MATH 578 - Advanced Probability Theory I

Aram Pooladian (260639736)
Department of Mathematics
McGill University
Montreal, QC, Canada

January 2018

Contents

Chapter 1

General Measure Spaces

1.1 Preliminaries

We start by presenting general notions and definitions of measure spaces. Unless otherwise specified, S will denote an arbitrary set.

Definition 1. A collection Σ_0 of subsets of S is called an **algebra** if (and only if) the following hold:

1. $S \in \Sigma_0$,
2. if $A \in \Sigma_0$, then $\overline{A} \in \Sigma_0$,
3. for all $n \geq 1$: if $A_i \in \Sigma_0$ (for all $i = 1, \ldots, n$), then $\bigcup_{i=1}^{n} A_i \in \Sigma_0$,
4. $\emptyset \in \Sigma_0$,
5. for all $n \geq 1$: if $A_i \in \Sigma_0$ (for all $i = 1, \ldots, n$), then $\bigcap_{i=1}^{n} A_i \in \Sigma_0$,
6. If $A, B \in \Sigma_0$, then $A \setminus B,\ B \setminus A,\ A \Delta B \in \Sigma_0$.

Note that the last three statements are implied from the first three and "Δ" here is the symmetric-difference operator. Essentially, an algebra is a set that is closed under finitely many intersections and unions. For infinitely many such operations, we have the following collection of set

Definition 2. Let S be a set. A collection Σ of subsets of S is called a **σ-algebra**, henceforth denoted s-algebra for brevity, if (and only if) the following hold:

1. $S \in \Sigma$,
2. if $A \in \Sigma$, then $\overline{A} \in \Sigma$,
3. if $A_i \in \Sigma_0$ (for all $i = 1, 2, \ldots$), then $\bigcup_{i=1}^{\infty} A_i \in \Sigma$,
4. $\emptyset \in \Sigma_0$,
5. if $A_i \in \Sigma_0$ (for all $i = 1, 2, \ldots$), then $\bigcap_{i=1}^{\infty} A_i \in \Sigma$

The trivial s-algebra of a set S is $\{\emptyset, S\}$; otherwise they can be huge.

Definition 3. Let $\mathcal{C}$ be a collection of subsets of S. The **s-algebra generated by** $\mathcal{C}$, denoted $\sigma(\mathcal{C})$, is the *smallest* s-algebra that contains $\mathcal{C}$.

Example 1. Take $A \subseteq S$, then $\sigma(\{A\}) = \sigma(A) = \{\emptyset, A, \overline{A}, S\}$

Definition 4. We call the pair (S, Σ) a measureable space.

Fact 1. Let I be an (arbitrary) index set, then $\bigcap_{i \in I} \Sigma_i$ is a s-algebra (where each Σ_i is a s-algebra). This is NOT true for unions of s-algebras, however.

Proposition 1. Take a collection $\mathcal{C} \subseteq S$, then

$$\sigma(\mathcal{C}) = \bigcap \left\{ \Sigma \mid \Sigma \text{ is an s-algebra and } \mathcal{C} \subseteq \Sigma \right\}$$

Proof. The RHS (right-hand side) is a s-algebra with $\mathcal{C} \subseteq$ RHS; thus $\sigma(\mathcal{C}) \subseteq$ RHS. Since $\sigma(\mathcal{C})$ is an element of the RHS, RHS $\subseteq \sigma(\mathcal{C})$ (due to intersections). □

Corollary 1.

1. $\sigma(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$,
2. $\mathcal{C}_1 \subseteq \mathcal{C}_2$, then $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$,
3. $\mathcal{C}_1 \subseteq \sigma(\mathcal{C}_2)$, then $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$.

Definition 5. We call $(S, \mathcal{T})$ a **topological space**, where S supports the notion of "open sets"; $\mathcal{T}$ is a collection of open sets of S. We call $\sigma(\mathcal{T})$ the **Borel s-algebra** on S, denoted $\mathcal{B}(S)$.

Example 2. Let $S = \mathbb{R}$, then $\mathcal{B}(\mathbb{R}) := \sigma(\{\text{open sets in } \mathbb{R}\})$

Proposition 2. $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b) \mid a < b; a, b \in \mathbb{R}\})$; the s-algebra generated by open intervals in $\mathbb{R}$.

Proof. Again, the proof is very "obvious". Let Σ_1 denote the RHS; then $\Sigma_1 \subseteq \mathcal{B}(\mathbb{R})$ by definition. For the other containment, take any open set in $\mathcal{B}(\mathbb{R})$ and recall that this can be written by a countable union of open intervals. So $\mathcal{B}(\mathbb{R}) \subseteq \Sigma_1$ and we have equality of the two sets. □

In fact $\mathcal{B}(\mathbb{R})$ has many equivalent representations, as seen in the following proposition.

Proposition 3.

$$\begin{aligned} B(\mathbb{R}) &= \sigma(\{(a, b] \mid a < b; a, b \in \mathbb{R}\}) \\ &= \sigma(\{[a, b] \mid a < b; a, b \in \mathbb{R}\}) \\ &= \sigma(\{[a, b) \mid a < b; a, b \in \mathbb{R}\}) \\ &= \sigma(\{(-\infty, b] \mid b \in \mathbb{R}\}) \\ &= \sigma(\{(a, +\infty) \mid a \in \mathbb{R}\}) \end{aligned}$$

We will not provide the proof of this proposition. However, notions of the following kind, in addition to the previous proposition, might be useful if you wish to do it yourself:

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

1.2 Set Theory

Definition 6. A collection $\mathcal{I}$ of subsets of S is called a ***π*-system** iff: $\forall A_1, A_2 \in \mathcal{I}$, $A_1 \cap A_2 \in \mathcal{I}$

Definition 7. A collection $\mathcal{D}$ of subsets of S is called a ***d*-system** iff

1. $S \in \mathcal{D}$
2. $\forall A_1, A_2 \in \mathcal{D}$, $A_1 \setminus A_2 \in \mathcal{D}$
3. If $A_i \in \mathcal{D}$ (for $i \geq 1$) and $A_i \uparrow A$, then $A \in \mathcal{D}$

When we say a sequence of sets is "increasing" or "decreasing" to a limiting set, we mean it in the following way:

1. $A_i \uparrow A$: $A_i \subseteq A_{i+1}$ and $A = \lim(A_i) = \bigcup_{i=1}^{\infty} A_i$
2. $A_i \downarrow A$: $A_i \supseteq A_{i+1}$ and $A = \lim(A_i) = \bigcap_{i=1}^{\infty} A_i$

Lemma 1. Σ is a σ-algebra $\Leftrightarrow$ Σ is a π-system and a d-system

Proof.
The forward direction follows by definition. For the converse, we just need to show that Σ satisfies the s-algebra criteria. From the d-system properties, we have that $S \in \Sigma$, and for any $A \in \Sigma, A^C \in \Sigma$. It remains to show that Σ is closed under arbitrary unions. Let $B := \bigcup_{i \geq 1} A_i$ and $B_n := \bigcup_{i=1}^{n} A_i$. If we can show that $B_n \in \Sigma$, then again due to d-system properties: $B_n \uparrow B \in \Sigma$, which would complete the proof. This amounts to some re-writing and invoking the π-system property:

$$B_n = \overline{\left(\bigcap_{i=1}^{n} \overline{A_i}\right)} \in \Sigma$$

□

Theorem 1. (Dykin's *π*-*d* Lemma) Suppose I is a π-system and D is a d-system and $I \subseteq D$. Then $\sigma(I) \subseteq D$

Proof. Let $d(I)$ be the smallest d-system generated by I. If we can show that $d(I)$ is also a pi-system, then by the previous lemma, we would have $d(I) = \sigma(I) \subseteq D$ and we would be done. To show it is a π-system, we want to show that $d(I)$ is closed under finite intersection. To do this, we create the following to sets:

$$D_1 := \{B \in d(I) \mid B \cap A \in d(I), \forall A \in I\},\ D_2 = \{A \in d(I) \mid A \cap B \in d(I), \forall B \in I\}$$

and implore the following logic: since $I \subseteq D_1$, then if we can show that D_1 is a d-system, we have that $d(I) = D_1$ (since $d(I)$ is the smallest d-system that contains I) and therefore, for all $B \in d(I)$ and any $A \in I$, $A \cap B \in d(I)$. The case for D_2 would follow symmetrically and thus we would have $D_1 = D_2 = d(I)$ and $d(I)$ would also be a π-system.

We show that D_1 is a d-system. The first property is satisfied readily. For the second one: Take $A_1, A_2 \in D_1$, we want to show their difference is in D_1. For all $A \in I$:

$$(A_1 \setminus A_2) \cap A = (A_1 \cap A) \setminus (A_2 \cap A) \in d(I).$$

For the final property, take $B_i \in D_1$ (for all $i \geq 1$) and $B_i \uparrow B$: we want to show that $B \in D_1$. So for all $A \in I$, since $(B_i \cap A) \in d(I)$, we have that

$$(B_i \cap A) \uparrow (B \cap A) \in d(I) \implies B \in D_1.$$

□

Chapter 2

Measure Theory

2.1 Definitions

Let μ_0 be a non-negative set function on an algebra Σ_0 of subsets of S i.e. $\mu_0 : \Sigma_0 \to [0, \infty]$.

Definition 8. μ_0 is called **additive** if: $\mu_0(\emptyset) = 0$ and for all disjoint $A, B \in \Sigma_0$, we have $\mu_0(A \cup B) = \mu_0(A) + \mu_0(B)$.

Definition 9. μ_0 is called **countably additive** if: $\mu_0(\emptyset) = 0$ and for all $A_i \in \Sigma_0$ such that $A_i \cap A_j = \emptyset$, and $\bigcup_{i=1}^{\infty} A_i \in \Sigma$, then $\mu_0\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} \mu_0(A_i)$.

Definition 10. The triple (S, Σ, μ) is a **measure space** if (S, Σ) is a measurable space and $\mu : \Sigma \to [0, \infty]$ is countably additive. Here, μ is called the "measure" on (S, Σ)

Definition 11. We call μ **finite** if $\mu(S) < \infty$ (for a triple (S, Σ, μ))

Definition 12. The triple (S, Σ, μ) is called a **probability space** if $\mu(S) = 1$

Definition 13. We call μ **$\boldsymbol{\sigma}$-finite** if there exist $S_i \in \Sigma$ (for all $i \geq 1$) such that $\bigcup_{i=1}^{\infty} S_i = S$ and $\mu(S_i) < \infty$ (for all $i \geq 1$)

Definition 14. A set $A \in \Sigma$ is called a **$\boldsymbol{\mu}$-null set** if $\mu(A) = 0$ (this does not mean $A = \emptyset$).

2.2 Properties of Measure

Theorem 2.

1. **(Monotonicity)** Let $A, B \in \Sigma$ and $A \subseteq B$. Then $\mu(A) \leq \mu(B)$
2. **(Subadditivity)** Let $A_i \in \Sigma$ (for all $i \geq 1$). Then $\mu\big(\bigcup_{i=1}^{\infty} A_i\big) \leq \sum_{i=1}^{\infty} \mu(A_i)$
3. **(Continuity from below)** If $A_n \in \Sigma$ such that $A_n \uparrow A$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$
4. **(Continuity from above)** If $A_n \in \Sigma$ such that $A_n \downarrow A$ **and** $\mu(A_m) < \infty$ for some $m \geq 1$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$

Proof. We show the proofs for the latter two properties. To prove continuity from below: Let $B_1 := A_1$ and $B_n := A_n \setminus A_{n-1}$. This forces the B_n's to be disjoint. Thus, $\bigcup_{j=1}^n B_j = A_n$ and thus $\bigcup_{j\geq 1} B_j = A$. Hence, by countable additivity:

$$\mu(A) = \sum_{j\geq 1} \mu(B_j) = \lim_n \sum_{j=1}^n \mu(B_j) = \lim_n \mu(A_n).$$

For continuity from above, WLOG we can assume that $\mu(A_1) < \infty$. Let $B_n = A_1 \setminus A_n$, thus $B_n \uparrow A_1 \setminus A$. Hence, by invoking continuity from below in the first equality:

$$\begin{aligned} &\mu(A_1 \setminus A) = \lim_n \mu(B_n) \\ &\implies \mu(A_1) - \mu(A) = \lim_n \mu(B_n) \\ &\implies \mu(A_1) - \mu(A) = \lim_n (\mu(A_1) - \mu(A_n)) \\ &\implies \mu(A) = \lim_n \mu(A_n) \end{aligned}$$

since we can subtract $\mu(A_1)$ from both sides because it is finite. □

The "eventually finite" assumption is crucial in continuity from above. Consider the following counter-example: Let λ be the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Let $A_n := (n, \infty)$. Then $\mu(A_n) = \infty$ for all n. However, $A_n \downarrow \emptyset$ and $\lambda(\emptyset) = 0$.

Theorem 3. Let Σ_0 be an algebra on a set S with $\mu_0 : \Sigma_0 \to [0, \infty)$ (non-negative and *real* valued set function). Then μ_0 is countably additive $\iff$

$$\mu_0 \text{ is finitely additive and if } A_n \in \Sigma\ (n \geq 1) \text{ such that } A_n \downarrow \emptyset, \lim_{n\to\infty} \mu_0(A_n) = 0$$

Proof. The forward implication follows from "continuity from above". For the other direction: take $\{B_n \mid n \geq 1\} \subseteq \Sigma_0$ disjoint. Set $B := \bigcup_{j\geq 1} B_j$ and $A_n := B \setminus \bigcup_{j=1}^n B_j = \bigcup_{j\geq n+1} B_j$ with $A_n \downarrow \emptyset$ (and thus $\lim(\mu(A_n)) = 0$). Then:

$$\mu(B) = \mu(A_n) + \mu\left(\bigcup_{j=1}^n B_j\right) = \mu(A_n) + \sum_{j=1}^n \mu(B_j) \xrightarrow{n\to\infty} 0 + \sum_{j=1}^\infty \mu(B_j)$$

□

We say that two measures, μ_1, μ_2 on a measurable space (S, Σ) are equal if for all $A \in \Sigma$, $\mu_1(A) = \mu_2(A)$. The following theorem is a useful characterization of this definition.

Theorem 4. Let S be a set and let I be a π-system (of S), with $\Sigma = \sigma(I)$. Suppose μ_1 and μ_2 are two measures on (S, Σ). Then

$$\mu_1 = \mu_2 \iff \mu_1(A) = \mu_2(A)\ \forall A \in \mathcal{I} \text{ and } \mu_1(S) = \mu_2(S) < \infty$$

Proof. The forward implication is trivial. For the reverse implication, consider the following set

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

Since $I \subseteq D$ i.e. D is already a π-system, it suffices to show that D is also a d-system (and thus a s-algebra where the two measures agree). The first and second d-system properties are trivially satisfied. The third property follows from the continuity of the two measures. □

Theorem 5. (Caratheodory's Extension Theorem) Let S be a set and Σ_0 is an algebra of the subsets of S. Let $\mu_0 : \Sigma_0 \to [0, \infty]$ be a countably additive map. Then, there exists a measure μ on $(S, \Sigma := \sigma(\Sigma_0))$ such that $\mu = \mu_0$ on Σ_0. Furthermore, if $\mu_0(S) < \infty$, then this *extension* is unique.

2.3 Construction of Lebesgue Measure

As the title indicates, here we outline the construction of Lebesgue measure on the measureable space $((0,1], \mathcal{B}((0,1])$. We start by defining the following collection of sets (check that this is in fact an algebra!)

$$\Sigma_0 := \left\{ F \subseteq (0,1] \;\middle|\; F = \bigcup_{i=0}^{k} (a_i, b_i],\ 0 \leq a_0 < b_0 < \cdots < a_k < b_k \leq 1 \right\}.$$

Now we define the measure on this algebra as $\mu_0(F) := \sum_{i=0}^{k}(b_i - a_i)$ (check that this is well-defined).

We ultimately want to invoke Caratheodory's Extension Theorem to finish the proof. Thus, we must show that μ_0 is countably additive. In light of Theorem 3, it suffices to show that μ_0 is finitely additive (which is very believable) and that it is "continuous at the emptyset". Suppose not, i.e. there exists a sequence of sets $\{F_n\} \subseteq \Sigma_0$, with $F_n \downarrow \emptyset$ such that, for some positive δ, $\mu_0(F_n) > \delta$ for all $n \geq 1$. then, for all $n \geq 1$, there exists a $C_n \subseteq F_n$ such that $\mathrm{cl}(C_n) \subseteq F_n$ and $\mu_0(F_n \backslash C_n) \leq 2^{-n-1}\delta$. In turn, for all $m \geq 1$, we have

$$F_m \setminus \left(\bigcap_{n=1}^{m} C_n \right) = \bigcup_{n=1}^{m} (F_m \setminus C_n) \subseteq \bigcup_{n=1}^{m} (F_n \setminus C_n)$$

which implies the following:

$$\begin{aligned} \mu_0 \left(F_m \setminus \left(\bigcap_{n=1}^{m} C_n \right) \right) &\leq \sum_{n=1}^{m} \mu_0(F_n \setminus C_n) \leq \sum_{n=1}^{m} (2^{-n-1}\delta) \leq \delta/2 \\ \implies \frac{\delta}{2} &\geq \mu_0(F_m) - \mu_0 \left(\bigcap_{n=1}^{m} C_n \right) \implies \mu_0 \left(\bigcap_{n=1}^{m} C_n \right) \geq \frac{\delta}{2}. \end{aligned}$$

Now define $K_m := \bigcap_{n=1}^{m} \mathrm{cl}(C_n)$, for all $m \geq 1$. From the above, we have that $\mu_0(K_m) \geq \delta 2^{-1}$, for all $m \geq 1$, and thus $K_m \neq \emptyset$. Hence, there exists at least one point $x_m \in K_m$ and thus $\{x_m\} \subseteq \mathrm{cl}(C_1)$, which is a closed and bounded set. Hence there exists a subsequence with a limiting point, denoted x_∞, inside $\mathrm{cl}(C_1)$. So for all $n \geq 1$, for $m_l > n$ (the former being the subsequence index), we get $\{x_{m_l}\} \subseteq \mathrm{cl}(C_n)$, thus $x_\infty \in \mathrm{cl}(C_n) \subseteq F_n$. Hence $x_\infty \in \bigcap_{n \geq 1} F_n = \emptyset$, a contradiction.

Since μ_0 is now countable additive, by Caratheodory's theorem, there exists a measure λ on $((0,1], \mathcal{B}(0,1]))$. For $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we use a similar construction on $(k, k+1]$ for all $k \in \mathbb{Z}$.

2.4 Completion of a Measure space

Chapter 3

Probability Spaces

3.1 Limsup/Liminf and BC1

Henceforth, we denote a probability space as the following triple: $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}(\Omega) = 1$. We call an element of Ω a "sample point" and $A \in \mathcal{F}$ an "event". So, for $A \in \mathcal{F}$ and $\mathbb{P}(A) = 1$, we say that event happens "almost surely" or "almost everywhere". Alternatively, if $\mathbb{P}(A) = 0$, it is said to happen "almost nowhere" or "almost never".

See the appendix for a recap of limsups/liminfs from real analysis. We now introduce "set-valued" imsups and iminfs. Let $\{E_n \mid n \geq 1\} \subseteq \mathcal{F}$ be a sequence of events. We define limsup and liminf analogously:

$$
\begin{aligned}
\limsup_{n\to\infty}(E_n) = \bigcap_{n=1}^{\infty}\left(\bigcup_{m\geq n} E_m\right) &= \{\omega : \text{ for every } m, \exists n(\omega) \geq m \text{ such that } \omega \in E_{n(\omega)}\} \\
&= \{\omega : \omega \in E_n \text{ for infinitely many } n\} \in \mathcal{F} \\
&= \text{"happens infinitely often"} \\
\liminf_{n\to\infty}(E_n) = \bigcup_{n=1}^{\infty}\left(\bigcap_{m\geq n} E_m\right) &= \{\omega : \text{ for some } n(\omega), \omega \in E_m \ \forall m \geq n(\omega)\} \\
&= \{\omega : \omega \in E_n \text{ for all large } n\} \in \mathcal{F} \\
&= \text{"eventually happens"}
\end{aligned}
$$

Remark 1. In general, $\liminf(E_n) \subseteq \limsup(E_n)$

Proposition 4. If $\limsup(E_n) = \liminf(E_n)$, then $\lim(E_n) = \liminf(E_n) = \limsup(E_n)$

Fact 2. Let $\{A_n : n \geq 1\}, \{B_n : n \geq 1\} \subseteq \mathcal{F}$

1. $\limsup(E_n^C) = (\liminf(E_n))^C$
2. $(\limsup(E_n)) \setminus (\liminf(E_n)) = \limsup(E_n \setminus E_{n+1})$
3. $(\limsup(A_n)) \cap (\limsup(B_n)) \supseteq \limsup(A_n \cap B_n)$
4. $(\limsup(A_n)) \cup (\limsup(B_n)) = \limsup(A_n \cup B_n)$
5. $(\liminf(A_n)) \cap (\liminf(B_n)) = \liminf(A_n \cap B_n)$

6. $(\liminf(A_n)) \cup (\liminf(B_n)) \subseteq \liminf(A_n \cup B_n)$

For those who have done analysis, the following proposition will maybe be more familiar as "Fatou's Lemma" and "Reverse Fatou's Lemma" respectively. The proofs will be done in more generality later.

