Probability Space

The standard notation for a probability space is $(\Omega, \mathcal{F}, \mathbb{P})$: **random trials.** The components of this tuple are given by:

1. Ω: this is the **sample space**. It's the collection of *all* possible outcomes.

$$\omega \in \Omega \text{ is a } \textbf{sample point}.$$

2. $\mathcal{F}$: this is a σ-algebra: this is a collection of events. For $A \in \mathcal{F}$, we say that A is an **event**. Hence, $A \subseteq \Omega$.
3. $\mathbb{P}$: this ia a function defined on the σ-algebra $\mathcal{F}$:

$$\mathcal{F} \to [0,1]$$
$$A \in \mathcal{F} \mapsto \mathbb{P}(A) \in [0,1].$$

This is called the **probability of the event** A.

Example 1. Flip a fair coin. Then,

$$\Omega = \{H, T\}.$$
$$\mathcal{F} = \{\{H\}, \{T\}, \{H, T\}, \emptyset\}.$$

Then, the probabilities are given by:

$$\mathbb{P}(H) = \frac{1}{2},\ \mathbb{P}(T) = \frac{1}{2},\ \mathbb{P}(\{H, T\}) = 1,\ \mathbb{P}(\emptyset) = 0.$$

1.2 Measure Theory

Measure theory is the foundation of modern probability theory. We will define things for a general measure space (S, Σ, μ) to replace $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1 (Algebra)**.** Let S be a set. A collection Σ_0 of subsets of S is called an **algebra** if:

1. $S \in \Sigma_0$.
2. **(Closed Under Complements)**: $A \in \Sigma_0 \Rightarrow A^c = S \setminus A \in \Sigma_0$.
3. **(Closed Under Finite Unions)**: $\forall\ n \in \mathbb{N}$, if $A_1, ..., A_n \in \Sigma_0$, then,

$$\bigcup_{j=1}^{n} A_j \in \Sigma_0.$$

Remarks: If Σ_0 is an algebra of S, then:

1. $\emptyset \in \Sigma_0$ (by **(1)** and **(2)**).
2. if $A, B \in \Sigma_0$, then $A \cup B$, $A \cap B$, $A \setminus B$, $A \triangle B$, $B \setminus A \in \Sigma_0$.
3. for all $n \in \mathbb{N}$, $A_1, ..., A_n \in \Sigma_0 \Rightarrow \bigcap_{j=1}^{n} A_j \in \Sigma_0$ **(Closed Under Finite Intersections)**.

Definition 2 (σ-algebra)**.** A collection Σ of S is a **sigma algebra** if:

1. Σ is an algebra.
2. **(Closed Under Countable Unions)**: $A_1, A_2, A_3, ... \in \Sigma \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \Sigma$.

Note that if Σ is a sigma algebra, then Σ satisfies **(1)-(6)**, and:

$$A_1, A_2, ... \in \Sigma \Rightarrow \bigcap_{j=1}^{\infty} A_j \in \Sigma.$$

Very often at this stage, if we want to prove something, we need to *go back to the definitions*. So at this stage, there will be lots of sets and logic.

Definition 3 (Measurable Space)**.** The pair (S, Σ) is a **measurable space**. A set $A \in \Sigma$ is a **measurable set**.

This means that there is a chance that it will turn measurable.

Definition 4 (σ-algebra generated by $\mathcal{C}$)**.** Let $\mathcal{C}$ be a collection of subsets of S. The **sigma algebra generated by** $\mathcal{C}$, denoted by $\sigma(\mathcal{C})$, is the *smallest* σ-algebra which is a superset of $\mathcal{C}$. So:

1. $\mathcal{C} \subseteq \sigma(\mathcal{C})$.
2. if Σ' is a σ-algebra containing $\mathcal{C}$, then $\sigma(\mathcal{C}) \subseteq \Sigma'$.

Proposition 1. (Properties of σ-algebra)

1. if $\mathcal{C}$ is a σ-algebra, then $\sigma(\mathcal{C}) = \mathcal{C}$.
2. $\sigma(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$.
3. If $\mathcal{C}_1 \subseteq \mathcal{C}_2$, then $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$.

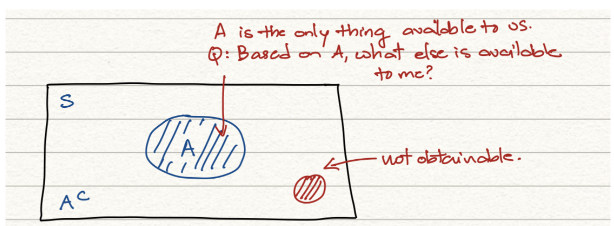
Proposition 2. Another way of getting the smallest σ-algebra:

$$\sigma(\mathcal{C}) = \bigcap \{\Sigma \mid \Sigma \text{ is a sigma algebra and } \mathcal{C} \in \Sigma\} \tag{1}$$

We have the following fact: for an index set I, if $\{\Sigma_\alpha \mid \alpha \in I\}$ is any collection of σ-algebras of subsets of S, then $\bigcap \Sigma_\alpha$ remains a σ-algebra, i.e. , the intersections of σ-algebras are still σ-algebras.

Exercise: prove **Proposition 1** and **Fact**.

Example 2. Let $A, B \subseteq S$. Then, what is $\sigma(\{A\})$?



Then, $\sigma(\{A\}) = \{A, A^c, \emptyset, S\}$. This means that these are the ONLY 4 sets which are measurable with respect to $\sigma(\{A\})$. Nothing else is measurable.

Exercise: find $\sigma(\{A, B\})$.

What does a σ-algebra mean for us? A σ-algebra contains the collection of events, so it tells me the information available to me (from the point of view of probability). If you're not in the σ-algebra, then you're not measurable with respect to the σ-algebra.

Example 3 (Borel σ-algebra)**.** Take $S = \mathbb{R}$. Then, $\mathcal{B}(\mathbb{R})$ is the **Borel Sigma Algebra**, which is defined as:

$$\mathcal{B}(\mathbb{R}) = \sigma(\{\text{open subsets of } \mathbb{R}\}) \tag{2}$$

This applies to any topological space. So, an equivalent condition for $\mathbb{R}$ is:

$$\mathcal{B}(\mathbb{R}) = \sigma(\{]a, b[\mid a < b,\ a, b \in \mathbb{R}\}) \tag{3}$$

If $B \in \mathcal{B}(\mathbb{R})$, then B is called a **Borel Set**.

1. $\mathcal{B}(\mathbb{R}) = \sigma(\{]a, b] \mid a < b,\ a, b \in \mathbb{R}\})$.
2. $\mathcal{B}(\mathbb{R}) = \sigma(\{[a, b[\mid a < b,\ a, b \in \mathbb{R}\})$.
3. $\mathcal{B}(\mathbb{R}) = \sigma(\{[a, b] \mid a < b,\ a, b \in \mathbb{R}\})$.

4. $\mathcal{B}(\mathbb{R}) = \sigma(\{]a, \infty[\mid a \in \mathbb{R}\})$.
5. $\mathcal{B}(\mathbb{R}) = \sigma(\{[a, \infty[\mid a \in \mathbb{R}\})$.

and so on... the message is that the generating class which gives us $\mathcal{B}(\mathbb{R})$ is *not* unique. Let's see how a proof of showing some of them are equivalent works. Let's show that $\Sigma_{],[} = \Sigma_{],]}$. For all $a, b \in \mathbb{R}$, for $a < b$, how do I construct $]a, b[$? We push $],]$ out from the inside:

$$]a, b[= \bigcup_{n=1}^{\infty} \left]a, b - \frac{1}{n}\right[$$

This shows the inclusion, $\Sigma_{],[} \subseteq \Sigma_{],]}$. Similarly, if I want to reach $]a, b]$, we take:

$$\bigcap_{n=1}^{\infty} \left]a, b + \frac{1}{n}\right] \tag{4}$$

This shows that $\Sigma_{],]} \subseteq \Sigma_{],[}$. Also, don't forget that for all $x \in \mathbb{R}$, $\{x\} \in \mathcal{B}(\mathbb{R})$.

Definition 5 (π-system)**.** Let S be a set. A collection I (of subsets of S) is called a **π-system** if for all $A, B \in I$, $A \cap B \in I$.

So, a π-system is a collection which is closed under intersections.

Definition 6. Let S be a set. A collection D (of subsets of S) is called a **d-System** if:

1. $S \in D$.
2. **(Closed Under Taking Differences)**: if $A, B \in D$ and if $A \subseteq B$, then $B \setminus A \in D$.
3. **(Closed Under Monotonic Limits)**: if $A_n \in D$ for $n \geq 1$, (some countable sequence of sets), and if $A_n \uparrow A$, then $A \in D$.

Definition 7 (Set-Theoretic Limits)**.** "$A_n \uparrow A$" means that:

- **Monotonic Increasing**: $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$.
- $\bigcup_{n=1}^{\infty} A_n = A$.

"$B_n \downarrow B$" means that:

- **Monotonic Decreasing**: $B_{n+1} \subseteq B_n$
- $\bigcap_{n=1}^{\infty} B_n = B$.

The reason we care about π-systems and d-Systems is that we can separate the properties of $\mathcal{B}(\mathbb{R})$ into a π-system and a d-system. This will let us further decode a σ-algebra.

Lemma 1. Let Σ be a collection of subsets of S. Then, Σ is a σ-algebra $\iff$ Σ is a π-system and a d-system

Proof. "$\Rightarrow$": trivial.
"$\Leftarrow$": we verify that Σ is a σ-algebra:

1. $S \in \Sigma$ ✓.
2. if $A \in \Sigma \Rightarrow A^c = S \setminus A \in \Sigma$ (d-System) ✓.
3. if $A_n \in \Sigma$ for $n1$, we need to check that $\bigcup_{n=1}^{\infty} A_n \in \Sigma$: this is the one that will need some work.

   Set $B_n := \bigcup_{j=1}^{n} A_j$. Then, B_n forms an increasing sequence, and $B_n \in \Sigma$ for all $n \in \mathbb{N}$ since Σ is a π-system and a d-System and by DeMorgan's Law. But, by **(III)** of a π-system, $\bigcup_{n=1}^{\infty} B_n \in \Sigma$. But,

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n \in \Sigma.$$

□

Theorem 1 (Dynkins π-d Lemma). *Suppose that I is a π-system of subsets of S and $d(I)$ is the π-system generated by I (i.e., $d(I)$ is a d-system and if d is another d-system such that $I \in d$, then $d(I) \subseteq D$). Then, $d(I) = \sigma(I)$.*

What this theorem is saying is the following: if you start with a π-system generating class, all you need is a d-system and you'll automatically get a σ-algebra.

Proof. We observe that it is sufficient to show that $d(I)$ is a π-system. Why?

1. By the lemma, we know that $d(I)$ is a σ-algebra. Hence, we get the inclusion $\sigma(I) \subseteq d(I)$.
2. Since $\sigma(I)$ is certainly a d-system, $d(I) \subseteq \sigma(I)$.

So the general sketch of the proof is: we start with a π-system, generate a d-system, and then get a π-system. The proof technique we will use, which is standard in set theory, is the **good set principle**: we collect all the items with the property that we want, argue that this collection satisfies a certain property, then show that this collection is actually the whole set. This proof will require two stages.

First, we set $\mathcal{D}_1 := \{B \in d(I) \mid B \cap A \in d(I)\ \forall A \in I\}$. This will be our "good set."
Claim: $\mathcal{D}_1$ is a d-system. To check:

1. $S \in \mathcal{D}_1$ ✓: since $A \in I$, $A \cap S = A \in d(I)$.
2. $A_1, A_2 \in D_1$, $A_1 \subseteq A_2$, we want to show that $A_2 \setminus A_1 \in \mathcal{D}_1$ ✓: because for all $A \in I$:

$$A \cap (A_2 \setminus A_1) = \underbrace{(A_2 \cap A)}_{\in d(I)} \setminus \underbrace{(A_1 \cap A)}_{\in d(I)} \in d(I) \text{ since } d(I) \text{ is a d-system.}$$

3. For $A_n \in \mathcal{D}_1$, for $n \geq 1$ and $A_n \uparrow A$, we need to show that $A_\infty \in \mathcal{D}_1$ ✓: because for all $A \in I$,

$$\underbrace{A_n \cap A}_{\in d(I)} \uparrow A_\infty \cap A.$$

This shows that $A_\infty \cap A \in d(I)$, which shows that $A_\infty \in \mathcal{D}_1$.

Hence, we have proven that $\mathcal{D}_1$ forms a d-System, and certainly $I \subseteq \mathcal{D}_1$. But, based on how $\mathcal{D}_1$ is defined, we get that $\mathcal{D}_1 \subseteq d(I)$. Hence, $d(I) = \mathcal{D}_1$. This means that for all $B \in d(I)$ and for all $A \in I$, $B \cap A \in d(I)$. Intermediate step complete! We need to now replace for all $A \in I$ with for all $A \in d(I)$. So we do the good set principle once more. Set:

$$\mathcal{D}_2 := \{C \in d(I) \mid B \cap C \in d(I)\ \forall\ B \in d(I)\}.$$

From our intermediate step conclusion, we know that $I \subseteq \mathcal{D}_2$. Next, we need to verify that $\mathcal{D}_2$ is a d-System. Exercise: go through the three conditions of a d-System.

Since $\mathcal{D}_2$ is a d-system and $I \subseteq \mathcal{D}_2$, this shows that $d(I) \subseteq \mathcal{D}_2$. Hence, $d(I) = \mathcal{D}_2$. Now we can conclude that $\forall C \in d(I)$, for all $B \in d(I)$, $B \cap C \in d(I)$. Hence, $d(I)$ is a π-system, which is what we wanted to show. □

This idea is very important in the study of measures. When constructing a measure, we only look at the π-system which generating the σ-algebra, which is why this theorem is important.

Definition 8 (Additive). Let S be a set, Σ_0 be an algebra of subsets of S. Let μ_0 be a non-negative set function defined on Σ_0, i.e.,

$$\mu_0 : \Sigma_0 \to [0, \infty].$$

We say that μ_0 is **additive** if:

1. $\mu_0(\emptyset) = 0$
2. $\forall A, B \in \Sigma_0$ and $A \cap B = \emptyset$,

$$\mu_0(A \cup B) = \mu_0(A) + \mu_0(B).$$

We say that μ_0 is **countably additive** if:

1. $\mu_0(\emptyset) = 0$.
2. $\forall A_n \in \Sigma_0$, for all $n \geq 1$ such that $A_i \cap A_j = \emptyset$ if $i \neq j$ and $\bigcup_{n=1}^{\infty} A_n \in \Sigma_0$, we require:

$$\mu_0\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu_0(A_n). \tag{5}$$

Definition 9 (Measure)**.** Let (S, Σ) be a measure space. If μ is a non-negative set function defined on Σ and μ is countably additive, then μ is called a **measure**. The triple (S, Σ, μ) is called a **measure space**.

- If $\mu(S) < \infty$, then μ is a **finite measure**.
- If $\mu(S) = 1$, then μ is a **probability measure**.
- if there exists a sequence $\{S_n \mid n \geq 1\} \subseteq \Sigma$ such that $\bigcup_{n=1}^{\infty} S_n = S$ and $\mu(S_n) < \infty$ for all $n \geq 1$, then μ is **σ-finite**.

Remark. All measures we will discuss in this course will be finite or σ-finite.

- If $N \in \Sigma$ such that $\mu(N) = 0$, then we say that N is a **null set**.
- If a statement holds everywhere except on a null set, then we say that the statement is true **almost everywhere (a.e.)** or **almost surely (a.s.)**.

1.3 Properties of a Measure μ

Proposition 3 (Monotonicity)**.** Let $A, B \in \Sigma$, $A \subseteq B$. Then, $\mu(A) \leq \mu(B)$.

Proof. Write $B = A \cup (B \setminus A)$. Then, $A \cap (B \setminus A) = \emptyset$. By the additivity of μ:

$$\mu(B) = \mu(A) + \mu(B \setminus A) \Rightarrow \mu(B) \leq \mu(A).$$

Caution! Do not take the difference, $\mu(B) - \mu(A)$, because both $\mu(A)$ and $\mu(B)$ might be infinite. That would be undefined. □

Proposition 4 (Subadditivity)**.** Let $A_n \in \Sigma$ for all $n \geq 1$. Then,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n). \tag{6}$$

Proof. Set $B_1 = A_1$, and:

$$B_n := A_n \setminus \left(\bigcup_{n=1}^{\infty} A_j\right).$$

The B_n's are disjoint, and $B_n \subseteq A_n$. So, by **(Monotonicity)**, $\mu(B_n) \leq \mu(A_n)$ for all $n \geq 1$. Furthermore,

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n.$$

Hence,

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

□

Note the other two properties we have:

- if $A_n \in \Sigma$ and $\mu(A_n) = 0$ for all $n \geq 1$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = 0$: the countable union of a null set is again a null set.
- we also have finite subadditivity: if Σ_0 is an algebra of subsets of S and $\mu_0 : \Sigma \to [0, \infty]$ is additive, then for all $n \geq 1$, and for all $A_j \in \Sigma_0$, we have:

$$\mu_0\left(\bigcup_{j=1}^{n} A_j\right) \leq \sum_{j=1}^{n} \mu_0(A_j).$$

Proposition 5 (Continuity from Below)**.** If $A_n \in \Sigma$ for all $n \geq 1$ and $A_n \uparrow$, then,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

Proof. Set $B_1 = A_1$, and $B_n = A_n \setminus A_{n+1}$. The B_n's are all disjoint, and

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n.$$

Hence,

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) = \lim_{n\to\infty} \sum_{j=1}^{\infty} \mu(B_j) = \lim_{n\to\infty} \mu\left(\bigcup_{j=1}^{n} B_j\right) = \lim_{n\to\infty} \mu(A_n).$$

□

Proposition 6 (Continuity from Above)**.** If $A_n \in \Sigma$ for all $n \geq 1$ and $A_n \downarrow$ and $\mu(A_n) < \infty$ for some $n \geq 1$, then:

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

Proof. We'll show this through continuity from below. WLOG, we assume that $\mu(A_1) < \infty$. This implies that $\mu(A_n)$ is finite for all $n \in \mathbb{N}$ (by monotonicity). So, starting from this property, things become finite. Set $B_n := A_1 \setminus A_n$ for all $n \geq 1$. Then, $B_n \uparrow$ and

$$\lim_{n\to\infty} \mu(B_n) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \mu\left(\bigcup_{n=1}^{\infty} (A_1 \setminus B_n)\right) = \mu\left(A_1 \setminus \bigcap_{n=1}^{\infty} B_n\right) = \mu(A_1) - \mu\left(\bigcap_{n=1}^{\infty} B_n\right).$$

Since everything is finite, write $\mu(B_n) = \mu(A_1) - \mu(A_n)$. Then,

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right).$$

□

We remark that in general, the assumption "$\mu(A_n) < \infty$" for some n is necessary. E.g., if $S = \mathbb{R}$, $A_n :=]n, \infty[$, then:

- $\mu_L(]n, \infty[) = \infty$. So, $\lim_{n\to\infty} \mu(A_n) = \infty$.
- Since $A_n \downarrow \emptyset$, $\mu_L\left(\bigcap_{n=1}^{\infty} A_n\right) = 0$.

This is a contradiction.

Theorem 2. *Given a set S and an algebra Σ_0, assume that μ is a non-negative set function:*

$$\mu : \Sigma_0 \to [0, \infty[,$$

is a set function (i.e., μ is non-negative and real-valued and μ is finitely additive. Then, μ is countably additive $\iff$ μ is continuous at the empty set. To be ***continuous at the empty set*** *means that if $A_n \in \Sigma$ and $A_n \downarrow \emptyset$ then $\lim_{n\to\infty} \mu(A_n) = 0$.*

Proof. "$\Rightarrow$": this is implied by continuity from above. The proof is exactly the same.
"$\Leftarrow$": we go straight back to the definition. Take $\{B_n \mid n \geq 1\} \subseteq \Sigma_0$ such that $B_i \cap B_j = \emptyset$ for all $i \neq j$ and assume that $B = \bigcup_{n=1}^{\infty} \in \Sigma_0$. Set $A_n := B \setminus \bigcup_{j=1}^{n} B_j$. Clearly, $A_n \downarrow \emptyset$. By continuity at the empty set, $\lim_{n\to\infty} \mu(A_n) = 0$. On the other hand, $A_n \cap \left(\bigcup_{j=1}^{n} B_j\right) = \emptyset$. Hence,

$$\begin{aligned}
\mu(B) &= \mu(A_n) + \mu\left(\bigcup_{j=1}^{n} B_j\right) \quad \textbf{(finite additivity)} \\
&= \mu(A_n) + \sum_{j=1}^{n} \mu(B_j) \quad \textbf{(finite additivity)} \underbrace{\to}_{n\to\infty} 0 + \sum_{n=1}^{\infty} \mu(B_n).
\end{aligned}$$

i.e.,

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

□

1.3.1 Existence and Uniqueness of Measure

Given a measure space (S, Σ) and two measures μ_1 and μ_2, we say that two measures are equal, $\mu_1 = \mu_2$, if for all $A \in \Sigma$, $\mu_1(A) = \mu_2(A)$ for all $A \in I$.

Theorem 3. *Given a set S and a π-system I of subsets of S, let μ_1 and μ_2 be two measures on $S, \Sigma := \sigma(I)$. Then, if $\mu_1(S) = \mu_2(S) < \infty$ and $\mu_1(A) = \mu_2(A)$ $\forall A \in I$, then $\mu_1 = \mu_2$ on the whole σ-algebra.*

Significance of this theorem:

- Only $\mathbb{R}$-valued measures allowed.
- We can extend this to any σ-finite space by breaking the space down.
- $\mu_1(S) = \mu_2(S) < \infty$ allows us to apply the theorem to σ-finite measures.

Proof. This will use "good set principle." Set:

$$\mathcal{D} := \{A \in \Sigma \mid \mu_1(A) = \mu_2(A)\}.$$

Then, $I \subseteq \mathcal{D}$, by assumption. We want to now show that $\mathcal{D}$ is a d-system. We check the properties.

1. $S \in \mathcal{D}$ is given.
2. If $A, B \in \mathcal{D}$, $A \subseteq B \Rightarrow B \setminus A \in \mathcal{D}$. This follows from the additivity of measure and the finiteness of it:

$$\begin{aligned}
\mu_1(A \setminus B) &= \mu_1(B) - \mu_1(A) \\
&= \mu_2(B) - \mu_2(A) \\
&= \mu_2(B \setminus A).
\end{aligned}$$

3. for $A_n \in \mathcal{D}$, for all $n \geq 1$, and $A_n \uparrow A$, we need to check that $A \in \mathcal{D}$. This follows from the continuity from below property:

$$\mu_1(A) = \lim_{n\to\infty} \mu_1(A_n) = \lim_{n\to\infty} \mu_2(A_n) = \mu_2(A).$$

Hence, $\mathcal{D} = \sigma(I) = \Sigma$.

$\square$

We can see that the built in properties of d-Systems are compatible with measures $\Rightarrow$ this is why we can focus on π-systems. Having established the uniqueness of measure, we can now move onto the existence of measures. We will briefly review the construction of the Lebesgue measure. We will use, but not prove, the following key result.

Theorem 4 (Caratheodory's Extension Theorem)**.** *Given a set S, suppose that Σ_0 is an algebra and $\mu_0 : \Sigma_0 \to [0, \infty]$ is countably additive. Then, there exists a measure μ defined on $\Sigma = \sigma(\Sigma_0)$ such that $\mu(A) = \mu_0(A)$ for all $A \in \Sigma$.*

This theorem tells us that we can extend a measure from Σ_0 to Σ. Moreover, if $\mu_0(S) < \infty$, then such an extension is unique.

Next, we will use Caratheodory's Extension Theorem to construct $\lambda = \lambda_{\text{Leb}}$ (the Lebesgue Measure on $(]0, 1], \mathcal{B}(]0, 1]))$.

1. We need to define a candidate measure on an algebra of subsets.

$$\Sigma_0 := \left\{ F \subseteq]0,1] \mid F = \bigcup_{i=1}^{k}]a_i, b_i] \text{ where } k \in \mathbb{N}, \text{ and } 0 \leq a_1 \leq b1 \leq a_2 \leq b_2 \leq ... \leq a_k \leq b_k \leq 1 \right\}$$

2. Define μ_0 on Σ_0: for

$$F = \bigcup_{i=1}^{k}]a_i, b_i] \in \Sigma_0,$$

then,

$$\mu_0(F) := \sum_{i=1}^{k} (b_i - a_i).$$

One can verify that μ_0 is well-defined, i.e., if

$$F = \bigcup_{i=1}^{k}]a_i, b_i] = \bigcup_{j=1}^{l}]c_j, d_j] \Rightarrow \sum_{i=1}^{k} (b_i - a_i) = \sum_{j=1}^{l} (d_j - c_j).$$

One can also verify that μ_0 is additive.

3. Now need to verify countable additivity: by the theorem, this means we need to check for continuity of μ_0 at the empty set.

Proof. Take $F_n \in \Sigma_0$ with $n \geq 1$ such that $F_n \downarrow \emptyset$. The goal is to show that $\lim_{n\to\infty} \mu_0(F_n) = 0$. To that end, assume that $\lim_{n\to\infty} \mu_0(F_n) > 0$. Then, there exists a $\delta > 0$ such that $\mu_0(F_n) \geq \delta$ for all $n \geq 1$. For each n, choose $C_n \in \Sigma_0$ such that $\overline{C}_n \subseteq F_n$ and $\mu_0(F_n \setminus C_n) \leq 2^{-(n+1)}\delta$. Define $K_n := \bigcap_{j=1}^n C_n$. Then, $K_n \downarrow$ and for all $m \geq 1$, by deMorgan's Law,

$$F_m \setminus K_m = \bigcup_{n=1}^m (F_m \setminus C_n).$$

Now, $\mu_0(F_m \setminus K_m) \leq \sum_{n=1}^m \mu_0(F_m \setminus C_n) \leq \sum_{n=1}^m \mu_0(F_n \setminus C_n) \leq \delta \sum_{n=1}^m 2^{-(n+1)} \leq \frac{\delta}{2}$ for all $m \geq 1$. The first inequality follows from finite subadditvity and the second inequality follows from monotonicity. Since $\mu_0(F_m) > \delta$ and for all $m \geq 1$, $\mu_0(K_m) \geq \frac{\delta}{2}$ for all $m \geq 1$.

$$\begin{aligned}
&\Rightarrow \exists\, x_m \in K_m \ \forall m \geq 1 \\
&\Rightarrow \{x_m\} \subseteq \overline{C}_1 \text{ (compact)} \\
&\Rightarrow \exists\, \{m_l \mid l \geq 1\} \text{ s.t. } x_{m_l} \to x_\infty \text{ as } l \to \infty.
\end{aligned}$$

For every n, when l is sufficiently large, $m_l > m$. Hence,

$$\begin{aligned}
&\Rightarrow x_{m_l} \in K_{m_l} \subseteq K_n \subseteq C_n \subseteq F_n \\
&\Rightarrow x_\infty \in \overline{C}_n = F_n \ \forall n \geq 1 \\
&\Rightarrow x_\infty \in \bigcap_{n=1}^\infty F_n.
\end{aligned}$$

This is a contradiction, so μ_0 is countably additive. □

(Will Continue Later).

1.4 Completion of a Measure / Measure Space

Sometimes, it's convenient to assume that subsets of null sets are measurable. Let (S, Σ, μ) be a measure space. Set $N := \{A \subseteq S \mid \exists\, B \in \Sigma \text{ and } \mu(B) = 0 \text{ s.t. } A \subseteq B\}$. Define a new sigma algebra:

$$\Sigma^* := \{F \subseteq S \mid \exists\, G, H \in \Sigma \text{ s.t. } G \subseteq F \subseteq H \text{ and } \mu(H \setminus G) = 0\}$$

Note that $\Sigma \subseteq \Sigma^*$.

Theorem 5. *Σ^* is a σ-algebra and Σ^* is the σ-algebra generated by:*

$$\Sigma^* = \sigma(\Sigma \cup N).$$

Proof. Exercise. □

Definition 10. Define μ^* to be a set function of Σ^* by: for all $F \in \Sigma^*$ if $G \subseteq F \subseteq H$ for some $G, H \in \Sigma$, with $\mu(G) = \mu(H)$, then,

$$\mu^*(F) := \mu(G) = \mu(H).$$

Proposition 7. μ^* is a measure on (S, Σ^*).

Example 4. Prove this statement.

Definition 11 (Complete Measure Space)**.** (S, Σ^*, μ^*) is a **complete measure space**, i.e., the completion of (S, Σ, μ) (we get this by "patching up the holes").

Note that if we complete the Lebesgue measure λ, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda_{\text{Leb}}$, then we obtain the **completed Lebesgue measure** and the **Lebesgue σ-algebra**: $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda^*)$.

1.5 Events and Independence

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Definition 12 (Set-theoretic limsup and liminf)**.** Let $\{A_n \mid n \geq 1\}$ be a sequence of events, i.e., $A_n \in \mathcal{F}$ for all $n \geq 1$. Then,

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \lim_{n\to\infty} \bigcup_{m=n}^{\infty} A_m. \tag{7}$$

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m-n}^{\infty} A_m = \lim_{n\to\infty} \bigcap_{m=n}^{\infty} A_m. \tag{8}$$

Note that $\limsup_n A_n \in \mathcal{F}$ and $\liminf_n A_n \in \mathcal{F}$. In words, what do these sets mean?

- if $\omega \in \limsup_n A_n$, then $\forall\ n \geq 1$, $\exists$ an $m_n > n$ such that $\omega \in A_{m_n} \iff \exists$ a sequence $\{m_n \mid n \geq 1\}$ such that $m_n \to \infty$ as $n \to \infty$ and $\omega \in A_{m_n}$ for all $n \geq 1 \iff \omega$ is in infinitely many $A_n's$. So, we write $\limsup_n A_n = A_n$ i.o. .
- if $\omega \in \liminf_n A_n$, then there exists an $n \geq 1$ such that $\omega \in A_m$ for all $m \geq n \iff \omega \in A_n$ eventually for all n. Hence, we say $\liminf_{n\to\infty} A_n = A_n$ "eventually always, (ea)"

Proposition 8. Properties of liminf/limsup.

1. Obviously, $\liminf_n A_n \subseteq \limsup_n A_n$. If $\liminf_n A_n = \limsup_n A_n$, then we say that the **set-theoretic limit**, $\lim_n A_n$, exists, and is defined to be:

$$\lim_n A_n = \liminf_n A_n = \limsup_n A_n.$$

   If $\{A_n\}$ is **monotonic**, then $\lim_n A_n$ exists.
2. If $\{A_n \mid n \in \mathbb{N}\}$ is a sequence of events and $\{B_n \mid n \in \mathbb{N}\} \subseteq \mathcal{F}$, and $A_n \subseteq B_n$ for all $n \in \mathbb{N}$, then:

$$\limsup_n A_n \subseteq \limsup_n B_n$$
$$\liminf_n A_n \subseteq \limsup_n B_n.$$

3. **(deMorgan's Law)**:

$$\limsup_n A_n^c = (\liminf_n A_n)^c$$
$$\liminf_n A_n^c = (\limsup_n A_n)^c$$

4. "Jumping between A_n and A_n^c":

$$(\limsup_n A_n) \setminus (\liminf_n A_n) = \limsup_n (A_n \setminus A_{n+1}) \tag{9}$$

5. Let $\{A_n \mid n \geq 1\}$ and $\{B_n \mid n \geq 1\}$ be two sequences of events. Then, in general:
   (a) $(\limsup_n A_n) \cap (\limsup_n B_n) \supseteq \limsup_n (A_n \cap B_n)$.
      i. Note that the converse inclusion is in general NOT true. For example, $A_n = \{(-1)^n\}$ and $B_n = \{(-1)^{n+1}\}$ for all $n \in \mathbb{N}$. They're out of phase, and so $A_n \cap B_n = \emptyset \Rightarrow \limsup_n (A_n \cap B_n) = \emptyset$. However, $\limsup_n A_n = \{-1, 1\} = \limsup_n B_n$.
   (b) $(\limsup_n A_n) \cup (\limsup_n B_n) = \limsup_n (A_n \cup B_n)$
   (c) $(\liminf_n A_n) \cap (\liminf_n B_n) = \liminf_n (A_n \cap B_n)$.
   (d) $(\liminf_n A_n) \cup (\liminf_n B_n) \subseteq \liminf_n (A_n \cup B_n)$. In general, the "$\supseteq$" is NOT true.

Theorem 6 (Borel-Cantelli 1 (BC1)). *If* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, *then* $\mathbb{P}(\limsup_n A_n) = 0$.

Verbally, this result tells us that if the probability of A_n decays sufficiently fast (i.e., summable), then the chance of the limsup happening will go to zero.

Proof.

$$\begin{aligned}
\mathbb{P}\left(\limsup_n A_n\right) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right). \\
&\leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m) \\
&= 0.
\end{aligned}$$

□

2 Random Variables. Independence

2.1 Independence

Definition 13. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sequence of events $\{E_n \mid n \geq 1\} \subseteq \mathcal{F}$ is called **(mutually) independent** if for all $k \in \mathbb{N}$, for all $1 \leq i_1 \leq_2 \leq ... \leq i_2$:

$$\mathbb{P}\left(\bigcap_{j=1}^{k} E_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}\left(E_{i_j}\right). \tag{10}$$

It's also possible to define independence for an uncountable family of events. For example, consider $\{E_\alpha \mid \alpha \in I\} \subseteq \mathcal{F}$, where I is an arbitrary index set. This is independent if, for all $k \in \mathbb{N}$, distinct $\{\alpha_1, ..., \alpha_k\} \subseteq I$,

$$\mathbb{P}\left(\bigcap_{j=1}^{k} E_{\alpha_j}\right) = \prod_{j=1}^{k} \mathbb{P}\left(E_{\alpha_j}\right).$$

Proposition 9. (Properties of Independence).

1. If $A, B \in \mathcal{F}$, we use the notation $A \perp B$, then the following are all equivalent:
   (a) $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
   (b) $\mathbb{P}(A^c \cap B) = \mathbb{P}(A^c)\mathbb{P}(B)$.
   (c) $\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c)\mathbb{P}(B^c)$.
   (d) $\mathbb{P}(A \cap B^c) = \mathbb{P}(A)\mathbb{P}(B^c)$.
2. Let A be an event ($A \in \mathcal{F}$). Then, $A \perp B$ for all $B \in \mathcal{F} \iff \mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.
   (a) Heuristically, if every event has nothing to do with A, then A must be quite extreme, i.e., it is extremely likely to happen or it is extremely unlikely to happen.

   Proof. "$\Rightarrow$": $A \perp A$ means that:

   $$\mathbb{P}(A) = \mathbb{P}(A \cap A) \Rightarrow \mathbb{P}(A) \cap \mathbb{P}(A).$$

The only solution to $x^2 = x$ is $x = 0$ and $x = 1$. Hence, $\mathbb{P}(A) \in \{0, 1\}$.

"$\Leftarrow$": if $\mathbb{P}(A) = 0$, then for all $B \in \mathcal{F}$, $A \cap B \in \mathcal{F}$ and $\mathbb{P}(A) \cap \mathbb{P}(B) = 0$. Hence,

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

□

3. We say that a sequence is pairwise independent if for all $i \neq j$,

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j)$$

i.e., $A_i \perp A_j$. Note that (mutual) independence $\Rightarrow$ pairwise independence, but the converse is not true. We can illustrate this with an example:

Example 5. Let $\Omega = \{HH, HT, TH, TT\}$. Let $\mathcal{F} = 2^{\Omega} = \{\text{all subsets of } \Omega\ \}$. For all $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = \frac{1}{4}$. Define the following events:

(a) $E_1 = \{HH, HT\}$.
(b) $E_2 = \{HH, TH\}$.
(c) $E_3 = \{HH, TT\}$.

The events are pairwise independent:

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2 \cap E_3) = \mathbb{P}(E_1 \cap E_3) = \frac{1}{4}.$$

However, the events themselves are not independent:

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) = \frac{1}{4} \neq \mathbb{P}(E_1)\,\mathbb{P}(E_2)\,\mathbb{P}(E_3).$$

4. If $A \subseteq B$, then A and B cannot be independent, unless in the trivial case:

$$\mathbb{P}(A) = 0 \text{ or } \mathbb{P}(B) = 1.$$

If $A \cap B = \emptyset$, again, A and B cannot be independent unless in the trivial case.

Definition 14 (Independent)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose that $\{G_n \mid n \in \mathbb{N}\}$ is a sequence of σ algebras, $G_n \subseteq \mathcal{F}$ for all $n \in \mathbb{N}$. Then, $\{G_n \mid n \in \mathbb{N}\}$ is **independent** if for any choice of $E_n \in G_n$ for $n \geq 1$, the sequence $\{E_n \mid n \in \mathbb{N}\}$ is independent.

Proposition 10. Given $\{E_n \mid n \in \mathbb{N}\} \subseteq \mathcal{F}$, we say that $\{E_n\}$ is independent $\iff$ $\{\sigma(E_n) \mid n \geq 1\} = \{\emptyset, \Omega, E_n, E_n^c\}$ is independent.

Theorem 7. *Given $(\Omega, \mathcal{F}, \mathbb{P})$, let $\{I_n\}$ be a sequence of π-systems (of subsets of Ω), $I_n \subseteq \mathcal{F}$ for all $n \geq 1$. Then, $\{I_n\}$ is independent $\iff$ $\{\sigma(I_n) \mid n \geq 1\}$ is independent.*

What this theorem tells us is that to check if two sigma algebras A and B are independent, we only need to show that the generating π-system is independent. This makes sense, as the behaviour of measure is determined through behaviour on a generating π-system.

Proof. "$\Leftarrow$": trivial.
"$\Rightarrow$": WLOG, we will assume that $\Omega \in I_n$ for all $n \geq 1$. Otherwise, just replace I_n with $I_n \cup \{\Omega\} = \tilde{I}_n$ and $\tilde{I}_n$ is a π-system and $\{\tilde{I}_n\}$ is independent. It's sufficient to show that for any fixed $N \geq 1$, the family $\{\sigma(I_n), \sigma(I_2), ..., \sigma(I_n)\}$ is independent. Choose and fix an arbitrary $G_n \in I_n$ for $n = 1, 2, 3, ..., N-1$. We will define two set functions on $\sigma(I_n)$, both $\sigma(I_n) \to [0, 1]$:

1. $E^N \in \sigma(I_N) \mapsto \mathbb{P}_N(E^N) := \mathbb{P}\left(G_1 \cap ... \cap G_{N-1} \cap E^N\right)$.
2. $E^N \in \sigma(I_N) \mapsto \mathbb{P}'_N(E^N) = \prod_{j=1}^{N-1} \mathbb{P}\left(G_j\right) \cdot \mathbb{P}\left(E^N\right)$.

Clearly, it's easy to check that both $\mathbb{P}_N$ and $\mathbb{P}'_N$ are measures on the same σ-algebra and

$$\mathbb{P}_N(\Omega) = \mathbb{P}'_N(\Omega).$$

Further, $\mathbb{P}_N = \mathbb{P}'_N$ on I_n and for all $G_N \in I_N$:

$$\mathbb{P}\left(\bigcap_{j=1}^{N} G_j\right) = \prod_{j=1}^{N} \mathbb{P}\left(G_j\right).$$

By the uniqueness of measures, we know that $\mathbb{P}_N = \mathbb{P}_{N'}$ on the sigma algebra generated by I_N. Therefore, for arbitrary $G_n \in I_n$ ($n = 1, ..., N-1$), and arbitrary $E^N \in \sigma(I_N)$:

$$\mathbb{P}\left(G_1 \cap G_2 \cap ... \cap G_{N-1} \cap E^N\right) = \prod_{j=1}^{N-1} \mathbb{P}\left(G_j\right) \mathbb{P}\left(E^N\right).$$

Stage 1 is done. Next, we choose and fix $G_j \in I_j$ for $1 \leq j \leq N-2$ and $E^N \in \sigma(I_n)$. Set two measures:

1. $\mathbb{P}_{N-1} := E^{N-1} \in \sigma(I_{N-1}) \mapsto \mathbb{P}_{N-1}(E^{N-1}) = \mathbb{P}\left(\bigcap_{j=1}^{N-2} G_j \cap E^{N-1} \cap E^N\right)$.
2. $\mathbb{P}'_{N-1} := E^{N-1} \in \sigma(I_n) \mapsto \mathbb{P}'_{N-1}(E^{N-1}) = \prod_{j=1}^{N-2} \mathbb{P}\left(G_j\right) \mathbb{P}\left(E^{N-1}\right) \mathbb{P}\left(E^N\right)$.

Moreover,

- $\mathbb{P}_{N-1}$ and $\mathbb{P}'_{N-1}$ are measures defined on $\sigma(I_{N-1})$.
- $\mathbb{P}_{N-1}(\Omega) = \mathbb{P}'_{N-1}(\Omega)$ by the conclusion of the previous step.
- $\mathbb{P}_{N-1} = \mathbb{P}'_{N-1}$ on I_{N-1} also by the previous step.

Hence, by the uniqueness of measure, $\mathbb{P}_{N-1} = \mathbb{P}'_{N-1}$ on the whole sigma algebra $\sigma(I_{N-1})$. Hence, $\forall G_j \in I_j$ for $j = 1, ..., N-2$, and for all $E^{N-1} \in \sigma(I_{N-1})$ and for all $E^N \in \sigma(I_N)$,

$$\mathbb{P}\left(G_1 \cap G_2 \cap ... \cap G_{N-2} \cap E^{N-1} \cap E^N\right) = \prod_{j=1}^{N-2} \mathbb{P}\left(G_j\right) \mathbb{P}\left(E^{N-1}\right) \mathbb{P}\left(E^N\right).$$

Repeating this procedure, we will eventually get: for all $E^j \in \sigma(I_j)$ for $i \leq j \leq N$,

$$\mathbb{P}\left(\bigcap_{j=1}^{N} E^j\right) = \prod_{j=1}^{N} \mathbb{P}\left(E^j\right).$$

Hence,

$$\{\sigma(I_1), \sigma(I_2), ..., \sigma(I_N)\}$$

is independent for all $N \geq 1$. Hence,

$$\{\sigma(I_n) \mid n \in \mathbb{N}\}$$

is independent. $\square$

Theorem 8 (Borel-Cantelli Lemma 2 **BC2**). *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\{E_n \mid n \in \mathbb{N}\}$ be an independent sequence of events. If*

$$\sum_{n=1}^{\infty} \mathbb{P}\left(E_n\right) = \infty,$$

then, $\mathbb{P}\left(\limsup_n E_n\right) = 1$.

Proof. It's enough to show that $\mathbb{P}\left(\liminf_n E_n^c\right) = 0$. We have:

$$\mathbb{P}\left(\liminf_n E_n^c\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{m=n}^{\infty} E_m^c\right).$$

Fix any $n \in N$. For every $n' > n$,

$$\mathbb{P}\left(\bigcap_{m=n}^{n'} E_m^c\right) \underbrace{=}_{\text{indep.}} \prod_{m=n}^{n'} \mathbb{P}\left(E_m^c\right) = \prod_{m=n}^{n'} (1 - \mathbb{P}\left(E_n\right)).$$

Next note that for every $t \in [0,1]$, $0 \leq 1 - t \leq e^{-t}$. Then,

$$\prod_{m=n}^{n'} (1 - \mathbb{P}\left(E_m\right)) \leq e^{-\sum_{m=n}^{n'} \mathbb{P}(E_m)}$$

which goes to zero as we send $n' \to \infty$ (as it's the tail of a divergent series). Hence,

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

$\square$

We remark that in **(BC2)** independence is a necessary condition. We can easily cook up counter examples for how this is false when we lose independence. For example, consider $([0,1], \mathcal{B}([0,1]), \lambda)$. Set $E_n :=]0, 1/n[$ for all $n \in \mathbb{N}$. Then:

$$\lambda(E_n) = \frac{1}{n}$$

and,

$$\sum_{n=1}^{\infty} \lambda(E_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

However, $\limsup_n E_n = \emptyset$ and $\lambda(\emptyset) = 0$. This fails since the $\{E_n\}$ is not independent, since the are nested.

Corrolary 1. Given $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{E_n \mid n \in \mathbb{N}\} \subseteq \mathcal{F}$ being independent. If $\{C_n \mid k \geq 1\}$ is a family of index sets such that $C_k \subseteq \mathbb{N}$ for all $k \geq 1$ and $C_k \cap C_l = \emptyset$ for all $k \neq l$, then $\{\sigma(E_n \mid n \in C_k) \mid k \geq 1\}$ is independent, i.e., $\{\sigma(\{E_n \mid n \in C_1\}), \sigma(\{E_n \mid n \in C_2\}), ...\}$ is independent.

Proof. For each $k \geq 1$, set:

$$I_k := \left\{ \bigcap_{j=1}^{p} E_{n_j} \mid p \geq 1,\ n_j \in C_k,\ 1 \leq n_1 \leq n_2 \leq ... \leq n_p \right\}$$

I_k is a π-system and $\sigma(I_k) = \sigma(\{E_n \mid n \in C_k\})$. We only need to show that I_k is independent. To this end, take $1 \leq k_1 \leq ... \leq k_l$ to be any finite set of indices, and $A_j \in I_{k_j}$ for $1 \leq j \leq l$. Assume that

$$A_j = \bigcap_{j=1}^{p_j} E_{n_i}^{(j)},$$

$p_j \in \mathbb{N}$ and $n_i^{(j)} \in C_{k_j}$. Then,

$$\mathbb{P}(A_1 \cap A_2 \cap ... \cap A_l) = \mathbb{P}\left(\bigcap_{j=1}^{l} E_{n_j^{(l)}}\right) = \prod_{j=1}^{l}\prod_{i=1}^{p_j} \mathbb{P}\left(E_{n_i}^{(j)}\right) = \prod_{j=1}^{l} \mathbb{P}(A_j).$$

where the second to last equality follows from the independence of $\{E_n\}$ and the disjointedness of all the $C_k's$. $\square$

Example 6. An example of **(BC2)** . Consider flipping a fair coin infinitely many times. Again let our probability space be $(\Omega, \mathcal{F}, \mathbb{P})$ as introduced before. Set:

$$E_n := \{\omega \in \Omega \mid \omega_n = 0\},$$

i.e., the nth flip results in a tail. It's easy to see that the $\{E_n\}$ are independent. We also notice that $\mathbb{P}(E_n) = \frac{1}{2}$ for all $n \geq 1$. Hence, by **(BC2)** :

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

In words, this means that almost surely tails appears infinitely many times. Now consider:

$$B_n := \{\omega_n < \omega_{n+1}\}$$

for all $n \in \mathbb{N}$. In words, this means $\omega_n = 0$ and $\omega_{n+1} = 1$. It's easy to see that $\mathbb{P}(B_n) = \frac{1}{4}$, since we are specifying outcomes for two coin flips. Note that $\{B_n \mid n \geq 1\}$ is NOT independent – there are outcomes overlapped in a flip, in particular, ω_{n+1} is involved in both flips. But, $\{B_{2n} \mid n \geq 1\}$ is independent by the previous corollary. Now we can use **(BC2)** .

$$\sum_{n=1}^{\infty} \mathbb{P}(B_{2n}) = \infty \Rightarrow \mathbb{P}(B_{2n} \text{ i.o. }) = 1$$

along a sub-sequence, so $\mathbb{P}(B_n \text{ i.o. }) = 1$.

2.2 Tail σ-algebra

Definition 15. Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of events $\{E_n \mid n \geq 1\}$, define:

$$T := \bigcap_{n=1}^{\infty} \sigma(\{E_n, E_{n+1}, E_{n+2}, ...\}). \tag{11}$$

This T is defined to be the **tail sigma algebra** associated with $\{E_n\}$. If an event $A \in T$, then A is called a **tail event** with respect to $\{E_n\}$.

Heuristically, the tail sigma algebra ignores the first finite length of a sequence of events. Let's see some examples.

Example 7. $\limsup_n E_n \in T$. Similarly, $\liminf_n E_n \in T$. Let's prove it, we will prove it for $\limsup_n E_n$ but you can do it for $\liminf_n E_n$.

Proof. For all $N \geq 1$:

$$\limsup_n E_n = \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{m=n}^{\infty} E_m}_{:=B_n \downarrow} = \bigcap_{n=N}^{\infty} \bigcup_{m=n}^{\infty} E_m \in \sigma(\{E_N, E_{N+1}, ...\})$$

Hence,

$$\limsup_n E_n \in \bigcap_{N=1}^{\infty} \sigma(\{E_N, E_{N+1}, ...\}) = T.$$

□

Theorem 9 (Kolmogorov's 0-1 Law)**.** *If $\{E_n\}$ is independent, and T is the tail σ-algebra associated with E_n, then for all $A \in T$, $\mathbb{P}(A) \in \{0, 1\}$.*

Heuristically, if $A \in T$, then it's not going to talk to finitely many of the E_n's, so it must be a somewhat trivial event.

Proof. Set:

$$I := \left\{ \bigcap_{j=1}^{k} E_{n_j} \mid k \in \mathbb{N}, n_j \in \mathbb{N}, 1 \leq j < k, \{1 \leq n_1 \leq 1 n_2 \leq ... \leq n_k\} \right\}$$

I is a π-system and $\sigma(I) = \sigma(\{E_n \mid n \in \mathbb{N}\})$. Now, given any $A \in T$, for all $N \geq 1$, $A \subseteq \sigma(\{E_{N+1}, E_{N+2}, ...\})$ This means that for all $N \geq 1$ and for all $B \in \sigma(\{E_1, ..., E_N\})$, one has that $A \perp B$, because $\{E_n\}$ is independent. Now take $E \in I$ and assume that $E = \bigcap_{j=1}^{k} E_{n_j}$. Then, $E \in \sigma(\{E_{n_1}, ..., E_{n_k})$ which shows that $A \perp E$. Hence,

$$\forall\ B \in \sigma(I) = \sigma(\{E_n \mid n \in \mathbb{N}\})\ A \perp B.$$

However, $A \in \sigma(\{E_n \mid n \in \mathbb{N}\}) \Rightarrow A \perp A \Rightarrow \mathbb{P}(A) \in \{0, 1\}$. □

Let's see an example of Kolmogorov 0-1 Law in action.

Example 8. Consider flipping a fair coin infinitely many times. Same probability space as usual, $(\Omega, \mathcal{F}, \mathbb{P})$. Set:

$$E_n := \{\omega_n = 0\}$$

for all $n \in \mathbb{N}$. We have that $\{E_n\}$ is independent.

1. Consider $E := \{\omega \in \Omega \mid \sum_{n=1}^{\infty} \omega_n < \infty\} \in \mathcal{F}$. This set is equivalent to $\{\omega \in \Omega \mid \omega_n = 0$ for all but finitely many n $\}$. We claim that $E \in T$. To see this, observe that for any $N \geq 1$:

$$\begin{aligned} E &= \{\omega \in \Omega \mid \sum_{n=1}^{\infty} \omega_n < \infty\} \\ &= \{\omega \in \Omega \mid \sum_{n=N}^{\infty} \omega_n < \infty\} \in \sigma(\{E_N, E_{N+1}, ...\}) \end{aligned}$$

   By **(K 0-1 Law)** , $\mathbb{P}(E) \in \{0, 1\}$. Very often, choosing which one is the trouble. In this case, we can see that $\mathbb{P}(E) = 0$, since $\mathbb{P}(E_n^c \text{ i.o.}) = 1$.

2. For $r \in [0,1]$, set

$$E_r := \{\omega \in \Omega \mid \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \omega_j = r\}.$$

We can confirm that $E_r \in T$, because for all $N \geq 1$:

$$\begin{aligned} E_r &= \{\omega \in \Omega \mid \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \omega_j = r\} \\ &= \{\omega \in \Omega \mid \lim_{n\to\infty} \frac{1}{n}\sum_{j=N}^{n} \omega_j = r\} \in \sigma(\{E_N, E_{N+1}, ...\}) \end{aligned}$$

Hence, by **(K 0-1 Law)** , $\forall\, r \in [0,1]$, $\mathbb{P}(E_r) \in \{0,1\}$. Since every limit is unique, $E_r \cap E_{r'} = \emptyset$ for all $r \neq r'$, and so there exists at most one r, r^*, such that $\mathbb{P}(E_{r^*}) = 1$.

2.3 Measurable Functions

Definition 16 (Measurable)**.** Let (S,Σ) be a measurable space and $h : S \to \mathbb{R}$ (in certain situations, could be $\overline{\mathbb{R}}$) be a function. We say that h is **Σ-measurable**, denoted by $h \in m\Sigma$ if: $\forall B \in \mathcal{B}(\mathbb{R})$, the pre-image of B under h is measurable, i.e., $h^{-1}(B) \in \Sigma$.

Proposition 11. (Properties of Measurable Functions)

1. If h is a measurable function, $h \in m\Sigma$, then $\{h = \infty\} = \{s \in S \mid h(s) = +\infty\} \in \Sigma$ and $\{h = -\infty\} = \{s \in S \mid h(s) = +\infty\} \in \Sigma$.
   (a) Because, for example $\{h = \infty\} = \bigcap_{n=1}^{\infty}\{h > n\} = \bigcap_{n=1}^{\infty} h^{-1}(]n,\infty[) \in \Sigma$.
2. More generally, $h : (S_1,\Sigma_1) \to (S_2,\Sigma_2)$, then we say that h is $\Sigma_1 \setminus \Sigma_2$-measurable if $\forall B \in \Sigma_2$, $h^{-1}(B) \in \Sigma_1$.
3. For al $A \subseteq \mathbb{R}$, $h^{-1}(A^c) = (h^{-1}(A))^c$. Moreover, for all $A_\alpha \in \mathbb{R}$,where $\alpha \in I$,

$$h^{-1}\left(\bigcup_{\alpha\in I} A_\alpha\right) = \bigcup_{\alpha\in I} h^{-1}(A_\alpha),$$
$$h^{-1}\left(\bigcap_{\alpha\in I} A_\alpha\right) = \bigcap_{\alpha\in I} h^{-1}(A_\alpha).$$

4. Suppose $\mathcal{C} \subseteq \mathcal{B}(\mathbb{R})$and $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$. Then, a function is measurable $\iff$ $\forall c \in \mathcal{C}$, $h^{-1}(c) \in \Sigma$.

   Proof. "$\Rightarrow$": trivial.
   "$\Leftarrow$": Assume that $h^{-1}(\mathcal{C}) \subseteq \Sigma$. Since $\mathcal{C}$ generates $\mathcal{B}(\mathbb{R})$, we can write:

$$h^{-1}(\mathcal{B}(\mathbb{R})) = h^{-1}(\sigma(\mathcal{C})) = \sigma(h^{-1}(\mathcal{C})) \subseteq \Sigma.$$

$\square$

   In particular, $h : S \to \mathbb{R}$ is measurable $\iff$ $\forall a \in \mathbb{R}$, $\{h \leq a\} \in \Sigma$.
5. Given $h : S \to \mathbb{R}$ measurable, $f : \mathbb{R} \to \mathbb{R}$ a Borel function, then $(f \circ h) \in m\Sigma$.
6. Given $h_1, h_2 \in m\Sigma$, $h_1 + h_2$, $h_1 - h_2$, $h_1 \cdot h_2$, $\frac{h_1}{h_2}$ (where $h_2 \neq 0$),..., $\in m\Sigma$.

(a) To see that $h_1 + h_2$ is a measurable function: for all $a \in \mathbb{R}$:

$$\begin{aligned}\{h_1 + h_2 > a\} &= \{h_1 > a - h_2\} \\ &= \bigcup_{q \in \mathbb{R}} \{h_1 > q > a - h_2\} \\ &= \bigcup_{q \in Q} \{h_1 > q\} \cap \{h_2 > a - q\}\end{aligned}$$

All those sets are in Σ.

7. Given $\{h_n \mid n \in \mathbb{N}\} \subseteq m\Sigma$, we have:

$$\inf_n h_n,\ \sup_n h_n,\ \liminf_n h_n,\ \limsup_n h_n \in m\Sigma.$$

(a) We only need to prove the first two: to see that $\liminf_n h_n \in m\Sigma$, for all $a \in \mathbb{R}$:

$$\{\inf_n h_n \geq a\} = \bigcap_{n=1}^{\infty} \{h_n \geq a\} \in \Sigma.$$

Note that if we were to consider $\{\inf_n h_n \leq a\}$, then $\{\inf_n h_n \leq a\} \neq \bigcup_{n=1}^{\infty}\{h_n \leq a\}$. The $\supseteq$ is correct, but $\subseteq$ is incorrect. To fix it, push it out by ε.

(b) In particular, $\{\limsup_n h_n = \infty\}, \{\liminf_n h_n = \infty\}, \{\lim h_n \text{ exists }\}, \{\lim_n h_n \text{ exists in } \mathbb{R}\}$ are all in Σ.

Definition 17 (Random Variable)**.** Consider a probability space $(\Sigma, \mathcal{F}, \mathbb{P})$. $X : \Sigma \to \mathbb{R}$ is a **Random Variable** if X is $\mathcal{F}$-measurable, i.e., $X \in m\mathcal{F}$.

Definition 18 (Sigma Algebra Generated by X)**.** Let X be a random variable. The σ-algebra generated by X, denoted by $\sigma(X)$, is:

$$\sigma(X) := \{X^{-1}(\mathcal{B}) \mid B \in \mathcal{B}(\mathbb{R})\}. \tag{12}$$

i.e.: $\sigma(X)$ is the smallest σ-algebra with respect to which the random variable X is measurable.

Heuristically, it's all the information available to us through X.

Example 9. Set $X = \chi_A$ for some $A \in \mathcal{F}$. Then, $\sigma(X) = \{A, A^c, \emptyset, \Omega\}$.

Proposition 12. Remarks:

1. Given $(\Sigma, \mathcal{F}, \mathbb{P})$ and $X : \Omega \to \mathbb{R}$. Then, X is a random variable $\iff \forall\, a \in \mathbb{R}\ \{X \leq a\} \in \mathcal{F}$.
   (a) Note that $\{\{X \leq a\} \mid a \in \mathbb{R}\}$ is the π system generating $\sigma(X)$.
2. Let $\{X_\alpha \mid \alpha \in I\}$ be a family of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the σ-algebra generated by $\{X_\alpha \mid \alpha \in I\}$ is

$$\sigma(\{X_\alpha \mid \alpha \in I\}) = \sigma(\{X_\alpha^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R}), \alpha \in I\}).$$

3. Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Set:

$$\mathcal{P} := \left\{ \bigcap_{j=1}^{k} \{x_{n_j} \leq a_j\} \right\}.$$

Then, $\mathcal{P}$ is a π-system and $\sigma(\mathcal{P}) = \sigma(\{X_n \mid n \in \mathbb{N}\})$.

Definition 19 (Independent Random Variables)**.** Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of random variables $\{X_n \mid n \in \mathbb{N}\}$, $\{X_n \mid n \in \mathbb{N}\}$ is **independent** if $\{\sigma(X_n) \mid n \in \mathbb{N}\}$ is independent.

Proposition 13. $\{X_n \mid n \in \mathbb{N}\}$ are independent $\iff \forall k \geq 1$, $\forall\ 1 \leq n_1 \leq n_2 \leq ... \leq n_k$, $\forall a_1, a_2, ..., a_k \in \mathbb{R}$:

$$\mathbb{P}\left(\bigcap_{j=1}^{k}\{X_{n_j} \leq a_j\}\right) = \prod_{j=1}^{k} \mathbb{P}\left(X_{n_j} \leq a_j\right). \tag{13}$$

Definition 20 (Tail Sigma Algebra)**.** Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of random variables. The **tail sigma-algebra** associated with $\{X_n \mid n \in \mathbb{N}\}$ is defined as:

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma(\{X_n, X_{n+1}, ...\}) \tag{14}$$

This sigma algebra looks at the asymptotic behaviour of X.

We remark that we can think of $\liminf_n X_n$ as a function:

$$\omega \in \Omega \mapsto \liminf_n X_n(\omega).$$

To see that $\liminf_n X_n \in mT$, for all $a \in \mathbb{R}$, $\{\liminf_n X_n \leq a\}$. Note that:

$$\liminf_n r_n = \sup_{n \geq 1} \inf_{m \geq n} r_m.$$

So,

$$\{\liminf_n X_n \leq a\} = \{\sup_{n \geq N} \inf_{m \geq n} X_m \leq a\} \in \sigma(\{X_N, X_{N+1}, ...\})$$

for all $N \geq 1$. Moreover, as in the case of sets, the sets $\{\limsup_n X_n = \infty\}$, $\{\liminf_n X_n = -\infty\}$, $\{\lim X_n \text{ exists}\}$, $\{\lim_n X_n \text{ exists in } \mathbb{R}\}$, ... $\in T$. This makes sense as this is all asymptotic behaviour. In addition, if $S_n := \sum_{j=1}^{n} X_j$ for all $n \in \mathbb{N}$, then S_n is a random variable and $S_n \in m\sigma(\{X_1, ..., X_n\}$. Moreover, given any sequence $\{b_n \mid n \in \mathbb{N}\} \subseteq \mathbb{R}^+$ with $b_n \uparrow \infty$, we have:

$$\limsup_n \frac{S_n}{b_n},\ \liminf_n \frac{S_n}{b_n} \in mT.$$

To see this, note that for all $n \in \mathbb{N}$,

$$\begin{aligned} \limsup_n \frac{S_n}{b_n} &= \limsup_n \frac{1}{b_n}\left(S_N + \sum_{j=N+1}^{n} X_j\right) \\ &= \limsup_n \frac{\sum_{j=N+1}^{n} X_j}{b_n} \in m\sigma(\{X_{N+1}, X_{N+2}, ...\}). \end{aligned}$$

Theorem 10 (Kolmogorov's 0-1 Law)**.** *Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of independent random variables and let $\mathcal{T}$ be the tail sigma-algebra associated with $\{X_n \mid n \in \mathbb{N}\}$. Then, for all $A \in \mathcal{T}$:*

$$\mathbb{P}(A) \in \{0, 1\} \tag{15}$$

If $X \in m\mathcal{T}$, then X is constant a.s., i.e., $\exists$ an $a \in \bar{\mathbb{R}}$ such that $\mathbb{P}(X = a) = 1$.

Proof. We will only prove the second statement. If $X \in mT$, then for all $x \in \mathbb{R}$, $\{X \leq x\} \in T$. Hence,

$$\mathbb{P}(X = x) \in \{0, 1\}.$$

Define: $a := \inf\{x \in \mathbb{R} \mid \mathbb{P}(X \leq x) = 1\}$. Then, there are three possibilities:

1. $a = +\infty$: the jump never happens, and so $\forall M > 0$, $\mathbb{P}(X \leq M) = 0 \Rightarrow \mathbb{P}(X = \infty) = 1$.
2. $a = -\infty$: then, $\mathbb{P}(X \leq -M) = 0 \Rightarrow \mathbb{P}(X = -\infty) = 1$.
3. $a \in \mathbb{R}$: $\forall\, n \geq 1$, $\mathbb{P}\left(X \leq a + \frac{1}{n}\right) = 1 \Rightarrow \mathbb{P}\left(X \leq a - \frac{1}{n}\right) = 0$. This implies that:

$$\mathbb{P}(X = a) = \lim_{n\to\infty} \mathbb{P}\left(a - \frac{1}{n} \leq X \leq a + \frac{1}{n}\right) = 1. \tag{16}$$

□

Example 10. Suppose that $\{X_n\}$ is independent. Set $S_n := \sum_{j=1}^n X_j$. Let $\{b_n\} \subseteq \mathbb{R}^+$ such that $b_n \uparrow \infty$. Then,

$$\limsup_n X_n,\ \liminf_n X_n,\ \limsup_n \frac{S_n}{b_n},\ \liminf_n \frac{S_n}{b_n} \in mT.$$

Then, **(K 0-1 Law)** they are all constant a.s.

There is a horrible proof with combinatorics and boxes which I am omitting.

2.4 Quick Review of Probability

Definition 21 (Law/Distribution)**.** Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$. The **law/distribution** of X, denoted by $\mathcal{L}_X$, id the probability distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$such that for all $B \in \mathcal{B}(\mathbb{R})$:

$$\mathcal{L}_X(B) = \mathbb{P}\left(X^{-1}(B)\right) = \mathbb{P}(X \in B). \tag{17}$$

The **distribution function** of X (of $\mathcal{L}_X$) is:

$$F_x : \mathbb{R} \to [0, 1]$$
$$x \in \mathbb{R} \mapsto F_X(x) := \mathbb{P}(X \leq x) = \mathcal{L}_X(]-\infty, x])$$

Proposition 14 (Properties of F_X)**.**
1. F_X is increasing.
2. $\lim_{x\to+\infty} F_X(x) = 1$.
3. $\lim_{x\to-\infty} F_X(x) = 0$.
4. F_X is **right continuous**: $\forall\, a \in \mathbb{R}$, $F_X(a^+) = \lim_{x\to a^+} F_X(x) = F_X(a)$.
5. For all $a > b$:
   (a) $F_X(a) - F_X(b) = \mathbb{P}(b < X \leq a) = \mathcal{L}_X(]b, a])$.
   (b) $F_X(a^-) - F_X(b) = \mathbb{P}(b < X < a) = \mathcal{L}_X(]b, a[)$.
   (c) $F_X(a) - F_X(a^-) = \mathbb{P}(X = a) = \mathcal{L}_X(\{a\})$.

In particular, if F_X is continuous at a, then $\mathbb{P}(X = a) = 0$.

Example 11. Some examples of distribution functions:

1. X is a uniform random variable on $]a, b[$, $a < b$, if:

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } x \in]a, b[\\ 1 & \text{if } x \geq b. \end{cases}$$

2. X is an exponential random variable with parameter $\lambda > 0$ if:

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-\lambda x} & \text{if } x > 0. \end{cases}$$

3. X is a Guassian random variable with parameters m, $\sigma^2 > 0$ if:

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t-m}{2\sigma^2}}\, dt \text{ for all } x \in \mathbb{R}.$$

Definition 22 (Independent and Identically Distributed)**.** Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of random variables. We say they are **independent and identically distributed (iid)** if $\{X_n \mid n \in \mathbb{N}\}$ is independent and for some probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\mathcal{L}_{X_n} = \mu$ for all $n \in \mathbb{N}$.

CAUTION! The outcomes may not always be the same.

Example 12. Let $\{X_n \mid n \in \mathbb{N}\}$ be iid on $(\Omega, \mathcal{F}, \mathbb{P})$ with common distribution being the exponential distribution with parameter L. Set:

$$L := \limsup_n \frac{X_n}{\ln(n)}. \tag{18}$$

By **(0-1 Law)** we know that L is constant a.s. Now the question is: what is that constant?

Claim: $L = 1$ a.s.