Proposition 5. $\{E_n : n \geq 1\} \subseteq \mathcal{F}$

1. $\mathbb{P}(\liminf(E_n)) \leq \liminf(\mathbb{P}(E_n))$
2. $\mathbb{P}(\limsup(E_n)) \geq \limsup(\mathbb{P}(E_n))$

Theorem 6. (Borel-Cantelli 1) Let $\{E_n : n \geq 1\} \subseteq \mathcal{F}$. If $\sum_{n\geq 1} \mathbb{P}(E_n) < \infty$, $\mathbb{P}(\limsup(E_n)) = 0$

Proof.

$$\mathbb{P}(\limsup E_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{m=n}^{\infty} E_m\right) \leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}\left(E_m\right) = 0$$

since $\sum_{n=1}^{\infty} \mathbb{P}(E_n)$ converges. □

3.2 Independence of events

Definition 15. Given $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{E_n : n \geq 1\} \subseteq \mathcal{F}$, we say that $\{E_n : n \geq 1\}$ are independent if, $\forall k \geq 1$: for all finite distinct subsequences $\{n_1, \ldots, n_k\} \subseteq \mathbb{N}$

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

Remark 2. We can also define independence for $\{E_\alpha : \alpha \in I\}$ (where I is potentially uncountable) such that, for all $k \geq 1$: for all $\alpha_1, \ldots, \alpha_k \in I$, the above holds.

Fact 3.

1. If A and B are independent (henceforth denoted $A \perp\!\!\!\perp B$) $\implies A^C \perp\!\!\!\perp B, A^C \perp\!\!\!\perp B^C, A \perp\!\!\!\perp B^C$
2. $\mathbb{P}(A) \in \{0,1\} \implies A \perp\!\!\!\perp B, \forall B \in \mathcal{F}$
3. $A \subseteq B \implies A \not\perp\!\!\!\perp B$ unless $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 1$
4. If $A \cap B = \emptyset \implies A \not\perp\!\!\!\perp B$ unless $\mathbb{P}(A) = 0$ (and/or) $\mathbb{P}(B) = 0$

Definition 16. $\{E_n : n \geq 1\}$ are called **pairwise independent** if, $\forall i \neq j \in \mathbb{N}$

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i)\mathbb{P}(E_j)$$

Remark 3. Independence $\implies$ Pairwise Independence

Definition 17. Let $\{\mathcal{G}_n : n \geq 1\}$ be a sequence of collections of subsets of Ω and $\mathcal{G}_n \subseteq \mathcal{F}$ for all $n \geq 1$. We call $\{\mathcal{G}_n : n \geq 1\}$ **mutually independent** if, for any choice of $E_n \in \mathcal{G}_n$, $\{E_n\}$ is independent

Remark 4. $\{E_n : n \geq 1\}$ is independent $\Leftrightarrow$ $\{\sigma(E_n) : n \geq 1\}$ is independent

Theorem 7. Let $\{\mathcal{G}_n : n \geq 1\}$ be a sequence of π-systems and $\mathcal{G}_n \subseteq \mathcal{F}$ for all $n \geq 1$. Then $\{\mathcal{G}_n : n \geq 1\}$ is independent if and only if $\{\sigma(\mathcal{G}_n) : n \geq 1\}$ is independent.

Proof. Left as an exercise for the reader (because I don't really get it myself). $\square$

Theorem 8. (Borel Cantelli 2) If $\{E_n : n \geq 1\}$ are *independent* and $\sum_{n\geq 1} \mathbb{P}(E_n) = \infty$ then

$$\mathbb{P}(E_n \text{ i.o}) = \mathbb{P}(\limsup E_n) = 1$$

Proof. Note that it is sufficient to show that $\mathbb{P}(\liminf(E_n^C)) = 0$. By definition,

$$\mathbb{P}(\liminf(E_n^C)) = \lim_n \mathbb{P}\left(\bigcup_{m=n}^{\infty} E_m^C\right) \quad \forall n \geq 1.$$

So for any $N > n$, we have:

$$\mathbb{P}\left(\bigcup_{m=n}^{N} E_m^C\right) = \prod_{m=n}^{N} \mathbb{P}(E_m^C) = \prod_{m=n}^{N} (1 - \mathbb{P}(E_m)) \leq \exp\left[-\prod_{m=n}^{N} \mathbb{P}(E_m)\right],$$

where we used the amazing inequality: $1 + x \leq e^x$. Taking the limit as $N \to \infty$, we're done. $\square$

Remark 5. BC2 can be used when $\{E_n : n \geq 1\}$ are pairwise independent too

Proposition 6. Let $\{E_n : n \geq 1\}$ be independent and $\{I_k : k \geq 1\} \subseteq \mathbb{N}$ such that $I_k \cap I_l = \emptyset$ for all $k \neq l$. Define the following

$$\mathcal{F}_{I_k} = \sigma\Big(\{E_n : n \in I_k\}\Big)$$

Then $\{\mathcal{F}_{I_k} : k \geq 1\}$ is independent

Definition 18. Given $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{E_n : n \geq 1\} \subseteq \mathcal{F}$; the following is called the **tail $\boldsymbol{\sigma}$-algebra** associated with $\{E_n : n \geq 1\}$

$$\mathcal{T} := \bigcap_{n\geq 1} \sigma(\{E_n, E_{n+1}, \ldots\})$$

Theorem 9. (Kolmogorov's 0-1 Law) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be your measure space. If $\{E_n : n \geq 1\}$ are independent and $\mathcal{T}$ is the tail σ-algebra associated with $\{E_n : n \geq 1\}$: for all $A \in \mathcal{T}$ (called a **tail set**), $\mathbb{P}(A) = 0$ or 1

Proof. $\{E_n : n \geq 1\}$ is independent and so $\{\sigma(\{E_1\}), \sigma(\{E_2\}), \ldots, \sigma(\{E_{n-1}\}), \sigma(\{E_n, E_{n+1}, \ldots, \})\}$ is also independent (for $n \geq 2$). Take an arbitrary $A \in \mathcal{T}$: since $A \in \sigma(\{E_n, E_{n+1}, \ldots\})$, then $\{A, E_1, \ldots, E_{n-1}\}$ is independent and thus $\{A, E_1, \ldots, E_{n-1}, E_n, \ldots\}$ is independent. This implies that $\sigma(\{A\}) \perp\!\!\!\perp \sigma(\{E_1, E_2, \ldots\})$. On the other hand, since $A \in \sigma(\{E_1, E_2, \ldots, E_n, \ldots\})$, then $A \perp\!\!\!\perp A$. And so:

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = (\mathbb{P}(A))^2 \implies \mathbb{P}(A) = 0 \text{ or } 1.$$

$\square$

Remark 6. $\limsup(E_n) \in \mathcal{T}$ and $\liminf(E_n) \in \mathcal{T}$

Corollary 2. $\{E_n : n \geq 1\}$ indep. $\mathbb{P}(E_n,\ \text{i.o}) = 0$ or $1 \iff \sum_{n\geq 1} \mathbb{P}(E_n) < \infty$ (FIX?!)

Chapter 4

Measurable functions

4.1 Preliminaries

Definition 19. Let (S, Σ) be a measurable space and $h : S \to \mathbb{R}$ is a function. Then h is called **($\boldsymbol{\Sigma}$-)measurable**, denoted $h \in m\Sigma$ if

$$\forall B \in \mathbb{B},\ h^{-1}(B) \in \Sigma$$

In addition, if $h \geq 0$, then $h \in (m\Sigma)^+$
If h is bounded, then $h \in b\Sigma$

Remark 7.
$h : S \to \overline{\mathbb{R}}$ is measurable if for all $B \in \overline{\mathbb{B}}$, $h^{-1}(B) \in \Sigma$ (for $\overline{\mathbb{R}}$, view $[-\infty, a)$ and $(b, \infty]$ as open sets)

Definition 20. Given (S_1, Σ_1) and (S_2, Σ_2) and $h : S_1 \to S_2$, then we say that
h is **($\boldsymbol{\Sigma_1 \setminus \Sigma_2}$-)measurable** if $\forall B \in \Sigma_2, h^{-1}(B) \in \Sigma_1$

Definition 21. If $h : \mathbb{R} \to \mathbb{R}$ is $\mathbb{B} \setminus \mathbb{B}$-measurable, then h is a **Borel function**

Definition 22. Let S be a topological space and let $\Sigma = \mathbb{B}(S)$. If $h : S \to \mathbb{R}$ is such that: $\forall G \in \mathbb{B}(\mathbb{R})$ and $h^{-1}(G)$ is open, then h is said to be **continuous**

Remark 8. h is continuous $\Rightarrow$ h is measurable

Fact 4.

1. h^{-1} preserves set operations (i.e. $h^{-1}(A^C) = (h^{-1}(A))^C$, etc)
2. Suppose $\mathcal{C}$ is a generating π-system of $\mathcal{B}$ and $h^{-1}(\mathcal{C}) \in \Sigma$, then $h \in m\Sigma$
3. Given (S, Σ) and $h \in m\Sigma$ and $f : \mathbb{R} \to \mathbb{R}$ is Borel, then $f \circ h \in m\Sigma$
4. Given (S, Σ) and $h_1, h_2 \in m\Sigma$: $h_1 \pm h_2$, $h_1 h_2$, h_1/h_2 $(h_2 \neq 0)$ are in $m\Sigma$
5. Given (S, Σ) and $\{h_n : n \geq 1\} \in m\Sigma$: $\lim(h_n)$, $\sup(h_n)$, $\liminf(h_n)$, $\limsup(h_n)$ are in $m\Sigma$

4.2 Random Variables

Definition 23. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. $X : \Omega \to \mathbb{R}$ is a **random variable** if $X \in m\mathcal{F}$. Note that the σ-algebra generated by X is

$$\sigma(X) := \{X^{-1}(B) \mid B \in \mathcal{B}\}$$

Example 3. Let $X \equiv c$ (a fixed, real constant). Then $\sigma(X) = \{\emptyset, \Omega\}$. Note that in this case,

$$\sigma(X) = \sigma(\{X \leq c\})$$

Example 4. Let $X = \mathbb{1}_A$, then $\sigma(X) = \sigma(A) = \{\emptyset, \Omega, A, A^C\}$.

Definition 24. If $\{X_n : n \geq 1\}$ is a sequence of random variables, then

$$\sigma(\{X_n : n \geq 1\}) = \sigma(\{X_n^{-1}(B) : B \in \mathcal{B}, \forall n \geq 1\}),$$

i.e. the smallest σ-algebra with respect to which X_n is measurable for all $n \geq$1. As a generating π-system, one would write this as:

$$\sigma(\{X_n : n \geq 1\}) = \sigma\left(\{\{X_1 \leq c_1, X_2 \leq c_2, \ldots, X_n \leq c_n\} : n \geq 1, c_i \in \mathbb{R}(i = 1, \ldots, n).\}\right)$$

Definition 25. Take $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\{X_n : n \geq 1\}$ is a sequence of random variables. Then $\{X_n : n \geq 1\}$ are independent *if* $\{\sigma(X_n) : n \geq 1\}$ are independent. In particular:

$$\{X_n : n \geq 1\} \text{ is independent } \Leftrightarrow \forall n \geq 1, \forall c_i \in \mathbb{R}, \quad \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \leq c_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq c_i)$$

Definition 26. Take $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\{X_n : n \geq 1\}$ is a sequence of random variables. The **tail $\boldsymbol{\sigma}$-algebra associated with $\boldsymbol{\{X_n\}}$** is

$$\mathcal{T} := \bigcap_{n \geq 1} \sigma(\{X_n, X_{n+1}, \ldots\})$$

Proposition 7. If $\{X_n : n \geq 1\}$ is independent, then $\forall A \in \mathcal{T}$, $\mathbb{P}(A) = 0$ or 1. In particular, if $X \in m\mathcal{T}$, then X is constant almost everywhere.

Proof. We only prove the latter statement. By Kolmogorov's 0-1 Law, $\mathbb{P}(\{X \leq b\}) = 0$ or 1, for any $b \in \mathbb{R}$. Let $c := \inf\{a \in \mathbb{R} \mid \mathbb{P}(\{X \leq a\}) = 1\}$ and suppose c is finite. So for all $a < c$, $\mathbb{P}(\{X \leq a\}) = 0$. Thus,

$$\forall k \geq, \mathbb{P}(\{X \leq c + k^{-1}\}) = 1 \implies \mathbb{P}(\{X = c\} = \lim_{k \to \infty} \mathbb{P}(\{c - k^{-1} < X \leq c + k^{-1}\}) = 1.$$

In the cases where $c = \pm\infty$: $\forall N > 0$, $\mathbb{P}(\{X \leq \pm N\}) = 0 \implies \mathbb{P}(\{X = \pm\infty\}) = 1$. $\square$

Remark 9. If $\{X_n : n \geq 1\}$ are independent and $S_n = \sum_{j=1}^{n} X_j$ and $\{b_n : n \geq 1\}$ is a non-negative sequence of numbers such that $b_n \uparrow \infty$, then

$$\limsup(S_n/b_n), \liminf(S_n/b_n) \in m\mathcal{T}$$

Remark 10. If $\{X_n : n \geq 1\}$ is a sequence of discrete rv, say $X_i \in S_i = \{s_1^{(i)}, s_2^{(i)}, \ldots\}$ then

$$\{X_n : n \geq 1\} \text{ is independent } \iff \mathbb{P}\left(\bigcap_{j=1}^{k} \{X_j = S_{n_j}^{(j)}\}\right) = \prod_{j=1}^{k} \mathbb{P}(X_j = S_{n_j}^{(i)})$$

Definition 27. Take $(\Omega, \mathcal{F}, \mathbb{P})$ and a real-valued random variable X. Define the set function $\mathcal{L}_X$, called the **distribution of $\boldsymbol{X}$** on $\mathcal{B}$ by (verify this is a measure on $(\mathbb{R}, \mathcal{B})$):

$$\forall B \in \mathcal{B}, \mathcal{L}_X = \mathbb{P}(\{X \in B\}) = \mathbb{P}(X^{-1}(B)) = (\mathbb{P} \circ X^{-1})(B)$$

Fact 5. Since $\{(-\infty, x] \mid x \in \mathbb{R}\}$ is a generating π-system of $\mathcal{B}$, $\mathcal{L}_X$ is determined by $\mathcal{L}_X((-\infty, x])$.

Definition 28. Define the **distribution function** (abbreviated **DF**) of $\mathcal{L}_X$ (or of X) by:

$$\forall x \in \mathbb{R}, F(x) := \mathcal{L}_X((-\infty, x]) = \mathbb{P}(X \leq x)$$

Fact 6. (Properties of DF)

1. F is non-decreasing: if $x \leq y$, $F(x) \leq F(y)$.
2. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.
3. F is right-continuous at every $x \in \mathbb{R}$: $\lim_{y \to x^+} F(y) = F(x)$.
4. $\forall x \in \mathbb{R}$, $F(x-) := \lim_{y \to x^-} F(y)$ exists! Then: $F(x) - F(x-) = \mathcal{L}_X(\{x\})$.

Proposition 8. If $F : \mathbb{R} \to [0, 1]$ and F satisfies the above properties 1, 2, and 3: then there probability measure $\mathcal{L}$ on $(\mathbb{R}, \mathcal{B})$ such that F is the DF of $\mathcal{L}$ and there exists $(\Omega, \mathcal{F}, \mathbb{P})$ and r.v. X such that F is the DF of X

Proof. To find $\mathcal{L}$, we mimick the construction of λ on $\mathbb{R}$. Consider the same algebra Σ_0 and replace "$|b_i - a_i|$" with $F(b_i) - F(a_i)$; the rest is the same. Then take $(\Omega, \mathcal{F}, \mathbb{P}) := (\mathbb{R}, \mathcal{B}, \mathcal{L})$ and let the random variable X be the identity map: $X(x) = x$. Then X has distribution $\mathcal{L}$ and distribution function F. $\square$

Definition 29. Let F be a DF. If F is *absolutely continuous*, i.e.

$$\exists f : \mathbb{R} \to [0, \infty) \text{ such that } F(x) = \int_{-\infty}^{x} f(t)dt$$

then f is the **probability distribution function** (abbreviated **pdf**) of F

Fact 7.

1. F is absolutely continuous $\implies$ F is continuous (the converse is not true).
2. F is absolutely continuous $\implies$ F' exists almost everywhere ($F' = f$ a.e).
3. If f is continuous at $x \implies F'(x) = f(x)$. If f is continuous, F is differentiable everywhere.

For the next few examples, we restrict our attention to the well-known probability space $(\mathbb{R}, \mathcal{B}, \mathcal{L})$.

Example 5. When random variable is said to be *uniformly distributed* on $[a, b]$, written $X \sim U(a, b)$, then the distribution function is given by:

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

It is easy to see that the corresponding pdf is given by: $f(x) = (b - a)^{-1} \mathbb{1}_{[a,b]}(x)$.

Example 6. When a random variable is said to be*exponentially distributed* with parameter $\lambda > 0$, sometimes written $X \sim \exp(\lambda)$, then the distribution function is given by:

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - \exp(-\lambda x) & \text{if } x \geq 0. \end{cases}$$

It is easy to see that the corresponding pdf is given by: $f(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \geq 0}(x)$.

Example 7. A random variable is *Gaussian (normally distributed)* with parameters $\mu \geq 0$ (representing the mean) and $\sigma^2 > 0$ (representing the variance), written $X \sim N(\mu, \sigma^2)$, with DF:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \, dy$$

with the pdf being the integrand. When X is normally distributed with $\mu = 0$ and $\sigma^2 = 1$, we refer to X as a "standard Gaussian" or "standard normal" random variable.

Definition 30. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $\{X_n : n \geq 1\}$ are independent and $\mathcal{L}_{X_n} = \mathcal{L}$ for some distribution $\mathcal{L}$ (for all $n \geq 1$), then we call the sequence of r.v **independent and identically distributed** (abbreviated to **iid**).

Lemma 2. Standard bounds for the Gaussian integral (see Chp 14.8 of "Williams" for a proof):

$$\forall x > 0, \; (2\pi)^{-1/2}(x + x^{-1})^{-1} e^{-x^2/2} \leq (2\pi)^{-1/2} \int_{x}^{\infty} e^{-y^2/2} \, dy \leq (2\pi)^{-1/2} x^{-1} e^{-x^2/2}.$$

Example 8. $\{X_n : n \geq 2\} \overset{\text{iid}}{\sim} N(0,1)$. Let $L := \limsup \left(\frac{X_n}{\sqrt{2\ln(n)}} \right)$. Claim: $\mathbb{P}(L = 1) = 1$.

Proof. Fix $\alpha > 0$ and consider the sequence of events $\{E_n : n \geq 2\}$ to be "$X_n / \sqrt{2\ln(n)} > \alpha$". The X_n's are iid and so the events E_n are independent. While α does not necessarily equal 1, we are still dealing with the "limsup" of a set of events so it is natural to consider using BC1 and BC2 for the proof. We use the lemma to bound the sum of $\mathbb{P}(E_n)$ and get conditions on α:

$$\begin{aligned} \mathbb{P}(E_n) = \mathbb{P}\left(\frac{X_n}{\sqrt{2\ln(n)}} > \alpha \right) &= \mathbb{P}\left(X_n > \alpha\sqrt{2\ln(n)} \right) = 1 - F(\alpha\sqrt{2\ln(n)}) \\ = \frac{1}{\sqrt{2\pi}} \int_{\alpha\sqrt{2\ln(n)}}^{\infty} e^{-y^2/2} \, dy &\leq \left(\alpha\sqrt{4\pi\ln(n)} \right)^{-1} \exp\left(-\frac{\alpha^2 2\ln(n)}{2} \right) = \frac{n^{-\alpha^2}}{2\alpha\sqrt{\pi\ln(n)}}. \end{aligned}$$

Now take the sum over all $n \geq 2$. Now notice (either by intuition/experience for series summations or by using the "integral test") that the following is also true:

$$\sum_{n \geq 2} \mathbb{P}(E_n) \leq \sum_{n=2}^{\infty} \frac{n^{-\alpha^2}}{2\alpha\sqrt{\pi\ln(n)}} = \begin{cases} < \infty & \text{if } \alpha > 1, \\ +\infty & \text{if } \alpha \leq 1. \end{cases}$$

Applying BC1 is straight-forward from definition. For BC2, we use the lower-bound from the lemma above to get that the sum is divergent. Thus:

$$\mathbb{P}(\limsup(E_n)) = \mathbb{P}\left(\frac{X_n}{\sqrt{2\ln(n)}} > \alpha, \text{ i.o} \right) = \begin{cases} 0 & \text{if } \alpha > 1, \\ 1 & \text{if } \alpha \leq 1. \end{cases}$$

Letting $\alpha = 1$, this implies $\mathbb{P}(L \geq 1) \geq \mathbb{P}(X_n/\sqrt{2\ln(n))} > 1, \text{ i.o}) = 1$. We complete the proof by showing $\mathbb{P}(L > 1) = 0$. Take $\alpha = 1 + (2k)^{-1}$ and use the above conditions (and continuity):

$$\mathbb{P}(L > 1 + k^{-1}) \leq \mathbb{P}(X_n/\sqrt{2\ln(n)} > 1 + (2k)^{-1}, \text{ i.o}) = 0$$
$$\implies \mathbb{P}(L > 1) = \mathbb{P}\left(\bigcup_{k\geq 1}\{L > 1 + k^{-1}\}\right) = 0.$$

□

4.3 Convergence

For the next few definitions and propositions, let $\{X_n : n \geq 1\}$ be a sequence of r.v and X be another random variable in a probability space.

Definition 31. We say "X_n converges to X **(almost surely)**", denoted $X_n \longrightarrow X$ a.s., when

$$\mathbb{P}(\lim(X_n) = X) = 1.$$

Note that X can also be replaced with $\pm\infty$.

Definition 32. We say "X_n converges to X **in probability**", denoted $X_n \xrightarrow{p} X$, when

$$\forall \epsilon > 0, \lim_{n\longrightarrow\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

Again, X can be replaced with $\pm\infty$ here. Formally the definition in these cases is:

$$X_n \longrightarrow +\infty \text{ in prob} \quad \text{if} \quad \forall M > 0, \lim_{n\longrightarrow\infty} \mathbb{P}(X_n \leq M) = 0,$$
$$X_n \longrightarrow -\infty \text{ in prob} \quad \text{if} \quad \forall M > 0, \lim_{n\longrightarrow\infty} \mathbb{P}(X_n \geq -M) = 0$$

Proposition 9.

$$X_n \longrightarrow X \text{ a.s.} \iff \forall \epsilon > 0, \ \mathbb{P}(|X_n - X| > \epsilon, \text{ i.o.}) = 0, \text{ or equivalently:}$$
$$X_n \longrightarrow X \text{ a.s.} \iff \forall \epsilon > 0, \ \mathbb{P}(|X_n - X| \leq \epsilon, \text{ e.v.}) = 1$$

Proof. We prove the "second" version of this proposition. First note that

$$\left\{\lim_{n\to\infty} X_n = X\right\} = \bigcap_{k\geq 1}\bigcup_{m\geq 1}\bigcap_{n\geq m}\{|X_n - X| \leq k^{-1}\}.$$

Forward direction: for any $\epsilon > 0$, there exists an $K \in \mathbb{N}$ such that $K^{-1} < \epsilon$. Then,

$$\mathbb{P}(\liminf\{|X_n - X| \leq \epsilon\}) \geq \mathbb{P}\left(\bigcup_{m\geq 1}\bigcap_{n\geq m}\{|X_n - X| \leq K^{-1}\}\right) = 1.$$

Backward direction:

$$\mathbb{P}\left(\bigcup_{m\geq 1}\bigcap_{n\geq m}\{|X_n - X| \leq \epsilon\}\right) = 1 \implies \mathbb{P}\left(\bigcup_{m\geq 1}\bigcap_{n\geq m}\{|X_n - X| \leq k^{-1}\}\right) = 1, \quad \forall k \in \mathbb{N}$$
$$\implies \mathbb{P}\left(\bigcap_{k\in\mathbb{N}}\bigcup_{m\geq 1}\bigcap_{n\geq m}\{|X_n - X| \leq k^{-1}\}\right) = 1$$
$$\implies \mathbb{P}(\lim_{n\to\infty} X_n = X) = 1.$$

□

Proposition 10. $X_n \longrightarrow X$ a.s. $\implies X_n \xrightarrow{p} X$.

Proof. Take $\epsilon > 0$. By "Reverse Fatou",

$$\limsup_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(\limsup_{n\to\infty}\{|X_n - X| > \epsilon\}) = 0 \implies \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

$\square$

Example 9. Counterexample to the converse. Take $\Omega = \{0,1\}$ with $\{X_n : n \geq 1\}$ independent such that $\mathbb{P}(X_n = 1) = n^{-1}$ and $\mathbb{P}(X_n = 0) = 1 - n^{-1}$. Then $X_n \xrightarrow{p} 0$. However,

$$\sum_{n\geq 1} \mathbb{P}(X_n = 1) = \infty \overset{\text{(BC1)}}{\implies} \mathbb{P}(X_n = 1,\ \text{i.o}) = 1 \implies X_n \not\longrightarrow 0 \text{ a.s.}$$

Proposition 11. $X_n \xrightarrow{p} X \implies \exists\ \{X_{n_k} : k \geq 1\}$ such that $X_{n_k} \longrightarrow X$ a.s.

Proof. For all $k \in \mathbb{N}$, $\mathbb{P}(|X_n - X| > k^{-1}) = 0$ as $n \to \infty$. Thus, we can find a subsequence $\{n_k\} \subseteq \mathbb{N}$ such that $n_k \to \infty$ (as $k \to \infty$) and $\mathbb{P}(|X_{n_k} - X| > k^{-1}) \leq k^{-2}$. Summing over all k, we notice that we can use BC1:

$$\sum_{k\geq 1} \mathbb{P}(|X_{n_k} - X| > k^{-1}) \leq \sum_{k\geq 1} k^{-2} < \infty \overset{\text{BC1}}{\implies} \mathbb{P}(|X_{n_k} - X| > k^{-1},\ \text{i.o}) = 0.$$

By Prop 9, the proof is complete. $\square$

Proposition 12. $X_n \xrightarrow{p} X \iff \forall\{n_k\}, \exists\{n_{k_l}\}$ (sub-subsequence) s.t $X_{n_{k_l}} \longrightarrow X$ a.s.

Proof.
Forward Direction: By Prop 11, there exists a subsequence that converges almost surely to X. By Prop 9, this subsequence converges in probability. Again by Prop 11, there is a subsubsequence that converges almost surely.
Backward Direction: Suppose X_n fails to converge to X in probability. Then there exists a subsequence $\{n_k\}$ and an $\epsilon, \delta > 0$ such that

$$\mathbb{P}(|X_{n_k} - X| > \epsilon) \geq \delta.$$

However, every subsequence is assumed to have a sub-subsequence $\{n_{k_l}\}$ which converges to X almost surely, hence in probability. But

$$\mathbb{P}(|X_{n_{k_l}} - X| > \epsilon) \geq \delta$$

contradicts convergence in probability. $\square$

Proposition 13. $X_n \xrightarrow{p} X$ and $f : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous $\implies f(X_n) \xrightarrow{p} f(X)$.

Proof. For any subsequence $\{n_k\}$, $X_{n_k} \xrightarrow{p} X$ (as $k \to \infty$). Then there exists $\{n_{k_l}\}$ such that $X_{n_{k_l}} \longrightarrow X$ almost surely. By continuity of $f(\cdot)$ and Prop 12 (in the last implication):

$$X_{n_{k_l}} \longrightarrow X \text{ a.s} \implies f(X_{n_{k_l}}) \longrightarrow f(X) \text{ a.s} \iff f(X_n) \xrightarrow{p} f(X)$$

$\square$

Chapter 5

Expectation and Integration

5.1 "Decomposition Steps"

Let (S, Σ, μ) be a measure space and assume $f : S \to \overline{\mathbb{R}}$ and $f \in m\Sigma$. Then:

1. $f^+ := \max\{0, f\}$ and $f^- := \max\{-f, 0\}$, then $f = f^+ - f^-$ and $f^\pm \in (m\Sigma)^+$.
2. For $f \in (m\Sigma)^+$: $\forall k \geq 1$, let $f_k := \min\{f, k\} \in (m\Sigma)^+ \cap (b\Sigma)$. Then $f_k \uparrow f$ as $k \to \infty$. Thus, any $f \in (m\Sigma)^+$ can be approximated by a sequence of bounded, non-negative functions.
3. Assume $f \leq k$ and $f \in (m\Sigma)^+ \cap (b\Sigma)$. Define

$$A(n,i) := \{s \in S \mid i2^{-n} \leq f(s) \leq (i+1)2^{-n}\} \text{ and } f_n(s) := \sum_{i=1}^{k2^n} \mathbb{1}_{A(n,i)}(s) i2^{-n}$$

   We have that for all $n \geq 1$, the $A(n,i)$'s are disjoint for $i = 0, 1, \ldots, k2^n$. Thus

$$S = \bigcup_{i=1}^{k2^n} A(n,i) \Rightarrow \text{ the } A(n,i) \text{ form a partition of } S$$

   As n grows, $f_n \uparrow f$. Thus, any $f \in (m\Sigma)^+ \cap (b\Sigma)$ can be approximated by a simple function.

Definition 33. $f : S \to \overline{\mathbb{R}}$ is a **simple function** if if can be written in the following form:

$$f = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}, \ n \in \mathbb{N}, \ a_k \in [0, \infty], \text{ disjoint } A_k \in \Sigma$$

Theorem 10. (Monotone Class Theorem 1) Let $\mathcal{H}$ be the following class of functions

$$\mathcal{H} := \{f \mid f : S \longrightarrow \mathbb{R}\}$$

such that the following are satisfied:

(I) $\mathcal{H}$ is a vector space over $\mathbb{R}$,

(II) $1 \in \mathcal{H}$ (the constant function),

(III) if $\{f_n : n \geq 1\} \subseteq \mathcal{H}$ and $f_n \geq 0$ and $f_n \uparrow f$, then $f \in \mathcal{H}$.

Then: if for some π-system, I, we have that $\{\mathbb{1}_A \mid A \in I\} \subseteq \mathcal{H}$, we have that $m\sigma(I) \subseteq \mathcal{H}$.

Proof. Let $D := \{B \in \sigma(I) \mid \mathbb{1}_B \in \mathcal{H}\}$. By assumption, $I \subseteq D$. We start by showing that D is a d-system:

- $S \in D$: since $\mathbb{1}_S \equiv 1 \in \mathcal{H}$ (by (II)).
- $\forall A, B \in D$ (such that $A \subseteq B$), $B \setminus A \in D$: $\mathbb{1}_{B\setminus A} = \mathbb{1}_B - \mathbb{1}_A \in D$ (by (I)).
- For $A_n \in D$ with $A_n \uparrow A$, $A \in D$: True (by (III)).

By Dykin's π-d Lemma, $\sigma(I) \subseteq D$ i.e. $\forall B \in \sigma(I), \mathbb{1}_B \in \mathcal{H}$. Now take any $f \in m\sigma(I)$. Performing the three "decomposition steps" above, we can write f as a limit of simple functions:

$$f \xrightarrow{(1)} f^{\pm} \xrightarrow{(2)} f_k^{\pm} \xrightarrow{(3)} f_{k,n}^{\pm} = \sum_{i=0}^{k2^n} \mathbb{1}_{A(n,i)} i 2^{-n},$$

where $\lim_n f_{k,n}^{\pm} = f_k^{\pm}$. We have that $f_{k,n}^{\pm} \in \mathcal{H}$ and also $f_k^{\pm} \uparrow f^{\pm} \in \mathcal{H}$ by (III). Finally, since $f = f^+ - f^-$ and $\mathcal{H}$ is a vector space by (I), $f \in \mathcal{H}$. □

Theorem 11. (Monotone Class Theorem 2) Let $\mathcal{H}$ be the following class of functions

$$\mathcal{H} := \{f \mid f : S \longrightarrow \mathbb{R}\}$$

such that the following are satisfied:

- $\mathcal{H}$ is a vector space over $\mathbb{R}$,
- $1 \in \mathcal{H}$ (the constant function),
- if $\{f_n : n \geq 1\} \subseteq \mathcal{H}$ and $f_n \geq 0$ and $f_n \uparrow f$ (for some bounded f), then $f \in \mathcal{H}$.

Then: if for some π-system, I, we have that $\{\mathbb{1}_A \mid A \in I\} \subseteq \mathcal{H}$, we have that $b\sigma(I) \subseteq \mathcal{H}$.

Proof. Left as an exercise. □

Proposition 14. Given two measurable spaces (S, Σ) and $(T, \mathcal{T})$, assume $X : S \to T$ with X being $\Sigma \setminus \mathcal{T}$-measurable and $Y : S \to \mathbb{R}$, with $Y \in m\Sigma$. Then

$$Y \in m\sigma(X) \iff \exists\, f \in m\mathcal{T} \text{ s.t. } Y = (f \circ X).$$

Proof. Forward direction: Let $\mathcal{H} := \{Y : S \to \mathbb{R} \mid \exists f \in m\mathcal{T} \text{ s.t. } Y = (f \circ X)\}$. We first want to show that $\mathcal{H}$ is a monotone class. Clearly, $\mathcal{H}$ is a vector space over $\mathbb{R}$ and $1 \in \mathcal{H}$. For the final criteria: suppose there exists $Y_n \in \mathcal{H}$ with $f_n \in m\mathcal{T}$ such that $Y_n = (f_n \circ X) \uparrow Y$. Letting $f := \limsup_n f_n$ asserts this claim, since

$$Y = \lim_n Y_n = \lim_n (f_n \circ X) = ((\limsup_n f_n) \circ X) = f \circ X;$$

with $f \in m\mathcal{T}$, thus $Y \in \mathcal{H}$. To invoke MTC1, take $I = \sigma(X)$ to be the desired π-system. Then for all $B \in I$, there exists $A \in \mathcal{T}$ such that $B = X^{-1}(A)$.

$$\mathbb{1}_B = \mathbb{1}_{X^{-1}(A)} = (\mathbb{1}_A \circ X) \in \mathcal{H} \overset{\text{MCT1}}{\Longrightarrow} m\sigma(X) \subseteq \mathcal{H}$$

and since $\mathcal{H} \subseteq m\sigma(X)$, we're done. The backwards direction is trivial. □

5.2 Integration

5.2.1 Define integral for $f \in SF^+$

Definition 34. Given (S, Σ, μ) and $f \in m\Sigma$, the **integral with respect to μ** is

$$\mu(f) = \int_S f d\mu = \int_S f(s)\mu(ds) = \int_S f(s)d\mu(s) \Rightarrow \mu(f; A) = \mu(\mathbb{1}_A f) = \int_A f d\mu$$

Definition 35. If $f \in SF^+$, then $f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$, then $\mu(f) = \sum_{k=1}^n a_k \mu(A_k)$ $(\bigstar)$

Proposition 15.

1. $\bigstar$ is *well-defined*: if $f = \sum_{k=1}^n a_k \mathbb{1}_{A_k} = \sum_{l=1}^m b_l \mathbb{1}_{B_l}$ then $\sum_{k=1}^n a_k \mu(A_k) = \sum_{l=1}^m b_l \mu(B_l)$
2. Let $f, g \in SF^+$. If $\mu(f \neq g) = 0$ then $\mu(f) = \mu(g)$
3. Linearity: Let $f, g \in SF^+$ and $c \geq 0$. Then $\mu(f + g) = \mu(f) + \mu(g)$ and $\mu(cf) = c\mu(f)$
4. Monotonicity: Let $f, g \in SF^+$ and $f \leq g$. Then $\mu(f) \leq \mu(g)$

Proof.
(2): Take $f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$ and $g = \sum_{l=1}^m b_l \mathbb{1}_{B_l}$. Then:

$$\begin{aligned} g = g\mathbb{1}_{\{f=g\}} + g\mathbb{1}_{\{f\neq g\}} &= \sum_{k=1}^n a_k \mathbb{1}_{A_k \cap \{f=g\}} + \sum_{l=1}^m b_l \mathbb{1}_{B_l \cap \{f \neq g\}} \\ \implies \mu(g) &= \sum_{k=1}^n a_k \mu(A_k \cap \{f = g\}) + \sum_{l=1}^m b_l \mu(B_l \cap \{f \neq g\}) = \mu(f), \end{aligned}$$

since $\mu(B_l \cap \{f \neq g\}) = 0$ (because $B_l \cap \{f \neq g\} \subseteq \{f \neq g\}$).

(4): Write any f and g in SF$^+$ as:

$$\begin{aligned} f &= \sum_{k=1}^n a_k \mathbb{1}_{A_k} = \sum_{k=1}^n \sum_{l=1}^m a_k \mathbb{1}_{A_k \cap B_l}, \\ g &= \sum_{l=1}^m b_l \mathbb{1}_{B_l} = \sum_{l=1}^m \sum_{k=1}^n b_k \mathbb{1}_{A_k \cap B_l}. \end{aligned}$$

with $f \leq g$. If $A_k \cap B_k \neq \emptyset$, then $a_k \leq b_l$ by assumption so it's okay. If $A_k \cap B_l = \emptyset$, then the measure does not change. So $\mu(f) \leq \mu(g)$ □

5.2.2 Define integral for $f \in (m\Sigma)^+$

Definition 36. Let $f \in (m\Sigma)^+$ then $\mu(f) = \sup\{\mu(h) \mid h \in SF^+, h \leq f\}$. (Δ)

Proposition 16. Let $f \in SF^+$, then $(\bigstar) \equiv (\Delta)$.

Proposition 17. If $f \in (m\Sigma)^+$ and $\mu(f) = 0$ then $\mu(f \neq 0) = 0$.

Proof. Assume $\mu(f > 0) > 0$. Then there exists $n > 0$ and $\epsilon > 0$ such that $\mu(f > n^{-1}) \geq \epsilon$.

$$n^{-1}\mathbb{1}_{\{f>n^{-1}\}} < f \implies \mu(n^{-1}\mathbb{1}_{\{f>n^{-1}\}}) \leq \mu(f) \implies n^{-1}\mu(f > n^{-1}) \leq \mu(f) = 0.$$

□

Theorem 12. Suppose $f \in (m\Sigma)^+$ and $h_n \in SF^+$ and $h_n \uparrow f$. Then $\mu(f) = \lim(\mu(h_n))$.

Proof. Let $c := \lim_n(\mu(h_n))$. By definition, $c \leq \mu(f)$. It remains to show $\mu(f) \leq c$. It suffices to show that, for all $g \in \mathrm{SF}^+$ with $g \leq f$, that $\mu(g) \leq c$. We can break this down more explicitly: take $g = \sum_{k=1}^m a_k \mathbb{1}_{A_k}$ with $g \leq f$ and the A_k's are disjoint; note that $h_n = \sum_{k=1}^m h_n \mathbb{1}_{A_k}$ too. So:

$$\mu(g) \leq c \iff \sum_{k=1}^m a_k\mu(A_k) \leq \lim_n \sum_{k=1}^m \mu(h_n\mathbb{1}_{A_k}) \iff \forall k = 1, \ldots, m : a_k\mu(A_k) \leq \lim_n \mu(h_n\mathbb{1}_{A_k}).$$

In the case where $a_k = 0$, we're done. For $a_k \in (0, \infty)$: take $\epsilon > 0$ and define the following set

$$C_n^\epsilon := \{s \in A_k \mid h_n(s) > (1-\epsilon)a_k\}.$$

Since $h_n(s) \uparrow f(s)$ for all $s \in A_k$, we have that $\lim_n h_n(s) \geq a_k$ for all $s \in A_k$. Thus $C_n^\epsilon \uparrow A_k$ (as $n \to \infty$). By definition of C_n^ϵ and its properties:

$$\begin{aligned}\mu(h_n\mathbb{1}_{A_k}) \geq \mu(h_n\mathbb{1}_{C_n^\epsilon}) \geq (1-\epsilon)a_k\mu(C_n^\epsilon) &\implies \lim_n \mu(h_n\mathbb{1}_{A_k}) \geq (1-\epsilon)a_k\mu(A_k) \\ &\implies \lim_n \mu(h_n\mathbb{1}_{A_k}) \geq a_k\mu(A_k)\end{aligned}$$

by continuity of μ and taking $\epsilon \to 0$. The final case is when $a_k = \infty$ but the approach is similar. Fix $M > 0$ and again define a nice set

$$C_n^M := \{s \in A_k \mid h_n(s) > M\},$$

with $C_n^M \uparrow A_k$ (as $n \to \infty$) once again.

$$\lim_n \mu(h_n\mathbb{1}_{A_k}) \geq \lim_n \mu(h_n\mathbb{1}_{C_n^M}) \geq M \lim_n \mu(C_n^M) \geq M\mu(A_k) \implies \lim_n \mu(h_n\mathbb{1}_{A_k}) = \infty$$

by continuity of μ and taking $M \to \infty$. □

Remark 11. Let $f \in (m\Sigma)^+$, then (Δ) does not depend on the choice of $\{h_n : n \geq 1\}$

Proposition 18.