Proof. For all $\alpha > 0$ and for all $n \geq 1$,

$$\mathbb{P}\left(X_n > \alpha \ln(n)\right) = 1 - F_{X_n}(\alpha \ln(n)) = n^{-\alpha}$$

By **(BC1)** and **(BC2)** we have:

$$\mathbb{P}\left(\frac{X_n}{\ln(n)} > \alpha \text{ i.o. }\right) = \begin{cases} 0 & \text{if } \alpha > 1 \\ 1 & \text{if } \alpha \leq 1. \end{cases}$$

In particular, when $\alpha = 1$, we get tat

$$\mathbb{P}\left(\frac{X_n}{\ln(n)} > 1 \text{ i.o.}\right) = 1.$$

For $\omega \in \Omega$, if

$$\frac{X_n(\omega)}{\ln(n)} > 1$$

for infinitely many n's, then $L(\omega) = \limsup_n \frac{X_n(\omega)}{\ln(n)} \geq 1$. Hence:

$$\left\{\frac{X_n}{\ln(n)} \geq 1 \text{ i.o. }\right\} \subseteq \{L \geq 1\} \Rightarrow \mathbb{P}\left(L \geq 1\right) = 1.$$

On the other hand, if:

$$\begin{aligned} \{L > 1\} &= \bigcup_{k=1}^{\infty} \left\{L \geq 1 + \frac{1}{k}\right\} \\ &\subseteq \bigcup_{k=1}^{\infty} \left\{\frac{X_n}{\ln(n)} > 1 + \frac{1}{2k} \text{ i.o. }\right\}. \end{aligned}$$

But, this is a null set for all $k \geq 1$. Hence, $\mathbb{P}\left(L > 1\right) = 0 \Rightarrow \mathbb{P}\left(L = 1\right) = 1.$ $\square$

2.5 Convergence of Random Variables

Definition 23 (Almost Sure Convergence)**.** Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{X_n\}$ a sequence of random variables.

1. We say that X_n converges to X **almost surely** and we write $X_n \to X$ a.s. (as $n \to \infty$) if:

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{19}$$

   If X_n and X are $\mathbb{R}$-valued, this is equivalent to:

$$\mathbb{P}\left(\lim_{n\to\infty} |X_n - X| = 0\right) = 1. \tag{20}$$

2. We write that "$X_n \to \infty$" a.s. if $\mathbb{P}\left(\lim_{n\to+\infty} X_n = 0\right) = 1$.

We also call this **pointwise convergence**.

Proposition 15. Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of $\mathbb{R}$-valued random variables and X be a $\mathbb{R}$-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then,

1. $X_n \to X$ a.s. $\iff \forall\, \varepsilon > 0$,

$$\mathbb{P}\left(|X_n - X| \le \varepsilon \text{ e.a.}\right) = 1 \iff \forall\, \varepsilon > 0\ \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0. \tag{21}$$

   This converts pointwise behaviour to something I can look at probabilistically.
2. $X_n \to \infty$ a.s. $\iff$

$$\forall M > 0\ \mathbb{P}\left(X_n \ge M \text{ e.a.}\right) = 1 \iff \forall M > 0\ \mathbb{P}\left(X_n < M \text{ i.o.}\right) = 0. \tag{22}$$

Proof. We will prove the first equivalence.
"$\Rightarrow$": Assume that $X_n \to X$ a.s. Write the set $\{\lim_{n\to\infty} X_n = X\}$ in terms of unions and intersections:

$$\left\{\lim_{n\to\infty} X_n = X\right\} = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{|X_n - X| \le \frac{1}{k}\right\}. \tag{23}$$

Notice that for all $\varepsilon > 0$, we can choose some $k_0 > 1$ such that $\frac{1}{k_0} < \varepsilon$. Then, we have the following set inclusions:

$$\left\{\lim_{n\to\infty} X_n = X\right\} \subseteq \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{|X_m - X| \le \frac{1}{k_0}\right\} \subseteq \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{|X_n - X| \le \varepsilon\} = \liminf \{|X_n - X| \le \varepsilon\} \tag{24}$$

Since $\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$, $\mathbb{P}\left(\liminf_{n\to\infty} \{|X_n - X| \le \varepsilon\}\right) = 1$, which is what we wanted to show. □

Definition 24 (Convergence in Probability)**.** We say that "X_n converges to X" **in probability**, we write "$X_n \to X$" in prob as $n \to \infty$ if $\forall\, \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0 \tag{25}$$

Similarly, we write "$X_n \to \infty$" in probability if $\forall M > 0$, $\lim_{n\to\infty} \mathbb{P}\left(X_n \le M\right) = 0$.

What is the relationship between these two modes of convergence?

Proposition 16. $X_n \to X$ a.s. $\Rightarrow X_n \to X$ in probability.

Proof. Assume that $X_n \to X$ a.s. Then, for all $\varepsilon > 0$, $\mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$. But recall that:

$$\limsup_n \mathbb{P}(|X_n - X| > \varepsilon) \le \mathbb{P}\left(\limsup_n \{|X_n - X| > \varepsilon\}\right) = 0.$$

□

The proof relies on the following lemma.

Lemma 2. Let $\{A_n \mid n \in \mathbb{N}\} \subseteq \mathcal{F}$. Then,

$$\mathbb{P}\left(\liminf_n A_n\right) \le \liminf_n \mathbb{P}(A_n) \text{ and } \mathbb{P}\left(\limsup_n A_n\right) \ge \limsup_n \mathbb{P}(A_n). \tag{26}$$

Proof. Set $B_n := \bigcap_{n=m}^{\infty} A_n$. Then, the B_n form an increasing sequence of sets. By construction ,for all $m \ge n$, $\mathbb{P}(B_n) \le \mathbb{P}(A_n)$, we have that $\mathbb{P}(B_n) \le \inf_{m \ge n} \mathbb{P}(A_n)$.

$$\begin{aligned}\mathbb{P}\left(\liminf_n A_n\right) &= \lim_{n\to\infty} \mathbb{P}(B_n) \\ &\le \liminf_{n\to\infty, m\ge n} \mathbb{P}(A_n) \\ &= \liminf_n \mathbb{P}(A_n).\end{aligned}$$

□

Remark. Convergence in probability in general does not imply convergence a.s. Consider the following example:

$$X_n : \Omega \to \{0, 1\}$$

with

$$\begin{aligned}\mathbb{P}(X_n = 0) &= 1 - \frac{1}{n} \\ \mathbb{P}(X_n = 1) &= \frac{1}{n}.\end{aligned}$$

Further, assume that the sequence is independent. Then, for all $0 < \varepsilon < 1$:

$$\lim_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) = \lim_{n\to\infty} \mathbb{P}(|X_n| = 1) = \lim_{n\to\infty} \frac{1}{n} = 0.$$

This implies that $X_n \to 0$ in probability. However, $\sum_{n=1}^{\infty} \mathbb{P}(X_n = 1) = \infty$. By Borel-Cantelli II, this means that $\mathbb{P}(X_n = 1 \text{ i.o}) = 1$. This means that $X_n \nrightarrow 0$ a.s.

Proposition 17. If $X_n \to X$ in probability, then there exists a subsequence $\{n_k \mid k \in \mathbb{N}\} \subseteq \mathbb{N}$ such that $X_{n_k} \to X$ a.s. as $k \to \infty$.

Proof. Suppose $X_n \to X$ in probability. Then, for all $k \in \mathbb{N}$, this means that $\lim_n \mathbb{P}\left(|X_n - X| > \frac{1}{k}\right) = 0$. This means we can construct a subsequence $\{n_k | k \in \mathbb{N}\}$ such that $\mathbb{P}\left(|X_{n_k} - X| > \frac{1}{k}\right) \le \frac{1}{k^2}$. Since $\frac{1}{k^2}$ is summable, by Borel-Cantelli Lemma 1,

$$\mathbb{P}\left(|X_{n_k} - X| > \frac{1}{k} \text{ i.o.}\right) = 0.$$

For all $\varepsilon > 0$, for k sufficiently large, $\frac{1}{k} < \varepsilon$. Hence, we have the following set inclusion:

$$\{|X_{n_k} - X| > \varepsilon \text{ i.o.}\} \subseteq \left\{|X_{n_k} - X| > \frac{1}{k} \text{ i.o.}\right\}.$$

By monotonicity of probability,

$$\mathbb{P}\left(|X_{n_k} - X| > \varepsilon \text{ i.o.}\right) = 0.$$

This proves that $X_{n_k} \to X$ a.s. as $k \to \infty$. $\square$

Proposition 18. $X_n \to X$ in probability $\iff$ for all subsequences $\{n_k \mid k \in \mathbb{N}\}$, there exists a subsequence $\{n_{k_l} \mid l \in \mathbb{N}\}$ such that $X_{n_{k_l}} \to X$ in probability as $l \to \infty$.

Proof. "$\Rightarrow$": Trivial.
"$\Leftarrow$": For a contradiction, assume that $X_n \nrightarrow X$ in probability. Then, there exists a $\varepsilon > 0$ such that for some $\delta > 0$, there exists a subsequence $\{n_k \mid k \in \mathbb{N}\}$ such that:

$$\mathbb{P}\left(|X_{n_k} - X| > \varepsilon\right) > \delta \ \forall\ k \in \mathbb{N}. \tag{27}$$

However, according to the assumption/hypothesis, there exists a subsequence of this, $\{n_{k_l} \mid l \in \mathbb{N}\}$, such that $X_{n_{k_l}} \to X$ in probability, which is a contradiction. Hence, $X_n \to X$ in probability. $\square$

Proposition 19. If $X_n \to X$ in probability, and if $f : \mathbb{R} \to \mathbb{R}$ is continuous, then $f(X_n) \to f(X)$ in probability.

Proof. Since $X_n \to X$ in probability, we have by the previous proposition that for all subsequences $\{n_k \mid k \in \mathbb{N}\}$ $X_{n_k} \to X$ in probability as $k \to \infty$. By an earlier proposition, we know that there exists a previous subsequence $X_{n_{k_l}} \to X$ a.s. as $l \to \infty$. With almost sure / pointwise behaviour, we can use the continuity of functions. This gives us that $f(X_{n_{k_l}}) \to f(X)$ almost surely as $l \to \infty$. Hence, $f(X_{n_{k_l}}) \to f(X)$ in probability. Since the mother subsequence X_{n_k} was arbitrary, this tells us that $f(X_n) \to f(X)$ in probability. $\square$

3 Constructing Integrals on General Measure Spaces

3.1 Preparation

We want to construct integrals of $h \in m\Sigma$. Given $h : S \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$, $h \in m\Sigma$.

1. Set $h = h^+ - h^-$, where

$$h^+ := \max\{h, 0\} \text{ and } h^- := \max\{-h, 0\},$$

where both h^+ and h^- are non-negative and measurable with respect to Σ, i.e., $h^\pm \in (m\Sigma)^+$
2. Suppose $h \in (m\Sigma)^+$. For each $k > 0$, set $h_k := \min\{h, k\}$. In other words, for all $s \in S$:

$$h_k(s) = \begin{cases} h(s) & \text{if } h(s) \le k \\ k & \text{if } h(s) > k \end{cases}$$

As $k \uparrow$, $h_k \uparrow h$. We are essentially approximating h with a sequence of bounded, non-negative functions. For each $k > 0$, h_k is non-negative and bounded, i.e., $h \in (m\Sigma)^+ \cap b\Sigma$.
3. Suppose $h \in (m\Sigma)^+ \cap b\Sigma$. Suppose that h is bounded by some value k. For each $n \ge 1$ and for every $i = 0, 1, ..., 2^n k$, set:

$$A^h(n, i) := \{s \in S \mid i2^{-n} \le h(s) \le (i+1)2^{-n}\} = h^{-1}([i2^{-n}, (i+1)2^{-n}])$$

Clearly these sets are disjoint for different indices. Define:

$$h_n := \sum_{i=0}^{2^n k} \chi_{A^n(n,i)} i2^{-n}. \tag{28}$$

i.e., $s \in A^h(n, i)$ then $h_n(s) = i2^{-n}$. For each $n \in \mathbb{N}$, h_n is non-negative, bounded, and simple (it's a linear combination of indicator functions). As $n \to \infty$ $h_n \uparrow h$.

The combination of (1), (2), and (3), $h \in m\Sigma$ means that $h = h^+ - h^-$ with $h^+ \in (m\Sigma)^+$.

Theorem 11 (Monotone Class Theorem)**.** *Let H be a class of bounded functions on some space S satisfying:*

1. *H is a vector space over $\mathbb{R}$.*
2. *the constant function $1 \in H$.*
3. *if $h_n \in H$ for $n \in \mathbb{N}$ such that $h_n \geq 0$ and $h_n \uparrow h$ for some bounded function h on S, then $h \in H$. We call this **closed under monotonic convergence**.*

*If H satisfies these properties, we call H a **monotone class**. If I is a π-system of subsets of S and for all $A \in I$, $\chi_A \in H$, then $b\sigma(I) \subseteq H$. In other words, if f is bounded and measurable with respect to $\sigma(I)$, then $f \in H$.*

Proof. General advice: when you see a π-system, try to make a d-system out of it. Let $D := \{F \subseteq S \mid \chi_F \in H\}$. To see this, we need to check the rules.

1. $S \in D$ because $1 \in H$ by (2).
2. if $A, B \in D$ and if $A \subseteq B$ then $B \setminus A \in D$ because:

$$\chi_{B\setminus A} = \chi_B - \chi_A \in H \text{ (by being a vector space)} \tag{29}$$

3. if $A_n \in D$ for $n \in \mathbb{N}$ and $A_n \uparrow A$, then $A \in D$ because $\chi_{A_n} \uparrow \chi_A$. Since monotonic classes are closed under monotonic convergence, this means that $\chi_A \in H$.

This shows that D is a d-system. Since $I \subseteq D$, by the π-d theorem, $\sigma(I) \subseteq D$. This shows that for all $B \in \sigma(I)$, $\chi_B \in H$.

Next, given any $h \in b\sigma(I)$, we apply steps (I)-(III) to h. There exists a sequence $h_n^\pm \in SF^+$ such that $h_n^\pm \uparrow h^\pm$. We have that $h^\pm \leq k$ and we can write it as:

$$h_n^\pm = \sum_{i=0}^{2^n k} i2^{-n} \underbrace{\chi_{\{i2^{-n} \leq h^\pm \leq (i+1)2^{-n}\}}}_{\in\sigma(I)} \in H \tag{30}$$

Since H is a monotone class, H is closed under limits. Hence, $h^\pm \in H$. Since H is a monotone class, it's a vector space, and so $h = h^+ - h^- \in H$. $\square$

Theorem 12 (Monotone Class Theorem – General Measurable Functions)**.** *Let H be a monotone class of (general) $\mathbb{R}$-valued functions on S. If I is a π-system and for all $A \in I$, $\chi_A \in H$, then $m\sigma(I) \subseteq H$.*

Proof. Almost identical to the one above, except we will apply steps (I)-(III) to a given function $h \in m\sigma(I)$. $\square$

The following proposition is a useful application of the MCT.

Proposition 20. Given (S, Σ_1) and (S, Σ_2) two measurable spaces. Then, $X : S_1 \to S_2$ and $Y : S_1 \to \mathbb{R}$. Assume that X is $\Sigma_1 \setminus \Sigma_2$-measurable.[1] Then, $Y \in m\sigma(X_1) \iff$ there exists an $f : S_2 \to \mathbb{R}$ such that $f \in m\Sigma_2$ such that $Y = f(X)$.

Proof. "$\Leftarrow$": trivial.
"$\Rightarrow$": Set:

$$H := \{Y : S_1 \to \mathbb{R} \mid \exists\, f \in m\Sigma_2 \text{ s.t. } Y = f(X)\}. \tag{31}$$

Claim: H is a monotone class of general functions. We need to check the conditions:

[1] i.e. for all $B \in \Sigma_2$, $X^{-1}(B) \in \Sigma_1$. $\sigma(X) = \{X^{-1}(B) \mid B \in \Sigma_2\}$ is a σ-algebra of subsets of S_1.

1. H is clearly a vector space.
2. $1 \in H$, just set $f \equiv 1$.
3. Since the limit may not exist, we need to go with the limsup. If $Y_n \in H$, $Y_n \geq 0$, $Y_n \uparrow Y$, then there exists an $f_n \in m\Sigma_2$ such that $Y_n = f_n(x)$. Just:
$$Y = \lim_{n\to\infty} f_n(x) = (\limsup_n f_n)(x). \tag{32}$$
$Y \in H$.

Moreover, for every $A \in \sigma(X)$ (a π-system), there exists a $B \in \Sigma_2$ such that $A = X^{-1}(B)$. By the monotone class theorem, $m\sigma(X) \subseteq H$. $\square$

We now start the construction of the integral.

Notation. Given (S, Σ, μ) and $h \in m\Sigma$, then the integral of h with respect to μ is denoted by $\mu(h)$. We write:
$$\mu(h) = \int_S h d\mu = \int_S h(s)\mu(ds). \tag{33}$$

For some $B \in \Sigma$,
$$\mu(\chi_B h) = \int_B h d\mu = \int_B h(s)\mu(ds). \tag{34}$$

Definition 25. Given $h \in SF^+$, assume that $h = \sum_{k=1}^n a_k \chi_{A_k}$ where all the A_k are measurable and $a_k \in [0, \infty]$ for every $k = 1, ..., n$. Define:
$$\mu(h) = \sum_{k=1}^n a_k \mu(A_k). \tag{35}$$

Remarks.

1. One should verify that $\mu(h)$ is well-defined, i.e., if $h = \sum_{k=1}^n a_k \chi_{A_k} = \sum_{l=1}^m b_l \chi_{B_l}$ for some $B_l \in \Sigma$, $b_l \in [0, \infty]$, then:
$$\sum_{k=1}^n a_k \mu(A_k) = \sum_{l=1}^n b_k \mu(B_l). \tag{36}$$
2. If $h_1, h_2 \in SF^+$ and if $\mu(h_1 \neq h_2) = 0$ then $\mu(h_1) = \mu(h_2)$. To see this, *do not* take $h_1 - h_2$!
$$h_1 = \sum_{k=1}^n a_k \chi_{A_k}$$
$$h_2 = \sum_{l=1}^n b_l \chi_{B_l}.$$
Then,
$$\mu(h_1) = \underbrace{\sum_{k=1}^n a_k \mu(A_k \cap \{h_1 \neq h_2\})}_{=0} + \sum_{k=1}^n a_k \mu(A_k \cap \{h_1 = h_2\})$$

We have:

$$\begin{aligned} h_2 &= h_2 \cdot \chi_{\{h_1 \neq h_2\}} + h_1 \cdot \chi_{\{h_1 = h_2\}\}} \\ &= \sum_{l=1}^{m} b_l \chi_{\{B_l \cap \{h_1 \neq h_2\}} + \sum_{k=1}^{n} a_k \chi_{\{A_k \cap \{h_1 = h_2\}\}}. \end{aligned}$$

This gives us:

$$\mu(h_2) = \underbrace{\sum_{l=1}^{m} b_l \mu(B_l \cap \{h_1 \neq h_2\})}_{=0} + \sum_{k=1}^{n} a_k \mu(A_k \cap \{h_1 = h_2\}).$$

This shows that they are equal.

3. Let h_1, h_2 be simple functions. Then:
   (a) $h_1 \vee h_2 = \max\{h_1, h_2\} \in SF^+$.
   (b) $h_1 \wedge h_2 = \min\{h_1, h_2\} \in SF^+$.
4. Given $h \in SF^+$, write $h = \sum_{k=1}^{n} A_k \chi_{A_k}$. Whenever convenient, necessary, or helpful, we assume that A_k is a partition of S, i.e., $\bigcup_{k=1}^{n} A_k = S$ and $A_k \cap A_{k'} = \emptyset$ whenever $k \neq k'$.

Properties of $\mu(h)$ for SF^+

1. **Linearity**: given $h_1, h_2 \in SF^+$, $c_1, c_2 \in [0, \infty]$:

$$\mu(c_1 h_1 + c_2 h_2) = c_1 \mu(h_1) + c_2 \mu(h_2). \tag{37}$$

2. **Monotonicity**: given $h_1, h_2 \in SF^+$, if $h_1 \leq h_2$, then $\mu(h_1) \leq \mu(h_2)$.

Definition 26 (Integral for General Non-Negative Measurable Functions)**.** Given $f \in (m\Sigma)^+$. Define:

$$\mu(f) := \sup\{\mu(h) \mid h \in SF^+ \; h \leq f\} \tag{38}$$

If $f \in SF^+ \subseteq (m\Sigma)^+$, say, $f = \sum_{k=1}^{n} \chi_{A_k} a_k$. Previously, we had that the integral was:

$$\mu(f) = \sum_{k=1}^{n} a_k \mu(A_k). \tag{39}$$

We ask ourselves: what is the relation between (38) and (39)? **Claim:** for $f \in SF^+$, it turns out that (38) = (39).

Proof. Since $f \in SF^+$, we clearly get (38) $\leq$ (39). Meanwhile, for all $h \in SF^+$ such that $h \leq f$, by monotonicity, $\mu(h) \leq \mu(f)$. This observation gives us the other inequality, (39) $\leq$ (38). □

The following theorem gives us a baby version of the monotonic convergence theorem.

Theorem 13. *Suppose $f \in (m\Sigma)^+$ and $h_n \in SF^+$ for $n \in \mathbb{N}$. Suppose that $h_n \uparrow f$. Then,*

$$\mu(f) = \lim_{n \to \infty} \mu(h_n) \text{ i.e. } \mu(h_n) \uparrow \mu(f). \tag{40}$$

Proof. To do. □

Properties of $\mu(f)$ for $f \in (m\Sigma)^+$

1. If $f \in (m\Sigma)^+$ and $\mu(f) = 0$, then $f = 0$ a.e.

Proof. Assume otherwise, i.e., $\mu(\{f > 0\}) = \delta > 0$. Since $\{f_n > \frac{1}{n}\} \uparrow \{f > 0\}$, there exists some $n \in \mathbb{N}$ such that

$$\mu\left(\left\{f > \frac{1}{n}\right\}\right) > \frac{\delta}{2} > 0 \tag{41}$$

Set $h = \frac{1}{n}\chi_{\{f > \frac{1}{n}\}}$. Then, $h \in SF^+$ and $h \leq f$. Hence, by monotonicity,

$$\mu(f) \geq \mu(h) = \frac{1}{n}\mu\left(\left\{f > \frac{1}{n}\right\}\right) \geq \frac{1}{n}\frac{\delta}{2} > 0$$

□

2. If $f \in (m\Sigma)^+$ and $\mu(f) < \infty$, then $f < \infty$ a.e. The proof of this one is an exercise.
3. If $f, g \in (m\Sigma)^+$ and $f = g$ a.e., then $\mu(f) = \mu(g)$.

   Proof. Reminder: do not take $f - g$! Take $f_n, g_n \in SF^+$ such that $f_n \uparrow f$ and $g_n \uparrow g$. Then, for every $n \in \mathbb{N}$,

   $$f_n\chi_{\{f=g\}} + g_n\chi_{\{f\neq g\}} \uparrow g$$

   Hence,

   $$\begin{aligned}\mu(g) &= \lim_{n\to\infty} \mu(f_n\chi_{\{f=g\}} + g_n\chi_{\{f\neq g\}}) \\ &= \lim_{n\to\infty} \mu(f_n\chi_{\{f=g\}}) + \underbrace{\lim_{n\to\infty} \mu(g_n\chi_{\{f\neq g\}})}_{=0}\end{aligned}$$

   Similarly, $\mu(f) = \lim_{n\to\infty} \mu(f_n\chi_{\{f=g\}})$. Hence, $\mu(f) = \mu(g)$. □

4. **Linearity:** let $f, g \in SF^+$. $c \in [0, \infty]$. Then:
   (a) $\mu(f + g) = \mu(f) + \mu(g)$.
   (b) $c\mu(f) = \mu(cf)$.

   Proof. The proof follows from baby MON. Take $f_n, g_n \in SF^+$ such that $f_n \uparrow f$ and $g_n \uparrow g$. Note that $f_n + g_n \uparrow f + g$ and $cf_n \uparrow cf$. Then,

   $$\mu(f + g) = \lim_n \mu(f_n + g_n) + \lim_n \mu(f_n) + \lim_n \mu(g_n) = \mu(f) + \mu(g).$$

   □

5. **Monotonicity:** if $f, g \in (m\Sigma)^+$, $f \leq g$. Then, $\mu(f) \leq \mu(g)$.

   Proof. For all $h \in SF^+$, $h \leq g$ implies that $h \leq f$, obviously. Then,

   $$\mu(f) = \sup\{\mu(h) \mid h \in SF^+,\ h \leq f\} \leq \sup\{\mu(h) \mid h \in SF^+,\ h \leq g\} = \mu(g).$$

   □

Now all that's left to do is define $\mu(f)$ for $f \in m\Sigma$.

Definition 27. Given (S, Σ, μ) and $f \in m\Sigma$. If at least one of $\mu(f^+)$, $\mu(f^-)$ is finite, then we define the integral as:

$$\mu(f) := \mu(f^+) - \mu(f^-). \tag{42}$$

Note that the existence of $\mu(f)$ is a quite "weak" property: e.g. if $\mu(f)$ and $\mu(g)$ exist, in general we may not be able to conclude that $\mu(f+g) = \mu(f) + \mu(g)$.

Definition 28 (Integrable Functions)**.** Given (S, Σ, μ), $f \in m\Sigma$. f is called μ-**integrable**, denoted by $f \in L^1(S, \Sigma, \mu)$ if $\mu(f^+) < \infty$ and $\mu(f^-) < \infty$, or equivalently, $\mu(|f|) < \infty$, or equivalently, $\mu(f^+) - \mu(f^-) \in \mathbb{R}$.

Properties of $\mu(f)$ for $f \in L^1(S, \Sigma, \mu)$

1. If $f \in L^1(S, \Sigma, \mu)$, then $f \in \mathbb{R}$ almost everywhere.
2. If $f, g \in L^1(S, \Sigma, \mu)$, and if $c \in \mathbb{R}$, then $f + g \in L^1(\mu)$, $cf \in L^1(\mu)$, and $\mu(f+g) = \mu(f) + \mu(g)$ and $\mu(cf) = c\mu(f)$.

   Proof. We will prove that $\mu(f+g) = \mu(f) + \mu(g)$. Set $h = f + g$. Since $|h| \le |f| + |g|$, $h \in L^1(\mu)$. Next, write out the positive and negative parts of h:

   $$h = h^+ - h^- = f^+ - f^- + g^+ - g^-.$$

   Note that with the integrability conditions, $h^\pm, g^\pm, f^\pm$ are almost everywhere finite. Re-arrange the above as:

   $$h^+ + f^- + g^- = f^+ + g^+ + h^-$$

   Apply linearity of integrals for $(m\Sigma)^+$, and use the fact that each integral is finite:

   $$\begin{aligned} &\mu(h^+) + \mu(f^-) + \mu(g^-) = \mu(f^+) + \mu(g^+) + \mu(h^-) \\ \Rightarrow &\mu(h^+) - \mu(h^-) = \mu(f^+) - \mu(f^-) + \mu(g^+) - \mu(g^-). \\ \Rightarrow &\mu(h) = \mu(f) + \mu(g). \end{aligned}$$

   This proves linearity. □

3. If $f, g \in L^1(S, \Sigma, \mu)$ and $f \le g$ almost everywhere, then $\mu(f) \le \mu(g)$.

3.2 Integral Convergence Theorems

Theorem 14 (Monotone Convergence Theorem)**.** *Given (S, Σ, μ), let $\{f_n\}_{n \in \mathbb{N}} \subseteq m\Sigma$ such that $f_n \uparrow f$ (this implies that $f \in m\Sigma$). Assume that $\mu(f_1^-) < \infty$ (Equivalently, $\mu(f_1)$ exists and $\mu(f_1) > -\infty$). Then,*

$$\mu(f_n) \uparrow \mu(f) \iff \lim_{n\to\infty} \mu(f_n) = \mu(\lim_{n\to\infty} f_n). \tag{43}$$

Proof. Since $f_n \uparrow f$ and $\mu(f_1^-) < \infty$, then $f_1^- < \infty$ almost everywhere. Hence, by the monotonicity of the sequence, $f_n \le f_n^- \le f_1^- < \infty$ and so by monotonicity of the integrals:

$$\mu(f^-) \le \mu(f_n^-) \le \mu(f_1^-) < \infty.$$

First Step: Assume that $f_1 \ge 0$. Then, $f \ge f_n \ge f_1 \ge 0$. Then, for each $n \ge 1$, take $\{h_{n,m} \mid m \in \mathbb{N}\} \subseteq SF^+$ such that $h_{n,m} \uparrow f_n$ as $m \to \infty$.

$$\begin{array}{ccccccc} f_1 & f_2 & f_3 & \cdots & f_n & \cdots & f \\ \vdots & & & \vdots & \vdots & & \vdots \\ h_{1,m} & h_{2,m} & h_{3,m} & \cdots & h_{n,m} & & \\ \vdots & \vdots & & & & & \\ h_{1,3} & h_{2,3} & & & & & \\ h_{1,2} & h_{2,2} & & & & & \\ h_{1,1} & h_{2,1} & & & & & \end{array}$$

For each $m \geq 1$, set $g_m := \max\{h_{1m}, h_{2m}, ..., h_{n,m}\}$. This is going to be a diagonalization argument. For each m, we have that $g_m \in SF^+$, $g_{m+1} \geq g_m$, and $g_m \uparrow f$. Why? Because for each $n \geq 1$:

$$f \geq \lim_{m\to\infty} g_m \geq \lim_{m\to} h_{n,m} = f_n. \tag{44}$$

Hence,

$$f \geq \lim_{m\to\infty} g_m \geq \lim_{n\to\infty} f_n = f \Rightarrow f = \lim_{n\to\infty} g_n.$$

This gives us that $\mu(g_n) \uparrow \mu(f)$ as $n \to \infty$.

Second Step: since $h_{1m}, h_{2m}, ..., h_{nm} \leq f_m$, we get that $g_m \leq f_m$. By monotonicity, $\mu(g_m) \leq \mu(f_m)$. We then obtain:

$$\mu(f) = \lim_{m\to\infty} \mu(g_m) \leq \lim_{n\to\infty} \mu(f_n) \leq \mu(f).$$

So, by the squeeze theorem, $\mu(f_n) \uparrow \mu(f)$ almost everywhere as $n \to \infty$.

Third Step: Set $\tilde{f}_n := f_n + f_1^-$. Then, $\tilde{f}_n \geq 0$. Set $\tilde{f} := f + f_1^-$. Then, $\tilde{f} \geq 0$. Then, $\tilde{f}_n \uparrow \tilde{f}$. Hence, $\mu(\tilde{f}_n) \uparrow \mu(\tilde{f})$. Now write it all out:

$$\tilde{f}_n = f_n + f_1^- = f_n^+ - f_n^- + f_1^- \Rightarrow \tilde{f}_n + f_n^- = f_n^+ + f_1^-$$

Taking the integral of both sides,

$$\mu(\tilde{f}_n) + \mu(f_n^-) = \mu(f_n^+) + \mu(f_1^-) \Rightarrow \mu(\tilde{f}_n) = \mu(f_n) + \mu(f_1^-).$$

Similarly, $\mu(\tilde{f}) = \mu(f) + \mu(f_1^-)$. Thus, $\mu(f_n) \uparrow \mu(f)$. $\square$

Theorem 15 (Monotone Convergence Theorem'). *Given a measure space (S, Σ, μ) and $\{f_n \mid n \in \mathbb{N}\} \subseteq m\Sigma$. If $f_n \downarrow f$ for some $f \in m\Sigma$ and $\mu(f_1^+) < \infty$, then $\mu(f)$ is defined an $\mu(f_n) \downarrow \mu(f)$.*

Proof. Consider $\overline{f}_n := f_1^+ - f_n$, and $\overline{f} = f_1^+ - f$. Then, $\overline{f}_n, f \in (m\Sigma)^+$, and $\overline{f}_n \uparrow \overline{f}$. By the previous theorem, Monotone Convergence Theorem, $\mu(\overline{f}_n) \uparrow \mu(\overline{f})$. Verify that $\mu(\overline{f}_n) = \mu(f_1^+) - \mu(f_n)$ and $\mu(\overline{f}) = \mu(f_1^+) - \mu(f)$. Therefore, $\mu(f_n) \downarrow \mu(f)$. $\square$

Remark that MON and MON', $f_n \uparrow f$ or $f_n \downarrow f$ can be replaced by $f_n \uparrow f$ almost everywhere or $f_n \downarrow f$ almost everywhere.