1. Let $f, g \in (m\Sigma)^+$. If $f = g$ a.s. (i.e. if $\mu(f \neq g) = 0$) then $\mu(f) = \mu(g)$
2. Linearity: Let $f, g \in (m\Sigma)^+$ and $c \geq 0$. Then $\mu(f + g) = \mu(f) + \mu(g)$ and $\mu(cf) = c\mu(f)$
3. Monotonicity: Let $f, g \in (m\Sigma)^+$ and $f \leq g$. Then $\mu(f) \leq \mu(g)$

5.2.3 Define integral for $f \in m\Sigma$

Definition 37. Let $f \in m\Sigma$, then $\mu(f)$ is defined if *at least* one of $\mu(f^+), \mu(f^-) < \infty$ with

$$\mu(f) := \mu(f^+) - \mu(f^-).$$

Definition 38. $f \in m\Sigma$ is called **integrable**, denoted $f \in L^1(S, \Sigma, \mu)$ if $\mu(|f|) < \infty$

Remark 12.

- $\mu(f)$ can be $\pm\infty$ or real-valued
- $|\mu(f)| \leq \mu(|f|)$
- $\mu(f) \in \mathbb{R} \iff \mu(f^+), \mu(f^-) < \infty \iff \mu(|f|) = \mu(f^+) + \mu(f^-) < \infty \iff f \in L^1(S, \Sigma, \mu)$

Proposition 19. Let $f, g \in L^1(S, \Sigma, \mu)$ and $c \in \mathbb{R}$

1. $\mu(\{f = \pm\infty\}) = 0$
2. $f + g \in L^1$ with $\mu(f + g) = \mu(f) + \mu(g)$ and $cf \in L^1$ with $\mu(cf) = c\mu(f)$
3. If $f \leq g$, then $\mu(f) \leq \mu(g)$

Proof. (2): Set $h = f + g$, then

$$h = h^+ - h^- = (f^+ - f^-) + (g^+ - g^-) \iff h^+ + f^- + g^+ = h^- + f^+ + g^-.$$

Since $h^\pm, f^\pm, g^\pm \in (m\Sigma)^+$ and so we can use linearity. Since $\mu(f), \mu(g) \in \mathbb{R}$, then $\mu(f^\pm), \mu(g^\pm) < \infty$. Furthermore, $h^\pm \leq |h| \leq |f| + |g| \implies \mu(h^\pm) \leq \mu(|f|) + \mu(|g|) < \infty$. Taking the measure of the above equation and (safely) re-rearranging, we get the result. □

5.3 Convergence Theorems

Theorem 13. (Monotone Convergence Theorem) Let $f_n \in m\Sigma$ with $f_n \leq f_{n+1}$ (non-decreasing) and $\mu(f_1^-) < \infty$ (floor bounded). Then if $f_n \uparrow f$, $\mu(f_n) \uparrow \mu(f)$

Corollary 3. Suppose $f_n : S \to \overline{\mathbb{R}} \in m\Sigma$ and $\mu(f_1^+) < \infty$. Then if $f_n \downarrow f$, $\mu(f_n) \downarrow \mu(f)$

Theorem 14. (Fatou's Lemma) Suppose $f_n : S \to \overline{\mathbb{R}} \in m\Sigma$ and $\exists\ g : S \to \overline{\mathbb{R}} \in m\Sigma$ such that $\mu(g^-) < \infty$ and $f_n \geq g$ for all $n \geq 1$. Then

$$\mu(\liminf(f_n)) \leq \liminf(\mu(f_n))$$

Theorem 15. (Reverse Fatou's Lemma) Suppose $f_n : S \to \overline{\mathbb{R}}$ and $f_n \leq g$ for some $g : S \to \overline{\mathbb{R}} \in m\Sigma$ and $\mu(g^+) < \infty$. Then

$$\mu(\limsup(f_n)) \geq \limsup(\mu(f_n))$$

Theorem 16. (Dominated Convergence Theorem) Suppose $f_n, f : S \to \mathbb{R} \in m\Sigma$ and $\lim(f_n) = f$ and $\exists\ g \in L^1(S, \Sigma, \mu)$ such that $|f_n| \leq |g|$ for all $n \geq 1$. Then

1. $f \in L^1(S, \Sigma, \mu)$
2. $f_n \to f$ "in L^1" i.e. $\mu(|f_n - f|) \to 0$ as $n \to \infty \Rightarrow \lim(\mu(f_n)) = \mu(f)$

Theorem 17. (Scheffe's Lemma) Suppose $f_n, f \in L^1(S, \Sigma, \mu)$ and $f_n \to f$ (pointwise convergence). Then

$$f_n \xrightarrow{L^1} f \Leftrightarrow \lim(\mu(|f_n|)) = \mu(|f|)$$

5.4 More Measure Theory

Definition 39. Let μ and λ be two measures on (S, Σ). Then λ is said to be **absolutely continuous with respect to $\boldsymbol{\mu}$**, denoted $\lambda << \mu$, if

$$\forall A \in \Sigma, \mu(A) = 0 \Rightarrow \lambda(A) = 0.$$

Definition 40. λ is said to be **singular to $\boldsymbol{\mu}$**, denoted $\lambda \perp\!\!\!\perp \mu$, if

$$\exists B \in \Sigma \text{ such that } \lambda(B) = \mu(B^C) = 0/$$

Definition 41. Given a measure space (S, Σ, μ) and $f \in (m\Sigma)^+$, then $f\mu$ is a set function on Σ such that:

$$\forall A \in \Sigma, (f\mu)(A) := \int_A f d\mu$$

and it is also a measure on Σ.

Proposition 20. $\forall h \in (m\Sigma)^+$, we have that $(f\mu)(h) = \mu(fh)$

Proof. Suppose $h = \mathbb{1}_A \in (m\Sigma)^+$ (for some $A \in \Sigma$); then this property obviously holds. By linearity, this would also hold for any function in SF^+. By (MON), this also holds for any function in $(m\Sigma)^+$. □

Proposition 21. $h \in L^1(S, \Sigma, f\mu) \iff fh \in L^1(S, \Sigma, \mu)$ and $(f\mu)(h) = \mu(fh)$

Proof.

$$\begin{aligned} h \in L^1(S, \Sigma, f\mu) &\iff (f\mu)(h^+) < \infty, (f\mu)(h^-) < \infty \\ &\iff \mu(fh^+) < \infty, \mu(fh^+) < \infty \\ &\iff \mu((fh)^+) < \infty, \mu((fh)^-) < \infty \iff fh \in L^1(S, \Sigma, \mu). \end{aligned}$$

Also: $(f\mu)(h) = (f\mu)(h^+) - (f\mu)(h^-) = \mu((fh)^+) - \mu((fh)^-) = \mu(fh)$. □

Theorem 18. (Radon-Nikodym) Let λ and μ be two σ-finite measures on (S, Σ) and assume $\lambda << \mu$. Then there exists an $f \in (m\Sigma)^+$ such that $\lambda = f\mu$.

Remark 13. In the above case, f is called the **Radon-Nikodym Derivative** of λ with respect to μ, denoted $f = d\lambda/d\mu$

Theorem 19. (Lebesgue Decomposition) Let λ and μ be finite measures. Then there exists a unique pair of measures λ_1 and λ_2 such that $\lambda_1 + \lambda_2 = \lambda$ with $\lambda_1 << \mu$ and $\lambda_2 \perp\!\!\!\perp \mu$

Chapter 6

Expectations and L^p Spaces

6.1 Expectation

Definition 42. We define the **expectation** of a random variable X as $\mathbb{E}[X] = \int_\Omega X d\mathbb{P}$.

6.1.1 Strengthen convergence theorems

(MON), (DOM), (SCH) can all be strengthened in this context. That is, instead of "$X_n \to X$ a.e.", we can simply replace it with "$X_n \xrightarrow{p} X$". In these theorems, $L^1 := L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem 20. (MON') Suppose $X_n \le X_{n+1}$ and $X_n \xrightarrow{p} X$ and $\mathbb{E}[X_1^-] < \infty$. Then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$

Theorem 21. (DOM') Suppose $X_n \xrightarrow{p} X$ and $\exists\ Y \in L^1$ with $|X_n| \le Y$ (for all n). Then $X_n \xrightarrow{L^1} X$.

Theorem 22. (SCH') Suppose $X_n, X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}[|X_n|] \longrightarrow \mathbb{E}[|X|]$. Then $X_n \xrightarrow{L^1} X$.

6.2 L^p Spaces

Definition 43. For $p \in [1, \infty]$, define $L^p := L^p(\Omega, \mathcal{F}, \mathbb{P})$ to be the collection of random variables such that $\mathbb{E}[|X|^p] < \infty$. We say that $X \in L^p \iff X$ is p-integrable. Furthermore:

$$\|X\|_p := \left(\int_\Omega |X|^p d\mathbb{P}\right)^{1/p} = (\mathbb{E}[|X|^p])^{1/p}$$

Fact 8. (Properties of L^p Spaces)

1. L^p is a vector space,
2. $\|\cdot\|_p$ is a norm,
3. **(Cauchy-Schwarz)** $\forall X, Y \in L^2 : XY \in L^1$ and $\mathbb{E}[|XY|] \le \|X\|_2 \|Y\|_2$,
4. **(Holder)** $\forall X \in L^p, Y \in L^q$ with $p^{-1} + q^{-1} = 1$: $XY \in L^1$ and $\mathbb{E}[|XY|] \le \|X\|_p \|Y\|_q$,
5. **(Monotonicity)** Suppose $1 \le p \le q \le \infty$, then: $X \in L^q \implies X \in L^p$ and $\|X\|_p \le \|X\|_q$,
6. $(L^p, \|\cdot\|_p)$ is a Banach Space (complete with respect to this norm),

7. If $p = 2$, $(L^2, ||\cdot||_2)$ is a **Hilbert Space** with inner product $\langle X, Y\rangle_2 = \mathbb{E}[XY]$,

8. $\forall X \in L^2$, $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Remark 14.

- $\text{Var}(X) := 0 \iff \exists c \in \mathbb{R}$ such that $\mathbb{P}(X = c) = 1$,
- If $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 0$, X and Y are uncorrelated but not necessarily independent,
- $(X, Y) \sim \text{N}(\bar{\mu}, \Sigma)$ then $X \perp\!\!\!\perp Y \Leftrightarrow X$ and Y are uncorrelated.

Theorem 23. (Markov's Inequality) Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a real-valued random variable X. If $g : \mathbb{R} \to [0, \infty)$ is a non-decreasing Borel measurable function then for all $c > 0$, if $g(c) \neq 0$:

$$\mathbb{E}(g(X)) \geq \mathbb{E}[g(X)\mathbb{1}_{\{X\geq c\}}] \geq g(c)\mathbb{P}(X \geq c) \implies \mathbb{P}(X \geq c) \leq \frac{1}{g(c)}\mathbb{E}[g(X)\mathbb{1}_{\{X\geq c\}}] \leq \frac{\mathbb{E}[g(X)]}{g(c)}.$$

Example 10.

- If $X \in L^1$ and $c > 0$: $\mathbb{P}(|X| \geq c) \leq c^{-1}\mathbb{E}[|X|\mathbb{1}_{\{|X|\geq c\}}] \leq c^{-1}\mathbb{E}[|X|]$,
- If $X \in L^p$ and $c > 0$: $\mathbb{P}(|X| \geq c) \leq c^{-p}\mathbb{E}[|X|^p; |X| \geq c] \leq c^{-p}\mathbb{E}[|X|^p]$,
- If $\exists \alpha > 0$ s.t. $e^{\alpha|X|} \in L^1$, then for all $c > 0$: $\mathbb{P}(|X| \geq c) \leq e^{-\alpha c}\mathbb{E}[e^{\alpha|X|}; |X| \geq c] \leq e^{-\alpha c}\mathbb{E}[e^{\alpha|X|}]$.

Proposition 22. $\{X_n : n \geq 1\}, X \in L^p$ $(p \geq 1)$.

$$X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{p} X.$$

Proof. For any $\epsilon > 0$, use Markov's Inequality:

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \epsilon^{-p}\mathbb{E}[|X_n - X|^p] \longrightarrow 0.$$

□

6.3 Uniform Integrability (UI)

Remark 15. There are counterexamples to show that convergence "almost surely" does not imply L^1 convergence; and the converse is also not true! Similarly, convergence "in probability" alone is not sufficient for L^1 convergence. For this, we need a little more structure.

Lemma 3. (Motivation) $X \in L^1 \iff \lim_{M\to\infty} \mathbb{E}[|X|; |X| > M] = 0$.

Proof. Forwards:

$$\lim_{M\to\infty} \mathbb{E}[|X|; |X| \leq M] = \mathbb{E}[|X|; |X| \leq \infty] = \mathbb{E}[|X|] < \infty \implies \lim_{M\to\infty} \mathbb{E}[|X|; |X| > M] = 0.$$

Backwards:

$$\mathbb{E}[|X|] = \mathbb{E}[|X|; |X| \leq M] + \mathbb{E}[|X|; |X| > M].$$

Taking the limit as $M \to \infty$, the last term drops and $X \in L^1$. □

Definition 44. Let $\{X_n : n \geq 1\}$ be a sequence of r.v. in a probability space. We say that $\{X_n\}$ is **uniformly integrable (UI)** if:

$$\lim_{M \to \infty} \sup_n \mathbb{E}[|X_n|; |X_n| > M] = 0$$

or: if for all $\epsilon > 0$, there exists an $M > 0$ such that $\mathbb{E}[|X_n|; |X_n| > M] < \epsilon$ (for all $n \geq 1$).

Proposition 23.

1. If $\{X_n\}$ is UI, then $\{X_n\}$ is bounded in L^1 i.e.

$$\{X_n\} \text{ is UI} \implies \sup_n \|X_n\|_1 = \sup_n \mathbb{E}[|X_n|] < \infty,$$

2. If $\{X_n\}$ is bounded in L^p (for $p > 1$), then $\{X_n\}$ is UI (not true for $p = 1$).

Proof.
(1):

$$\begin{aligned}\sup_n \mathbb{E}[|X_n|] &= \sup_n \mathbb{E}[|X_n|; |X_n| > M] + \sup_n \mathbb{E}[|X_n|; |X_n| \leq M] \\ &\leq \sup_n \mathbb{E}[|X_n|; |X_n| > M] + M < \epsilon + M < \infty.\end{aligned}$$

(2): For $0 < k \leq v$ (scalars), we have $k^p v \leq v^p k$, which implies $v \leq v^p k^{1-p}$. Using this inequality; let $k := M$ and $v := \mathbb{E}[|X_n|; |X_n| > M]$:

$$\begin{aligned}&\mathbb{E}[|X_n|; |X_n| > M] \leq \mathbb{E}\left[\frac{|X_n|^p}{M^{p-1}}; |X_n| > M\right] \leq M^{1-p}\mathbb{E}[|X_n|^p] \\ \implies &\sup_n \mathbb{E}[|X_n|; |X_n| > M] \leq M^{1-p} \underbrace{\sup_n \mathbb{E}[|X_n|^p]}_{<\infty} \longrightarrow 0 \quad (M \to \infty).\end{aligned}$$

$\square$

Theorem 24.

1. $\{X_n : n \geq 1\}, X \in L^1$ and $X_n \xrightarrow{L^1} X \implies X_n \xrightarrow{p} X$ and $\{X_n : n \geq 1\}$ is UI,

2. $\{X_n : n \geq 1\}$ is UI and $X_n \xrightarrow{p} X \implies X \in L^1$ and $X_n \xrightarrow{L^1} X$.

Proof.
(1): We already know L^1 convergence implies convergence in prob. For UI property:

$$\begin{aligned}\mathbb{E}[|X_n|; |X_n| > M] &\leq \mathbb{E}[|X_n - X|; |X_n| > M] + \mathbb{E}[|X|; |X_n| > M] \\ &\leq \mathbb{E}[|X_n - X|] + \mathbb{E}[|X|; |X_n| > M; |X| \leq \sqrt{M}] + \mathbb{E}[|X|; |X_n| > M; |X| > \sqrt{M}] \\ &\leq \mathbb{E}[|X_n - X|] + \sqrt{M}\mathbb{P}(|X_n| > M) + \mathbb{E}[|X|; |X| > \sqrt{M}] \\ &\leq \mathbb{E}[|X_n - X|] + \sqrt{M}\frac{\mathbb{E}[|X_n|]}{M} + \mathbb{E}[|X|; |X| > \sqrt{M}],\end{aligned}$$

where the last step used Markov's Inequality and currently $M > 0$ (unspecified). Now fix $\epsilon > 0$. For the first term, choose $N > 0$ such that $\mathbb{E}[|X_n - X|] < 3^{-1}\epsilon$ (due to L^1 convergence). Since $\{X_n\}$ are in L^1, choose M large enough such that $M^{-1/2} \sup \mathbb{E}[|X_n|] < 3^{-1}\epsilon$. If need be, take M even larger such that the last term is bounded above by $3^{-1}\epsilon$ too. Taking the supremum of both sides of the above expansion, we get that

$$\sup_n \mathbb{E}[|X_n|; |X_n| > M] < \epsilon.$$

(Note that there is a short argument to be made for $n \leq N$).

(2): We first show $X \in L^1$. Since $X_n \xrightarrow{p} X$, then there exists a subsequence that converges almost surely and since $f(\cdot) = |\cdot|$ is continuous, $|X_{n_k}| \longrightarrow |X|$ almost surely. Using Fatou's Inequality:

$$\mathbb{E}[|X|] = \mathbb{E}[\lim_k |X_{n_k}|] = \mathbb{E}[\liminf_k |X_{n_k}|] \leq \liminf_k \mathbb{E}[|X_{n_k}|] \leq \limsup_k \mathbb{E}[|X_{n_k}|] \leq \sup_n \mathbb{E}[|X_n|] < \infty.$$

Showing convergence in L^1 will take some work. **Claim:** $\{X_n - X\}$ is UI. We prove this as usual:

$$\begin{aligned} \mathbb{E}[|X_n - X|; |X_n - X| > M] &\leq \mathbb{E}[|X_n| + |X|; |X_n| > 2^{-1}M \text{ or } |X| > 2^{-1}M] \\ &\leq \mathbb{E}[|X_n|; |X_n| > 2^{-1}M] + \underbrace{\mathbb{E}[|X_n|; |X| > 2^{-1}M]}_{\text{(I)}} \\ &+ \underbrace{\mathbb{E}[|X|; |X_n| > 2^{-1}M]}_{\text{(II)}} + \mathbb{E}[|X|; |X| \geq 2^{-1}M]. \end{aligned}$$

We show the bound for (I); the one for (II) is analogous:

$$\begin{aligned} \mathbb{E}[|X_n|; |X| > 2^{-1}M] &\leq \mathbb{E}[|X_n|; |X| > 2^{-1}M; |X_n| \leq \sqrt{M}] + \mathbb{E}[|X_n|; |X| > 2^{-1}M; |X_n| > \sqrt{M}] \\ &\leq \sqrt{M}\mathbb{P}(|X| > 2^{-1}M) + \mathbb{E}[|X_n|; |X_n| > \sqrt{M}] \\ &\leq \frac{2}{\sqrt{M}}\mathbb{E}[|X|] + \mathbb{E}[|X_n|; |X_n| > \sqrt{M}]. \end{aligned}$$

Putting these terms together, taking the supremum and $M \longrightarrow \infty$, we prove the claim. To finally show convergence in L^1: fixing $\epsilon > 0$

$$\begin{aligned} \mathbb{E}[|X_n - X|] &= \mathbb{E}[|X_n - X|; |X_n - X| > M] + \mathbb{E}[|X_n - X|; |X_n - X| < \epsilon] \\ &+ \mathbb{E}[|X_n - X|; \epsilon \leq |X_n - X| \leq M] \\ &< \mathbb{E}[|X_n - X|; |X_n - X| > M] + \epsilon + M\mathbb{P}(|X_n - X| > \epsilon) < 3\epsilon \end{aligned}$$

(taking M sufficiently large, we bound the first term by ϵ; for n sufficiently large, last term is bounded above by $M^{-1}\epsilon$. Add them up and we're done). □

Proposition 24. (Jensen's Inequality) Let $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ be convex and $X, \varphi(X) \in L^1$. Then:

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

Example 11. Common examples of convex functions that might come up in problem are: $|\cdot|$, $|\cdot|^2$, $\exp(\cdot)$, $\max\{0, \cdot\}$, and $\max\{-\cdot, 0\}$.

Theorem 25. Suppose $f : \mathbb{R} \longrightarrow \mathbb{R}$ is Borel and X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution $\mathcal{L}_X$ such that for all $B \in \mathcal{B}$, $\mathcal{L}_X(B) = \mathbb{P}(X \in B)$. Then:

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

In particular, this means

$$\mathbb{E}[f(X)] = \int_\Omega f(X)\, d\mathbb{P} \overset{(\Delta)}{=} \int_\mathbb{R} f(x)\, d\mathcal{L}_X(x).$$

Proof. Define the class $\mathcal{H} := \{f \in b\mathcal{B} \text{ s.t. } (\Delta) \text{ holds}\}$; it is a monotone class (the last criteria is by (MON)). Take any $B \in \mathcal{B}$, then:

$$\int_\Omega \mathbb{1}_B \; d\mathbb{P} = \int_B \; d\mathbb{P} = \mathbb{P}(X \in B) = \mathcal{L}_X(B) \implies \mathbb{1}_B \in \mathcal{H} \overset{\text{MCT2}}{\implies} b\mathcal{B} = \mathcal{H}.$$

Now take $f \in \mathcal{H}$ and write $f = f^+ - f^-$. For each $k \geq 1$, define

$$f_k^\pm = f^\pm \mathbb{1}_{\{f^\pm \leq k\}} \in \mathcal{H} \implies f_k^\pm \uparrow f^\pm \quad ((\text{MON}))$$

and so (Δ) holds for $f^\pm$. Finally:

$$f(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \iff \mathbb{E}[f^\pm(X)] < \infty \iff \mathcal{L}_X(f^\pm) < \infty \iff f \in L^1(\mathbb{R}, \mathcal{B}, \mathcal{L}_X).$$

□

Definition 45. A random variable X has a **probability density function**, denoted by f_X if $\mathcal{L}_X \ll \lambda$, where λ is the Lebesgue measure and f_X is the R-N derivative $f_X = d\mathcal{L}_X/d\lambda$.

Remark 16. If X has pdf f_X and $h : \mathbb{R} \longrightarrow \mathbb{R}$ is Borel, then

$$h(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \; iff h \in L^1(\mathbb{R}, \mathcal{B}, \mathcal{L}_X) \iff f_X h \in L^1(\mathbb{R}, \mathcal{B}, \lambda).$$

Further, we have

$$\mathbb{E}[h(X)] = \int_\mathbb{R} h(x) \; d\mathcal{L}_X(x) = \int_\mathbb{R} h(x) f_X(x) \; d\lambda(x).$$

Remark 17. If a pdf exists, then the distribution function $F(x) = \mathbb{P}(X \leq x) = \mathcal{L}_X((-\infty, x])$ is absolutely continuous i.e. F is continuous. However, if F is continous then a pdf does not have to exist (e.g. Cantor function).

Lemma 4. Assume $X \perp\!\!\!\perp Y$ and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then: for all $B \in \mathcal{B}$,

$$\mathbb{E}[X; Y \in B] \overset{(\star)}{=} \mathbb{E}[X]\mathbb{P}(Y \in B).$$

Proof. Let $\Sigma = \sigma(X)$ then $\mathcal{H} := \{W \in b\Sigma \text{ s.t. } (\star) \text{ holds}\}$. It is a monotone class (third condition again via (MON)). Now taking an arbitrary $A \in \Sigma$, we know there exists a $E \in \mathcal{B}$ such that $A = \{X \in E\}$. Then $\mathbb{1}_A \in \mathcal{H}$ by independence, so $b\Sigma = \mathcal{H}$ by MCT2.

So for all $k \geq 1$, let $X_k := X\mathbb{1}_{\{|X| \leq k\}} \in \mathcal{H}$ and $X_k \uparrow X$. By (DOM):

$$\mathbb{E}[X; Y \in B] = \lim_k \mathbb{E}[X_k; Y \in B] = \lim_k \mathbb{E}[X_k]\mathbb{P}(Y \in B) = \mathbb{E}[X]\mathbb{P}(Y \in B).$$

□

Theorem 26. If $X \perp\!\!\!\perp Y$ and $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ then $XY \in L^1$ and

$$\mathbb{E}[XY] \overset{(\star\star)}{=} \mathbb{E}[X]\mathbb{E}[Y].$$

Proof. Hint: $\mathcal{H} := \{W \in b\sigma(Y) \text{ s.t. } (\star\star) \text{ holds}\}$. □

Corollary 4. If $X \perp\!\!\!\perp Y$ and f, g are Borel functions such that $f(X), g(Y) \in L^1$, then

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

Corollary 5. If $X \perp\!\!\!\perp Y$ and $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ then the covariance of X and Y is zero.

Chapter 7

Law of Large Numbers (LLN)

Definition 46. Let $\{X_n : n \geq 1\}$ be a sequence of r.v and define $S_n := \sum_{i=1}^n X_i$ (assumed to always mean this definition unless otherwise specified). The **Weak Law of Large Numbers**, abbreviated to **WLLN**, is said to hold for the sequence if

$$\frac{S_n - \mathbb{E}[S_n]}{n} \xrightarrow{p} 0.$$

Similarly, the **Strong Law of Large Numbers**, or **SLLN**, is said to hold when

$$\frac{S_n - \mathbb{E}[S_n]}{n} \longrightarrow 0 \quad \text{a.s.}$$

Remark 18. There are several WLLNs and SLLNs. They differentiate from each other by either the *moment conditions* or *distribution conditions*.

7.1 WLLN

Lemma 5. (Chebyshev's Inequality) If $X \in L^2$, then for all $c > 0$:

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

Theorem 27. (WLLN (1)) Let $\{X_n : n \geq 1\}$ be a sequence of *uncorrelated* random variables and suppose $\sup_n \mathbb{E}[X_n^2]$ is finite. Then the WLLN holds.

Proof. WLOG, assume $\mathbb{E}[X_n] = 0$, then the uncorrelated criteria means $\mathbb{E}[X_i X_j] = 0$ ($\forall i \neq j$). With $\mathbb{E}[S_n] = 0$, we fix $\epsilon > 0$ (and use Chebyshev's inequality at the start):

$$\begin{aligned}
\mathbb{P}\left(\frac{|S_n|}{n} > \epsilon\right) &= \mathbb{P}(|S_n| > n\epsilon) \\
&\leq (n\epsilon)^{-2}\mathbb{E}[S_n^2] = (n\epsilon)^{-2}\mathbb{E}\left[\left(\sum_{j=1}^n X_j\right)^2\right] \\
&= (n\epsilon)^{-2}\mathbb{E}\left[\sum_{j=1}^n X_j^2 + \sum_{i\neq j} X_i X_j\right] = (n\epsilon)^{-2}\mathbb{E}\left[\sum_{j=1}^n X_j^2\right] + (n\epsilon)^{-2}\mathbb{E}\left[\sum_{i\neq j} X_i X_j\right] \\
&= (n\epsilon)^{-2}\mathbb{E}\left[\sum_{j=1}^n X_j^2\right] \leq (n\epsilon)^{-2} n \sup_n \mathbb{E}[X_n^2] \longrightarrow 0 \quad (n \longrightarrow \infty),
\end{aligned}$$

where we bounded the "sum of n terms" by "n times the largest term". $\square$

Lemma 6. $\{X_n : n \geq 1\}$ *uncorrelated* and $X_n \in L^2$ $(\forall n \geq 1)$. Then $\operatorname{Var}(S_n) = \sum_{j=1}^n \operatorname{Var}(X_j)$.

Theorem 28. (Cantelli WLLN/SLLN) $\{X_n : n \geq 1\}$ uncorrelated and $\sup_n \mathbb{E}[X_n^4] < \infty$. Then the WLLN and SLLN hold.

Definition 47. Suppose $\{X_n : n \geq 1\}$ and $\{Y_n : n \geq 1\}$ are two sequences of r.v. They are said to be **equivalent** if

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

Remark 19. If $\{X_n\}$ is equivalent to $\{Y_n\}$

1. by BC1, we have $\mathbb{P}(X_n \neq Y_n, \text{ i.o}) = 0$. More explicitly, for almost every $\omega \in \Omega$:
$$\exists N_\omega \in \mathbb{N} \text{ such that } X_n(\omega) = Y_n(\omega) \ (\forall n \geq N_\omega).$$
2. For almost every $\omega \in \Omega$: $\sum_{n\geq 1} X_n(\omega)$ converges iff $\sum_{n\geq 1} Y_n(\omega)$ converges.
3. Let $\{b_n\} \subseteq \mathbb{R}_+$ (with $b_n \uparrow \infty$). For almost every $\omega \in \Omega$:
$$\frac{\sum_{j=1}^n X_j(\omega)}{b_n} \longrightarrow 0 \iff \frac{\sum_{j=1}^n Y_n(\omega)}{b_n} \longrightarrow 0.$$

Lemma 7. Let X be a non-negative random variable. Then

$$\sum_{n\geq 1} \mathbb{P}(X \geq n) < \infty \iff \mathbb{E}[X] < \infty.$$

Remark 20. (From real analysis) If $\{a_n\}$ is a sequence such that $a_n \longrightarrow 0$, then $n^{-1}\sum_{j=1}^n a_j \longrightarrow 0$.

Theorem 29. (WLLN (2)) Suppose $\{X_n\}$ be *pairwise* iid with $\mathbb{E}[X_1] = m$, with m finite. Then the WLLN holds i.e.

$$\frac{S_n}{n} \xrightarrow{p} m.$$

Proof. When there is no sufficient integrability assumption, we resort to a "truncation" method and then use the notion of equivalent sequences. Define $Y_n := X_n \mathbb{1}_{\{|X_n|\leq n\}}$ and let $T_n := \sum_{j=1}^n Y_j$ (note that $Y_n \in L^2$). We first show $\{X_n\}$ and $\{Y_n\}$ are equivalent:

$$\sum_{n\geq 1} \mathbb{P}(X_n \neq Y_n) = \sum_{n\geq 1} \mathbb{P}(|X_n| > n) \overset{\text{(iid)}}{=} \sum_{n\geq 1} \mathbb{P}(|X_1| > n) < \infty$$

by the above lemma. Furthermore, the pairwise independence of $\{X_n\}$ implies $\{Y_n\}$ are also pairwise independent, hence *uncorrelated*. Instead of directly showing the WLLN like in the previous case, we break this up into smaller terms using the triangle inequality:

$$\frac{|S_n - nm|}{n} \leq \frac{|S_n - T_n|}{n} + \frac{|T_n - \mathbb{E}[T_n]|}{n} + \frac{|\mathbb{E}[T_n] - nm|}{n}.$$

The two sequence $\{X_n\}$ and $\{Y_n\}$ are equivalent, so the first term converges to zero almost surely (hence in probability). The last term goes to zero (as $n \longrightarrow \infty$) by (re-writing and) using Lemma 7, Remark 20, Jensen's Inequality, and Lemma 3:

$$\frac{\left|\mathbb{E}\left[\sum_{j=1}^n X_j \mathbb{1}_{\{|X_j|\leq j\}}\right] - \sum_{j=1}^n \mathbb{E}[X_j]\right|}{n} \leq \frac{\sum_{j=1}^n \mathbb{E}[|X_j|; |X_j| > j]}{n} \overset{\text{iid}}{=} \frac{1}{n}\sum_{j=1}^n \mathbb{E}[|X_1|; |X_1| > j] \longrightarrow 0.$$

It remains to show that the middle term goes to zero in probability. Fix $\epsilon > 0$:

$$\mathbb{P}(n^{-1}|T_n - \mathbb{E}[T_n]| > \epsilon) \leq (n\epsilon)^{-2}\mathrm{Var}(T_n) = (n\epsilon)^{-2}\sum_{j=1}^{n}\mathrm{Var}(Y_j)$$
$$\leq (n\epsilon)^{-2}\sum_{j=1}^{n}\mathbb{E}[Y_j^2] = (n\epsilon)^{-2}\sum_{j=1}^{n}\mathbb{E}[X_j^2; |X_j| \leq j];$$

the first inequality is Chebyshev's and the second is a naive bound on the variance. Now choose $\{c_n\} \subseteq \mathbb{R}_+$ with $c_n \sim \sqrt{n}$ and split the sum accordingly (ignore potential "floors" and "ciels"). Note also, using the pairwise iid-ness of X_n, we write $\mathbb{E}[X_j^2; |X_j| \leq j] = j\mathbb{E}[|X_1|]$ (similarly with the terms $\mathbb{E}[X_j^2; |X_j| \leq c_n]$ and $\mathbb{E}[X_j^2; c_n < |X_j| \leq j]$). To this end:

$$\begin{aligned}
&\mathbb{P}(n^{-1}|T_n - \mathbb{E}[T_n]| > \epsilon) \leq (n\epsilon)^{-2}\sum_{j=1}^{n}\mathbb{E}[X_j^2; |X_j| \leq j] \\
&\leq (n\epsilon)^{-2}\left[\sum_{j=1}^{c_n}\mathbb{E}[X_j^2; |X_j| \leq j] + \sum_{j=c_n+1}^{n}\mathbb{E}[X_j^2; |X_j| \leq j]\right] \\
&\overset{\text{iid}}{=} (n\epsilon)^{-2}\left[\sum_{j=1}^{c_n} j\mathbb{E}[|X_1|] + \sum_{j=c_n+1}^{n}\mathbb{E}[X_j^2; c_n < |X_j| \leq j] + \sum_{j=c_n+1}^{n}\mathbb{E}[X_j^2; |X_j| \leq c_n]\right] \\
&\leq (n\epsilon)^{-2}\left[m\sum_{j=1}^{c_n} j + \sum_{j=c_n+1}^{n} j\,\underbrace{\mathbb{E}[|X_1|; |X_1| > c_n]}_{\longrightarrow 0} + \sum_{j=c_n+1}^{n} c_n\mathbb{E}[|X_1|]\right] \\
&= (n\epsilon)^{-2}\left[O(c_n^2) + o(n^2) + O(nc_n)\right] = (n\epsilon)^{-2}\left[O(n) + o(n^{-2}) + O(n^{3/2}\right] \longrightarrow 0,
\end{aligned}$$

where we used Lemma 3 again to get the necessary "little o" term. □

Theorem 30. (WLLN (3)) Suppose $\{X_n : n \geq 1\}$ are pairwise independent and $\{b_n\} \subseteq \mathbb{R}_+$ (with $b_n \uparrow \infty$) such that

1. $\sum_{j=1}^{n}\mathbb{P}(|X_j| > b_n) \longrightarrow 0$ (as $n \longrightarrow \infty$),
2. $(b_n)^{-2}\sum_{j=1}^{n}\mathbb{E}[|X_j|^2; |X_j| \leq b_n] \longrightarrow 0$ (as $n \longrightarrow \infty$).

Then, if $a_n := \sum_{j=1}^{n}\mathbb{E}[X_j; |X_j| \leq b_n]$, we have $(b_n)^{-1}(S_n - a_n) \xrightarrow{p} 0$.

Proof. Define the r.v $Y_{n,j} := X_j\mathbb{1}_{\{|X_j| \leq b_n\}}$ ($\forall n \geq 1$ with $j = 1, \ldots, n$) that is pairwise independent with respect to j and let $T_n := \sum_{j=1}^{n} Y_{n,j}$. To complete the proof, fix $\epsilon > 0$ and start with the triangle inequality. The first term follows immediately from Assumption 1. For the second term: we note that $a_n \equiv \mathbb{E}[T_n]$ and, since the $Y_{n,j}$'s are pairwise independent, we bound the variance after using Chebyshev's inequality (the resulting term happens to be Assumption 2).

$$\begin{aligned}
&\mathbb{P}((b_n)^{-1}|S_n - a_n| > \epsilon) \leq \mathbb{P}((b_n)^{-1}|S_n - T_n| > \epsilon/2) + \mathbb{P}((b_n)^{-1}|T_n - a_n| > \epsilon/2) \\
&\leq \mathbb{P}(S_n \neq T_n) + \mathbb{P}((b_n)^{-1}|T_n - a_n| > \epsilon/2) \leq \sum_{j=1}^{n}\mathbb{P}(|X_j| > b_n) + \mathbb{P}((b_n)^{-1}|T_n - a_n| > \epsilon/2) \\
&\leq \sum_{j=1}^{n}\mathbb{P}(|X_j| > b_n) + (2/\epsilon)(b_n)^{-2}\mathrm{Var}(T_n) \leq \sum_{j=1}^{n}\mathbb{P}(|X_j| > b_n) + (2/\epsilon)(b_n)^{-2}\sum_{j=1}^{n}\mathbb{E}[Y_{n,j}^2] \\
&= \sum_{j=1}^{n}\mathbb{P}(|X_j| > b_n) + (2/\epsilon)(b_n)^{-2}\sum_{j=1}^{n}\mathbb{E}[|X_j|^2; |X_j| \leq b_n] \longrightarrow 0.
\end{aligned}$$

$\square$

7.2 SLLN

Theorem 31. (SLLN (1)) Let $\{X_n : n \geq 1\}$ be a sequence of *uncorrelated* random variables and suppose $\sup_n \mathbb{E}[X_n^2]$ is finite. Then the WLLN holds.