Lemma 3 (Fatou's Lemma). Given (S, Σ, μ) and $\{f_n\} \subseteq m\Sigma$, if there exists a $g \in m\Sigma$ such that $\mu(g^-) < \infty$ and $f_n \geq g$ for all $n \in \mathbb{N}$. Then,

$$\mu(\liminf_n f_n) \leq \liminf_n \mu(f_n) \tag{45}$$

Proof. For every $n \in \mathbb{N}$, set $g_n := \inf_{m\geq n} f_m$. Then, $g_n \uparrow \liminf_n f_n$. We want to use the Monotone Convergence Theorem; since $g_1 \geq g$, $\mu(g_1^-) \leq \mu(g^-) < \infty \Rightarrow$ I have a floor. By the Monotone Convergence Theorem, $\mu(g_n) \uparrow \mu(\liminf_n f_n)$. Meanwhile, $g_n + g \geq 0$ and $g_n + g \leq f_m + g^-$ for all $m \geq n$. Hence, by monotonicity for integrals of non-negative functions,

$$\mu(g_n + g) \leq \mu(f_m + g^-) \text{ for all } m \geq n.$$

Verify that $\mu(g_n + g^-) = \mu(g_n) + \mu(g^-)$ and $\mu(f_m + g^-) = \mu(f_n) + \mu(g^-)$. This gives us that:

$$\mu(g_n) \leq \inf_{m\geq n} \mu(f_m) \Rightarrow \mu(\liminf_n f_n) = \lim_n \mu(g_n) \leq \liminf_n \mu(f_n).$$

$\square$

Theorem 16 (Fatou'). *If $\{f_n \mid n \in \mathbb{N}\} \subseteq m\Sigma$ and there exists a $g \in m\Sigma$ such that $\mu(g^+) < \infty$ and $f_n \leq g$ for all $n \in \mathbb{N}$, Then:*

$$\mu(\limsup_n f_n) \geq \limsup_n \mu(f_n). \tag{46}$$

Theorem 17 (Dominated Convergence Theorem (DOM)). *Suppose $\{f_n \mid n \in \mathbb{N}\} \subseteq m\Sigma$ and $g \in L^1(S, \Sigma, \mu)$ and $|f_n| \leq |g|$ for all $n \in \mathbb{N}$. If $f_n \to f$ for some $f \in m\Sigma$ (i.e., for all $s \in S$, $f(s) = \lim_n f_n(s)$), then $f_n \to f$ in $L^1(S, \Sigma, \mu)$, then $f_n \to f$ in $L^1(S, \Sigma, \mu)$, i.e., $\lim_{n\to\infty} \mu(|f_n - f|) = 0$ and in particular,*

$$\lim_{n\to\infty} \mu(f_n) = \mu(f). \tag{47}$$

Proof. Since $g \in L^1$, $|f_n| \leq |g|$ for all $n \in \mathbb{N}$. By monotonicity, $\mu(|f_n|) \leq \mu(|g|) < \infty$. Hence, $f_n \in L^1$ for each $n \in \mathbb{N}$. Moreover, $|f| = \lim_{n\to\infty} |f_n|$. By Fatou's Lemma, $\mu(|f|) \leq \liminf_{n\to\infty} \mu(|f_n|) < \infty \Rightarrow f \in L^1$. Observe that $|f_n - f| \leq 2|g|$. By Fatou's Lemma Prime,

$$0 = \mu(\limsup_n |f_n - f|) \geq \limsup_n \mu(|f_n - f|).$$

This establishes that $\lim_n \mu(|f_n - f|) = 0$, i.e., $f_n \to f$ in $L^1(\mu)$. By the triangle inequality,

$$\begin{aligned} |\mu(f_n) - \mu(f)| &= |\mu(f_n - f)| \\ &\leq \mu(|f_n - f|) \to 0. \end{aligned}$$

□

Note that the existence of $g \in L^1$ such that $|f_n| \leq |g|$ for all $n \in \mathbb{N}$ is necessary to apply dominated convergence theorem. To see why, consider:

$$f_n = \chi_{[n,2n]}.$$

Then, $f_n \geq 0$ for all $n \in \mathbb{N}$, and $\lim_n f_n = 0$. However, for λ the Lebesgue measure, for every $n \in \mathbb{N}$, $\mu(f_n) = n$. Hence,

$$\lambda(f_n) \nrightarrow \lambda(\lim_n f_n).$$

Lemma 4 (Scheffé's Lemma). Suppose we have a sequence $\{f_n \mid n \in \mathbb{N}\} \subseteq L^1(S, \Sigma, \mu)$, $f \in L^1(S, \Sigma, \mu)$ and $f_n \to f$. Then,

$$f_n \to f \text{ in } L^1(S, \Sigma, \mu) \iff \lim_{n\to\infty} \mu(|f_n|) = \mu(|f|). \tag{48}$$

Proof. "⇒": Using the inequality, $||f_n| - |f|| \leq |f_n - f|$:

$$\begin{aligned} |\mu(|f_n|) - \mu(|f|)| &= |\mu(|f_n| - |f|)| \\ &\leq \mu(||f_n| - |f||) \\ &\leq \mu(|f_n - f|) \to 0. \end{aligned}$$

"⇐": Let $g_n := |f_n| + |f| - |f_n - f|$ for all $n \in \mathbb{N}$. Then, $g_n \in (m\Sigma)^+$ and $g_n \in L^1(S, \Sigma, \mu)$. By Fatou's lemma,

$$\mu(\liminf_n g_n) \leq \liminf_n \mu(g_n).$$

Hence,

$$
\begin{aligned}
\underbrace{\mu(\liminf_n(|f_n|+|f|-|f_n-f|))}_{2|f|} &\le \underbrace{\liminf_n(\mu(|f_n|)+\mu(|f|)-\mu(|f_n-f|))}_{2\mu(|f|)-\limsup_n \mu(|f_n-f|)} \\
\Rightarrow & 2\mu(|f|) \le 2\mu(|f|) - \limsup_n(|f_n-f|) \\
\Rightarrow & \lim_{n\to\infty}\mu(|f_n-f|)=0.
\end{aligned}
$$

□

Remark that in (DOM) and (Scheffé), "$f_n \to f$" can be replaced by "$f_n \to f$" a.e.

Next we discuss **Radon-Nikodym (RN)** theorem.

Definition 29. Given a measure space (S, Σ, μ) and $f \in (m\Sigma)^+$, we define a measure f_μ on Σ as follows: for all $A \in \Sigma$:

$$f_\mu(A) := \int_A f d\mu = \mu(\chi_A \cdot f). \tag{49}$$

Proposition 21. f_μ is a measure.

Proof. Need to show the rules.

1. $f_\mu(\emptyset) = 0$.
2. Given a sequence of sets $\{A_n \mid n \in \mathbb{N}\}$ such that $A_i \cap A_j = \emptyset$ for $i \neq j$,

$$
\begin{aligned}
f_\mu\left(\bigcup_{n=1}^\infty A_n\right) &= \mu\left(f\sum_{n=1}^\infty \chi_{A_n}\right) \\
&= \mu\left(\lim_{N\to\infty}\sum_{n=1}^N f\chi_{A_n}\right) \\
&= \lim_{N\to\infty}\mu\left(\sum_{n=1}^\infty f\chi_{A_n}\right) \quad \text{(by MON)} \\
&= \lim_{N\to\infty}\sum_{n=1}^N \mu(f\chi_{A_n}) \\
&= \sum_{n=1}^\infty \mu(f\chi_{A_n}) \\
&= \sum_{n=1}^\infty f\mu(A_n)
\end{aligned}
$$

This shows that f is countably additive.

□

Proposition 22. For $g \in (m\Sigma)^+$, $(f_\mu)(g) = \mu(f \cdot g)$.

Proof. By definition, this holds for every χ_A for all $A \in \Sigma$. By linearity, we know that this holds for all $g \in SF^+$. Next, for a general $g \in (m\Sigma)^+$, we take $)g_n \in SF^+$ for $n \in \mathbb{N}$ such that $g_n \uparrow g$. By MON:

$$(f_\mu)(g_n) \uparrow (f_\mu)(g) \text{ and } \mu(f \cdot g_n) \uparrow \mu(f \cdot g).$$

We then obtain:

$$(f_\mu)(g) \underbrace{=}_{(MON)} \lim_n (f_\mu)(g_n) = \lim_n \mu(f \cdot g_n) \underbrace{=}_{(MON)} \mu(f \cdot g).$$

□

Proposition 23. Given $f \in (m\Sigma)^+$, $h \in m\Sigma$, $h \in L^1(S, \Sigma, f_\mu) \iff f \cdot h \in L^1(S, \Sigma, \mu)$. If $h \in L^1(S, \Sigma, f_\mu)$ then $(f_\mu)(h) = \mu(fh)$

Proof. $h \in L^1(f_\mu) \iff (f_\mu)(f^\pm) < \infty \iff \mu(fh^\pm) < \infty \iff \mu((fh)^\pm) < \infty \iff fh \in L^1(\mu)$. □

Theorem 18 (Radon-Nikodym Theorem). *Let γ and μ be two measures on (S, Σ). We say that γ is **absolutely continuous** with respect μ if for all $A \in \Sigma$ $\mu(A) = 0 \Rightarrow \gamma(A) = 0$. If γ is absolutely continuous with respect to μ, then there exists an $f \in (m\Sigma)^+$ such that $\gamma = f_\mu$ i.e., for all $A \in \Sigma$,*

$$\gamma(A) = \int_A f d\mu. \tag{50}$$

*We call f the **RN-Derivative** of γ with respect to μ, denoted by $f = \frac{d\gamma}{d\mu}$.*

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and X is a random variable. Let $\mathcal{L}_X$ be the law or distribution of X. If $\mathcal{L}_X$ is absolutely continuous with respect to λ_{Leb} then by the RN theorem, there exists an $f_X \in (mB(\mathbb{R}))^+$ such that $\mathcal{L}_X = f_X \lambda_{\text{Leb}}$, i.e., f_X is the RN-derivative of $\mathcal{L}_X$ with respect to λ_{Leb}. We call f_X the probability density function of X. For example, $X \sim N(\mu, \sigma^2)$ a random variable, then:

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

If f_X is the probability density function of X, then for all $A \in B(\mathbb{R})$,

$$\mathbb{P}(X \in A) = \mathcal{L}_X(A) = \int_A f_X(x) dx. \tag{51}$$

Then,

$$\int_\Omega X d\mathbb{P} = \int_\mathbb{R} X d\mathcal{L}_X = \int_\mathbb{R} x f_X(x) dx.$$

Definition 30 (Expectation). The **expectation** of X is

$$\mathbb{E}[X] := \int_\Omega X d\mathbb{P} \tag{52}$$

For $A \in \mathcal{F}$,

$$\mathbb{E}[X; A] = \mathbb{E}[X \chi_A] = \int_A X d\mathbb{P}.$$

Then, $X \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \iff \mathbb{E}[|X|] < \infty \iff \mathbb{E}[X] \in \mathbb{R}$. (MON) and (MON)', (Fatou) and (Fatou)', (DOM), and (Scheffé) still apply in the setting of $(\Omega, \mathcal{F}, \mathbb{P})$. In fact, some of those results can be strengthened. In other words, $X_n \to X$ a.s. can be replaced by $X_n \to X$ in probability.

Theorem 19 (pMON). *Let X_n, X be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. If $X_n \leq X_{n+1}$ and $X_n \to X$ in probability and $\mathbb{E}\left[X_1^-\right] < \infty$. Then,*

$$\mathbb{E}[X_n] \uparrow \mathbb{E}[X].$$

Proof. Suppose $X_n \to X$ in probability. Then, I can find a subsequence $X_{n_k} \to X$ as $k \to \infty$. Since $X_n \leq X_{n+1}$ it turns out that I'll need $X_n \to X$ a.s. The result now follows from the regular (MON). $\square$

Theorem 20 (pDOM)**.** *Suppose X_n, $n \in \mathbb{N}$, X are random variables on Ω such that $X_n \to X$ in probability. Assume that there exists a $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $|X_n| \leq |Y|$ for all $n \geq 1$. Then, $X_n, X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $X_n \to X$ in $L^1(\mathbb{P})$.*

Proof. Since for all $n \geq 1$, $|X_n| \leq |Y|$ and $X_n \in L^1(\mathbb{P})$. Then, since $X_n \to X$ in probability, there exists a subsequence X_{n_k} such that $X_{n_k} \to X$ a.s. as $k \to \infty$. Hence,

$$|X| = \lim_{k\to\infty} |X_{n_k}| \leq |Y| \Rightarrow X \in L^1(\mathbb{P}).$$

For a contradiction, assume that the statement does not hold. In other words, assume that $X_n \nrightarrow X$ in $L^1(\mathbb{P})$. Then, $\mathbb{E}[|X_n - X|] \nrightarrow 0$ a.s. as $n \to \infty$. Then, there exists a $\delta > 0$ and a subsequence $\{n_l\}$ such that $\mathbb{E}[|X_{n_l} - X|] \geq \delta$. Since $X_n \to X$ in probability, there is a subsequence $\{n_{l_p}\}$ such that $X_{n_{l_p}} \to X$ a.s. as $p \to \infty$ and $|X_{n_{l_p}}| \to X$ in $L^1(\mathbb{P})$ (by the standard DCT). $\square$

Theorem 21 (pScheffé)**.** *Let X_n, $n \in \mathbb{N}$, X be random variables on Ω, $X \in L^1(\mathbb{P})$ and $X_n \to X$ in probability. Then, $X_n \to X$ in $L^1(\mathbb{P}) \iff \lim_{n\to\infty} \mathbb{E}[|X_n|] = \mathbb{E}[|X|]$.*

Proof. Exercise. $\square$

3.3 Review of L^p Spaces

Definition 31 (pth Moment)**.** For $1 \leq p < \infty$, we say that $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ if $|X|^p \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. If $X \in L^p(\mathbb{P})$, then $\mathbb{E}[X^p]$ is the **pth moment of** X.

Facts about L^p spaces.

1. L^p is a vector space over $\mathbb{R}$: i.e., if $X, Y \in L^p$, $a, b \in \mathbb{R}$, then $aX + bY \in L^p$.
2. Given $X \in L^p$, $||X||_p := (\mathbb{E}[|X|^p])^{1/p}$.
   (a) $||X||_p$ is a norm. We have:
      i. $||\cdot||_p \geq 0$ and $||X||_p = 0 \iff X = 0$ a.s.
      ii. For all $c \in \mathbb{R}$, $||cX||_p = |c|||X||_p$.
      iii. For all $X, Y \in L^p$, $||X + Y||_p \leq ||X||_p + ||Y||_p$ **(Minkowski's Inequality)**.
3. **Cauchy-Schwarz Inequality:** if $X, Y \in L^2(\mathbb{P})$, then $XY \in L^1(\mathbb{P})$ and
$$\mathbb{E}[|XY|] \leq ||X||_2||Y||_2. \tag{53}$$
4. **Hölder's Inequality**: assume $1 \leq p < \infty$, $1 \leq q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. If $X \in L^p(\mathbb{P})$ and $Y \in L^1(\mathbb{P})$, then $XY \in L^1(\mathbb{P})$ and:
$$\mathbb{E}[|XY|] \leq ||X||_p||Y||_q.$$
5. **Monotonicity:** if $1 \leq p \leq q < \infty$, then $X \in L^1 \Rightarrow X \in L^p$. Hence, $||X||_p \leq ||X||_p$. This tells us that L^p spaces are nested.
6. L^p is a **Banach Space.**
   (a) $(L^p, ||\cdot||_p)$ is a Banach Space.
   (b) $(L^2, ||\cdot||_2)$ is a Hilbert space with inner product $(X, Y)_2 = \mathbb{E}[XY]$.
7. If $X, Y \in L^2(\mathbb{P})$ (we call this condition "if X and Y are square integrable"). Then, we can define the **variance** of X and the **covariance** of X and Y as:
   (a) $\text{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$.
   (b) $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

   By Cauchy-Schwarz, we have the following inequality:
$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)} \tag{54}$$

3.4 Concentration Inequalities

Theorem 22 (Markov's Inequality). *Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X on Ω. If $g : \mathbb{R} \to [0, \infty[$ is a* ***non-negative****,* ***increasing****,* ***Borel*** *function, then for all $c \in \mathbb{R}$:*

$$\mathbb{P}(X > c) \leq \frac{\mathbb{E}[g(x); X \geq c]}{g(c)} \leq \frac{\mathbb{E}[g(X)]}{g(c)} \tag{55}$$

(assuming that $g(c) \neq 0$).

Proof. The first inequality follows from the monotonicity of integrals of non-negative functions. This is because $g(x) \geq g(c)$ since g is increasing:

$$\mathbb{E}[g(X)] \geq \mathbb{E}[g(X); X \geq c] \geq \mathbb{E}[g(c); X \geq c] = g(c)\mathbb{P}(X \geq c).$$

Re-arranging this gives the desired result. □

Theorem 23 (Chebychev's Inequality). *Given $X \in L^2(\mathbb{P})$, for all $c > 0$,*

$$\mathbb{P}(|X - \mathbb{E}[X]| > c) \leq \frac{1}{c^2}\mathrm{Var}(X). \tag{56}$$

- If $X \in L^p(\mathbb{P})$ for some $p \geq 1$, then $\forall c \geq 0$:

$$\mathbb{P}(|X| \geq c) < \frac{1}{c^p}\mathbb{E}[|X|^p; |X| \geq c] \leq \frac{1}{c^p}\mathbb{E}[|X|^p].$$

- If $e^{\alpha|X|} \in L^1(\mathbb{P})$ for some $\alpha > 0$, then for all $c > 0$,

$$\mathbb{P}(|X| \geq c) \leq e^{-\alpha X}\mathbb{E}\left[e^{\alpha|X|}; |X| \geq c\right] \leq e^{-\alpha c}\mathbb{E}\left[e^{\alpha|X|}\right]$$

Essentially, better integrability will give us better decay.

Proposition 24. Given X_n, $n \in \mathbb{N}$, $X \in L^p(\mathbb{P})$ for some $p \geq 1$. If $X_n \to X$ in $L^p(\mathbb{P})$ i.e., $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0$. Then, $X_n \to X$ in probability.

Proof. The proof follows directly from Markov's Inequality,.

$$\forall \varepsilon > 0 \ \mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p}\mathbb{E}[|X_n - X|^p] \to 0 \text{ as } n \to \infty.$$

□

1. Convergence in $L^p \Rightarrow$ convergence in probability.
2. Convergence almost surely $\Rightarrow$ convergence in probability.
3. Convergence in $L^p \nRightarrow$ convergence almost surely
4. Convergence almost surely $\nRightarrow$ Convergence in L^p

Example 13. Take $([0,1], B([0,1]), \lambda_{\text{Leb}})$. Set:

$$X_n := \chi_{]0,1/n[}n^{2/p}$$

for $n \geq 1$. Then, $X_n \to 0$ pointwise almost surely. But:

$$\int_{]0,1[} |X_n|^p d\lambda = n^2\frac{1}{n} = n \to \infty$$

which shows that $X_n \nrightarrow 0$ in L^p.

Example 14. Take again $([0,1], B([0,1]), \lambda_{\text{Leb}})$. For $k \in \mathbb{N}$. and $j = 1, 2, ..., k$, set

$$\varphi_{k_j} := \chi_{]\frac{j-1}{k}, \frac{1}{k}]}.$$

Re-order the φ_{k_j}'s in lexographical order:

$$\begin{aligned}
&\varphi_{11} := X_1 \\
&\varphi_{21} := X_2 \ \varphi_{22} := X_3 \\
&\varphi_{31} := X_4 \ \varphi_{32} := X_5 \ \varphi_{33} := X_6
\end{aligned}$$

and re-name them $X_1, X_2, ...$ For all $n \geq 1$, $\exists_1\ k_n, j_n$ such that $X_n = \varphi_{k_n j_n}$. For all $n \geq 1$, there exists one k_n, j_n such that $X_n = \varphi_{k_n j_n}$. For every $p \geq 1$,

$$\mathbb{E}\left[|X_n|^p\right] = \frac{1}{k_n} \to 0 \text{ as } n \to \infty \Rightarrow X_n \to 0 \text{ in } L^p(\mathbb{P}).$$

However, for all $x \in [0,1]$, for all $k \geq 1$, there exists $j = 1, ..., k$ such that $\varphi_{k_j}(x) = 1$. Hence, $X_n(x) = 1$ for infinitely many n. This implies that $\mathbb{P}(X_n = 1 \text{ i.o.}) = 1$ which tells us that $X_n \nrightarrow 0$ almost surely.

Theorem 24 (Jensen's Inequality)**.** *If $X \in L^1(\mathbb{P})$ and if $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $\varphi(x) \in L^1(\mathbb{P})$, then:*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)] \tag{57}$$

Proof. The proof relies on using a convenient fact about convex functions: if φ is convex, then for all $x \in \mathbb{R}$, there exists a line l passing through $(x, \varphi(x))$ and l lies entirely below the graph of φ. This line is called the **supporting line** at $(x, \varphi(x))$. Assume that $y = \alpha x + \beta$ is the supporting line at $(\mathbb{E}[X], \varphi(\mathbb{E}[X])$. So I know that everywhere, it's true that:

$$\varphi(\mathbb{E}[X]) = \alpha\mathbb{E}[X] + \beta \leq \varphi(x).$$

By integrability and monotonicity, this implies that:

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)],$$

which is what we wanted to show. $\square$

As a corollary, we obtain the following helpful inequalities:

- If $\mathbb{E}[|X|^p] < \infty$ for some $p \geq 1$ then,

$$(\mathbb{E}[|X|])^p \leq \mathbb{E}[|X|^p].$$

- If $\mathbb{E}[e^{\alpha x}] < \infty$ for some $\alpha \in \mathbb{R}$, then

$$e^{\alpha\mathbb{E}[X]} \leq \mathbb{E}\left[e^{\alpha X}\right]$$

- If $\mathbb{E}[X^+] < \infty$, then

$$(\mathbb{E}[X])^+ \leq \mathbb{E}\left[X^+\right]$$

  Here, $\varphi = \max\{0, x\}$.
- If $\mathbb{E}[X^-] < \infty$, then

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

  Here, $\varphi = \max\{0, -x\}$.

Theorem 25. *Suppose X is an $\mathbb{R}$-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution $\mathcal{L}_X$ and $f : \mathbb{R} \to \mathbb{R}$ a Borel function. Then,*

$$f(x) \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \iff f \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L}_X) \tag{58}$$

In this case,

$$\mathbb{E}\left[f(x)\right] = \int_{\mathbb{R}} f d\mathcal{L}_X \ (*). \tag{59}$$

Proof. We'll use the Monotone Class Theorem. First, assume that f is bounded and Borel $f \in b\mathcal{B}(\mathbb{R})$ (and hence $f \in L^1(\mathcal{L}_X), f(x) \in L^1(\mathbb{P})$). Set:

$$H := \{f \in b\mathcal{B}(\mathbb{R}) \mid (*) \text{ holds for } f\}.$$

We need to check that H is a monotone class.

1. H is clearly a vector space over R (because of linearity).
2. $1 \in H$ also clear.
3. If $f_n \in H$, $f_n \geq 0$, and $f_n \uparrow f \in b\mathcal{B}(\mathbb{R})$, then f will satisfy $(*)$ because of (MON) or (DOM).

Moreover, for all $A \in \mathcal{B}(\mathbb{R})$, χ_A is in H. By the monotone class theorem, $b\mathcal{B}(\mathbb{R}) \subseteq H$ which implies that $b\mathcal{B}(\mathbb{R}) = H$. Hence, $(*)$ holds for all bounded Borel functions.

Now, for general Borel functions f, set $f_k := f\chi_{\{|f| \leq k\}}$ for all $k \in \mathbb{N}$. For sure, we know that $|f_k| \uparrow |f|$ and $f_k \to f$ as $k \to \infty$. By Montone convergence theorem,

$$\mathbb{E}\left[|f(x)|\right] = \lim_{k\to\infty} \mathbb{E}\left[|f_k(x)|\right] \underbrace{=}_{(*)} \lim_{k\to\infty} \int_{\mathbb{R}} |f_k| d\mathcal{L}_X = \int_{\mathbb{R}} |f| d\mathcal{L}_X.$$

Hence, $f(x) \in L^1(\mathbb{P}) \iff f \in L^1(\mathcal{L}_X)$. Now we can show it for general f by using DOM with dominating function $|f|$:

$$\mathbb{E}\left[|f(x)|\right] = \lim_{k\to\infty} = \lim_{k\to\infty} \int_{\mathbb{R}} f_k d\mathcal{L}_X \underbrace{=}_{(DOM)} \int_{\mathbb{R}} f d\mathcal{L}_X.$$

$\square$

Some remarks on the proof and this theorem:

1. It's clear from the proof that $(*)$ holds for $f \in (m\mathcal{B}(\mathbb{R}))^+$.
2. If $\mathcal{L}_X << \lambda_{\text{Leb}}$, (i.e., $\mathcal{L}_X << dx$), and the probability density function is given by f_X. Then, for every Borel function $g \in m\mathcal{B}(\mathbb{R})$, $g(x) \in L^1(\mathbb{P}) \iff g \in L^1(\mathcal{L}_X) \iff g \cdot f_X \in L^1(\lambda_{\text{Leb}})$. In this case,

$$\mathbb{E}\left[g(X)\right] = \int_{\mathbb{R}} g d\mathcal{L}_X = \int_{\mathbb{R}} g(x)\mathcal{L}_X(x)dx.$$

Theorem 26. *Given $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. Let X and Y be two independent random variables on Ω. Then, for every $W \in L^1(\mathbb{P}) \cap m\sigma(X)$ and $Z \in L^1(\mathbb{P}) \cap m\sigma(Y)$. Then,*

$$W \cdot Z \in L^1(\mathbb{P}) \text{ and } \mathbb{E}\left[WZ\right] \underbrace{=}_{(*)} \mathbb{E}\left[W\right]\mathbb{E}\left[Z\right]. \tag{60}$$

Before the proof, we need a Lemma.

Lemma 5. Assume $f(X) \in L^1(\mathbb{P})$ and $B \in \mathcal{B}(\mathbb{R})$. Then,

$$\mathbb{E}\left[f(X); Y \in B\right] \underbrace{=}_{(**)} \mathbb{E}\left[f(X)\right] \mathbb{P}\left(Y \in B\right).$$

In this case, $g = \chi_B$,

Proof. Set $H = \{f \in b\mathcal{B}(\mathbb{R}) \mid (**) \text{ holds for } f\}$. First check that H is a monotone class. We need to check these three things:

1. H is a vector space over R.
2. $1 \in H$ is obvious.
3. If $f_n \in \mathcal{H}$, $f_n \geq 0$, $f_n \uparrow f \in b\mathcal{B}(\mathbb{R}) \Rightarrow f \ \in H$.

H is a monotone class ✓. Now, for all $A \in \mathcal{B}(\mathbb{R})$, set $f = \chi_A$. Then,

$$\begin{aligned}
\mathbb{E}\left[\chi_A(X); Y \in B\right] &= \mathbb{P}\left(X \in A, Y \in B\right) \\
&= \mathbb{P}\left(X \in A\right) \mathbb{P}\left(Y \in B\right) \text{ (by independence).} \\
&= \mathbb{E}\left[\chi_A(X)\right] \mathbb{P}\left(Y \in B\right).
\end{aligned}$$

This shows that $\chi_A \in H$ and hence by the Monotone class theorem, $b\mathcal{B}(\mathbb{R}) = H$. Net, for a general $f \in m\mathcal{B}(\mathbb{R})$ such that $f(x) \in L^1(\mathbb{P})$. Set $f_k := f\chi_{\{|f| \leq k\}}$. Then, $f_k(x) \to f(x)$ as $k \to \infty$ and $|f_k(x)| \leq |f(x)|$. Hence,

$$\mathbb{E}\left[f(x); Y \in B\right] \underbrace{=}_{(DOM)} \lim_{k \to \infty} \mathbb{E}\left[f_k(X); Y \in B\right] \underbrace{=}_{(**)} \lim_{k \to \infty} \mathbb{E}\left[f_k(X)\right] \mathbb{P}\left(Y \in B\right) \underbrace{=}_{(MON)} \mathbb{E}\left[f(X)\right] \mathbb{P}\left(Y \in B\right).$$

Which proves the lemma. □

Proof. The theorem is equivalent to proving that for every $f, g \in m\mathcal{B}(\mathbb{R})$ such that $f(X), g(Y) \in L^1(\mathbb{P})$ we must have that $f(X), g(Y) \in L^1(\mathbb{P})$ and $\mathbb{E}\left[f(X)g(Y)\right] = \mathbb{E}\left[f(X)\right] \mathbb{E}\left[g(Y)\right]$.

By the lemma, we know that $(*)$ holds for $g = \chi_B$ (we call these root functions) for all $b \in \mathcal{B}(\mathbb{R})$. **Exercise**: set $G := \{g \in b\mathcal{B}(\mathbb{R}) \mid (*) \text{ holds for } g\}$. Complete the proof by following similar steps as above. □

Corrolary 2. If X and Y are independent random variables and $X, Y \in L^1(\mathbb{P})$. Then, the **covariance of** X **and** Y, denoted $\text{Cov}(X, Y)$ exists and $\text{Cov}(X, Y) = 0$.

3.5 Uniform Integrability

Proposition 25. Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then,

$$X \in L^1(\mathbb{P}) \iff \lim_{M \to \infty} \mathbb{E}\left[|X|; |X| \geq M\right] = 0$$

Proof. "$\Leftarrow$": Choose M sufficiently large such that $\mathbb{E}\left[|X|; |X| \geq M\right] \leq 1$. Since $|X|$ is non-negative, we can apply linearity:

$$\begin{aligned}
\mathbb{E}\left[|X|\right] &= \mathbb{E}\left[|X|; |X| < M\right] + \mathbb{E}\left[|X|; |X| \geq M\right] \\
&\leq M + 1 \\
&\leq \infty \Rightarrow X \in L^1(\mathbb{P}).
\end{aligned}$$

"$\Rightarrow$": Since $X \in L^1(\mathbb{P})$, $\mathbb{E}[|X|] < \infty$. Using (DOM) or (MON), we get:

$$\begin{aligned}\infty &> \mathbb{E}[|X|]\\ &= \lim_{M\to\infty} \mathbb{E}[|X|; |X| < M]\\ &= \lim_{M\to\infty} (\mathbb{E}[|X|] - \mathbb{E}[|X|; |X| \geq M])\\ &= \mathbb{E}[|X|] - \lim_{M\to\infty} \mathbb{E}[|X|; |X| \geq M].\end{aligned}$$

This implies that $\lim_{M\to\infty} \mathbb{E}[|X|; |X| \geq M] = 0$, which is what we wanted to show. □

Definition 32. Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, $\{X_n\}$ is **uniformly integrable** if

$$\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|; |X_n| \geq M] = 0. \tag{61}$$

i.e., for all $\varepsilon > 0$, there exists a $M > 0$ such that $\mathbb{E}[|X_n|; |X_n| \geq M] \leq \varepsilon$ for all $n \in \mathbb{N}$. In other words, the concentration of X_n happens in a uniform way in N.

Proposition 26. (Helpful Properties about Uniform Integrability)

1. If $\{X_n\}$ is uniformly integrable, then $\{X_n\}$ is bounded in $L^1(\mathbb{P})$. In other words,

$$\sup_n \mathbb{E}[|X_n|] < \infty \tag{62}$$

2. If $\{X_n\}$ is bounded in $L^1(\mathbb{P})$ for some $p > 1$, then $\{X_n | \ n \in \mathbb{N}\}$ is uniformly integrable.
   (a) We have: L^p boundedness for $p > 1 \Rightarrow$ uniform integrability $\Rightarrow L^1$ boundedness.

Proof. Let's prove (2). Assume that $\{X_n \mid n \in \mathbb{N}\}$ is bounded in L^1 for some $p > 1$. Then, for all $M \geq 0$,

$$\begin{aligned}\mathbb{E}[|X_n|; |X_n| \geq M] &\leq \mathbb{E}\left[|X_n|\frac{|X_n|^{p-1}}{M^{p-1}}; |X_n| \geq M\right]\\ &\leq \mathbb{E}[|X_n|^p]\frac{1}{M^{p-1}} \ \forall n \in \mathbb{N}.\end{aligned}$$

Taking the sup of both sides,

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|; |X_n| \geq M] \leq \frac{1}{M^{p-1}} \sup_n \mathbb{E}[|X_n|^p] \to 0 \text{ as } M \to \infty.$$

This proves that $\{X_n\}$ is uniformly integrable.

Let's prove (1). Assume that $\{X_n\}$ is uniformly integrable. Choose $M > 0$ large such that,

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|; |X_n| \geq M] \leq 1.$$

Break this up exactly as was done in a previous proof:

$$\begin{aligned}\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] &= \sup_{n\in\mathbb{N}} (\mathbb{E}[|X_n|; |X_n| \leq M] + \mathbb{E}[|X_n|; |X_n| \geq M])\\ &\leq M + 1\\ &< \infty.\end{aligned}$$

This shows that $\{X_n\}$ is bounded in $L^1(\mathbb{P})$. □

Proposition 27. Given a sequence $\{X_n\}$, X random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the following statements are equivalent:

1. $X_n \in L^1(\mathbb{P})$ for all $n \in \mathbb{N}$, $X \in L^1(\mathbb{P})$ and $X_n \to X$ in $L^1(\mathbb{P})$.
2. The sequence $\{X_n \mid n \in \mathbb{N}\}$ is uniformly integrable and $X_n \to X$ in probability.

What this proposition tells us is that what's missing between convergence in probability and convergence in L^1 is uniform integrability.

Proof. "(1) $\Rightarrow$ (2)": Assume that everything is integrable: $X_n \in L^1(\mathbb{P})$, $X \in L^1(\mathbb{P})$, and $X_n \to X$ in $L^1(\mathbb{P})$. Obviously, $X_n \to X$ in probability. We need to show that $\{X_n\}$ is uniformly integrable. So, we go back to the definition. Note that the definition has nothing to do with the X, so the only technique we can do is to force an X to appear and bound the pieces. We have that for all $M > 0$:

$$\begin{aligned}\mathbb{E}\left[|X_n|; |X_n| \geq M\right] &\leq \mathbb{E}\left[|X_n - X|; |X_n| \geq M\right] + \mathbb{E}\left[|X|; |X_n| \geq M\right] \\ &\leq \underbrace{\mathbb{E}\left[|X_n - X|\right]}_{(A)} + \underbrace{\mathbb{E}\left[|X|; |X| \leq \sqrt{M} \text{ and } |X_n| \geq M\right]}_{(B2)} + \underbrace{\mathbb{E}\left[|X|; |X| > \sqrt{M} \text{ and } |X_n| \geq M\right]}_{(B2)}\end{aligned}$$

1. $\lim_{n\to\infty}(A) = 0$ by assumption since $X_n \to X$ in L^1.
2. We can bound (B1) using Markov's Inequality:

$$\begin{aligned}\mathbb{E}\left[|X|; |X| \leq \sqrt{M} \text{ and } |X_n| \geq M\right] &\leq \sqrt{M}\mathbb{P}\left(|X| \leq M; |X_n| \geq M\right) \\ &\leq \sqrt{M}\mathbb{P}\left(|X_n| \geq M\right) \\ &\leq \sqrt{M}\frac{\mathbb{E}\left[|X_n|\right]}{M} \\ &\frac{1}{\sqrt{M}}\sup_{n\in\mathbb{N}}\mathbb{E}\left[|X_n|\right] \to 0 \text{ as } M \to \infty \text{ because } X_n \to X \text{ in } L^p1.\end{aligned}$$

3. We can bound (B2) as follows:

$$\mathbb{E}\left[|X|; |X| > \sqrt{M} \text{ and } |X_n| \geq M\right] \leq \mathbb{E}\left[|X|; |X| > \sqrt{M}\right] \to 0 \text{ as } M \to \infty \text{ bcoz } X \in L^1(\mathbb{P}).$$

For all $\varepsilon > 0$, choose $N > 0$ large enough such that for all $n \geq N$, $(A) \leq \frac{\varepsilon}{3}$. Then, choose $M > 0$ large such that $(B1) \leq \frac{\varepsilon}{3}$ and $(B2) \leq \frac{\varepsilon}{3}$ and

$$\mathbb{E}\left[|X_j|; |X_j| \geq M\right] \leq \varepsilon,$$

for all $j = 1, 2, ..., N-1$. Combining all of this together,

$$\sup_{n\in\mathbb{N}}\mathbb{E}\left[|X_n|; |X_n| \geq M\right] \leq \varepsilon,$$

which shows that $\{X_n\}$ is uniformly integrable.

(2) $\Rightarrow$ (1): Now assume that $\{X_n\}$ is uniformly integrable. Then, $X_n \in L^1$ for all $n \in \mathbb{N}$. Let $A := \sup_n \mathbb{E}\left[|X_n|\right] < \infty$. Since $X_n \to X$ in probability, there exists a sub sequence $X_{n_k} \to X$ almost surely as $k \to \infty$. By Fatou's Lemma,

$$\mathbb{E}\left[|X|\right] \leq \liminf_k \mathbb{E}\left[|X_{n_k}|\right] \leq A < \infty \Rightarrow X \in L^1.$$

To show L^1 convergence, we'll need a lemma.

Lemma 6. If $\{X_n\}$ is uniformly integrable and if $X \in L^1$, then $\{X_n - X \mid n \in \mathbb{N}\}$ is uniformly integrable.

Proof. For all $M > 0$, we can bound:

$$\begin{aligned}\mathbb{E}\left[|X_n - X|; |X_n - X| \geq M\right] &\leq \mathbb{E}\left[|X_n| + |X|; |X_n| \geq \frac{M}{2} \text{ or } |X| \geq \frac{M}{2}\right] \\ &\leq \underbrace{\mathbb{E}\left[|X_n|; |X_n| \geq \frac{M}{2}\right]}_{(A1)} + \underbrace{\mathbb{E}\left[|X|; |X| \geq \frac{M}{2}\right]}_{(A2)} + \underbrace{\mathbb{E}\left[|X_n|; |X| \geq \frac{M}{2}\right]}_{(A3)} + \underbrace{\mathbb{E}\left[|X|; |X_n| \geq \frac{M}{2}\right]}_{(A4)}\end{aligned}$$

We can bound (A1) - (A4) as follows:

1. (A1): $\sup_n (A1) \to 0$ as $M \to \infty$ since $\{X_n\}$ is uniformly integrable.
2. $(A2) : (A2) \to 0$ as $M \to 0$ since $X \in L^1$.
3. (A3): We will bound this using Markov's inequality:

$$\begin{aligned}\mathbb{E}\left[|X_n|; |X| \geq \frac{M}{2}\right] &= \mathbb{E}\left[|X_n|; |X_n| \geq \sqrt{M} \text{ and } |X| \geq \frac{M}{2}\right] + \mathbb{E}\left[|X_n|; |X_n| < \sqrt{M} \text{ and } |X| \geq \frac{M}{2}\right] \\ &\leq \mathbb{E}\left[|X_n|; |X_n| \geq \sqrt{M}\right] + \sqrt{M}\mathbb{P}\left(|X| \geq \frac{M}{2}\right) \\ &\leq \mathbb{E}\left[|X_n|; |X_n| \geq \sqrt{M}\right] + \sqrt{M}\frac{2\mathbb{E}\left[|X|\right]}{M} \to 0 \text{ as } M \to \infty.\end{aligned}$$

4. (A4): exactly the same. Exercise: prove that $\lim_{M\to\infty} \sup_n (A_n) = 0$.

This shows that $\{X_n - X \mid n \in \mathbb{N}\}$ is uniformly integrable. □

Now we are ready to go back to the proof of the theorem. Assume that $\{X_n | \; n \in \mathbb{N}\}$ is uniformly integrable and $X_n \to X$ in probability. Then, $\{X_n - X\}$ is uniformly integrable. We want to show that $X_n \to X$ in L^1. For all $M > 0$ and for all $\varepsilon > 0$:

$$\mathbb{E}\left[|X_n - X|\right] = \underbrace{\mathbb{E}\left[|X_n - X|; |X_n - X| \geq M\right]}_{(B1)} + \underbrace{\mathbb{E}\left[|X_n - X|; \varepsilon \leq |X_n - X| < M\right]}_{(B2)} + \underbrace{\mathbb{E}\left[|X_n - X|; |X_n - X| \leq \varepsilon\right]}_{(B3)}$$

Again, bound (B1)-(B3):

1. (B1): Since $\{X_n - X | n \in \mathbb{N}\}$ is uniformly integrable, we can take M sufficiently large such that $\sup_n (B1) < \varepsilon$.
2. (B2): We can bound:

$$\mathbb{E}\left[|X_n - X|; \varepsilon \leq |X_n - X| < M\right] \leq M\mathbb{P}\left(|X_n - X| \geq \varepsilon\right) \to 0 \text{ as } n \to \infty.$$

3. (B3): Already done $\leq \varepsilon$.

Therefore, $\mathbb{E}\left[|X_n - X|\right]$ can be made arbitrarily small for all sufficiently large n. □

Hence, we have an equivalence: convergence in probability and uniform integrability $\iff$ L^1 convergence.

4 Laws of Large Numbers (LLN)

4.1 Terminology

Given a sequence of random variables $\{X_n \mid n \in \mathbb{N}\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$. For all $n \geq 1$, set:

$$S_n := \sum_{j=1}^{n} X_j.$$

We say that **LLN** holds for $\{X_n\}$ if:

1. in the **classical** setting (assuming $\mathbb{E}[X_n]$) exists if:

$$\frac{S_n - \mathbb{E}[S_n]}{n} \to 0. \tag{63}$$

   Depending on the nature of the convergence above, this is either a **weak** law of large numbers or a **strong** law of large numbers:

   (a) if the convergence is in probability, then it's a **Weak Law of Large Numbers.**
   (b) if the convergence is almost surely, then it's a **Strong Law of Large Numbers.**

2. in the **general** setting: if there exists a sequence $\{a_n \mid n \in \mathbb{N}\} \subseteq \mathbb{R}$, $\{b_n \mid n \in \mathbb{N}\}$ with $b_n \uparrow \infty$ such that:

$$\frac{S_n - a_n}{b_n} \to 0. \tag{64}$$

   As in the classical setting, depending on the nature of the convergence, this is either a WLLN (if the convergence is in probability) or a SLLN (if the convergence is almost surely).

To establish a WLLN/SLLN, we in general need two types of conditions.

1. Conditions on integrability.
2. Conditions on joint distributions.

Theorem 27 (WLLN 1 (Chebychev))**.** *Let $\{X_n\}$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\{X_n\}$ is bounded in $L^2(\mathbb{P})$ i.e.: $A := \sup_n \mathbb{E}[X_n^2] < \infty$ and $\{X_n \mid n \in \mathbb{N}\}$ is uncorrelated, i.e., $Cov(X_i, X_j) = 0$ for $i \neq j$. Then, WLLN holds for $\{X_n\}$ i.e.,*

$$\frac{S_n - \mathbb{E}[S_n]}{n} \to 0 \tag{65}$$

in probability.

Proof. WLOG, we can assume that $\mathbb{E}[X_n] = 0$ for all $n \in \mathbb{N}$. Otherwise, replace X_n by $X_n - \mathbb{E}[X_n]$. So, in this case: $\mathbb{E}[X_i X_j] = 0$ for all $i \neq j$. Hence,

$$S_n^2 = \mathbb{E}\left[\left(\sum_{j=1}^{n} X_j\right)^2\right] = \sum_{i,j=1}^{n} \mathbb{E}[X_i X_j] \underbrace{=}_{(*)} \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] \leq An$$

The uncorrelation in the step $(*)$ is *critical.* Hence, we get

$$\mathbb{E}\left[S_n^2\right] = O(n) = o(n^2). \tag{66}$$

When proving laws of large numbers, we're always after something like (66)! The rest follows from Chebychev / Markov: for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leq \frac{\mathbb{E}\left[S_n^2\right]}{n^2\varepsilon^2} \leq \frac{An}{n^2\varepsilon^2} \to 0 \text{ as } n \to \infty.$$

Hence, $\frac{S_n}{n} \to 0$ in probability. $\square$

In fact, under exactly the same assumptions, we can have the **Strong Law of Large Numbers**.

Theorem 28 (SLLN 1 (Raychman))**.** *Under the same assumptions as WWLN 1: bounded in $L^2(\mathbb{P})$ and uncorrelated, SSLN holds for $\{X_n\}$, i.e.,*

$$\frac{S_n - \mathbb{E}\left[S_n\right]}{n} \to 0 \ \text{a.s.} \tag{67}$$

Proof. Again, we can assume that all the expectations are zero: $\mathbb{E}\left[X_n\right] = 0$ for all $n \in \mathbb{N}$. From the previous proof, we know that for all $\varepsilon > 0$:

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leq \frac{A}{\varepsilon^2 n} \text{ for all } n \in \mathbb{N}. \tag{68}$$

Since we want to use Borelli-Cantelli to get almost sure convergence, we need summability, so we have:

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| > \varepsilon\right) \leq \frac{A}{\varepsilon^2 n^2}$$

By (BC1), for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| > \varepsilon \text{ i.o}\right) = 0 \Rightarrow \frac{S_{n^2}}{n^2} = 0 \text{ a.s. (E1)}$$

Now we need to control the fluctuations. For $n \geq 1$:

$$D_n := \underbrace{\max_{n^2 \leq k \leq (n+1)^2} |S_k - S_{n^2}|}_{\text{largest fluctuation}} \Rightarrow D_n^2 = \max_{n^2 \leq k \leq (n+1)^2} |S_k - S_{n^2}|^2 \leq \sum_{k=n^2+1}^{(n+1)^2-1} |S_k - S_{n^2}|^2.$$

So,

$$\begin{aligned}
\mathbb{E}\left[D_n^2\right] &\leq \sum_{k=n^2+1}^{(n+1)^2-1} \mathbb{E}\left[|S_k - S_{n^2}|^2\right] \\
&= \sum_{k=n^2+1}^{(n+1)^2-1} \mathbb{E}\left[|X_{n^2+1} + X_{n^2+2} + \ldots + X_k|^2\right] \\
&= \sum_{k=n^2+1}^{(n+1)^2-1} \left(\sum_{j=n^2+1}^{k} \underbrace{\mathbb{E}\left[X_j^2\right]}_{\leq A}\right) \\
&\leq AO(n^2) \\
&\leq Cn^2,
\end{aligned}$$

where C is some constant. By Chebychev/Markov's inequality, for all $\varepsilon > 0$:

$$\mathbb{P}\left(\frac{D_n}{n^2} > \varepsilon\right) \leq \frac{\mathbb{E}\left[D_n^2\right]}{\varepsilon^2 n^4} \leq \frac{Cn^2}{\varepsilon^2 n^4} = \frac{C}{\varepsilon^2 n^2}.$$

By (BC1),

$$\mathbb{P}\left(\frac{D_n}{n^2} > \varepsilon \text{ i.o.}\right) = 0 \Rightarrow \frac{D_n}{n^2} \to 0 \text{ a.s. (E2)}$$

Both (E1) and (E2) have probability 1. For all ω such that (E1) and (E2) occur at ω, we have the following: for all $k \geq 1$, there exists a unique n_k such that $n_k^2 \leq k \leq (n_k+1)^2$ and $n_k \uparrow \infty$ as $k \to \infty$, with:

$$\begin{aligned}\left|\frac{S_k(\omega)}{k}\right| &\leq \frac{|S_k(\omega) - S_{n_k^2}|}{n_k^2} + \frac{|S_{n_k^2}(\omega)|}{n_k^2} \\ &\leq \frac{D_{n_k}(\omega)}{n_k^2} + \frac{S_{n_k^2}(\omega)}{n_k^2} \to 0 \text{ as } k \to \infty.\end{aligned}$$

Hence, $(E1) \cap (E2) \subseteq \left\{\frac{S_n}{n} \to 0\right\}$. Therefore, by the monotonicity of probability, $\frac{S_n}{n} \to 0$ almost surely. □

4.2 More Preparation

We would like to establish LLNs without assumptions on the second moment. In this case, Chebyvhev et al will fail. So, we have to do truncations.

Definition 33 (Equivalent Sequences)**.** Assume that $\{X_n \mid n \in \mathbb{N}\}$ and $\{Y_n \mid n \in \mathbb{N}\}$ are two sequences of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that two sequences are **equivalent** if

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) < \infty.$$

Note that if $\{X_n\}$ and $\{Y_n\}$ are equivalent, then by (BC1),

$$\mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0.$$

Pointwise, this means that for almost every $\omega \in \Omega$, there exists an $N_\omega > 0$ such that $X_n(\omega) = Y_n(\omega)$ for all $n \geq N_\omega$. In this case:

1. $\sum_{n=1}^{\infty}(X_n - Y_n)$ converges almost surely.
2. For all $\{b_n \mid \in \mathbb{N}\} \subseteq \mathbb{R}^+$ such that $b_n \uparrow \infty$,

$$\frac{1}{b_n}\sum_{j=1}^{n}(X_j - Y_j) \to 0 \text{ a.s.}$$

If $\{X_n \mid n \in \mathbb{N}\}$ and $\{Y_n \mid n \in \mathbb{N}\}$ are equivalent, then set $S_n := \sum_{j=1}^n X_j$ and $T_n := \sum_{j=1}^n Y_j$. Take a sequence of real numbers $\{b_n \mid n \in \mathbb{N}\} \subseteq \mathbb{R}^+$ with $b_n \uparrow \infty$. Then,

$$\lim_{n\to\infty} \frac{1}{b_n}(S_n - T_n) = 0 \text{ a.s.} \tag{69}$$

Theorem 29 (WLLN 2)**.** *Let $\{X_n\}$ be a sequence of random variables on Ω such that the X_n's are identically distributed and pairwise independent and $\mathbb{E}[X_1] = m \in \mathbb{R}$ i.e. $X_n \in L^1(\mathbb{P})$ for all $n \in \mathbb{N}$. Then, WLLN holds for the sequence, i.e.:*

$$\frac{S_n}{n} \to m \tag{70}$$

in probability.

Proof. We cannot talk about the variance, since the second moment may not exist. So, we need to truncate the sequence: For all $n \in \mathbb{N}$, set

$$Y_n := \chi_{\{|X_n| \le n\}} \cdot X_n = \begin{cases} X_n & \text{if } |X_n| \le n \\ 0 & \text{otherwise.} \end{cases} \tag{71}$$

Now, $Y_n \in L^2(\mathbb{P})$ and $\{Y_n \mid n \in \mathbb{N}\}$ are pairwise independent $\Rightarrow$ the Y_n are uncorrelated. For all $n \in \mathbb{N}$, set:

$$T_n := \sum_{j=1}^n Y_j.$$

We want to first show that WLLN holds for $\{Y_n \mid n \in \mathbb{N}\}$.

$$\begin{aligned} \mathrm{Var}[T_n] &= \mathbb{E}\left[(T_n - \mathbb{E}[T_n])^2\right] \\ &= \sum_{j=1}^n \mathbb{E}\left[(Y_j - \mathbb{E}[Y_j])^2\right] \text{ (because the covariance is zero for } i \neq j \\ &\le \sum_{j=1}^n \mathbb{E}\left[Y_j^2\right] \end{aligned}$$

The goal is to show that $\sum_{j=1}^n \mathbb{E}\left[Y_j^2\right]$ is $o(n^2)$. Let's see some approaches that will fail:

1. Try:

$$\sum_{j=1}^n \mathbb{E}\left[Y_j^2\right] \le \sum_{j=1}^n j^3 = O(n^3).$$

   This one is not good enough.
2. Try:

$$\sum_{j=1}^n \mathbb{E}\left[Y_j^2\right] \le \sum_{j=1}^n j\mathbb{E}\left[Y_j^2\right] \le \sum_{j=1}^n j\mathbb{E}\left[|X_j|\right] = \mathbb{E}\left[|X_1|\right] \sum_{j=1}^n j = O(n^2).$$

   This is better, but still not good enough.

We are not there yet. In particular, we are not using in the second inequality that all of the X_j's are integrable, so we know that the tails are shrinking. This is how we're supposed to use that.

$$\begin{aligned} \sum_{j=1}^n \mathbb{E}\left[Y_j^2\right] &\le \sum_{j=1}^n \mathbb{E}\left[X_j^2; |X_j| \le j\right] \\ &= \underbrace{\sum_{j=1}^{\sqrt{n}} \mathbb{E}\left[X_j^2; |X_j| \le j\right]}_{(A1)} + \underbrace{\sum_{j=\sqrt{n}+1}^n \mathbb{E}\left[X_j^2; |X_j| \le \sqrt{n}\right]}_{(A2)} + \underbrace{\sum_{j=\sqrt{n}+1}^n \mathbb{E}\left[X_j^2; \sqrt{n} < |X_j| \le j\right]}_{(A3)} \end{aligned}$$

To bound (A1)-(A3):

$$(A1) \leq \sum_{j=1}^{\sqrt{n}} j\mathbb{E}\left[|X_1|\right] = O(n) = o(n^2).$$

$$(A2) \leq \sum_{j=\sqrt{n}+1}^{n} \sqrt{n}\mathbb{E}\left[|X_1|\right] = O(n^{3/2}) = o(n^2).$$

$$(A3) \leq \sum_{j=\sqrt{n}+1}^{n} j\mathbb{E}\left[|X_j|; |X_j| > \sqrt{n}\right] = \underbrace{\mathbb{E}\left[|X_1|; |X_1| \geq \sqrt{n}\right]}_{\to 0 \text{ as } n\to\infty} \underbrace{\left(\sum_{j=\sqrt{n}+1}^{n} j\right)}_{O(n^2)} = o(n^2)$$

Therefore $\text{Var}[T_n] = o(n^2)$. By Chebychev's inequality, for all $\varepsilon > 0$:

$$\mathbb{P}\left(\frac{(T_n - \mathbb{E}\left[T_n\right]}{n} > \varepsilon\right) \to 0 \text{ as } n \to \infty.$$

Hence,

$$\frac{T_n - \mathbb{E}\left[T_n\right]}{n} \to 0 \text{ in probability.}$$

Now, we need to TT_n to show the desired result:

$$\begin{aligned}\left|\frac{S_n}{n} - m\right| &= \frac{S_n - nm}{n} \\ &\leq \underbrace{\frac{|S_n - T_n|}{n}}_{(B1)} + \underbrace{\frac{|T_n - \mathbb{E}\left[T_n\right]|}{n}}_{(B2)} + \underbrace{\frac{|\mathbb{E}\left[T_n\right] - nm|}{n}}_{(B3)}\end{aligned}$$

Claim: X_n and Y_n are equivalent. The proof follows from the result from Homework 3.

$$\sum_{n=1}^{\infty} \mathbb{P}\left(X_n \neq Y_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(X_n > n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(|X_1| > n\right) < \infty$$

because $\mathbb{E}\left[|X_1|\right] < \infty$. Since X_n and Y_n are equivalent, $(B1) \to 0$ almost surely and $(B2) \to 0$ in probability as proven. All we need to do is work on (B3):

$$\begin{aligned}(B3) &= \frac{1}{n}\left|\sum_{j=1}^{n}(\mathbb{E}\left[Y_j\right] - \mathbb{E}\left[X_j\right])\right| \\ &= \frac{1}{n}\left|\sum_{j=1}^{n}\mathbb{E}\left[Y_j - X_j\right]\right| \\ &= \frac{1}{n}\left|\sum_{j=1}^{n}\mathbb{E}\left[|X_j|; |X_j| > j\right]\right| \\ &\leq \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[|X_j|; |X_j| > j\right] \\ &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[|X_1|; |X_1| > j\right].\end{aligned}$$

Now, if $\{a_n \mid n \in \mathbb{N}\} \subseteq \mathbb{R}$ such that $\lim_{n\to\infty} a_n = 0$, then $\frac{1}{n}\sum_{j=1}^n a_j \to 0$ as $n \to \infty$. To see this, for all $\varepsilon > 0$, there exists an $N > 0$ such that $|a_n| < \varepsilon$ for all $n \in \mathbb{N}$:

$$\left|\frac{1}{n}\sum_{j=1}^n a_j\right| \le \underbrace{\frac{1}{n}\sum_{j=1}^N |a_j|}_{\to 0} + \underbrace{\frac{1}{n}\sum_{j=N+1}^n |a_j|}_{\le \varepsilon},$$

which can be made arbitrarily small when n is large. Since $X_1 \in L^1(\mathbb{P})$, $\lim_{n\to\infty} \mathbb{E}\left[|X_1|; |X_1| > n\right] = 0 \Rightarrow \frac{1}{n}\sum_{j=1}^n \mathbb{E}\left[|X_1|; |X_1| > j\right] \to 0$ as $n \to \infty$. Thus, $(B3) \to 0$ as $n \to \infty$. Therefore,

$$\left|\frac{S_n}{n} - m\right| \to 0$$

in probability. □

Theorem 30 (WLLN 3 (Kolmogorov & Feller)). *Let X_n be a sequence of pairwise independent random variables on Ω. Assume that there exists a sequence $\{b_n \mid n \in \mathbb{N}\}$ with $b_n \uparrow \infty$ such that:*

1. $\sum_{j=1}^n \mathbb{P}\left(|X_j| > b_n\right) \to 0$ *as* $n \to \infty$.
2. $\frac{1}{b_n^2}\sum_{j=1}^n \mathbb{E}\left[X_j^2; |X_j| \le b_n\right] \to 0$ *as* $n \to \infty$.

Then, WLLN holds in the general setting for X_n i.e., if

$$a_n = \sum_{j=1}^n \mathbb{E}\left[X_j; |X_j| \le b_n\right],$$

then $\frac{S_n - a_n}{b_n} \to 0$ in probability.

Proof. This is HW 4 Problem 2. Here's a hint: For all $n \ge 1$, $\forall\ j = 1, ..., n$, set

$$Y_{n,j} = \chi_{\{|X_j| \le b_n\}} X_j = \begin{cases} X_j & \text{if } |X_j| \le b_n \\ 0 & \text{otherwise} \end{cases}$$

Set $T_n = \sum_{j=1}^n Y_{n,j}$. First prove that $\frac{T_n - \mathbb{E}[T_n]}{b_n} \to 0$ in probability. □

Lemma 7 (Kronecker's Lemma). Let $\{x_k \mid k \in \mathbb{N}\} \subseteq \mathbb{R}$, $\{a_k \mid k \in \mathbb{N}\} \subseteq \mathbb{R}^+$ with $a_k \uparrow \infty$. Then,

$$\sum_{n=1}^\infty \frac{x_n}{a_n} \text{ converges } \Rightarrow \frac{1}{a_n}\sum_{j=1}^n x_j \to 0 \text{ as } n \to \infty.$$

Proof. Do later. □

Lemma 8 (Kolmogorov's Inequality). Let X_n be a sequence of independent random variables with $\mathbb{E}[X_n] = 0$ and $\mathbb{E}\left[X_n^2\right] < \infty$ for all $n \in \mathbb{N}$. Then,

$$\mathbb{P}\left(\max_{1\le j\le n} |S_n| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\mathbb{E}\left[S_n^2\right] \tag{72}$$

Proof. Write $A := \{\max_{1\leq j\leq n}|S_j| > \varepsilon\}$. Then,

$$A = \bigcup_{j=1}^{n} A_j,$$

where $A_j = \{|S_i| < \varepsilon \text{ for all } 1 \leq i \leq j-1, |S_j| > \varepsilon\}$. In words, A_j means the first time that $|S_n|$ goes above ε happens at $n = j$. The A_j's are disjoint. Now,

$$\begin{aligned}
\mathbb{E}\left[S_n^2\right] &\geq \mathbb{E}\left[S_n^2; A\right] \\
&= \sum_{j=1}^{n} \mathbb{E}\left[S_n^2; A_j\right] \\
&= \sum_{j=1}^{n} \mathbb{E}\left[(S_n - S_j + S_j)^2; A_j\right] \\
&= \sum_{j=1}^{n} \mathbb{E}\left[(S_n - S_j)^2; A_j\right] + 2\sum_{j=1}^{n} \mathbb{E}\left[(S_n - S_j)S_j\chi_{A_j}\right] + \sum_{j=1}^{n} \mathbb{E}\left[S_j^2; A_j\right] \\
&\geq \sum_{j=1}^{n} \mathbb{E}\left[S_j^2; A_j\right] \\
&\geq \varepsilon^2 \sum_{j=1}^{n} \mathbb{P}\left(A_j\right) \\
&= \varepsilon \mathbb{P}\left(A\right).
\end{aligned}$$

This proves that

$$\mathbb{P}\left(A\right) = \mathbb{P}\left(\max_{1\leq j\leq n}|S_j| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[S_n^2\right]$$

□

Theorem 31. *If $\{Y_n \mid n \in \mathbb{N}\}$ is a sequence of independent random variables and $\mathbb{E}\left[Y_n\right] = 0$ and $\sum_{n=1}^{\infty} \mathbb{E}\left[Y_n^2\right] < \infty$. Then, $\sum_{n=1}^{\infty} Y_n$ converges almost surely. That is, if*

$$T_n := \sum_{j=1}^{n} Y_j, \tag{73}$$

then $\lim_{n\to\infty} T_n$ exists in $\mathbb{R}$ almost surely.

Proof. We'll use Kolmogorov's Inequality. Fix $N > 0$, and apply Kolmogorov's inequality to the sequence starting from N: $\{Y_{N+j} \mid j \in \mathbb{N}\}$. Set:

$$K_m := \sum_{j=1}^{m} Y_{N+j} \text{ for } m \in \mathbb{N}.$$

By Kolmogorov's Inequality, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{1\leq j\leq n}|K_j| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[K_n^2\right] = \frac{1}{\varepsilon^2}\sum_{k=1}^{n} \mathbb{E}\left[Y_{N+j}^2\right]$$

Notice that $\max_{1\le j\le n}|K_j| \uparrow \sup_{j\ge 1}|K_j|$. So, we will use the continuity of probability to change the sup from the max, and apply the bounds above. So, for all $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{j\ge 1}|K_j| > \varepsilon\right) &= \lim_{n\to\infty}\mathbb{P}\left(\max_{1\le j\le n}|K_j| > \varepsilon\right)\\ &\le \frac{1}{\varepsilon^2}\sum_{j=1}^{\infty}\mathbb{E}\left[Y_{N+j}^2\right]\\ &= \frac{1}{\varepsilon^2}\sum_{j=N+1}^{\infty}\mathbb{E}\left[Y_j^2\right]\end{aligned}$$

this goes to zero as $N\to\infty$ by the second integrability condition. Thus, set:

$$W_N := \sup_{n\ge N+1}\left|\sum_{k=N+1}^{n} Y_k\right|.$$

Then, for all $\varepsilon > 0$, $\mathbb{P}(W_n > \varepsilon) \to 0$ as $N\to\infty$. Next set:

$$W'_N := \sup_{n'\ge n\ge N+1}\underbrace{\left|\sum_{k=n}^{n'} Y_k\right|}_{|T_{n'}-T_n|}.$$

By the triangle inequality $W'_N \le 2W_N$. Therefore,

$$W'_N \to 0 \text{ in probability.} \tag{74}$$

Since W'_N is decreasing in N, we have a monotonic sequence converging in probability which means that it converges almost surely. That is $\{T_n \mid n\in\mathbb{N}\}$ is a Cauchy sequence almost surely. Hence,

$$\lim_{n\to\infty} T_n \text{ exists in } \mathbb{R} \text{ almost surely.}$$

□

Theorem 32 (SLLN2 (Kolmogorov))**.** *Let $\{X_n \mid n\in\mathbb{N}\}$ be a sequence of independent random variables with $\mathbb{E}\left[X_n^2\right] < \infty$ for all $n\in\mathbb{N}$ and if there exists a sequence $\{b_n \mid n\in\mathbb{N}\}\subseteq\mathbb{R}^+$ with $b_n\uparrow\infty$ such that*

$$\sum_{n=1}^{\infty}\frac{\mathrm{Var}[X_n]}{b_n^2} < \infty.$$

Then, the SLLN holds in the sense that:

$$\frac{S_n - \mathbb{E}[S_n]}{b_n} \to 0 \text{ almost surely.} \tag{75}$$

Before the proof, we remark that if $\mathrm{Var}[X_n]$ is bounded in n, then we could take b_n to be n, which reduces us to (SLLN 1).

Proof. Set $Y_n := \frac{X_n - \mathbb{E}[X_n]}{b_n}$ for all $n\in\mathbb{N}$. Then, $\{Y_n \mid n\in\mathbb{N}\}$ is independent and $\mathbb{E}[Y_n] = 0$ and $\sum_{n=1}^{\infty}\mathbb{E}\left[Y_n^2\right] = \sum_{n=1}^{\infty}\frac{\mathrm{Var}[X_n]}{b_n} < \infty$. By the previous theorem, this tells us that

$$\sum_{n=1}^{\infty} Y_n = \sum_{n=1}^{\infty}\frac{X_n - \mathbb{E}[X_n]}{b_n}$$

converges almost surely. By Kronecker's Lemma,

$$\lim_{n\to\infty}\frac{1}{b_n}\sum_{j=1}^{n}(X_j-\mathbb{E}[X_j])=0 \text{ almost surely.}$$

□

Theorem 33 (SLLN 3 (Kolmogorov))**.** *Let $\{X_n\}$ be a sequence of iid random variables. Then,*

1. *If $\mathbb{E}[|X_1|]<\infty$, then,*

$$\frac{S_n}{n}\to\mathbb{E}[X_1] \text{ almost surely.} \tag{76}$$

2. *If $\mathbb{E}[|X_1|]=\infty$ then:*

$$\limsup_{n\to\infty}\frac{|S_n|}{n}=\infty \text{ almost surely} \tag{77}$$

Proof. (ii) Assume that $\mathbb{E}[|X_1|]=\infty$. Then, for all $A>0$,

$$\mathbb{E}\left[\frac{|X_1|}{A}\right]=0. \tag{78}$$

Then, by the homework, this implies that:

$$\begin{aligned}&\Rightarrow\sum_{n=1}^{\infty}\mathbb{P}(|X_1|>An)=\infty\\&\Rightarrow\sum_{n=1}^{\infty}\mathbb{P}(|X_n|>An)=\infty \text{ (by the iid condition)}\end{aligned}$$

Since the X_n's are independent, by (BC2), we can conclude that $\mathbb{P}(|X_n|>An \text{ i.o.})=1$. We have:

$$\begin{aligned}\{|X_n|>An\}&=\{|S_n-S_{n-1}|\}\\&\subseteq\left\{|S_n|>\frac{An}{2}\right\}\cup\left\{|S_{n-1}|>\frac{An}{2}\right\}\\&\subseteq\left\{|S_n|>\frac{A}{2}n\right\}\cup\left\{|S_{n-1}|>\frac{A}{2}(n-1)\right\}\end{aligned}$$

Since we matched the indices, this means that

$$\begin{aligned}&\Rightarrow\{|X_n|>An \text{ i.o. }\}\subseteq\left\{|S_n|>\frac{A}{2}n \text{ i.o.}\right\}\\&\Rightarrow\mathbb{P}\left(\frac{|S_n|}{n}>\frac{A}{2} \text{ i.o.}\right)\end{aligned}$$

Hence, $\limsup_{n\to\infty}\frac{|S_n|}{n}\geq\frac{A}{2}$ almost surely. Since A is arbitrary, this means that $\limsup_{n\to\infty}\frac{|S_n|}{n}=0$ almost surely.