Proof. WLOG, assume $\mathbb{E}[X_n] = 0$, then the uncorrelated criteria means $\mathbb{E}[X_i X_j] = 0$ $(\forall i \neq j)$. With $\mathbb{E}[S_n] = 0$, we fix $\epsilon > 0$ (and use Chebyshev's inequality at the start):

$$
\begin{aligned}
\mathbb{P}\left(\frac{|S_n|}{n} > \epsilon\right) &= \mathbb{P}(|S_n| > n\epsilon) \\
&\leq (n\epsilon)^{-2}\mathbb{E}[S_n^2] = (n\epsilon)^{-2}\mathbb{E}\left[\left(\sum_{j=1}^{n} X_j\right)^2\right] \\
&= (n\epsilon)^{-2}\mathbb{E}\left[\sum_{j=1}^{n} X_j^2 + \sum_{i\neq j} X_i X_j\right] \\
&= (n\epsilon)^{-2}\mathbb{E}\left[\sum_{j=1}^{n} X_j^2\right] + (n\epsilon)^{-2}\mathbb{E}\left[\sum_{i\neq j} X_i X_j\right] \\
&= (n\epsilon)^{-2}\mathbb{E}\left[\sum_{j=1}^{n} X_j^2\right] \leq (n\epsilon)^{-2} n \sup_n \mathbb{E}[X_n^2] \longrightarrow 0 \quad (n \longrightarrow \infty),
\end{aligned}
$$

where we bounded the "sum of n terms" by "n times the largest term". $\square$

Lemma 8. (Kronecker's Lemma) Let $\{x_n\} \subseteq \mathbb{R}$, $\{a_n\} \subseteq \mathbb{R}_+$ (with $a_n \uparrow \infty$). Then

$$
\sum_{n\geq 1} \frac{x_n}{a_n} < \infty \implies \frac{1}{a_n}\sum_{j=1}^{n} x_j \longrightarrow 0.
$$

Theorem 32. (Kolmogorov's Inequality) Let $\{X_n : n \geq 1\}$ be a sequence of independent r.v with $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] < \infty$. Then: $\forall \epsilon > 0$,

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

Proof. Let $A := \{\max_{1\leq n} |S_j| > \epsilon\}$ and $A_j := \{|S_i| < \epsilon, 1 \leq i \leq j-1; |S_j| > \epsilon\}$. The A_j's form a partition of A and so $\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}(A_j)$. Note that for $j = 1, \ldots, n$: $S_n - S_j \perp\!\!\!\perp S_j$). To this end:

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

$\square$

Theorem 33. Let $\{X_n : n \geq 1\}$ be independent r.v with $\mathbb{E}[X_n] = 0$ and $\sum_{n\geq 1} \mathbb{E}[X_n^2] < \infty$. Then

$$\sum_{n\geq 1} X_n \text{ converges almost surely.}$$

Proof. Fix $N \geq 1$ and consider the sequence $\{X_{N+n} : n \geq 1\}$. Define $T_m := S_{m+N} - S_N = \sum_{j=1}^m X_{N+j}$. For any $\epsilon > 0$, apply Kolmogorov's inequality to this random variable gives

$$\mathbb{P}\left(\max_{1\leq j\leq m} |T_j| > \epsilon\right) \leq \frac{\mathbb{E}[T_m^2]}{\epsilon^2} = \frac{\sum_{j=1}^m \mathbb{E}[X_{N+j}^2]}{\epsilon^2},$$

and so

$$\mathbb{P}\left(\sup_{j\geq 1} |T_j| > \epsilon\right) = \mathbb{P}\left(\bigcup_{m\geq 1}\left\{\max_{1\leq j\leq m} |T_j| > \epsilon\right\}\right) = \lim_m \mathbb{P}\left(\max_{1\leq j\leq m} |T_j| > \epsilon\right) \leq \frac{\sum_{j=1}^\infty \mathbb{E}[X_{N+j}^2]}{\epsilon^2}.$$

Re-writing the sum and using the second assumption:

$$\mathbb{P}\left(\sup_{j\geq 1} |S_{N+j} - S_n| > \epsilon\right) \leq \frac{\sum_{l=N+1}^\infty \mathbb{E}[X_l^2]}{\epsilon^2} \longrightarrow 0 \quad (\bigstar)$$

as $N \longrightarrow \infty$. To prove the result, we show that the opposite cannot happen. Note that

$$\{\{S_n\} \text{ does not converge}\} \subseteq \{\{S_n\} \text{ is not Cauchy}\} \subseteq \bigcup_{m\geq 1}\bigcap_{N\geq 1}\left\{\sup_{j\geq 1} |S_{N+j} - S_N| > \frac{1}{m}\right\}.$$

Thus, for all $m \geq 1$:

$$\begin{aligned}\mathbb{P}\left(\bigcap_{N\geq 1}\left\{\sup_{j\geq 1} |S_{N+j} - S_N| > \frac{1}{m}\right\}\right) &\leq \mathbb{P}\left(\liminf_{N\longrightarrow\infty}\left\{\sup_{j\geq 1} |S_{N+j} - S_N| > \frac{1}{m}\right\}\right)\\ &\overset{\text{Fatou}}{\leq} \liminf_{N\longrightarrow\infty} \mathbb{P}\left(\left\{\sup_{j\geq 1} |S_{N+j} - S_N| > \frac{1}{m}\right\}\right) \overset{\bigstar}{=} 0\\ &\implies \mathbb{P}(\{S_N\} \text{ does not converge}) = 0.\end{aligned}$$

$\square$

Theorem 34. (SLLN (2)) Let $\{Y_n : n \geq 1\}$ be a sequence of independent r.v with (for some $\{b_n\} \subseteq \mathbb{R}_+$ $(b_n \uparrow \infty)$) $\sum_{n\geq 1}(b_n)^{-2}\mathrm{Var}(Y_n) < \infty$. Then

$$\frac{1}{b_n}\sum_{j=1}^n (Y_j - \mathbb{E}[Y_j]) \longrightarrow 0 \text{ a.s.}$$

Proof. Letting $X_n := \dfrac{Y_n - \mathbb{E}[Y_n]}{b_n}$, we have that $\{X_n : n \geq 1\}$ is also a sequence of independent r.v with $\mathbb{E}[X_n] = 0$ and $\sum_{n\geq 1} \mathbb{E}[X_n^2]$ finite. By Theorem 33, $\sum_{n\geq 1} X_n$ converges almost surely. Using Kronecker's Lemma, the proof is complete:

$$\sum_{n\geq 1} \frac{Y_n - \mathbb{E}[Y_n]}{b_n} < \infty \text{ a.s} \implies \frac{1}{b_n}\sum_{j=1}^n (Y_j - \mathbb{E}[Y_j]) \longrightarrow 0 \text{ a.s.}$$

$\square$

Theorem 35. (SLLN (3)) Suppose $\{X_n : n \geq 1\}$ are iid r.v. Then:

1. $\mathbb{E}[|X_1|] < \infty \implies \dfrac{S_n}{n} \longrightarrow \mathbb{E}[X_1]$ a.s.

2. $\mathbb{E}[|X_1|] = \infty \implies \limsup \left(\dfrac{S_n}{n}\right) = \infty$ a.s.

Proof.
(1): Let $Y_n := X_n \mathbb{1}_{\{|X_n| \leq n\}}$. From the proof of WLLN (2), we know that $\{Y_n\}$ and $\{X_n\}$ are equivalent sequences. Using the triangle inequality we get:

$$\frac{\left|\sum_{j=1}^n (X_j - \mathbb{E}[X_j])\right|}{n} \leq \frac{\left|\sum_{j=1}^n (X_j - Y_j)\right|}{n} + \frac{\left|\sum_{j=1}^n (Y_j - \mathbb{E}[Y_j])\right|}{n} + \frac{\left|\sum_{j=1}^n \mathbb{E}[Y_j - X_j]\right|}{n}.$$

The first term goes to zero because the sequences are equivalent. The last term goes to zero by using Lemma 3 and Remark 20:

$$\frac{\left|\sum_{j=1}^n \mathbb{E}[Y_j - X_j]\right|}{n} \leq \frac{\left|\sum_{j=1}^n \mathbb{E}[X_j; |X_j| > j]\right|}{n} \leq \frac{1}{n}\sum_{j=1}^n \mathbb{E}[|X_1|; |X_1| > j] \longrightarrow 0.$$

It remains to show the middle term goes to zero. In order to use SLLN (2), it remains to check the variance condition (since $\{Y_n\}$ is independent):

$$\begin{aligned}
\sum_{n\geq 1} \frac{\mathrm{Var}(Y_n)}{n^2} \leq \sum_{n\geq 1} \frac{\mathbb{E}[Y_n^2]}{n^2} &= \sum_{n\geq 1} \frac{\mathbb{E}[X_n^2; |X_n| \leq n]}{n^2} \\
&= \sum_{n\geq 1}\sum_{j=1}^n \frac{1}{n^2}\mathbb{E}[X_n^2; j-1 \leq |X_n| < j] \\
&\overset{\text{iid}}{=} \sum_{n\geq 1}\sum_{j=1}^n \frac{1}{n^2}\mathbb{E}[X_1^2; j-1 \leq |X_1| < j] \\
&= \sum_{j\geq 1}\left(\sum_{n=j}^{\infty} n^{-2}\right)\mathbb{E}[X_1^2; j-1 \leq |X_1| < j] \\
&= \sum_{j\geq 1} O(j^{-1}) j \mathbb{E}[|X_1|; j-1 \leq |X_1| < j] \\
&= C\sum_{j\geq 1} \mathbb{E}[|X_1|; j-1 \leq |X_1| < j] < \infty.
\end{aligned}$$

(2): For all $A > 0$, via Lemma 7, the iid-ness of $\{X_n\}$ and BC2, we get:

$$\begin{aligned}
\mathbb{E}[|X_1|/A] = \infty \implies \infty &= \sum_{n\geq 1} \mathbb{P}(|X_1| \geq An) = \sum_{n\geq 1} \mathbb{P}(|X_n| \geq An) \\
&\overset{\text{BC2}}{\implies} \mathbb{P}(|X_n| \geq An, \text{ i.o}) = 1 \implies \mathbb{P}(|S_n - S_{n-1}| \geq An, \text{ i.o}) = 1.
\end{aligned}$$

Note the following set inclusions:

$$\begin{aligned}
&\{|S_n - S_{n-1}| \geq An\} \subseteq \{|S_n| \geq An/2\} \cup \{|S_{n-1}| \geq An/2\} \subseteq \{|S_n| \geq An/2\} \cup \{|S_{n-1}| \geq A(n-1)/2\} \\
&\implies 1 = \mathbb{P}(|S_n - S_{n-1}| \geq An, \text{ i.o}) \leq \mathbb{P}(|S_n| \geq An/2, \text{ i.o}) \\
&\implies \mathbb{P}(|S_n|/n \geq A, \text{ i.o}) = 1 \implies \limsup(|S_n|/n) = \infty \text{ a.s.}
\end{aligned}$$

$\square$

7.3 Applications

7.3.1 Weierstrass Approximation Theorem

Theorem 36. Let f be a continuous function on $[0,1]$. For each $n \geq 1$, let

$$p_n(x) = \sum_{k=0}^{n} f(k/n)\binom{n}{k}x^k(1-x)^{n-k} \quad \text{(Bernstein polynomial)}$$

for $x \in [0,1]$. Then: for all $x \in [0,1]$, $p_n(x) \longrightarrow f(x)$ (as $n \longrightarrow \infty$) (the convergence is uniform!)

Proof. Fix $x \in [0,1]$ and let $\{X_m : m \geq 1\}$ be a sequence of iid Bernoulli random variables i.e.

$$\mathbb{P}(X_m = 1) = x,\ \mathbb{P}(X_m = 0) = 1 - x.$$

Let $S_n := \sum_{j=1}^n X_j$, then $S_n \sim \text{bin}(n,x)$ i.e.

$$\mathbb{P}\left(\frac{S_n}{n} = \frac{k}{n}\right) = \binom{n}{k}x^k(1-x)^{n-k}$$

for all $k = 1, 2, \ldots$. By definition, we have $p_n(x) = \mathbb{E}[f(S_n/n)]$. By the WLLN, $S_n/n \xrightarrow{p} x$ and, by continuity of f: $f(S_n/n) \xrightarrow{p} f(x)$. Since f is continuous on a compact set, it is bounded. Using (DOM'), we have:

$$f(S_n/n) \xrightarrow{L^1} f(x) \implies p_n(x) = \mathbb{E}[f(S_n/n)] \longrightarrow f(x).$$

We now show uniform convergence. Fixing $\epsilon > 0$:

$$\begin{aligned}
&|p_n(x) - f(x)| \leq \mathbb{E}[|f(S_n/n) - f(x)|] \\
&= \mathbb{E}[|f(S_n/n) - f(x)|; |S_n/n - x| \leq \delta] + \mathbb{E}[|f(S_n/n) - f(x)|; |S_n/n - x| > \delta] \\
&< \epsilon + \mathbb{E}[|f(S_n/n) - f(x)|; |S_n/n - x| > \delta] \\
&\leq \epsilon + 2\|f\|_\infty \mathbb{P}(|S_n/n - x| > \delta) \leq \epsilon + 2\|f\|_\infty \frac{\text{Var}(S_n)}{n^2\delta^2} \\
&= \epsilon + 2\|f\|_\infty \frac{x(1-x)}{n^2\delta^2} \leq \epsilon + 2\|f\|_\infty \frac{(1/4)}{n^2\delta^2} \longrightarrow \epsilon \quad (n \longrightarrow \infty)
\end{aligned}$$

(the first inequality is Jensen's, the strict inequality is due to the uniform continuity of f (continuous on a compact set). From there it's Chebyshev's inequality and noting that $\max\{x(1-x)\} = 1/4$). $\square$

Theorem 37. (Levy's Equivalence) Let $\{X_n\}$ be independent r.v with $S_n := \sum_{j=1}^n X_j$. Then

$$S_n \xrightarrow{p} S \text{ (for some r.v } S) \iff S_n \longrightarrow S \text{ a.s} \iff \text{“}S_n \longrightarrow S \text{ in distribution”}.$$

Chapter 8

More Measure Theory Results

8.1 Return to General Measure Space

Let (S_1, Σ_1) and (S_2, Σ_2) be two measure spaces. Let

$$S := S_1 \times S_2 = \{(s_1, s_2) \mid s_1 \in S_1, s_2 \in S_2\}, \quad \Sigma := \Sigma_1 \times \Sigma_2 = \sigma(\{B_1 \times B_2 \mid B_i \in \Sigma_i\});$$

then (S, Σ) is again a measure space.

Lemma 9. Let $f : S \longrightarrow \mathbb{R}$ is measurable ($f \in m\Sigma$), then:

1. For all $s_1 \in S_1$, the function $s_2 \in S_2 \longmapsto f(s_1, s_2) \in \mathbb{R}$ is measurable w.r.t Σ_2 ($f(s_1, \cdot) \in m\Sigma_2$),
2. For all $s_2 \in S_2$, the function $s_1 \in S_1 \longmapsto f(s_1, s_2) \in \mathbb{R}$ is measurable w.r.t Σ_1 ($f(\cdot, s_2) \in m\Sigma_1$).

Assume μ_i is a finite measure on (S_i, Σ_i). For any $f \in (m\Sigma)^+ \cup (b\Sigma)$, we can define the following:

$$\forall s_1 \in S_1,\ I_1^f(s_1) := \int_{S_2} f(s_1, s_2)\mu_2(ds_2), \quad \forall s_2 \in S_2,\ I_2^f(s_2) := \int_{S_1} f(s_1, s_2)\mu_1(ds_1).$$

Lemma 10. If $f \in b\Sigma$, then $I_i^f(\cdot) \in b\Sigma_i$ (for $i = 1, 2$) and

$$\int_{S_1} I_1^f(s_1)\mu_1(ds_1) = \int_{S_2} I_2^f(s_2)\mu_2(ds_2).$$

Corollary 6. If $f \in (m\Sigma)^+$, then $I_i^f(\cdot) \in (m\Sigma)^+$ (for $i = 1, 2$) and

$$\int_{S_1} I_1^f(s_1)\mu_1(ds_1) = \int_{S_2} I_2^f(s_2)\mu_2(ds_2).$$

Theorem 38. (Fubini's Theorem) Let (S_i, Σ_i, μ_i) as above. Define $\mu : \Sigma \longrightarrow [0, \infty)$ such that:

$$\forall A \in \Sigma, \mu(A) = \int_{S_1} I_1^{\mathbb{1}_A}(s_1)\mu(ds_1) = \int_{S_2} I_2^{\mathbb{1}_A}(s_2)\mu(ds_2).$$

Then:

1. μ is a finite measure on (S, Σ). 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, Σ) such that $\mu(B_1 \times B_2) = \mu(B_1)\mu(B_2)$,

3. if $f \in (m\Sigma)^+$ then $\mu(f) = \int_{S_1} I_1^f(s_1)\mu_1(ds_1) = \int_{S_2} I_2^f(s_2)\mu_2(ds_2)$ i.e.

$$\begin{aligned}\iint_{S_1\times S_2} f(s_1,s_2)d\mu(s_1,s_2) &= \int_{S_2}\left(\int_{S_1} f(s_1,s_2)\mu_1(ds_1)\right)\mu_2(ds_2)\\ &\overset{\dagger}{=} \int_{S_1}\left(\int_{S_2} f(s_1,s_2)\mu_2(ds_2)\right)\mu_1(ds_1),\end{aligned}$$

4. $f \in L^1(S,\Sigma,\mu) \implies I_i^f \in L^1(S_i,\Sigma_i,\mu_i)$ and (†) holds.

8.2 Return to Probability Space

Definition 48. Given X, Y r.v on $(\Omega,\mathcal{F},\mathbb{P})$, then $F_{(X,Y)}$ is the **joint distribution**:

$$F_{(X,Y)} : (x,y) \in \mathbb{R}^2 \longmapsto F_{(X,Y)}(x,y) = \mathbb{P}(X \le x, Y \le y).$$

Definition 49. The **joint distribution** or **joint law** of (X,Y), denoted $\mathcal{L}_{(X,Y)}$ is the probability measure on $(\mathbb{R}^2,\mathcal{B}^2)$.

Definition 50. If $\mathcal{L}_{(X,Y)} \ll \lambda^2_{\text{Leb}}(:= dxdy)$, then $f_{(X,Y)} := \dfrac{d\mathcal{L}_{(X,Y)}}{dxdy}$ is the **joint pdf**.

Fact 9. By Fubini, if (X,Y) has the joint pdf $f_{(X,Y)}$, then X and Y have the following pdfs

$$f_X(x) := \int_{\mathbb{R}} f_{(X,Y)}(x,y)\ dy, \quad f_Y(y) := \int_{\mathbb{R}} f_{(X,Y)}(x,y)\ dx.$$

Fact 10. If X, Y have distribution laws $\mathcal{L}_X, \mathcal{L}_Y$ (resp.) and distribution functions F_X, F_Y (resp.), TFAE:

1. $X \perp\!\!\!\perp Y$
2. $F_{(X,Y)}(x,y) = F_X(x)F_Y(y)$
3. $\mathcal{L}_{(X,Y)} = \mathcal{L}_X \times \mathcal{L}_Y$
4. If X and Y have pdfs, $f_{(X,Y)}(x,y) = f_X(x)f_Y(y)$.

Lemma 11. Assume $X \perp\!\!\!\perp Y$, then for all $c \in \mathbb{R}$: $\mathbb{P}(X+Y\le c) = (\mathcal{L}_X \star \mathcal{L}_Y)((-\infty,c])$

Proof.

$$\begin{aligned}\mathbb{P}(X+Y\le c) &= \iint_{\{x+y\le c\}} \mathcal{L}_{(X,Y)}\ (dxdy) = \iint_{\{x+y\le c\}} \mathcal{L}_X(dx)\mathcal{L}_Y(dy)\\ &= \int_{\mathbb{R}}\int_{\{y\le c-x\}} \mathcal{L}_Y(dy)\mathcal{L}_X(dx) = \int_{\mathbb{R}} \mathcal{L}_Y((-\infty,c-x))\mathcal{L}_X(dx) = (\mathcal{L}_X \star \mathcal{L}_Y)((-\infty,c]).\end{aligned}$$

□

8.2.1 Cylinder Sets

Chapter 9

Conditioning

Definition 51. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with $X \in L^1$ and $\mathcal{G} \subseteq \mathcal{F}$ (a sub σ-algebra). Then $Y \in L^1$ is called the **conditional expectation** of X given $\mathcal{G}$, denoted $Y = \mathbb{E}[X|\mathcal{G}]$ if:

1. $Y \in m\mathcal{G}$,
2. $\forall A \in \mathcal{G}$, $\int_A X \; d\mathbb{P} = \int_A Y \; d\mathbb{P}$.

Remark 21. (2) in the above definition can be replaced with:

$$(2') : \forall A \in I, \text{ such that } \sigma(I) = \mathcal{G}, \int_A X \; d\mathbb{P} = \int_A Y \; d\mathbb{P},$$

where I is a π-system.

Proposition 25. Let $X_1, X_2 \in L^1$ and $\mathcal{G} \subseteq \mathcal{F}$. If $X_1 \leq X_2$ almost surely, then

$$\mathbb{E}[X_1|\mathcal{G}] \leq \mathbb{E}[X_2|\mathcal{G}] \text{ a.s.}$$

Proof. Let $Y_1 := \mathbb{E}[X_1|\mathcal{G}]$, $Y_2 := \mathbb{E}[X_2|\mathcal{G}]$ and $A := \{Y_1 > Y_2\} \in \mathcal{G}$.

$$\int_A Y_1 \; d\mathbb{P} = \int_A X_1 \; d\mathbb{P} \leq \int_A X_2 \; d\mathbb{P} = \int_A Y_2 \; d\mathbb{P} \implies \int_A (Y_1 - Y_2) \; d\mathbb{P} \leq 0 \implies \mathbb{P}(A) = 0.$$

□

Theorem 39. Suppose $X \in L^1$ and $\mathcal{G} \subseteq \mathcal{F}$. Then $\mathbb{E}[X|\mathcal{G}]$ always exists and is unique.

Example 12. If $\mathcal{G} = \sigma(A)$, for some $A \in \mathcal{F}$ [i.e. $\mathcal{G} = \{\emptyset, \Omega, A, A^C\}$]. Given $X \in L^1$, we have

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{1}_A \frac{\mathbb{E}[X; A]}{\mathbb{P}(A)} + \mathbb{1}_{A^c} \frac{\mathbb{E}[X; A]}{\mathbb{P}(A^C)}.$$

In particular, if $X = \mathbb{1}_E$, for some $E \in \mathcal{F}$, then:

$$\mathbb{E}[\mathbb{1}_E|\sigma(A)] = \mathbb{1}_A \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(A)} + \mathbb{1}_{A^C} \frac{\mathbb{P}(E \cap A^C)}{\mathbb{P}(A^C)} = \mathbb{1}_A \mathbb{P}(E|A) + \mathbb{1}_{A^C} \mathbb{P}(E|A^C).$$

Example 13. If $\mathcal{G} = m\sigma(Y)$ for some r.v $Y \in m\mathcal{F}$, we write $\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$.

Definition 52. If (X, Y) has a joint pdf $f_{(X,Y)}$, then define the **conditional pdf**, $f_{(X|Y)}$ as follows:

$$f_{(X|Y)}(x|y) := \begin{cases} \dfrac{f_{(X,Y)}(x,y)}{f_Y(y)} & \text{if } f_Y(y) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

with $f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)\ dx$.