(i). Assume that $\mathbb{E}[|X_1|]<\infty$. Truncate the X_n:

$$Y_n=X_n\chi_{\{|X_n|\leq n\}}$$

Let's check that they are equivalent:

$$\sum_{n=1}^{\infty} \mathbb{P}\left(X_n \neq Y_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(|X_n| > n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(|X_1| > n\right) < \infty,$$

because $\mathbb{E}\left[|X_1|\right] < \infty$. This shows that $\{Y_n\}$ is equivalent to $\{X_n\}$. $\{Y_n\}$ is also independent, and $\mathbb{E}\left[Y_n^2\right] < \infty$. Now bound:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{\mathrm{Var}[Y_n]}{n^2} &\leq \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[Y_n^2\right]}{n^2} \\
&= \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[X_n^2; |X_n| \leq n\right]}{n^2} \\
&\leq \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=1}^{n} \mathbb{E}\left[|X_n|^2; j-1 \leq |X_n| \leq j\right] \\
&= \sum_{j=1}^{\infty} \frac{1}{n^2} \sum_{j=1}^{n} \mathbb{E}\left[|X_1|^2; j-1 \leq |X_1| \leq j\right] \\
&= \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} \frac{1}{n^2} \mathbb{E}\left[|X_1|^2; j-1 \leq |X_1| \leq j\right] \\
\Rightarrow \ \exists C > 0 \text{ s.t. } &\leq C \sum_{j=1}^{\infty} \frac{1}{j} \mathbb{E}\left[|X_1|^2; j-1 \leq |X_1| \leq j\right] \\
&\leq C \sum_{j=1}^{\infty} \frac{1}{j} j \mathbb{E}\left[|X_1|; j-1 \leq |X_1| \leq j\right] \\
&= C\mathbb{E}\left[|X_1|\right] \\
&< \infty.
\end{aligned}$$

Hence, $\sum_{n=1}^{\infty} \frac{\mathrm{Var}[Y_n]}{n^2} < \infty$. Since the Y_n's are independent, the (SLLN) holds for the Y_n, i.e., if $T_n = \sum_{j=1}^{n} Y_j$, then by SSLN 2,

$$\frac{T_n - \mathbb{E}\left[T_n\right]}{n} \to 0 \text{ almost surely}$$

Now we do the same thing we did last time we truncated.

$$\frac{|S_n - \mathbb{E}\left[S_n\right]|}{n} \leq \frac{|S_n - T_n|}{n} + \frac{|T_n - \mathbb{E}\left[T_n\right]|}{n} + \frac{|\mathbb{E}\left[T_n\right] - \mathbb{E}\left[S_n\right]|}{n}$$

The first term in the bound will go to 0 almost surely because $\{Y_n\}$ and $\{X_n\}$ are equivalent. The second term we just proved goes to 0 almost surely. For the third term,

$$\begin{aligned}
\frac{|\mathbb{E}\left[T_n\right] - \mathbb{E}\left[S_n\right]|}{n} &\leq \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}\left[|X_j|; |X_j| > j\right] \\
&= \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}\left[|X_1|; |X_1| > j\right].
\end{aligned}$$

This will tend to zero, since $\mathbb{E}\left[|X_1|; |X_1| > n\right] \to 0$ as $n \to \infty$. Combining all of this, this shows that $\frac{S_n}{n} \to \mathbb{E}\left[X_1\right]$ almost surely. $\square$

Theorem 34 (SLLN 4)**.** *Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of independent random variables and assume that $\mathbb{E}\left[|X_n|\right] < \infty$ for all $n \in \mathbb{N}$. Assume that there exists a function $\varphi : \mathbb{R} \to \mathbb{R}^+$ continuous, even, such that $\frac{\varphi(x)}{x}$ is increasing on $]0, \infty[$ and $\frac{\varphi(x)}{x^2}$ is decreasing on $]0, \infty[$, and there exists a sequence $\{b_n\} \subseteq \mathbb{R}^+$ with $b_n \uparrow \infty$ such that:*

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\left[\varphi(X_n)\right]}{\varphi(b_n)} < \infty.$$

Then,

$$\sum_{n=1}^{\infty} \frac{X_n - \mathbb{E}\left[X_n\right]}{b_n}$$

converges almost surely and hence $\frac{S_n - \mathbb{E}[S_n]}{b_n} \to 0$ almost surely.

Proof. Homework. □

Theorem 35 (Weierstrass Theorem)**.** *Let $f : [0, 1] \to \mathbb{R}$ be a continuous function. For every $n \in \mathbb{N}$, define:*

$$p_n(x) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}, \tag{79}$$

for $x \in [0, 1]$. $p_n(x)$ is called the ***Bernstein Polynomial Associated with*** *f. Then,*

$$\lim_{n\to\infty} \sup_{x\in[0,1]} |p_n(x) - f(x)| = 0. \tag{80}$$

In other words, f is uniformly *approximated by Bernstein's Polynomials.*

Proof. First let's establish pointwise convergence. The convergence at the end points is trivial:

1. $\lim_{n\to\infty} p_n(0) = f(0)$.
2. $\lim_{n\to\infty} p_n(1) = f(1)$.

Fix an $x \in]0, 1[$. Consider the following sequence of iid Bernoulli random variables $\{X_n \mid n \in \mathbb{N}\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}\left(X_1 = 1\right) = x$ and $\mathbb{P}\left(X_1 = 0\right) = 1 - x$. If $S_n = \sum_{j=1}^{n} X_j$, then

$$\mathbb{P}\left(S_n = k\right) = \binom{n}{k} x^k (1-x)^{n-k}.$$

Hence, taking the expected value, we get:

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^{n} f\left(\frac{n}{k}\right)\mathbb{P}\left(S_n = k\right) = p_n(x).$$

Since f is continuous, we know that by the SLLN (or could use WLLN),

$$\frac{S_n}{n} \to x \text{ a.s. } \Rightarrow f\left(\frac{S_n}{n}\right) \to f(x) \text{ a.s}$$

Now use the Dominated Convergence Theorem (which we can use since f is continuous on $[0, 1]$ and so it's bounded on $[0, 1]$. Hence,

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] \to f(x). \tag{81}$$

Hence, for all $x \in [0,1]$, we have $\lim_{n\to\infty} p_n(x) = f(x)$ i.e., we've established pointwise convergence. Now note that f is uniformly continuous on $[0,1]$. Hence,

$$\forall\, \varepsilon > 0,\ \exists\, \delta > 0 \text{ s.t. } |x-y| \le \delta \Rightarrow |f(x) - f(y)| \le \varepsilon.$$

Hence,

$$\begin{aligned}
|p_n(x) - f(x)| &= \left|\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] - f(x)\right| \\
&\le \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right] \\
&= \underbrace{\mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|; \left|\frac{S_n}{n} - x\right| > \delta\right]}_{(A1)} + \underbrace{\mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|; \left|\frac{S_n}{n} - x\right| \le \delta\right]}_{(A2)}
\end{aligned}$$

(A2) is automatically less than ε by the uniform continuity of f on $[0,1]$. For **(A1)** we do the same trick as before:

$$\textbf{(A1)} \le 2M\mathbb{P}\left(\left|\frac{S_n}{n} - x\right| > \delta\right)$$

where $M = \sup_{x\in[0,1]}$. By Chebychev, we can bound this as:

$$M = \sup_{x\in[0,1]} |f(x)| \le \frac{2M\text{Var}[\frac{S_n}{n}]}{\delta^2} = \frac{2M\text{Var}[S_n]}{n^2\delta^2} = 2M\frac{nx(1-x)}{n^2\delta^2}.$$

This quantity goes to 0 as $n \to \infty$. Hence, we can conclude:

$$\begin{aligned}
\sup_{x\in[0,1]} |p_n(x) - f(x)| &\le \frac{2M}{n\delta^2}\left(\sup_{x\in[0,1]} x(1-x)\right) + \varepsilon \\
&\le \frac{M}{2n\delta^2} + \varepsilon,
\end{aligned}$$

which again can be made arbitrarily small when n is sufficiently large. Hence,

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

□

Example 15. Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

$$\mathbb{P}(X_1 = 1) = p \text{ and } \mathbb{P}(X_1 = 0) = 1 - p,$$

for $p \in]0,1[$. Next, define a random variable $T : \Omega \to [0,1]$ as follows:

$$T(\omega) = \sum_{n=1}^{\infty} \frac{X_n(\omega)}{2^n}.$$

Denote by μ_p the distribution of T, i.e., μ_p is a probability measure on $([0,1], \mathcal{B}([0,1]))$. Denote by F_p the distribution function of T.

Claim: F_p is continuous on $\mathbb{R}$, i.e., the map $\mu_p(\{x\}) = 0$ for all $c \in [0,1]$. Set Q to be the set of Dyadic rationals:

$$Q = \{m2^n \mid n \in \mathbb{N}, m = 0, ..., 2^n\}. \tag{82}$$

There are two cases:

1. $c \in Q$ and $T = c$ for every $\omega \in \{T(\omega) = c\}$, then either:
   (a) $X_n(\omega) = 1$ for all but finitely many n.
   (b) $X_n(\omega) = 0$ for all but finitely many n.

   But, $\mathbb{P}\left(\text{ either of the above events happen}\right) = 0$ because $\lim_{n\to\infty}(p \vee (1-p))^n = 0$. Hence, $\mathbb{P}\left(T = c\right) = 0$ which shows that $\mu_p(\{c\}) = 0$.
2. Suppose c is not a dyadic rational. Then, for a given $n \in \mathbb{N}$, $\exists_1\ m \in \{0, ..., 2^n\}$ such that:

$$(m-1)2^{-n} < c < m2^{-n}.$$

Hence,

$$\{T = c\} \subseteq \{\omega \in \Omega \mid X_1(\omega) = \xi_1, ..., X_n(\omega) = \xi_n\},$$

for a unique $(\xi_1, ..., \xi_n) \in \{0, 1\}^n$. Hence,

$$\mathbb{P}\left(T = c\right) \leq (p \vee (1-p))^n \to 0 \text{ as } n \to \infty.$$

Hence, $\mu_p(\{c\}) = 0$ which shows that F_p is continuous.
(Do the rest of this example later).

Theorem 36 (Levy's Equivalence Theorem)**.** *Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of independent random variables. Then, $S_n \to S$ in probability for some random variable $S \iff S_n \to S$ almost surly.*

Proof. The "$\Leftarrow$" direction is trivial, so all we need to do is the "$\Rightarrow$" direction. Assume that $S_n \to S$ in probability. Then, for all $\varepsilon > 0$, there exists an $N \geq 0$ such that for all $n \geq N$,

$$\mathbb{P}\left(|S_n - S| > \varepsilon/2\right) \leq \frac{\varepsilon}{2}.$$

Hence, by the triangle inequality for probabilities, for all $m, n \geq N$,

$$\mathbb{P}\left(|S_m - S_n| > \varepsilon\right) \leq \mathbb{P}\left(|S_n - S| > \varepsilon/2\right) + \mathbb{P}\left(|S_m - S| > \varepsilon/2\right) \leq \varepsilon.$$

Hence, $\{S_n \mid n \in \mathbb{N}\}$ is a Cauchy Sequence in the sense of probability. Now assume that $m > n$. Then,

$$\begin{aligned}
\varepsilon &\geq \mathbb{P}\left(|S_m - S_n| > \varepsilon\right) \\
&\geq \mathbb{P}\left(|S_m - S_n| > \varepsilon,\ \max_{n \leq k \leq m} |S_k - S_n| \geq 2\varepsilon\right) \\
&= \sum_{k=n+1}^{m} \mathbb{P}\left(|S_m - S_n| > \varepsilon,\ \max_{n \leq j \leq k-1} |S_j - S_n| < 2\varepsilon,\ |S_k - S_n| > 2\varepsilon\right)
\end{aligned}$$

We want to make use of independence, but the indices overlap. So we need a smaller event. Hence,

$$\begin{aligned}
&\geq \sum_{k=n+1}^{m} \mathbb{P}\left(|S_m - S_k| > \varepsilon,\ \max_{n \leq j \leq k-1} |S_j - S_n| < 2\varepsilon,\ |S_k - S_n| > 2\varepsilon\right) \\
&= \sum_{k=1}^{m} \mathbb{P}\left(|S_m - S_k| < \varepsilon\right) \mathbb{P}\left(\max_{n \leq j \leq k+1} |S_j - S_n| \leq 2\varepsilon, |S_k - S_n| > 2\varepsilon\right) \\
&\geq (1-\varepsilon)\mathbb{P}\left(\max_{n \leq j \leq k+1} |S_j - S_n| \leq 2\varepsilon, |S_k - S_n| > 2\varepsilon\right) \\
&= (1-\varepsilon)\mathbb{P}\left(\max_{n \leq k \leq m} |S_k - S_n| > 2\varepsilon\right).
\end{aligned}$$

Re-arranging, this gives us:

$$\mathbb{P}\left(\max_{n\leq\leq m}|S_k - S_n| > 2\varepsilon\right) \leq \frac{\varepsilon}{1-\varepsilon} \Rightarrow \mathbb{P}\left(\sup_{k\geq n}|S_k - S_n| > 2\varepsilon\right) \leq \frac{\varepsilon}{1-\varepsilon} \tag{83}$$

$$\Rightarrow \sup_{k\geq n}|S_k - S_n| \to 0 \text{ in probability.} \tag{84}$$

Following the same proof as in the case of a previous theorem, we can then conclude that S_n is a Cauchy Sequence almost surely. Hence, $S_n \to S$ almost surely. □

5 Product Space

Let (S_1, Σ_1) and (S_2, Σ_2) be two measurable spaces. Define $S = S_1 \times S_2 = \{s = (s_1, s_2) \mid s_i \in S_i, i = 1, 2\}$. Define the **product sigma algebra** as:

$$\Sigma = \Sigma_1 \otimes \Sigma_2 := \sigma(\{B_1 \times B_2 \mid B_i \in \Sigma_i, i = 1, 2\}).$$

Here, $B_1 \times B_2$. are rectangles. This set of rectangles is a generating π-system. For $i = 1, 2$, $\rho_i : S \to S_i$ is the **coordinate map**, $\rho_i(s_1, s_2) = s_i$ for all $(s_1, s_2) \in S$. This reduces the dimension. Some remarks:

1. $\rho_i : S \to S_i$ is $\Sigma \setminus \Sigma_i$-measurable. In other words, one has:

$$\forall\ B_1 \in \Sigma_1, \varphi_1^{-1}(B_1) = B_1 \times S_2 \in \Sigma.$$

   and,

$$\forall\ B_2 \in \Sigma_2, \varphi^{-1}(B_2) = S_1 \times B_2 \in \Sigma.$$

2. $\Sigma = \sigma(\rho_1, \rho_2)$ the product sigma algebra is the smallest sigma-algebra with respect to which ρ_1 and ρ_2 are measurable. To quickly see this, note that LHS $\supseteq$ RHS, as seen in the remark above. For the other inclusion, we need to show that all the elements of the generating sigma-algebra are in RHS. For all $B_i \in \Sigma_i$, $i = 1, 2$,

$$B_1 \times B_2 = \rho_1^{-1}(B_1) \cap \rho_2^{-1}(B_2) \in \sigma(\rho_1, \rho_2).$$

   This is what we wanted to show.

Lemma 9. If $f : S \to \mathbb{R}$ is Σ-measurable, then:

1. $\forall\ s_1 \in S_1$, the function $s_2 \in S_2 \mapsto f(s_1, s_2) \in \mathbb{R}$ is Σ_2-measurable.
2. $\forall\ s_2 \in S_2$, the function $s_1 \in S_1 \mapsto f(s_1, s_2) \in \mathbb{R}$ is Σ_1-measurable.

In other words, measurability with respect to the product σ-algebra $\Rightarrow$ measurability with respect to the coordinate σ-algebra.

Proof. We prove this using the Monotone Class Theorem. Set:

$$H := \{f \in m\Sigma \mid \text{ (i) and (ii) hold for } f\}. \tag{85}$$

Claim: H is a monotone class of measurable functions. To see this:

1. H is a vector space $\to$ obvious ✓.
2. $1 \in H$: $\to$ measurable with respect to everything ✓.
3. $f_n \in H$, $f_n \geq 0$, $f_n \uparrow f \Rightarrow f \in H \to$ limit of measurable functions is measurable.

Next, we need to show that the basic roots are in H. Call:

$$I := \{B_1 \times B_2 \mid B_i \in \Sigma_i \text{ for } i = 1, 2\}.$$

Clearly, I is a π-system generating Σ. Now, for all $A \in I$, assume that $A = B_1 \times B_2$. Then, it's the product of indicator functions:

$$\chi_A(s_1, s_2) = \chi_{B_1}(s_1) \cdot \chi_{B_2}(s_2).$$

Obviously, $\chi_A \in H$. So, by the monotone class theorem, we may conclude that $m\Sigma \subseteq H$. Hence, $m\Sigma = H$. □

Preparation: for $i = 1, 2$, assume that μ_i is a finite measure on (S_i, μ_i). Given $f : S \to \mathbb{R}$, if $f \in (m\Sigma)^+$ or $f \in b\Sigma$, define the following two functions:

1. for $s_1 \in S_1$, define the integral where we integrate the second coordinate out:
$$I_1^f(s_1) := \int_{S_2} f(s_1, s_2)\mu_2(dS_2). \tag{86}$$
2. for $s_2 \in S_2$, define the integral where we integrate the first coordinate out:
$$I_2^f(s_2) := \int_{S_1} f(s_1, s_2)\mu_1(dS_1). \tag{87}$$

We are ready to state and prove the first version of Fubini's Theorem.

Lemma 10. If $f \in (b\Sigma)$, then $I_i^f \in b\Sigma_i$ for $i = 1, 2$. Moreover,

$$\int_{S_1} I_1^f(s_1)\mu_1(dS_1) = \int_{S_2} I_2^f(s_2)\mu_2(dS_2). \tag{88}$$

Explicitly, the order of doing the integral doesn't matter:

$$\int_{S_1} \left(\int_{S_2} f(s_1, s_2)\mu_2(dS_2) \right) \mu_1(dS_1) \underbrace{=}_{(*)} \int_{S_2} \left(\int_{S_1} f(s_1, s_2)\mu_1(dS_1) \right) \mu_2(dS_2). \tag{89}$$

Proof. The proof will follow the same strategy as before. Set:

$$S := \{f \in b\Sigma \mid I_i^f \in b\Sigma_i \text{ for } i = 1, 2, \text{ and } (*) \text{ holds for } f\}. \tag{90}$$

We need to check that H is a Monotone Class of bounded functions (we do not need to worry about the integral being infinite).

1. H is a vector space: ✓by the linearity of integrals.
2. $1 \in H$: ✓since both sides of $(*)$ are equal to $\mu_1(S_1)\mu_2(S_2)$.
3. If $f_n \in H$ and $f_n \geq 0$, $f_n \uparrow f$, then $f \in H$. This one follows from the Monotone Convergence Theorem (MON).

Let's check if we have the right roots. Let $A = B_1 \times B_2$. Then,

$$\chi_A(s_1, s_2) = \chi_{B_1}(s_1)\chi_{B_2}(s_2). \tag{91}$$

Then, for all $s_1 \in S_1$,

$$I_1^{\chi_A}(s_1) = \int_{S_2} \chi_A(s_1, s_2)\mu_2(dS_2) = \chi_{B_1}(s_2)\mu_2(B_2). \tag{92}$$

Similarly, for all $s_2 \in S_2$:

$$I_2^{\chi_A}(s_2) = \int_{S_1} \chi_A(s_1, s_2)\mu_1(dS_1) = \mu_1(B_1)\chi_{B_2}(s_2). \tag{93}$$

This shows that $I_1^{\chi_A} \in b\Sigma_i$ and (*) holds. Hence, $\chi_A \in H$. By the MCT, $b\Sigma =$. □

Corrolary 3. If $f \in (m\Sigma)^+$, then $I_i^f \in (m\Sigma_i)^+$ for $i = 1, 2$ and (*) holds for f.

Theorem 37 (Fubini's Theorem)**.** *Let (S_i, Σ_i, μ_i), $i = 1, 2$ be two measure spaces. Let μ_1 and μ_2 be finite masures. We define the following set function on Σ:*

$$\mu : \Sigma \to [0, \infty[$$

by: for all $A \in \Sigma$,

$$\mu(A) := \int_{S_1} S_1^{\chi_A}(s_1)\mu_1(dS_1) = \int_{S_2} I_2^{\chi_A}(s_2)\mu_2(dS_2). \tag{94}$$

Then,

1. *μ is a finite measure on (S, Σ) denoted by $\mu = \mu_1 \times \mu_2$. This is called the* ***product measure****.*
   (a) *We write $(S, \Sigma, \mu) = (S_1, \Sigma_1, \mu_1) \times (S_2, \Sigma_2, \mu_2)$.*
2. *μ is the unique measure on S such that:*
$$\forall B_i \in \Sigma_i,\ i = 1, 2,\ \mu(B_1 \times B_2) = \mu_1(B_1) \cdot \mu_2(B_2).$$
3. ***(Tonelli's Theorem):*** *if $f \in (m\Sigma)^+$, then:*
$$\boxed{\mu(f) \underbrace{=}_{(\dagger)} \int_{S_1} I_1^f(S_1)\mu_1(dS_1) \underbrace{=}_{(*)} \int_{S_2} I_2^f(S_2)\mu_2(dS_2).} \tag{95}$$
4. ***(Fubini's Theorem):*** *if $f \in L^1(\mu)$, then $I_1^f \in L^1(\mu_1)$ and $I_2^f \in L^2(\mu_2)$. Moreover,*
$$\boxed{\mu(f) \underbrace{=}_{(\dagger)} \mu_1(I_1^f) \underbrace{=}_{(\triangle)} \mu_2(I_2^f)} \tag{96}$$
(This statement means that integrability with respect to the product measure $\Rightarrow$ integrability with respect to the coordinate measures (product measure is stronger than coordinate measures).

Proof. 1. We need to show that μ is a measure first. It's clear that $\mu(\emptyset) = 0$. Let $\{A_n \mid n \in \mathbb{N}\} \subseteq \Sigma$ be disjoint. For $m \geq 1$, set $B_m := \bigcup_{u=1}^m A_n$. Then, by the disjointness,

$$\chi_{B_m} = \sum_{n=1}^m \chi_{A_n}.$$

Moreover,

$$\chi_{B_m} \uparrow \sum_{n=1}^\infty \chi_{A_n} = \chi_{\bigcup_{n=1}^\infty A_n}.$$

So, by definition, for all $s_1 \in S_1$, we have:

$$I_1^{\chi_{\bigcup_{n=1}^\infty A_n}}(s_1) = \int_{S_2} \chi_{\bigcup_{n=1}^\infty A_n}(s_1, s_2)\mu_2(dS_2).$$

By the Monotone Convergence Theorem (MON) and by the linearity of integration, this becomes:

$$= \sum_{n=1}^\infty \int_{S_2} \chi_{A_n}(s_1, s_2)\mu_2(dS_2).$$

Therefore,

$$\mu\left(\bigcup_{n=1}^\infty A_n\right) = \int_{S_1} I_1^{\chi_{\bigcup_{n=1}^\infty A_n}}(s_1)\mu_1(dS_1) \underbrace{=}_{(MON)} \sum_{n=1}^\infty \int_{S_1}\int_{S_2} \chi_{A_n}(s_1, s_2)\mu_2(dS_2) = \sum_{n=1}^\infty \mu(A_n).$$

2. Now we need to prove that μ is the *unique* measure. Suppose μ' is another measure on (S, Σ) such that:

$$\mu'(B_1 \times B_2) = \mu_1(B_1) \cdot \mu_2(B_2),$$

for all $B_i \in \Sigma_i$ for $i = 1, 2$. Note that

$$\mu(S) = \mu(S_1 \times S_2) = \mu_1(S_1) \cdot \mu_2(S_2) = \mu'(S).$$

Since $\mu(S) = \mu(S_1 \times S_2) = \mu_1(S_1)\mu_2(S_2)$ on a π-system, $I = \{B_1 \times B_2 \mid B_i \in \Sigma_i \text{ for } i = 1, 2\}$, by the Uniqueness of Measure Theorem, $\mu = \mu'$ on the entire sigma algebra $\Sigma = \sigma(I)$.

3. **Tonelli's Theorem**: $(\triangle)$ has already been proven; all that's left to establish is $(\dagger)$. First, $(\dagger)$ holds for indicator functions. Next, by the linearity of integrals, $(\dagger)$ holds for $(SF)^+$. Next, by **(MON)**, $(\dagger)$ holds for all $f \in (m\Sigma)^+$.
4. If $f \in L^1(\mu)$, then $\mu(|f|) < \infty$. Hence, $\mu(f^{\pm}) < \infty$. By **(iii)**, $(\dagger)$ holds for $f^{\pm}$, i.e.,

$$\mu(f^{\pm}) = \int_{S_1} I_1^{f^{\pm}}(s_1)\mu_1(dS_1).$$

Now,

$$\mu(f^{\pm}) < \infty \iff I_1^{f^{\pm}} \in L^1(\mu_1) \Rightarrow I_1^{f^{\pm}} < \infty \ \mu_1\text{-a.e..}$$

Similarly, $I_2^{f^{\pm}} \in L^1(\mu_2)$ and $I_2^{\pm} < \infty$ μ_2-a.e. Now write $f = f^+ - f^-$. Then, $(\dagger)$ and $(\triangle)$ follow from **(Linearity)**.

$\square$

Some Remarks.

1. Fubini's Theorem also applies to μ_i being σ-finite. Just cut the space up. For example, take a sequence of rectangles $\{T_1^n \times T_2^n \mid n \in \mathbb{N}\}$ such that $T_i^n \in \Sigma_i$ for $i = 1, 2$ such that $T_1^n \times T_2^n \uparrow S_1 \times S_2$ and μ_i is finite on T_i^n for all $n \in \mathbb{N}$ for $i = 1, 2$.
2. $f \in m\Sigma \Rightarrow f(s_1, \cdot) \in m\Sigma_2$ and $f(\cdot, s_2) \in m\Sigma_1$. In general, the opposite implication does not necessarily hold.
3. $f \in L^1(\mu) \Rightarrow I_1^f \in L^1(\mu_1)$ and $I_2^f \in L^2(\mu_2)$. In general, the reverse implication does not necessarily hold.

Consider the following counter-example to illustrate this point: set $S = \{(x, y) \mid x - 1 < y < x + 1\}$. Then,

$$f(x, y) = \begin{cases} 1 & \text{if } x < y < x + 1 \\ -1 & \text{if } x - 1 \le y \le x \\ 0 & \text{otherwise.} \end{cases}$$

Then, for all $x \in \mathbb{R}$, $\int_{\mathbb{R}} f(x, y)dy = 0 \Rightarrow I_1^f \equiv 0 \Rightarrow I_1^f \in L^1(dx)$. Similarly, for all $y \in \mathbb{R}$, $\int_{\mathbb{R}} f(x, y)dy = 0 \Rightarrow I_2^f \equiv 0 \Rightarrow I_2^f \in L^1(dy)$. But, $f \notin L^1(dxdy)$ because:

$$\int_{\mathbb{R}^2} |f(x, y)|dxdy = \infty. \tag{97}$$

Hence, integrability on the product space is a strictly stronger condition than integrability on the coordinates. Before going back to probability spaces, remark that

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda_{\text{Leb}}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)^n.$$

5.1 Back to Probability Space

Definition 34 (Joint Distribution Function)**.** Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then, $F_{(X,Y)}$ is the **joint distribution function** of (X, Y) if $\forall (x, y) \in \mathbb{R}^2$,

$$F_{(X,Y)}(x, y) = \mathbb{P}(X \leq x, Y \leq y). \tag{98}$$

Some terminology:

1. $\mathcal{L}_{(X,Y)}$ is the joint distribution of (X, Y). It's a probability measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ such that for all $A \in \mathcal{B}(\mathbb{R}^2)$,
$$\mathcal{L}_{(X,Y)}(A) = \mathbb{P}((X, Y) \in A)$$
2. If $\mathcal{L}_{(X,Y)}$ is absolutely continuous with respect to $\lambda^2_{\text{Leb}} = dxdy$, then the RN derivative,
$$f_{(X,Y)} = \frac{d\mathcal{L}(X, Y)}{dxdy}, \tag{99}$$
is the joint probability density function of (X, Y).

Proposition 28. If (X, Y) has the joint probability function $f_{(X,Y)}$ then for all $x \in \mathbb{R}$,

$$f_X(x) := \int_{\mathbb{R}} f_{(X,Y)}(x, y)dy \tag{100}$$

is a probability density of X. Moreover, for all $y \in \mathbb{R}$,

$$f_Y(y) := \int_{\mathbb{R}} f_{(X,Y)}(x, y)dx \tag{101}$$

is a probability density of Y.

Proof. These are marginal densities. For all $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
\mathcal{L}_X(B) &= \mathbb{P}(X \in B) \\
&= \mathbb{P}(X \in B, Y \in \mathbb{R}) \\
&= \iint_{B\times\mathbb{R}} f_{(X,Y)}(x, y)dxdy \text{ (by the existence of the joint density)} \\
&= \int_B \left(\int_{\mathbb{R}} f_{(X,Y)}(x, y)dy \right) dx \text{ (Tonelli's theorem)}
\end{aligned}$$

Therefore, for all $B \in \mathcal{B}(\mathbb{R})$,

$$\mathcal{L}_X(B) = \int_B f_X(x)dx.$$

□

Proposition 29. Assume X and Y have distributions $\mathcal{L}_{(X,Y)}$, X has distribution $\mathcal{L}_X$ with distribution function F_X, and Y has distribution $\mathcal{L}_Y$ with distribution function F_Y. Then, TFAE:

1. X and Y are independent.
2. For all $x, y \in \mathbb{R}$,
$$F_{(X,Y)}(x, y) = F_X(x)F_Y(y). \tag{102}$$
3. $\mathcal{L}_{(X,Y)} = \mathcal{L}_X \times \mathcal{L}_Y$.

In particular, if (X, Y) has a joint density function $f_{(X,Y)}$, then (i)-(iii) are equivalent to

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y).$$

Proof. The key idea for the proof of "(ii) $\Rightarrow$ (iii)" is that $\{]-\infty, x]\times]-\infty, y] \mid x, y \in \mathbb{R}\}$ is a generating π-system of $\mathcal{B}(\mathbb{R}^2)$. □

5.1.1 Product of Infinitely Many Spaces

Let $\{(\Omega_i, \mathcal{F}_i, \mathbb{P}_i) \mid i \in \mathbb{N}\}$ be a sequence of probability spaces. Set $\Omega = \prod_{i=1}^{\infty} \Omega_i$. Then, for all $\omega \in \Omega$, ω is an infinite vector: $\omega = (\omega_1, \omega_2, \omega_3, ...)$ where $\omega_n \in \Omega_n$. Consider the following "cylinder sets":

$$E := \prod_{n=1}^{\infty} F_n, \text{ where } F_n \in \mathcal{F}_n \ \forall n \in \mathbb{N},$$

and $F_n = \Omega_n$ for all but finitely many n, i.e.,

$$E = F_1 \times F_2 \times F_3 \times ... \times F_n \times \Omega_{n+1} \times \Omega_{n+2} \times ...$$

Set:

$$\Sigma_0 = \left\{ \bigcup_{k=1}^{K} E^{(k)} \;\middle|\; K \in \mathbb{N} \text{ and } E^{(k)} \text{ are disjoint cylinder sets} \right\}$$

Σ_0 is an algebra. Set $\mathcal{F} = \sigma(\Sigma_0)$. Let's define a candidate measure. Define the following set function of Σ_0:

$$E \in \Sigma_0 \text{ set: } E := \bigcup_{k=1}^{K} E^{(k)},$$

where $E^{(k)} = \prod_{n=1}^{\infty} F_n^{(k)}$. Then,

$$\mathbb{P}(E) = \sum_{k=1}^{K} \prod_{n=1}^{\infty} \mathbb{P}_n(F_n^{(k)})$$

One can verify that indeed $\mathbb{P}$ can be extended to $\mathcal{F}$ as a measure. Check that $\mathbb{P}$ is **additive** on Σ_0 and that it's **continuous at the empty set**. One would use Caratheodory's Extension Theorem to check this (if we wanted to). The first one to check would be relatively easy, but checking the second condition requires more work.

Theorem 38 (Kolmogorov Extension Theorem)**.** *Let $\{\mu_n \mid n \in \mathbb{N}\}$ be a sequence of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $\{X_n \mid n \in \mathbb{N}\}$ on Ω such that $\forall\ n \in \mathbb{N}$,*

$$\mathcal{L}_{X_n} = \mu_n,$$

and the $\{X_n \mid n \in \mathbb{N}\}$ are independent.

What's significant about this is that we can put the probability measures on the *same* probability space *and* they can be independent. This has applications in optimal control.

Proof. For all $n \in \mathbb{N}$, there exists a probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ and random variables Y_n on Ω_n such that $\mathcal{L}_{Y_n} = \mu_n$. Set our candidate probability space to be the infinite product,

$$(\Omega, \mathcal{F}, \mathbb{P}) = \prod_{n=1}^{\infty} (\Omega_n, \mathcal{F}_n, \mathbb{P}_n).$$

For all random vectors $\omega \in \Omega$, it has components $\omega = (\omega_1, \omega_2, \omega_3, ...)$ with components $\omega_n \in \Omega_n$. For $n \in \mathbb{N}$, define:

$$X_n : \Omega \to \mathbb{R},$$
$$\omega = (\omega_1, \omega_2, \omega_3, ...) \mapsto X_n(\omega) = Y_n(\omega_n).$$

Hence, X_n is a random variable, and for all $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} X_n^{-1}(B) &= \{\omega \in \Omega \mid X_n(\omega) \in B\} \\ &= \Omega_1 \times \Omega_2 \times ... \times \Omega_{n-1} \times Y_n^{-1}(B) \times \Omega_{n+1} \times ... \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{P}\left(X_n^{-1}(B)\right) &= \mathbb{P}\left(\prod_{j=1}^{n-1} \Omega_j \times Y_n^{-1}(B) \times \prod_{j \geq n+1} \Omega_j\right) \\ &= \mathbb{P}_n(Y_n^{-1}(B)) \\ &= \mu_n(B). \end{aligned}$$

This shows that $\mathcal{L}_{X_n} = \mu_n$. Now we need to show independence. For all $k \in \mathbb{N}$, take $B_1, B_2, ..., B_k \in \mathcal{B}(\mathbb{R})$. Then, for all $1 \leq n_1 < n_2 < ... < n_k \in \mathbb{N}$, one has:

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^{k} X_{n_i}^{-1}(B_i)\right) &= \mathbb{P}\left(\Omega_1 \times \Omega_2 \times ... \times \Omega_{n_1-1} \times Y_{n_1}^{-1}(B_1) \times \Omega_{n_1+1} \times ... \times \Omega_{n_2-1} \times Y_{n_2}^{-1}(B_2) \times ...\right) \\ &= \prod_{i=1}^{k} \mathbb{P}_{n_i}(Y_{n_i}^{-1}(B_i)) \\ &= \prod_{i=1}^{k} \mu_i(B_i) \\ &= \prod_{i=1}^{k} \mathbb{P}\left(X_{n_i} \in B_i\right). \end{aligned}$$

□

Theorem 39 (Kolmogorov's Extension Theorem)**.** *For $n \in \mathbb{N}$, let $\mu^{(n)}$ be a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. For $1 \leq m \leq n$, let $\prod_{m,n}$ be the extension map from $\mathcal{B}(\mathbb{R}^m)$ to $\mathcal{B}(\mathbb{R}^n)$ for all $B \in \mathcal{B}(\mathbb{R}^m)$:*

$$\prod_{n,m}(B) := \{(x_1, ..., x_n) \mid (x_1, ..., x_m) \in B\} \tag{103}$$

$$= B \times \underbrace{\mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R}}_{n-m \text{ copies}} \tag{104}$$

If $\{\mu^{(n)} \mid n \in \mathbb{N}\}$ is **consistent**, i.e., for all $n \in \mathbb{N}$, $\forall 1 \leq m \leq n$,

$$\mu^{(n)} \circ \prod_{m,n} = \mu^{(m)},$$

then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{X_n \mid n \in \mathbb{N}\}$ on Ω such that:

$$\mathcal{L}_{(X_1,...,X_n)} = \mu^{(n)}$$

for all $n \in \mathbb{N}$.

Remark: remark that the previous theorem is a particular case of Kolmogorov's Extension Theorem with

$$\mu^{(n)} = \mu_1 \times ... \times \mu_n \text{ for all } n \in \mathbb{N}.$$

In this case, the $\{X_n\}$ are independent.

Example 16. Let $\mu^{(n)}$ be a centred Gaussian probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ with covariance matrix,

$$C^{(n)} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2 \\ \vdots & & \ddots & 3 \\ 1 & 2 & \dots & n \end{bmatrix} = i \wedge j \text{ for } 1 \leq i, j \leq n.$$

i.e., $\mu^{(n)} = N(0, C^{(n)})$. Verify that $\{\mu^{(n)} \mid n \in \mathbb{N}\}$ is consistent. By Kolmogorov's Extension theorem, there exists $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{X_n \mid n \in \mathbb{N}\}$ on Ω such that $\mathcal{L}_{(X_1,\dots,X_N)} = \mu^{(n)}$ (this is the discrete Brownian motion).

6 Conditioning and Martingales

6.1 Conditional Expectation

Recall conditional probability: Given $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, the conditional expectation was defined as:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

This is going to be the general reasoning, and now we will try to extend this idea.

Definition 35 (Conditional Expectation)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X \in L^1$ and $\mathcal{G} \subseteq \mathcal{F}$ be a sub-sigma algebra. Then, a random variable Y on Ω is called a **conditional expectation** of X given $\mathcal{G}$, denoted by $Y = \mathbb{E}[X|\mathcal{G}]$ if:

1. $Y \in L^1$.
2. $Y \in m\mathcal{G}$.
3. For every event $A \in \mathcal{G}$,

$$\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}. \tag{105}$$

This is *not* an expectation; it's a random variable! The information, encoded by $\mathcal{G}$, will change the randomness. The new random variable, which is given by the conditional expectation, will re-predict the randomness. It will be a new prediction on X based on the new information in $\mathcal{G}$.

Remark. Some remarks:

1. It's possible to define $\mathbb{E}[X|\mathcal{G}]$ even if $X \notin L^1$. In fact, we only need $\int_A X d\mathbb{P}$ to exist for all $A \in \mathcal{G}$.
2. Condition **(iii)** can be replaced by condition **(iii)'**: for all $A \in I$,

$$\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}, \tag{106}$$

   where I is a π-system and $\mathcal{G} = \sigma(I)$.

Proposition 30. Assume $X_1, X_2 \in L^1$ and $X_1 \leq X_2$ almost surely. Suppose $\mathcal{G} \subseteq \mathcal{F}$ is a sub-sigma algebra. Then, if $Y_1 = \mathbb{E}[X_1|\mathcal{G}]$ and $Y_2 = \mathbb{E}[X_2|\mathcal{G}]$, then $Y_1 \leq Y_2$ almost surely.

Proof. Everything is integrable, so everything is almost everywhere finite. Hence, we can subtract with no issues. Set $A := \{Y_1 > Y_2\}$. First, $A \in \mathcal{G}$. Second,

$$\int_A Y_1 d\mathbb{P} = \int_A X_1 d\mathbb{P} \leq \int_A X_2 d\mathbb{P} = \int_A Y_2 d\mathbb{P},$$

where the inequality follows from standard monotonicity. Hence,

$$\begin{aligned} \int_A (Y_1 - Y_2) d\mathbb{P} \leq 0 &\Rightarrow \mathbb{P}(A) = 0 \\ &\Rightarrow Y_1 \leq Y_2 \text{ (almost surely)} \end{aligned}$$

□

Corrolary 4. Given $X \in L^1$, $\mathcal{G} \subseteq \mathcal{F}$ a sub sigma-algebra. If the conditional expectation of X given $\mathcal{G}$ exists, then it must be unique almost surely, and it will be denoted by $\mathbb{E}[X|\mathcal{G}]$.

Proof. **1. Existence**: We'll show two proofs: the first method is the standard proof which is seen in textbooks, and the second method will look at the problem from a different angle (from a functional analysis point of view). The second angle will be very nice to know, since it will help a lot with **Q3** on the homework.

1. Method #1: define two set functions $\mu_g^{\pm}$ by:

$$\forall\ A \in \mathcal{G},\ \mu_g^{\pm}(A) := \int_A X^{\pm} d\mathbb{P}.$$

It's easy to verify that $\mu_g^{\pm}$ are two measures on $(\Omega, \mathcal{G})$. Furthermore, these two measures are both absolutely continuous with respect to $\mathbb{P}|_{\mathcal{G}}$. Therefore, by the RN-theorem, there exists a $Y^{\pm} \in m\mathcal{G}$ such that $Y^{\pm} = \frac{d\mu_g^{\pm}}{d\mathbb{P}}$. Hence, for all $A \in \mathcal{G}$,

$$\int_A X^{\pm} d\mathbb{P} = \mu_g^{\pm}(A) = \int_A Y^{\pm} d\mathbb{P}.$$

Obviously, $Y^{\pm} \in L^1$ (by taking $A = \Omega$ in the above since $X \in L^1$). Hence, $Y^{\pm} = \mathbb{E}[X^{\pm}|\mathcal{G}]$. By the linearity of integrals, we see that $\mathbb{E}[X|G] = Y^+ - Y^-$.

2. Method #2: look at the problem from a functional analysis point of view. First, assume that $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. This is a Hilbert Space with an inner product given by the L^2-norm. We consider:

$$L^2(\mathcal{G}) = \{Z \in m\mathcal{G} \mid \mathbb{E}\left[Z^2\right] < \infty\}.$$

Obviously, $L^2(\mathcal{G}) \subseteq L^2(\mathcal{F})$ is a sub-Hilbert Space. Set,

$$P_g : L^2(\mathcal{F}) \to L^2(\mathcal{G})$$

to be the projection onto the subspace $L^2(\mathcal{G})$. (We are projecting elements from the bigger space onto the smaller space). For all $X \in L^2(\mathcal{F})$, $X - P_g(X)$ is perpendicular to $L^2(\mathcal{G})$. Hence, for all $W \in L^2(\mathcal{F})$,

$$(X - P_g(X), W)_{L^2} = 0.$$

Since for all $A \in \mathcal{G}$, $\chi_A \in L^2(\mathcal{G})$, one has:

$$\int_\Omega (X - P_g(X))\chi_A d\mathbb{P} = 0 \Rightarrow \int_A X d\mathbb{P} = \int_A P_g(X) d\mathbb{P}.$$

For the integrability conditions, since $P_g(X) \in L^2(\mathcal{G})$, $P_g(X) \in m\mathcal{G}$. Therefore, when $X \in L^2$,

$$\mathbb{E}[X|\mathcal{G}] = P_g(X).$$

Now, for a general $X \in L^1(\mathbb{P})$, take $X_k^{\pm} = X^{\pm} \wedge k$ for every $k \in \mathbb{N}$. Then, $X_k^{\pm} \in L^2$ and $X_k^{\pm} \uparrow X^{\pm}$. Set:

$$Y_k^{\pm} := \mathbb{E}\left[X_k^{\pm}|\mathcal{G}\right] = P_g(X_k^{\pm}).$$

Last time we proved **(cMON)** (monotonicity for conditional expectation), so this let's us set

$$Y^{\pm} := \lim_{k\to\infty} Y_k^{\pm}.$$

Finally, as an exercise, verify that for general $X \in L^1$,

$$\mathbb{E}\left[X|\mathcal{G}\right] = Y^+ - Y^-.$$

$\square$

Some examples now.

Example 17. If $\mathcal{G} = \sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$, with $A \in \mathcal{F}$ and $\mathbb{P}(A) \in]0,1[$. Then, heuristically, we'd expect that for every $X \in L^1$,

$$\mathbb{E}\left[X|\mathcal{G}\right] = \chi_A \frac{\mathbb{E}\left[X;A\right]}{\mathbb{P}(A)} + \chi_{A^c}\frac{\mathbb{E}\left[X;A^c\right]}{\mathbb{P}(A^c)},$$

where the $\mathbb{E}\left[X;A\right]$ corresponds to making a prediction, and the $\mathbb{P}(A)$ corresponds to normalizing.

Example 18. If $\mathcal{G} = \sigma(Y)$ (the information of $\mathcal{G}$ is provided by another random variable) for some $Y \in m\mathcal{F}$. Let X be a random variable and let h be a Borel function such that $h(X) \in L^1$. Write:

$$\mathbb{E}\left[h(X)|\mathcal{G}\right] = \mathbb{E}\left[h(X)|Y\right].$$

First, since $\mathbb{E}\left[h(X)|Y\right] \in m\sigma(Y)$, there exists an $H : \mathbb{R} \to \mathbb{R}$ Borel such that

$$\mathbb{E}\left[h(X)|Y\right] = H(Y).$$

Now assume that the pair (X,Y) has the joint PDF function $f_{(X,Y)}$ and let f_Y be the probability density function of Y. Next, we define the following probability function:

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

This is called the **conditional probability density**. This does not always exist! If this exists, then if $H : \mathbb{R} \to \mathbb{R}$ is given by,

$$H(y) := \int_{\mathbb{R}} h(x) f_{X|Y}(x,y)dx$$

for all $y \in \mathbb{R}$, then $\mathbb{E}\left[h(X)|Y\right] = H(Y)$.

Proof. 1. H is a Borel function, by the preparation for Fubini theorem section, so we have $H(Y) \in m\Sigma(Y)$ ✓

2. To see that $H(Y) \in L^1(\mathbb{P})$, it's sufficient to check that $H \cdot f_y \in L^1(dy)$. By Tonelli's theorem,

$$\begin{aligned}
\int_{\mathbb{R}} |H(y)| d_Y(y) dy &\leq \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |h(x)| f_{X|Y}(x,y) dx \right) f_Y(y) dy \\
&= \iint_{\mathbb{R}^2} |h(x)| \underbrace{f_{X|Y}(x,y) f_Y(y)}_{\text{joint density}} dx dy \\
&= \iint_{\mathbb{R}^2} |h(x)| f_{(X,Y)}(x,y) dx dy \\
&= \mathbb{E}\left[|h(X)|\right] \\
&< \infty \text{ which gives us integrability.}
\end{aligned}$$

3. For all $A \in \sigma(Y)$, there exists a $B \in \mathcal{B}(\mathbb{R})$ such that $A = Y^{-1}(B)$. Knowing this way of representing A, we write:

$$\begin{aligned}
\mathbb{E}\left[h(X); A\right] &= \mathbb{E}\left[h(X)\chi_B(Y)\right] \\
&= \iint_{\mathbb{R}^2} h(x)\chi_B(y) f_{(X,Y)}(x,y) dx dy \\
&= \int_B \left(\int_{\mathbb{R}} h(x) f_{X|Y}(x,y) dx \right) f_Y(y) dy \\
&= \int_B H(y) f_Y(y) dy \\
&= \mathbb{E}\left[H(Y); A\right].
\end{aligned}$$

where we can do the third step by Fubini's theorem, which we can use since integrability has been checked.

□

Hence, a conditional expectation has randomness, which entirely comes from Y.

6.1.1 Properties of Conditional Expectation $\mathbb{E}\left[X|\mathcal{G}\right]$

Assume $X \in L^1$, and $\mathcal{G} \subseteq \mathcal{F}$ is a sub-sigma algebra. Then,

1. if $Y = \mathbb{E}\left[X|\mathcal{G}\right]$, then $\mathbb{E}\left[Y\right] = \mathbb{E}\left[X\right]$ i.e.:

$$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{G}\right]\right] = \mathbb{E}\left[X\right].$$

2. if $X \in m\mathcal{G}$, then $\mathbb{E}\left[X|G\right] = X$ (you can make the perfect prediction).
3. **cLin (conditional linearity)**: if $X, Y \in L^1$ and $a, b \in \mathbb{R}$, then:

$$\mathbb{E}\left[aX + bY|\mathcal{G}\right] = a\mathbb{E}\left[X|\mathcal{G}\right] + b\mathbb{E}\left[Y|\mathcal{G}\right].$$

4. **cMON (conditional monotonicity)**: if $X, Y \in L^1$ and $X \leq Y$ a.s., then:

$$\mathbb{E}\left[X|\mathcal{G}\right] \leq \mathbb{E}\left[Y|\mathcal{G}\right].$$

5. **cJen (Conditional Jensen's Inequality)**: Suppose that φ is a convex function such that $\varphi(x) \in L^1$. Then,

$$\mathbb{E}\left[\varphi(x)|\mathcal{G}\right] \geq \varphi(\mathbb{E}\left[X|\mathcal{G}\right]).$$

The proof is basically the same as the standard Jensen's inequality with the support line. For almost every $\omega \in \Omega$ fixed, take the supporting line at the point $(\mathbb{E}[X|\mathcal{G}](\omega), \varphi(\mathbb{E}[X|\mathcal{G}](\omega))$. For $y = ax+b$, $\varphi(x) \geq ax+b$ for all $x \in \mathbb{R}$. So, by **(cMON)**:

$$\mathbb{E}[\varphi(x)|\mathcal{G}] \geq a\mathbb{E}[X|\mathcal{G}] + b.$$

So, pointwise, this means:

$$\mathbb{E}[\varphi(x)|\mathcal{G}](\omega) \geq a\mathbb{E}[X|\mathcal{G}](\omega) + b = \varphi(\mathbb{E}[X|\mathcal{G}](\omega)).$$

6. **(cMON) Conditional Monotonicity)**: assume $\{X_n \mid n \in \mathbb{N}\} \subseteq L^1$ and $X \in L^1$. If $X_n \uparrow X$, then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$.

   Proof. Since $\mathbb{E}[X_n|\mathcal{G}]$ is a monotonically increasing sequence, the limit $Y := \lim_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}]$ exists. By Fatou's Lemma,

$$\begin{aligned}\mathbb{E}[|Y|] &\leq \liminf_{n\to\infty} \mathbb{E}[|\mathbb{E}[X_n|\mathcal{G}]|] \\ &\leq \liminf_{n\to\infty} \mathbb{E}[|X_n|] \\ &\leq \mathbb{E}\left[X^+\right] + \mathbb{E}\left[X_1^-\right] \\ &< \infty \Rightarrow Y \in L^1.\end{aligned}$$

   The rest of the proof follows from the classical monotone convergence theorem. We need to check that for all $A \in \mathcal{G}$:

$$\begin{aligned}\int_A X d\mathbb{P} \underbrace{=}_{(MON)} \lim_{n\to\infty} \int_A X_n d\mathbb{P} \\ &= \lim_{n\to\infty} \int_A \mathbb{E}[X_n \mid \mathcal{G}] d\mathbb{P} \\ &\underbrace{=}_{(MON)} \int_A Y d\mathbb{P}.\end{aligned}$$

   □

7. **(cMON')**: if $\{X_n \mid n \in \mathbb{N}\} \subseteq L^1$ and $X \in L^1$. If $X_n \downarrow X$, then $\mathbb{E}[X_n \mid \mathcal{G}] \downarrow \mathbb{E}[X|\mathcal{G}]$.
8. **(cFatou)**: if $\{X_n \mid n \in \mathbb{N}\} \subseteq L^1$ and assume that there exists a $Y \in L^1$ such that $X_n \geq Y$ for all $n \in \mathbb{N}$. Then,

$$\mathbb{E}\left[\liminf_n X_n|\mathcal{G}\right] \leq \liminf_n \mathbb{E}[X_n \mid \mathcal{G}]. \tag{107}$$

   Proof. Recall in the classical setting how we proved Fatou's Lemma: set $Z_m := \inf_{n\geq m} X_n$ for all $m \in \mathbb{N}$. Then, $Z_m \uparrow \liminf_n X_n$. Moreover, $Y \leq Z_m \leq X_n$ for all $n \geq m$. So, $Z_m \in L^1$ for all $m \in \mathbb{N}$. Then apply **(cMON)** to the expected values to get that:

$$\mathbb{E}[Z_m \mid \mathcal{G}] \uparrow \mathbb{E}\left[\liminf_n X_n \mid \mathcal{G}\right].$$

   Moreover, by **(cMonotonicity)**:

$$\mathbb{E}[Z_m|\mathcal{G}] \leq \inf_{n\geq m} \mathbb{E}[X_n \mid \mathcal{G}].$$

   □

9. **(cFatou')**: if $\{X_n \mid n \in \mathbb{N}\} \subseteq L^1$, $Y \in L^1$ such that $X_n \leq Y$ for all $n \in \mathbb{N}$, then:

$$\mathbb{E}\left[\limsup_n X_n \mid \mathcal{G}\right] \geq \limsup_n \mathbb{E}\left[X_n \mid \mathcal{G}\right]. \tag{108}$$

10. **(cDOM)**: if $\{X_n \mid n \in \mathbb{N}\} \subseteq L^1$ and $Y \in L^1$ such that $|X_n| \leq Y$ for all $n \in \mathbb{N}$, and $X_n \to X$ a.s. for some random variable X, then $X \in L^1$ and:

$$\mathbb{E}\left[X_n \mid \mathcal{G}\right] \to \mathbb{E}\left[X|\mathcal{G}\right]$$

a.s. as well as in L^1.

Exercise: prove **(cDOM)** using **(cFatou)** and **(cFatou')**. We can see here that conditional expectation behaves in a similar way to classical expectation.

11. **(Tower Property)**: If $\mathcal{H}$ and $\mathcal{G}$ are two sub σ-algebras which are ordered, i.e., $\mathcal{H} \subseteq \mathcal{G}$, then for all $X \in L^1$,

$$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{H}\right]|\mathcal{G}\right] \underbrace{=}_{\triangle 1} \mathbb{E}\left[X|\mathcal{H}\right] \underbrace{=}_{\triangle 2} \mathbb{E}\left[\mathbb{E}\left[X|\mathcal{G}\right] \mid \mathcal{H}\right]. \tag{109}$$

Proof. Your constraint is going to be your smaller σ-algebra: your smallest σ-algebra determines everything.

(a) ($\triangle 1$): this is trivial because $\mathbb{E}\left[X|\mathcal{H}\right] \in m\mathcal{H} \subseteq m\mathcal{G}$.

(b) ($\triangle 2$): We only need to check this, so let's check against the definition we want to study. For all $A \in \mathcal{H} \subseteq \mathcal{G}$, we have:

$$\int_A \mathbb{E}\left[X|\mathcal{G}\right] d\mathbb{P} \underbrace{=}_{(1)} \int_A X d\mathbb{P} \underbrace{=}_{(2)} \int_A \mathbb{E}\left[X|\mathcal{H}\right] d\mathbb{P},$$

where in (1) we view $A \in \mathcal{G}$ and in (2) we view $A \in \mathcal{H}$. Thus,

$$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{G}\right]|\mathcal{H}\right] = \mathbb{E}\left[X|\mathcal{H}\right].$$

□

12. Suppose $Z \in m\mathcal{G}$ and $XZ \in L^1$. Then, $\mathbb{E}\left[XZ|\mathcal{G}\right] = Z \cdot \mathbb{E}\left[X|\mathcal{G}\right]$.

Proof. We want to verify that for all $A \in \mathcal{G}$,

$$\int_A X \cdot Z d\mathbb{P} \underbrace{=}_{(\dagger)} \int_A Z \cdot \mathbb{E}\left[X|\mathcal{G}\right] d\mathbb{P}.$$

(a) Go to the root of integrals: see how they behave on indicator functions. ($\dagger$) holds for $Z = \chi_B$ for all $B \in \mathcal{G}$ and $X \geq 0$.

(b) Use **(cLin)** and **(cMON)** to extend the statement to a general case.

□

13. If $\mathcal{H} \subseteq \mathcal{F}$ is another sub σ-algebra, and $\mathcal{H}$ is independent of $\sigma(\sigma(X) \cup \mathcal{G})$ (all the information available by using either X or $\mathcal{G}$ or both), then,

$$\mathbb{E}\left[X|\sigma(\mathcal{G} \cup \mathcal{H})\right] = \mathbb{E}\left[X|\sigma(\mathcal{G})\right].$$

Proof. Consider a π-system:

$$I := \{G \cap H \mid G \in \mathcal{G},\ H \in \mathcal{H}\}.$$

This is a generating π-system. It generates $\sigma(\mathcal{H} \cup \mathcal{G})$. Let $A \in I$ be of the form $A = G \cap H$. Then,

$$\int_A X d\mathbb{P} = \int_G \chi_X d\mathbb{P} \underbrace{=}_{(indep)} \mathbb{P}(H) \int_G X d\mathbb{P} = \mathbb{P}(H) \int_G \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} \underbrace{=}_{(indep)} \int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P}.$$

$\square$

Corr. if $\sigma(X)$ is independent of $\mathcal{H}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$.

Proof. In **(13)** , take $\mathcal{G} = \{\emptyset, \Omega\}$. Then, $\mathbb{E}[X|\sigma(\mathcal{G} \cup H)] = \mathbb{E}[X|\mathcal{H}]$ and $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$. $\square$

14. Assume $X_1, ..., X_N$ are independent and X_j has distribution $\mathcal{L}_{X_j}$ for $1 \leq j \leq n$. Let $h : \mathbb{R}^n \to \mathbb{R}$ be Borel such that $h(X_1, ..., X_n) \in L^1$. Then,

$$\mathbb{E}[h(X_1, ..., X_n)|X_1] = \gamma(X_1),$$

where we get γ by integrating out the $X_2, ..., X_n$:

$$\gamma(X) = \iint ... \iint_{\mathbb{R}^{n-1}} h(x_1, ..., x_n)(\mathcal{L}_{X_2} \times \mathcal{L}_{X_3} \times ... \times \mathcal{L}_{X_n})(dx_2 \cdots dx_n).$$

Exercise. Prove it using Fubini's Theorem and Independence.

6.2 Martingales

Definition 36 (Filtration)**.** Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, if $\{\mathcal{F}_n \mid n \geq 0\}$ is a sequence of sub σ-algebras such that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq ... \subseteq \mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq ... \subseteq \mathcal{F}$. Then, $\{\mathcal{F}_n \mid n \geq 1\}$ is called a **filtration**. The triple $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ is called a **filtered space**.

Definition 37 (Adapted with respect to a filtration)**.** Let $\{X_n \mid n \in \mathbb{N}\}$ be a sequence of random variables on a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$. $\{X_n\}$ is called **adapted with respect to the filtration** $\{\mathcal{F}_n\}$ if $X_n \in m\mathcal{F}_n$ for all $n \in \mathbb{N}$.

Definition 38 (Martingale)**.** Given a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ and a process $\{X_n \mid n \in \mathbb{N}\}$, the process is called a **martingale** with respect to the filtration $\{\mathcal{F}_n \mid n \geq 0\}$ if:

1. $\{X_n\}$ adapted with respect to $\{\mathcal{F}_n \mid n \geq 0\}$.
2. $X_n \in L^1$ for all $n \geq 0$.
3. $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$ for all $n \geq 0$.

- **Submartingale**: (1) and (2) are the same. For (3):

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n\ \forall n \geq 0. \tag{110}$$

- **Supermartingale**: (1) and (2) are the same. For (3):

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n\ \forall n \geq 0. \tag{111}$$

Remark that if $\{X_n \mid n \geq 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n \mid n \geq 0\}$, then for all $m \geq n$,

$$\mathbb{E}[X_m|\mathcal{F}_n] = X_n. \tag{112}$$

How does one prove this? You run a tower property argument and apply the definition of a martingale:

$$\mathbb{E}[X_m|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X_m|\mathcal{F}_{m-1}]\,|\mathcal{F}_n] = \mathbb{E}[X_{m-1}|\mathcal{F}_n] = ... = X_n.$$

Example 19. Let $\{Y_n \mid n \in \mathbb{N}\}$ be a sequence of indepndent random variables on a probability space such that $\mathbb{E}[Y_n] = 0$ for all $n \in \mathbb{N}$. Set $S_0 \equiv 0$, and set $S_n := \sum_{j=1}^{n} Y_j$. Set $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma(\{Y_1, ..., Y_n\})$. Then, $\{S_n \mid n \geq 0\}$ is a martingale with respect to $\{\mathcal{F}_n \mid n \geq 0\}$.

Proof. We can see this by using **(clin)** to break the sum up:

$$\begin{aligned}
\mathbb{E}[S_{n+1}|\mathcal{F}_n] &= \mathbb{E}[S_n + Y_{n+1} \mid \mathcal{F}_n] \\
&= \underbrace{\mathbb{E}[S_n|\mathcal{F}_n]}_{\text{measurable}} + \underbrace{\mathbb{E}[Y_{n+1}|\mathcal{F}_n]}_{\text{indep}} \\
&= S_n + \mathbb{E}[Y_{n+1}] \\
&= S_n \checkmark.
\end{aligned}$$

$\square$

In addition, if $\mathbb{E}\left[X_n^2\right] = 1$ for all $n \geq 1$, then we get a second martingale: $\{S_n^2 \ - n| \ n \geq 0\}$ is a martingale with respect to $\{\mathcal{F}_n \mid n \geq 0\}$. There will be no issues with (1) and (2), those are pretty straightforward to see. For (3), we just need to do some work:

$$\begin{aligned}
\mathbb{E}\left[S_{n+1}^2 - (n+1)|\mathcal{F}_n\right] &= \mathbb{E}\left[(S_n + Y_{n+1})^2|\mathcal{F}_n\right] - (n+1) \\
&= \mathbb{E}\left[S_n^2 \mid \mathcal{F}_n\right] + 2\mathbb{E}[S_n \cdot Y_n \mid \mathcal{F}_n] + \mathbb{E}\left[Y_{n+1}^2 \mid \mathcal{F}_{n+1}\right] - (n+1) \\
&= S_n^2 + 2S_{n+1}\mathbb{E}[Y_{n+1}] + \mathbb{E}\left[Y_{n+1}^2\right] - (n+1) \\
&= S_n^2 - n.
\end{aligned}$$

In addition, if $Y_n \sim N(0,1)$ random variable for all $n \geq 1$, then:

$$\left\{e^{tS_n - \frac{t^2}{2}n} \mid n \geq 0\right\},$$

is a martingale with respect to the same filtration $\{\mathcal{F}_n\}$.

Exercise: verify it yourself.

Example 20. If $X \in L^1$ and $\{\mathcal{F}_n \mid n \geq 0\}$ is a filtration and $X_n := \mathbb{E}[X|\mathcal{F}_n]$ for all $n \geq 0$. Then, this sequence of conditional expectations $\{X_n\}$ is a martingale with respect to $\mathcal{F}_n$. This follows from the Tower Property:

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[X|\mathcal{F}_n] = X_n \checkmark.$$

We remark that if $\{X_n\}$ is a sub-martingale $\iff$ $\{-X_n \mid n \geq 0\}$ is a super-martingale.

Theorem 40. *Given a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ consider a process $\{X_n\}$ adapted with respect to $\{\mathcal{F}_n\}$. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function such that $f(X_n) \in L^1$ for all $n \geq 0$. If either $\{X_n \mid n \geq 0\}$ is a martingale or $\{X_n \mid n \geq 0\}$ is a sub-martingale and f is increasing, then $\{f(X_n) \mid n \in \mathbb{N}\}$ is a sub-martingale.*

Proof. $\{f(x_n) \mid n \in \mathbb{N}\}$ is an adapted process. Then, for all $n \geq 0$, by **(cJensen)** :

$$\mathbb{E}[f(x_{n+1})|\mathcal{F}_n] \geq f(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]) = \begin{cases} = f(x_n) & \text{if martingale} \\ \geq f(x_n) & \text{if sub-martingale} \end{cases}$$

In either case, this shows that $\{f(x_n) \mid n \geq 0\}$ is a sub-martingale. $\square$

Corrolary 5. From this theorem, we get...

- If $\{X_n\}$ is a martingale, then $\{|X_n|^p\}$ is a sub-martingale for all $p \geq 1$.
- If $\{X_n\}$ is a sub-martingale, then $\{|X_n|^p\}$ $(p \geq 1)$ is a sub-martingale if $X_n \geq 0$ for all $n \geq 0$.
- If $\{X_n \mid n \geq 0\}$ is a sub-martingale, then $\{X_n^+ \mid n \geq 0\}$ is a sub-martingale.

The next theorem tells us how to extract the part of the game which makes it unfair to get a fair game.

Theorem 41 (Doob's Decomposition Theorem)**.** *Given $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ and a sub-martingale $\{X_n \mid n \geq 0\}$, there exists a process $\{Y_n\}$ such that:*

1. *$Y_0 \equiv 0$, $Y_n \in L^1$, and $Y_{n+1} \in m\mathcal{F}_n$ for all $n \geq 0$.*
   (a) *We call this property $\{Y_n\}$ being* ***pre-visible*** *with respect to the filtration $\{\mathcal{F}_n\}$.*
2. *$Y_{n+1} \geq Y_n$ for all $n \geq 0$.*
3. *$\{M_n := X_n - Y_n \mid n \geq 0\}$ is a martingale with respect to $\{\mathcal{F}_n\}$.*

Furthermore, such a process is almost surely unique.

Proof. First, we prove the uniqueness of $\{Y_n \mid n \geq 0\}$. Assume that there exists another process $\{W_n\}$ satisfying **(i)-(iii)**. Set:

$$\begin{aligned}
&\Delta_n := Y_n - W_n \ \forall n \in \mathbb{N} \\
&\Delta_0 \equiv 0 \text{ since they both start at zero} \\
&\Delta_{n+1} \in m\mathcal{F}_n \ \forall n \geq 0.
\end{aligned}$$

Then, $\Delta_n = (X_n - W_n) - (X_n - Y_n)$. So, $\{\Delta_n \mid n \geq 0\}$ is a martingale with respect to $\{\mathcal{F}_n\}$. Thus,

$$\Delta_{n+1} = \mathbb{E}\left[\Delta_{n+1} \mid \mathcal{F}_n\right] = \Delta_n,$$

for all $n \geq 0$. This shows that $\Delta_n = 0$ a.s. for all $n \geq 0$.

Now sum up the increment at every step. Set $Y_0 \equiv 0$. Then,

$$Y_n := \sum_{j=0}^{n-1} \left(\mathbb{E}\left[X_{j+1} | \mathcal{F}_j\right] - X_j\right) \text{ for all } n \geq 1.$$

We have that $Y_n \geq 0$, $Y_{n+1} \geq Y_n$, and $Y_{n+1} \in m\mathcal{F}_n$ for all $n \geq 0$. Now let's see what happens when I remove Y_n from X_n:

$$\begin{aligned}
\mathbb{E}\left[M_{n+1} | \mathcal{F}_n\right] &= \mathbb{E}\left[X_{n+1} - Y_{n+1} | \mathcal{F}_n\right] \\
&= \mathbb{E}\left[X_{n+1} | \mathcal{F}_n\right] - Y_{n+1} \\
&= \mathbb{E}\left[X_{n+1} | \mathcal{F}_n\right] - \left(\mathbb{E}\left[X_{n+1} | \mathcal{F}_n\right] - X_n\right) - Y_n \\
&= X_n - Y_n \\
&= M_n \ \forall n \in \mathbb{N}.
\end{aligned}$$

□

Hence, to conclude: given any sub-martingale, I can isolate the growing part from the martingale part.

6.3 Stopping Times

Definition 39 (Stopping Time)**.** Given $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$, a random variable $\tau : \Omega \to \{0, 1, 2, 3, ...\}$ $0 \leq \tau \leq \infty$ is a **stopping time** if for all $0 \leq n \leq \infty$, $\{\tau \leq n\} \in \mathcal{F}_n$. Note that $\mathcal{F}_\infty = \sigma\left(\bigcup_{n \geq 0} \mathcal{F}_n\right)$.

The motivation or heuristic meaning of this is that τ is the "stopping strategy": τ tells me the moment that I quit the game. The measurability condition in the definition tells me that I shouldn't need to look into the future to decide when I'm quitting the game: all that information should be contained in $\mathcal{F}_n$.