Proposition 26. Let $h : \mathbb{R} \longrightarrow \mathbb{R}$ be Borel such that $h(X) \in L^1$. Define $g : \mathbb{R} \longrightarrow \mathbb{R}$ as follows

$$g(y) = \int_{\mathbb{R}} h(x) f_{(X|Y)}(x|y)\ dx.$$

(Note by Fubini, g is Borel). Then: $\mathbb{E}[h(X)|Y] = g(Y)$ almost surely.

Proof. Since $g(Y) \in m\sigma(Y)$, we have that for all $A \in \sigma(Y)$, there exists $B \in \mathcal{B}$ such that $A = \{Y \in B\}$ and $\int_{\{Y\in B\}} g(Y)\ d\mathbb{P}$ is defined. To this end:

$$\begin{aligned}
\int_{\{Y\in B\}} g(Y)\ d\mathbb{P} &= \int_B g(y) f_Y(y)\ dy = \int_B \left(\int_{\mathbb{R}} h(x) f_{(X|Y)}(x|y)\ dx \right) f_Y(y)\ dy \\
&= \int_B \int_{\mathbb{R}} h(x) f_{(X|Y)}(x|y) f_Y(y)\ dxdy = \int_B \int_{\mathbb{R}} h(x) f_{(X,Y)}(x,y)\ dxdy \\
&= \iint_{\mathbb{R}\times\mathbb{R}} h(x) \mathbb{1}_B(y) f_{(X,Y)}(x,y)\ dxdy = \mathbb{E}[h(X)\mathbb{1}_B(Y)] = \int_{\{Y\in B\}} h(X)\ d\mathbb{P}.
\end{aligned}$$

□

Proposition 27.

1. If $Y = \mathbb{E}[X|\mathcal{G}]$, then $\mathbb{E}[X] = \mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]]$
2. If $X \in m\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = X$
3. $\mathbb{E}[aX_1 + bX_2|\mathcal{G}] = a\mathbb{E}[X_1|\mathcal{G}] + b\mathbb{E}[X_2|\mathcal{G}]$
4. **(Monotonicity)** $X_1 \leq X_2$, then $\mathbb{E}[X_1|\mathcal{G}] \leq \mathbb{E}[X_2|\mathcal{G}]$
5. **(cMON)** If $X_n \uparrow X$, then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$
6. **(cFatou)** If $X_n \geq 0$ and $X_n \in L^1$: $\mathbb{E}[\liminf X_n|\mathcal{G}] \leq \liminf \mathbb{E}[X_n|\mathcal{G}]$
7. **(cDOM)** If $|X_n| \leq Y$ for some $Y \in L^1$ and $X_n \xrightarrow{p} X$: $\lim \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]$
8. **(cJensen)** Let φ be convex, $X \in L^1$ and $\varphi(X) \in L^1$. Then: $\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}]$
9. **(Tower Property)** If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathbb{F}$: $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$
10. Let $Z \in m\mathcal{G}$ and $XZ \in L^1$. Then $\mathbb{E}[XZ|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$
11. If $\mathcal{H} \subseteq \mathcal{F}$ and $\mathcal{H} \perp\!\!\!\perp \sigma(\sigma(X) \cup \mathcal{G})$, then $\mathbb{E}[X|\sigma(\mathcal{G} \cup \mathcal{H})] = \mathbb{E}[X|\mathcal{G}]$

12. Suppose $X_1, \dots, X_N$ be independent with distributions $\mathcal{L}_1, \dots, \mathcal{L}_N$. Suppose $h : \mathbb{R}^N \to \mathbb{R}$ is Borel and $h(X_1, \dots, X_N) \in L^1$. Define:

$$x_1 \in \mathbb{R} \longmapsto \gamma^h(x_1) := \mathbb{E}[h(x_1, X_2, \dots, X_N)] = \int_\Omega h(x_1, X_2, \dots, X_N)\ d\mathbb{P}.$$

Then: $\mathbb{E}[h(X_1, \dots, X_N)|X_1] = \gamma^h(X_1)$.

Proof.

(5): Let $\mathbb{E}[X_n|\mathcal{G}] \uparrow Y := \lim_n \mathbb{E}[X_n|\mathcal{G}]$. For any $A \in \mathcal{G}$:

$$\int_A Y\ d\mathbb{P} \overset{\text{(MON)}}{=} \lim_n \int_A \mathbb{E}[X_n|\mathcal{G}]\ d\mathbb{P} = \lim_n \int_A X_n\ d\mathbb{P} = \int_A X\ d\mathbb{P}.$$

(6): $\mathbb{E}[\liminf_n X_n|\mathcal{G}] \overset{\text{(cMON)}}{=} \lim_m \mathbb{E}[\inf_{n\geq m} X_n|\mathcal{G}] \leq \lim_m \inf_{n\geq m} \mathbb{E}[X_n|\mathcal{G}] = \liminf_n \mathbb{E}[X_n|\mathcal{G}]$.

(7): By (DOM'), $X_n \xrightarrow{L^1} X$ then for all $A \in \mathcal{G}$: $\int_A X_n\ d\mathbb{P} \longrightarrow \int_A X\ d\mathbb{P}$. Now define $Y_n := Y - X_n$:

$$\begin{aligned}
&\mathbb{E}[\liminf Y_n|\mathcal{G}] \overset{\text{(cFatou)}}{\leq} \liminf \mathbb{E}[Y_n|\mathcal{G}] \\
&\implies \mathbb{E}[Y - X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}] - \limsup \mathbb{E}[X_n|\mathcal{G}] \\
&\implies \mathbb{E}[X|\mathcal{G}] \geq \limsup \mathbb{E}[X_n|\mathcal{G}].
\end{aligned}$$

Similarly, using $Z_n := Y + X_n$, we get $\mathbb{E}[X|\mathcal{G}] \leq \liminf \mathbb{E}[X_n|\mathcal{G}]$.

(9): The first equality is trivial because $\mathbb{E}[X|\mathcal{H}] \in m\mathcal{H} \subseteq m\mathcal{G}$ via (2). For the second equality: take any $A \in \mathcal{H}$

$$\int_A \mathbb{E}[X|\mathcal{G}]\ d\mathbb{P} = \int_A X\ d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{H}]\ d\mathbb{P} \implies \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}].$$

(10): First suppose $X \geq 0$ and let $Z = \mathbb{1}_A$ (for $A \in \mathcal{G}$). For all $B \in \mathcal{G}$:

$$\int_B XZ\ d\mathbb{P} = \int_{A\cap B} X\ d\mathbb{P} = \int_{A\cap B} \mathbb{E}[X|\mathcal{G}]\ d\mathbb{P} = \int_B Z\mathbb{E}[X|\mathcal{G}]\ d\mathbb{P}.$$

Thus the statement holds for $X \geq 0$ and $Z = \mathbb{1}_A$. By linearity, it holds true for all $Z \in \text{SF}^+$. By (cMON), it holds true for all $X \geq 0$ and $Z \geq 0$ ($X \in m\mathcal{F}^+$ and $Z \in m\mathcal{G}^+$). Finally, since $XZ \in L^1$, then $|XZ| = (X^+ - X^-)(Z^+ - Z^-)$; the statements holds for general X and Z.

(11): Let $I := \{G \cap H \mid G \in \mathcal{G}, H \in \mathcal{H}\}$. For all $A \in I$:

$$\int_A X\ d\mathbb{P} = \int_\Omega \mathbb{1}_G \mathbb{1}_H X\ d\mathbb{P} = \mathbb{P}(H) \int_G X\ d\mathbb{P} = \mathbb{P}(H) \int_G \mathbb{E}[X|\mathcal{G}]\ d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}]\ d\mathbb{P}.$$

Since I is a π-system, $\sigma(I) = \sigma(\mathcal{G} \cup \mathcal{H})$ and we're done.

(12): For all $A \in \sigma(X_1)$, there exists a $B \in \mathcal{B}$ such that $A \in \{X_1 \in B\}$. Then:

$$\begin{aligned}
\int_A h(X_1, \dots, X_N)\ d\mathbb{P} &= \int \cdots \int_{B\times\mathbb{R}\times\cdots\times\mathbb{R}} h(x_1, \dots, x_N) \mathcal{L}_{(X_1,\dots,X_N)}\ (dx_1, \dots, dx_N) \\
&= \int_B \left(\int \cdots \int_{\mathbb{R}\times\cdots\times\mathbb{R}} h(x_1, \dots, x_N) \mathcal{L}_{X_2}(dx_2) \times \cdots \times \mathcal{L}_{X_N}(dx_N) \right) \mathcal{L}_{X_1}(dx_1) \\
&\overset{\text{(Fubini)}}{=} \int_B \gamma^h(x_1) \mathcal{L}_{X_1}(dx_1) = \int_A \gamma^h(X_1)\ d\mathbb{P}.
\end{aligned}$$

□

Chapter 10

Martingale Theory

10.1 Introduction

Definition 53. Given $(\Omega, \mathcal{F}, \mathbb{P})$, $\{\mathcal{F}_n\}_{n=0}^{\infty}$ is called a **filtration** if $\mathcal{F}_n \subseteq \mathcal{F}$ and $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ $(\forall n \geq 0)$.

Definition 54. $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geq 0}, \mathbb{P})$ is called a **filtered space**.

Remark 22. Let $\mathcal{F}_\infty := \sigma\left(\bigcup_{n\geq 0} \mathcal{F}_n\right)$.

Definition 55. A sequence of r.v $\{X_n\}_{n\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is called an **adaptive process** if $X_n \in m\mathcal{F}_n$.

Remark 23. A sequence $\{X_n\}_{n\geq 0}$ is always adaptive w.r.t its **natural filtration**: $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$.

Definition 56. Given a filtered space, a process $\{X_n\}_{n\geq 0}$ is called a **martingal** w.r.t $\{\mathcal{F}_n\}_{n\geq 0}$ if:

1. $\{X_n\}_{n\geq 0}$ is adaptive,
2. $X_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$,
3. $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$.

Definition 57. The process is called a **submartingale** when $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$.

Definition 58. The process is called a **supermartingale** when $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$.

Remark 24. If $\{X_n\}_{n\geq 0}$ is a submartingale, then for all $m > n \geq 0$:

$$\mathbb{E}[X_m|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X_m|\mathcal{F}_{m-1}]|\mathcal{F}_n] \geq \mathbb{E}[X_{m-1}|\mathcal{F}_n] \geq \cdots \geq X_n$$

(we get all equalities in the case of a martingale)

Remark 25. If $\{X_n\}_{n\geq 0}$ is a (sub)martingale, then $\{X_n - X_0\}_{n\geq 0}$ is a (sub)martingale. WLOG, we can assume $X_0 \equiv 0$.

Example 14. Let $\{X_n\}$ be a sequence of independent r.v such that $\mathbb{E}[X_n] = 0$, $S_0 \equiv 0$, and $S_n = \sum_{j=1}^{n} S_j$ $(\forall n \geq 1)$. We show that $\{S_n\}$ is a martingale with respect (to the natural filtration) $\mathcal{F}_n := \sigma(S_0, S_1, \ldots, S_n)$. The first two conditions are obviously satisfied. For the third:

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \mathbb{E}[S_n + X_{n+1}|\mathcal{F}_n] = S_n + \mathbb{E}[X_{n+1}] = S_n.$$

Example 15. Let $\{X_n\}$ be a sequence of independent r.v such that $\mathbb{E}[X_n] = 0$, $\mathbb{E}[X_n^2] = 1$, $S_0 \equiv 0$, and $S_n = \sum_{j=1}^n S_j$ $(\forall n \geq 1)$. We show that $\{S_n^2 - n\}$ is a martingale with respect (to the natural filtration). As before, it only remains to show the third condition:

$$\begin{aligned}\mathbb{E}[S_{n+1}^2 - (n+1)|\mathcal{F}_n] &= \mathbb{E}[(S_n + X_{n+1})^2 - (n+1)|\mathcal{F}_n] \\ &= -n + 1 + \mathbb{E}[S_n^2 + 2S_nX_{n+1} + X_{n+1}^2|\mathcal{F}_n] \\ &= -n + 1 + S_n^2 + 2S_n\mathbb{E}[X_{n+1}] + \mathbb{E}[X_{n+1}^2] = S_n^2 - n.\end{aligned}$$

Example 16. If $\{X_n\}$ are idd standard normal r.v, then for all $\lambda \in \mathbb{R}$: $\{\exp[\lambda S_n - \lambda^2 n/2] : n \geq 0\}$ is a martingale (w.r.t the natural filtration).

Example 17. Take $\{X_n\}$ independent such that $\mathbb{E}[X_n] = 1$ $(\forall n \geq 1)$. Set $T_0 \equiv 1$ and $T_n = \prod_{j=1}^n X_j$. We show that T_n is a martingale w.r.t the natural filtration: $\mathcal{F}_n = \sigma(T_0, \ldots, T_n) = \sigma(X_1, \ldots, X_n)$. We check the third condition:

$$\mathbb{E}[T_{n+1}|\mathcal{F}_n] = \mathbb{E}[T_nX_{n+1}|\mathcal{F}_n] = T_n\mathbb{E}[X_{n+1}] = T_n.$$

Example 18. Take $X \in L^1$ and $\{\mathcal{F}_n\}$ to be any filtration. The sequence of r.v $X_n := \mathbb{E}[X|\mathcal{F}_n]$ is a martingale w.r.t $\{\mathcal{F}_n\}$:

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[X|\mathcal{F}_n] = X_n.$$

Theorem 40. Consider a filtered space with process $\{X_n\}_{n\geq 0}$. Further suppose $f : \mathbb{R} \to \mathbb{R}$ is convex with $f(X_n) \in L^1$. If *either* of the following hold:

1. $\{X_n\}$ is a martingale

2. $\{X_n\}$ is a submartingale and f is non-decreasing

then $\{f(X_n)\}$ is a submartingale.

Proof. We only check the third condition for a submartingale. Under assumption (2) and (cJensen):

$$\mathbb{E}[f(X_{n+1})|\mathcal{F}_n] \geq f(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) \geq f(X_n).$$

The proof is the same with assumption (1) but with equality at the last step. □

Theorem 41. (Doob's Decomposition Theorem) Given a submartingale $\{X_n\}_{n\geq 0}$ in a filtered space, there exists a unique process $\{Y_n\}_{n\geq 0}$ such that

1. $Y_0 \equiv 0, Y_n \in L^1$, and $Y_n \in m\mathcal{F}_{n-1}$ $(\forall n \geq 1)$ (called **previsible**)

2. $\{Y_n\}$ is non-decreasing i.e. $Y_n \leq Y_{n+1}$ a.s

3. $\{M_n := X_n - Y_n\}_{n\geq 0}$ is a martingale.

Proof. We first show uniqueness: Suppose $\{Z_n\}$ is another process satisfying the above criteria. Let $\Delta_n := Z_n - Y_n$; $\Delta_0 \equiv 0$ and $\Delta_n \in m\mathcal{F}_{n-1}$. Note also that we can write $\Delta_n = (X_n - Y_n) - (X_n - Z_n)$. Using the previsible property:

$$\begin{aligned}\Delta_{n+1} = \mathbb{E}[\Delta_{n+1}|\mathcal{F}_n] &= \mathbb{E}[(X_{n+1} - Y_{n+1}) - (X_{n+1} - Z_{n+1})|\mathcal{F}_n] \\ &= \mathbb{E}[(X_{n+1} - Y_{n+1})|\mathcal{F}_n] - \mathbb{E}[(X_{n+1} - Z_{n+1})|\mathcal{F}_n] \\ &= (X_n - Y_n) - (X_n - Z_n) = \Delta_n \\ &\implies \Delta_n \equiv 0 \ (\forall n \geq 0) \implies Z_n \equiv Y_n.\end{aligned}$$

For existence: Define $Y_0 \equiv 0$ and for all $n \geq 0$: $Y_{n+1} := \sum_{j=0}^{n} (\mathbb{E}[X_{j+1}|\mathcal{F}_j] - X_j) \in m\mathcal{F}_n$. Note that $Y_n \in L^1$ and non-decreasing (because $\{X_n\}$ is a submartingale),

$$Y_{n+1} = Y_n + (\mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n) \geq Y_n.$$

For the martingale property:

$$\mathbb{E}[X_{n+1} - Y_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_{n+1}|\mathcal{F}_n] - Y_{n+1} = X_n - \sum_{j=0}^{n=1} (\mathbb{E}[X_{j+1}|\mathcal{F}_j] - X_j) = X_n - Y_n.$$

□

10.2 Stopping Time

Definition 59. In a filtered space, a r.v $\tau : \Omega \longmapsto \{0, 1, 2, \ldots\} \cup \{\infty\}$ is called a **stopping time** (w.r.t the filtration $\{\mathcal{F}_n\}$) if, for all $n \in [0, \infty]$:

$$\{\tau \leq n\} \in \mathcal{F}_n \quad \text{or equivalently} \quad \{\tau = n\} \in \mathcal{F}_n.$$

Example 19. Consider $\{X_n\}$ adapted w.r.t $\{\mathcal{F}_n\}$ and $a, b \in \mathbb{R}$ with $a < b$. Then the following random variables are stopping times:

$$\tau_1 := \inf\{n \geq 0 \mid X_n \leq a\},$$
$$\tau_2 := \inf\{n \geq \tau_1 \mid X_n \geq b\}.$$

Fact 11.

1. If τ is a stopping time, then $(\forall n \geq 0)$: $\{\tau > n\}, \{\tau \leq n\}, \{\tau < n\} \in \mathcal{F}_n$
2. Let τ be a stopping time and $N \in \mathbb{N}$. Then $\tau \wedge N := \min(\tau, N)$ is a stopping time
3. Let τ and π be two stopping times. Then $\tau \wedge \pi$, $\tau \vee \pi$, and $\tau + \pi$ are stopping times.

Definition 60. Suppose τ is a stopping time on a filtered space. We define the following σ-algebra for the stopping time:

$$\begin{aligned}\mathcal{F}_\tau &:= \{A \in \mathcal{F} \mid \{\tau \leq n\} \cap A \in \mathcal{F}_n, \forall\ 0 \leq n \leq \infty\} \\ &= \{A \in \mathcal{F} \mid \{\tau = n\} \cap A \in \mathcal{F}_n, \forall\ 0 \leq n \leq \infty\}.\end{aligned}$$

Proposition 28.

1. $\tau \in m\mathcal{F}_\tau$
2. If τ and π are two stopping times, $\tau \leq \pi$: $\mathcal{F}_\tau \subseteq \mathcal{F}_\pi$
3. For any two stopping times τ and π: $\mathcal{F}_{\tau \wedge \pi} = \mathcal{F}_\tau \cap \mathcal{F}_\pi$.

Proof.
(2): Take $A \in \mathcal{F}_\tau$. Then for all $0 \leq n \leq \infty$,

$$A \cap \{\pi \leq n\} = A \cap \{\tau \leq n\} \cap \{\pi \leq \infty\} \in \mathcal{F}_n \implies A \in \mathcal{F}_\pi.$$

□

Definition 61. Let $\{X_n\}$ be adapted w.r.t $\{\mathcal{F}_n\}$ and let τ be a stopping time.

(I) If $\tau < \infty$ a.s, then $X_\tau(\omega) = \begin{cases} X_{\tau(\omega)}(\omega) & \text{if } \tau(\omega) < \infty, \\ \text{any constant} & \text{if } \tau(\omega) = \infty. \end{cases}$

(II) If $\mathbb{P}(\tau = \infty) > 0$ and $X_\infty := \lim_n X_n$ exists a.s, then $X_\tau(\omega) = \begin{cases} X_{\tau(\omega)}(\omega) & \text{if } \tau(\omega) < \infty, \\ X_\infty(\omega) & \text{if } \tau(\omega) = \infty. \end{cases}$

(III) If neither of the above hold, X_τ is not defined on $\{\tau = \infty\}$.

Proposition 29. If X_τ is defined, $X_\tau \in m\mathcal{F}_\tau$.

10.3 Convergence (and other) theorems

Theorem 42. (Doob's Stopping Time Theorem) Consider a filtered space with stopping time τ. If $\{X_n\}$ is a (sub)martingale, then $\{X_{\tau\wedge n}\}$ is a (sub)martingale. As an immediate consequence:

$$\forall\ 0 \le n \le \infty, \mathbb{E}[X_{\tau\wedge n}] \ge \mathbb{E}[X_0] \quad \text{(equality with martingale)}.$$

Proof. We go through all three conditions for (sub)martingales.

1. Adapted: $X_{n\wedge\tau} \in m\mathcal{F}_{\tau\wedge n} \subseteq m\mathcal{F}_n$

2. $\{X_{\tau\wedge n}\} \in L^1$:

$$|X_{n\wedge\tau}| = \sum_{j=0}^{n} |X_{n\wedge\tau}| \mathbb{1}_{\{\tau=j\}} + \mathbb{1}_{\{\tau\ge n+1\}} |X_{n\wedge\tau}| = \sum_{j=0}^{n} |X_j| \mathbb{1}_{\{\tau=j\}} + \mathbb{1}_{\{\tau\ge n+1\}} |X_{n\wedge\tau}| \in L^1.$$

3. For all $0 \le n \le \infty$, For all $A \in \mathcal{F}_n$:

$$\begin{aligned} \int_A X_{(n+1)\wedge\tau}\ d\mathbb{P} &= \int_{A\cap\{\tau\le n\}} X_\tau\ d\mathbb{P} + \int_{A\cap\{\tau>n\}} X_{n+1}\ d\mathbb{P} \\ &\ge \int_{A\cap\{\tau\le n\}} X_\tau\ d\mathbb{P} + \int_{A\cap\{\tau>n\}} X_n\ d\mathbb{P} \quad \text{(equality if martingale)} \\ &= \int_A X_{n\wedge\tau}\ d\mathbb{P}. \end{aligned}$$

□

Theorem 43. (Hunt's Theorem I) Consider a filtered space with $\{X_n\}$ as a (sub)martingale, $\tau_1 \le \tau_2$ are two stopping times, and there exists $T \in \mathbb{N}$ such that $\tau_1 \le \tau_2 \le T$. Then: $X_{\tau_1}, X_{\tau_2} \in L^1$ and

$$\mathbb{E}[X_{\tau_2}|\mathcal{F}_{\tau_1}] \ge X_{\tau_1} \text{ a.s} \quad \text{(equality if martingale)}.$$

In particular, $\mathbb{E}[X_{\tau_2}] \ge \mathbb{E}[X_{\tau_1}]$ (equality if martingale).

Proof. Because $\{X_n\}$ is a (sub)martingale, for $i = 1, 2$ we have:

$$|X_{\tau_i}| = \sum_{j=0}^{T} |X_j| \mathbb{1}_{\tau_i = j} \in L^1.$$

Now suppose $\{X_n\}$ is a submartingale. Using Doob's decomposition, $X_n = Y_n + M_n$, with Y_n being previsible and non-decreasing and M_n is a martingale. To this end:

$$\mathbb{E}[X_{\tau_2}|\mathcal{F}_{\tau_1}] = \mathbb{E}[Y_{\tau_2}|\mathcal{F}_{\tau_1}] + \mathbb{E}[M_{\tau_2}|\mathcal{F}_{\tau_1}] \geq \mathbb{E}[Y_{\tau_1}|\mathcal{F}_{\tau_1}] + \mathbb{E}[M_{\tau_2}|\mathcal{F}_{\tau_1}] = Y_{\tau_1} + M_{\tau_1} = X_{\tau_1}.$$

If $\{X_n\}$ is a martingale, take $A \in \mathcal{F}_{\tau_1}$:

$$\begin{aligned}
\int_A X_{\tau_2} \, d\mathbb{P} &= \sum_{j=0}^{T} \int_{A \cap \{\tau_1 = j\}} X_{\tau_2} \, d\mathbb{P} \\
&= \sum_{j=0}^{T} \sum_{k=j}^{T} \int_{A \cap \{\tau_1 = j\} \cap \{\tau_2 = k\}} X_k \, d\mathbb{P} \\
&= \sum_{j=0}^{T} \sum_{k=j}^{T} \int_{A \cap \{\tau_1 = j\} \cap \{\tau_2 = k\}} X_T \, d\mathbb{P} = \int_A X_T \, d\mathbb{P}.
\end{aligned}$$

Similarly,

$$\int_A X_{\tau_1} \, d\mathbb{P} = \sum_{j=0}^{T} \int_{A \cap \{\tau_1 = j\}} X_j \, d\mathbb{P} = \sum_{j=0}^{T} \int_{A \cap \{\tau_1 = j\}} X_T \, d\mathbb{P} = \int_A X_T \, d\mathbb{P},$$

which completes the claim. $\square$

Theorem 44. (Doob's Upcrossing Inequality) Consider a submartingale $\{X_n\}$ and $a, b \in \mathbb{R}$ such that $a < b$. Define the sequence of stopping times:

$$\begin{aligned}
\tau_1 &= \inf\{n \geq 0 \mid X_n \leq a\} \\
\tau_2 &= \inf\{n \geq \tau_1 \mid X_n \geq b\} \\
&\vdots \\
\tau_{2k-1} &= \inf\{n \geq \tau_{2k-1} \mid X_n \leq a\} \\
\tau_{2k} &= \inf\{n \geq \tau_{2k-1} \mid X_n \geq a\},
\end{aligned}$$

where the last two above denote the "k-th uprising". Furthermore, for every $n \geq 1$,

$$U_{(a,b)}^{(n)} := \max\{k \geq 1 \mid \tau_{2k} \leq n\} = \text{ number of completed upcrossings,}$$

and let

$$U_{(a,b)} := \lim_{n \to \infty} U_{(a,b)}^{(n)} = \text{ total number of upcrossings.}$$

Then:

$$\mathbb{E}[U_{(a,b)}^{(n)}] \leq \frac{\mathbb{E}[(X_n - a)^+]}{b - a} \quad (\forall \ n \geq 1).$$

In particular, if $\sup \mathbb{E}[X_n^+] < \infty$, then $U_{(a,b)} < \infty$ a.s.

Proof. WLOG, we can assume $\tau_1 < n$ (otherwise $U^{(n)}_{(a,b)} = 0$). Define $Y_n := (X_n - a)^+$, which is a non-negative submartingale with $Y_{\tau_{2k-1}} = 0$ and $Y_{\tau_{2k}} \geq b - a$. This implies that

$$1 \leq \frac{Y_{\tau_{2k}} - Y_{\tau_{2k-1}}}{b-a}.$$

To this end, letting n be sufficiently large such that the rest of the terms are zero after:

$$\begin{aligned} U^{(n)}_{(a,b)} = \sum_{k=1}^{U^{(n)}_{(a,b)}} 1 &\leq \sum_{k=1}^{U^{(n)}_{(a,b)}} \frac{Y_{\tau_{2k}} - Y_{\tau_{2k-1}}}{b-a} \\ &\leq \sum_{k=1}^{n} \frac{Y_{\tau_{2k}\wedge n} - Y_{\tau_{2k-1}\wedge n}}{b-a} \\ &= (b-a)^{-1}\left[Y_{\tau_{2k}\wedge n} - \sum_{k=2}^{n-1}\left(Y_{\tau_{2k-1}\wedge n} - Y_{\tau_{2k-2}\wedge n}\right) + Y_{\tau_1 \wedge n}\right] \\ &= (b-a)^{-1}\left[Y_{\tau_{2k}\wedge n} - \sum_{k=2}^{n-1}\left(Y_n - Y_{\tau_{2k-2}\wedge n}\right) + 0\right]. \end{aligned}$$

Taking expectations, we get:

$$\mathbb{E}[U^{(n)}_{(a,b)}] \leq (b-a)^{-1}\mathbb{E}[Y_n] - (b-a)^{-1}\sum_{k=2}^{n-1}\left(\mathbb{E}[Y_{\tau_{2k-1}\wedge n}] - \mathbb{E}[Y_{\tau_{2k-1}\wedge n}]\right).$$

Since $\{Y_n\}$ is a submartingale and $\tau_{2k-2} \wedge n \leq \tau_{2k-1} \wedge n \leq n$ are "two bounded stopping times", we can apply (HUNT1), which says:

$$\mathbb{E}[Y_{\tau_{2k-1}\wedge n}] \geq \mathbb{E}[Y_{\tau_{2k-2}\wedge n}] \implies \mathbb{E}[U^{(n)}_{(a,b)}] \leq \frac{\mathbb{E}[Y_n]}{b-a} = \frac{\mathbb{E}[(X_n - a)^+]}{b-a}.$$

If $\sup \mathbb{E}[X_n^+] < \infty$, then

$$\sup_n \mathbb{E}[U^{(n)}_{(a,b)}] \leq \frac{\sup \mathbb{E}[X_n^+] + |a|}{b-a} < \infty \overset{\text{(MON)}}{\implies} \mathbb{E}[U_{(a,b)}] < \infty,$$

so $U_{(a,b)}$ is finite almost surely. $\square$

Theorem 45. (Martingale Convergence Theorem 1) Let $\{X_n\}$ be a submartingale and $\sup \mathbb{E}[X_n^+]$ is finite. Then there exists an $X \in L^1$ such that $X_n \longrightarrow X$ (a.s).

Proof. By Doob's upcrossing inequality, there exists $\Omega' \subseteq \Omega$, with $\mathbb{P}(\Omega') = 1$ such that for all $\omega \in \Omega'$,

$$U^{(n)}_{(a,b)}(\omega) < \infty,\ \forall a, b \in \mathbb{Q}\ (a < b) \implies X := \lim_n X_n \ \text{ exists.}$$

It remains to show that $X \in L^1$:

$$\begin{aligned} \mathbb{E}[|X_n|] = 2\mathbb{E}[X_n^+] - \mathbb{E}[X_n] \leq 2\mathbb{E}[X_n^+] - \mathbb{E}[X_1] &\implies \sup \mathbb{E}[|X_n|] < \infty \\ &\implies \{X_n\} \text{ is bounded in } L^1. \end{aligned}$$

By Fatou's Lemma, $\mathbb{E}[|X|] \leq \liminf \mathbb{E}[|X_n|] < \infty$. $\square$

Theorem 46. (Martingale Convergence Theorem 2) Let $\{X_n\}$ be a (sub)martingale. Then $\{X_n\}$ is UI if and only if there exists an $X \in L^1$ such that $X_n \longrightarrow X$ (a.s and in L^1) and

$$\mathbb{E}[X|\mathcal{F}_n] \geq X_n \text{ a.s} \quad \text{(equality with martingale)}$$

Theorem 47. (Doob's Maximum Inequality) $\{X_n\}$ is a submartingale then for all $N \geq 1$, for all $\epsilon > 0$:

$$\mathbb{P}\left(\max_{1\leq n\leq N} X_n > \epsilon\right) \leq (\epsilon)^{-1}\mathbb{E}\left[X_N; \max_{1\leq n\leq N} X_n > \epsilon\right].$$

In particular, if $X_n \geq 0$ (for all $n \geq 0$), then for all $p \in (1, \infty)$:

$$\|\max_{1\leq n\leq N} X_n\|_p \leq \frac{p}{p-1}\|X_N\|_p.$$

Taking $N \longrightarrow \infty$,

$$\left(\mathbb{E}\left[|\sup_n X_n|^p\right]\right)^{1/p} \leq \frac{p}{p-1}\sup_n \left(\mathbb{E}[|X_n|^p]\right)^{1/p}$$

Theorem 48. (Martingale Convergence Theorem 3) Assume $\{X_n\}$ is either a martingale or a non-negative submartingale and $p \in (1, \infty)$. Then

$\{X_n\}$ is bounded in $L^p \iff X_n \longrightarrow X$ a.s (and in L^p) and $\mathbb{E}[X|\mathcal{F}_n] \geq X_n$ (equality for martingale)

Proposition 30. Consider a filtered space and $X \in L^1$. Set $X_n := \mathbb{E}[X|\mathcal{F}_n]$ for all $n \geq 0$. Then $\{X_n\}$ is a UI martingale and there exists $Y := \mathbb{E}[X|\mathcal{F}_\infty]$ such that

$$X_n \longrightarrow Y \quad \text{(a.s and in } L^1).$$

Proposition 31. (Generalized 0-1 Law) Consider a filtered space with $A \in \mathcal{F}_\infty$. If $A \perp\!\!\!\perp \mathcal{F}_n$ for all $0 \leq n \leq \infty$, then $\mathbb{P}(A) = 0$ or 1.

Proof. Take $X = \mathbb{1}_A$ and let $X_n = \mathbb{E}[X|\mathcal{F}_n]$, which converges almost surely (and in L^1) to $\mathbb{E}[X|\mathcal{F}_\infty] = X$ (because $A \in \mathcal{F}_\infty$). To this end:

$$\mathbb{P}(A) = \int_\Omega (\mathbb{1}_A)^2 \, d\mathbb{P} \overset{\text{(DOM)}}{=} \lim_n \int_\Omega X_n \mathbb{1}_A \, d\mathbb{P} = \lim_n \mathbb{E}[X_n]\mathbb{P}(A) = (\mathbb{P}(A))^2.$$

□

Lemma 12. Let $\{X_n\}$ be a UI martingale and let τ be a stopping time. Then $\{X_{\tau\wedge n}\}$ is a UI martingale.

Theorem 49. (Hunt's Theorem 2) Let $\{X_n\}$ be a (sub)martingale and let τ_1 and τ_2 be stopping times. If $\{X_n\}$ is UI, then:

$$\mathbb{E}[X_{\tau_2}|\mathcal{F}_{\tau_1}] \geq X_{\tau_1} \text{ a.s} \quad \text{(equality if martingale)}.$$

In particular, $\mathbb{E}[X_{\tau_2}] \geq \mathbb{E}[X_{\tau_1}]$ (equality if martingale).

Lemma 13. Let $\{X_n\}$ be a UI submartingale and $X_n = M_n + Y_n$ (that follow from Doob's Decomposition theorem). Then $\{M_n\}$ and $\{Y_n\}$ are UI.

Theorem 50. (Hunt's Theorem 3) Let $\{X_n\}$ be a (sub)martingale and let τ_1 and τ_2 be stopping times. Further suppose that there exists a $K > 0$ such that, for all $n \geq 0$,

$$|X_{n+1} - X_n| \leq K$$

and $\mathbb{E}[\tau_i] < \infty$ (for $i = 1, 2$). Then

$$\mathbb{E}[X_{\tau_2}|\mathcal{F}_{\tau_1}] \geq X_{\tau_1} \text{ a.s} \quad \text{(equality if martingale)}.$$

In particular, $\mathbb{E}[X_{\tau_2}] \geq \mathbb{E}[X_{\tau_1}]$ (equality if martingale).

Corollary 7. (Wald's Inequality) Let $\{X_n\}$ be iid with $\mathbb{E}[X_1] \in \mathbb{R}$. Define $S_0 \equiv 0$ and $S_n = \sum_{j=1}^n X_j$ for all $n \geq 1$ and consider the natural filtration $\mathcal{F}_n = \sigma(S_0, \dots, S_n)$ (for all $n \geq 0$). Further, let τ be a stopping time with finite expectation. Then:

$$\mathbb{E}[S_\tau] = \mathbb{E}[\tau]\mathbb{E}[X_1].$$

10.4 Hitting Time

10.4.1 Random Walk

Definition 62. Suppose $\{X_n\}$ are iid random variables with $\mathbb{P}(X_n = 1) = p$, $\mathbb{P}(X_n = -1) = 1 - p = q$. Let $S_0 \equiv 0$ and $S_n = \sum_{j=1}^n X_j \in \mathbb{Z}$. Then $\{S_n\}$ is called the **(p-q) random walk**.For the remainder of the definitions and examples, we assume a *symmetric random walk* i.e. $p = 0.5$.

For the remainder of the definitions and examples, we assume a *symmetric random walk* i.e. $p = 0.5$.

Definition 63. For all $a \in \mathbb{Z}$, let $\tau_a := \inf\{n \geq 0 \mid S_n = a\}$ is the **hitting time** at a.

Example 20. (Part 1) Let $a, b \in \mathbb{Z}$ such that $a < 0 < b$. Define the following hitting time: $\tau := \tau_a \wedge \tau_b = \inf\{n \geq 0 \mid S_n = a \text{ or } S_n = b\}$. Then $\mathbb{P}(\tau < \infty) = 1$ and so $\mathbb{E}[\tau] < \infty$.

Example 21. (Part 2) Compute $\mathbb{P}(S_\tau = a)$ and $\mathbb{P}(S_\tau = b)$ (such that $\mathbb{P}(S_\tau = a) + \mathbb{P}(S_\tau = b) = 1$).

Proof. In addition to $\{S_n\}$ being a martingale, we know that $\mathbb{E}[\tau]$ is finite and $|S_{n+1} - S_n| = 1$. We can apply (HUNT3) and so $\mathbb{E}[S_\tau] = \mathbb{E}[S_0] = 0$. That is to say:

$$\mathbb{E}[S_\tau] = 0 = a\mathbb{P}(S_\tau = a) + b\mathbb{P}(S_\tau = b).$$

From the other constraint on this problem we know that

$$\mathbb{P}(S_\tau = a) = \frac{b}{b-a}, \quad \mathbb{P}(S_\tau = b) = \frac{-a}{b-a}.$$

□

Example 22. (Part 3) Explicitly, what is $\mathbb{E}[\tau]$?

Proof. We know from the examples at the start of this chapter that $\{S_n^2 - n\}$ is a martingale. By Doob's Stopping Time Theorem, we know that for any fixed n, $\{S_{n\wedge\tau}^2 - (n \wedge \tau)\}$ is a martingale as well. Thus, for all $n \geq 1$, $\mathbb{E}[n \wedge \tau] = \mathbb{E}[S_{n\wedge\tau}^2]$. Using (MON), we get the following:

$$\mathbb{E}[\tau] = \lim_n \mathbb{E}[S_{n\wedge\tau}^2] \overset{\text{(DOM)}}{=} \mathbb{E}[S_\tau^2] = a^2\frac{b}{b-a} + b^2\frac{(-a)}{b-a} = -ab.$$

□

Example 23. (Part 4) For fixed $s \leq 0$, compute $\mathbb{E}[\exp(s\tau)]$.

Proof. For now, replace s with the dummy-variable t. We want to use some form of HUNT's theorem so finding a suitable martingale to deal with is priority. Computing

$$\begin{aligned}
\mathbb{E}[\exp(tS_{n+1})|\mathcal{F}_n] &= \mathbb{E}[\exp(tS_n + tX_{n+1})|\mathcal{F}_n] \\
&= \exp(tS_n)\mathbb{E}[tX_{n+1}] \\
&= \exp(tS_n)\left(0.5e^t + 0.5e^{-t}\right) = (\cosh(t))\exp(tS_n).
\end{aligned}$$

So, after proper correction. we have that $\{\exp(tS_n)(\cosh(t)^{-n})\}$ is a martingale for all $t \in \mathbb{R}$. By Doob's Stopping Time theorem, we have that $\left\{\exp(tS_{n\wedge\tau})\left(\cosh(t)^{-(n\wedge\tau)}\right)\right\}$is also a martingale. Also this martingale is UI because, for all $n \geq 1$, it is bounded:

$$\exp(tS_{n\wedge\tau})\left(\cosh(t)^{-(n\wedge\tau)}\right) \leq \exp[|t|\max\{a,b\}].$$

By (HUNTII), $\mathbb{E}[\exp(tS_\tau)(\cosh(t))^{-\tau}] = \mathbb{E}[\exp(tS_0)(\cosh(t))^0] = 1$ i.e.

$$\begin{aligned}
&\mathbb{E}[\exp(tS_\tau(\cosh(t))^{-\tau}] = e^{ta}\int_{\{S_\tau=a\}}(\cosh(t))^{-\tau}\ d\mathbb{P} + e^{tb}\int_{\{S_\tau=b\}}(\cosh(t))^{-\tau}\ d\mathbb{P} = 1 \\
&\overset{\text{“}t=-t\text{”}}{\Longrightarrow}\ e^{-ta}\int_{\{S_\tau=a\}}(\cosh(t))^{-\tau}\ d\mathbb{P} + e^{-tb}\int_{\{S_\tau=b\}}(\cosh(t))^{-\tau}\ d\mathbb{P} = 1 \\
&\Longrightarrow\ \text{Solve for}\int_{\{S_\tau=a\}}(\cosh(t))^{-\tau}\ d\mathbb{P},\ \int_{\{S_\tau=b\}}(\cosh(t))^{-\tau}\ d\mathbb{P}.
\end{aligned}$$

Using the above, we can compute $\mathbb{E}[(\cosh(t))^{-\tau}]$:

$$\begin{aligned}
\mathbb{E}[\cosh(t)^{-\tau}] &= \int_{\{S_\tau=a\}}(\cosh(t))^{-\tau}\ d\mathbb{P} + \int_{\{S_\tau=b\}}(\cosh(t))^{-\tau}\ d\mathbb{P} \\
&= \frac{e^{t(a+b)}[e^{-tb} - e^{tb} + e^{ta} - e^{-at}]}{e^{2ta} - e^{2tb}} \quad (\bigstar).
\end{aligned}$$

Let $s := -\ln(\cosh(t))$ then we get $\mathbb{E}[\exp(s\tau)] = (\bigstar)$ but all the “t” are replaced with $\operatorname{arccosh}(e^{-s})$ $\square$