Remark. τ is a stopping time $\iff$ for all $0 \leq n \leq \infty$, $\{\tau = n\} \in \mathcal{F}_n$. We can very quickly give a justification to this:

Proof. "$\Rightarrow$": for all $n \geq 0$, we can write:

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n - 1\}$$

which is in $\mathcal{F}_n$ since $\mathcal{F}$ is a sigma algebra. For the infinite case:

$$\{\tau = \infty\} = \bigcap_{n \geq 0} \{\tau \geq n\} \in \mathcal{F}_\infty.$$

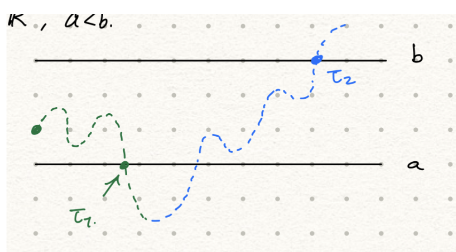
"$\Leftarrow$": for all $0 \leq n \leq \infty$:

$$\{\tau \leq n\} = \bigcup_{j=0}^{n} \{\tau = j\} \in \mathcal{F}_n.$$

$\square$

Example 21. Suppose $\{X_n \mid n \geq 0\}$ is adapted with respect to $\{\mathcal{F}_n \mid n \geq 0\}$. Let $a, b \in \mathbb{R}$ be such that $a < b$. Set:

- $\tau_0 \equiv 0$.
- $\tau_1 := \inf\{n \geq 0 \mid X_n \leq a\}$.
- $\tau_2 := \inf\{n \geq \tau_1 \mid X_n \geq b\}$.



τ_1 is a stopping time, because if $n \geq 0$, then:

$$\{\tau_1 = n\} = \{X_0 > a, X_1 > a, ..., X_{n-1} > a, X_n \leq a\} \in \mathcal{F}_n.$$

In the infinite case,

$$\{\tau_1 = \infty\} = \{X_n > a \ \forall n \geq 0\} \in \mathcal{F}_\infty.$$

τ_2 is also a stopping time; τ_1 and τ_2 is also a stopping time. For example, if $n \geq 0$:

$$\{\tau_2 = n\} = \bigcup_{j=0}^{n} \{\tau_1 = j, X_{j+1} < b, X_{j+2} < b, ..., X_{n-1} < b, X_n \geq b\} \in \mathcal{F}_n.$$

From τ_1 to τ_2, $\{X_n \mid n \geq 0\}$ completes the **first upcrossing from a to b**. Similarly, we can define:

- **"Start" of the kth upcrossing**: $\tau_{2k-1} := \inf\{n \geq \tau_{2k-2} \mid X_n \leq a\}$.
- **"End" of the kth upcrossing**: $\tau_{2k} := \inf\{n \geq \tau_{2k-1} \mid X_n \geq b\}$.

This gives us that $\{\tau_k \mid k \geq 0\}$ is a family of ordered stopping times.

Proposition 31. (Properties of Stopping Time)

1. τ is a stopping time. For all $0 \leq n \leq \infty$, the following are all stopping times: $\{\tau = n\}$, $\{\tau \geq n\}$, $\{\tau \leq n\}$, $\{\tau > n\}$, and $\{\tau < n\} \in \mathcal{F}_n$.
2. τ is a stopping time: for all $N > 0$ constant, $\tau \wedge N := \min\{\tau, N\}$ is a stopping time.
3. If τ_1 and τ_2 are two stopping times, then the following are all stopping times:
   (a) $\tau_1 \vee \tau_2 = \max\{\tau_1, \tau_2\}$
   (b) $\tau_1 + \tau_2$

Exercise: prove (2) and (3), by breaking up into the integer case and the infinite case.

Definition 40 (Sigma Algebra Generated By Stopping Time)**.** Given a $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ and a stopping time τ: set:

$$\mathcal{F}_\tau := \{A \in \mathcal{F} \mid A \cap \{\tau \leq n\} \in \mathcal{F}_n \ \forall 0 \leq n \leq \infty\}, \tag{113}$$

or equivalently,

$$\mathcal{F}_\tau := \{A \in \mathcal{F} \mid A \cap \{\tau = n\} \in \mathcal{F}_n \ \forall 0 \leq n \leq \infty\}.$$

Exercise: verify that $\mathcal{F}_\tau$ is a σ-algebra. We remark that $\sigma(\tau) \subset \mathcal{F}_\tau$, but in general this inclusion is *strict*. We can verify this inclusion as follows:

Proof. For all $0 \leq m \leq \infty$ $\{\tau = m\} \in \mathcal{F}_\tau$. For all $0 \leq n \leq \infty$:

$$\{\tau = m\} \cap \{\tau = n\} = \begin{cases} \{\tau = n\} \in \mathcal{F}_n & \text{if } n = m. \\ \emptyset & \text{otherwise} \end{cases}$$

□

Definition 41. Let $\{X_n \mid n \geq 0\}$ be an adapted process with respect to $\{\mathcal{F}_n\}$ and let τ be a stopping time. We define for every sample point $\omega \in \Omega$:

$$X_\tau(\omega) := \begin{cases} X_n(\omega) & \text{if } \tau = n \text{ for } n \geq 0. \\ \lim_{n\to\infty} X_n(\omega) & \text{if } \tau = \infty \text{ and } \lim_{n\to\infty} X_n(\omega) \text{ exists.} \\ \text{ undefined} & \text{if } \tau = \infty \text{ and limit DNE.} \end{cases} \tag{114}$$

Proposition 32. Given $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$, let X_n be adapted. Then,

$$X_\tau \cdot \chi_{\{X_\tau \text{ is defined}\}} \in m\mathcal{F}_\tau.$$

(We basically want to rule out the last case in this construction).

Proof. Write it out in terms of the possible categories in the set:

$$X_\tau \cdot \chi_{\{X_\tau \text{ is defined}\}} = \sum_{n \geq 0} X_n \cdot \chi_{\tau = n} + \lim_{n\to\infty} X_n \cdot \chi_{\{\tau = \infty\}} \cdot \chi_{\{\lim_{n\to\infty} X_n \text{ exists.}\}}$$

Verify that, for all $B \in \mathcal{B}(\mathbb{R})$:

$$\left\{X_\tau \chi_{\{X_\tau \text{ is defined}\}} \in B\right\} \in \mathcal{F}_\tau.$$

Exercise: finish this proof. □

Proposition 33. If τ_1 and τ_2 are two stopping times on $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ then:

1. If $\tau_1 \leq \tau_2$, then:
$$\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}. \tag{115}$$
2. $\tau_1 \vee \tau_2 := \min\{\tau_1, \tau_2\}$.
$$\mathcal{F}_{\tau_1 \wedge \tau_2} = \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}. \tag{116}$$

Proof. 1. $\forall A \in \mathcal{F}_{\tau_1}$ for all $0 \leq n \leq \infty$:
$$A \cap \{\tau_2 \leq n\} = \underbrace{A \cap \{\tau_1 \leq n\}}_{\in \mathcal{F}_n} \cap \underbrace{\{\tau_2 \leq n\}}_{\in \mathcal{F}_n}$$
Hence, $A \in \mathcal{F}_{\tau_2}$.

2. $\tau_1 \wedge \tau_2 \leq \tau_1$ and $\tau_1 \wedge \tau_2 \leq \tau_2$. This shows that:
$$\mathcal{F}_{\tau_1 \wedge \tau_2} \subseteq \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$$
Now we need to show the other inclusion:
$$\begin{aligned} \forall\ A \in \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}\ \forall\ 0 \leq n \leq \infty\ A \cap \{\tau_1 \wedge \tau_2 \leq n\} &= A \cap (\{\tau_1 \leq n\} \cup \{\tau_2 \leq n\}) \\ &= \underbrace{(A \cap \{\tau_1 \leq n\})}_{\in \mathcal{F}_n} \cup \underbrace{(A \cap \{\tau_2 \leq n\})}_{\in \mathcal{F}_n} \in \mathcal{F}_n. \end{aligned}$$
Hence, $A \in \mathcal{F}_{\tau_1 \wedge \tau_2}$.

$\square$

Proposition 34. Given a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ and an adapted process $\{X_n \mid n \geq 0\}$, let τ be a stopping time. Then, $\{X_{n\wedge\tau} \mid n \geq 0\}$ is again adapted. Furthermore, if $X_n \in L^1$ for all $n \geq 0$, then $X_{n\wedge\tau} \in L^1$ for all $n \geq 0$. This is called a **stopped process**.

Proof. For all $n \geq 0$,
$$X_{n\wedge\tau} = \chi_{\{\tau \geq n+1\}} \cdot X_n + \sum_{j=0}^{n} \chi_{\{\tau=j\}} \cdot X_j \in m\mathcal{F}_n.$$
It's also clear that if $X_n \in L^1$ for all $n \geq 0$, then $X_{n\wedge\tau} \in L^1$ since it's a finite sum consisting of X_j's times an indicator function. $\square$

The next theorem tells us that this stopping process does not change if a game is favourable or not.

Theorem 42 (Doob's Stopping Time Theorem)**.** *Consider a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ and a stopping time τ. If $\{X_n \mid n \in \mathbb{N}\}$ is a (sub)-martingale with respect to $\{\mathcal{F}_n\}$, then $\{X_{n\wedge\tau}\} \in L^1$ is again a (sub)-martingale.*

Proof. We already know that $\{X_{n\wedge\tau} \mid n \geq 0\}$ is adapted and $X_{n\wedge\tau} \in L^1$ for all $n \geq 0$. We need to check the (sub)-martingale property now: for all $n \geq 0$, $\forall A \in \mathcal{F}_n$:
$$\begin{aligned} \int_A X_{(n+1)\wedge\tau} d\mathbb{P} &= \int_{A\cap\{\tau>n\}} X_{n+1} d\mathbb{P} + \int_{A\cap\{\tau\leq n\}} X_\tau d\mathbb{P} \\ &= (\geq) \int_{A\cap\{\tau>n\}} X_n d\mathbb{P} + \int_{A\cap\{\tau\leq n\}} X_\tau d\mathbb{P} \\ &= \int_A X_{n\wedge\tau} d\mathbb{P}. \end{aligned}$$
Since $A \cap \{\tau > n\} \in \mathcal{F}_n$, $\mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right](\geq) = X_n$, this shows that $\mathbb{E}\left[X_{(n+1)\wedge\tau}|\mathcal{F}_n\right](\geq) = X_{n\wedge\tau}$. $\square$

Corrolary 6. If $\{X_n \mid n \geq 0\}$ is a (sub)-martingale and τ is a stopping time, then:

$$\mathbb{E}\left[X_{n\wedge\tau}\right](\geq) = \mathbb{E}\left[X_0\right] \;\; \forall n \geq 0.$$

Caution! This does NOT imply that $\mathbb{E}\left[X_\tau\right] = \mathbb{E}\left[X_0\right]$!! We can see this illustrated with an example:

Example 22. Let $\{Y_n\}$ be a sequence of iid random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

$$\mathbb{P}\left(Y_1 = 1\right) = \frac{1}{2} = \mathbb{P}\left(Y_1 = -1\right).$$

Set $S_0 \equiv 0$, $S_n = \sum_{j=1}^n Y_j$ for all $n \geq 1$. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then,

$$\mathcal{F}_n := \sigma(\{Y_1, ..., Y_n\})$$

for all $n \geq 1$. $\{S_n \mid n \geq 0\}$ is a martingale with respect to $\{\mathcal{F}_n\}$. Set $\tau := \inf\{n \geq 0 \mid S_n = 1\}$. One can easily check that τ is a stopping time. $\{S_{n\wedge\tau} \mid n \geq 0\}$ is a martingale. This shows that for all $n \geq 0$, $\mathbb{E}\left[S_{n\wedge\tau}\right] = \mathbb{E}\left[S_0\right] = 0$. However, $S_\tau = 1$. So,

$$\mathbb{E}\left[S_\tau\right] = 1 \neq \mathbb{E}\left[S_0\right].$$

Theorem 43 (Hunt's Theorem (I))**.** *Given $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ let $\{X_n\}$ be a (sub)-martingale and τ_1 and τ_2 be two stopping times with $\tau_1 \leq \tau_2 \leq T$ for some constant $T > 0$ (i.e., $\tau_1 \leq \tau_2$ and τ_1 and τ_2 are bounded by T). Then, $X_{\tau_i} \in L^1$ for $i = 1, 2$ and $\mathbb{E}\left[X_{\tau_2} | \mathcal{F}_{\tau_1}\right](\geq) = X_{\tau_1}$. In particular, $\mathbb{E}\left[X_{\tau_2}\right](\geq) = \mathbb{E}\left[X_{\tau_1}\right]$.*

Corrolary 7. If τ is a bounded stopping time, and $\{X_n \mid n \geq 0\}$ is a (sub)-martingale, then

$$\mathbb{E}\left[X_\tau\right](\geq) = \mathbb{E}\left[X_0\right]. \tag{117}$$

Proof. 1. **Integrability**: For $i = 1, 2$:

$$X_{\tau_i} = \sum_{j=0}^{T} \chi_{\{\tau_i = j\}} \cdot X_j \in L^1,$$

since a finite sum of of random variables will be finite since T is finite.

Next, let's assume first that $\{X_n \mid n \geq 0\}$ is a martingale. Then, for all $A \in \mathcal{F}_{\tau_1}$:

$$\int_A X_{\tau_2} d\mathbb{P} = \sum_{j=0}^{T} \int_{A\cap\{\tau_2 = j\}} X_j d\mathbb{P}.$$

Note that for all $0 \leq j \leq T$, $A \cap \{\tau_1 = k\} \cap \{\tau_2 = j\} \in \mathcal{F}_j$. So,

$$\begin{aligned}
&= \sum_{j=0}^{T} \sum_{k=0}^{j} \int_{A\cap\{\tau_2 = k\}\cap\{\tau_2 = j\}} X_j d\mathbb{P} \\
&= \sum_{j=0}^{T} \sum_{k=0}^{j} \int_{A\cap\{\tau_1 = k\}\cap\{\tau_2 = j\}} X_T d\mathbb{P} \\
&= \int_A X_T d\mathbb{P} (T \text{ is a constant.})
\end{aligned}$$

Re-run the same argument for τ_2: similarly:

$$\int_A X_{\tau_1} d\mathbb{P} = \sum_{k=0}^{T} \int_{A\cap\{\tau_1 = k\}} X_k d\mathbb{P}.$$

Thus, $\int_A X_{\tau_2} d\mathbb{P} = \int_A X_{\tau_1} d\mathbb{P}$ for all $A \in \mathcal{F}_{\tau_1}$. Recall that $X_{\tau_1} \in m\mathcal{F}_{\tau_1}$. Therefore,

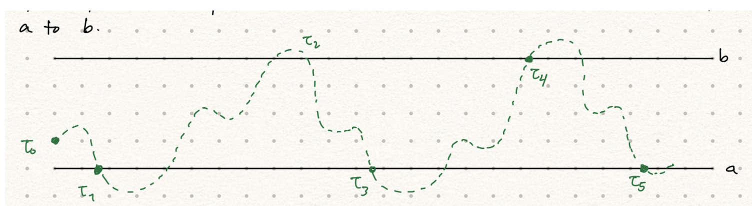
$$\mathbb{E}\left[X_{\tau_2} | \mathcal{F}_{\tau_1}\right] = X_{\tau_1}.$$

Now, let's assume that $\{X_n \mid n \geq 0\}$ is a sub-martingale. Assume Doob's Decomposition of $X_n = M_n + Y_n$, where $\{Y_n \mid n \geq 0\}$ is a non-negative, increasing, and pre-visible process and $\{M_n\}$ is a martingale, and the decomposition is unique. Then,

$$\begin{aligned} \mathbb{E}\left[X_{\tau_2} | \mathcal{F}_{\tau_1}\right] &= \mathbb{E}\left[M_{\tau_2} \mid \mathcal{F}_{\tau_1}\right] + \mathbb{E}\left[Y_{\tau_2} \mid \mathcal{F}_{\tau_1}\right] \\ &\geq M_{\tau_1} + Y_{\tau_1} \\ &= X_{\tau_1}, \end{aligned}$$

where we got M_{τ_1} from the previous step and we got Y_{τ_1} because $Y_{\tau_2} \geq Y_{\tau_1}$ and $Y_{\tau_1} \in m\mathcal{F}_{\tau_1}$. □

Recall upcrossings:



Given $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ let $\{X_n\}$ be adapted. Consider real numbers $a, b \in \mathbb{R}$, $a < b$, $\tau_0 \equiv 0$. Then, $\{\tau_k \mid k \geq 0\}$ is the upcrossing time from a to b.

Theorem 44 (Doob's Upcrossing Inequality)**.** *Given $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ let $\{X_n \mid n \in \mathbb{N}\}$ be a sub-martingale. For every $a, b \in \mathbb{R}$, $a < b$, let $\{\tau_k\}$ be the upcrossing time from a to be, and for every $n \geq 1$, set:*

$$U_{a,b}^{(n)} := \text{ \# upcrossings from a to be completed by time } n. \tag{118}$$

In other words, this is $\max\{k \mid \tau_{2k} \geq n\}$. *Then,*

$$\mathbb{E}\left[U_{a,b}^{(n)}\right] \leq \frac{\mathbb{E}\left[(X_n - a)^+\right]}{b - a}, \tag{119}$$

for all $n \geq 1$. In particular, if $\sup_n \mathbb{E}\left[X_n^+\right] < \infty$, then $U_{a,b} := \lim_{n\to\infty} U_{a,b}^{(n)} < \infty$ a.s.

This theorem tells us that the oscillations are bounded by a terminal index n. It's reminiscent of the Kolmogorov 0-1 Law. Also note that $U_{a,b}^{(n)} \uparrow U_{a,b} = \lim_{n\to\infty} U_{a,b}^{(n)}$ exists and is the total number of upcrossings from a to b.

Proof. Set $Y_n := (X_n - a)^+$ for all $n \geq 0$. Note that $\{Y_n \mid n \geq 0\}$ is again a sub-martingale, because $\varphi(x) = (x - a)^+$ is increasing and convex. In addition, $Y_{\tau_{2k-1}} = 0$ and $Y_{\tau_{2k}} \geq b - a > 0$ is true for all $k \geq 1$. This implies that

$$1 \leq \frac{Y_{\tau_{2k}} - Y_{\tau_{2k-1}}}{b - a}.$$

WLOG, assume that $\tau_1 < n$. Otherwise, if $\tau_1 > n$, the first upcrossing hasn't begun yet, i.e., $U_{a,b}^{(n)} = 0$. Then,

$$\begin{aligned} U_{a,b}^{(n)} &= \sum_{k=1}^{U_{a,b}^{(n)}} 1 \\ &\leq \sum_{k=1}^{U_{a,b}^{(n)}} \frac{Y_{\tau_{2k}} - Y_{\tau_{2k-1}}}{b-a} \\ &= \sum_{k=1}^{n} \frac{Y_{\tau_{2k\wedge n}} - Y_{\tau_{(2k-1)\wedge n}}}{b-a} \\ &= \frac{1}{b-a}(\underbrace{Y_{\tau_{2n\wedge n}}}_{=Y_n} - \sum_{k=2}^{n}(Y_{\tau_{2k-1\wedge n}} - Y_{\tau_{2k-2}\wedge n}) - \underbrace{Y_{\tau_1 \wedge n}}_{=Y_{\tau_1}=0}). \end{aligned}$$

Taking the expectation of both sides yields:

$$\mathbb{E}\left[U_{a,b}^{(n)}\right] \leq \underbrace{\frac{1}{b-a}\mathbb{E}\left[Y_n\right]}_{\frac{\mathbb{E}[(X_n-a)^+]}{b-a}} - \frac{1}{b-a}\underbrace{\sum_{k=2}^{n}(\mathbb{E}\left[Y_{\tau_{2k-1}\wedge n}\right] - \mathbb{E}\left[Y_{\tau_{2k-2}\wedge n}\right])}_{n \geq \tau_{2k-1}\wedge n \geq \tau_{2k-2}\wedge n (2)}$$

so (2) implies $\mathbb{E}\left[Y_{\tau_{2k-1}\wedge n}\right] \geq \mathbb{E}\left[Y_{\tau_{2k-2}\wedge n}\right]$ by **(Hunt's Theorem)** . Hence,

$$\mathbb{E}\left[U_{a,b}^{(n)}\right] \leq \frac{\mathbb{E}\left[(X_n - a)^+\right]}{b-a}.$$

Finally, by **(MON)** ,

$$\mathbb{E}\left[U_{a,b}\right] = \lim_{n\to\infty} \mathbb{E}\left[U_{a,b}^{(n)}\right] \leq \sup_n \frac{\mathbb{E}\left[(X_n - a)^+\right]}{b-a} < \infty,$$

since $\sup_n \mathbb{E}\left[(X_n - a)^+\right] \leq \sup_n \mathbb{E}\left[X_n^+\right] + |a|$. Hence, $U_{a,b} < \infty$ a.s. ☐

6.4 Martingale Convergence Theorems

Theorem 45 (Martingale Convergence Theorem 1)**.** *Given* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ *let* $\{X_n \mid n \geq 0\}$ *be a sub-martingale with* $\sup_n \mathbb{E}\left[X_n^+\right] < \infty$*. Then, there exists an* $X \in L^1$ *such that* $X_n \to X$ *almost surely.*

Note that the condition on the expectation, $\sup_n \mathbb{E}\left[X_n^+\right] < \infty$ is quivalent to saying that X_n is bounded in L^1 in the case of being a sub-martingale.

Proof. By **(Doob's Upcrossing Inequality)** we can write:

$$\mathbb{P}\left(\bigcup_{a,b\in\mathbb{Q},\ a<b} \{U_{a,b} = \infty\}\right) = 0.$$

Hence,

$$\mathbb{P}\left(\bigcap_{a,b\in\mathbb{Q},\ a<b} \{U_{a,b} < \infty\}\right) = 1 \Rightarrow \lim_{n\to\infty} X_n =: X.$$

Hence, we've found that the limit exists almost surely. It CAN be infinity, however. Next, by **(Fatou)** ,

$$\begin{aligned}\mathbb{E}\left[|X|\right] &\leq \liminf_n \mathbb{E}\left[|X_n|\right] \\ &= \liminf_n (2\mathbb{E}\left[X_1^+\right] - \mathbb{E}\left[X_n\right]) \\ &\leq 2\sup_n \mathbb{E}\left[X_n^+\right].\end{aligned}$$

Now use that X_n is a sub-martingale:

$$\begin{aligned}&\leq 2\sup_n \mathbb{E}\left[X_n^+\right] - \mathbb{E}\left[X_0\right] \\ &< \infty.\end{aligned}$$

□

We remark that **(Martingale CV Thm 1)** does NOT imply that $X_n \to X$ in $L61$. Consider the following example to see why.

Example 23. $\{Y_n \mid n \in \mathbb{N}\}$ be iid random variables with $Y_n > 0$ for all $n \geq 1$ and $\mathbb{E}\left[Y_1\right] = 1$ and $\ln(Y_n) \in L^1$. Set $T_0 \equiv 1$ and:

$$T_j := \prod_{j=1}^n Y_j \ \forall n \geq 1.$$

Set $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and $\mathcal{F}_n = \sigma(\{Y_1, ..., Y_n\})$. Then, $\{T_n \mid n \geq 0\}$ is a martingale with respect to $\{\mathcal{F}_n\}$. Quick check of this:

$$\mathbb{E}\left[T_{n+1}|\mathcal{F}_n\right] = \mathbb{E}\left[T_n \cdot Y_{n+1}|\mathcal{F}_n\right] = T_n\mathbb{E}\left[Y_{n+1}\right] = T_n.$$

Hence, for all $n \geq 0$, $\mathbb{E}\left[T_n\right] = 1$. So, there exists a $T \in L^1$ such that $T_n \to T$ almost surely. **What is T?** Well, consider a new sequence $\{\ln(Y_n) \mid n \geq 1\}$ iid random variables with $\ln(Y_n) \in L^1$ for all $n \geq 1$. By **(SLLN 3)** we have:

$$\frac{1}{n}\sum_{j=1}^n \ln(Y_j) \to \mathbb{E}\left[\ln(Y_1)\right] \text{ almost surely.}$$

Assume that Y_1 is not almost everywhere constant. Then, by **(Jensen)** we have:

$$\mathbb{E}\left[\ln(Y_1)\right] < \ln(\mathbb{E}\left[Y_1\right]) < 0.$$

The inequality is strict, since Y_1 is not constant. Hence,

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n \lim(Y_j) = c < 0 \text{ a.s.} \\ \Rightarrow &\lim_{n\to\infty} \frac{1}{n}\ln(T_n) = c \text{ a.s.}\end{aligned}$$

This shows that $T_n \sim e^{cn}$ when n is sufficiently large. Hence, $T = \lim_{n\to\infty} T_n = 0$ almost surely which shows that T_n does NOT converge to T in L^1.

Theorem 46 (Martingale Convergence Theorem II)**.** *Given a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ let $\{X_n \mid n \geq 0\}$ be a (sub)-martingale and let $\{X_n \mid n \geq 0\}$ be uniformly integrable. Then, there exists an $X \in L^1$ such that $X_n \to X$ almost surely and in L^1. IN addition, $\mathbb{E}\left[X|\mathcal{F}_n\right](\geq) = X_n$ for all $n \geq 0$.*

Proof. By **(Martingale CV Thm I)** , there exists an $X \in L^1$ such that $X_n \to X$ almost surely.

$$\Rightarrow \{X_n \, | n \geq 0\} \text{ is uniformly integrable}$$
$$\Rightarrow X_n \to X \text{ in } L^1.$$

Now to show the second statement:

$$\forall \, n \geq 0, \; \forall \, A \; \in \mathcal{F}_n \; \int_A X_n d\mathbb{P}(\leq) = \int_A X_m d\mathbb{P} \; \forall \; m \geq n \underbrace{\to}_{m \to \infty} \int_A X d\mathbb{P}.$$

This shows that

$$X_n(\leq) = \mathbb{E}\left[X | \mathcal{F}_n\right].$$

□

Lemma 11 (Doob's Maximal Inequality)**.** Let $\{X_n \mid n \geq 0\}$ be a sub-martingale on a filtered space. Then, for all $N > 0$, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{0 \leq n \leq N} X_n > \varepsilon\right) \leq \frac{1}{\varepsilon} \mathbb{E}\left[X_n ; \max_{0 \leq n \leq N} X_n > \varepsilon\right]. \tag{120}$$

In particular, if $X_n \geq 0$ for all $n \geq 0$ and $\sup_n ||X_n||_p < \infty$ for some $p > 1$, then we can conclude that:

$$|| \max_{0 \leq n \leq N} X_n ||_p \leq \frac{p}{p-1} ||X_N||_p. \tag{121}$$

And, furthermore,

$$|| \sup_n X_n || \leq \frac{p}{p-1} \sup_n ||X_n||. \tag{122}$$

This is a very intrinsic property about martingales; we're trying to control something that involves the whole process.

Proof. We will only show the first inequality. Set $M_n := \max_{0 \leq n \leq N} X_n$ and $\tau = \inf\{n \geq 0 \mid X_n > \varepsilon\}$. τ is a stopping time. We have the following:

$$\{M_n > \varepsilon\} = \{\tau \leq N\} \tag{123}$$

Since τ reports the time when the process goes above ε.

$$\begin{aligned}
\mathbb{P}\left(\tau \leq N\right) &= \sum_{j=0}^{N} \mathbb{P}\left(\tau = j\right) \\
&\leq \sum_{j=0}^{N} \frac{1}{\varepsilon} \int_{\{\tau = j\}} X_j d\mathbb{P} \\
&\leq \sum_{j=0}^{N} \frac{1}{\varepsilon} \int_{\{\tau = j\}} X_N d\mathbb{P} \\
&= \frac{1}{\varepsilon} \mathbb{E}\left[X_n ; \tau \leq N\right]
\end{aligned}$$

□

Theorem 47 (Martingale Convergence Theorem III)**.** *Given a filtered space* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ *let* $\{X_n | n \geq 0\}$ *be a martingale (or a non-negative sub-martingale) with* $\sup_n ||X_n||_p < \infty$ *for some* $p > 1$*. Then, there exists an* $X \in L^p$ *such that* $X_n \to X$ *almost surely and in* L^p *and*

$$\mathbb{E}\left[X|\mathcal{F}_n\right](\geq) = X_n \ \forall n \geq 0. \tag{124}$$

Proof. $\{X_n\}$ is bounded in L^p. Hence,

$$\begin{aligned}&\Rightarrow \{X_n \mid n \geq 0\} \text{ is uniformly integrable}\\&\Rightarrow \exists\, X \in L^1 \text{ s.t. } X_n \to X \text{ a.s. and in } L^1 \text{ and } \mathbb{E}\left[X|\mathcal{F}_n\right](\geq) = X_n.\end{aligned}$$

By **(Fatou)** ,

$$\begin{aligned}\mathbb{E}\left[|X|^p\right] &\leq \liminf_n \mathbb{E}\left[|X_n|^p\right]\\&\leq \sup_n \mathbb{E}\left[|X_n|^p\right]\\&< \infty \Rightarrow X \in L^p.\end{aligned}$$

Now, we need to use one of the integral convergence theorems. $\{|X_n| \mid n \geq 0\}$ is a sub-martingale. Applying **(Doob's Maximal Inequality)** on $\{|X_n|\}$ we get that $\sup_n |X_n| \in L^p$ and thus by **(DOM)** $X_n \to X$ in L^p. ☐

Theorem 48 (Hunt's Theorem II)**.** *Given a filtered space* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ *and a uniformly integrable (sub)-martingale* $\{X_n\}$*, let* τ_1, τ_2 *be two stopping times with* $\tau_1 \leq \tau_2$*. Then, for* $i = 1, 2$*,* $X_{\tau_i} \in L^1$ *and*

$$\mathbb{E}\left[X_{\tau_2}|\mathcal{F}_{\tau_1}\right](\geq) = X_{\tau_1}. \tag{125}$$

Proof. If $\{X_n \mid n \geq 0\}$ is uniformly integrable, then there exists an $X_\infty \in L^1$ such that $X_n \to X$ almost surely and in L^1. Hence, X_{τ_i} is well-defined almost surely for $i = 1, 2$.

1. Let $\{X_n = M_n + Y_n\}$ by **(Doob's Decomposition)** of $\{X_n\}$. We have $0 \leq Y_n$ and $Y_n \uparrow$ to $Y := \lim_n Y_n$. We have:

$$\begin{aligned}\mathbb{E}\left[Y\right] &\leq \liminf_n \mathbb{E}\left[Y_n\right]\\&= \liminf_n (\mathbb{E}\left[X_n\right] - \mathbb{E}\left[M_n\right])\\&\leq \sup_n \mathbb{E}\left[|X_n|\right] - \mathbb{E}\left[M_n\right]\end{aligned}$$

   Hence, $Y \in L^1$ and $Y_n \to Y$ almost surely and in L^1 (by **(DOM)**); the dominating function is Y. Hence, $\{Y_n \mid n \geq 0\}$ is uniformly integrable. Set $M_\infty := X_\infty - Y_\infty$ which shows that $M_n \to M$ almost surely and in L^1 and $\{M_n \mid n \geq 0\}$ is uniformly integrable.
2. In this step we claim that if τ is a stopping time, then $\{X_{n\wedge\tau} \mid n \geq 0\}$ is uniformly integrable. **Exercise:** prove this statement.
3. We know that for $i = 1, 2$, $\{X_{n\wedge\tau_i} \mid i = 1, 2\}$ is a uniformly integrable sub-martingale. Hence,

$$X_{n\wedge\tau_i} \to X_{\tau_i}$$

almost surely and in L^1 by uniform integrability. Hence, for all $A \in \mathcal{F}_{\tau_1}$:

$$\begin{aligned}
\int_A X_2 d\mathbb{P} &= \lim_{n\to\infty} \int_A X_{\tau_2\wedge n} d\mathbb{P} \text{ (by } L^1 \text{ convergence)} \\
&= \lim_{n\to\infty} \left(\int_{A\cap\{\tau_1\le n\}} X_{\tau_2\wedge n} d\mathbb{P} + \int_{A\cap\{\tau_1>n\}} X_n d\mathbb{P} \right) \\
&= (\ge) \lim_{n\to\infty} \left(\int_{A\cap\{\tau_1\le n\}} X_{\tau_1\wedge n} d\mathbb{P} + \int_{A\cap\{\tau_1>n\}} X_n d\mathbb{P} \right) \\
&= \lim_{n\to\infty} \int_A X_{\tau_1\wedge n} d\mathbb{P} \\
&= \int_A X_{\tau_1} d\mathbb{P} \text{ (by } L^1 \text{ convergence).}
\end{aligned}$$

Hence,

$$\mathbb{E}\left[X_{\tau_2\wedge n} | \mathcal{F}_{\tau_1\wedge n}\right] (\ge) = X_{\tau_1\wedge n}.$$

□

Theorem 49 (Hunt's Theorem III)**.** *Let $\{X_n \mid n \ge 0\}$ be a (sub)-martingale, τ_1, τ_2 be stopping times such that $\mathbb{E}[\tau_1] \le \mathbb{E}[\tau_2] < \infty$. Assume that there exists an $k > 0$ such that $|X_{n+1} - X_n| \le k$ for all $n \ge 0$. Then, for $i = 1, 2$, $X_{\tau_i} \in L^1$ and $\mathbb{E}[X_{\tau_2} | \mathcal{F}_{\tau_1}] (\ge) = X_{\tau_1}$.*

Corrolary 8 (Wald's Inequality)**.** Let $\{W_n \mid n \ge 1\}$ be iid random variables with $\mathbb{E}[W_1] \in \mathbb{R}$. Set $X_0 \equiv 0$, $X_n := \sum_{j=1}^n W_j$ for all $n \ge 1$. Set $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n := \sigma(\{W_1, ..., W_n\})$. Let τ be a stopping time with $\mathbb{E}[\tau] < \infty$. Then,

$$\mathbb{E}[X_\tau] = \mathbb{E}[\tau]\,\mathbb{E}[W_1].$$